

대한민국의 학문 발전에 기여하고자 하는 세경사 이해연사장의 결단으로 본서 “다변량시계열분석”의 PDF를 웹에 올립니다. 원하시는 모든 분이 다운로드를 하실 수 있습니다.

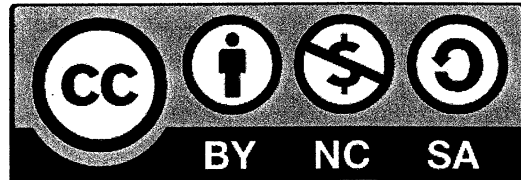
2015. 09. 30.

세경사 대표 이해연 *이해연*
저자 최병선 *崔秉善*

이 저작물에는 크리에이티브 커먼즈 저작자표시-비영리 4.0 국제 라이선스가 적용 되어 있습니다. 이 라이선스의 설명을 보고 싶으시면 다음 웹사이트를 참조하세요.

<http://creativecommons.org/licenses/by-nc/4.0/>

 **creative commons**



* 높은 해상도의 출력과 제본을 원하시면, 현대문화사 (hun1249@hanmail.net) 로 연락하십시오.

SAS를 이용한 시계열분석시리즈 4

다변량시계열분석

世經社

◆ SAS를 이용한 시계열분석시리즈 ◆

- ① PC SAS입문(최병선 지음, 박영사)
초판(1991), 개정판(1992)
- ② 단변량시계열분석 1(최병선 지음, 세경사)
초판(1992)
- ③ 회귀분석(예정)
- ④ 다변량시계열분석(최병선 지음, 세경사)

1. The SAS[®] System is an integrated system of software providing complete control over data access, management, analysis, and presentation. Base SAS software is the foundation of the SAS System. Products within the SAS System include SAS/ACCESS[®], SAS/AF[®], SAS/ASSIST[®], SAS/CPE[®], SAS/DMI[®], SAS/ETS[®], SAS/FSP[®], SAS/GRAPH[®], SAS/IML[®], SAS/IMS-DL/I[®], SAS/OR[®], SAS/QC[®], SAS/REPLAYCICS[®], SAS/SHARE[®], SAS/STAT[®], SAS/CONNECT[™], SAS/DB2[™], SAS/INSIGHT[™], SAS/SQL-DS[™], and SAS/TOOLKIT[™] software. Other SAS Institute products are SYSTEM 2000[®], Data Management Software, with basic SYSTEM 2000, CREATE[™], Multi-User[™], QueX[™], Screen Writer[™], and CICS interface software; Neo Visuals[®], software; JMP[®], JMP IN[®], and JMP Serve[™] software; SAS/RTERM[®] software; and the SAS/C[®] Compiler. MultiVendor Architecture[™] and MVA[™] are trademarks of SAS Institute Inc. *SAS Communications*[®], *SAS Training*[®], *SAS Views*[®], and the SAS ware Ballot[®] are published by SAS Institute Inc. All trademarks above are registered trademarks or trademarks, as indicated by their mark, of SAS Institute Inc.

2. 부록 ㉔의 SAS명령어 목록은 SAS주식회사와 한국SAS판매주식회사의 관대한 허락을 받아 전재한다.

3. 이 저서는 1990년도 연세대학교 학술연구비에 의하여 연구된 것이다.

東然에게

머 리 글

본서의 목적은 SAS패키지의 하나인 SAS/ETS를 사용하여 다변량시계열데이터를 분석하는 방법들을 설명하는 것이다. 그러므로 단순히 SAS/ETS의 여러 프로시저들을 사용하는 방법을 기술하는 지침서의 형태가 아니라, 다변량시계열데이터를 분석하기 위해서 SAS/ETS를 사용하는 방법을 설명하고자 한다. 즉, SAS/ETS의 여러 프로시저들을 상세히 나열하기보다는 다변량시계열분석의 이론과 기법들을 설명하고 그 이론들을 실제로 적용시키기 위해서 SAS/ETS를 어떻게 활용하는가를 보이고자 한다.

다변량시계열분석은 아직도 연구되어야 할 부분이 많은 분야로서 지금도 새로운 이론들이 많이 발표되고 있다. 그러나 이론을 전개하는 데 사용되는 수학의 어려움과 실제 다변량시계열데이터들이 다변량시계열분석기법들의 기본적인 가정들을 잘 만족하지 못하고 있다는 점들이 다변량시계열분석의 발전을 더디게 하고 있다. 그러나 최근에는 컴퓨터 하드웨어와 소프트웨어의 급속한 발전에 힘입어 다변량시계열분석이 활발히 연구되고 있다.

물론 본서에서 모든 현존하는 다변량시계열분석의 기법들을 다루지는 못한다. 본 저자는 본서에서 기존의 다변량시계열분석기법들 중에서 가장 널리 쓰이고 있는 유용한 기법들을 소개하고자 한다. 또한 다변량시계열분석 분야의 발전 가능성을 염두에 두고 비록 SAS/ETS에서 다루지 않는 이론이라 하더라도 앞으로 많이 사용될 가능성이 있는 이론은 SAS/IML을 사용해서 설명하고자 한다.

본 저자는 본서가 출판될 1994년 말을 기준으로 할 때 본서를 학부 수준에서 다루는 것은 무리라고 생각한다. 본서는 다변량시계열데이터를 분석해야 하는 통계학, 경제학, 경영학, 공학 등 여러 학문 분야의 대학원생들과 교수들 그리고 실제 현장에서 나오는 다변량시계열데이터들을 분석해야 하는 연구원들을 위해서 쓴 것이다. 따라서 본서를 구입하고자 하는 예비독자들은 우선 제 1장을 읽고 정말 본인에게 필요하다고 판단되는 경우에 한해서 구입할 것을 권한다.

책을 빠른 속도로 쓰는 본 저자로서도 본서를 쓰는 데는 오랜 시간이 걸렸다. 원고를 작성하는 데 약 2년이 소요되었고, 그로부터 책을 출판하는 데 약 1년 반이 소요되었다. 본서는 지난 4년간 연세대학교 응용통계학과 대학원에서 강의한 강의노트를 바탕으로 한 것이다. 우선 본 저자의 강의를 수강한 학생들에게 감사한다. 때로는 강의 내용이 수강하는 학생들의 수준을 넘어서 이해하기 힘들어 하기도 했고, 또 때로는 강의 내용을 전혀 이해할 수 없어서 '나는 왜 이럴까' 하고 자기비하를 하기도 한 학생들에게 미안하기도 하고 고맙

4 머 리 글

기도 하다. 본서를 만드는 데는 세경사 이해연 씨의 정성이 큰 도움이 되었다. 이해연 씨는 그냥 지내기도 어렵게 무더웠던 1994년의 여름을 본서를 만드는 데 할애하였다. 이해연 씨에게 원고만 넘겨주고 본 저자는 미국에서 여름을 보냈기 때문에 더욱 미안함을 금할 수 없다. 제 5.6절의 공누적에 관한 원고를 읽고 좋은 조언을 해 주신 연세대학교 경제학과의 유병삼 교수님께 감사드린다. 바쁜 사회생활에도 불구하고 교정을 도와 준 노현주(포스코 경영연구소), 양현조(한국신용평가주식회사), 정은이(동서경제연구소), 정영옥(한국신용정보주식회사)에게 감사한다. 또한 교정에 도움을 준 강윤정, 신석인, 이현주, 오연정, 유승연, 이승원에게 감사하며, 주부역할과 사회활동으로 시간적인 여유가 없음에도 불구하고 제 6장과 제 7장의 원고를 입력시켜 준 안나영에게 감사한다. 본서를 만드는 데 이해연 씨와의 연락 등 전반적인 행정처리와 교정에 도움을 준 김명정의 수고에 감사한다. 끝으로 본서의 출판을 맡아주신 김재은 사장님께 감사를 드립니다.

1995. 1

동연의 이름을 지어 주신
도올 선생님께 감사드리며.....

최 병 선

차 례

제 1 장 서 론	11
1.1 다변량시계열분석이란 무엇인가? / 13	
1.2 SAS/ETS / 14	
1.3 본서의 목적 / 15	
1.4 본서의 구성 / 15	
1.5 참고문헌 / 16	
제 2 장 시계열데이터	17
2.1 단변량시계열데이터 M1 : BJ_A / 19	
2.2 다변량시계열데이터 M2 : BJ_J / 21	
2.3 다변량시계열데이터 M3 : GRUNFELD / 25	
2.4 다변량시계열데이터 M4 : HOUSECON / 29	
2.5 다변량시계열데이터 M5 : HOUSING / 33	
2.6 단변량시계열데이터 M6 : LEADPROD / 38	
2.7 단변량시계열데이터 M7 : MA1A / 40	
2.8 다변량시계열데이터 M8 : MINKMUSK / 41	
2.9 다변량시계열데이터 M9 : RIVER / 44	
2.10 다른 시계열데이터들 / 47	
제 3 장 SAS/IML의 사용법.....	49
3.1 SAS/IML 맛보기 / 51	
3.2 SAS/IML의 기초 / 56	
3.2.1 SAS/IML이란 무엇인가? / 56	
3.2.2 SAS/IML의 기초적 사용법 / 57	
3.2.3 SAS/IML의 बै치처리 / 59	
3.2.4 RESET스테이트먼트 / 60	
3.3 행렬의 연산 / 70	
3.3.1 할당스테이트먼트 / 70	
3.3.2 행렬기준연산자들 / 71	

6 차 례

- 3.3.3 원소기준연산자들 / 77
- 3.3.4 부행렬들 / 84
- 3.3.5 연산자들의 우선순위 / 90
- 3.4 SAS/IML 함수 / 91
 - 3.4.1 스칼라함수 / 92
 - 3.4.2 행렬함수 / 96
 - 3.4.3 서브루틴 / 117
 - 3.4.4 SAS/IML 모듈 / 123
- 3.5 제어스테이트먼트 / 130
 - 3.5.1 조건처리와 스테이트먼트들의 그룹화 / 130
 - 3.5.2 반복처리 / 132
 - 3.5.3 건너뛰기 / 136
 - 3.5.4 중단과 재실행 / 138
- 3.6 일반목적명령어들 / 140
 - 3.6.1 현재상태의 파악 / 140
 - 3.6.2 메모리의 효과적 사용 / 141
 - 3.6.3 SAS/IML 저장라이브러리 / 143
 - 3.6.4 출력의 조절 / 146
- 3.7 SAS/IML에서의 데이터처리 / 149
 - 3.7.1 현재SAS데이터세트 / 149
 - 3.7.2 SAS데이터세트의 열기 / 149
 - 3.7.3 SAS데이터세트의 닫기 / 153
 - 3.7.4 접근명령어들의 개관 / 154
 - 3.7.5 범위피연산자들 / 157
 - 3.7.6 VAR구문 / 161
 - 3.7.7 WHERE구문 / 163
 - 3.7.8 행렬을 사용한 데이터처리 / 166
 - 3.7.9 접근명령어들 / 174
 - 3.7.10 범위피연산자의 디폴트 / 186
 - 3.7.11 SAS/IML과 SAS의 DATA스텝 / 186
- 3.8 외부파일의 입출력 / 189
 - 3.8.1 파일참조이름 / 189
 - 3.8.2 텍스트파일과 이진파일 / 190
 - 3.8.3 외부파일의 입력 / 191
 - 3.8.4 외부파일로 출력 / 193
 - 3.8.5 기 타 / 195

제 4 장 전이함수모형분석 197

- 4.1 전이함수모형 / 199
 - 4.1.1 일반적 전이함수모형 / 199
 - 4.1.2 단일입력단일출력전이함수모형 / 200
 - 4.1.3 충격반응가중함수 / 201
 - 4.1.4 예 제 들 / 205
 - 4.1.5 교차상관함수 / 212
 - 4.1.6 표본교차상관함수 / 213
- 4.2 충격반응가중함수의 추정 / 219
 - 4.2.1 최소제곱추정법 / 219
 - 4.2.2 2단계최소제곱추정법 / 226
 - 4.2.3 사전백색화추정법 / 232
 - 4.2.4 능형회귀추정법 / 242
- 4.3 전이함수모형의 식별 / 244
 - 4.3.1 Box-Jenkins방법 / 245
 - 4.3.2 모통이방법 / 251
 - 4.3.3 잡음시계열데이터의 식별 / 259
 - 4.3.4 전이함수모형식별의 유의점들 / 273
- 4.4 전이함수모형의 추정 / 273
- 4.5 전이함수모형의 검진 / 281
- 4.6 전이함수모형의 예측 / 287
- 4.7 예 제 들 / 292
 - 4.7.1 다변량시계열데이터 BJ_{LJ} / 292
 - 4.7.2 시뮬레이션데이터 / 306
 - 4.7.3 다변량시계열데이터 RIVER / 318
 - 4.7.4 다변량시계열데이터 HOUSECON / 333

제 5 장 잡음항이 ARMA과정인 모형..... 347

- 5.1 잡음항이 정상성을 갖는 회귀모형 / 349
 - 5.1.1 잡음항이 ARMA과정인 회귀모형 / 349
 - 5.1.2 예 제 들 / 350
 - 5.1.3 PROC ARIMA의 사용법 / 401
 - 5.1.4 PROC AUTOREG의 사용법 / 413
- 5.2 간접시계열모형 / 419

- 5.2.1 동태모형과 잡음항모형 / 421
- 5.2.2 Box-Tiao의 간섭시계열모형 / 422
- 5.2.3 간섭시계열모형의 추정 / 432
- 5.2.4 PROC ARIMA를 사용한 간섭시계열모형화 / 434
- 5.2.5 예 제 들 / 436
- 5.2.6 참고사항 / 470
- 5.3 공 누 적 / 471
 - 5.3.1 과다차분 / 471
 - 5.3.2 공누적의 개념 / 494
 - 5.3.3 공누적의 정의 / 496
 - 5.3.4 공누적시스템의 추정 / 498

제 6 장 다변량시계열모형 509

- 6.1 자기공분산행렬과 자기상관행렬 / 511
- 6.2 다변량시계열의 일반모형들 / 516
 - 6.2.1 다변량백색잡음과정 / 516
 - 6.2.2 VMA(∞)모형 / 519
 - 6.2.3 VAR(∞)모형 / 520
- 6.3 VAR모형 / 521
 - 6.3.1 작인적 VAR모형 / 521
 - 6.3.2 예 제 들 / 532
 - 6.3.3 VAR과정의 자기공분산행렬 / 544
 - 6.3.4 Whittle알고리즘 / 556
 - 6.3.5 블럭토우폴리츠행렬*/ 572
- 6.4 VMA모형 / 576
 - 6.4.1 가역적 VMA모형 / 576
 - 6.4.2 가역성 / 588
 - 6.4.3 예 제 들 / 590
 - 6.4.4 VMA계수행렬 / 599
- 6.5 VARMA모형 / 605
 - 6.5.1 VARMA모형의 평균벡터 / 605
 - 6.5.2 VARMA모형의 작인성 / 606
 - 6.5.3 VARMA모형의 가역성 / 610
 - 6.5.4 예 제 들 / 613
 - 6.5.5 자기공분산행렬함수 / 625
 - 6.5.6 확장Whittle알고리즘 / 644

6.5.7 VARMA모형의 비유일성 / 658

6.5.8 VMA계수행렬 / 664

제 7 장 VAR모형의 추정과 식별..... 673

7.1 자기공분산행렬과 자기상관행렬의 추정 / 675

7.2 표본자기공분산행렬과 표본자기상관행렬의 성질 / 687

7.3 VAR계수행렬의 추정 / 690

7.3.1 Yule-Walker추정법 / 690

7.3.2 최소제곱추정법 / 700

7.3.3 최우추정법 / 729

7.4 VAR계수추정행렬들의 성질 / 731

7.4.1 최소제곱추정량의 성질 / 732

7.4.2 Yule-Walker추정량의 성질 / 734

7.4.3 최우추정량의 성질 / 735

7.5 VAR모형의 식별 / 737

7.5.1 공분산추정행렬의 편의 / 737

7.5.2 부분자기상관계수의 확장 / 738

7.5.3 우도비검정방법 / 745

7.5.4 벌칙함수식별방법 / 753

7.5.5 VAR모형 식별방법들의 비교 / 761

7.6 VAR모형의 검진 / 762

7.6.1 다변량백색잡음과정의 표본자기상관행렬함수 / 762

7.6.2 잔차벡터들의 표본자기상관행렬함수 / 764

7.6.3 퍼트맨토우검정통계량 / 766

제 8 장 상태공간모형 775

8.1 상태공간모형 / 777

8.2 상태공간모형화의 기초 / 778

8.2.1 PROC STATESPACE의 기초적 사용법 / 778

8.2.2 RESTRICT스테이트먼트 / 791

8.2.3 상태공간모형을 사용한 예측 / 805

8.3 상태공간모형화 / 815

8.3.1 상태공간모형과 VARMA모형 / 816

8.3.2 VAR모형단계 / 821

8.3.3 정준상관분석단계 / 823

10 차례

8.3.4	최종추정단계 /	829
8.4	PROC STATESPACE의 자세한 사용법 /	831
8.4.1	PROC STATESPACE의 사용법 /	831
8.4.2	단변량시계열데이터 BJ_A /	839
8.4.3	다변량시계열데이터 BJ_J /	851
8.5	수리적 배경 /	876
8.5.1	VARMA모형의 상태공간모형화 /	876
8.5.2	VAR모형단계의 이론적 배경 /	879
8.5.3	정준상관분석단계의 이론적 배경 /	882
8.5.4	Akaike의 최종추정단계 /	889
8.5.5	Kalman필터링 /	890
8.6	상태공간모형화방법의 예제들 /	897
8.6.1	시뮬레이션데이터 MA1A /	897
8.6.2	단변량시계열 LEADPROD /	906
8.6.3	다변량시계열데이터 MINKMUSK /	909
8.6.4	다변량시계열데이터 HOUSECON /	931
부록 A	점화식 연구	951
부록 B	SAS/ETS® Software Reference Card	1031
부록 C	SAS/IML™ Software Reference Card.....	1035
색인	1039

서 론

- 1.1 다변량시계열분석이란 무엇인가?
- 1.2 SAS/ETS
- 1.3 본서의 목적
- 1.4 본서의 구성
- 1.5 참고문헌

제 1 장 서 론

본서의 제목이 나타내듯이 본서의 목적은 SAS시스템을 사용해서 다변량시계열데이터를 분석하는 방법과 그 배경이론을 설명하고자 하는 것이다. 이 장에서는 다변량시계열데이터의 의미, 다변량시계열분석에서 SAS/ETS의 역할, 본서를 효과적으로 읽는 방법, 또한 본서를 읽는 데 도움이 되는 참고문헌 등을 소개하고자 한다.

1.1 다변량시계열분석이란 무엇인가?

본서를 읽는 독자의 수준이면 시계열분석이 무엇인가를 알고 있으리라 생각한다. 만일, 시계열분석이 무엇인가를 전혀 모르는 독자가 있다면, 본서를 읽기 전에 'SAS를 이용한 시계열분석시리즈'의 제2권 『단변량시계열분석1』(최병선, 1992, 세경사, 이하 [SAS4TSA2]라 부를 것임)을 먼저 읽을 것을 권한다.

[SAS4TSA2](p. 14)에서 시계열데이터를 다음과 같이 두 종류로 나누었다. 연세대학교 상경대학 511호실의 월별 평균온도라든가 포항제철주식의 일별 증가 등과 같이 각 시점에서 분석대상이 되는 변수가 하나인 시계열데이터를 단변량시계열데이터(univariate time series)라고 한다. 반면에 연세대학교 상경대학 511호실의 월별 평균온도와 평균습도를 동시에 분석한다든가 포항제철주식의 일별 증가와 일별 거래량을 동시에 분석하는 것처럼 각 시점에서 두 개 이상의 변수들을 취급하는 경우가 있다. 이렇게 분석대상이 되는 변수들이 두 개 이상인 시계열데이터를 다변량시계열데이터(multivariate time series)라고 한다.

다변량시계열데이터를 분석하는 목적 중의 하나는 좀더 신뢰할 수 있는 예측시스템(forecasting system)을 얻기 위한 것이다. 포항제철주식의 미래값들을 예측하기 위해서 이 주식의 현재가격과 과거가격들만을 사용하는 단변량시계열분석을 할 수도 있다. 주식시세를

분석하는 사람들은 주식거래량을 주식가격의 선행지수의 하나로 보고 있다. 즉, 주식거래량은 주식가격을 예측하는 시스템이 가질 수 있는 추가적인 정보인 것이다. 이 추가적인 정보를 잘 활용한다면 좀더 정확히 주식가격을 예측할 수 있을 것이다. 이렇게 주식가격과 주식거래량을 하나의 확률시스템에 넣어서 분석하는 것을 다변량시계열분석이라 한다. 즉, 다변량시계열분석의 목적 중의 하나는 미래값을 잘 예측할 수 있도록 연관된 변수들이 구성하고 있는 시스템을 통계적으로 구축하는 것이다. 다변량시계열데이터를 분석하는 다른 목적은 여러 변수들 사이의 동적 관계(dynamic interrelationships)를 규명하는 것이다. 예를 들어, 한국의 노동수요시장을 분석할 때 사용되는 변수들은 고용량, 명목임금, 스트라이크 횟수, 총자본, GNP성장률, 월평균노동시간, 기대물가상승률, 그리고 노동의 한계생산력 등이다. 이 변수들 중에서 한 변수의 값의 변화가 다른 변수들에 미치는 영향을 알기 위해서는 변수들 사이의 동적 관계를 알아야 한다. 다변량시계열분석의 목적 중의 하나는 이러한 동적 관계를 모형화하는 것이다.

다변량시계열분석은 아직도 연구되어야 할 부분이 많은 분야로서 오늘날도 새로운 이론들이 발표되고 있다. 그러나 이론을 전개하는 데 사용되는 수학기법들이 어려운 점과 실제 다변량시계열데이터들이 기본적인 가정들을 잘 만족하지 못하고 있는 점 등이 이 분야의 발전을 더디게 하고 있다. 그러나 최근에는 컴퓨터 하드웨어와 소프트웨어의 급속한 발전에 힘입어 다변량시계열분석이 활발히 연구되고 있다.

1.2 SAS/ETS

[SAS4TSA2](p. 14)나 'SAS를 이용한 시계열분석시리즈'의 제1권인 『PC SAS 입문(개정판)』(최병선, 1992, 박영사, 이하 [SAS4TSA1]이라 부를 것임)의 제 2 장에서도 언급했듯이 SAS는 아주 방대한 정보처리용 소프트웨어다(최근 SAS사는 SAS가 Strategic Application Software를 뜻한다고 발표하였다). 이 SAS시스템에서 계량경제학(Econometrics)과 시계열분석(Time Series analysis)을 위한 소프트웨어가 SAS/ETS이다. 본서에서는 SAS/ETS의 DOS 버전 6.04를 바탕으로 설명하고자 한다.

SAS/ETS는 다음과 같은 프로시저들을 포함하고 있다.

ARIMA, AUTOREG, CITIBASE, COMPUTAB, EXPAND,
FORECAST, MODEL, MORTGAGE, PDLREG,
SIMLIN, SPECTRA, STATESPACE, SYSLIN, X11

이 프로시저들 중에서 본서에서 주로 다루게 될 것은 ARIMA, AUTOREG, STATESPACE이다.

1.3 본서의 목적

본서의 목적은 SAS/ETS를 사용해서 다변량시계열데이터를 분석하는 방법들을 설명하고자 하는 것이다. 그러나 단순히 SAS/ETS의 프로시저들을 사용하는 방법들을 기술하는 지침서(manual) 형태가 아니라, 다변량시계열데이터를 분석하기 위해서 SAS/ETS를 사용하는 방법을 설명하고자 한다. 따라서 SAS/ETS의 여러 프로시저들을 상세히 나열하기 보다는 다변량시계열분석의 이론과 기법을 설명하고 이 이론들을 실제적으로 적용시키기 위해서 SAS/ETS를 어떻게 활용하는가를 보이고자 한다.

본서를 읽기 위해서는 [SAS4TSA2]의 내용을 이해하고 있어야 한다. 만일, 본서를 읽고 있는 독자가 단변량시계열분석에 대해서 잘 알지 못한다면 본서를 덮어 두고 [SAS4TSA2]를 먼저 읽을 것을 권한다. 또한 본서를 읽기 위해서는 SAS의 기초적인 사용법도 알고 있어야 한다. 만일, 본서를 읽고 있는 독자가 기초적인 SAS사용법을 잘 알지 못한다면 본서를 덮어 놓고 [SAS4TSA1]을 먼저 읽을 것을 권한다.

1.4 본서의 구성

본서는 8개의 장으로 구성되어 있다. 제 2장에서는 본서에서 분석할 시계열데이터들을 설명하고 이 시계열데이터들의 시계열산포도들을 그린다. SAS/ETS의 프로시저들만으로는 다변량시계열분석의 특수한 이론이나 기법을 잘 설명할 수 없는 경우가 있다. 이러한 경우에는 SAS/IML을 사용해서 프로그램을 만들어야 한다. 제 3장에서는 이러한 프로그램을 만드는 데 필요한 SAS/IML의 기초적인 사용법을 설명한다. 제 4장에서는 전이함수 모형(transfer function model)을 사용해서 다변량시계열데이터를 분석하는 방법을 설명한다. 제 5장에서는 잡음항이 ARMA모형을 갖는 회귀모형을 사용해서 다변량시계열데이터를 분석하는 방법을 설명한다. 특히, 제 5장에서는 간섭시계열모형(intervention time series model)과 공능적(cointegration)에 대해서 설명한다. 제 6장에서는 벡터자기회귀이동평균모형(vector autoregressive moving-average model, VARMA model)을 설명하고, 제 7장에서는 벡터자기회귀모형(vector autoregressive model, VAR model)의 식별, 추정, 그리고 검진에 대해서 설명한다. 제 8장에서는 상태공간모형(state space model)을 사용해서 다변량시계열데이터를 분석하는 방법을 설명한다.

1.5 참고문헌

본서에서 자주 인용하게 될 참고문헌들은 다음과 같다. 각 참고문헌의 끝에 []로 표시된 것은 해당 참고문헌의 약자이다.

- 최병선(1992), 『PC SAS입문(개정판) - SAS를 이용한 시계열분석시리즈 1』, 박영사, 서울. [SAS4TSA1]
- 최병선(1992), 『단변량시계열분석 1 - SAS를 이용한 시계열분석시리즈 2』, 세경사, 서울. [SAS4TSA2]
- Brocklebank, J. C. & D. A. Dickey(1986), *SAS System for Forecasting Time Series*, SAS Institute Inc., Cary. [BD86]
- Brockwell, P. J. & R. A. Davis(1990), *Time Series: Theory and Methods*, Second Edition, Springer-Verlag, New York. [BD90]
- Choi, B. S. (1992), *ARMA Model Identification*, Springer-Verlag, New York. [Choi92]
- SAS Institute Inc. (1988), *SAS/ETS User's Guide*, Version 6, First Edition, SAS Institute Inc., Cary. [ETS6.1]
- SAS Institute Inc. (1988), *SAS/IML User's Guide*, Release 6.03 Edition, SAS Institute Inc., Cary. [IML6.03]
- SAS Institute Inc. (1990), *Advanced Forecasting Techniques Using SAS/ETS Software Course Notes*, SAS Institute Inc., Cary. [CNote2]
- SAS Institute Inc. (1991), *Forecasting Techniques Using SAS/ETS Software Course Notes*, SAS Institute Inc., Cary. [CNote1]
- SAS Institute Inc. (1991), *SAS/ETS Software: Applications Guide 1*, Version 6, First Edition, SAS Institute Inc., Cary. [ApplGuide]

본서를 쓰는 데는 이 참고문헌들 이외에도 다음과 같은 최근에 나온 다변량시계열분석에 관한 문헌들을 참조하였다.

- Lütkepohl, H. (1991), *Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.
- Reinsel, G. C. (1993), *Elements of Multivariate Time Series Analysis*, Springer-Verlag, Berlin.

이 절에서 언급된 문헌들 이외에도 각 장의 끝에 좀더 자세한 참고문헌들의 목록을 수록할 것이다.

시계열데이터

- 2.1 단변량시계열데이터 M1 : BJ_A
- 2.2 다변량시계열데이터 M2 : BJ_J
- 2.3 다변량시계열데이터 M3 : GRUNFELD
- 2.4 다변량시계열데이터 M4 : HOUSECON
- 2.5 다변량시계열데이터 M5 : HOUSING
- 2.6 단변량시계열데이터 M6 : LEADPROD
- 2.7 단변량시계열데이터 M7 : MA1A
- 2.8 다변량시계열데이터 M8 : MINKMUSK
- 2.9 다변량시계열데이터 M9 : RIVER
- 2.10 다른 시계열데이터들

제 2 장 시계열데이터

이 장에서는 다변량시계열분석기법들을 설명하기 위해서 본서에서 주로 사용하게 될 시계열데이터들을 소개하고자 한다.

2.1 단변량시계열데이터 M1 : BJ_A

이 단변량시계열데이터는 Box & Jenkins(1976, p. 525)의 Series A로서 일명 Chemical Process Readings Data라고 한다. 이 단변량시계열데이터의 변수 X는 어느 화학공정의 매 두 시간마다의 농도(chemical process concentration readings : every two hours)를 측정된 것이고, 변수 ID는 관찰점들의 순서를 나타낸다. 이 단변량시계열데이터는 197개의 관찰점들로 구성되었다. 이 시계열데이터를 영구SAS데이터세트로 만들기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 BJ_A.PRG에 수록되어 있다).

```
/* */
/*          UNIVARIATE TIME SERIES DATA BJ_ADAT          */
/*          MAKING SASDATASET BJA.SSD                      */
/* */
LIBNAME TSA4DATA 'A:\TSA4DATA';
DATA TSA4DATA.BJA;
    INPUT ID X:3.1 @@ ;
    CARDS;
1 170    2 166    3 163    4 161    5 171    6 169    7 168
8 174    9 171   10 170   11 167   12 174   13 172   14 174
```

20 제 2 장 시계열데이터

```

15 174 16 170 17 173 18 172 19 174 20 168 21 171
22 174 23 174 24 175 25 174 26 176 27 174 28 173
29 170 30 178 31 175 32 181 33 175 34 174 35 174
36 171 37 176 38 177 39 174 40 178 41 176 42 175
43 165 44 178 45 173 46 173 47 171 48 174 49 169
50 173 51 176 52 169 53 167 54 168 55 168 56 172
57 168 58 176 59 172 60 166 61 171 62 169 63 166
64 180 65 172 66 173 67 170 68 169 69 173 70 168
71 173 72 174 73 177 74 168 75 169 76 170 77 169
78 170 79 166 80 167 81 168 82 167 83 164 84 165
85 164 86 166 87 165 88 167 89 164 90 164 91 162
92 164 93 163 94 164 95 170 96 169 97 171 98 171
99 167 100 169 101 165 102 172 103 164 104 170 105 170
106 167 107 162 108 166 109 169 110 165 111 166 112 166
113 170 114 171 115 171 116 167 117 168 118 163 119 166
120 168 121 169 122 171 123 168 124 170 125 172 126 173
127 172 128 173 129 172 130 172 131 175 132 169 133 169
134 169 135 170 136 165 137 167 138 168 139 167 140 167
141 166 142 165 143 170 144 167 145 167 146 169 147 174
148 171 149 170 150 168 151 172 152 172 153 174 154 172
155 169 156 168 157 170 158 174 159 172 160 172 161 171
162 171 163 171 164 174 165 172 166 169 167 169 168 170
169 167 170 169 171 173 172 178 173 178 174 176 175 175
176 170 177 169 178 171 179 172 180 174 181 175 182 179
183 170 184 170 185 170 186 172 187 173 188 174 189 174
190 170 191 180 192 182 193 176 194 178 195 177 196 172
197 174

```

:

```

RUN;
TITLE1 'DATASET BJ_A.DAT';
SYMBOL1 C=BLACK V=NONE L=1 I=JOIN;
PROC GPLOT;
PLOT X*ID;
RUN;

```

이 SAS프로그램을 수행하면 영구SAS데이터세트 BJA가 디스켓 SAS4TSA4의 서브디렉토리 TSA4DATA의 파일 BJA.SSD에 저장된다. 이 SAS프로그램의 PROC GPLOT를 수행하면 그림 2.1의 시계열산포도를 얻는다. 이 시계열산포도의 실선이 변수 X를 나타낸다.

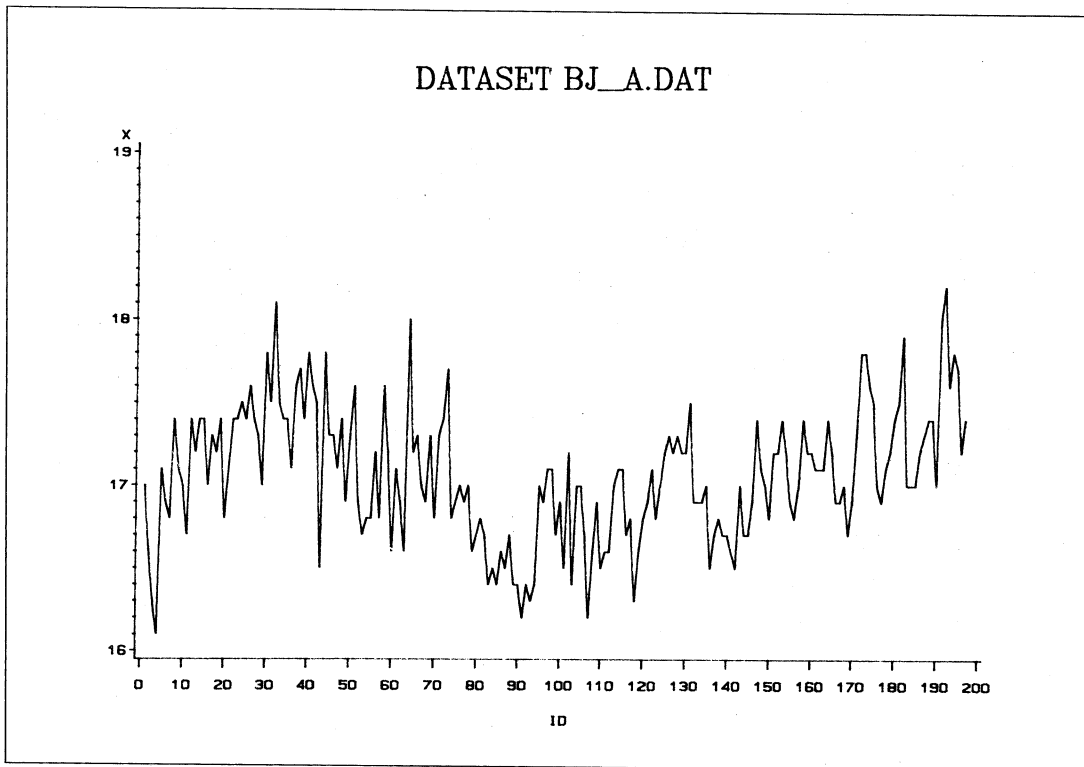


그림 2.1 단변량시계열데이터 BJ_A의 시계열산포도

2.2 다변량시계열데이터 M2 : BJ_J

이 2변량시계열데이터는 Box & Jenkins(1976, pp. 532~533)의 Series J로서 일명 Gas Furnace Data라고 한다. 이 2변량시계열데이터의 변수 X는 입력개스비율(input gas rate)을 나타내고, 변수 Y는 출력개스내의 이산화탄소의 퍼센트(CO₂ percent in output gas)를 나타낸다. 이 다변량시계열데이터는 296개의 관찰점들로 구성되어 있다. 이 시계열데이터를 영구SAS데이터세트로 만들기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 BJ_J.PRG에 수록되어 있다).

```

/*                               */
/*           MULTIVARIATE TIME SERIES DATA BJ_J.DAT           */
/*           MAKING SASDATASET BJJ.SSD                         */
/*                               */
LIBNAME TSA4DATA 'A:\TSA4DATA';
DATA TSA4DATA.BJJ;

```

INPUT T X Y @@ :

LABEL X = 'INPUT GAS RATE' Y = 'OUTPUT CO2':

CARDS:

1	-0.109	53.8	2	0.000	53.6	3	0.178	53.5	4	0.339	53.5
5	0.373	53.4	6	0.441	53.1	7	0.461	52.7	8	0.348	52.4
9	0.127	52.2	10	-0.180	52.0	11	-0.588	52.0	12	-1.055	52.4
13	-1.421	53.0	14	-1.520	54.0	15	-1.302	54.9	16	-0.814	56.0
17	-0.475	56.8	18	-0.193	56.8	19	0.088	56.4	20	0.435	55.7
21	0.771	55.0	22	0.866	54.3	23	0.875	53.2	24	0.891	52.3
25	0.987	51.6	26	1.263	51.2	27	1.775	50.8	28	1.976	50.5
29	1.934	50.0	30	1.866	49.2	31	1.832	48.4	32	1.767	47.9
33	1.608	47.6	34	1.265	47.5	35	0.790	47.5	36	0.360	47.6
37	0.115	48.1	38	0.088	49.0	39	0.331	50.0	40	0.645	51.1
41	0.960	51.8	42	1.409	51.9	43	2.670	51.7	44	2.834	51.2
45	2.812	50.0	46	2.483	48.3	47	1.929	47.0	48	1.485	45.8
49	1.214	45.6	50	1.239	46.0	51	1.608	46.9	52	1.905	47.8
53	2.023	48.2	54	1.815	48.3	55	0.535	47.9	56	0.122	47.2
57	0.009	47.2	58	0.164	48.1	59	0.671	49.4	60	1.019	50.6
61	1.146	51.5	62	1.155	51.6	63	1.112	51.2	64	1.121	50.5
65	1.223	50.1	66	1.257	49.8	67	1.157	49.6	68	0.913	49.4
69	0.620	49.3	70	0.255	49.2	71	-0.280	49.3	72	-1.080	49.7
73	-1.551	50.3	74	-1.799	51.3	75	-1.825	52.8	76	-1.456	54.4
77	-0.944	56.0	78	-0.570	56.9	79	-0.431	57.5	80	-0.577	57.3
81	-0.960	56.6	82	-1.616	56.0	83	-1.875	55.4	84	-1.891	55.4
85	-1.746	56.4	86	-1.474	57.2	87	-1.201	58.0	88	-0.927	58.4
89	-0.524	58.4	90	0.040	58.1	91	0.788	57.7	92	0.943	57.0
93	0.930	56.0	94	1.006	54.7	95	1.137	53.2	96	1.198	52.1
97	1.054	51.6	98	0.595	51.0	99	-0.080	50.5	100	-0.314	50.4
101	-0.288	51.0	102	-0.153	51.8	103	-0.109	52.4	104	-0.187	53.0
105	-0.255	53.4	106	-0.229	53.6	107	-0.007	53.7	108	0.254	53.8
109	0.330	53.8	110	0.102	53.8	111	-0.423	53.3	112	-1.139	53.0
113	-2.275	52.9	114	-2.594	53.4	115	-2.716	54.6	116	-2.510	56.4
117	-1.790	58.0	118	-1.346	59.4	119	-1.081	60.2	120	-0.910	60.0
121	-0.876	59.4	122	-0.885	58.4	123	-0.800	57.6	124	-0.544	56.9
125	-0.416	56.4	126	-0.271	56.0	127	0.000	55.7	128	0.403	55.3
129	0.841	55.0	130	1.285	54.4	131	1.607	53.7	132	1.746	52.8
133	1.683	51.6	134	1.485	50.6	135	0.993	49.4	136	0.648	48.8
137	0.577	48.5	138	0.577	48.7	139	0.632	49.2	140	0.747	49.8
141	0.900	50.4	142	0.993	50.7	143	0.968	50.9	144	0.790	50.7

145	0.399	50.5	146	-0.161	50.4	147	-0.553	50.2	148	-0.603	50.4
149	-0.424	51.2	150	-0.194	52.3	151	-0.049	53.2	152	0.060	53.9
153	0.161	54.1	154	0.301	54.0	155	0.517	53.6	156	0.566	53.2
157	0.560	53.0	158	0.573	52.8	159	0.592	52.3	160	0.671	51.9
161	0.933	51.6	162	1.337	51.6	163	1.460	51.4	164	1.353	51.2
165	0.772	50.7	166	0.218	50.0	167	-0.237	49.4	168	-0.714	49.3
169	-1.099	49.7	170	-1.269	50.6	171	-1.175	51.8	172	-0.676	53.0
173	0.033	54.0	174	0.556	55.3	175	0.643	55.9	176	0.484	55.9
177	0.109	54.6	178	-0.310	53.5	179	-0.697	52.4	180	-1.047	52.1
181	-1.218	52.3	182	-1.183	53.0	183	-0.873	53.8	184	-0.336	54.6
185	0.063	55.4	186	0.084	55.9	187	0.000	55.9	188	0.001	55.2
189	0.209	54.4	190	0.556	53.7	191	0.782	53.6	192	0.858	53.6
193	0.918	53.2	194	0.862	52.5	195	0.416	52.0	196	-0.336	51.4
197	-0.959	51.0	198	-1.813	50.9	199	-2.378	52.4	200	-2.499	53.5
201	-2.473	55.6	202	-2.330	58.0	203	-2.053	59.5	204	-1.739	60.0
205	-1.261	60.4	206	-0.569	60.5	207	-0.137	60.2	208	-0.024	59.7
209	-0.050	59.0	210	-0.135	57.6	211	-0.276	56.4	212	-0.534	55.2
213	-0.871	54.5	214	-1.243	54.1	215	-1.439	54.1	216	-1.422	54.4
217	-1.175	55.5	218	-0.813	56.2	219	-0.634	57.0	220	-0.582	57.3
221	-0.625	57.4	222	-0.713	57.0	223	-0.848	56.4	224	-1.039	55.9
225	-1.346	55.5	226	-1.628	55.3	227	-1.619	55.2	228	-1.149	55.4
229	-0.488	56.0	230	-0.160	56.5	231	-0.007	57.1	232	-0.092	57.3
233	-0.620	56.8	234	-1.086	55.6	235	-1.525	55.0	236	-1.858	54.1
237	-2.029	54.3	238	-2.024	55.3	239	-1.961	56.4	240	-1.952	57.2
241	-1.794	57.8	242	-1.302	58.3	243	-1.030	58.6	244	-0.918	58.8
245	-0.798	58.8	246	-0.867	58.6	247	-1.047	58.0	248	-1.123	57.4
249	-0.876	57.0	250	-0.395	56.4	251	0.185	56.3	252	0.662	56.4
253	0.709	56.4	254	0.605	56.0	255	0.501	55.2	256	0.603	54.0
257	0.943	53.0	258	1.223	52.0	259	1.249	51.6	260	0.824	51.6
261	0.102	51.1	262	0.025	50.4	263	0.382	50.0	264	0.922	50.0
265	1.032	52.0	266	0.866	54.0	267	0.527	55.1	268	0.093	54.5
269	-0.458	52.8	270	-0.748	51.4	271	-0.947	50.8	272	-1.029	51.2
273	-0.928	52.0	274	-0.645	52.8	275	-0.424	53.8	276	-0.276	54.5
277	-0.158	54.9	278	-0.033	54.9	279	0.102	54.8	280	0.251	54.4
281	0.280	53.7	282	0.000	53.3	283	-0.493	52.8	284	-0.759	52.6
285	-0.824	52.6	286	-0.740	53.0	287	-0.528	54.3	288	-0.204	56.0
289	0.034	57.0	290	0.204	58.0	291	0.253	58.6	292	0.195	58.5
293	0.131	58.3	294	0.017	57.8	295	-0.182	57.3	296	-0.262	57.0

```

RUN;
TITLE1 'DATASET BJ_J.DAT';
SYMBOL1 C=BLACK V=NONE L=1 I=JOIN;
SYMBOL2 C=BLACK V=NONE L=2 I=JOIN;
AXIS1 ORDER=(-2 TO 2.5 BY 0.5);
AXIS2 ORDER=(40 TO 65 BY 3);
AXIS3 LABEL=('TIME')
OFFSET=(1,1)
ORDER=(1 TO 301 BY 20);
PROC GLOT:
PLOT X*T / HAXIS=AXIS3 VAXIS=AXIS1;
PLOT2 Y*T / HAXIS=AXIS3 VAXIS=AXIS2;
RUN;
    
```

이 SAS프로그램을 수행하면 영구SAS데이터세트 BJJ가 디스켓 SAS4TSA4의 서브디렉토리 TSA4DATA의 파일 BJJ.SSD에 저장된다. 이 SAS프로그램의 PROC GLOT를 수행하면 그림 2.2의 시계열산포도를 얻는다. 이 시계열산포도의 실선은 입력개소비율인 변수 X를 나타내고, 점선은 출력개스내의 이산화탄소의 퍼센트인 변수 Y를 나타낸다.

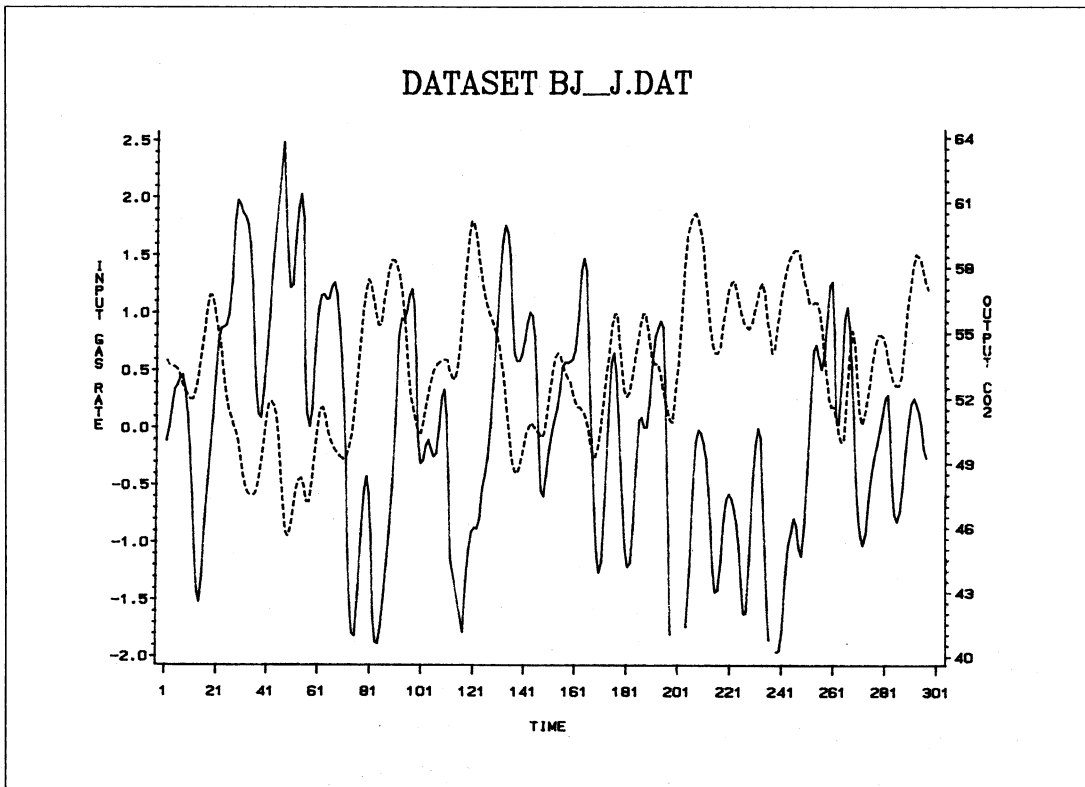


그림 2.2 다변량시계열데이터 BJ_의 시계열산포도

2.3 다변량시계열데이터 M3 : GRUNFELD

이 6변량시계열데이터 GRUNFELD는 Grunfeld(1958)가 그의 박사논문에서 제안한 '투자함수의 추정에 관한 이론'을 예증하기 위해서 사용한 다변량시계열데이터이다. Grunfeld는 두 개의 독립된 투자함수모형들을 추정할 때 각각의 투자모형을 개별적으로 추정하는 것보다는 두 개의 모형들이 보유하고 있는 정보를 이용해서 두 개의 모형들을 동시에 추정하는 것이 효율적임을 보였다. 이 이론을 예증하기 위해서 사용한 다변량시계열데이터들을 Grunfeld의 시계열데이터들이라고 부른다. Grunfeld의 시계열데이터들은 General Motors, U.S. Steel, General Electrics, Chrysler, Atlantic Richfield, IBM, Union Oil, Westinghouse, Goodyear, Diamond Match 등 미국의 대기업들의 투자비용에 관한 것이다. 이 Grunfeld의 시계열데이터들은 Grunfeld & Griliches(1960), Maddala(1977, pp. 214~215), 그리고 Greene(1991, pp.462~463) 등에 수록되어 있다. [ETS6.1](pp.148~152, pp.185~190)에서는 이 중에서 General Electrics(GE)와 Westinghouse(W)에 관한 변수들을 포함한 6변량시계열데이터를 다루었다. 이 6변량시계열데이터는 다음과 같은 변수들을 포함하고 있다.

- GEI = GE의 총투자액(gross investment)
- GEC = GE의 자본금(stock of plant and equipment)
- GEF = GE의 자산가치(value of the firm)
- WI = W의 총투자액
- WC = W의 자본금
- WF = W의 자산가치

이 6변량시계열데이터는 1935년부터 1954년까지의 20개의 관찰점들로 구성되어 있다. 이 6변량시계열데이터를 영구SAS데이터세트로 만들기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 GRUNFELD.PRG에 수록되어 있다).

```

/*                                                    */
/*          MULTIVARIATE TIME SERIES DATA GRUNFELD.DAT          */
/*          MAKING SASDATASET GRUNFELD.SSD                      */
/*                                                    */
LIBNAME TSA4DATA 'A:\TSA4DATA';
DATA TSA4DATA.GRUNFELD;

```

```

INPUT YEAR GEI GEF GEC WI WF WC;
LABEL GEI='Gross Investment GE'
      GEC='Capital Stock Lagged GE'
      GEF='Value of Shares GE Lagged'
      WI ='Gross Investment W'
      WC ='Capital Stock Lagged W'
      WF ='Value of Shares Lagged W';

```

CARDS:

1935	33.1	1170.6	97.8	12.93	191.5	1.8
1936	45.0	2015.8	104.4	25.90	516.0	.8
1937	77.2	2803.3	118.0	35.05	729.0	7.4
1938	44.6	2039.7	156.2	22.89	560.4	18.1
1939	48.1	2256.2	172.6	18.84	519.9	23.5
1940	74.4	2132.2	186.6	28.57	628.5	26.5
1941	113.0	1834.1	220.9	48.51	537.1	36.2
1942	91.9	1588.0	287.8	43.34	561.2	60.8
1943	61.3	1749.4	319.9	37.02	617.2	84.4
1944	56.8	1687.2	321.3	37.81	626.7	91.2
1945	93.6	2007.7	319.6	39.27	737.2	92.4
1946	159.9	2208.3	346.0	53.46	760.5	86.0
1947	147.2	1656.7	456.4	55.56	581.4	111.1
1948	146.3	1604.4	543.4	49.56	662.3	130.6
1949	98.3	1431.8	618.3	32.04	583.8	141.8
1950	93.5	1610.5	647.4	32.24	635.2	136.7
1951	135.2	1819.4	671.3	54.38	723.8	129.7
1952	157.3	2079.7	726.1	71.78	864.1	145.5
1953	179.5	2371.6	800.3	90.08	1193.5	174.8
1954	189.6	2759.9	888.9	68.60	1188.9	213.5

;

RUN;

```

TITLE1 'DATASET GRUNFELD.DAT';
SYMBOL1 C=BLACK V=NONE L=1 I=JOIN;
SYMBOL2 C=BLACK V=NONE L=2 I=JOIN;
AXIS1 ORDER=(1935 TO 1954 BY 1) MINOR=NONE;
AXIS2 ORDER=(10 TO 190 BY 30);
TITLE2 'GROSS INVESTMENT';
PROC GPLOT;
  PLOT (GEI WI)*YEAR
      / OVERLAY HAXIS=AXIS1 VAXIS=AXIS2;

```

```

RUN;
AXIS3 ORDER=(100 TO 2900 BY 400);
TITLE2 'VALUE OF OUTSTANDING SHARES';
PROC Gplot;
  PLOT (GEF WF)*YEAR
    / OVERLAY HAXIS=AXIS1 VAXIS=AXIS3;
RUN;
AXIS4 ORDER=(0 TO 900 BY 100);
TITLE2 'CAPITAL STOCK';
PROC Gplot;
  PLOT (GEC WC)*YEAR
    / OVERLAY HAXIS=AXIS1 VAXIS=AXIS4;
RUN;
TITLE:

```

이 SAS프로그램을 수행하면 영구SAS데이터세트 GRUNFELD가 디스켓 SAS4TSA4의 서브디렉토리 TSA4DATA의 파일 GRUNFELD.SSD에 저장된다. 이 SAS프로그램의 첫 번째 PROC Gplot를 수행하면 그림 2.3.1의 시계열산포도를 얻는다. 이 시계열산포도의 실선은 변수 GEI, 즉 GE의 총투자액을 나타내고, 점선은 변수 WI, 즉 Westinghouse의 총투자액을 나타낸다. 두 번째 PROC Gplot를 수행하면 그림 2.3.2의 시계열산포도를

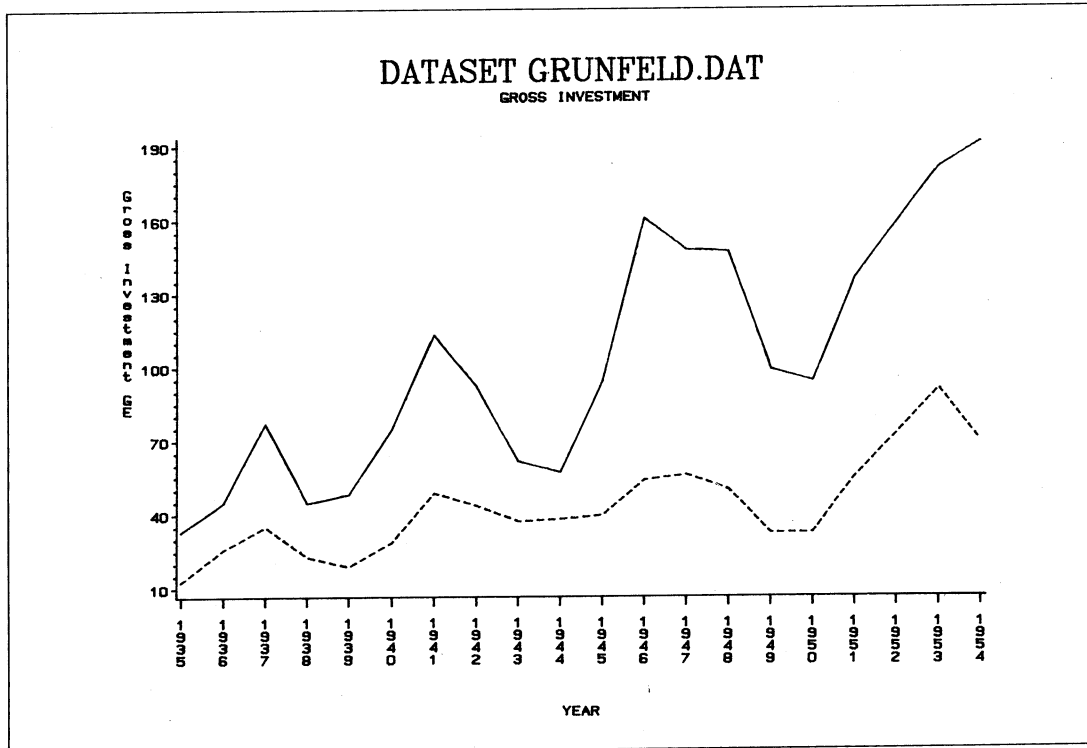


그림 2.3.1 다변량시계열데이터 GRUNFELD의 변수들 GEI와 WI의 시계열산포도

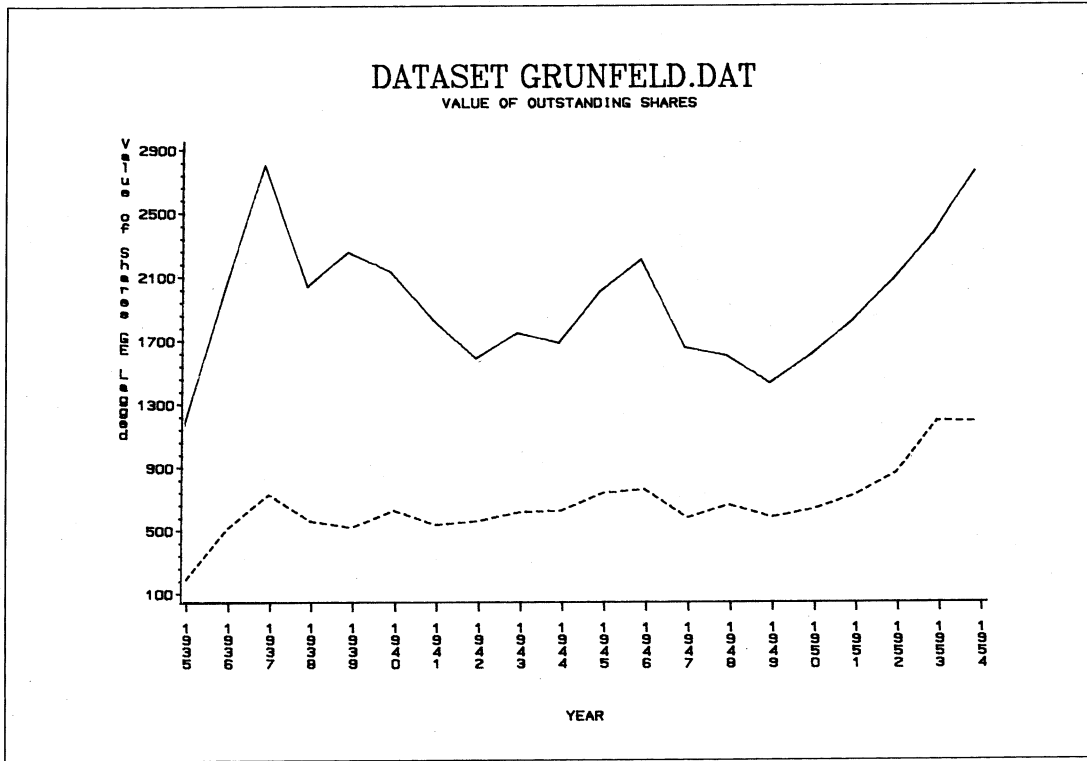


그림 2.3.2 다변량시계열데이터 GRUNFELD의 변수들 GEF와 WF의 시계열산포도

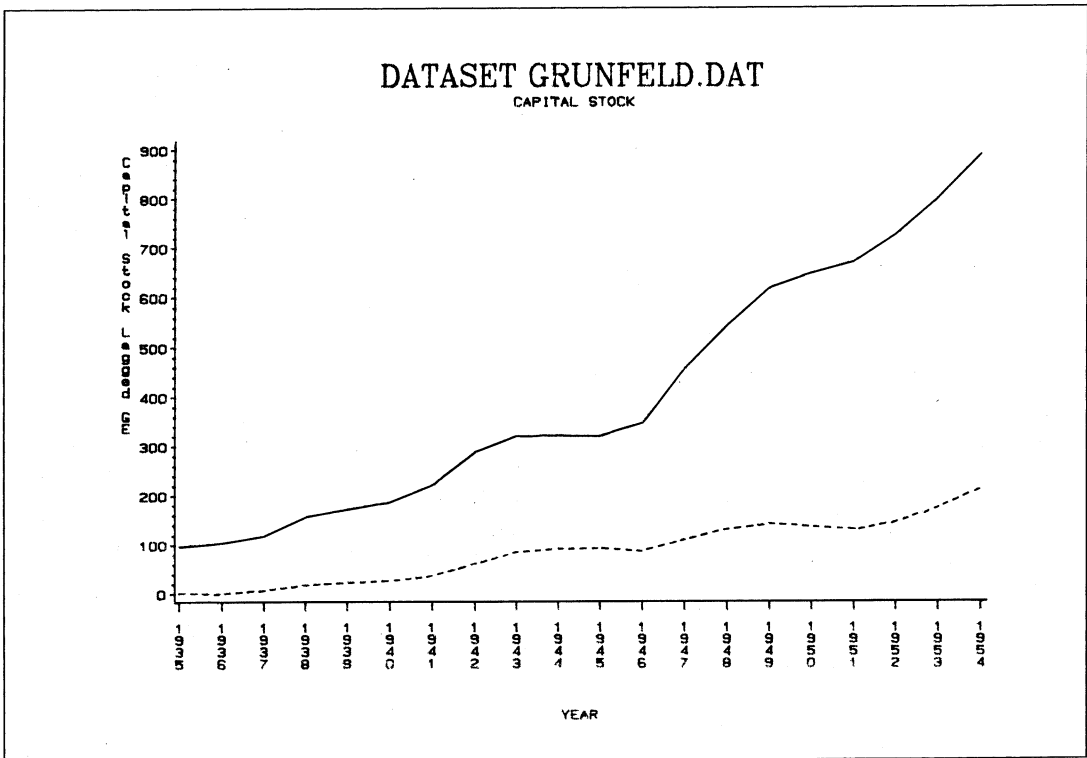


그림 2.3.3 다변량시계열데이터 GRUNFELD의 변수들 GEC와 WC의 시계열산포도

얻는다. 이 시계열산포도의 실선은 변수 GEF, 즉 GE가 가지고 있는 자산가치를 나타내고, 점선은 변수 WF, 즉 Westinghouse가 가지고 있는 자산가치를 나타낸다. 세 번째 PROC GPLOT를 수행하면 그림 2.3.3의 시계열산포도를 얻는다. 이 시계열산포도의 실선은 변수 GEC, 즉 GE의 자본금을 나타내고, 점선은 변수 WC, 즉 Westinghouse의 자본금을 나타낸다.

2.4 다변량시계열데이터 M4 : HOUSECON

다변량시계열데이터 HOUSECON은 미국의 상무성(U. S. Department of Commerce)에서 발간한 『*Survey of Current Business*』(1990)에서 발췌된 것으로 1983년 1월부터 1989년 10월까지의 월별 건설도금액(construction contracts in million of dollars)인 변수 CONSTR, 착공한 주택의 수(private housing starts in thousands of units)인 변수 HSTARTS, 그리고 월평균 새 주택자금이자율(average new home mortgage rate)인 변수 INTRATE를 기록한 것이다. 이 다변량시계열데이터의 관찰점들의 개수는 82이다. 이 시계열데이터를 영구SAS데이터세트로 만들기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 HOUSECON.PRG에 수록되어 있다).

```

/* */
/*          MULTIVARIATE TIME SERIES DATA HOUSECON.DAT          */
/*          MAKING SASDATASET HOUSECON.SSD                        */
/* */
LIBNAME TSA4DATA 'A:\TSA4DATA';
DATA TSA4DATA.HOUSECON;
    INPUT DATE:MONYY5. CONSTR INTRATE HSTARTS @@;
    FORMAT DATE MONYY5.;
CARDS:
JAN83 11358 13.00 91.3      FEB83 11355 12.62 96.3
MAR83 16100 12.97 134.6    APR83 16315 12.02 135.8
MAY83 19205 12.21 174.9    JUN83 20263 11.90 173.2
JUL83 16885 12.02 161.6    AUG83 19441 12.01 176.8
SEP83 17379 12.08 154.9    OCT83 16028 11.80 159.3
NOV83 15401 11.82 136.0    DEC83 13518 11.94 108.3
JAN84 14023 11.80 109.1    FEB84 14442 11.78 130.0
MAR84 17916 11.56 137.5    APR84 17655 11.55 172.7
MAY84 21990 11.68 180.7    JUN84 20036 11.61 184.0

```

JUL84	19224	11.91	162.1	AUG84	19367	11.89	147.4
SEP84	16923	12.03	148.5	OCT84	18413	12.27	152.3
NOV84	16616	12.27	126.2	DEC84	14220	12.05	98.9
JAN85	15154	11.77	105.4	FEB85	13652	11.74	95.8
MAR85	20004	11.42	145.2	APR85	20692	11.55	176.0
MAY85	22532	11.55	170.5	JUN85	20043	11.31	163.4
JUL85	22047	10.94	160.7	AUG85	21055	10.78	160.7
SEP85	20541	10.69	147.7	OCT85	21715	10.64	173.0
NOV85	17691	10.55	124.1	DEC85	16276	10.47	120.5
JAN86	15417	10.40	115.6	FEB86	16152	10.21	107.2
MAR86	19617	10.04	151.0	APR86	23754	9.87	188.2
MAY86	23050	9.84	186.6	JUN86	23740	9.74	183.6
JUL86	23621	9.89	172.0	AUG86	21884	9.84	163.8
SEP86	21763	9.74	154.0	OCT86	21862	9.57	154.8
NOV86	17998	9.45	115.6	DEC86	17982	9.28	113.0
JAN87	16694	9.14	105.1	FEB87	15729	8.87	102.8
MAR87	22622	8.77	141.2	APR87	23077	8.84	159.3
MAY87	22054	8.99	158.0	JUN87	25703	9.05	162.9
JUL87	24567	9.01	152.4	AUG87	23836	9.01	143.6
SEP87	22418	9.03	152.0	OCT87	23360	8.86	139.1
NOV87	18663	8.92	118.8	DEC87	19224	8.78	85.4
JAN88	15113	8.75	78.2	FEB88	17496	8.76	90.2
MAR88	22257	8.77	128.8	APR88	22344	8.76	153.2
MAY88	24138	8.59	140.2	JUN88	26940	8.90	150.2
JUL88	22309	8.80	137.0	AUG88	24826	8.68	136.8
SEP88	22670	8.90	131.1	OCT88	22223	8.77	135.1
NOV88	19767	9.05	113.0	DEC88	19125	9.04	94.2
JAN89	15776	9.20	100.1	FEB89	15086	9.46	85.8
MAR89	21080	9.63	117.8	APR89	21725	9.88	129.4
MAY89	23796	9.82	131.7	JUN89	24650	10.09	143.2
JUL89	22330	10.06	134.7	AUG89	24128	9.83	122.4
SEP89	23371	9.87	109.3	OCT89	22669	9.77	130.3

;

RUN;

TITLE1 'DATASET HOUSECON.DAT';

SYMBOL C=BLACK V=NONE L=1 I=JOIN;

AXIS1 ORDER=(10000 TO 28000 BY 3000);

AXIS2 ORDER=(8.0 TO 13.5 BY 0.5);

AXIS3 ORDER=(75 TO 200 BY 25);


```

AXIS4 ORDER=( '01JAN83'D TO '01JAN90'D BY YEAR);
PROC GPLOT;
  TITLE2 'CONSTRUCTION CONTRACTS';
  PLOT CONSTR*DATE / HAXIS=AXIS4 VAXIS=AXIS1;
  RUN;
PROC GPLOT;
  TITLE2 'NEW HOME MORTGAGE RATE';
  PLOT INTRATE*DATE / HAXIS=AXIS4 VAXIS=AXIS2;
  RUN;
PROC GPLOT;
  TITLE2 'PRIVATE HOUSING STARTS';
  PLOT HSTARTS*DATE / HAXIS=AXIS4 VAXIS=AXIS3;
  RUN;
TITLE;

```

이 SAS프로그램을 수행하면 영구SAS데이터세트 HOUSECON이 디스켓 SAS4TSA4의 서브디렉토리 TSA4DATA의 파일 HOUSECON.SSD에 저장된다. 이 SAS프로그램의 첫 번째 PROC GPLOT를 수행하면 그림 2.4.1의 시계열산포도를 얻는다. 이 시계열산포도의 실선은 변수 CONSTR, 즉 월별 건설도급액을 분기별로 나타낸다. 두 번째 PROC GPLOT를 수행하면 그림 2.4.2의 시계열산포도를 얻는다. 이 시계열산포도의 실선은 변수

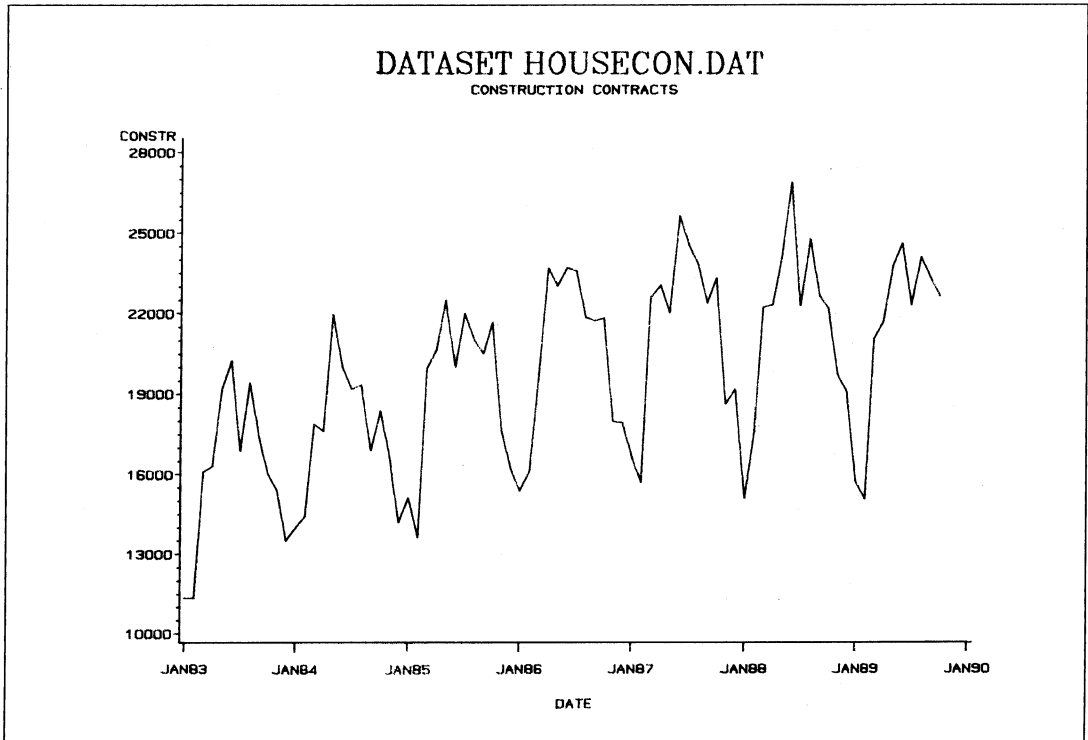


그림 2.4.1 다변량시계열데이터 HOUSECON의 변수 CONSTR의 시계열산포도

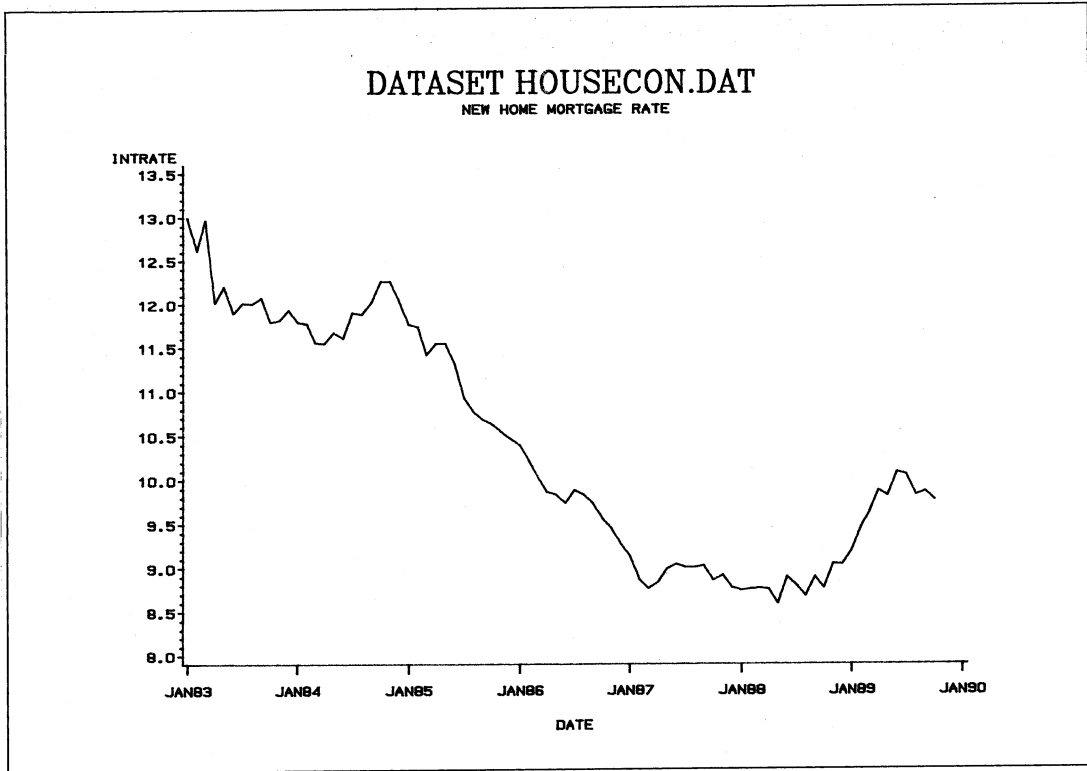


그림 2.4.2 다변량시계열데이터 HOUSECON의 변수 INTRATE의 시계열산포도

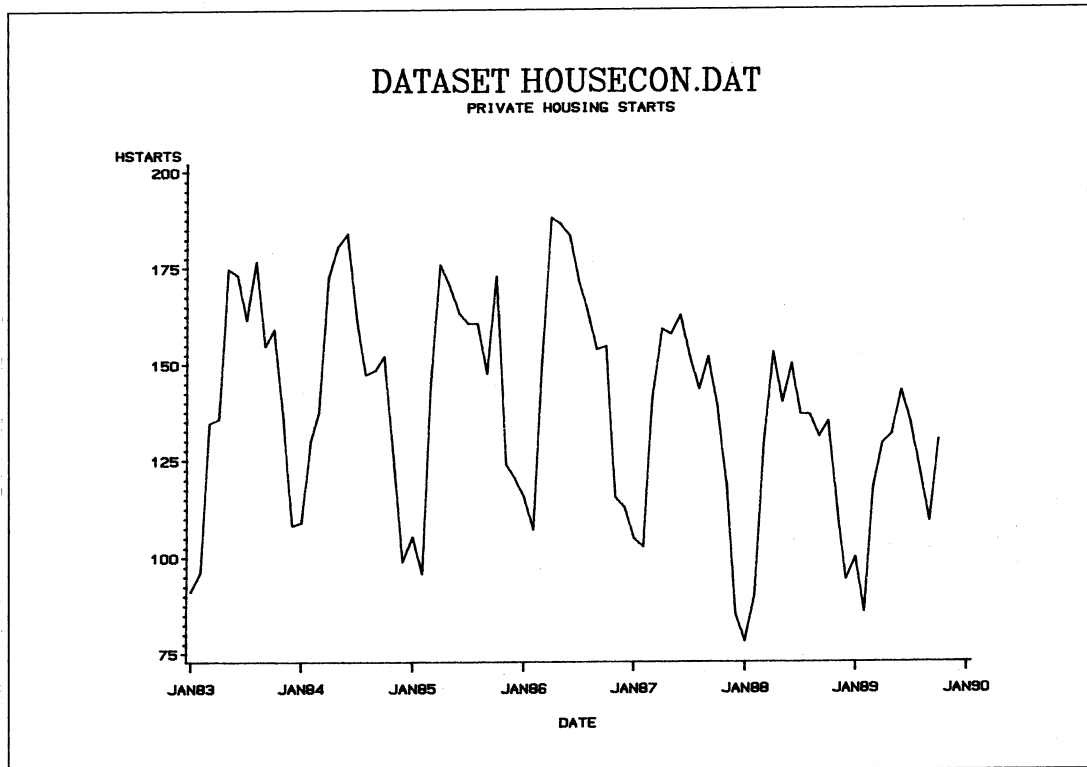


그림 2.4.3 다변량시계열데이터 HOUSECON의 변수 HSTARTS의 시계열산포도

INTRATE, 즉 분기별 주택자금이자율을 나타낸다. 세 번째 PROC GPLOT를 수행하면 그림 2.4.3의 시계열산포도를 얻는다. 이 시계열산포도의 실선은 변수 HSTARTS, 즉 월별 착공한 주택의 수를 나타낸다.

2.5 다변량시계열데이터 M5 : HOUSING

이 4변량시계열데이터는 미국의 한 건축재료업자가 1974년 1사분기부터 1984년 1사분기까지의 분기별 건축자재판매액을 나타내는 변수 SALES, 분기별 착공한 주택의 수를 나타내는 변수 STARTS, 분기별 처분가능한 개인소득(disposable personal income)을 나타내는 변수 DPIC, 그리고 분기별 주택자금이자율(mortgage rates)을 나타내는 변수 MORT의 관찰값들을 기록한 것이다. 이 시계열데이터의 관찰점들의 개수는 41이다. 이 시계열데이터를 영구SAS데이터세트로 만들기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 HOUSING.PRG에 수록되어 있다).

```

/*                                                    */
/*          MULTIVARIATE TIME SERIES DATA HOUSING.DAT          */
/*          MAKING SASDATASET HOUSING.SSD                      */
/*                                                    */
LIBNAME TSA4DATA 'A:\TSA4DATA';
DATA TSA4DATA.HOUSING;
  INPUT DATE YYQ4. DPIC STARTS MORT;
  LABEL DPIC='DIS PERS INCOME'
        STARTS='HOUSING STARTS'
        MORT='MORTGAGE RATES'
        SALES='SALES';
  FORMAT DATE YYQ4. ;
  T=_N_;
  E=NORMAL(72766133);
  EAR=-.35*EAR1+E-.25*E1;
  IF T=1 THEN EAR=2*NORMAL(72766133);
  NSERR=NSERR+EAR;
  IF T=1 THEN NSERR=EAR;
  DSTART=STARTS-START1;
  DMORT=MORT-MORT1;
  DINC=DPIC-DPIC1;

```

```

DSALES=5*DSTART-250*DMORT+3*DINC+300*EAR;
SALES=SALES+DSALES;
IF T=1 THEN SALES=17000;
IF T=41 THEN SALES=19950.3;
OUTPUT;
RETAIN;
DPIC1=DPIC;
START1=STARTS;
MORT1=MORT;
EAR1=EAR;
E1=E;
CARDS;
74Q1 949.5 1628 8.59 74Q2 973.1 1527 8.75 74Q3 999.7 1209 9.08
74Q4 1016.2 1001 9.27 75Q1 1025.4 995 9.17 75Q2 1092.2 1050 8.94
75Q3 1095.7 1258 8.91 75Q4 1124.1 1365 9.01 76Q1 1153.3 1400 8.95
76Q2 1174.1 1428 8.93 76Q3 1193.3 1570 9.02 76Q4 1222.6 1770 9.07
77Q1 1248.0 1756 9.00 77Q2 1285.3 1937 8.96 77Q3 1319.1 2041 9.02
77Q4 1359.6 2146 9.08 78Q1 1398.0 1721 9.20 78Q2 1440.7 2104 9.38
78Q3 1482.1 2044 9.67 78Q4 1531.0 2078 9.91 79Q1 1580.2 1615 10.23
79Q2 1612.8 1815 10.58 79Q3 1663.8 1809 11.05 79Q4 1710.1 1593 11.51
80Q1 1765.1 1263 12.37 80Q2 1784.1 1053 13.53 80Q3 1840.6 1412 12.41
80Q4 1897.0 1535 14.23 81Q1 1947.3 1391 14.56 81Q2 1985.6 1170 15.50
81Q3 2042.0 966 18.00 81Q4 2086.4 903 17.00 82Q1 2115.3 919 17.55
82Q2 2148.9 909 16.37 82Q3 2229.2 1063 16.88 82Q4 2279.6 1239 15.94
83Q1 2323.6 1314 15.76 83Q2 2366.2 1388 15.38 83Q3 2449.8 1377 14.99
83Q4 2497.7 1366 14.78 84Q1 2535.8 1366 15.01
;
RUN;
TITLE1 'DATASET HOUSING.DAT';
AXIS1 ORDER=( '01JAN74'D TO '01JAN85'D BY YEAR);
SYMBOL1 C=BLACK V=NONE L=1 I=JOIN;
TITLE2 'DISPOSABLE PERSONAL INCOME';
PROC GPLOT;
PLOT DPIC*DATE / HAXIS=AXIS1;
RUN;
TITLE2 'U.S. HOUSING STARTS';
PROC GPLOT;
PLOT STARTS*DATE / HAXIS=AXIS1;
RUN;

```

```

TITLE2 'MORTGAGE RATES';
PROC GPLOT;
    PLOT MORT*DATE / HAXIS=AXIS1;
    RUN;
TITLE 'SALES OF BUILDING SUPPLIES';
PROC GPLOT;
    PLOT SALES*DATE / HAXIS=AXIS1;
    RUN;

/ *
/ *          MULTIVARIATE TIME SERIES DATA HOUSCOMB.DAT          */
/ *          MAKING SASDATASET HOUSCOMB.SSD                       */
/ *
/ *
DATA HOUSEEND;
    INPUT DATE YYQ4. STARTS;
    CARDS;
84Q2 1373.2418      84Q3 1387.6686      84Q4 1369.1916
85Q1 1369.1916      85Q2 1371.2554      85Q3 1375.3670
85Q4 1370.1011      86Q1 1370.1011
;
    RUN;
TITLE 'DATA WITH FORECASTS OF STARTS APPENDED AND SALES =.';
DATA TSADATA.HOUSCOMB;
    SET TSA4DATA.HOUSING HOUSEEND;
    RUN;
TITLE:

```

이 SAS프로그램의 첫 번째 DATA스텝을 수행하면 영구SAS데이터세트 HOUSING이 디스켓 SAS4TSA4의 서브디렉토리 TSA4DATA의 파일 HOUSING.SSD에 저장된다. 이 SAS프로그램의 첫 번째 PROC GPLOT를 수행하면 그림 2.5.1의 시계열산포도를 얻는다. 이 시계열산포도의 실선은 변수 DPIC, 즉 분기별 처분가능한 개인소득을 나타낸다. 두 번째 PROC GPLOT를 수행하면 그림 2.5.2의 시계열산포도를 얻는다. 이 시계열산포도의 실선은 변수 STARTS, 즉 분기별 착공한 주택의 수를 나타낸다. 세 번째 PROC GPLOT를 수행하면 그림 2.5.3의 시계열산포도를 얻는다. 이 시계열산포도의 실선은 변수 MORT, 즉 분기별 주택자금이자율을 나타낸다. 네 번째 PROC GPLOT를 수행하면 그림 2.5.4의 시계열산포도를 얻는다. 이 시계열산포도의 실선은 변수 SALES, 즉 분기별 건축자재판매액을 나타낸다.

이 SAS프로그램의 두 번째 DATA스텝을 수행하면 SAS데이터세트 HOUSEEND가 만들어진다. 이 SAS데이터세트에는 1984년 2사분기부터 1986년 1사분기까지의 변수 STA-

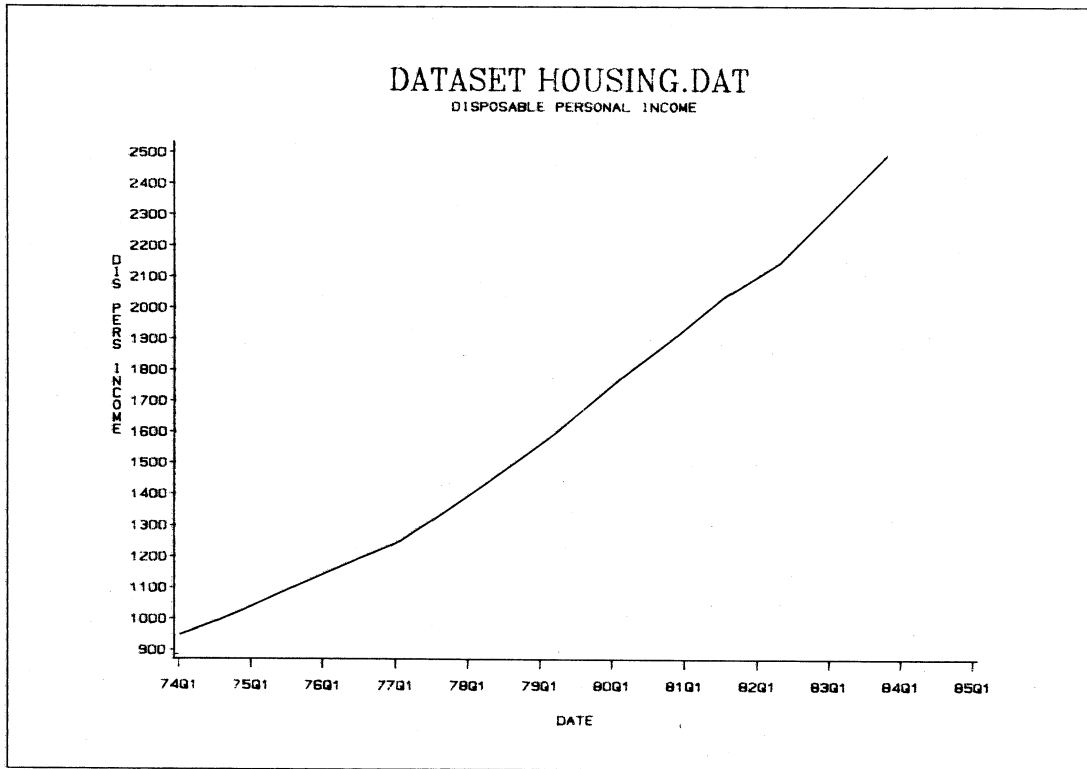


그림 2.5.1 다변량시계열데이터 HOUSING의 변수 DPIC의 시계열산포도

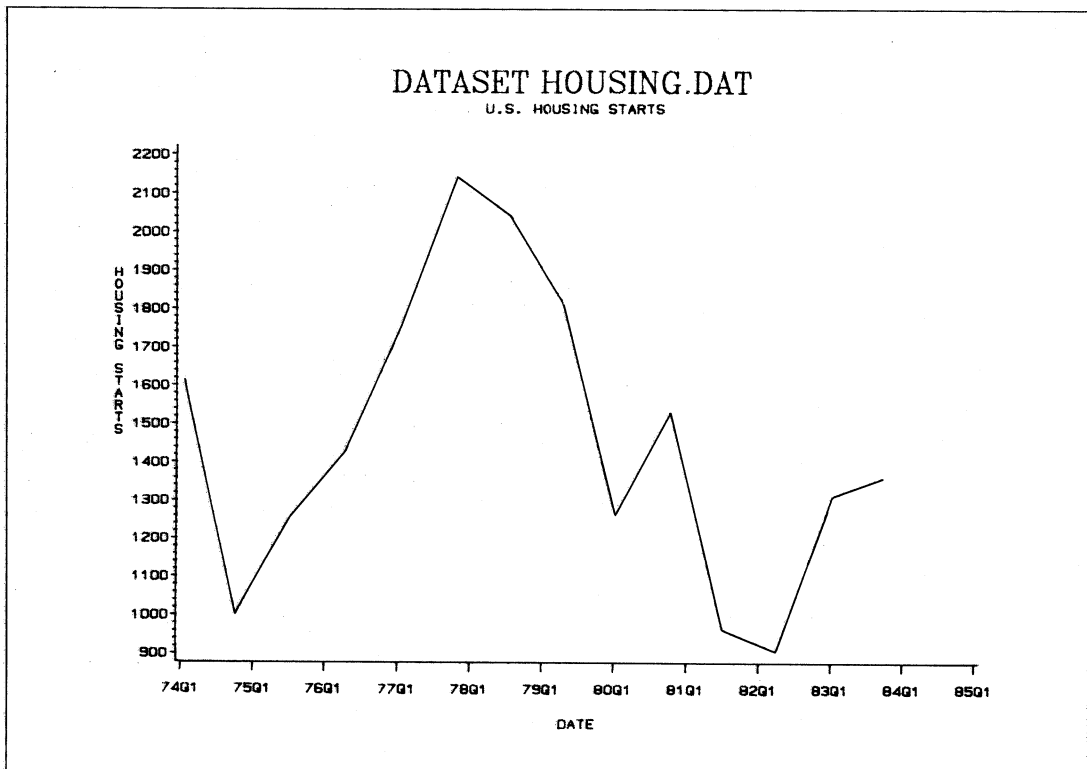


그림 2.5.2 다변량시계열데이터 HOUSING의 변수 STARTS의 시계열산포도

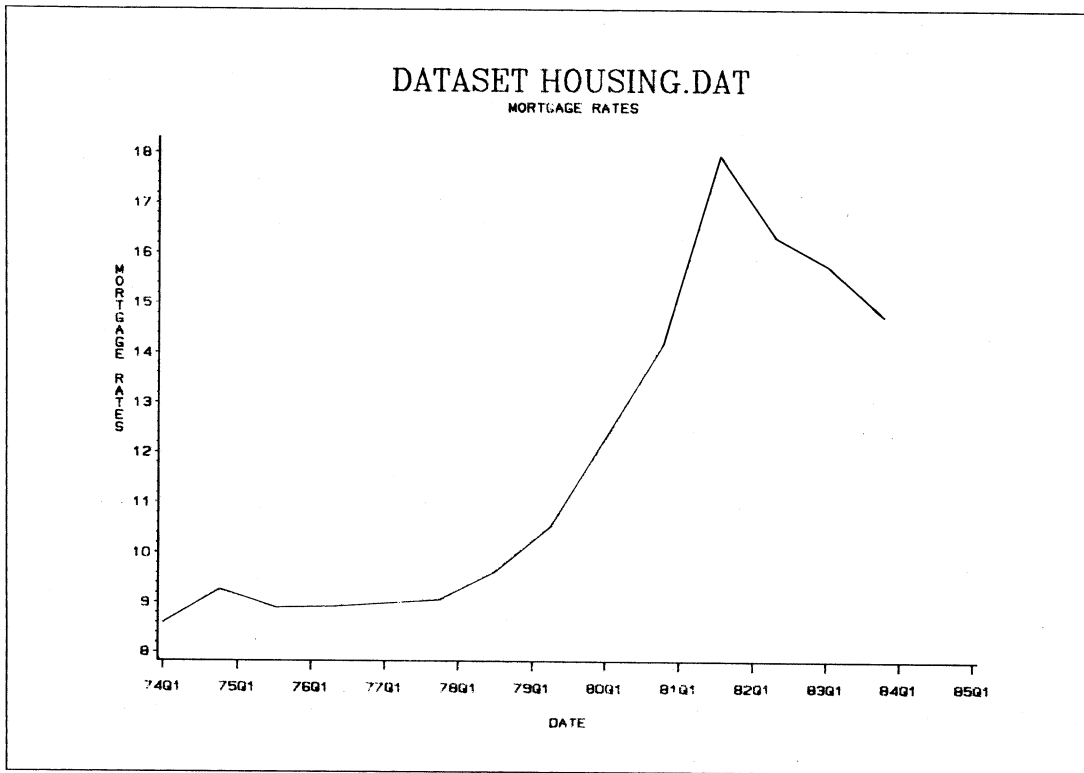


그림 2.5.3 다변량시계열데이터 HOUSING의 변수 MORT의 시계열산포도

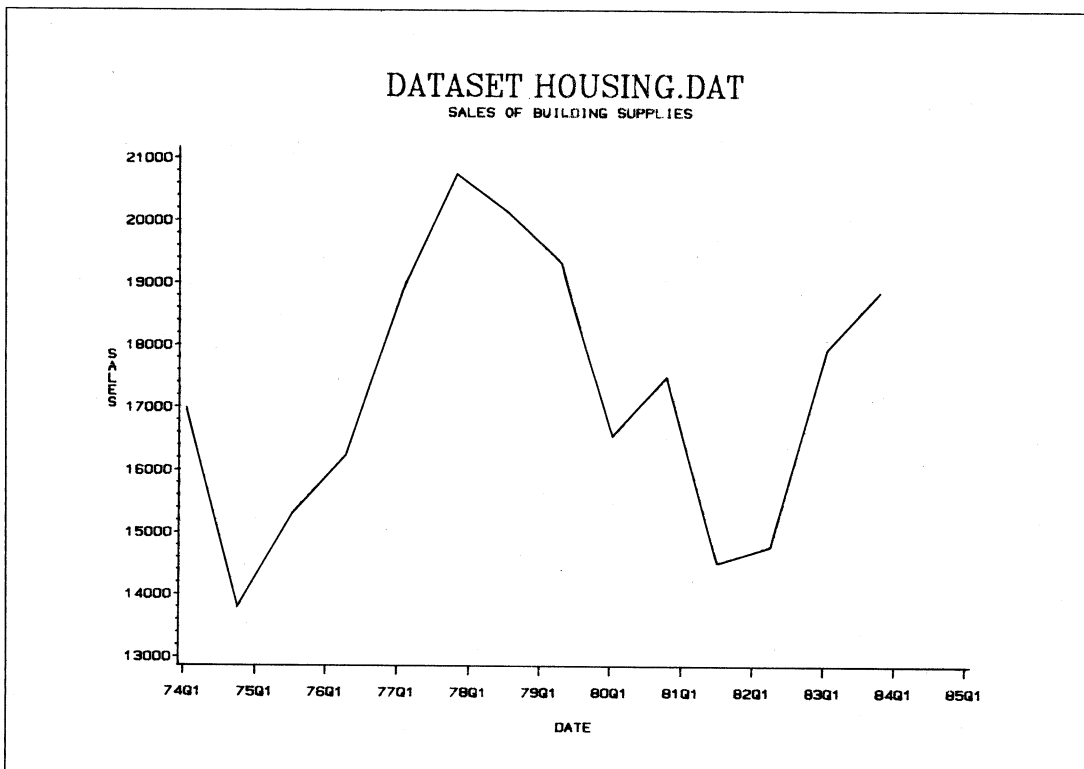


그림 2.5.4 다변량시계열데이터 HOUSING의 변수 SALES의 시계열산포도

RTS의 관찰값들이 수록되어 있다. 이 관찰값들은 변수 STARTS를 입력변수로 하는 모형에서 미래값들을 예측할 때 사용하게 될 것이다. 이 SAS프로그램의 세 번째 DATA스텝을 수행하면 영구SAS데이터세트 HOUSCOMB이 디스켓 SAS4TSA4의 서브디렉토리 TSA4DATA의 파일 HOUSCOMB.SSD에 저장된다. 이 SAS데이터세트는 영구SAS데이터세트 TSA4DATA.HOUSING과 임시SAS데이터세트 HOUSEEND를 세로로 결합한 것이다.

2.6 단변량시계열데이터 M6 : LEADPROD

이 단변량시계열데이터의 변수 LEADPROD는 1981년 1월부터 1985년 12월까지의 미국의 월별 납생산량을 기록한 것이다. 이 시계열데이터의 관찰점들의 개수는 100이다. 이 시계열데이터를 영구SAS데이터세트로 만들기 위해서 다음 SAS프로그램을 수행해 보자 (이 SAS프로그램은 디스켓 SAS4TSA4의 파일 LEADPROD.PRG에 수록되어 있다).

```

/* */
/*          UNIVARIATE TIME SERIES DATA LEADPROD.DAT          */
/*          MAKING SASDATASET LEADPROD.SSD                      */
/* */
LIBNAME TSA4DATA 'A:\TSA4DATA';
DATA TSA4DATA.LEADPROD;
    INPUT DATE:MONYY5. LEADPROD @@ ;
    FORMAT DATE MONYY. ;
    CARDS:
JAN81 57984    FEB81 53569    MAR81 58170    APR81 52741
MAY81 35038    JUN81 24378    JUL81 33027    AUG81 45097
SEP81 43704    OCT81 46082    NOV81 39412    DEC81 44036
JAN82 50257    FEB82 45193    MAR82 43994    APR82 36111
MAY82 54323    JUN82 52974    JUL82 44224    AUG82 46362
SEP82 49201    OCT82 50574    NOV82 50816    DEC82 48585
JAN83 57029    FEB83 50943    MAR83 52825    APR83 48938
MAY83 47230    JUN83 42045    JUL83 39525    AUG83 45283
SEP83 41226    OCT83 49972    NOV83 50886    DEC83 46381
JAN84 51500    FEB84 51400    MAR84 43000    APR84 40000
MAY84 39200    JUN84 26200    JUL84 28400    AUG84 29100
SEP84 25300    OCT84 40800    NOV84 28000    DEC84 26500

```



```
JAN85 43100    FEB85 50300    MAR85 56300    APR85 52700
MAY85 53400    JUN85 50900    JUL85 32500    AUG85 46600
SEP85 44900    OCT85 39300    NOV85 30100    DEC85 44500
```

;

RUN;

TITLE1 'DATASET LEADPROD.DAT';

SYMBOL C=BLACK V=NONE L=1 I=JOIN;

AXIS1 ORDER=(20000 TO 60000 BY 5000);

AXIS2 ORDER=('01JAN81'D TO '01JAN86'D BY YEAR);

PROC GPLOT;

TITLE2 '(LEAD PRODUCTION IN TONS)';

PLOT LEADPROD*DATE / HAXIS=AXIS2 VAXIS=AXIS1;

RUN;

TITLE;

이 SAS프로그램이 수행되면 영구SAS데이터세트 LEADPROD가 디스켓 SAS4TSA4의 서브디렉토리 TSA4DATA의 파일 LEADPROD.SSD에 저장된다. 이 SAS프로그램의 PROC GPLOT를 수행하면 그림 2.6의 시계열산포도를 얻는다. 이 시계열산포도의 실선은 월별 납생산량인 변수 LEADPROD를 나타낸다.

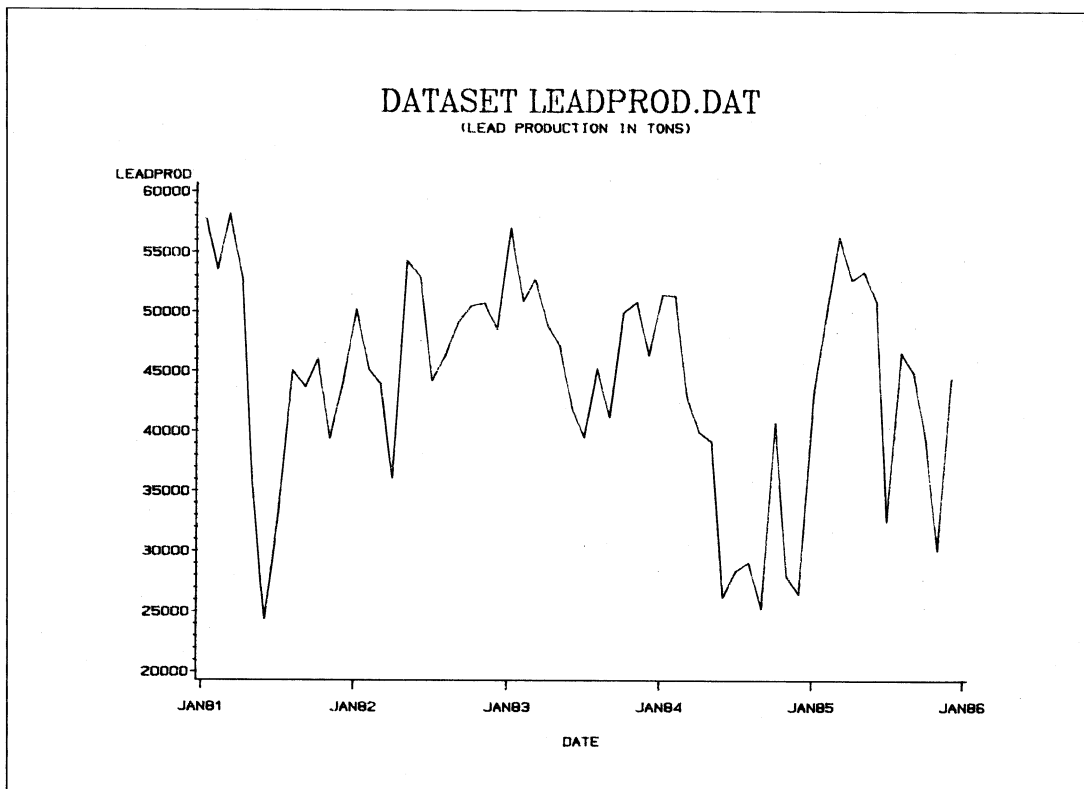


그림 2.6 단변량시계열데이터 LEADPROD의 시계열산포도

2.7 단변량시계열데이터 M7 : MA1A

단변량시계열데이터 MA1A는 다음 MA(1)모형에서 100개의 관찰점들을 발생시킨 것이다.

$$y_t = v_t + 0.8v_{t-1}$$

여기서 $\{v_t\}$ 는 평균이 0이고 분산이 1인 정규백색잡음과정이다. 이 시계열데이터를 발생시키고 또한 영구SAS데이터세트로 만들기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 MA1A.PRG에 수록되어 있다).

```

/*                                                    */
/*          MAKING UNIVARIATE TIME SERIES MA1A      */
/*          MAKING SASDATASET MA1A.SSD              */
/*                                                    */
LIBNAME TSA4DATA 'A:\TSA4DATA';
DATA DUM1;
  START:T+1;
  ERR = RANNOR(123456789);
  ERR1 = LAG(ERR);
  GROUP = 3;
  IF T < 152 THEN GROUP = 2;
  IF T < 52 THEN GROUP = 1;
  Y = ERR + 0.8*ERR1;
  OUTPUT;
  IF T < 751 THEN GO TO START;
RUN;
DATA TSA4DATA.MA1A;
  SET DUM1;
  T = _N_;
  IF GROUP = 2;
  KEEP Y T;
RUN;
TITLE1 'DATASET MA1A';
SYMBOL1 C=BLACK V=NONE L=1 I=JOIN;
PROC GPLOT;
  PLOT Y*T;

```

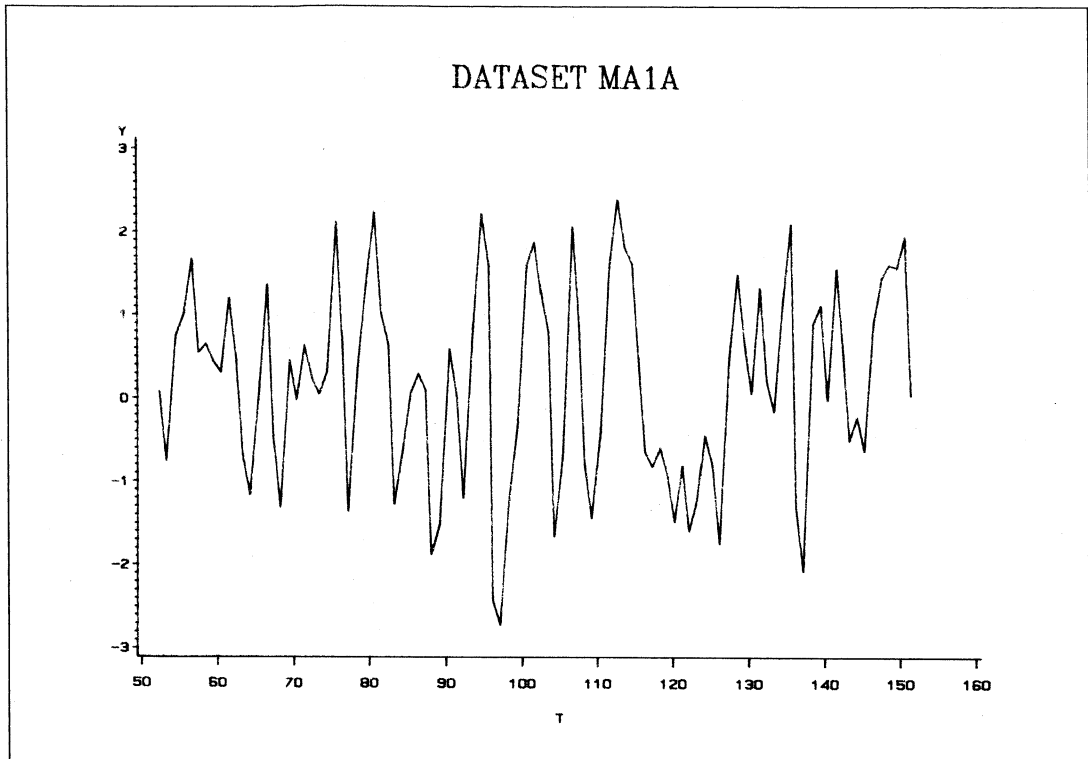


그림 2.7 단변량시계열데이터 MA1A의 시계열산포도

```
SYMBOL C=BLACK V=NONE I=JOIN;
```

```
RUN;
```

```
TITLE1;
```

이 SAS프로그램을 수행하면 영구SAS데이터세트 MA1A가 디스켓 SAS4TSA4의 서브디렉토리 TSA4DATA의 파일 MA1A.SSD에 저장된다. 이 SAS프로그램의 PROC GPLOT를 수행하면 그림 2.7의 시계열산포도를 얻는다. 이 시계열산포도의 실선은 변수 Y를 나타낸다.

2.8 다변량시계열데이터 M8 : MINKMUSK

다변량시계열데이터 MINKMUSK는 Canada의 Hudson's Bay Company가 1842년부터 1890년까지 Hudson 강가에서 잡힌 밍크(mink)와 사향쥐(muskrat)의 털가죽(pelt)을 유럽에 수출한 양을 기록한 것이다. 이 2변량시계열데이터의 변수 MINK는 수출한 밍크가죽의 수를 기록한 것이고, 변수 MUSKRAT는 수출한 사향쥐가죽의 수를 기록한 것이다. 이 다변량시계열데이터의 관찰점들의 개수는 49이다. 이 시계열데이터를 영구SAS데이

42 제 2 장 시계열데이터

터세트로 만들기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 MINKMUSK.PRG에 수록되어 있다).

```

/ *                                                                 */
/ *          MULTIVARIATE TIME SERIES DATA MINKMUSK.DAT          */
/ *          MAKING SASDATASET MINKMUSK.SSD                      */
/ *                                                                 */
LIBNAME TSA4DATA 'A:\TSA4DATA';
DATA TSA4DATA.MINKMUSK;
    INPUT YEAR MINK MUSKRAT @@;
    FORMAT MINK MUSKRAT COMMA10. ;
    CARDS;
1842 17780 549577      1843 25382 543155
1844 24855 265177      1845 32031 295617
1846 53264 303172      1847 36621 248710
1848 37123 224347      1849 35712 179075
1850 24772 192261      1851 17827 291281
1852 27413 488238      1853 39686 527161
1854 49373 319444      1855 58628 260805
1856 54924 295847      1857 65522 311953
1858 73066 243862      1859 55720 206156
1860 32548 205471      1861 38306 330527
1862 45534 356789      1863 59599 429304
1864 63724 367302      1865 49349 424875
1866 52113 312543      1867 73752 610280
1868 81769 469775      1869 35081 272963
1870 28184 436509      1871 35660 590916
1872 44352 711174      1873 55496 659159
1874 63810 474942      1875 83319 626711
1876 79206 583319      1877 90080 464297
1878 63318 511993      1879 36360 519963
1880 38828 830100      1881 40834 1028187
1882 56882 1081489     1883 52258 1082999
1884 110610 817003     1885 76393 347050
1886 64215 380022     1887 82941 344818
1888 43641 223615     1889 35400 322360
1890 29363 574742
;
    RUN;
```

```
TITLE1 'DATASET MINKMUSK.DAT';
TITLE2 'MINK & MUSKRAT TRAPPING 1842-1890';
SYMBOL1 C=BLACK V=NONE L=1 I=JOIN;
SYMBOL2 C=BLACK V=NONE L=2 I=JOIN;
AXIS1 ORDER=(1842 TO 1892 BY 5);
AXIS2 ORDER=(10000 TO 120000 BY 10000);
AXIS3 ORDER=(100000 TO 1100000 BY 100000);
PROC GPLOT;
    PLOT MINK*YEAR
        / HAXIS=AXIS1 HM=4 VAXIS=AXIS2;
    PLOT2 MUSKRAT*YEAR
        / HAXIS=AXIS1 HM=4 VAXIS=AXIS3;
RUN;
TITLE;
```

이 SAS프로그램을 수행하면 영구SAS데이터세트 MINKMUSK가 디스켓 SAS4TSA4의 서브디렉토리 TSA4DATA의 파일 MINKMUSK.SSD에 저장된다. 이 SAS프로그램의 PROC GPLOT를 수행하면 그림 2.8의 시계열산포도를 얻는다. 이 시계열산포도의 실선은 수출한 밍크가죽의 수 MINK를 나타내고, 점선은 수출한 사향쥐가죽의 수 MUSKRAT를 나타낸다.

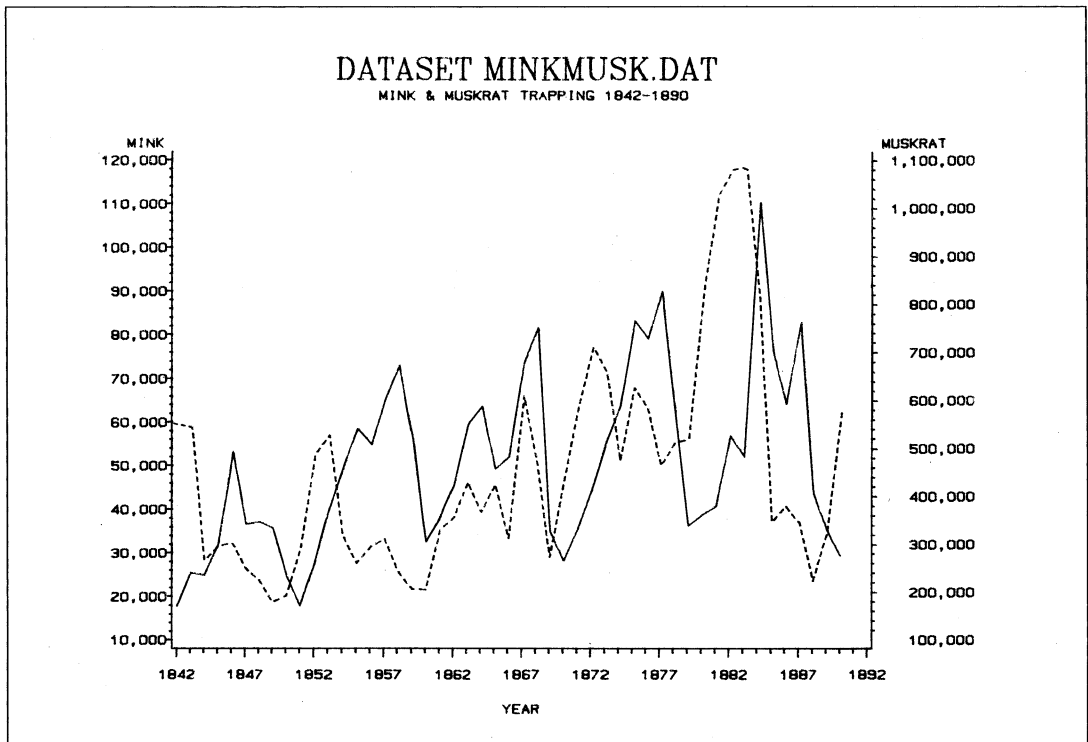


그림 2.8 다변량시계열데이터 MINKMUSK의 시계열산포도

2.9 다변량시계열데이터 M9 : RIVER

이 2변량시계열데이터는 미국 North Carolina 주의 Neuse 강변의 도시 Goldsboro에서의 유속과 이 도시로부터 30마일 하류에 있는 도시 Kinston에서의 유속을 일별로 기록한 것이다. 변수 GOLD는 Goldsboro에서 측정한 유속을 나타내고, 변수 KINS는 Kinston에서 측정한 유속을 나타낸다. 이 다변량시계열데이터는 400개의 관찰점들로 구성되어 있다. 이 2변량시계열데이터를 영구SAS데이터세트로 만들기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 RIVER.PRG에 수록되어 있다).

```

/*                                     */
/*          MULTIVARIATE TIME SERIES DATA RIVER.DAT          */
/*          MAKING SASDATASET RIVER.SSD                       */
/*                                     */

LIBNAME TSA4DATA 'A:\TSA4DATA';
DATA TSA4DATA.RIVER;
  INPUT GOLD KINS @@;
  RETAIN DATE '30SEP70'D;
  LGOLD=LOG(GOLD);
  LKINS=LOG(KINS);
  DATE=INTNX('DAY',DATE,1);
  FORMAT DATE DATE7. ;
  LABEL LGOLD='LOG OF GOLDSBORO'
        LKINS='LOG OF KINSTON';
  CARDS:
525  548  616  615  361  716  298  555  298  433  160  415
153  342  182  271  212  280  262  295  160  349  184  289
143  286  190  265  156  280  146  280  138  286  187  271
361  262  271  373  316  433  382  471  416  583  539  653
556  643  429  709  475  646  464  594  367  622  271  562
528  534  493  759 1030  875 1460 1050 1330 1480 1100 1540
895 1390  753 1190  620 1020  592  864  606  777  620  852
810 1000 1520 1140 1810 1520 1880 1880 2060 1990 2250 2100
1770 2250 1810 2090 1350 1950 1110 1720 1010 1410  912 1260
808 1160  765 1060  736  997  701  952  696  919  689  887

```

605	876	591	811	702	751	619	826	563	821	613	738
612	752	593	768	573	751	562	738	539	723	509	704
523	676	574	655	512	687	488	682	640	652	962	878
1150	1240	1710	1360	1770	1670	1600	1880	1390	1830	1400	1690
1540	1710	1790	1800	2050	1910	2120	2100	1890	2230	1700	2160
1360	1940	1210	1700	1130	1550	1110	1450	1190	1380	1350	1410
1490	1520	1730	1640	2770	1830	3500	2620	4000	3080	4400	3610
4900	4110	5420	4560	5480	5000	5130	5380	4560	5820	3870	5990
3230	5800	2750	5230	2470	4490	2200	3790	1980	3210	1800	2760
1690	2460	1900	2310	2540	2580	3130	3040	3560	3420	3670	3780
3370	4040	2880	4150	2750	4220	2820	4140	3130	4050	3320	3970
3250	4050	3300	4160	3740	4270	4660	4430	5560	4740	6480	5310
7200	5950	7870	6480	8630	7000	9510	7830	10300	8670	10700	9350
10600	10100	10300	10700	9770	11000	8970	11100	7570	10800	5990	10200
4290	9380	4630	8450	5540	7460	6440	6860	7040	6670	7070	6860
6070	7200	4750	7460	4230	7390	5560	7370	7790	8220	8410	9170
9400	9300	10400	9510	11400	10000	12200	10700	12300	11500	11800	12400
10700	12900	8570	13000	6260	12500	4080	11300	3500	9070	3410	7010
3530	5530	3560	4930	3280	5620	2810	4280	2520	3940	2350	3480
2230	3090	2040	2850	2080	2710	2470	2760	2940	2980	3300	3240
3900	3560	4510	3920	4930	4320	5190	4770	5360	5200	5080	5560
3930	5770	4450	5980	5280	5950	6020	6110	6580	6390	6880	6670
6930	6940	6850	7130	6640	7250	5700	7260	3930	7140	2700	6790
2350	5530	2060	4160	1840	3160	1690	2660	1560	2260	1600	2060
1630	1990	1620	2020	1670	2010	1620	1990	1530	1970	1700	1980
2030	2220	2670	2390	3070	2660	3040	3000	2630	3230	2130	3190
1850	2830	1600	2350	1360	2000	1240	1730	1180	1540	1100	1440
1110	1350	0999	1310	1120	1300	1220	1460	2160	1560	2760	2210
3020	2750	3550	3070	3910	3330	4110	3620	4280	4100	4240	4430
3620	4450	1900	4370	1370	3520	1180	2110	1030	1520	902	1310
874	1410	1070	1270	1910	1270	2580	1760	3000	2280	3170	2700
3350	2940	2880	3110	1490	3180	1020	2320	846	1360	782	1060
711	946	525	872	630	732	753	764	634	856	634	804
651	753	623	773	623	763	641	757	683	756	634	779
578	766	606	844	571	975	588	826	585	758	634	747
585	762	477	729	407	649	394	580	519	561	842	675
606	972	560	821	439	722	616	680	749	897	854	1030
918	999	854	1110	1030	1080	1190	1140	942	1260	627	1160
532	855	442	708	400	627	314	552	362	506	445	524

```

455 588 690 596 752 690 602 825 458 768 388 722
506 688 602 699 1350 720 1480 1250 1370 1550 1140 1580
914 1410 814 1170 763 987 588 921 850 784 1090 828
926 1080 606 1050 529 830 483 670 423 625 350 574
347 511 445 535 543 718 602 995 609 1090 806 954
874 968 763 1060 676 997 496 920 648 773 858 779
958 1030 822 1330 954 1360 1040 1300 1430 1300 1520 1480
1000 1670 641 1370 557 998 435 807 338 715 296 603
376 553 474 570 423 654 359 672 376 703 407 939
439 926 1620 728 2310 1340 2590 2030 2370 2420 1090 2480
651 1660 451 1060 496 883 464 697 557 648 539 634
529 658 416 641 376 578 362 504 567 655 790 1840
906 2360 2400 2680 3470 3080 4040 3580 4420 3990 4490 4350
4420 4560 4400 4700 4630 4890 4920 5130 5110 5470 5240 5680
5180 5710 5060 5710 5040 5660 5020 5600 4820 5530 3420 5440
2070 5240 1650 4250 1530 3410 2330 3740 3470 4440 4600 5130
5330 5380 6140 5630 6870 5990 7470 6420 7830 6860 7890 7330
7720 7780 7580 8100 7630 8320 7580 8280

```

```

:
      RUN;
      TITLE1 'DATASET RIVER.DAT';
      SYMBOL1 C=BLACK V=NONE L=1 I=JOIN;
      SYMBOL2 C=BLACK V=NONE L=2 I=JOIN;
      AXIS1 LABEL=('DATE') OFFSET=(1,1)
            ORDER=('01SEP70'D TO '31DEC71'D BY MONTH);
      PROC GPLOT;
            PLOT GOLD*DATE / HAXIS=AXIS1;
            PLOT2 KINS*DATE / HAXIS=AXIS1;
      RUN;
      TITLE1:

```

이 SAS프로그램을 수행하면 영구SAS데이터세트 RIVER가 디스켓 SAS4TSA4의 서브 디렉토리 TSA4DATA의 파일 RIVER.SSD에 저장된다. 이 SAS프로그램의 PROC GPLOT를 수행하면 그림 2.9의 시계열산포도를 얻는다. 이 시계열산포도의 실선은 Goldsboro에서 측정한 유속 GOLD를 나타내고, 점선은 Kinston에서 측정한 유속 KINS를 나타낸다.

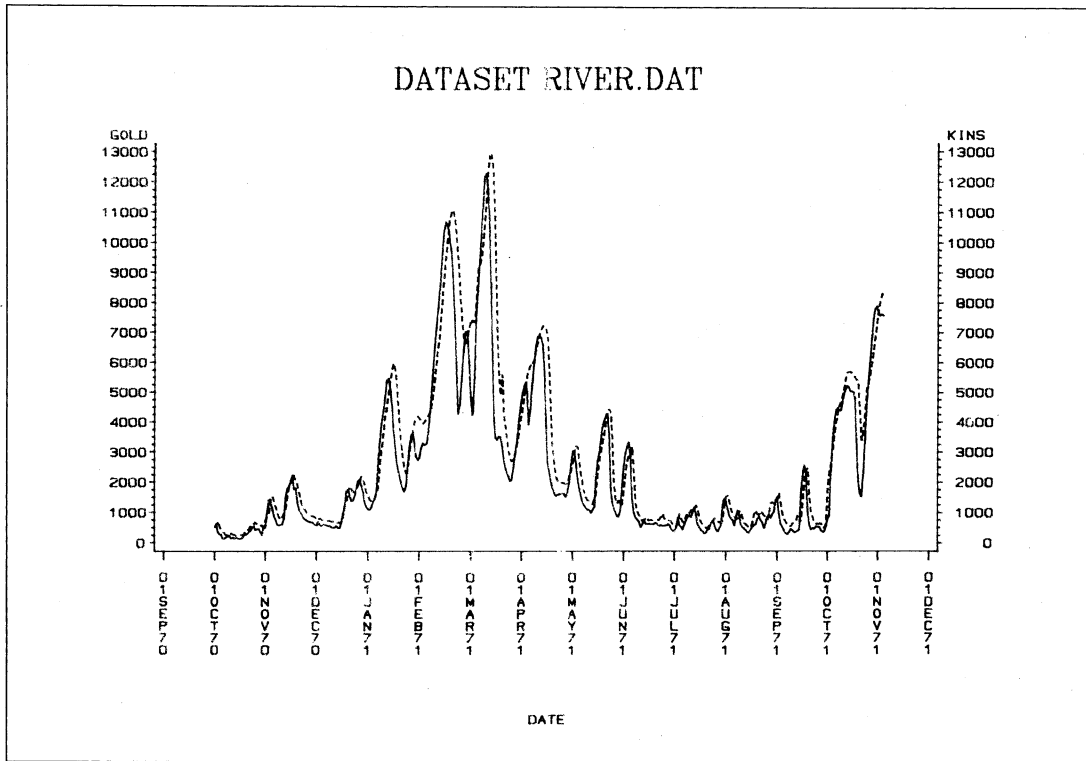


그림 2.9 다변량시계열데이터 RIVER의 시계열산포도

2.10 다른 시계열데이터들

본서에서는 이 장에서 언급된 시계열데이터들 이외에도 여러 시계열데이터들이 분석될 것이다. 이 중에는 모의실험에 의해서 만들어진 시계열데이터들도 있고 측정된 시계열데이터들도 있다. 특히 제 5.2절에서 다루게 될 간접시계열데이터들은 편의상 제 5.2절에 수록하였다.

SAS / IML의 사용법

- 3.1 SAS / IML 맛보기
- 3.2 SAS / IML의 기초
- 3.3 행렬의 연산
- 3.4 SAS / IML 함수
- 3.5 제어스태이트먼트
- 3.6 일반목적명령어들
- 3.7 SAS / IML에서의 데이터처리
- 3.8 외부파일의 입출력

제 3 장 SAS / IML의 사용법

유감스럽게도 SAS/ETS의 프로시저들만으로는 현재 널리 사용되고 있는 다변량시계열 분석의 기법들을 적용하는데 한계가 있다. 따라서, SAS시스템을 사용해서 다변량시계열 분석을 할 때에는 SAS의 프로그래밍언어인 SAS/IML을 사용해서 적당한 프로그램을 만들어야 할 때가 있다. 이 장에서는 이러한 SAS프로그램을 만들 수 있는 정도의 SAS/IML의 기초적인 사용법을 설명하고자 한다. 그러나 본서에서 소개하는 다변량시계열분석기법들을 이해하기 위해서 이 장을 반드시 읽어야 하는 것은 아니다. 만일, 독자 자신이 SAS/IML프로그래밍을 하지 않으면서 단지 본서의 SAS/IML프로그램들을 이해하고자 한다면 이 장을 읽지 않아도 된다. 제4장부터는 독자들이 이 장을 읽지 않았다고 가정하고 SAS/IML프로그램들을 설명할 것이다. SAS/IML에 관한 좀더 자세한 내용은 [IML6.03]을 참조하라.

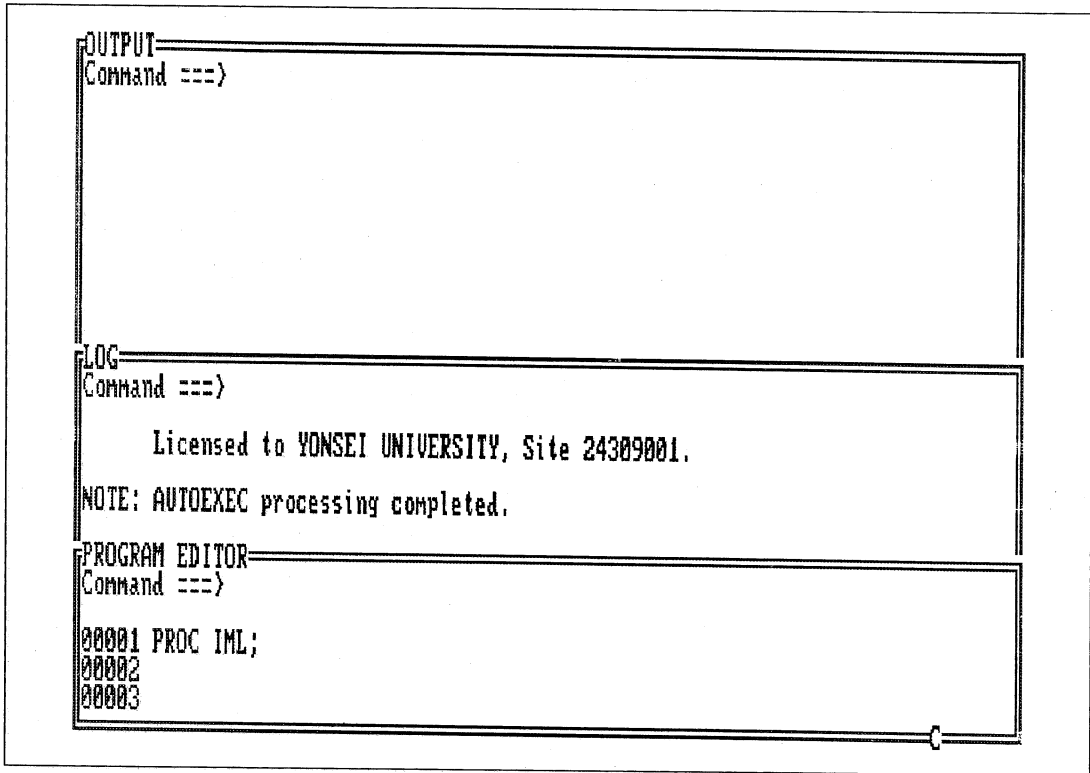
3.1 SAS/IML 맛보기

SAS/IML을 사용하기 위해서는 일단 SAS를 부팅시켜야 한다. 본서의 독자들은 당연히 이 과정을 알고 있으리라 생각한다. 만일, 이 과정을 모른다면 [SAS4TSA1]을 일독한 다음에 본서를 읽어야 할 것이다. 일단 SAS가 부팅되면 PGM윈도우에 다음과 같이 입력한다.

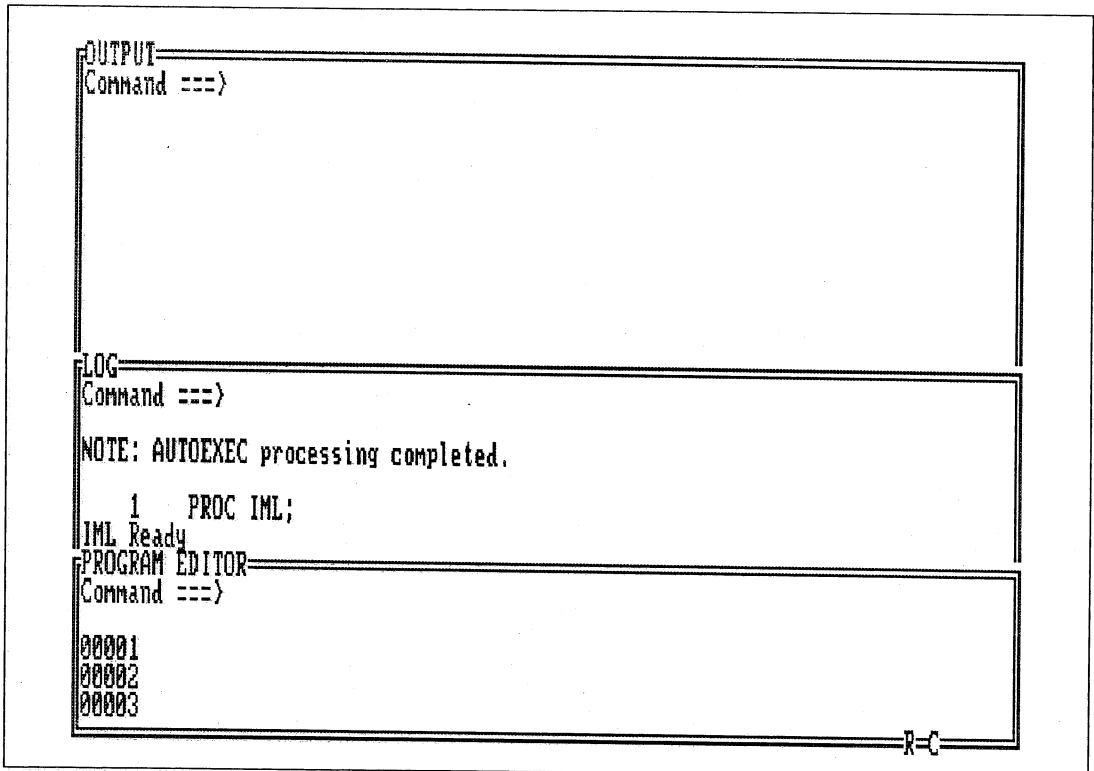
```
PROC IML;
```

이 PROC IML스테이트먼트가 입력된 컴퓨터의 디스플레이(display)의 화면은 화면그림 3.1과 같다.

이 PROC IML스테이트먼트를 실행하기 위해서 기능키 F10을 누른 뒤의 디스플레이의 화면은 화면그림 3.2와 같고, 이는 SAS/IML이 작동되고 있는 상태를 보여준다. 즉, SAS



화면그림 3.1 PROC IML스테이트먼트를 PGM윈도우에 입력



화면그림 3.2 SAS/IML환경

```

OUTPUT
Command ==>

LOG
Command ==>
NOTE: AUTOEXEC processing completed.

1 PROC IML;
IML Ready
PROGRAM EDITOR
Command ==>

00001 QUIT;
00002
00003

```

화면그림 3.3 QUIT상태먼트의 입력

시스템이 SAS/IML환경(SAS/IML environment)으로 바뀐 것이다.

어떤 소프트웨어를 작동시키는 과정을 아는 것 못지않게 이 소프트웨어의 작동상태에서 벗어나는 과정을 아는 것도 매우 중요하다. SAS/IML환경에서 벗어나기 위해서는 PGM 윈도우에 다음과 같이 입력한다.

QUIT;

이 QUIT상태먼트가 입력된 디스플레이의 화면은 화면그림 3.3과 같다.

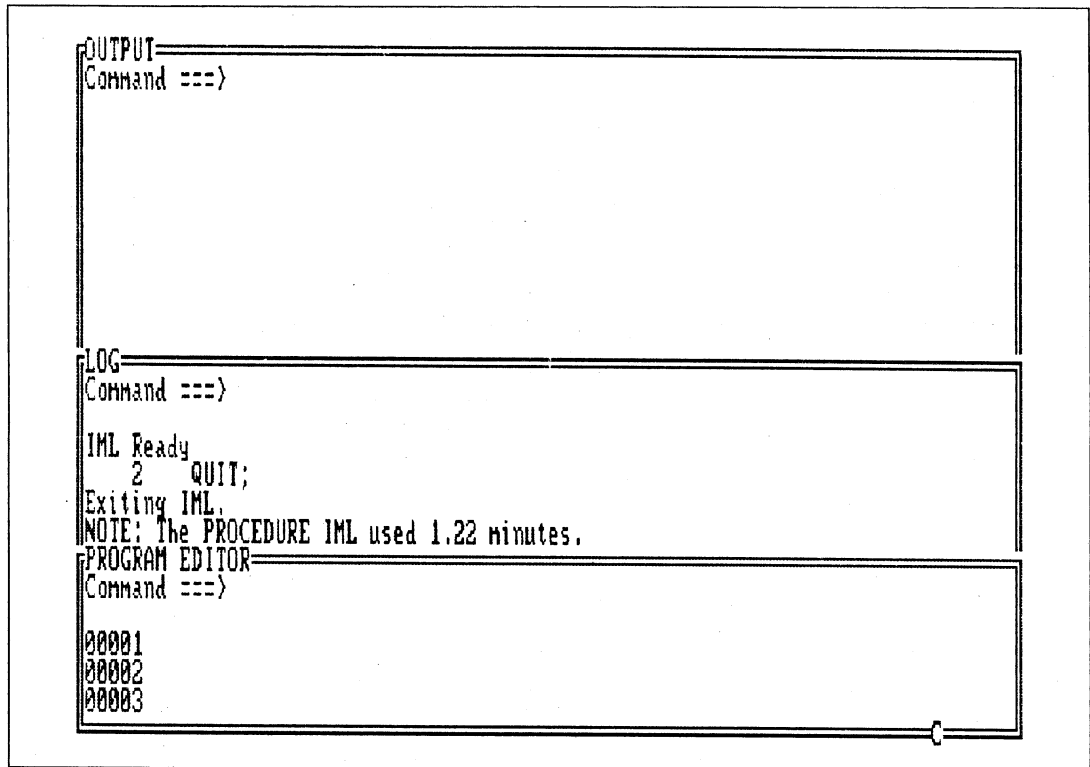
이 QUIT상태먼트를 실행하기 위해서 기능키 F10을 누르면 디스플레이의 화면은 화면그림 3.4와 같다. 이는 SAS/IML이 작동되지 않는 상태이다. 즉, SAS시스템이 SAS/IML환경에서 벗어난 것이다.

지금부터는 SAS/IML을 사용해서 간단한 행렬연산을 해보기로 하자. 우선 행렬들 A 와 B 를 다음과 같이 정의하자.

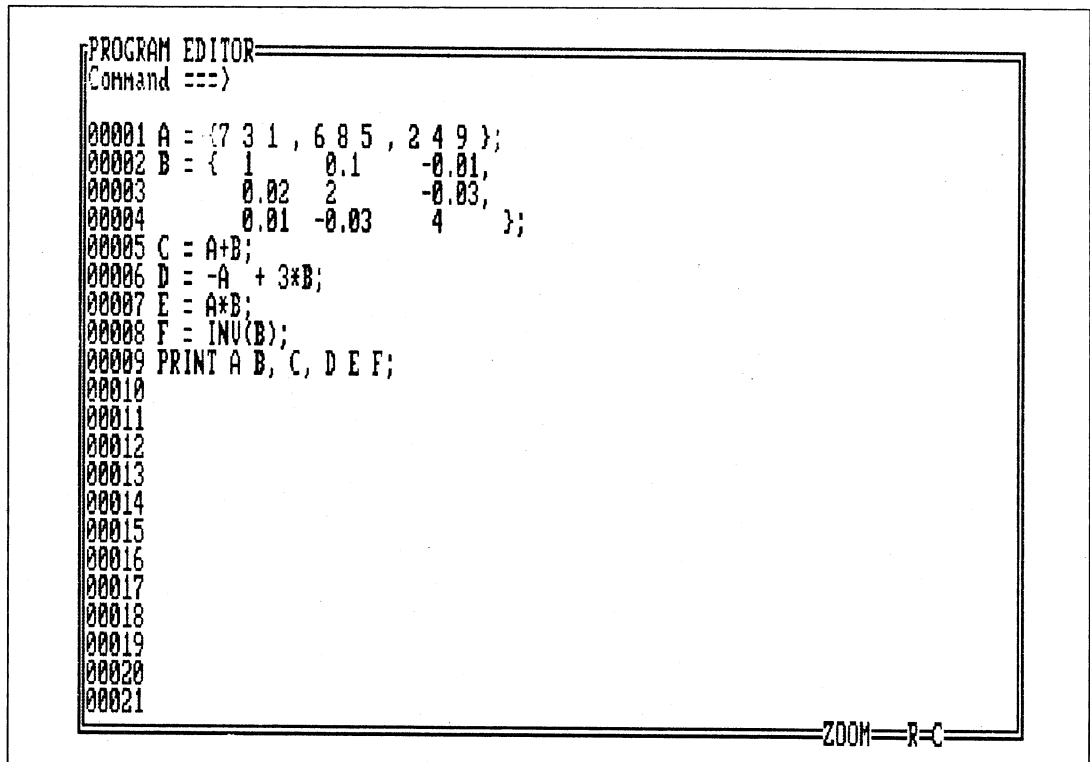
$$A = \begin{bmatrix} 7 & 3 & 1 \\ 6 & 8 & 5 \\ 2 & 4 & 9 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0.1 & -0.01 \\ 0.02 & 2 & -0.03 \\ 0.01 & -0.03 & 4 \end{bmatrix}$$

SAS/IML을 사용해서 다음 행렬들을 계산하기로 하자.

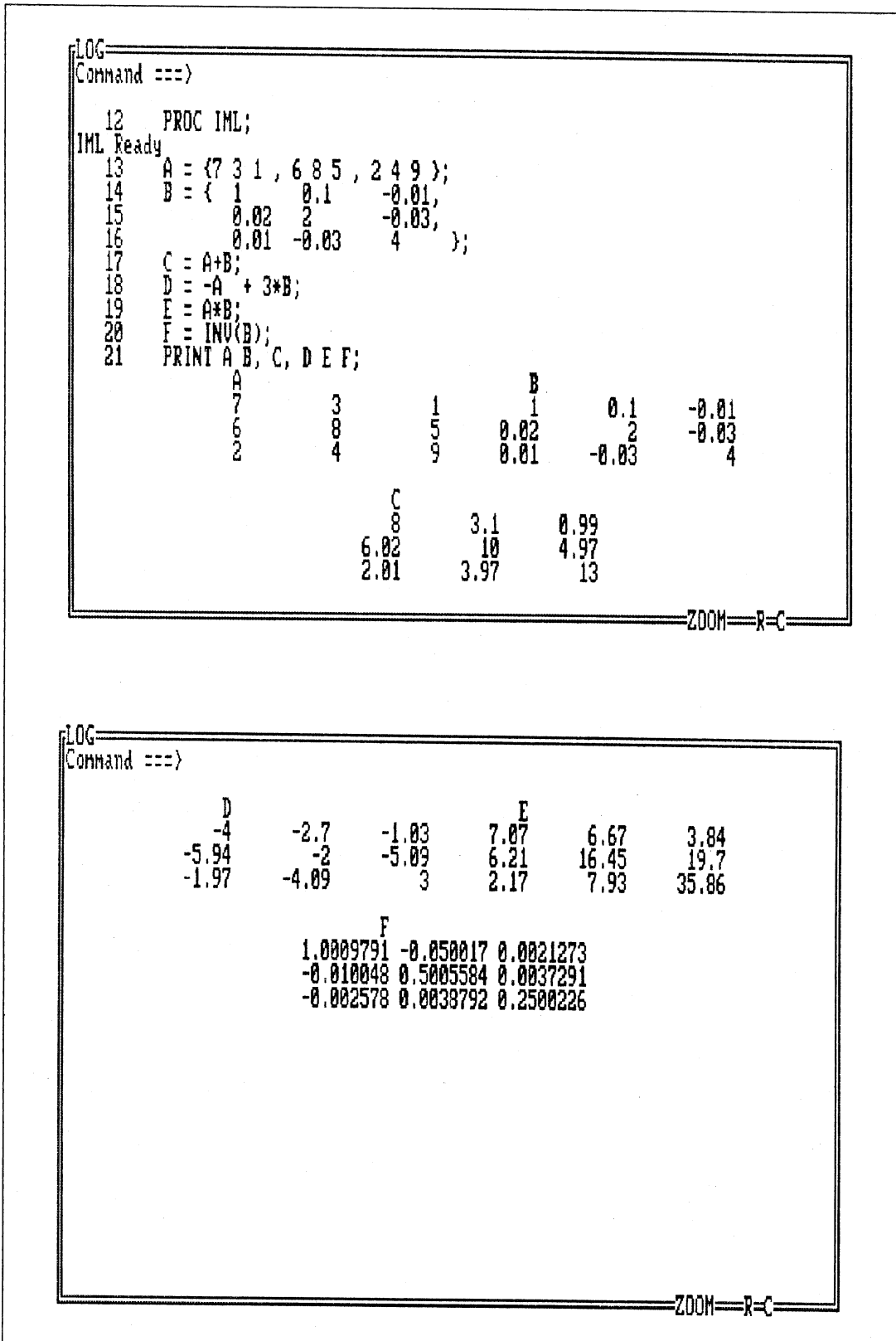
$$C = A + B, \quad D = -A + 3B, \quad E = AB, \quad F = B^{-1}$$



화면그림 3.4 SAS/IML환경에서 벗어남



화면그림 3.5 SAS/IML프로그램



화면그림 3.6 SAS/IML 프로그램을 수행한 결과

지금 SAS시스템은 SAS/IML환경에서 벗어나 있으므로 우선 SAS시스템을 SAS/IML 환경으로 만들어야 한다. 따라서 PGM윈도우에 PROC IML스태이트먼트를 입력시킨 뒤 기능키 F10을 눌러서 디스플레이의 화면을 화면그림 3.2처럼 만든다. 이 SAS/IML환경하의 PGM윈도우에 화면그림 3.5처럼 스타이트먼트들을 입력시킨다. 지금 독자들이 구태여 이 SAS/IML프로그램의 각 스타이트먼트를 이해하려고 노력할 필요는 없다.

화면그림 3.5의 SAS/IML프로그램을 실행하기 위해서 기능키 F10을 누르면 LOG윈도우가 화면그림 3.6처럼 된다. 여기서 유의할 점은 SAS/IML프로그램을 실행한 결과가 OUTPUT윈도우가 아닌 LOG윈도우에 나타난다는 것이다. 화면그림 3.6으로부터 행렬연산의 결과가 다음과 같다는 것을 알 수 있다.

$$C = \begin{bmatrix} 8 & 3.1 & 0.99 \\ 6.02 & 10 & 4.97 \\ 2.01 & 3.97 & 13 \end{bmatrix}$$

$$D = \begin{bmatrix} -4 & -2.7 & -1.03 \\ -5.94 & -2 & -5.09 \\ -1.97 & -4.09 & 3 \end{bmatrix}, \quad E = \begin{bmatrix} 7.07 & 6.67 & 3.84 \\ 6.21 & 16.45 & 19.7 \\ 2.17 & 7.93 & 35.86 \end{bmatrix}$$

$$F = \begin{bmatrix} 1.0009791 & -0.050017 & 0.0021273 \\ -0.010048 & 0.5005584 & 0.0037291 \\ -0.002578 & 0.0038792 & 0.2500226 \end{bmatrix}$$

PASCAL, FORTRAN, BASIC 등의 컴퓨터프로그래밍언어를 사용해 본 독자들은 화면그림 3.6의 결과에 비해서 화면그림 3.5의 SAS/IML프로그램이 아주 간략하다고 생각할 것이며, GAUSS, MATHEMATICA, MATLAB 등을 사용해 본 독자들은 SAS/IML이 컴퓨터프로그래밍언어들과 유사하다고 생각할 것이다. SAS/IML은 행렬연산을 하는데 있어서 다른 프로그래밍언어에 비해서 비교적 간편하고 유연하면서도 강력한 능력을 가진 프로그래밍언어이다.

3.2 SAS/IML의 기초

3.2.1 SAS/IML이란 무엇인가?

SAS/IML의 IML은 Interactive Matrix Language의 두문자어(acronym)로 SAS시스템에서 행렬연산을 담당하는 컴퓨터프로그래밍언어이다. SAS/IML은 PROC IML의 형태를 갖기 때문에 SAS의 한 프로시저라고 할 수도 있으나, SAS/IML은 하나의 프로시저라고 하기에는 지나치게 방대하므로 차라리 하나의 컴퓨터프로그래밍언어로 이해하는 것

이 좋을 듯하다. 오늘날 널리 사용되는 GAUSS, MATLAB, MATHEMATICA 등이 그러하듯이 SAS/IML도 행렬연산을 간편하게 처리할 수 있다. 그러나 지금 널리 사용되고 있는 버전 6.03, 또는 버전 6.04는 복소수를 취급하지 못하는 등 몇 가지 약점을 가지고 있다.

SAS/IML의 문자 I가 Interactive를 뜻한다고 했듯이 SAS/IML은 대화형시스템(interactive system)을 전제로 만들어졌으나, 본 저자의 경험으로는 시계열분석을 하는 경우라면 SAS/IML프로그램을 बै치(batch)로 처리하는 것이 더 효율적이라 생각한다. 따라서 이 장에서 SAS/IML의 기초적인 사용법을 설명할 때만 대화형시스템을 사용하고 이후 시계열분석의 이론을 설명하거나 실제 시계열데이터를 분석할 때는 बै치처리를 하기로 한다.

3.2.2 SAS/IML의 기초적 사용법

이 소절에서는 제 3.1 절의 예제를 가지고 SAS/IML의 기초적인 사용법을 다시 설명하고자 한다.

SAS/IML을 사용하기 위해서는 일단 SAS를 부팅시켜야 한다. 일단 SAS가 부팅되면 PGM윈도우에 다음과 같이 입력한다.

PROC IML;

이 PROC IML스테이트먼트가 입력된 컴퓨터의 디스플레이화면은 화면그림 3.1과 같다. 이 PROC IML스테이트먼트를 실행하기 위해서 기능키 F10을 누르면 디스플레이의 화면은 화면그림 3.2와 같아진다. 즉, LOG윈도우에 다음과 같은 문장이 출력된다.

IML Ready

따라서 독자의 디스플레이화면이 화면그림 3.2와 같다면 독자의 컴퓨터는 SAS/IML을 사용하는 환경, 즉 SAS/IML환경(SAS/IML environment)에 있는 것이다.

제 3.1 절의 예제를 다시 수행하기 위해서, 우선 SAS/IML환경에서 다음 행렬들 A와 B를 정의하기로 하자.

$$A = \begin{bmatrix} 7 & 3 & 1 \\ 6 & 8 & 5 \\ 2 & 4 & 9 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0.1 & -0.01 \\ 0.02 & 2 & -0.03 \\ 0.01 & -0.03 & 4 \end{bmatrix}$$

제 3.1 절에서는 이 행렬들을 SAS/IML환경에서 정의하기 위해 다음과 같은 스테이트먼트들을 사용하였다.

$$\begin{aligned} A &= \{ 7 \ 3 \ 1, \ 6 \ 8 \ 5, \ 2 \ 4 \ 9 \}; \\ B &= \{ 1 \quad 0.1 \quad -0.01, \end{aligned}$$

$$\begin{matrix} 0.02 & 2 & -0.03, \\ 0.01 & -0.03 & 4 \end{matrix} \};$$

대부분의 독자들이 짐작을 했겠지만 SAS/IML 환경에서 행렬을 정의할 때는 다음과 같은 규칙을 따른다.

- ① 정의할 행렬의 이름을 등호의 좌변에 놓는다. 행렬에 이름을 붙이는 방법은 SAS에서 변수의 이름을 붙이는 방법과 같다.
- ② 등호의 우변에는 행렬의 원소들을 중괄호 사이에 나열한다.
- ③ 행렬의 각 행의 원소들 사이에는 적어도 하나 이상의 빈 칸이 있어야 한다.
- ④ 행렬의 행(row)이 바뀔 때는 쉼표를 사용한다.
- ⑤ 다른 SAS스태이트먼트에서와 마찬가지로 한 행렬을 정의하는 데 PGM윈도우의 여러 행들을 사용해도 무방하나, 한 행렬의 정의를 끝내기 위해서는 반드시 마지막에 세미콜론을 사용해야 한다.
- ⑥ 다른 SAS스태이트먼트에서와 마찬가지로 PGM윈도우의 한 행에 여러 행렬들을 정의할 수 있다. 그러나 각 행렬의 정의를 끝낼 때마다 반드시 세미콜론을 사용해야 한다.

지금까지 나열한 규칙을 어기지만 않는다면 SAS/IML 환경에서 행렬을 정의할 수 있지만, SAS/IML 환경에서 편리하고 보편적인 방법으로 행렬을 정의하기 위해서는 행렬의 각 행을 PGM윈도우의 한 행으로 입력시킨다.

지금부터는 SAS/IML을 사용해서 다음 행렬들을 계산하기로 하자.

$$C = A + B, \quad D = -A + 3B, \quad E = AB, \quad F = B^{-1}$$

이러한 계산을 한다는 것은 새로운 행렬들 C, D, E, F 를 정의한다는 것을 의미한다. SAS/IML 환경에서 이 행렬들을 정의하기 위해서는 다음과 같은 스타이트먼트들을 수행한다.

```
C = A + B;
D = -A + 3*B;
E = A*B;
F = INV(B);
```

이 스타이트먼트들은 독자들이 선형대수학책에서 읽은 표기법과 거의 동일한 표기법으로 작성되었음을 알 수 있을 것이다. 이제 독자들은 SAS/IML이 행렬연산을 하는 데 편리하고 강력한 컴퓨터프로그래밍언어임을 느끼기 시작했을 것이다.

이 스타이트먼트들을 PGM윈도우에 입력시킨 다음 기능키 F10을 눌러보자. 유감스럽게도 OUTPUT윈도우나 LOG윈도우에 이 행렬들을 계산한 결과가 출력되지 않을 것이다. 이 스타이트먼트들을 실행하면 행렬들 C, D, E, F 가 계산되지만 아무런 출력명령이 없기 때문에 OUTPUT윈도우나 LOG윈도우에는 이 행렬들이 출력되지 않는다. 이 행렬들을 출

력시키기 위해서 다음과 같은 PRINT스테이트먼트를 PGM윈도우에 입력시킨다.

```
PRINT A B, C, D E F;
```

이제 다시 기능키 F10을 눌러서 PRINT스테이트먼트를 실행하면 LOG윈도우가 화면그림 3.6처럼 될 것이다. 즉, 행렬들 A, B, C, D, E, F 가 LOG윈도우에 출력될 것이다. 화면그림 3.6을 보면 행렬들이 PRINT스테이트먼트에 지정된 순서대로 출력되었다. PRINT스테이트먼트에 사용된 쉼표는 행을 바꾸도록 하는 역할을 한다. 화면그림 3.6에서 볼 수 있듯이 행렬들 A 와 B 를 한 행에 출력한 다음, 쉼표에 의해서 행을 바꾸어 행렬 C 를 출력하고, 다시 쉼표에 의해서 행을 바꾸어 행렬들 D 와 E 를 출력하고, 다시 행을 바꾸어 행렬 F 를 출력한다. 원래대로라면 행렬들 D, E, F 는 같은 행에 출력되어야 하나 LOG윈도우의 폭이 좁아서 나누어 출력된 것이다.

3.2.3 SAS/IML의 बै치처리

SAS/IML의 बै치처리란 PROC IML스테이트먼트와 QUIT스테이트먼트 사이의 스테이트먼트들을 모두 입력시킨 다음에 이 스테이트먼트들을 한꺼번에 실행시키는 방법이다.

SAS/IML프로그램의 बै치처리의 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-1.PRG에 수록되어 있다).

```

/*                                                    */
/*                EXAM3-1.PRG                        */
/*                                                    */
/*                BATCH PROCESSING OF A SAS/IML PROGRAM    */
/*                                                    */
OPTIONS PAGESIZE = 60;
PROC IML;
A = {1 2 , 3 4};
B = {56 78,
     90 11};
C = A-4.0*B;
D = (A+C)*B;
E = INV(2*A + 3.1*B*C);
PRINT A B, C, D E ;
QUIT;

```

이 SAS/IML프로그램을 수행하면 LOG윈도우에 그림 3.1과 같은 결과가 나타난다. 또한 이 SAS/IML프로그램의 수행이 끝나면 SAS시스템은 SAS/IML환경에서 벗어난다.

```

NOTE: Copyright(c) 1985, 86, 87 SAS Institute Inc., Cary, NC 27512-8000, U.S.A.
NOTE: SAS (r) Proprietary Software Release 6.04
      Licensed to YONSEI UNIVERSITY, Site 24309001.

NOTE: AUTOEXEC processing completed.

1  /*
2  /*          EXAM3-1.PRG
3  /*
4  /*          BATCH PROCESSING OF A SAS/IML PROGRAM
5  /*
6  OPTIONS PAGESIZE=60;
7  PROC IML;
IML Ready
8  A = {1 2 , 3 4};
9  B = {56 78,
10     90 11};
11 C = A-4.0 * B;
12 D = (A + C) * B;
13 E = INV(2 * A + 3.1 * B * C);
14 PRINT A B, C, D E;

           A          B
           1          2          56          78
           3          4          90          11

           C
          -223          -310
          -357          -40

           D          E
        -40152 -20704 -0.000014 0.0000101
        -23064 -28008 0.0000119 -0.00002

15  QUIT;

Exiting IML.

NOTE: The PROCEDURE IML used 3.00 seconds.

```

그림 3.1 SAS/IML 프로그램을 बै치 처리한 결과

3.2.4 RESET 스테이트먼트

지금까지는 SAS/IML 프로그램을 수행한 결과가 LOG 윈도우에 출력되었다. 이는 SAS/IML 환경에서는 SAS/IML 프로그램을 수행한 결과가 LOG 윈도우에 출력되는 것이 디폴트이기 때문이다. 또한, LOG 윈도우에 출력된 행렬들이 LOG 윈도우의 좌측이 아닌 가운데에 정렬되어 있음을 알 수 있다. 이는 SAS/IML 환경에서 PRINT 스테이트먼트를 수행한 결과가 LOG 윈도우의 가운데를 기준으로 출력되는 것이 디폴트이기 때문이다. SAS 시스템에서는 SAS/IML 프로그램이 아닌 다른 SAS 프로그램을 수행한 결과가 OUTPUT 윈도우에 출력되는 것이 일반적이다. 반면에 지금까지 설명했듯이 SAS/IML 프로그램을 수행한 결과는 디폴트로 LOG 윈도우에 출력된다. 따라서 PROC IML과 다른 프로시저들을 포함하는 SAS 프로그램을 수행하면, PROC IML의 결과는 LOG 윈도우에 나타나고 다른 프로시저들의 결과는 OUTPUT 윈도우에 나타나는 기현상이 일어난다. 따라서 PROC IML을 수행한 결과도 OUTPUT 윈도우에 출력시키는 것이 좋을 것이다. SAS/IML을 수행한 결

과를 OUTPUT윈도우에 출력시키기 위해서는 다음과 같이 RESET스테이트먼트의 옵션 NOLOG를 사용한다.

RESET NOLOG;

이 RESET스테이트먼트가 실행되면 이후 수행되는 모든 SAS/IML의 스테이트먼트들의 수행결과와 이에 부가적으로 발생하는 에러메시지들(error messages)이 OUTPUT윈도우에 출력된다. 옵션 NOLOG에 반대되는 옵션은 LOG이다. 만일, 다음과 같이 RESET스테이트먼트에 옵션 LOG를 사용하면 이후 출력되는 행렬들이 LOG윈도우에 출력된다.

RESET LOG;

RESET스테이트먼트의 옵션 NOLOG를 사용하는 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-2.PRG에 수록되어 있다).

```

/ *                               */
/ *           EXAM3-2.PRG         */
/ *                               */
/ *           RESET OPTION NOLOG  */
/ *                               */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOLOG;
A = { 1 2 , 3 4 };
B = { 56 78,
      90 11 };
C = A-4.0*B;
D = (A+C)*B;
E = INV(2*A + 3.1*B*C);
PRINT A B, C, D E ;
QUIT;

```

이 SAS/IML프로그램을 수행하면 OUTPUT윈도우에 그림 3.2와 같은 결과가 나타난다. 그림 3.2가 그림 3.1과 다른 점은 그림 3.1에는 SAS/IML프로그램 부분이 들어가 있다는 것이다. 즉, 그림 3.1에는 보통 SAS프로그램을 수행했을 때 OUTPUT윈도우에 나타나는 부분과 LOG윈도우에 나타나는 부분 모두가 나타난 반면에 그림 3.2에는 보통 SAS프로그램을 수행했을 때 OUTPUT윈도우에 나타나는 부분만이 나타났다.

그림 3.1과 그림 3.2에서 알 수 있듯이 SAS/IML환경에서 PRINT스테이트먼트를 수행한 결과로 출력된 행렬들은 윈도우의 가운데를 기준으로 정렬되어 있다. 만일, 이 행렬들

```

SAS 20:20 Tuesday, February 15, 1994 1

      A          B
      1          2          56          78
      3          4          90          11

      C
      -223        -310
      -357        -40

      D          E
-40152 -20704 -0.000014 0.0000101
-23064 -28008 0.0000119 -0.00002

Exiting IML.

```

그림 3.2 RESET스테이트먼트의 옵션 NOLOG를 사용한 결과물

을 윈도우의 좌측을 기준으로 정렬시키기 위해서는 다음과 같이 RESET스테이트먼트에 옵션 NOCENTER를 사용한다.

RESET NOCENTER;

옵션 NOCENTER에 반대되는 옵션은 CENTER이다. 만일, 다음과 같이 RESET스테이트먼트에 옵션 CENTER를 사용하면 이후 출력되는 행렬들은 윈도우의 가운데를 기준으로 정렬된다.

RESET CENTER;

RESET스테이트먼트의 옵션 NOCENTER를 사용하는 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-3.PRG에 수록되어 있다).

```

/* */
/*          EXAM3-3.PRG          */
/* */
/*          RESET OPTION NOCENTER          */
/* */
OPTIONS PAGESIZE = 60;
PROC IML;
RESET NOCENTER;
A = {1 2, 3 4};
B = {56 78,
     90 11};
C = A-4.0*B;
D = (A+C)*B;
E = INV(2*A + 3.1*B*C);

```



```

NOTE: Copyright(c) 1985, 86, 87 SAS Institute Inc., Cary, NC 27512-8000, U.S.A.
NOTE: SAS (r) Proprietary Software Release 6.04
      Licensed to YONSEI UNIVERSITY, Site 24309001.

NOTE: AUTOEXEC processing completed.

1  /*                                     */
2  /*           EXAM3-3.PRG              */
3  /*                                     */
4  /*           RESET OPTION NOCENTER   */
5  /*                                     */
6  OPTIONS PAGESIZE=60;
7  PROC IML;
IML Ready
8  RESET NOCENTER;
9  A = { 1 2 , 3 4 };
10 B = { 56 78,
11       90 11 };
12 C = A-4.0 * B;
13 D = (A + C) * B;
14 E = INV(2 * A + 3.1 * B * C);
15 PRINT A B, C, D E;
      A           B
      1           2           56           78
      3           4           90           11

      C
      -223         -310
      -357         -40

      D           E
      -40152  -20704  -0.000014  0.0000101
      -23064  -28008  0.0000119  -0.00002

16  QUIT;
Exiting IML.
NOTE: The PROCEDURE IML used 3.00 seconds.

```

그림 3.3 RESET스테이트먼트의 옵션 NOCENTER를 사용한 결과물

```

PRINT A B, C, D E;
QUIT;

```

이 SAS/IML 프로그램을 수행하면 LOG 윈도우에 그림 3.3과 같은 결과가 나타난다.

한 RESET 스테이트먼트에는 여러 옵션을 동시에 사용할 수 있다. 한 예로 SAS/IML 프로그램을 수행한 결과를 OUTPUT 윈도우에 출력시키고, 또한 OUTPUT 윈도우에 출력된 행렬들을 OUTPUT 윈도우의 좌측을 기준으로 정렬시키기 위해서는 다음 RESET 스테이트먼트를 사용한다.

```
RESET NOLOG NOCENTER;
```

지금까지 수행된 행렬연산의 결과는 SAS 시스템에 저장되어 있다가 PRINT 스테이트먼트를 사용해야 지정된 윈도우에 출력되었다. 일일이 PRINT 스테이트먼트를 사용하지 않고 자동적으로 행렬들을 지정된 윈도우에 출력시키기 위해서는 다음과 같이 RESET 스테이트먼트에 옵션 PRINT를 사용한다.

```
RESET PRINT;
```

이 RESET스태이트먼트가 수행되면 SAS/IML환경에서 정의되는 모든 행렬들이 자동적으로 지정된 윈도우로 출력된다. 옵션 PRINT를 작동하면 출력되는 행렬의 이름, 이 행렬의 행들의 개수와 열들의 개수가 행렬이 지정된 윈도우의 좌측을 기준으로 출력된다. 다음과 같은 RESET스태이트먼트를 실행하자.

RESET PRINT CENTER;

이 RESET스태이트먼트에는 옵션 PRINT와 옵션 CENTER를 동시에 사용하였다. 유의할 점은 이렇게 RESET스태이트먼트에 옵션 PRINT와 옵션 CENTER를 동시에 사용해도 옵션 PRINT에 의해 출력되는 행렬의 이름, 이 행렬의 행들의 개수와 열들의 개수는 행렬이 지정된 윈도우의 가운데를 기준으로 하지 않고 좌측을 기준으로 출력된다는 것이다. 옵션 PRINT에 반대되는 옵션은 NOPRINT이다. 다음과 같이 RESET스태이트먼트에 옵션 NOPRINT를 사용하면 이후 PRINT스태이트먼트에 지정하는 행렬들만이 지정된 윈도우에 출력된다.

RESET NOPRINT;

RESET스태이트먼트의 옵션 PRINT를 사용하는 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-4.PRG에 수록되어 있다).

```

/* */
/*          EXAM3-4.PRG          */
/* */
/*          RESET OPTION PRINT & NOLOG          */
/* */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOLOG PRINT;
A = { 1 2 , 3 4 };
B = { 56 78,
      90 11 };
C = A-4.0*B;
D = (A+C)*B;
E = INV(2*A + 3.1*B*C);
PRINT A B, C, D E;
QUIT;

```

이 SAS/IML프로그램을 수행하면 OUTPUT윈도우에 그림 3.4와 같은 결과가 나타난다. 만일, 출력되는 행렬뿐만 아니라 이 행렬이 계산되는 중간단계의 행렬들까지 지정된 윈

```

SAS 20:20 Tuesday, February 15, 1994 2

A      2 rows 2 columns
      1      2
      3      4

B      2 rows 2 columns
      56     78
      90     11

C      2 rows 2 columns
      -223   -310
      -357   -40

D      2 rows 2 columns
      -40152 -20704
      -23064 -28008

E      2 rows 2 columns
      -0.000014 0.0000101
      0.0000119 -0.00002

              A      B
              1      2      56      78
              3      4      90      11

              C
              -223   -310
              -357   -40

              D      E
      -40152 -20704 -0.000014 0.0000101
      -23064 -28008 0.0000119 -0.00002

Exiting IML.

```

그림 3.4 RESET스테이트먼트의 옵션 PRINT를 사용한 결과물

도구에 출력시키기 위해서는 다음과 같이 RESET스테이트먼트에 옵션 ALL을 사용한다.

RESET ALL;

옵션 ALL에 반대되는 옵션은 NOALL이다. 만일, 다음과 같이 RESET스테이트먼트에 옵션 NOALL을 사용하면 최종적으로 계산된 행렬들만이 지정된 윈도우에 출력된다.

RESET NOALL;

RESET스테이트먼트의 옵션 ALL을 사용하는 예제로 다음 SAS/IML 프로그램을 수행해 보자(이 SAS/IML 프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-5.PRG에 수록되어 있다).

```

/* */
/*          EXAM3-5.PRG          */
/* */
/*          RESET OPTION ALL & NOLOG          */
/* */
OPTIONS PAGESIZE=60;
PROC IML;

```

RESET NOLOG ALL;

A = { 1 2 , 3 4 };

B = { 56 78,
90 11 };

C = A-4.0*B;

```

SAS 20:20 Tuesday, February 15, 1994 3

A          2 rows 2 columns
  1         2
  3         4

B          2 rows 2 columns
 56        78
 90        11

# TEM1001  2 rows 2 columns
 224       312
 360       44

C          2 rows 2 columns
 -223      -310
 -357      -40

# TEM1001  2 rows 2 columns
 -222      -308
 -354      -36

D          2 rows 2 columns
-40152     -20704
-23064     -28008

# TEM1001  2 rows 2 columns
 2         4
 6         8

# TEM1002  2 rows 2 columns
 173.6     241.8
 279       34.1

# TEM1003  2 rows 2 columns
-125035.4  -63488
-74390.7   -87854

# TEM1004  2 rows 2 columns
-125033.4  -63484
-74384.7   -87846

E          2 rows 2 columns
-0.000014  0.0000101
0.0000119  -0.00002

          A          B
          1          2          56          78
          3          4          90          11

          C
          -223      -310
          -357      -40

          D          E
-40152 -20704 -0.000014 0.0000101
-23064 -28008 0.0000119 -0.00002

Exiting IML.
```

그림 3.5 RESET스테이트먼트의 옵션 ALL을 사용한 결과물

```

D = (A+C)*B;
E = INV(2*A + 3.1*B*C);
PRINT A B, C, D E;
QUIT;

```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.5와 같은 결과가 나타난다. 그림 3.5에서 알 수 있듯이 SAS/IML 환경에서 출력되는 행렬의 위에는 이 행렬의 이름이 출력된다. 때때로 이 행렬 이름의 출력을 억제시킬 필요가 있는데, 이러한 목적을 위해서는 다음과 같이 RESET 스테이트먼트에 옵션 NONAME을 사용한다.

```
RESET NONAME;
```

옵션 NONAME에 반대되는 옵션은 NAME이다. 다음과 같이 RESET 스테이트먼트에 옵션 NAME을 사용하면 이후 출력되는 행렬의 위에는 행렬 이름이 출력된다.

```
RESET NAME;
```

RESET 스테이트먼트의 옵션 NONAME을 사용하는 예제로 다음 SAS/IML 프로그램을 수행해 보자(이 SAS/IML 프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-6.PRG에 수록되어 있다).

```

/* */
/*          EXAM3-6.PRG          */
/* */
/*          RESET OPTION NONAME & NOLOG          */
/* */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOLOG NONAME;
A = { 1 2 , 3 4 };
B = { 56 78,
      90 11 };
C = A-4.0*B;
D = (A+C)*B;
E = INV(2*A + 3.1*B*C);
PRINT A B, C, D E;
QUIT;

```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.6과 같은 결과가 나타난다. 앞에서 언급된 옵션들 이외에도 RESET 스테이트먼트에는 다음과 같은 옵션들을 사용할 수가 있다. 기호 |는 둘 중에서 하나를 선택한다는 것을 뜻하며, 앞의 것이 SAS/IML

```

SAS 20:20 Tuesday, February 15, 1994 4

      1      2      56      78
      3      4      90      11

      -223     -310
      -357     -40

-40152 -20704 -0.000014 0.0000101
-23064 -28008 0.0000119 -0.00002

Exiting IML.
    
```

그림 3.6 RESET스테이트먼트의 옵션 NONAME을 사용한 결과물

환경에서의 디폴트이다.

(1) FW = 숫자

옵션 FW=에는 SAS/IML환경에서 결과를 출력할 때 한 숫자가 차지하는 칸의 수 (field width for printing numeric values)를 지정한다. 즉, 행렬의 각 원소가 차지하는 칸의 수를 지정한다. 이 옵션의 디폴트는 9이다.

(2) NODetails|Details

옵션 Details를 사용하면 파일들을 열 때, 혹은 닫을 때(when files are opened or closed)와 같은 추가적인 정보를 출력한다. 반대로 옵션 NODetails를 사용하면 이런 정보가 출력되지 않는다.

(3) NOFlow|Flow

옵션 Flow를 사용하면 실행되는 연산의 순서(operations as executed)가 출력된다. 옵션 Flow는 SAS/IML프로그램에 잘못이 있을 때 이 잘못된 부분을 찾는 목적에 사용된다. 옵션 NOFlow를 사용하면 연산의 순서가 출력되지 않는다.

(4) Linesize = 숫자

옵션 Linesize=에는 SAS/IML의 결과물의 한 행에 쓸 수 있는 문자의 수를 지정한다. 이 옵션의 디폴트는 78이다.

(5) Space = 숫자

옵션 Space=에는 PRINT스테이트먼트에 의해서 출력되는 결과물에서 숫자들 사이의 빈 칸의 수(the number of spaces printed results for the PRINT statement)를 지정한다. 즉, 각 행렬의 원소들 사이의 빈 칸의 수를 지정한다. 이 옵션의 디폴트는 1이다.

(6) Spill

옵션 Spill을 사용하면 SAS/IML에서 메모리가 모자랄 때 사용하지 않는 행렬들을 보조기억장치인 디스크에 저장한다. 따라서 이 행렬들이 차지하던 메모리를 당장 필요한 행렬연산에 사용할 수 있다. 그러나 이 옵션을 사용하면 SAS/IML의 수행속도가 떨어진다.

(7) BREAK

옵션 BREAK를 사용하면 합성키 Ctrl-Break를 눌러서 SAS/IML프로그램의 실행을 중단시킬 수 있다. 그러나 SAS/IML에서 이 옵션이 작동되는 것이 디폴트이므로 구태여 이 옵션을 사용할 필요는 없다. 이 옵션을 사용하면 이 옵션이 작동되는 것에 대한 사전경고를 할 따름이다(will just give a warning to that effect).

RESET스테이트먼트에는 지금까지 설명한 옵션들 이외에도 여러 옵션들을 사용할 수 있다. 이 옵션들에 대한 자세한 내용은 [IML6.1](pp. 296~297)을 참조하라.

현재 사용되고 있는 RESET스테이트먼트의 옵션들을 알고 싶으면 다음과 같은 SHOW스테이트먼트를 실행한다.

SHOW OPTIONS;

이 SHOW스테이트먼트를 사용하는 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-7.PRG에 수록되어 있다).

```

/ *                               */
/ *           EXAM3-7.PRG         */
/ *                               */
/ *           SHOW RESET OPTIONS  */
/ *                               */

OPTIONS PAGESIZE=60;
PROC IML;
RESET NOLOG NOPRINT NONAME;
SHOW OPTIONS;
QUIT;

```

이 SAS/IML프로그램을 수행하면 OUTPUT윈도우에 그림 3.7과 같은 결과가 나타난다. SHOW스테이트먼트에 관한 자세한 내용은 제 3.6.1소절을 참조하라.

```

SAS 20:20 Tuesday, February 15, 1994 5

Options: noprint nodetails noflow nodebug nolog nocase center
         noall noname break nospill linesize=78 pagesize=60
         fw=9 spaces=1 storage=SASUSER. IMLSTORE deflib=

Exiting IML.

```

그림 3.7 SHOW스테이트먼트를 실행한 결과물

3.3 행렬의 연산

3.3.1 할당스테이트먼트

제 3.2.2 소절에서 SAS/IML 환경에서 행렬을 정의하는 양식을 설명하였으므로 여기서는 필요한 내용을 보충하고자 한다. 다음은 SAS/IML 환경에서 행렬을 정의하는 행렬정의 스테이트먼트들의 예들이다.

```

ABC = { 32.4 };
D = { 3 45, -23 };
E1234567 = { "CHOI" "KIM" , "PARK" "LEE" };
FFF = -1234.5;
GHIJ = .;
K9 = 'CHOP';
LZXC = "PARK'S";

```

행렬정의 스테이트먼트의 좌변에서 행렬의 이름을 정하는데, 이 예제들에서도 알 수 있듯이 행렬의 이름을 정하는 규칙은 SAS에서 변수의 이름을 정하는 규칙과 동일하다. 즉, 행렬의 이름은 영문자로 시작하는 여덟 자 이내의 영숫자(alphanumeric)로 정한다. 행렬 정의 스테이트먼트의 우변에서 중괄호를 사용하면 이는 행렬을 정의하는 것이고 중괄호를 사용하지 않으면 행렬의 특수한 경우인 스칼라를 정의하는 것이다. 즉, 첫 3개의 예제들은 행렬을 정의하는 것이고, 나머지 예제들은 스칼라를 정의하는 것이다. 행렬의 원소들과 스칼라로는 숫자 또는 문자열(character string)을 사용할 수 있고, 만일 스칼라 또는 원소가 문자열이면 작은따옴표(')나 큰따옴표(") 사이에 이 문자열을 놓아야 한다. 여기서 유의할 점은 원소로 사용되는 문자열에 작은따옴표(apostrophe)가 포함되어 있는 경우에는 이 문자열을 큰따옴표 사이에 놓아야 혼동이 없을 것이다. 즉, 일곱 번째 예제에서 행렬 LZXC는 PARK'S라는 원소를 갖는 1×1 행렬이다. 특히 유의할 점은 행렬의 원소로 숫자 또는 문자열을 사용할 수 있으나, 한 행렬의 원소들로 숫자와 문자열을 동시에 사용할 수는 없다. 즉, 한 행렬의 모든 원소들은 숫자로 통일되거나 또는 문자열로 통일되어야 한다. 원소들이 숫자인 행렬을 숫자행렬, 원소들이 문자열인 행렬을 문자행렬이라 한다.

FORTRAN, BASIC, 그리고 SAS 등 일반적인 컴퓨터 프로그래밍 언어에서 등호는 이 등호의 우변의 식을 계산한 결과로 얻어진 값(또는 행렬)을 좌변에 지정된 이름으로 저장시키라는 할당스테이트먼트(assignment statement)이다. SAS/IML 환경에서 할당스테이트먼트의 사용양식은 다음과 같다.

행렬 = 행렬연산식;

독자들은 할당스테이트먼트에서 등호 대신에 할당기호 \rightarrow 를 사용하는 것이 더 타당하다고 생각할 것이다. 그러나 관습상 등호가 사용되고 있다. 엄격히 말해서 앞에서 배운 행렬 정의스테이트먼트도 우변의 행렬을 좌변의 행렬이름으로 할당하는 할당스테이트먼트이다.

SAS/IML환경에서 할당스테이트먼트의 우변에는 연산자들(operators)을 사용한 행렬 연산식, SAS/IML함수, 그리고 함수모듈을 사용할 수 있다. 지금부터는 할당스테이트먼트의 우변에 사용하는 행렬연산자들을 다루고자 한다. 행렬연산자들은 행렬단위로 계산하는 행렬기준연산자들, 원소기준으로 계산하는 원소기준연산자들, 그리고 부행렬을 만드는 첨자연산자들로 나눌 수 있다. 제 3.3.2소절에서는 행렬기준연산자들을, 제 3.3.3소절에서는 원소기준연산자들을, 그리고 제 3.3.4소절에서는 첨자연산자들을 다루기로 한다. 이 절에서 다루는 연산자들에 대한 자세한 내용은 [IML6.03]의 (pp. 211~222)을 참조하라.

3.3.2 행렬기준연산자들

독자들이 알다시피 벡터나 스칼라도 행렬의 특수한 경우들로 생각할 수 있다. 그러나 SAS/IML환경에서의 연산을 설명하려면 행렬, 벡터, 그리고 스칼라를 구분해야 할 경우가 있다. 이러한 경우에는 행렬들을 M_1, M_2, M_3 , 벡터들을 V_1, V_2, V_3 , 그리고 스칼라들을 S_1, S_2, S_3 로 표기하기로 하자.

1) 더하기: +

더하기연산자 +의 사용양식은 다음과 같다.

$$M_1 + M_2$$

$$M_1 + S_1$$

첫 번째 사용양식은 선형대수에서 행렬 M_1 에 행렬 M_2 를 더하는 것이다. 두 번째 사용양식은 선형대수에서는 성립되지 않는 식이나 SAS/IML에서는 행렬 M_1 의 각 원소에 스칼라 S_1 을 더하는 것으로 정의한다.

2) 부호바꾸기: -

부호바꾸기연산자 -의 사용양식은 다음과 같다.

$$-M_1$$

이 부호바꾸기연산을 실행하면 행렬 M_1 의 각 원소에 (-1) 을 곱한 원소로 구성된 행렬

$-M_1$ 으로 바뀐다.

3) 빼기: -

빼기연산자 -의 사용양식은 다음과 같다.

$$\begin{aligned} M_1 - M_2 \\ M_1 - S_1 \end{aligned}$$

첫 번째 사용양식은 선형대수에서 행렬 M_1 에서 행렬 M_2 를 빼는 것이다. 두 번째 사용양식은 선형대수에서는 성립되지 않는 식이나 SAS/IML에서는 행렬 M_1 의 각 원소에서 스칼라 S_1 을 빼는 것으로 정의한다.

4) 행렬곱하기: *

행렬곱하기연산자 *의 사용양식은 다음과 같다.

$$M_1 * M_2$$

행렬곱하기연산을 실행하면 선형대수에서 정의되는 두 행렬들의 곱 $M_1 M_2$ 가 계산된다.

5) 행렬먹: **

행렬먹연산자 **의 사용양식은 다음과 같다.

$$M_1 ** S_1$$

이 행렬먹연산을 수행하면 행렬 M_1 에 스칼라 S_1 만큼의 먹을 취한다. 여기서 행렬 M_1 은 정방행렬이고 스칼라 S_1 은 -1보다 크거나 같은 정수이다. 만일, 스칼라 S_1 이 실수이면 정수부분만이 사용된다. 스칼라 S_1 이 -1이면 행렬먹연산을 수행한 결과는 M_1^{-1} 이다.

6) 전치: `

전치연산자 `의 사용양식은 다음과 같다.

$$M_1`$$

이 전치연산(transpose operator)을 실행하면 행렬 M_1 을 전치시킨다. 전치연산자 `는 뒤따옴표(backquote character)라 불리우는데, 어떤 키보드에는 이 뒤따옴표가 없다. 이 경우에는 행렬 M_1 의 전치행렬로 $t(M_1)$ 을 사용한다.

7) Kronecker 곱: @

Kronecker 곱연산자 @의 사용양식은 다음과 같다.

$$M_1 @ M_2$$

Kronecker 곱연산자 @는 선형대수에서 정의한 Kronecker 곱(Kronecker product, 또는 direct product)을 수행하는 연산자이다.

지금까지 설명한 행렬기준연산자들에 대한 예제로 다음 SAS/IML 프로그램을 수행해 보자(이 SAS/IML 프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-8.PRG에 수록되어 있다).

```

/* */
/*          EXAM3-8.PRG          */
/* */
/*          OPERATORS +, -, *, **, `,@          */
/* */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOCENTER NOLOG NOPRINT FW=6;
M1 = { 1 2 3,
      4 5 6,
      7 8 0};
M2 = { 0 8 7,
      6 5 4,
      3 2 1};
PRINT M1 " " M2;
S1 = 2;
S2 = -1;
PRINT S1 " " S2;
M3 = M1 + M2;
M4 = M1 + S1;
PRINT M3 " " M4;
M5 = -M1;
PRINT M5;
M6 = M1 - M2;
M7 = M1 - S1;
PRINT M6 " " M7;
M8 = M1 * M2;
PRINT M8;
M9 = M1 ** S1;

```

```

SAS
12:18 Saturday, March 5, 1994 1

M1
1 2 3
4 5 6
7 8 0

M2
0 8 7
6 5 4
3 2 1

S1
2
S2
-1

M3
1 10 10
10 10 10
10 10 1

M4
3 4 5
6 7 8
9 10 2

M5
-1 -2 -3
-4 -5 -6
-7 -8 0

M6
1 -6 -4
-2 0 2
4 6 -1

M7
-1 0 1
2 3 4
5 6 -2

M8
21 24 18
48 69 54
48 96 81

M9
30 36 15
66 81 42
39 54 69

M10
-1.7778 0.88889 -0.1111
1.55556 -0.7778 0.22222
-0.1111 0.22222 -0.1111

M11
1 555E-19 139E-19
-61E-17 1 -83E-18
0 0 0

M12
1 4 7
2 5 8
3 6 0

M13
1 555E-19 139E-19 2 111E-18 278E-19 3 167E-18 416E-19
-61E-17 1 -83E-18 -12E-16 2 -17E-17 -18E-16 3 -25E-17
0 0 1 0 0 2 0 0 3
4 222E-18 555E-19 5 278E-18 694E-19 6 333E-18 833E-19
-24E-16 4 -33E-17 -31E-16 5 -42E-17 -37E-16 6 -5E-16
0 0 4 0 0 5 0 0 6
7 389E-18 971E-19 8 444E-18 111E-18 0 0 0
-43E-16 7 -58E-17 -49E-16 8 -67E-17 0 0 0
0 0 7 0 0 8 0 0 0

M14
1 2 3 555E-19 111E-18 167E-18 139E-19 278E-19 416E-19
4 5 6 222E-18 278E-18 333E-18 555E-19 694E-19 833E-19
7 8 0 389E-18 444E-18 0 971E-19 111E-18 0
-61E-17 -12E-16 -18E-16 1 2 3 -83E-18 -17E-17 -25E-17
-24E-16 -31E-16 -37E-16 4 5 6 -33E-17 -42E-17 -5E-16
-43E-16 -49E-16 0 7 8 0 -58E-17 -67E-17 0
0 0 0 0 0 0 1 2 3
0 0 0 0 0 0 4 5 6
0 0 0 0 0 0 7 8 0

Exiting IML.

```

그림 3.8 행렬기준연산자들을 실행한 결과

```

M10 = M1 ** S2;
M11 = M1 * M10;
      PRINT M9, M10 " " M11,;
M12 = M1 `;
      PRINT M12,;
M13 = M1 @ M11;
M14 = M11 @ M1;
      PRINT M13 " " M14,;
QUIT;

```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.8과 같은 결과가 나타난다.

8) 가로결합 : ||

가로결합연산자 ||의 사용양식은 다음과 같다.

$$M_1 \parallel M_2 \parallel M_3$$

가로결합연산자(horizontal concatenation operator)는 행의 수가 같은 두 개 이상의 행렬들을 가로로 결합시키는 연산자이다. 이 가로결합연산의 결과로 다음 행렬이 만들어진다.

$$[M_1 \ M_2 \ M_3]$$

9) 세로결합 : //

세로결합연산자 //의 사용양식은 다음과 같다.

$$M_1 // M_2 // M_3$$

세로결합연산자(vertical concatenation operator)는 열의 수가 같은 두 개 이상의 행렬들을 세로로 결합시키는 연산자이다. 이 세로결합연산의 결과로 다음 행렬이 만들어진다.

$$\begin{bmatrix} M_1 \\ M_2 \\ M_3 \end{bmatrix}$$

결합연산자들에 대한 예제로 다음 SAS/IML 프로그램을 수행해 보자(이 SAS/IML 프로그램은 디스크 SAS4TSA4의 파일 EXAM3-9.PRG에 수록되어 있다).

```

/ *                               */
/ *                               EXAM3-9.PRG */
/ *                               */

```

```

/*          CONCATENATION OPERATORS          */
/*                                          */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOCENTER NOLOG FW=6 NOPRINT;
M1 = { 1 2 3, 4 5 6, 7 8 9 };
M2 = { 11 22 33, 44 55 66, 77 88 99 };
M3 = { -1 -2 -3, -4 -5 -6, -7 -8 -9 };
      PRINT M1 " " M2 " " M3;
M4 = M1 || M2;
M5 = M1 || M2 || M3;
PRINT M4 " " M5;
M6 = M1 // M2;
M7 = M1 // M2 // M3;
PRINT M6 " " M7;
QUIT;

```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.9와 같은 결과가 나타난다.

SAS			7:46 Saturday, March 5, 1994 12					
M1			M2					
1	2	3	11	22	33			
4	5	6	44	55	66			
7	8	9	77	88	99			
M3								
-1	-2	-3						
-4	-5	-6						
-7	-8	-9						
M4								
1	2	3	11	22	33			
4	5	6	44	55	66			
7	8	9	77	88	99			
M5								
1	2	3	11	22	33	-1	-2	-3
4	5	6	44	55	66	-4	-5	-6
7	8	9	77	88	99	-7	-8	-9
M6			M7					
1	2	3	1	2	3			
4	5	6	4	5	6			
7	8	9	7	8	9			
11	22	33	11	22	33			
44	55	66	44	55	66			
77	88	99	77	88	99			
			-1	-2	-3			
			-4	-5	-6			
			-7	-8	-9			
Exiting IML.								

그림 3.9 결합연산자들을 실행한 결과

3.3.3 원소기준연산자들

원소기준연산자들(elementwise operators)은 선형대수학에서 일반적으로 사용되지 않는 연산자들이다. 예를 들어, 두 $m \times n$ 행렬들에 원소곱하기연산을 하면 $m \times n$ 행렬이 발생하고 이 발생된 행렬의 (i, j) 원소는 두 피연산행렬들의 (i, j) 원소들의 곱이다. 이 원소곱하기연산자처럼 대응하는 원소들만을 사용하는 연산자를 원소기준연산자라고 한다.

SAS/IML에서 원소기준연산자는 원칙적으로 차원이 같은 두 행렬들에 적용된다. 반면에 행렬과 벡터, 또는 행렬과 스칼라 사이의 원소기준연산도 가능한데 이 경우에는 먼저 주어진 벡터와 스칼라를 주어진 행렬과 같은 차원으로 확장한 다음에 연산을 행한다.

1) 원소곱하기 :

원소곱하기연산자 #의 사용양식은 다음과 같다.

$$M_1 \# M_2$$

$$M_1 \# S_1$$

$$M_1 \# V_1$$

앞에서 설명했듯이 차원이 같은 두 행렬들에 원소곱하기를 해서 얻은 행렬의 (i, j) 원소는 이 두 행렬들의 (i, j) 원소들의 곱이다. 행렬과 벡터, 또는 행렬과 스칼라 사이의 원소곱하기에서는 먼저 벡터와 스칼라를 주어진 행렬과 같은 차원으로 확장한 다음에 연산을 행한다. 다음과 같은 예들을 살펴보자.

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \# \begin{bmatrix} 10 \\ 100 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \# 3$$

이 원소곱하기를 한 결과는 각각 다음 연산을 한 결과와 같다.

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \# \begin{bmatrix} 10 & 10 \\ 100 & 100 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \# \begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix}$$

2) 원소나누기: /

원소나누기연산자 /의 사용양식은 다음과 같다.

$$M_1 / M_2$$

$$M_1 / S_1$$

차원이 같은 두 행렬들에 원소나누기를 해서 얻은 행렬의 (i, j) 원소는 첫 번째 행렬의

(i, j) 원소를 두 번째 행렬의 (i, j) 원소로 나눈 것이다. 반면에 행렬과 스칼라 사이의 원소 나누기연산에서는 먼저 주어진 스칼라를 주어진 행렬과 같은 차원으로 확장한 다음에 연산을 행한다.

3) 원소곱:

원소곱연산자 ##의 사용양식은 다음과 같다.

$$M_1 \## M_2$$

$$M_1 \## S_1$$

차원이 같은 두 행렬들에 원소곱연산을 한 결과로 발생된 행렬의 (i, j) 원소는 첫 번째 행렬의 (i, j) 원소에 두 번째 행렬의 (i, j) 원소만큼의 곱을 취한 것이다. 행렬과 스칼라 사이의 원소곱연산에서는 행렬의 각 원소에 주어진 스칼라만큼의 곱을 취한다.

4) 원소최대: <>

원소최대연산자 <>의 사용양식은 다음과 같다.

$$M_1 <> M_2$$

$$M_1 <> S_1$$

차원이 같은 두 행렬들에 원소최대연산을 한 결과로 발생된 행렬의 (i, j) 원소는 첫 번째 행렬의 (i, j) 원소와 두 번째 행렬의 (i, j) 원소 중에서 큰 것을 고른 것이다. 반면에 행렬과 스칼라 사이의 원소최대연산에서는 행렬의 각 원소와 주어진 스칼라를 비교해서 큰 값을 택한다.

5) 원소최소: ><

원소최소연산자 ><의 사용양식은 다음과 같다.

$$M_1 >< M_2$$

$$M_1 >< S_1$$

차원이 같은 두 행렬들에 원소최소연산을 한 결과로 발생된 행렬의 (i, j) 원소는 첫 번째 행렬의 (i, j) 원소와 두 번째 행렬의 (i, j) 원소 중에서 작은 것을 고른 것이다. 반면에 행렬과 스칼라 사이의 원소최소연산에서는 행렬의 각 원소와 주어진 스칼라를 비교해서 작은 값을 택한다.

지금까지 설명한 원소기준연산자들에 대한 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-10.PRG에 수록되어 있다).


```

/* */
/*          EXAM3-10.PRG          */
/* */
/*          ELEMENTWISE OPERATORS  */
/* */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOCENTER NOLOG NOPRINT;
M1 = {1 2, 3 4};
M2 = {4 3, 2 1};
      PRINT M1 " " M2,;
S1 = 2;
      PRINT S1,;
V1 = {1 1};
      PRINT V1,;
M3 = M1 # M2;
M4 = M1 # 3;
M5 = M1 # S1;
M6 = M1 # V1;
      PRINT M3 " " M4, M5 " " M6,;
M7 = M1 / M2;
M8 = M1 / 2;
      PRINT M7 " " M8,;
M9 = M1 ## M2;
M10 = M1 ## S1;
      PRINT M9 " " M10,;
M11 = M1 <> M2;
M12 = M1 <> S1;
      PRINT M11 " " M12,;
M13 = M1 >< M2;
M14 = M1 >< S1;
      PRINT M13 " " M14,;
QUIT;

```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.10과 같은 결과가 나타난다.

```

SAS
7:46 Saturday, March 5, 1994 13

M1          M2
1           4           3
3           2           1
           4

S1
2

V1
1           1

M3          M4
4           3           6
6           4           9           12

M5          M6
2           4           1           2
6           8           3           4

M7          M8
0.25 0.6666667 0.5           1
1.5           4           1.5           2

M9          M10
1           8           1           4
9           4           9           16

M11         M12
4           3           2           2
3           4           3           4

M13         M14
1           2           1           2
2           1           2           2

Exiting IML.
    
```

그림 3.10 원소기준연산자들을 실행한 결과

6) 원소비교: $\langle \rangle = \langle = \rangle = \hat{=}$

원소비교연산자들(elementwise comparison operators) $\langle, \rangle, =, \langle =, \rangle =, \hat{=}$ 의 사용양식은 다음과 같다.

$$\begin{aligned}
 M_1 < M_2, & \quad M_1 < S_1 \\
 M_1 > M_2, & \quad M_1 > S_1 \\
 M_1 = M_2, & \quad M_1 = S_1 \\
 M_1 \langle = M_2, & \quad M_1 \langle = S_1 \\
 M_1 \rangle = M_2, & \quad M_1 \rangle = S_1 \\
 M_1 \hat{=} M_2, & \quad M_1 \hat{=} S_1
 \end{aligned}$$

두 $m \times n$ 행렬들을 원소비교하면 $m \times n$ 행렬이 발생한다. 이 발생된 행렬의 (i, j) 원소는 첫 번째 행렬의 (i, j) 원소와 두 번째 행렬의 (i, j) 원소를 비교한 결과가 참(true)이면 1, 거짓(false)이면 0이다. 또한, 행렬과 스칼라 사이의 원소비교연산에서는 먼저 주어진 스칼라를 주어진 행렬과 같은 차원으로 확장한 다음에 연산을 행한다.

원소비교연산자들에 대한 예제로 다음 SAS/IML 프로그램을 수행해 보자(이 SAS/IML 프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-11.PRG에 수록되어 있다).

```

/* */
/* */          EXAM3-11.PRG          /* */
/* */          /* */
/* */          ELEMENTWISE COMPARISON OPERATORS          /* */
/* */          /* */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOCENTER NOLOG NOPRINT;
M1 = {1 2, 3 4};
M2 = {4 2, 3 1};
S1 = 2;
      PRINT M1 " " M2 " " S1,;
M3 = M1 < M2;
M4 = M1 < S1;
      PRINT M3 " " M4,;
M5 = M1 > M2;
M6 = M1 > S1;
      PRINT M5 " " M6,;
M7 = M1=M2;
M8 = M1=S1;
      PRINT M7 " " M8,;
M9 = M1<=M2;
M10 = M1<=S1;
      PRINT M9 " " M10,;
M11 = M1>=M2;
M12 = M1>=S1;
      PRINT M11 " " M12,;
M13 = M1^=M2;
M14 = M1^=S1;
      PRINT M13 " " M14,;
QUIT;

```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.11과 같은 결과가 나타난다.

7) 원소논리: & | ^

원소논리연산자들(elementwise logical operators)인 &(and), |(or), ^(not)의 사용양

```

SAS
7:46 Saturday, March 5, 1994 14

M1      M2      S1
1      4      2
2      2      2
3      3      1

M3      M4
1      1      0
0      0      0

M5      M6
0      0      0
0      1      1

M7      M8
0      0      1
1      0      0

M9      M10
1      1      1
1      0      0

M11     M12
0      0      1
1      1      1

M13     M14
1      1      0
0      1      1

Exiting IML.
    
```

그림 3.11 원소비교연산자들을 실행한 결과

식은 다음과 같다.

$$\begin{aligned}
 &M_1 \& M_2, \quad M_1 \& S_1 \\
 &M_1 | M_2, \quad M_1 | S_1 \\
 &\hat{M}_1
 \end{aligned}$$

이 원소논리연산자들을 적용하는 행렬의 원소들과 스칼라는 0 또는 1이어야 한다. 원소논리곱연산자(elementwise logical-and operator) &는 차원이 같은 두 행렬들의 원소논리곱을 한다. 만일, 첫 번째 행렬의 (i, j) 원소, 또는 두 번째 행렬의 (i, j) 원소가 0이라면 원소논리곱된 행렬의 (i, j) 원소는 0이고, 만일 첫 번째 행렬의 (i, j) 원소와 두 번째 행렬의 (i, j) 원소가 모두 1이라면 원소논리곱된 행렬의 (i, j) 원소는 1이다. 또한, 행렬과 스칼라 사이의 원소논리곱연산에서는 먼저 주어진 스칼라를 주어진 행렬과 같은 차원으로 확장한 다음에 원소논리곱연산을 행한다.

원소논리합연산자(elementwise logical-or operator) |는 차원이 같은 두 행렬들의 원소논리합을 한다. 만일, 첫 번째 행렬의 (i, j) 원소와 두 번째 행렬의 (i, j) 원소가 모두 0이라면 원소논리합된 행렬의 (i, j) 원소는 0이고, 만일 첫 번째 행렬의 (i, j) 원소, 또는 두 번째 행렬의 (i, j) 원소가 1이라면 원소논리합된 행렬의 (i, j) 원소는 1이다. 또한, 행렬과 스칼라 사이의 원소논리합연산에서는 먼저 주어진 스칼라를 주어진 행렬과 같은 차원으로 확장한 다음에 원소논리합연산을 행한다.

원소논리부연산자(elementwise logical-not operator) ^는 행렬의 원소논리부연산을 행한다. 만일, 행렬의 (i, j) 원소가 0이라면 원소논리부연산이 수행된 행렬의 (i, j) 원소는 1이고, 만일 행렬의 (i, j) 원소가 1이라면 원소논리부연산이 수행된 행렬의 (i, j) 원소는 0이다.

원소논리연산자들에 대한 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-12.PRG에 수록되어 있다).

```

/*                                     */
/*                                     EXAM3-12.PRG                             */
/*                                     */
/*                                     ELEMENTWISE LOGICAL OPERATORS           */
/*                                     */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOCENTER NOLOG NOPRINT;
M1 = {1 0 1, 0 1 1, 1 1 1};
M2 = {0 0 0, 0 0 1, 1 0 1};
S1 = 0;
      PRINT M1 " " M2, S1;
M3 = M1 & M2;
M4 = M1 & S1;
      PRINT M3 " " M4;
M5 = M1 | M2;
M6 = M1 | S1;
      PRINT M5 " " M6;

```

SAS			7:46 Saturday, March 5, 1994 15			
M1			M2			S1
1	0	1	0	0	0	0
0	1	1	0	0	1	
1	1	1	1	0	1	
M3			M4			
0	0	0	0	0	0	
0	0	1	0	0	0	
1	0	1	0	0	0	
M5			M6			
1	0	1	1	0	1	
0	1	1	0	1	1	
1	1	1	1	1	1	
M7						
0	1	0				
1	0	0				
0	0	0				

Exiting IML.

그림 3.12 원소논리연산자들을 실행한 결과

```
M7 = ^M1;
      PRINT M7;
QUIT;
```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.12와 같은 결과가 나타난다.

3.3.4 부행렬들

이 소절에서는 부행렬들을 만드는 데 사용하는 첨자연산자들(subscript operators)을 다루고자 한다. 다음과 같은 행렬을 사용해서 첨자연산자들을 설명하기로 하자.

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \\ 16 & 17 & 18 & 19 & 20 \end{bmatrix}$$

이 소절에서 다루는 첨자연산자들에 대한 자세한 내용은 [IML6.03](pp. 59~65)을 참조하라.

1) 첨자창조: :

첨자창조연산자(index creation operator) : 의 사용양식은 다음과 같다.

정수 1 : 정수 2

첨자창조연산자는 첨자들을 나타내는 행벡터를 생성하는 데 사용된다. 즉, 첫 번째 원소를 정수 1, 두 번째 원소를 (정수 1 + 1), 세 번째 원소를 (정수 1 + 2), ..., 마지막 원소를 정수 2로 하는 행벡터를 생성한다. 다음과 같은 예를 살펴보자.

$I = 7 : 10 ;$

이 스테이트먼트를 실행하면 결과는 다음과 같은 행벡터가 발생한다.

$I = [7\ 8\ 9\ 10]$

만일, 원소들의 증가 크기가 1이 아니면 제 3.4.2 소절에서 설명할 DO 함수를 사용한다. 만일, 정수 1이 정수 2보다 크면 행의 수가 0이고 열의 수가 0인 빈 행렬(empty matrix)이 생성된다.

2) 첨자: []

첨자연산자(subscript operator)는 주어진 $p \times q$ 행렬 M 의 부행렬들을 만드는 데 사용

한다. 여기서는 부행렬을 1×1 행렬인 원소, $p \times 1$ 행렬인 열벡터, 또는 $1 \times q$ 행렬인 행벡터, 그리고 $r \times s$ 행렬인 부행렬(단, $r > 1$ 또는 $s > 1$)로 나누어 생각하기로 한다.

(1) 원 소

행렬 M 의 한 원소로 구성되는 부행렬을 만들기 위해서는 첨자연산자를 다음과 같이 사용한다.

$$M[i, j];$$

이 첨자연산을 수행하면 행렬 M 의 (i, j) 원소로 이루어진 1×1 행렬을 만든다. 다음과 같은 예를 살펴보자.

$$A1 = A[2, 3];$$

이 첨자스테이트먼트를 실행한 결과는 다음과 같다.

$$A1 = [8];$$

(2) 행벡터 또는 열벡터

행렬 M 의 제 i 번째 행으로 구성되는 행벡터를 만들기 위해서는 다음과 같이 첨자연산자를 사용한다.

$$M[i,];$$

이 첨자연산을 수행하면 행렬 M 의 제 i 번째 행으로 이루어진 $1 \times q$ 행렬을 만든다. 다음과 같은 예를 살펴보자.

$$A2 = A[3,];$$

이 첨자스테이트먼트를 실행한 결과는 다음과 같다.

$$A2 = [11 \ 12 \ 13 \ 14 \ 15]$$

행렬의 제 j 번째 열로 구성되는 열벡터를 만들기 위해서는 다음과 같이 첨자연산자를 사용한다.

$$M[, j];$$

이 첨자연산을 수행하면 행렬의 제 j 번째 열로 이루어진 $p \times 1$ 행렬을 만든다. 다음과 같은 예를 살펴보자.

$$A3 = A[, 4];$$

이 첨자스테이트먼트를 실행한 결과는 다음과 같다.

$$A3 = \begin{bmatrix} 4 \\ 9 \\ 14 \\ 19 \end{bmatrix}$$

(3) 부행렬

행렬 M 의 제 i_1 번째, 제 i_2 번째, ..., 제 i_s 번째 행들과 제 j_1 번째, 제 j_2 번째, ..., 제 j_t 번째 열들의 공통부분에 있는 원소들로 이루어진 부행렬을 만들기 위해서는 다음과 같이 첨자연산자를 사용한다.

$$M[\{i_1 \ i_2 \ \dots \ i_s\}, \{j_1 \ j_2 \ \dots \ j_t\}];$$

이 양식에서 유의할 점은 열번호들과 행번호들 사이에는 쉼표를 사용하지 않고 하나 이상의 빈 칸만을 놓는다는 것이다. 이 첨자연산을 수행하면 행렬 M 의 제 i_1 번째, 제 i_2 번째, ..., 제 i_s 번째 행들과 제 j_1 번째, 제 j_2 번째, ..., 제 j_t 번째 열들의 공통부분에 있는 원소들로 이루어진 $s \times t$ 행렬을 만든다. 다음과 같은 예를 살펴보자.

$$A4 = A[\{1 \ 3\}, \{2 \ 3 \ 5\}];$$

이 첨자스테이트먼트를 실행한 결과는 다음과 같다.

$$A4 = \begin{bmatrix} 2 & 3 & 5 \\ 12 & 13 & 15 \end{bmatrix}$$

첨자스테이트먼트에 첨자벡터들을 사용할 수가 있다. 다음과 같은 예를 살펴보자.

$$I1 = \{1 \ 3\};$$

$$I2 = \{2 \ 3 \ 5\};$$

$$A4 = A[I1, I2];$$

이 첨자스테이트먼트를 실행한 결과는 앞에서 얻은 행렬 $A4$ 와 같다.

$$A4 = \begin{bmatrix} 2 & 3 & 5 \\ 12 & 13 & 15 \end{bmatrix}$$

부행렬은 지정된 행들의 순서와 열들의 순서에 따라 만들어진다. 다음과 같은 예를 살펴보자.

$$A5 = A[\{3 \ 1\}, \{2 \ 5 \ 3\}];$$

이 첨자스테이트먼트를 실행한 결과는 다음과 같다.

$$A5 = \begin{bmatrix} 12 & 15 & 13 \\ 2 & 5 & 3 \end{bmatrix}$$

또한, 부행렬이 열벡터인 경우에는 열번호를 둘러싸는 중괄호를 생략할 수 있고, 행벡터인 경우에는 행번호를 둘러싸는 중괄호를 생략할 수 있다. 예를 들면, 다음의 두 스테이트먼트들은 동일한 결과를 출력한다.

$$A6 = A[\{3\}, \{2\ 5\ 3\}];$$

$$A6 = A[3, \{2\ 5\ 3\}];$$

3) 첨자감축연산: + # <> >< <:> >:< :

첨자감축연산자들의 사용양식은 다음과 같다.

$$M[+,], \quad M[,+], \quad M[+,+] = M[+]$$

$$M[\#,], \quad M[,\#], \quad M[\#, \#] = M[\#]$$

$$M[\langle \rangle,], \quad M[,\langle \rangle], \quad M[\langle \rangle, \langle \rangle] = M[\langle \rangle]$$

$$M[\rangle \langle,], \quad M[,\rangle \langle], \quad M[\rangle \langle, \rangle \langle] = M[\rangle \langle]$$

$$M[\langle : \rangle,], \quad M[,\langle : \rangle], \quad M[\langle : \rangle, \langle : \rangle] = M[\langle : \rangle]$$

$$M[\rangle : \langle,], \quad M[,\rangle : \langle], \quad M[\rangle : \langle, \rangle : \langle] = M[\rangle : \langle]$$

$$M[:,], \quad M[,:], \quad M[:, :] = M[:]$$

$$M[\#\#\#,], \quad M[,\#\#\#], \quad M[\#\#\#, \#\#\#] = M[\#\#\#]$$

첨자의 위치에 첨자대신 연산기호들 +, #, <>, ><, <:>, >:<, :, ### 를 사용하여 첨자감축연산자들(subscript reduction operators)로 사용할 수 있다. 이 첨자감축연산자들을 사용하면 모든 행들, 또는 모든 열들에 따라 감축(reductions across all rows and columns)을 행할 수 있다.

예를 들어, M 이 $p \times q$ 행렬일 때, $M[+,]$ 는 행에 대해서 더하기감축을 실시한 $1 \times q$ 행렬로서 이 행렬의 $(1, j)$ 원소는 $\sum_{k=1}^p m_{kj}$ 이고, $M[,+]$ 는 열에 대한 더하기감축을 실시한 $p \times 1$ 행렬로서 이 행렬의 $(i, 1)$ 원소는 $\sum_{k=1}^q m_{ik}$ 이며, $M[+,+]$ 는 행에 대한 더하기감축을 실시한 다음에 열에 대한 더하기감축을 실시한 1×1 행렬로서 이 행렬의 $(1, 1)$ 원소는 $\sum_{i=1}^p \sum_{j=1}^q m_{ij}$ 이다. 즉, 각 행에 대해서 모든 열들의 원소들의 합을 구하기 위해서는 첨자감축연산자 $M[,+]$ 를 사용하고, 각 열에 대해서 모든 행들의 원소들의 합을 구하기 위해서는 첨자감축연산자 $M[+,]$ 를 사용하고, 행렬의 모든 원소들의 합을 구하기 위해서는 첨자감축연산자 $M[+,+]$, 또는 $M[+]$ 를 사용한다. 다음과 같은 예들을 살펴보자.

$$A7 = A[+,];$$

$$A8 = A[,+];$$

$$A9 = A[+,+];$$

이 스테이트먼트들을 실행한 결과는 다음과 같다.

$$A7 = [34 \ 38 \ 42 \ 46 \ 50], \quad A8 = \begin{bmatrix} 15 \\ 40 \\ 65 \\ 90 \end{bmatrix}, \quad A9 = [210]$$

첨자의 위치에 첨자대신 기입하는 연산기호들 +, #, <>, ><, <:>, >:<, :, ## 에 따라 감축된 행렬의 각 원소는 합, 곱, 최대값, 최소값, 최대값을 갖는 첨자, 최소값을 갖는 첨자, 표본평균, 제곱합을 나타낸다. 행과 열에 대해서 함께 감축을 실시할 때는 행에 대해서 먼저 감축을 실시한 다음에 열에 대해서 감축을 실시한다. 한 예로 각 행별 최대값들의 합을 구하려면 $M[,<>][+,]$ 을 사용해야 한다. 만일, $M[+,<>]$ 을 사용하면 먼저 각 열별 합을 구하고 다음 단계로 이 합들 중에서 최대값을 출력한다. 이 첨자감축연산자들에 관한 자세한 내용은 [IML6.03] (pp. 63~64)을 참조하라.

첨자연산자들과 첨자감축연산자들에 대한 예제로 다음 SAS/IML 프로그램을 수행해 보자(이 SAS/IML 프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-13.PRG에 수록되어 있다).

```

/*                               */
/*           EXAM3-13.PRG         */
/*                               */
/*           SUBSCRIPT OPERATORS  */
/*                               */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOCENTER NOLOG PRINT;
I1 = 7:10;
A = { 1 2 3 4 5,
      6 7 8 9 10,
      11 12 13 14 15,
      16 17 18 19 20};
A1 = A[2,3];
A2 = A[3, ];
A3 = A[ ,4];
A4 = A[{1 3}, {2 3 5}];
I1 = {1 3};
I2 = {2 3 5};
A4TWO = A[I1, I2];
A5 = A[{3 1}, {2 5 3}];
A6 = A[3, {2 5 3}];
A7 = A[+, ];

```

SAS		21:49 Thursday, February 17, 1994 11				
I1	1 rows 4 columns 7 8 9 10					
A	4 rows 5 columns 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20					
A1	1 rows 1 columns 8					
A2	1 rows 5 columns 11 12 13 14 15					
A3	4 rows 1 columns 4 9 14 19					
A4	2 rows 3 columns 2 3 5 12 13 15					
I1	1 rows 2 columns 1 3					
I2	1 rows 3 columns 2 3 5					
A4TWO	2 rows 3 columns 2 3 5 12 13 15					
A5	2 rows 3 columns 12 15 13 2 5 3					
A6	1 rows 3 columns 12 15 13					
A7	1 rows 5 columns 34 38 42 46 50					
A8	4 rows 1 columns 15 40 65 90					
A9	1 rows 1 columns 210					
A10	1 rows 1 columns 50					
A11	1 rows 1 columns 50					
A12	2 rows 1 columns 40 65					
A13	1 rows 1 columns 90					
A14	1 rows 1 columns 34					

```

A15      4 rows 1 columns
         5
         5
         5
         5

A16      1 rows 5 columns
         1      1      1      1      1

A17      1 rows 1 columns
         10.5

Exiting IML.
    
```

그림 3.13 첨자연산자들을 실행한 결과

```

A8 = A[ ,+];
A9 = A[+,+];
A10 = A[ ,<>][+, ];
A11 = A[+, <>];
A12 = A[{2 3}, +];
A13 = A[<>, +];
A14 = A[ ,><][+, ];
A15 = A[ ,<:>];
A16 = A[>:<, ];
A17 = A[:];
QUIT;
    
```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.13과 같은 결과가 나타난다.

3.3.5 연산자들의 우선순위

지금까지 여러 연산자들을 설명하였다. 이 연산자들을 사용해서 합성표현(compound expression)을 할 수가 있다. 예를 들어, 다음과 같은 스테이트먼트들은 올바른 것들이다.

```

A = X+Y+Z;
A = X+Y*Z;
A = X/Y/Z;
A = X/(Y/Z);
    
```

이 스테이트먼트들의 우측에 있는 합성표현들을 계산할 때는 다음과 같은 규칙을 따른다.

- ① 표 3.1에 수록된 연산자들의 우선순위(operator precedence)에 따라 연산이 수행된다. 그룹 I의 연산자들이 최우선순위에 있고, 그룹 II의 연산자들이 다음 우선순위에 있으며, 그룹 VII의 연산자가 최하위순위에 있다. 다음과 같은 예를 살펴보자.

$$A = X + Y * Z;$$

이 스테이트먼트를 수행할 때는 우선 행렬들 Y 와 Z 의 곱인 행렬이 계산되고, 이 곱행렬에 행렬 X 가 더해진 다음, 이 결과가 행렬 A 에 저장된다.

- ② 같은 그룹의 연산자들 사이에서는 그룹 I인 경우에는 우측에서 좌측으로의 순서에 따라 연산이 수행되고 나머지 그룹에서는 좌측에서 우측으로의 순서에 따라 연산이 수행된다.
- ③ 괄호가 있으면 이 안에 포함되어 있는 연산들이 가장 먼저 수행된다.

그룹 I	^ ` 침자 -(부호바꿈) ## **
그룹 II	* # <>>< / @
그룹 III	+ -(빼기)
그룹 IV	// :
그룹 V	<< => = = ^ =
그룹 VI	&
그룹 VII	

표 3.1 연산자들의 우선순위

3.4 SAS/IML함수

수학에서 그러하듯이 SAS/IML환경에서 사용되는 SAS/IML함수는 연산기호에 의해서가 아니라 고유한 함수의 이름에 의해서 작동되는 연산자이다. SAS/IML함수는 다음과 같이 함수이름 다음에 있는 소괄호 안에 매개변수들(arguments)을 나열해서 정의한다.

결과 = 함수이름(매개변수 1, 매개변수 2, ..., 매개변수 n);

여기서 매개변수 1, 매개변수 2, ..., 매개변수 n 으로는 행렬이름들(matrix names), 문자들(literals), 그리고 수식들(expressions)을 사용할 수 있다. 제 3.3절에서 보았듯이 연산자에는 원소를 기준으로 하는 연산자와 행렬을 기준으로 하는 연산자가 있다. 이와 마찬가지로 SAS/IML함수에는 원소기준의 연산을 하는 스칼라함수와 행렬기준의 연산을 하는 행렬함수가 있다. 제 3.4.1소절에서는 스칼라함수를 설명하고 제 3.4.2소절에서는 행렬함수를 설명한다. SAS/IML함수와 같은 역할을 하는 것으로 서브루틴(subroutine)이 있다. 서브루틴을 호출하기 위해서는 CALL스테이트먼트를 사용한다. 이 서브루틴들에 대해서는 제 3.4.3소절에서 설명할 것이다.

3.4.1 스칼라함수

SAS/IML의 스칼라함수는 원소기준의 함수이다. 즉, 매개변수로 행렬을 사용하고 결과도 행렬로 출력된다. 스칼라함수들은 다음과 같다.

ABS, ARCOS, ARSIN, ATAN, COS, EXP, INT, LOG, MOD,
 PROBCHI, PROBF, PROBIT, PROBNORM, PROBT,
 NORMAL, SIN, TAN, UNIFORM

독자들이 이 스칼라함수들의 의미를 잘 알고 있으리라고 생각되기 때문에 구태여 이 함수들을 여기에서 설명할 필요성을 느끼지 않는다. 이 함수들이 생성한 독자는 표 3.2와 [SAS4TSA1](p. 61)을 참조하라.

스칼라함수들의 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-14.PRG에 수록되어 있다).

함수	결과	사용법
ABS	절대값행렬	ABS(행렬)
ARCOS	역코사인행렬	ARCOS(행렬)
ARSIN	역사인행렬	ARSIN(행렬)
ATAN	역탄젠트행렬	ATAN(행렬)
COS	코사인행렬	COS(행렬)
EXP	지수행렬	EXP(행렬)
INT	정수화행렬	INT(행렬)
LOG	자연로그행렬	LOG(행렬)
MOD	나머지행렬	MOD(행렬, 나눗수)
PROBCHI	카이제곱분포행렬	PROBCHI(행렬, 자유도)
PROBF	F분포행렬	PROBF(행렬, 분자자유도, 분모자유도)
PROBIT	역정규분포행렬	PROBIT(행렬)
PROBNORM	정규분포행렬	PROBNORM(행렬)
PROBT	t분포행렬	PROBT(행렬, 자유도)
NORMAL	정규난수행렬	NORMAL(행렬)
SIN	사인행렬	SIN(행렬)
SQRT	제곱근행렬	SQRT(행렬)
TAN	탄젠트행렬	TAN(행렬)
UNIFORM	일양난수행렬	UNIFORM(행렬)

표 3.2 SAS/IML의 스칼라함수들

```
/* */
/*          EXAM3-14.PRC          */
/* */
/*          SCALAR FUNCTIONS      */
/* */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOCENTER NOLOG NOPRINT;
M = {-1 0 1, 0 0.2 -0.2, 0.3 0.12 -0.4};
PRINT, 'MATRIX = 'M,;
A1 = ABS(M);
PRINT, 'ABS FUNCTION', A1;
A2 = ARCOS(M);
PRINT, 'ARCOS FUNCTION', A2;
A3 = ARSIN(M);
PRINT, 'ARSIN FUNCTION', A3;
A4 = ATAN(M);
PRINT, 'ATAN FUNCTION', A4;
A5 = COS(M);
PRINT, 'COS FUNCTION', A5;
A6 = EXP(M);
PRINT, 'EXP FUNCTION', A6;
A7 = INT(5*M);
PRINT, 'INT FUNCTION', A7;
A8 = LOG(A1+100);
PRINT, 'LOG FUNCTION', A8;
A9 = MOD(100*M, 7);
PRINT, 'MOD FUNCTION', A9;
A10 = PROBCHI(ABS(M), 3);
PRINT, 'PROBCHI FUNCTION', A10;
A11 = PROBF(ABS(M), 3, 2);
PRINT, 'PROBF FUNCTION', A11;
A12 = PROBIT(A11+0.01);
PRINT, 'PROBIT FUNCTION', A12;
A13 = PROBNORM(ABS(M));
PRINT, 'PROBNORM FUNCTION', A13;
A14 = PROBT(ABS(M), 15);
PRINT, 'PROBT FUNCTION', A14;
```

SAS

14:39 Monday, February 21, 1994 36

MATRIX =

	M		
	-1	0	1
	0	0.2	-0.2
	0.3	0.12	-0.4

ABS FUNCTION

A1		
1	0	1
0	0.2	0.2
0.3	0.12	0.4

ARCOS FUNCTION

A2			
3.1415927	1.5707963		0
1.5707963	1.3694384	1.7721542	
1.2661037	1.4505064	1.9823132	

ARSIN FUNCTION

A3			
-1.570796	0	1.5707963	
0	0.2013579	-0.201358	
0.3046927	0.1202899	-0.411517	

ATAN FUNCTION

A4			
-0.785398	0	0.7853982	
0	0.1973956	-0.197396	
0.2914568	0.1194289	-0.380506	

COS FUNCTION

A5			
0.5403023	1	0.5403023	
1	0.9800666	0.9800666	
0.9553365	0.9928086	0.921061	

EXP FUNCTION

A6			
0.3678794	1	2.7182818	
1	1.2214028	0.8187308	
1.3498588	1.1274969	0.67032	

INT FUNCTION

A7			
-5	0	5	
0	1	-1	
1	0	-2	

LOG FUNCTION

A8			
4.6151205	4.6051702	4.6151205	
4.6051702	4.6071682	4.6071682	
4.6081657	4.6063695	4.6091622	

MOD FUNCTION

A9			
-2	0	2	
0	6	-6	
2	5	-5	


```

PROBCHI FUNCTION
      A10
0.198748      0      0.198748
      0      0.0224107      0.0224107
0.0399715      0.0106662      0.0597575

PROBF FUNCTION
      A11
0.464758      0      0.464758
      0      0.110858      0.110858
0.1728888      0.059578      0.2296397

PROBIT FUNCTION
      A12
-0.063315      -2.326348      -0.063315
-2.326348      -1.170709      -1.170709
-0.904411      -1.478941      -0.707462

PROBNORM FUNCTION
      A13
0.8413447      0.5      0.8413447
      0.5      0.5792597      0.5792597
0.6179114      0.5477584      0.6554217

PROBT FUNCTION
      A14
0.8334149      0.5      0.8334149
      0.5      0.5779166      0.5779166
0.6158527      0.5469622      0.6526025

NORMAL FUNCTION
      A15
0.1328326      -0.375777      -2.120378
2.4217131      -1.21159      0.8889208
-0.086855      -0.014762      0.1533778

SIN FUNCTION
      A16
-0.841471      0      0.841471
      0      0.1986693      -0.198669
0.2955202      0.1197122      -0.389418

SQRT FUNCTION
      A17
      1      0      1
      0      0.4472136      0.4472136
0.5477226      0.3464102      0.6324555

TAN FUNCTION
      A18
-1.557408      0      1.5574077
      0      0.20271      -0.20271
0.3093362      0.1205793      -0.422793

UNIFORM FUNCTION
      A19
0.2215869      0.2760926      0.4888256
0.707499      0.5120657      0.194007
0.5847473      0.3178335      0.2082261

Exiting IML.

```

그림 3.14 스칼라함수들을 실행한 결과

```

A15 = NORMAL(ABS(M)*1234567);
PRINT, 'NORMAL FUNCTION', A15;
A16 = SIN(M);
PRINT, 'SIN FUNCTION', A16;
A17 = SQRT(ABS(M));
PRINT, 'SQRT FUNCTION', A17;
A18 = TAN(M);
PRINT, 'TAN FUNCTION', A18;
A19 = UNIFORM(ABS(M)*1234567);
PRINT, "UNIFORM FUNCTION ", A19;
QUIT;

```

이 SAS/IML 프로그램의 PRINT 스테이트먼트들에서는 작은따옴표(')와 큰따옴표(")가 사용되었다. 단순히 결과물에 출력시키고 싶은 문자들을 이 따옴표들 사이에 입력시킨다. 이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.14와 같은 결과가 나타난다.

3.4.2 행렬함수

[IML6.03] (pp. 202~209)에서는 행렬함수들을 행렬조회함수(matrix inquiry function), 축약함수(reduction function), 행렬계산함수(matrix arithmetic function), 행렬조작함수(manipulation and reshaping function), 수열함수(sequence function), 순서함수(ordering function), 문자조작함수(character manipulation function), 집합함수(set function), 접촉조정함수(interface function), 선형대수함수(linear algebraic function), 통계함수(statistical function), 시계열함수(time series function) 등으로 구분한다. 이 소절에서는 이 분류에 따라 행렬함수들을 설명하고자 한다. 각 함수에 대한 자세한 설명은 [IML6.03] (pp. 222~318)을 참조하라.

1) 행렬조회함수

(1) ALL(행렬)

매개변수로 쓰인 행렬의 어느 한 원소라도 0이면 결과를 0으로 출력하고 그렇지 않은 경우에는 1을 출력한다. 따라서 이 행렬조회함수 ALL의 매개변수는 행렬이지만 함수값은 0 또는 1이다. 이 함수는 주로 비교연산의 결과를 확인할 목적으로 사용한다.

(2) ANY(행렬)

매개변수로 쓰인 행렬의 어느 한 원소라도 0이 아니면 결과를 1로 출력하고 그렇지 않은

경우에는 0을 출력한다. 따라서 이 행렬조회함수 ANY의 매개변수는 행렬이지만 함수값은 0 또는 1이다. 이 함수도 주로 비교연산의 결과를 확인할 목적으로 사용한다.

(3) LOC(벡터)

함수 LOC의 매개변수로는 벡터가 사용된다. 행벡터, 또는 열벡터에서 0이 아닌 원소들의 위치를 행벡터로 출력한다. 함수 LOC의 매개변수로 행렬이 사용되면 이 행렬의 첫째 행, 둘째 행, ..., 마지막 행을 가로결합한 행벡터에서 0이 아닌 원소들의 위치를 행벡터로 출력한다.

(4) NCOL(행렬)

매개변수로 쓰인 행렬의 열의 수를 출력한다.

(5) NLENG(행렬)

SAS/IML환경에서 한 행렬의 모든 원소들의 크기(the size in bytes of each element of a matrix)는 같다. 함수 NLENG는 매개변수로 쓰인 행렬의 원소들이 문자열일 때 원소의 크기를 출력한다. 따라서 이 함수값은 스칼라이다.

(6) NROW(행렬)

매개변수로 쓰인 행렬의 행의 수를 출력한다.

행렬조회함수들의 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-15.PRG에 수록되어 있다).

```

/* */
/*          EXAM3-15.PRG          */
/* */
/*          MATRIX INQUIRY FUNCTIONS          */
/* */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOCENTER NOLOG NOPRINT;
M = {1 0 1, 0 1 1, 1 1 0};
PRINT, 'MATRIX = 'M,;
A1 = ALL(M);
PRINT, 'ALL FUNCTION', A1;
A2 = ANY(M);
PRINT, 'ANY FUNCTION', A2;
A3 = LOC(M);
PRINT, 'LOC FUNCTION', A3;

```

```

SAS
7:46 Saturday, March 5, 1994 19

MATRIX =      M
              1      0      1
              0      1      1
              1      1      0

ALL FUNCTION

A1
0

ANY FUNCTION

A2
1

LOC FUNCTION

A3
1      3      5      6      7      8

NCOL FUNCTION

A4
3

NROW FUNCTION

A5
3

Exiting IML.
    
```

그림 3.15 행렬조치함수들을 실행한 결과

```

A4 = NCOL(M);
PRINT, 'NCOL FUNCTION', A4;
A5 = NROW(M);
PRINT, 'NROW FUNCTION', A5, ;
QUIT;
    
```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.15와 같은 결과가 나타난다.

2) 축약함수

(1) MAX(행렬) 또는 MAX(행렬 1, 행렬 2, ...)

함수 MAX는 매개변수로 주어진 행렬들의 원소들 중에서 최대값을 출력한다. 따라서 이 함수는 스칼라를 출력한다. 매개변수로는 15개까지의 행렬들을 사용할 수 있으며 이 행렬들의 차원들이 꼭 같을 필요는 없다.

(2) MIN(행렬) 또는 MIN(행렬 1, 행렬 2, ...)

함수 MIN은 매개변수로 주어진 행렬들의 원소들 중에서 최소값을 출력한다. 따라서 이 함수는 스칼라를 출력한다. 매개변수로는 15개까지의 행렬들을 사용할 수 있으며 이 행렬

들의 차원들이 꼭 같을 필요는 없다.

(3) SSQ(행렬) 또는 SSQ(행렬 1, 행렬 2, ...)

함수 SSQ는 매개변수로 주어진 행렬들의 원소들의 제곱합을 계산한다. 따라서 이 함수는 스칼라를 출력한다. 매개변수로는 15개까지의 행렬들을 사용할 수 있으며 이 행렬들의 차원들이 꼭 같을 필요는 없다.

(4) SUM(행렬) 또는 SUM(행렬 1, 행렬 2, ...)

함수 SUM은 매개변수로 주어진 행렬들의 원소들의 합을 계산한다. 따라서 이 함수는 스칼라를 출력한다. 매개변수로는 15개까지의 행렬들을 사용할 수 있으며 이 행렬들의 차원들이 꼭 같을 필요는 없다.

축약함수들의 예제로 다음 SAS/IML 프로그램을 수행해 보자(이 SAS/IML 프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-16.PRG에 수록되어 있다).

```

/*                                     */
/*           EXAM3-16.PRG             */
/*                                     */
/*           REDUCTION FUNCTIONS      */
/*                                     */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOCENTER NOLOG NOPRINT;
M1 = {1 0 1, 0 1 1, 1 1 0};
M2 = {-1 0 -1, 0 -1 -1, -1 -1 0};
PRINT, M1 " " M2, ;
A1 = MAX(M1, M2);
PRINT, 'MAX FUNCTION', A1;
A2 = MIN(M1, M2);
PRINT, 'MIN FUNCTION', A2;
A3 = SSQ(M1, M2);
PRINT, 'SSQ FUNCTION', A3;
A4 = SUM(M1, M2);
PRINT, 'SUM FUNCTION', A4,;
QUIT;

```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.16과 같은 결과가 나타난다.

```

SAS
7:46 Saturday, March 5, 1994 20

M1      M2
  1      0      1      -1      0      -1
  0      1      1      0      -1      -1
  1      1      0      -1      -1      0

MAX FUNCTION
A1
  1

MIN FUNCTION
A2
 -1

SSQ FUNCTION
A3
 12

SUM FUNCTION
A4
  0

Exiting IML.
    
```

그림 3.16 축약함수들을 실행한 결과

3) 행렬계산함수

(1) CUSUM(행렬)

매개변수로 주어진 행렬의 원소들의 누적합을 출력한다. 따라서 이 함수는 매개변수와 같은 차원의 행렬을 출력한다. 누적합은 행을 우선하는(in row-major order) 규칙에 따라 계산된다.

(2) HDIR(행렬 1, 행렬 2)

함수 HDIR은 행의 수가 같은 두 행렬들을 수평직적(horizontal direct product)한다. 수평직적은 행렬 1과 행렬 2의 각 행들을 Kronecker 곱을 해서 얻은 열벡터들을 쌓아서(stacking) 만든 행렬이다.

(3) TRACE(행렬)

함수 TRACE는 매개변수로 주어진 정방행렬의 대각원소들의 합을 계산한다. 따라서 이 함수는 스칼라를 출력한다.

행렬계산함수들의 예제로 다음 SAS/IML 프로그램을 수행해 보자(이 SAS/IML 프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-17.PRG에 수록되어 있다).

```

/ *                               */
/ *                               EXAM3-17.PRG                               */
/ *                               */
    
```

```

SAS
14:39 Monday, February 21, 1994 43

M1=      M1      M2=      M2
      1  0  1      -1  0  -1
      0  1  1      0  -1  -1
      1  1  0      -1  -1  0

CUMSUM FUNCTION

A1
  1  1  2
  2  3  4
  5  6  6

HDIR FUNCTION

A2
-1  0  -1  0  0  0  -1  0  -1
 0  0  0  0  -1  -1  0  -1  -1
-1 -1  0  -1  -1  0  0  0  0

TRACE FUNCTION

A3
  2

Exiting IML.

```

그림 3.17 행렬계산함수들을 실행한 결과

```

/*          MATRIX ARITHMATIC FUNCTIONS          */
/*
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOCENTER NOLOG NOPRINT FW=6;
M1 = {1 0 1, 0 1 1, 1 1 0};
M2 = {-1 0 -1, 0 -1 -1, -1 -1 0};
PRINT, 'M1 = ' M1 " " 'M2 = ' M2;
A1 = CUSUM(M1);
PRINT, 'CUMSUM FUNCTION', A1;
A2 = HDIR(M1, M2);
PRINT, 'HDIR FUNCTION', A2;
A3 = TRACE(M1);
PRINT, 'TRACE FUNCTION', A3;
QUIT;

```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.17과 같은 결과가 나타난다.

4) 행렬조작함수

(1) BLOCK(행렬 1, 행렬 2, ...)

함수 BLOCK은 매개변수로 주어진 행렬들을 대각행렬들로 하는 블록대각행렬(block-

diagonal matrix)을 만든다. 따라서 이 함수는 행렬을 출력한다. 매개변수로는 15개까지의 행렬들을 사용할 수 있으며 이 행렬들의 차원들이 꼭 같을 필요는 없다.

(2) DIAG(정방행렬) 또는 DIAG(벡터)

함수 DIAG는 매개변수로 주어진 정방행렬, 또는 벡터로부터 대각행렬을 만든다. 매개변수가 정방행렬인 경우에는 출력되는 대각행렬의 대각원소들이 이 정방행렬의 대각원소들이고, 매개변수가 벡터인 경우에는 출력되는 대각행렬의 대각원소들이 이 벡터의 원소들이다.

(3) I(자연수)

함수 I는 매개변수로 주어진 자연수를 차원으로 하는 단위행렬(identity matrix)을 만든다.

(4) INSERT(행렬 1, 행렬 2, i) 또는 INSERT(행렬 1, 행렬 2, i, j)

함수 INSERT는 첫 번째 매개변수인 행렬 1의 지정된 위치에 두 번째 매개변수인 행렬 2를 삽입한 행렬을 만든다. 만일, 행렬 1이 $p \times q$ 행렬이면 행을 나타내는 i 는 0에서 $(p+1)$ 사이의 값을 갖고 열을 나타내는 j 는 0에서 $(q+1)$ 사이의 값을 갖는다. 유의할 점은 i 와 j 중에서 하나가 반드시 0이어야 하고, 둘 다 0이어서는 안 된다. 매개변수 j 가 사용되지 않으면 $j=0$ 이라는 뜻이다.

(5) J(p), J(p, q), 또는 J(p, q, 스칼라)

함수 J는 모든 원소들이 동일한 값을 갖는 $p \times q$ 행렬을 만든다. 만일, 매개변수 q 가 지정이 되지 않으면 $q=p$ 이다. 또한, 스칼라의 디폴트는 1이다. 함수 REPEAT와 함수 SHAPE가 함수 J보다는 좀더 일반적인 기능을 가지고 있다.

(6) REMOVE(행렬, 첨자들)

함수 REMOVE는 두 번째 매개변수에 지정된 첨자들에 대응하는 첫 번째 매개변수인 행렬의 원소들을 제거한 다음 나머지 원소들을 사용해서 행벡터를 만든다.

(7) REPEAT(행렬, r, s)

함수 REPEAT는 첫 번째 매개변수에 지정된 행렬을 행방향으로 r 번 반복한 다음 다시 열방향으로 s 번 반복한 행렬을 만든다. 따라서 만일 주어진 행렬의 차원이 $p \times q$ 이면 이 함수는 $pr \times qs$ 행렬을 출력한다.

(8) SHAPE(행렬, p), SHAPE(행렬, p, q) 또는 SHAPE(행렬, p, q, 스칼라)

함수 SHAPE는 첫 번째 매개변수로 주어진 행렬의 원소들을 사용해서 차원이 다른 새로운 행렬을 만든다. 두 번째 매개변수 p 와 세 번째 매개변수 q 가 새로이 만들어지는 행렬의 차원이다. 네 번째 매개변수인 스칼라는 새로이 만들어지는 행렬의 마지막 원소들로 사용될 여백을 메꾸는 값(padvalue)이다.

(9) SQRSYM(벡터)

함수 SQRSYM는 함수 SYMSQR가 출력하는 형태의 벡터를 매개변수로 한다. 이 벡터의 원소들을 정방행렬의 하위삼각부분에 있는 원소들로 하는 행렬을 만든 뒤 이 행렬을 대칭시켜서 만든 대칭행렬을 출력한다.

(10) SYMSQR(행렬)

함수 SYMSQR는 매개변수로 주어진 정방행렬의 하위삼각부분에 있는 원소들로 열벡터를 만든다.

(11) T(행렬)

함수 T는 매개변수로 주어진 행렬의 전치행렬을 만든다.

(12) VECDIAG(정방행렬)

함수 VECDIAG는 매개변수로 주어진 정방행렬의 대각원소들을 원소들로 하는 열벡터를 만든다.

행렬조작함수들의 예제로 다음 SAS/IML 프로그램을 수행해 보자(이 SAS/IML 프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-18.PRG에 수록되어 있다).

```

/ *                               */
/ *                               EXAM3-18.PRG                               */
/ *                               */
/ *                               MANIPULATION & RESHAPING FUNCTIONS          */
/ *                               */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOCENTER NOLOG NOPRINT;
M1 = { 1 2, 3 4 };
M2 = { 5 6, 7 8 };
PRINT, 'M1 = 'M1 'M2 = 'M2;
A1 = BLOCK(M1, M2);
PRINT, 'BLOCK FUNCTION', A1;
A2 = DIAG(M1);
PRINT, 'DIAG FUNCTION', A2;
A3 = I(4);
PRINT, 'I FUNCTION', A3;
A4 = INSERT(M1, M2, 2, 0);
PRINT, 'INSERT FUNCTION', A4;
A5 = J(2, 3, 100);
PRINT, 'J FUNCTION', A5;

```

SAS

14:39 Monday, February 21, 1994 45

M1= M1
 1 2
 3 4

M2= M2
 5 6
 7 8

BLOCK FUNCTION

A1
 1 2 0 0
 3 4 0 0
 0 0 5 6
 0 0 7 8

DIAG FUNCTION

A2
 1 0
 0 4

I FUNCTION

A3
 1 0 0 0
 0 1 0 0
 0 0 1 0
 0 0 0 1

INSERT FUNCTION

A4
 1 2
 5 6
 7 8
 3 4

J FUNCTION

A5
100 100 100
100 100 100

REMOVE FUNCTION

A6
 2 4

REPEAT FUNCTION

A7
 1 2 1 2 1 2
 3 4 3 4 3 4
 1 2 1 2 1 2
 3 4 3 4 3 4

SHAPE FUNCTION

A8
 1 2 3
 4 100 100
100 100 100

SYMSQR FUNCTION

A9
 1
 2
 3

SQRSYM FUNCTION

A10
 1 2
 2 3

```

T FUNCTION
      A11
        1   3
        2   4

VECDIAG FUNCTION
      A12
        1
        4

Exiting IML.

```

그림 3.18 행렬조작함수들을 실행한 결과

```

A6 = REMOVE(M1, {1 3});
PRINT, 'REMOVE FUNCTION', A6;
A7 = REPEAT(M1, 2, 3);
PRINT, 'REPEAT FUNCTION', A7;
A8 = SHAPE(M1, 3, 3, 100);
PRINT, 'SHAPE FUNCTION', A8;
A9 = SYMSQR({1 2, 2 3});
PRINT, 'SYMSQR FUNCTION', A9;
A10 = SQRSYM(A9);
PRINT, 'SQRSYM FUNCTION', A10;
A11 = T(M1);
PRINT, 'T FUNCTION', A11;
A12 = VECDIAG(A11);
PRINT, 'VECDIAG FUNCTION', A12;
QUIT;

```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.18과 같은 결과가 나타난다.

5) 수열함수

(1) DO(시작값, 끝값, 증가값)

함수 DO는 첫 번째 매개변수로 주어진 시작값으로부터 세 번째 매개변수로 주어진 증가값만큼씩 더한 값들을 두 번째 매개변수로 주어진 끝값까지 만들어서 이 값들을 원소로 하는 행벡터를 출력한다.

함수 DO의 예제로 다음 SAS/IML 프로그램을 수행해 보자(이 SAS/IML 프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-19.PRG에 수록되어 있다).

```

/*                                     */
/*           EXAM3-19.PRG             */
/*                                     */

```

```

SAS
14:39 Monday, February 21, 1994 47
IND
5      5.5      6      6.5      7      7.5      8
Exiting IML.
    
```

그림 3.19 함수 DO를 실행한 결과

```

/ *          SEQUENCE FUNCTION          */
/ *          */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOCENTER NOLOG NOPRINT;
IND = DO(5, 8, 0.5);
PRINT IND,;
QUIT;
    
```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.19와 같은 결과가 나타난다.

6) 순서함수

(1) RANK(행렬)

함수 RANK는 매개변수로 주어진 행렬의 원소들의 랭크를 원소로 하는 행렬을 만든다.

(2) RANKTIE(행렬)

함수 RANKTIE도 매개변수로 주어진 행렬의 원소들의 랭크를 원소로 하는 행렬을 만든다. 함수 RANK와 다른 점은 랭크가 동점(tie)인 경우에는 평균랭크를 출력한다.

순서함수들의 예제로 다음 SAS/IML 프로그램을 수행해 보자(이 SAS/IML 프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-20.PRG에 수록되어 있다).

```

/ *          */
/ *          EXAM3-20.PRG          */
/ *          */
/ *          ORDERING FUNCTION          */
/ *          */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOCENTER NOLOG NOPRINT;
M = {9 8 7, 6 5 5, 5 2 1};
PRINT, 'MATRIX = 'M;
A1 = RANK(M);
    
```

```

SAS
14:59 Monday, February 21, 1994 49

MATRIX =
      M
      9      8      7
      6      5      5
      5      2      1

RANK FUNCTION

      A1
      9      8      7
      6      5      4
      3      2      1

RANKTIE FUNCTION

      A2
      9      8      7
      6      4      4
      4      2      1

Exiting IML.

```

그림 3.20 순서함수들을 실행한 결과

```

PRINT, 'RANK FUNCTION', A1;
A2 = RANKTIE(M);
PRINT, 'RANKTIE FUNCTION', A2;
QUIT;

```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.20과 같은 결과가 나타난다.

7) 문자조작함수

(1) BYTE(행렬)

함수 BYTE는 매개변수로 주어진 행렬의 원소들의 ASCII에 해당하는 문자들을 원소들로 하는 행렬을 출력한다.

(2) CHAR(행렬, w, d), CHAR(행렬, w) 또는 CHAR(행렬)

함수 CHAR는 매개변수로서 원소들이 수치인 수치행렬을 입력해서 각 원소를 문자열로 하는 행렬을 출력한다. 두 번째 매개변수 w 는 출력되는 문자행렬의 원소의 필드크기(field width)이고 세 번째 매개변수 d 는 출력되는 문자행렬의 원소에서 소수점(decimal point)의 위치이다.

(3) CHOOSE(조건, 행렬 1, 행렬 2) 또는 CHOOSE(조건, 행렬, 스칼라)

함수 CHOOSE는 첫 번째 매개변수인 조건이 행렬로 나타났을 때 (i, j) 원소가 참(true)이면 두 번째 매개변수인 행렬 1의 (i, j) 원소를 출력하고 거짓(false)이면 세 번째 매개변수인 행렬 2의 (i, j) 원소를 출력한다. 따라서 조건을 나타내는 행렬, 행렬 1, 그리고 행렬 2의 차원이 같아야 한다. 만일, 세 번째 매개변수가 스칼라인 경우에는 이 스칼라를 같은 차원의 행렬로 확장한 다음에 함수를 계산한다.

(4) CONCAT(행렬 1, 행렬 2) 또는 CONCAT(행렬, 스칼라)

함수 CONCAT는 첫 번째 매개변수인 행렬 1의 (i, j) 원소인 문자열과 두 번째 매개변수인 행렬 2의 (i, j) 원소인 문자열을 결합(concatenation)해서 출력한다. 만일, 두 번째 매개변수가 스칼라인 경우에는 이 스칼라를 같은 차원의 행렬로 확장한 다음에 결합한다.

(5) CSHAPE(행렬, p, q, s)

함수 CSHAPE는 주어진 문자열을 반복사용해서 행렬의 차원을 바꾼다. 이 함수에 대한 자세한 내용은 [IML6.03] (pp. 233~234)을 참조하라.

(6) LENGTH(행렬)

함수 LENGTH는 매개변수인 문자행렬의 각 원소인 문자열의 크기를 출력한다. 만일, 어떤 원소에 문자열이 없으면 문자열의 크기를 1이라 한다. 따라서 출력된 행렬의 차원은 매개변수인 문자행렬의 차원과 같다.

(7) NAME(매개변수 1, 매개변수 2, ...)

함수 NAME은 매개변수들의 이름들을 열벡터로 출력한다.

(8) NUM(행렬)

함수 NUM은 함수 CHAR의 반대되는 역할을 하는 함수이다. 매개변수로서 원소들이 문자열들이 문자행렬을 입력해서 각 원소를 수치로 바꾼 행렬을 출력한다.

(9) ROWCAT(행렬, p, q), ROWCAT(행렬, p) 또는 ROWCAT(행렬)

함수 ROWCAT는 매개변수로서 원소들이 문자열들이 문자행렬을 입력해서 각 행의 모든 원소들을 결합해서 한 문자열을 만든다. 이 함수에 대한 자세한 내용은 [IML6.03] (p. 298)을 참조하라.

(10) ROWCATC(행렬, p, q), ROWCATC(행렬, p) 또는 ROWCATC(행렬)

함수 ROWCATC는 함수 ROWCAT와 마찬가지로 매개변수로서 원소들이 문자열들이 문자행렬을 입력해서 각 행의 모든 원소들을 결합하여 한 문자열을 만든다. 단, 함수 ROWCATC가 함수 ROWCAT와 다른 점은 각 행의 모든 원소들을 결합해서 한 문자열을 만들 때 빈공간(blank)은 제거한다는 것이다. 이 함수에 대한 자세한 내용은 [IML6.03] (pp. 298~299)을 참조하라.

(11) SUBSTR(행렬, 위치, 길이) 또는 SUBSTR(행렬, 위치)

함수 SUBSTR는 첫 번째 매개변수로서 원소들이 문자열인 문자행렬을 입력해서 이 행렬의 (i, j) 원소의 문자열에서 두 번째 매개변수인 위치에 있는 문자부터 세 번째 매개변수인 길이 만큼의 문자들을 (i, j) 원소로 하는 행렬을 출력한다. 이 함수에 대한 자세한 내용은 [IML6.03] (pp. 306~307)을 참조하라.

8) 집합함수

(1) SETDIF(행렬 1, 행렬 2)

함수 SETDIF는 첫 번째 매개변수인 행렬 1의 원소들의 집합에서 두 번째 매개변수인 행렬 2의 원소들의 집합을 뺀 차집합에서 서로 다른 원소들을 오름차순으로 배열한 행벡터를 출력한다.

(2) UNION(행렬 1, 행렬 2, ...) 또는 UNIQUE(행렬 1, 행렬 2, ...)

함수 UNION과 함수 UNIQUE는 매개변수들인 행렬들의 원소들의 합집합에서 서로 다른 원소들을 오름차순으로 배열한 행벡터를 출력한다. 매개변수로 사용할 수 있는 행렬들의 개수는 최대 15이다.

(3) XSECT(행렬 1, 행렬 2, ...)

함수 XSECT는 매개변수들인 행렬들의 원소들의 교집합(intersect)에서 서로 다른 원소들을 오름차순으로 배열한 행벡터를 출력한다. 매개변수로 사용할 수 있는 행렬들의 개수는 최대 15이다.

집합함수들의 예제로 다음 SAS/IML 프로그램을 수행해 보자(이 SAS/IML 프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-21.PRG에 수록되어 있다).

```

/ *                               */
/ *          EXAM3-21.PRG         */
/ *                               */
/ *          SET FUNCTION         */
/ *                               */
OPTIONS PAGESIZE=60 LINESIZE=72;
PROC IML;
RESET NOCENTER NOLOG NOPRINT FW=5;
M1 = {9 8 7, 6 5 5, 5 2 1};
M2 = {-3 -2 -1, 0 1 2, 3 4 5};
PRINT, M1 " " M2;
A1 = SETDIF(M1, M2);
PRINT, 'SETDIF FUNCTION', A1;
A2 = UNION(M1, M2);
PRINT, 'UNION FUNCTION', A2;
A3 = UNIQUE(M1, M2);
PRINT, 'UNIQUE FUNCTION', A3;
A4 = XSECT(M1, M2);
PRINT, 'XSECT FUNCTION', A4;
QUIT;

```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.21과 같은 결과가 나타난다.

```

SAS
7:46 Saturday, March 5, 1994 23

M1          M2
  9   8   7   -3   -2   -1
  6   5   5    0    1    2
  5   2   1    3    4    5

SETDIF FUNCTION

A1
  6   7   8   9

UNION FUNCTION

A2
-3   -2   -1   0   1   2   3   4   5   6   7
:  8   9

UNIQUE FUNCTION

A3
-3   -2   -1   0   1   2   3   4   5   6   7
:  8   9

XSECT FUNCTION

A4
  1   2   5

Exiting IML.
    
```

그림 3.21 집합함수들을 실행한 결과

9) 접촉조정함수

(1) VALUE(문자)

함수 VALUE의 매개변수로는 행렬의 이름을 나타내는 한 문자(a single character containing the name of a matrix)를 사용한다. 이 문자는 행렬의 이름을 나타내고 함수 VALUE는 이 행렬을 출력한다.

10) 선형대수함수

(1) DET(행렬)

함수 DET는 매개변수로서 행렬을 입력해서 이 행렬의 행렬식을 출력한다.

(2) ECHELON(행렬)

함수 ECHELON은 매개변수로서 행렬을 입력해서 이 행렬의 간소화된 사다리꼴형태(row reduced echelon form)를 만든다.

(3) EIGVAL(대칭행렬)

함수 EIGVAL은 매개변수로서 대칭행렬을 입력해서 이 대칭행렬의 고유값들을 구한 다음 오름차순으로 배열해서 열벡터로 출력한다.

(4) EIGVEC(대칭행렬)

함수 EIGVEC은 매개변수로서 대칭행렬을 입력해서 이 대칭행렬의 정규직교 고유벡터들(orthonormal eigenvectors)을 구하여 이 고유벡터들을 열벡터로 하는 행렬을 출력한다.

(5) GINV(행렬)

함수 GINV는 매개변수로서 행렬을 입력해서 이 행렬의 Moore-Penrose 일반역행렬(Moore-Penrose's generalized inverse matrix)을 출력한다. 이 함수에 대한 자세한 내용은 [IML6.03](p. 247)을 참조하라.

(6) HALF(행렬)

매개변수로서 비음정치행렬(nonnegative definite matrix)을 입력해서 이 행렬의 Cholesky 분해를 출력한다. 이 함수는 함수 ROOT와 동일하다.

(7) HERMITE(행렬)

함수 HERMITE는 매개변수로 행렬을 입력해서 이 행렬의 Hermite 정규형태(Hermite normal form)를 출력한다. 이 함수에 대한 자세한 내용은 [IML6.03](p. 255)을 참조하라.

(8) HOMOGEN(행렬)

함수 HOMOGEN은 매개변수로 행렬 A 를 입력해서 다음 방정식의 해 \mathbf{x} 를 출력한다.

$$A\mathbf{x} = 0$$

이 함수에 대한 자세한 내용은 [IML6.03](pp. 255~256)을 참조하라.

(9) INV(행렬)

함수 INV는 매개변수로서 행렬을 입력해서 이 행렬의 역행렬을 출력한다.

(10) INVUPDT(행렬, 벡터, 스칼라)

함수 INVUPDT는 어떤 역행렬을 최신화(update)한다. 다음 스테이트먼트를 살펴보자.

$$R = \text{INVUPDT}(A, \mathbf{x}, w);$$

여기서 A 는 $n \times n$ 양정치행렬(positive definite matrix)이고, \mathbf{x} 는 $n \times 1$ 행렬 또는 $1 \times n$ 행렬이며 w 는 스칼라이다. 만일, w 가 주어지지 않으면 디폴트는 1이다. 이 스테이트먼트를 수행해서 얻은 행렬 R 은 다음과 같다.

$$R = A - wA\mathbf{x}(1 + w\mathbf{x}'A\mathbf{x})^{-1}\mathbf{x}'A^{-1}\mathbf{x}'A$$

이 함수에 대한 자세한 내용은 [IML6.03](p. 262)을 참조하라.

(11) ORPOL(벡터, 최대차수, 가중벡터), ORPOL(벡터, 최대차수) 또는 ORPOL(벡터)

함수 ORPOL은 첫 번째 매개변수인 벡터에 대해서 직교다항식들(orthogonal polynomials)을 열벡터들로 만들어 이 열벡터들을 행렬로 출력한다. 생성하는 직교다항식들의

개수는 두 번째 매개변수인 최대차수에 1을 더한 것이고, 최대차수가 지정되지 않은 경우에는 모든 가능한 직교다항식들을 만든다. 세 번째 매개변수인 가중벡터는 행벡터, 또는 열벡터로서 모든 원소들이 양수 또는 0이어야 한다. 출력되는 행렬을 P , 가중벡터를 w 라 할 때 다음 식이 만족된다.

$$P' \text{diag}(w) P = I$$

(12) ROOT(행렬)

함수 ROOT는 매개변수로서 비음정치행렬(nonnegative definite matrix)을 입력해서 이 행렬의 Cholesky분해를 출력한다. 즉, 함수 ROOT의 입력행렬을 A 라 하면 다음 식을 만족하는 상삼각행렬(upper triangular matrix)을 출력한다.

$$A = U'U$$

(13) SOLVE(행렬 1, 행렬 2)

함수 SOLVE는 선형연립방정식의 해를 제공한다. 다음 방정식을 풀기로 하자.

$$AX = B$$

함수 SOLVE를 사용해서 이 연립방정식의 해를 구하기 위해서는 다음과 같은 스테이트먼트를 사용한다.

$$X = \text{SOLVE}(A, B)$$

(14) SWEEP(행렬, 첨자벡터) 또는 SWEEP(행렬)

회귀분석, 다변량분석, 시계열분석 등 통계학의 여러 분야에서 이론적인 전개나 실제적인 계산에 널리 사용되는 개념의 하나가 쓸기연산자(sweep operator)이다. 이 쓸기연산자에 대한 자세한 내용은 Kennedy & Gentle(1980, pp. 288~294)과 Newton(1988, pp. 361~363)을 참조하라. 일부 사람들이 알고 있듯이 쓸기연산자를 이용한 기법이 최근에 발표된 것이 아니라 선형대수학에서는 이미 경계기법(boardering method)으로 널리 알려진 것이다. 요약해서 설명하면 쓸기연산자는 역행렬이나 선형연립방정식의 해를 계산하는 연산자라고 할 수 있다. 함수 SWEEP는 첨자벡터에 주축원소들(pivot elements)을 기준으로 주어진 행렬을 쓸기(sweeping)를 하여 새로운 행렬을 만든다. 함수 SWEEP에 관한 자세한 내용은 [IML6.03](pp. 310~311)을 참조하라.

선형대수함수들의 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-22.PRG에 수록되어 있다).

```

/*                                     */
/*                                     EXAM3-22.PRG                             */
/*                                     */

```

```

/*          LINEAR ALGEBRAIC FUNCTIONS          */
/*          */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOCENTER NOLOG NOPRINT FW=7;
M1 = { 10 -2 3, -2 40 -5, 3 -5 60};
M2 = { 3 -2 3, -2 3 -2, 3 -2 3};
PRINT, 'M1 = ' M1 " " 'M2 = 'M2;
A1 = DET(M2);
PRINT, 'DET FUNCTION', A1;
A2 = ECHELON(M1);
PRINT, 'ECHELON FUNCTION', A2;
A3 = EIGVAL(M1);
PRINT, 'EIGVAL FUNCTION', A3;
A4 = EIGVEC(M2);
PRINT, 'EIGVEC FUNCTION', A4;
A5 = GINV(M2);
PRINT, 'GINV FUNCTION', A5;
A6 = HALF(M2);
PRINT, 'HALF FUNCTION', A6;
A7 = HERMITE(M2);
PRINT, 'HERMITE FUNCTION', A7;
A8 = HOMOGEN(M2);
PRINT, 'HOMOGEN FUNCTION', A8;
A9 = INV(M1);
PRINT, 'INV FUNCTION', A9;
X = {3, 2, 1};
W = 1.23;
A10 = INVUPDT(M1, X, W);
PRINT, 'INVUPDT FUNCTION', X " " W " " A10;
A11 = ORPOL(X);
PRINT, 'ORPOL FUNCTION', X " " A11;
A12 = ROOT(M2);
PRINT, 'ROOT FUNCTION', A12;
A13 = SOLVE(M1, M2);
PRINT, 'SOLVE FUNCTION', A13;
A14 = SWEEP(M1, X);
PRINT, 'SWEEP FUNCTION', A14;
QUIT;

```

```

SAS
15:45 Monday, February 21, 1994 1

M1 =      M1      M2
      10      -2      3      M2 =      3      -2      3
      -2      40     -5      M2 =     -2      3      -2
      3      -5      60      M2 =      3      -2      3

DET FUNCTION
      A1
      0

ECHELON FUNCTION
      A2
      1      0      0
      0      1      0
      0      0      1

EIGVAL FUNCTION
      A3
      61.4028
      38.8735
      972375

EIGVEC FUNCTION
      A4
      0.60591  0.36451  -0.7071
      -0.5155  0.85689  508E-18
      0.60591  0.36451  0.70711

GINV FUNCTION
      A5
      0.15      0.2      0.15
      0.2      0.6      0.2
      0.15      0.2      0.15

HALF FUNCTION
      A6
      1.73205  -1.1547  1.73205
      0      1.29099  -86E-18
      0      0      0

HERMITE FUNCTION
      A7
      1      0      1
      0      1      0
      0      0      0

HOMOGEN FUNCTION
      A8
      -0.7071
      -12E-17
      0.70711

INV FUNCTION
      A9
      0.10233  0.00452  -0.0047
      0.00452  0.02546  0.0019
      -0.0047  0.0019  0.01706

INVUPDT FUNCTION
      X      W      A10
      3      1.23  7.04719  -9.0257  -3.0075
    
```

```

      2          -9.0257   23.2838   -19.294
      1          -3.0075   -19.294   47.7779

ORPOL FUNCTION

      X          A11
      3          0.57735   0.70711   0.40825
      2          0.57735   0          -0.8165
      1          0.57735  -0.7071   0.40825

ROOT FUNCTION

      A12
      1.73205   -1.1547   1.73205
      0          1.29099  -86E-18
      0          0          0

SOLVE FUNCTION

      A13
      0.28371   -0.1816   0.28371
      -0.0317   0.06355  -0.0317
      0.03318   -0.019   0.03318

SWEEP FUNCTION

      A14
      0.10233   0.00452  -0.0047
      0.00452   0.02546   0.0019
      -0.0047   0.0019   0.01706

Exiting IML.

```

그림 3.22 선형대수함수들을 실행한 결과

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.22와 같은 결과가 나타난다.

11) 통계함수

(1) DESIGN(열벡터)

함수 DESIGN은 매개변수로 열벡터를 입력해서 계획행렬(design matrix)을 만든다. 이 열벡터를 \mathbf{c} , 계획행렬을 A 라고 하자. 계획행렬 A 의 행의 개수는 열벡터 \mathbf{c} 의 행의 개수와 같고, 열의 개수는 열벡터 \mathbf{c} 의 서로 다른 원소들의 개수와 같다. 계획행렬 A 의 제 i 행에서 1인 원소가 단 한 개 있고 나머지 원소들은 0이다. 열벡터 \mathbf{c} 의 제 i 행의 원소 c_i 가 계획행렬 A 의 제 i 행에서 1을 갖는 열의 번호이다. 즉, 다음 식이 성립한다.

$$a_{ij} = \begin{cases} 1 & (j = c_i) \\ 0 & (j \neq c_i) \end{cases}$$

이 함수에 대한 자세한 내용은 [IML6.03] (p. 236)을 참조하라.

(2) DESIGNF(열벡터)

함수 DESIGNF는 함수 DESIGN과 마찬가지로 매개변수로 열벡터를 입력해서 계획행렬을 만든다. 함수 DESIGNF가 함수 DESIGN과 다른 점은 만들어지는 계획행렬이 완전랭크(full rank)를 갖도록 한다는 것이다. 즉, 함수 DESIGNF로 만들어지는 계획행렬은

함수 DESIGN으로 만들어지는 계획행렬의 모든 열에서 마지막 열을 뺀 다음에 마지막 열을 제거한 행렬이다. 이 함수에 대한 자세한 내용은 [IML6.03] (p. 236)을 참조하라.

통계함수들의 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-23.PRG에 수록되어 있다).

```

/* */
/*          EXAM3-23.PRG          */
/* */
/*          STATISTICAL FUNCTIONS  */
/* */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOCENTER NOLOG NOPRINT;
IND = {2, 4, 3, 1, 1, 2, 3};
A1 = DESIGN(IND);
PRINT, 'IND = ' IND " " A1;
A2 = DESIGNF(IND);
PRINT, 'IND = ' IND " " A2,;
QUIT;

```

이 SAS/IML프로그램을 수행하면 OUTPUT윈도우에 그림 3.23과 같은 결과가 나타난다.

SAS		14:39 Monday, February 21, 1994 59			
IND =	IND	A1			
	2	0	1	0	0
	4	0	0	0	1
	3	0	0	1	0
	1	1	0	0	0
	1	1	0	0	0
	2	0	1	0	0
	3	0	0	1	0
IND =	IND	A2			
	2	0	1	0	
	4	-1	-1	-1	
	3	0	0	1	
	1	1	0	0	
	1	1	0	0	
	2	0	1	0	
	3	0	0	1	
Exiting IML.					

그림 3.23 통계함수들을 실행한 결과

12) 시계열함수

SAS/IML에서 시계열분석에 쓰이는 시계열함수로는 BTRAN, COVLAG, FFT, HAN-

KEL, IFFT, PRODUCT, RATIO, TOEPLITZ 등이 있다. 이 함수들을 이해하기 위해서는 시계열분석을 어느 정도 이해하는 것이 필수적이므로 본서에서는 각 함수가 필요한 때 해당하는 함수를 설명하기로 한다.

3.4.3 서브루틴

SAS/IML에서 행렬연산을 하는 다른 방법은 서브루틴(subroutine)을 사용하는 것이다. 서브루틴을 호출하기 위해서는 CALL스테이트먼트를 사용한다. CALL스테이트먼트의 사용양식은 다음과 같다.

```
CALL 서브루틴이름(매개변수 1, 매개변수 2, ...);
CALL 서브루틴이름;
```

여러 매개변수들을 사용할 때는 이들 사이에 쉼표를 사용한다. 만일, 입력모수들과 출력결과가 동시에 매개변수들로 지정되는 경우에는 항상 출력결과를 먼저 지정하고 다음에 입력모수들을 지정한다. CALL스테이트먼트를 사용하는 예제로 다음과 같은 스테이트먼트들을 생각해 보자.

```
X = {1 2 3, 2 4 5, 3 5 6};
CALL EIGEN(EVAL, EVEC, X);
```

서브루틴 EIGEN은 대칭행렬의 고유값들과 고유벡터들을 계산하기 위한 것이다. 이 스테이트먼트들을 실행하면 대칭행렬 X의 고유값들이 열벡터 EVAL에, 고유벡터들은 행렬 EVEC에 저장된다.

많은 독자들이 서브루틴의 개념을 어렵듯이 알고 있으리라 생각한다. 다른 컴퓨터 프로그래밍언어에서와 마찬가지로 SAS/IML의 서브루틴들도 SAS/IML의 함수들과는 달리 한 번 불리어져서 하나가 아닌 두 개 이상의 행렬들을 출력하는 것이 일반적이다. 즉, SAS/IML의 서브루틴들은 SAS/IML의 함수들보다 약간 더 복잡한 문제들을 다루는 것들이라고 생각하면 된다.

[IML6.03] (pp. 202~209)에서는 서브루틴들을 문자조작서브루틴(character manipulation subroutine), 접촉조정서브루틴(interface subroutine), 선형대수서브루틴(linear algebraic subroutine), 통계서브루틴(statistical subroutine), 시계열서브루틴(time series subroutine), 그래픽서브루틴(graphics subroutine), 콘트롤프로그램서브루틴(control program subroutine) 등으로 구분한다. 이 소절에서는 이 분류에 따라 서브루틴들을 간략히 설명하고자 한다. 각 서브루틴에 대한 자세한 설명은 [IML6.03] (pp. 222~318)을 참조하라.

1) 문자조작서브루틴

(1) CHANGE(행렬, 옛 문자열, 새 문자열, 반복 횟수) 또는 CHANGE(행렬, 옛 문자열, 새 문자열)

서브루틴 CHANGE는 첫 번째 매개변수인 문자행렬의 원소에 있는 두 번째 매개변수인 옛 문자열을 세 번째 매개변수인 새 문자열로 교환한다. 문자열이 교환되는 횟수가 네 번째 매개변수에 지정된 반복 횟수이다. 네 번째 매개변수에 반복 횟수가 지정되지 않으면 반복 횟수를 1로 한다. 이 서브루틴에 관한 자세한 내용은 [IML6.03] (p. 229)을 참조하라.

2) 접촉조정서브루틴

(1) SOUND(주파수, 시간) 또는 SOUND(주파수)

서브루틴 SOUND는 첫 번째 매개변수에 지정된 주파수(frequency in hertz)를 갖는 음(tone)을 두 번째 매개변수에 지정된 시간(duration) 동안 발생시킨다. 만일, 두 번째 매개변수에 시간이 지정되지 않으면 0.25초 동안 음을 발생시킨다. 이 서브루틴에 관한 자세한 내용은 [IML6.03] (p. 304)을 참조하라.

(2) VALSET(단일문자, 매개변수)

서브루틴 VALSET은 첫 번째 매개변수에 지정된 단일문자열(single character string)이 나타내는 행렬을 찾아서 두 번째 매개변수에 지정된 값을 이 행렬에 복사한다. 이 서브루틴에 관한 자세한 내용은 [IML6.03] (p. 315)을 참조하라.

3) 선형대수서브루틴

(1) EIGEN(열벡터, 행렬, 대칭행렬)

서브루틴 EIGEN은 세 번째 매개변수에 할당되는 대칭행렬의 고유값들과 고유벡터들을 계산한 뒤, 이 고유값들을 내림차순으로 배열한 열벡터를 만들어서 첫 번째 매개변수에 지정된 열벡터로 출력시키고, 각 고유값에 해당하는 고유벡터를 계산한 다음 이 고유벡터들을 가로결합시켜 행렬로 만들어서 두 번째 매개변수에 지정된 행렬로 출력한다. 이 서브루틴에 관한 자세한 내용은 [IML6.03] (pp. 241~242)을 참조하라.

(2) GENEIG(열벡터, 행렬, 대칭행렬, 양정치행렬)

다음과 같은 일반화된 고유문제(generalized eigenproblem)를 살펴보자.

$$AE = BE \text{diag}(m)$$

여기서 A 는 대칭행렬, B 는 양정치행렬(positive definite matrix)로서 주어진 행렬들이라고 가정하자. 이 방정식의 해인 열벡터 m 의 원소들을 일반화된 고유값들, 행렬 E 의 열들을 일반화된 고유벡터들이라고 한다. 이 일반화된 고유문제는 다음과 같은 서브루틴 GENEIG을 사용해서 해결할 수 있다.

CALL GENEIG(M, E, A, B):

즉, 서브루틴 GENEIG는 세 번째와 네 번째 매개변수들에 해당하는 일반화된 고유문제를 풀어서 고유값들과 고유벡터들을 계산한 뒤, 이 고유값들을 내림차순으로 배열한 열벡터를 첫 번째 매개변수로, 각 고유값에 해당하는 고유벡터를 열벡터로 계산한 다음 이 고유벡터들을 가로결합시킨 행렬을 두 번째 매개변수로 출력한다. 이 서브루틴에 관한 자세한 내용은 [IML6.03] (p. 246)을 참조하라.

(3) GSORTH(행렬 1, 상삼각행렬, 스칼라, 행렬 2)

주어진 행렬 A 를 다음과 같이 분해하는 것을 Gram-Schmidt의 정규직교분해(orthonormal decomposition)라고 한다.

$$A = PT$$

여기서 P 는 열정규직교행렬(column-orthonormal matrix)이고, T 는 상삼각행렬이다. 이 Gram-Schmidt분해를 하기 위해서는 다음과 같이 서브루틴 GSORTH를 사용한다.

CALL GSORTH(P, T, LINDEP, A):

즉, 서브루틴 GSORTH는 네 번째 매개변수에 할당된 행렬을 Gram-Schmidt분해를 한다. 만일, 이 행렬이 완전랭크(full rank)를 가지면 세 번째 매개변수에는 0이 할당되고, 불완전랭크를 가지면 세 번째 매개변수에는 1이 할당된다. 이 서브루틴에 관한 자세한 내용은 [IML6.03] (p. 251)을 참조하라.

(4) IPF(FIT, STATUS, DIM, TABLE, CONFIG, INITAB, MOD)

서브루틴 IPF는 분할표(contingency table)에서 주변합들(marginal totals)을 구할 때 반복비교적합(Iterative Proportional Fit)을 수행한다. 이 서브루틴에 관한 자세한 내용은 [IML6.03] (pp. 263~265)을 참조하라.

(5) LCP(코드, 열벡터 1, 열벡터 2, 행렬, 열벡터 3, 상수):

다음과 같은 선형보완문제(linear complementarity problem)를 살펴보자.

$$w = Mz + q, \quad w'z = 0 \quad (w \geq 0, z \geq 0)$$

여기서 M 은 $m \times m$ 행렬, q 는 $m \times 1$ 행렬로서 주어진 행렬들이고, w 와 z 는 $m \times 1$ 행렬들로서 미지의 벡터들(unknown vectors)이다. 이 방정식에서 w 와 z 를 구하기 위해서는 다음과 같이 서브루틴 LCP를 사용한다.

$$\text{EPSILON} = 1.0\text{E-}9;$$

CALL LCP(RC, W, Z, M, Q, EPSILON):

여기서 어떤 스칼라의 절대값이 여섯 번째 매개변수 EPSILON에 지정되는 상수보다 작

으면 이 스칼라를 0으로 간주한다. 또한 첫 번째 매개변수 RC는 코드(a scalar return code)로서 이 코드의 값이 0이면 이 방정식의 해를 찾은 것, 1이면 해가 존재하지 않는 것, 5이면 해가 수치적으로 불안정한(numerically unstable) 것, 6이면 이 서브루틴이 충분한 메모리를 확보할 수 없는 것을 뜻한다.

(6) LP(코드, 열벡터 1, 열벡터 2, 행렬, 열벡터 3, 행벡터, 열벡터 4, 열벡터 5, 열벡터 6):
다음과 같은 선형계획문제(linear programming problem)를 살펴보자.

$$\begin{aligned} & \max(0, \dots, 1, 0, \dots, 0)\mathbf{x} \\ & \text{subject to } A\mathbf{x} = \mathbf{b}, \mathbf{l} \leq \mathbf{x} \leq \mathbf{u} \end{aligned}$$

여기서 A 는 $m \times n$ 행렬, \mathbf{b} 는 $m \times 1$ 행렬, \mathbf{u} 와 \mathbf{l} 은 $m \times 1$ 행렬들로서 주어진 행렬들이고, \mathbf{x} 는 $n \times 1$ 행렬들로서 미지의 벡터들(unknown vectors)이다. 이 선형계획 문제에서 최적의 \mathbf{x} 를 구하기 위해서는 다음과 같이 서브루틴 LP를 사용한다.

CALL LP(RC, X, DUAL, A, B, U, L);

이 서브루틴에 관한 자세한 내용은 [IML6.03] (pp. 271~274)을 참조하라.

(7) MARG(LOCMAR, MARGINAL, DIM, TABLE, CONFIG):

서브루틴 MARG는 다원분할표(multiway contingency table)에서 주변합들(marginal totals)을 출력한다. 이 서브루틴에 관한 자세한 내용은 [IML6.03] (pp. 274~275)을 참조하라.

(8) SVD(직교행렬 1, 특이벡터, 직교행렬 2, 행렬)

주어진 $m \times n$ 행렬 A 를 다음과 같이 분해하는 것을 특이값분해(singular value decomposition)라고 한다(단, $m \geq n$).

$$A = U \text{diag}(Q) V'$$

여기서 U 는 $m \times n$ 행렬, V 는 $n \times n$ 행렬로서 다음 식들을 만족한다.

$$U'U = V'V = VV' = I$$

또한, Q 는 $n \times 1$ 행렬로서 이 행렬의 각 원소는 행렬 A 의 특이값들(singular values)이다. 이 특이값분해를 하기 위해서는 다음과 같이 서브루틴 SVD를 사용한다.

CALL SVD(U, Q, V, A);

이 서브루틴에 관한 자세한 내용은 [IML6.03] (p. 310)을 참조하라.

선형대수서브루틴들의 예제로 다음 SAS/IML 프로그램을 수행해 보자(이 SAS/IML 프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-24.PRG에 수록되어 있다).

```

/* */
/*          EXAM3-24.PRG          */
/* */
/*          LINEAR ALGEBRAIC SUBROUTINES          */
/* */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOCENTER NOLOG NOPRINT;
M1 = { 10 2 3 1 1,
       2 12 1 2 1,
       3 1 11 1 -1,
       1 2 1 9 1,
       1 1 -1 1 15};
M2 = { 12 1 -1 2 1,
       1 14 1 -1 1,
       -1 1 16 -1 1,
       2 -1 -1 12 -1,
       1 1 1 -1 11};
PRINT,, 'M1 = 'M1, 'M2 = 'M2;
PRINT,, "*** MODULE EIGEN ***", ;
CALL EIGEN(EVAL, EVEC, M1);
PRINT, EVAL " " EVEC;
PRINT,, "*** MODULE GENEIG ***", ;
CALL GENEIG(M, E, M1, M2);
PRINT, M " " E;
PRINT,, "*** MODULE GSORTH ***", ;
CALL GSORTH(P, T, LINDEP, M1);
PRINT, P " " T, LINDEP;
PRINT,, "*** MODULE LCP ***", ;
Q = {3, 2, 1, 0, 1};
CALL LCP(RC, W, Z, M2, Q, 1.0E-9);
PRINT, W " " Z, RC;
PRINT,, "*** MODULE SVD ***", ;
M3 = { 1 2 3, 4 5 7, 3 2 1, -1 -9 -3, 5 -2 6};
PRINT, 'M3 = 'M3;
CALL SVD(U, Q, V, M3);
PRINT, U " " Q " " V,;
QUIT;

```

```

SAS
14:39 Monday, February 21, 1994 66

M1 =
      M1
10    2    3    1    1
      2    12   1    2    1
      3    1    11   1   -1
      1    2    1    9    1
      1    1   -1   1    15

M2 =
      M2
12    1   -1    2    1
      1   14    1   -1    1
     -1    1   16   -1    1
      2   -1   -1   12   -1
      1    1    1   -1   11

*** MODULE EIGEN ***
      EVAL      EVEC
16.561367  0.4192823  0.3270494  0.3216068  -0.322773  -0.713882
14.782719  0.5145935  0.2591637  -0.7161  -0.312666  0.2397271
10.591529  0.2612374  0.5480092  0.5247791  0.1289698  0.5825931
 8.0537327  0.3092815  0.1172415  -0.207488  0.883815  -0.257719
 7.0106527  0.6288846  -0.715411  0.2555897  -0.01719  0.1645275

*** MODULE GENEIG ***
      M      E
1.4923532  -0.076387  0.142012  0.19171  -0.08292  -0.134591
1.1092845  0.017098  0.14242  -0.158991  -0.153148  0.0612947
 0.943859  -0.066665  0.1209976  0.0748391  0.1186037  0.1579026
 0.6636627  0.086048  0.125531  -0.137469  0.182813  -0.109466
 0.4327872  0.2894334  0.0076922  0.0889779  -0.003562  0.041473

*** MODULE GSORTH ***
      P
0.9325048  -0.20415  -0.267204  -0.060531  -0.117026
 0.186501  0.9678241  -0.028215  -0.156541  -0.056858
 0.2797514  -0.026464  0.948854  -0.050321  0.134843
 0.0932505  0.1361003  0.0422219  0.9805128  -0.097944
 0.0932505  0.0491473  -0.160309  0.0888448  0.9773927

      T
10.723805  4.662524  6.0612813  2.517763  2.331262
 0  11.500473  0.1512225  2.9790836  1.6634477
 0  0  9.8100969  0.8449076  -3.606691
 0  0  0  8.4895266  2.1464337
 0  0  0  0  14.254219

      LINDEP
0

*** MODULE LCP ***
      W      Z
3      0
2      0
1      0
0      0
1      0

      RC
0

*** MODULE SVD ***
M3 =
      M3
1    2    3
4    5    7
3    2    1
-1   -9   -3
5    -2    6

      U      Q
0.2588702  0.023913  -0.354126  14.115291

```

0.6678207	0.1150033	-0.211455	8.3611903
0.2232568	0.0282148	0.9097654	2.2020594
-0.564267	0.6278418	-0.031438	
0.3446172	0.768909	0.0349266	
V			
0.417085	0.4527195	0.7880896	
0.6158233	-0.778492	0.1212907	
0.6684323	0.4347354	-0.603493	
Exiting IML.			

그림 3.24 선형대수서브루틴들을 실행한 결과

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.24와 같은 결과가 나타난다.

4) 시계열서브루틴

SAS/IML에서 시계열분석에 쓰이는 시계열서브루틴으로는 ARMACOV와 ARMA-LIK가 있다. 이 서브루틴을 이해하기 위해서는 시계열분석을 어느 정도 이해하고 있어야 한다. 따라서 이 서브루틴들은 필요할 때 설명하기로 한다.

5) 기타 서브루틴들

지금까지 설명한 서브루틴들 이외에도 그래픽(graphics)에 이용되는 서브루틴들과 콘트롤프로그램실행(control program execution)에 사용하는 서브루틴들 등이 있으나 본서를 읽는 데는 별로 필요가 없다. 관심있는 독자들은 [IML6.03] (pp. 208~209)을 참조하라.

3.4.4 SAS/IML 모듈

제 3.4.3 소절에서는 SAS/IML이 내장하고 있는 서브루틴들을 설명하였다. 이 소절에서는 사용자가 만드는 서브루틴인 모듈(module)을 작성하고 사용하는 방법을 설명하고자 한다. 모든 컴퓨터프로그래밍언어가 그러하듯이 SAS/IML이 내장하고 있는 서브루틴들은 표준적인 프로그램들이다. 따라서 SAS/IML의 사용자가 어떤 특수한 문제를 반복적으로 해결해야 할 경우에는 스스로 모듈을 만들어서 사용하는 것이 좋다.

SAS/IML의 모듈은 다음과 같은 특성을 갖는다.

- ① 나중에 수행될 스테이트먼트들을 저장한다. 즉, 스테이트먼트들의 수행을 연기한다.
- ② 여러 스테이트먼트들을 한꺼번에 수행할 수 있도록 한 묶음으로 만든다.
- ③ SAS/IML 프로그램의 여러 곳에서 이 스테이트먼트들의 묶음을 실행할 수 있게 한다.
- ④ 주프로그램과 분리된 환경(environment)을 만든다.

하나의 모듈은 START 스테이트먼트로 시작하여 FINISH 스테이트먼트로 끝나며 이 두 스테이트먼트들 사이에 있는 SAS/IML 스테이트먼트들이 모듈을 구성한다. 주프로그램

(main program) 안에서 모듈을 수행하기 위해서는 RUN스테이트먼트를 사용한다. 즉, 모듈을 만들고 수행하기 위한 SAS/IML스테이트먼트들은 다음과 같다.

```

PROC IML:
    START 모듈이름(매개변수 1, 매개변수 2, ...);
        SAS/IML스테이트먼트들
    FINISH;
    SAS/IML스테이트먼트들
    RUN 모듈이름(매개변수값 1, 매개변수값 2, ...);
    SAS/IML스테이트먼트들
QUIT;

```

모듈은 START스테이트먼트로 시작하고 FINISH스테이트먼트로 끝난다. 만일, START스테이트먼트에 모듈이름을 사용하지 않으면 모듈은 자동적으로 MAIN이라는 이름을 갖는다. RUN스테이트먼트는 START스테이트먼트에서 정의한 매개변수들에 대응하는 매개변수값들을 가지고 있어야 한다. 당연한 말이지만 어떤 모듈을 정의하는 START스테이트먼트가 이 모듈을 실행하기 위한 RUN스테이트먼트보다 뒤에 있어서는 안 된다.

모듈을 사용하는 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-25.PRG에 수록되어 있다).

```

/* */
/*          EXAM3-25.PRG          */
/* */
/*          SAS/IML MODULE 1      */
/* */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOCENTER NOLOG NOPRINT;
START ADD;
    HAP = X+Y;
    CHA = X-Y;
FINISH;
X = { 1 2, 3 4 };
Y = { 0 1, 2 3 };
RUN ADD;
PRINT, X " " Y, HAP " " CHA,;
QUIT;

```

SAS		16:32 Monday, February 21, 1994 1	
X		Y	
1	2	0	1
3	4	2	3
HAP		CHA	
1	3	1	1
5	7	1	1
Exiting IML.			

그림 3.25 SAS/IML 모듈

이 SAS/IML 프로그램의 모듈 ADD에는 매개변수들이 사용되지 않았다. 따라서 모듈 ADD에서도 직접환경(immediate environment)과 같은 변수들을 사용한다. 즉, 주프로그램과 서브루틴이 같은 변수들을 사용한다. 이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.25와 같은 결과가 나타난다.

매개변수들을 사용한 모듈의 예제로 다음 SAS/IML 프로그램을 수행해 보자(이 SAS/IML 프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-26.PRG에 수록되어 있다).

```

/* */
/*          EXAM3-26.PRG          */
/* */
/*          SAS/IML MODULE 2      */
/* */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOCENTER NOLOG NOPRINT;
  START ADD(HAP, CHA, X, Y);
    HAP = X+Y;
    CHA = X-Y;
  FINISH;
XX = {1 2, 3 4};
YY = {0 1, 2 3};
  RUN ADD(SUMM, DIFF, XX, YY);
PRINT, XX " " YY, SUMM " " DIFF,;
QUIT;

```

이 SAS/IML 프로그램의 모듈 ADD에는 매개변수들이 사용되었다. 따라서 모듈 ADD에서는 독자적인 국부적 환경(local environment)을 만들어 새로운 변수들을 사용한다. 즉, 주프로그램과 서브루틴이 다른 변수들을 사용한다. 이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.26과 같은 결과가 나타난다. 그림 3.26은 그림 3.25와 동일함을

SAS		16:32 Monday, February 21, 1994 2	
XX		YY	
1	2	0	1
3	4	2	3
SUMM		DIFF	
1	3	1	1
5	7	1	1
Exiting IML.			

그림 3.26 매개변수들을 사용한 SAS/IML모듈

알 수 있다.

한 SAS/IML프로그램에서 한 모듈을 여러 번 불러 사용할 수 있다. 예로써 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-27.PRG에 수록되어 있다).

```

/*                               */
/*          EXAM3-27.PRG          */
/*                               */
/*          SAS/IML MODULE 3     */
/*                               */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOCENTER NOLOG NOPRINT;
  START ADD(HAP, CHA, X, Y);
    HAP = X+Y;
    CHA = X-Y;
  FINISH;
XX = { 1 2 , 3 4 };
YY = { 0 1 , 2 3 };
  RUN ADD(SUMM, DIFF, XX, YY);
PRINT, XX " " YY, SUMM " " DIFF,;
  RUN ADD(SUM2, DIF2, { 33 44 }, { 55 66 });
PRINT, SUM2 " " DIF2,;
QUIT;

```

이 SAS/IML프로그램에서는 모듈 ADD를 두 번 불러 사용하였다. 이 SAS/IML프로그램을 수행하면 OUTPUT윈도우에 그림 3.27과 같은 결과가 나타난다.

한 SAS/IML프로그램에서는 여러 개의 모듈들을 정의할 수 있고, 하나의 모듈 안에서 다른 모듈을 부를 수도 있다. 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-28.PRG에 수록되어 있다).

SAS		16:32 Monday, February 21, 1994 3	
XX		YY	
1	2	0	1
3	4	2	3
SUM1		DIFF	
1	3	1	1
5	7	1	1
SUM2		DIF2	
88	110	-22	-22
Exiting IML.			

그림 3.27 한 모듈을 여러 번 사용한 경우

```

/*                               */
/*           EXAM3-28.PRG        */
/*                               */
/*           SAS/IML MODULE 4   */
/*                               */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOCENTER NOLOG PRINT;
START ADD(X, Y, HAP);
START MINUS(Z, W, M);
M = Z - W;
FINISH;
RUN MINUS(X, Y, CHA1);
RUN MINUS(CH A1, Y, CHA2);
RUN MINUS(CH A2, Y, CHA3);
HAP = CHA3 + 4*Y;
FINISH;
XX = { 1 2 , 3 4 };
YY = { 0 1 , 2 3 };
RUN ADD(XX, YY, SM);
PRINT, XX " " YY " " SM,;
QUIT;

```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.28과 같은 결과가 나타난다. 지금까지는 서브루틴 형태의 모듈을 설명했다. SAS/IML에서는 함수 형태의 모듈을 정의할 수 있는데 이를 함수모듈(function module)이라고 한다. 함수모듈 안의 RETURN 스테이트먼트에 출력시킬 함수값을 지정한다. 함수모듈의 예제로 다음 SAS/IML 프로그램을 수행해 보자(이 SAS/IML 프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-29.PRG에

```

SAS
16:32 Monday, February 21, 1994 13

XX      2 rows 2 columns
      1  2
      3  4

YY      2 rows 2 columns
      0  1
      2  3

M       2 rows 2 columns
      1  1
      1  1

M       2 rows 2 columns
      1  0
     -1 -2

M       2 rows 2 columns
      1 -1
     -3 -5

HAP     2 rows 2 columns
      1  3
      5  7

      XX      YY
      1  2      0  1
      3  4      2  3

      SM
      1  3
      5  7

Exiting IML.

```

그림 3.28 중첩된 모듈들을 사용

수록되어 있다).

```

/ *                               */
/ *           EXAM3-29.PRG        */
/ *                               */
/ *           SAS/IML MODULE 5    */
/ *                               */

OPTIONS PAGESIZE=60;
PROC IML;
RESET NOCENTER NOLOG NOPRINT;
START ADD(X, Y);
  HAP = X+Y;
  RETURN(HAP);
FINISH;
XX = { 1 2, 3 4};
YY = { 0 1, 2 3};
PM = ADD(XX, YY);

```

SAS		16:32 Monday, February 21, 1994 14		
XX			YY	
1	2	0	1	
3	4	2	3	
PM				
1	3			
5	7			
Exiting IML.				

그림 3.29 SAS/IML의 함수모듈

```
PRINT, XX " " YY, PM,;
QUIT;
```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.29와 같은 결과가 나타난다.

함수모듈도 중첩해서(nested) 사용할 수 있다. 다음 SAS/IML 프로그램을 수행해 보자 (이 SAS/IML 프로그램이 디스켓 SAS4TSA4의 파일 EXAM3-30.PRG에 수록되어 있다).

```
/* */
/*          EXAM3-30.PRG          */
/* */
/*          SAS/IML MODULE 6      */
/* */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOCENTER NOLOG NOPRINT;
START ADD(X, Y);
  HAP = X+Y;
  RETURN(HAP);
FINISH;
XX = { 1 2 , 3 4 };
YY = { 0 1 , 2 3 };
PM = ADD(XX, YY);
PRINT, XX " " YY, PM,;
  PM2 = ADD(ADD(PM, XX), ADD(PM, YY));
PRINT, PM2,;
QUIT;
```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.30과 같은 결과가 나타난다.

```

SAS
                                     16:32 Monday, February 21, 1994 15

      XX
      1      2
      3      4
      YY
      0      1
      2      3

      PM
      1      3
      5      7

      PM2
      3      9
      15     21

Exiting IML.
    
```

그림 3.30 함수모듈의 중첩사용

3.5 제어스테이트먼트

SAS/IML 프로그램에서 제어스테이트먼트들은 다른 스테이트먼트들의 실행 흐름(the flow of execution of the other statements)을 조절한다. SAS/IML의 제어스테이트먼트들은 다른 컴퓨터 프로그래밍 언어의 제어스테이트먼트들과 거의 비슷한 역할을 하며, 특히 [SAS4TSA1]의 제 4 장에서 설명한 DATA 스텝의 제어스테이트먼트들과 비슷하다.

3.5.1 조건처리와 스테이트먼트들의 그룹화

1) IF-THEN/ELSE 스테이트먼트

IF-THEN/ELSE 스테이트먼트는 여러 스테이트먼트들을 조건부로 처리하는 데 사용한다. 이 스테이트먼트의 사용양식은 다음과 같다.

```

IF 조건 THEN 스테이트먼트 1;
      ELSE 스테이트먼트 2;
    
```

SAS 키워드 IF 다음의 조건이 참(true)이면 SAS 키워드 THEN 다음의 스테이트먼트 1을 실행하고, 거짓(false)이면 SAS 키워드 ELSE 다음의 스테이트먼트 2를 실행한다. 즉, SAS/IML에서는 SAS 키워드 IF 다음에 있는 행렬의 모든 원소들이 1이면 SAS 키워드 THEN 다음의 스테이트먼트 1이 수행되고, SAS 키워드 IF 다음에 있는 행렬의 어느 한 원소라도 0이면 SAS 키워드 ELSE 다음의 스테이트먼트 2가 수행된다. 물론 ELSE 스테이

트먼트가 필요하지 않다면 ELSE스태이트먼트를 생략해도 되나 THEN스태이트먼트는 생략할 수 없다. IF-THEN/ELSE스태이트먼트는 중첩해서 사용할 수 있다.

2) DO-END스태이트먼트

DO-END스태이트먼트의 사용양식은 다음과 같다.

```
DO;
    스테이트먼트들
END;
```

여러 개의 스테이트먼트들을 한 단위로 처리하기 위해서는 이 스테이트먼트들을 DO스태이트먼트와 END스태이트먼트 사이에 놓아서 하나의 DO그룹으로 만든다. DO그룹은 일반적으로 다음과 같은 두 가지 목적으로 사용한다. 첫 번째 목적은 여러 스테이트먼트들을 한 단위로 만들어 스테이트먼트 단위의 대화형(interactive)처리를 하지 못하도록 하고 이 스테이트먼트들 모두를 한꺼번에 처리하게 하는 것이다. 두 번째 목적은 IF-THEN/ELSE스태이트먼트에서 조건이 참이냐 거짓이냐 각각에 따라서 여러 스테이트먼트들을 수행하게 하는 것이다. 이 DO-END스태이트먼트에 대해서는 [SAS4TSA1](p. 64)을 참조하라.

이 스테이트먼트들의 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-31.PRG에 수록되어 있다).

```
/* */
/*          EXAM3-31.PRG          */
/* */
/*          DO-END STATEMENT      */
/*          & IF-THEN/ELSE STATEMENT */
/* */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOCENTER NOLOG NOPRINT;
PRINT "DO-END INCLUDING IF-THEN-ELSE";
X = 1;
Y = 2;
DO;
    IF X = Y THEN PRINT, "X = Y",;
    ELSE PRINT, "X ^ = Y",;
END;
```

```

SAS
22:26 Tuesday, February 22, 1994 22

DO-END INCLUDING IF-THEN-ELSE

X = Y

IF-THEN/ELSE USING DO-END

      X      Y      Z      W
      1      2      3      0.6666667

Exiting IML.

```

그림 3.31 DO그룹과 IF-THEN/ELSE스테이트먼트

```

PRINT "IF-THEN/ ELSE USING DO-END";
DO:
  IF X=1 THEN DO:
    Z = 3;
    W = Y/Z;
    PRINT, X " " Y " " Z " " W,;
  END;
  ELSE DO:
    Z = 4;
    W = Y*Z;
    PRINT, X " " Y " " Z " " W,;
  END;
END;
QUIT;

```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.31과 같은 결과가 나타난다.

3.5.2 반복처리

DO스테이트먼트를 사용하면 다른 스테이트먼트들을 반복처리할 수 있다. 즉, DO스테이트먼트를 사용하면 어떤 조건이 만족되는 동안 여러 스테이트먼트들로 구성된 단위를 반복수행할 수 있다. SAS/IML에서 DO스테이트먼트를 사용해서 반복처리를 하는 데는 다음과 같은 네 가지 방법이 있다.

1) DO-END스테이트먼트

DO-END스테이트먼트의 사용양식은 다음과 같다.

```
DO 변수=시작값 TO 끝값;
```

```
    스테이트먼트들
```

```
END;
```

```
또는 DO 변수=시작값 TO 끝값 BY 증가값;
```

```
    스테이트먼트들
```

```
END;
```

첫 번째 사용양식에서 증가값(increment)의 디폴트는 1이다.

2) DO UNTIL-END스테이트먼트

DO UNTIL-END스테이트먼트의 사용양식은 다음과 같다.

```
DO UNTIL (조건);
```

```
    스테이트먼트들
```

```
END;
```

```
또는 DO 변수=시작값 TO 끝값 UNTIL (조건);
```

```
    스테이트먼트들
```

```
END;
```

```
또는 DO 변수=시작값 TO 끝값 BY 증가값 UNTIL (조건);
```

```
    스테이트먼트들
```

```
END;
```

DO UNTIL-END스테이트먼트는 SAS키워드 UNTIL 다음의 조건이 참이 될 때까지 반복처리를 계속한다. 두 번째 사용양식에서 증가값의 디폴트는 1이다.

3) DO WHILE-END스테이트먼트

이 DO WHILE-END스테이트먼트의 사용양식은 다음과 같다.

```
DO WHILE (조건);
```

```
    스테이트먼트들
```

```
END;
```

```
또는 DO 변수=시작값 TO 끝값 WHILE (조건);
```


SAS		7:46 Saturday, March 5, 1994 25			
A. DO-END STATEMENT,					
A1	A2	A3	A4		
1	0.841471	0.5403023	1		
A1	A2	A3	A4		
3	0.14112	-0.989992	1		
A1	A2	A3	A4		
5	-0.958924	0.2836622	1		
A1	A2	A3	A4		
7	0.6569866	0.7539023	1		
A1	A2	A3	A4		
9	0.4121185	-0.91113	1		
A1	A2	A3	A4		
11	-0.99999	0.0044257	1		
B. DO-UNTIL-END STATEMENT,					
B1	B2	B3	B4		
1	0.841471	0.5403023	1		
B1	B2	B3	B4		
3	0.14112	-0.989992	1		
B1	B2	B3	B4		
5	-0.958924	0.2836622	1		
B1	B2	B3	B4		
7	0.6569866	0.7539023	1		
B1	B2	B3	B4		
9	0.4121185	-0.91113	1		
B1	B2	B3	B4		
11	-0.99999	0.0044257	1		
C. DO-WHILE-END STATEMENT,					
C1	C2	C3	C4		
1	0.841471	0.5403023	1		
C1	C2	C3	C4		
3	0.14112	-0.989992	1		
C1	C2	C3	C4		
5	-0.958924	0.2836622	1		
C1	C2	C3	C4		
7	0.6569866	0.7539023	1		
C1	C2	C3	C4		
9	0.4121185	-0.91113	1		
C1	C2	C3	C4		
11	-0.99999	0.0044257	1		
Exiting IML.					

그림 3.32 DO스태이트먼트들을 실행한 결과

```

PROC IML;
RESET NOCENTER NOLOG NOPRINT;
PRINT , "A. DO-END STATEMENT,";
DO A1 = 1 TO 11 BY 2;
  A2 = SIN(A1);
  A3 = COS(A1);
  A4 = A2*A2 + A3*A3;
  PRINT A1 " " A2 " " A3 " " A4;
END;
PRINT , "B. DO UNTIL-END STATEMENT,";
DO B1 = 1 TO 11 BY 2 UNTIL ( (B1 - 13) >= 0 );

```

```

B2 = SIN(B1);
B3 = COS(B1);
B4 = B2*B2 + B3*B3;
PRINT B1 " " B2 " " B3 " " B4;
END;
PRINT , "C. DO WHILE-END STATEMENT,";
DO C1 = 1 TO 11 BY 2 WHILE ( (C1 - 11) <= 11 );
  C2 = SIN(C1);
  C3 = COS(C1);
  C4 = C2*C2 + C3*C3;
  PRINT C1 " " C2 " " C3 " " C4;
END;
QUIT;

```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.32와 같은 결과가 나타난다.

3.5.3 건너뛰기

GOTO 스테이트먼트와 LINK 스테이트먼트는 주어진 레이블(the given label)을 갖는 스테이트먼트로 건너뛰기(jump)를 하도록 하고, 이 레이블을 갖는 스테이트먼트부터 수행을 계속하도록 한다. RETURN 스테이트먼트는 LINK 스테이트먼트의 바로 다음 스테이트먼트로 되돌아 가거나, 제 3.4.4 소절에서 설명했듯이 어떤 함수모듈에서 이 함수모듈을 부른 곳으로 되돌아 가도록 한다.

1) GOTO 스테이트먼트

GOTO 스테이트먼트의 사용양식은 다음과 같다.

GOTO 레이블;

GOTO 스테이트먼트는 이 스테이트먼트에서 지정된 레이블이 붙은 스테이트먼트로 SAS/IML 프로그램의 실행을 건너뛰도록 한다. 유의할 점은 목적지의 레이블 다음에 반드시 콜론을 붙여야 한다는 것이다. GOTO 스테이트먼트는 IF 스테이트먼트의 뒤를 따르는 것이 일반적이고, 대부분의 GOTO 스테이트먼트는 DO 그룹과 더불어 사용된다.

2) LINK-RETURN 스테이트먼트

이 LINK-RETURN 스테이트먼트의 사용양식은 다음과 같다.

LINK 레이블;
 스테이트먼트들
 레이블: 스테이트먼트들
 RETURN;

LINK스테이트먼트는 GOTO스테이트먼트와 마찬가지로 레이블이 표시된 스테이트먼트로 건너뛰기를 할 것을 지시한다. 그러나 LINK스테이트먼트는 GOTO스테이트먼트와 달리 지정된 레이블이 있는 곳으로 건너뛰어 스테이트먼트들을 수행하다가 RETURN스테이트먼트를 만나면 LINK스테이트먼트 다음에 있는 스테이트먼트로 되돌아 온다. 즉, LINK-RETURN스테이트먼트는 모듈과 같은 역할을 하는데, [IML6.03] (p. 268)에서는 LINK-RETURN스테이트먼트 대신에 모듈을 사용할 것을 권한다. LINK-RETURN스테이트먼트도 IF스테이트먼트의 뒤를 따르는 것이 일반적이고, 대부분의 LINK-RETURN스테이트먼트는 DO그룹과 더불어 사용한다.

건너뛰기의 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-33.PRG에 수록되어 있다).

```

/* */
/*          EXAM3-33.PRG          */
/* */
/*          JUMPING STATEMENTS    */
/* */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOCENTER NOLOG NOPRINT;
  X = 2;
  Y = 1;
PRINT, "USING GOTO STATEMENT";
DO;
  IF X > Y THEN GOTO SKIP;
  Z = LOG(Y-X);
  ZZ = Z-20;
  PRINT, Z, ZZ,;
STOP;
SKIP: ZZZ = SIN(X);
  PRINT, ZZZ,;
END;
```

```

SAS
22:26 Tuesday, February 22, 1994 14

USING GOTO STATEMENT

      ZZZ
0.9092974

USING LINK-RETURN STATEMENT

      ZZZ
0.9092974

      Z
      0

      ZZ
      -17

Exiting IML.

```

그림 3.33 건너뛰기를 실행한 결과

```

PRINT, "USING LINK-RETURN STATEMENT";
DO;
  IF X > Y THEN LINK SKIP2;
  Z = LOG(Y-X);
  ZZ = Y-20;
  PRINT, Z, ZZ,;
  STOP;
SKIP2: ZZZ = SIN(X);
  PRINT, ZZZ,;
  Y = Y+ABS(X);
  RETURN;
END;
QUIT;

```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.33과 같은 결과가 나타난다.

3.5.4 중단과 재실행

SAS/IML 프로그램의 수행을 중단시키는 스테이트먼트로는 STOP 스테이트먼트, ABORT 스테이트먼트, PAUSE 스테이트먼트, 그리고 QUIT 스테이트먼트가 있다. 이 중에서 QUIT 스테이트먼트가 실행되면 즉시 SAS/IML 환경에서 벗어나게 되나, STOP 스테이트먼트, ABORT 스테이트먼트, 그리고 PAUSE 스테이트먼트는 SAS/IML 환경에서 SAS/IML 프로그램의 나머지 스테이트먼트들을 수행한다. RESUME 스테이트먼트는 중단된 수행을 다시 시작하도록 한다.

1) STOP스태이트먼트

STOP스태이트먼트의 사용양식은 다음과 같다.

```
STOP;
```

STOP스태이트먼트는 SAS/IML프로그램의 수행을 중지시키고 직접환경(direct environment)으로 돌아가서, 사용자가 입력시키는 새로운 스타이트먼트들을 처리한다.

2) PAUSE스태이트먼트

PAUSE스태이트먼트의 사용양식은 다음과 같다.

```
PAUSE;
```

PAUSE스태이트먼트는 STOP스태이트먼트와 마찬가지로 SAS/IML프로그램의 수행을 중지시킨다. 그러나 RESUME스태이트먼트에 의해서 SAS/IML프로그램의 수행이 다시 시작될 수 있도록 수행이 중지된 곳을 기억한다.

3) ABORT스태이트먼트

ABORT스태이트먼트의 사용양식은 다음과 같다.

```
ABORT;
```

ABORT스태이트먼트는 QUIT스태이트먼트와 마찬가지로 SAS/IML프로그램의 수행을 중지시키고 열려진 모든 파일들을 닫으면서 SAS/IML환경에서 빠져 나온다. 그러나 ABORT스태이트먼트가 QUIT스태이트먼트와 다른 점은 수행가능하고(executable) 프로그램화(programmable)할 수 있다는 것이다.

4) RESUME스태이트먼트

RESUME스태이트먼트의 사용양식은 다음과 같다.

```
RESUME;
```

RESUME스태이트먼트는 PAUSE스태이트먼트에 의해서 중단된 SAS/IML프로그램의

수행을 재개하도록 한다.

3.6 일반목적명령어들

일반목적명령어들(general purpose commands)은 다음과 같은 것들이 있다.

FINISH, FREE, LOAD, REMOVE, RESET, PRINT, QUIT,
RUN, SHOW, START, STORE, MORE, MATTRIB

이 중에서 지금까지 설명되지 않은 명령어들을 이 절에서 설명하고자 한다.

3.6.1 현재상태의 파악

SAS/IML환경에서, 특히 대화형(interactive mode)으로 SAS/IML을 사용하는 경우에는 사용자가 현재 사용되고 있는 옵션들, 정의된 행렬들과 모듈들 등을 알고자 할 때 SHOW스테이트먼트를 사용한다.

1) SHOW스테이트먼트

SHOW스테이트먼트의 사용양식은 다음과 같다.

SHOW 항목들;

SHOW스테이트먼트에 사용할 수 있는 항목들(items)과 각 항목이 가지고 있는 정보는 다음과 같다.

- OPTIONS : RESET스테이트먼트에서 지정해서 현재 설정된 옵션들
- SPACE : 작업공간(workspace)의 크기와 사용내역
- MODULES : 정의된 모듈들
- ALL : OPTIONS, SPACES, DATASETS, FILES, MODULES가 포함한 모든 정보
- STORAGE : 현재 사용하고 있는 SAS/IML의 저장라이브러리에 포함된 모든 행렬들
- NAMES : 값을 가지고 있는 모든 행렬들의 이름, 크기와 속성
- ALLNAMES : NAMES가 가지고 있는 정보와 값을 가지고 있지 않은 행렬들의 이

름, 크기와 속성

- DATASETS : 현재 열려 있는 SAS데이터세트들의 이름들
- CONTENTS : 현재 사용하고 있는 SAS데이터세트의 모든 변수들의 이름들과 속성
- FILES : 현재 열려 있는 외부파일들
- WINDOWS : 현재 열려 있는 윈도우들
- PAUSE : 중지된 모듈들
- 행 렬 : 지정된 행렬의 크기와 속성
- MEMORY : 메모리에서 사용할 수 있는 가장 큰 부분(the largest chunk of the main memory)의 크기

3.6.2 메모리의 효과적 사용

메모리를 효과적으로 사용하기 위해서 FREE스테이트먼트를 사용한다.

1) FREE스테이트먼트

이 FREE스테이트먼트의 사용양식은 다음과 같다.

```
FREE 행렬들;
FREE /;
FREE / 행렬들;
```

첫 번째 사용양식의 FREE스테이트먼트는 지정된 행렬들을 메모리에서 소거한다. 두 번째 사용양식의 FREE스테이트먼트는 메모리에 있는 모든 행렬들을 소거한다. 세 번째 사용양식의 FREE스테이트먼트는 슬러시(/) 다음에 지정된 행렬들을 제외한 나머지 행렬들을 메모리에서 소거한다. FREE스테이트먼트를 이용하면 작업공간(workspace)이 넓어지므로 적극적으로 사용하기를 권한다.

FREE스테이트먼트의 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-34.PRG에 수록되어 있다).

```
/* */
/*          EXAM3-34.PRG          */
/* */
/*          FREE STATEMENT        */
/* */
OPTIONS PAGESIZE=60;
LIBNAME TSA4DATA 'A:\TSA4DATA';
```

```

SAS
8:14 Wednesday, February 23, 1994 6

W/O FREE STATEMENT
A      2 rows      2 cols num      8
B      1 rows      1 cols num      8
C      2 rows      1 cols num      8
D      1 rows      2 cols num      8
Number of symbols = 4 (includes those without values)

FREE STATEMENT 1
B      1 rows      1 cols num      8
C      2 rows      1 cols num      8
D      1 rows      2 cols num      8
Number of symbols = 4 (includes those without values)

FREE STATEMENT 2
D      1 rows      2 cols num      8
Number of symbols = 4 (includes those without values)

FREE STATEMENT 3
Number of symbols = 4 (includes those without values)

Exiting IML.

```

그림 3.34 FREE스태이트먼트를 실행한 결과

```

PROC IML;
RESET NOCENTER NOLOG NOPRINT;
A = {1 2, 3 4};
B = 5;
C = {6, 7};
D = {8 9};
PRINT , "W/O FREE STATEMENT";
SHOW NAMES;
PRINT , "FREE STATEMENT 1";
FREE A;
SHOW NAMES;
PRINT , "FREE STATEMENT 2";
FREE /D;
SHOW NAMES;
PRINT , "FREE STATEMENT 3";
FREE / ;
SHOW NAMES;
QUIT;

```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.34와 같은 결과가 나타난다.

3.6.3 SAS/IML저장라이브러리

SAS에서 데이터세트들을 디스크의 SAS데이터라이브러리에 저장시킨 뒤에 다시 사용할 수 있는 것과 마찬가지로, SAS/IML환경에서는 행렬들을 디스크의 저장라이브러리에 저장시킨 뒤 후에 다시 꺼내 사용할 수 있다. 이 저장라이브러리의 이름을 지정하기 위해서는 다음과 같은 RESET스테이트먼트를 사용한다.

```
RESET STORAGE="member";
RESET STORAGE="libref. member";
RESET STORAGE=operand;
```

첫 번째 사용양식의 RESET스테이트먼트에 사용한 member는 SAS/IML저장라이브러리의 이름이다. 두 번째 사용양식의 RESET스테이트먼트에 사용한 libref는 SAS데이터라이브러리이고 member는 SAS/IML저장라이브러리의 이름이다. SAS데이터라이브러리의 디폴트는 SASUSER이고 SAS/IML저장라이브러리의 디폴트는 IMLSTORE이다. 세 번째 사용양식의 RESET스테이트먼트에 사용한 operand는 문자행렬의 이름, 또는 괄호(parentheses) 안의 수식(expression)이다.

한 SAS데이터라이브러리는 여러 개의 SAS/IML저장라이브러리들을 가질 수 있고, 한 SAS/IML저장라이브러리는 여러 개의 행렬들을 포함할 수 있다. SAS/IML환경에서 몇 개의 SAS/IML저장라이브러리들을 정의할 수 있으나 어느 한 시점에서는 단 한 개의 SAS/IML저장라이브러리만이 열려 있어야 한다. 따라서 새로운 SAS/IML저장라이브러리를 열기 전에 반드시 열려 있는 SAS/IML저장라이브러리를 닫아야 한다. SAS/IML저장라이브러리에 관한 명령어들에는 STORE, LOAD, 그리고 REMOVE가 있다.

1) STORE스테이트먼트

이 STORE스테이트먼트의 사용양식은 다음과 같다.

```
STORE;
STORE 행렬들;
```

STORE스테이트먼트는 메모리에 있는 행렬들을 SAS/IML저장라이브러리에 복사한다. 첫 번째 사용양식처럼 행렬들을 지정하지 않으면 메모리의 모든 행렬들을 SAS/IML저장라이브러리에 복사한다. 이 STORE스테이트먼트가 실행되어도 메모리에 있는 행렬들이 지워지지 않는다. 행렬들을 메모리에서 지우기 위해서는 FREE스테이트먼트를 사용한다.

2) LOAD스테이트먼트

LOAD스테이트먼트의 사용양식은 다음과 같다.

```
LOAD;
LOAD 행렬들;
```

LOAD스테이트먼트는 SAS/IML저장라이브러리에 있는 행렬들을 메모리에 복사한다. 첫 번째 사용양식처럼 행렬들을 지정하지 않으면 SAS/IML저장라이브러리의 모든 행렬들을 메모리로 복사한다.

3) REMOVE스테이트먼트

REMOVE스테이트먼트의 사용양식은 다음과 같다.

```
REMOVE 행렬들;
```

REMOVE스테이트먼트는 SAS/IML저장라이브러리에 있는 행렬들을 지운다.

SAS/IML저장라이브러리를 사용하는 예제로 다음 SAS/IML프로그램을 수행해 보자 (이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-35.PRG에 수록되어 있다).

```

/ *                               */
/ *           EXAM3-35.PRG        */
/ *                               */
/ *           STORAGE STATEMENTS */
/ *                               */
OPTIONS PAGESIZE=60;
LIBNAME TSA4DATA 'A:\TSA4DATA';
PROC IML;
RESET NOCENTER NOLOG NOPRINT;
RESET STORAGE="TSA4DATA.MYIML";
A = {1 2, 3 4};
B = 5;
C = {6, 7};
D = {8 9};
PRINT , "W/O STORAGE STATEMENT";
SHOW NAMES STORAGE;
PRINT , "STORE STATEMENT 1";
STORE A B;
```

```

SAS
8:14 Wednesday, February 23, 1994 11

NOTE: New storage = TSA4DATA.MYIML

W/O STORAGE STATEMENT
A      2 rows      2 cols num      8
B      1 rows      1 cols num      8
C      2 rows      1 cols num      8
D      1 rows      2 cols num      8
Number of symbols = 4 (includes those without values)
Contents of storage = TSA4DATA.MYIML
A      B

STORE STATEMENT 1
A      2 rows      2 cols num      8
B      1 rows      1 cols num      8
C      2 rows      1 cols num      8
D      1 rows      2 cols num      8
Number of symbols = 4 (includes those without values)
Contents of storage = TSA4DATA.MYIML
A      B

STORE STATEMENT 2
A      2 rows      2 cols num      8
B      1 rows      1 cols num      8
C      2 rows      1 cols num      8
D      1 rows      2 cols num      8
Number of symbols = 4 (includes those without values)
Contents of storage = TSA4DATA.MYIML
A      B      C      D

LOAD STATEMENT 1
A      2 rows      2 cols num      8
B      1 rows      1 cols num      8
C      2 rows      1 cols num      8
D      1 rows      2 cols num      8
Number of symbols = 4 (includes those without values)
Contents of storage = TSA4DATA.MYIML
A      B      C      D

LOAD STATEMENT 2
A      2 rows      2 cols num      8
B      1 rows      1 cols num      8
C      2 rows      1 cols num      8
D      1 rows      2 cols num      8
Number of symbols = 4 (includes those without values)
Contents of storage = TSA4DATA.MYIML
A      B      C      D

REMOVE STATEMENT 1
A      2 rows      2 cols num      8
B      1 rows      1 cols num      8
C      2 rows      1 cols num      8
D      1 rows      2 cols num      8
Number of symbols = 4 (includes those without values)
Contents of storage = TSA4DATA.MYIML
A      B

REMOVE STATEMENT 2
A      2 rows      2 cols num      8
B      1 rows      1 cols num      8
C      2 rows      1 cols num      8
D      1 rows      2 cols num      8
Number of symbols = 4 (includes those without values)
Contents of storage = TSA4DATA.MYIML

Exiting IML.

```

그림 3.35 SAS/IML저장라이브러리의 이용

```
SHOW NAMES STORAGE;  
PRINT , "STORE STATEMENT 2";  
STORE;  
SHOW NAMES STORAGE;  
PRINT , "LOAD STATEMENT 1";  
LOAD C D;  
SHOW NAMES STORAGE;  
PRINT , "LOAD STATEMENT 2";  
LOAD;  
SHOW NAMES STORAGE;  
PRINT , "REMOVE STATEMENT 1";  
REMOVE C D;  
SHOW NAMES STORAGE;  
PRINT , "REMOVE STATEMENT 2";  
REMOVE A B;  
SHOW NAMES STORAGE;  
QUIT;
```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.35와 같은 결과가 나타난다.

3.6.4 출력의 조절

SAS/IML 환경에서 출력을 좀더 멋있게 하기 위해서는 PRINT 스테이트먼트와 MATTRIB 스테이트먼트를 잘 사용해야 한다.

1) PRINT 스테이트먼트

이 PRINT 스테이트먼트의 사용양식은 다음과 같다.

PRINT 항목들;

PRINT 스테이트먼트에 사용할 수 있는 항목들(items)과 각 항목의 기능은 다음과 같다. 한 PRINT 스테이트먼트에 여러 항목을 동시에 사용할 수 있다.

- 행렬 : 행렬의 이름
- (수식) : 괄호 안의 수식의 연산결과
- "메시지" : 인용기호 안의 메시지
- , : 항목들을 분리, 다음 줄로 넘김

- / : 다음 페이지로 넘김
- [옵션들] : 행렬이름 다음에 표시되는 옵션들로 FORMAT=, ROWNAME=과 COLNAME=

마지막 항목인 옵션들은 행렬의 출력형태를 조절하는 데 사용된다. 옵션 FORMAT=에는 출력되는 행렬의 각 원소의 포맷을 지정한다. 옵션 ROWNAME=과 옵션 COLNAME=에는 각각 행의 이름과 열의 이름을 지정한다. 행의 이름과 열의 이름은 원소들이 문자들인 행벡터, 또는 열벡터로 정의되어야 하나, SAS키워드들 ROWNAME=과 COLNAME= 다음에 따옴표로 둘러싼 문자열을 사용할 수는 없다.

PRINT스태이트먼트의 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML 프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-36.PRG에 수록되어 있다).

```

/ *                               */
/ *           EXAM3-36.PRG        */
/ *                               */
/ *           STORAGE STATEMENTS */
/ *                               */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOCENTER NOLOG NOPRINT;
IND = 1:5;
ROOTIND = SQRT(IND);
LOGIND = LOG(IND);
EXPIND = EXP(IND);
FTNIND = ROOTIND // LOGIND // EXPIND;
RNAME = { "ROOT" "LOG" "EXP" };
CNAME = { "1" "2" "3" "4" "5" };
PRINT, "FUNCTION VALUES", ,
      FTNIND[FORMAT = 8.4 ROWNAME = RNAME COLNAME = CNAME], ;
QUIT;

```

이 SAS/IML프로그램을 수행하면 OUTPUT윈도우에 그림 3.36과 같은 결과가 나타난다.

SAS		18:44 Wednesday, February 23, 1994 7				
FUNCTION VALUES						
FTNIND	1	2	3	4	5	
ROOT	1.0000	1.4142	1.7321	2.0000	2.2361	
LOG	0.0000	0.6931	1.0986	1.3863	1.6094	
EXP	2.7183	7.3891	20.0855	54.5982	148.4132	
Exiting IML.						

그림 3.36 PRINT스태이트먼트를 실행한 결과

2) MATTRIB스테이트먼트

MATTRIB스테이트먼트의 사용양식은 다음과 같다.

MATTRIB 행렬 ROWNAME = 벡터 COLNAME = 벡터;

MATTRIB스테이트먼트에는 출력할 행렬의 행의 이름과 열의 이름을 지정한다. MATTRIB스테이트먼트는 실행문이 아니라 선언문이므로 대응되는 PRINT스테이트먼트보다 앞에만 있다면 SAS/IML프로그램내의 어느 곳에 위치해도 좋다. PRINT스테이트먼트와 마찬가지로 옵션 ROWNAME = 과 옵션 COLNAME = 에는 각각 행의 이름과 열의 이름을 지정한다. 행의 이름과 열의 이름은 원소들이 문자들인 행벡터, 또는 열벡터로 정의되어야 하나, SAS키워드들 ROWNAME = 과 COLNAME = 다음에 따옴표로 둘러싼 문자열을 사용할 수는 없다.

MATTRIB스테이트먼트의 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-37.PRG에 수록되어 있다).

```

/ *                               */
/ *          EXAM3-37.PRG         */
/ *                               */
/ *          STORAGE STATEMENTS  */
/ *                               */

OPTIONS PAGESIZE=60;
PROC IML;
RESET NOCENTER NOLOG NOPRINT;
MATTRIB FTNIND ROWNAME = RNAME COLNAME = CNAME;
IND = 1:5;
ROOTIND = SQRT(IND);
LOGIND = LOG(IND);
EXPIND = EXP(IND);
    
```

SAS	18:44 Wednesday, February 23, 1994 4				
FUNCTION VALUES					
FTNIND	1	2	3	4	5
ROOT	1.0000	1.4142	1.7321	2.0000	2.2361
LOG	0.0000	0.6931	1.0986	1.3863	1.6094
EXP	2.7183	7.3891	20.0855	54.5982	148.4132
Exiting IML.					

그림 3.37 MATTRIB스테이트먼트를 실행한 결과

```
FTNIND = ROOTIND // LOGIND // EXPIND;
RNAME = { "ROOT" "LOG" "EXP" };
CNAME = { "1" "2" "3" "4" "5" };
PRINT, "FUNCTION VALUES", , FTNIND[FORMAT=8.4],;
QUIT;
```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.37과 같은 결과가 나타난다.

3.7 SAS/IML에서의 데이터처리

이 절에서는 데이터처리명령어들(data processing commands)에 대해서 설명한다. SAS/IML 환경에서 처리하는 데이터세트는 크게 두 종류로 나눌 수 있다. 하나는 SAS의 DATA 스텝에서 만들어진 SAS 데이터세트이고 다른 하나는 에스키(ASCII)로 기록된 텍스트 파일이다. 흔히들 전자를 내부파일(internal file)이라 하고 후자를 외부파일(external file)이라고 한다. 이 절에서는 SAS/IML에서 SAS 데이터세트를 처리하는 방법을 설명하고, 제 3.8 절에서는 SAS/IML에서 텍스트 파일을 처리하는 방법을 설명한다.

3.7.1 현재 SAS 데이터세트

모든 데이터처리명령들은 현재 SAS 데이터세트(current SAS dataset)에 적용된다. 따라서 각 연산명령을 내릴 때 일일이 대상이 되는 SAS 데이터세트를 지정할 필요가 없다. 현재 SAS 데이터세트에는 두 종류가 있는데 하나는 입력을 위한 현재입력 SAS 데이터세트(current input SAS dataset)이고 다른 하나는 출력을 위한 현재출력 SAS 데이터세트(current output SAS dataset)이다. SAS/IML 환경에서는 어떤 SAS 데이터세트가 열리면(open) 그 SAS 데이터세트를 현재 SAS 데이터세트로 한다. 열려 있는 SAS 데이터세트들, 현재출력 SAS 데이터세트 그리고 현재입력 SAS 데이터세트를 알고자 하면 다음과 같은 SHOW 스테이트먼트를 사용한다.

```
SHOW DATASETS;
```

3.7.2 SAS 데이터세트의 열기

모든 컴퓨터 프로그래밍 언어에서와 마찬가지로 어느 파일을 사용하기 위해서는 그 파일

을 열어야(open) 한다. SAS/IML환경에서는 어떤 SAS데이터세트가 열리면 그 SAS데이터세트를 현재SAS데이터세트로 한다. SAS/IML에서 이미 만들어져 있는 SAS데이터세트를 열기 위해서, 즉 현재입력SAS데이터세트를 만들기 위해서는 USE스테이트먼트를 사용하고, 새로운 SAS데이터세트를 만들기 위해서는, 즉 현재출력SAS데이터세트를 만들기 위해서는 CREATE스테이트먼트를 사용한다. 또한, 한 SAS데이터세트를 현재입출력SAS데이터세트(current input/output SAS dataset)로 만들기 위해서는 EDIT스테이트먼트를 사용한다. 또한, 여러 개의 SAS데이터세트들이 열려 있을 때 이 중에서 한 SAS데이터세트를 현재입력SAS데이터세트로 만들기 위해서는 SETIN스테이트먼트를 사용하고, 한 SAS데이터세트를 현재출력SAS데이터세트로 만들기 위해서는 SETOUT스테이트먼트를 사용한다.

1) USE스테이트먼트

USE스테이트먼트를 사용하면 SAS/IML환경으로 읽어들이는 SAS데이터세트가 열려져 있지 않은 경우에 SAS데이터세트를 열어서 현재입력SAS데이터세트로 한다. 또한, USE스테이트먼트는 여러 SAS데이터세트들이 열려 있을 때 어떤 SAS데이터세트를 현재입력SAS데이터세트로 만든다. USE스테이트먼트는 선언문이지 실행문이 아니다. 따라서 USE스테이트먼트는 지정된 SAS데이터세트를 사용하겠다고 선언하는 것 뿐이다. USE스테이트먼트의 사용양식은 다음과 같다.

```
USE SAS데이터세트;
USE SAS데이터세트 VAR(변수이름 1, 변수이름 2, ... );
USE SAS데이터세트 WHERE(수식);
USE SAS데이터세트 VAR(변수이름 1, 변수이름 2, ... ) WHERE(수식);
```

첫 번째 사용양식의 USE스테이트먼트는 SAS데이터세트를 열거나, 두 개 이상의 SAS데이터세트들이 열려 있을 때 그 중의 한 SAS데이터세트를 현재입력SAS데이터세트로 만들 때 사용한다. 두 번째 사용양식의 USE스테이트먼트는 어떤 SAS데이터세트를 열어서 이 SAS데이터세트가 포함하고 있는 변수들 중에서 몇 개의 변수들만을 읽어들이고자 할 때 사용한다. 읽어들이는 변수들을 VAR구문(VAR clause)에 지정한다. 즉, SAS키워드 VAR 다음의 중괄호 안에 읽어들이는 변수들을 지정한다. VAR구문에 대한 자세한 내용은 제 3.7.6 소절을 참조하라. 세 번째 사용양식의 USE스테이트먼트는 어떤 SAS데이터세트를 열어서 이 SAS데이터세트가 포함하고 있는 관찰점들 중에서 일부분만을 읽어들이고자 할 때 사용한다. 읽어들이는 관찰점들을 WHERE구문(WHERE clause)에 지정한다. WHERE구문에 대한 자세한 내용은 제 3.7.7 소절을 참조하라. 이 USE스테이트먼트의 사용양식들에서 VAR구문과 WHERE구문은 옵션들이고 이들은 어떤 순서로 사용해도 좋다.

2) CREATE스태이트먼트

CREATE스태이트먼트는 SAS/IML환경에서 새로운 SAS데이터세트를 만들기 위해서 이 SAS데이터세트를 연다. 즉, 현재출력SAS데이터세트를 연다. USE스태이트먼트와 마찬가지로 CREATE스태이트먼트는 선언문이지 실행문이 아니다. 따라서 CREATE스태이트먼트는 어떤 SAS데이터세트를 만들겠다고 선언하는 것이지 실제 SAS데이터세트를 만드는 것은 아니다. CREATE스태이트먼트의 사용양식은 다음과 같다.

```
CREATE SAS데이터세트;
CREATE SAS데이터세트 VAR(변수이름 1, 변수이름 2, ...);
CREATE SAS데이터세트 FROM 행렬 ;
CREATE SAS데이터세트 FROM 행렬 COLNAME=백터이름 1
                                ROWNAME=백터이름 2
```

첫 번째 사용양식의 CREATE스태이트먼트는 SAS데이터세트를 열거나, 두 개 이상의 SAS데이터세트들이 열려 있을 때 그 중의 한 SAS데이터세트를 현재출력SAS데이터세트로 만들 때 사용한다. 두 번째 사용양식의 CREATE스태이트먼트는 어떤 SAS데이터세트를 열어서 이 SAS데이터세트가 포함하고 있는 변수들 중에서 몇 개의 변수들만을 읽어들이고자 할 때 사용한다. 읽어들이는 변수들을 VAR구문(VAR clause)에 지정한다. 즉, SAS 키워드 VAR 다음의 중괄호 안에 읽어들이는 변수들을 지정한다. 세 번째와 네 번째 사용양식들의 CREATE스태이트먼트들은 행렬데이터를 SAS데이터세트로 만들 때 사용하는 데 이들에 대한 자세한 사용법은 제 3.7.8소절을 참조하라.

3) EDIT스태이트먼트

EDIT스태이트먼트는 SAS데이터세트가 열려져 있지 않은 경우에 SAS데이터세트를 열어서 현재입출력SAS데이터세트로 한다. 또한, 이 EDIT스태이트먼트는 여러 SAS입출력 데이터세트들이 열려 있을 때 지정된 SAS데이터세트를 현재입출력SAS데이터세트로 만든다. 따라서 EDIT스태이트먼트는 USE스태이트먼트와 CREATE스태이트먼트를 한 SAS데이터세트에 사용한 것과 같다. EDIT스태이트먼트는 선언문이지 실행문이 아니다. 즉, EDIT스태이트먼트는 지정된 SAS데이터세트를 사용하겠다고 선언하는 것 뿐이다. EDIT스태이트먼트의 사용양식은 다음과 같다.

```
EDIT SAS데이터세트;
EDIT SAS데이터세트 VAR(변수이름 1, 변수이름 2, ...);
EDIT SAS데이터세트 WHERE(수식);
EDIT SAS데이터세트 VAR(변수이름 1, 변수이름 2, ...} WHERE(수식);
```

어떤 SAS데이터세트를 읽고 이를 최신화(updating)하기 위해서는, 즉 SAS데이터세트

를 편집하기 위해서는 우선 EDIT스테이트먼트를 사용해서 이 SAS데이터세트를 열어야 한다. 어떤 SAS데이터세트를 EDIT스테이트먼트를 사용해서 열면 이 SAS데이터세트의 수록된 관찰점들을 읽고, 데이터값들을 수정하고, 새로운 관찰점들을 추가할 수 있다. 첫 번째 사용양식의 EDIT스테이트먼트는 어떤 SAS데이터세트를 열어서 입출력SAS데이터세트를 만들거나, 두 개 이상의 SAS데이터세트들이 열려 있을 때 그 중의 한 SAS데이터세트를 현재입출력SAS데이터세트로 만들 때 사용한다. 두 번째 사용양식의 EDIT스테이트먼트는 어떤 SAS데이터세트를 열어서 이 SAS데이터세트가 포함하고 있는 변수들 중에서 몇 개의 변수들만을 읽어 최신화하고자 할 때 사용한다. 읽어들일 변수들을 VAR구문에 지정한다. 즉, SAS키워드 VAR 다음의 중괄호 안에 읽어들일 변수들을 지정한다. 세 번째 사용양식의 EDIT스테이트먼트는 지정된 SAS데이터세트를 열어서 이 SAS데이터세트가 포함하고 있는 관찰점들 중에서 일부분만을 읽어들여 최신화하고자 할 때 사용한다. 읽어들일 관찰점들을 WHERE구문에 지정한다. EDIT스테이트먼트의 사용양식들에서 VAR구문과 WHERE구문은 옵션들이고 이들은 어떤 순서로 사용해도 좋다.

4) SETIN스테이트먼트

이미 USE스테이트먼트, 또는 EDIT스테이트먼트를 사용해서 연 SAS데이터세트들 중에서 한 SAS데이터세트를 현재입력SAS데이터세트로 만들기 위해서는 SETIN스테이트먼트를 사용한다. SETIN스테이트먼트의 사용양식은 다음과 같다.

SETIN SAS데이터세트;
 SETIN SAS데이터세트 NOBS 변수;
 SETIN SAS데이터세트 POINT 번호;
 SETIN SAS데이터세트 NOBS 변수 POINT 번호;

첫 번째 사용양식의 SETIN은 이미 열려져 있는 SAS데이터세트들 중에서 지정된 SAS데이터세트를 현재입력SAS데이터세트로 만든다. 두 번째 사용양식의 SETIN은 이미 열려져 있는 SAS데이터세트들 중에서 지정된 SAS데이터세트를 현재입력SAS데이터세트로 만들 뿐 아니라, 옵션 NOBS에 지정된 변수에 이 SAS데이터세트의 관찰점들의 개수를 저장한다. 세 번째 사용양식의 SETIN은 이미 열려져 있는 SAS데이터세트들 중에서 지정된 SAS데이터세트를 현재입력SAS데이터세트로 만들 뿐 아니라, 옵션 POINT에 지정된 번호의 관찰점을 현재관찰점(current observation)으로 한다. 네 번째 사용양식의 SETIN은 이미 열려져 있는 SAS데이터세트들 중에서 지정된 SAS데이터세트를 현재입출력SAS데이터세트로 만들 뿐 아니라, 옵션 NOBS에 지정된 변수에 이 SAS데이터세트의 관찰점들의 개수를 저장하고, 옵션 POINT에 지정된 번호의 관찰점을 현재관찰점으로 한다.

5) SETOUT스테이트먼트

이미 CREATE스테이트먼트, 또는 EDIT스테이트먼트를 사용해서 연 SAS데이터세트들 중에서 한 SAS데이터세트를 현재출력SAS데이터세트로 만들기 위해서는 SETOUT스테이트먼트를 사용한다. SETOUT스테이트먼트의 사용양식은 다음과 같다.

```
SETOUT SAS데이터세트;
SETOUT SAS데이터세트 NOBS 변수;
SETOUT SAS데이터세트 POINT 번호;
SETOUT SAS데이터세트 NOBS 변수 POINT 번호;
```

첫 번째 사용양식의 SETOUT은 이미 열려져 있는 SAS데이터세트들 중에서 지정된 SAS데이터세트를 현재출력SAS데이터세트로 만든다. 두 번째 사용양식의 SETOUT은 이미 열려져 있는 SAS데이터세트들 중에서 지정된 SAS데이터세트를 현재출력SAS데이터세트로 만들 뿐 아니라, 옵션 NOBS에 지정된 변수에 이 SAS데이터세트의 관찰점들의 개수를 저장한다. 세 번째 사용양식의 SETOUT은 이미 열려져 있는 SAS데이터세트들 중에서 지정된 SAS데이터세트를 현재출력SAS데이터세트로 만들 뿐 아니라, 옵션 POINT에 지정된 번호의 관찰점을 현재관찰점으로 한다. 네 번째 사용양식의 SETOUT은 이미 열려져 있는 SAS데이터세트들 중에서 지정된 SAS데이터세트를 현재출력SAS데이터세트로 만들 뿐 아니라, 옵션 NOBS에 지정된 변수에 이 SAS데이터세트의 관찰점들의 개수를 저장하고, 옵션 POINT에 지정된 번호의 관찰점을 현재관찰점으로 한다.

3.7.3 SAS데이터세트의 닫기

모든 프로그래밍언어에서와 마찬가지로 SAS/IML환경에서도 큰 프로그램을 수행할 때는 더 이상 사용하지 않을 파일들을 닫는 것이 좋다.

1) CLOSE스테이트먼트

이 CLOSE스테이트먼트는 USE스테이트먼트, CREATE스테이트먼트, 그리고 EDIT스테이트먼트에 의해서 열린 SAS데이터세트들을 닫는다. 물론 SAS/IML환경에서 벗어나면 모든 파일들이 저절로 닫히므로 구태여 CLOSE스테이트먼트를 쓰지 않아도 되는 경우가 많다. 그러나 SAS/IML환경에서 큰 프로그램을 수행할 때는 더 이상 사용하지 않을 파일들을 닫아놓는 것이 좋다. CLOSE스테이트먼트의 사용양식은 다음과 같다.

```
CLOSE SAS데이터세트들;
```

3.7.4 접근명령어들의 개관

지금까지 배운 USE스테이트먼트, CREATE스테이트먼트, 그리고 EDIT스테이트먼트는 SAS데이터세트를 여는 선언문들이지 직접 데이터값들을 SAS/IML환경으로 끌어들이는 역할을 하지는 못한다. SAS/IML환경에서 SAS데이터세트의 데이터값들에 직접 접근하는 명령어들을 접근명령어들(access commands)이라 한다. 이 접근명령어들이 데이터값들의 입력과 출력을 관장한다. 접근명령어들에는 다음과 같은 것들이 있다.

LIST, READ, REPLACE, APPEND, FIND,
DELETE, FORCE, PURGE

이 소절에서는 APPEND스테이트먼트, READ스테이트먼트, 그리고 LIST스테이트먼트를 간단히 설명하고자 한다.

1) APPEND

APPEND스테이트먼트는 CREATE스테이트먼트를 사용해서 연 SAS데이터세트에 데이터값들을 쓰기 위한 실행문이다. APPEND스테이트먼트의 가장 간단한 사용양식은 다음과 같다.

```
APPEND;
```

SAS/IML을 사용해서 새로운 SAS데이터세트를 만드는 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-38.PRG에 수록되어 있다).

```

/* */
/*          EXAM3-38.PRG          */
/* */
/*          CREATING A SAS DATASET          */
/* */
LIBNAME CBSIML "A:\IMLDATA";
PROC IML;
  RESET NOCENTER NOLOG NOPRINT PAGESIZE=60;
  IRUM = {"KIM", "PARK", "CHOI", "LEE", "AHN", "CHUNG"};
  Y = {11, 22, 33, 44, 55, 66};
  X1 = {1, 2, 3, 4, 5, 6};

```

```

X2 = { 0.72, 0.33, -0.34, 0.81, 0.12, 0.01 };
CREATE CBSIML.DS1;
APPEND;
CLOSE CBSIML.DS1;
QUIT;
PROC PRINT DATA=CBSIML.DS1;
RUN;

```

이 SAS/IML 프로그램을 수행하기 전에 드라이브 A:에 있는 디스켓에 서브디렉토리/IMLDATA를 만들어야 한다. LIBNAME 스테이트먼트에서 이 서브디렉토리를 SAS 데이터 라이브러리 CBSIML로 할당하였다. 이 SAS/IML 프로그램에서는 CREATE 스테이트먼트를 사용해서 CBSIML.DS1을 현재 출력 SAS 데이터세트로 선언했다. APPEND 스테이트먼트를 사용해서 이 SAS 데이터세트에 변수들 IRUM, Y, X1, X2의 데이터값들을 써넣는다. 이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.38과 같은 결과가 나타난다.

SAS					18:29 Saturday, February 26, 1994 6
Exiting IML.					
SAS					18:29 Saturday, February 26, 1994 7
OBS	IRUM	X1	X2	Y	
1	KIM	1	0.72	11	
2	PARK	2	0.33	22	
3	CHOI	3	-0.34	33	
4	LEE	4	0.81	44	
5	AHN	5	0.12	55	
6	CHUNG	6	0.01	66	

그림 3.38 SAS/IML을 사용한 SAS 스테이트먼트의 창조

2) READ

READ 스테이트먼트는 현재 입력 SAS 데이터세트에서 데이터값들을 읽어서 각 변수를 열벡터로 만들기 위한 실행문이다. 생성된 열벡터의 이름은 SAS 데이터세트의 변수이름과 같다. READ 스테이트먼트의 가장 간단한 사용양식은 다음과 같다.

```
READ;
```

SAS/IML을 사용해서 기존의 SAS 데이터세트를 불러들이는 예제로 다음 SAS/IML 프로그램을 수행해 보자(이 SAS/IML 프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-39.PRG에 수록되어 있다).

SAS					18:29 Saturday, February 26, 1994 8
Exiting IML.					
SAS					18:29 Saturday, February 26, 1994 9
OBS	IRUM	X1	X2	Y	
1	KIM	1	0.72	11	
2	PARK	2	0.33	22	
3	CHOI	3	-0.34	33	
4	LEE	4	0.81	44	
5	AHN	5	0.12	55	
6	CHUNG	6	0.01	66	

그림 3.39 SAS/IML을 사용해서 기존의 SAS데이터세트를 읽어들이

```

/ *                               */
/ *           EXAM3-39.PRG        */
/ *                               */
/ *           USING A SAS DATASET */
/ *                               */
LIBNAME CBSIML "A:\IMLDATA";
PROC IML;
RESET NOCENTER NOLOG NOPRINT PAGESIZE=60;
USE CBSIML.DS1;
READ;
CLOSE CBSIML.DS1;
QUIT;
PROC PRINT DATA=CBSIML.DS1;
RUN;

```

이 SAS/IML 프로그램에서는 USE 스테이트먼트를 사용해서 CBSIML.DS1을 현재 입력 SAS 데이터세트로 선언하고, READ 스테이트먼트를 사용해서 이 SAS 데이터세트의 변수들 IRUM, Y, X1, X2의 데이터값들을 읽어들이는다. 이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.39와 같은 결과가 나타난다.

3) LIST

이 LIST 스테이트먼트는 현재 SAS 데이터세트에서 데이터값들을 읽어서 지정된 윈도우에 출력한다. LIST 스테이트먼트의 가장 간단한 사용양식은 다음과 같다.

```
LIST;
```

SAS/IML을 사용해서 지정된 윈도우에 SAS 데이터세트를 출력시키는 예제로 다음 SAS/IML 프로그램을 수행해 보자(이 SAS/IML 프로그램은 디스켓 SAS4TSA4의 파일

EXAM3-40.PRG에 수록되어 있다).

```

/ *                               */
/ *           EXAM3-40.PRG        */
/ *                               */
/ *           LISTING A SAS DATASET */
/ *                               */
LIBNAME CBSIML "A:\IMLDATA";
PROC IML;
RESET NOCENTER NOLOG NOPRINT PAGESIZE=60;
USE CBSIML.DS1;
LIST;
CLOSE CBSIML.DS1;
QUIT;

```

이 SAS/IML프로그램에서는 USE스테이트먼트를 사용해서 CBSIML.DS1을 현재입력 SAS데이터세트로 선언하고, LIST스테이트먼트를 사용해서 이 SAS데이터세트의 변수들 IRUM, Y, X1, X2의 현재관찰점(current observation)의 데이터값들을 지정된 윈도우에 출력시킨다. 이 SAS/IML프로그램에서는 RESET스테이트먼트에 옵션 NOLOG를 사용하였으므로 지정된 윈도우는 OUTPUT윈도우이다. 이 SAS/IML프로그램을 수행하면 OUTPUT윈도우에 그림 3.40과 같은 결과가 나타난다.

SAS		18:29 Saturday, February 26, 1994 12			
OBS	IRUM	X1	X2	Y	
1	KIM	1.0000	0.7200	11.0000	
Exiting IML.					

그림 3.40 LIST스테이트먼트를 실행한 결과

3.7.5 범위피연산자들

접근명령어들의 첫 번째 피연산자(operand)는 보통 처리대상의 관찰점들의 범위(range)를 지정하는 범위피연산자(range operand)이다. 이 범위피연산자들과 이들의 의미는 다음과 같다.

- ALL : 모든 관찰점
- CURRENT : 현재관찰점
- NEXT : 다음 관찰점
- AFTER : 현재관찰점 다음의 모든 관찰점

· POINT 번호 : 지정된 번호의 관찰점

관찰값들을 지정된 윈도우에 출력시키기 위해서는 LIST스테이트먼트를 사용한다. 범위 피연산자들을 예를 들어 설명하기 위해서 다음과 같은 LIST스테이트먼트를 사용한 예제들을 살펴보기로 하자.

- ① LIST;
이 LIST스테이트먼트는 현재관찰점을 출력한다.
- ② LIST ALL;
이 LIST스테이트먼트는 모든 관찰점들을 출력한다.
- ③ LIST NEXT;
이 LIST스테이트먼트는 현재관찰점의 다음 관찰점을 출력한다.
- ④ LIST POINT 10;
이 LIST스테이트먼트는 제10번째 관찰점을 출력한다.
- ⑤ LIST POINT {10 25};
이 LIST스테이트먼트는 제10번째와 제25번째의 관찰점들을 출력한다.
- ⑥ LIST POINT 스칼라변수;
이 LIST스테이트먼트는 스칼라변수에 수록된 번호의 관찰점을 출력한다.
- ⑦ LIST POINT (20:25);
이 LIST스테이트먼트는 제20번에서 제25번까지의 관찰점들을 출력한다.
- ⑧ LIST POINT ((20:25)|| (30:35));
이 LIST스테이트먼트는 제20번에서 제25번까지의 관찰점들과 제30번에서 제35번까지의 관찰점들을 출력한다.

SAS스테이트먼트를 처음 열면 현재관찰점의 번호는 0으로 고정된다. 이 경우에 LIST스테이트먼트, 또는 READ스테이트먼트를 실행하려고 하면 현재관찰점의 번호가 1로 바뀐다. 즉, 다음의 두 LIST스테이트먼트들은 모두 제1번째 관찰점을 지정된 윈도우에 출력시킨다.

LIST NEXT;
LIST CURRENT;

범위피연산자들을 사용하는 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-41.PRG에 수록되어 있다).

```

/*                               */
/*           EXAM3-41.PRG        */
/*                               */
/*           RANGE OPERANDS     */
/*                               */

```


SAS		21:14 Saturday, February 26, 1994 7		
OBS	IRUM	X1	X2	Y
1	KIM	1.0000	0.7200	11.0000
OBS	IRUM	X1	X2	Y
1	KIM	1.0000	0.7200	11.0000
OBS	IRUM	X1	X2	Y
1	KIM	1.0000	0.7200	11.0000
2	PARK	2.0000	0.3300	22.0000
3	CHOI	3.0000	-0.3400	33.0000
4	LEE	4.0000	0.8100	44.0000
5	AHN	5.0000	0.1200	55.0000
6	CHUNG	6.0000	0.0100	66.0000
OBS	IRUM	X1	X2	Y
2	PARK	2.0000	0.3300	22.0000
OBS	IRUM	X1	X2	Y
3	CHOI	3.0000	-0.3400	33.0000
OBS	IRUM	X1	X2	Y
2	PARK	2.0000	0.3300	22.0000
5	AHN	5.0000	0.1200	55.0000
OBS	IRUM	X1	X2	Y
3	CHOI	3.0000	-0.3400	33.0000
OBS	IRUM	X1	X2	Y
2	PARK	2.0000	0.3300	22.0000
3	CHOI	3.0000	-0.3400	33.0000
4	LEE	4.0000	0.8100	44.0000
5	AHN	5.0000	0.1200	55.0000
OBS	IRUM	X1	X2	Y
1	KIM	1.0000	0.7200	11.0000
2	PARK	2.0000	0.3300	22.0000
3	CHOI	3.0000	-0.3400	33.0000
5	AHN	5.0000	0.1200	55.0000
6	CHUNG	6.0000	0.0100	66.0000

Exiting IML.

그림 3.41 범위피연산자들

```

LIBNAME CBSIML "A:\IMLDATA";
PROC IML;
RESET NOCENTER NOLOG NOPRINT
      PAGESIZE=60 DEFLIB='CBSIML';
NN = 3;
USE DS1;
LIST NEXT;
LIST CURRENT;
LIST ALL;
LIST POINT 2;
LIST NEXT;
LIST POINT {2 5};

```

```
LIST POINT NN;
LIST POINT (2:5);
LIST POINT ((1:3)||{(5:6)});
CLOSE DS1;
QUIT;
```

이 SAS/IML 프로그램에서는 RESET 스테이트먼트의 옵션 DEFLIB=을 사용해서 디폴트 SAS 데이터 라이브러리를 CBSIML로 할당하였다. 이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.41과 같은 결과가 나타난다.

현재범위(current range)는 입출력을 수행한 마지막 연산에 의해서 결정(set)된다. 만일, 입출력을 수행하지 않고 현재관찰점을 정하기 위해서는 제 3.7.2 소절에서 설명하였듯이 SETIN 스테이트먼트, 또는 SETOUT 스테이트먼트를 범위피연산자 POINT와 함께 사용한다.

```
SETIN SAS 스테이트먼트 POINT 번호;
SETOUT SAS 스테이트먼트 POINT 번호;
```

이 SETIN 스테이트먼트의 예제로 다음 SAS/IML 프로그램을 수행해 보자(이 SAS/IML 프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-42.PRG에 수록되어 있다).

```
/* */
/*          EXAM3-42.PRG          */
/* */
/*          LISTING A SAS DATASET  */
/* */
LIBNAME CBSIML "A:\IMLDATA";
PROC IML;
RESET NOCENTER NOLOG NOPRINT
      PAGESIZE=60 DEFLIB='CBSIML';
USE DS1;
SETIN DS1 POINT 3;
LIST CURRENT;
CLOSE DS1;
QUIT;
```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.42와 같은 결과가 나타난다.

SAS		21:14 Saturday, February 26, 1994 8		
OBS	IRUM	X1	X2	Y
3	CHOI	3.0000	-0.3400	33.0000
Exiting IML.				

그림 3.42 SETIN 스테이트먼트를 사용한 현재관찰점의 지정

3.7.6 VAR구문

SAS스테이트먼트를 열기 위한 명령어들, 그리고 접근명령어들 중에서 다음 명령어들은 VAR구문(VAR clause)을 사용해서 처리대상이 되는 변수들을 지정할 수 있다.

USE, EDIT, CREATE, LIST, READ, REPLACE, APPEND

VAR구문을 스코프구문(scope clause)이라고도 한다. VAR구문은 SAS키워드 VAR 다음에 다음과 같은 방법으로 변수들을 지정한다.

- ① 작은따옴표들 사이, 또는 큰따옴표들 사이, 또는 중괄호 사이에 한 변수의 이름을 지정한다.
- ② 중괄호 사이에 여러 변수의 이름을 지정한다.
- ③ 변수들의 이름을 원소들로 하는 문자행렬을 지정한다.
- ④ SAS키워드들 `_ALL_`, `_NUM_`, `_CHAR_`를 사용한다. SAS키워드 `_ALL_`은 모든 변수를 지정하는 것이고, SAS키워드 `_NUM_`은 숫자변수들만을 지정하는 것이며, SAS키워드 `_CHAR_`는 문자변수들만을 지정하는 것이다.

VAR구문을 설명하기 위해서 다음과 같은 LIST스테이트먼트를 사용한 예제들을 살펴보기로 하자.

- ① LIST:
이 LIST스테이트먼트는 모든 변수들의 현재관찰점을 출력한다.
- ② LIST VAR "X";
LIST VAR 'X';
LIST VAR {X};
이 LIST스테이트먼트들은 변수 X의 현재관찰점을 출력한다.
- ③ LIST VAR {X Y};
이 LIST스테이트먼트는 변수들 X와 Y의 현재관찰점을 출력한다.
- ④ NAME1 = {X Y}; LIST VAR NAME1;
이 LIST스테이트먼트도 변수들 X와 Y의 현재관찰점을 출력한다.
- ⑤ NAME1 = {X Y}; NAME2 = {Z W}; LIST VAR (NAME1 || NAME2);
이 LIST스테이트먼트는 변수들 X, Y, Z와 W의 현재관찰점을 출력한다.
- ⑥ LIST VAR _ALL_ ;
이 LIST스테이트먼트는 모든 변수들의 현재관찰점을 출력한다.
- ⑦ LIST VAR _NUM_ ;

이 LIST스테이트먼트는 모든 숫자변수들의 현재관찰점을 출력한다.

⑧ LIST VAR _CHAR _;

이 LIST스테이트먼트는 모든 문자변수들의 현재관찰점을 출력한다.

VAR구문의 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-43.PRG에 수록되어 있다).

```

/* */
/* EXAM3-43.PRG */
/* */
/* VAR CLAUSES */
/* */
LIBNAME CBSIML "A:\IMLDATA";
PROC IML;
RESET NOCENTER NOLOG NOPRINT
      PAGESIZE=60 DEFLIB='CBSIML';
FREE;
USE DS1;
LIST;

```

SAS		23:51 Saturday, February 26, 1994 1			
OBS	IRUM	X1	X2	Y	
1	KIM	1.0000	0.7200	11.0000	
OBS	X1				
1	1.0000				
OBS	X1				
1	1.0000				
OBS	X1				
1	1.0000		X2		
OBS	X1		X2		
1	1.0000	0.7200		Y	
OBS	X1	X2		Y	
1	1.0000	0.7200	11.0000	Y	
OBS	IRUM	X1	X2	Y	
1	KIM	1.0000	0.7200	11.0000	
OBS	X1	X2		Y	
1	1.0000	0.7200	11.0000		
OBS	IRUM				
1	KIM				

Exiting IML.

그림 3.43 VAR구문

```

LIST VAR "X1";
LIST VAR 'X1';
LIST VAR {X1};
LIST VAR {X1 X2};
NAMVA1 = {X1 X2}; LIST VAR NAMVA1;
NAMVA2 = {Y}; LIST VAR (NAMVA1 || NAMVA2);
LIST VAR _ALL_;
LIST VAR _NUM_;
LIST VAR _CHAR_;
CLOSE DS1;
QUIT;

```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.43과 같은 결과가 나타난다.

3.7.7 WHERE구문

WHERE구문(WHERE clause)은 처리할 범위(range)의 부분집합을 정의한다. WHERE구문은 SAS키워드 WHERE와 소괄호 안에 비교연산자들로 이루어진 수식으로 구성되어 있다. 즉, 다음과 같은 사용양식을 갖는다.

```

WHERE (변수 비교연산자 글자값);
WHERE (변수 비교연산자(수식));
WHERE (변수 비교연산자 변수의 현재값);

```

독자들은 다음과 같은 비교연산자들에 대해서 잘 알고 있으리라 생각한다.

< < = = ^= > >=

다음과 같이 독자들에게 익숙하지 않은 비교연산자들도 있다.

- ? 문자열을 포함한다.
- ^? 문자열을 포함하지 않는다.
- =: 이 값으로 시작한다.
- =* 소리가 유사하다(sounds like).
- & AND연산자
- | OR연산자

다음과 같은 비교연산자들은 피비교연산자들로 스칼라뿐만 아니라 행렬도 사용할 수 있다. 이 비교연산자들을 사용하면 비교연산이 참이 되는 원소(a matching value)를 찾기

위해서 행렬을 뒤진다(scan).

```
= ^= ? ^? =: =*
```

WHERE구문을 설명하기 위해서 다음과 같은 LIST스테이트먼트를 사용한 예제들을 살펴보기로 하자.

① LIST ALL WHERE(IRUM="CHOI");

이 LIST스테이트먼트는 변수 IRUM의 데이터값이 CHOI인 모든 관찰점들을 출력한다.

② LIST ALL WHERE(IRUM="CHOI" | IRUM="AHN");

```
LIST ALL WHERE(IRUM={"CHOI" "AHN"});
```

```
IRUMLIST={"CHOI" "AHN"}; LIST ALL WHERE(IRUM=IRUMLIST);
```

이 LIST스테이트먼트들은 변수 IRUM의 데이터값이 CHOI, 또는 AHN인 모든 관찰점들을 출력한다.

WHERE구문의 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-44.PRG에 수록되어 있다).

```

/*                               */
/*           EXAM3-44.PRG         */
/*                               */
/*           WHERE CLAUSES       */
/*                               */
LIBNAME CBSIML "A:\IMLDATA";
PROC IML;
RESET NOCENTER NOLOG NOPRINT
      PAGESIZE=80 DEFLIB='CBSIML';
FREE;
USE DS1;
LIST ALL WHERE (Y < 50);
LIST ALL WHERE (Y <= 55);
LIST ALL WHERE (Y = 44);
LIST ALL WHERE (Y ^= 44);
LIST ALL WHERE (Y > 44);
LIST ALL WHERE (Y >= 44);
LIST ALL WHERE (IRUM ? "A");
LIST ALL WHERE (IRUM ^? "A");
LIST ALL WHERE (IRUM =: "A");
LIST ALL WHERE (IRUM =* "PARC");
LIST ALL WHERE ( (Y>11) & (X1<4) );

```

SAS

0:42 Sunday, February 27, 1994 5

OBS	IRUM	X1	X2	Y
1	KIM	1.0000	0.7200	11.0000
2	PARK	2.0000	0.3300	22.0000
3	CHOI	3.0000	-0.3400	33.0000
4	LEE	4.0000	0.8100	44.0000
OBS	IRUM	X1	X2	Y
1	KIM	1.0000	0.7200	11.0000
2	PARK	2.0000	0.3300	22.0000
3	CHOI	3.0000	-0.3400	33.0000
4	LEE	4.0000	0.8100	44.0000
5	AHN	5.0000	0.1200	55.0000
OBS	IRUM	X1	X2	Y
4	LEE	4.0000	0.8100	44.0000
OBS	IRUM	X1	X2	Y
1	KIM	1.0000	0.7200	11.0000
2	PARK	2.0000	0.3300	22.0000
3	CHOI	3.0000	-0.3400	33.0000
5	AHN	5.0000	0.1200	55.0000
6	CHUNG	6.0000	0.0100	66.0000
OBS	IRUM	X1	X2	Y
5	AHN	5.0000	0.1200	55.0000
6	CHUNG	6.0000	0.0100	66.0000
OBS	IRUM	X1	X2	Y
4	LEE	4.0000	0.8100	44.0000
5	AHN	5.0000	0.1200	55.0000
6	CHUNG	6.0000	0.0100	66.0000
OBS	IRUM	X1	X2	Y
2	PARK	2.0000	0.3300	22.0000
5	AHN	5.0000	0.1200	55.0000
OBS	IRUM	X1	X2	Y
1	KIM	1.0000	0.7200	11.0000
3	CHOI	3.0000	-0.3400	33.0000
4	LEE	4.0000	0.8100	44.0000
6	CHUNG	6.0000	0.0100	66.0000
OBS	IRUM	X1	X2	Y
5	AHN	5.0000	0.1200	55.0000
OBS	IRUM	X1	X2	Y
2	PARK	2.0000	0.3300	22.0000
OBS	IRUM	X1	X2	Y
2	PARK	2.0000	0.3300	22.0000
3	CHOI	3.0000	-0.3400	33.0000
OBS	IRUM	X1	X2	Y
1	KIM	1.0000	0.7200	11.0000
2	PARK	2.0000	0.3300	22.0000
3	CHOI	3.0000	-0.3400	33.0000
4	LEE	4.0000	0.8100	44.0000
5	AHN	5.0000	0.1200	55.0000
6	CHUNG	6.0000	0.0100	66.0000
OBS	IRUM	X1	X2	Y
2	PARK	2.0000	0.3300	22.0000
OBS	IRUM	X1	X2	Y
2	PARK	2.0000	0.3300	22.0000
3	CHOI	3.0000	-0.3400	33.0000

OBS	IRUM	X1	X2	Y
2	PARK	2.0000	0.3300	22.0000
3	CHOI	3.0000	-0.3400	33.0000
OBS	IRUM	X1	X2	Y
2	PARK	2.0000	0.3300	22.0000
3	CHOI	3.0000	-0.3400	33.0000

Exiting IML.

그림 3.44 WHERE구문

```
LIST ALL WHERE ( (Y>11) | (X1<4) );
LIST ALL WHERE (IRUM = "PARK");
LIST ALL WHERE (IRUM = "PARK" | IRUM = "CHOI");
LIST ALL WHERE (IRUM = {"PARK" "CHOI"});
ILIST={"PARK" "CHOI"}; LIST ALL WHERE (IRUM = ILIST);
CLOSE DS1;
QUIT;
```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.44와 같은 결과가 나타난다.

3.7.8 행렬을 사용한 데이터처리

SAS/IML 환경에서는 SAS 데이터 세트의 관찰값들을 한 행렬로 처리해야 할 필요가 있고, 또한 한 행렬을 여러 개의 변수들로 나누어 처리해야 할 경우도 있다. 이러한 목적을 위해서는 READ 스테이트먼트, APPEND 스테이트먼트, 그리고 CREATE 스테이트먼트에서 FROM 구문 (FROM clause) 과 INTO 구문 (INTO clause) 을 사용한다.

1) INTO 구문

예를 들어 다음 READ 스테이트먼트를 살펴보자.

```
READ ALL INTO M1;
```

이 READ 스테이트먼트를 실행하면 현재 입력 SAS 데이터 세트의 각 변수를 한 열로 하고 각 관찰점을 한 행으로 하는 행렬 M1을 만든다. 즉, INTO 구문은 변수들을 가로로 결합해서 주어진 SAS 데이터 세트를 하나의 행렬로 만들도록 한다.

변수들과 관찰점들의 부분집합을 하나의 행렬로 만들기 위해서는 범위 피연산자들, VAR 구문, 그리고 WHERE 구문을 사용한다. 예를 들어, 다음 READ 스테이트먼트를 살펴보자.

```
READ ALL VAR _CHAR_ INTO M1;
```

이 READ 스테이트먼트를 실행하면 현재 입력 SAS 데이터 세트에서 문자 변수들을 열로 하는 행렬 M1을 만든다.

SAS		18:39 Saturday, March 5, 1994 1		
USING INTO CLAUSES				
M1		M2	M3	
1	0.72	KIM	2	0.33 22
2	0.33	PARK	5	0.12 55
3	-0.34	CHOI		
4	0.81	LEE		
5	0.12	AHN		
6	0.01	CHUNG		
Exiting IML.				

그림 3.45 INTO구문

INTO구문의 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-45.PRG에 수록되어 있다).

```

/*                               */
/*          EXAM3-45.PRG          */
/*                               */
/*          INTO CLAUSES         */
/*                               */
LIBNAME CBSIML "A:\IMLDATA";
PROC IML;
RESET NOCENTER NOLOG NOPRINT
      PAGESIZE=80 DEFLIB='CBSIML' FW=7;
FREE;
USE DS1;
READ ALL VAR {X1 X2} INTO M1;
READ ALL VAR _CHAR_ INTO M2;
READ ALL VAR _NUM_ INTO M3 WHERE( IRUM ? "A" );
CLOSE DS1;
PRINT, "USING INTO CLAUSES", M1 " " M2 " " M3;
QUIT;

```

이 SAS/IML프로그램을 수행하면 OUTPUT윈도우에 그림 3.45와 같은 결과가 나타난다.

2) FROM구문

예를 들어 다음 APPEND스테이트먼트를 살펴보자.

```
APPEND FROM M1;
```

이 APPEND스테이트먼트를 실행하면 현재출력SAS데이터세트의 끝(to the end of a dataset)에 관찰점들을 추가한다. 행렬 M1의 열의 개수를 ncol이라 하자. 행렬 M1의 열들

이 첫 ncol개의 변수들이 된다. 또한, 행렬 M1의 각 행이 하나의 관찰점이 된다. 즉, FROM 구문은 행렬을 열벡터들로 나누어서 변수들을 만들게 한다.

다음 CREATE스테이트먼트를 살펴보자.

```
CREATE ABC FROM M1;
```

이 CREATE스테이트먼트 앞에서 행렬 M1이 정의되어야 하며, 이 행렬의 열의 개수를 ncol이라 하자. 다시 한 번 강조하지만 CREATE스테이트먼트는 선언문이지 실행문이 아니다. 따라서 이 CREATE스테이트먼트가 선언되면 새로운 SAS데이터세트 ABC가 열린다. 이 SAS데이터세트 ABC의 변수들의 개수는 ncol이며 각 변수의 특성은 행렬 M1의 대응하는 열의 특성과 같다. 즉, SAS데이터세트 ABC는 행렬 M1의 구조를 빌려 쓰는 것이다. 다음과 같은 스테이트먼트들을 살펴보자.

```
CREATE ABC FROM M1;
APPEND FROM M2;
```

이 CREATE스테이트먼트 앞에서 행렬 M1이 정의되어야 하며, SAS데이터세트 ABC는 행렬 M1의 구조를 빌려 쓰고, 실제 데이터값들은 행렬 M2의 것들을 사용한다. 한 가지 유의할 점은 VAR구문은 FROM구문과 함께 사용할 수 없으나 INTO구문과는 함께 사용할 수 있다는 것이다.

FROM구문의 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-46.PRG에 수록되어 있다).

```
/* */
/*      EXAM3-46.PRG      */
/* */
/*      FROM CLAUSES      */
/* */
LIBNAME CBSIML "A:\IMLDATA";
PROC IML;
RESET NOCENTER NOLOG NOPRINT
      PAGESIZE=80 DEFLIB='CBSIML';
FREE;
M1 = { 10.1 10.2, 10.3 10.4, 10.5 10.6,
      10.7 10.8, 10.9 10.0, -10.1 -10.2};
USE DS1;
READ ALL VAR {X1 X2} INTO M2;
CLOSE DS1;
CREATE DS2 FROM M1;
APPEND FROM M1;
```

SAS			16:08 Sunday, February 27, 1994 12
Exiting IML.			
SAS			16:08 Sunday, February 27, 1994 13
OBS	COL1	COL2	
1	10.1	10.2	
2	10.3	10.4	
3	10.5	10.6	
4	10.7	10.8	
5	10.9	10.0	
6	-10.1	-10.2	
OBS	COL1	COL2	
1	1	0.72	
2	2	0.33	
3	3	-0.34	
4	4	0.81	
5	5	0.12	
6	6	0.01	

그림 3.46 FROM구문

```

CLOSE DS2;
CREATE DS3 FROM M1;
APPEND FROM M2;
CLOSE DS3;
QUIT;
PROC PRINT DATA=CBSIML.DS2;
RUN;
PROC PRINT DATA=CBSIML.DS3;
RUN;

```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.46과 같은 결과가 나타난다.

3) 옵션 COLNAME=

CREATE 스테이트먼트와 READ 스테이트먼트에서는 옵션 COLNAME=을 사용해서 COLNAME 벡터를 지정할 수 있다. 우선 다음과 같은 CREATE 스테이트먼트를 살펴보자.

```
CREATE ABC FROM M1 [COLNAME=COLIRUM];
```

이 CREATE 스테이트먼트에서는 새로운 SAS 데이터세트 ABC에서 사용할 변수들의 이름들로 구성된 문자벡터 COLIRUM을 옵션 COLNAME=에 할당했다. 물론, 문자벡터 COLIRUM은 CREATE 스테이트먼트 앞에서 정의되어야 하며 원소들을 갖고 있어야 한다. 즉, 이 문자벡터로 원소들을 갖지 않고 이름만 정해진 벡터를 사용해서는 안 된다. 만일, CREATE 스테이트먼트에서 옵션 COLNAME=이 사용되지 않으면, SAS 데이터세트

ABC의 변수들의 이름들은 COL1, COL2, ...으로 정해진다. 다음과 같은 READ스태이트먼트를 살펴보자.

```
READ INTO M2 [COLNAME=COLIRUM2];
```

이 READ스태이트먼트에서는 현재입력SAS데이터세트의 변수들의 이름들로 구성된 문자벡터를 만들어 옵션 COLNAME=에 정의된 벡터이름 COLIRUM2로 저장한다. 즉, 옵션 COLNAME=에는 새로이 만들어지는 행렬 M2의 열들의 이름들로 구성된 문자벡터의 이름을 지정한다.

옵션 COLNAME=의 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-47.PRG에 수록되어 있다).

```

/*                               */
/*           EXAM3-47.PRG         */
/*                               */
/*           USING OPTION COLNAME= */
/*                               */
LIBNAME CBSIML "A:\IMLDATA";
PROC IML;
  RESET NOCENTER NOLOG NOPRINT
        PAGESIZE=80 DEFLIB='CBSIML';
FREE;
M1 = { 10.1 10.2, 10.3 10.4, 10.5 10.6,
       10.7 10.8, 10.9 10.0, -10.1 -10.2 };
COLIRUM1 = { X11 X12 };
CREATE DS4 FROM M1 [COLNAME=COLIRUM1];
APPEND FROM M1;
CLOSE DS4;
COLIRUM2 = { X21 X22 };
USE DS1;
READ ALL VAR { X1 X2 } INTO M2 [COLNAME=COLIRUM2];
CLOSE DS1;
PRINT, "USING OPTION COLNAME=", M2, COLIRUM2;
QUIT;
PROC PRINT DATA=CBSIML.DS4;
RUN;

```

이 SAS/IML프로그램을 수행하면 OUTPUT윈도우에 그림 3.47과 같은 결과가 나타난다.

SAS		7:46 Saturday, March 5, 1994 26	
USING OPTION COLNAME=			
	M2		
	1	0.72	
	2	0.33	
	3	-0.34	
	4	0.81	
	5	0.12	
	6	0.01	
COLIRUM2			
X1	X2		
Exiting IML.			
SAS		7:46 Saturday, March 5, 1994 27	
OBS	X11	X12	
1	10.1	10.2	
2	10.3	10.4	
3	10.5	10.6	
4	10.7	10.8	
5	10.9	10.0	
6	-10.1	-10.2	

그림 3.47 옵션 COLNAME=을 사용한 결과

4) 옵션 ROWNAME=

CREATE스테이트먼트, APPEND스테이트먼트와 READ스테이트먼트에서는 옵션 ROWNAME=을 사용해서 ROWNAME벡터를 지정할 수 있다.

다음과 같은 스테이트먼트들을 살펴보자.

```
CREATE ABC FROM M1 [ROWNAME=ROWIRUM];
APPEND FROM M1 [ROWNAME=ROWIRUM];
```

이 CREATE스테이트먼트에서는 새로운 SAS데이터세트 ABC에서 사용할 관찰점들의 이름들로 구성된 문자벡터 ROWIRUM을 옵션 ROWNAME=에 할당했다. 물론, 문자벡터 ROWIRUM은 CREATE스테이트먼트 앞에서 정의되어야 하며 원소들을 갖고 있어야 한다. 즉, 이 문자벡터로 원소들을 갖지 않고 이름만 정해진 벡터를 사용해서는 안 된다. CREATE스테이트먼트와 APPEND스테이트먼트에서 옵션 ROWNAME=을 사용하기 위해서는 이 두 스테이트먼트들에서 옵션 ROWNAME=에 지정하는 문자벡터는 같은 것이어야 한다. 따라서 CREATE스테이트먼트에서 옵션 ROWNAME=을 사용하지 않는다면 APPEND스테이트먼트에서도 옵션 ROWNAME=을 사용할 수 없다.

다음과 같은 READ스테이트먼트를 살펴보자.

```
READ INTO M2 [ROWNAME=IRUM];
```

READ스테이트먼트에서는 현재입력SAS데이터세트의 문자변수들 중의 하나를 옵션

ROWNAME=에 할당한다. 이 READ스테이트먼트의 옵션 ROWNAME=에는 기존의 문자변수들 중의 하나인 IRUM을 할당하였다.

옵션 ROWNAME=의 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML 프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-48.PRG에 수록되어 있다).

```

/* */
/*          EXAM3-48.PRG          */
/* */
/*          USING OPTION ROWNAME=  */
/* */
LIBNAME CBSIML "A:\IMLDATA";
PROC IML;
RESET NOCENTER NOLOG NOPRINT
      PAGESIZE=80 DEFLIB='CBSIML';
FREE;
M1 = { 10.1 10.2, 10.3 10.4, 10.5 10.6,
       10.7 10.8, 10.9 10.0, -10.1 -10.2};
ROWIRUM1 = { NO1, NO2, NO3, NO4, NO5, NO6 };
CREATE DS5 FROM M1 [ROWNAME=ROWIRUM1];
APPEND FROM M1 [ROWNAME=ROWIRUM1];
CLOSE DS5;
USE DS1;
READ ALL VAR {X1 X2} INTO M2 [ROWNAME=IRUM];
CLOSE DS1;

```

SAS			7:46 Saturday, March 5, 1994 30
USING OPTION ROWNAME=			
	M2		IRUM
	1	0.72	KIM
	2	0.33	PARK
	3	-0.34	CHOI
	4	0.81	LEE
	5	0.12	AHN
	6	0.01	CHUNG
Exiting IML.			
SAS			7:46 Saturday, March 5, 1994 31
ROWIRUM1	COL1	COL2	
NO1	10.1	10.2	
NO2	10.3	10.4	
NO3	10.5	10.6	
NO4	10.7	10.8	
NO5	10.9	10.0	
NO6	-10.1	-10.2	

그림 3.48 옵션 ROWNAME=을 사용한 결과

```

PRINT, "USING OPTION ROWNAME=", M2 " " IRUM,;
QUIT;
PROC PRINT DATA=CBSIML.DS5;
    ID ROWIRUM1;
RUN;

```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.48과 같은 결과가 나타난다.

옵션 ROWNAME=과 옵션 COLNAME=을 동시에 사용하는 예제로 다음 SAS/IML 프로그램을 수행해 보자(이 SAS/IML 프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-49.PRG에 수록되어 있다).

```

/*                                                    */
/*                EXAM3-49.PRG                        */
/*                                                    */
/*                USING OPTIONS COLNAME= & ROWNAME=  */
/*                                                    */
LIBNAME CBSIML "A:\IMLDATA";
PROC IML;
RESET NOCENTER NOLOG NOPRINT
    PAGESIZE=80 DEFLIB='CBSIML';
FREE;
M1 = { 10.1 10.2, 10.3 10.4, 10.5 10.6,
      10.7 10.8, 10.9 10.0, -10.1 -10.2};
ROWIRUM1 = {NO1, NO2, NO3, NO4, NO5, NO6};
COLIRUM1 = {SUBJ1 SUBJ2};
CREATE DS6 FROM M1
    [COLNAME=COLIRUM1 ROWNAME=ROWIRUM1];
APPEND FROM M1 [ROWNAME=ROWIRUM1];
CLOSE DS6;
COLIRUM3 = {X31 X32};
USE DS1;
READ ALL VAR {X1 X2} INTO M2
    [COLNAME=COLIRUM3 ROWNAME=IRUM];
CLOSE DS1;
PRINT, "USING OPTIONS ROWNAME & COLNAME =",
      M2 " " IRUM, COLIRUM3;
QUIT;
PROC PRINT DATA=CBSIML.DS6;
    ID ROWIRUM1;
RUN;

```

```

SAS
7:46 Saturday, March 5, 1994 34

USING OPTIONS ROWNAME= & COLNAME=

      M2
      1      0.72      IRUM
      2      0.33      KIM
      3     -0.34      PARK
      4      0.81      CHOI
      5      0.12      LEE
      6      0.01      AHN
                        CHUNG

COLIRUM3
X1      X2

Exiting IML.

SAS
7:46 Saturday, March 5, 1994 35

ROWIRUM1  SUBJ1      SUBJ2

      NO1      10.1      10.2
      NO2      10.3      10.4
      NO3      10.5      10.6
      NO4      10.7      10.8
      NO5      10.9      10.0
      NO6     -10.1     -10.2
    
```

그림 3.49 옵션들 COLNAME=과 ROWNAME=을 사용한 결과

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.49와 같은 결과가 나타난다.

3.7.9 접근명령어들

제 3.7.4 소절에서 SAS/IML 환경에서 SAS 데이터 세트의 데이터 값들에 직접 접근하는 접근명령어들 중에서 APPEND, READ, LIST를 간단히 설명했다. 여기서는 모든 접근명령어들을 설명하고자 한다. 이 소절에서는 대소부호들(<>) 사이의 기능은 옵션을 의미하기로 한다.

1) APPEND

APPEND 스테이트먼트는 CREATE 스테이트먼트를 사용해서 연 SAS 데이터 세트에 데이터 값들을 쓰기 위한 실행문이다. APPEND 스테이트먼트의 사용양식은 다음과 같다.

```

APPEND <VAR구문><FROM구문 <[ROWNAME = 문자벡터]>>;
    
```

제 3.7.8 소절에서 설명하였듯이 APPEND 스테이트먼트의 사용양식에서 FROM 구문을 사용하면 주어진 행렬로부터 데이터 값들을 읽어서 지정한 SAS 데이터 세트에 쓴다.

SAS					20:24 Monday, February 28, 1994 2
Exiting IML.					
SAS					20:24 Monday, February 28, 1994 3
RNAME	C1	C2	C3	C4	
R1	1	2	3	4	
R2	5	6	7	8	
R3	9	10	11	12	
R4	13	14	15	16	
R5	17	18	19	20	

그림 3.50 APPEND스테이트먼트를 실행한 결과

APPEND스테이트먼트의 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-50.PRG에 수록되어 있다).

```

/* */
/*          EXAM3-50.PRG          */
/* */
/*          APPEND STATEMENT      */
/* */

LIBNAME CBSIML "A:\IMLDATA";
PROC IML;
RESET NOCENTER NOLOG NOPRINT
      PAGESIZE=80 DEFLIB='CBSIML';
FREE;
MAT1 = { 1 2 3 4,
         5 6 7 8,
         9 10 11 12,
         13 14 15 16,
         17 18 19 20};
RNAME = {R1 R2 R3 R4 R5};
CNAME = {C1 C2 C3 C4};
CREATE DS7 FROM MAT1
      [COLNAME=CNAME ROWNAME=RNAME];
APPEND FROM MAT1 [ROWNAME=RNAME];
CLOSE DS7;
QUIT;
PROC PRINT DATA=CBSIML.DS7;
      ID RNAME;
RUN;

```

이 SAS/IML프로그램을 수행하면 OUTPUT윈도우에 그림 3.50과 같은 결과가 나타난다.

2) READ

READ스태이트먼트는 현재입력SAS데이터세트에서 데이터값들을 읽어서 각 변수를 열 벡터로 만들거나 행렬을 만들기 위한 실행문이다. READ스태이트먼트의 사용양식은 다음과 같다.

```
READ <범위> <VAR구문> <WHERE구문> ;
READ <범위> <VAR구문> <WHERE구문>
    INTO구문 <[ROWNAME=문자벡터 COLNAME=문자벡터]>;
```

제 3.7.8소절에서 설명하였듯이 READ스태이트먼트의 사용양식에서 INTO구문을 사용하면 주어진 데이터값들을 읽어서 지정한 행렬을 만든다.

READ스태이트먼트의 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML 프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-51.PRG에 수록되어 있다).

```

/*                                     */
/*           EXAM3-51.PRG             */
/*                                     */
/*           READ STATEMENTS         */
/*                                     */
LIBNAME CBSIML "A:\IMLDATA";
PROC IML;
RESET NOCENTER NOLOG NOPRINT
      PAGESIZE=80 DEFLIB='CBSIML';
FREE;
USE DS7;
READ ALL VAR { C1 C3 C4 } WHERE (C2 < 9);
PRINT , "READ STATEMENT 1", C1 C3, C4,;
RIRUM = { "RN1", "RN2", "RN3", "RN4", "RN5" };
CREATE DS8;
APPEND;
CLOSE DS7 DS8;
USE DS8;
READ ALL VAR { C1 C3 } WHERE (C4 < 14)
      INTO MAT2 [ROWNAME=RIRUM COLNAME=CIRUM];
PRINT , "READ STATEMENT 2",
      CIRUM , RIRUM MAT2,;
CLOSE DS8;
QUIT;
```

이 SAS/IML프로그램을 수행하면 OUTPUT윈도우에 그림 3.51과 같은 결과가 나타난다.

```

SAS
20:24 Monday, February 28, 1994 11

READ STATEMENT 1

      C1      C3
      1      3
      5      7

      C4
      4
      8

READ STATEMENT 2

CIRUM
C1      C3

RIRUM      MAT2
RN1      1      3
RN2      5      7

Exiting IML.
    
```

그림 3.51 READ스테이트먼트를 실행한 결과

3) LIST

LIST스테이트먼트는 현재SAS데이터세트에서 데이터값들을 읽어서 지정된 윈도우에 출력한다. LIST스테이트먼트의 사용양식은 다음과 같다.

```

LIST <범위> <VAR구문> <WHERE구문> ;
    
```

LIST스테이트먼트의 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-52.PRG에 수록되어 있다).

```

/* */
/*      EXAM3-52.PRG      */
/* */
/*      LIST STATEMENT      */
/* */
LIBNAME CBSIML "A:\IMLDATA";
PROC IML;
RESET NOCENTER NOLOG NOPRINT
      PAGESIZE=80 DEFLIB='CBSIML';
FREE;
USE DS7;
LIST ALL VAR {C1 C3 C4} WHERE (C2<9);
CLOSE DS7;
QUIT;
    
```

이 SAS/IML프로그램을 수행하면 OUTPUT윈도우에 그림 3.52와 같은 결과가 나타난다.

SAS		20:24 Monday, February 28, 1994 15		
OBS	C1	C3	C4	
1	1.0000	3.0000	4.0000	
2	5.0000	7.0000	8.0000	
Exiting IML.				

그림 3.52 LIST스테이트먼트를 실행한 결과

4) FIND

FIND스테이트먼트는 WHERE구문의 조건을 만족하는 관찰점들의 번호들을 INTO구문에 지정한 열벡터로 출력한다. FIND스테이트먼트의 사용양식은 다음과 같다.

FIND 범위 WHERE구문 INTO구문 ;

다음과 같은 예들을 살펴보자.

```
FIND ALL WHERE( IRUM="C" ) INTO P;
FIND NEXT WHERE( AGE > 30 ) INTO P2;
```

첫 번째 FIND스테이트먼트는 변수 IRUM의 첫자가 C인 관찰점들의 번호들을 열벡터 P에 저장한다. 두 번째 FIND스테이트먼트는 변수 AGE가 30이 넘는 관찰점들의 번호들을 열벡터 P2에 저장한다.

FIND스테이트먼트의 예제로 다음 SAS/IML 프로그램을 수행해 보자(이 SAS/IML 프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-53.PRG에 수록되어 있다).

```
/* */
/*          EXAM3-53.PRG          */
/* */
/*          FIND STATEMENT        */
/* */
LIBNAME CBSIML "A:\IMLDATA";
PROC IML;
RESET NOCENTER NOLOG NOPRINT
      PAGESIZE=80 DEFLIB='CBSIML';
FREE;
V1 = { 1, 5, 9, 13, 17 };
V2 = { 2, 6, 10, 14, 18 };
V3 = { 3, 7, 11, 15, 19 };
V4 = { 4, 8, 12, 16, 20 };
```

SAS		20:24 Monday, February 28, 1994 24			
RIRUM	V1	V2	V3	V4	
R1	1	2	3	4	
R2	5	6	7	8	
R3	9	10	11	12	
R4	13	14	15	16	
R5	17	18	19	20	
RIRUM	V1	V2	V3	V4	
R1	1	2	3	4	
R2	5	6	7	8	
R3	9	10	11	12	
R4	13	14	15	16	
R5	17	18	19	20	
OBS NO. OF RIRUM=R3 IS			POSIT	3	
Exiting IML.					

그림 3.53 FIND스테이트먼트를 실행한 결과

```

RIRUM = { "R1", "R2", "R3", "R4", "R5" };
CREATE DS9;
APPEND;
PRINT, RIRUM V1 V2 V3 V4, ;
FIND ALL WHERE (RIRUM="R3") INTO POSIT;
PRINT, RIRUM V1 V2 V3 V4,
      "OBS NO. OF RIRUM=R3 IS" POSIT,;
CLOSE DS9;
QUIT;

```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.53과 같은 결과가 나타난다.

5) REPLACE

REPLACE 스테이트먼트는 SAS 데이터 세트의 어떤 변수의 값을 작업공간(workspace)의 같은 변수의 현재값(current value)으로 대체한다. REPLACE 스테이트먼트의 사용양식은 다음과 같다.

```
REPLACE <범위><VAR구문><WHERE구문>;
```

다음과 같은 예를 살펴보자.

```

STATE = "IL";
REPLACE ALL VAR{ STATE } WHERE(STATE="ILL");

```

이 REPLACE 스테이트먼트를 실행하면 변수 STATE의 데이터값들 중에서 ILL이 IL

SAS		9:58 Tuesday, March 1, 1994 28			
OBS	RIRUM	V1	V2	V3	V4
1	R1	1.0000	2.0000	3.0000	4.0000
2	R2	5.0000	6.0000	7.0000	8.0000
3	R3	9.0000	10.0000	11.0000	12.0000
4	R4	13.0000	14.0000	15.0000	16.0000
5	R5	17.0000	18.0000	19.0000	20.0000

OBS	RIRUM	V1	V2	V3	V4
1	R1	1.0000	2.0000	3.0000	4.0000
2	R2	5.0000	6.0000	7.0000	8.0000
3	R3	9.0000	10.0000	11.0000	12.0000
4	R4	13.0000	14.0000	15.0000	16.0000
5	R5	17.0000	18.0000	19.0000	20.0000

RIRUM	V1	V2	V3	V4
RR	1	2	3	4
	5	6	7	8
	9	10	11	12
	13	14	15	16
	17	18	19	20

Exiting IML.

SAS		9:58 Tuesday, March 1, 1994 29			
OBS	RIRUM	V1	V2	V3	V4
1	R1	1	2	3	4
2	R2	5	6	7	8
3	RR	9	10	11	12
4	R4	13	14	15	16
5	R5	17	18	19	20

그림 3.54 REPLACE스테이트먼트를 실행한 결과

로 바뀐다.

REPLACE스테이트먼트의 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-54.PRG에 수록되어 있다).

```

/* */
/*          EXAM3-54.PRG          */
/* */
/*          REPLACE STATEMENT     */
/* */
LIBNAME CBSIML "A:\IMLDATA";
PROC IML;
RESET NOCENTER NOLOG NOPRINT
      PAGESIZE=80 DEFLIB='CBSIML';
FREE;

```

```

V1 = { 1, 5, 9, 13, 17 };
V2 = { 2, 6, 10, 14, 18 };
V3 = { 3, 7, 11, 15, 19 };
V4 = { 4, 8, 12, 16, 20 };
RIRUM = { "R1", "R2", "R3", "R4", "R5" };
CREATE DS9;
APPEND;
CLOSE DS9;
EDIT DS9;
READ ALL;
LIST ALL;
RIRUM = "RR";
REPLACE ALL VAR { RIRUM } WHERE (RIRUM="R3");
LIST ALL;
PRINT, RIRUM V1 V2 V3 V4, ;
CLOSE DS9;
QUIT;
PROC PRINT DATA=CBSIML.DS9;
RUN;

```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.54와 같은 결과가 나타난다.

6) DELETE

이 DELETE 스테이트먼트는 지정된 범위내에서 WHERE 구문의 조건을 만족하는 관찰점들을 지운다. DELETE 스테이트먼트의 사용양식은 다음과 같다.

DELETE 범위 WHERE 구문 ;

다음과 같은 예들을 살펴보자.

```

DELETE;
DELETE POINT 34;
DELETE ALL WHERE( AGE < 21 );

```

첫 번째 DELETE 스테이트먼트는 현재 관찰점을 지운다. 두 번째 DELETE 스테이트먼트는 제34번째의 관찰점을 지운다. 세 번째 DELETE 스테이트먼트는 변수 AGE가 21미만인 관찰점들을 지운다.

DELETE 스테이트먼트를 수행해서 관찰점들을 지운다는 것은 이 관찰점들을 SAS 데이

SAS		8:28 Tuesday, March 1, 1994 1			
OBS	RIRUM	V1	V2	V3	V4
1	R1	1.0000	2.0000	3.0000	4.0000
2	R2	5.0000	6.0000	7.0000	8.0000
3	R3	9.0000	10.0000	11.0000	12.0000
4	R4	13.0000	14.0000	15.0000	16.0000
5	R5	17.0000	18.0000	19.0000	20.0000
1 observations were deleted.					
OBS	RIRUM	V1	V2	V3	V4
1	R1	1.0000	2.0000	3.0000	4.0000
2	R2	5.0000	6.0000	7.0000	8.0000
4	R4	13.0000	14.0000	15.0000	16.0000
5	R5	17.0000	18.0000	19.0000	20.0000
1 observations were deleted.					
OBS	RIRUM	V1	V2	V3	V4
1	R1	1.0000	2.0000	3.0000	4.0000
2	R2	5.0000	6.0000	7.0000	8.0000
4	R4	13.0000	14.0000	15.0000	16.0000
Exiting IML.					

그림 3.55 DELETE스테이트먼트를 실행한 결과

터세트에서 완전히 지운다는 것이 아니라, 관찰점들에 제거표시(mark record for deletion)를 하고 이 제거표시가 된 관찰점들에는 이후 수행되는 스테이트먼트들이 적용되지 않는다는 것이다. 이 관찰점들을 SAS데이터세트에서 완전히 지우려면 PURGE스테이트먼트를 사용한다.

DELETE스테이트먼트의 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-55.PRG에 수록되어 있다).

```

/* */
/*          EXAM3-55.PRG          */
/* */
/*          DELETE STATEMENT      */
/* */
LIBNAME CBSIML "A:\IMLDATA";
PROC IML;
RESET NOCENTER NOLOG NOPRINT
      PAGESIZE=80 DEFLIB='CBSIML';
FREE;
V1 = {1, 5, 9, 13, 17};
V2 = {2, 6, 10, 14, 18};
V3 = {3, 7, 11, 15, 19};
V4 = {4, 8, 12, 16, 20};
RIRUM = {"R1", "R2", "R3", "R4", "R5"};

```



```

CREATE DS9;
APPEND;
CLOSE DS9;
EDIT DS9;
READ ALL;
LIST ALL;
DELETE POINT 3;
LIST ALL;
DELETE ALL WHERE (V3 > 18);
LIST ALL;
CLOSE DS9;
QUIT;

```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.55와 같은 결과가 나타난다.

7) PURGE

PURGE 스테이트먼트는 DELETE 스테이트먼트에 의해 제거표시가 된 관찰점들을 SAS 데이터세트에서 지우고 남은 관찰점들에 새로운 관찰점번호(observation number)를 붙인다.

PURGE :

PURGE 스테이트먼트의 예제로 다음 SAS/IML 프로그램을 수행해 보자(이 SAS/IML 프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-56.PRG에 수록되어 있다).

```

/ *                               */
/ *          EXAM3-56.PRG         */
/ *                               */
/ *          PURGE STATEMENT     */
/ *                               */
LIBNAME CBSIML "A:\IMLDATA";
PROC IML;
RESET NOCENTER NOLOG NOPRINT
      PAGESIZE=80 DEFLIB='CBSIML';
FREE;
V1 = { 1, 5, 9, 13, 17 };
V2 = { 2, 6, 10, 14, 18 };
V3 = { 3, 7, 11, 15, 19 };
V4 = { 4, 8, 12, 16, 20 };

```

SAS						8:28 Tuesday, March 1, 1994 3
OBS	RIRUM	V1	V2	V3	V4	
1	R1	1.0000	2.0000	3.0000	4.0000	
2	R2	5.0000	6.0000	7.0000	8.0000	
3	R3	9.0000	10.0000	11.0000	12.0000	
4	R4	13.0000	14.0000	15.0000	16.0000	
5	R5	17.0000	18.0000	19.0000	20.0000	
1 observations were deleted.						
OBS	RIRUM	V1	V2	V3	V4	
1	R1	1.0000	2.0000	3.0000	4.0000	
2	R2	5.0000	6.0000	7.0000	8.0000	
3	R4	13.0000	14.0000	15.0000	16.0000	
4	R5	17.0000	18.0000	19.0000	20.0000	
1 observations were deleted.						
OBS	RIRUM	V1	V2	V3	V4	
1	R1	1.0000	2.0000	3.0000	4.0000	
2	R2	5.0000	6.0000	7.0000	8.0000	
3	R4	13.0000	14.0000	15.0000	16.0000	
Exiting IML.						

그림 3.56 PURGE스테이트먼트를 실행한 결과

```

RIRUM = {"R1", "R2", "R3", "R4", "R5"};
CREATE DS9;
APPEND;
CLOSE DS9;
EDIT DS9;
READ ALL;
LIST ALL;
DELETE POINT 3;
PURGE;
LIST ALL;
DELETE ALL WHERE (V3 > 18);
PURGE;
LIST ALL;
CLOSE DS9;
QUIT;
    
```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.56과 같은 결과가 나타난다.

8) FORCE

이 FORCE 스테이트먼트는 현재 열려 있는 모든 출력 SAS 데이터 세트들과 파일들을 닫은 다음 다시 연다(closing and then reopening the files). FORCE 스테이트먼트를 사용하는 목적은 현재 출력 버퍼에 있는(residing in output buffer) 데이터값들을 디스크에 저장하기 위한 것이다. 따라서 이 FORCE 스테이트먼트는 SAS/IML 작업을 하는 도중에 우연

히 또는 실수로 컴퓨터가 꺼질 때를 대비하는 것이다. FORCE스테이트먼트의 사용양식은 다음과 같다.

FORCE:

FORCE스테이트먼트의 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML 프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-57.PRG에 수록되어 있다).

```

/*                               */
/*          EXAM3-57.PRG          */
/*                               */
/*          FORCE STATEMENT       */
/*                               */
LIBNAME CBSIML "A:\IMLDATA";
PROC IML;
  RESET NOCENTER NOLOG NOPRINT
        PAGESIZE=80 DEFLIB='CBSIML';
FREE;
V1 = { 1, 5, 9, 13, 17 };
V2 = { 2, 6, 10, 14, 18 };
V3 = { 3, 7, 11, 15, 19 };
V4 = { 4, 8, 12, 16, 20 };
RIRUM = { "R1", "R2", "R3", "R4", "R5" };
CREATE DS9;
APPEND;
FORCE;
READ ALL;
LIST ALL;
DELETE POINT 3;
LIST ALL;
DELETE ALL WHERE (V3 > 18);
LIST ALL;
PURGE;
LIST ALL;
CLOSE DS9;
QUIT;

```

이 SAS/IML프로그램을 수행하면 OUTPUT윈도우에 그림 3.57과 같은 결과가 나타난다.

SAS		8:28 Tuesday, March 1, 1994 8			
OBS	RIRUM	V1	V2	V3	V4
1	R1	1.0000	2.0000	3.0000	4.0000
2	R2	5.0000	6.0000	7.0000	8.0000
3	R3	9.0000	10.0000	11.0000	12.0000
4	R4	13.0000	14.0000	15.0000	16.0000
5	R5	17.0000	18.0000	19.0000	20.0000
1 observations were deleted.					
OBS	RIRUM	V1	V2	V3	V4
1	R1	1.0000	2.0000	3.0000	4.0000
2	R2	5.0000	6.0000	7.0000	8.0000
4	R4	13.0000	14.0000	15.0000	16.0000
5	R5	17.0000	18.0000	19.0000	20.0000
1 observations were deleted.					
OBS	RIRUM	V1	V2	V3	V4
1	R1	1.0000	2.0000	3.0000	4.0000
2	R2	5.0000	6.0000	7.0000	8.0000
4	R4	13.0000	14.0000	15.0000	16.0000
1 observations were deleted.					
OBS	RIRUM	V1	V2	V3	V4
1	R1	1.0000	2.0000	3.0000	4.0000
2	R2	5.0000	6.0000	7.0000	8.0000
3	R4	13.0000	14.0000	15.0000	16.0000
Exiting IML.					

그림 3.57 FORCE스태이트먼트를 실행한 결과

3.7.10 범위피연산자의 디폴트

접근명령스태이트먼트들에 범위피연산자들을 지정하지 않은 경우의 디폴트는 다음과 같다.

- LIST : 현재 관찰점
- READ : 현재 관찰점
- FIND : 모든 관찰점
- REPLACE : 현재 관찰점
- APPEND : 항상 SAS스태이트먼트의 마지막에 위치(always at the end)
- DELETE : 현재 관찰점

3.7.11 SAS/IML과 SAS의 DATA스텝

SAS/IML환경에서 SAS데이터세트를 만드는 것과 SAS의 DATA스텝을 사용하는 것은 다음과 같은 차이가 있다.

- ① SAS/IML환경에서는 SAS데이터세트를 만들기 전에 변수의 속성(right attribute)에

맞추어 변수를 설정하여야 한다. 그러나 DATA스텝은 그 안에서 변수의 속성을 결정한다. 외부파일이나 SAS데이터세트는 숫자변수들과 문자변수들을 포함할 수 있다. SAS/IML환경에서는 사전에 아무런 지시가 없으면 읽어들이는 모든 변수들을 숫자변수들로 가정한다. 따라서 SAS/IML환경에서 새로운 SAS데이터세트를 만들 때에는 CREATE스테이트먼트를 사용해서 이 SAS데이터세트를 현재출력SAS 데이터세트로 선언하기 전에 문자변수들의 속성을 지정해야 한다.

- ② SAS데이터세트를 만들기 위해서 SAS/IML환경에서는 CREATE스테이트먼트로부터 시작하고 DATA스텝에서는 DATA스테이트먼트로부터 시작한다.
- ③ SAS/IML환경에서 한 관찰점을 SAS데이터세트로 보내기 위해서는 반드시 APPEND스테이트먼트를 사용해야 한다. 반면에 DATA스텝에서 한 관찰점을 SAS데이터세트로 보내기 위해서는 OUTPUT스테이트먼트를 사용한다. 그러나 DATA스텝에서는 OUTPUT스테이트먼트를 사용하지 않아도 자동적으로(automatically) 관찰점을 SAS데이터세트로 보내기도 한다.
- ④ DATA스텝에서는 한 관찰점의 데이터값들을 읽어들이어 이 데이터값들을 변환시키거나 새로운 변수들의 데이터값들을 계산해서 작성중인 SAS데이터세트에 그 관찰점의 데이터값들을 추가한다. 만일, 다음 관찰점이 있으면 이와 같은 과정을 반복하고 관찰점이 더 없으면 DATA스텝을 마친다. 이에 대한 자세한 내용은 [SAS4 TSA1] (pp. 59~60)을 참조하라. 반면에 SAS/IML환경에서는 이러한 반복과정을 DO DATA-END스테이트먼트를 사용해서 실행한다.
- ⑤ SAS/IML환경에서 QUIT스테이트먼트를 사용해서 SAS/IML환경에서 빠져나오는 경우가 아니면 CLOSE스테이트먼트를 사용해서 SAS데이터세트를 닫아야 한다. 그러나 DATA스텝에서는 DATA스텝의 마지막에서 자동적으로 SAS데이터세트를 닫는다.
- ⑥ 일반적으로 DATA스텝을 사용하는 것이 SAS/IML환경에서 SAS데이터세트를 작성하는 것보다 수행속도가 빠르다.

결론적으로 DATA스텝은 문제를 단순화해서 간단한 프로그램으로 해결한다. 반면에 SAS/IML은 대화적(interactive)이고 행렬을 다루는 데 있어서 강력한 능력이 있으므로 SAS/IML환경에서 SAS데이터세트를 만드는 데는 유연성(flexibility)이 있다.

DATA스텝을 사용해서 SAS데이터세트를 만드는 것과 SAS/IML환경에서 SAS데이터세트를 만드는 것을 비교하는 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-58.PRG에 수록되어 있다).

```

/* */
/*          EXAM3-58.PRG          */
/* */
/*          DATA STEP & SAS/IML  */
/* */

```

```

LIBNAME CBSIML "A:\IMLDATA";
PROC IML;
RESET NOCENTER NOLOG NOPRINT
      PAGESIZE=80 DEFLIB='CBSIML';
FREE;
V1 = { 1, 5, 9, 13, 17 };
V2 = { 2, 6, 10, 14, 18 };
V3 = { 3, 7, 11, 15, 19 };
V4 = { 4, 8, 12, 16, 20 };
RIRUM = { "R1", "R2", "R3", "R4", "R5" };
CREATE DS9;
APPEND;
CLOSE DS9;
QUIT;
/* USING DATA STEP */
TITLE1 "USING DATA STEP";
DATA CBSIML.DS9A;
  SET CBSIML.DS9;
  V5 = V1+V2+V3+V4;
  IF RIRUM="R5";
  RUN;
PROC PRINT DATA=CBSIML.DS9A;
  RUN;
/* USING SAS/IML */
TITLE1 "USING SAS/IML";
PROC IML;
RESET NOCENTER NOLOG NOPRINT
      PAGESIZE=80 DEFLIB='CBSIML';
FREE;
RIRUM = "12";
CREATE DS9B VAR{ RIRUM V1 V2 V3 V4 V5 };
USE DS9;
DO DATA;
  READ NEXT;
  V5 = V1+V2+V3+V4;
  IF RIRUM="R5" THEN APPEND;
END;
CLOSE DS9 DS9B;
QUIT;

```

USING SAS/IML							19:31 Tuesday, March 1, 1994 9
Exiting IML.							
USING DATA STEP							19:31 Tuesday, March 1, 1994 10
OBS	RIRUM	V1	V2	V3	V4	V5	
1	R5	17	18	19	20	74	
USING SAS/IML							19:31 Tuesday, March 1, 1994 11
End of file							
Exiting IML.							
USING SAS/IML							19:31 Tuesday, March 1, 1994 12
OBS	RIRUM	V1	V2	V3	V4	V5	
1	R5	17	18	19	20	74	

그림 3.58 DATA스텝과 SAS/IML의 비교

```
PROC PRINT DATA=CBSIMLDS9B;
  RUN;
```

이 SAS/IML 프로그램을 수행하면 OUTPUT 윈도우에 그림 3.58과 같은 결과가 나타난다.

3.8 외부파일의 입출력

SAS/IML 환경에서는 외부파일(external file)에서 데이터를 읽거나 외부파일에 데이터를 쓸 수가 있다. SAS/IML 환경에서 외부파일에 접근하는 스테이트먼트들은 SAS의 DATA스텝의 스테이트먼트들과 매우 비슷하다. SAS/IML 환경에서 외부파일로부터 데이터를 읽기 위해서는 INFILE 스테이트먼트, INPUT 스테이트먼트, 그리고 CLOSEFILE 스테이트먼트를 사용하고, 외부파일에 데이터를 쓰기 위해서는 FILE 스테이트먼트, PUT 스테이트먼트, 그리고 CLOSEFILE 스테이트먼트를 사용한다. 이 절에서는 이 스테이트먼트들의 사용법을 설명하고자 한다.

3.8.1 파일참조이름

[SAS4TSA1] (pp. 297~298)에서 설명하였듯이 SAS에서 ASCII 데이터세트의 대명사를 지정하기 위해서는 FILENAME 스테이트먼트를 사용한다. 처리하고 싶은 데이터세트가 SAS 스테이트먼트가 아닌 ASCII 데이터세트인 경우, FILENAME 스테이트먼트를 사용하면 INFILE 스테이트먼트에서 지정되는 파일이름을 간단히 표현할 수 있다. FILENAME

스테이트먼트로 지정되는 파일참조이름(file reference name)은 여덟 자 이하의 문자열로 구성된다. 다음 예를 살펴보자.

```
LIBNAME CBSLIB 'C:\SAS\SASCBS';
FILENAME ABC 'A:\CBSDATA\XYZ.DAT';
DATA CBSLIB.ABCD;
    INFILE ABC;
    INPUT X1 X2 X3;
RUN;
```

이 SAS프로그램을 실행하면 드라이브 A:의 서브디렉토리 CBSDATA의 ASCII데이터 세트 XYZ.DAT를 읽어서 드라이브 C:의 서브디렉토리 SAS\SASCBS에 영구SAS스테이트먼트 CBSLIB.ABCD를 만든다. FILENAME스테이트먼트는 SAS프로그램 중 어디에 있어도 좋으나, SAS스테이트먼트를 작성하는 DATA스텝 앞에 FILENAME스테이트먼트들을 모아놓는 것이 편리할 것이다.

SAS에서와 마찬가지로 SAS/IML환경에서도 ASCII데이터세트의 대명사인 파일참조이름을 지정하기 위해서는 FILENAME스테이트먼트를 사용한다. 처리하고 싶은 데이터세트가 SAS스테이트먼트가 아닌 ASCII데이터세트인 경우, 파일참조이름들을 사용하면 INFILE스테이트먼트, 또는 FILE스테이트먼트에서 지정되는 파일이름들을 간단히 표기할 수 있다. 다음과 같은 예를 살펴보자.

```
INFILE "A:\CBSDATA\ABC.DAT";
```

이 INFILE스테이트먼트는 다음 두 스테이트먼트를 수행한 것과 같다.

```
FILENAME ABC "A:\CBSDATA\ABC.DAT";
INFILE ABC;
```

FILENAME스테이트먼트에서 정의하지 않아도 SAS에 의해서 이미 지정되어 있는 파일참조이름들 PRINT와 LOG를 FILE스테이트먼트에 사용할 수 있다. 키워드 PRINT는 표준적 출력(standard print output)의 파일참조이름이고 키워드 LOG는 LOG윈도우출력(log output)의 파일참조이름이다.

3.8.2 텍스트파일과 이진파일

앞에서도 설명했듯이 에디터(editor)를 사용해서 만든 파일은 텍스트파일(text file)이다. 이 텍스트파일은 어떤 특수한 컴퓨터프로그래밍언어를 사용하지 않아도 디스플레이화면에 인간이 읽을 수 있는 형태로 나타나고 또한 편집을 할 수가 있다. 텍스트파일은 각 레코드(record)를 구분하는 특수한 문자들을 가지고 있다. SAS는 이 레코드들이 어떻게

정의되었는지를(how records are defined) 알기 때문에 독자들이 읽어들이야 할 텍스트파일이 있으면 간단한 INFILE스테이트먼트와 INPUT스테이트먼트를 사용해서 이 텍스트파일을 SAS시스템으로 읽어들이 수 있다.

텍스트파일이 아닌 파일은 어떤 특수한 컴퓨터프로그래밍언어를 사용하지 않으면 디스플레이화면에 인간이 이해하지 못할 수 있는 형태로 나타나거나, 아예 화면에 나타낼 수 없는 제어문자들(control characters)을 포함하고 있다. 이러한 파일을 이진파일(binary file)이라고 부른다. 특히 이진파일은 레코드분리자들(record separators)을 가지고 있지 않으며, 화면에 나타낼 수 없는 제어문자들을 포함하는 이진코드(binary codes)를 사용하기도 한다. 이진파일을 SAS시스템에 읽어들이고자 하면 INFILE스테이트먼트와 INPUT스테이트먼트의 특별한 옵션들을 사용해야 한다. 또한, SAS시스템의 데이터를 외부의 이진파일로 출력시키고자 하면 FILE스테이트먼트와 PUT스테이트먼트의 특별한 옵션들을 사용해야 한다. SAS시스템에서 이진파일을 다루는 방법에 대한 자세한 내용은 [IML6.03] (pp. 102~104)을 참조하라.

3.8.3 외부파일의 입력

외부파일을 SAS/IML환경으로 읽어들이려면 우선 INFILE스테이트먼트를 사용해서 입력파일(input file)을 열어야 한다. 또한 여러 개의 입력파일들이 열려 있는 경우에는 INFILE스테이트먼트를 사용해서 원하는 파일을 현재입력파일(current input file)로 만든다. INPUT스테이트먼트는 현재입력파일로부터 데이터를 읽어들이 것을 지시하며 또한 입력방식을 지시한다. 이 INFILE스테이트먼트나 INPUT스테이트먼트의 사용법은 SAS의 DATA스텝의 INFILE스테이트먼트나 INPUT스테이트먼트의 사용법과 같다. 따라서 INFILE스테이트먼트와 INPUT스테이트먼트의 자세한 사용법은 [SAS4TSA1]의 제 3장을 참조하라. 이 INFILE스테이트먼트와 INPUT스테이트먼트의 사용양식은 다음과 같다.

INFILE 파일이름 또는 파일참조이름 ;
INPUT 항목들;

INPUT스테이트먼트에 사용할 수 있는 항목들 중에서 본서를 읽는 독자들에게 필요한 것들은 다음과 같다.

- INPUT피연산자 : 현재의 위치(current position in the record)에서 읽어들이고자 하는 변수를 지정한다.
- 인포맷(informat) : 인포맷(input format)을 지정한다. 숫자데이터값의 표준적 인포맷은 w.d이고 문자데이터값의 표준적 인포맷은 w. 이다. 여기서 w는 필드의 크기이고 d는 소수점 이하의 자릿수이다.

- +n : 포인터를 n열만큼 앞으로 나가게 한다.
- @n : 현재의 데이터행의 제n번째 열로 포인터를 이동시킨다.
- / : 포인터를 다음 행의 맨 앞으로 이동시킨다.

SAS시스템은 데이터값들을 읽어들이거나 쓸 때 PUT스테이트먼트와 PUT스테이트먼트에서 포인터(pointer)라는 개념을 사용한다. 포인터는 어떤 열의 데이터값을 읽을 것인가를 가리키는 것이다. 포인터를 지정하는 곳으로 움직이기 위해서는 포인터컨트롤들을 사용한다. 포인터컨트롤들에 대한 자세한 내용은 [SAS4TSA1] (pp. 304~305)을 참조하라.

INFILE스테이트먼트로 연 외부파일을 닫기 위해서는 CLOSEFILE스테이트먼트를 사용한다. 모든 열린 파일들은 SAS/IML환경에서 벗어나면 저절로 닫히므로 CLOSE스테이트먼트와 마찬가지로 CLOSEFILE스테이트먼트도 반드시 사용해야 하는 것은 아니다. 그러나 사용하지 않는 파일을 열어놓는 것은 메모리를 낭비하는 것이다. 큰 프로그램, 또는 큰 파일들을 사용할 때는 반드시 사용하지 않는 파일들을 닫는 습관을 갖는 것이 좋겠다. CLOSEFILE스테이트먼트의 사용양식은 다음과 같다.

CLOSEFILE 파일이름 또는 파일참조이름 ;

외부파일을 읽어들이는 예제로 다음 SAS/IML프로그램을 수행해 보자(이 SAS/IML 프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-59.PRG에 수록되어 있다).

```

/* */
/*          EXAM3-59.PRG          */
/* */
/*          USING EXTERNAL FILE   */
/* */
LIBNAME CBSIML "A:\IMLDATA";
PROC IML;
RESET NOCENTER NOLOG NOPRINT
      PAGESIZE=80 DEFLIB="CBSIML";
FILENAME IN1 "A:\CHAPTER3\EXAM3-59.DAT";
FREE;
LAST = "1234567890";
FIRST = "12345";
CREATE DS10 VAR{ LAST FIRST ENGL MATH TOT };
INFILE IN1;
DO DATA;
      INPUT LAST $ FIRST $ ENGL MATH;

```

SAS						20:22 Tuesday, March 1, 1994 23
Exiting IML.						
SAS						20:22 Tuesday, March 1, 1994 24
OBS	LAST	FIRST	ENGL	MATH	TOT	
1	CHOI	BS	45	56	101	
2	KIM	HY	55	65	120	
3	PARK	CS	46	48	94	
4	CHUNG	WK	44	55	99	
5	LEE	KEM	46	51	97	

그림 3.59 SAS/IML환경에서 외부파일 읽기

```
TOT = ENGL + MATH;
APPEND;
END;
CLOSE DS10;
CLOSEFILE IN1;
QUIT;
PROC PRINT DATA=CBSIML.DS10;
RUN;
```

이 SAS/IML 프로그램을 수행한 결과는 그림 3.59와 같다.

3.8.4 외부파일로 출력

SAS/IML 환경에서 외부파일로 데이터를 출력시키려면 우선 FILE 스테이트먼트를 사용하여 출력파일(output file)을 열어야 한다. 또한, 여러 개의 출력파일들이 열려 있는 경우에는 FILE 스테이트먼트를 사용하여 원하는 파일을 현재출력파일(current output file)로 만든다. PUT 스테이트먼트는 현재출력파일에 데이터를 쓸 것을 지시하며 또한 출력방식을 지시한다. 이 FILE 스테이트먼트나 PUT 스테이트먼트의 사용법은 SAS의 DATA 스템의 FILE 스테이트먼트나 PUT 스테이트먼트의 사용법과 같다. 따라서 이 스테이트먼트들의 자세한 사용법은 [IML6.03] (pp. 108~110)을 참조하라. 이 FILE 스테이트먼트와 PUT 스테이트먼트의 사용양식은 다음과 같다.

FILE 파일이름 또는 파일참조이름 ; PUT 항목들 ;

PUT 스테이트먼트에 사용할 수 있는 항목들 중에서 본서를 읽는 독자들에게 필요한 것들은 다음과 같다.

- PUT피연산자 : 현재의 위치(current position in the record)에서 출력데이터세트에 쓰고자 하는 값을 지정한다. PUT피연산자로는 변수이름, 문자값(literal value), 그리고 괄호 안의 식(expression in parenthesis)이 있다.
- 포맷 : 포맷(format)을 지정한다. 숫자데이터값의 표준적 포맷은 w.d이고 문자데이터값의 표준적 포맷은 w.이다. 여기서 w는 필드의 폭이고 d는 소수점 이하의 자릿수이다.
- +n : 포인터를 n열만큼 앞으로 나가게 한다.
- @n : 현재의 데이터행의 제 n번째 열로 포인터를 이동시킨다.
- / : 포인터를 다음 행의 맨 앞으로 이동시킨다.

FILE스테이트먼트로 연 외부파일을 닫기 위해서도 CLOSEFILE스테이트먼트를 사용한다. CLOSEFILE스테이트먼트의 사용양식은 다음과 같다.

CLOSEFILE 파일이름 또는 파일참조이름 ;

SAS/IML환경에서 외부파일을 만드는 예제로 다음 SAS/IML프로그램을 수행해 보자 (이 SAS/IML프로그램은 디스켓 SAS4TSA4의 파일 EXAM3-60.PRG에 수록되어 있다).

```

/*                               */
/*           EXAM3-60.PRG         */
/*                               */
/*           WRITING AN EXTERNAL  */
/*                               */
LIBNAME CBSIML "A:\IMLDATA";
PROC IML;
RESET NOCENTER NOLOG NOPRINT
      PAGESIZE=80 DEFLIB="CBSIML";
FILENAME IN1 "A:\CHAPTER3\EXAM3-59.DAT";
FILENAME OUT1 "A:\CHAPTER3\EXAM3-60.DAT";
FREE;
INFILE IN1;
FILE OUT1;
DO DATA;
  INPUT LAST $ FIRST $ ENGL MATH;
  TOT = ENGL + MATH;
  PUT LAST $8. FIRST $8. ENGL MATH TOT;
END;

```

CHOI	BS	45	56	101
KIM	HY	55	65	120
PARK	CS	46	48	94
CHUNG	WK	44	55	99
LEE	KEN	46	51	97

그림 3.60 SAS/IML환경에서 만든 외부파일

```
CLOSEFILE IN1 OUT1;
QUIT;
```

이 SAS/IML 프로그램을 수행하면 드라이브 A:의 서브디렉토리 CHAPTER3에 EXAM3-60.DAT라는 파일이 생긴다. 이 파일을 보기 위해서 PGM윈도우의 커맨드행에 다음과 같은 명령을 내린다.

```
INC 'A:\CHAPTER3\EXAM3-60.DAT';
```

이 명령문을 실행하기 위해서 기능키 F10을 누르면 PGM윈도우에 그림 3.60과 같은 데이터세트가 나타난다.

3.8.5 기 타

[IML6.03]에 나오는 SAS/IML의 사용법에서 지금까지 설명하지 않은 것은 그래픽명령어들이다. 이 그래픽명령어들을 설명하는 것은 본서의 목적에도 어울리지 않을 뿐 아니라 지나치게 많은 지면을 차지할 것이기 때문이다. 이 그래픽명령어들의 자세한 사용법에 대해서는 [IML6.03]의 제 14장을 참조하라.

▣ 참고문헌 ▣

- Kennedy, W. J. & J. E. Gentle(1980), *Statistical Computing*, Marcel Dekker, New York.
- Newton, H. J. (1988), *Time Series: A Time Series Analysis Laboratory*, Wadsworth & Brooks, Pacific Grove.

전이함수모형분석

- 4.1 전이함수모형
- 4.2 충격반응가중함수의 추정
- 4.3 전이함수모형의 식별
- 4.4 전이함수모형의 추정
- 4.5 전이함수모형의 검진
- 4.6 전이함수모형의 예측
- 4.7 예 제 들

제 4 장 전이함수모형분석

단변량시계열분석에서는 시계열데이터 $\{y_t\}$ 가 자신의 과거항들과 오차항들에 의해서 설명되어진다. 그러나 분석대상의 시계열데이터 $\{y_t\}$ 와 밀접한 관계가 있는 한 개 이상의 다른 시계열데이터들이 있을 때에는 이 다른 시계열데이터들에 포함되어 있는 정보를 이용해서 시계열데이터 $\{y_t\}$ 를 분석하면 좀더 나은 결과를 얻을 수 있다. 이렇게 한 시계열데이터를 분석하기 위해서 다른 시계열데이터들을 이용하는 모형들 중의 하나가 전이함수모형(transfer function model)이다. 이 장에서는 전이함수모형화에 대해서 다루고자 한다.

4.1 전이함수모형

4.1.1 일반적 전이함수모형

전이함수모형은 주된 분석대상의 시계열데이터 $\{y_t\}$ 를 다른 시계열데이터들을 사용해서 좀더 효과적으로 분석하기 위한 모형들 중의 하나이다. 분석대상의 시계열데이터 $\{y_t\}$ 를 출력시계열데이터(output time series)라 하고 보조가 되는 시계열데이터를 입력시계열데이터(input time series)라 한다. 전이함수(transfer function)는 입력시계열데이터들이 가지고 있는 정보를 출력시계열데이터로 보내 주는 필터(filter)와 같은 역할을 한다. 출력시계열데이터에서 전이함수를 통하여 입력시계열데이터들의 현재항과 과거항들로 설명되지 못한 부분을 잡음항(noise term)이라 한다. 이 전이함수모형을 사용하는 주된 목적은 출력시계열데이터와 출력시계열데이터의 선행척도(leading indicator)라고 할 수 있는 입력시계열데이터들과의 동적 관계(dynamic relation)를 분석하여, 출력시계열데이터 자체만으로 구성된 단변량시계열모형을 개선하는 데 있다.

전이함수모형은 변수들 사이에 작인적 구조(causal structure)를 갖는 시계열데이터들

을 분석하는데 널리 적용되고 있다. 예를 들면, 어느 물품의 현재 매출량은 현재와 과거의 광고비에 영향을 받으므로, 매출량을 출력시계열데이터로 하고 광고비를 입력시계열데이터로 하는 전이함수모형을 구축함으로써 매출량만을 가지고 분석한 단변량시계열모형보다 좋은 분석을 할 수 있다.

4.1.2 단일입력단일출력전이함수모형

이후 이 장에서는 입력시계열데이터와 출력시계열데이터가 각각 하나인 경우만을 다루기로 하자. 즉, 입력시계열데이터도 하나이고 출력시계열데이터도 하나인 단일입력단일출력시스템(single input single output system)을 다루기로 하자. 우선 출력시계열데이터 $\{y_t\}$ 와 입력시계열데이터 $\{x_t\}$ 가 정상적이라고 가정하자. 이 단일입력단일출력시스템은 다음과 같은 전이함수모형으로 표현할 수 있다.

$$Y_t = h(x_t, x_{t-1}, \dots)$$

$$y_t = Y_t + n_t$$

여기서 $\{x_t\}$ 는 입력시계열데이터, $\{Y_t\}$ 는 실제 출력시계열데이터(real output time series), $\{y_t\}$ 는 관찰된 출력시계열데이터(observed output time series), $\{n_t\}$ 는 잡음을 나타내고 함수 $h(\cdot)$ 는 시스템의 동적 관계를 나타낸다. 일반적으로 함수 $h(\cdot)$ 를 알지 못하므로 유한개의 관찰점들이 주어졌을 때 이 전이함수모형을 구축하기 위해서는 함수 $h(\cdot)$ 를 좀더 간단한 형태로 가정할 필요가 있다. 이후 함수 $h(\cdot)$ 는 x_t, x_{t-1}, \dots 의 선형함수이고, 또한 시간에 대해 불변(time invariant)이라고 가정하자. 즉, 모든 시점 t 에 대해서 현재의 실제 출력시계열데이터는 입력시계열데이터의 현재와 과거의 관찰값들의 선형결합이라고 가정한다.

일반적인 단일입력단일출력전이함수모형은 다음과 같이 나타낼 수 있다.

$$y_t = \nu(B)x_t + n_t$$

$$\text{단, } \nu(z) = \sum_{j=-\infty}^{\infty} \nu_j z^j$$

여기서 B 는 후진연산자(backshift operator)이고, $\nu(z)$ 는 전이함수이며, 잡음시계열데이터 $\{n_t\}$ 와 입력시계열데이터 $\{x_t\}$ 는 서로 독립이다. 전이함수의 계수 ν_j 는 시차 j 에서의 충격반응가중값(impulse response weight)이라 부른다. 실제 상황에서 얻어지는 시계열데이터를 나타내는 전이함수모형이 다음과 같이 정의되는 안정성(stability)과 작인성(causality)을 만족한다고 가정하는 것은 일반적이다.

정의 4.1 충격반응가중값들의 무한급수가 절대수렴하면 전이함수모형은 안정적(stable)이라고 한다. 즉, 다음 식이 만족되면 이에 해당하는 전이함수모형은 안정적이다.

$$\sum_{j=-\infty}^{\infty} |\nu_j| < \infty \quad \square$$

정의 4.2 모든 $j(<0)$ 에 대해서 $\nu_j=0$ 이면, 이 전이함수모형은 작인적(causal)이라 정의한다. □

안정적인 시스템에서는 입력시계열데이터의 관찰값들이 유한하면 출력시계열데이터의 관찰값들도 유한하며, 작인성조건을 만족하는 시스템에서는 출력시계열데이터의 현재값이 입력시계열데이터의 현재와 과거의 관찰값들에만 영향을 받는다. 이후 이 장에서는 아래와 같이 안정적이고 작인적인 단일입력단일출력전이함수모형만을 다루기로 한다.

$$\begin{aligned} y_t &= \nu_0 x_t + \nu_1 x_{t-1} + \nu_2 x_{t-2} + \dots + n_t \\ &= \nu(B)x_t + n_t \\ \text{단, } \nu(z) &= \sum_{j=0}^{\infty} \nu_j z^j, \quad \sum_{j=0}^{\infty} |\nu_j| < \infty \end{aligned}$$

또한, 입력시계열데이터 $\{x_t\}$ 와 잡음시계열데이터 $\{n_t\}$ 는 서로 독립이라고 가정한다.

주어진 시계열데이터의 관찰점들이 유한개인 데 반하여 전이함수 $\nu(z)$ 는 무한개의 충격 반응가중값들을 갖고 있기 때문에 이 전이함수모형을 단순하게 만들 필요가 있다. 전이함수 $\nu(z)$ 가 무한차수의 멱함수이므로, 이 전이함수를 유한개의 모수를 갖는 유리함수형태로 근사시키는 것은 자연스러운 것이다. 따라서 이후 이 장에서는 다음과 같은 유리함수형태의 전이함수모형을 다루기로 한다.

$$\begin{aligned} y_t &= \nu(B)x_t + n_t \\ \nu(z) &= \frac{\omega_s(z)}{\delta_r(z)} z^b \\ \text{단, } \omega_s(z) &= \omega_0 - \omega_1 z - \dots - \omega_s z^s \\ \delta_r(z) &= 1 - \delta_1 z - \dots - \delta_r z^r \\ &\text{b는 지연모수(delay parameter)} \end{aligned}$$

이후 $\nu(z)$ 는 차수들이 (b, r, s) 인 전이함수라 부르기로 한다. 지연모수 b 는 입력시계열데이터의 충격이 출력시계열데이터에 영향을 미치는 데 걸리는 시차를 나타낸다. 앞에서도 언급했듯이 이 전이함수모형이 안정적이고 작인적이라고 가정한다. 이 시스템이 안정적이기 위해서 방정식 $\delta_r(z)=0$ 의 모든 근들이 단위원 밖에 존재해야 한다. 또한 잡음항 $\{n_t\}$ 는 ARMA모형을 만족한다고 가정하는 것이 일반적이다.

4.1.3 충격반응가중함수

전이함수의 $\omega_s(z)$ 와 $\delta_r(z)$, 그리고 지연모수 b 가 주어지면, 충격반응가중함수(impulse

response function) $\{\nu_j\}$ 는 다음 항등식으로부터 구할 수 있다.

$$\delta_r(z)\nu(z) = \omega_s(z)z^b$$

이 식은 다음과 같다.

$$\begin{aligned} (1 - \delta_1 z - \dots - \delta_r z^r)(\nu_0 + \nu_1 z + \nu_2 z^2 + \dots) \\ = (\omega_0 - \omega_1 z - \dots - \omega_s z^s)z^b \end{aligned}$$

이 식에 항등식의 원리를 적용해서 다음 식들을 얻는다.

$$\begin{aligned} \nu_j &= 0 & (j < b) \\ \nu_j &= \delta_1 \nu_{j-1} + \delta_2 \nu_{j-2} + \dots + \delta_r \nu_{j-r} + \omega_0 & (j = b) \\ \nu_j &= \delta_1 \nu_{j-1} + \delta_2 \nu_{j-2} + \dots + \delta_r \nu_{j-r} - \omega_{j-b} & (j = b+1, b+2, \dots, b+s) \\ \nu_j &= \delta_1 \nu_{j-1} + \delta_2 \nu_{j-2} + \dots + \delta_r \nu_{j-r} & (j > b+s) \end{aligned}$$

정리 4.1 안정적이고 작인적이며 차수들이 (b, r, s) 인 유리함수형태의 전이함수의 충격반응가중함수는 초기값들을 r 개의 충격반응가중값들 $\nu_{b+s}, \nu_{b+s-1}, \dots, \nu_{b+s-r+1}$ 로 하는 다음과 같은 r 차차분방정식을 만족한다.

$$\delta_r(B)\nu_j = 0 \quad (j > b+s) \quad \blacksquare$$

정리 4.1에 의해서 유리함수형태의 전이함수모형의 충격반응가중값들은 다음과 같은 성질들을 만족한다는 것을 알 수 있다.

- ① 첫 b 개의 충격반응가중값들 $\nu_0, \nu_1, \dots, \nu_{b-1}$ 은 0이다.
- ② 충격반응가중값들 $\nu_b, \nu_{b+1}, \dots, \nu_{b+s-r}$ 은 어떤 특별한 형태를 갖지 않는다.
- ③ 충격반응가중값들 $\nu_{b+s-r+1}, \nu_{b+s-r+2}, \dots, \nu_{b+s}$ 는 차분방정식 $\delta_r(B)\nu_j = 0$ 의 초기값들이다.
- ④ 임의의 $j (> b+s)$ 에 대해 차분방정식 $\delta_r(B)\nu_j = 0$ 을 만족한다.

이 성질들을 이용해서 유리함수형태의 전이함수의 차수들 (b, r, s) 를 다음과 같이 구할 수 있다.

- ① 임의의 $j (< b)$ 에 대해 $\nu_j = 0$ 이고 $\nu_b \neq 0$ 라는 성질을 이용해서 지연모수 b 를 결정한다.
- ② 지연모수 b 가 정해지면, ARMA모형에서 AR차수와 MA차수를 정하는 모형식별방법들과 비슷한 방법들을 사용해서 차수들 r 과 s 를 정한다.

전이함수의 $\omega_s(z)$ 와 $\delta_r(z)$, 그리고 지연모수 b 가 주어진 경우, SAS/IML을 사용해서 충격반응가중함수 $\{\nu_j\}$ 를 구해 보자. 첫 번째 예제로써 다음과 같은 전이함수의 충격반응가중함수를 구해 보자.

$$\nu(z) = \frac{1.0 + 0.4z}{1 - 0.8z} z^2$$

이 전이함수의 충격반응가중합수를 구하기 위해 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM4-1.PRG에 수록되어 있다).

```

/* */
/* */          EXAM4-1.PRG          */
/* */
/* */          CALCULATING IMPULSE RESPONSE WEIGHTS          */
/* */          MODEL = IRF1          */
/* */
OPTIONS PAGESIZE=60;
LIBNAME TSA4DATA 'A:\TSA4DATA';
PROC IML;
RESET NOLOG NONAME NOCENTER;
/* */   INPUT DATA   */
DELTA1 = { 1.0 -0.8 };
OMEGA1 = { 0.0 0.0 1.0 0.4 };
NUMAT = RATIO(DELTA1, OMEGA1, 31, 1);
PRINT , 'NU =', NUMAT,;
NUMAT = NUMAT';
NUNAME = { NU };
CREATE TSA4DATA.IRF1 FROM NUMAT [COLNAME=NUNAME];
APPEND FROM NUMAT;
CLOSE TSA4DATA.IRF1;
QUIT;
DATA EXAM41;
    SET TSA4DATA.IRF1;
    NULAG = _N_ - 1;
    OUTPUT;
    RUN;
PROC PLOT DATA=EXAM41;
    PLOT NU*NULAG/ VPOS=25 HAXIS=0 TO 30 BY 1
        VREF=0 VAXIS=-0.2 TO 1.4 BY 0.2;
    RUN;

```

이 SAS프로그램을 수행한 결과가 그림 4.1이다. 그림 4.1의 부분 A에는 PROC IML을 수행해서 구한 충격반응가중합수가 실려 있다. 첫 번째 DATA스텝에서 이 충격반응가중합수를 SAS데이터라이브러리 TSA4DATA의 영구SAS데이터세트 IRF1에 저장한다. PROC PLOT를 사용해서 이 충격반응가중합수를 그린 결과가 그림 4.1의 부분 B이다.


```

/*          CALCULATING IMPULSE RESPONSE WEIGHTS          */
/*          MODEL = IRF2                                   */
/*                                                                 */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOLOG NOCENTER NONAME;
/*          INPUT DATA          */
DELTA2 = { 1.0 -0.5 -0.3 };
OMEGA1 = { 0.0 0.0 0.0 1.0 0.7 };
NUMAT = RATIO(DELTA2, OMEGA1, 31, 1);
PRINT , 'NU = ' , NUMAT,;
NUMAT = NUMATt ;
NUNAME = { NU };
CREATE TSA4DATA.IRF2 FROM NUMAT [COLNAME=NUNAME];
APPEND FROM NUMAT;
CLOSE TSA4DATA.IRF2;
QUIT;
DATA EXAM42;
    SET TSA4DATA.IRF2;
    NULAG=_N_-1;
    OUTPUT;
    RUN;
PROC PLOT DATA=EXAM42;
    PLOT NU*NULAG/ VPOS=25 HAXIS=0 TO 30 BY 1
        VREF=0 VAXIS=-0.2 TO 1.4 BY 0.2;
    RUN;

```

이 SAS프로그램을 수행한 결과가 그림 4.2이다. 그림 4.2의 부분 A에는 PROC IML을 수행해서 구한 충격반응가중함수가 실려 있다. DATA스텝에서 이 충격반응가중함수를 SAS데이터라이브러리 TSA4DATA의 영구SAS데이터세트 IRF2에 저장한다. PROC PLOT를 사용해서 이 충격반응가중함수를 그린 결과가 그림 4.2의 부분 B이다.

4.1.4 예 제 들

이 소절에서는 두 개의 유리함수형태의 전이함수모형들에서 시계열데이터들을 발생시키기로 한다. 우선 두 예제들 모두에서 잡음항 $\{n_t\}$ 는 다음과 같은 ARMA(1, 2)모형에서 생성시키기로 한다.

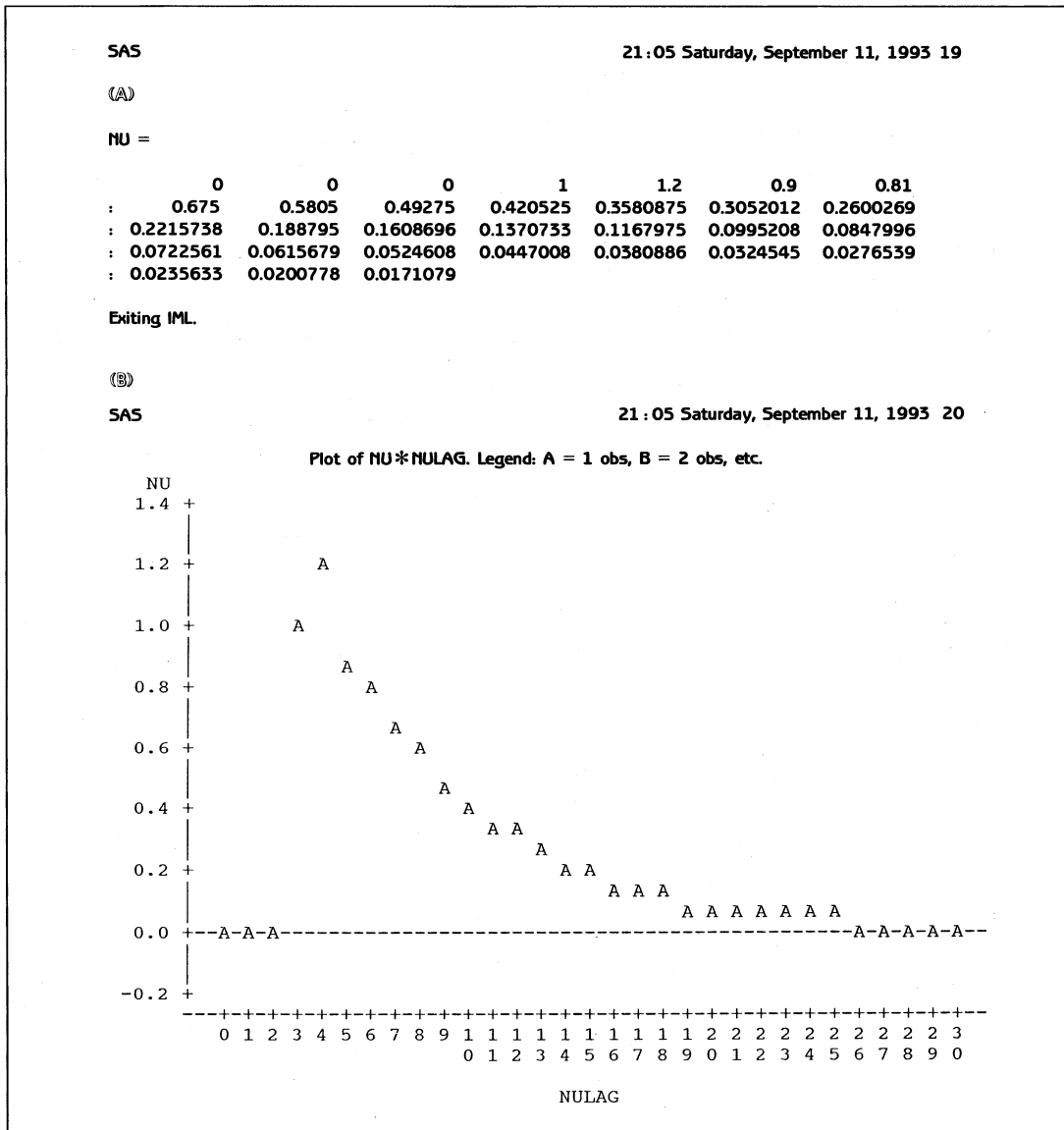


그림 4.2 차수들이 (3, 2, 1)인 전이함수모형의 충격반응가중함수

$$n_t = -0.7n_{t-1} + v_t - 0.5v_{t-1} - 0.3v_{t-2}$$

여기서 $\{v_t\}$ 는 평균이 0이고 분산이 1인 정규백색잡음과정이다. 또한 입력시계열데이터 $\{x_t\}$ 는 다음과 같은 ARMA(1, 1)모형에서 생성시킨다.

$$x_t = 0.71x_{t-1} + a_t + 0.2a_{t-1}$$

여기서 $\{a_t\}$ 는 평균이 0이고 분산이 1인 정규백색잡음과정이다.

첫 번째 예제는 다음과 같은 전이함수를 갖는 전이함수모형에서 시계열데이터를 발생시키기로 한다.

$$v(z) = \frac{1+0.4z}{1-0.8z} z^2$$

이 전이함수는 EXAM4-1.PRG에서 사용한 것이다. 이 전이함수모형에서 시계열데이터를 발생시키기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM4-3.PRG에 수록되어 있다).

```

/*
/*
/*
/*
/*
/*
/*
/*
LIBNAME TSA4DATA 'A:\TSA4DATA';
DATA EXAM43;
    DUMT=0;
    START:DUMT+1;
        ERR=RANNOR(7654321);
        ERR1 = LAG1(ERR);
        ERR2 = LAG1(ERR1);
        IF ERR1=. THEN ERR1=0;
        IF ERR2=. THEN ERR2=0;
        IF NOISE1=. THEN NOISE1=0;
        NOISE = -0.7*NOISE1+ERR-0.5*ERR1-0.3*ERR2;
        T = DUMT;
        OUTPUT;
        NOISE1=NOISE;
        IF DUMT<310 THEN GO TO START;
    KEEP T NOISE;
RUN;
DATA EXAM43B;
    DO T1 = 1 TO 310;
        ERROR=3.0*RANNOR(1234567);
        ERROR1 = LAG1(ERROR);
        IF ERROR1=. THEN ERROR1=0;
        IF X1=. THEN X1=0;
        X=0.71*X1+ERROR+0.20*ERROR1;
        OUTPUT;
        X1=X;
    END;

```

```

KEEP X;
RUN;
DATA EXAM43C;
MERGE EXAM43 EXAM43B;
X2 = LAG2(X);
X3 = LAG1(X2);
NOISE1 = LAG1(NOISE);
IF X2=. THEN X2=0;
IF X3=. THEN X3=0;
IF NOISE1=. THEN NOISE1=0;
IF T <= 310
THEN DO;
IF Y1=. THEN Y1=0.0;
Y = 0.8*Y1 + X2 + 0.4*X3 + NOISE - 0.8*NOISE1;
OUTPUT;
Y1=Y;
END;
KEEP T X Y;
RUN;
DATA TSA4DATA.TFM1;
SET EXAM43C;
T = T - 110;
IF (T < 0) OR (T > 120) THEN DELETE;
RUN;
TITLE 'TRANSFER FUNCTION MODEL 1';
SYMBOL1 C=BLACK V=NONE L=1 I=JOIN;
SYMBOL2 C=BLACK V=NONE L=4 I=JOIN;
AXIS1 ORDER=( 0 TO 120 BY 10 ) LABEL=('TIME');
AXIS2 ORDER=( -15 TO 35 BY 5 );
AXIS3 ORDER=( -35 TO 15 BY 5 );
PROC GPLOT;
PLOT X*T/ HAXIS=AXIS1 VAXIS=AXIS2;
PLOT2 Y*T/ HAXIS=AXIS1 VAXIS=AXIS3;
RUN;

```

이 SAS프로그램의 첫 번째 DATA스텝은 잡음시계열데이터 $\{n_t\}$ 를 생성하기 위한 것이고, 두 번째 DATA스텝은 입력시계열데이터 $\{x_t\}$ 를 생성하기 위한 것이다. 또한 세 번째 DATA스텝은 출력시계열데이터 $\{y_t\}$ 를 생성하기 위한 것이고, 네 번째 DATA스텝은 이 입력 시계열데이터와 출력시계열데이터에서 가운데에 있는 120개의 관찰점들을 SAS데이


```

LIBNAME TSA4DATA 'A:\TSA4DATA';
DATA EXAM44;
  DUMT=0;
  START:DUMT+1;
    ERR=RANNOR(7654321);
    ERR1 = LAG1(ERR);
    ERR2 = LAG1(ERR1);
    IF ERR1=. THEN ERR1=0;
    IF ERR2=. THEN ERR2=0;
    IF NOISE1=. THEN NOISE1=0;
    NOISE = -0.7*NOISE1+ERR-0.5*ERR1-0.3*ERR2;
    T = DUMT;
    OUTPUT;
    NOISE1=NOISE;
    IF DUMT<310 THEN GO TO START;
  KEEP T NOISE;
RUN;
DATA EXAM44B;
  DO T1 = 1 TO 310;
    ERROR=3.0*RANNOR(1234567);
    ERROR1 = LAG1(ERROR);
    IF ERROR1=. THEN ERROR1=0;
    IF X1=. THEN X1=0;
    X=0.71*X1+ERROR+0.20*ERROR1;
    OUTPUT;
    X1=X;
  END;
  KEEP X;
RUN;
DATA EXAM44C;
  MERGE EXAM44 EXAM44B;
  X3 = LAG3(X);
  X4 = LAG1(X3);
  NOISE1 = LAG1(NOISE);
  NOISE2 = LAG1(NOISE1);
  IF X3=. THEN X3=0;
  IF X4=. THEN X4=0;
  IF NOISE1=. THEN NOISE1=0;
  IF NOISE2=. THEN NOISE2=0;

```

```

IF T <= 310
  THEN DO;
    IF Y1=. THEN Y1=0.0;
    IF Y2=. THEN Y2=0.0;
    Y = 0.5*Y1 + 0.3*Y2 + X3 + 0.7*X4
      + NOISE - 0.5*NOISE1 - 0.3*NOISE2;
    OUTPUT;
    Y2=Y1;
    Y1=Y;
  END;
KEEP T X Y;
RUN;

DATA TSA4DATA.TFM2;
SET EXAM44C;
  T = T-110;
  IF (T < 0) OR (T > 100) THEN DELETE;
RUN;

TITLE 'TRANSFER FUNCTION MODEL 2';
SYMBOL1 C=BLACK V=NONE L=1 I=JOIN;
SYMBOL2 C=BLACK V=NONE L=4 I=JOIN;
AXIS1 ORDER=( 0 TO 100 BY 10 ) LABEL=('TIME');
AXIS2 ORDER=( -15 TO 35 BY 5 );
AXIS3 ORDER=( -35 TO 15 BY 5 );
PROC GLOT;
  PLOT X*T / HAXIS=AXIS1 VAXIS=AXIS2;
  PLOT2 Y*T / HAXIS=AXIS1 VAXIS=AXIS3;
RUN;

```

SAS프로그램 EXAM4-3.PRG에서와 마찬가지로 이 SAS프로그램의 첫 번째 DATA스텝은 잡음시계열데이터 $\{n_t\}$ 를 생성하기 위한 것이고, 두 번째 DATA스텝은 입력시계열데이터 $\{x_t\}$ 를 생성하기 위한 것이다. 또한 세 번째 DATA스텝은 출력시계열데이터 $\{y_t\}$ 를 생성하기 위한 것이고, 네 번째 DATA스텝은 이 입력시계열데이터와 출력시계열데이터에서 가운데에 있는 100개의 관찰점들을 SAS데이터라이브러리 TSA4DATA의 영구SAS데이터세트 TFM2에 저장하기 위한 것이다. 첫 번째 PROC스텝은 입력시계열데이터 $\{x_t\}$ 와 출력시계열데이터 $\{y_t\}$ 의 시계열산포도를 그리기 위한 것이다. 이 SAS프로그램을 수행한 결과가 그림 4.4이다. 그림 4.4에서 실선은 입력시계열데이터 $\{x_t\}$ 를 표시하고 점선은 출력시계열데이터 $\{y_t\}$ 를 나타낸다.

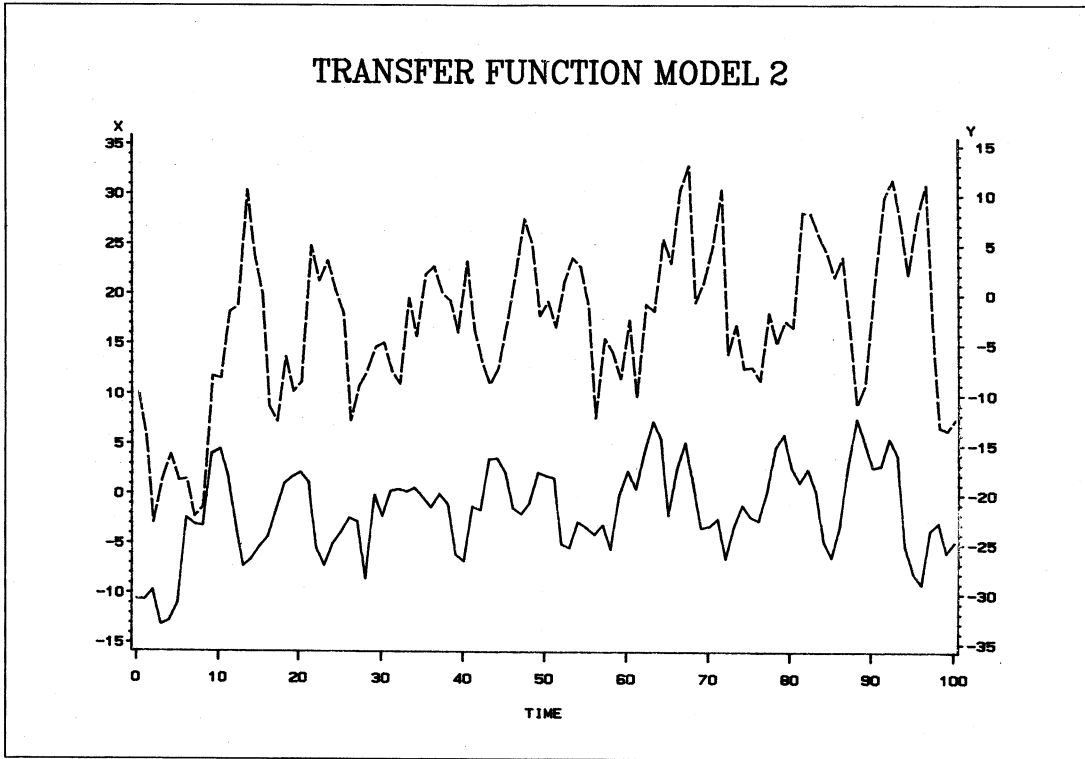


그림 4.4 차수들이 (3, 2, 1)인 전이함수모형에서 생성한 시계열데이터

4.1.5 교차상관함수

교차상관함수(cross-correlation function, CCRF)는 두 시계열데이터들 사이의 상관관계의 정도(strength)와 방향(direction)을 측정하는 척도이다. 두 시계열데이터들 $\{x_i\}$ 와 $\{y_i\}$ 가 주어졌다고 하자. 만일, $\{x_i\}$ 와 $\{y_i\}$ 가 각각 정상적이며, x_t 와 y_t 의 공분산이 시차 $(s-t)$ 만의 함수이면 이 두 시계열데이터 $\{x_i\}$ 와 $\{y_i\}$ 는 결합정상적(joint stationary)이라고 한다. 물론 이 결합정상성은 약한 의미의 결합정상성을 의미한다. 만일, 두 시계열데이터들 $\{x_i\}$ 와 $\{y_i\}$ 가 결합정상적이면 $\{x_i\}$ 와 $\{y_i\}$ 사이의 교차공분산함수(cross-covariance function, CCVF)는 다음과 같이 정의된다.

$$\sigma_{xy}(k) = E[(x_{t+k} - \mu_x)][(y_t - \mu_y)] \quad (k=0, \pm 1, \pm 2, \dots)$$

여기서 μ_x 와 μ_y 는 각각 $\{x_i\}$ 와 $\{y_i\}$ 의 평균들이다. 또한, 두 시계열데이터들 $\{x_i\}$ 와 $\{y_i\}$ 사이의 교차상관함수는 다음과 같이 정의된다.

$$\rho_{xy}(k) = \frac{\sigma_{xy}(k)}{\sqrt{\sigma_{xx}(0) \sigma_{yy}(0)}} \quad (k=0, \pm 1, \pm 2, \dots)$$

교차공분산함수 $\{\sigma_{xy}(k)\}$ 와 교차상관함수 $\{\rho_{xy}(k)\}$ 는 각각 자기공분산함수와 자기상관함

수를 일반화시킨 것이다. 왜냐하면 $\{\sigma_{xx}(k)\}$ 는 단변량시계열데이터 $\{x_t\}$ 의 자기공분산함수이고, $\{\rho_{yy}(k)\}$ 는 단변량시계열데이터 $\{y_t\}$ 의 자기상관함수이기 때문이다. 자기상관함수 $\{\rho_{xx}(k)\}$ 는 기점 0에 대해서 대칭이다. 즉, 다음 식이 성립한다.

$$\rho_{xx}(k) = \rho_{xx}(-k) \quad (k = 0, \pm 1, \pm 2, \dots)$$

그러나 교차상관함수는 대칭이 아니다.

$$\rho_{xy}(k) \neq \rho_{xy}(-k)$$

반면에 교차상관함수는 다음과 같은 성질을 갖고 있다.

$$\rho_{xy}(k) = \rho_{yx}(-k) \quad (k = 0, \pm 1, \pm 2, \dots)$$

따라서 교차상관함수는 확률변수들 $\{x_t\}$ 와 $\{y_t\}$ 의 상관관계의 정도뿐만 아니라 방향도 나타낸다. 따라서 시계열데이터들 $\{x_t\}$ 와 $\{y_t\}$ 사이의 관계를 알기 위해서는 양의 시차 ($k > 0$)뿐만 아니라 음의 시차 ($k < 0$)에서도 교차상관함수 $\{\rho_{xy}(k)\}$ 를 조사해야 한다.

4.1.6 표본교차상관함수

입력시계열데이터 $\{x_t | t=1, \dots, T\}$ 와 출력시계열데이터 $\{y_t | t=1, \dots, T\}$ 가 주어졌을 때, 즉, T 개의 관찰점들이 주어졌을 때, 교차공분산함수는 다음과 같은 표본교차공분산함수로 추정한다.

$$\hat{\sigma}_{xy}(k) = \begin{cases} \frac{1}{T} \sum_{t=1}^{T-k} (x_{t+k} - \bar{x})(y_t - \bar{y}) & (k \geq 0) \\ \frac{1}{T} \sum_{t=1-k}^T (x_{t+k} - \bar{x})(y_t - \bar{y}) & (k < 0) \end{cases}$$

여기서 \bar{x} 와 \bar{y} 는 각각 입력시계열데이터 $\{x_t\}$ 와 출력시계열데이터 $\{y_t\}$ 의 표본평균이다. 또한, 교차상관함수는 다음과 같은 표본교차상관함수로 추정한다.

$$\hat{\rho}_{xy}(k) = \frac{\hat{\sigma}_{xy}(k)}{\sqrt{\hat{\sigma}_{xx}(0) \hat{\sigma}_{yy}(0)}} \quad (k=0, \pm 1, \pm 2, \dots)$$

SAS에서는 SAS/IML의 내장함수 COVLAG를 사용하거나 SAS/ETS의 PROC ARIMA의 IDENTIFY스태이트먼트의 옵션 OUTCOV=을 사용해서 교차공분산함수와 교차상관함수를 추정할 수 있다. 첫 번째 방법으로 SAS/IML의 내장함수 COVLAG를 사용해서 교차공분산함수와 교차상관함수를 추정하기 위해서 다음 SAS프로그램을 수행해 보자 (이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM4-5.PRG에 수록되어 있다).

```

/* */
/*          EXAM4-5.PRG          */
/* */
/*          GENERATING A SAMPLE CCRF          */
/*          USING SAS/IML FUNCTION COVLAG          */
/*          MODEL = TFM1          */
/* */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOLOG NOCENTER;
/* CALL TSA4DATA.TFM1 */
USE TSA4DATA.TFM1;
READ ALL VAR{X Y} INTO XY;
CLOSE TSA4DATA.TFM1;
/* MAKE CCVF */
ACVMF = COVLAG(XY, 6);
SIGYX0 = ACVMF[2,1];
SIGYX1 = ACVMF[2,3];
SIGYX2 = ACVMF[2,5];
SIGYX3 = ACVMF[2,7];
SIGYX4 = ACVMF[2,9];
SIGYX5 = ACVMF[2,11];
SIGXY0 = ACVMF[1,2];
SIGXY1 = ACVMF[1,4];
SIGXY2 = ACVMF[1,6];
SIGXY3 = ACVMF[1,8];
SIGXY4 = ACVMF[1,10];
SIGXY5 = ACVMF[1,12];
/* PRINT CCVF */
PRINT , 'CROSS-COVARIANCES', ,
      SIGXY5 " " SIGXY4 " " SIGXY3 " " SIGXY2,
      SIGXY1 " " SIGXY0 " " SIGYX0 " " SIGYX1,
      SIGYX2 " " SIGYX3 " " SIGYX4 " " SIGYX5,;
/* CALCULATE ACRMF */
RHO0 = ACVMF[ , {1 2}];
Q = INV(SQRT(DIAG(RHO0)));
RHO0 = Q*RHO0*Q;
RHO1 = Q*ACVMF[ , {3 4}]*Q;

```


SAS		19:30 Sunday, September 12, 1993 2	
CROSS-COVARIANCES			
SIGXY5	SIGXY4	SIGXY3	SIGXY2
-5.445837	-3.644728	-0.341121	3.2297292
SIGXY1	SIGXY0	SIGYX0	SIGYX1
5.1665853	9.7679878	9.7679878	15.917971
SIGYX2	SIGYX3	SIGYX4	SIGYX5
23.716057	20.078212	11.698369	5.0080414
CROSS-CORRELATIONS			
RHOXY5	RHOXY4	RHOXY3	RHOXY2
-0.19245	-0.128801	-0.012055	0.1141351
RHOXY1	RHOXY0	RHOYX0	RHOYX1
0.1825815	0.3451901	0.3451901	0.5625239
RHOYX2	RHOYX3	RHOYX4	RHOYX5
0.8380998	0.7095423	0.4134077	0.1769788
Exiting IML.			

그림 4.5 SAS/IML을 사용해서 얻은 교차상관함수

```

RHO2 = Q*ACVMF[ , {5 6}]*Q;
RHO3 = Q*ACVMF[ , {7 8}]*Q;
RHO4 = Q*ACVMF[ , {9 10}]*Q;
RHO5 = Q*ACVMF[ , {11 12}]*Q;
/* CALCULATE CCRF */
RHOYX0 = RHO0[2, 1];
RHOYX1 = RHO1[2, 1];
RHOYX2 = RHO2[2, 1];
RHOYX3 = RHO3[2, 1];
RHOYX4 = RHO4[2, 1];
RHOYX5 = RHO5[2, 1];
RHOXY0 = RHO0[1, 2];
RHOXY1 = RHO1[1, 2];
RHOXY2 = RHO2[1, 2];
RHOXY3 = RHO3[1, 2];
RHOXY4 = RHO4[1, 2];
RHOXY5 = RHO5[1, 2];
/* PRINT CCRF */
PRINT, 'CROSS-CORRELATIONS' , ,
      RHOXY5 " " RHOXY4 " " RHOXY3 " " RHOXY2,
      RHOXY1 " " RHOXY0 " " RHOYX0 " " RHOYX1,
      RHOYX2 " " RHOYX3 " " RHOYX4 " " RHOYX5,;
QUIT;

```

SAS/IML의 함수 COVLAG의 자세한 사용법은 본서의 제7.1절과 [IML6.03](p. 232)을 참조하라. 이 SAS프로그램을 수행한 결과가 그림 4.5이다.

두 번째 방법으로 SAS/ETS의 PROC ARIMA의 IDENTIFY스테이트먼트의 옵션 OUTCOV=을 사용해서 교차공분산함수와 교차상관함수를 추정하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM4-6.PRG에 수록되어 있다).

```

/* */
/* */
/* */
/* */
/* */
/* */
/* */
/* */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=72;
DATA EXAM46;
    SET TSA4DATA.TFM1;
    RUN;
PROC ARIMA;
    IDENTIFY VAR=Y CROSS=(X)
            OUTCOV=EXAM46B NOPRINT;
    RUN;
PROC PRINT DATA=EXAM46B;
    RUN;

```

이 SAS프로그램의 PROC ARIMA의 IDENTIFY스테이트먼트의 옵션 VAR=에는 출력시계열데이터 Y를 할당하고 옵션 CROSS=에는 입력시계열데이터 X를 할당하였으며 옵션 OUTCOV=에는 표본교차상관함수를 저장할 SAS데이터세트 EXAM46B를 할당하였다. PROC ARIMA의 IDENTIFY스테이트먼트의 옵션 OUTCOV=에 지정한 SAS데이터세트에는 다음 변수들이 포함된다.

- ① 변수 LAG는 공분산변수(covariance variable)들의 값에 해당하는 시차들을 나타낸다.
- ② 변수 VAR는 IDENTIFY스테이트먼트의 옵션 VAR=에 지정된 변수의 이름을 나타낸다.
- ③ 변수 CROSSVAR는 IDENTIFY스테이트먼트의 옵션 CROSS=에 지정되어 교차공분산함수를 계산하도록 하는 변수의 이름을 나타낸다. 만일, 자기공분산함수가 계산되는 경우에는 이 변수의 값이 공란이다.

19:30 Sunday, September 12, 1993 3

LAG	VAR	CROSSVAR	N	COV	CORR	STDERR	INVCORR	PARTCORR
1	0	Y	121	42.9210	1.00000	0.00000	1.00000	1.00000
2	1	Y	120	21.2031	0.49400	0.09091	-0.30588	0.49400
3	2	Y	119	20.4637	0.47678	0.11090	-0.34780	0.30787
4	3	Y	118	5.3966	0.12573	0.12671	0.27981	-0.27674
5	4	Y	117	5.8656	0.13666	0.12774	-0.11166	0.04394
6	5	Y	116	-0.0965	-0.00225	0.12894	0.02107	0.01351
7	6	Y	115	-4.3304	-0.10089	0.12894	0.01122	-0.22207
8	7	Y	114	-5.7328	-0.13357	0.12959	-0.03214	0.00548
9	8	Y	113	-5.1230	-0.11936	0.13072	0.10887	0.08736
10	9	Y	112	0.1027	0.00239	0.13162	-0.06881	0.09910
11	10	Y	111	2.8972	0.06750	0.13162	0.01596	0.08489
12	11	Y	110	3.3452	0.07794	0.13191	-0.03540	-0.05483
13	12	Y	109	4.2126	0.09815	0.13229	-0.05493	0.01542
14	13	Y	108	4.5233	0.10539	0.13289	0.17380	0.05052
15	14	Y	107	6.4912	0.15124	0.13358	-0.10679	0.04258
16	15	Y	106	5.7960	0.13504	0.13498	-0.08303	0.04306
17	16	Y	105	-0.3969	-0.00925	0.13610	0.11064	-0.18754
18	17	Y	104	-0.5262	-0.01226	0.13610	-0.04355	0.04454
19	18	Y	103	-2.4128	-0.05621	0.13611	0.08205	0.08220
20	19	Y	102	-0.0285	-0.00066	0.13630	-0.10749	-0.04279
21	20	Y	101	-0.7004	-0.01632	0.13630	-0.06325	0.04377
22	21	Y	100	-3.9848	-0.09284	0.13632	0.19773	-0.12823
23	22	Y	99	2.0682	0.04819	0.13684	-0.07423	0.19589
24	23	Y	98	2.1003	0.04893	0.13698	-0.03809	0.06898
25	24	Y	97	9.0045	0.20979	0.13712	-0.00151	0.00264
26	-24	Y	X 97	4.3382	0.15331	0.09091	.	.
27	-23	Y	X 98	5.4896	0.19400	0.09091	.	.
28	-22	Y	X 99	5.0048	0.17687	0.09091	.	.
29	-21	Y	X 100	3.5890	0.12683	0.09091	.	.
30	-20	Y	X 101	0.7119	0.02516	0.09091	.	.
31	-19	Y	X 102	-2.8080	-0.09923	0.09091	.	.
32	-18	Y	X 103	-1.1251	-0.03976	0.09091	.	.
33	-17	Y	X 104	-0.4250	-0.01502	0.09091	.	.
34	-16	Y	X 105	-0.2375	-0.00839	0.09091	.	.
35	-15	Y	X 106	-1.9778	-0.06989	0.09091	.	.
36	-14	Y	X 107	0.3751	0.01326	0.09091	.	.
37	-13	Y	X 108	2.4154	0.08536	0.09091	.	.
38	-12	Y	X 109	4.4751	0.15814	0.09091	.	.
39	-11	Y	X 110	3.6768	0.12994	0.09091	.	.
40	-10	Y	X 111	2.9779	0.10524	0.09091	.	.
41	-9	Y	X 112	2.9817	0.10537	0.09091	.	.
42	-8	Y	X 113	2.6976	0.09533	0.09091	.	.
43	-7	Y	X 114	0.4927	0.01741	0.09091	.	.
44	-6	Y	X 115	-1.8305	-0.06469	0.09091	.	.
45	-5	Y	X 116	-5.4458	-0.19245	0.09091	.	.
46	-4	Y	X 117	-3.6447	-0.12880	0.09091	.	.
47	-3	Y	X 118	-0.3411	-0.01205	0.09091	.	.
48	-2	Y	X 119	3.2297	0.11414	0.09091	.	.
49	-1	Y	X 120	5.1666	0.18258	0.09091	.	.
50	0	Y	X 121	9.7680	0.34519	0.09091	.	.
51	1	Y	X 120	15.9180	0.56252	0.09091	.	.
52	2	Y	X 119	23.7161	0.83810	0.09091	.	.
53	3	Y	X 118	20.0782	0.70954	0.09091	.	.
54	4	Y	X 117	11.6984	0.41341	0.09091	.	.
55	5	Y	X 116	5.0080	0.17698	0.09091	.	.
56	6	Y	X 115	0.9475	0.03348	0.09091	.	.
57	7	Y	X 114	-1.2686	-0.04483	0.09091	.	.
58	8	Y	X 113	-4.9262	-0.17409	0.09091	.	.
59	9	Y	X 112	-5.8788	-0.20775	0.09091	.	.
60	10	Y	X 111	-5.1888	-0.18337	0.09091	.	.
61	11	Y	X 110	-1.8169	-0.06421	0.09091	.	.
62	12	Y	X 109	1.5283	0.05401	0.09091	.	.
63	13	Y	X 108	3.4756	0.12282	0.09091	.	.
64	14	Y	X 107	4.2385	0.14979	0.09091	.	.
65	15	Y	X 106	3.3682	0.11903	0.09091	.	.
66	16	Y	X 105	3.9414	0.13928	0.09091	.	.
67	17	Y	X 104	2.6140	0.09238	0.09091	.	.

68	18	Y	X	103	0.0527	0.00186	0.09091	.	.
69	19	Y	X	102	-1.30216	-0.046017	0.090909	.	.
70	20	Y	X	101	-2.52252	-0.089143	0.090909	.	.
71	21	Y	X	100	-0.69642	-0.024611	0.090909	.	.
72	22	Y	X	99	0.04108	0.001452	0.090909	.	.
73	23	Y	X	98	-0.87381	-0.030880	0.090909	.	.
74	24	Y	X	97	-0.28184	-0.009960	0.090909	.	.

그림 4.6 옵션 OUTCOV=을 사용해서 얻은 교차상관함수

- ④ 변수 COV는 변수 CROSSVAR의 값에 따라 자기공분산이나 교차공분산을 나타낸다. 즉, 변수 COV는 변수 CROSSVAR의 값이 공란이면 IDENTIFY스테이트먼트의 옵션 VAR=에 지정되어 있는 출력시계열데이터의 자기공분산을 나타내며, 변수 CROSSVAR의 값이 공란이 아니면 변수 CROSSVAR의 값으로 지정된 변수, 즉 입력시계열데이터와 IDENTIFY스테이트먼트의 옵션 VAR=에 지정되어 있는 출력시계열데이터의 교차공분산을 나타낸다.
- ⑤ 변수 CORR는 변수 CROSSVAR의 값에 따라 자기상관계수나 교차상관계수를 나타낸다. 즉, 변수 CORR는 변수 CROSSVAR의 값이 공란이면 IDENTIFY스테이트먼트의 옵션 VAR=에 지정되어 있는 출력시계열데이터의 자기상관계수를 나타내며, 변수 CROSSVAR의 값이 공란이 아니면 변수 CROSSVAR의 값으로 지정된 변수, 즉 입력시계열데이터와 IDENTIFY스테이트먼트의 옵션 VAR=에 지정되어 있는 출력시계열데이터 사이의 교차상관계수를 나타낸다.
- ⑥ 비록 [ETS6.1]에는 설명이 없지만 본 저자의 판단으로는 변수 STDERR는 변수 CORR에 해당하는 표준오차를 나타내는 것 같다.
- ⑦ 변수 INVCORR는 변수 CROSSVAR의 값이 공란이면 IDENTIFY스테이트먼트의 옵션 VAR=에 지정되어 있는 출력시계열데이터의 역자기상관계수(inverse autocorrelation)를 나타내며, 변수 CROSSVAR의 값이 공란이 아니면 결측값을 갖는다.
- ⑧ 변수 PARTCORR는 변수 CROSSVAR의 값이 공란이면 IDENTIFY스테이트먼트의 옵션 VAR=에 지정되어 있는 출력시계열데이터의 부분자기상관계수를 나타내며, 변수 CROSSVAR의 값이 공란이 아니면 결측값을 갖는다.

이 SAS프로그램의 마지막 PROC스텝은 PROC PRINT를 사용해서 옵션 OUTCOV=에 지정한 SAS데이터세트 EXAM46B를 출력시키는 것이다. 이 PROC스텝을 수행한 결과가 그림 4.6이다. 그림 4.6과 그림 4.5가 같은 결과를 보임을 알 수 있다.

Bartlett(1946)은 정규성가정하에서 표본교차상관계수들의 점근분산과 점근공분산이 다음과 같음을 보였다.

$$\begin{aligned} & \text{var} [\hat{\rho}_{xy}(k)] \\ & \approx \frac{1}{T-k} \sum_{j=-\infty}^{\infty} [\rho_{xx}(j)\rho_{yy}(j) + \rho_{xy}(-j)\rho_{xy}(j+2k)] \\ & \quad + \frac{1}{2}\rho_{xy}^2(k)\{\rho_{xx}^2(j) + \rho_{yy}^2(j) + 2\rho_{xy}^2(j)\} \end{aligned}$$

$$\begin{aligned}
 & -2\rho_{xy}(k)\{\rho_{xx}(j)\rho_{xy}(j+k) + \rho_{xy}(-j)\rho_{yy}(j+k)\}] \\
 \text{cov}[\hat{\rho}_{xy}(k), \hat{\rho}_{xy}(k+i)] \\
 \approx & \frac{1}{T-k} \sum_{j=-\infty}^{\infty} [\rho_{xx}(j)\rho_{yy}(j+i) + \rho_{xy}(-j)\rho_{xy}(j+2k+i) \\
 & + \frac{1}{2}\rho_{xy}(k)\rho_{xy}(k+i)\{\rho_{xx}^2(j) + \rho_{yy}^2(j) + 2\rho_{xy}^2(j)\} \\
 & - \rho_{xy}(k)\{\rho_{xx}(j)\rho_{xy}(j+k+i) + \rho_{xy}(-j)\rho_{yy}(j+k+i)\} \\
 & - \rho_{xy}(k+i)\{\rho_{xx}(j)\rho_{xy}(j+k) + \rho_{xy}(-j)\rho_{yy}(j+k)\}]
 \end{aligned}$$

입력시계열데이터 $\{x_t\}$ 가 백색잡음이고 입력시계열데이터 $\{x_t\}$ 와 출력시계열데이터 $\{y_t\}$ 사이에 상관관계가 없다는 가정 하에서 표본교차상관계수들의 점근분산과 점근공분산은 각각 다음과 같다.

$$\begin{aligned}
 \text{var}[\hat{\rho}_{xy}(k)] & \approx \frac{1}{T-k} \\
 \text{cov}[\hat{\rho}_{xy}(k), \hat{\rho}_{xy}(k+i)] & \approx \frac{1}{T-k} \rho_{yy}(i)
 \end{aligned}$$

입력시계열데이터 $\{x_t\}$ 가 백색잡음과 동일 때는 이 점근식들을 이용해서 입력시계열데이터 $\{x_t\}$ 와 출력시계열데이터 $\{y_t\}$ 의 교차상관계수가 0이라는 가설을 검정할 수 있다. 이에 대한 자세한 내용은 제4.2.3소절을 참조하라.

4.2 충격반응가중함수의 추정

주어진 시계열데이터에 맞는 전이함수모형을 구축하기 위해서는 우선 전이함수의 충격반응가중값들을 추정하여야 한다. 이 절에서는 최소제곱추정법, 2단계최소제곱추정법, 사전백색화추정법, 그리고 능형회귀추정법을 사용해서 충격반응가중함수를 추정하는 과정을 다루고자 한다.

4.2.1 최소제곱추정법

작인성가정을 만족하는 유리함수형태의 전이함수모형은 다음과 같이 무한차원의 선형전이함수모형으로 나타낼 수 있다.

$$y_t = \nu_0 x_t + \nu_1 x_{t-1} + \dots + n_t \quad (t=1, 2, \dots, T)$$

충분히 큰 수 k 를 사용하면 이 무한차원의 선형전이함수모형은 다음과 같은 유한차원의

선형전이함수모형으로 근사시킬 수 있다.

$$y_t = \nu_0 x_t + \nu_1 x_{t-1} + \dots + \nu_k x_{t-k} + n_t \quad (t = 1, 2, \dots, T)$$

이 유한차원의 전이함수모형에서도 입력시계열데이터 $\{x_t\}$ 와 잡음항 $\{n_t\}$ 가 독립이라고 가정하면, 최소제곱추정법을 적용해서 충격반응가중값들로 구성된 벡터 $\nu = (\nu_0, \nu_1, \dots, \nu_k)'$ 를 다음과 같이 추정할 수 있다.

$$\hat{\nu} = (X'X)^{-1}X'y$$

$$\text{단, } X = \begin{bmatrix} x_k & x_{k-1} & \dots & x_2 & x_1 \\ x_{k+1} & x_k & \dots & x_3 & x_2 \\ x_{k+2} & x_{k+1} & \dots & x_4 & x_3 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ x_{T-1} & x_{T-2} & \dots & x_{T-k+1} & x_{T-k} \\ x_T & x_{T-1} & \dots & x_{T-k+2} & x_{T-k+1} \end{bmatrix}, \quad y = \begin{bmatrix} y_k \\ y_{k+1} \\ \vdots \\ y_{T-1} \\ y_T \end{bmatrix}$$

SAS에서 PROC REG를 사용하면 이 최소제곱추정값들을 얻을 수 있다.

실제 시계열데이터를 분석할 때에는 유한차원인 멱함수의 차원 k 를 모른다. 따라서 여러 k 에 대해서 충격반응가중값들을 추정할 필요가 있다. [Choi](1993)는 여러 k 에 대해서 충격반응가중값들을 축차적으로 최소제곱추정하는 Levinson-Durbin류의 알고리즘을 제시했다. 다음과 같은 차수 k 인 선형전이함수모형을 살펴보자.

$$y_t = \nu_{k0} x_t + \nu_{k1} x_{t-1} + \dots + \nu_{kk} x_{t-k} + n_{kt} \quad (t = 1, 2, \dots, T)$$

이 전이함수모형에서도 입력시계열데이터 $\{x_t\}$ 와 잡음항 $\{n_{kt}\}$ 가 독립이라 가정하고 잡음항 $\{n_{kt}\}$ 의 분산을 다음과 같이 정의하자.

$$\kappa_k = E(y_t - \sum_{j=0}^k \nu_{kj} x_{t-j})^2$$

입력시계열데이터 $\{x_t\}$ 를 AR(k)모형에 적합시켜서 얻은 AR계수들을 $l_{k,1}, l_{k,2}, \dots, l_{k,k}$ 라고 하면 다음 Yule-Walker방정식들이 성립한다.

$$\sigma_{xx}(j) = \sum_{m=1}^k l_{k,m} \sigma_{xx}(j-m) \quad (j = 1, 2, \dots, k)$$

입력시계열데이터 $\{x_t\}$ 를 AR(k)모형에 적합시켰을 때의 백색잡음과정의 분산 λ_k 는 다음과 같이 정의된다.

$$\lambda_k = E(x_t - \sum_{j=1}^k l_{k,j} x_{t-j})^2$$

정의로부터 다음 식들이 성립한다.

$$\kappa_k = \sigma_{yy}(0) - \sum_{j=0}^k \nu_{k,j} \sigma_{yx}(j)$$


```

DATA EXAM47;
  START:T+1;
    IF X1=. THEN X1=0;
    X=0.4*X1+RANNOR(7654321);
    OUTPUT;
    X1=X;
    IF T<510 THEN GO TO START;
  RUN;
DATA EXAM47B;
  SET EXAM47;
  X1 = LAG(X);
  X2 = LAG2(X);
  IF X1=. THEN X1=0. ;
  IF X2=. THEN X2=0. ;
  Y = X - 0.8*X1 + 0.5*X2;
  T=T - 10;
  IF T<1 THEN DELETE;
  KEEP Y X;
  RUN;
/* */
/* A LEVINSON-TYPE ALGORITHM FOR */
/* SINGLE INPUT SINGLE OUTPUT MODEL */
/* MAKE THE DATASET */
PROC IML;
  RESET NOLOG NONAME FW=6;
/* */
/* MODULE FOR TRANSFER FUNCTION MODEL */
/* */
START CHOITFM(YX);
/* FORMING MATRICES */
  MAXORD = 8;
  MAXORD1 = MAXORD+1;
  MAXORD2 = MAXORD+2;
  COVMAT = COVLAG(YX,MAXORD2);
  FREE YX;
  SIGXX1 = SHAPE(0,MAXORD2,1);
  SIGYX1 = SHAPE(0,MAXORD2,1);
  SIGYY1 = SHAPE(0,MAXORD2,1);
  DO ILAG = 1 TO MAXORD2;

```



```

EVENLAG = 2*ILAG;
ODDLAG = 2*ILAG - 1;
SIGXX1[ILAG,1] = COVMAT[2,EVENLAG];
SIGYX1[ILAG,1] = COVMAT[1,EVENLAG];
SIGYY1[ILAG,1] = COVMAT[1,ODDLAG];
END;
PRINT 'SIGYX' 'SIGXX' 'SIGYY',
      SIGYX1 SIGXX1 SIGYY1,.;
FREE COVMAT;
L1 = SHAPE(0,MAXORD2,MAXORD2);
NU1 = SHAPE(0,MAXORD1,MAXORD1);
LAMBDA1 = SHAPE(0,MAXORD2,1);
KAPPA1 = SHAPE(0,MAXORD1,1);
/* INITIAL VALUES FOR RECURSION */
K = 0;
NU1[1,1] = SIGYX1[1,1] / SIGXX1[1,1];
LAMBDA1[1] = SIGXX1[1,1];
KAPPA1[1] = SIGYY1[1,1] - NU1[1,1]*SIGYX1[1,1];
L1[2,2] = SIGXX1[2,1] / SIGXX1[1,1];
LAMBDA1[2] = SIGXX1[1,1] * (1 - L1[2,2]*L1[2,2]);
/* RECURSIVE STEP */
K = 1;
DO WHILE (K <= MAXORD);
  K1 = K+1;
  K2 = K+2;
  L1[K2,K2] = SIGXX1[K2,1];
  M=1;
  DO WHILE (M <= K);
    L1[K2,K2] = L1[K2,K2] - L1[K1,M+1]*SIGXX1[K2-M,1];
    M = M+1;
  END;
  L1[K2,K2] = L1[K2,K2]/LAMBDA1[K1];
  NU1[K1,K1] = SIGYX1[K1,1];
  M=1;
  DO WHILE (M <= K);
    NU1[K1,K1] = NU1[K1,K1] - L1[K1,M+1]*SIGYX1[K1-M,1];
    M = M+1;
  END;
  NU1[K1,K1] = NU1[K1,K1]/LAMBDA1[K1];

```

```

LAMBDA1[K2] = LAMBDA1[K1] * (1 - L1[K2, K2] * L1[K2, K2]);
KAPPA1[K1] = KAPPA1[K] - NU1[K1, K1] * NU1[K1, K1] * LAMBDA1[K1];
J=1;
DO WHILE (J <= K);
    J1 = J + 1;
    L1[K2, J1] = L1[K1, J1] - L1[K2, K2] * L1[K1, K2 - J];
    NU1[K1, J] = NU1[K, J] - NU1[K1, K1] * L1[K1, K2 - J];
    J = J + 1;
END;
K = K + 1;
END;
L = L1[2:MAXORD2, 2:MAXORD2];
FREE L1;
PRINT 'NU' ' ' ' ' 'KAPPA', NU1 KAPPA1,,
      'L' ' ' ' ' 'LAMBDA' L LAMBDA1,;
FINISH;
/* END OF THE MODULE CHOITFM */
/* USE THE MODULE CHOITFM */
USE EXAM47B;
READ ALL VAR{Y X} INTO YX;
CLOSE EXAM47B;
RUN CHOITFM(YX);
QUIT;

```

이 SAS프로그램의 첫 번째 DATA스텝은 다음과 같은 AR(1)모형에서 입력시계열데이터를 발생시키기 위한 것이다.

$$x_t = 0.4x_{t-1} + v_t$$

두 번째 DATA스텝은 다음과 같은 선형전이함수모형에서 출력시계열데이터를 발생시키기 위한 것이다.

$$y_t = x_t - 0.8x_{t-1} + 0.5x_{t-2}$$

이 선형모형에는 잡음항이 없음을 유의하라. 이러한 잡음항이 없는 전이함수모형에서 출력시계열데이터를 발생시킨 이유는 알고리즘 4.1의 유용성을 보이기 위해서다. 이 SAS 프로그램을 수행한 결과가 그림 4.7이다. 그림 4.7에서 다음과 같은 결과를 알 수 있다.

$$\nu_{k,3} = \nu_{k,4} = \dots = \nu_{k,k} = 0 \quad (k = 3, 4, \dots, 8)$$

$$\kappa_2 = \kappa_3 = \dots = \kappa_8$$

따라서 이 출력시계열데이터는 차수가 2차인 선형전이함수모형에서 발생되었음을 알 수

Mallows' C_p , MSE, AIC, BIC 등 벌칙함수들을 이용한 방법들을 사용할 수 있다. 특히 벌칙함수들을 이용해서 선형전이함수를 식별하는 경우에 알고리즘 4.1이 유용하게 사용될 것이다. 이 식별방법들에 대한 자세한 내용은 [SAS4TSA2](pp. 172~178)를 참조하라.

4.2.2 2단계최소제곱추정법

다음과 같은 유한차원의 선형전이모형에서 T 개의 관찰점들을 얻었다고 하자.

$$y_t = \nu_0 x_t + \nu_1 x_{t-1} + \dots + \nu_k x_{t-k} + n_t \quad (t = 1, 2, \dots, T)$$

이 시계열데이터에 최소제곱추정법을 적용한다는 것은 잠재적으로 잡음항 $\{n_t\}$ 가 백색잡음과정이라는 Gauss-Markov가정을 바탕으로 한 것이다. 일반적으로 전이함수모형의 잡음항은 백색잡음과정이 아니라 유색잡음(colored noise)이다. 제4.1.2소절에서 기술했듯이 전이함수모형에서는 일반적인 잡음항 $\{n_t\}$ 는 ARMA과정이라고 가정한다. 그러나 2단계최소제곱추정값(2 stage least squares estimate, 2SLS estimate)들을 사용하기 위해서는 잡음항 $\{n_t\}$ 는 다음의 AR(m)모형을 만족한다고 가정한다(단, $m < k$ 이다).

$$n_t = \phi_1 n_{t-1} + \phi_2 n_{t-2} + \dots + \phi_m n_{t-m} + v_t$$

여기서 $\{v_t\}$ 는 평균이 0이고 분산이 σ_v^2 인 백색잡음과정이다. 잡음항 $\{n_t\}$ 가 AR(m)모형을 만족하면 잡음항으로 이루어진 벡터 $(n_1, n_2, \dots, n_T)'$ 의 공분산행렬 $\sigma_v^2 \Omega$ 가 대각행렬이 아니므로 Gauss-Markov조건을 만족하지 못한다. 따라서 이 최소제곱추정량들은 여러 가지 문제점들을 안고 있다. 이에 대한 자세한 내용은 [SAS4TSA2](pp. 261~262)를 참조하라. 이 최소제곱추정법의 문제점을 해결하기 위해서 사용하는 방법이 2단계최소제곱추정법이다.

선형전이함수모형의 충격반응가중값들의 2단계최소제곱추정법은 다음과 같다.

제1 단계 우선 선형전이함수모형을 최소제곱추정하여 잔차 $\hat{n}_{k+1}, \hat{n}_{k+2}, \dots, \hat{n}_T$ 를 구한 뒤 이 잔차들의 표본자기상관함수를 다음과 같이 계산한다.

$$r_j(\hat{n}) = \frac{\sum_{t=k+j+1}^T v_t \hat{n}_{t-j}}{\sum_{t=k+1}^T v_t^2} \quad (j = 0, 1, \dots, m)$$

이 잔차들의 표본자기상관함수로 구성된 Yule-Walker방정식들은 다음과 같다.

$$\begin{bmatrix} r_0(\hat{n}) & r_1(\hat{n}) & r_2(\hat{n}) & \dots & r_{m-1}(\hat{n}) \\ r_1(\hat{n}) & r_0(\hat{n}) & r_1(\hat{n}) & \dots & r_{m-2}(\hat{n}) \\ r_2(\hat{n}) & r_1(\hat{n}) & r_0(\hat{n}) & \dots & r_{m-3}(\hat{n}) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ r_{m-1}(\hat{n}) & r_{m-2}(\hat{n}) & r_{m-3}(\hat{n}) & \dots & r_0(\hat{n}) \end{bmatrix} \begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \hat{\phi}_3 \\ \vdots \\ \hat{\phi}_m \end{bmatrix} = \begin{bmatrix} r_1(\hat{n}) \\ r_2(\hat{n}) \\ r_3(\hat{n}) \\ \vdots \\ r_m(\hat{n}) \end{bmatrix}$$

이 Yule-Walker방정식을 풀어서 $\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_m$ 을 구하면 2단계최소제곱추정법의 제1단계가 완성된다. 이 Yule-Walker방정식들을 푸는 방법은 여러 가지가 있으나 Levinson-Durbin알고리즘을 사용하는 것이 가장 일반적이다. 이 Levinson-Durbin알고리즘에 대한 자세한 내용은 [SAS4TSA2] (pp. 409~410)를 참조하라.

제2단계 잡음시계열데이터 $\{n_t\}$ 의 AR계수추정값들 $\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_m$ 을 상관행렬 Ω 에 대입해서 추정행렬 $\hat{\Omega}$ 을 구한 다음 이 추정행렬을 사용해서 다음과 같이 충격반응가중함수들의 가능일반화최소제곱추정값(feasible generalized least squares estimate, FGLS estimate)들을 구한다.

$$\hat{v} = (X' \hat{\Omega}^{-1} X)^{-1} X' \hat{\Omega}^{-1} y$$

2단계최소제곱추정법의 제2단계에서 모든 관찰점들을 사용해서 가능일반화최소제곱추정을 하면 이 방법을 Prais-Winsten방법이라 하고, 첫 m 개의 관찰점들을 제외하고 가능일반화최소제곱추정을 하면 Cochrane-Orcutt방법이라 한다. 관찰점들의 개수가 큰 경우에는 이 두 방법들이 비슷한 결과를 제시하나, 관찰점들의 개수가 크지 않은 경우에는 Prais-Winsten방법을 사용하는 것이 좋다. 특히, 관찰점들의 개수가 작으면서 설명변수가 추세를 나타내는 경우에는 Cochrane-Orcutt방법의 효율성이 떨어질 수도 있다. 2단계최소제곱추정법에 관한 자세한 내용은 [SAS4TSA2] (pp. 267~277)를 참조하라.

SAS/ETS의 PROC AUTOREG를 사용하면 Prais-Winsten형의 2단계최소제곱추정을 할 수 있다. 제4.1.4소절에서 만든 시계열데이터 TFM1에 2단계최소제곱추정법을 적용하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM4-8.PRG에 수록되어 있다).

```

/* */
/*          EXAM4-8.PRG          */
/* */
/*          2SLS ESTIMATE        */
/*          DATASET = TFM1       */
/* */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS LINESIZE = 72;
DATA EXAM48;
  SET TSA4DATA.TFM1;
  X1 = LAG1(X);
  X2 = LAG1(X1);
  X3 = LAG1(X2);
  X4 = LAG1(X3);

```

```

IF X4 = . THEN DELETE;
RUN;
TITLE1 '2SLS ESTIMATES BY PROC AUTOREG';
PROC AUTOREG DATA = EXAM48;
MODEL Y = X X1 X2 X3 X4 / NLAG = 3;
RUN;

```

이 SAS프로그램을 수행한 결과가 그림 4.8이다. 이 결과물에 대한 자세한 해설은 [SAS4TSA2](pp. 267~283)를 참조하라. 그림 4.8의 부분 A에는 2단계최소제곱추정법의 제 1단계로 최소제곱추정법을 사용해서 잡음항의 AR계수들을 추정한 결과가 수록되어 있다. 우선 종속변수의 이름이 Y임을 나타낸 뒤, 오차제곱합(SSE), 오차제곱합의 자유도(DFE), 평균오차제곱합(MSE), 평균오차제곱합의 제곱근(Root MSE), 베이잔정보판단기준값(SBC), Akaike정보판단기준값(AIC), 회귀결정계수(Reg Rsq), 총결정계수(Total Rsq), Durbin-Watson통계값(Durbin-Watson)이 수록되어 있다. 비록 총결정계수가 0.795로 최소제곱추정식이 이 시계열데이터를 잘 나타내고 있는 것 같지만, Durbin-Watson통계량값이 3.63으로 잔차들의 시차 1에서의 자기상관계수가 -1에 가깝다. 따라서 가능일반화최소제곱추정법을 적용할 필요가 있다. 그 다음에는 설명변수이름(Variable), 각 변수의 자유도(DF), 최소제곱추정값(B Value), 최소제곱추정값의 표준오차(Std Error), 이에 해당하는 t 통계량값(t Ratio), 최소제곱추정값의 모수가 0이라는 귀무가설을 기각하는 유의확률(Approx Prob), 즉 p 값이 수록되어 있다. 본 저자는 회귀분석에서 t 검정을 지나치게 신뢰해서는 안 된다는 의견을 가지고 있다. 직교회귀분석(orthogonal regression)의 경우를 제외하고는 대부분의 회귀분석문제에서는 다공선성이 발생하기 때문에 어느 한 설명변수의 회귀계수가 0이라는 귀무가설을 검정하기 위해서는 t 검정법 대신 부분 F 검정법(partial F test)을 사용하는 것이 타당할 것이다. MODEL스테이트먼트의 옵션 NLAG = 에는 잡음항이 갖는 AR모형의 차수 m 을 지정한다. 자기상관계수들의 추정부분(Estimates of Autocorrelations)에는 시차(Lag)가 0, 1, ..., m 인 경우의 잔차들의 자기공분산(Covariance)들과 자기상관계수(Correlation)들이 수록되어 있다. 그림 4.8의 부분 B에는 2단계최소제곱추정법의 제2단계를 수행한 결과가 수록되어 있다. 앞에서 설명한 2단계최소제곱추정법의 제2단계를 PROC AUTOREG에서는 Yule-Walker추정법이라 부른다. 따라서 제2단계에서 얻은 추정값들을 Yule-Walker추정값들이라고 부른다. 부분 B의 통계값들과 부분 A의 통계값들을 비교해 보면, Yule-Walker추정에 의한 오차제곱합(SSE), 평균오차제곱합(MSE), 평균오차제곱합의 제곱근(Root MSE), 베이잔정보판단기준값(SBC), 그리고 Akaike정보판단기준값(AIC)이 최소제곱추정에 의한 값들에 비해 현저히 작아졌고, Yule-Walker추정에 의한 회귀결정계수(Reg Rsq)와 총결정계수(Total Rsq)는 최소제곱추정에 의한 값들에 비해 현저히 커졌다. 또한, Yule-Walker추정에 의한 회귀결정계수 0.994는 Yule-Walker추정에 의한 총결정계수 0.958에 비해 크다. 이는 이 시계열데이터의 잡음항이 백색잡음과정이라는 가정보다는 AR(3)모형을 갖는다는 가정이 더 타당함을

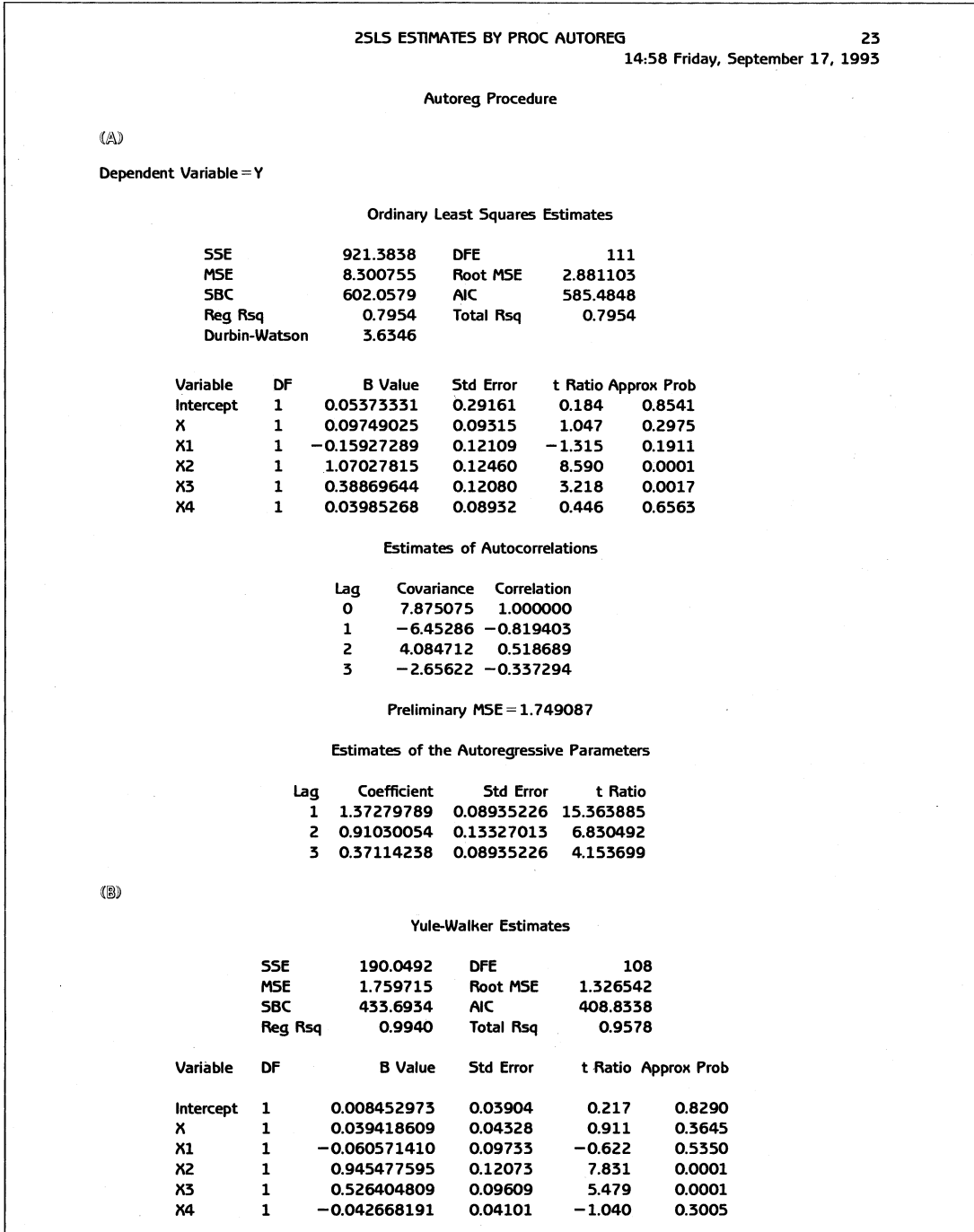


그림 4.8 PROC AUTOREG를 사용한 전이함수모형의 추정

뜻한다.

그림 4.8에서 얻은 최종추정모형은 다음과 같다.

$$y_t = 0.008 + 0.039x_t - 0.061x_{t-1} + 0.945x_{t-2} + 0.526x_{t-3} - 0.043x_{t-4} + n_t$$

$$n_t = 1.373n_{t-1} + 0.910n_{t-2} + 0.371n_{t-3} + v_t$$

앞에서는 2단계최소제곱추정법의 제2단계의 가능일반화최소제곱추정방법으로 Yule-Walker추정법을 사용했다. PROC AUTOREG에서는 이 제2단계의 추정법으로 Yule-Walker추정법 이외에도 비조건부최소제곱추정법이나 최우추정법을 사용할 수 있다. 한 예로 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM4-9.PRG에 수록되어 있다).

```

/*                                                    */
/*                                                    EXAM4-9.PRG          */
/*                                                    */
/*          2SLS ESTIMATES                          */
/*          USING UNCONDITIONAL LEAST SQUARE METHOD  */
/*          DATASET = TFM1                            */
/*                                                    */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS LINESIZE = 72;
DATA EXAM49;
    SET TSA4DATA.TFM1;
    X1 = LAG1(X);
    X2 = LAG1(X1);
    X3 = LAG1(X2);
    IF X3 = . THEN DELETE;
RUN;
TITLE1 '2SLS ESTIMATES BY PROC AUTOREG';
PROC AUTOREG DATA = EXAM49;
    MODEL Y = X X1 X2 X3
          / NLAG = 2 METHOD = ULS;
RUN;

```

그림 4.8의 부분 B에서 알 수 있듯이 시차가 4인 변수 X4의 회귀계수추정값이 -0.043 이고 이에 해당하는 p 값이 0.301 이므로 이 SAS프로그램에서는 이 설명변수를 모형에서 제거하였다. 또한, MODEL스테이트먼트의 옵션 METHOD = 에 ULS를 할당하였다. 이는 제2단계에서 Yule-Walker추정법 대신에 비조건부최소제곱추정법을 사용하기 위한 것이다. 만일, 옵션 METHOD = 에 ML을 할당하면 최우추정법이 사용된다. 이 SAS프로그램을 수행한 결과가 그림 4.9이다. 그림 4.9의 부분 A에는 2단계최소제곱추정법의 제1단계로 최소제곱추정법을 사용해서 잡음항의 AR계수들을 추정한 결과가 수록되어 있다. 이 부분은 그림 4.8의 부분 A와 같은 방법으로 설명할 수 있다. PROC AUTOREG스테이트먼트의 옵션 NLAG = 에 2를 할당하였다. 따라서 자기상관계수들의 추정부분에는 시차(Lag)가 0, 1, 2인 경우의 잔차들의 자기공분산(Covariance)들과 자기상관계수(Correlation)들이 수록되어 있다. 그림 4.9의 부분 B는 2단계최소제곱추정법의 제2단계를 수행한 결과이다.

2SLS ESTIMATES BY PROC AUTOREG 1
 15:33 Friday, September 17, 1993
 Autoreg Procedure

Dependent Variable = Y

(A)

Ordinary Least Squares Estimates

SSE	930.1392	DFE	113
MSE	8.23132	Root MSE	2.869028
SBC	602.3516	AIC	588.4981
Reg Rsq	0.7986	Total Rsq	0.7986
Durbin-Watson	3.6133		

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	0.02402015	0.28570	0.084	0.9331
X	1	0.07644120	0.09032	0.846	0.3991
X1	1	-0.14519163	0.11966	-1.213	0.2275
X2	1	1.04875147	0.11987	8.749	0.0001
X3	1	0.41858791	0.08874	4.717	0.0001

Estimates of Autocorrelations

Lag	Covariance	Correlation
0	7.882535	1.000000
1	-6.39804	-0.811673
2	4.017301	0.509646

Preliminary MSE = 2.175355

Estimates of the Autoregressive Parameters

Lag	Coefficient	Std Error	t Ratio
1	1.16653596	0.08536390	13.665449
2	0.43719971	0.08536390	5.121600

(B)

Unconditional Least Squares Estimates

SSE	237.483	DFE	111
MSE	2.139486	Root MSE	1.462698
SBC	452.4638	AIC	433.069
Reg Rsq	0.9889	Total Rsq	0.9486
Durbin-Watson	2.3034		

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	0.01487111	0.05513	0.270	0.7879
X	1	0.03225969	0.04065	0.794	0.4291
X1	1	-0.07552796	0.07687	-0.983	0.3279
X2	1	0.99598668	0.07661	13.001	0.0001
X3	1	0.44958519	0.03896	11.539	0.0001
A(1)	1	1.22270422	0.08350	14.644	0.0001
A(2)	1	0.48632462	0.08341	5.831	0.0001

Autoregressive parameters assumed given.

SSE	237.483	DFE	111
MSE	2.139486	Root MSE	1.462698
SBC	452.4638	AIC	433.069
Reg Rsq	0.9889	Total Rsq	0.9486
Durbin-Watson	2.3034		

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	0.014871112	0.05513	0.270	0.7879
X	1	0.032259694	0.04065	0.794	0.4291
X1	1	-0.075527965	0.07687	-0.983	0.3279
X2	1	0.995986676	0.07661	13.001	0.0001
X3	1	0.449585187	0.03896	11.539	0.0001

그림 4.9 최우추정법을 사용한 2단계최소제곱추정

앞에서 설명했듯이 MODEL스태이트먼트의 옵션 METHOD =에 ULS를 할당했으므로 제2단계에서 비조건부최소제곱추정법이 적용되었다. 이 예제에서도 부분 B의 통계값들과 부분 A의 통계값들을 비교해 보면, 비조건부최소제곱추정에 의한 오차제곱합(SSE), 평균 오차제곱합(MSE), 평균오차제곱합의 제곱근(Root MSE), 베이지안정보판단기준값(SBC), Akaike정보판단기준값(AIC), 그리고 Durbin-Watson통계값이 최소제곱추정에 의한 값들에 비해 현저히 작아졌고, 비조건부최소제곱추정에 의한 회귀결정계수(Reg Rsq)와 총결정계수(Total Rsq)는 최소제곱추정에 의한 값들에 비해 현저히 커졌다. 또한, 비조건부최소제곱추정에 의한 회귀결정계수는 총결정계수에 비해 크다. 따라서 이 시계열 데이터의 잡음향이 백색잡음과정이라는 가정보다는 AR(2)모형을 갖는다는 가정이 더 타당하다. 그림 4.9에서 얻은 최종추정모형은 다음과 같다.

$$y_t = 0.015 + 0.032x_t - 0.076x_{t-1} + 0.996x_{t-2} + 0.450x_{t-3} + n_t$$

$$n_t = 1.223n_{t-1} + 0.486n_{t-2} + v_t$$

4.2.3 사전백색화추정법

이 장에서는 전이함수모형이 작인적이라고 가정했으므로 t 대신 $(t+k)$ 를 대입하면 다음 식이 성립한다.

$$y_{t+k} = \nu_0 x_{t+k} + \nu_1 x_{t+k-1} + \nu_2 x_{t+k-2} + \dots + n_{t+k}$$

이 식의 양변에 x_t 를 곱한 뒤 기대값을 취하면, 안정성가정에 의해서 다음 식을 얻을 수 있다.

$$\sigma_{yx}(k) = \nu_0 \sigma_{xx}(k) + \nu_1 \sigma_{xx}(k-1) + \nu_2 \sigma_{xx}(k-2) + \dots$$

이 식으로부터 다음 식을 얻는다.

$$\rho_{yx}(k) = \frac{\sqrt{\sigma_{xx}(0)}}{\sqrt{\sigma_{yy}(0)}} \{ \nu_0 \rho_{xx}(k) + \nu_1 \rho_{xx}(k-1) + \nu_2 \rho_{xx}(k-2) + \dots \}$$

이 식으로부터 교차상관함수 $\{\rho_{yx}(k)\}$ 는 충격반응가중함수 $\{\nu_j\}$ 와 입력시계열데이터 $\{x_t\}$ 의 자기상관함수에 의해 결정된다는 것을 알 수 있다. 그러나 자기상관함수나 교차상관함수가 주어졌다고 하더라도 이 식을 이용해서 충격반응가중값들을 구하는 것은 쉬운 일이 아니다. 유리함수형태의 전이함수모형에서 분모의 차수 r 이 0인 경우, 즉 전이함수가 유한개의 충격반응가중값들을 갖는 경우라 하더라도 $\{\rho_{yx}(k)\}$ 와 $\{\rho_{xx}(k)\}$ 가 주어졌을 때 이 방정식을 풀어서 충격반응가중함수 $\{\nu_j\}$ 를 구한다는 것은 쉬운 일이 아니다.

만일, 입력시계열데이터 $\{x_t\}$ 가 백색잡음과정이면, 즉 $k \neq 0$ 에 대해서 $\rho_{xx}(k) = 0$ 이면 교차상관함수 $\{\rho_{yx}(k)\}$ 의 식은 다음과 같이 간단해진다.

$$\nu_k = \frac{\sqrt{\sigma_{yy}(0)}}{\sqrt{\sigma_{xx}(0)}} \rho_{yx}(k) \quad (k = 0, 1, 2, \dots)$$

즉, 입력시계열데이터 $\{x_t\}$ 가 백색잡음과정이면, 충격반응가중값 ν_k 는 교차상관계수 $\rho_{yx}(k)$ 에 비례한다. 따라서 입력시계열데이터 $\{x_t\}$ 가 백색잡음과정이 아니라면 이 입력시계열데이터를 사전백색화(whitening)하면 쉽게 전이함수모형화를 할 수 있다.

입력시계열데이터 $\{x_t\}$ 가 다음과 같은 ARMA모형을 만족한다고 가정하자.

$$\phi_x(B)x_t = \theta_x(B)z_t$$

이 경우에 입력시계열데이터 $\{x_t\}$ 를 사전백색화한 시계열데이터 $\{z_t\}$ 는 다음과 같다.

$$z_t = \frac{\phi_x(B)}{\theta_x(B)} x_t$$

단, $\{z_t\}$ 는 평균 0, 분산 σ_z^2 인 백색잡음과정

사전백색화과정을 적용해서 충격반응가중함수를 추정하기 위해서는 다음과 같은 3단계 과정을 수행한다.

제1단계 입력시계열데이터가 백색잡음과정이 아니면 이 입력시계열데이터를 사전백색화시킨다. 입력시계열데이터 $\{x_t\}$ 가 ARMA모형을 따른다고 가정하고, 이 입력시계열데이터를 다음과 같이 ARMA모형화한다.

$$\phi_x(B)x_t = \theta_x(B)z_t$$

사전백색화된 입력시계열데이터 $\{z_t\}$ 는 다음과 같이 나타낼 수 있다.

$$z_t = \frac{\phi_x(B)}{\theta_x(B)} x_t$$

단, $\{z_t\}$ 는 평균 0, 분산 σ_z^2 인 백색잡음시계열데이터

제2단계 입력시계열데이터 $\{x_t\}$ 에 적용한 사전백색화과정의 필터를 출력시계열데이터 $\{y_t\}$ 에 적용하여 다음과 같은 여과된 출력시계열데이터 $\{w_t\}$ 를 생성한다.

$$w_t = \frac{\phi_x(B)}{\theta_x(B)} y_t$$

제3단계 백색잡음시계열데이터 $\{z_t\}$ 와 여과된 출력시계열데이터 $\{w_t\}$ 사이의 표본교차상관함수 $\{\rho_{zw}(k)\}$ 를 계산한 뒤, 이를 이용해서 다음과 같이 충격반응가중함수 $\{\nu_k\}$ 를 추정한다.

$$\hat{v}_k = \frac{\sqrt{\hat{\sigma}_{uw}(0)}}{\sqrt{\hat{\sigma}_z(0)}} \hat{\rho}_{uz}(k) \quad (k = 0, 1, 2, \dots)$$

SAS/ETS에서는 PROC ARIMA를 사용해서 사전백색화추정법을 적용할 수 있다. 이 방법을 사용해서 시계열데이터 TFM1의 충격반응가중함수를 추정하기 위해서 다음 SAS 프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM4-10.PRG에 수록되어 있다).

```

/ *                               */
/ *           EXAM4-10.PRG        */
/ *                               */
/ *           2SLS ESTIMATE       */
/ *           USING PREWHITENING */
/ *           DATASET = TFM1      */
/ *                               */

LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS LINESIZE = 72;
OPTIONS PAGESIZE = 80;
DATA EXAM410;
    SET TSA4DATA.TFM1;
    RUN;
TITLE1 'PREWHITENING INPUT & OUTPUT SERIES';
PROC ARIMA DATA = EXAM410;
    I VAR=X NOPRINT;
    E P=1 Q=1 NOPRINT;
    I VAR=Y CROSS=(X) OUTCOV = EXAM410B;
    RUN;
PROC PRINT DATA = EXAM410B;
    RUN;
DATA EXAM410C;
    SET EXAM410B;
    IF INVCORR NE.THEN DELETE;
    IF LAG < 0 THEN DELETE;
    NU = SQRT(48.04037/ 9.135349) * CORR;
    RUN;
PROC PRINT DATA = EXAM410C;
    RUN;
DATA TSA4DATA.NU1;

```

```

SET EXAM410C;
KEEP LAG N CORR STDERR NU;
RUN;

```

이 SAS프로그램을 수행한 결과가 그림 4.10이다. 이 SAS프로그램에서 PROC ARIMA의 첫 번째 IDENTIFY스테이트먼트는 입력시계열데이터 X를 식별하기 위한 것이다. 단변량시계열분석에서의 ARIMA모형식별에 관한 자세한 내용은 [SAS4TSA2](제10장, 제11장)를 참조하라. 이 SAS프로그램에서는 IDENTIFY스테이트먼트의 옵션 NOPRINT를 사용해서 IDENTIFY스테이트먼트의 수행결과와 출력을 억제했다. 그러나 이 입력시계열데이터를 Box-Jenkins방법으로 식별하면 ARMA(1,1)모형이나 AR(2)모형을 선택한다. 또한, AIC방법이나 BIC방법을 적용하면 ARMA(1,1)모형을 선택한다. 첫 번째 ESTIMATE스테이트먼트는 입력시계열데이터를 이렇게 선택된 ARMA(1,1)모형에 적합시켜 추정을 하기 위한 것이다. 이 SAS프로그램에서는 ESTIMATE스테이트먼트의 옵션 NOPRINT를 사용해서 ESTIMATE스테이트먼트의 수행결과와 출력을 억제했다. 그러나 이 ESTIMATE스테이트먼트를 수행한 결과로 다음과 같은 ARMA(1,1)모형의 추정식을 얻었다.

$$(1-0.5113B)(x_t+1.6646) = (1+0.3825B)z_t$$

여기서 $\{z_t\}$ 는 사전백색화된 입력시계열데이터이다. 두 번째 IDENTIFY스테이트먼트를 수행한 결과가 그림 4.10의 부분 A이다. 부분 A에는 IDENTIFY스테이트먼트의 옵션 VAR=에 지정된 변수 Y를 식별하기 위한 표본자기상관함수, 표본역자기상관함수, 그리고 표본부분자기상관함수가 그려져 있고, 이 출력시계열데이터가 백색잡음과정에서 출력되었는가를 검정하기 위한 카이제곱통계량값들, 즉 퍼트맨도우통계량값들이 수록되어 있다. IDENTIFY스테이트먼트의 옵션 CROSS=에 입력변수를 X로 지정했으므로, 앞에서 추정된 입력시계열데이터 $\{x_t\}$ 의 ARMA모형을 사용해서 출력시계열데이터를 사전백색화한다. 즉, 다음 식을 만족하는 여과된 출력시계열데이터 $\{w_t\}$ 를 구한다.

$$(1-0.5113B)(y_t+2.2671) = (1+0.3825B)w_t$$

여기서 유의할 점은 출력시계열데이터를 여과시키기 위해서는 반드시 사전에 입력시계열데이터를 식별하고 추정하는 단계를 거쳐야 한다는 것이다. 이렇게 입력시계열데이터를 사전백색화하는 필터와 똑같은 필터로 여과된 출력시계열데이터의 표본교차상관함수가 그림 4.10의 부분 A2에 수록되어 있다. 부분 A2에서 시차들 0과 1의 표본교차상관계수들이 유의적으로 0과 다르지 않음을 알 수 있다. 부분 A3에는 사전백색화된 입력시계열데이터와 여과된 출력시계열데이터가 서로 독립인가를 검정하기 위해서, 표본교차상관함수를 사용해서 만든 카이제곱통계량값들이 실려 있다. 부분 A4는 입력시계열데이터를 사전백색잡음화하고 출력시계열데이터를 여과시키기 위한 필터이다. 이 필터는 입력시계열데이터를 ARMA모형에 적합시키는 ESTIMATE스테이트먼트의 추정결과이다. 이 예제에서 필

2	9.478615	0.45246	.	*****
3	2.952861	0.14095	.	***.
4	0.630208	0.03008	.	* .
5	0.224725	0.01073	.	. .
6	-1.046883	-0.04997	.	* .
7	1.415279	0.06756	.	* .
8	-2.401570	-0.11464	.	** .
9	-0.475427	-0.02269	.	. .
10	-2.181261	-0.10412	.	** .
11	0.101516	0.00485	.	. .
12	0.639952	0.03055	.	* .
13	0.402815	0.01923	.	. .
14	2.113301	0.10088	.	** .
15	-1.328605	-0.06342	.	* .
16	1.609718	0.07684	.	** .
17	0.925347	0.04417	.	* .
18	-1.308870	-0.06248	.	* .
19	1.157621	0.05526	.	* .
20	-2.498093	-0.11925	.	** .
21	0.880805	0.04204	.	* .
22	0.699933	0.03341	.	* .
23	-1.261067	-0.06020	.	* .
24	0.446041	0.02129	.	. .

“.” marks two standard errors

(A3)

Crosscorrelation Check Between Series

To Lag	Chi Square	DF	Prob	Crosscorrelations					
5	28.84	6	0.000	0.100	-0.052	0.452	0.141	0.030	0.011
11	32.67	12	0.001	-0.050	0.068	-0.115	-0.023	-0.104	0.005
17	35.49	18	0.008	0.031	0.019	0.101	-0.063	0.077	0.044
23	38.84	24	0.028	-0.062	0.055	-0.119	0.042	0.033	-0.060

(A4)

Both variables have been prewhitened by the following filter:
Prewhitening Filter

Autoregressive Factors
Factor 1:1 -0.5113B**(1)

Moving Average Factors
Factor 1:1 +0.38252B**(1)

(B)

PREWHITENING INPUT & OUTPUT SERIES

38

23:12 Saturday, September 18, 1993

OBS	LAG	VAR	CROSSVAR	N	COV	CORR	STDERR	INVCORR	PARTCORR
1	0	Y		121	42.9210	1.00000	0.00000	1.00000	1.00000
2	1	Y		120	21.2031	0.49400	0.09091	-0.30588	0.49400
3	2	Y		119	20.4637	0.47678	0.11090	-0.34780	0.30787
4	3	Y		118	5.3966	0.12573	0.12671	0.27981	-0.27674
5	4	Y		117	5.8656	0.13666	0.12774	-0.11166	0.04394
6	5	Y		116	-0.0965	-0.00225	0.12894	0.02107	0.01351
7	6	Y		115	-4.3304	-0.10089	0.12894	0.01122	-0.22207
8	7	Y		114	-5.7328	-0.13557	0.12959	-0.03214	0.00548
9	8	Y		113	-5.1230	-0.11936	0.13072	0.10887	0.08736
10	9	Y		112	0.1027	0.00239	0.13162	-0.06881	0.09910
11	10	Y		111	2.8972	0.06750	0.13162	0.01596	0.08489
12	11	Y		110	3.3452	0.07794	0.13191	-0.03540	-0.05483
13	12	Y		109	4.2126	0.09815	0.13229	-0.05493	0.01542
14	13	Y		108	4.5233	0.10539	0.13289	0.17380	0.05052
15	14	Y		107	6.4912	0.15124	0.13358	-0.10679	0.04258
16	15	Y		106	5.7960	0.13504	0.13498	-0.08303	0.04306
17	16	Y		105	-0.3969	-0.00925	0.13610	0.11064	-0.18754
18	17	Y		104	-0.5262	-0.01226	0.13610	-0.04355	0.04454
19	18	Y		103	-2.4128	-0.05621	0.13611	0.08205	0.08220
20	19	Y		102	-0.0285	-0.00066	0.13630	-0.10749	-0.04279
21	20	Y		101	-0.7004	-0.01632	0.13630	-0.06325	0.04377
22	21	Y		100	-3.9848	-0.09284	0.13632	0.19773	-0.12823
23	22	Y		99	2.0682	0.04819	0.13684	-0.07423	0.19589
24	23	Y		98	2.1003	0.04893	0.13698	-0.03809	0.06898

25	24	Y		97	9.0045	0.20979	0.13712	-0.00151	0.00264
26	-24	Y	X	97	1.7115	0.08170	0.09091		
27	-23	Y	X	98	0.6622	0.03161	0.09091		
28	-22	Y	X	99	1.2181	0.05814	0.09091		
29	-21	Y	X	100	0.2525	0.01205	0.09091		
30	-20	Y	X	101	1.8854	0.09000	0.09091		
31	-19	Y	X	102	-4.5738	-0.21833	0.09091		
32	-18	Y	X	103	2.4509	0.11699	0.09091		
33	-17	Y	X	104	-1.5577	-0.07436	0.09091		
34	-16	Y	X	105	2.7825	0.13282	0.09091		
35	-15	Y	X	106	-3.9908	-0.19050	0.09091		
36	-14	Y	X	107	1.9551	0.09333	0.09091		
37	-13	Y	X	108	-1.3106	-0.06256	0.09091		
38	-12	Y	X	109	2.6722	0.12756	0.09091		
39	-11	Y	X	110	0.1357	0.00648	0.09091		
40	-10	Y	X	111	-0.1043	-0.00498	0.09091		
41	-9	Y	X	112	0.4227	0.02018	0.09091		
42	-8	Y	X	113	1.3692	0.06536	0.09091		
43	-7	Y	X	114	-0.7744	-0.03697	0.09091		
44	-6	Y	X	115	1.6429	0.07842	0.09091		
45	-5	Y	X	116	-3.8167	-0.18219	0.09091		
46	-4	Y	X	117	-0.1817	-0.00868	0.09091		
47	-3	Y	X	118	-0.3810	-0.01819	0.09091		
48	-2	Y	X	119	1.9055	0.09096	0.09091		
49	-1	Y	X	120	-1.3078	-0.06243	0.09091		
50	0	Y	X	121	2.1018	0.10033	0.09091		
51	1	Y	X	120	-1.0908	-0.05207	0.09091		
52	2	Y	X	119	9.4786	0.45246	0.09091		
53	3	Y	X	118	2.9529	0.14095	0.09091		
54	4	Y	X	117	0.6302	0.03008	0.09091		
55	5	Y	X	116	0.2247	0.01073	0.09091		
56	6	Y	X	115	-1.0469	-0.04997	0.09091		
57	7	Y	X	114	1.4153	0.06756	0.09091		
58	8	Y	X	113	-2.4016	-0.11464	0.09091		
59	9	Y	X	112	-0.4754	-0.02269	0.09091		
60	10	Y	X	111	-2.1813	-0.10412	0.09091		
61	11	Y	X	110	0.1015	0.00485	0.09091		
62	12	Y	X	109	0.6400	0.03055	0.09091		
63	13	Y	X	108	0.4028	0.01923	0.09091		
64	14	Y	X	107	2.1133	0.10088	0.09091		
65	15	Y	X	106	-1.3286	-0.06342	0.09091		
66	16	Y	X	105	1.6097	0.07684	0.09091		
67	17	Y	X	104	0.9253	0.04417	0.09091		
68	18	Y	X	103	-1.3089	-0.06248	0.09091		
69	19	Y	X	102	1.1576	0.05526	0.09091		
70	20	Y	X	101	-2.4981	-0.11925	0.09091		
71	21	Y	X	100	0.8808	0.04204	0.09091		
72	22	Y	X	99	0.6999	0.03341	0.09091		
73	23	Y	X	98	-1.2611	-0.06020	0.09091		
74	24	Y	X	97	0.4460	0.02129	0.09091		

(C)

PREWHITENING INPUT & OUTPUT SERIES

39

23:12 Saturday, September 18, 1995

OBS	LAG	VAR	CROSSVAR	N	CORR	STDERR	NU
1	0	Y	X	121	0.10033	0.090909	0.23007
2	1	Y	X	120	-0.05207	0.090909	-0.11941
3	2	Y	X	119	0.45246	0.090909	1.03758
4	3	Y	X	118	0.14095	0.090909	0.32323
5	4	Y	X	117	0.03008	0.090909	0.06899
6	5	Y	X	116	0.01073	0.090909	0.02460
7	6	Y	X	115	-0.04997	0.090909	-0.11460
8	7	Y	X	114	0.06756	0.090909	0.15492
9	8	Y	X	113	-0.11464	0.090909	-0.26289
10	9	Y	X	112	-0.02269	0.090909	-0.05204
11	10	Y	X	111	-0.10412	0.090909	-0.23877
12	11	Y	X	110	0.00485	0.090909	0.01111
13	12	Y	X	109	0.03055	0.090909	0.07005
14	13	Y	X	108	0.01923	0.090909	0.04409
15	14	Y	X	107	0.10088	0.090909	0.23133
16	15	Y	X	106	-0.06342	0.090909	-0.14544
17	16	Y	X	105	0.07684	0.090909	0.17621
18	17	Y	X	104	0.04417	0.090909	0.10129
19	18	Y	X	103	-0.06248	0.090909	-0.14328
20	19	Y	X	102	0.05526	0.090909	0.12672

21	20	Y	X	101	-0.11925	0.090909	-0.27345
22	21	Y	X	100	0.04204	0.090909	0.09642
23	22	Y	X	99	0.03341	0.090909	0.07662
24	23	Y	X	98	-0.06020	0.090909	-0.13804
25	24	Y	X	97	0.02129	0.090909	0.04883

그림 4.10 사전백색화과정을 사용한 충격반응가중함수의 추정

터는 다음과 같다.

$$AR\text{필터} = 1 - 0.5113B$$

$$MA\text{필터} = 1 + 0.3825B$$

IDENTIFY스테이트먼트의 옵션 OUTCOV = 에는 표본교차상관함수를 저장할 SAS데이터세트 EXAM410B를 지정했다. 이 옵션에 대해서는 제4.1.6소절에 자세히 설명했다.

그림 4.10의 부분 B는 옵션 OUTCOV = 에 지정한 SAS데이터세트를 PROC PRINT를 사용해서 OUTPUT윈도우에 출력시킨 것이다. 그림 4.10의 부분 B와 그림 4.6은 약간 다른 결과를 보여준다. 그림 4.10의 부분 B에서나 그림 4.6에서나 변수 CROSSVAR가 공란인 경우에는 필터로 여과되기 이전의 출력시계열데이터의 표본자기공분산함수와 표본자기상관함수가 수록되어 있다. 변수 CROSSVAR가 공란이 아닌 경우에는 그림 4.10에는 사전백색화된 입력시계열데이터와 필터로 여과된 출력시계열데이터의 표본교차공분산과 표본교차상관함수가 수록되어 있으나, 그림 4.6에서는 필터로 여과되기 이전의 입력시계열데이터와 출력시계열데이터의 표본교차공분산과 표본교차상관함수가 수록되어 있다.

그림 4.10의 부분 C는 옵션 OUTCOV = 에 지정한 SAS데이터세트 EXAM410B의 변수 CORR를 다음과 같이 변형해서 충격반응가중함수를 추정한 것이다.

$$\hat{\nu}_k = \frac{\sqrt{\hat{\sigma}_{uu}(0)}}{\sqrt{\hat{\sigma}_z(0)}} \hat{\rho}_{uz}(k) \quad (k=0, 1, 2 \dots)$$

여기서 $\hat{\sigma}_{uu}(0)$ 와 $\hat{\sigma}_z(0)$ 는 PROC ARIMA의 두 번째 IDENTIFY스테이트먼트에서 얻을 수 있다. 이 예제에서는 그림 4.10의 부분 A2의 앞 부분에서 $\hat{\sigma}_{uu}(0)$ 와 $\hat{\sigma}_z(0)$ 의 값들이 각각 48.040과 9.135임을 알 수 있다. 이 값들을 사용해서 충격반응가중함수를 추정하는 과정이 이 SAS프로그램의 두 번째 DATA스텝이다. 이렇게 만들어진 SAS데이터세트를 두 번째 PROC PRINT를 사용해서 OUTPUT윈도우에 출력시킨 것이 그림 4.10의 부분 C이다. 이 SAS프로그램의 세 번째 DATA스텝은 이렇게 추정된 충격반응가중함수를 SAS데이터라이브러리 TSA4DATA의 파일 NU1에 저장하기 위한 것이다.

시계열데이터 TFM2의 충격반응가중함수를 사전백색화추정법으로 추정하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM 4-11.PRG에 수록되어 있다).

```

/*                               */
/*                               EXAM4-11.PRG                               */
/*                               */

```

```

/*          2SLS ESTIMATE          */
/*          USING PREWHITENING     */
/*          DATASET = TFM2         */
/*          */                       */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS LINESIZE = 72;
OPTIONS PAGESIZE = 80;
DATA EXAM411;
    SET TSA4DATA.TFM2;
    RUN;
TITLE1 'PREWHITENING INPUT & OUTPUT SERIES';
PROC ARIMA DATA = EXAM411;
    I VAR=X NOPRINT;
    E P=1 Q=1 NOPRINT;
    I VAR=Y CROSS=(X) NOPRINT
    OUTCOV = EXAM411B;
    RUN;
DATA EXAM411C;
    SET EXAM411B;
    IF INVCORR NE . THEN DELETE;
    IF LAG < 0 THEN DELETE;
    NU = SQRT(32.46778/9.185686)*CORR;
    KEEP LAG VAR CROSSVAR N CORR STDERR NU;
    RUN;
PROC PRINT DATA = EXAM411C;
    RUN;
DATA TSA4DATA.NU2;
    SET EXAM411C;
    KEEP LAG N CORR STDERR NU;
    RUN;

```

이 SAS 프로그램을 수행한 결과가 그림 4.11이다. 이 SAS 프로그램에서 PROC ARIMA의 첫 번째 IDENTIFY 스테이트먼트는 입력시계열데이터 X를 식별하기 위한 것이고, 첫 번째 ESTIMATE 스테이트먼트는 입력시계열데이터를 선택된 ARMA(1,1) 모형에 적합시켜 추정하기 위한 것이다. 여기까지는 앞의 예제와 동일하다. 따라서 이 ESTIMATE 스테이트먼트를 수행한 결과로 다음과 같은 ARMA(1,1) 모형의 추정식을 얻는다.

$$(1 - 0.5113B)(x_t + 1.6646) = (1 + 0.3825B)z_t$$

여기서 $\{z_t\}$ 는 사전백색화된 입력시계열데이터이다. 두 번째 IDENTIFY 스테이트먼트를

PREWHITENING INPUT & OUTPUT SERIES								40
								23:12 Saturday, September 18, 1993
OBS	LAG	VAR	CROSSVAR	N	CORR	STDERR	NU	
1	0	Y	X	101	0.13572	0.099504	0.25516	
2	1	Y	X	100	-0.09658	0.099504	-0.18157	
3	2	Y	X	99	0.13128	0.099504	0.24681	
4	3	Y	X	98	0.41452	0.099504	0.77931	
5	4	Y	X	97	0.44153	0.099504	0.85011	
6	5	Y	X	96	0.00070	0.099504	0.00131	
7	6	Y	X	95	-0.06966	0.099504	-0.13096	
8	7	Y	X	94	0.12328	0.099504	0.23177	
9	8	Y	X	93	-0.11874	0.099504	-0.22323	
10	9	Y	X	92	-0.00338	0.099504	-0.00635	
11	10	Y	X	91	-0.19804	0.099504	-0.37252	
12	11	Y	X	90	-0.01054	0.099504	-0.01981	
13	12	Y	X	89	-0.05770	0.099504	-0.10849	
14	13	Y	X	88	-0.01133	0.099504	-0.02130	
15	14	Y	X	87	0.09444	0.099504	0.17755	
16	15	Y	X	86	0.05246	0.099504	0.09862	
17	16	Y	X	85	0.08883	0.099504	0.16700	
18	17	Y	X	84	0.06950	0.099504	0.13066	
19	18	Y	X	83	-0.06325	0.099504	-0.11892	
20	19	Y	X	82	0.07861	0.099504	0.14779	
21	20	Y	X	81	-0.07302	0.099504	-0.13728	
22	21	Y	X	80	-0.02900	0.099504	-0.05452	
23	22	Y	X	79	0.04407	0.099504	0.08286	
24	23	Y	X	78	0.00108	0.099504	0.00203	
25	24	Y	X	77	0.01556	0.099504	0.02925	

그림 4.11 사전백색화방법을 이용한 충격반응가중함수의 추정

수행한 결과를 옵션 OUTCOV = 에 지정한 SAS데이터세트 EXAM411B에 저장한다. 다음의 DATA스텝은 SAS데이터세트 EXAM411B의 변수 CORR를 다음과 같이 변형해서 충격반응가중함수를 추정하기 위한 것이다.

$$\hat{\nu}_k = \frac{\sqrt{\hat{\sigma}_{uw}(0)}}{\sqrt{\hat{\sigma}_z(0)}} \hat{\rho}_{wz}(k) \quad (k=0, 1, 2, \dots)$$

여기서 $\hat{\sigma}_{uw}(0)$ 와 $\hat{\sigma}_z(0)$ 는 PROC ARIMA의 두 번째 IDENTIFY스테이트먼트에서 얻을 수 있는데, 이 예제에서는 $\hat{\sigma}_{uw}(0)$ 와 $\hat{\sigma}_z(0)$ 의 값들이 각각 32.468과 9.186이다. 이렇게 만들어진 SAS데이터세트를 PROC PRINT를 사용해서 OUTPUT윈도우에 출력시킨 것이 그림 4.11이다. 이 SAS프로그램의 세 번째 DATA스텝은 이렇게 추정된 충격반응가중함수를 SAS데이터라이브러리 TSA4DATA의 파일 NU2에 저장하기 위한 것이다.

4.2.4 능형회귀추정방법

Edlund(1990)는 유한차원의 선형전이함수모형의 충격반응가중함수를 추정하기 위해서 능형회귀추정법(ridge regression method)을 사용할 것을 제안하였다. 능형회귀추정량의 편의(bias)를 갖는에도 불구하고 이 추정량을 사용하는 이유는 입력시계열데이터의 현재항과 과거항들 사이에는 다공선성(multicollinearity)이 존재하기 때문이다. 이 다공선성은 최소제곱추정량의 분산을 지나치게 크게 만든다. 이것이 약간의 편의를 가지나 분산이

최소제곱추정량의 분산보다 훨씬 작은 능형회귀추정법을 사용하는 이유이다. 유한차원의 선형전이함수모형에서 능형회귀추정식은 다음과 같다.

$$v = (X'X + cI)^{-1}X'y$$

여기서 상수 $c(>0)$ 는 수축요인(shrinkage factor)이다.

수축요인 c 를 정하는 방법은 여러 가지가 있다. Edlund(1990)는 모의실험을 해서 여러 종류의 능형회귀추정량들을 비교하였다. 행렬 $X'X$ 의 고유값들을 $\lambda_1, \lambda_2, \dots, \lambda_k$ 라 하고, 편의상 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$ 라고 하자. 행렬 Λ 가 이 고유값들을 대각원소들로 하는 대각행렬이고, 제 i 번째 고유값 λ_i 에 대한 고유벡터를 제 i 번째 열로 하는 행렬을 U 라 하고, k 차원 벡터 α 를 다음과 같이 정의하자.

$$\begin{aligned}\hat{\alpha} &= (\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_k)' \\ &= \Lambda^{-1}XU'y\end{aligned}$$

또한 $\hat{\sigma}^2$ 는 잡음항 $\{n_i\}$ 의 분산의 추정값이라고 하자. Edlund가 비교한 능형회귀추정량들의 수축요인들은 다음과 같다.

- ① Hoerl, Kennard & Baldwin(1975)이 제안한 수축요인

$$c_1 = \frac{k\hat{\sigma}^2}{\sum_{i=1}^k \hat{\alpha}_i^2}$$

- ② Hocking, Speed & Lynn(1976)이 제안한 수축요인

$$c_2 = \frac{\sigma^2 \sum_{i=1}^k \lambda_i^2 \hat{\alpha}_i^2}{\sum_{i=1}^k \lambda_i^2 \hat{\alpha}_i^4}$$

- ③ Lawless & Wang(1976)이 제안한 수축요인

$$c_3 = \frac{k\hat{\sigma}^2}{\sum_{i=1}^k \lambda_i \hat{\alpha}_i^2}$$

Edlund(1990)는 이 능형회귀추정량들 중에서 표준오차가 가장 작은 Lawless-Wang추정량을 사용하여 충격반응가중값들을 추정할 것을 제안하였다. 그러나 능형회귀추정법에서 수축요인을 정하는 것은 주관적인 판단에 의존하기 때문에 능형회귀추정법은 조심해서 사용해야 한다. 특히 충격반응가중함수를 추정하는 데 능형회귀추정법이 널리 사용되는 것은 아니다.

4.3 전이함수모형의 식별

이 절에서는 유리함수형태의 전이함수모형의 차수들 (b, r, s)를 정하는, 즉 전이함수모형을 식별하는 방법들에 대해서 살펴보자. 전이함수모형을 식별하기 위해서는 우선 제4.2 절에서 소개된 방법등을 사용해서 충격반응가중함수를 추정하고, 이 추정된 충격반응가중함수를 사용해서 전이함수모형을 식별한다.

사전백색화추정법을 적용하면 충격반응가중함수는 사전백색화된 입력시계열데이터와 여과된 출력시계열데이터의 교차상관함수에 일정한 상수를 곱한 것과 같다. 따라서 사전백색화방법을 적용하는 경우에는 충격반응가중함수 대신에 교차상관함수를 분석해도 된다. SAS/ETS의 PROC ARIMA를 사용하면, 사전백색화방법을 적용해서 얻은 입력시계열데이터와 출력시계열데이터의 표본교차상관함수가 계산되고 따라서 쉽게 충격반응가중함수의 형태를 조사할 수 있다. 그러므로 이 절에서는 사전백색화추정법을 사용하기로 한다. 우선 여러 변환방법들이나 차분 등을 이용해서 출력시계열데이터 $\{y_t\}$ 와 입력시계열데이터 $\{x_t\}$ 가 결합정상적이 되도록 만든 뒤, 다음과 같은 3단계 방법으로 표본충격반응가중함수를 구한다.

제1단계 입력시계열데이터가 백색잡음과정이 아니면 이 입력시계열데이터를 사전백색화시킨다. 입력시계열데이터 $\{x_t\}$ 가 ARMA모형을 따른다고 가정하고, 이 입력시계열데이터를 다음과 같이 ARMA모형화한다.

$$\phi_x(B)x_t = \theta_x(B)z_t$$

사전백색화된 입력시계열데이터 $\{z_t\}$ 는 다음과 같이 나타낼 수 있다.

$$z_t = \frac{\phi_x(B)}{\theta_x(B)} x_t$$

단, $\{z_t\}$ 는 평균 0, 분산 σ_z^2 인 백색잡음과정

제2단계 사전백색화하기 위해서 입력시계열데이터 $\{x_t\}$ 에 적용한 필터를 출력시계열데이터 $\{y_t\}$ 에 적용하여 다음과 같은 여과된 출력시계열데이터 $\{w_t\}$ 를 생성한다.

$$w_t = \frac{\phi_x(B)}{\theta_x(B)} y_t$$

제3단계 백색잡음시계열데이터 $\{z_t\}$ 와 여과된 출력시계열데이터 $\{w_t\}$ 사이의 표본교차상관함수 $\{\rho_{zw}(k)\}$ 를 계산한 뒤, 이를 이용해서 다음과 같이 충격반응가중함수 $\{\nu_k\}$ 를 추정한다.

$$\hat{\nu}_k = \frac{\sqrt{\hat{\sigma}_{uw}(0)}}{\sqrt{\hat{\sigma}_z(0)}} \hat{\rho}_{wz}(k)$$

제3단계에서 볼 수 있듯이 사전백색화추정법을 적용하면 충격반응가중함수는 사전백색화된 입력시계열데이터와 여과된 출력시계열데이터의 교차상관함수에 상수를 곱한 것과 같다. 따라서 실제 시계열데이터를 분석할 때는 표본충격반응가중함수의 형태를 조사하는 대신에 사전백색화과정을 거친 시계열데이터들의 표본교차상관함수의 형태를 조사한다.

유리함수형태의 전이함수모형을 식별하는 데는 다양한 방법들을 적용할 수 있다. 이 방법들은 근본적으로 ARIMA모형을 식별하는 방법들을 변형한 것들이다. 이 절에서는 Box-Jenkins방법과 모퉁이방법을 소개한다. 이 방법들 이외에도 [Choi92] (pp. 102~112)의 3배열방법이나 McIntire(1977)와 Kelly(1977)의 R배열과 S배열 방법 등을 사용해서 유리함수형태의 전이함수모형을 식별할 수 있다.

4.3.1 Box-Jenkins방법

현존하는 시계열분석에 관한 문헌들 중에서 전이함수모형의 식별에 가장 널리 쓰이고 있는 것이 Box & Jenkins(1976, pp. 377~388)에 의한 방법이다. 이 Box-Jenkins방법은 표본충격반응가중함수의 형태와 제4.1.3소절에서 기술된 이론적 충격반응가중함수의 형태를 비교해서 전이함수모형을 식별하는 것이다. 즉, 사전백색화과정을 거친 시계열데이터들의 표본교차상관함수의 형태를 이론적 충격반응함수와 비교해서 전이함수모형을 식별한다.

실제로 시계열데이터를 분석하는 데 있어서 전이함수모형의 분모의 차수 r 과 분자의 차수 s 가 2를 초과하는 경우는 드물다. 따라서 널리 사용되는 충격반응가중함수들의 형태는 다음과 같다.

1) $r=0$

이 경우 충격반응가중함수는 표 4.1에서 볼 수 있듯이 시차 b 에서 충격반응가중값 $\nu_b = \omega_0$ 과 시차 b 에서 충격반응가중값 $\nu_{b+s} = -\omega_0$ 가 0이 아니고, 이 두 시차들 사이의 충격반응가중값들은 0이 아닐 수 있으며, 그 외의 시차들에서는 충격반응가중값들이 모두 0이다.

2) $r=1$

이 경우 충격반응가중함수는 표 4.2에서 볼 수 있듯이, 만일 $s=0$ 이면 시차 b 에서 충격반

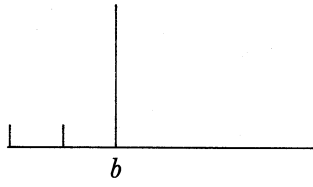
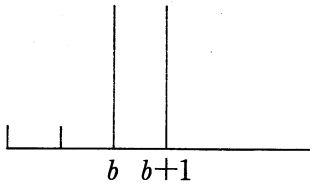
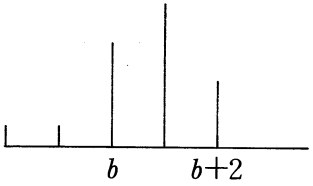
(b, r, s)	전이함수모형	충격반응가중함수
$(2, 0, 0)$	$y_t = \omega_0 x_{t-2}$	
$(2, 0, 1)$	$y_t = (\omega_0 - \omega_1 B) x_{t-2}$	
$(2, 0, 2)$	$y_t = (\omega_0 - \omega_1 B - \omega_2 B^2) x_{t-2}$	

표 4.1 충격반응가중함수의 형태($r=0$)

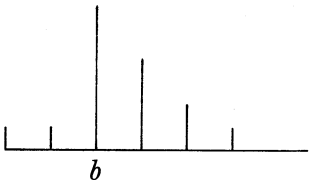

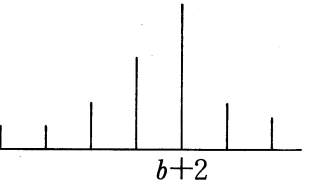
(b, r, s)	전이함수모형	충격반응가중함수
$(2, 1, 0)$	$y_t = \frac{\omega_0}{(1 - \delta_1 B)} x_{t-2}$	
$(2, 1, 1)$	$y_t = \frac{(\omega_0 - \omega_1 B)}{(1 - \delta_1 B)} x_{t-2}$	
$(2, 1, 2)$	$y_t = \frac{(\omega_0 - \omega_1 B - \omega_2 B^2)}{(1 - \delta_1 B)} x_{t-2}$	

표 4.2 충격반응가중함수의 형태($r=1$)

증가증값 ν_b 가 0이 아니고 이 시차로부터 충격반응가중값들이 지수적으로 감소한다. 또한, $s=1$ 이면 시차 $(b+1)$ 에서 충격반응값 ν_{b+1} 이 0이 아니고 이 시차로부터 충격반응가중값들이 지수적으로 감소하며, $s=2$ 이면 시차 $(b+2)$ 에서 충격반응가중값 ν_{b+2} 가 0이 아니고 이 시차로부터 충격반응가중값들이 지수적으로 감소한다.

3) $r=2$

이 경우 충격반응가중함수는 표 4.3에서 볼 수 있듯이, 충격반응가중값들은 전이함수의 분모의 다항식 $\delta_2(z)=0$ 의 근들에 의존하는 시나브로형태를 보인다. 시나브로형태란 시차가 커짐에 따라 크기가 지수적으로 감소하는 형태나 지수적으로 감소하는 사인함수형태를 말한다. 만일, 다항식 $\delta_2(z)=0$ 의 근이 실근이면 충격반응가중함수의 절대값은 지수적으로 감소하는 형태를 보이고, 복소수근이면 충격반응가중함수의 절대값은 지수적으로 감소하는 사인함수형태를 보인다.

전이함수모형의 첫 번째 예제로써 시계열데이터 TFM1을 Box-Jenkins방법으로 식별하기로 하자. 그림 4.10의 부분 A2에는 사전백색화과정을 거친 입력시계열데이터와 출력시계열데이터의 표본교차상관함수가 실려 있다. 이 부분에서 시차 2와 시차 3의 표본교차상관계수들이 유의적으로 0과 같지 않음을 알 수 있고, 시차 4 이후의 표본교차상관계수들이 유의적으로 0과 같다고 할 수 있다. 반면에 시차 2 이후의 표본교차상관계수들이 시나브로형태를

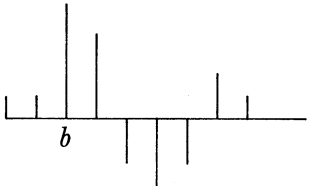
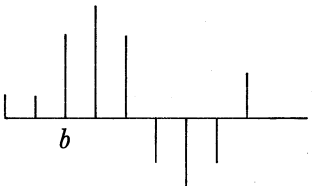
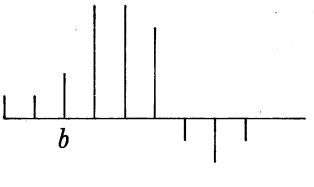
(b, r, s)	전이함수모형	충격반응가중함수
$(2, 2, 0)$	$y_t = \frac{\omega_0}{(1 - \delta_1 B - \delta_2 B^2)} x_{t-2}$	
$(2, 2, 1)$	$y_t = \frac{(\omega_0 - \omega_1 B)}{(1 - \delta_1 B - \delta_2 B^2)} x_{t-2}$	
$(2, 2, 2)$	$y_t = \frac{(\omega_0 - \omega_1 B - \omega_2 B^2)}{(1 - \delta_1 B - \delta_2 B^2)} x_{t-2}$	

표 4.3 충격반응가중함수의 형태($r=2$)

보인다고 할 수도 있다. 따라서 Box-Jenkins식별방법에 의하면 이 시계열데이터는 차수들이 (2, 0, 1), 또는 (2, 1, 0)인 전이함수모형으로 식별할 수 있다.

두 번째 예제로써 시계열데이터 TFM2를 Box-Jenkins방법을 사용해서 식별하기 위해서, 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM4-12.PRG에 수록되어 있다).

```

/ * * /
/ *          EXAM4-12.PRG          * /
/ * * /
/ *          TRANSFER FUNCTION MODEL IDENTIFICATION          * /
/ *          USING BOX-JENKINS METHOD          * /
/ *          DATASET = TFM2          * /
/ * * /

LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS LINESIZE = 72;
OPTIONS PAGESIZE = 80;
DATA EXAM412;
    SET TSA4DATA.TFM2;
    RUN;
TITLE1 'IDENTIFICATION BY BOX-JENKINS METHOD';
PROC ARIMA DATA = EXAM412;
    I VAR=X NOPRINT;
    E P=1 Q=1 NOPRINT;
    I VAR=Y CROSS=(X);
    RUN;

```

이 SAS프로그램을 수행한 결과가 그림 4.12이다. 그림 4.12에는 사전백색화과정을 거친 입력시계열데이터와 같은 필터로 여과된 출력시계열데이터의 표본교차상관함수가 실려 있다. 이 표본교차상관함수에서 시차 3과 시차 4의 표본교차상관계수들이 유의적으로 0과 같지 않고 또한 시차 5 이후의 표본교차상관계수들이 시나브로형태를 보인다고 할 수도 있다. 따라서 이 시계열데이터는 Box-Jenkins방법에 의해서 차수들이 (3, 1, 1)인 전이함수모형으로 식별할 수 있다.

이 예제들에서 알 수 있듯이 Box-Jenkins방법에 의한 전이함수모형의 식별에서도 Box-Jenkins방법에 의한 ARIMA모형의 식별에서 생기는 문제점들이 발생한다. 이러한 문제점들에 대해서는 [SAS4TSA2] (pp. 433~434)를 참조하라.

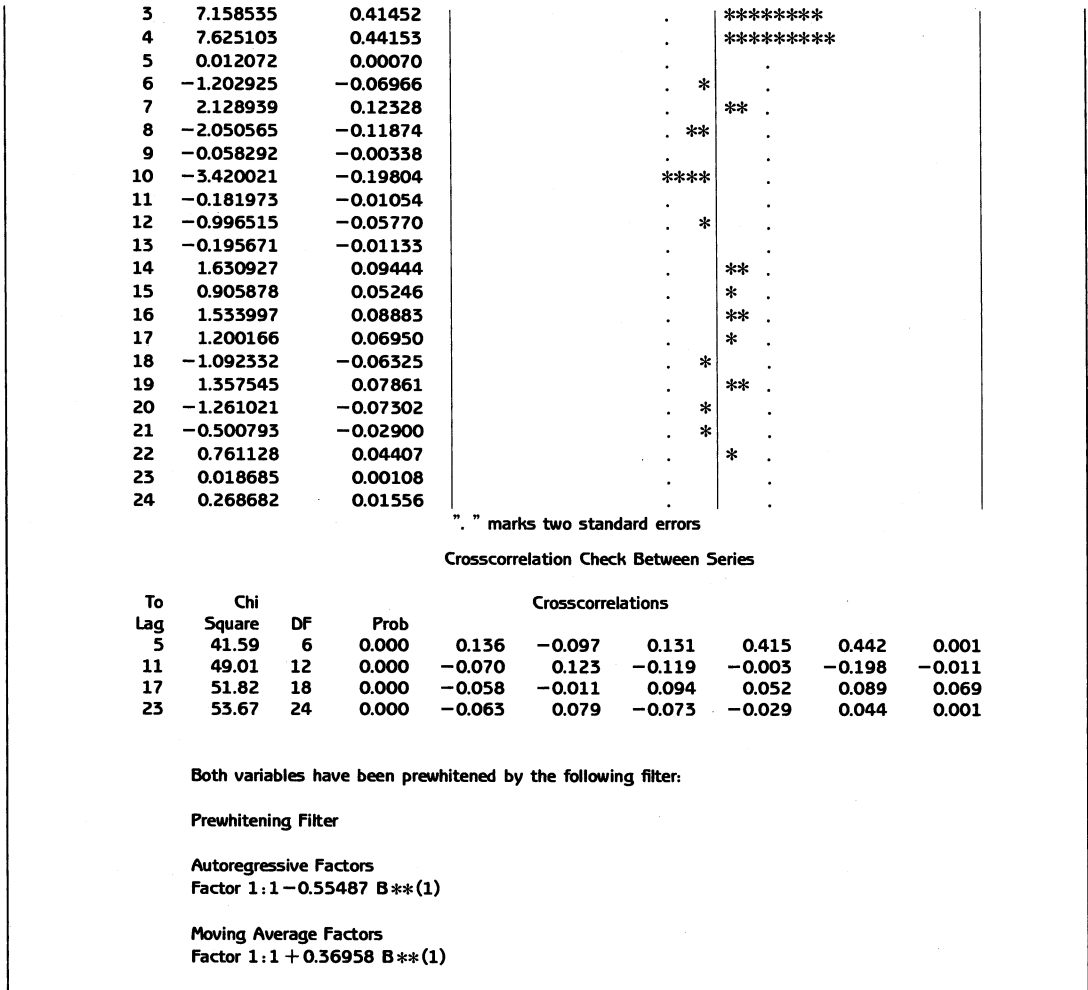


그림 4.12 Box-Jenkins방법에 의한 시계열데이터 TFM2의 식별

4.3.2 모퉁이방법

Liu & Hanssens(1982)은 Beguin, Gourieroux & Monfort(1980)가 ARIMA모형의 식별을 위해서 제안한 모퉁이방법(corner method)을 변형시켜 유리함수형태의 전이함수모형을 식별에 적용할 것을 제안하였다.

모퉁이방법을 설명하기 위해서, $k=1, 2, 3, \dots, i=0, 1, 2, \dots$ 에 대해서 다음과 같이 행렬 $B(k, i)$ 를 정의하자.

$$B(k, i) = \begin{bmatrix} \nu_i & \nu_{i-1} & \nu_{i-2} & \dots & \nu_{i-k+1} \\ \nu_{i+1} & \nu_i & \nu_{i-1} & \dots & \nu_{i-k+2} \\ \nu_{i+2} & \nu_{i+1} & \nu_i & \dots & \nu_{i-k+3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \nu_{i+k-1} & \nu_{i+k-2} & \nu_{i+k-3} & \dots & \nu_i \end{bmatrix}$$

$$\text{단, } \nu_{-k} = 0 \quad (k=1, 2, \dots)$$

또한, 행렬 $B(k, i)$ 의 행렬식을 $D(k, i)$ 라 하자. 즉, 다음 식이 성립한다.

$$D(k, i) = \det[B(k, i)]$$

제 4.1.3소절의 정리 4.1의 차분방정식으로부터 다음 정리를 유도할 수 있다.

정리 4.2 유리함수형태의 전이함수모형의 차수들이 (b, r, s) 이기 위한 필요충분조건은 다음과 같다.

$$D(k, i) = 0 \quad (k \geq r+1 \text{ and } i \geq s+b+1)$$

$$D(k, i) = 0 \quad (i \leq b-1)$$

$$D(r, i) \neq 0 \quad (i \geq s+b)$$

$$D(k, s+b) \neq 0 \quad (k \geq r)$$



정리 4.2의 행렬식들을 배열한 D 배열은 표 4.4와 같다.

$k \setminus i$	0	1	...	$b-1$	b	...	$s+b$	$s+b+1$...	M
1	0	0	...	0	$D(1, b)$...	$D(1, s+b)$	$D(1, s+b+1)$...	$D(1, M)$
2	0	0	...	0	$D(2, b)$...	$D(2, s+b)$	$D(2, s+b+1)$...	$D(2, M)$
:	:	:	:	:	:	:	:	:	:	:
r	0	0	...	0	$D(r, b)$...	X	X	...	X
$r+1$	0	0	...	0	$D(r+1, b)$...	X	0	...	0
$r+2$	0	0	...	0	$D(r+2, b)$...	X	0	...	0
:	:	:	:	:	:	:	:	:	:	:
M	0	0	...	0	$D(r+s, b)$...	X	0	...	0

표 4.4 D 배열(여기서 X는 0이 아님을 뜻한다)

정리 4.2를 예증하기 위해서 두 개의 유리함수들의 D 배열들을 살펴보자. 즉, 잡음항이 없는 전이함수모형들의 D 배열들을 살펴보자. 첫 번째 예제로써 다음 전이함수모형을 살펴보자.

$$\nu(z) = \frac{1+0.4z}{1-0.8z} z^2$$

SAS프로그램 EXAM4-1.PRG에서 이 전이함수의 충격반응가중함수를 영구SAS데이터 세트 TSA4DATA.IRF1에 수록하였다. 이 전이함수의 D 배열을 구하기 위해서 다음 SAS 프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM4-13.PRG에 수록되어 있다).

```

/* */
/*          EXAM4-13.PRG          */
/* */
/*          CALCULATING CORNER ARRAY          */
/*          FOR TRANSFER FUNCTION MODEL          */
/*          DATASET = IRF1          */
/* */
LIBNAME TSA4DATA 'A:\TSA4DATA';
PROC IML;
RESET NOLOG NONAME NOCENTER FW = 6;
/* MODULE FOR CORNER METHOD */
START CORNER(NU);
  MAXORDER = 11;
  /* MAKE THE LARGEST B MATRIX */
  NU = NU;
  BIGMAT = TOEPLITZ(NU);
  DIMBIG = NROW(BIGMAT);
  KDUM = 1;
  DO WHILE(KDUM <= DIMBIG);
    IDUM = 1;
    DO WHILE(IDUM <= DIMBIG);
      IF IDUM - KDUM > 0
        THEN BIGMAT[KDUM, IDUM] = 0;
      IDUM = IDUM + 1;
    END;
    KDUM = KDUM + 1;
  END;
  /* MAKE D ARRAY */
  CORNER = SHAPE(0, MAXORDER - 1, MAXORDER);
  I = 0;
  DO WHILE(I < MAXORDER);
    I1 = I + 1;
    K = 1;
    DO WHILE(K < MAXORDER);
      K1 = K + 1;
      MAT = BIGMAT[I1:I1 + K1 - 2, 1:K1 - 1];
      IF ABS(DET(MAT)) < 1.0E - 10
        THEN CORNER[K, I1] = 0;
    END;
    I = I + 1;
  END;
END;

```

SAS			10:34 Sunday, September 26, 1993 15							
CORNER ARRAY										
0	0	1	1.2	0.96	0.768	0.6144	0.4915	0.3952	0.3146	0.2517
0	0	1	0.48	0	0	0	0	0	0	0
0	0	1	0.192	0	0	0	0	0	0	0
0	0	1	0.0768	0	0	0	0	0	0	0
0	0	1	0.0307	0	0	0	0	0	0	0
0	0	1	0.0123	0	0	0	0	0	0	0
0	0	1	0.0049	0	0	0	0	0	0	0
0	0	1	0.002	0	0	0	0	0	0	0
0	0	1	0.0008	0	0	0	0	0	0	0
0	0	1	0.0005	0	0	0	0	0	0	0
5 workspace compresses.										
Exiting IML.										

그림 4.13 전이함수모형의 D배열

```

ELSE CORNER[K, I1] = DET(MAT);
K = K1;
END;
I = I1;
END;
PRINT, 'CORNER ARRAY' , CORNER,;
FINISH;
/* END OF MODULE CORNER */
USE TSA4DATA.IRF1;
READ ALL;
CLOSE TSA4DATA.IRF1;
RUN CORNER(NU);
QUIT;

```

이 SAS프로그램을 수행한 결과가 그림 4.13이다. 그림 4.13의 D배열과 정리 4.2의 D배열을 비교하면 다음 식들을 얻을 수 있다.

$$\begin{aligned}
 b - 1 &= 1 \\
 s + b + 1 &= 4 \\
 r + 1 &= 2
 \end{aligned}$$

따라서 모퉁이방법을 사용하면 이 전이함수모형은 $(b, r, s) = (2, 1, 1)$ 로 식별할 수 있다. 오차항이 없는 전이함수모형의 두 번째 예제로써 다음 전이함수모형을 살펴보자.

$$\nu(z) = \frac{1 + 0.7z}{1 - 0.5z - 0.3z^2} z^3$$

SAS		11:55 Sunday, September 26, 1993 1									
CORNER ARRAY											
0	0	0	1	1.2	0.9	0.81	0.675	0.5805	0.4927	0.4205	
0	0	0	1	0.54	-0.162	0.0486	-0.015	0.0044	-0.001	0.0004	
0	0	0	1	0.378	0	0	0	0	0	0	
0	0	0	1	0.2646	0	0	0	0	0	0	
0	0	0	1	0.1852	0	0	0	0	0	0	
0	0	0	1	0.1297	0	0	0	0	0	0	
0	0	0	1	0.0908	0	0	0	0	0	0	
0	0	0	1	0.0635	0	0	0	0	0	0	
0	0	0	1	0.0445	0	0	0	0	0	0	
0	0	0	1	0.0311	0	0	0	0	0	0	
5 workspace compresses.											
Exiting IML.											

그림 4.14 전이함수모형의 D 배열

SAS프로그램 EXAM4-2.PRG에서 이 전이함수의 충격반응가중함수를 영구SAS데이터 세트 TSA4DATA.IRF2에 수록하였다. 이 전이함수의 D 배열을 구하기 위해서 다음 SAS 프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM4-14.PRG에 수록되어 있다).

```

/*                                     */
/*           EXAM4-14.PRG              */
/*                                     */
/*           CALCULATING CORNER ARRAY  */
/*           FOR TRANSFER FUNCTION MODEL*/
/*           DATASET = IRF2            */
/*                                     */
LIBNAME TSA4DATA 'A:\TSA4DATA';
PROC IML;
RESET NOLOG NONAME NOCENTER FW = 6;

/*                                     */
/*           PUT MODULE 'CORNER'       */
/*           OF EXAM4-13.PRG HERE     */
/*                                     */

USE TSA4DATA.IRF2;
READ ALL;
CLOSE TSA4DATA.IRF2;
RUN CORNER(NU);
QUIT;

```

이 SAS프로그램을 수행한 결과가 그림 4.14이다. 그림 4.14의 D 배열과 정리 4.2의 D 배열

을 비교하면 다음 식들을 얻을 수 있다.

$$\begin{aligned} b-1 &= 2 \\ s+b+1 &= 5 \\ r+1 &= 3 \end{aligned}$$

따라서 모퉁이방법을 사용하면 이 전이함수모형은 $(b, r, s) = (3, 2, 1)$ 로 식별할 수 있다.

실제 시계열데이터분석에서 모퉁이방법의 유용성을 보이기 위해서 두 종류의 시계열데이터들에 모퉁이방법을 적용하기로 하자. 첫 번째 예제로 시계열데이터 TFM1을 모퉁이방법으로 식별하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM4-15.PRG에 수록되어 있다).

```

/*                                     */
/*           EXAM4-15.PRG              */
/*                                     */
/*           CALCULATING CORNER ARRAY  */
/*           FOR TRANSFER FUNCTION MODEL*/
/*           DATASET = TFM1            */
/*                                     */
LIBNAME TSA4DATA 'A:\TSA4DATA';
PROC IML;
  RESET NOLOG NONAME NOCENTER FW = 6;

      /*                               */
      /* PUT MODULE 'CORNER'           */
      /* OF EXAM4-13.PRG HERE          */
      /*                               */

USE TSA4DATA.NU1;
READ ALL;
CLOSE TSA4DATA.NU1;
RUN CORNER(NU);
QUIT;

```

시계열데이터 TFM1의 표본충격반응가중함수가 SAS데이터라이브러리 TSA4DATA의 파일 NU1에 수록되어 있다. 이 SAS프로그램에서는 표본충격반응가중함수에 모퉁이방법을 적용한다. 이 SAS프로그램을 수행한 결과가 그림 4.15이다. 그림 4.15의 D배열과 정리 4.2의 D배열을 비교하면 다음 식들을 얻을 수 있다.

$$b-1=1$$

SAS		11:35 Sunday, September 26, 1993 2								
CORNER ARRAY										
0.2301	-0.119	1.0376	0.3232	0.069	0.0246	-0.115	0.1549	-0.263	-0.052	-0.239
0.0529	-0.224	1.1152	0.0329	-0.003	0.0085	0.0093	-0.006	0.0772	-0.06	0.0576
0.0122	0.0724	1.2057	0.0144	-0.004	0.0042	-0.001	-0.004	-0.021	0.0161	-0.018
0.0028	0.0421	1.3026	0.1496	0.0139	0.0015	0.0021	0.001	0.0068	0.0021	0.005
0.0006	-0.026	1.4021	0.2978	0.0068	-0.007	-0.002	0.003	-0.002	-0.002	-85E-5
0.0001	-0.005	1.5151	0.5285	0.1441	0.0407	0.0121	0.0039	0.0014	0.0007	0.0001
341E-7	0.0075	1.6393	0.2047	-0.107	0.0141	0.0046	-73E-5	0.0004	-2E-4	-26E-6
785E-8	-83E-6	1.7726	0.4106	0.0592	0.0169	0.0026	-33E-5	645E-8	0.0001	0.0001
181E-8	-0.002	1.9169	0.3109	0.032	0.0094	0.0027	-13E-5	581E-7	-21E-6	482E-7
416E-9	0.0004	2.0731	0.086	-0.032	0.0002	0.0032	0.0004	0.0001	-33E-6	414E-7
Exiting IML.										

그림 4.15 시계열데이터 TFM1의 전이함수모형 식별

$$s + b + 1 = 4$$

$$r + 1 = 2$$

따라서 모퉁이방법을 사용하면 이 전이함수모형은 $(b, r, s) = (2, 1, 1)$ 로 식별할 수 있다. 제4.1.3소절에서 알 수 있듯이 이 시계열데이터는 차수들이 $(b, r, s) = (2, 1, 1)$ 인 모형에서 발생되었음을 유의하라.

모퉁이방법을 실제 시계열분석에 적용하기 어려운 점은 D 배열의 각 원소가 0인지 아닌지를 결정하는 것이다. [Choi](1991)의 결과를 이용하면 이 문제를 통계적으로 해결할 수 있다. 그러나 이 방법은 많은 계산을 필요로 할 뿐 아니라 이론적으로도 쉬운 내용이 아니다. 본 저자는 경험을 바탕으로 다음과 같은 주관적인 판단기준을 가지고 있다. 만일, D 배열의 한 원소의 절대값이 좌측의 원소의 절대값, 또는 위쪽의 원소의 절대값의 0.1배 정도이거나 그 이하이면 그 원소를 0으로 간주한다. 그러나 이는 어디까지나 본 저자의 주관적 기준임을 기억하라.

두 번째 예제로 시계열데이터 TFM2를 모퉁이방법으로 식별하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM4-16.PRG에 수록되어 있다).

```

/* */
/*          EXAM4-16.PRG          */
/* */
/*          CALCULATING CORNER ARRAY          */
/*          FOR TRANSFER FUNCTION MODEL          */
/*          DATASET = TFM2          */
/* */
LIBNAME TSA4DATA 'A:\TSA4DATA';
PROC IML;
RESET NOLOG NONAME NOCENTER FW = 6;

```

SAS		11:35 Sunday, September 26, 1993 3									
CORNER ARRAY											
0.2552	-0.182	0.2468	0.7793	0.8301	0.0013	-0.131	0.2318	-0.223	-0.006	-0.372	
0.0651	-0.03	0.2024	0.4024	0.6881	0.1087	0.0168	0.0245	0.0513	-0.083	0.1385	
0.0166	0.0676	0.2149	0.0291	0.5176	0.1731	0.0182	-0.001	-0.021	0.0322	-0.043	
0.0042	-0.033	0.2185	-0.274	0.3821	0.1892	0.0313	0.0156	0.0092	-0.002	0.0134	
0.0011	0.0028	0.1795	-0.282	0.3823	0.1378	-0.108	0.0413	-0.005	-0.004	-0.005	
0.0003	0.0056	0.1511	-0.041	0.4844	0.3187	0.1918	0.0721	0.0199	0.0061	0.0023	
0.0001	-0.004	0.1284	0.2533	0.6475	0.063	-0.128	0.0334	0.0076	0.0026	-59E-5	
18E-6	0.0009	0.1155	0.4682	0.8326	0.2717	0.0739	0.029	-0.001	0.0018	0.0001	
458E-8	0.0003	0.1005	0.4856	0.8742	0.1946	0.0188	0.0282	-0.007	0.0013	-21E-5	
117E-8	-38E-5	0.086	0.3161	0.8043	0.0787	-0.069	0.0319	-0.005	0.0002	-24E-6	
Exiting IML.											

그림 4.16 시계열데이터 TFM2의 전이함수모형 식별

```

/ *                               */
/ * PUT MODULE 'CORNER'          */
/ * OF EXAM4-13.PRG HERE         */
/ *                               */

USE TSA4DATA.NU2;
READ ALL;
CLOSE TSA4DATA.NU2;
RUN CORNER(NU);
QUIT;

```

시계열데이터 TFM2의 표본충격반응가중합수가 SAS데이터라이브러리 TSA4DATA의 파일 NU2에 수록되어 있다. 이 SAS프로그램에서는 표본충격반응가중합수에 모퉁이방법을 적용한다. 이 SAS프로그램을 수행한 결과가 그림 4.16이다. 그림 4.16의 D배열과 정리 4.2의 D배열을 비교하면 다음 식들을 얻을 수 있다.

$$\begin{aligned}
 b - 1 &= 2 \\
 s + b + 1 &= 6 \\
 r + 1 &= 2
 \end{aligned}$$

따라서 모퉁이방법을 사용하면 이 전이함수모형은 $(b, r, s) = (3, 1, 1)$ 로 식별할 수 있다. 이 모퉁이방법을 좀더 통계적으로 적용하기 위해서는 추정된 D배열의 원소들의 점근분포를 구해야 한다. 이는 간단한 문제가 아니다. 모퉁이방법과 같은 방법을 형태인식에 의한 전이함수모형 식별방법이라고 하는데, 다른 형태인식에 의한 전이함수모형 식별방법으로는 [Choi](1991, 1993)의 3배열방법이나 Kelly(1977)와 McIntire(1977)의 R배열과 S배열을 이용한 방법 등이 있다. 이에 대한 자세한 내용은 장승욱(1992)을 참조하라.

4.3.3 잡음시계열데이터의 식별

제4.2절에서 추정한 표본충격반응가중값들과 지금까지 다룬 유리함수형태의 전이함수모형의 식별방법들에 의해서 선택된 차수들 (b, r, s) 를 이용해서, 유리함수형태의 전이함수모형의 분자식 $\omega_s(z)$ 와 분모식 $\delta_r(z)$ 를 임시로(temporarily) 추정할 수 있다. 즉, 다음 항등식으로부터 전이함수모형의 분자식과 분모식을 임시로 추정할 수 있다.

$$\hat{v}(z)\hat{\delta}_r(z) \equiv \hat{\omega}_s(z)z^b$$

그러나 유리함수형태의 전이함수모형의 차수들 (b, r, s) 가 주어졌을 때, SAS/ETS의 PROC ARIMA에서 전이함수의 분자식 $\omega_s(z)$ 와 분모식 $\delta_r(z)$ 를 임시로 추정하는 방법은 다음과 같은 오차제곱합을 최소화하는 것이다.

$$\sum_t^T \left(y_t - \frac{\omega_s(B)}{\delta_r(B)} B^b x_t \right)^2$$

즉, 최소제곱추정법을 적용하는 것이다. 현 단계에서는 전이함수의 분자식 $\omega_s(z)$ 와 분모식 $\delta_r(z)$ 를 임시로 추정하는 것이므로 처음 몇 개의 관찰점들에서 입력시계열데이터의 값들이 주어졌다고 가정하는 조건부최소제곱추정법을 적용하는 것이 타당하다. 이렇게 임시로 추정된 유리함수형태의 전이함수를 얻으면, 잡음시계열데이터 $\{n_t\}$ 의 적합값을 다음과 같이 정의한다.

$$\hat{n}_t = y_t - \frac{\hat{\omega}_s(B)}{\hat{\delta}_r(B)} B^b x_t$$

앞에서도 언급했듯이 전이함수모형의 잡음항이 ARMA모형을 갖는다고 가정하자. 이 잡음항의 ARMA모형은 잡음시계열데이터의 적합값들을 분석함으로써 식별할 수 있다.

임시로 추정된 유리함수형태의 전이함수와 잡음시계열데이터 $\{n_t\}$ 의 적합값들로부터 잡음시계열데이터의 ARMA모형을 식별하는 예제들을 살펴보자. 첫 번째 예제로 시계열데이터 TFM1의 잡음항의 ARMA모형을 식별하기 위해서 다음 SAS프로그램을 수행해 보자 (이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM4-17.PRGM에 수록되어 있다).

```

/*
/*
/*
/*          EXAM4-17.PRGM
/*
/*          IDENTIFYING ERROR TERM OF TRANSFER FUNCTION MODEL
/*          DATASET = TFM1
/*

```

```

LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS LINESIZE = 72;
OPTIONS PAGESIZE = 80;
DATA EXAM417;
    SET TSA4DATA.TFM1;
RUN;
PROC ARIMA DATA = EXAM417;
    I VAR=X NOPRINT;
    E P=1 Q=1 NOPRINT;
    I VAR=Y CROSS=(X) NOPRINT;
    E INPUT=(2$ (1)/(1)X) PLOT;
RUN;

```

이 SAS프로그램의 PROC ARIMA의 첫 번째 IDENTIFY스테이트먼트에서는 입력시계열데이터 X를 식별한다. 이 IDENTIFY스테이트먼트의 옵션 NOPRINT를 사용해서 IDENTIFY스테이트먼트를 수행한 결과의 출력을 억제한다. 첫 번째 ESTIMATE스테이트먼트에서는 입력시계열데이터 X를 ARMA(1, 1)모형에 적합시킨다. 이 첫 번째 IDENTIFY스테이트먼트와 첫 번째 ESTIMATE스테이트먼트가 잡음항을 식별하는데 필요한 것은 아니나, 전체적인 전이함수모형화를 하는데 필요하므로 이 장의 체제를 맞추기 위해서 이 SAS프로그램에 넣었다. 두 번째 PROC ARIMA의 IDENTIFY스테이트먼트에서는 입력시계열데이터가 X이고 출력시계열데이터가 Y임을 지정한다. 이 두 번째 IDENTIFY스테이트먼트에서도 옵션 NOPRINT를 사용해서 IDENTIFY스테이트먼트를 수행한 결과의 출력을 억제한다. ESTIMATE스테이트먼트의 옵션 INPUT =에서는 전이함수를 다음과 같이 지정한다.

$$\nu(z) = \frac{1 - \omega z}{1 - \delta z} z^2$$

이 ESTIMATE스테이트먼트는 이 전이함수를 조건부최소제곱추정하도록 한다. ESTIMATE스테이트먼트의 옵션 PLOT는 이렇게 추정된 전이함수모형의 잔차들, 즉 잡음시계열데이터의 적합값들의 표본자기상관함수, 표본역자기상관함수, 그리고 표본부분자기상관함수의 그림들과 이 잡음시계열데이터의 적합값들이 백색잡음과정에서 나왔는지를 검정하기 위한 카이제곱통계량값들인 퍼트맨토우통계량값들을 OUTPUT윈도우에 출력시킨다. 이 표본자기상관함수, 표본역자기상관함수, 그리고 표본부분자기상관함수의 그림들과 퍼트맨토우통계량값들을 이용해서 잡음시계열데이터의 ARMA모형을 식별한다.

이 SAS프로그램을 수행한 결과가 그림 4.17이다. 그림 4.17의 부분 A에서 ESTIMATE스테이트먼트를 수행해서 추정된 전이함수모형은 다음과 같음을 알 수 있다.

$$y_t = 0.0439 + \frac{0.961 + 0.392B}{1 - 0.0391B} x_{t-2} + \hat{n}_t$$

22	1.984334	0.24017
23	-2.240337	-0.27116
24	1.741478	0.21078

```

. *****
. *****
. *****
. "marks two standard errors
    
```

Inverse Autocorrelations

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
1	0.79467												*****										
2	0.46407												*****										
3	0.20155												*****										
4	0.01725																						
5	-0.05503										*												
6	-0.10401									**													
7	-0.11773								**														
8	-0.11004								**														
9	-0.13575								***														
10	-0.10343								**														
11	-0.01829																						
12	0.05593											*											
13	0.07675											**											
14	0.00827																						
15	-0.04428										*												
16	-0.03739										*												
17	-0.02224																						
18	0.01956																						
19	0.05885											*											
20	0.08387											**											
21	0.08807											**											
22	0.06105											*											
23	0.04361											*											
24	0.02008												*										

Partial Autocorrelations

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
1	-0.81395												*****										
2	-0.42868									*****													
3	-0.31311									*****													
4	-0.27547									*****													
5	-0.01700																						
6	-0.09076									**													
7	0.03470										*												
8	0.13463										***												
9	-0.10165									**													
10	-0.01230																						
11	0.09524												**										
12	0.10172												**										
13	0.17252												***										
14	-0.23177									*****													
15	-0.04428									*													
16	0.06632										*												
17	-0.08407									**													
18	0.09781										**												
19	-0.05581									*													
20	0.09045										**												
21	-0.04652									*													
22	-0.06446									*													
23	0.04563										*												
24	-0.08930									**													

Crosscorrelation Check of Residuals with Input X

To Lag	Chi Square	DF	Prob	Crosscorrelations					
5	1.49	4	0.828	0.015	-0.004	0.037	0.033	-0.082	0.059
11	5.09	10	0.885	-0.110	0.101	-0.071	0.054	-0.022	-0.013
17	9.00	16	0.913	0.032	-0.041	0.009	0.060	-0.119	0.115
23	11.26	22	0.971	-0.082	0.055	-0.017	-0.027	0.054	-0.076

Model for variable Y
 Estimated Intercept = 0.04392965
 Input Number 1 is X with a shift of 2.

The Numerator Factors are
 Factor 1: 0.9607 + 0.39226 B**(1)
 The Denominator Factors are
 Factor 1: 1 - 0.039078 B**(1)

그림 4.17 잡음시계열데이터의 식별

잡음시계열데이터의 적합값들 $\{\hat{n}_t\}$ 의 퍼트맨토우통계량값들로부터 잡음항 $\{n_t\}$ 가 백색잡음과정에서 나오지 않았음을 알 수 있다. 그림 4.17의 부분 B에는 ESTIMATE스테이트먼트의 옵션 PLOT를 사용함으로써 출력된 잡음시계열데이터의 적합값들의 표본자기상관함수, 표본역자기상관함수, 그리고 표본부분자기상관함수의 그림들이 실려 있다. 이 표본자기상관함수는 시나브로형태를 보이고 표본역자기상관함수는 시차 4 이후에서 절단형태를 보이며 표본부분자기상관함수는 시차 5 이후에서 절단형태를 보인다. 따라서 Box-Jenkins의 ARIMA모형식별방법을 사용하면 잡음시계열데이터를 AR(4)모형으로 식별할 수 있다.

앞의 예제에서 보았듯이 PROC ARIMA의 ESTIMATE스테이트먼트의 옵션 INPUT=은 전이함수모형이나 다음 장에서 다루게 될 간접시계열모형의 전이함수의 형태를 지정하는데 사용한다. 이 옵션 INPUT=의 사용양식은 다음과 같다.

INPUT = (형태 1 변수 1 형태 2 변수 2 ...)

여기서 변수 1, 변수 2, ... 들은 이 옵션 INPUT=이 사용되는 ESTIMATE스테이트먼트의 바로 앞에 사용된 IDENTIFY스테이트먼트의 옵션 CROSS=에 지정된 변수들이어야 한다. 만일 입력시계열데이터가 단변량시계열데이터라면 옵션 INPUT=에 하나의 변수만이 지정되고, 입력시계열데이터가 다변량시계열데이터라면 옵션 INPUT=에 두 개 이상의 변수들이 지정된다. 만일, 옵션 CROSS=에 어떤 변수의 차분이 지정되면, 이 차분된 변수가 전이함수의 입력시계열데이터가 된다. 옵션 INPUT=에 지정되는 형태 1, 형태 2, ...의 양식은 다음과 같다.

$bS(L_{1,1}, L_{1,2}, \dots)(L_{i,1}, L_{i,2}, \dots) / (L_{j,1}, L_{j,2}, \dots)(L_{k,1}, L_{k,2}, \dots)$

여기서 b 는 지연모수이고, 각 괄호는 멱함수인자(polynomial factor)를 뜻한다. 이 멱함수인자들 중에서 슬러시(slash, /) 이전의 멱함수인자들은 분자의 인자들이고, 슬러시 이후의 멱함수인자들은 분모의 인자들이다. 만일 멱함수인자 (L_1, L_2, \dots, L_k) 가 분자의 첫 번째 멱함수인자가 아니라면 다음과 같은 멱함수를 의미한다.

$$(1 - \alpha_{L_1} z^{L_1} - \alpha_{L_2} z^{L_2} \dots - \alpha_{L_k} z^{L_k})$$

만일, (L_1, L_2, \dots, L_k) 가 분자의 첫 번째 멱함수인자라면 이 멱함수인자는 다음과 같은 멱함수를 의미한다.

$$(\alpha_0 - \alpha_{L_1} z^{L_1} - \alpha_{L_2} z^{L_2} - \dots - \alpha_{L_k} z^{L_k})$$

만일, ESTIMATE스태이트먼트의 옵션 ALTPARM을 사용하면 분자의 첫 번째 멱함수인자 ($L1, L2, \dots, Lk$)는 다음과 같은 멱함수를 의미한다.

$$\alpha_0(1 - \alpha_{L1}z^{L1} - \alpha_{L2}z^{L2} - \dots - \alpha_{Lk}z^{Lk})$$

만일, 지연모수가 0이면 b 부분을 생략하고, 만일 전이함수의 분자가 상수이거나 또는 분모가 1이면 이에 해당하는 멱함수인자들을 생략한다. 옵션 INPUT = 의 몇 가지 예들과 이에 해당하는 전이함수들을 살펴보자.

① INPUT = (X)

$$\omega_0 X_t$$

② INPUT = ((1) X)

$$(\omega_0 - \omega_1 B) X_t$$

③ INPUT = (2\$ / (1) X)

$$\frac{\omega_0}{1 - \delta_1 B} X_{t-2}$$

④ INPUT = (3\$ (1) / (1 2) X)

$$\frac{\omega_0 - \omega_1 B}{1 - \delta_1 B - \delta_2 B^2} X_{t-3}$$

⑤ INPUT = (2\$ (1)(1 3) / (2)(1 2) X)

$$\frac{(\omega_{1,0} - \omega_{1,1} B)(1 - \omega_{2,1} B - \omega_{2,2} B^2)}{(1 - \delta_{1,1} B^2)(1 - \delta_{2,1} B - \delta_{2,2} B^2)} X_{t-2}$$

PROC ARIMA를 사용해서 잡음시계열데이터의 ARMA모형을 식별할 때 반드시 Box-Jenkins방법이나 Cleveland방법을 이용해야 하는 것은 아니다. 잡음시계열데이터의 적합값들을 새로운 SAS데이터세트에 저장한 뒤, 임의의 ARMA모형식별방법을 써서 잡음시계열데이터의 ARMA모형을 식별할 수 있다. 시계열데이터 TFM2의 잡음시계열데이터의 ARMA모형을 식별하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM4-18.PRG에 수록되어 있다).

```

/* */
/*          EXAM4-18.PRG          */
/* */
/*          IDENTIFYING THE ERROR ARMA MODEL          */
/*          OF TRANSFER FUNCTION MODEL          */
/*          DATASET = TFM2          */
/* */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS LINESIZE = 72;
OPTIONS PAGESIZE = 80;

```

```

DATA EXAM418;
  SET TSA4DATA.TFM2;
  RUN;
PROC ARIMA DATA = EXAM418;
  I VAR = Y CROSS = (X) NOPRINT;
  E INPUT = (3$ (1)/ (1, 2)X) MAXIT = 30 NOPRINT;
  F LEAD = 0 OUT = EXAM418B NOPRINT;
  RUN;
PROC ARIMA DATA = EXAM418B;
  I VAR = RESIDUAL;
  E P=0 Q=0 NOINT MAXIT = 30;
  E P=0 Q=1 NOINT MAXIT = 30;
  E P=0 Q=2 NOINT MAXIT = 30;
  E P=1 Q=0 NOINT MAXIT = 30;
  E P=1 Q=1 NOINT MAXIT = 30;
  E P=1 Q=2 NOINT MAXIT = 30;
  E P=2 Q=0 NOINT MAXIT = 30;
  E P=2 Q=1 NOINT MAXIT = 30;
  E P=2 Q=2 NOINT MAXIT = 30;
  RUN;

```

이 SAS프로그램의 첫 번째 PROC ARIMA에는 입력시계열데이터 X의 식별을 위한 IDENTIFY스테이트먼트와 입력시계열데이터 X의 추정을 위한 ESTIMATE스테이트먼트가 생략되어 있다. 따라서 입력시계열데이터의 사전백색잡음화과정이 생략된다. 결과적으로 추정된 전이함수에 의한 잔차시계열데이터가 계산되어도 이 잔차시계열데이터와 사전백색잡음화된 입력시계열데이터 사이의 표본교차상관함수가 계산되지 않는다.

이 SAS프로그램을 수행한 결과가 그림 4.18이다. 이 SAS프로그램의 첫 번째 PROC ARIMA의 IDENTIFY스테이트먼트에 NOPRINT를 할당하였으므로 이 IDENTIFY스테이트먼트를 수행한 결과물의 출력이 억제된다. 이 ESTIMATE스테이트먼트의 옵션 INPUT = 에는 전이함수를 다음과 같이 지정한다.

$$\nu(z) = \frac{\omega_0 - \omega_1 z}{1 - \delta_1 z - \delta_2 z^2} z^3$$

이 ESTIMATE스테이트먼트의 옵션 MAXIT = 에 30을 할당하였다. 이는 이 전이함수모형을 추정할 때 알고리즘의 반복과정을 최대 30번 수행할 수 있도록 한다. 이 ESTIMATE스테이트먼트에도 옵션 NOPRINT를 사용하였으므로 이 ESTIMATE스테이트먼트를 수행한 결과물의 출력이 억제된다. PROC ARIMA의 FORECAST스테이트먼트의 옵션 OUT = 에 EXAM418B를 할당했으므로, 추정된 전이함수에 의한 적합값(Forecast), 표준오차(Std), 잔차값(Residual), 신뢰도 95%의 예측구간의 상한값(U95)과 하한값

Partial Autocorrelations

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1		
1	-0.68378								*****															
2	-0.39599								*****															
3	-0.41496								*****															
4	-0.34913								*****															
5	-0.16674								***															
6	-0.25316								*****															
7	-0.01978																							
8	0.00386																							
9	-0.26270								*****															
10	-0.07261									*														
11	-0.07697									**														
12	0.09950										**													
13	0.11514										**													
14	-0.20254								*****															
15	0.01585																							
16	0.01571																							
17	-0.00364																							
18	0.06371											*												
19	0.01330																							
20	0.21275											****												
21	-0.03860									*														
22	0.02345																							
23	0.10883										**													
24	-0.12462									**														

Autocorrelation Check for White Noise

To Lag	Chi Square	DF	Prob	Autocorrelations						
6	58.08	6	0.000	-0.684	0.257	-0.161	0.130	-0.016	-0.088	
12	69.15	12	0.000	0.158	-0.166	0.028	0.128	-0.133	0.121	
18	79.41	18	0.000	-0.165	0.075	0.104	-0.146	0.104	-0.111	
24	98.55	24	0.000	0.111	-0.020	-0.122	0.227	-0.238	0.120	

(B)

Variance Estimate = 4.48627069
 Std Error Estimate = 2.11808184
 AIC = 416.534289*
 SBC = 416.534289*
 Number of Residuals = 96
 *Does not include log determinant.

Autocorrelation Check of Residuals

To Lag	Chi Square	DF	Prob	Autocorrelations						
6	58.08	6	0.000	-0.684	0.257	-0.161	0.130	-0.016	-0.088	
12	69.15	12	0.000	0.158	-0.166	0.028	0.128	-0.133	0.121	
18	79.41	18	0.000	-0.165	0.075	0.104	-0.146	0.104	-0.111	
24	98.55	24	0.000	0.111	-0.020	-0.122	0.227	-0.238	0.120	

Model for variable RESIDUAL
 No mean term in this model.

(C)

Conditional Least Squares Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag
MA1,1	0.94355	0.03397	27.78	1

Variance Estimate = 1.68976406
 Std Error Estimate = 1.29990925
 AIC = 323.791489*
 SBC = 326.355837*
 Number of Residuals = 96
 * Does not include log determinant.

Correlations of the Estimates

Parameter	MA1,1
MA1,1	1.000

Autocorrelation Check of Residuals

To Lag	Chi Square	DF	Prob	Autocorrelations					
6	22.64	5	0.000	-0.422	0.108	-0.099	0.155	0.045	-0.037
12	31.54	11	0.001	0.132	-0.135	0.065	0.185	-0.059	0.071
18	40.98	17	0.001	-0.139	0.126	0.170	-0.088	0.075	-0.063
24	54.64	23	0.000	0.121	-0.001	-0.084	0.187	-0.195	0.112

Model for variable RESIDUAL

No mean term in this model.

Moving Average Factors

Factor 1:1 -0.94355 B**(1)

(D)

Conditional Least Squares Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag
MA1,1	1.50076	0.08461	17.74	1
MA1,2	-0.57483	0.08490	-6.77	2

Variance Estimate = 1.27334333
 Std Error Estimate = 1.12842515
 AIC = 297.613085*
 SBC = 302.741782*
 Number of Residuals = 96
 * Does not include log determinant.

Correlations of the Estimates

Parameter	MA1,1	MA1,2
MA1,1	1.000	-0.953
MA1,2	-0.953	1.000

Autocorrelation Check of Residuals

To Lag	Chi Square	DF	Prob	Autocorrelations					
6	2.32	4	0.677	0.002	0.033	-0.085	0.095	0.024	-0.069
12	8.40	10	0.590	-0.008	-0.157	0.042	0.170	-0.020	-0.018
18	19.10	16	0.264	-0.106	0.162	0.228	0.002	0.033	-0.044
24	25.52	22	0.273	0.069	-0.030	-0.097	0.058	-0.154	0.091

Model for variable RESIDUAL

No mean term in this model.

Moving Average Factors

Factor 1:1 - 1.5008 B**(1)+0.57483 B**(2)

(E)

Conditional Least Squares Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag
AR1,1	-0.68872	0.07490	-9.20	1

Variance Estimate = 2.39852206
 Std Error Estimate = 1.54871626
 AIC = 357.416817*
 SBC = 359.981165*
 Number of Residuals = 96
 * Does not include log determinant.

Correlations of the Estimates

Parameter	AR1,1
AR1,1	1.000

Autocorrelation Check of Residuals

To Lag	Chi Square	DF	Prob	Autocorrelations					
6	24.65	5	0.000	-0.274	-0.380	0.048	0.139	0.027	-0.069
12	36.77	11	0.000	0.099	-0.215	0.030	0.223	-0.045	-0.059
18	51.23	17	0.000	-0.205	0.141	0.201	-0.138	-0.044	-0.043
24	64.50	23	0.000	0.143	-0.058	-0.072	0.163	-0.219	0.033

Model for variable RESIDUAL

No mean term in this model.

Autoregressive Factors

Factor 1:1 + 0.68872 B**(1)

(F)

Conditional Least Squares Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag
MA1,1	0.90523	0.04527	20.00	1
AR1,1	-0.45925	0.09543	-4.81	1

Variance Estimate = 1.37019538
 Std Error Estimate = 1.17055345
 AIC = 304.650592*
 SBC = 309.779289*
 Number of Residuals = 96
 * Does not include log determinant.

Correlations of the Estimates

Parameter	MA1,1	AR1,1
MA1,1	1.000	0.268
AR1,1	0.268	1.000

Autocorrelation Check of Residuals

To Lag	Chi Square	DF	Prob	Autocorrelations					
6	8.15	4	0.086	-0.070	-0.201	-0.061	0.155	0.085	-0.018
12	18.02	10	0.055	0.042	-0.147	0.077	0.240	0.007	-0.059
18	31.10	16	0.013	-0.147	0.174	0.243	-0.040	-0.016	-0.027
24	38.51	22	0.016	0.118	-0.009	-0.059	0.093	-0.163	0.075

Model for variable RESIDUAL

No mean term in this model.

Autoregressive Factors

Factor 1:1 + 0.45925 B**(1)

Moving Average Factors

Factor 1:1 - 0.90523 B**(1)

(G)

Conditional Least Squares Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag
MA1,1	1.50610	0.14799	10.18	1
MA1,2	-0.57982	0.14396	-4.03	2
AR1,1	0.0073735	0.18150	0.04	1

Variance Estimate = 1.28702288
 Std Error Estimate = 1.13447051
 AIC = 299.612167*
 SBC = 307.305212*
 Number of Residuals = 96
 * Does not include log determinant.

Correlations of the Estimates

Parameter	MA1,1	MA1,2	AR1,1
MA1,1	1.000	-0.984	0.820
MA1,2	-0.984	1.000	-0.807
AR1,1	0.820	-0.807	1.000

Autocorrelation Check of Residuals

To Lag	Chi Square	DF	Prob	Autocorrelations					
6	2.34	3	0.505	-0.000	0.036	-0.084	0.095	0.023	-0.070
12	8.38	9	0.496	-0.008	-0.157	0.041	0.169	-0.021	-0.017
18	19.02	15	0.213	-0.106	0.162	0.227	0.002	0.034	-0.044
24	25.48	21	0.227	0.069	-0.031	-0.097	0.059	-0.154	0.091

Model for variable RESIDUAL

No mean term in this model.

Autoregressive Factors

Factor 1:1 - 0.0073735 B**(1)

Moving Average Factors

Factor 1:1 - 1.5061 B**(1) + 0.57982 B**(2)

(H)

Conditional Least Squares Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag
AR1,1	-0.96235	0.09471	-10.16	1
AR1,2	-0.40057	0.09492	-4.22	2

Variance Estimate = 2.0379282
 Std Error Estimate = 1.42756023
 AIC = 342.760707*
 SBC = 347.889403*
 Number of Residuals = 96
 * Does not include log determinant.

Correlations of the Estimates

Parameter	AR1,1	AR1,2
AR1,1	1.000	0.685
AR1,2	0.685	1.000

Autocorrelation Check of Residuals

To Lag	Chi Square	DF	Prob	Autocorrelations					
6	25.24	4	0.000	-0.169	-0.339	-0.214	0.238	0.070	-0.019
12	38.60	10	0.000	-0.035	-0.179	0.046	0.247	0.007	-0.161
18	54.85	16	0.000	-0.206	0.160	0.230	-0.088	-0.104	-0.031
24	61.63	22	0.000	0.108	-0.003	-0.055	0.059	-0.182	0.046

Model for variable RESIDUAL

No mean term in this model.

Autoregressive Factors

Factor 1:1 + 0.96235 B**(1) + 0.40057 B*(2)

(I)

Conditional Least Squares Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag
MA1,1	0.88584	0.05342	16.58	1
AR1,1	-0.54785	0.11200	-4.89	1
AR1,2	-0.17106	0.11149	-1.53	2

Variance Estimate = 1.34867712
 Std Error Estimate = 1.16132559
 AIC = 304.104247*
 SBC = 311.797292*
 Number of Residuals = 96

* Does not include log determinant.

Correlations of the Estimates

Parameter	MA1,1	AR1,1	AR1,2
MA1,1	1.000	0.407	0.392
AR1,1	0.407	1.000	0.550
AR1,2	0.392	0.550	1.000

Autocorrelation Check of Residuals

To Lag	Chi Square	DF	Prob	Autocorrelations					
6	6.37	3	0.095	-0.037	-0.102	-0.146	0.146	0.090	-0.017
12	15.26	9	0.084	0.006	-0.135	0.083	0.229	0.007	-0.063
18	27.93	15	0.022	-0.138	0.172	0.245	-0.014	-0.007	-0.038
24	34.26	21	0.034	0.097	0.003	-0.068	0.061	-0.153	0.093

Model for variable RESIDUAL

No mean term in this model.

Autoregressive Factors

Factor 1:1 + 0.54785 B**(1) + 0.17106 B**(2)

Moving Average Factors

Factor 1:1 - 0.88584 B**(1)

(J)

Conditional Least Squares Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag
MA1,1	1.88769	0.05556	33.98	1
MA1,2	-0.92596	0.05246	-17.65	2
AR1,1	0.38755	0.11731	3.30	1
AR1,2	0.27538	0.11292	2.44	2

Variance Estimate = 1.28137628
 Std Error Estimate = 1.13197892
 AIC = 300.152209*
 SBC = 310.409601*
 Number of Residuals = 96

* Does not include log determinant.

Correlations of the Estimates

Parameter	MA1,1	MA1,2	AR1,1	AR1,2
MA1,1	1.000	-0.986	0.505	0.343
MA1,2	-0.986	1.000	-0.483	-0.289
AR1,1	0.505	-0.483	1.000	-0.208
AR1,2	0.343	-0.289	-0.208	1.000

Autocorrelation Check of Residuals

To Lag	Chi Square	DF	Prob	Autocorrelations					
6	3.40	2	0.183	-0.008	-0.044	-0.018	0.155	0.065	-0.051
12	10.63	8	0.223	-0.004	-0.186	0.006	0.142	-0.063	-0.089
18	20.21	14	0.124	-0.175	0.112	0.177	-0.067	-0.026	-0.059
24	26.55	20	0.148	0.078	-0.027	-0.079	0.076	-0.162	0.069

```

Model for variable RESIDUAL

No mean term in this model.

Autoregressive Factors
Factor 1:1 - 0.38755 B**(1) + 0.27538 B**(2)

Moving Average Factors
Factor 1:1 - 1.8877 B**(1) + 0.92596 B**(2)

```

그림 4.18 AIC방법과 BIC방법에 의한 잔차항의 ARMA모형식별

(L95) 등이 SAS데이터세트 EXAM418B에 수록된다. 옵션 OUT = 에 대한 자세한 내용은 [ETS6.1] (pp. 113~114)을 참조하라. 이 FORECAST스테이트먼트에도 옵션 NOPRINT 를 사용하였으므로 이 FORECAST스테이트먼트를 수행한 결과물이 OUTPUT윈도우에 출력되는 것이 억제된다.

이 SAS프로그램의 두 번째 PROC ARIMA에서는 앞의 FORECAST스테이트먼트의 옵션 OUT = 에 지정한 SAS데이터세트 EXAM418B를 분석한다. IDENTIFY스테이트먼트의 옵션 VAR = 에 변수 RESIDUAL을 지정하였으므로, 이 IDENTIFY스테이트먼트가 수행되면 조건부최소제곱추정에 의한 전이함수모형의 잔차값들을 나타내는 변수 RESIDUAL, 즉 잡음시계열데이터의 적합값들의 표본자기상관함수, 표본역자기상관함수, 그리고 표본부분자기상관함수의 그림들과 이 잡음시계열데이터의 적합값들이 백색잡음과정에서 나왔는가를 검정하기 위한 카이제곱통계량값들인 퍼트맨토우통계량값들이 OUTPUT윈도우에 출력된다. 그림 4.18의 부분 A에서 볼 수 있듯이 잡음시계열데이터의 적합값들의 표본자기상관함수, 표본역자기상관함수, 그리고 표본부분자기상관함수가 모두 시나브로형태를 보인다. 따라서 Box-Jenkins식별방법이나 Cleveland식별방법을 사용해서는 잡음시계열데이터에 타당한 ARMA모형을 식별할 수 없다. 따라서 AIC방법과 BIC방법을 사용해서 이 잡음시계열데이터에 타당한 ARMA모형을 식별하기 위하여 여러 ESTIMATE스테이트먼트들을 사용한다. 그림 4.18의 부분 B는 잡음시계열데이터를 ARMA(0, 0)모형에 적합시킨 결과이고, 부분 C는 잡음시계열데이터를 MA(1)모형에 적합시킨 결과이며, 부분 D는 잡음시계열데이터를 MA(2)모형에 적합시킨 결과이다. 부분 E는 잡음시계열데이터를 AR(1)모형에 적합시킨 결과이고, 부분 F는 잡음시계열데이터를 ARMA(1, 1)모형에 적합시킨 결과이며, 부분 G는 잡음시계열데이터를 ARMA(1, 2)모형에 적합시킨 결과이다. 또한 부분 H는 잡음시계열데이터를 AR(2)모형에 적합시킨 결과이고, 부분 I는 잡음시계열데이터를 ARMA(2, 1)모형에 적합시킨 결과이며, 부분 J는 잡음시계열데이터를 ARMA(2, 2)모형에 적합시킨 결과이다. 이 ARMA모형들 중에서 AIC값과 BIC값을 가장 작게 하는 모형은 MA(2)모형이다.

4.3.4 전이함수모형식별의 유의점들

잡음시계열데이터의 ARMA모형을 다음과 같이 표현하자.

$$\phi(B)n_t = \theta(B)v_t$$

여기서 $\{v_t\}$ 는 백색잡음과정이다. 따라서 지금까지의 분석을 거쳐서 식별한 유리함수형태의 전이함수모형은 다음과 같다.

$$y_t = \frac{\omega_s(B)}{\delta_r(B)} x_{t-b} + \frac{\theta(B)}{\phi(B)} v_t$$

이 유리함수형태의 전이함수모형을 식별하는 데 유의할 점들은 다음과 같다. 첫째로 출력시계열데이터 $\{y_t\}$ 와 입력시계열데이터 $\{x_t\}$, 그리고 잡음시계열데이터 $\{n_t\}$ 는 모두 정상적이라고 가정한다. 따라서 비정상시계열데이터들은 변환이나 차분 등을 이용해서 정상시계열데이터들로 만든 다음에 전이함수모형을 식별해야 한다. 둘째로 전이함수 $\nu(z)$ 를 식별하는 과정에서 입력시계열데이터를 사전백색화한다. 이 사전백색화모형은 출력시계열데이터를 여과하는 데 사용된다. 그러나 항상 사전백색화작업이 필요한 것은 아니다. 만일, 입력시계열데이터가 백색잡음과정이면 사전백색화작업이 필요없다. 이 사전백색화작업은 인과적인 관계를 갖는 입력시계열데이터와 출력시계열데이터의 전이함수모형을 구축하는 데 널리 사용된다. 그러나 만일 $\{y_t\}$ 가 $\{x_t\}$ 에 의해서 영향을 받고 또한 $\{x_t\}$ 가 $\{y_t\}$ 에 의해 영향을 받는 피드백(feedback) 현상이 존재하는 비작인적 시스템을 모형화할 때에는 교차상관함수를 조사하기 전에 두 시계열데이터 모두를 사전백색화해야 한다. 이것을 이중사전백색화(double prewhitening)라고 한다. 그러나 비작인적인 시계열데이터들을 모형화할 때는 제 8장에서 다룰 상태공간모형을 사용하는 것이 좋다.

4.4 전이함수모형의 추정

유리함수형태의 전이함수모형을 다음과 같이 식별했다고 하자.

$$y_t = \frac{\omega_s(B)}{\delta_r(B)} x_{t-b} + \frac{\theta(B)}{\phi(B)} v_t$$

이 식은 다음과 같이 쓸 수 있다.

$$\delta_r(B)\phi(B)y_t = \phi(B)\omega_s(B)x_{t-b} + \delta_r(B)\theta(B)v_t$$

이 식을 다시 쓰면 다음과 같다.

$$c(B)y_t = d(B)x_{t-b} + e(B)v_t$$

$$\text{단, } c(z) = \delta_r(z)\phi(z)$$

$$= (1 - \delta_1 z - \dots - \delta_r z^r)(1 - \phi_1 z - \dots - \phi_p z^p)$$

$$= 1 - c_1 z - c_2 z^2 - \dots - c_{p+r} z^{p+r}$$

$$d(z) = \omega_s(z)\phi(z)$$

$$= (\omega_0 - \omega_1 z - \dots - \omega_s z^s)(1 - \phi_1 z - \dots - \phi_p z^p)$$

$$= d_0 - d_1 z - d_2 z^2 - \dots - d_{p+s} z^{p+s}$$

$$e(z) = \delta_r(z)\theta(z)$$

$$= (1 - \delta_1 z - \dots - \delta_r z^r)(1 - \theta_1 z - \dots - \theta_q z^q)$$

$$= 1 - e_1 z - e_2 z^2 - \dots - e_{r+q} z^{r+q}$$

따라서 백색잡음항 $\{v_t\}$ 는 다음과 같이 나타날 수 있다.

$$v_t = y_t - c_1 y_{t-1} - \dots - c_{p+r} y_{t-p-r} - d_0 x_{t-b} + d_1 x_{t-b-1} \\ + \dots + d_{p+s} x_{t-b-s} + e_1 v_{t-1} + \dots + e_{r+q} v_{t-r-q}$$

여기서 $\{c_i\}$, $\{d_j\}$, 그리고 $\{e_k\}$ 는 $\{\delta_i\}$, $\{\omega_j\}$, $\{\phi_k\}$, 그리고 $\{\theta_l\}$ 의 함수들이다. 모수벡터 δ , ω , ϕ , 그리고 θ 를 다음과 같이 정의하자.

$$\delta = (\delta_1, \dots, \delta_r)'$$

$$\omega = (\omega_0, \omega_1, \dots, \omega_s)'$$

$$\phi = (\phi_1, \dots, \phi_p)'$$

$$\theta = (\theta_1, \dots, \theta_q)'$$

또한, σ_v^2 는 백색잡음과정 $\{v_t\}$ 의 분산이다. 만일 $\{v_t\}$ 가 정규분포를 갖는 백색잡음과정이라면 초기값들이 주어졌다는 조건하에서 조건부우도함수는 다음과 같다.

$$L(\delta, \omega, \phi, \theta, \sigma_v^2 | x, y, x_0, y_0, v_0) = (2\pi\sigma_v^2)^{-T/2} \exp\left\{-\sum_{t=1}^T v_t^2 / (2\sigma_v^2)\right\}$$

여기서 x_0 , y_0 , 그리고 v_0 는 ARMA모형을 조건부최소제곱추정할 때 필요한 초기값들과 마찬가지로 백색잡음항 $\{v_t\}$ 를 계산하기 위해서 적절히 선택된 초기벡터들이다. 모수들을 추정하기 위해서는 알지 못하는 $\{v_t\}$ 의 초기항들 대신에 이 항들의 기대값인 0으로 놓고 다음 식을 최소화시킴으로써 이 모수들의 조건부최소제곱추정값들을 얻을 수 있다.

$$S(\delta, \omega, \phi, \theta | x_0, y_0, v_0 = 0) = \sum_{t=t_0}^T v_t^2$$

$$\text{단, } t_0 = \text{Max}\{p+r+1, b+p+s+1\}$$

만일, $\{v_t\}$ 가 정규분포를 갖는 백색잡음과정이라면 비조건부우도함수는 다음과 같다.

$$L(\delta, \omega, \phi, \theta, \sigma_v^2 | x, y) = (2\pi\sigma_v^2)^{-T/2} J(\delta, \omega, \phi, \theta) \exp \left\{ -\frac{1}{2\sigma_v^2} \sum_{t=-\infty}^T v_t^2 \right\}$$

여기서 $J(\delta, \omega, \phi, \theta)$ 는 Jacobian에 의존하는 함수이며, t 가 1보다 작은 경우의 v_t 값은 주어진 x_1, x_2, \dots, x_T 와 y_1, y_2, \dots, y_T 를 바탕으로 해서 후향으로 예측한 값들이다. 이 예측값들을 후향예측값들이라 부른다. 이 비조건부우도함수를 최대화하는 추정량을 최우추정량이라 부른다.

이 관찰점들의 개수가 큰 경우에는 $J(\delta, \omega, \phi, \theta)$ 가 비조건부우도함수에 큰 영향을 미치지 못한다. 따라서 이 비조건부우도함수를 최대화하는 대신에 다음 비조건부제곱합 $S_*(\delta, \omega, \phi, \theta)$ 를 최소화하기도 한다.

$$S_*(\delta, \omega, \phi, \theta) = \sum_{t=-\infty}^T v_t^2$$

이 식을 최소화하는 추정량을 비조건부최소제곱추정량이라고 한다.

이 추정법들에 대한 자세한 내용은 [SAS4TSA2]의 제II.1절과 [ETS6.1](pp. 116~118)을 참조하라. PROC ARIMA의 ESTIMATE스테이트먼트의 옵션 METHOD=에 CLS, ML, 그리고 ULS를 할당하면 각각 조건부최소제곱추정법, 최우추정법, 그리고 비조건부최소제곱추정법이 적용된다. 이 옵션의 디폴트는 CLS이다. 옵션 METHOD=에 대한 자세한 내용은 [SAS4TSA2](p. 511)를 참조하라. PROC ARIMA에서는 비선형최적화알고리즘을 사용하기 때문에, 만일 주어진 시계열데이터가 식별한 모형과 잘 부합되지 않으면 문제가 있는 결과(problematic result)를 발생시킬 수도 있다. 따라서 추정의 결과를 주의 깊게 검토해야 한다. ESTIMATE스테이트먼트의 옵션 GRID를 사용해서 추정의 결과를 검토하는 것도 한 가지 방법일 것이다. PROC ARIMA를 사용해서 ARIMA모형을 추정할 때 발생할 수 있는 계산상의 문제들은 다음과 같다.

- ① MA모수들의 예비추정값(preliminary MA estimate)들이 수렴하지 않을 수도 있다. 이러한 경우에는 ESTIMATE스테이트먼트의 옵션 MA=을 사용해서 MA모수들의 주관적인 예비추정값들을 공급한다.
- ② 모수들의 추정값들이 불안정적(unstable)인 시계열모형을 이룰 수도 있다. 더구나 이렇게 추정된 시계열모형에 의한 예측값들의 절대값이 폭발적으로 커질 수도 있다.
- ③ Jacobian행렬이 특이행렬일 수 있다. 일반적으로 모든 모수들이 식별 가능하지 않은 경우에 이러한 현상이 일어난다. 이 경우에는 몇 개의 모수들을 제거하거나 좀 더 많은 관찰점들을 사용해서 문제를 해결할 수도 있다.
- ④ 추정의 반복과정(iterative process)이 수렴하지 않을 수도 있다. PROC ARIMA의 추정과정은 옵션 MAXIT=에 지정한 숫자만큼 반복을 한 뒤 멈춘다. 만일 추정의 각 단계가 오차제곱합(SSE)을 감소시키지 않으면, 작은 오차제곱합을 얻을 때까지 Marquadt모수가 10배씩 증가된다.

- ⑤ 만일, 옵션 METHOD=에 디폴트인 CLS를 지정해서 조건부최소제곱추정법을 사용하면, 추정값들이 수렴하는 값들은 비조건부최소제곱추정값들이 아닐 수도 있다. 이 경우에는 조건부제곱합이 최소값(global minimum)을 갖는 것이 아니라 극소값(local minimum)을 갖는다. 제곱합의 표면(sum-of-squares surface)이 평활(smooth)하지 않거나, 추정값들이 작인성조건을 만족하는 구역, 또는 가역성조건을 만족하는 구역 밖에 있기 때문에 이러한 현상이 일어날 수도 있다.
- ⑥ 만일, 시계열데이터를 차분하고 MA항이 있는 모형을 사용하면, 추정값들은 가역성조건이 만족되는 구역내에서 수렴하려고 한다. 이 경우에는 표준오차의 추정값들이 정확하지 않을 수도 있다.

전이함수모형을 추정하는 첫 번째 예제로써 시계열데이터 TFM1을 조건부최소제곱추정법으로 추정하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM4-19.PRG에 수록되어 있다).

```

/* */
/* */
/* */
/* */
/* */
/* */
/* */
/* */
/* */
/* */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS LINESIZE=72;
OPTIONS PAGESIZE=80;
DATA EXAM419;
    SET TSA4DATA.TFM1;
RUN;
PROC ARIMA DATA=EXAM419;
    I VAR=X NOPRINT;
    E P=1 Q=1 NOPRINT;
    I VAR=Y CROSS=(X) NOPRINT;
    E P=4 INPUT=(2$ (1)/(1)X);
RUN;

```

이 SAS프로그램을 수행한 결과가 그림 4.19이다. 그림 4.19의 부분 A와 부분 D에서 볼 수 있듯이 조건부최소제곱추정법에 의해 추정된 모형은 다음과 같다.

$$y_t = -0.0474 + \frac{0.8675 + 0.7606B}{1 + 0.1784B} x_{t-2} + \frac{1}{1 + 1.4800B + 1.1822B^2 + 0.8166B^3 + 0.3401B^4} v_t$$

SAS 9:46 Friday, October 1, 1993 15

ARIMA Procedure

(A)

Conditional Least Squares Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag	Variable	Shift
MU	-0.00985	0.02869	-0.34	0	Y	0
AR1,1	-1.47997	0.08987	16.47	1	Y	0
AR1,2	-1.18218	0.14844	-7.96	2	Y	0
AR1,3	-0.81655	0.14854	-5.50	3	Y	0
AR1,4	-0.34005	0.09051	-3.76	4	Y	0
NUM1	0.86750	0.03765	23.04	0	X	2
NUM1,1	-0.76059	0.13689	-5.56	1	X	2
DEN1,1	-0.17837	0.07936	-2.25	1	X	2

Constant Estimate = -0.0474552
 Variance Estimate = 1.74963213
 Std Error Estimate = 1.32273661
 AIC = 408.595247*
 SBC = 430.760724*
 Number of Residuals = 118
 * Does not include log determinant.

Correlations of the Estimates

Variable	Parameter	Y	Y	Y	Y	Y
		MU	AR1,1	AR1,2	AR1,3	AR1,4
Y	MU	1.000	0.011	0.011	0.013	0.007
Y	AR1,1	0.011	1.000	0.825	0.537	0.358
Y	AR1,2	0.011	0.825	1.000	0.826	0.537
Y	AR1,3	0.013	0.537	0.826	1.000	0.823
Y	AR1,4	0.007	0.358	0.537	0.823	1.000
X	NUM1	0.220	0.048	0.058	0.062	0.041
X	NUM1,1	0.218	0.046	0.051	0.049	0.028
X	DEN1,1	0.240	0.045	0.049	0.046	0.025

Variable	Parameter	X	X	X
		NUM1	NUM1,1	DEN1,1
X	MU	0.220	0.218	0.240
X	AR1,1	0.048	0.046	0.045
X	AR1,2	0.058	0.051	0.049
X	AR1,3	0.062	0.049	0.046
X	AR1,4	0.041	0.028	0.025
X	NUM1	1.000	0.888	0.817
X	NUM1,1	0.888	1.000	0.988
X	DEN1,1	0.817	0.988	1.000

(B)

Autocorrelation Check of Residuals

To Lag	Chi Square	DF	Prob	Autocorrelations					
6	2.43	2	0.297	-0.024	-0.072	-0.087	-0.051	-0.050	0.036
12	8.09	8	0.425	0.051	-0.090	0.051	0.145	-0.073	-0.061
18	18.25	14	0.196	-0.167	0.033	0.190	-0.065	0.036	-0.060
24	26.50	20	0.150	-0.036	-0.056	-0.091	0.055	-0.177	0.093

(C)

Crosscorrelation Check of Residuals with Input X

To Lag	Chi Square	DF	Prob	Crosscorrelations					
5	5.34	4	0.255	0.071	-0.012	-0.045	0.157	0.060	0.103
11	8.61	10	0.569	-0.091	-0.100	-0.073	0.006	0.067	-0.005
17	12.84	16	0.684	-0.015	-0.078	-0.068	0.150	-0.054	0.009
23	14.19	22	0.895	-0.005	0.004	0.092	-0.041	-0.007	-0.038

```

(D)

Model for variable Y

Estimated Intercept = -0.009848

Autoregressive Factors
Factor 1: 1+1.48 B**(1)+1.1822 B**(2)+0.81655 B**(3)
          +0.34005 B**(4)

Input Number 1 is X with a shift of 2.

The Numerator Factors are
Factor 1: 0.8675+0.76059 B**(1)

The Denominator Factors are
Factor 1: 1+0.17857 B**(1)
    
```

그림 4.19 조건부최소제곱추정법을 사용한 전이함수모형의 추정

전이함수모형을 추정하는 두 번째 예제로써 시계열데이터 TFM2를 최우추정법으로 추정하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM4-20.PRG에 수록되어 있다).

```

/ * * /
/ * * EXAM4-20.PRG * /
/ * * * /
/ * * TRANSFER FUNCTION MODEL ESTIMATION * /
/ * * USING MLE * /
/ * * DATASET = TFM2 * /
/ * * * /

LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS LINESIZE=72;
OPTIONS PAGESIZE=80;
DATA EXAM420;
    SET TSA4DATA.TFM2;
RUN;

PROC ARIMA DATA=EXAM420;
    I VAR=X NOPRINT;
    E P=1 Q=1 NOPRINT;
    I VAR=Y CROSS=(X) NOPRINT;
    E P=0 Q=2 INPUT=(3$ (1)/(1 2)X) NOINT
    ALTPARM METHOD=ML PRINTALL MAXIT=30;
RUN;
    
```

이 SAS프로그램을 수행한 결과가 그림 4.20이다. 그림 4.20의 부분 A에는 잡음항모형의 예비추정결과가 수록되어 있다. 부분 B에는 전이함수모형을 조건부최소제곱추정한 결과가

SAS 9:46 Friday, October 1, 1993 17

ARIMA Procedure

(A)

Preliminary Estimation

Initial Moving Average Estimates

1 -0.77566
 2 -0.51900
 White Noise Variance Est = 33.714

(B)

Conditional Least Squares Estimation

Iteration	SSE	MA1,1 DEN1,2	MA1,2 Lambda	SCALE1	NUM1,1	DEN1,1
0	1997.6	-0.7757 0.1000	-0.5190 0.00001	0.7793	0.1000	0.1000
1	848.11002	-0.4511 0.02218	-0.3307 0.001	0.8416	-1.6788	-0.5020
2	384.19267	0.1240 0.09160	-0.05218 0.0001	0.9340	-1.2475	-0.2504
3	299.35552	0.3273 0.04516	0.01549 0.0001	0.9229	-1.2000	-0.1959
4	232.71974	0.5529 0.01312	0.07683 0.0001	0.9141	-1.1670	-0.1564

(C)

Maximum Likelihood Estimation

Iter	Loglike	MA1,1 DEN1,2	MA1,2 Lambda	SCALE1	NUM1,1	DEN1,1
0	-178.94716	0.5529 0.01312	0.07683 0.00001	0.9141	-1.1670	-0.1564
1	-167.63212	0.7905 -0.00492	0.1117 0.00001	0.9073	-1.1497	-0.1336
2	-160.25005	0.9613 -0.00699	0.03552 0.00001	0.9028	-1.1520	-0.1311
3	-160.15149	0.9647 -0.00696	0.03368 0.00001	0.9028	-1.1522	-0.1312
4	-160.12673	0.9660 -0.00696	0.03323 0.00001	0.9027	-1.1522	-0.1312
5	-160.12061	0.9664 -0.00695	0.03312 0.00001	0.9027	-1.1522	-0.1312
6	-160.11908	0.9666 -0.00695	0.03310 0.00001	0.9027	-1.1522	-0.1312
7	-160.11756	0.9668 -0.00695	0.03307 0.00001	0.9027	-1.1522	-0.1312
8	-160.11756	0.9668 -0.00695	0.03307 0.00001	0.9027	-1.1522	-0.1312

(D)

WARNING: The model defined by the new estimates is unstable.
 The iteration process has been terminated.
 WARNING: Estimates may not have converged.

ARIMA Estimation Optimization Summary

Estimation Method:	Maximum Likelihood
Parameters Estimated:	6
Termination Criteria:	Maximum Relative Change in Estimates
Iteration Stopping Value:	0.001
Criteria Value:	9.55261786
Maximum Absolute Value of Gradient:	439.710033
R-Square (Relative Change in Regression SSE) from Last Iteration Step:	0.26344536
Objective Function:	Log Gaussian Likelihood

Objective Function Value: -160.11756
 Marquardt's Lambda Coefficient: 0.00001
 Numerical Derivative Perturbation Delta: 0.001
 Iterations: 8
 Warning Message: Estimates may not have converged.

(E)

Maximum Likelihood Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag	Variable	Shift
MA1,1	0.96681	12.61129	0.08	1	Y	0
MA1,2	0.03307	0.45398	0.07	2	Y	0
SCALE1	0.90273	0.04633	19.48	0	X	3
NUM1,1	-1.15225	0.25854	-4.46	1	X	3
DEN1,1	-0.13122	0.13693	-0.96	1	X	3
DEN1,2	-0.0069534	0.04215	-0.16	2	X	3

Variance Estimate = 1.67458095
 Std Error Estimate = 1.29405601
 AIC = 332.23511
 SBC = 347.621199
 Number of Residuals = 96

Correlations of the Estimates

Variable	Parameter	Y	Y	X	X	X
Y	MA1,1	1.000	0.972	-0.010	0.051	0.088
Y	MA1,2	0.972	1.000	0.000	0.060	0.096
X	SCALE1	-0.010	0.000	1.000	0.754	0.375
X	NUM1,1	0.051	0.060	0.754	1.000	0.881
X	DEN1,1	0.088	0.096	0.375	0.881	1.000
X	DEN1,2	-0.140	-0.147	-0.133	-0.651	-0.904

Variable	Parameter	X
X	MA1,1	-0.140
X	MA1,2	-0.147
X	SCALE1	-0.133
X	NUM1,1	-0.651
X	DEN1,1	-0.904
X	DEN1,2	1.000

(F)

Autocorrelation Check of Residuals

To Lag	Chi Square	DF	Prob	Autocorrelations					
6	30.06	4	0.000	-0.501	0.096	-0.144	0.130	0.016	-0.065
12	38.35	10	0.000	0.113	-0.168	0.037	0.154	-0.091	0.051
18	48.40	16	0.000	-0.176	0.102	0.140	-0.111	0.061	-0.098
24	62.16	22	0.000	0.109	-0.023	-0.095	0.177	-0.215	0.094

(G)

Crosscorrelation Check of Residuals with Input X

To Lag	Chi Square	DF	Prob	Crosscorrelations					
5	4.73	2	0.094	0.023	-0.024	0.062	-0.115	0.114	-0.140
11	7.45	8	0.489	0.099	-0.072	0.098	0.036	0.007	0.057
17	10.35	14	0.736	-0.016	-0.111	0.093	-0.060	0.066	-0.045
23	14.10	20	0.826	-0.043	0.069	-0.008	0.072	-0.103	0.134

(H)

Model for variable Y

No mean term in this model.

```

Moving Average Factors
Factor 1: 1-0.96681 B**(1)-0.033075 B**(2)

Input Number 1 is X with a shift of 3.
Overall Regression Factor = 0.902734

The Numerator Factors are
Factor 1: 1 + 1.1522 B**(1)

The Denominator Factors are
Factor 1: 1 + 0.13122 B**(1) + 0.0069534 B**(2)

```

그림 4.20 최우추정법을 사용한 전이함수모형의 추정

수록되어 있다. 이 조건부최소제곱추정값들이 최우추정, 또는 비조건부최소제곱추정의 초기값들로 사용된다. 이 SAS프로그램의 ESTIMATE스테이트먼트에서는 옵션 METHOD=에 ML을 할당했으므로 최종적인 추정방법으로 최우추정법이 사용된다. 부분 C에는 전이함수모형을 최우추정한 결과의 각 단계가 수록되어 있다. ESTIMATE스테이트먼트의 옵션 MAXIT=에는 최우추정을 위한 알고리즘의 반복횟수의 최대값을 지정한다. 이 옵션의 디폴트는 15이다. 이렇게 각 반복단계의 결과가 출력된 이유는 ESTIMATE스테이트먼트의 옵션 PRINTALL을 사용했기 때문이다. 부분 D에서 이 최우추정값들을 추정하는 알고리즘이 수렴하지 않을 수도 있음을 알려 준다. ARMA모형을 추정하기 위한 일반적인 알고리즘의 수렴성은 AR특성식이 작인성조건을 만족하고 MA특성식이 가역성조건을 만족하는 것을 전제로 한다. 여기서 이 알고리즘이 잘 수렴하지 않는 이유는 부분 E에서 알 수 있듯이 잡음시계열데이터의 MA특성방정식의 한 근이 단위원 가까이에 있으며 전이함수의 분자의 방정식의 근이 단위원 안에 존재하기 때문이다. 그림 4.20의 부분 F와 부분 H에서 볼 수 있듯이 최우추정법에 의해 추정된 모형은 다음과 같다.

$$y_t = 0.9027 \frac{1+1.1522B}{1+0.1312B+0.0070B^2} x_{t-3} + (1-0.9668B-0.0331B^2)v_t$$

ESTIMATE스테이트먼트에서는 옵션 ALTPARM이 사용되었음을 유의하라.

4.5 전이함수모형의 검진

통계분석에서는 모형을 식별하고 모수를 추정해서 얻은 추정모형을 예측, 제어 등 여러 목적에 이용하기에 앞서서 이 추정모형이 주어진 데이터와 잘 부합되는가를 조사해 보아야 한다. 이러한 과정을 모형의 검진(diagnostic checking)이라고 한다. 유리함수형태의 전이함수모형에서는 잡음시계열데이터 $\{n_t\}$ 가 입력시계열데이터 $\{x_t\}$ 와 독립적이라고 가정한다. 따라서 잡음시계열데이터 $\{n_t\}$ 와 사전백색잡음화된 입력시계열데이터 $\{z_t\}$ 는 서로 독립적이다. 또한, 잡음시계열모형의 오차항 $\{v_t\}$ 는 백색잡음과정이라고 가정한다. 선택된 전

이함수모형이 주어진 시계열데이터와 잘 부합되는지를 검진하기 위해서는 이 가정들이 만족되는지 여부를 조사해야 한다. 따라서 전이함수모형의 검진은 다음과 같은 2단계의 과정으로 나눌 수 있다. 첫 번째 단계로는 사전백색잡음화된 입력시계열데이터 $\{z_t\}$ 와 잡음시계열모형의 오차항 $\{v_t\}$ 의 독립성을 조사하고, 두 번째 단계로 잡음시계열모형의 오차항 $\{v_t\}$ 의 백색잡음성을 조사한다. 각 단계를 자세히 설명하면 다음과 같다.

1) 교차상관관계의 검진

잡음시계열모형의 오차항의 적합값들 $\{\hat{v}_t\}$ 와 사전백색잡음화된 입력시계열데이터 $\{z_t\}$ 가 서로 독립적인지 여부를 검진한다. 선택된 전이함수모형이 주어진 시계열데이터와 잘 부합된다면, $\{\hat{v}_t\}$ 와 $\{z_t\}$ 사이의 표본교차상관계수들의 절대값들은 모든 시차에서 표준오차의 두 배보다 작아야 한다. 이를 검정하기 위해서 다음과 같은 퍼트맨토우통계량을 사용한다.

$$Q_0 = m(m+2) \sum_{j=0}^K \frac{1}{(m-j)} \hat{\rho}_m^2(j)$$

여기서 m 은 계산된 잔차들의 개수인 $(T-t_0+1)$ 이다. 전이함수의 모수들 $\{\delta_i\}$ 와 $\{\omega_j\}$ 의 개수를 M 으로 나타내면, 잡음시계열모형의 잔차들 $\{\hat{v}_t\}$ 와 사전백색잡음화된 입력시계열데이터 $\{z_t\}$ 가 서로 독립적이라는 귀무가설하에서 이 퍼트맨토우통계량은 점근적으로 자유도 $(K+1-M)$ 인 카이제곱분포를 따른다. 이 퍼트맨토우통계량의 자유도는 잡음시계열데이터의 ARMA모형의 지수들 p 와 q 에 의존하지 않는다.

2) 자기상관관계의 검진

이 단계에서는 선택된 잡음시계열데이터의 ARMA모형이 주어진 시계열데이터와 잘 부합되는지를 검진한다. 선택된 ARMA모형이 주어진 시계열데이터와 잘 부합된다면, 잡음시계열모형의 잔차들 $\{\hat{v}_t\}$ 의 표본자기상관함수는 시차 1 이후에서 절단형태를 보여야 한다. 이를 검정하기 위해서 다음과 같은 퍼트맨토우통계량을 사용한다.

$$Q_1 = m(m+2) \sum_{j=1}^K \frac{1}{(m-j)} \hat{\rho}_m^2(j)$$

여기서 m 은 계산된 잔차들의 개수인 $(T-t_0+1)$ 이다. 만일, 잡음시계열데이터의 ARMA모형의 차수들을 p 와 q 라고 하면, 잡음시계열모형의 잔차들 $\{\hat{v}_t\}$ 가 백색잡음과정에서 나왔다는 가설하에서 이 퍼트맨토우통계량은 점근적으로 자유도 $(K-p-q)$ 인 카이제곱분포를 따른다.

교차상관관계의 검진과 자기상관관계의 검진의 결과는 다음과 같은 세 가지 경우로 나눌 수 있다.

- ① 교차상관관계의 검진결과로 잡음시계열모형의 잔차들 $\{\hat{v}_t\}$ 와 사전백색잡음화된 입력시계열데이터 $\{z_t\}$ 가 서로 독립이라는 가정을 채택하고, 자기상관관계의 검진결

과로 선택된 잡음시계열데이터의 ARMA모형이 주어진 시계열데이터와 잘 부합된다는 가정을 채택한다.

- ② 교차상관관계의 검진결과로 잡음시계열모형의 잔차들 $\{\hat{v}_i\}$ 와 사전백색잡음화된 입력시계열데이터 $\{z_i\}$ 가 서로 독립이라는 가정을 채택하나, 자기상관관계의 검진결과로 선택된 잡음항의 ARMA모형이 주어진 시계열데이터와 잘 부합된다는 가정을 기각한다. 이는 추정된 전이함수는 시계열데이터와 잘 부합되나, 추정된 잡음시계열데이터의 ARMA모형은 시계열데이터와 잘 부합되지 않는 경우이다. 따라서 모든 k 에 대해서 $\rho_w(k)=0$ 이지만 어떤 $i(\neq 0)$ 에 대해서는 $\rho_w(i)\neq 0$ 이다. 이 경우에는 잡음시계열데이터의 ARMA모형을 다시 식별하고 추정한다.
- ③ 교차상관관계의 검진결과로 잡음시계열모형의 잔차들 $\{\hat{v}_i\}$ 와 사전백색잡음화된 입력시계열데이터 $\{z_i\}$ 가 서로 독립이라는 가정을 기각한다. 이는 선택된 전이함수모형이 주어진 시계열데이터와 잘 부합되지 않는 경우이다. 이 경우에는 잡음시계열데이터의 ARMA모형의 타당성과 관계없이 어떤 k 에 대해서 $\rho_w(k)\neq 0$ 이고, 어떤 $i(\neq 0)$ 에 대해서 $\rho_w(i)\neq 0$ 이다. 이를 자세히 살펴보기 위해서 원래의 전이함수모형이 다음과 같다고 가정하자.

$$y_i = \nu(B)x_i + \psi(B)v_i$$

이 전이함수모형을 다음과 같은 모형으로 잘못 추정했다고 하자.

$$y_i = \nu_0(B)x_i + \psi_0(B)b_i$$

이 결과 잡음시계열모형의 오차항은 다음과 같다.

$$b_i = \psi_0^{-1}(B)\{\nu(B) - \nu_0(B)\}x_i + \psi_0^{-1}(B)\psi(B)v_i$$

이 식에서 알 수 있듯이 $\{b_i\}$ 는 $\{x_i\}$ 및 $\{v_i\}$ 와 교차상관되어 있을 뿐 아니라 자기상관도 되어 있다. 이러한 현상은 다음과 같이 잡음항의 ARMA모형이 정확한 경우에도 마찬가지로 일어난다.

$$\psi_0(z) = \psi(z)$$

따라서 이러한 경우의 처방책은 전이함수 $\nu(z)$ 를 다시 식별 및 추정한 다음 잡음항의 모형을 수정하는 것이다.

잡음시계열데이터의 모형은 부정확한 전이함수모형에 의해서 오염될(contaminated) 수 있으므로 전이함수모형을 검진할 때는 우선 교차상관함수를 신중하게 조사해 보아야 한다. 이는 분산분석모형에서 주효과(main effect)를 검정하기 전에 상호작용효과(interaction effect)를 검정하는 것과 유사하다. 교차상관관계의 검진결과로 잡음시계열모형의 잔차들 $\{\hat{v}_i\}$ 와 사전백색잡음화된 입력시계열데이터 $\{z_i\}$ 가 서로 독립이라는 가정을 기각하면, 자기상관관계의 검진을 하는 것은 의미가 없다. 이러한 경우의 처방책으로는 전이함수모형을 다시 식별하고 추정한 모형

이 검진의 단계에서 주어진 시계열데이터와 잘 부합된다고 판단될 때까지 모형화 작업을 반복한다.

PROC ARIMA의 ESTIMATE스태이트먼트를 사용해서 최종적으로 전이함수모형을 추정하면, 잡음시계열데이터의 추정된 ARMA모형이 주어진 시계열데이터와 잘 부합되는가를 조사하기 위한, 즉 잔차값들의 자기상관검진을 위한 퍼트맨토우통계량 Q_1 이 출력된다. 또한, 한 PROC ARIMA 안에서 입력시계열데이터를 식별하기 위한 IDENTIFY스태이트먼트와 추정하기 위한 ESTIMATE스태이트먼트를 사용한 다음에 출력시계열데이터를 전이함수모형에 적합시키기 위한 ESTIMATE스태이트먼트를 사용하면 추정된 전이함수모형에 의한 잡음시계열모형의 잔차 $\{\hat{v}_i\}$ 와 사전백색잡음화된 입력시계열데이터 $\{z_i\}$ 가 서로 독립이라는 가설을 검진하기 위한 퍼트맨토우통계량 Q_0 가 출력된다.

전이함수모형을 검진하는 첫 번째 예제로써 SAS프로그램 EXAM4-19.PRG에서 추정한 전이함수모형을 분석하기로 하자. 그림 4.19의 부분 B의 퍼트맨토우통계량값들에서 볼 수 있듯이 추정된 전이함수모형의 잡음시계열모형의 잔차값들이 주어진 시계열데이터와 잘 부합된다는 가설을 채택하는 데 큰 무리가 없다. 부분 C의 퍼트맨토우통계량값들로부터 잡음시계열모형의 잔차값들과 사전백색잡음화된 입력시계열데이터 $\{z_i\}$ 가 서로 독립이라는 가설을 채택하는 데 큰 무리가 없음을 알 수 있다.

전이함수모형을 검진하는 두 번째 예제로써 SAS프로그램 EXAM4-20.PRG에서 추정한 전이함수모형을 살펴보자. 그림 4.20의 부분 F의 퍼트맨토우통계량값들에서 알 수 있듯이 최종적으로 추정된 전이함수모형의 잡음시계열모형의 잔차값들이 주어진 시계열데이터와 잘 부합된다는 가설을 채택하기가 어렵다. 반면에 부분 G의 퍼트맨토우통계량값들로부터 잡음시계열모형의 잔차값들과 사전백색잡음화된 입력시계열데이터 $\{z_i\}$ 가 서로 독립이라는 가설을 채택하는 데 큰 무리가 없음을 알 수 있다. 앞에서 언급하였듯이 이러한 경우 추정된 전이함수는 시계열데이터와 잘 부합되나, 추정된 잡음시계열데이터의 ARMA모형은 시계열데이터와 잘 부합되지가 않는 경우이다. 주어진 시계열데이터에 좀더 잘 부합되는 전이함수모형을 구하기 위해서 다시 식별하고 추정한 전이함수모형이 검진의 단계에서 주어진 시계열데이터와 잘 부합된다고 판단될 때까지 모형화작업을 반복하기로 한다. SAS 데이터세트 TFM2에 좀더 잘 부합되는 전이함수모형을 구하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM4-21.PRG에 수록되어 있다).

```

/* */
/*          EXAM4-21.PRG          */
/* */
/*          DIAGNOSTIC CHECKING OF */
/*          TRANSFER FUNCTION MODEL */
/*          DATASET = TFM2         */
/* */

```

```

LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS LINESIZE=72;
OPTIONS PAGESIZE=80;
DATA EXAM421;
    SET TSA4DATA.TFM2;
    RUN;
PROC ARIMA DATA=EXAM421;
    I VAR=X NOPRINT;
    E P=1 Q=1 NOPRINT;
    I VAR=Y CROSS=(X) NOPRINT;
    E P=4 INPUT=(3$ (1)/(1) X) NOINT
    ALTPARM METHOD=ML MAXIT=30;
    RUN;

```

그림 4.20의 부분 Ⅱ에서 알 수 있듯이 SAS프로그램 EXAM4-20.PRG에서 시계열데이터 TFM2의 잡음시계열데이터에 MA(2)모형을 적합시키면 MA특성방정식의 한 근이 1에 가깝기 때문에 전이함수모형의 최종적인 추정모형이 검진단계에서 좋지 않은 모형으로 판단되었다. 따라서 SAS프로그램 EXAM4-21.PRG에서는 잡음시계열데이터의 모형으로 AR(4)모형을 사용한다. 또한 전이함수의 분모의 이차항의 계수추정값이 -0.007로 0과 유의적으로 다르지 않다. 따라서 전이함수의 분모에 1차식을 사용한다. 이 SAS프로그램을 수행한 결과가 그림 4.21이다. 그림 4.21의 부분 Ⅰ과 부분 Ⅲ에서 볼 수 있듯이 최우추정법에 의해 추정된 모형은 다음과 같다.

$$y_t = 0.8548 \frac{1+1.3251B}{1+0.1714B} x_{t-3} + \frac{1}{1+1.3406B+1.1821B^2+0.9209B^3+0.4225B^4} v_t$$

그림 4.21의 부분 Ⅱ의 퍼트맨토우통계량값들에서 볼 수 있듯이 잡음시계열모형의 잔차값들이 주어진 시계열데이터와 잘 부합된다는 가설을 기각할 수도 있고 채택할 수도 있다. 그러나 이 퍼트맨토우통계량값들이 그림 4.20의 부분 Ⅱ의 퍼트맨토우통계량값들보다는 작음을 알 수 있다. 그림 4.21의 부분 Ⅲ의 퍼트맨토우통계량값들로부터 잡음시계열모형의 잔차값들과 사전백색화된 입력시계열데이터 $\{z_t\}$ 가 서로 독립이라는 가설을 채택하는 데 큰 무리가 없음을 알 수 있다. 더구나 잡음시계열데이터를 AR모형에 적합시킴으로써 MA특성식을 제거했기 때문에 최우추정법의 알고리즘이 수렴한다. 따라서 그림 4.21에서 구축한 전이함수모형이 그림 4.20에서 구축한 전이함수모형보다 나음을 알 수 있다.

SAS 14:38 Friday, October 1, 1993 166

ARIMA Procedure

(A)

Maximum Likelihood Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag	Variable	Shift
AR1,1	-1.34059	0.09576	-14.00	1	Y	0
AR1,2	-1.18211	0.14084	-8.39	2	Y	0
AR1,3	-0.92086	0.14055	-6.55	3	Y	0
AR1,4	-0.42252	0.09606	-4.40	4	Y	0
SCALE1	0.85481	0.03963	21.57	0	X	3
NUM1,1	-1.32508	0.19460	-6.81	1	X	3
DEN1,1	-0.17141	0.05081	-3.37	1	X	3

Variance Estimate = 1.40619638
 Std Error Estimate = 1.18583152
 AIC = 317.510643
 SBC = 335.53362
 Number of Residuals = 97

Correlations of the Estimates

Variable	Parameter	Y	Y	Y	Y	X
		AR1,1	AR1,2	AR1,3	AR1,4	SCALE1
Y	AR1,1	1.000	0.796	0.584	0.453	0.126
Y	AR1,2	0.796	1.000	0.822	0.582	0.166
Y	AR1,3	0.584	0.822	1.000	0.796	0.107
Y	AR1,4	0.453	0.582	0.796	1.000	0.088
X	SCALE1	0.126	0.166	0.107	0.088	1.000
X	NUM1,1	0.127	0.158	0.088	0.071	0.961
X	DEN1,1	0.118	0.138	0.064	0.050	0.848

Variable	Parameter	X	X
		NUM1,1	DEN1,1
X	AR1,1	0.127	0.118
X	AR1,2	0.158	0.138
X	AR1,3	0.088	0.064
X	AR1,4	0.071	0.050
X	SCALE1	0.961	0.848
X	NUM1,1	1.000	0.959
X	DEN1,1	0.959	1.000

(B)

Autocorrelation Check of Residuals

To Lag	Chi Square	DF	Prob	Autocorrelations					
6	9.07	2	0.011	-0.095	-0.205	-0.133	-0.065	-0.091	0.088
12	19.37	8	0.013	0.008	-0.100	0.111	0.207	-0.062	-0.156
18	31.11	14	0.005	-0.210	0.112	0.189	0.015	-0.001	-0.093
24	37.20	20	0.011	0.083	-0.045	-0.113	0.037	-0.143	0.064

(C)

Crosscorrelation Check of Residuals with Input X

To Lag	Chi Square	DF	Prob	Crosscorrelations					
5	2.08	3	0.557	-0.016	-0.009	0.025	-0.026	0.045	-0.136
11	3.15	9	0.958	0.036	-0.046	0.048	0.049	0.046	0.034
17	7.16	15	0.953	-0.014	-0.185	0.002	-0.050	0.060	0.044
23	9.18	21	0.988	-0.055	0.051	0.048	0.056	-0.085	0.057

(D)

Model for variable Y

No mean term in this model.


```

Autoregressive Factors
Factor 1: 1 + 1.3406 B**(1) + 1.1821 B**(2) + 0.92086 B**(3) + 0.42252 B**(4)

Input Number 1 is X with a shift of 3.
Overall Regression Factor = 0.85481

The Numerator Factors are
Factor 1: 1 + 1.3251 B**(1)

The Denominator Factors are
Factor 1: 1 + 0.17141 B**(1)

```

그림 4.21 추정된 전이함수모형의 검진과 재모형화

4.6 전이함수모형의 예측

PROC ARIMA를 사용해서 식별하고 모수를 추정한 전이함수모형이 주어진 시계열데이터와 잘 부합되면, 이 선택된 전이함수모형을 사용해서 미래의 값들을 예측할 수 있다. 즉, PROC ARIMA의 FORECAST스태이트먼트를 사용하면 출력시계열데이터의 미래값들을 추정된 전이함수모형에 의해서 예측할 수 있다. 이러한 예측을 하기 위해서는 우선 PROC ARIMA에 입력시계열데이터를 식별하기 위한 IDENTIFY스태이트먼트와 입력시계열데이터를 추정하기 위한 ESTIMATE스태이트먼트를 사용한 다음, 출력시계열데이터를 유리함수형태의 전이함수모형에 적합시키기 위한 IDENTIFY스태이트먼트와 ESTIMATE스태이트먼트를 사용하고, 그 다음에 출력시계열데이터를 예측하기 위한 FORECAST스태이트먼트를 사용한다. 이렇게 구성된 PROC ARIMA에서는 우선 입력시계열데이터를 분석하기 위한 IDENTIFY스태이트먼트와 ESTIMATE스태이트먼트에 의해서 구축된 입력시계열데이터의 ARIMA모형에 의해서 입력시계열데이터의 미래값들을 예측한 뒤, 출력시계열데이터를 분석하기 위한 IDENTIFY스태이트먼트와 ESTIMATE스태이트먼트에 의해서 구축된 출력시계열데이터의 전이함수모형에 이 입력시계열데이터의 미래값들을 적용해서 출력시계열데이터의 미래값들을 예측한다. 그러나 PROC ARIMA에 입력시계열데이터를 식별하기 위한 IDENTIFY스태이트먼트와 입력시계열데이터를 추정하기 위한 ESTIMATE스태이트먼트가 사용되지 않는 경우에는, 입력시계열데이터의 미래값들이 주어진 데이터세트에 포함된 경우에 한해서 출력시계열데이터의 미래값들을 예측할 수 있다. ARIMA모형을 이용한 예측에 관한 자세한 내용은 [SAS4TSA2] (pp. 601~620)를 참조하라.

전이함수모형을 사용한 예측의 첫 번째 예제로써 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM4-22.PRG에 수록되어 있다).

```

/* */
/*          EXAM4-22.PRG          */
/* */
/*          TRANSFER FUNCTION MODEL FORECASTING          */
/*          DATASET = TFM1          */
/* */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS LINESIZE=72;
OPTIONS PAGESIZE=80;
DATA EXAM422;
    SET TSA4DATA.TFM1;
    RUN;
PROC ARIMA DATA=EXAM422;
    I VAR=X NOPRINT;
    E P=1 Q=1 NOPRINT;
    I VAR=Y CROSS=(X) NOPRINT;
    E P=4 INPUT=(2$(1)/(1)X) NOPRINT;
    F LEAD=25 ID=T OUT=EXAM422B;
    RUN;
PROC PLOT DATA=EXAM422B(FIRSTOBS=116);
    PLOT FORECAST*T='F' Y*T='Y' L95*T='L'
        U95*T='U' / VPOS=25 OVERLAY;
    RUN;

```

이 SAS 프로그램을 수행한 결과가 그림 4.22이다. 그림 4.22의 부분 A에는 예측에 사용된 전이함수모형이 수록되어 있다. 이 전이함수모형은 그림 4.19의 부분 D에서 알 수 있듯이 최종적으로 추정된 전이함수모형이다. 부분 B에는 예측값(Forecast), 표준오차(Std

SAS 8:04 Saturday, October 2, 1993
1
ARIMA Procedure
(A)
Model for variable Y
Estimated Intercept = -0.009848
Autoregressive Factors
Factor 1: $1 + 1.48 B^{**}(1) + 1.1822 B^{**}(2) + 0.81655 B^{**}(3) + 0.34005 B^{**}(4)$
Input Number 1 is X with a shift of 2.
The Numerator Factors are
Factor 1: $0.8675 + 0.76059 B^{**}(1)$
The Denominator Factors are
Factor 1: $1 + 0.17837 B^{**}(1)$

(B)

Forecasts for variable Y

Obs	Forecast	Std Error	Lower 95%	Upper 95%
122	0.5862	1.3227	-2.0063	3.1787
123	5.3292	2.3626	0.6986	9.9598
124	5.0886	3.7950	-2.3493	12.5266
125	1.9309	5.7271	-9.2940	13.1559
126	0.2510	6.2999	-12.0966	12.5986
127	-1.6323	6.4413	-14.2569	10.9924
128	-1.3533	6.4750	-14.0440	11.3374
129	-2.3403	6.4841	-15.0489	10.3683
130	-1.9154	6.4877	-14.6310	10.8003
131	-2.3805	6.4895	-15.0996	10.3387
132	-2.2114	6.4918	-14.9352	10.5123
133	-2.2946	6.4942	-15.0230	10.4338
134	-2.3261	6.4953	-15.0566	10.4043
135	-2.2582	6.4955	-14.9892	10.4728
136	-2.3607	6.4956	-15.0919	10.3705
137	-2.2607	6.4957	-14.9919	10.4706
138	-2.3453	6.4957	-15.0765	10.3860
139	-2.2844	6.4957	-15.0157	10.4470
140	-2.3247	6.4957	-15.0561	10.4067
141	-2.3037	6.4958	-15.0351	10.4278
142	-2.3090	6.4958	-15.0405	10.4225
143	-2.3142	6.4958	-15.0458	10.4173
144	-2.3039	6.4958	-15.0354	10.4276
145	-2.3159	6.4958	-15.0475	10.4156
146	-2.3043	6.4958	-15.0358	10.4272

(C)

Plot of FORECAST*T. Symbol used is 'F'.
 Plot of Y*T. Symbol used is 'Y'.
 Plot of L95*T. Symbol used is 'L'.
 Plot of U95*T. Symbol used is 'U'.

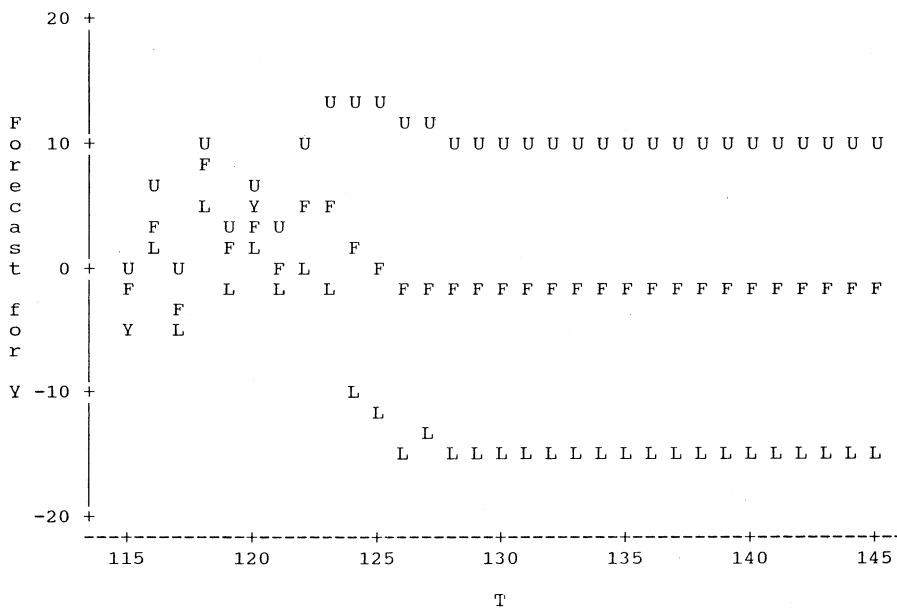


그림 4.22 전이함수모형을 사용한 시계열데이터 TFM1의 예측

Error), 신뢰도 95%의 예측구간의 하한값(Lower 95%)과 상한값(Upper 95%)이 실려 있다. 이들은 각각 변수들 FORECAST, STD, L95와 U95에 실려 있다. 이 PROC ARIMA의 FORECAST스태이트먼트의 옵션 OUT=에 SAS데이터세트 EXAM422B를 할당하였다. 이 SAS데이터세트에는 FORECAST스태이트먼트를 수행한 결과가 수록된다. 이 SAS데이터세트에 저장된 변수들은 만일 BY변수가 있다면 BY변수, 만일 ID변수가 있다면 ID변수, IDENTIFY스태이트먼트의 옵션 VAR=에 지정한 시계열데이터, FORECAST, STD, L95, U95, 그리고 잔차들을 수록한 RESIDUAL이다. 옵션 OUT=에 대한 자세한 내용은 [ETS6.1](pp. 113~114)을 참조하라. 부분 ㉔는 PROC PLOT를 사용해서 SAS데이터세트 EXAM422B에 저장된 출력시계열데이터의 관찰, 예측, 예측구간의 하한, 그리고 상한의 시계열산포도를 출력시킨 것이다.

전이함수모형을 사용한 예측의 두 번째 예제로써 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM4-23.PRG에 수록되어 있다).

```

/* */
/* */          EXAM4-23.PRG          /* */
/* */
/* */      TRANSFER FUNCTION MODEL FORECASTING      /* */
/* */      DATASET = TFM2                      /* */
/* */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS LINESIZE=72;
OPTIONS PAGESIZE=80;
DATA EXAM423;
    SET TSA4DATA.TFM2;
    RUN;
PROC ARIMA DATA=EXAM423;
    I VAR=X NOPRINT;
    E P=1 Q=1 NOPRINT;
    I VAR=Y CROSS=(X) NOPRINT;
    E P=4 INPUT=(3$ (1)/ (1) X) NOINT
        ALTPARM METHOD=ML MAXIT=30 NOPRINT;
    F LEAD=20 OUT=EXAM423B;
    RUN;
DATA EXAM423C;
    SET EXAM423B;
    T=_N_;
    RUN;
PROC PLOT DATA=EXAM423C(FIRSTOBS=91);

```

```
PLOT FORECAST*T='F' Y*T='Y' L95*T='L'
      U95*T='U' / VPOS=25 OVERLAY:
RUN:
```

이 SAS프로그램을 수행한 결과가 그림 4.23이다. 그림 4.23의 부분 A에는 예측에 사용된 전이함수모형이 수록되어 있다. 이 전이함수모형은 그림 4.21의 부분 D에서 알 수 있듯이 최종적으로 추정된 전이함수모형이다. 부분 B에는 예측, 표준오차, 신뢰도 95%의 예측구간의 하한과 상한이 실려 있다. 부분 C는 PROC PLOT를 사용해서 이 출력시계열데이터의 관찰, 예측, 예측구간의 하한, 그리고 상한의 시계열산포도를 출력시킨 것이다.

SAS 21:21 Friday, October 1, 1993 21				
ARIMA Procedure				
(A)				
Model for variable Y				
No mean term in this model.				
Autoregressive Factors				
Factor 1: $1 + 1.3406 B^{**}(1) + 1.1821 B^{**}(2) + 0.92086 B^{**}(3) + 0.42252 B^{**}(4)$				
Input Number 1 is X with a shift of 3.				
Overall Regression Factor = 0.85481				
The Numerator Factors are				
Factor 1: $1 + 1.3251 B^{**}(1)$				
The Denominator Factors are				
Factor 1: $1 + 0.17141 B^{**}(1)$				
(B)				
Forecasts for variable Y				
Obs	Forecast	Std Error	Lower 95%	Upper 95%
102	-2.3972	1.1858	-4.7213	-0.0730
103	-9.0554	1.9833	-12.9426	-5.1683
104	-8.7577	2.1131	-12.8994	-4.6161
105	-6.9475	3.3754	-13.5631	-0.3319
106	-4.9416	6.4251	-17.5345	7.6512
107	-4.3367	7.3825	-18.8062	10.1327
108	-4.5787	7.6286	-19.5305	10.3730
109	-3.5005	7.7118	-18.6153	11.6142
110	-3.9427	7.7356	-19.1042	11.2187
111	-3.5428	7.7429	-18.7186	11.6330
112	-3.8177	7.7468	-19.0012	11.3658
113	-3.6230	7.7483	-18.8095	11.5634
114	-3.5475	7.7486	-18.7345	11.6395
115	-3.6877	7.7489	-18.8752	11.4999
116	-3.5928	7.7489	-18.7803	11.5948
117	-3.6732	7.7489	-18.8607	11.5144
118	-3.5621	7.7491	-18.7501	11.6259
119	-3.6340	7.7492	-18.8221	11.5542
120	-3.6293	7.7492	-18.8175	11.5588
121	-3.6158	7.7492	-18.8039	11.5724
(C)				
Plot of FORECAST*T. Symbol used is 'F'.				
Plot of Y*T. Symbol used is 'Y'.				
Plot of L95*T. Symbol used is 'L'.				
Plot of U95*T. Symbol used is 'U'.				

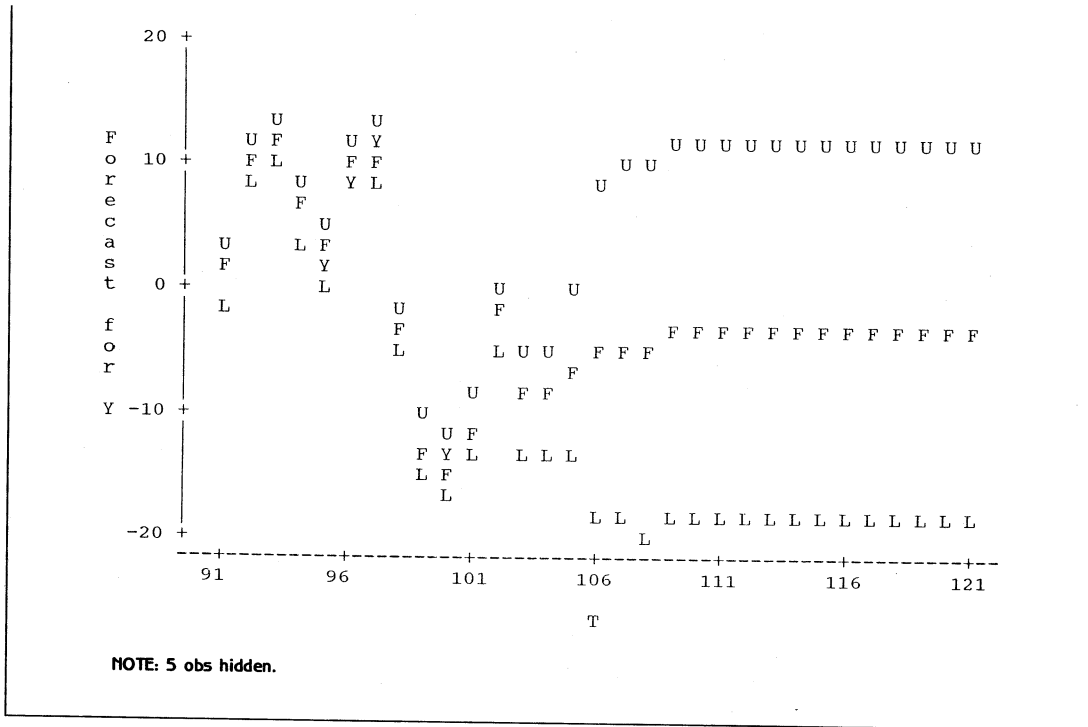


그림 4.23 전이함수모형을 사용한 시계열데이터 TFM2의 예측

4.7 예 제 들

이 절에서는 지금까지 다룬 전이함수모형화방법을 여러 시계열데이터들에 적용해서 이 방법의 유용성을 보이코자 한다. 전이함수모형을 사용해서 시계열데이터를 분석한 국내문헌으로는 남기석(1992)과 양현조(1992)가 있다.

4.7.1 다변량시계열데이터 BJ_J

이 소절에서는 Box & Jenkins(1976, pp. 532~533)의 Series J인 시계열데이터 BJ_J를 유리함수형태의 전이함수모형화 한다. 이 2변량시계열데이터의 입력시계열데이터 X는 입력개스비율(input gas rate)이고 출력시계열데이터 Y는 출력개스내의 이산화탄소의 퍼센트(CO₂ percent in outlet gas)이다. 이 시계열데이터는 296개의 관찰점들로 구성되어 있다. 시계열데이터 BJ_J는 제2.2절에 수록되어 있고 디스켓 SAS4TSA4의 SAS데이터라이브러리 TSA4DATA의 파일 BJJ.SSD에 저장되어 있다. 이 시계열데이터의 시계열산포

도가 그림 2.2이다. [ETS6.1] (pp. 153~162)에서 이 시계열데이터를 전이함수모형을 사용해서 분석하였고, 이 소절에서는 이 분석을 따르고자 한다.

시계열데이터 BJ_J를 유리함수형태의 전이함수모형화하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM4-24.PRG에 수록되어 있다).

```

/* */
/*          EXAM4-24.PRG          */
/* */
/*          TRANSFER FUNCTION MODELING          */
/*          DATASET = BJ_J          */
/* */

LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS LINESIZE=72;
OPTIONS PAGESIZE=80;
DATA EXAM424;
    SET TSA4DATA.BJJ;
RUN;
TITLE1 'BOX-JENKINS SERIES J';
TITLE2 'GAS FURNACE DATA';
PROC ARIMA DATA=EXAM424;
    I VAR=X;
    E P=3;
    I VAR=Y CROSS=(X);
    E INPUT=(3$ (1,2)/(1,2) X)
    PLOT MAXIT=30;
    E P=2 INPUT=(3$ (1,2)/(1,2) X);
    F LEAD=20 OUT=EXAM424B;
RUN;
DATA EXAM424C;
    SET EXAM424B;
    T= _N_ ;
RUN;
PROC PLOT DATA=EXAM424C(FIRSTOBS=291);
    PLOT FORECAST*T='F' Y*T='Y' L95*T='L'
        U95*T='U' / VPOS=25 OVERLAY;
RUN;

```

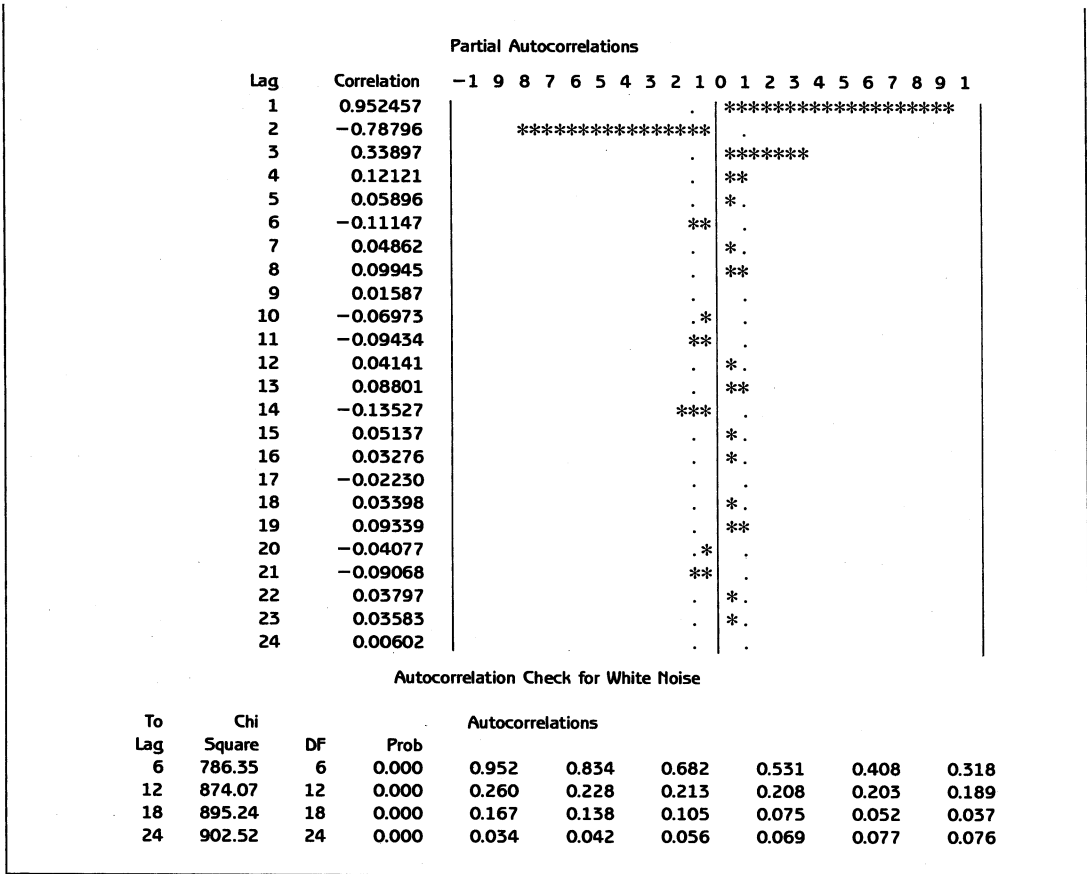



그림 4.24.1 입력시계열데이터의 식별

이 SAS프로그램의 PROC ARIMA의 첫 번째 IDENTIFY스테이트먼트는 다음과 같다.

IDENTIFY VAR=X;

이 스테이트먼트는 입력시계열데이터 X를 식별하기 위한 것이다. 이 스테이트먼트를 수행한 결과가 그림 4.24.1이다. 그림 4.24.1에서 표본자기상관함수는 시나브로형태를 보이고 표본부분자기상관함수는 시차 4 이후에서 절단형태를 보인다. 따라서 입력시계열데이터 X를 AR(3)모형으로 식별한다.

PROC ARIMA의 첫 번째 ESTIMATE스테이트먼트는 다음과 같다.

ESTIMATE P=3;

이 스테이트먼트는 입력시계열데이터 X를 AR(3)모형에 적합시키기 위한 것이다. 이 스테이트먼트를 수행한 결과가 그림 4.24.2이다. 그림 4.24.2에서 입력시계열데이터의 추정모형이 다음과 같음을 알 수 있다.

$$x_t = -0.0068 + 1.9761x_{t-1} - 1.3750x_{t-2} + 0.3434x_{t-3} + v_t$$

이 AR(3)모형에 의한 입력시계열데이터의 잔차값들의 퍼트맨트우통계량값들로부터 추

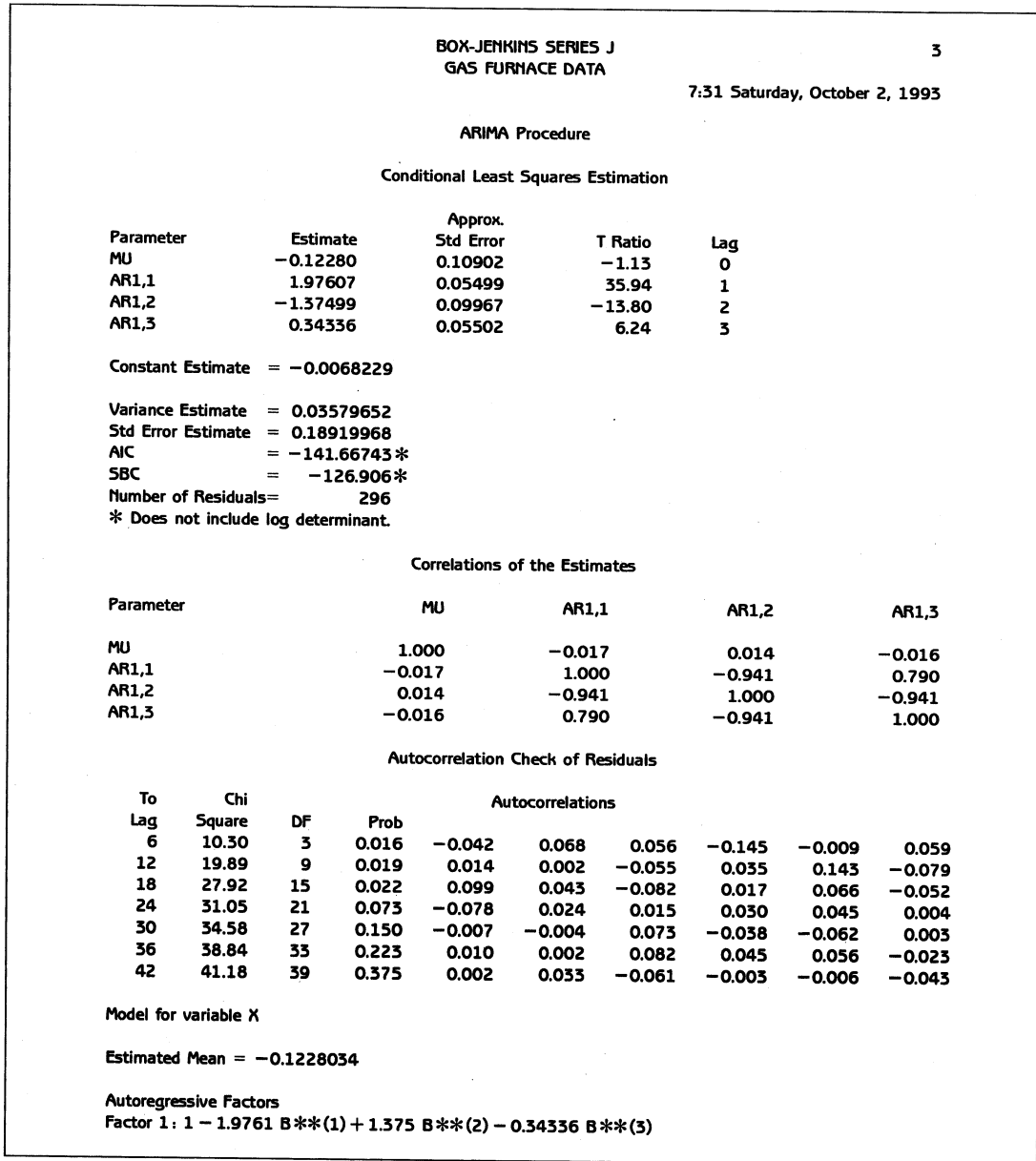


그림 4.24.2 입력시계열데이터의 추정

정된 AR(3)모형이 입력시계열데이터와 잘 부합됨을 알 수 있다.

PROC ARIMA의 두 번째 IDENTIFY스태이트먼트는 다음과 같다.

IDENTIFY VAR=Y CROSS=(X);

이 스타이트먼트는 입력시계열데이터가 X이고 출력시계열데이터가 Y인 전이함수를 식별하기 위한 것이다. 이 스타이트먼트를 수행한 결과가 그림 4.24.3이다. 그림 4.24.3에는 사전백색화된 입력시계열데이터와 같은 필터로 여과된 출력시계열데이터의 표본교차상관함수가 수록되어 있다. 이 표본교차상관함수에서 시차들 3, 4, 그리고 5의 표본상관계수들

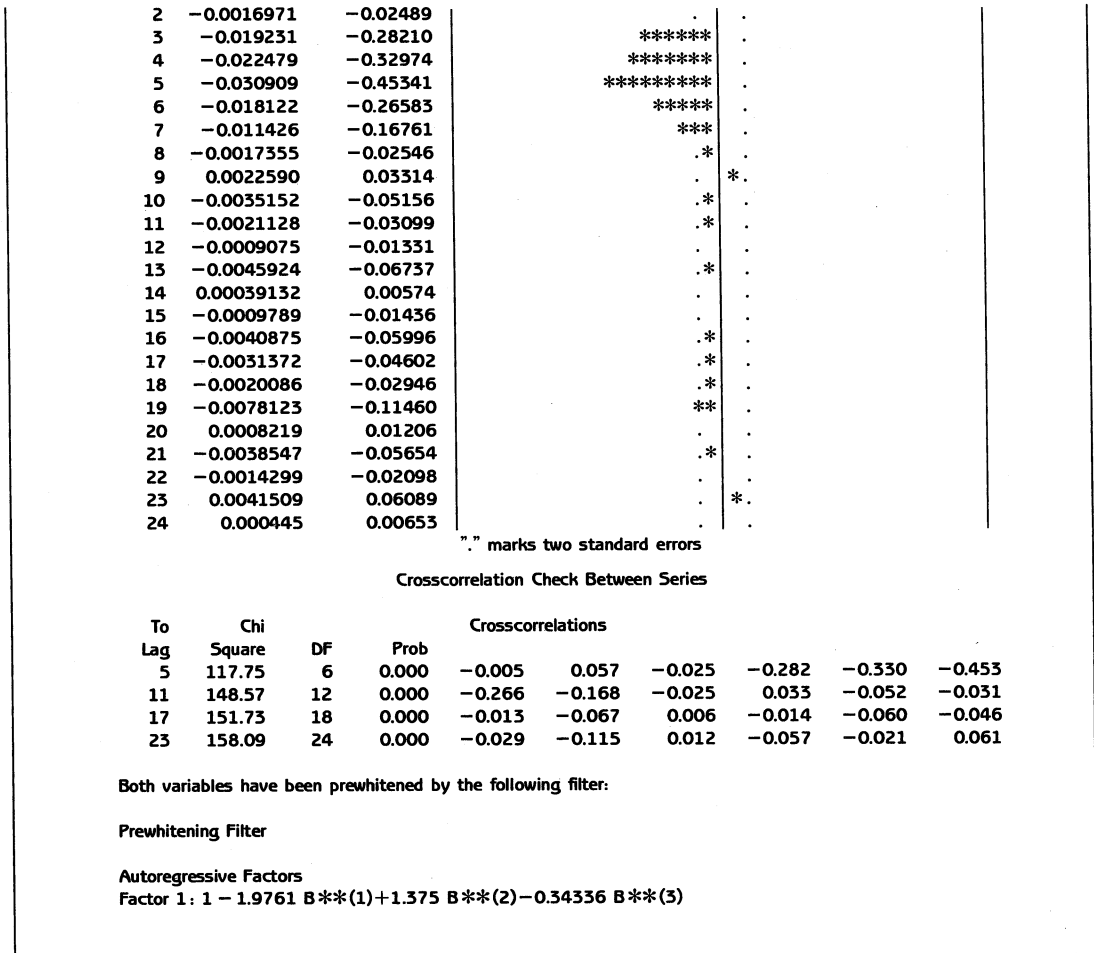


그림 4.24.3 전이함수의 식별

이 유의적으로 0과 같지 않고 또한 시차 5 이후의 표본상관계수들이 시나브로형태를 보인다 고 할 수 있다. 따라서 이 시계열데이터는 Box-Jenkins방법에 의해서 차수들이 (3, 2, 2) 인 전이함수모형으로 식별할 수 있다.

PROC ARIMA의 두 번째 ESTIMATE스테이트먼트는 다음과 같다.

ESTIMATE INPUT=(3\$ (1, 2)/(1, 2) X) PLOT MAXIT=30;

이 ESTIMATE스테이트먼트는 잡음시계열데이터의 ARMA모형을 식별하기 위한 것이다. 따라서 ESTIMATE스테이트먼트의 옵션 PLOT를 사용하였다. 이 스테이트먼트를 수행한 결과가 그림 4.24.4이다. 그림 4.24.4에서 다음과 같은 추정식을 얻는다.

$$y_t = 53.3224 + \frac{-0.6287 - 0.4726B - 0.7366B^2}{1 - 0.1541B - 0.2777B^2} x_{t-3} + \hat{n}_t$$

이 추정모형에 의한 잡음시계열데이터의 적합값들 { \hat{n}_t }의 표본자기상관함수는 시나브로 형태를 보이고, 표본역자기상관함수는 시차 2 이후에서 절단형태를 보이며, 표본부분자기

10	-0.028378	-0.04030
11	-0.015417	-0.02189
12	-0.014712	-0.02089
13	-0.031434	-0.04463
14	-0.058300	-0.08278
15	-0.092307	-0.13107
16	-0.119741	-0.17003
17	-0.138025	-0.19599
18	-0.137245	-0.19488
19	-0.114550	-0.16266
20	-0.071587	-0.10165
21	-0.018429	-0.02617
22	0.033281	0.04726
23	0.074909	0.10637
24	0.099810	0.14173

.	.	*	.
.	.	.	.
.	.	*	.
.	.	**	.
.	.	***	.
.	.	***	.
.	.	***	.
.	.	***	.
.	.	***	.
.	.	***	.
.	.	**	.
.	.	*	.
.	.	.	*
.	.	.	**
.	.	.	***

"," marks two standard errors

Inverse Autocorrelations

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
1	-0.56485
2	0.01608
3	0.02586
4	0.05091
5	-0.01536
6	-0.00213
7	0.00488
8	-0.07603
9	0.10261
10	-0.00500
11	-0.03629
12	-0.03521
13	0.07135
14	-0.06096
15	0.06492
16	-0.05600
17	0.01508
18	0.00089
19	0.02587
20	-0.01283
21	-0.02295
22	0.02728
23	-0.01522
24	0.00257

Partial Autocorrelations

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
1	0.89294
2	-0.42765
3	-0.13463
4	0.02199
5	0.03891
6	-0.02219
7	-0.02249
8	0.01538
9	0.00634
10	0.07737
11	-0.05640
12	-0.10978
13	-0.07122
14	-0.00207
15	-0.06097
16	0.00550
17	-0.02745
18	0.05558
19	0.07093
20	0.06301
21	0.00920
22	-0.00171
23	0.02343
24	-0.00666

Crosscorrelation Check of Residuals with Input X

To Lag	Chi Square	DF	Prob	Crosscorrelations					
5	0.48	2	0.785	-0.009	-0.005	0.026	0.013	-0.017	-0.022

11	0.93	8	0.999	-0.006	0.008	0.022	0.023	-0.017	-0.013
17	2.63	14	1.000	0.012	0.035	0.037	0.039	-0.005	-0.040
23	19.19	20	0.509	-0.076	-0.108	-0.122	-0.122	-0.094	-0.041
29	20.12	26	0.786	-0.039	-0.013	0.010	-0.020	-0.031	-0.005
35	24.22	32	0.836	-0.022	-0.031	-0.074	-0.036	0.014	0.076
41	30.66	38	0.795	0.108	0.091	0.046	0.018	0.003	0.009

Model for variable Y

Estimated Intercept = 53.3223702

Input Number 1 is X with a shift of 3.

The Numerator Factors are
Factor 1: -0.6287 -0.47258 B**(1)-0.7366 B**(2)

The Denominator Factors are
Factor 1: 1-0.15411 B**(1)-0.27774 B**(2)

그림 4.24.4 입력시계열데이터의 식별

상관함수는 시차 4 이후에서 절단형태를 보인다. 따라서 잡음시계열데이터를 AR(1)모형, AR(2)모형, 또는 AR(3)모형으로 식별할 수 있다. 그러나 시차 3에서의 표본부분자기상관 계수의 절대값이 아주 작으므로 AR(2)모형으로 식별하기로 한다.

PROC ARIMA의 세 번째 ESTIMATE스테이트먼트는 다음과 같다.

ESTIMATE P=2 INPUT=(3\$ (1, 2)/(1, 2) X);

이 스테이트먼트는 출력시계열데이터 Y를 전이함수모형에 적합시키기 위한 것이다. 이 스테이트먼트를 수행한 결과가 그림 4.24.5이다. 그림 4.24.5에서 추정된 전이함수모형이 다음과 같음을 알 수 있다.

$$y_t = 5.3174 + \frac{-0.5362 - 0.3638B - 0.5074B^2}{1 - 0.5711B - 0.0165B^2} x_{t-3} + \frac{1}{1 - 1.5330B + 0.6329B^2} v_t$$

잡음시계열모형의 잔차 $\{\hat{v}_t\}$ 와 사전백색화된 입력시계열데이터 $\{z_t\}$ 의 표본교차상관함수에 의한 퍼트맨토우통계량값들로부터 잡음시계열모형의 잔차 $\{\hat{v}_t\}$ 와 사전백색화된 입력시계열데이터 $\{z_t\}$ 가 서로 독립이라는 가설을 채택한다. 추정된 전이함수모형의 잡음시계열모형의 잔차 $\{\hat{v}_t\}$ 의 퍼트맨토우통계량값들로부터 추정된 전이함수모형이 주어진 시계열데이터와 잘 부합됨을 알 수 있다.

PROC ARIMA의 첫 번째 FORECAST스테이트먼트는 다음과 같다.

FORECAST LEAD=20 OUT=EXAM424B;

이 FORECAST스테이트먼트는 바로 앞의 ESTIMATE스테이트먼트에서 추정된 전이함수모형을 이용해서 출력시계열데이터의 미래값들을 예측하기 위한 것이다. 이 FORECAST스테이트먼트를 수행한 결과가 그림 4.24.6이다. 그림 4.24.6의 첫 번째 부분에는 예

BOX-JENKINS SERIES J

11

GAS FURNACE DATA

23:22 Saturday, October 2, 1993

ARIMA Procedure

Conditional Least Squares Estimation

Parameter	Estimate	Approx.		Lag	Variable	Shift
		Std Error	T Ratio			
MU	53.26337	0.11988	444.30	0	Y	0
AR1,1	1.53303	0.04763	32.19	1	Y	0
AR1,2	-0.63286	0.05017	-12.62	2	Y	0
NUM1	-0.53617	0.07552	-7.10	0	X	3
NUM1,1	0.36379	0.14940	2.43	1	X	3
NUM1,2	0.50742	0.15762	3.22	2	X	3
DEN1,1	0.57112	0.20930	2.73	1	X	3
DEN1,2	-0.01645	0.14486	-0.11	2	X	3

Constant Estimate = 5.31741842

Variance Estimate = 0.0590327

Std Error Estimate = 0.24296647

AIC = 10.2780726*

SBC = 39.6646587*

Number of Residuals = 291

* Does not include log determinant.

Correlations of the Estimates

Variable	Parameter	Y		X		
		MU	AR1,1	AR1,2	NUM1	NUM1,1
Y	MU	1.000	-0.063	0.047	-0.013	-0.041
Y	AR1,1	-0.063	1.000	-0.927	-0.002	0.009
Y	AR1,2	0.047	-0.927	1.000	0.022	-0.009
X	NUM1	-0.013	-0.002	0.022	1.000	0.584
X	NUM1,1	-0.041	0.009	-0.009	0.584	1.000
X	NUM1,2	-0.018	0.003	-0.002	-0.049	0.290
X	DEN1,1	0.031	-0.003	0.004	-0.117	-0.709
X	DEN1,2	-0.041	0.006	-0.008	0.117	0.718

Variable	Parameter	X		X
		NUM1,2	DEN1,1	DEN1,2
X	MU	-0.018	0.031	-0.041
X	AR1,1	0.003	-0.003	0.006
X	AR1,2	-0.002	0.004	-0.008
X	NUM1	-0.049	-0.117	0.117
X	NUM1,1	0.290	-0.709	0.718
X	NUM1,2	1.000	-0.804	0.726
X	DEN1,1	-0.804	1.000	-0.983
X	DEN1,2	0.726	-0.983	1.000

Autocorrelation Check of Residuals

To Lag	Chi Square	DF	Prob	Autocorrelations					
6	8.62	4	0.071	0.024	0.056	-0.073	-0.055	-0.054	0.118
12	15.45	10	0.116	0.031	0.029	-0.080	0.048	0.023	0.107
18	21.16	16	0.173	-0.038	0.053	-0.094	-0.013	-0.072	-0.005
24	27.54	22	0.191	-0.117	-0.002	-0.007	0.076	0.024	-0.005
30	37.01	28	0.119	0.035	-0.021	0.019	0.093	-0.119	0.065
36	44.36	34	0.110	-0.025	-0.055	0.114	0.022	0.030	0.065
42	45.73	40	0.246	-0.018	-0.036	-0.029	-0.014	-0.032	0.018

Crosscorrelation Check of Residuals with Input X

To Lag	Chi Square	DF	Prob	Crosscorrelations					
5	0.88	2	0.643	0.008	0.005	0.006	0.012	-0.044	0.029

11	6.63	8	0.578	0.074	-0.027	-0.021	-0.028	-0.111	0.013
17	13.88	14	0.459	0.049	0.042	0.015	0.013	-0.141	-0.027
23	18.63	20	0.546	-0.074	-0.078	0.023	-0.016	0.021	0.060
29	28.05	26	0.356	-0.071	-0.000	0.038	-0.155	0.031	0.035
35	35.23	32	0.318	-0.014	0.015	-0.040	0.028	0.046	0.141
41	37.20	38	0.506	0.032	-0.029	-0.069	-0.007	0.011	-0.004

Model for variable Y

Estimated Intercept=53.2633739

Autoregressive Factors
Factor 1: $1 - 1.533 B^{**}(1) + 0.63286 B^{**}(2)$

Input Number 1 is X with a shift of 3.

The Numerator Factors are
Factor 1: $-0.5362 - 0.36379 B^{**}(1) - 0.50742 B^{**}(2)$

The Denominator Factors are
Factor 1: $1 - 0.57112 B^{**}(1) + 0.016452 B^{**}(2)$

그림 4.24.5 전이함수모형의 추정

BOX-JENKINS SERIES J
GAS FURNACE DATA

13

23:22 Saturday, October 2, 1993

ARIMA Procedure

Model for variable Y

Estimated Intercept = 53.2633739

Autoregressive Factors
Factor 1: $1 - 1.533 B^{**}(1) + 0.63286 B^{**}(2)$

Input Number 1 is X with a shift of 3.

The Numerator Factors are
Factor 1: $-0.5362 - 0.36379 B^{**}(1) - 0.50742 B^{**}(2)$

The Denominator Factors are
Factor 1: $1 - 0.57112 B^{**}(1) + 0.016452 B^{**}(2)$

Forecasts for variable Y

Obs	Forecast	Std Error	Lower 95%	Upper 95%
297	56.5279	0.2430	56.0517	57.0041
298	56.0475	0.4447	55.1759	56.9192
299	55.6078	0.6098	54.4125	56.8030
300	55.2180	0.7385	53.7706	56.6653
301	54.8199	0.8824	53.0905	56.5493
302	54.4261	1.1472	52.1776	56.6746
303	54.0715	1.5437	51.0459	57.0972
304	53.7854	1.9805	49.9037	57.6671
305	53.5812	2.3729	48.9303	58.2321
306	53.4570	2.6799	48.2045	58.7096
307	53.4005	2.8959	47.7246	59.0765
308	53.3944	3.0354	47.4452	59.3436
309	53.4207	3.1191	47.3073	59.5340
310	53.4640	3.1667	47.2575	59.6705
311	53.5127	3.1926	47.2554	59.7700
312	53.5588	3.2064	47.2745	59.8432
313	53.5981	3.2138	47.2993	59.8969
314	53.6286	3.2177	47.3219	59.9352
315	53.6502	3.2200	47.3392	59.9612
316	53.6640	3.2213	47.3505	59.9775

그림 4.24.6 전이함수모형을 사용한 예측

측에 사용된 전이함수모형이 수록되어 있다. 이 전이함수모형은 그림 4.24.5에서 알 수 있듯이 최종적으로 추정된 전이함수모형이다. 그림 4.24.6의 두 번째 부분에는 예측값(Forecast), 표준오차(Std Error), 신뢰도 95%의 예측구간의 하한값(Lower 95%)과 상한값(Upper 95%)이 수록되어 있다. 이들은 각각 변수들 FORECAST, STD, L95와 U95에 저장되어 있다. 이 FORECAST스테이트먼트의 옵션 OUT=에 SAS데이터세트 EXAM424B를 할당하였다. 이 SAS데이터세트에는 FORECAST스테이트먼트를 수행한 결과로 얻어진 예측의 결과가 수록된다. 이 SAS데이터세트에는 출력시계열데이터, FORECAST, STD, L95, U95, 그리고 RESIDUAL이 수록되어 있다.

이 SAS프로그램의 두 번째 PROC스텝은 다음과 같다.

```
PROC PLOT DATA=EXAM424C(FIRSTOBS=291);
PLOT FORECAST*T='F' Y*T='Y' L95*T='L'
      U95*T='U' / VPOS=25 OVERLAY;
RUN;
```

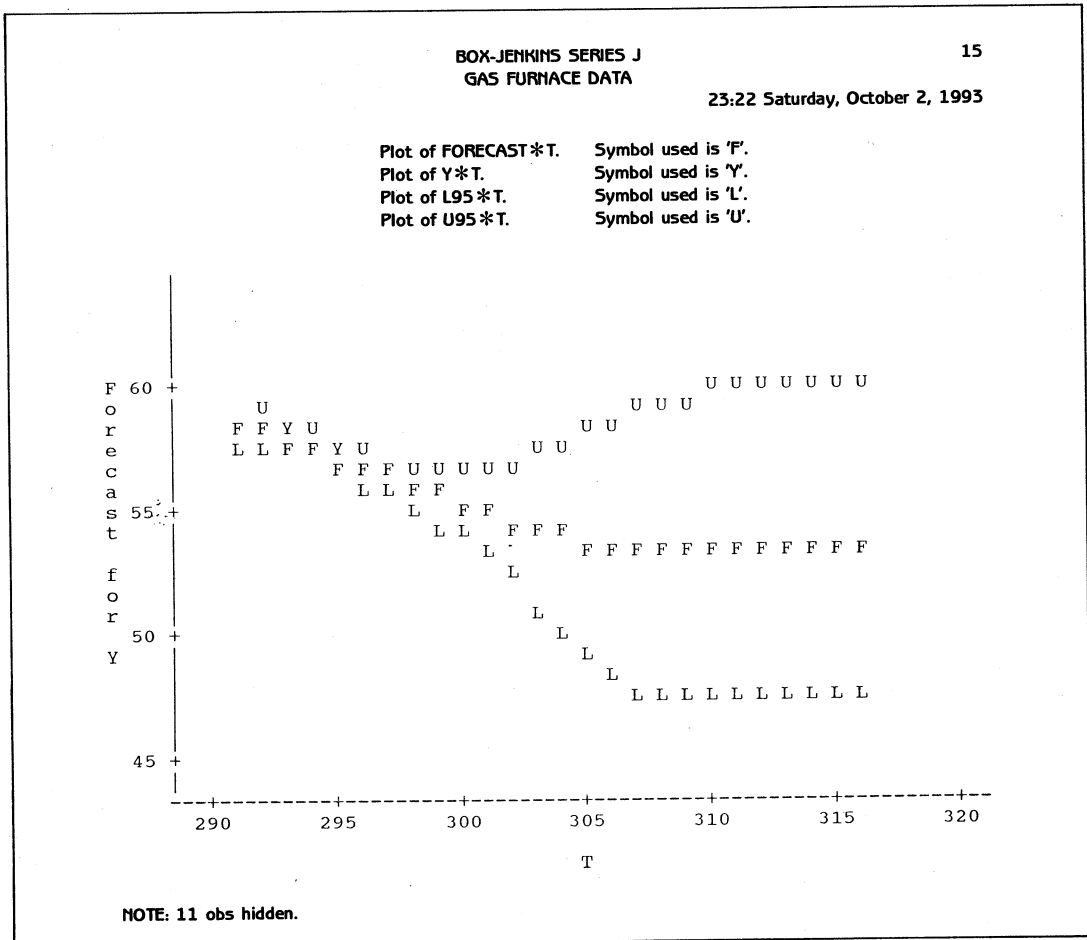


그림 4.24.7 전이함수모형을 사용한 예측에 의한 시계열산포도

이 PROC스텝은 PROC PLOT를 사용해서 SAS데이터세트 EXAM424C에 저장된 출력 시계열데이터의 관찰, 예측, 예측구간의 하한, 그리고 상한의 시계열산포도를 출력하는 것이다. 이 PROC스텝을 수행한 결과가 그림 4.24.7이다.

4.7.2 시물레이션데이터

이 소절에서는 [BD86] (pp. 139~151)에서 분석한 시물레이션데이터를 다루기로 하자. 이 2변량시계열데이터의 입력시계열데이터 X는 다음과 같은 AR(1)모형에서 발생시킨다.

$$x_t - 5 = 0.5(x_{t-1} - 5) + \varepsilon_t$$

여기서 $\{\varepsilon_t\}$ 는 평균이 0이고 분산이 1인 정규백색잡음과정이다. 출력시계열데이터 Y는 다음과 같은 전이함수모형에서 발생시킨다.

$$y_t = 3 \frac{1-0.4B}{1-0.8B} x_{t-2} + \frac{1+0.6B}{1-0.8B} v_t$$

여기서 $\{v_t\}$ 는 $\{\varepsilon_t\}$ 와 독립인 백색잡음과정이다. 이 시계열데이터는 500개의 관찰점들로 구성되어 있다. 이 시계열데이터를 생성한 뒤, 유리함수형태의 전이함수모형화를 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM 4-25.PRG에 수록되어 있다).

```

/ *                               */
/ *                               EXAM4-25.PRG                               */
/ *                               */
/ *                               TRANSFER FUNCTION MODELING                               */
/ *                               DATASET = SIMULATION DATA                               */
/ *                               */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS LINESIZE=72;
OPTIONS PAGESIZE=80;
DATA EXAM425;
  ONE:
  T+1;
  YR=. 8*YR1+3*XR2-1.2*XR3+NORMAL(1234567);
  IF YR=. THEN YR=0;
  XR=. 5*XR1+NORMAL(1234567);
  IF XR=. THEN XR=0;
  X=XR+5;

```

```

Y=YR+10;
OUTPUT;
XR3=XR2;
XR2=XR1;
XR1=XR;
YR1=YR;
IF T<600 THEN GO TO ONE;
RUN;
DATA EXAM425B;
SET EXAM425;
T=T-100;
IF T>0;
RUN;
TITLE1 'SIMULATED TIME SERIES';
TITLE2 'BY BROCKLEBANK & DICKEY(1986)';
PROC ARIMA DATA=EXAM425B;
  I VAR=X;
  E P=1;
  I VAR=Y CROSSCOR=(X);
  E INPUT=(2$(1)/(1) X) MAXIT=30
  ALTPARM PLOT METHOD=ML;
  E P=1 INPUT=(2$(1)/(1)X) ALTPARM METHOD=ML;
  F LEAD=20 OUT=EXAM425C ID=T;
RUN;
PROC PLOT DATA=EXAM425C(FIRSTOBS=491);
  PLOT FORECAST*T='F' Y*T='Y' L95*T='L'
  U95*T='U' / VPOS=25 OVERLAY;
RUN;

```

이 SAS프로그램의 첫 번째와 두 번째 DATA스텝들은 위에서 주어진 모형들에 의해서 입력시계열데이터와 출력시계열데이터를 발생시킨다. 이 발생된 시계열데이터를 SAS데이터세트 EXAM425B에 수록한다.

이 SAS프로그램의 PROC ARIMA의 첫 번째 IDENTIFY스테이트먼트는 다음과 같다.

```
IDENTIFY VAR=X;
```

이 스테이트먼트는 입력시계열데이터 X를 식별하기 위한 것이다. 이 스테이트먼트를 수행한 결과가 그림 4.25.1이다. 그림 4.25.1에서 표본자기상관함수는 시나브로형태를 보이고 표본역자기상관함수와 표본부분자기상관함수는 각각 시차 2 이후에서 절단형태를 보인다. 따라서 입력시계열데이터 X를 AR(1)모형으로 식별한다.

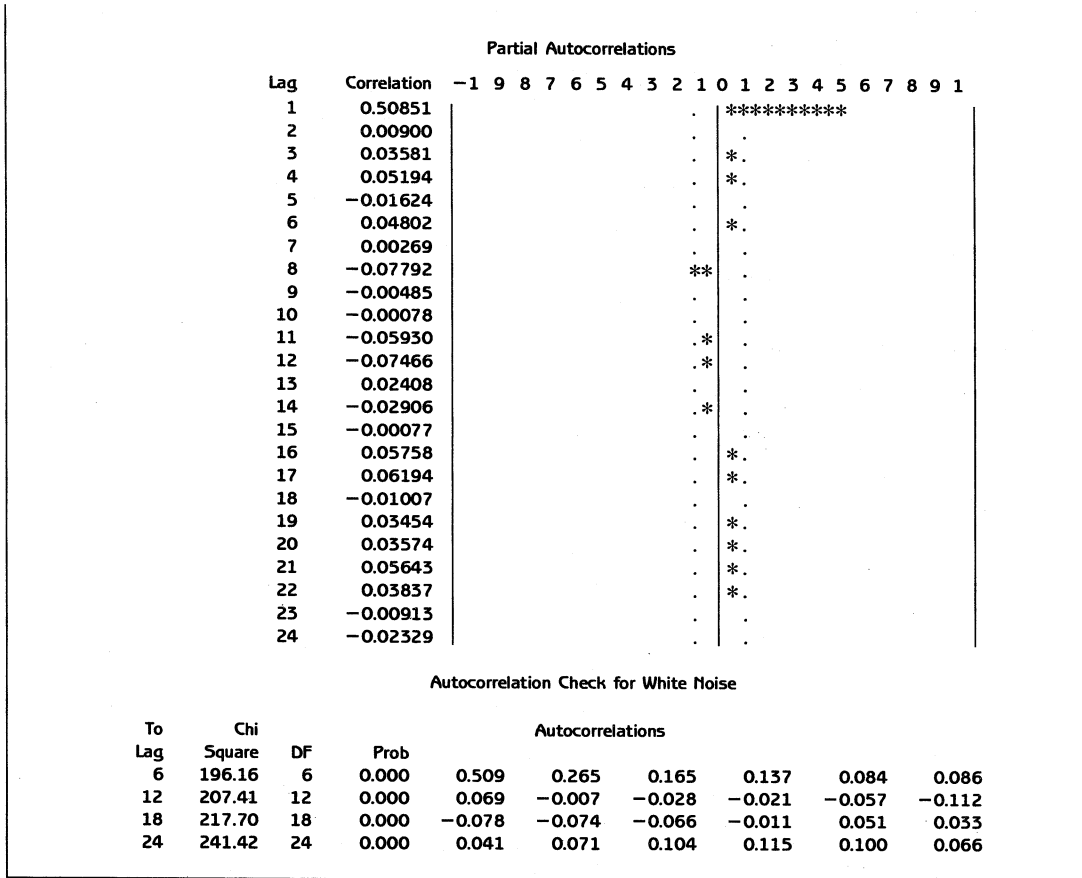


그림 4.25.1 입력시계열데이터의 식별

PROC ARIMA의 첫 번째 ESTIMATE스태이트먼트는 다음과 같다.

ESTIMATE P=1;

이 스타이트먼트는 입력시계열데이터 X를 AR(1)모형에 적합시키기 위한 것이다. 이 스타이트먼트를 수행한 결과가 그림 4.25.2이다. 그림 4.25.2에서 입력시계열데이터의 추정모형이 다음과 같음을 알 수 있다.

$$x_t = 2.4601 + 0.5085x_{t-1} + \epsilon_t$$

이 추정모형에 의한 입력시계열데이터의 잔차값들의 퍼트맨토우통계량값들로부터 추정된 AR(1)모형이 입력시계열데이터와 잘 부합됨을 알 수 있다.

PROC ARIMA의 두 번째 IDENTIFY스태이트먼트는 다음과 같다.

IDENTIFY VAR=Y CROSS=(X);

이 스타이트먼트는 입력시계열데이터가 X이고 출력시계열데이터가 Y인 전이함수를 식별하기 위한 것이다. 이 스타이트먼트를 수행한 결과가 그림 4.25.3이다. 그림 4.25.3에는 사전백색화된 입력시계열데이터와 같은 필터로 여과된 출력시계열데이터의 표본교차상관

22	0.098758	0.02566							*
23	0.228171	0.05928							*
24	0.295370	0.07674							**

"." marks two standard errors

Crosscorrelation Check Between Series

To Lag	Chi Square	DF	Prob	Crosscorrelations									
5	418.85	6	0.000	0.023	0.009	0.793	0.331	0.239	0.202				
11	474.27	12	0.000	0.210	0.131	0.150	0.146	0.066	0.038				
17	479.81	18	0.000	0.055	0.060	-0.035	-0.003	-0.025	-0.051				
23	483.03	24	0.000	-0.027	0.032	0.004	0.022	0.026	0.059				

Both variables have been prewhitened by the following filter:

Prewhitening Filter

Autoregressive Factors
 Factor 1: 1 - 0.50854 B***(1)

그림 4.25.3 전이함수의 식별

함수가 수록되어 있다. 이 표본교차상관함수에서 시차들 2의 표본상관계수가 유의적으로 0과 같지 않고 또한 시차 2 이후의 표본상관계수들이 시나브로형태를 보인다고 할 수 있다. 따라서 이 시계열데이터는 Box-Jenkins방법에 의해서 차수들이 (2, 1, 0), 또는 (2, 1, 1)인 전이함수모형들로 식별할 수 있다. 여기서는 차수들이 (2, 1, 1)인 전이함수모형으로 식별하기로 한다.

PROC ARIMA의 두 번째 ESTIMATE스태이트먼트는 다음과 같다.

```
ESTIMATE INPUT=(2$ (1)/(1) X)
MAXIT=30 ALTPARM PLOT METHOD=ML;
```

이 ESTIMATE스태이트먼트는 잡음시계열데이터의 ARMA모형을 식별하기 위한 것이므로 ESTIMATE스태이트먼트의 옵션 PLOT를 사용하였다. 이 스타이트먼트를 수행한 결과가 그림 4.25.4이다. 그림 4.25.4에서 다음과 같은 추정식을 얻는다.

$$y_t = -33.9977 + 3.0627 \frac{1 - 0.3987B}{1 - 0.7907B} x_{t-2} + \hat{n}_t$$

이 추정모형에 의한 잡음시계열데이터의 적합값들 $\{\hat{n}_t\}$ 의 표본자기상관함수는 시나브로 형태를 보이고, 표본역자기상관함수는 시차 2 이후에서 절단형태를 보이며, 표본부분자기상관함수는 시차 2 이후에서 절단형태를 보인다. 따라서 잡음시계열데이터를 AR(1)모형으로 식별한다.

SIMULATED TIME SERIES 37
 BY BROCKLEBANK & DICKEY (1986)
 7:51 Saturday, October 2, 1993

ARIMA Procedure
 Maximum Likelihood Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag	Variable	Shift
MU	-33.99765	0.77548	-43.84	0	Y	0
SCALE1	3.06271	0.07035	43.53	0	X	2
NUM1,1	0.39865	0.02465	16.17	1	X	2
DEN1,1	0.79069	0.0079342	99.66	1	X	2

Constant Estimate = -33.997654
 Variance Estimate = 2.70215333
 Std Error Estimate = 1.64382278
 AIC = 1908.45106
 SBC = 1925.28542
 Number of Residuals = 497

Correlations of the Estimates

Variable	Parameter	Y	X	X	X
		MU	SCALE1	NUM1,1	DEN1,1
Y	MU	1.000	-0.328	-0.347	-0.634
X	SCALE1	-0.328	1.000	0.689	0.291
X	NUM1,1	-0.347	0.689	1.000	0.821
X	DEN1,1	-0.634	0.291	0.821	1.000

Autocorrelation Check of Residuals

To Lag	Chi Square	DF	Prob	Autocorrelations					
6	731.66	6	0.000	0.780	0.607	0.489	0.365	0.265	0.201
12	751.66	12	0.000	0.165	0.084	0.043	0.048	0.031	0.002
18	771.14	18	0.000	-0.027	-0.047	-0.096	-0.096	-0.095	-0.086
24	784.80	24	0.000	-0.087	-0.071	-0.062	-0.075	-0.059	-0.025
30	802.67	30	0.000	-0.021	-0.019	-0.048	-0.083	-0.111	-0.107
36	862.27	36	0.000	-0.116	-0.141	-0.147	-0.143	-0.142	-0.125
42	892.19	42	0.000	-0.108	-0.093	-0.106	-0.088	-0.090	-0.088

Autocorrelations Plot of Residuals

Lag	Covariance	Correlation	Autocorrelations Plot of Residuals																				
0	2.702153	1.00000	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1
1	2.107649	0.77999	*****																				
2	1.640982	0.60729	*****																				
3	1.321664	0.48912	*****																				
4	0.985110	0.36456	*****																				
5	0.715252	0.26470	*****																				
6	0.543255	0.20105	*****																				
7	0.447150	0.16548	*****																				
8	0.226082	0.08367	*****																				
9	0.115414	0.04271	*****																				
10	0.129205	0.04782	*****																				
11	0.083770	0.03100	*****																				
12	0.0044361	0.00164	*****																				
13	-0.073574	-0.02723	*****																				
14	-0.127236	-0.04709	*****																				
15	-0.258879	-0.09580	*****																				
16	-0.259761	-0.09613	*****																				
17	-0.257315	-0.09523	*****																				
18	-0.231280	-0.08559	*****																				
19	-0.235694	-0.08722	*****																				
20	-0.192890	-0.07138	*****																				
21	-0.166829	-0.06174	*****																				
22	-0.202506	-0.07494	*****																				
23	-0.159551	-0.05905	*****																				
24	-0.066475	-0.02460	*****																				

"." marks two standard errors

Inverse Autocorrelations

Lag	Correlation	Inverse Autocorrelations																				
1	-0.51566	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1
2	0.08327	*****																				
3	-0.11757	*****																				
4	0.08143	*****																				
5	-0.03172	*****																				
6	0.07093	*****																				

7	-0.12684	***	.
8	0.05891	.	*
9	0.06996	.	*
10	-0.06601	.	*
11	0.02191	.	.
12	-0.05976	.	*
13	0.07577	.	*
14	-0.08846	**	.
15	0.12159	.	*
16	-0.10937	**	.
17	0.08898	.	*
18	-0.06655	.	*
19	0.05656	.	*
20	-0.03798	.	*
21	-0.01557	.	.
22	0.01789	.	.
23	0.03962	.	*
24	-0.03402	.	*

Partial Autocorrelations

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
1	0.77999	*****
2	-0.00280
3	0.04161
4	-0.07432
5	-0.01456
6	0.02003
7	0.03808
8	-0.13172
9	0.03341
10	0.06932
11	-0.02508
12	-0.04722
13	-0.04741
14	-0.01045
15	-0.07563
16	0.06077
17	-0.03741
18	0.04020
19	-0.03639
20	0.03207
21	-0.02876
22	-0.03333
23	0.02596
24	0.05989

Crosscorrelation Check of Residuals with Input X

To Lag	Chi Square	DF	Prob	Crosscorrelations						
5	0.57	4	0.967	-0.031	0.008	-0.004	-0.007	-0.006	0.001	
11	1.93	10	0.997	0.007	-0.009	0.004	-0.002	-0.018	0.048	
17	7.25	16	0.968	0.056	0.042	0.037	0.043	0.044	0.027	
23	15.71	22	0.830	0.032	0.072	0.055	0.058	0.057	0.035	
29	16.35	28	0.960	0.019	-0.003	0.016	0.012	-0.019	0.013	
35	19.87	34	0.974	-0.075	-0.022	-0.002	0.029	-0.013	0.003	
41	23.89	40	0.980	0.002	0.002	-0.019	0.026	0.070	0.047	

Model for variable Y

Estimated Intercept = -33.997654

Input Number 1 is X with a shift of 2.

Overall Regression Factor = 3.06271

The Numerator Factors are

Factor 1: 1 - 0.39865 B**(1)

The Denominator Factors are

Factor 1: 1 - 0.79069 B**(1)

그림 4.25.4 입력시계열데이터의 식별

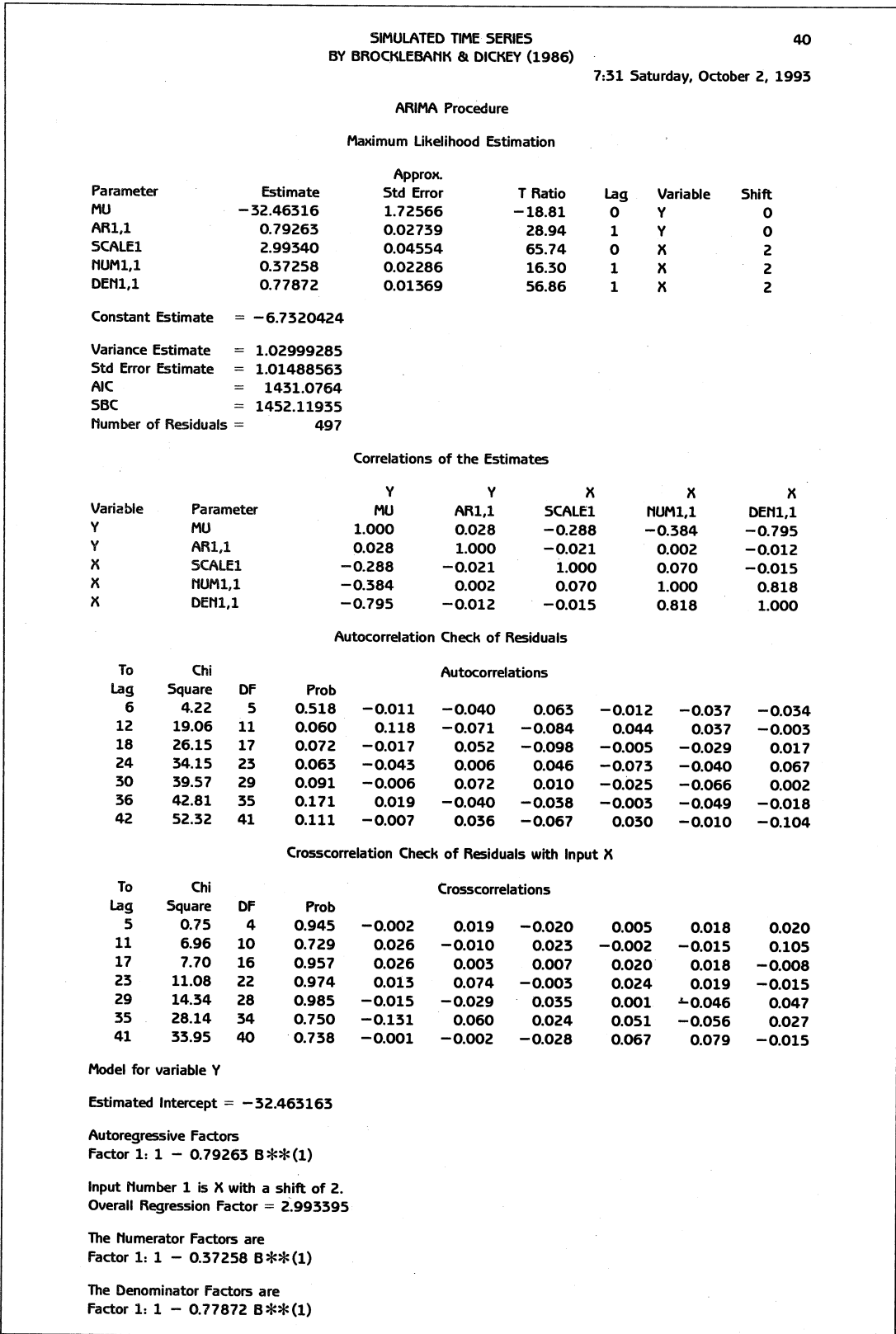


그림 4.25.5 전이함수모형의 추정

SAS프로그램의 PROC ARIMA의 세 번째 ESTIMATE스테이트먼트는 다음과 같다.

```
ESTIMATE P=1 INPUT=(2$ (1)/(1) X)
ALTPARM METHOD=ML;
```

이 스테이트먼트는 출력시계열데이터 Y를 전이함수모형에 적합시키기 위한 것이다. 이 스테이트먼트를 수행한 결과가 그림 4.25.5이다. 그림 4.25.5에서 출력시계열데이터의 전이함수모형의 추정식이 다음과 같음을 알 수 있다.

$$y_t = -6.7320 + 2.9934 \frac{1-0.3726B}{1-0.7787B} x_{t-2} + \frac{1}{1-0.7926B} v_t$$

잡음시계열모형의 잔차 $\{\hat{v}_t\}$ 와 사전백색화된 입력시계열데이터 $\{z_t\}$ 의 표본교차상관함수에 의한 퍼트맨토우통계량값들로부터 잡음시계열데이터의 오차항과 입력시계열데이터가 서로 독립이라는 가설을 채택한다. 추정된 전이함수모형의 잡음시계열모형의 잔차 $\{\hat{v}_t\}$ 의 퍼트맨토우통계량값들로부터 추정된 전이함수모형이 주어진 시계열데이터와 잘 부합됨을 알 수 있다.

이 PROC ARIMA의 첫 번째 FORECAST스테이트먼트는 다음과 같다.

```
FORECAST LEAD=20 OUT=EXAM425C ID=T;
```

이 FORECAST스테이트먼트는 바로 앞의 ESTIMATE스테이트먼트에서 추정된 전이함수모형을 이용해서 출력시계열데이터의 미래값들을 예측하기 위한 것이다. 이 FORECAST스테이트먼트를 수행한 결과가 그림 4.25.6이다. 그림 4.25.6의 첫 부분에는 예측에 사용된 전이함수모형이 수록되어 있다. 이 전이함수모형은 그림 4.25.5에서 알 수 있듯이 최종적으로 추정된 전이함수모형이다. 그림 4.25.6의 뒷부분에는 예측값(Forecast), 표준오차(Std Error), 신뢰도 95%의 예측구간의 하한값(Lower 95%)과 상한값(Upper 95%)이 수록되어 있다. 이들은 각각 변수들 FORECAST, STD, L95와 U95에 저장되어 있다. 이 PROC ARIMA의 FORECAST스테이트먼트의 옵션 OUT=에 SAS데이터세트 EXAM425C를 할당하였다. 이 SAS데이터세트에는 FORECAST스테이트먼트를 수행한 결과로 얻어진 예측의 결과가 수록된다. 이 SAS데이터세트에는 ID변수 T, 출력시계열데이터, FORECAST, STD, L95, U95, 그리고 RESIDUAL이 수록되어 있다.

이 SAS프로그램의 두 번째 PROC스텝은 다음과 같다.

```
PROC PLOT DATA = EXAM425C(FIRSTOBS = 491);
PLOT FORECAST*T = 'F' Y*T = 'Y' L95*T = 'L'
U95*T = 'U' / VPOS = 25 OVERLAY;
RUN;
```

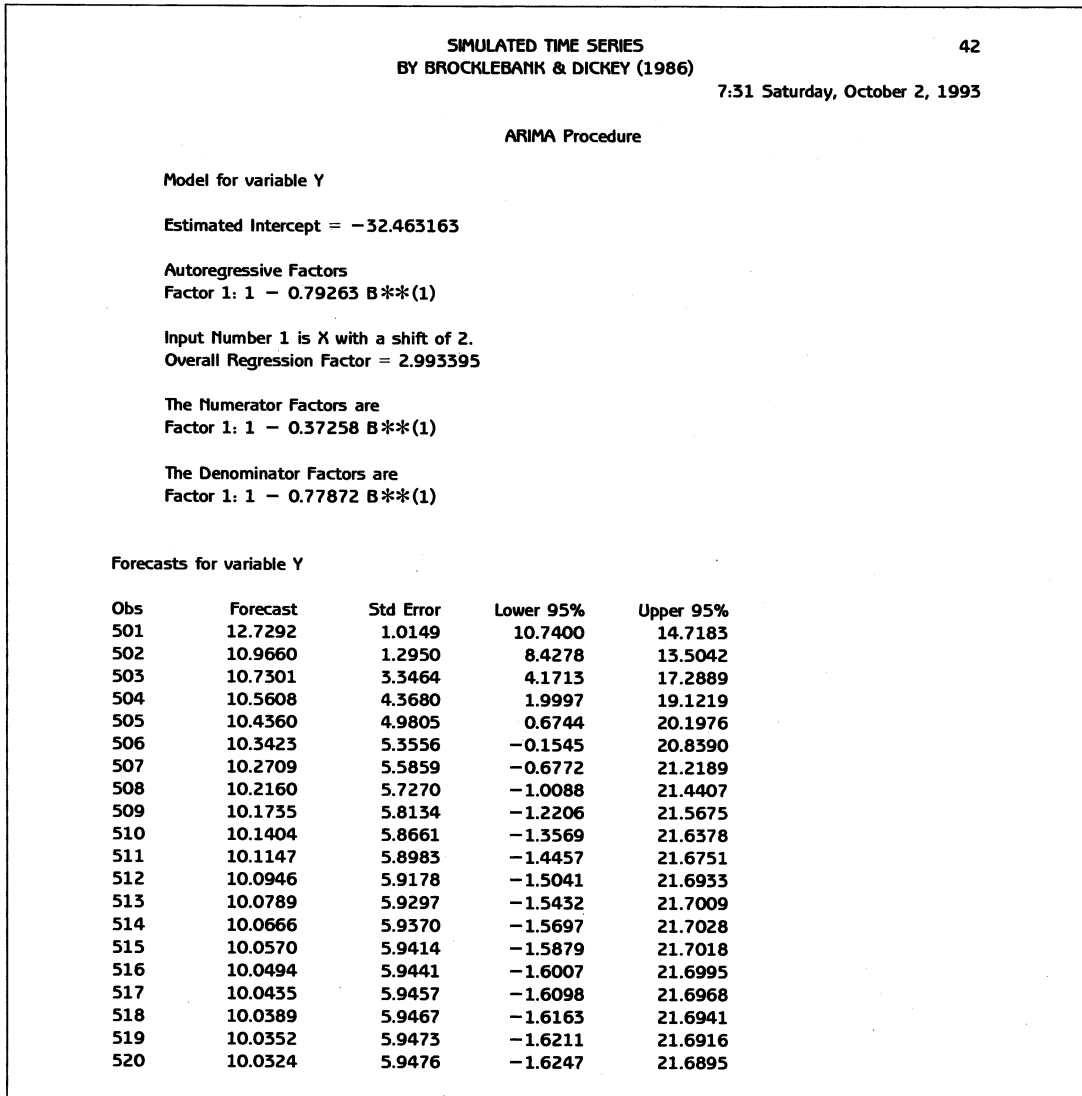


그림 4.25.6 전이함수모형을 사용한 예측

이 PROC스텝은 PROC PLOT를 사용해서 SAS데이터세트 EXAM425C에 저장된 출력 시계열데이터의 관찰, 예측, 예측구간의 하한, 그리고 상한의 시계열산포도를 출력하기 위한 것이다. 이 PROC스텝을 수행한 결과가 그림 4.25.7이다.

4.7.3 다변량시계열데이터 RIVER

이 소절에서는 시계열데이터 RIVER를 유리함수형태의 전이함수모형화 한다. 이 2변량 시계열데이터는 미국 North Carolina 주의 Neuse 강변의 도시 Goldsboro에서의 유속 (flow rates)과 이 도시로부터 30마일 하류에 있는 도시 Kinston에서의 유속을 일별(daily)로 기록한 것이다. 당연히 Goldsboro에서의 유속을 나타내는 GOLD가 입력시계열데이

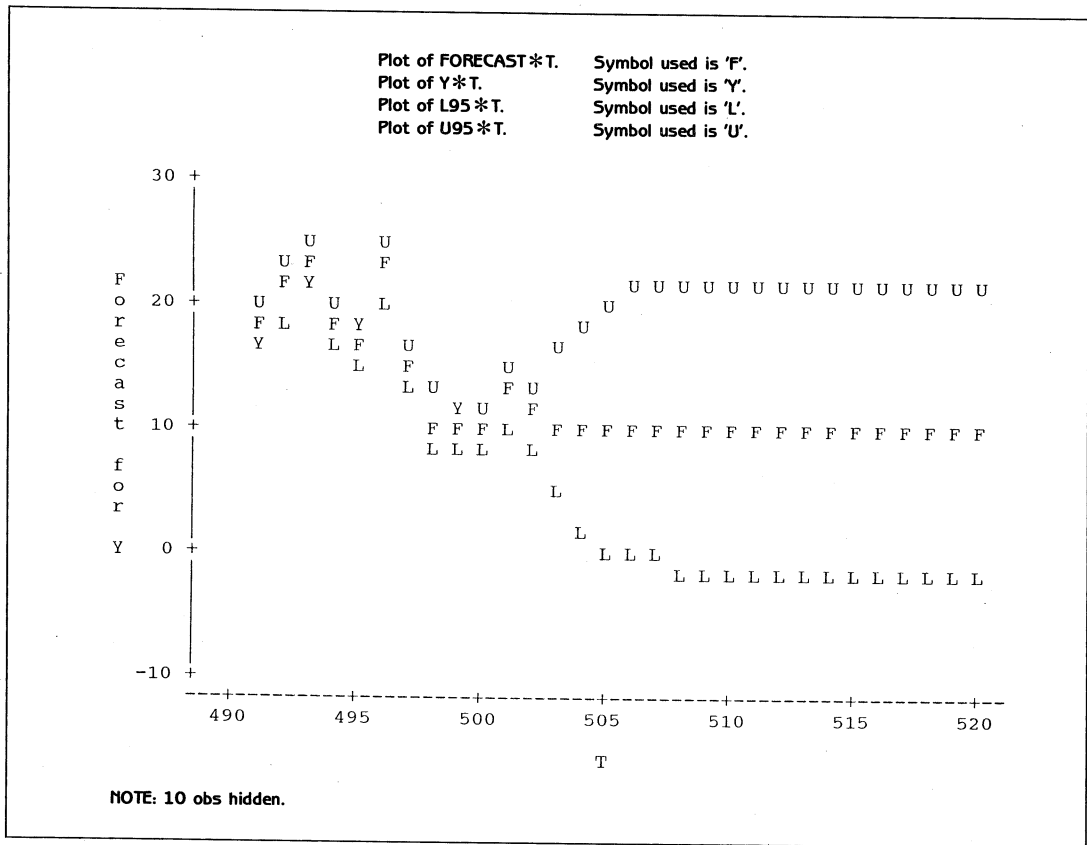


그림 4.25.7 전이함수모형을 이용한 예측값들의 시계열산포도

터이고 Kinston에서의 유속을 나타내는 KINS가 출력시계열데이터이다. 이 시계열데이터는 400개의 관찰점들로 구성되어 있다. 이 시계열데이터는 제 2.9절에 수록되어 있으며, 또한 디스켓 SAS4TSA4의 SAS데이터라이브러리 TSA4DATA의 파일 RIVER.SSD에 저장되어 있다. 시계열데이터 RIVER의 시계열산포도가 그림 2.9에 수록되어 있다. 이 시계열데이터를 [BD86] (pp. 151~159)에서 전이함수모형을 사용해서 분석하였고, 이 소절에서는 이 분석을 따르고자 한다.

시계열데이터 RIVER를 유리함수형태의 전이함수모형화하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM4-26.PRG에 수록되어 있다).

```

/*                               */
/*                               EXAM4-26.PRG                               */
/*                               */
/*                               TRANSFER FUNCTION MODELING                */
/*                               DATASET = RIVER                            */
/*                               */
LIBNAME TSA4DATA 'A:\TSA4DATA';
    
```

```

OPTIONS LINESIZE=72;
OPTIONS PAGESIZE=80;
DATA EXAM426;
    SET TSA4DATA.RIVER;
    LGOLD=LOG(GOLD);
    LKINS=LOG(KINS);
    T=_N_;
    RUN;
TITLE 'FLOW RATES OF NEUSE RIVER';
TITLE2 'AT GOLDSBORO AND KINSTON';
PROC ARIMA DATA=EXAM426;
    I VAR=LGOLD(1);
    E Q=1 P=3 METHOD=ML NOCONSTANT MAXIT=100;
    I VAR=LKINS(1) CROSSCOR=(LGOLD(1));
    E INPUT=(1$(1) LGOLD) METHOD=ML PLOT;
    E P=2 Q=1 INPUT=(1$(1) LGOLD) METHOD=ML
    NOCONSTANT ALTPARM;
    F LEAD=20 ID=T OUT=EXAM426B;
    RUN;
PROC PLOT DATA=EXAM426B(FIRSTOBS=391);
    PLOT FORECAST*T='F' LKINS*T='Y' L95*T='L'
        U95*T='U' / VPOS=26 OVERLAY;
    RUN;

```

그림 2.9에서 볼 수 있듯이 입력시계열데이터 GOLD와 출력시계열데이터 KINS는 시간이 흐름에 따라 분산이 증가되는 분산이질성을 가지고 있다. 따라서 이 SAS프로그램의 첫 번째 DATA스텝에서 각각을 대수변환해서 새로운 입력시계열데이터 LGOLD와 출력시계열데이터 LKINS를 만들었다. 이후 이 대수변환된 시계열데이터를 분석하기로 한다.

PROC ARIMA의 첫 번째 IDENTIFY스테이트먼트는 다음과 같다.

```
IDENTIFY VAR=LGOLD(1);
```

입력시계열데이터 LGOLD와 출력시계열데이터 LKINS는 당연히 계절주기 365를 가지므로 자기상관함수는 아주 느린 속도로 감소할 것이다. 따라서 입력시계열데이터 LGOLD와 출력시계열데이터 LKINS의 차분들을 취한 차분시계열데이터들이 정상성조건을 만족할 것이다. 이 IDENTIFY스테이트먼트에서는 차분된 입력시계열데이터를 식별하기 위해서 옵션 VAR=에 LGOLD(1)을 할당하였다. 이 스테이트먼트를 수행한 결과가 그림 4.26.1이다. 그림 4.26.1에서 표본자기상관함수는 시나브로형태를 보이며, 표본역자기상관함수와 표본부분자기상관함수는 시차 2 이후에서 절단형태를 보인다고 할 수 있다. 따라

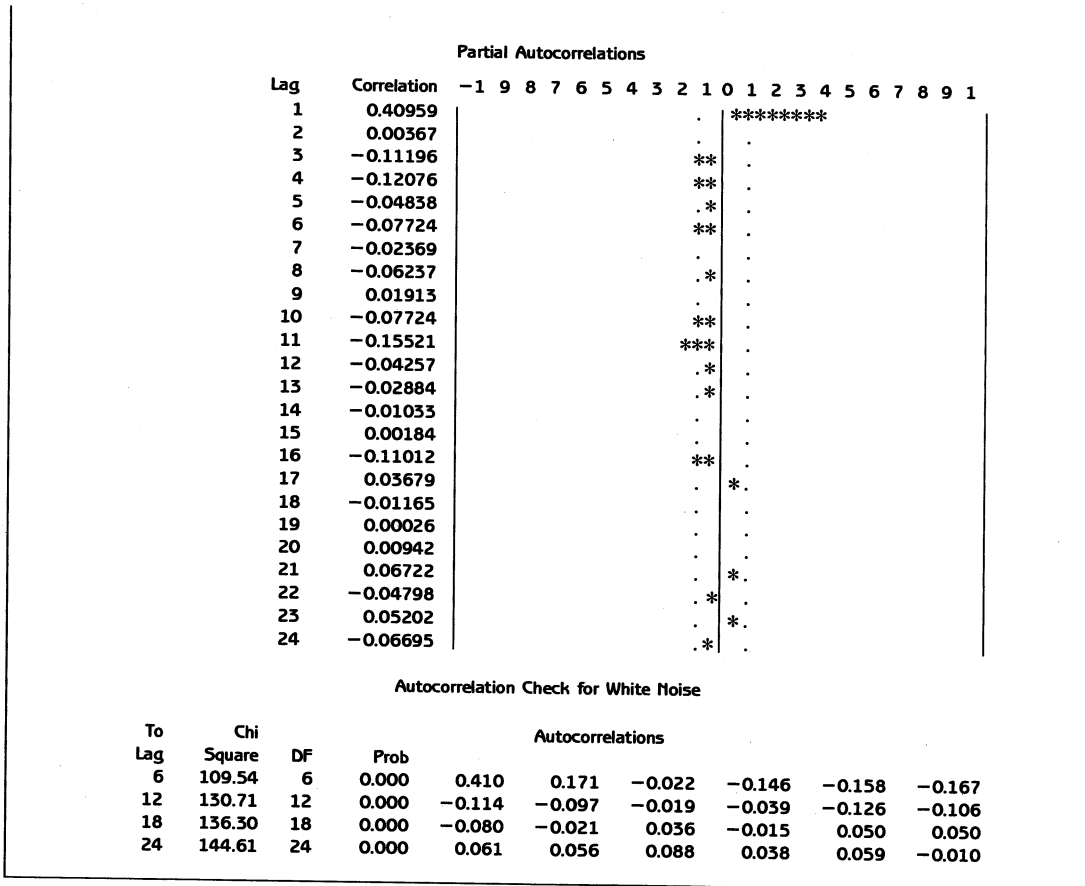


그림 4.26.1 입력시계열데이터의 식별

서 차분된 입력시계열데이터 LGOLD(1)은 AR(1)모형으로 식별할 수 있다. 그러나 시차들 4, 8, 16 등에서의 표본역자기상관계수들이 유의적으로 0과 다르고, 시차 3, 4, 12 등에서의 표본부분자기상관계수들이 유의적으로 0과 다르다. 즉, 표본자기상관함수, 표본역자기상관함수와 표본부분자기상관함수 모두가 시나브로형태를 보인다고 할 수 있으므로 차분된 입력시계열데이터 LGOLD(1)은 혼합ARMA모형을 갖는다고 할 수도 있다. 따라서 이 입력시계열데이터를 식별하기 위해서는 다른 ARMA모형 식별방법들을 사용해야 한다. [BD86] (p. 152)은 이 차분된 입력시계열데이터 LGOLD(1)을 ARMA(3, 1)모형으로 식별하였다.

PROC ARIMA의 첫 번째 ESTIMATE스테이트먼트는 다음과 같다.

```
ESTIMATE Q=1 P=3 METHOD=ML NOCONSTANT MAXIT=100;
```

이 스테이트먼트는 차분된 입력시계열데이터 LGOLD(1)을 ARMA(3, 1)모형에 적합시키기 위한 것이다. 이 스테이트먼트를 수행한 결과가 그림 4.26.2이다. 그림 4.26.2에서 입력시계열데이터의 추정모형이 다음과 같음을 알 수 있다.

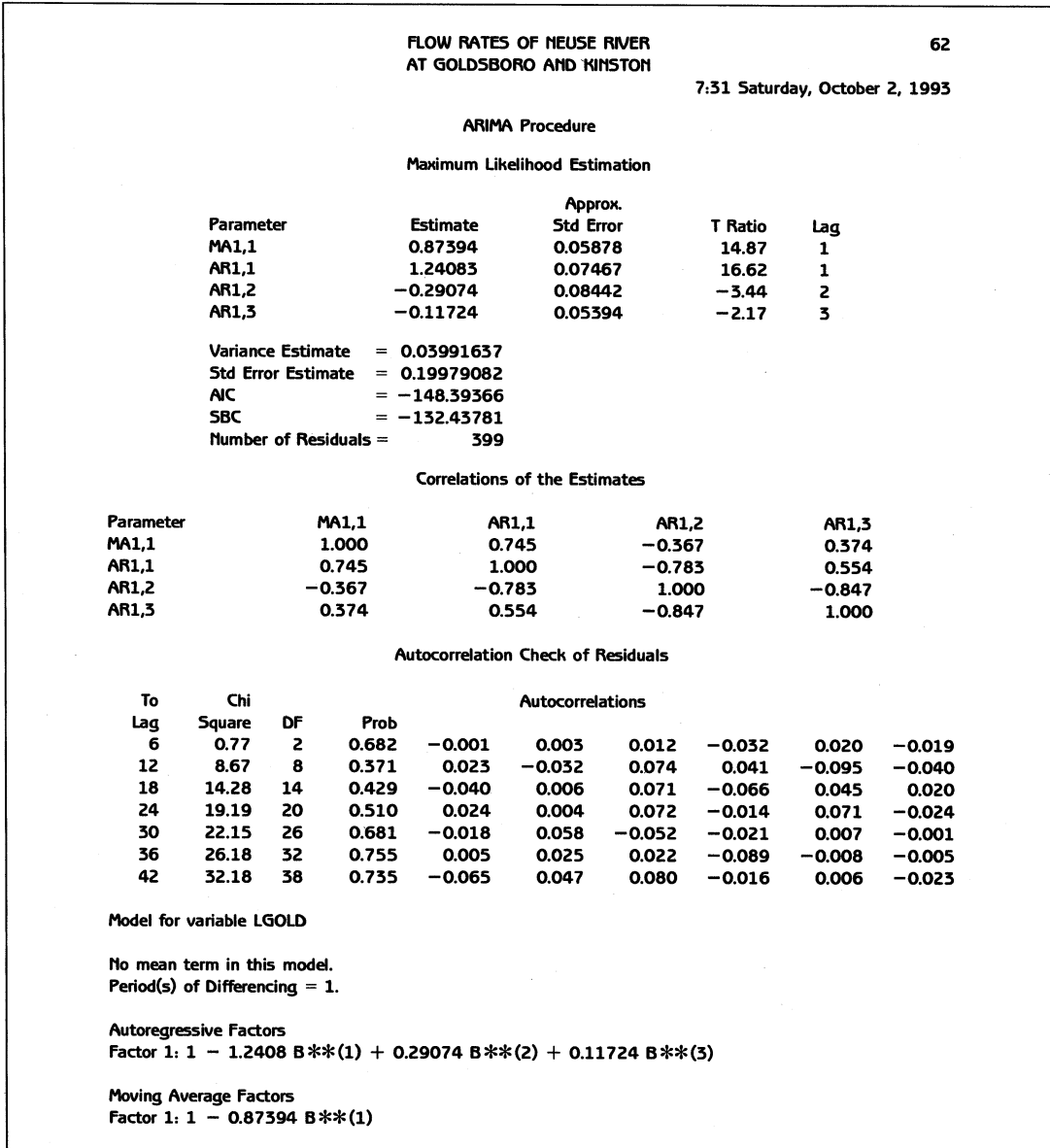


그림 4.26.2 입력시계열데이터의 추정

$$x_t = \text{LGOLD}_t - \text{LGOLD}_{t-1}$$

$$x_t = \frac{1 - 0.8739B}{1 - 1.2408B + 0.2907B^2 + 0.1172B^3} u_t$$

차분된 입력시계열데이터의 추정모형에 의한 잔차값들의 퍼트맨토우통계량값들로부터 추정된 ARMA(3, 1)모형이 입력시계열데이터와 잘 부합됨을 알 수 있다.

PROC ARIMA의 두 번째 IDENTIFY스태이트먼트는 다음과 같다.

IDENTIFY VAR=LKINS(1) CROSSCOR=(LGOLD(1));

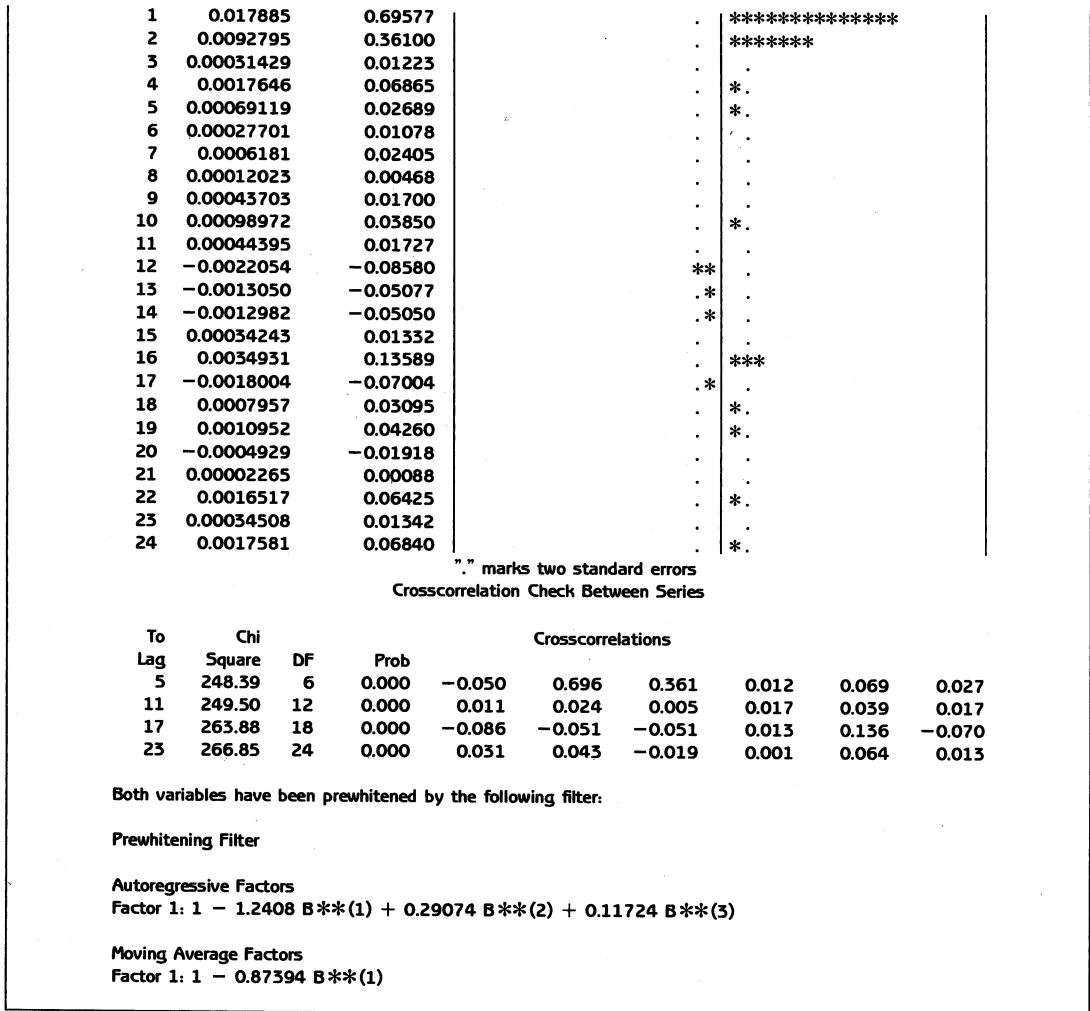


그림 4.26.3 전이함수의 식별

이 스테이트먼트는 입력시계열데이터가 LGOLD(1)이고 출력시계열데이터가 LKINS(1)인 전이함수를 식별하기 위한 것이다. 이 스테이트먼트를 수행한 결과가 그림 4.26.3이다. 그림 4.26.3에 사전백색화된 입력시계열데이터와 같은 필터로 여과된 출력시계열데이터의 표본교차상관함수가 수록되어 있다. 이 표본교차상관함수에서 시차 1과 시차 2에서의 표본상관계수들이 유의적으로 0과 같지 않고, 또한 시차 3 이후의 표본상관계수들이 절단형태를 보인다고 할 수 있다. 따라서 이 시계열데이터는 Box-Jenkins방법에 의해서 차수들이 (1, 0, 1)인 전이함수모형으로 식별할 수 있다.

PROC ARIMA의 두 번째 ESTIMATE스테이트먼트는 다음과 같다.

ESTIMATE INPUT=(1\$(1) LGOLD) METHOD=ML PLOT;

이 스테이트먼트를 수행한 결과가 그림 4.26.4이다. 그림 4.26.4에서 다음과 같은 추정식을 얻는다.

FLOW RATES OF NEUSE RIVER
AT GOLDSBORO AND KINSTON

67

7:31 Saturday, October 2, 1993

ARIMA Procedure

Maximum Likelihood Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag	Variable	Shift
MU	0.0018976	0.0043054	0.44	0	LKINS	0
NUM1	0.43109	0.02107	20.46	0	LGOLD	1
NUM1,1	-0.22837	0.02106	-10.84	1	LGOLD	1

Constant Estimate = 0.00189759

Variance Estimate = 0.00735015

Std Error Estimate = 0.085733

AIC = -820.84902

SBC = -808.89721

Number of Residuals = 397

Correlations of the Estimates

Variable	Parameter	LKINS MU	LGOLD NUM1	LGOLD NUM1,1
LKINS	MU	1.000	-0.018	0.020
LGOLD	NUM1	-0.018	1.000	0.410
LGOLD	NUM1,1	0.020	0.410	1.000

Autocorrelation Check of Residuals

To Lag	Chi Square	DF	Prob	Autocorrelations						
6	76.40	6	0.000	0.315	-0.088	-0.195	-0.169	-0.106	-0.073	
12	87.15	12	0.000	-0.105	-0.081	0.062	0.042	0.054	-0.013	
18	97.51	18	0.000	-0.023	-0.005	-0.033	-0.140	-0.010	0.059	
24	114.97	24	0.000	0.092	0.150	0.059	0.036	-0.033	-0.069	
30	120.35	30	0.000	-0.044	-0.053	-0.056	-0.062	-0.023	-0.019	
36	131.49	36	0.000	0.031	0.042	0.087	0.099	0.063	-0.039	
42	137.18	42	0.000	-0.049	-0.042	-0.041	-0.004	-0.048	-0.068	

Autocorrelations Plot of Residuals

Lag	Covariance	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
0	0.0073501	1.00000												*****										
1	0.0023175	0.31531												*****										
2	-0.0006451	-0.08776												**
3	-0.0014358	-0.19534												***
4	-0.0012412	-0.16886												***
5	-0.0007808	-0.10623												**
6	-0.000535	-0.07279												.	*
7	-0.0007739	-0.10530												**
8	-0.0005976	-0.08130												**
9	0.00045435	0.06182												.	*
10	0.00031116	0.04233												.	*
11	0.00039753	0.05408												.	*
12	-0.0000927	-0.01261											
13	-0.00017	-0.02312											
14	-0.00004	-0.00545											
15	-0.0002424	-0.03298												.	*
16	-0.0010321	-0.14042												***
17	-0.0000751	-0.01021											
18	0.00043128	0.05868												.	*
19	0.00067524	0.09187												.	**
20	0.0011004	0.14972												.	***
21	0.00043052	0.05857												.	*
22	0.00026751	0.03640												.	*
23	-0.0002427	-0.03301												.	*
24	-0.0005082	-0.06914												.	*

." marks two standard errors

Inverse Autocorrelations

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1		
1	-0.28165									*****														
2	0.20889												*****											
3	0.12155													**										
4	0.07801													**										
5	0.10685													**										
6	0.05649													*										
7	0.03509													*										
8	0.17435													*										
9	-0.06121													*										
10	0.09295													*										
11	-0.00055													*										
12	0.03401													*										
13	0.03761													*										
14	0.01513													*										
15	-0.07029													*										
16	0.15677													*										
17	-0.09352													*										
18	0.00402													*										
19	0.02305													*										
20	-0.09855													*										
21	0.06347													*										
22	-0.07294													*										
23	0.02384													*										
24	0.00998													*										

Partial Autocorrelations

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1		
1	0.31531												*****											
2	-0.20784												****											
3	-0.11186												**											
4	-0.09195												**											
5	-0.07202												*											
6	-0.08240												**											
7	-0.13441												**											
8	-0.07821												**											
9	0.04706												*											
10	-0.08698												**											
11	0.01795												*											
12	-0.07777												**											
13	-0.01386												*											
14	-0.01931												*											
15	-0.07046												*											
16	-0.15748												***											
17	0.07068												*											
18	-0.03508												*											
19	0.03301												*											
20	0.08679												**											
21	-0.00999												*											
22	0.07693												**											
23	-0.03975												*											
24	-0.01307												*											

Crosscorrelation Check of Residuals with Input LGOLD

To Lag	Chi Square	DF	Prob	Crosscorrelations																					
5	4.08	5	0.537	0.011	-0.012	-0.010	0.071	0.065	0.027																
11	11.11	11	0.434	0.036	-0.007	0.019	-0.008	-0.076	-0.101																
17	21.14	17	0.220	-0.033	-0.025	0.027	0.140	-0.007	0.056																
23	24.04	23	0.401	0.036	-0.048	-0.041	-0.015	-0.002	0.042																
29	27.77	29	0.530	-0.024	-0.065	-0.063	-0.014	-0.006	0.020																
35	33.12	35	0.559	0.023	0.007	0.011	-0.063	0.016	0.093																
41	44.70	41	0.319	0.126	0.045	-0.038	-0.067	-0.047	0.055																

Model for variable LKINS

Estimated Intercept = 0.00189759

Period(s) of Differencing = 1.

Input Number 1 is LGOLD with a shift of 1.
Period(s) of Differencing = 1.

The Numerator Factors are
Factor 1: 0.4311 + 0.22837 B**(1)

그림 4.26.4 입력시계열데이터의 식별

$$y_t = \text{LKINS}_t - \text{LKINS}_{t-1}$$

$$y_t = 0.0019 + (0.4311 + 0.2284B) x_{t-1} + \hat{n}_t$$

이 추정모형에 의한 잡음시계열데이터의 적합값들 $\{\hat{n}_t\}$ 의 표본자기상관함수, 표본역자기상관함수, 그리고 표본부분자기상관함수가 모두 시나브로형태를 보이므로 Box-Jenkins 방법으로는 잡음항모형을 식별할 수가 없다. 따라서 다른 식별방법들을 사용해야 할 것이다. [BD86] (p. 158)은 이 잡음시계열데이터를 ARMA(2, 1)모형으로 식별하였다.

PROC ARIMA의 세 번째 ESTIMATE스테이트먼트는 다음과 같다.

```
ESTIMATE P=2 Q=1 INPUT=(1$(1) LGOLD)
METHOD=ML NOCONSTANT ALTPARM;
```

이 스테이트먼트는 출력시계열데이터 LKINS(1)을 전이함수모형에 적합시키기 위한 것이다. 이 스테이트먼트를 수행한 결과가 그림 4.26.5이다. 그림 4.26.5에서 출력시계열데이터의 전이함수모형의 추정식이 다음과 같음을 알 수 있다.

$$y_t = 0.4954(1 + 0.5503B) x_{t-1} + \frac{1 - 0.8878B}{1 - 1.1632B + 0.4796B^2} v_t$$

잡음시계열데이터의 잔차값들 $\{\hat{v}_t\}$ 와 사전백색화된 입력시계열데이터의 표본교차상관함수에 의한 퍼트맨트우통계량값들로부터 잡음시계열모형의 오차항과 입력시계열데이터가 서로 독립이라는 가설을 채택한다. 추정된 전이함수모형의 잡음시계열모형의 잔차값들의 퍼트맨트우통계량값들로부터 추정된 전이함수모형이 주어진 시계열데이터와 잘 부합됨을 알 수 있다.

PROC ARIMA의 첫 번째 FORECAST스테이트먼트는 다음과 같다.

```
FORECAST LEAD=20 ID=T OUT=EXAM426B;
```

이 FORECAST스테이트먼트는 바로 앞의 ESTIMATE스테이트먼트에서 추정된 전이함수모형을 이용해서 출력시계열데이터의 미래값들을 예측하는 것이다. 이 FORECAST스테이트먼트를 수행한 결과가 그림 4.26.6이다. 그림 4.26.6의 첫 부분에는 예측에 사용된 전이함수모형이 수록되어 있다. 이 전이함수모형은 그림 4.26.5에서 알 수 있듯이 최종적으로 추정된 전이함수모형이다. 그림 4.26.6의 뒷부분에는 예측값(Forecast), 표준오차(Std Error), 신뢰도 95%의 예측구간의 하한값(Lower 95%)과 상한값(Upper 95%)이 수록되어 있다. 이들은 각각 변수들 FORECAST, STD, L95와 U95에 저장되어 있다. 이 FORE-

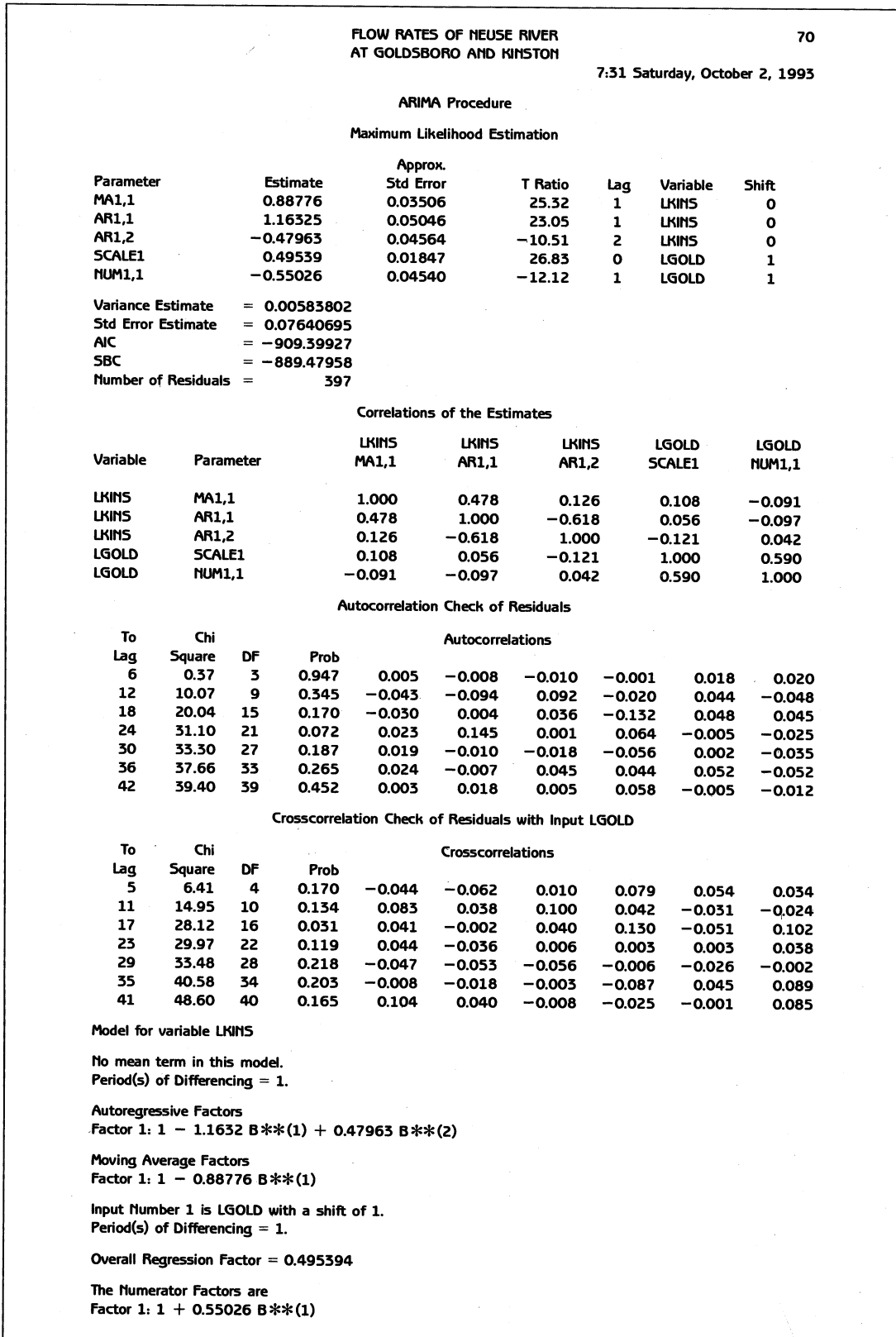


그림 4.26.5 전이함수모형의 추정

FLOW RATES OF NEUSE RIVER AT GOLDSBORO AND KINSTON		72		
7:31 Saturday, October 2, 1993				
ARIMA Procedure				
Model for variable LKINS				
No mean term in this model.				
Period(s) of Differencing = 1.				
Autoregressive Factors				
Factor 1: 1 - 1.1632 B**(1) + 0.47963 B**(2)				
Moving Average Factors				
Factor 1: 1 - 0.88776 B**(1)				
Input Number 1 is LGOLD with a shift of 1.				
Period(s) of Differencing = 1.				
Overall Regression Factor = 0.495394				
The Numerator Factors are				
Factor 1: 1 + 0.55026 B**(1)				
Forecasts for variable LKINS				
Obs	Forecast	Std Error	Lower 95%	Upper 95%
401	9.0139	0.0764	8.8642	9.1637
402	8.9879	0.1585	8.6772	9.2986
403	8.9516	0.2616	8.4390	9.4642
404	8.9157	0.3510	8.2277	9.6037
405	8.8848	0.4231	8.0556	9.7141
406	8.8611	0.4777	7.9248	9.7975
407	8.8447	0.5177	7.8299	9.8594
408	8.8346	0.5467	7.7631	9.9060
409	8.8295	0.5678	7.7165	9.9424
410	8.8279	0.5840	7.6833	9.9725
411	8.8287	0.5970	7.6586	9.9989
412	8.8308	0.6082	7.6387	10.0229
413	8.8333	0.6185	7.6211	10.0454
414	8.8357	0.6282	7.6044	10.0671
415	8.8378	0.6379	7.5876	10.0881
416	8.8394	0.6476	7.5701	10.1088
417	8.8405	0.6575	7.5518	10.1292
418	8.8411	0.6676	7.5327	10.1495
419	8.8414	0.6777	7.5131	10.1697
420	8.8414	0.6879	7.4931	10.1897

그림 4.26.6 전이함수모형을 사용한 예측

CAST스태이트먼트의 옵션 OUT=에 SAS데이터세트 EXAM426B를 할당하였다. 이 SAS 데이터세트에는 FORECAST스태이트먼트를 수행한 결과로 얻어진 예측의 결과가 수록된다. 즉, 이 SAS데이터세트에는 ID변수 T, 출력시계열데이터, FORECAST, STD, L95, U95, 그리고 RESIDUAL이 수록되어 있다.

유의할 점은 입력시계열데이터 GOLD와 출력시계열데이터 KINS가 갖는 분산이질성을 해결하기 위해서 이 시계열데이터를 대수변환해서 새로운 입력시계열데이터 LGOLD와 출력시계열데이터 LKINS를 만들었다는 점이다. 따라서 입력시계열데이터 GOLD와 출력시계열데이터 KINS의 예측값들을 얻기 위해서는 대수변환된 입력시계열데이터 LGOLD와 출력시계열데이터 LKINS의 예측값들을 역대수변환해야 한다. 그러나 대수변환된 시계열데이터의 최소평균제곱오차예측을 역대수변환한 것이 원래의 시계열데이터의 최소평

관계잡오차예측은 아니다. 따라서 입력시계열데이터 GOLD와 출력시계열데이터 KINS의 예측값들을 얻기 위해서는 대수변환된 입력시계열데이터 LGOLD와 출력시계열데이터 LKINS의 예측값들을 변형된 역대수변환을 해야 한다. 자세한 내용은 [SAS4TSA2] (pp. 609~610)를 참조하라.

이 SAS프로그램의 두 번째 PROC스텝은 다음과 같다.

```
PROC PLOT DATA=EXAM426B(FIRSTOBS=391);
  PLOT FORECAST*T='F' LKINS*T='Y' L95*T='L'
      U95*T='U' / VPOS=26 OVERLAY;
RUN;
```

이 PROC스텝은 PROC PLOT를 사용해서 SAS데이터세트 EXAM426B에 저장된 출력시계열데이터의 관찰, 예측, 예측구간의 하한과 상한의 시계열산포도를 출력하는 것이다. 이 PROC스텝을 수행한 결과가 그림 4.26.7이다.

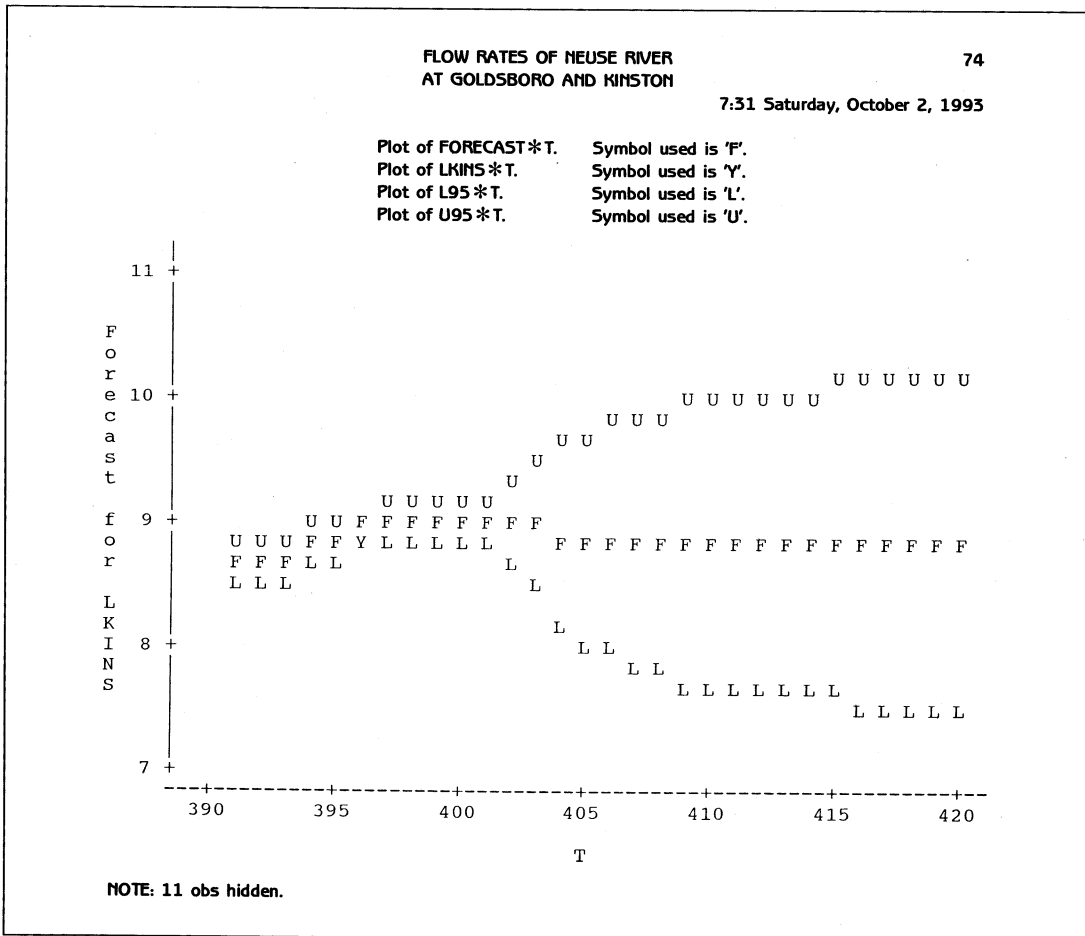


그림 4.26.7 전이함수모형에 의한 예측결과와 시계열산포도

4.7.4 다변량시계열데이터 HOUSECON

이 소절에서는 시계열데이터 HOUSECON을 유리함수형태의 전이함수모형화 한다. 시계열데이터 HOUSECON은 1983년 1월부터 1989년 10월까지의 미국의 월별 건설도금액(construction contracts in million of dollars)과 착공한 주택의 수(private housing starts in thousands of units)를 기록한 것이다. 이 시계열데이터의 관찰점들의 개수는 82이다. 시계열데이터 HOUSECON은 제 2.4절에 수록되어 있고 또한 SAS데이터라이브러리 TSA4 DATA의 영구SAS데이터세트 HOUSECON.SSD에 수록되어 있다. 입력시계열데이터 CONSTR는 월별 건설도금액이고 출력시계열데이터 HSTARTS는 주택의 수이다. 그러나 입력시계열데이터 CONSTR가 원인이(causal) 되고 출력시계열데이터 HSTARTS가 결과가 되는 인과관계를 갖는다고 확실히 말하기는 어렵다. 즉, 이 시계열데이터에 전이함수모형을 적용할 수 있다고 단언할 수는 없다. 이 시계열데이터를 [AppGuide] (pp. 178~186)에서는 전이함수모형을 사용해서 분석하였고, 이 소절에서는 이 분석을 따르고자 한다.

시계열데이터 HOUSECON을 유리함수형태의 전이함수모형화하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM4-27.PRG에 수록되어 있다).

```

/* */
/*          EXAM4-27.PRG          */
/* */
/*          TRANSFER FUNCTION MODELING          */
/*          DATASET = HOUSECON          */
/* */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS LINESIZE=72;
OPTIONS PAGESIZE=80;
DATA EXAM427;
    SET TSA4DATA.HOUSECON;
    T=_N_;
    RUN;
TITLE1 'CONSTRUCTION DATA';
TITLE2 'APPLICATION GUIDE 1 (1991)';
PROC ARIMA DATA=EXAM427;
    I VAR=CONSTR(12) NLAG=15;
    E P=(2, 3) METHOD=ML;
    I VAR=HSTARTS(12) CROSS=CONSTR(12) NLAG=15;

```

```

E INPUT=(/(1) CONSTR) METHOD=ML PLOT;
E P=(2, 5) INPUT=(/(1) CONSTR) METHOD=ML;
F LEAD=20 ID=T OUT=EXAM427B;
RUN;
PROC PLOT DATA=EXAM427B(FIRSTOBS=73);
PLOT FORECAST*T='F' HSTARTS*T='Y' L95*T='L'
      U95*T='U' / VPOS=27 OVERLAY;
RUN;

```

그림 2.4.1의 입력시계열데이터 CONSTR의 시계열산포도와 그림 2.4.3의 출력시계열데이터 HSTARTS의 시계열산포도에서 이 두 변수들은 시간이 흐름에 따라 분산들이 증가하는 분산이질성을 보임을 알 수 있다. 따라서 이 변수들을 대수변환해서 새로운 입력시계열데이터와 출력시계열데이터를 만들어야 한다. 그러나 [ApplGuide]에서 이 과정을 수행하지 않았으므로 이 소절에서도 대수변환과정을 생략하기로 한다.

PROC ARIMA의 첫 번째 IDENTIFY스테이트먼트는 다음과 같다.

```
IDENTIFY VAR=CONSTR(12) NLAG=15;
```

이 IDENTIFY스테이트먼트의 옵션 NLAG=은 OUTPUT윈도우에 출력할 표본자기상관계수들, 표본역자기상관계수들, 그리고 표본부분자기상관계수들의 최대시차를 지정한다. 입력시계열데이터 CONSTR와 출력시계열데이터 HSTARTS는 월별로 측정된 시계열데이터이므로 당연히 계절주기 12를 갖는다. 따라서 입력시계열데이터 CONSTR와 출력시계열데이터 HSTARTS의 계절차분들을 취한 계절차분시계열데이터들이 정상성조건을 더 잘 만족할 것이다. 이 IDENTIFY스테이트먼트에서는 계절차분된 입력시계열데이터를 식별하기 위해서 옵션 VAR=에 CONSTR(12)를 할당하였다. 이 스테이트먼트를 수행한 결과가 그림 4.27.1이다. 그림 4.27.1에서 표본자기상관함수는 시나브로형태를 보이고, 표본역자기상관함수와 표본부분자기상관함수는 시차 4 이후에서 절단형태를 보이므로 계절차분된 입력시계열데이터 CONSTR(12)는 AR(3)모형을 갖는다고 할 수 있다. [ApplGuide] (p. 178)는 이 차분된 입력시계열데이터 CONSTR(12)를 시차 1에서 AR계수가 0인 AR(3)모형으로 식별하였다.

이 SAS프로그램의 PROC ARIMA의 첫 번째 ESTIMATE스테이트먼트는 다음과 같다.

```
ESTIMATE P=(2, 3) METHOD=ML;
```

이 스테이트먼트는 계절차분된 입력시계열데이터 CONSTR(12)를 시차 1에서의 AR계수를 0으로 하는 AR(3)모형에 적합시키기 위한 것이다. 이 스테이트먼트를 수행한 결과가 그림 4.27.2이다. 그림 4.27.2에서 입력시계열데이터의 추정모형이 다음과 같음을 알 수 있다.

Autocorrelation Check for White Noise									
To Lag	Chi Square	DF	Prob	Autocorrelations					
6	24.46	6	0.000	0.101	0.277	0.307	0.172	0.139	0.302
12	38.76	12	0.000	0.068	0.158	0.252	-0.030	0.274	-0.045

그림 4.27.1 입력시계열데이터의 식별

$$x_t = \text{CONSTR}_t - \text{CONSTR}_{t-12}$$

$$x_t = 394.1908 + 0.2667 x_{t-2} + 0.3024 x_{t-3} + v_t$$

입력시계열데이터의 이 추정모형에 의한 잔차값들의 퍼트맨토우통계량값들로부터 추정된 AR(3)모형이 입력시계열데이터와 잘 부합됨을 알 수 있다.

CONSTRUCTION DATA					3				
APPLICATION GUIDE 1 (1991)									
					7:53 Saturday, October 2, 1993				
ARIMA Procedure									
Maximum Likelihood Estimation									
Parameter	Estimate	Approx. Std Error	T Ratio	Lag					
MU	914.76193	353.99721	2.58	0					
AR1,1	0.26672	0.10932	2.44	2					
AR1,2	0.30236	0.11124	2.72	3					
Constant Estimate = 394.190804									
Variance Estimate = 1793066.1									
Std Error Estimate = 1339.05418									
AIC = 1210.02961									
SBC = 1216.77509									
Number of Residuals = 70									
Correlations of the Estimates									
Parameter	MU	AR1,1	AR1,2						
MU	1.000	-0.024	-0.038						
AR1,1	-0.024	1.000	-0.136						
AR1,2	-0.038	-0.136	1.000						
Autocorrelation Check of Residuals									
To Lag	Chi Square	DF	Prob	Autocorrelations					
6	2.49	4	0.647	-0.043	-0.044	-0.076	0.040	-0.025	0.143
12	12.57	10	0.249	-0.028	0.019	0.173	-0.046	0.167	-0.239
18	20.03	16	0.219	-0.100	0.178	0.140	-0.075	0.122	0.011
24	23.09	22	0.397	-0.150	0.049	-0.000	-0.009	0.073	0.006
Model for variable CONSTR									
Estimated Mean = 914.761935									
Period(s) of Differencing = 12.									
Autoregressive Factors									
Factor 1: 1 - 0.26672 B**(2) - 0.30236 B**(3)									

그림 4.27.2 입력시계열데이터의 추정

Autocorrelation Check for White Noise

To Lag	Chi Square	DF	Prob	Autocorrelations					
6	45.16	6	0.000	0.465	0.435	0.287	0.210	0.233	0.141
12	53.24	12	0.000	0.076	-0.008	0.109	-0.132	-0.060	-0.237

Correlation of HSTARTS and CONSTR
 Variable CONSTR has been differenced.
 Period(s) of Differencing = 12.
 Both series have been prewhitened.
 Variance of transformed series = 189.086 and 1751098
 Number of observations = 70
 NOTE: The first 12 observations were eliminated by differencing.

Crosscorrelations

Lag	Covariance	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
-15	4544.451	0.24974	*****
-14	2079.400	0.11428	**
-13	428.557	0.02355
-12	-3093.897	-0.17003	***
-11	28.318916	0.00156
-10	-529.224	-0.02908	*
-9	3386.869	0.18613	****
-8	-626.411	-0.03443	*
-7	-724.580	-0.03982	*
-6	3350.027	0.18410	****
-5	-1913.770	-0.10517	**
-4	-1071.472	-0.05888	*
-3	-2455.364	-0.13494	*
-2	-3464.675	-0.19040	***
-1	2753.660	0.15133	****
0	8006.569	0.44001	*****
1	4243.793	0.23322	*****
2	2253.832	0.12386	**
3	337.487	0.01855
4	-1025.246	-0.05634
5	59.565205	0.00327	*
6	-883.550	-0.04856	*
7	-295.396	-0.01623
8	4179.921	0.22971	****
9	2454.341	0.13488	***
10	-83.215086	-0.00457
11	-2361.591	-0.12978	***
12	-3157.227	-0.17351	***
13	-814.905	-0.04478	*
14	1778.136	0.09772	**
15	1513.845	0.08319	**

"." marks two standard errors

Crosscorrelation Check Between Series

To Lag	Chi Square	DF	Prob	Crosscorrelations					
5	18.68	6	0.005	0.440	0.233	0.124	0.019	-0.056	0.003
11	25.01	12	0.015	-0.049	-0.016	0.230	0.135	-0.005	-0.130

Both variables have been prewhitened by the following filter:

Prewhitening Filter

Autoregressive Factors

Factor 1: 1 - 0.26672 B**(2) - 0.30236 B**(3)

그림 4.27.3 전이함수의 식별

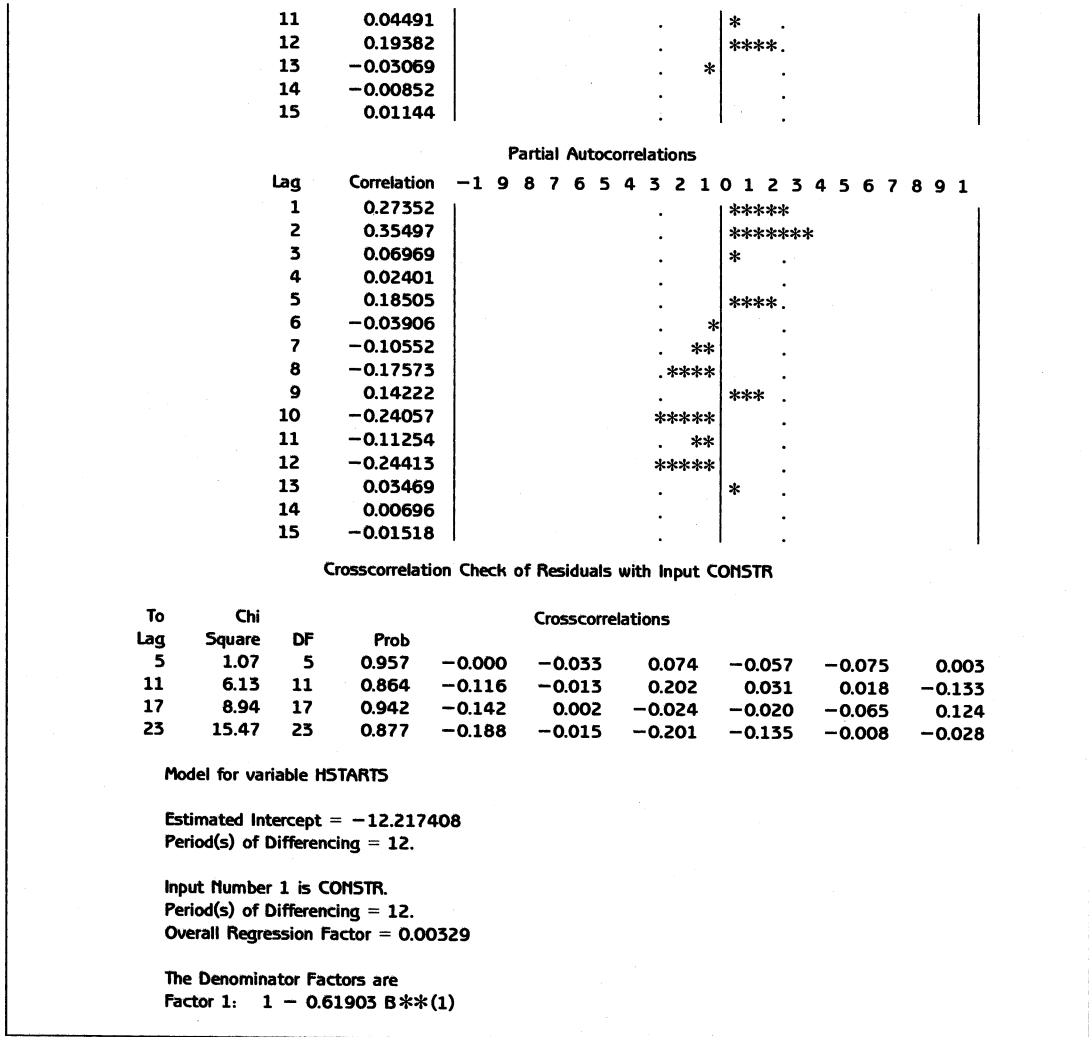


그림 4.27.4 입력시계열데이터의 식별

PROC ARIMA의 두 번째 IDENTIFY스테이트먼트는 다음과 같다.

IDENTIFY VAR=HSTARTS(12) CROSS=CONSTR(12) NLAG=15;

이 스테이트먼트는 입력시계열데이터가 CONSTR(12)이고 출력시계열데이터가 HSTARTS(12)인 전이함수를 식별하기 위한 것이다. 이 스테이트먼트를 수행한 결과가 그림 4.27.3이다. 그림 4.27.3에는 사전백색화된 입력시계열데이터와 같은 필터로 여과된 출력시계열데이터의 표본교차상관함수가 수록되어 있다. 이 표본교차상관함수에서 시차 0에서의 표본교차상관계수가 유의적으로 0과 같지 않고 또한 시차 0 이후의 표본교차상관계수들이 시나브로형태를 보인다고 할 수 있다. 따라서 이 시계열데이터는 Box-Jenkins방법에 의해서 차수들이 (0, 1, 0)인 전이함수모형으로 식별할 수 있다.

PROC ARIMA의 두 번째 ESTIMATE스테이트먼트는 다음과 같다.

ESTIMATE INPUT=(/ (1) CONSTR) METHOD=ML PLOT:

이 ESTIMATE스테이트먼트는 잡음시계열데이터의 ARMA모형을 식별하기 위한 것이므로 ESTIMATE스테이트먼트의 옵션 PLOT를 사용하였다. 이 스테이트먼트를 수행한 결과가 그림 4.27.4이다. 그림 4.27.4에서 다음과 같은 추정식을 얻는다.

$$y_t = \text{HSTARTS}_t - \text{HSTARTS}_{t-12}$$

$$y_t = -12.2174 + \frac{0.0033}{1-0.6190B} x_t + \hat{n}_t$$

이 추정모형에 의한 잡음시계열데이터의 적합값들 $\{\hat{n}_t\}$ 의 표본자기상관함수는 시나브로 형태를 보이고, 표본역자기상관함수와 표본부분자기상관함수는 시차들 2와 5에서 유의적으로 0과 다르다. 따라서 잡음시계열데이터를 시차들 1, 3, 4에서 AR계수들이 0인 AR모형으로 식별할 수 있다.

이 SAS프로그램의 PROC ARIMA의 세 번째 ESTIMATE스테이트먼트는 다음과 같다.

ESTIMATE P=(2, 5) INPUT=(/ (1) CONSTR) METHOD=ML:

이 스테이트먼트는 계절차분된 출력시계열데이터 HSTARTS(12)를 전이함수모형에 적합시키기 위한 것이다. 이 스테이트먼트를 수행한 결과가 그림 4.27.5이다. 그림 4.27.5에서 출력시계열데이터의 전이함수모형의 추정식이 다음과 같음을 알 수 있다.

$$y_t = -4.9234 + \frac{0.0041}{1-0.6361B} x_t + \frac{1}{1-0.3548B^2 - 0.2854B^5} v_t$$

잡음시계열모형의 잔차값들과 사전백색화된 입력시계열데이터의 표본교차상관함수에 의한 퍼트맨토우통계량값들로부터 잡음시계열모형의 잔차와 입력시계열데이터가 서로 독립이라는 가설을 채택한다. 추정된 전이함수모형의 잡음시계열데이터의 잔차값들의 퍼트맨토우통계량값들로부터 추정된 전이함수모형이 주어진 시계열데이터와 잘 부합됨을 알 수 있다.

PROC ARIMA의 첫 번째 FORECAST스테이트먼트는 다음과 같다.

FORECAST LEAD=20 ID=T OUT=EXAM427B:

이 FORECAST스테이트먼트는 바로 앞의 ESTIMATE스테이트먼트에서 추정된 전이함수모형을 이용해서 출력시계열데이터의 미래값들을 예측하는 것이다. 이 FORECAST스테이트먼트를 수행한 결과가 그림 4.27.6이다. 그림 4.27.6의 첫 부분에는 예측에 사용된 전이함수모형이 수록되어 있다. 이 전이함수모형은 그림 4.27.5에서 알 수 있듯이 최종적으로 추정된 전이함수모형이다. 그림 4.27.6의 뒷부분에는 예측값(Forecast), 표준오차(Std Error), 신뢰도 95%의 예측구간의 하한값(Lower 95%)과 상한값(Upper 95%)이 수록되

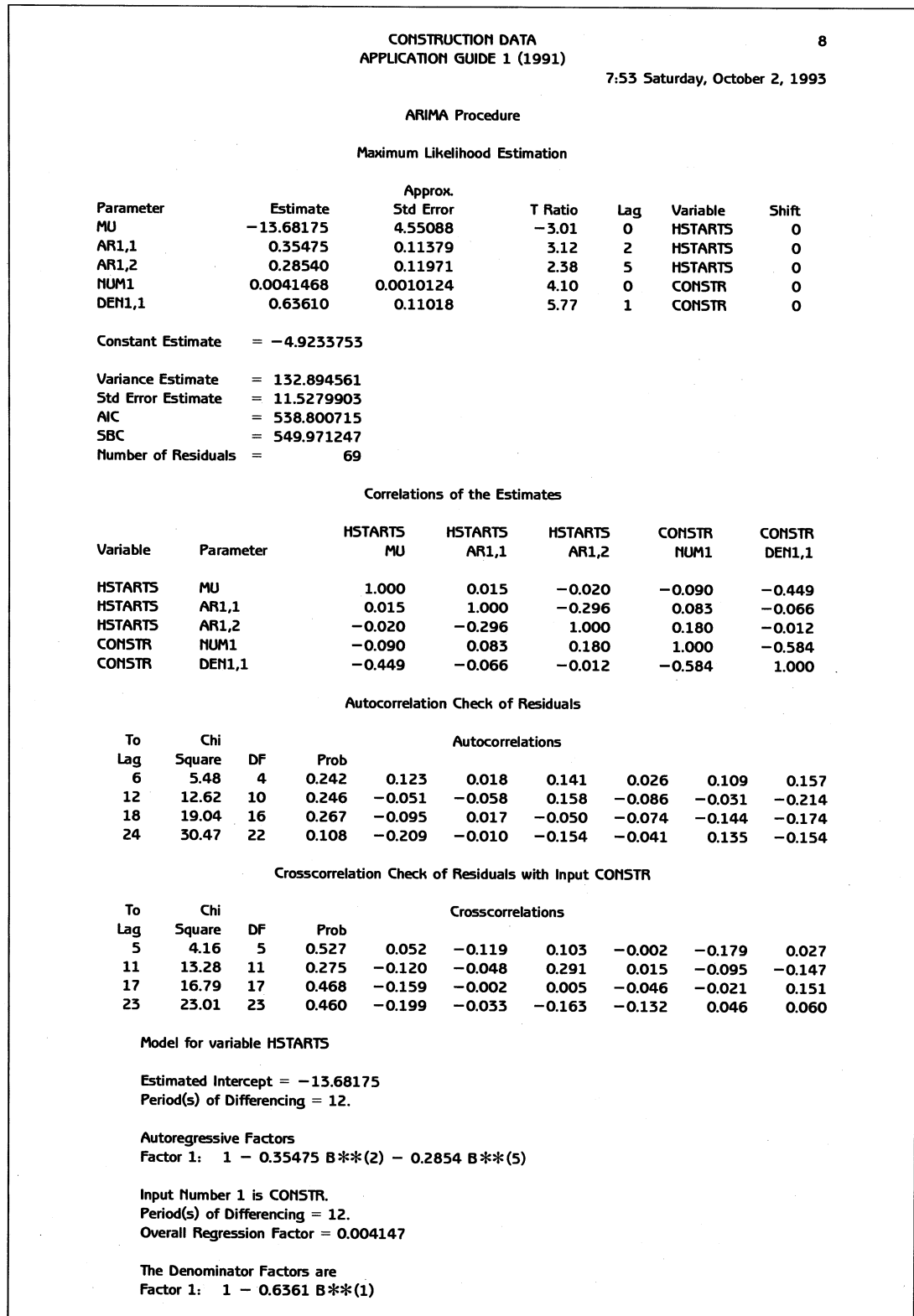


그림 4.27.5 전이함수모형의 추정

CONSTRUCTION DATA		9		
APPLICATION GUIDE 1 (1991)		7:53 Saturday, October 2, 1993		
ARIMA Procedure				
Model for variable HSTARTS				
Estimated Intercept = -13.68175				
Period(s) of Differencing = 12.				
Autoregressive Factors				
Factor 1: 1 - 0.35475 B**(2) - 0.2854 B**(5)				
Input Number 1 is CONSTR.				
Period(s) of Differencing = 12.				
Overall Regression Factor = 0.004147				
The Denominator Factors are				
Factor 1: 1 - 0.6361 B**(1)				
Forecasts for variable HSTARTS				
Obs	Forecast	Std Error	Lower 95%	Upper 95%
83	104.9165	12.7956	79.8375	129.9954
84	95.4232	13.2742	67.4064	119.4401
85	95.2224	14.3814	67.0354	123.4094
86	79.8512	14.9409	50.5677	109.1347
87	115.0437	15.3024	85.0516	145.0358
88	125.1381	15.8980	93.9786	156.2976
89	130.1762	16.0842	98.6519	161.7005
90	139.8876	16.3605	107.8217	171.9535
91	131.8860	16.4422	99.6599	164.1122
92	119.7064	16.5415	87.2857	152.1271
93	106.2907	16.6059	73.7438	138.8375
94	128.0272	16.6396	95.4141	160.6402
95	101.9600	21.7612	59.3089	144.6112
96	90.8298	22.1463	47.4239	134.2356
97	92.3791	23.0464	47.2091	137.5492
98	77.0181	23.4699	31.0180	123.0182
99	112.3093	23.7708	65.7195	158.8991
100	122.1923	24.2493	74.6646	169.7200
101	127.3545	24.3973	79.5368	175.1722
102	136.9075	24.6291	88.6354	185.1795

그림 4.27.6 전이함수모형을 사용한 예측

어 있다. 이들은 각각 변수들 FORECAST, STD, L95와 U95에 저장되어 있다. 이 PROC ARIMA의 FORECAST스태이트먼트의 옵션 OUT=에 SAS데이터세트 EXAM427B를 할당하였다. 이 SAS데이터세트에는 FORECAST스태이트먼트를 수행한 결과로 얻어진 예측의 결과가 수록된다. 이 SAS데이터세트에는 ID변수 T, 출력시계열데이터, FORECAST, STD, L95, U95, 그리고 RESIDUAL이 수록되어 있다.

이 SAS프로그램의 두 번째 PROC스텝은 다음과 같다.

```
PROC PLOT DATA=EXAM427B(FIRSTOBS=73);
  PLOT FORECAST*T='F' HSTARTS*T='Y' L95*T='L'
      U95*T='U' / VPOS=27 OVERLAY;
RUN;
```

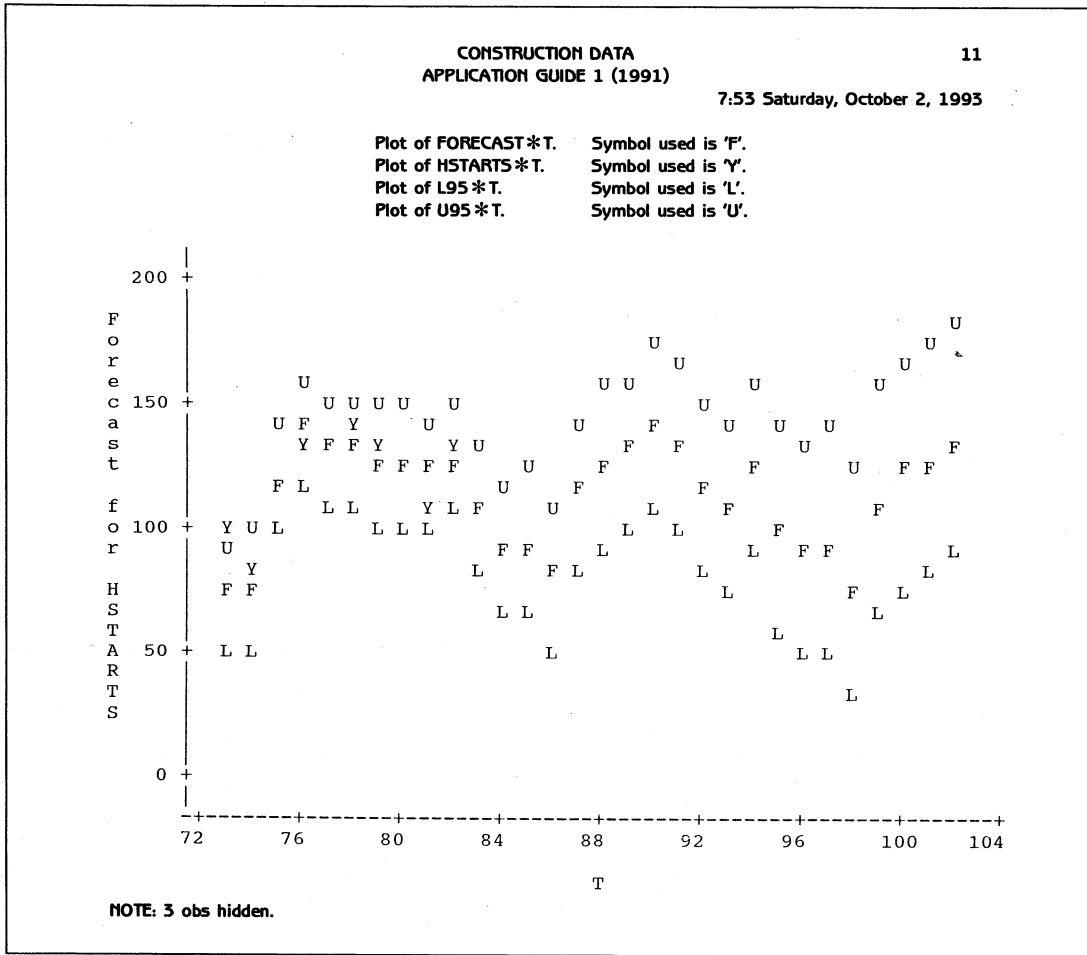


그림 4.27.7 전이함수모형을 사용한 예측결과의 시계열산포도

이 PROC스텝은 PROC PLOT를 사용해서 SAS데이터세트 EXAM427B에 저장된 출력 시계열데이터의 관찰, 예측, 예측구간의 하한과 상한의 시계열산포도를 출력하는 것이다. 이 PROC스텝을 수행한 결과가 그림 4.27.7이다.

▣ 참고문헌 ▣

- 남기석(1992), 「변형된 ARIMA모형을 사용한 한국의 주가분석」, 석사논문, 연세대학교 대학원, 서울.
- 양현조(1992), 「전이함수모형과 상태공간모형의 비교연구」, 석사논문, 연세대학교 대학원, 서울.
- 장승욱(1992), 「3배열방법을 사용한 전이함수모형의 식별」, 석사논문, 연세대학교 대학원, 서울.

원, 서울.

- Bartlette, M. S. (1946), On the theoretical specification of sampling properties of autocorrelated time series, *J. Roy. Statist. Soc. Ser. B*, vol. 8, pp. 27~41.
- Beguín, J. M., C. Gourieroux & A. Monfort(1980), Identification of a mixed ARMA process: the corner method, *in Time Series*, O. D. Anderson, Ed., North-Holland, Amsterdam.
- Box, G. E. P. & G. M. Jenkins(1976), *Time Series Analysis: Forecasting and Control*, Holden-Day, New York.
- Choi, B. S. (1991), The asymptotic joint distribution of the correlation determinant estimates of an ARMA process, *Commun. Statist. - Theory and Methods*, vol. 20, pp. 2823~2835.
- Choi, B. S. (1992), *ARMA Model Identification*, Springer-Verlag, New York.
- Choi, B. S. (1993), Two chi-square statistics for determining orders of an ARMA process, *IEEE Trans. Signal Processing*, vol. 41, pp. 2165~2176.
- Choi, B. S. (1994), On the covariance matrix estimators of the white noise process of a vector autoregressive model, *Commun. Statist. - Theory and Methods*, vol. 23, pp. 249~256.
- Edlund, P. O. (1990), Ridge estimation of transfer function weights, *Commun. Statist. - Simulation and Computation*, vol. 19, pp. 451~468.
- Hocking, R. R., F. M. Speed & M. J. Lynn(1976), A class of biased estimators in linear regression, *Technometrics*, vol. 18, pp. 425~437.
- Hoerl, A. E., R. W. Kennard & K. F. Baldwin(1975), Ridge regression: some simulations, *Commun. Statist. - Theory and Methods*, vol. 4, pp. 105~123.
- Kelley, G. D. (1977), A new approach to ARMA modeling - the nonstationary case, Ph. D. Dissertation, Southern Methodist Univ., Texas.
- Lawless, J. F. & P. Wang(1976), A simulation study of ridge and other regression estimators, *Commun. Statist. - Theory and Methods*, vol. 5, pp. 307~323.
- Liu, L. M. & D. M. Hanssens(1982), Identification of the multiple-input transfer function models, *Commun. Statist. - Theory and Methods*, vol. 11, pp. 297~314.
- McIntire, D. D. (1977), A new approach to ARMA modeling, Ph. D. Dissertation, Southern Methodist Univ., Texas.

잡음항이 ARMA과정인 모형

- 5.1 잡음항이 정상성을 갖는 회귀모형
- 5.2 간섭시계열모형
- 5.3 공 누 적

제 5 장 잡음항이 ARMA과정인 모형

5.1 잡음항이 정상성을 갖는 회귀모형

5.1.1 잡음항이 ARMA과정인 회귀모형

시계열데이터 $\{y_t\}$ 를 분석할 때 이 시계열데이터와 관련있는 시계열데이터들 $\{x_{t,1}, x_{t,2}, \dots, x_{t,k}\}$ 를 입력변수들로 사용함으로써 모형의 신뢰성과 예측의 정확도를 높일 수 있다. 회귀분석에서 일반적으로 가정되는 회귀모형은 다음과 같다.

$$y_t = \beta_0 + \beta_1 x_{t,1} + \beta_2 x_{t,2} + \dots + \beta_k x_{t,k} + v_t$$

여기서 $\{x_{t,1}, x_{t,2}, \dots, x_{t,k}\}$ 는 입력변수들이고 오차항 $\{v_t\}$ 는 백색잡음과정이다. 이 회귀모형에서 종속변수와 입력변수들이 시계열데이터들이므로 오차항 $\{v_t\}$ 가 서로 독립이고 같은 분포를 갖는지 여부를 검정해야 한다. 독립성을 검정하기 위해서는 Durbin-Watson검정, 런검정, 그리고 스펙트럴검정 등을 사용한다. 이 검정들에 대한 자세한 내용은 [SAS4TSA2] (pp. 238~251)를 참조하라. 이러한 검정들을 수행한 결과로 오차항들이 독립이라는 귀무가설이 기각되면 이 시계열데이터 $\{y_t\}$ 는 다음과 같은 회귀모형으로 나타내는 것이 타당할 것이다.

$$y_t = \beta_0 + \beta_1 x_{t,1} + \beta_2 x_{t,2} + \dots + \beta_k x_{t,k} + n_t$$

여기서 잡음항 $\{n_t\}$ 는 정상성을 가지며 시계열데이터 $\{y_t\}$ 는 입력변수들의 시차변수들, 즉 $\{x_{t-j,1}, x_{t-j,2}, \dots, x_{t-j,k} \mid j=1, 2, \dots\}$ 에는 영향을 받지 않는다고 가정한다. 만일, 입력변수들의 시차변수들도 y_t 에 영향을 미친다면, 잡음항이 정상성을 갖는 회귀모형보다는 제 4장에서 다룬 전이함수모형을 사용해서 시계열데이터 $\{y_t\}$ 를 분석하는 것이 타당할 것이다.

설명변수들(explanatory variables), 또는 입력변수들(input variables) x_1, x_2, \dots, x_k 는 두 종류로 나눌 수 있다. 하나는 결정적(deterministic) 변수이고, 다른 하나는 비결정적(nondeterministic), 또는 확률적(stochastic) 변수이다. 결정적 변수란 이 변수의 미래값

들을 확실성을 가지고(with certainty) 알 수 있는 변수를 의미한다. 한 예로 연도별 국민총생산액을 종속변수로, 연도(year)를 입력변수로 회귀분석할 때에는 입력변수 연도는 오차없이 예측할 수 있으므로 연도는 결정적 변수이다. 결정적 변수의 대표적인 예가 보통 0과 1을 관찰값들로 갖는 지시변수(indicator variable)이다. 비결정적 변수란 이 변수의 미래값들을 확실성을 가지고 알 수가 없는 변수를 의미한다. 한 예로 연도별 국민총생산액을 종속변수로, 연도별 노동인구수를 입력변수로 회귀분석할 때에는 입력변수 노동인구수를 오차없이 예측할 수 없으므로 노동인구수는 비결정적 변수이다. 대부분의 입력변수들은 비결정적이다. 물론 때때로 비결정적 입력변수의 예측된 미래값들을 확실성을 가지고 알고 있는 값들이라고 가정하고 종속변수의 미래값들을 예측하기도 한다. 그러나 이렇게 얻어진 결과를 지나치게 믿어서는 안 된다. 미래값을 예측할 때 입력변수들이 결정적이면 이 절에서 다루는 잡음항이 정상성을 갖는 회귀모형을 사용한다. 반면에 미래값을 예측할 때 입력변수들이 비결정적이면 제 4장에서 다루는 전이함수모형을 사용한다.

입력변수들 x_1, x_2, \dots, x_k 가 결정적인 회귀모형에서 잡음항 $\{n_t\}$ 가 정상성을 갖는다면 이 잡음항이 ARMA모형을 따른다고 가정하는 것은 자연스러운 일이다. SAS/ETS의 PROC ARIMA를 사용하면 잡음항 $\{n_t\}$ 가 ARMA모형을 갖는 시계열데이터 $\{y_t\}$ 의 회귀모형을 추정할 수 있다. 또한 PROC AUTOREG를 사용하면 잡음항 $\{n_t\}$ 가 AR모형을 갖는 시계열데이터 $\{y_t\}$ 의 회귀모형을 추정할 수 있다.

5.1.2 예 제 들

시계열데이터 $\{y_t\}$ 의 회귀모형의 잡음항 $\{n_t\}$ 가 백색잡음과정은 아니지만 정상성을 갖는 경우에 이 시계열데이터를 분석하는 첫 번째 예제로써 다음 SAS프로그램을 수행해 보자 (이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM5-1.PRG에 수록되어 있다).

```

/*                                                    */
/*                                                    EXAM5-1.PRG          */
/*                                                    */
/* REGRESSION WITH TIME SERIES ERROR                */
/* DATASET = SIMULATED LINEAR MODEL WITH AR(5) ERROR */
/*                                                    */
OPTIONS LINESIZE=72 PAGESIZE=60;
DATA EXAM51;
  X=0;
  DO I=1 TO 820;
    X=X+RANUNI(387213);
    E=RANNOR(387317);
    N=SUM(E, 0.8*N1, 0.8*N4, -0.64*N5);
  END;

```



```

Y=20+X+N;
N5=N4; N4=N3; N3=N2; N2=N1; N1=N;
IF I > 500 THEN OUTPUT;
END;
KEEP Y X;
RUN;
DATA EXAM51B;
SET EXAM51;
T=_N_;
IF T GT 300 THEN Y=. ;
RUN;
DATA EXAM51C;
SET EXAM51B;
IF T GT 300 THEN DELETE;
RUN;
PROC PLOT DATA=EXAM51C;
PLOT Y*T='Y' X*T='X' / OVERLAY VPOS=30;
RUN;
PROC REG DATA=EXAM51C;
MODEL Y=X / DW;
RUN;
PROC ARIMA DATA=EXAM51C;
I VAR=Y CROSSCOR=(X) NOPRINT;
E P=(1 4 5) INPUT=(X) METHOD=ML PLOT;
RUN;
PROC ARIMA DATA=EXAM51B;
I VAR=Y CROSSCOR=(X) NOPRINT;
E P=(1 4 5) INPUT=(X) METHOD=ML NOPRINT;
F LEAD=20 ID=T OUT=FORE51A;
RUN;
PROC PLOT DATA=FORE51A(FIRSTOBS=276);
PLOT Y*T='Y' FORECAST*T='P' L95*T='L'
      U95*T='U' / OVERLAY VPOS=27
      VAXIS= 405 TO 445 BY 5;
RUN;
PROC AUTOREG DATA=EXAM51C PARTIAL;
MODEL Y=X / METHOD=ML NLAG=(1 4 5);
RUN;
PROC AUTOREG DATA=EXAM51B;

```

```

MODEL Y=X / METHOD=ML NLAG=(1 4 5) NOPRINT;
OUTPUT OUT=FORE51B P=PRED LCL=L95 UCL=U95;
RUN;
PROC PRINT DATA=FORE51B(FIRSTOBS=301);
RUN;
PROC PLOT DATA=FORE51B(FIRSTOBS=276);
PLOT Y*T='Y' PRED*T='P' L95*T='L'
      U95*T='U' / OVERLAY VPOS=27
      VAXIS = 405 TO 445 BY 5;
RUN;

```

첫 번째 DATA스텝은 다음과 같다. 이 DATA스텝은 [ETS6.1] (pp. 183~185)의 예제를 약간 변형시킨 것이다.

```

DATA EXAM51;
X=0;
DO I=1 TO 820;
X=X+RANUNI(387213);
E=RANNOR(387317);
N=SUM(E, 0.8*N1, 0.8*N4, -0.64*N5);
Y=20+X+N;
N5=N4; N4=N3; N3=N2; N2=N1; N1=N;
IF I>500 THEN OUTPUT;
END;
KEEP Y X;
RUN;

```

이 DATA스텝에서는 구간 (0, 1)에서 발생한 일양난수들(uniform random numbers)인 확률보행과정 $\{x_t\}$ 와 다음 회귀모형을 만족하는 확률과정 $\{y_t\}$ 에서 관찰점들이 320개인 시계열데이터를 발생시켜서 SAS데이터세트 EXAM51에 저장한다.

$$y_t = 20 + x_t + n_t$$

$$n_t = v_t + 0.8v_{t-1} + 0.8v_{t-4} - 0.64v_{t-5}$$

여기서 $\{v_t\}$ 는 평균이 0이고 분산이 1인 정규백색잡음과정이다. 두 번째와 세 번째 DATA스텝들은 다음과 같다.

```

DATA EXAM51B;
SET EXAM51;
T=_N_;
IF T GT 300 THEN Y = . ;

```

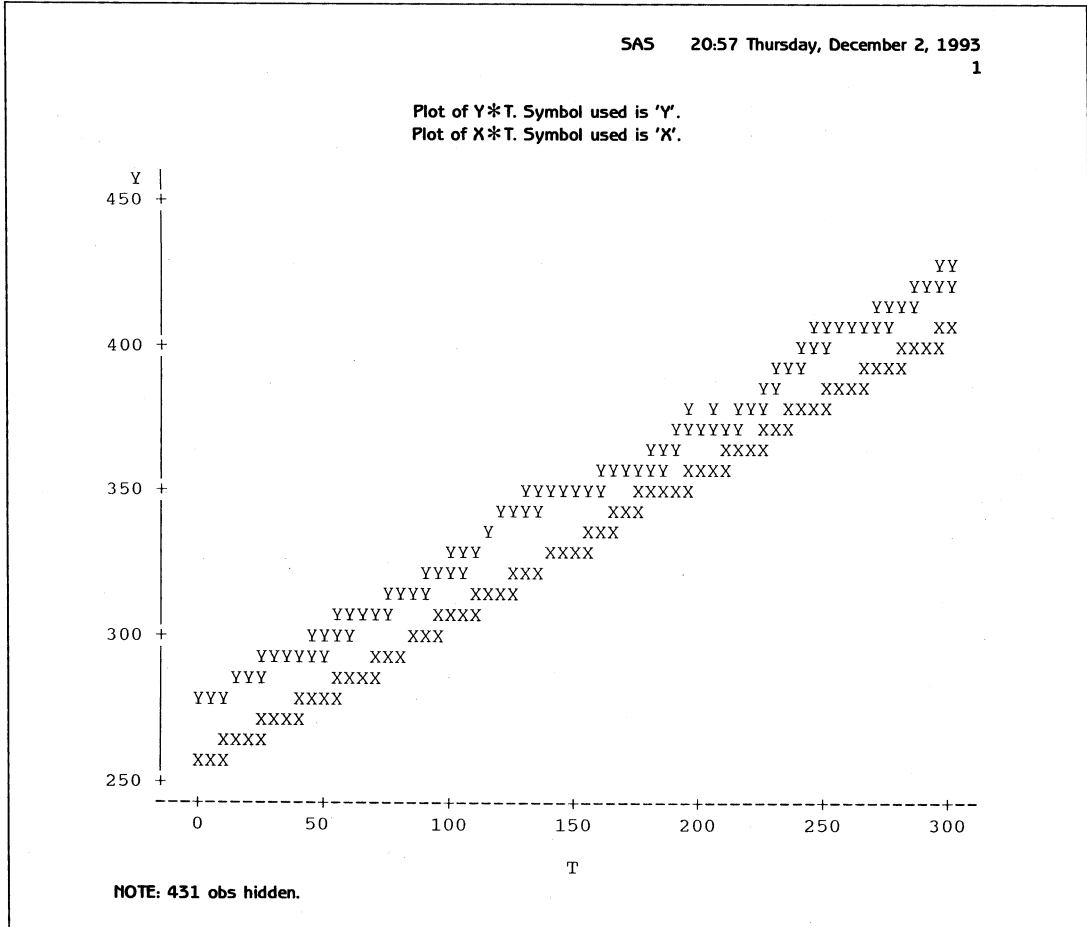


그림 5.1.1 추세를 갖는 시계열데이터들의 시계열산포도

```

RUN;
DATA EXAM51C;
SET EXAM51B;
IF T GT 300 THEN DELETE;
RUN;
    
```

두 번째 DATA스텝에서는 SAS데이터세트 EXAM51의 제301번 이후의 관찰점들에서 변수 Y의 관찰값들을 결측으로 만들어 SAS데이터세트 EXAM51B에 저장한다. 세 번째 DATA스텝에서는 SAS데이터세트 EXAM51에서 제301번 이후의 관찰점들을 제거한 뒤 SAS데이터세트 EXAM51C에 저장한다.

이 SAS데이터세트 EXAM51C의 시계열산포도를 그리기 위해서 다음 PROC PLOT를 수행한다.

```

PROC PLOT DATA=EXAM51C;
PLOT Y*T='Y' X*T='X' / OVERLAY VPOS=30;
RUN;
    
```

SAS 20:57 Thursday, December 2, 1993
2

Model: MODEL1
Dependent Variable: Y

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	552979.18563	552979.18563	49115.073	0.0000
Error	298	3355.13697	11.25885		
C Total	299	556334.32260			

Root MSE	3.35542	R-square	0.9940
Dep Mean	350.11319	Adj R-sq	0.9939
C.V.	0.95838		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	25.341608	1.47819806	17.144	0.0001
X	1	0.979827	0.00442122	221.619	0.0000

Durbin-Watson D (For Number of Obs.)	0.235 300
1st Order Autocorrelation	0.882

그림 5.1.2 보통최소제곱추정법을 사용한 회귀모형의 추정

이 PROC PLOT를 수행한 결과가 그림 5.1.1이다. 이 시계열산포도로부터 $\{x_t\}$ 와 $\{y_t\}$ 가 선형추세를 갖는 시계열데이터임을 알 수 있다.

이 절에서 다루는 회귀모형은 다음과 같다.

$$y_t = \beta_0 + \beta_1 x_{t,1} + \beta_2 x_{t,2} + \dots + \beta_k x_{t,k} + n_t$$

여기서 잡음항 $\{n_t\}$ 는 백색잡음과정이 아니다. 그러나 경우에 따라서는 보통최소제곱추정법(ordinary least square estimation method)으로 이 선형회귀모형을 추정할 필요가 있다. 보통최소제곱추정법으로 이 선형회귀모형을 추정하기 위해서 다음 PROC REG를 수행한다.

```
PROC REG DATA=EXAM51C;
MODEL Y=X /DW;
RUN;
```

이 PROC REG를 수행한 결과가 그림 5.1.2이다. 그림 5.1.2에서 이 잡음항이 백색잡음과정이라는 가정을 바탕으로 할 때 널리 사용되는 보통최소제곱추정법을 사용해서 추정한 회귀모형의 결정계수가 0.994로 상당히 크다. 그러나 Durbin-Watson통계량값이 0.235이고 시차 1에서 표본자기상관계수가 0.882이므로 계열상관(serial correlation)이 존재한다. 따라서 주어진 시계열데이터를 잡음항이 ARMA과정인 회귀모형에 적합시키면 더 좋은 결과를 얻을 수도 있을 것이다.

우리가 추정하고자 하는 회귀모형은 다음과 같다.

$$y_t = \beta_0 + \beta_1 x_{t,1} + \beta_2 x_{t,2} + \cdots + \beta_k x_{t,k} + n_t$$

주어진 시계열데이터를 잡음항 $\{n_t\}$ 가 백색잡음과정이 아니고 ARMA과정인 회귀모형에 적합시키기 위해서 다음과 같은 PROC ARIMA를 수행한다.

```
PROC ARIMA DATA=EXAM51C;
  I VAR=Y CROSSCOR=(X) NOPRINT;
  E P=(1 4 5) INPUT=(X) METHOD=ML PLOT;
  RUN;
```

이 PROC ARIMA의 IDENTIFY스테이트먼트의 옵션 VAR=에는 종속변수를 지정하고 옵션 CROSSCOR=에는 입력변수들을 지정한다. IDENTIFY스테이트먼트의 옵션 NOPRINT를 사용한 이유는 종속변수들을 사전백색화하지 않으면 IDENTIFY스테이트먼트의 결과가 의미없기 때문이다. 즉, 잡음항이 ARMA과정인 회귀모형을 추정하기 위해서 PROC ARIMA를 사용할 때 IDENTIFY스테이트먼트는 단지 종속변수와 입력변수들, 또는 이들의 차분차수들을 지정하는 데만 사용한다. ESTIMATE스테이트먼트는 다음 모형을 추정하기 위한 것이다.

$$y_t = \beta_0 + \beta_1 x_t + \frac{1}{1 - \phi_1 B - \phi_4 B^4 - \phi_5 B^5} u_t$$

ESTIMATE스테이트먼트의 옵션 P=에는 잡음항의 AR특성식의 형태를 지정한다. 뒤에서 다시 설명하겠지만 잡음항의 특성식의 형태를 정하는 것은 그리 간단한 일이 아니다. 여기서는 앞의 DATA스텝에서 시계열데이터를 발생시킬 때 사용한 잡음항의 AR특성식의 형태를 그대로 사용하였다. 옵션 INPUT=에는 입력변수들을 지정한다. 이 입력변수들은 해당되는 ESTIMATE스테이트먼트 앞의 IDENTIFY스테이트먼트의 옵션 CROSSCOR=에 지정되어 있어야 한다. 옵션 METHOD=에는 추정방법을 지정한다. 만일, 이 옵션에 CLS가 지정되면 조건부최소제곱추정법이 사용되고, ULS가 지정되면 비조건부최소제곱추정법이 사용되며, ML이 지정되면 최우추정법이 사용된다. 이 추정법들에 대한 자세한 내용은 [SAS4TSA2] (pp. 503~509)를 참조하라. 옵션 PLOT는 이 추정된 회귀모형의 잔차들의 표본자기상관함수, 표본역자기상관함수, 그리고 표본부분자기상관함수를 출력시키게 한다.

이 PROC ARIMA를 수행한 결과가 그림 5.1.3이다. 그림 5.1.3의 부분 A에서 최우추정된 회귀모형이 다음과 같음을 알 수 있다.

$$y_t = 24.482 + 0.982 x_t + \frac{1}{1 - 0.834 B - 0.773 B^4 + 0.681 B^5} u_t$$

또한, 추정모형의 추정계수들이 유의적으로 0과 다름을 알 수 있다. 부분 B의 카이제곱

19	-0.061677	-0.05929	.	*	.
20	-0.017483	-0.01681	.	.	.
21	0.034815	0.03347	.	.	*
22	-0.011498	-0.01105	.	.	.
23	0.031140	0.02993	.	.	*
24	0.096174	0.09245	.	.	**

"." marks two standard errors

Inverse Autocorrelations

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
1	-0.03412	*
2	-0.08691	**
3	0.02999	*
4	-0.07868	**
5	0.07554	*
6	0.07247	*
7	0.03072	*
8	0.01924	*
9	-0.07719	*
10	-0.06937	*
11	0.04684	*
12	0.06665	*	.	.	.
13	-0.04089	*	.	.
14	0.00896	*	.
15	-0.06802	*
16	-0.05419	*
17	0.11653	**
18	-0.04685	*
19	0.05151	*
20	0.02246
21	-0.07859	**
22	0.03192	*
23	-0.00881
24	-0.06865	*

Partial Autocorrelations

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
1	0.02611	*
2	0.07873	**
3	-0.04036	*
4	0.01851	*
5	-0.05879	*
6	-0.05525	*
7	-0.05396	*
8	-0.02129	*
9	0.07777	*	**
10	0.05148	*
11	-0.02425	*
12	-0.04967	*
13	0.03617	*
14	0.02603	*
15	0.05540	*
16	0.04839	*
17	-0.13047	*	*
18	0.02621	*	*
19	-0.05076	*
20	-0.01786	*
21	0.08079	*	**
22	-0.02121	*
23	0.01385	*
24	0.07412	*	*

(D)

Model for variable Y

Estimated Intercept = 24.4820662

Autoregressive Factors

Factor 1: $1 - 0.83402 B^{**}(1) - 0.77299 B^{**}(4) + 0.68135 B^{**}(5)$

Input Number 1 is X.

Overall Regression Factor = 0.982357

그림 5.1.3 PROC ARIMA를 사용한 회귀모형의 추정

통계량값들, 즉 퍼트먼트우통계량값들로부터 이 추정된 회귀모형의 잔차들이 백색잡음과정에서 나왔다고 할 수 있다. 부분 C의 표본자기상관함수, 표본역자기상관함수, 그리고 표본부분자기상관함수로부터 이 추정된 회귀모형의 잔차들이 백색잡음과정에서 나왔다는 결론을 재확인할 수 있다.

주어진 시계열데이터를 잡음항이 ARMA과정인 회귀모형에 적합시켜서 미래값을 예측하기 위해서 다음 PROC ARIMA를 수행한다.

```
PROC ARIMA DATA=EXAM51B;
  I VAR=Y CROSSCOR=(X) NOPRINT;
  E P=(1 4 5) INPUT=(X) METHOD=ML NOPRINT;
  F LEAD=20 ID=T OUT=FORE51A;
  RUN;
```

이 PROC ARIMA의 IDENTIFY스테이트먼트와 ESTIMATE스테이트먼트는 각각 앞의 PROC ARIMA의 IDENTIFY스테이트먼트와 ESTIMATE스테이트먼트와 거의 같다. 다만 ESTIMATE스테이트먼트에 옵션 NOPRINT를 사용한 것이 다르다. 이 옵션 NOPRINT는 ESTIMATE스테이트먼트를 수행한 결과를 OUTPUT윈도우에 출력시키지 않도록 한다. FORECAST스테이트먼트는 추정모형을 사용해서 미래값들을 예측하기 위한 것이다. FORECAST스테이트먼트의 옵션 LEAD=은 예측할 미래시점들을 지정하는 것이다. 이 예제에서는 이 옵션에 20이 지정되었으므로 20개의 미래시점들에 대해서 예측을 한다. 옵션 OUT=에는 예측된 결과를 저장할 SAS데이터세트 FORE51A를 지정하였다. 여기서 강조할 점은 PROC ARIMA를 사용해서 예측을 할 때에는 예측할 미래시점들에서 입력변수들의 값들이 PROC ARIMA스테이트먼트의 옵션 DATA=에 지정된 SAS데이터세트에 수록되어 있어야 한다는 것이다. 그렇지 않은 경우에는 종속변수의 모형을 추정하는 ESTIMATE스테이트먼트를 수행하기 전에 입력변수들을 추정하는 ESTIMATE스테이트먼트들을 수행하여야 한다. 즉, 전이함수모형을 사용해야 한다. 이 PROC ARIMA를 수행한 결과가 그림 5.1.4이다.

잡음항이 ARMA과정인 회귀모형을 사용해서 얻은 예측값들의 시계열산포도를 그리기 위해서 다음 PROC PLOT를 수행한다.

```
PROC PLOT DATA = FORE51A(FIRSTOBS = 276);
  PLOT Y*T='Y' FORECAST*T='P' L95*T='L'
  U95*T='U' / OVERLAY VPOS=27
  VAXIS = 405 TO 445 BY 5;
  RUN;
```

PLOT스테이트먼트에서 변수 FORECAST는 예측, 변수 L95와 변수 U95는 각각 신뢰수준 95%의 예측구간의 상한과 하한이다. 이 PROC PLOT를 수행한 결과가 그림 5.1.5이다.

만일, 회귀모형의 잡음항이 ARMA모형이 아닌 AR모형을 갖는다면 시계열데이터 $\{y_t\}$

SAS 20:57 Thursday, December 2, 1993
9

ARIMA Procedure

Forecasts for variable Y

Obs	Forecast	Std Error	Lower 95%	Upper 95%
301	425.3517	1.0200	423.3526	427.3508
302	426.9943	1.3281	424.3912	429.5974
303	424.8606	1.5057	421.9094	427.8118
304	425.2982	1.6178	422.1273	428.4691
305	426.8166	2.0641	422.7709	430.8622
306	427.9139	2.3076	423.3910	432.4368
307	426.2938	2.4521	421.4877	431.0999
308	427.1842	2.5409	422.2042	432.1643
309	428.6219	2.7861	423.1612	434.0825
310	429.5073	2.9210	423.7823	435.2323
311	428.4450	2.9968	422.5715	434.3186
312	429.1811	3.0396	423.2237	435.1385
313	430.6744	3.1572	424.4865	436.8623
314	431.6192	3.2158	425.3163	437.9220
315	430.7444	3.2444	424.3854	437.1034
316	431.3558	3.2578	424.9706	437.7410
317	431.9427	3.3056	425.4639	438.4214
318	433.1004	3.3249	426.5839	439.6170
319	433.0010	3.3317	426.4709	439.5311
320	433.9323	3.3337	427.3984	440.4662

그림 5.1.4 PROC ARIMA를 사용한 회귀모형의 예측

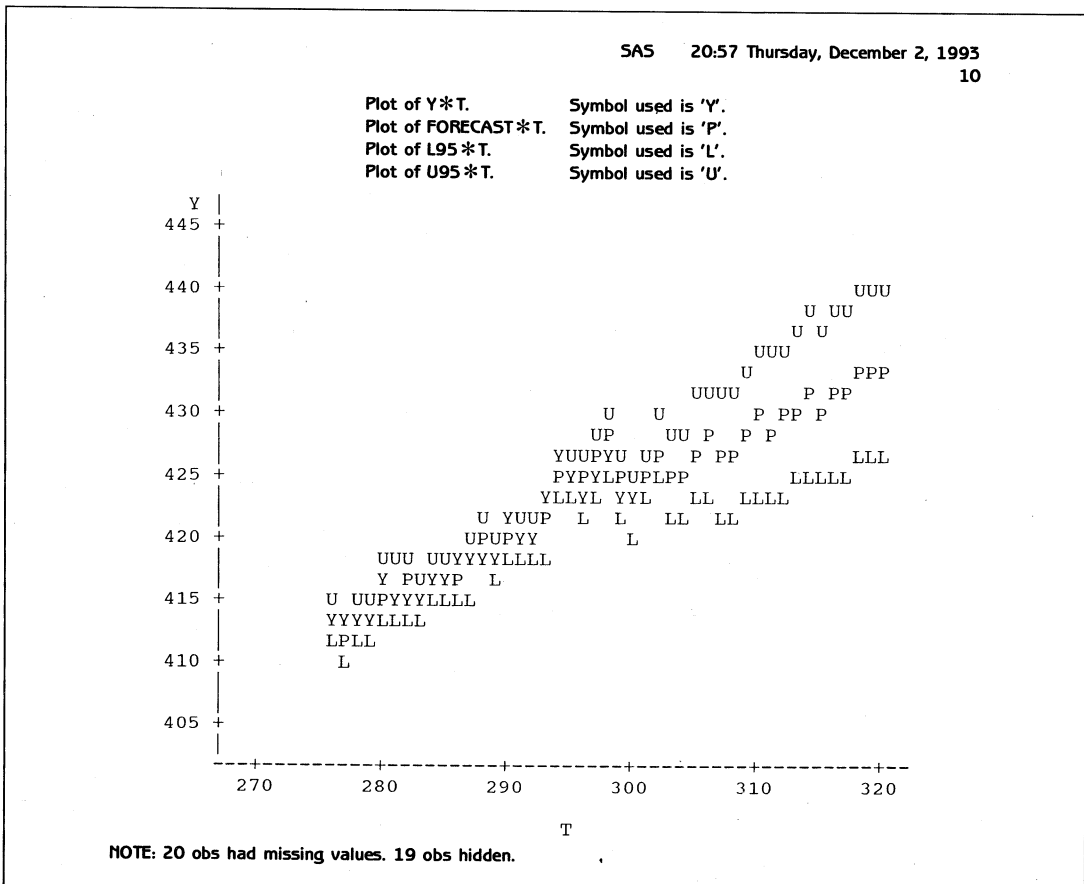


그림 5.1.5 PROC ARIMA에 의한 예측값들의 시계열산포도

는 다음과 같은 회귀모형으로 나타내는 것이 타당할 것이다.

$$y_t = \beta_0 + \beta_1 x_{t,1} + \beta_2 x_{t,2} + \dots + \beta_k x_{t,k} + n_t$$

여기서 잡음항 $\{n_t\}$ 는 AR모형을 따르며 시계열데이터 $\{y_t\}$ 는 설명변수들의 시차변수들, 즉 $\{x_{t-j,1}, x_{t-j,2}, \dots, x_{t-j,k} \mid j=1, 2, \dots\}$ 에는 영향을 받지 않는다고 가정한다. 이러한 경우에는 PROC ARIMA 대신에 PROC AUTOREG를 사용할 수가 있다. PROC AUTOREG의 자세한 사용법은 [SAS4TSA2] (pp. 266~297)를 참조하라. 잡음항이 AR모형을 갖는다는 가정하에 회귀분석하기 위해서 다음 PROC AUTOREG를 수행한다.

```
PROC AUTOREG DATA=EXAM51C PARTIAL;
MODEL Y=X / METHOD=ML NLAG=(1 4 5);
RUN;
```

이 PROC AUTOREG의 MODEL스태이트먼트는 다음 모형을 추정하기 위한 것이다.

$$y_t = \beta_0 + \beta_1 x_t + \frac{1}{1 + \alpha_1 B + \alpha_4 B^4 + \alpha_5 B^5} v_t$$

즉, MODEL스태이트먼트의 옵션 NLAG=에는 잡음항의 AR특성식의 형태를 지정한 다. PROC AUTOREG에서 잡음항의 특성식의 형태를 정하는 것은 PROC ARIMA에서 잡음항의 특성식의 형태를 정하는 것보다는 쉽다. 여기서는 앞의 PROC ARIMA에서 사용한 잡음항의 AR특성식의 형태를 그대로 사용하였다. 옵션 METHOD=에는 추정방법을 지정한다. 만일, 옵션 METHOD=에 YW가 지정되면 Yule-Walker추정법이 사용되고, ULS가 지정되면 비조건부최소제곱추정법이 사용되며, ML이 지정되면 최우추정법이 사용된다. 이 추정법들에 대한 자세한 내용은 [SAS4TSA2] (pp. 503~509)를 참조하라. PROC AUTOREG스태이트먼트의 옵션 PARTIAL은 표본부분자기상관함수를 OUTPUT윈도우에 출력시키기 위한 것으로서 MODEL스태이트먼트에서 사용할 수도 있다.

이 PROC AUTOREG를 수행한 결과가 그림 5.1.6이다. 그림 5.1.6의 부분 A는 보통최소제곱추정법을 적용해서 회귀모형을 추정한 것이다. 따라서 부분 A는 PROC REG를 수행한 결과인 그림 5.1.2와 동일하다. 그림 5.1.2에서와 마찬가지로 그림 5.1.6의 부분 A에서도 추정된 회귀모형의 결정계수가 0.994로 상당히 크나 Durbin-Watson통계량값이 0.2351이므로 주어진 시계열데이터를 잡음항이 AR과정인 회귀모형에 적합시키면 더 좋은 결과를 얻을 것이다. 부분 A2의 'Estimates of Autocorrelations'에는 보통최소제곱추정법으로 추정된 회귀모형에 의한 잔차의 표본자기공분산함수와 표본자기상관함수가 출력되었다. 또한 'Partial Autocorrelations'에는 잔차의 표본부분자기상관함수가 출력되었는데 이는 PROC AUTOREG스태이트먼트의 옵션 PARTIAL을 사용했기 때문에 출력된 것이다. 부분 A2의 'Estimate of Autoregressive Parameters'는 보통최소제곱추정법에 의한 잔차 $\{\hat{n}_t\}$ 의 표본자기상관함수를 다음과 같은 부분AR(5)모형의 Yule-Walker방정식들에 대입해서 구한 $\alpha_1, \alpha_4, \alpha_5$ 의 Yule-Walker추정값들이다.

$$\hat{n}_t = v_t - \alpha_1 \hat{n}_{t-1} - \alpha_4 \hat{n}_{t-4} - \alpha_5 \hat{n}_{t-5}$$

이 부분 AR모형의 오차항 $\{v_t\}$ 의 분산 σ_v^2 의 추정값이 부분 A2의 'Preliminary MSE'에 실려 있다. 이 예제에서 잡음항의 추정모형은 다음과 같다.

$$\hat{n}_t = 0.834\hat{n}_{t-1} + 0.759\hat{n}_{t-4} - 0.668\hat{n}_{t-5} + v_t$$

여기서 오차항의 분산 σ_v^2 의 추정값은 1.075이다. 또한 잡음항의 부분AR모형의 AR계수들이 모두 유의적으로 0과 다름을 알 수 있다. 부분 A2의 'Expected Autocorrelations'에 대해서는 SAS의 공식지침서 [ETS6.1]에 설명이 나오지 않지만 본 저자의 생각으로는 추정된 부분AR모형의 자기상관함수인 것 같다. 부분 A에 관한 자세한 설명은 [SAS4TSA2] (pp. 267~271)를 참조하라.

그림 5.1.6의 부분 B는 최우추정법을 사용해서 다음 모형을 추정한 결과이다.

$$y_t = \beta_0 + \beta_1 x_t + \frac{1}{1 - \alpha_1 B + \alpha_4 B^4 + \alpha_5 B^5} v_t$$

부분 B에서 추정된 회귀모형이 다음과 같음을 알 수 있다.

$$y_t = 24.481 + 0.982 x_t + \frac{1}{1 - 0.834 B - 0.773 B^4 + 0.681 B^5} v_t$$

이 추정모형은 PROC ARIMA를 사용해서 회귀모형을 최우추정한 그림 5.1.2의 추정식과 거의 동일하다. 여기서 다시 강조할 점은 Durbin-Watson통계량값이 1.9451로 2에 가깝다는 것이다. 따라서 이 추정된 회귀모형의 잔차는 계열상관을 갖지 않는다는 것을 알 수 있다. 부분 B2의 'Expected Autocorrelations'에 대해서도 SAS의 공식지침서 [ETS6.1]에 설명이 나오지 않지만 본 저자의 생각으로는 잡음항의 추정된 부분AR모형의 자기상관함수인 것 같다. 부분 B2는 잡음항의 AR계수들이 주어졌다고 가정하고 최우추정법을 사용해서 회귀모형을 추정한 결과이다. 즉, 다음 모형을 추정한 결과이다.

$$y_t = \beta_0 + \beta_1 x_t + \frac{1}{1 - 0.834 B - 0.773 B^4 + 0.681 B^5} v_t$$

이 회귀모형을 추정한 결과는 부분 B의 AR계수들이 주어졌다고 가정하지 않고 최우추정법을 사용해서 회귀모형을 추정한 결과와 동일하다. 부분 B에 관한 자세한 설명은 [SAS4TSA2] (pp. 277~281)를 참조하라.

주어진 시계열데이터를 잡음항이 AR과정인 회귀모형에 적합시켜서 미래값들을 예측하기 위해서 다음 PROC AUTOREG를 수행한다.

```
PROC AUTOREG DATA=EXAM51B;
  MODEL Y=X/METHOD=ML NLAG=(1 4 5) NOPRINT;
  OUTPUT OUT=FORE51B P=PRED LCL=L95 UCL=U95;
  RUN;
```

Autoreg Procedure

(A)

Dependent Variable = Y

(A1)

Ordinary Least Squares Estimates

SSE	3355.137	DFE	298
MSE	11.25885	Root MSE	3.355421
SBC	1587.11	AIC	1579.703
Reg Rsq	0.9940	Total Rsq	0.9940
Durbin-Watson	0.2351		

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	25.3416078	1.4782	17.144	0.0001
X	1	0.9798267	0.0044	221.619	0.0001

(A2)

Estimates of Autocorrelations

Lag	Covariance	Correlation
0	11.18379	1.000000
1	9.866538	0.882218
2	9.346773	0.835743
3	8.933757	0.798813
4	9.358056	0.836752
5	7.832201	0.700317

Partial Autocorrelations

1	0.882218
4	0.364812
5	-0.667793

Preliminary MSE = 1.075248

Estimates of the Autoregressive Parameters

Lag	Coefficient	Std Error	t Ratio
1	-0.83441999	0.03391926	-24.600179
4	-0.75934488	0.03944765	-19.249434
5	0.66779276	0.04333757	15.409095

Expected Autocorrelations

Lag	Autocorr
0	1.0000
1	0.8865
2	0.8246
3	0.8106
4	0.8437
5	0.7094

(B)

(B1)

Maximum Likelihood Estimates

SSE	306.8895	DFE	295
MSE	1.040303	Root MSE	1.019953
SBC	892.2232	AIC	873.7043
Reg Rsq	0.9199	Total Rsq	0.9994
Durbin-Watson	1.9451		

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	24.4809064	5.6595	4.326	0.0001
X	1	0.9823607	0.0169	58.075	0.0001
A(1)	1	-0.8340241	0.0334	-24.939	0.0001
A(4)	1	-0.7729780	0.0385	-20.073	0.0001
A(5)	1	0.6813238	0.0428	15.909	0.0001

Expected Autocorrelations					
	Lag	Autocorr			
	0	1.0000			
	1	0.8820			
	2	0.8189			
	3	0.8068			
	4	0.8449			
	5	0.7051			

(B2)

Autoregressive parameters assumed given.

Maximum Likelihood Estimates					
SSE	306.8895	DFE	295		
MSE	1.040303	Root MSE	1.019953		
SBC	892.2232	AIC	873.7043		
Reg Rsq	0.9199	Total Rsq	0.9994		
Durbin-Watson	1.9451				

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	24.4809064	5.6595	4.326	0.0001
X	1	0.9823607	0.0169	58.075	0.0001

그림 5.1.6 PROC AUTOREG를 사용한 회귀모형의 추정

이 PROC AUTOREG스테이트먼트와 MODEL스테이트먼트는 각각 앞의 PROC AUTOREG스테이트먼트와 MODEL스테이트먼트와 거의 같지만 MODEL스테이트먼트에 옵션 NOPRINT를 사용한 것이 다르다. 이 옵션 NOPRINT는 MODEL스테이트먼트를 수행한 결과를 OUTPUT윈도우에 출력을 시키지 않도록 한다. OUTPUT스테이트먼트는 추정모형을 사용해서 미래값들을 예측한 결과를 저장하기 위한 것이다. OUTPUT스테이트먼트의 옵션 OUT=에는 예측된 결과를 저장할 SAS데이터세트 FORE51B를 지정하였다. 또한 옵션 P=, 옵션 LCL=, 그리고 옵션 UCL=은 각각 예측(또는 적합), 신뢰수준 95%의 예측구간의 하한과 상한의 이름들을 지정하는 데 사용한다. 여기서 강조할 점은 PROC AUTOREG를 사용해서 예측할 때에는 예측할 미래시점들에서 입력변수들의 값들이 PROC AUTOREG스테이트먼트의 옵션 DATA=에 지정된 SAS데이터세트에 수록되어 있어야 한다는 것이다. 이 PROC AUTOREG를 수행하면 아무런 결과도 OUTPUT윈도우에 나타나지 않는다.

이 예측의 결과를 OUTPUT윈도우에 나타내기 위해서 다음 PROC PRINT와 PROC PLOT를 수행한다.

```

PROC PRINT DATA=FORE51B(FIRSTOBS=301);
RUN;
PROC PLOT DATA=FORE51B(FIRSTOBS=276);
PLOT Y*T='Y' PRED*T='P' L95*T='L'
U95*T='U'/OVERLAY VPOS=27
VAXIS= 405 TO 445 BY 5;
RUN;
    
```

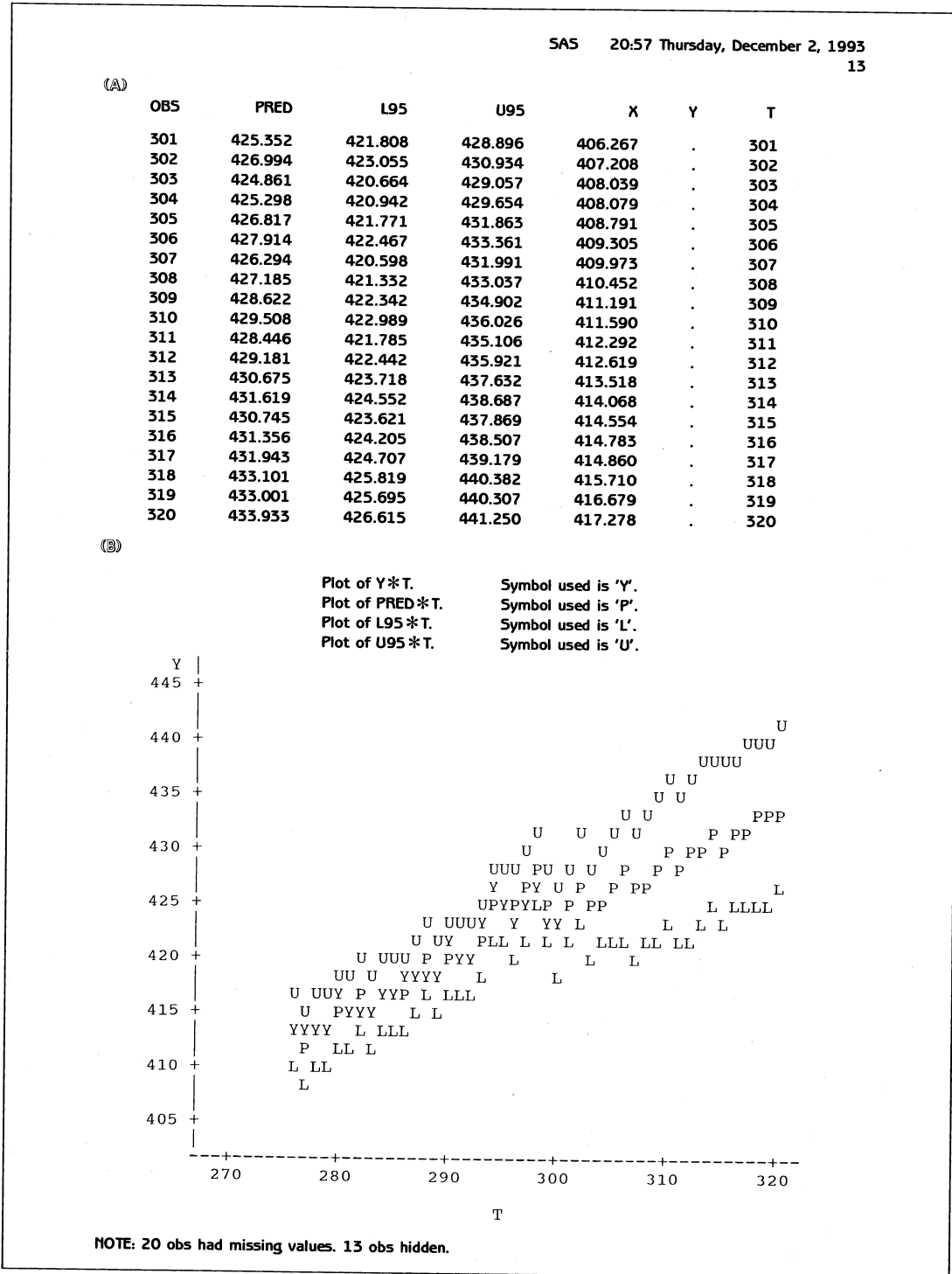


그림 5.1.7 PROC AUTOREG에 의한 예측결과

이 프로시저들을 수행한 결과가 그림 5.1.7이다. 그림 5.1.7의 부분 A에는 PROC PRINT를 수행한 결과가 수록되어 있고, 부분 B에는 PROC PLOT를 수행한 결과가 수록되어 있다. 그림 5.1.7의 부분 A는 그림 5.1.4와 거의 동일하고, 부분 B는 그림 5.1.5와 거의 동일하다.

SAS/ETS를 이용해서 시계열데이터 형태의 잡음항을 갖는 회귀모형을 추정할 때, 잡음항이 AR과정이면 PROC AUTOREG를 사용하고 잡음항이 ARMA과정이면 PROC ARIMA를 사용한다는 것을 알았다. 따라서 PROC AUTOREG로 분석할 수 있는 시계열데이터는 PROC ARIMA로도 분석할 수 있다. 더구나 PROC AUTOREG에서는 차분(differencing)을 허용하지 않기 때문에 PROC ARIMA의 적용범위가 PROC AUTOREG의 적용범위보다 넓다. 물론 변칙적인 기법을 사용하면 PROC AUTOREG에서도 차분된 시계열데이터를 분석할 수도 있다. 즉, 우선 DATA스텝에서 차분된 시계열데이터를 만든 뒤 PROC AUTOREG를 적용한다. 그러나 이 경우에는 예측값들과 표준오차들을 다루는 데 어려움이 있다. 반면에 PROC AUTOREG에서는 옵션 BACKSTEP과 옵션 SLS-TAY=을 사용하면 잡음항의 AR모형을 자동적(automatic)으로 정할 수 있다. 또한, PROC AUTOREG에서는 결측(missing)이 있는 시계열데이터를 분석할 수 있다. 즉, 종속변수의 값만이 결측인 관찰점들에서는 OUTPUT스테이트먼트를 사용해서 종속변수의 적합값을 얻을 수 있다. 이에 대한 자세한 내용은 [SAS4TSA2] (pp. 293~297)를 참조하라.

두 번째 예제는 시계열데이터 HOUSECON을 잡음항이 백색잡음과정이 아닌 ARMA과정을 따르는 회귀모형으로 분석하는 것이다. 시계열데이터 HOUSECON은 1983년 1월부터 1989년 10월까지의 미국의 월별 건설도금액을 나타내는 변수 CONSTR, 월별 착공한 주택의 수를 나타내는 변수 HSTARTS, 그리고 월평균 신주택자금이자율(average new home mortgage rates)을 나타내는 변수 INTRATE의 관찰값들을 기록한 것이다. 이 시계열데이터의 관찰점들의 개수는 82이다. 시계열데이터 HOUSECON은 제 2.4절에 수록되어 있으며, 또한 SAS데이터라이브러리 TSA4DATA의 파일 HOUSECON.SSD에 수록되어 있다. 이 시계열데이터는 [AppGuide] (pp. 151~172)에서 잡음항이 ARMA모형을 따르는 회귀모형으로 분석하였다. 본서의 제 4.7.4소절에서는 이 시계열데이터를 유리함수형태의 전이함수모형화하였다.

다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM5-2.PRG에 수록되어 있다).

```

/*                                     */
/*                                     EXAM5-2.PRG                               */
/*                                     */
/*                                     REGRESSION WITH TIME SERIES ERROR        */
/*                                     DATASET = HOUSECON                        */
/*                                     */
LIBNAME TSA4DATA 'A:TSA4DATA';
OPTIONS LINESIZE=72 PAGESIZE=60;
DATA EXAM52:
    SET TSA4DATA.HOUSECON;
    TIME=_N_;
RUN;
```

DATA EXAM52B:

INPUT DATE: MONYY5. CONSTR INTRATE @@ ;

FORMAT DATE MONYY5.;

CARDS:

NOV89	20137.12	9.7576
DEC89	19850.10	9.7458
JAN90	16403.76	9.7091
FEB90	15785.50	9.6725
MAR90	21860.87	9.6358
APR90	22495.57	9.5992
MAY90	24609.96	9.5625
JUN90	25485.82	9.5259
JUL90	23174.28	9.4892
AUG90	24991.23	9.4525
SEP90	24243.09	9.4159
OCT90	23548.71	9.3792

RUN;

DATA EXAM52C:

SET EXAM52 EXAM52B:

TIME = _N_ ;

RUN;

PROC ARIMA DATA=EXAM52:

I VAR=HSTARTS(12) CROSS=(CONSTR(12) INTRATE(1))

NLAG=15 NOPRINT;

E INPUT=(CONSTR INTRATE) PLOT METHOD=ML;

E P=2 INPUT=(CONSTR INTRATE) METHOD=ML;

E INPUT=CONSTR METHOD=ML PLOT;

E P=2 INPUT=CONSTR METHOD=ML;

E INPUT=INTRATE PLOT METHOD=ML;

E P=2 INPUT=INTRATE METHOD=ML;

RUN;

PROC PRINT DATA=EXAM52B(OBS=10);

RUN;

PROC ARIMA DATA=EXAM52C:

I VAR=HSTARTS(12) CROSS=CONSTR(12)

NLAG=15 NOPRINT;

E P=2 INPUT=CONSTR METHOD=ML NOPRINT;

F LEAD=12;

RUN;


```

PROC AUTOREG DATA=EXAM52;
  MODEL HSTARTS=CONSTR INTRATE
    /NLAG=4 BACKSTEP SLSTAY=0.06 METHOD=ML;
RUN;
PROC AUTOREG DATA=EXAM52;
  MODEL HSTARTS=CONSTR INTRATE TIME
    /NLAG=4 BACKSTEP SLSTAY=0.06 METHOD=ML;
RUN;
PROC AUTOREG DATA=EXAM52;
  MODEL HSTARTS=CONSTR TIME
    /NLAG=4 BACKSTEP SLSTAY=0.06 METHOD=ML;
RUN;
PROC AUTOREG DATA=EXAM52C;
  MODEL HSTARTS=CONSTR TIME /METHOD=ML NOPRINT;
  OUTPUT OUT=FORE52A P=PRED LCL=L95 UCL=U95;
RUN;
PROC PRINT DATA=FORE52A(FIRSTOBS=83);
RUN;

```

첫 번째 DATA스텝에서는 시계열데이터 HOUSECON의 82개 관찰점들을 읽어 SAS데이터세트 EXAM52를 만든다. 두 번째 DATA스텝에서는 변수 HSTARTRS의 미래값들을 예측하기 위해서 제83번째부터 제94번째까지의 관찰점들의 입력변수들 CONSTR와 INTRATE의 관찰값들을 읽어 SAS데이터세트 EXAM52B에 저장한다. 세 번째 DATA스텝에서는 SAS데이터세트들 EXAM52와 EXAM52B를 세로결합해서 새로운 SAS데이터세트 EXAM52C를 만든다.

시계열데이터 HOUSECON의 변수 CONSTR의 시계열산포도가 그림 2.4.1에 수록되어 있고, 변수 INTRATE의 시계열산포도가 그림 2.4.2에 수록되어 있으며, 변수 HSTARTS의 시계열산포도가 그림 2.4.3에 수록되어 있다. 그림 2.4.1의 시계열산포도로부터 건설도금액은 여름에 많고 겨울에는 적은 계절성을 가지고 있으며 시간이 흐름에 따라 장기적으로 증가하는 추세를 보임을 알 수 있다. 또한, 그림 2.4.3의 시계열산포도로부터 착공주택수도 여름에 많고 겨울에는 적은 계절성을 가지고 있고, 1986년까지는 시간이 흐름에 따른 수준(level)의 변화가 없었으나 그 이후로는 감소하는 추세를 보임을 알 수 있다. 이러한 현상은 인플레이션에 따라 달러의 가치가 하락했음에도 불구하고 건설도금액을 달러의 명목가액(unadjusted price)으로 기록했고, 반면에 착공주택수는 실제 착공된 주택의 수를 기록한 때문이다. 그림 2.4.2의 시계열산포도로부터 신주택자금이자율은 1988년까지 감소하다가 그 이후에는 증가하고 있음을 알 수 있다. 착공주택수는 건설도금액을 결정하는 하나의 요소이다. 그림 2.4.1과 그림 2.4.3의 시계열산포도들에서 두 시계열데이터들이 매우 비슷한 패턴을 가지고 있음을 알 수 있다. 따라서 착공주택수와 건설도금액은 양의 상관관계에 있

다는 가설을 세울 수 있다. 또한, 신주택자금이자율이 낮으면 착공주택수가 증가할 것이다. 따라서 신주택자금이자율과 착공주택수는 음의 상관관계가 있다는 가설을 세울 수 있다. 그러나 착공주택수가 건설도금액과 상관관계가 높기 때문에 착공주택수의 변동량의 대부분을 건설도금액으로 설명할 수 있을 것이다. 따라서 착공주택수를 종속변수로 건설도금액과 신주택자금이자율을 설명변수들로 하는 회귀모형에서는 신주택자금이자율이 의미있는 설명변수가 아닐 것이다. 이 예제에서는 이러한 가설들을 분석하게 될 것이다.

시계열데이터 HOUSECON을 잡음항이 ARMA과정인 회귀모형에 적합시키기 위해서 다음 PROC ARIMA를 수행한다.

```
PROC ARIMA DATA=EXAM52;
  I VAR=HSTARTS(12) CROSS=(CONSTR(12) INTRATE(1))
  NLAG=15 NOPRINT;
  E INPUT=(CONSTR INTRATE) PLOT METHOD=ML;
  E P=2 INPUT=(CONSTR INTRATE) METHOD=ML;
  E INPUT=CONSTR METHOD=ML PLOT;
  E P=2 INPUT=CONSTR METHOD=ML;
  E INPUT=INTRATE PLOT METHOD=ML;
  E P=2 INPUT=INTRATE METHOD=ML;
  RUN;
```

이 PROC ARIMA의 IDENTIFY스태이트먼트의 옵션 VAR=에는 종속변수를 지정하고 옵션 CROSSCOR=에는 입력변수들을 지정한다. 여기서 종속변수는 변수 HSTARTS를 계절주기 12로 계절차분한 HSTARTS(12)이고, 입력변수들은 변수 CONSTR를 계절주기 12로 계절차분한 CONSTR(12)와 변수 INTRATE를 차분한 INTRATE(1)이다. 여기서 변수들을 계절차분이나 차분을 하는 이유는 각각의 변수를 정상적인 변수로 변형하기 위한 것이다. 이렇게 차분이나 계절차분을 사용해서 정상적 시계열데이터로 변형시키는 과정에 대한 자세한 내용은 [SAS4TSA2] (pp. 414~422, pp. 623~637)를 참조하라. IDENTIFY스태이트먼트의 옵션 NOPRINT를 사용했기 때문에 이 IDENTIFY스태이트먼트를 수행한 결과가 OUTPUT윈도우에 나타나지 않는다. 만일, 옵션 NOPRINT를 사용하지 않은 경우에는 우선 옵션 VAR=에 지정한 종속변수 HSTARTS(12)의 표본자기상관함수, 표본역자기상관함수, 그리고 표본부분자기상관함수와 옵션 CROSS=에 지정한 입력변수들 CONSTR(12)와 INTRATE(1) 각각과 옵션 VAR=에 지정한 변수 HSTARTS(12)의 표본교차상관함수들이 OUTPUT윈도우에 출력된다. 표본교차상관함수는 해당하는 입력변수와 종속변수 사이의 종속성(interdependency)을 조사하는데 유용하다. 따라서 입력변수의 시차변수들(lagged variables)과 종속변수가 독립이라는 가정하에서는 표본교차상관함수는 그리 유용하지가 않다. 즉, 전이함수모형을 가정하지 않으면 표본교차상관함수를 OUTPUT윈도우에 출력시킬 필요가 없다. 전이함수모형을 가정하는 경우에도 종속변수들을 사전백색화하지 않으면 표본교차상관함수는 모형의 식별에 별 도움이 되지

않는다. 결과적으로 잡음항이 ARMA과정인 회귀모형을 추정하기 위해서 PROC ARIMA를 사용할 때 IDENTIFY스테이트먼트는 단지 종속변수와 입력변수들 또는 이들의 차분 차수들을 지정하는 데 사용한다.

이후 이 PROC ARIMA를 수행한 결과를 설명하는 동안에는 편의상 다음과 같은 기호들을 사용하기로 하자.

$$y_t = \nabla_{12} \text{HSTARTS}_t = \text{HSTARTS}_t - \text{HSTARTS}_{t-12}$$

$$x_{t,1} = \nabla_{12} \text{CONSTR}_t = \text{CONSTR}_t - \text{CONSTR}_{t-12}$$

$$x_{t,2} = \nabla \text{INTRATE}_t = \text{INTRATE}_t - \text{INTRATE}_{t-1}$$

이 PROC ARIMA의 첫 번째 ESTIMATE스테이트먼트는 다음 모형을 추정하기 위한 것이다.

$$y_t = \beta_0 + \beta_1 x_{t,1} + \beta_2 x_{t,2} + v_t$$

이 ESTIMATE스테이트먼트는 보통최소제곱추정법을 사용해서 이 회귀모형의 회귀계수들을 추정하기 위한 것이다. 옵션 PLOT는 이렇게 추정된 회귀모형의 잔차의 표본자기상관함수, 표본역자기상관함수, 그리고 표본부분자기상관함수를 OUTPUT윈도우에 출력시킨다. 이 표본자기상관함수들을 사용해서 잡음항의 AR차수와 MA차수를 식별한다. ESTIMATE스테이트먼트의 옵션 INPUT=에는 입력변수들을 지정한다. 앞서서도 언급했듯이 이 입력변수들은 이 ESTIMATE스테이트먼트 바로 앞의 IDENTIFY스테이트먼트의 옵션 CROSSCOR=에 지정되어 있어야 한다. 그러나 ESTIMATE스테이트먼트에서는 종속변수나 입력변수의 차분차수나 계절차분차수를 따로 지정하지 않고 이 ESTIMATE스테이트먼트에 해당하는 IDENTIFY스테이트먼트에서 지정한 차수들을 그대로 사용한다. 앞서서도 언급했듯이 옵션 METHOD=에는 추정방법을 지정한다. 만일, 이 옵션에 CLS가 지정되면 조건부최소제곱추정법이 사용되고, ULS가 지정되면 비조건부최소제곱추정법이 사용되며, ML이 지정되면 최우추정법이 사용된다. 이 ESTIMATE스테이트먼트를 수행한 결과가 그림 5.2.1이다. 그림 5.2.1의 부분 A에서 최우추정된 회귀모형이 다음과 같음을 알 수 있다.

$$y_t = -8.222 - 0.00466x_{t,1} - 1.815x_{t,2} + v_t$$

그러나 이 회귀모형은 큰 의미를 지니지 않는다는 것을 독자들은 알고 있을 것이다. 부분 B에는 퍼트맨토우통계량값들로부터 이 추정된 회귀모형의 잔차가 백색잡음과정에서 나온 것이 아님을 알 수 있다. 즉, 이 회귀모형이 타당한 것이 아님을 확인할 수 있다. 부분 C에는 이렇게 추정된 회귀모형의 잔차의 표본자기상관함수, 표본역자기상관함수, 그리고 표본부분자기상관함수들이 수록되어 있다. 이 표본자기상관함수들은 ESTIMATE스테이트먼트의 옵션 PLOT를 사용함으로써 출력된 것이다. 표본자기상관함수는 시나브로형태를 보이고, 표본역자기상관함수와 표본부분자기상관함수는 시차 3 이후에서 절단형태를 보인다. 따라서 Box-Jenkins식별방법에 의하면 이 잔차는 AR(2)모형에서 출력되었다고 할 수 있다.

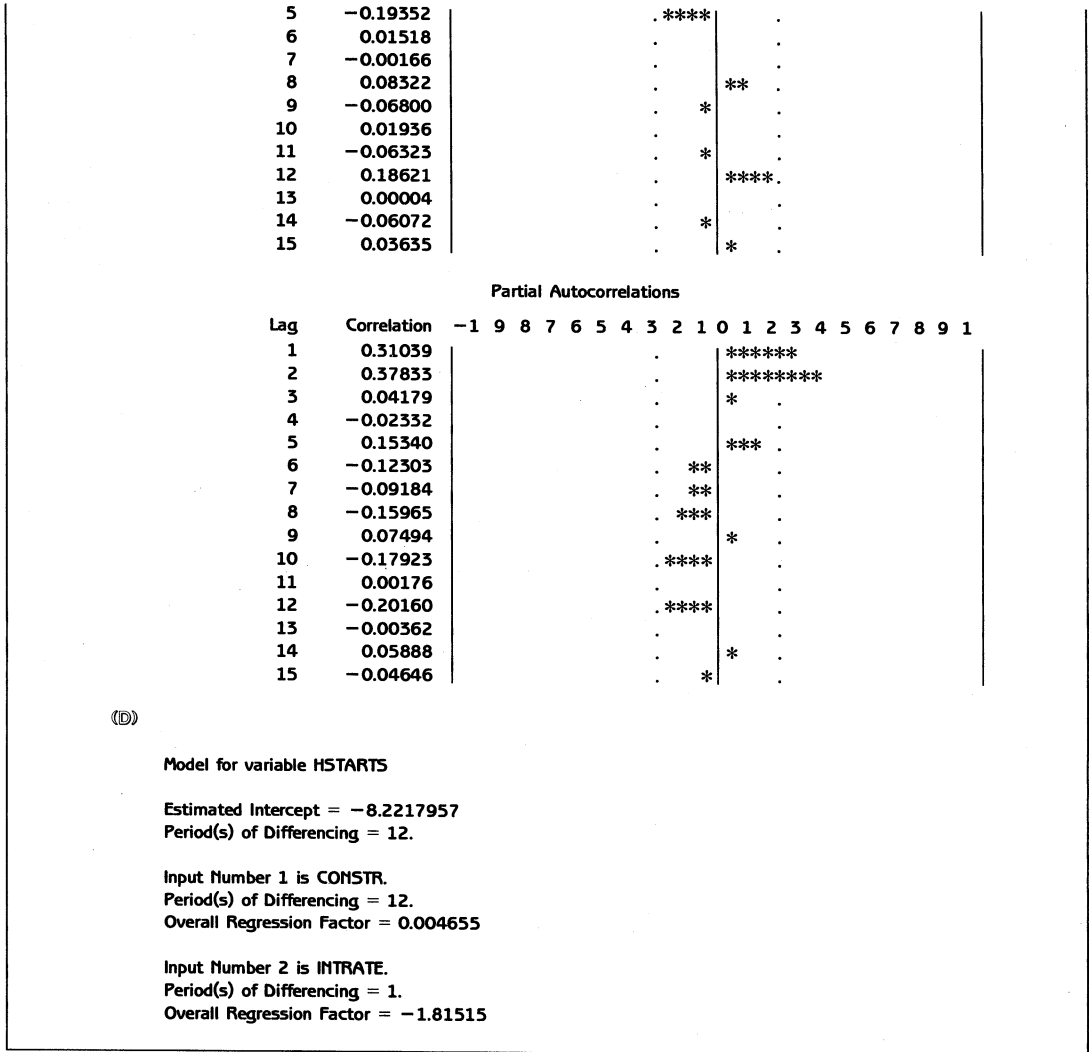


그림 5.2.1 보통최소제곱추정법을 사용한 회귀모형 추정 결과

그림 5.2.1에서 잡음항이 AR(2)모형을 갖는다는 것을 알 수 있다. 두 번째 ESTIMATE스 테이트먼트는 다음과 같은 잡음항이 AR(2)모형을 갖는 회귀모형을 추정하기 위한 것이다.

$$y_t = \beta_0 + \beta_1 x_{t,1} + \beta_2 x_{t,2} + \frac{1}{1 - \phi_1 B - \phi_2 B^2} v_t$$

이 ESTIMATE스 테이트먼트를 수행한 결과가 그림 5.2.2이다. 그림 5.2.2에서 최우추정 된 회귀모형이 다음과 같음을 알 수 있다.

$$y_t = -6.775 + 0.00356 x_{t,1} - 3.007 x_{t,2} + \frac{1}{1 - 0.229 B - 0.399 B^2} v_t$$

이 추정모형에서 $x_{t,2}$ 의 추정계수가 유의적으로 0과 다르지 않음을 알 수 있고, 카이제곱 통계량값들, 즉 퍼트맨토우통계량값들로부터 이 추정된 회귀모형의 잔차는 백색잡음과정

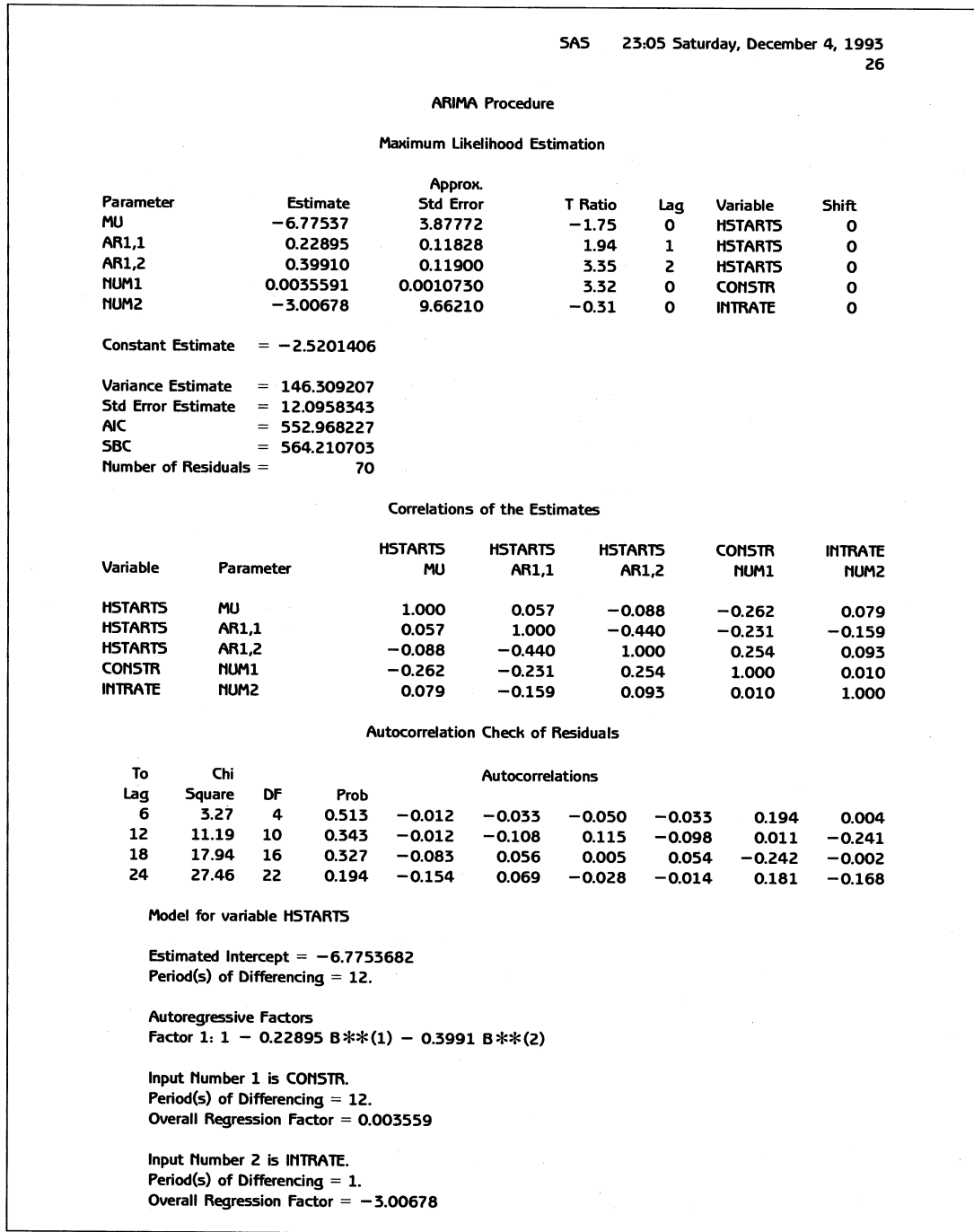


그림 5.2.2 입력변수들이 CONSTR(12)와 INTRATE(1)인 회귀모형

에서 나왔다고 할 수 있다.

앞에서 설명했듯이 착공주택수가 건설도금액과 상관관계가 높기 때문에 착공주택수의 변동량의 대부분을 건설도금액으로 설명할 수 있을 것이다. 따라서 계절차분된 착공주택 수 HSTARTS(12)를 종속변수로 계절차분된 건설도금액 CONSTR(12)와 차분된 신주택 자금이자율 INTRATE(1)을 설명변수들로 하는 회귀모형에서는 변수 INTRATE(1)이

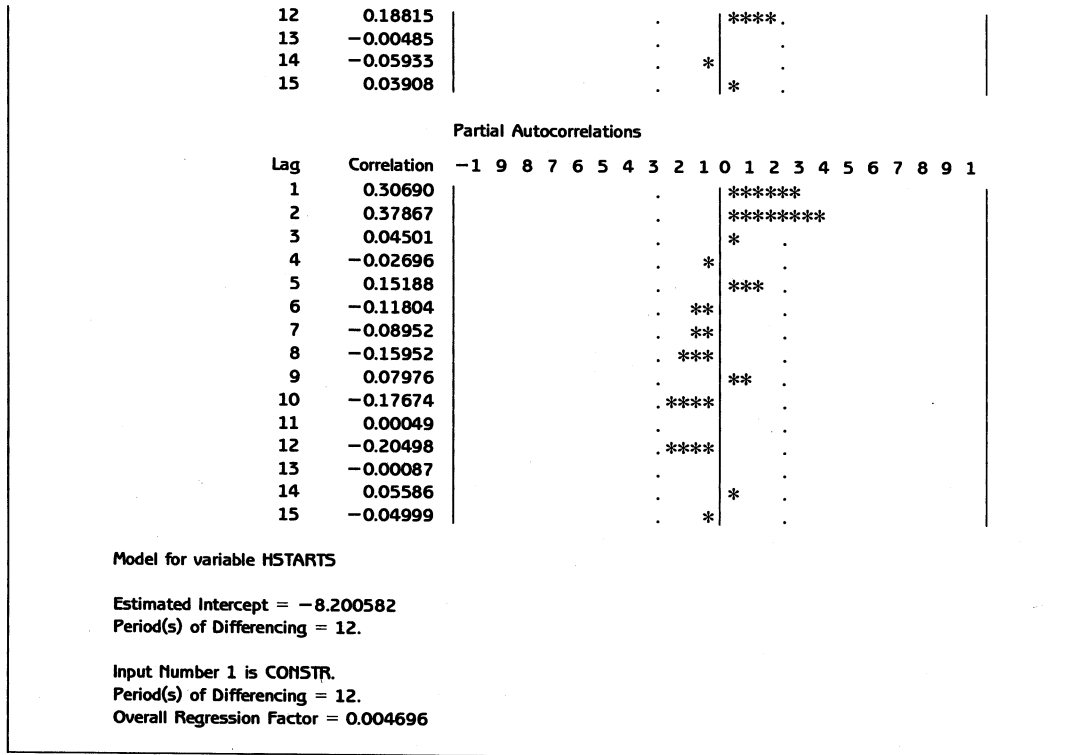


그림 5.2.3 입력변수를 CONSTR(12)로 하는 보통최소제곱추정법의 결과

의미있는 설명변수가 아닐 것이라는 가설을 세울 수 있다. 그림 5.2.2에서 이 가설이 사실임을 알 수 있다. 세 번째 ESTIMATE스테이트먼트는 계절차분된 건설도금액 CONSTR(12)만을 설명변수로 하는 다음 모형을 추정하기 위한 것이다.

$$y_t = \beta_0 + \beta_1 x_{t,1} + v_t$$

이 ESTIMATE스테이트먼트를 수행한 결과가 그림 5.2.3이다. 그림 5.2.3에서 최우추정된 회귀모형이 다음과 같음을 알 수 있다.

$$y_t = -8.201 - 0.00470 x_{t,1} + v_t$$

이 추정모형의 퍼트맨도우통계량값들로부터 추정된 회귀모형의 잔차가 백색잡음과정에서 나온 것이 아님을 알 수 있다. 즉, 이 회귀모형이 타당한 것이 아님을 확인할 수 있다. 또한, 잔차의 표본자기상관함수는 시나브로형태를 보이고, 표본역자기상관함수와 표본부분자기상관함수는 시차 3 이후에서 절단형태를 보인다. 따라서 Box-Jenkins식별방법에 의하면 이 잔차는 AR(2)모형에서 출력되었다고 할 수 있다.

네 번째 ESTIMATE스테이트먼트는 계절차분된 착공주택수 HSTARTS(12)를 종속변수로, 계절차분된 건설도금액 CONSTR(12)를 설명변수로 하고 오차항이 AR(2)모형인 다음과 같은 회귀모형을 추정하기 위한 것이다.

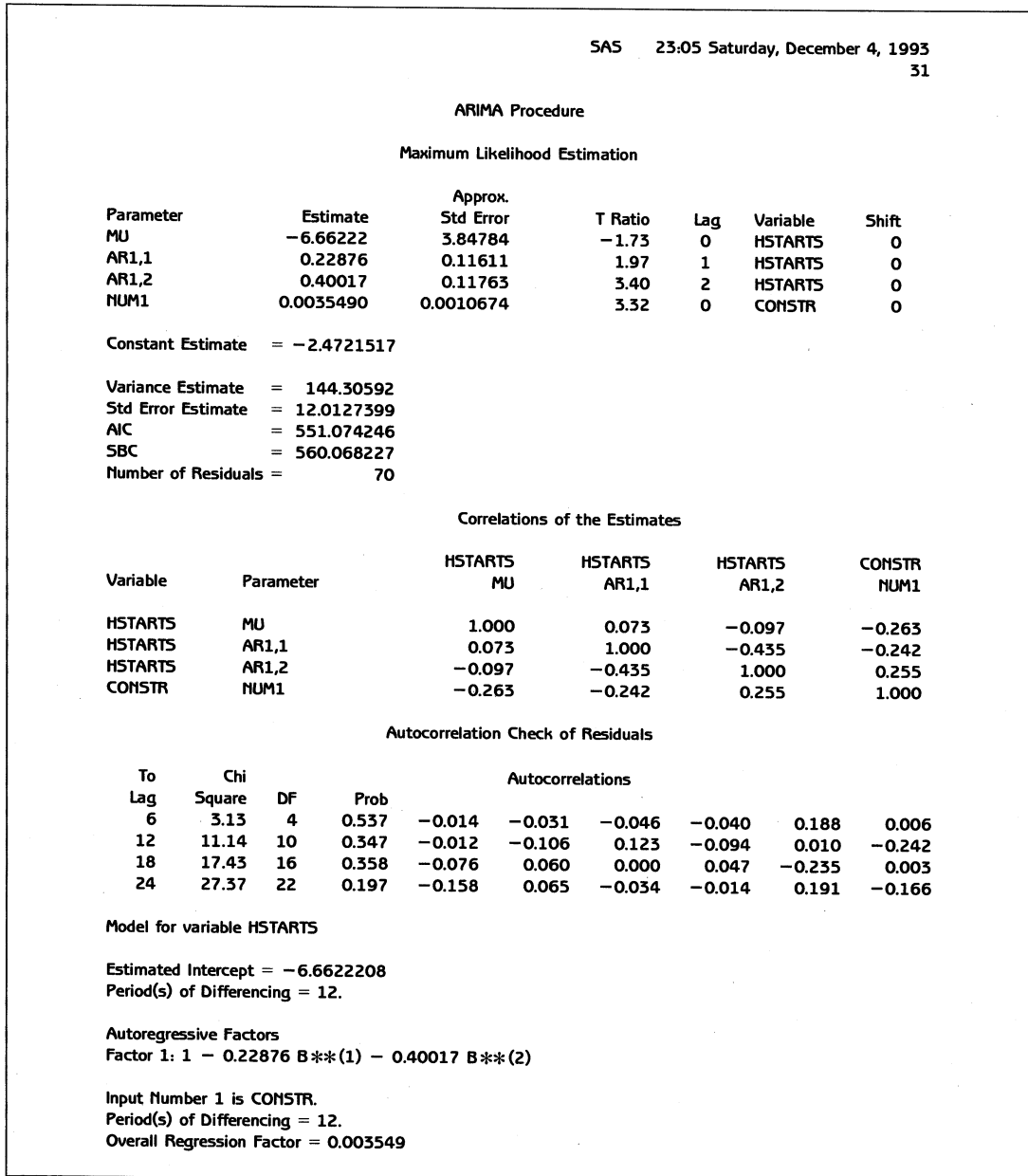


그림 5.24 입력변수가 CONSTR(12)이고 잡음항이 AR(2)과정인 회귀모형

$$y_t = \beta_0 + \beta_1 x_{t,1} + \frac{1}{1 - \phi_1 B - \phi_2 B^2} v_t$$

이 ESTIMATE스태이트먼트를 수행한 결과가 그림 5.2.4이다. 그림 5.2.4에서 최우추정된 회귀모형이 다음과 같음을 알 수 있다.

$$y_t = -6.662 + 0.00355 x_{t,1} + \frac{1}{1 - 0.229 B - 0.400 B^2} v_t$$

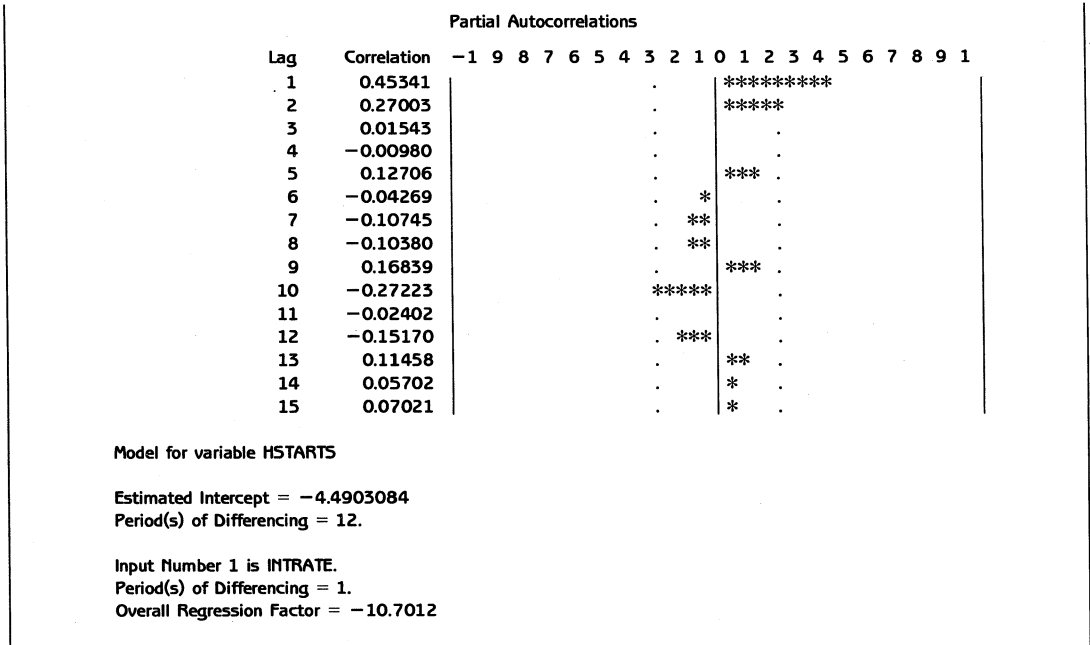


그림 5.2.5 입력변수를 INTRATE(1)로 하는 회귀모형

카이제곱통계량값들, 즉 퍼트맨토우통계량값들로부터 이 추정된 회귀모형의 잔차가 백색잡음과정에서 나왔다고 할 수 있다. 그림 5.2.2의 입력변수가 2개인 추정모형의 AIC와 BIC는 각각 553.0과 564.2이고 그림 5.2.3의 입력변수가 1개인 추정모형의 AIC와 BIC는 각각 551.1과 560.1이다. 따라서 정보이론의 관점에서 보면 입력변수가 CONSTR(12)와 INTRATE(1)인 추정모형보다는 입력변수가 CONSTR(12)인 추정모형이 종속변수 HSTARTS(12)를 설명하는 모형으로 더 타당하다는 것을 알 수 있다. 두 모형의 퍼트맨토우통계량값들의 크기를 비교해 봄으로써 이 결론을 확인할 수 있다.

다섯 번째 ESTIMATE스테이트먼트는 차분된 신주택자금이자율 INTRATE(1)만을 설명변수로 하는 다음 회귀모형을 추정하기 위한 것이다.

$$y_t = \beta_0 + \beta_2 x_{t,2} + v_t$$

이 ESTIMATE스테이트먼트를 수행한 결과가 그림 5.2.5이다. 그림 5.2.5에서 최우추정된 회귀모형이 다음과 같음을 알 수 있다.

$$y_t = -4.490 - 10.701x_{t,2} + v_t$$

이 추정모형의 퍼트맨토우통계량값들로부터 추정된 회귀모형의 잔차가 백색잡음과정에서 나온 것이 아님을 알 수 있다. 즉, 이 회귀모형이 타당한 것이 아님을 확인할 수 있다. 또한, 잔차의 표본자기상관함수는 시나브로형태를 보이고, 표본역자기상관함수는 시차 2 이후에서, 그리고 표본부분자기상관함수는 시차 3 이후에서 절단형태를 보인다. 따라서 Box-Jenkins식별방법에 의하면 이 잔차는 AR(1)모형 또는 AR(2)모형에서 출력되었다고 할 수 있다. 앞의 모형들과 균형을 맞추기 위해서 잡음항이 AR(2)모형에서 나왔다고 하자.

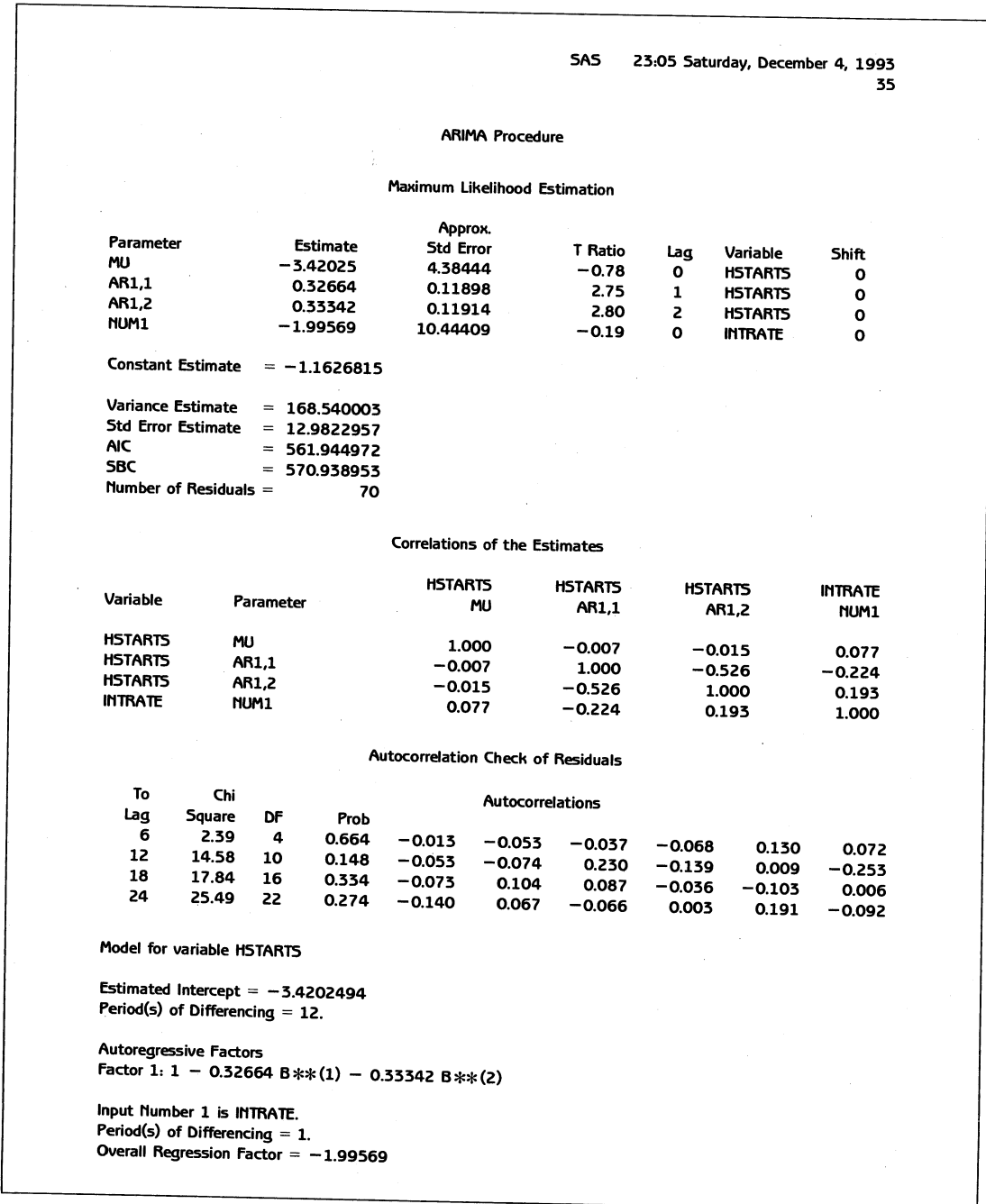


그림 5.2.6 입력변수가 INTRATE(1)이고 잡음항이 AR(2)과정인 회귀모형

여섯 번째 ESTIMATE 스테이트먼트는 계절차분된 착공주택수 HSTARTS(12)를 종속 변수로, 차분된 신주택자금이자율 INTRATE(1)을 설명변수로 하고 잡음항이 AR(2)모형인 다음과 같은 회귀모형을 추정하기 위한 것이다.

$$y_t = \beta_0 + \beta_2 x_{t,2} + \frac{1}{1 - \phi_1 B - \phi_2 B^2} v_t$$

이 ESTIMATE스태이트먼트를 수행한 결과가 그림 5.2.6이다. 그림 5.2.6에서 최우추정된 회귀모형은 다음과 같음을 알 수 있다.

$$y_t = -3.420 + 1.996x_{t,2} + \frac{1}{1 - 0.327B - 0.333B^2} v_t$$

카이제곱통계량값들, 즉 퍼트맨토우통계량값들로부터 이 추정된 회귀모형의 잔차가 백색잡음과정에서 나왔다고 할 수 있다. 이 추정된 회귀모형의 AIC와 BIC는 각각 561.9와 570.9이다. 따라서 정보이론의 관점에서 보면 입력변수가 INTRATE(1)인 추정모형보다는 입력변수가 CONSTR(12)인 추정모형이 종속변수 HSTARTS(12)를 설명하는 모형으로 더 타당하다는 것을 알 수 있다. 두 모형의 퍼트맨토우통계량값들의 크기를 비교해 봄으로써 이 결론을 재확인 할 수 있다.

시계열데이터 HOUSECON을 잡음항이 ARMA과정인 회귀모형에 적합시켜서 미래값들을 예측하기 위해서 다음 PROC PRINT와 PROC ARIMA를 수행한다.

```
PROC PRINT DATA=EXAM52B(OBS=10);
  RUN;
PROC ARIMA DATA=EXAM52C;
  I VAR=HSTARTS(12) CROSS=CONSTR(12)
  NLAG=15 NOPRINT;
  E P=2 INPUT=CONSTR METHOD=ML NOPRINT;
  F LEAD=12;
  RUN;
```

이 PROC PRINT의 옵션 DATA=에는 SAS데이터세트 EXAM52B가 지정되었다. 이 SAS데이터세트에는 입력변수 CONSTR의 미래값들이 12개 수록되어 있다. PROC PRINT스태이트먼트의 옵션 OBS=은 첫 번째 관찰점부터 이 옵션에 지정된 개수만큼의 관찰점만을 OUTPUT윈도우에 출력시키도록 한다. 이 PROC PRINT를 실행한 결과가 그림 5.2.7의 부분 A이다. 이 PROC ARIMA의 PROC ARIMA스태이트먼트의 옵션 DATA=에는 SAS데이터세트 EXAM52C가 지정되었다. 이 SAS데이터세트 EXAM52C는 SAS데이터세트 EXAM52와 SAS데이터세트 EXAM52B를 세로결합한 것이다. 이 PROC ARIMA에서는 종속변수 HSTARTS(12)를 가장 잘 설명하고 있는 그림 5.2.4의 추정모형을 사용한다. 즉, 입력변수를 CONSTR(12)로 하고 잡음항을 AR(2)과정으로 하는 회귀모형을 사용한다. FORECAST스태이트먼트의 옵션 LEAD=은 예측할 미래시점들을 지정하는 것이다. 이 예제에서는 옵션 LEAD=에 12가 지정되었으므로 12개의 미래시점들에 대해서 예측을 한다. 여기서 유의할 점은 비록 IDENTIFY스태이트먼트에서 종속변수 HSTARTS를 계절차분하였으나 FORECAST스태이트먼트에서 예측된 값들은 계절차분된 변수 HSTARTS(12)의 값들이 아니라 변수 HSTARTS의 값들이라는 것이다. 차분의 경우도 마찬가지로 유의하라. 이 PROC ARIMA를 실행한 결과가 그림 5.2.7의 부분 B이다.

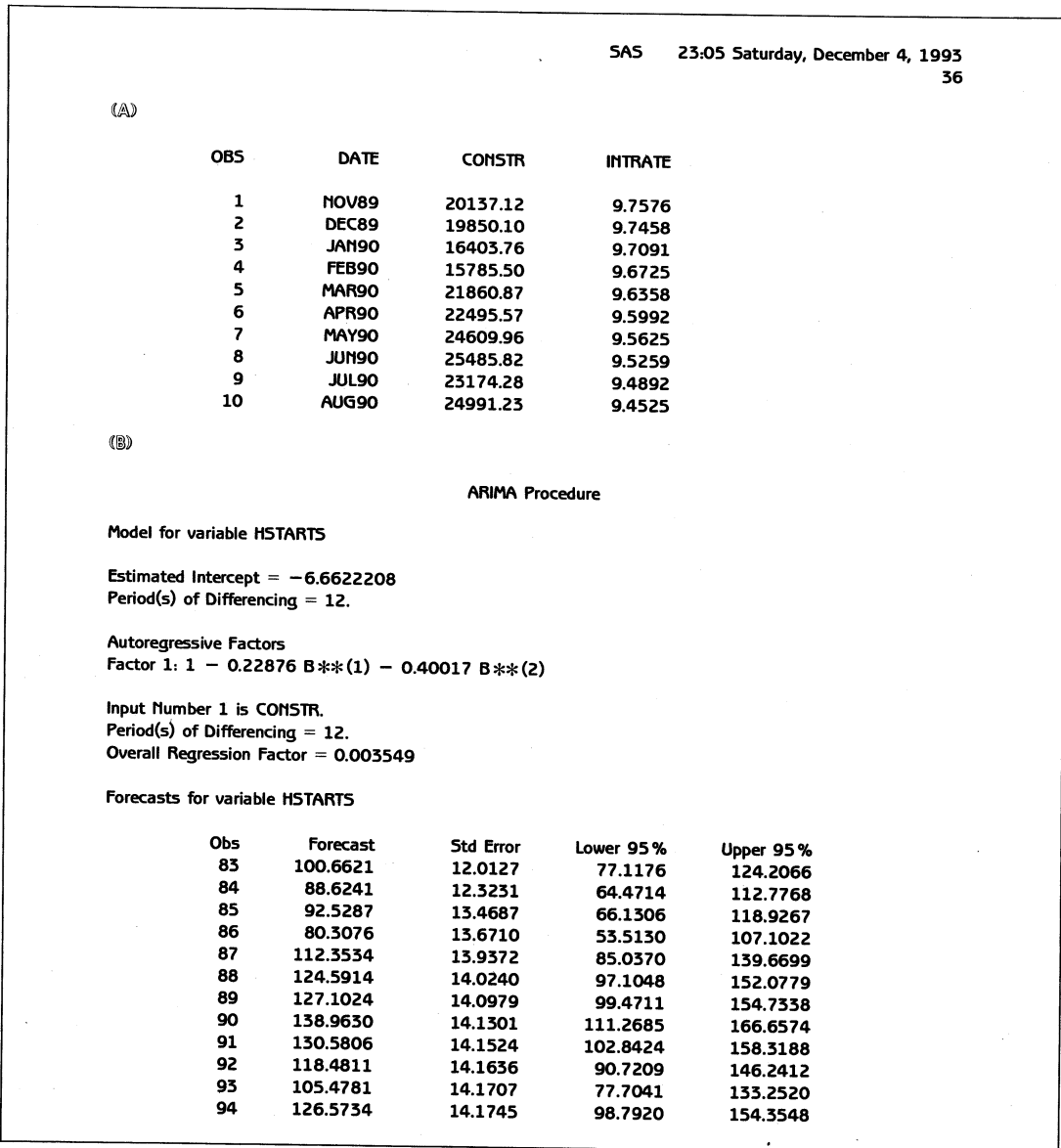


그림 5.2.7 PROC ARIMA를 사용한 예측

잡음항이 AR모형을 갖는다는 가정하에서 시계열데이터 HOUSECON을 회귀분석하기 위해서 PROC AUTOREG를 수행한다. PROC AUTOREG에서는 차분이나 계절차분을 할 수가 없다. 이후 PROC AUTOREG의 결과를 설명하는 동안에는 다음과 같은 표기를 사용하자.

$$\begin{aligned}
 y_t &= \text{HSTARTS}_t \\
 x_{t,1} &= \text{CONSTR}_t \\
 x_{t,2} &= \text{INTRATE}_t \\
 t &= \text{TIME}_t
 \end{aligned}$$

첫 번째 PROC AUTOREG는 다음과 같다.

```
PROC AUTOREG DATA=EXAM52;
  MODEL HSTARTS=CONSTR INTRATE
    / NLAG=4 BACKSTEP SLSTAY=0.06 METHOD=ML;
  RUN;
```

이 PROC AUTOREG의 MODEL스태이트먼트는 다음 모형을 추정하기 위한 것이다.

$$y_t = \beta_0 + \beta_1 x_{t,1} + \beta_2 x_{t,2} + n_t$$

MODEL스태이트먼트의 옵션 NLAG=에는 잡음항의 AR특성식의 형태를 지정한다. 이 PROC AUTOREG에는 옵션 NLAG=에 4가 지정되었으며 또한 옵션 BACKSTEP이 사용되었으므로 잡음항 $\{n_t\}$ 는 AR모형을 가지며 이 AR모형의 최대차수는 4이다. 옵션 BACKSTEP을 사용하면 잡음항의 AR모형이 후향제거방법(backward elimination method)에 의해서 자동적으로 정해진다. 좀더 자세히 설명하면 잡음항의 AR모형에서 유의적이지 않은(nonsignificant), 즉 유의적으로 0과 같은 AR계수들에 해당하는 항들이 제거된다. 이 항들이 제거되는 순서는 0에서 가까운(in order of least significance) AR계수의 순서이다. 이 후향제거방법은 초기의 보통최소제곱추정단계가 수행된 직후 Yule-Walker 추정단계에서 단 한 번 실행된다. 옵션 SLSTAY=에는 옵션 BACKSTEP이 수행될 때 추정된 AR계수들을 제거하는 기준이 되는 값을 지정한다. 이 옵션에는 보통 0.001과 0.2 사이의 값을 지정하며 디폴트는 0.05이다.

이 PROC AUTOREG를 수행한 결과가 그림 5.2.8이다. 그림 5.2.8의 부분 A는 보통최소제곱추정법을 적용해서 회귀모형을 추정한 결과이다. 그림 5.2.8의 부분 A에서 알 수 있듯이 추정된 회귀모형의 결정계수가 0.652이고 Durbin-Watson통계량값이 0.8420이다. 따라서 주어진 시계열데이터를 잡음항이 AR과정인 회귀모형에 적합시키면 더 좋은 결과를 얻을 것이다. 부분 A2의 'Estimates of Autocorrelations'에는 보통최소제곱추정법으로 추정된 회귀식에 의한 잔차의 표본자기공분산함수와 표본자기상관함수가 출력되었다. 부분 A3의 'Backward Elimination of Autoregressive Terms'는 옵션 BACKSTEP에 의해 출력된 것으로서 옵션 SLSTAY=에 지정된 유의수준 0.06보다 큰 p 값을 갖는 AR계수에 해당하는 항을 잡음항의 AR모형에서 제거한다. 이 예제에서는 시차들이 3과 4인 항들이 제거되고 시차들이 1과 2인 항들이 AR모형에 남는다. 부분 A4의 'Estimate of Autoregressive Parameters'는 보통최소제곱추정법에 의한 잔차 $\{\hat{n}_t\}$ 의 표본자기상관함수를 다음과 같은 AR(2)모형의 Yule-Walker방정식들에 대입해서 구한 α_1, α_2 의 Yule-Walker 추정값들이다.

$$\hat{n}_t = v_t - \alpha_1 \hat{n}_{t-1} - \alpha_2 \hat{n}_{t-2}$$

이 AR모형의 오차항 $\{v_t\}$ 의 분산 σ_v^2 의 추정값이 부분 A4의 'Preliminary MSE'에 실려 있다. 이 예제에서 추정된 잡음항의 모형은 다음과 같다.

SAS

37

9:09 Wednesday, December 8, 1993

Autoreg Procedure

(A)

Dependent Variable = HSTARTS

(A1)

Ordinary Least Squares Estimates

SSE	21621.29	DFE	79
MSE	273.6873	Root MSE	16.5435
SBC	703.0527	AIC	695.8325
Reg Rsq	0.6515	Total Rsq	0.6515
Durbin-Watson	0.8420		

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	-161.707523	26.403	-6.125	0.0001
CONSTR	1	0.0074270744	0.000626	11.859	0.0001
INTRATE	1	14.794782	1.675	8.832	0.0001

(A2)

Estimates of Autocorrelations

Lag	Covariance	Correlation
0	263.6743	1.000000
1	146.5879	0.555943
2	130.0086	0.493065
3	96.36173	0.365457
4	62.07979	0.235441

(A3)

Backward Elimination of Autoregressive Terms

Lag	Estimate	t-Ratio	Prob
3	-0.057559	-0.4643	0.6438
4	0.060324	0.5677	0.5719

(A4)

Preliminary MSE = 169.2606

Estimates of the Autoregressive Parameters

Lag	Coefficient	Std Error	t Ratio
1	-0.40789663	0.10984555	-3.713365
2	-0.26629788	0.10984555	-2.424294

(B)

(B1)

Maximum Likelihood Estimates

SSE	12813.32	DFE	77
MSE	166.4067	Root MSE	12.89987
SBC	669.7847	AIC	657.7511
Reg Rsq	0.6607	Total Rsq	0.7934
Durbin-Watson	2.0451		

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	-114.225728	46.033	-2.481	0.0153
CONSTR	1	0.007203297	0.000607	11.873	0.0001
INTRATE	1	10.418202	3.943	2.642	0.0100

A(1)	1	-0.414704	0.110	-3.768	0.0003
A(2)	1	-0.359573	0.116	-3.103	0.0027
(B2)					
Autoregressive parameters assumed given.					
Maximum Likelihood Estimates					
SSE	12813.32	DFE	77		
MSE	166.4067	Root MSE	12.89987		
SBC	669.7847	AIC	657.7511		
Reg Rsq	0.6607	Total Rsq	0.7934		
Durbin-Watson	2.0451				
Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	-114.225728	46.033	-2.481	0.0153
CONSTR	1	0.007203297	0.000607	11.873	0.0001
INTRATE	1	10.418202	3.943	2.642	0.0100

그림 5.2.8 입력변수들이 CONSTR와 INTRATE인 PROC AUTOREG를 수행한 결과

$$\hat{n}_t = 0.408\hat{n}_{t-1} + 0.266\hat{n}_{t-2} + v_t$$

여기서 오차항의 분산 σ_v^2 의 추정값은 169.261이다. 이 잡음항의 AR(2)모형의 AR계수들이 모두 유의적으로 0과 다름을 알 수 있다.

그림 5.2.8의 부분 B는 최우추정법을 사용해서 다음 모형을 추정한 결과이다.

$$y_t = \beta_0 + \beta_1 x_{t,1} + \beta_2 x_{t,2} + \frac{1}{1 + \alpha_1 B + \alpha_2 B^2} v_t$$

부분 B에서 추정된 회귀모형이 다음과 같음을 알 수 있다.

$$y_t = -114.226 + 0.00720 x_{t,1} + 10.418 x_{t,2} + \frac{1}{1 - 0.415 B - 0.360 B^2} v_t$$

여기서 강조할 점은 이 추정모형의 Durbin-Watson통계량값이 2.045로 2에 가깝다는 것이다. 따라서 이 추정된 회귀모형의 잔차는 계열상관을 갖지 않는다는 것을 알 수 있다. 부분 B는 잡음항의 AR계수들이 주어졌다고 가정하고 최우추정법을 사용해서 회귀모형을 추정한 결과이다. 즉, 다음 모형을 추정한 결과이다.

$$y_t = \beta_0 + \beta_1 x_{t,1} + \beta_2 x_{t,2} + \frac{1}{1 - 0.415 B - 0.360 B^2} v_t$$

이 회귀모형을 추정한 결과는 AR계수들이 주어졌다고 가정하지 않고 최우추정법을 사용해서 회귀모형을 추정한 결과, 즉 부분 B와 동일하다.

앞에서 PROC ARIMA를 사용해서 시계열데이터 HOUSECON을 잡음항이 ARMA모형을 갖는 회귀모형에 적합시킬 때는 종속변수 HSTARTS와 입력변수 CONSTR는 계절 주기 12인 계절차분을, 입력변수 INTRATE는 1차차분을 취했었다. 앞에서도 언급했듯이

PROC AUTOREG 안에서는 이러한 차분이나 계절차분을 취할 수가 없다. 따라서 이러한 차분과 계절차분의 효과를 얻기 위해서 PROC AUTOREG의 입력변수의 하나로 추세를 나타내는 시간의 변수 TIME을 추가하기로 하고 다음과 같은 두 번째 PROC AUTOREG를 수행해 보자.

```
PROC AUTOREG DATA=EXAM52;
    MODEL HSTARTS=CONSTR INTRATE TIME
        /NLAG=4 BACKSTEP SLSTAY=0.06 METHOD=ML;
RUN;
```

이 PROC AUTOREG의 MODEL스태이트먼트는 다음 모형을 추정하기 위한 것이다.

$$y_t = \beta_0 + \beta_1 x_{t,1} + \beta_2 x_{t,2} + \beta_3 t + n_t$$

SAS		19			
		11:19 Wednesday, December 8, 1993			
Autoreg Procedure					
Dependent Variable = HSTARTS					
(A)					
Ordinary Least Squares Estimates					
SSE	10893.1	DFE	78		
MSE	139.6551	Root MSE	11.81757		
SBC	651.2443	AIC	641.6175		
Reg Rsq	0.8244	Total Rsq	0.8244		
Durbin-Watson	1.8298				
Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	15.5407563	27.653	0.562	0.5757
CONSTR	1	0.0082019134	0.000456	17.986	0.0001
INTRATE	1	0.2500196	2.046	0.122	0.9031
TIME	1	-1.0024952	0.114	-8.765	0.0001
(B)					
Estimates of Autocorrelations					
Lag	Covariance	Correlation			
0	132.8426	1.000000			
1	8.503733	0.064014			
2	26.24816	0.197588			
3	15.34353	0.115502			
4	-4.7152	-0.035495			
(C)					
Backward Elimination of Autoregressive Terms					
Lag	Estimate	t-Ratio	Prob		
1	-0.041209	-0.3559	0.7230		
4	0.083514	0.7271	0.4694		
3	-0.103276	-0.9217	0.3596		
2	-0.197588	-1.7687	0.0809		

그림 5.2.9 입력변수들을 CONSTR, INTRATE와 TIME으로 하는 PROC AUTOREG

MODEL스태이트먼트의 옵션 NLAG=에는 4가 지정되었고 또한 옵션 BACKSTEP이 사용되었으므로 잡음항 $\{n_t\}$ 는 AR모형을 가지며 이 AR모형의 최대차수는 4이다. 이 PROC AUTOREG를 수행한 결과가 그림 5.2.9이다. 그림 5.2.9의 부분 A는 보통최소제곱 추정법을 적용해서 회귀모형을 추정한 것이다. 그림 5.2.9의 부분 A에서 알 수 있듯이 추정된 회귀모형의 결정계수가 0.8244이고 Durbin-Watson통계량값이 1.8298이다. 따라서 주어진 시계열데이터를 잡음항이 AR과정인 회귀모형에 적합시킨다고 해서 더 좋은 결과를 얻을 수 있을지는 의문이다. 부분 B의 'Estimates of Autocorrelations'에는 보통최소제곱추정법으로 추정된 회귀식에 의한 잔차의 표본자기공분산함수와 표본자기상관함수가 출력되어 있다. 시차 1 이후의 표본자기상관계수들은 유의적으로 0과 다르지 않음을 알 수 있다. 즉, 잔차들은 백색잡음과정에서 나왔다고 할 수 있다. 부분 C의 'Backward Elimination of Autoregressive Terms'는 옵션 BACKSTEP에 의해 출력된 것으로서 옵션 SLSTAY=에 지정된 유의수준 0.06보다 큰 p 값을 갖는 AR계수에 해당하는 항을 잡음항의 AR모형에서 제거한다. 이 예제에서는 시차들이 1, 2, 3과 4인 항들이 제거된다. 이는 잡음항이 AR(0)모형이라는 결론을 확인해 준다. 따라서 이 PROC AUTOREG에서는 잡음항이 AR모형을 갖는 회귀모형에 최우추정법을 적용할 기회를 갖지 못한다.

그림 5.2.9에서 변수 INTRATE의 회귀계수의 t 통계량값이 0.122이므로 이 회귀계수가 유의적으로 0과 다르지 않다. 따라서 입력변수 INTRATE를 제거한 회귀모형을 분석하기 위해서 같은 세 번째 PROC AUTOREG를 수행해 보자.

```
PROC AUTOREG DATA=EXAM52;
  MODEL HSTARTS=CONSTR TIME
    /NLAG=4 BACKSTEP SLSTAY=0.06 METHOD=ML;
  RUN;
```

이 PROC AUTOREG의 MODEL스태이트먼트는 다음 모형을 추정하기 위한 것이다.

$$y_t = \beta_0 + \beta_1 x_{t,1} + \beta_3 t + n_t$$

MODEL스태이트먼트의 옵션 NLAG=에는 4가 지정되었으며 또한 옵션 BACKSTEP이 사용되었으므로 잡음항 $\{n_t\}$ 는 AR모형을 가지며 이 AR모형의 최대차수는 4이다. 이 PROC AUTOREG를 수행한 결과가 그림 5.2.10이다. 그림 5.2.10의 부분 A는 보통최소제곱추정법을 적용해서 회귀모형을 추정한 것이다. 그림 5.2.10의 부분 A에서 알 수 있듯이 추정된 회귀모형의 결정계수가 0.824이고 Durbin-Watson통계량값이 1.830이다. 따라서 주어진 시계열데이터를 잡음항이 AR과정인 회귀모형에 적합시킨다고 해서 더 좋은 결과를 얻을 수 있을지는 의문이다. 부분 B의 'Estimates of Autocorrelations'에는 보통최소제곱추정법으로 추정된 회귀식에 의한 잔차의 표본자기공분산함수와 표본자기상관함수가 출력되어 있다. 시차 1 이후의 표본자기상관계수들은 유의적으로 0과 크게 다르지 않음을 알 수 있다. 즉, 잔차들은 백색잡음과정에서 나왔다고 할 수 있다. 부분 C의 'Backward Elimination of Autoregressive Terms'는 옵션 BACKSTEP에 의해 출력된 것으로서 옵션

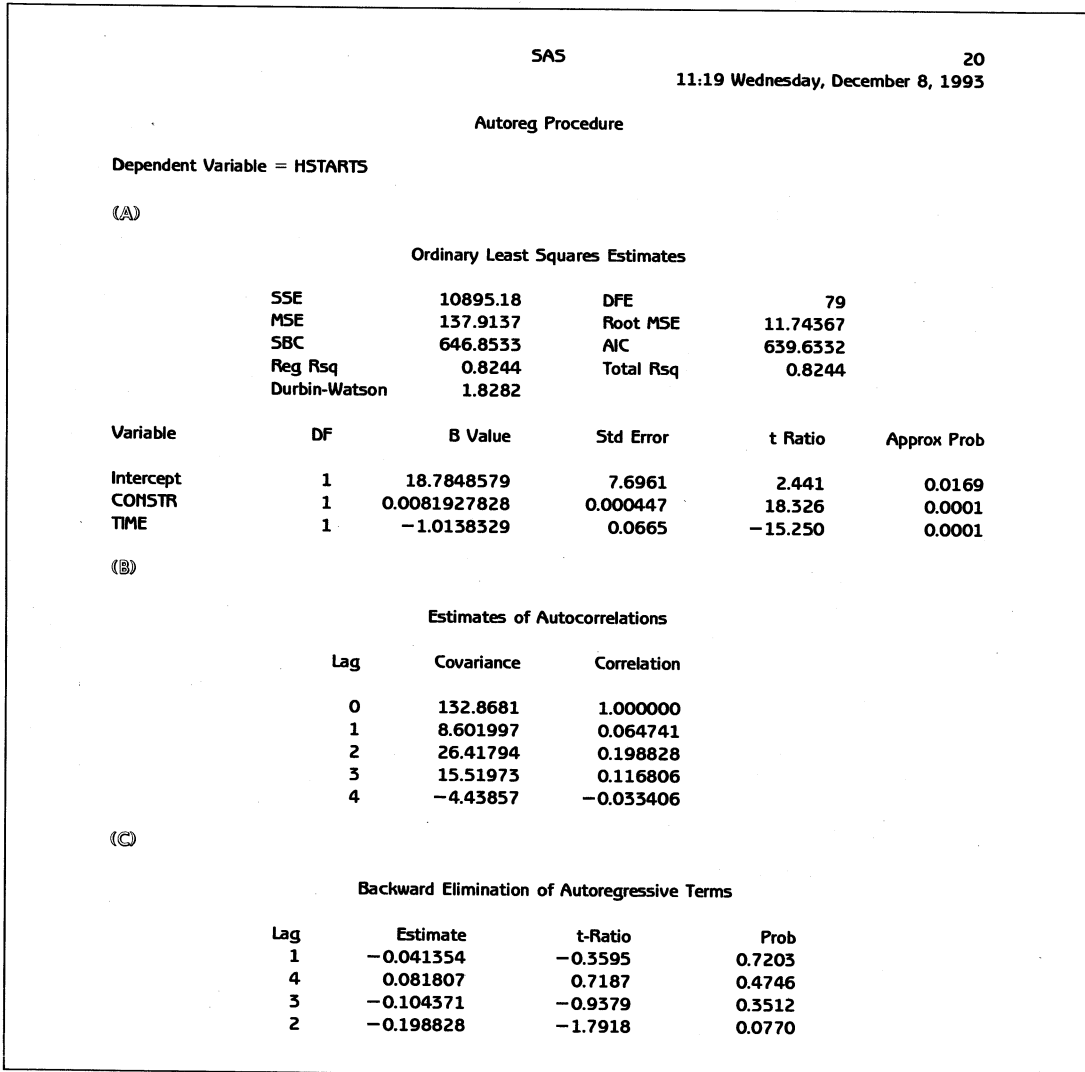


그림 5.2.10 입력변수들을 CONSTR와 TIME으로 한 PROC AUTOREG

SLSTAY=에 지정된 유의수준 0.06보다 큰 p 값을 갖는 AR계수에 해당하는 항을 잡음항의 AR모형에서 제거한다. 이 예제에서는 시차들이 1, 2, 3과 4인 항들이 제거된다. 이는 잡음항은 AR(0)모형, 즉 백색잡음모형을 갖는다는 결론을 확인해 준다. 따라서 이 PROC AUTOREG에서도 잡음항이 AR모형을 갖는 회귀모형에 최우추정법을 적용할 기회를 갖지 못한다. 그러나 이 입력변수들로 CONSTR와 TIME을 사용한 회귀모형은 입력변수들로 CONSTR와 INTRATE를 사용한 회귀모형이나 입력변수들로 CONSTR, INTRATE와 TIME을 사용한 회귀모형보다 AIC값이나 BIC값이 작다. 따라서 입력변수들로 CONSTR와 TIME을 사용한 회귀모형이 다른 2개의 회귀모형들보다 좋다고 할 수 있다.

주어진 시계열데이터의 미래값들을 예측하기 위해서 다음 PROC AUTOREG와 PROC PRINT를 수행해 보자.

SAS								21
								11:19 Wednesday, December 8, 1993
OBS	PRED	L95	U95	DATE	CONSTR	INTRATE	HSTARTS	TIME
83	99.616	75.514	123.717	NOV89	20137.12	9.7576	.	83
84	96.250	72.089	120.412	DEC89	19850.10	9.7458	.	84
85	67.002	42.228	91.775	JAN90	16403.76	9.7091	.	85
86	60.922	35.964	85.880	FEB90	15785.50	9.6725	.	86
87	109.683	85.603	133.763	MAR90	21860.87	9.6358	.	87
88	113.869	89.791	137.947	APR90	22495.57	9.5992	.	88
89	130.178	106.083	154.273	MAY90	24609.96	9.5625	.	89
90	136.340	112.187	160.493	JUN90	25485.82	9.5259	.	90
91	116.388	92.251	140.524	JUL90	23174.28	9.4892	.	91
92	130.260	106.088	154.432	AUG90	24991.23	9.4525	.	92
93	123.117	98.937	147.296	SEP90	24243.09	9.4159	.	93
94	116.414	92.205	140.623	OCT90	23548.71	9.3792	.	94

그림 5.2.11 PROC AUTOREG를 사용한 예측

```

PROC AUTOREG DATA=EXAM52C;
  MODEL HSTARTS=CONSTR TIME
    /NLAG=2 METHOD=ML NOPRINT;
  OUTPUT OUT=FORE52A P=PRED LCL=L95 UCL=U95;
RUN;
PROC PRINT DATA=FORE52A(FIRSTOBS=83);
RUN;

```

이 PROC AUTOREG의 PROC AUTOREG스태이트먼트와 MODEL스태이트먼트는 각각 바로 앞의 PROC AUTOREG의 PROC AUTOREG스태이트먼트와 MODEL스태이트먼트와 거의 같다. 다만 MODEL스태이트먼트에 옵션 NOPRINT를 사용한 것이 다르다. 이 옵션 NOPRINT는 MODEL스태이트먼트를 수행한 결과를 OUTPUT윈도우에 출력시키지 않도록 한다. OUTPUT스태이트먼트는 추정모형을 사용해서 미래값을 예측한 결과를 SAS데이터세트 FORE52A에 저장하기 위한 것이다. 또한 옵션 P=, 옵션 LCL=, 그리고 옵션 UCL=은 각각 예측(또는 적합), 신뢰수준 95%의 예측구간의 하한과 상한의 이름들을 지정하는 데 사용한다. 여기서 다시 한 번 강조하는 것은 PROC AUTOREG를 사용해서 예측을 할 때에는 예측할 미래시점들에서 입력변수들의 값들이 PROC AUTOREG스태이트먼트의 옵션 DATA=에 지정된 SAS데이터세트에 수록되어 있어야 한다는 것이다. 이 PROC AUTOREG를 수행하면 OUTPUT윈도우에 아무런 결과도 나타나지 않는다. 이 예측의 결과를 OUTPUT윈도우에 나타내기 위해서 PROC PRINT를 수행한다. 이 프로시저들을 수행한 결과가 그림 5.2.11이다.

세 번째 예제는 Grunfeld(1958)가 박사논문에서 제안한 투자함수의 추정모형에서 사용한 예제이다. Grunfeld는 두 개의 독립된 투자함수모형을 추정할 때 각각의 투자모형을 개별적으로 추정하는 것보다는 두 개의 모형이 공동으로 보유하고 있는 정보를 이용해서 두 개의 모형을 동시에 추정하는 것이 효율적임을 보였다. 이 이론을 예증하기 위해서 사용한 시계열데이터들을 Grunfeld의 시계열데이터들이라고 부른다. Grunfeld의 시계열데이터들

은 General Motors, U. S. Steel, General Electrics, Chrysler, Atlantic Richfield, IBM, Union Oil, Westinghouse, Goodyear, Diamond Match 등의 미국의 대기업들의 투자비용에 관한 것이다. 이 Grunfeld의 시계열데이터들은 Grunfeld & Griliches(1960), Maddala(1977, pp. 214~215, [ETS6.1]에서는 pp. 280~281이라고 했으나 이는 잘못된 것임), Greene(1991, pp. 462~463)에 수록되어 있다. 또한 본서의 제 2.3 절에도 수록되어 있다. [ETS6.1](pp. 148~152, pp. 185~190)에서는 이 중에서 General Electric(GE)과 Westinghouse(W)에 관한 변수들을 포함한 시계열데이터를 다루었다. 이 시계열데이터는 다음과 같은 변수들을 포함하고 있다.

- GEI = GE의 총투자액(gross investment)
- GEC = GE의 자본금(stock of plant and equipment)
- GEF = GE의 자산가치(value of the firm)
- WI = W의 총투자액
- WC = W의 자본금
- WF = W의 자산가치

[ETS6.1]은 이 시계열데이터를 PROC AUTOREG(pp. 185~190)를 사용해서 잡음향이 AR모형들을 갖는 회귀모형들에 적합시키고 PROC ARIMA(pp. 148~152)를 사용해서 잡음향이 ARMA모형들을 갖는 회귀모형들에 적합시켰다. 이 예제에서는 [ETS6.1]의 분석을 그대로 따르고자 한다. SAS사에서 공급하는 ETS의 디스켓 중에는 'Sample Library'라는 이름이 붙은 디스켓이 있다. 이 디스켓에는 [ETS6.1]에 수록되어 있는 예제들의 SAS 프로그램들의 코드(code)가 수록되어 있다. 이 예제에서는 이 디스켓의 코드를 읽어서 [ETS6.1]의 분석을 그대로 다시 수행하기로 한다. 단, DATA스텝은 제 2.3 절에서 만든 영구SAS데이터세트 TSA4DATA.GRUNFELD를 사용하기로 한다. 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM5-3.PRG에 수록되어 있다).

```

/* */
/*          EXAM5-3.PRG          */
/* */
/*          REGRESSION WITH TIME SERIES ERROR          */
/*          GRUNFELD'S INVESTMENT MODELS          */
/*          PROC ARIMA & PROC AUTOREG          */
/* */
OPTIONS LINESIZE=72 PAGESIZE=60;
/* */
/*          SAS  SAMPLE  LIBRARY          */
/* */
/*          NAME  : ARIMEX3          */

```

```

/ *      TITLE   : ARIMA User's Guide Example 3, Version 6      */
/ * PRODUCT   : ETS                                           */
/ *      KEYS    : TIME                                         */
/ *      PROCS   : ARIMA                                        */
/ *      DATA   : Taken from SAS/ ETS User's Guide, V6 Edition */
/ *                                                    */
/ *      Grunfeld: Maddala, pp. 280~281.                       */
/ *                                                    */
/ *      REF    :                                              */
/ *                                                    */
/ * Maddala, G. S. (1977). Econometrics. New York: McGraw-Hill. */
/ *      Chapter 12, pp. 280~281.                               */
/ *                                                    */
/ * SAS Institute Inc. (1988). SAS/ ETS User's Guide, Version 6 */
/ *      Edition. Cary, North Carolina: SAS Institute Inc.     */
/ *                                                    */
/ *      MISC: See also example in PROC AUTOREG documentation. */
/ *                                                    */
/ *                                                    */
title 'Grunfeld's Investment Models Fit With Autoregressive Errors';
/ *      NEW DATA STEP      */
LIBNAME TSA4DAT 'A:\TSA4DATA';
DATA GRUNFELD;
    SET TSA4DATA.GRUNFELD;
RUN;
/ *---- From ARIMA Documentation ----*/
proc arima data=grunfeld;
    identify var=gei crosscor=(gef gec) noprint;
    estimate p=1 method=ml input=(gef gec) plot;
    estimate p=2 method=ml input=(gef gec);
    estimate q=1 method=ml input=(gef gec);
    estimate q=1 p=1 method=ml input=(gef gec);
run;
/ *                                                    */
/ *      S A S      S A M P L E      L I B R A R Y      */
/ *                                                    */
/ *      NAME    : autoreg2                                     */
/ *      TITLE   : Example 2 from chapter 8, SAS/ ETS User's Guide */
/ *      PRODUCT : SAS/ ETS                                     */

```

```

/* SYSTEM : ALL */
/* KEYS : */
/* PROCS : AUTOREG */
/* DATA : */
/*
/* REF : SAS/ ETS User's Guide, Version 6, First Edition */
/* MISC : */
/*
proc autoreg data=grunfeld;
  model gei=gef gec / nlag=1 method=uls;
  model gei=gef gec / nlag=1;
  model gei=gef gec / nlag=1 method=ml;
  model gei=gef gec / nlag=2 method=ml;
run;

```

주어진 시계열데이터를 잡음항이 ARMA과정인 회귀모형에 적합시키기 위해서 다음 PROC ARIMA를 수행한다.

```

proc arima data=grunfeld;
  identify var=gei crosscor=(gef gec) noprint;
  estimate p=1 method=ml input=(gef gec) plot;
  estimate p=2 method=ml input=(gef gec);
  estimate q=1 method=ml input=(gef gec);
  estimate q=1 p=1 method=ml input=(gef gec);
run;

```

이 PROC ARIMA의 IDENTIFY스테이트먼트의 옵션 VAR=에는 종속변수 GEI를 지정하고 옵션 CROSSCOR=에는 입력변수들 GEF와 GEC를 지정한다. IDENTIFY스테이트먼트의 옵션 NOPRINT를 사용했기 때문에 이 IDENTIFY스테이트먼트를 수행한 결과가 OUTPUT윈도우에 나타나지 않는다. 이후 이 PROC ARIMA의 결과를 설명하는 동안에는 편의상 다음과 같은 기호를 사용하자.

$$\begin{aligned}
 y_t &= \text{GEI}_t \\
 x_{t,1} &= \text{GEF}_t \\
 x_{t,2} &= \text{GEC}_t
 \end{aligned}$$

첫 번째 ESTIMATE스테이트먼트는 다음과 같은 잡음항이 AR(1)모형을 갖는 회귀모형을 추정하기 위한 것이다.

$$y_t = \beta_0 + \beta_1 x_{t,1} + \beta_2 x_{t,2} + \frac{1}{1 - \phi_1 B} v_t$$

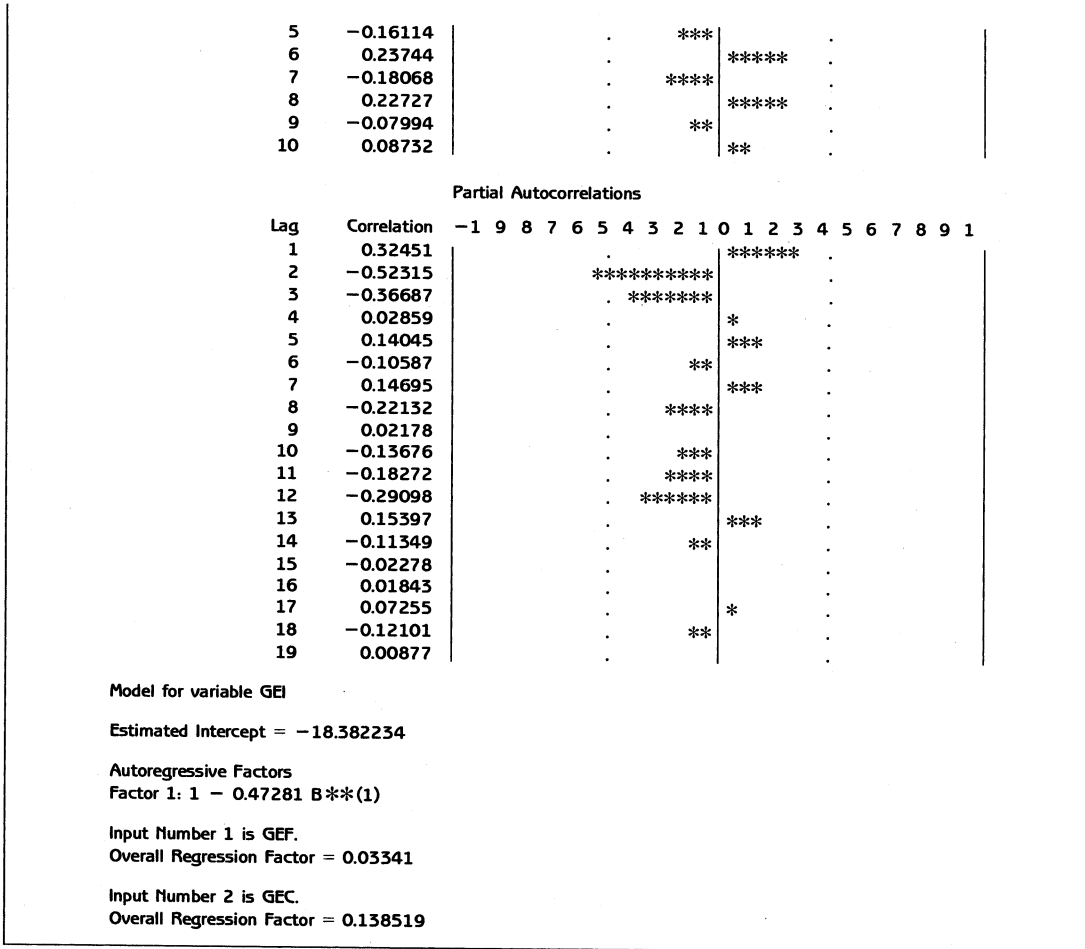


그림 5.3.1 잡음항이 AR(1)모형을 갖는 회귀모형

이 ESTIMATE스테이트먼트를 수행한 결과가 그림 5.3.1이다. 그림 5.3.1에서 최우추정된 회귀모형이 다음과 같음을 알 수 있다.

$$y_t = -18.382 + 0.0334 x_{t,1} + 0.139 x_{t,2} + \frac{1}{1 - 0.473B} v_t$$

또한, 카이제곱통계량값들, 즉 퍼트맨투우통계량값들로부터 이 추정된 회귀모형의 잔차가 백색잡음과정에서 나왔다고 할 수가 없다. 또한 잔차의 표본자기상관함수는 시나브로 형태를 보이고, 표본역자기상관함수는 특정한 형태를 보이지 않으며, 시차 3에서의 표본부분자기상관계수들은 유의적으로 0과 다르다.

두 번째 ESTIMATE스테이트먼트는 다음과 같은 잡음항이 AR(2)모형을 갖는 회귀모형을 추정하기 위한 것이다.

$$y_t = \beta_0 + \beta_1 x_{t,1} + \beta_2 x_{t,2} + \frac{1}{1 - \phi_1 B - \phi_2 B^2} v_t$$

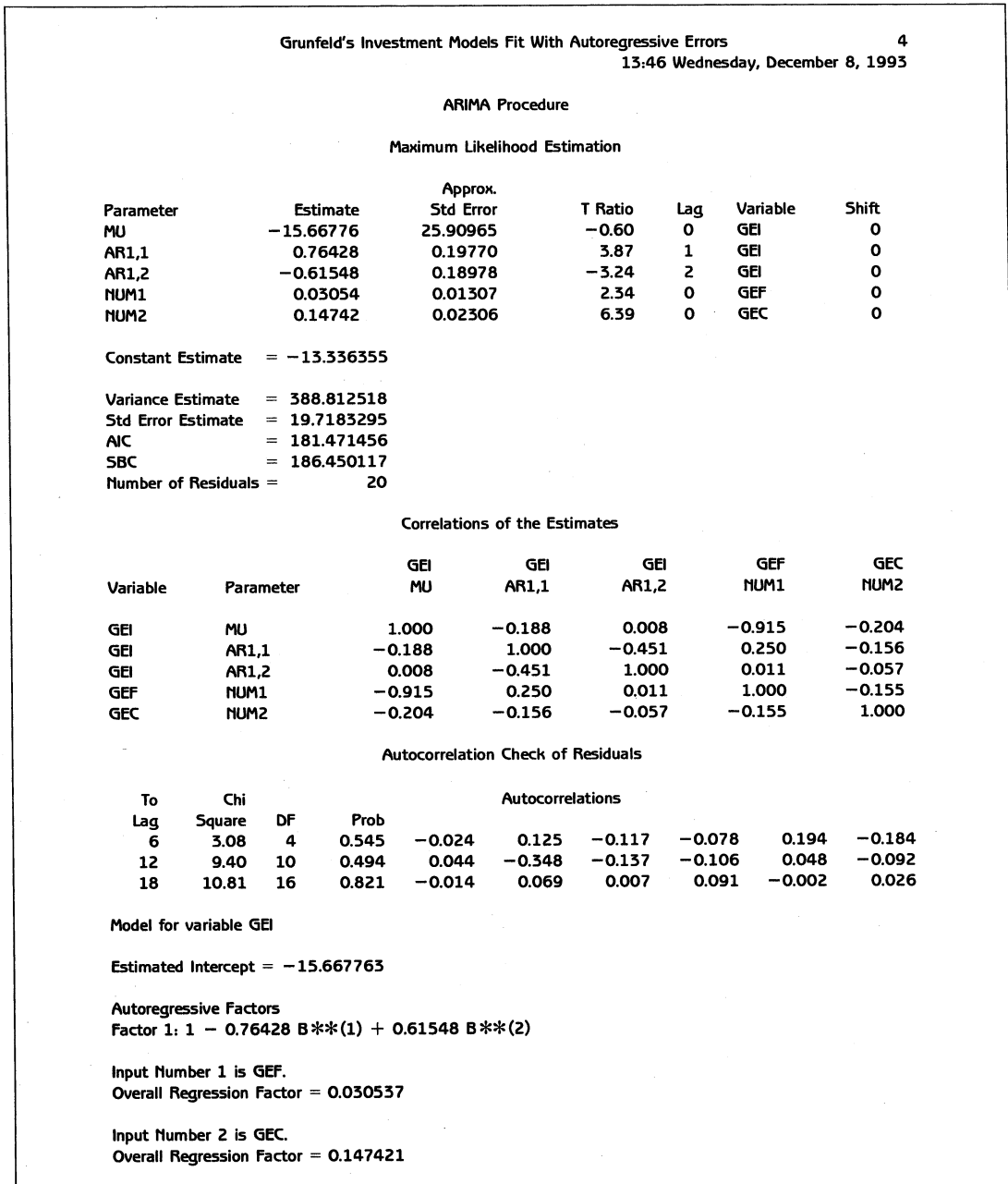


그림 5.3.2 잡음항이 AR(2)모형을 갖는 회귀모형

이 ESTIMATE스테이트먼트를 수행한 결과가 그림 5.3.2이다. 그림 5.3.2에서 최우추정된 회귀모형이 다음과 같음을 알 수 있다.

$$y_t = -15.668 + 0.0305x_{t,1} + 0.147x_{t,2} + \frac{1}{1 - 0.764B + 0.615B^2} v_t$$

또한, 카이제곱통계량값들, 즉 퍼트맨토우통계량값들로부터 이 추정된 회귀모형의 잔차가 백색잡음과정에서 나왔다고 할 수 있다.

세 번째 ESTIMATE스태이트먼트는 다음과 같은 잡음항이 MA(1)모형을 갖는 회귀모형을 추정하기 위한 것이다.

$$y_t = \beta_0 + \beta_1 x_{t,1} + \beta_2 x_{t,2} + v_t - \theta_1 v_{t-1}$$

이 ESTIMATE스태이트먼트를 수행한 결과가 그림 5.3.3이다. 그림 5.3.3에서 최우추정된 회귀모형이 다음과 같음을 알 수 있다.

$$y_t = -24.951 + 0.0371 x_{t,1} + 0.138 x_{t,2} + v_t + 0.785 v_{t-1}$$

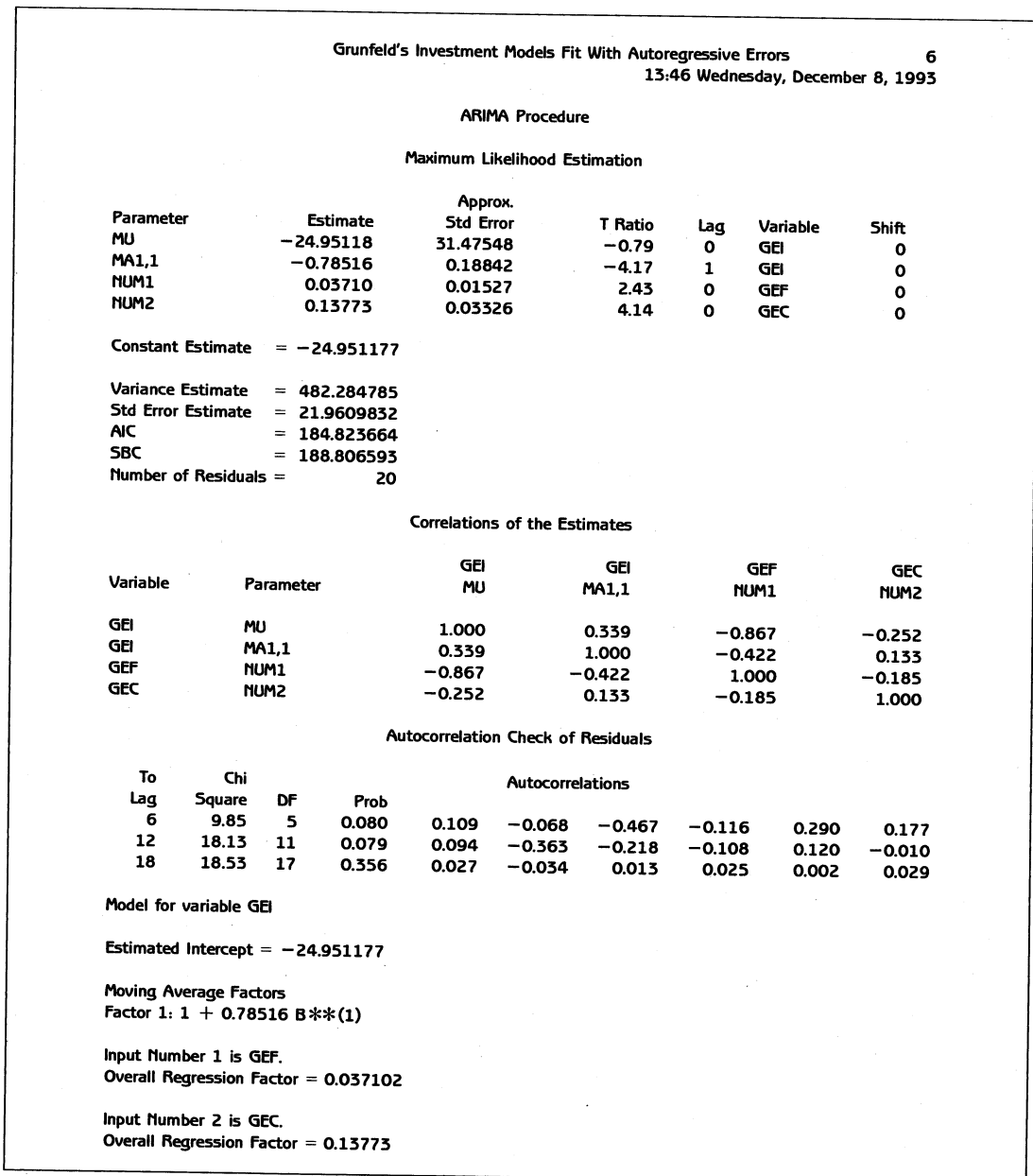


그림 5.3.3 잡음항이 MA(1)모형을 갖는 회귀모형

카이제곱통계량값들, 즉 퍼트맨토우통계량값들로부터 이 추정된 회귀모형의 잔차가 백색잡음과정에서 나왔다고 할 수 있다.

네 번째 ESTIMATE스태이트먼트는 다음과 같은 잡음항이 ARMA(1, 1)모형을 갖는 회귀모형을 추정하기 위한 것이다.

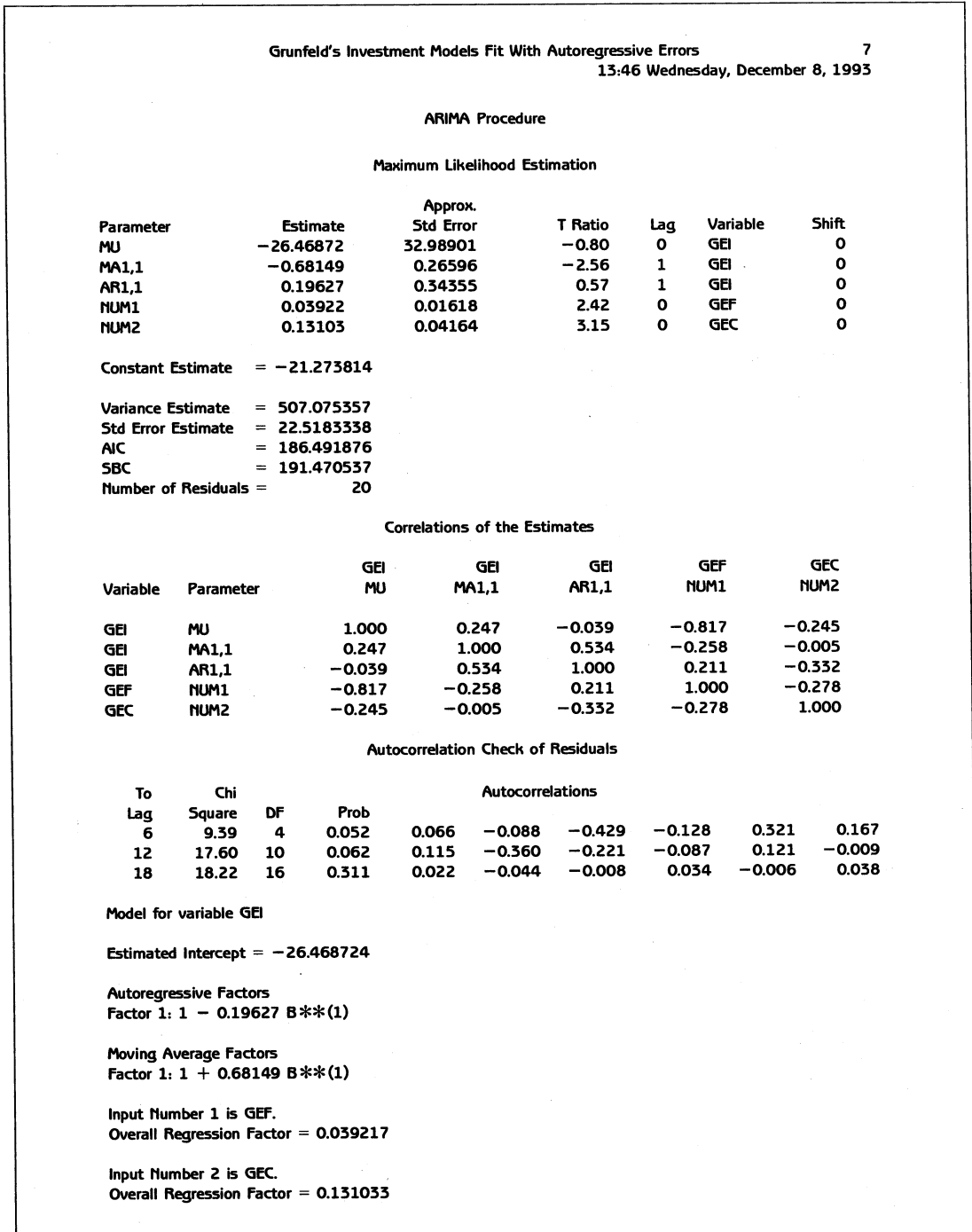


그림 5.34 잡음항이 MRMA(1, 1)모형을 갖는 회귀모형

$$y_t = \beta_0 + \beta_1 x_{t,1} + \beta_2 x_{t,2} + \frac{1 - \theta_1 B}{1 - \phi_1 B} v_t$$

이 ESTIMATE 스테이트먼트를 수행한 결과가 그림 5.3.4이다. 그림 5.3.4에서 최우추정된 회귀모형이 다음과 같음을 알 수 있다.

$$y_t = -26.496 + 0.0392 x_{t,1} + 0.131 x_{t,2} + \frac{1 + 0.681 B}{1 - 0.196 B} v_t$$

이 추정모형의 카이제곱통계량값들, 즉 퍼트맨토우통계량값들로부터 추정된 회귀모형의 잔차가 백색잡음과정에서 나왔다고 할 수 있다.

이 네 개의 회귀모형들 중에서 두 번째 모형, 즉 잡음항이 AR(2)모형을 갖는 회귀모형의 AIC값과 BIC값이 가장 작다. 또한, 이 회귀모형의 퍼트맨토우통계량값들도 다른 회귀모형들의 퍼트맨토우통계량값들보다 작다. 따라서 이 시계열데이터에는 잡음항이 AR(2)모형인 회귀모형이 가장 적합하다고 할 수 있다.

잡음항이 AR모형을 갖는다는 가정하에서 Grunfeld의 시계열데이터를 회귀분석하기 위해서 다음 PROC AUTOREG를 수행한다.

```
proc autoreg data=grunfeld;
  model gei=gef gec / nlag=1 method=uls;
  model gei=gef gec / nlag=1;
  model gei=gef gec / nlag=1 method=ml;
  model gei=gef gec / nlag=2 method=ml;
run;
```

이 PROC AUTOREG의 처음 세 개의 MODEL 스테이트먼트들은 다음 모형을 추정하기 위한 것이다.

$$y_t = \beta_0 + \beta_1 x_{t,1} + \beta_2 x_{t,2} + n_t$$

$$n_t = -\alpha_1 n_{t-1} + v_t$$

단, $\{v_t\}$ 는 백색잡음과정

첫 번째 MODEL 스테이트먼트를 수행한 결과가 그림 5.3.5이다. 그림 5.3.5의 부분 A는 보통최소제곱추정법을 적용해서 회귀모형을 추정한 것이다. 부분 A에서 알 수 있듯이 추정된 회귀모형의 결정계수가 0.705이고 Durbin-Watson통계량값이 1.072이다. 따라서 주어진 시계열데이터를 잡음항이 AR과정인 회귀모형에 적합시키면 더 좋은 결과를 얻을 수도 있을 것이다. 부분 A2의 'Estimates of Autocorrelations'에는 보통최소제곱추정법으로 추정된 회귀식에 의한 잔차의 표본자기공분산함수와 표본자기상관함수가 출력되었다. 부분 A3의 'Estimate of Autoregressive Parameters'는 보통최소제곱추정법에 의한 잔차 $\{\hat{n}_t\}$ 의 표본자기상관함수를 다음과 같은 AR(1)모형의 Yule-Walker방정식들에 대입해서 구한 α_1 의 Yule-Walker추정값이다.

Grunfeld's Investment Models Fit With Autoregressive Errors 9
 13:46 Wednesday, December 8, 1993

Autoreg Procedure

(A)

Dependent Variable = GEI Gross Investment GE

(A1)

Ordinary Least Squares Estimates

SSE	13216.59	DFE	17
MSE	777.4463	Root MSE	27.88272
SBC	195.6147	AIC	192.6275
Reg Rsq	0.7053	Total Rsq	0.7053
Durbin-Watson	1.0721		

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	-9.95630645	31.374	-0.317	0.7548
GEF	1	0.02655119	0.016	1.706	0.1063
GEC	1	0.15169387	0.026	5.902	0.0001

(A2)

Estimates of Autocorrelations

Lag	Covariance	Correlation
0	660.8294	1.000000
1	304.5546	0.460867

(A3)

Preliminary MSE = 520.4701

Estimates of the Autoregressive Parameters

Lag	Coefficient	Std Error	t Ratio
1	-0.46086726	0.22186727	-2.077221

(B)

Unconditional Least Squares Estimates

SSE	10220.85	DFE	16
MSE	638.8028	Root MSE	25.27455
SBC	193.7567	AIC	189.7738
Reg Rsq	0.5511	Total Rsq	0.7721
Durbin-Watson	1.3530		

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	-18.6581577	34.810	-0.536	0.5993
GEF	1	0.0338724	0.018	1.891	0.0769
GEC	1	0.1368992	0.045	3.052	0.0076
A(1)	1	-0.4996199	0.259	-1.928	0.0718

(C)

Unconditional Least Squares Estimates

SSE	10220.85	DFE	16
MSE	638.8028	Root MSE	25.27455
SBC	193.7567	AIC	189.7738
Reg Rsq	0.5511	Total Rsq	0.7721
Durbin-Watson	1.3530		

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	-18.6581577	34.810	-0.536	0.5993
GEF	1	0.0338724	0.018	1.891	0.0769
GEC	1	0.1368992	0.045	3.052	0.0076

그림 5.3.5 잡음항이 AR(1)모형인 회귀모형의 비조건부최소제곱추정

Grunfeld's Investment Models Fit With Autoregressive Errors					
					11
					13:46 Wednesday, December 8, 1993
Autoreg Procedure					
Dependent Variable = GEI			Gross Investment GE		
Yule-Walker Estimates					
	SSE	10238.3	DFE	16	
	MSE	639.8934	Root MSE	25.29612	
	SBC	193.7424	AIC	189.7595	
	Reg Rsq	0.5717	Total Rsq	0.7717	
Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	-18.2317775	33.251	-0.548	0.5911
GEF	1	0.0331996	0.016	2.096	0.0523
GEC	1	0.1391942	0.038	3.632	0.0022

그림 5.3.6 잡음항이 AR(1)모형인 회귀모형의 Yule-Walker추정

$$\hat{n}_i = v_i - \alpha_1 \hat{n}_{i-1}$$

이 AR모형의 오차항 $\{v_i\}$ 의 분산 σ_v^2 의 추정값이 부분 A의 'Preliminary MSE'에 실려 있다. 여기서 α_1 의 Yule-Walker추정값은 -0.461 이고 오차항의 분산 σ_v^2 의 추정값은 520.5 이다. 그림 5.3.5의 부분 B는 비조건부최소제곱추정법을 사용해서 다음 모형을 추정한 결과이다.

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \frac{1}{1 + \alpha_1 B} v_i$$

부분 B에서 추정된 회귀모형이 다음과 같음을 알 수 있다.

$$y_i = -18.658 + 0.0339 x_{i,1} + 0.137 x_{i,2} + \frac{1}{1 - 0.500 B} v_i$$

이 추정모형의 Durbin-Watson통계량값이 1.353 으로 보통최소제곱추정법의 Durbin-Watson통계량값보다는 2 에 가깝지만 유의적으로 2 에 가깝다고 쉽게 결론을 내릴 수는 없다. 부분 C는 잡음항의 AR계수가 주어졌다고 가정하고 비조건부최소제곱추정법을 사용해서 회귀모형을 추정한 결과이다. 즉, 다음 모형을 추정한 결과이다.

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \frac{1}{1 - 0.500 B} v_i$$

부분 C는 부분 B와 동일하다.

두 번째 MODEL스테이트먼트를 수행한 결과의 앞부분은 보통최소제곱추정법을 적용해서 회귀모형을 추정한 것으로 그림 5.3.5의 부분 A와 동일하다. 따라서 옵션 MET-HOD=의 디폴트인 Yule-Walker추정법을 사용해서 다음 모형을 추정한 결과만을 그림 5.3.6에 수록했다.

Grunfeld's Investment Models Fit With Autoregressive Errors					
					12
					13:46 Wednesday, December 8, 1993
Autoreg Procedure					
Dependent Variable = GEI			Gross Investment GE		
Maximum Likelihood Estimates					
SSE	10229.23	DFE	16		
MSE	639.3269	Root MSE	25.28491		
SBC	193.7389	AIC	189.7559		
Reg Rsq	0.5656	Total Rsq	0.7719		
Durbin-Watson	1.3389				
Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	-18.3750736	34.594	-0.531	0.6026
GEF	1	0.0334051	0.018	1.870	0.0799
GEC	1	0.1385257	0.043	3.233	0.0052
A(1)	1	-0.4727583	0.258	-1.831	0.0858

그림 5.3.7 잡음항이 AR(1)모형인 회귀모형의 최우추정

$$y_t = \beta_0 + \beta_1 x_{t,1} + \beta_2 x_{t,2} + \frac{1}{1 + \alpha_1 B} v_t$$

부분 B에서 추정된 회귀모형이 다음과 같음을 알 수 있다.

$$y_t = -18.232 + 0.0332 x_{t,1} + 0.139 x_{t,2} + \frac{1}{1 - 0.500 B} v_t$$

세 번째 MODEL스테이트먼트를 수행한 결과의 앞부분은 보통최소제곱추정법을 적용해서 회귀모형을 추정한 것으로 그림 5.3.5의 부분 A와 동일하다. 따라서 옵션 METHOD=에 지정한 최우추정법을 사용해서 다음 모형을 추정한 결과만을 그림 5.3.7에 수록했다.

$$y_t = \beta_0 + \beta_1 x_{t,1} + \beta_2 x_{t,2} + \frac{1}{1 + \alpha_1 B} v_t$$

그림 5.3.7에서 최우추정된 회귀모형이 다음과 같음을 알 수 있다.

$$y_t = -18.375 + 0.0334 x_{t,1} + 0.139 x_{t,2} + \frac{1}{1 - 0.473 B} v_t$$

이 PROC AUTOREG의 네 번째 MODEL스테이트먼트를 수행한 결과의 앞부분은 보통최소제곱추정법을 적용해서 회귀모형을 추정한 것으로 그림 5.3.5의 부분 A와 거의 동일하다. 이 MODEL스테이트먼트를 수행한 결과의 뒷부분은 최우추정법을 사용해서 다음 모형을 추정한 결과이다. 이 결과가 그림 5.3.8이다.

$$y_t = \beta_0 + \beta_1 x_{t,1} + \beta_2 x_{t,2} + \frac{1}{1 + \alpha_1 B + \alpha_2 B^2} v_t$$

Grunfeld's Investment Models Fit With Autoregressive Errors						14
						13:46 Wednesday, December 8, 1993
Autoreg Procedure						
Dependent Variable = GEI			Gross Investment GE			
(A)						
Maximum Likelihood Estimates						
SSE	5832.272	DFE	15			
MSE	388.8181	Root MSE	19.71847			
SBC	186.4501	AIC	181.4715			
Reg Rsq	0.7674	Total Rsq	0.8700			
Durbin-Watson	2.0227					
Variable	DF	B Value	Std Error	t Ratio	Approx Prob	
Intercept	1	-15.6668160	26.701	-0.587	0.5661	
GEF	1	0.0305363	0.013	2.267	0.0386	
GEC	1	0.1474205	0.024	6.204	0.0001	
A(1)	1	-0.7642446	0.204	-3.751	0.0019	
A(2)	1	0.6154026	0.196	3.147	0.0066	
(B)						
Autoregressive parameters assumed given.						
Maximum Likelihood Estimates						
SSE	5832.272	DFE	15			
MSE	388.8181	Root MSE	19.71847			
SBC	186.4501	AIC	181.4715			
Reg Rsq	0.7674	Total Rsq	0.8700			
Durbin-Watson	2.0227					
Variable	DF	B Value	Std Error	t Ratio	Approx Prob	
Intercept	1	-15.6668160	26.701	-0.587	0.5661	
GEF	1	0.0305363	0.013	2.267	0.0386	
GEC	1	0.1474205	0.024	6.204	0.0001	

그림 5.3.8 잡음항이 AR(2)모형인 회귀모형의 최우추정

그림 5.3.8의 부분 A에서 알 수 있듯이 최우추정된 회귀모형은 다음과 같다.

$$y_t = -15.667 + 0.0305x_{t,1} + 0.147x_{t,2} + \frac{1}{1 - 0.764B + 0.615B^2} v_t$$

이 추정모형의 Durbin-Watson통계량값이 2.023으로 이 추정된 회귀모형의 잔차가 백색 잡음과정에서 나왔다고 할 수 있다. 따라서 잡음항이 AR(2)모형을 갖는 회귀모형이 잡음항이 AR(1)모형을 갖는 회귀모형보다 시계열데이터 GRUNFELD를 더 잘 적합시킨다. 그림 5.3.8의 부분 B는 잡음항의 AR계수들이 주어졌다고 가정하고 최우추정법을 사용해서 회귀모형을 추정한 결과이다. 이 결과는 그림 5.3.2의 PROC ARIMA를 사용한 결과와 거의 동일함을 알 수 있다.

5.1.3 PROC ARIMA의 사용법

본 소절에서는 [SAS4TSA2]의 제 9장부터 제 13장까지와 지금까지 본서에서 다룬 PROC ARIMA의 사용양식을 요약하고자 한다. 본 소절의 내용은 [ETS6.1]의 제 7장 'The ARIMA Procedure'를 요약한 것이다.

SAS/ETS의 PROC ARIMA는 Box & Jenkins(1976, pp. 495~516)의 Programs 1-7과 비슷하게 프로그램되었다. PROC ARIMA는 적당한 크기(moderate size)의 시계열 데이터를 분석하는 데 적합하다. 즉, 최소 30개에서 최대 2,000개의 관찰점들을 가진 시계열 데이터를 분석할 수 있다. 만일, 관찰점들의 개수가 30 이하이면 모수들의 추정값들을 신뢰하기 힘들 것이고, 관찰점들의 개수가 몇천 개이면 긴 컴퓨터 사용시간과 큰 컴퓨터 기억용량을 필요로 할 것이다. 만일, 주어진 시계열데이터가 이러한 조건을 만족하지 않으면 PROC FORECAST나 PROC AUTOREG를 사용할 것을 고려해 보아야 한다. PROC ARIMA에서는 PROC REG나 PROC AUTOREG에서 다루는 모형들을 포함하는 좀더 일반적인 모형을 다룬다. 따라서 PROC ARIMA가 PROC REG나 PROC AUTOREG와 같은 모형을 다룰 때는 좀더 긴 컴퓨터 사용시간과 큰 컴퓨터 기억용량을 필요로 할 뿐 아니라 PROC REG나 PROC AUTOREG가 출력시키는 모든 통계량들을 제공하지 않을 수도 있다. SAS의 다른 프로시저들을 사용해서 같은 데이터셋을 같은 모형에 적합시켰을 때 다른 추정값들을 얻는다는 것은 프로시저들이 채택하고 있는 추정방법들의 차이와 장단점을 알려 준다. 따라서 주어진 데이터셋에 가장 적당한(most appropriate) 추정방법을 찾기 위해서는 SAS의 여러 프로시저들을 적용해 보는 것도 좋을 것이다.

PROC REG나 PROC GLM과 마찬가지로 PROC ARIMA는 대화형(interactive)으로 수행할 수 있다. 이를 좀더 자세히 설명하면 다음과 같다. 일반적으로 SAS프로그램은 DATA스텝, 또는 PROC스텝 단위로 수행된다. 즉, 각 스텝에서 RUN스태이트먼트, 또는 다음 스텝의 시작을 나타내는 SAS키워드 DATA 또는 PROC이 나오면 그 스텝이 수행된다. 그러나 PROC ARIMA 안에는 두 개 이상의 RUN스태이트먼트를 사용할 수 있다. SAS시스템은 이 RUN스태이트먼트를 만나면 바로 전 RUN스태이트먼트 이후의 스텤타이트먼트들을 수행한다. 이 PROC ARIMA는 새로운 DATA스태이트먼트, 또는 PROC스태이트먼트, ENDSAS스태이트먼트, 그리고 QUIT스태이트먼트가 나오면 끝이 난다. 이러한 스텤타이트먼트들을 종착스태이트먼트(terminal statement)라고 한다.

1) PROC ARIMA의 모형

PROC ARIMA에서 다루는 모형은 다음과 같다.

$$w_t = \mu + \sum_j \frac{\omega_j(B)}{\delta_j(B)} x_{t,j} + \frac{\theta(B)}{\phi(B)} \frac{\Theta(B^s)}{\Phi(B^s)} v_t$$

이 식의 각 성분은 다음과 같다.

- ① $\{w_t\}$ 는 반응(response)을 나타내는 변수, 즉 종속변수의 시점 t 에서의 관찰값, 또는 종속변수의 시점 t 에서의 차분(difference)이다.
- ② 잡음항의 $\phi(x)$ 는 AR특성식이다.
- ③ 잡음항의 $\theta(x)$ 는 MA특성식이다.
- ④ $\{v_t\}$ 는 백색잡음과정으로서 독립적 교란항(independent disturbance), 또는 오차항(random error)이라고 부른다.
- ⑤ $\{x_{t,j}\}$ 는 제 j 번째 입력변수 또는 제 j 번째 입력변수의 차분(difference)이다.
- ⑥ $\omega_j(x)/\delta_j(x)$ 는 제 j 번째 입력변수의 전이함수(transfer function)이다. 여기서 $\omega_j(x)$ 와 $\delta_j(x)$ 는 각각 멱함수(polynomial)이다.

PROC ARIMA의 사용양식은 다음과 같다.

```
PROC ARIMA 옵션들;
  BY 옵션들;
  IDENTIFY VAR=변수 옵션들;
  ESTIMATE 옵션들;
  FORECAST 옵션들;
```

2) PROC ARIMA스태이트먼트

PROC ARIMA스태이트먼트는 PROC ARIMA의 시작을 알려 준다. 이 PROC ARIMA스태이트먼트의 사용양식은 다음과 같다.

```
PROC ARIMA 옵션들;
```

PROC ARIMA스태이트먼트의 옵션들은 다음과 같다.

(1) DATA = SAS데이터세트

이 옵션 DATA=에는 분석할 시계열데이터를 포함하는 SAS데이터세트를 지정한다. 만일, IDENTIFY스태이트먼트의 옵션 DATA=에 지정한 SAS데이터세트와 PROC ARIMA스태이트먼트의 옵션 DATA=에 지정한 SAS데이터세트가 다른 경우에는 IDENTIFY스태이트먼트의 옵션 DATA=에 지정한 SAS데이터세트를 분석한다. 만일 IDENTIFY스태이트먼트에서나 PROC ARIMA스태이트먼트에서 옵션 DATA=을 사용하지 않으면 마지막으로 만들어진 SAS데이터세트를 분석한다.

(2) OUT = SAS데이터세트

이 옵션 OUT=에는 예측의 결과를 저장할 SAS데이터세트를 지정한다. 만일, FORE-

CAST스태이트먼트의 옵션 OUT=에 지정한 SAS데이터세트와 PROC ARIMA스태이트먼트의 옵션 OUT=에 지정한 SAS데이터세트가 다른 경우에는 FORECAST스태이트먼트의 옵션 OUT=에 지정한 SAS데이터세트를 사용한다.

3) IDENTIFY스태이트먼트

IDENTIFY스태이트먼트는 추정할 모형을 식별하기 위한 것이다. 모형화하기 위한 각 시계열데이터에 대해서 하나의 IDENTIFY스태이트먼트를 사용한다. 만일, 뒤따르는 ESTIMATE스태이트먼트의 옵션 INPUT=에 변수들이 지정되면 이는 입력시계열데이터들을 나타낸다. 이 입력변수들은 반드시 IDENTIFY스태이트먼트의 옵션 CROSSCOR=에 지정되어야 한다. 이 IDENTIFY스태이트먼트의 사용양식은 다음과 같다.

IDENTIFY VAR=종속변수 옵션들;
또는 I VAR=종속변수 옵션들;

IDENTIFY스태이트먼트의 옵션들은 다음과 같다.

(1) VAR = 변수(d_1, d_2, \dots, d_k)

IDENTIFY스태이트먼트에는 반드시 이 옵션 VAR=을 사용해야 한다. 따라서 옵션이라는 단어를 사용하는 것이 부적절하지만 편의상 이 단어를 사용하기로 한다. 여기서 변수는 식별할 시계열데이터, 즉 종속변수이다. 괄호 안의 숫자들은 이 시계열데이터를 차분할 주기(periods)들을 순서대로 나열한 것이다. 만일 시계열데이터를 차분할 필요가 없다면, 주기들을 나열할 괄호를 사용할 필요가 없다. 차분의 예들과 이에 해당하는 식들은 다음과 같다.

- ① Y : y_t
- ② Y(1) : $\nabla y_t = y_t - y_{t-1}$
- ③ Y(4) : $\nabla_4 y_t = y_t - y_{t-4}$
- ④ Y(1, 1) : $\nabla \nabla y_t = \nabla y_t - \nabla y_{t-1}$
- ⑤ Y(1, 12) : $\nabla_{12} \nabla y_t = \nabla_{12} y_t - \nabla_{12} y_{t-1}$
- ⑥ Y(5, 3, 2) : $\nabla_2 \nabla_3 \nabla_5 y_t = \nabla_2 \nabla_3 y_t - \nabla_2 \nabla_3 y_{t-5}$

(2) CENTER

이 옵션을 사용하면 각 시계열데이터에서 표본평균을 뺀 중심화된(centered) 시계열데이터를 만든다. 분석(analysis)을 할 때는 이 중심화된 시계열데이터를 사용한다. 그러나 이러한 분석결과로 예측된 값에 표본평균을 더해서 예측값을 계산한다. 차분을 취한 시계열데이터에서는 차분을 취한 다음에 중심화(centering)를 한다. 옵션 CENTER와 ESTIMATE스태이트먼트의 옵션 NOCONSTANT가 함께 사용되면 계산이 빨라질 것이다.

(3) CLEAR

앞에서 사용된 모든 모형들(all old models)을 제거한다. 이 옵션은 앞에서 어떤 입력시계열데이터를 사전백색화한(prewhitened) 경우에 이 사전백색화를 무효로 하고 원래의 시계열데이터를 분석하기 위해서 사용한다.

(4) CROSSCORR = (변수(d_1, \dots, d_{k1}), 변수(d_1, \dots, d_{k2}), ...)

CROSSCOR = (변수(d_1, \dots, d_{k1}), 변수(d_1, \dots, d_{k2}), ...)

CROSS = (변수(d_1, \dots, d_{k1}), 변수(d_1, \dots, d_{k2}), ...)

이 옵션에는 옵션 VAR=에 지정된 종속변수와 교차상관된 입력변수들을 지정한다. 괄호 안의 숫자들은 이 입력시계열데이터를 차분할 주기들을 순서대로 나열한 것이다. 만일, 이 입력시계열데이터를 차분할 필요가 없다면 주기들을 나열할 괄호를 사용할 필요가 없다. 이러한 차분들은 IDENTIFY스테이트먼트와 ESTIMATE스테이트먼트에서 이 입력시계열데이터를 분석하기 전에 수행된다.

또한 PROC ARIMA가 IDENTIFY스테이트먼트, ESTIMATE스테이트먼트, IDENTIFY스테이트먼트로 구성되었고, 첫 번째 IDENTIFY스테이트먼트에는 옵션 CROSSCOR=을 사용하지 않았고, ESTIMATE스테이트먼트에는 옵션 INPUT=을 사용하지 않았다고 가정해 보자. 또한, 두 번째 IDENTIFY스테이트먼트의 옵션 CROSSCOR=에 지정된 입력변수가 첫 번째 IDENTIFY스테이트먼트의 옵션 VAR=에 종속변수로 지정되고, 또한 ESTIMATE스테이트먼트에 의해서 추정되었다고 가정해 보자. 두 번째 IDENTIFY스테이트먼트에서 옵션 CROSSCOR=에 지정한 입력시계열데이터들과 옵션 VAR=에 지정한 출력시계열데이터 사이의 표본교차상관함수들을 계산하기 위해서는 다음과 같은 과정을 거친다. 우선 두 번째 IDENTIFY스테이트먼트의 옵션 VAR=과 옵션 CROSSCOR=에 지정된 모든 변수들을 이 옵션들에서 지정된 차분차수들에 맞추어 차분을 한 다음, ESTIMATE스테이트먼트에서 추정된 입력시계열데이터의 추정모형으로 여과한 뒤, 이 여과된 시계열데이터들의 표본교차상관함수들을 계산한다. 여기서 유의할 점은 두 번째 IDENTIFY스테이트먼트의 옵션 CROSSCOR=에 지정된 차분차수들이 첫 번째 IDENTIFY스테이트먼트에서 지정된 차분차수들을 다 포함하고 있지 않으면, 사전백색화 과정에서 적용되는 차분차수들은 두 번째 IDENTIFY스테이트먼트의 옵션 CROSSCOR=에 지정된 차분차수들이 아니라 첫 번째 IDENTIFY스테이트먼트의 옵션 VAR=에 지정된 차분차수들이라는 것이다.

한 예로 다음과 같은 SAS프로그램을 살펴보자.

```
PROC ARIMA DATA=SAMPLE;
  IDENTIFY VAR=X(1, 4);
  ESTIMATE P=1 Q=1;
  IDENTIFY VAR=Y(1) CROSS=(X(1));
```

이 SAS프로그램을 수행하면 다음과 같은 결과를 얻는다. 우선 첫 번째 IDENTIFY스테

이트먼트에서는 다음과 같은 시계열데이터 $\{u_t\}$ 가 계산된다.

$$u_t = \nabla_4 \nabla x_t$$

첫 번째 ESTIMATE스테이트먼트는 다음 식을 추정한다.

$$(1 - \phi_1 B)u_t = (1 - \theta_1 B)z_t$$

두 번째 IDENTIFY스테이트먼트를 수행하면 우선 다음과 같은 여과된 출력시계열데이터 $\{w_t\}$ 를 만든다.

$$w_t = \frac{1 - \hat{\phi}_1 B}{1 - \hat{\theta}_1 B} \nabla_4 \nabla y_t$$

또한, 이 사전백색화된 입력시계열데이터 $\{z_t\}$ 와 여과된 출력시계열데이터 $\{w_t\}$ 사이의 표본교차상관함수를 계산한다. 그러나 새로운 차분들을 지정하는 IDENTIFY스테이트먼트가 사용될 때까지 이 두 번째 IDENTIFY스테이트먼트 이후의 ESTIMATE스테이트먼트와 FORECAST스테이트먼트에서는 두 번째 IDENTIFY스테이트먼트에서 지정한 차분을 사용한다. 즉, 시계열데이터들 $\{\nabla x_t\}$ 와 $\{\nabla y_t\}$ 가 분석된다.

다음과 같은 SAS프로그램을 살펴보자.

```
PROC ARIMA DATA=SAMPLE;
  IDENTIFY VAR=X(1);
  ESTIMATE P=1 Q=1;
  IDENTIFY VAR=Y(1, 4) CROSS=(X(1, 4));
```

이 SAS프로그램을 수행하면 다음과 같은 결과를 얻는다. 우선 첫 번째 IDENTIFY스테이트먼트의 결과는 다음과 같다.

$$u_t = \nabla x_t$$

첫 번째 ESTIMATE스테이트먼트는 다음 식을 추정한다.

$$(1 - \phi_1 B)u_t = (1 - \theta_1 B)z_t$$

두 번째 IDENTIFY스테이트먼트를 수행하면 우선 다음과 같은 백색잡음화된 입력시계열데이터 $\{b_t\}$ 와 여과된 출력시계열데이터 $\{w_t\}$ 를 만든다.

$$b_t = \frac{1 - \hat{\phi}_1 B}{1 - \hat{\theta}_1 B} \nabla_4 \nabla x_t$$

$$w_t = \frac{1 - \hat{\phi}_1 B}{1 - \hat{\theta}_1 B} \nabla_4 \nabla y_t$$

또한, 이 사전백색화된 입력시계열데이터 $\{b_t\}$ 와 $\{w_t\}$ 사이의 표본교차상관함수를 계산한

다. 그러나 이 경우에 입력시계열데이터 $\{x_t\}$ 를 한 번 차분한 $\{\nabla x_t\}$ 가 정상성을 갖는다면, 추가적으로 계절주기 4의 계절차분한 입력시계열데이터를 추정한 결과로 얻어진 잔차가 백색잡음성을 갖지 않을 것이다. 따라서 제 4.3.1소절에서 제시한 Box-Jenkins의 전이함수 모형 식별방법을 적용할 수 없다.

차분을 취해서 비정상시계열데이터를 정상시계열데이터로 만들 때는 옵션 CROSSCOR = 을 주의해서 사용해야 한다. 한 예로 다음과 같은 모형을 살펴보자.

$$y_t = \frac{\omega_0}{1 - \delta_1 B} x_t + \frac{1 - \omega_1 B}{(1 - B)(1 - B^{12})} u_t$$

이 모형을 추정하기 위해서는 다음 SAS프로그램을 수행해야 한다.

```
PROC ARIMA DATA=SAMPLE;
  I VAR=X(1, 12) CROSSCOR=(X(1, 12));
  E Q=1 INPUT=(/ (1) X);
```

다음과 같은 모형을 살펴보자.

$$y_t = \frac{\omega_0}{(1 - \delta_1 B)(1 - B)(1 - B^{12})} x_t + \frac{1 - \omega_1 B}{(1 - B)(1 - B^{12})} u_t$$

이 모형을 추정하기 위해서는 다음 SAS프로그램을 수행해야 한다.

```
PROC ARIMA DATA=SAMPLE;
  I VAR=Y(1, 12) CROSSCOR=(X);
  E Q=1 INPUT=(/ (1) X);
```

(5) DATA = SAS데이터세트

분석할 시계열데이터를 포함한 SAS데이터세트를 지정한다.

(6) NLAG = 정수

계산할 표본자기상관계수들, 표본부분자기상관계수들, 표본역자기상관계수들, 표본교차상관계수들의 시차들의 개수를 지정한다. 옵션 NLAG=에 지정하는 값은 관찰점들의 개수보다 커야 하며, ARIMA(p, d, q)의 예비추정값들(preliminary estimates)을 구하기 위해서는 옵션 NLAG=에 ($p+d+q$)보다 크거나 같은 값을 지정해야 한다. 옵션 NLAG=의 디폴트는 24이다.

(7) NOMISS

만일, 주어진 시계열데이터에서 첫 결측점이 있는 시점 바로 전까지의 관찰점들(the first contiguous sequence of data with no missing values)을 사용하기 위해서는 옵션 NOMISS를 사용한다. 디폴트는 모든 관찰점들을 사용하는 것이다.

(8) NOPRINT

IDENTIFY스태이트먼트를 수행한 결과의 출력을 억제한다.

(9) OUTCOV = SAS데이터세트

표본자기공분산함수, 표본역자기상관함수, 표본부분자기상관함수, 표본교차상관함수를 저장할 SAS데이터세트를 지정한다. 옵션 OUTCOV=에 지정한 SAS데이터세트에 관한 자세한 내용은 제 4.1.5소절, 또는 [ETS6.1] (pp. 114~115)을 참조하라.

4) ESTIMATE스태이트먼트

ESTIMATE스태이트먼트는 선행하는 IDENTIFY스태이트먼트에 명시된 변수들의 일부, 또는 전체를 포함하는 모형을 추정한다. 추정하고자 하는 모형 하나에 하나의 ESTIMATE스태이트먼트를 사용한다.

이 ESTIMATE스태이트먼트의 사용양식은 다음과 같다.

ESTIMATE 옵션들;
또는 E 옵션들;

ESTIMATE스태이트먼트의 옵션들은 다음과 같다.

(1) ALTPARM

입력시계열데이터의 유리함수형태 전이함수의 분자의 첫 번째 인자(factor)는 다음과 같은 형태이다.

$$\omega_0 - \omega_1 z^1 - \dots - \omega_k z^k$$

그러나 만일 옵션 ALTPARM이 사용되면 유리함수형태 전이함수의 분자의 첫 번째 인자는 다음과 같은 형태이다.

$$\omega_0(1 - \omega_1 z^1 - \dots - \omega_k z^k)$$

옵션 ALTPARM에 대해서는 제 4.3.3소절을 참조하라.

(2) GRID

옵션 GRID를 사용하면 최종추정값들 주변의 모수공간의 격자점들에서(in a small grid of the parameter space around the final estimates) 계산된 오차제곱합(error sum of squares), 또는 대수우도함수값(concentrated log-likelihood surface)을 출력한다. 즉, 각 쌍의 모수들에 대해서 최종추정값들을 중심으로 하고 옵션 GRIDVAL=에 지정된 값을 간격으로 하는 9개의 격자점들에서 계산된 오차제곱합들을 출력한다. 이 옵션 GRID는 처음 접한 시계열데이터를 분석하는 데 적당하다. 이는 우도함수를 최대화하거나 오차제곱합을 최소화하는 알고리즘이 항상 최적값에 수렴하는 것은 아니기 때문에 SAS시스템이

제공하는 최종추정값들이 정말로 최적인가를 조사해 보는 것은 중요하다.

(3) GRIDVAL = 숫자

옵션 GRIDVAL=에는 옵션 GRID를 사용해서 오차제곱합들을 출력시킬 때 모수들의 계자점들 사이의 간격을 지정한다. 옵션 GRIDVAL=의 디폴트는 0.005이다. 일반적으로 이 값은 최종추정값들의 최적성 여부를 조사하는데 유용하다.

(4) INPUT = (형태 1, 변수 1, 형태 2, 변수 2, ...)

옵션 INPUT=은 전이함수모형이나 간섭모형(intervention model)을 사용할 때 입력 변수들과 각 입력변수에 해당하는 전이함수를 명시하는데 사용한다. 옵션 INPUT=에 명시되는 입력변수들, 즉 변수 1, 변수 2, ... 들은 이 옵션 INPUT=이 사용되는 ESTIMATE스테이트먼트의 바로 앞에 사용된 IDENTIFY스테이트먼트의 옵션 CROSSCOR=에 지정된 변수들이어야 한다. 만일, 이 IDENTIFY스테이트먼트의 옵션 CROSSCOR=의 어떤 변수에 차분이 지정되면, 이 차분된 변수가 전이함수의 입력시계열데이터가 된다. 옵션 INPUT=에 지정되는 형태 1, 형태 2, ... 의 양식은 다음과 같다.

$$b(L_{1,1}, L_{1,2}, \dots) \dots (L_{i,1}, L_{i,2}, \dots) / (L_{j,1}, L_{j,2}, \dots) \dots (L_{k,1}, L_{k,2}, \dots)$$

여기서 b 는 지연모수이고, 각 괄호는 멱함수인자(polynomial factor)를 뜻한다. 이 멱함수인자들 중에서 슬러시(slash, /) 이전의 멱함수인자들은 전이함수의 분자의 인자들이고, 슬러시 이후의 멱함수인자들은 분모의 인자들이다. 만일, 멱함수인자 (L_1, L_2, \dots, L_k) 가 분자의 첫 번째 멱함수인자가 아니라면, 즉 분자의 두 번째 멱함수인자이거나 분모의 멱함수인자라면 다음과 같은 멱함수를 의미한다.

$$(1 - \alpha_{L_1} z^{L_1} - \alpha_{L_2} z^{L_2} - \dots - \alpha_{L_k} z^{L_k})$$

만일, (L_1, L_2, \dots, L_k) 가 분자의 첫 번째 멱함수인자라면 이 멱함수인자는 다음과 같은 멱함수를 의미한다.

$$(\omega_0 - \omega_{L_1} z^{L_1} - \omega_{L_2} z^{L_2} - \dots - \omega_{L_k} z^{L_k})$$

만일, ESTIMATE스테이트먼트의 옵션 ALTPARM을 사용하면 분자의 첫 번째 멱함수인자 (L_1, L_2, \dots, L_k) 는 다음과 같은 멱함수를 의미한다.

$$\omega_0(1 - \omega_{L_1} z^{L_1} - \omega_{L_2} z^{L_2} - \dots - \omega_{L_k} z^{L_k})$$

만일, 지연모수가 0이면 b 부분을 생략하고, 만일 전이함수의 분자가 상수이거나 또는 분모가 1이면 이에 해당하는 멱함수인자들을 생략한다. 옵션 INPUT=의 자세한 사용법은 제 4.3.3소절을 참조하라. 다음과 같은 옵션 INPUT=의 예제들과 이에 해당하는 전이함수들을 살펴보자.

$$\textcircled{1} \text{ INPUT} = (3\$X)$$

$$\omega_0 x_{t-3}$$

$$\textcircled{2} \text{ INPUT} = ((1 \ 4) X)$$

$$(\omega_0 - \omega_1 B - \omega_4 B^4) x_t$$

$$\textcircled{3} \text{ INPUT} = (3\$ / (1) X)$$

$$\frac{1}{1 - \delta_1 B} x_{t-3}$$

$$\textcircled{4} \text{ INPUT} = (1\$ (1 \ 4) / (1 \ 2) X)$$

$$\frac{\omega_0 - \omega_1 B - \omega_4 B^4}{1 - \delta_1 B - \delta_2 B^2} x_{t-1}$$

$$\textcircled{5} \text{ INPUT} = (5\$ (1)(1 \ 3)(2 \ 4) / (2)(1 \ 2)(1 \ 2 \ 5) X)$$

$$\frac{(\omega_{1,0} - \omega_{1,1} B)(1 - \omega_{2,1} B - \omega_{2,2} B^2)(1 - \omega_{3,1} B - \omega_{3,2} B^2)}{(1 - \delta_{1,1} B^2)(1 - \delta_{2,1} B - \delta_{2,2} B^2)(1 - \delta_{3,1} B - \delta_{3,2} B^2 - \delta_{3,3} B^3)} x_{t-5}$$

$$(5) \text{ METHOD} = \text{CLS} | \text{ML} | \text{ULS}$$

옵션 METHOD=에 CLS를 할당하면 조건부최소제곱추정법, ULS를 할당하면 비조건부최소제곱추정법, 그리고 ML을 할당하면 최우추정법을 적용해서 모형을 추정한다. 옵션 METHOD=의 디폴트는 CLS이다. 옵션 METHOD=은 예측방법의 종류를 결정한다. 옵션 METHOD=에 CLS를 할당하면 무한히 먼 시점들까지의 과거값들(infinite past of the time series)을 주어진 시계열데이터의 표본평균으로 가정하고 예측하는 무한기억예측(infinite memory forecast), 또는 조건부예측(conditional forecast)을 한다. 반면에 옵션 METHOD=에 ULS 또는 ML을 할당하면 과거시점들의 값들을 주어진 관찰값들의 선형결합들로 가정하고 예측하는 유한기억예측(finite memory forecast), 또는 비조건부예측(unconditional forecast)을 한다. 최우추정법을 적용한 경우에 백색잡음과정의 분산의 추정값은 비조건부오차제곱합(unconditional sum of squares)을 잔차의 개수에서 모수들의 개수를 뺀 수로 나눈 것이다. 반면에 최우추정법에 의한 일반적인 백색잡음과정의 분산의 추정값을 얻기 위해서는 옵션 METHOD=에 ML을 지정하면서 ESTIMATE스태이트먼트의 옵션 NODF를 사용한다.

$$(6) \text{ NOCONSTANT 또는 NOINT}$$

옵션 NOCONSTANT, 또는 옵션 NOINT를 사용하면 주어진 시계열데이터를 상수항(constant), 또는 절편항(intercept)이 없는 모형에 적합시킨다. 모든 시계열데이터들이 차분되었거나 IDENTIFY스태이트먼트의 옵션 CENTER가 사용된 경우에 이 옵션 NOCONSTANT가 자주 사용된다. 그러나 차분된 시계열데이터의 표본평균이 반드시 0이 아니기 때문에 IDENTIFY스태이트먼트의 옵션 CENTER가 사용된 경우가 아니면 옵션 NOCONSTANT를 사용하는 것에 신중을 기하여야 한다. 시계열데이터가 중심화(center-

ing)가 되지 않아서 표본평균이 0이 아닌 경우에 옵션 NOCONSTANT를 사용해서 상수항이 없는 모형에 이 시계열데이터를 적합시킨다고 가정해 보자. 이러한 경우에는 모수들의 추정값들이 조악하고(poor), 다단계전방예측값들(multistep forecast values)은 선시차가 증가함에 따라 표본평균에 수렴하는 것이 아니라 0에 수렴한다.

(7) NODF

최우추정법에 의한 일반적인 백색잡음과정의 분산의 추정값, 즉 오차제곱합을 잔차의 개수로 나눈 값을 얻기 위해서는 옵션 METHOD=에 ML을 지정하면서 옵션 NODF를 사용한다. 이 옵션 NODF를 사용하지 않은 경우, 즉 디폴트의 경우에는 백색잡음과정의 분산의 추정값은 비조건부오차제곱합을 잔차의 개수에서 모수들의 개수를 뺀 수로 나눈 것이다.

(8) NOLS

옵션 NOLS를 사용하면 최우추정법이나 비조건부최소제곱추정법을 적용하기 위한 초기값들로 조건부최소제곱추정값들을 사용하는 것이 아니라 예비추정값들(preliminary estimates)을 사용한다.

(9) NOPRINT

옵션 NOPRINT를 사용하면 ESTIMATE스테이트먼트의 실행결과의 출력을 억제한다.

(10) P = 시차

$$P = (\text{시차}, \dots, \text{시차}) \dots (\text{시차}, \dots, \text{시차})$$

이 옵션들에는 모형의 AR특성식의 차수들을 지정한다. 디폴트는 0인데 이는 모형에 AR부분이 없음을 뜻한다. 괄호 (시차₁, 시차₂, ..., 시차_k)는 시차₁, 시차₂, ..., 시차_k에서 AR모수들을 갖는다는 것을 의미한다. 한 예로 P=(1, 3, 12)는 다음과 같은 AR특성식을 의미한다.

$$\phi(x) = 1 - \phi_{1,1}x - \phi_{1,2}x^3 - \phi_{1,3}x^{12}$$

또한, P=r은 P=(1, 2, ..., r)을 의미한다. 괄호 (시차₁, 시차₂, ..., 시차_k)들의 곱은 이에 상응하는 AR특성식들의 곱을 의미한다. 다음과 같은 예들을 살펴보자.

① P=(1, 2)(4, 8)

$$\phi(x) = (1 - \phi_{1,1}x - \phi_{1,2}x^2)(1 - \phi_{2,1}x^4 - \phi_{2,2}x^8)$$

② P=(1)(3)(4, 8)

$$\phi(x) = (1 - \phi_{1,1}x)(1 - \phi_{2,1}x^3)(1 - \phi_{3,1}x^4 - \phi_{3,2}x^8)$$

③ P=(3)(1, 4, 5)(4, 8)

$$\phi(x) = (1 - \phi_{1,1}x^3)(1 - \phi_{2,1}x - \phi_{2,2}x^4 - \phi_{2,3}x^5)(1 - \phi_{3,1}x^4 - \phi_{3,2}x^8)$$

(11) PLOT

옵션 PLOT를 사용하면 잔차의 표본자기상관함수, 표본역자기상관함수, 그리고 표본부분자기상관함수를 출력한다.

(12) PRINTALL

옵션 PRINTALL을 사용하면 예비추정결과와 최종추정의 각 단계의 결과를 출력한다.

(13) Q = 시차

$$Q = (\text{시차}_1, \dots, \text{시차}_k) \dots (\text{시차}_1, \dots, \text{시차}_k)$$

이 옵션들에는 모형의 MA특성식의 차수들을 지정한다. 디폴트는 0인데 이는 모형에 MA부분이 없음을 뜻한다. 괄호 (시차₁, 시차₂, ..., 시차_k)는 시차₁, 시차₂, ..., 시차_k에서 MA모수들을 갖는다는 것을 의미한다. 또한, Q=r은 Q=(1, 2, ..., r)을 의미한다. 괄호 (시차₁, 시차₂, ..., 시차_k)들의 곱은 이에 상응하는 MA특성식들의 곱을 의미한다. 자세한 사용법은 옵션 P=과 동일하다.

(14) 초기추정값들에 관한 옵션들

ESTIMATE스테이트먼트에는 초기추정값들을 지정하기 위한 옵션들로 AR=, INITVAL=, MA=, MU=, NOEST가 있다. 이에 관한 자세한 내용은 [ETS6.1] (pp. 106~107)을 참조하라.

(15) 추정과정을 조율하기 위한 옵션들

ESTIMATE스테이트먼트에는 추정과정(estimation process)에서 생기는 문제점들을 조율(tuning)하기 위한 옵션들로 BACKLIM=, CONVERGE=, DELTA=, MAXITER= 또는 MAXIT=, NOSTABLE, SINGULAR=이 있다. 이에 관한 자세한 내용은 [ETS6.1] (p. 108)을 참조하라.

5) FORECAST스테이트먼트

FORECAST스테이트먼트는 앞의 ESTIMATE스테이트먼트에 의해서 추정된 모형을 사용해서 시계열데이터의 미래값들을 예측하기 위한 것이다. 이 예측값들을 새로운 SAS데이터세트에 저장할 수도 있다. 만일, 모형에 입력변수들이 있는 경우에 입력변수들의 관찰값들이 주어지지 않은 미래시점들에서 출력시계열데이터의 예측값들을 구하기 위해서는 이 FORECAST스테이트먼트 앞에서 각 입력변수를 단변량ARIMA모형으로 식별하는 IDENTIFY스테이트먼트와 추정하는 ESTIMATE스테이트먼트가 수행되어야 한다. 이 경우에는 추정된 모형의 오차항들이 서로 독립적이라고 가정한다. 만일 이 가정이 성립되지 않으면 추정된 모형을 신뢰하기가 어렵다.

FORECAST스테이트먼트에서 예측값들은 ESTIMATE스테이트먼트에서 적용한 추정방법과 같은 방법으로 계산된다. 만일, ESTIMATE스테이트먼트에서 옵션 METHOD=에 CLS가 할당되면 무한기억예측, 또는 조건부예측을 한다. 반면에 옵션 METHOD=에

ULS 또는 ML을 할당하면 유한기억예측, 또는 비조건부예측을 한다. 모수들이 주어졌다고 가정하면 이 유한기억예측은 최량선형예측(best linear predictor)이다. 그러나 종속변수가 차분이 된 경우에는 비조건부최소제곱추정법, 또는 최우추정법에 의한 전방다단계예측(multi-step-ahead forecast)이 진짜 유한기억예측은 아니다. 이 FORECAST스태이트먼트의 사용양식은 다음과 같다.

FORECAST 옵션들;
또는 F 옵션들;

FORECAST스태이트먼트의 옵션들은 다음과 같다.

(1) BACK = 정수

옵션 BACK=에는 전방다단계예측을 시작하는 시점을 지정한다. 즉, 주어진 시계열데이터의 마지막 관찰점으로부터 뒤로 몇 번째 관찰점부터 전방다단계예측을 시작하는가를 지정한다. 이 전방다단계예측을 시작하는 시점 이전에서는 전방일단계예측(one-step-ahead forecast)을 한다. 주어진 시계열데이터의 마지막 관찰점이 잡음항의 값(noise value)을 계산할 수 있는 마지막 관찰점이다. 만일, 이 시계열모형에 입력변수가 없다면 이 관찰점은 종속변수가 결측값을 갖지 않는 마지막 관찰점(the last nonmissing value of the time series)일 것이고, 또한 이 시계열모형에 입력변수들이 있고 이 입력변수들이 결측값을 하나도 갖지 않는다면 이 관찰점은 종속변수가 결측값을 갖지 않는 마지막 관찰점일 것이며, 반면에 이 시계열모형에 입력변수들이 있고 어떤 입력변수가 결측값을 갖는다면 이 관찰점은 종속변수가 결측값을 갖지 않는 마지막 관찰점 앞에 있을 것이다. 옵션 BACK=의 디폴트는 0이다. 이 경우에는 전방다단계예측을 주어진 시계열데이터의 마지막 관찰점으로부터 시작한다. 만일, 옵션 BACK=에 자연수가 지정되면 PROC ARIMA는 실제값과 예측값을 비교한다. 옵션 BACK=에 지정되는 정수는 관찰점들의 개수에서 모수들의 개수를 뺀 수보다는 작아야 한다.

(2) ID변수 = 변수

입력되는 SAS데이터세트의 변수들 중에서 출력되는 SAS데이터세트의 관찰점들을 구분하는 데(identify) 쓰이는 변수를 지정한다. 입력되는 SAS데이터세트의 여러 변수들 중에서 ID변수와 종속변수만이 출력되는 SAS데이터세트로 복사된다.

(3) INTERVAL = 시간구간(time interval) 또는 정수

옵션 INTERVAL=에는 관찰점들 사이의 시간구간을 지정한다. 이에 대한 자세한 내용은 [SAS4TSA2]의 제2장을 참조하라.

(4) LEAD = 정수

옵션 LEAD=에는 전방다단계예측을 할 미래시점들의 개수를 지정한다. 예를 들어, 관찰점들의 개수가 T이고 옵션 LEAD=에 m이 지정되고 옵션 BACK=에 n이 지정되면, 전

방다단계예측은 시점 ($T-n$)에서 m 개의 미래시점들에 대해 수행된다. 즉, 시점 ($T-n+1$)부터 시점 ($T-n+m$)까지의 전방다단계예측값들이 계산된다. 전이함수모형에서 미래시점들에 대한 입력변수들의 관찰값들이 충분히 주어지지 않으면 옵션 LEAD=에 지정한 모든 미래시점들에서 전방다단계예측값들이 계산되지 않을 수도 있다. 이 옵션 LEAD=의 디폴트는 24이다.

(5) NOOUTALL

출력되는 SAS데이터세트에 전방다단계예측값들을 갖는 관찰점들만을 출력시킨다. 예를 들어, 관찰점들의 개수가 T 이고 옵션 BACK=에 n 이 지정되면 시점 ($T-n+1$) 이후의 관찰점들만을 출력시킨다.

(6) NOPRINT

옵션 NOPRINT를 사용하면 FORECAST스테이트먼트의 수행결과를 출력하지 않는다.

(7) OUT = SAS데이터세트

옵션 OUT=에는 FORECAST스테이트먼트를 수행한 결과를 저장할 SAS데이터세트를 지정한다. 이 SAS데이터세트에 저장되는 변수들에 관해서는 [SAS4TSA2] (pp. 613~614) 와 [ETS6.1] (pp. 113~114)을 참조하라. 만일, FORECAST스테이트먼트에 옵션 OUT=이 사용되지 않으면 PROC ARIMA스테이트먼트의 옵션 OUT=에 지정된 SAS데이터세트에 예측의 결과가 저장된다. 만일, FORECAST스테이트먼트의 옵션 OUT=이 사용되지 않고 PROC ARIMA스테이트먼트의 옵션 OUT=도 사용되지 않으면 출력SAS데이터세트가 만들어지지 않는다.

(8) PRINTALL

출력되는 SAS데이터세트에 모든 관찰점들을 출력시킨다. 즉, 전방다단계예측값들을 갖는 관찰점들뿐만 아니라 전방일단계예측값들을 갖는 관찰점들도 출력시킨다.

5.1.4 PROC AUTOREG의 사용법

이 소절에서는 [SAS4TSA2]의 제 7장과 지금까지 본서에서 다룬 PROC AUTOREG의 사용양식을 요약하고자 한다. 본 소절의 내용은 [ETS6.1]의 제 8장 'The AUTOREG Procedure'를 요약한 것이다.

앞에서 여러 번 언급된 것처럼 잡음항이 백색잡음과정인 AR모형을 가지면 PROC REG 대신에 PROC AUTOREG를 사용해야 한다. 많은 경우에 PROC AUTOREG를 수행한 결과로 얻어진 회귀계수들의 추정값들은 PROC REG를 수행한 결과로 얻어진 회귀계수들의 추정값들과 비슷하다. 그러나 두 프로시저들이 제공하는 표준오차들은 상당히 다를 수 있다. 이 표준오차들의 차이는 회귀계수의 유의성 검정에 큰 영향을 미친다. PROC AUTOREG에서 예측을 할 때는 모형의 시계열데이터부분(time series part of

the model)과 함께 예측값들을 발생시키는 체계적 부분(systematic part in generating predicted values)을 사용하기 때문에 모형의 시계열데이터부분만을 사용하는 PROC REG보다 더 정확한 예측을 할 수 있다.

1) PROC AUTOREG의 모형

PROC AUTOREG에서 다루는 모형은 다음과 같다.

$$y_t = \mu + \sum_{j=1}^k \beta_j x_{t,j} + n_t$$

$$n_t = v_t - \alpha_1 n_{t-1} - \dots - \alpha_p n_{t-p}$$

이 식의 각 성분은 다음과 같다.

- ① $\{y_t\}$ 는 반응(response)을 나타내는 변수, 즉 종속변수이다.
- ② $\{x_{t,j}\}$ 는 제 j 번째 입력변수이다.
- ③ $\{n_t\}$ 는 잡음항으로서 AR(p)모형을 갖는다고 가정한다.
- ④ $\{v_t\}$ 는 백색잡음과정으로서 평균이 0이고 분산이 σ_v^2 이며 오차항(random error)이라 부른다.

PROC AUTOREG의 사용양식은 다음과 같다.

```
PROC AUTOREG 옵션들;
MODEL 종속변수 = 입력변수들 / 옵션들;
BY 변수들;
OUTPUT OUT = SAS데이터세트 옵션들;
```

PROC AUTOREG에는 적어도 하나의 MODEL스태이트먼트가 필요하고, 각 MODEL스태이트먼트에 하나의 OUTPUT스태이트먼트가 따를 수 있다.

2) PROC AUTOREG스태이트먼트

PROC AUTOREG스태이트먼트는 PROC AUTOREG스태이트먼트의 시작을 알려 준다. 이 PROC AUTOREG스태이트먼트의 사용양식은 다음과 같다.

```
PROC AUTOREG 옵션들;
```

PROC AUTOREG스태이트먼트의 옵션들은 다음과 같다.

(1) DATA = SAS데이터세트

이 옵션 DATA=에는 분석할 SAS데이터세트를 지정한다. 만일, PROC AUTOREG스태이트먼트에서 옵션 DATA=을 사용하지 않으면 마지막으로 만들어진 SAS데이터세트를

분석한다.

(2) OUTEST = SAS데이터세트

이 옵션 OUTEST=에는 추정값들을 저장할 SAS데이터세트를 지정한다. 옵션 OUTEST=에 관한 자세한 내용은 [ETS6.1] (p. 176)을 참조하라.

(3) MODEL스테이트먼트와 공유하는 옵션들

다음과 같은 MODEL스테이트먼트의 옵션들은 PROC AUTOREG스테이트먼트의 옵션들로 사용할 수 있다.

ALL, BACKSTEP, COEF, CORRB, COVB, GINV, ITER, ITPRINT,
MAXITER, METHOD=, NOINT, NOMISS, NOPRINT, PARTIAL

3) MODEL스테이트먼트

MODEL스테이트먼트에는 추정할 회귀모형의 종속변수와 입력변수들을 지정한다. PROC AUTOREG에는 적어도 하나의 MODEL스테이트먼트가 있어야 한다. 이 MODEL스테이트먼트의 사용양식은 다음과 같다.

MODEL 종속변수=입력변수들 / 옵션들;

만일, MODEL스테이트먼트에 입력변수들이 하나도 지정되지 않으면 회귀모형은 상수항(constant term)만을 갖는다. 이러한 모형은 시계열데이터의 표본자기상관함수를 구하는 데 사용할 수 있다.

MODEL스테이트먼트의 옵션들은 다음과 같다.

(1) NLAG = 시차

NLAG = (시차, 시차, ..., 시차)

이 옵션에는 잡음항의 AR특성식의 차수들을 지정한다. 디폴트는 이 옵션을 사용하지 않는 것이다. 이는 회귀모형의 잡음항이 백색잡음과정이라고 가정하는 것이다. 괄호 (시차₁, 시차₂, ..., 시차_k)는 잡음항의 AR모형이 시차₁, 시차₂, ..., 시차_k에서 AR모수들을 갖는 것을 의미한다. 한 예로 NLAG=(1, 2, 12)를 지정하면 잡음항은 다음과 같은 AR모형을 갖는다.

$$n_t = v_t - \alpha_1 n_{t-1} - \alpha_2 n_{t-2} - \alpha_{12} n_{t-12}$$

또한, NLAG=r은 NLAG=(1, 2, ..., r)을 의미한다.

(2) ALL

이 옵션 ALL을 사용하면 옵션들 COEF, CORRB, COVB, GINV, 그리고 PARTIAL을 동시에 사용한 것과 같다.

(3) COEF

이 옵션 COEF를 사용하면 첫 p 개의 관찰점들의 변환계수들(transformation coefficients)을 출력시킨다. 이 계수들은 자기공분산함수의 Toeplitz행렬의 Cholesky제곱근의 역행렬에 상수를 곱한 형태이다. 자세한 내용은 [SAS4TSA2] (p. 272)를 참조하라.

(4) CORRB

이 옵션 CORRB를 사용하면 모수들의 최종추정값들 사이의 상관계수들의 추정값들을 출력시킨다.

(5) COVB

이 옵션 COVB를 사용하면 모수들의 최종추정값들 사이의 공분산들의 추정값들을 출력시킨다.

(6) GINV

이 옵션 GINV를 사용하면 Yule-Walker해에 의한 자기공분산값들의 Toeplitz행렬의 역행렬을 출력시킨다.

(7) NOPRINT

이 옵션 NOPRINT를 사용하면 PROC AUTOREG를 수행한 결과의 출력을 억제시킨다.

(8) PARTIAL

이 옵션 PARTIAL을 사용하면 표본부분자기상관계수들을 출력시킨다.

(9) BACKSTEP

이 옵션 BACKSTEP을 사용하면 잡음항의 AR모형이 후향제거방법(backward elimination method)에 의해서 자동적으로 정해진다. 좀더 자세히 설명하면 잡음항의 AR모형에서 유의적이지 않은(nonsignificant), 즉 유의적으로 0과 같은 AR계수들에 해당하는 항들이 제거된다. 이 항들이 제거되는 순서는 0에서 가까운(in order of least significance) AR계수의 순서이다. 이 후향제거방법은 초기의 보통최소제곱추정단계가 수행된 직후 Yule-Walker추정단계에서 단 한 번 실행된다.

(10) SLSTAY = 숫자

이 옵션 SLSTAY=에는 옵션 BACKSTEP이 수행될 때 추정된 AR계수들을 제거하는 기준이 되는 p 값을 지정한다. 이 옵션에는 보통 0.001과 0.2 사이의 값을 지정하며, 디폴트는 0.05이다.

(11) CONVERGE = 숫자

이 옵션 CONVERGE=에는 알고리즘의 수렴성을 판단하는 기준을 지정한다. 만일, 알고리즘이 수행되는 단계에서 AR계수들의 추정값들의 변화량들의 절대값의 최대값(the maximum absolute value of the change in the AR parameter estimates between it-

erations)이 이 옵션에 지정한 값보다 작다면 알고리즘이 수렴한다고 가정한다. 이 옵션 CONVERGE=의 디폴트는 0.001이다.

(12) ITER

이 옵션 ITER은 Yule-Walker추정법의 알고리즘을 반복적용하게 한다. 만일, 옵션 METHOD=에 YW가 아닌 ULS, 또는 ML이 지정되면 이 옵션 ITER은 유효하지가 않다. 디폴트는 Yule-Walker추정법의 알고리즘을 반복적용하지 않는 것이다.

(13) ITPRINT

이 옵션 ITPRINT는 각 반복단계(iteration)의 결과를 자세하게 출력시킨다.

(14) METHOD = YW | ML | ULS

옵션 METHOD=에 YW를 할당하면 Yule-Walker추정법, ULS를 할당하면 비조건부최소제곱추정법, 그리고 ML을 할당하면 최우추정법을 적용해서 모형을 추정한다. 옵션 METHOD=의 디폴트는 YW이다.

(15) NOMISS

이 옵션 NOMISS를 사용하면 결측값들을 가지고 있지 않는 첫 연속된 관찰점들(the first contiguous sequence of data with no missing values)을 사용해서 모형을 추정한다. 디폴트는 모든 관찰점들을 사용해서 모형을 추정하는 것이다.

4) OUTPUT스테이트먼트

OUTPUT스테이트먼트에는 PROC AUTOREG를 수행한 결과를 저장할 SAS데이터세트를 지정한다. OUTPUT스테이트먼트의 사용양식은 다음과 같다.

```
OUTPUT OUT=SAS데이터세트 옵션들
SAS키워드=변수;
```

OUTPUT스테이트먼트의 옵션들은 다음과 같다.

(1) OUT = SAS데이터세트

이 옵션 OUT=에는 예측되거나 변환된 값들을 저장할 SAS데이터세트를 지정한다. 만일, PROC AUTOREG스테이트먼트에서 옵션 OUT=을 사용하지 않으면 PROC AUTOREG는 새로운 SAS데이터세트의 이름을 명명하는 SAS의 규칙(convention)에 따라 DATA n 이라 이름짓는다.

(2) ALPHA CLI = 숫자

이 옵션 ALPHA CLI=에는 반응시계열데이터의 현재 실현(current realization)의 적합값, 또는 예측값의 신뢰구간의 유의수준을 지정한다. 즉, 시계열데이터와 구조적

잔차(structural residuals)의 선행값들을 사용한 신뢰구간을 구하기 위한 유의수준을 지정한다. 이 신뢰구간에 대한 자세한 내용은 [SAS4TSA2] (p. 291)를 참조하라. 결과적으로 얻어지는 신뢰구간은 (1-숫자)의 신뢰성을 갖는다. 물론 지정되는 숫자는 0 보다는 크고 1보다는 작아야 한다. 이 옵션 ALPHACLI=의 디폴트는 0.05이다.

(3) ALPHACLM = 숫자

이 옵션 ALPHACLM=에는 시계열데이터와 구조적 부분(structural part), 즉 회귀부분(regression part)을 사용한 신뢰구간을 구하기 위한 유의수준을 지정한다. 이 신뢰구간에 대한 자세한 내용은 [SAS4TSA2] (p. 291)를 참조하라. 결과적으로 얻어지는 신뢰구간은 (1-숫자)의 신뢰성을 갖는다. 물론 지정되는 숫자는 0보다는 크고 1보다는 작아야 한다. 이 옵션 ALPHACLM=의 디폴트는 0.05이다.

OUTPUT스태이트먼트에서는 'SAS키워드=변수'라는 옵션들을 사용할 수 있는데 이 'SAS키워드=변수'라는 옵션을 사용하면 해당하는 통계량이 출력시계열데이터에 수록되고, 할당된 변수가 출력시계열데이터에서 이 통계량의 변수이름이 된다. OUTPUT스태이트먼트에서 사용할 수 있는 SAS키워드들은 다음과 같다.

(1) CONSTANT = 변수

이 SAS키워드를 사용하면 우선 다음과 같은 변수를 만든다.

$$z_t = 1 + \hat{\alpha}_1 + \dots + \hat{\alpha}_p$$

이렇게 계산된 시계열데이터를 SAS키워드 CONSTANT=에 지정한 변수로 출력SAS데이터세트에 저장한다.

(2) LCL = 변수

이 SAS키워드를 사용하면 예측값의 신뢰구간의 하한(the lower confidence limit for the predicted value)을 이 옵션에 지정한 변수로 출력SAS데이터세트에 저장한다.

(3) LCLM = 변수

이 SAS키워드를 사용하면 구조적 예측값의 신뢰구간의 하한(the lower confidence limit for the structural predicted value)을 이 옵션에 지정한 변수로 출력SAS데이터세트에 저장한다.

(4) PREDICTED = 변수 또는 P = 변수

이 SAS키워드를 사용하면 예측값들(the predicted values)을 이 옵션에 지정한 변수로 출력SAS데이터세트에 저장한다. 이 예측값들은 모형의 구조부분과 AR부분으로부터 얻어진다.

(5) PREDICTEDM = 변수 또는 PM = 변수

이 SAS키워드를 사용하면 구조적 예측값들(the structural predicted values)을 이 옵션

션에 지정한 변수로 출력SAS데이터세트에 저장한다. 이 예측값들은 모형의 구조부분만을 사용해서 얻는다.

(6) RESIDUAL = 변수 또는 R = 변수

이 SAS키워드를 사용하면 잔차값들(the residuals)을 이 옵션에 지정한 변수로 출력 SAS데이터세트에 저장한다. 이 잔차값은 관찰값과 모형의 구조부분과 AR부분으로부터 얻어진 예측값의 차이이다.

(7) RESIDUALM = 변수 또는 RM = 변수

이 SAS키워드를 사용하면 잔차를 이 옵션에 지정한 변수로 출력SAS데이터세트에 저장한다. 이 잔차값은 관찰값과 모형의 구조부분만을 사용해서 얻어진 예측값의 차이이다.

(8) TRANSFORM = 변수들

이 SAS키워드를 사용하면 우선 입력시계열데이터들을 다음과 같이 변환한다.

$$z_t = x_t + \hat{\alpha}_1 x_{t-1} + \dots + \hat{\alpha}_p x_{t-p}$$

이렇게 변환된 시계열데이터들을 SAS키워드 TRANSFORM=에 지정한 변수들로 출력 SAS데이터세트에 저장한다.

(9) UCL = 변수

이 SAS키워드를 사용하면 예측값의 신뢰구간의 상한(the upper confidence limit for the predicted value)을 이 옵션에 지정한 변수로 출력SAS데이터세트에 저장한다.

(10) UCLM = 변수

이 SAS키워드를 사용하면 구조적 예측값의 신뢰구간의 상한(the upper confidence limit for the structural predicted value)을 이 옵션에 지정한 변수로 출력SAS데이터세트에 저장한다.

5.2 간섭시계열모형

어떤 시계열데이터에 영향을 미치는 외부요인의 변화에 의해서 시계열데이터의 형태가 급격한 변화를 보이거나 또는 급격한 변화가 있는 것처럼 보이는 경우가 있다. 예를 들면, 백화점에서 어떤 상품을 일정한 기간 동안 할인판매하였을 경우 할인기간에는 일별판매량이 급격히 증가했다가 할인기간이 끝나면 감소하는 형태를 보일 것이고, 자동차의 안전띠 착용을 법으로 의무화했을 때 교통사고 사망자의 수가 갑자기 감소할 것이며, 강의 상류에 댐이 생겼을 때는 강의 수위에 급격한 감소가 있을 것이다. 또한, 천재지변, 전쟁, 석유파동

등과 같은 정치적·사회적, 또는 경제적 여건의 급격한 변화는 대부분의 경제시계열데이터들의 흐름에 커다란 영향을 미칠 수가 있다. 어떤 시계열데이터를 발생시키는 확률구조에 대한 이러한 외부의 개입(intrusion)을 간섭(intervention)이라 부른다. 또한 이렇게 간섭된(intervened) 시계열데이터를 간섭시계열데이터, 이 시계열데이터의 모형을 간섭시계열모형(intervention model)이라 부른다. 본 저자는 intervention이라는 단어를 형태변화, 국면조정, 국면변환, 간섭, 간여 등 여러 가지로 번역해 보았다. 수동태 문장을 잘 쓰지 않는 우리말의 특성에도 불구하고 ‘간섭된’이라는 단어가 가장 가까운 뜻을 가진 것 같아서 intervention을 간섭으로 번역하기로 한다. 간섭시계열데이터는 어느 시점에서 시계열데이터의 수준(level, 여기서는 평균이라는 의미로 해석함)이 변화하는 것을 나타낼 뿐만 아니라, 어느 시점 이후에서 시계열데이터가 상승 또는 하강하는 형태로 변화하는 것을 나타내기도 한다. 이러한 간섭이 일어났을 때 이에 대응하는 시계열데이터의 변화가 일어났는지, 그리고 변화가 일어났다면 그 변화의 형태는 어떤 것인지를 분석할 필요가 있다. 이러한 분석을 간섭시계열분석이라 한다.

간섭시계열데이터의 간섭시점 전후의 평균의 변화를 검정하기 위해서 전통적으로는 t 검정법이 사용되었다. 그러나 다음과 같은 이유들 때문에 t 검정법은 간섭시계열데이터를 분석하는데 적당하지가 않다.

- ① 연속적인 시점에서 관찰되는 시계열데이터는 계열상관되어(serially correlated) 있는 경우가 일반적이며, 이러한 시계열데이터는 안정성조건을 만족시키지 못하는 경우가 많다. Abraham(1987)은 계열상관이 존재할 때 수준의 변화를 검정하기 위해서 t 검정법을 사용하는 것이 부적절함을 보였다.
- ② 간섭시계열데이터가 발생하는 간섭의 종류에는 t 검정법의 적용을 고려해 볼 수도 있는 수준의 변화가 생긴 경우 뿐만 아니라 간섭시점 이후 시계열데이터가 직선 형태로 증가하거나 감소하는 경우, 시계열데이터가 지수적으로 증가하거나 또는 지수적으로 감소하는 경우, 시계열데이터가 수준의 변화를 일으킨 뒤 이 수준선(level line)을 중심으로 진동하는 경우 등 여러 가지가 있다.

외부의 간섭에 의해서 확률구조가 변화된 시계열데이터를 분석하는 과정에서 t 검정법이 갖는 이러한 문제점들을 해결하는 방법의 하나로 Box & Tiao(1975)는 시계열데이터의 형태변화가 일어났을 때 이를 모형에 반영하는 간섭시계열모형을 사용할 것을 제안하였다. 간섭시계열분석은 간섭시점 이전의 안정상태에서 간섭시점 이후의 새로운 균형상태에 이르는 경우, 또는 간섭시점 이후에 균형상태에 이르지 못하는 경우의 시계열데이터의 구조의 변화를 나타내는 잠정적인 확률모형을 설정하고 Box-Jenkins의 ARIMA모형화와 마찬가지로 식별, 적합, 검진의 3단계 과정을 반복 실시해서 주어진 시계열데이터에 적합한 시계열모형을 찾는 것이다.

이 절에서는 Box & Tiao(1975)가 사용한 기호를 사용하기 위해서 관찰점의 개수를 T 가 아닌 n 이라 표기하기로 하고 간섭시점을 T 로 표기한다. 간섭시계열분석을 하기 위해서는 다음과 같은 가정을 필요로 한다.

- ① 외부의 간섭이 발생한 간섭시점 T 를 알고 있다.
- ② 간섭시점 T 에서 발생한 동태모형(dynamic model)의 형태를 알고 있다. 즉, 간섭의 속성과 크기를 알고 있다.
- ③ 간섭시점 T 이전과 이후의 잡음항모형이 동일하다.

5.2.1 동태모형과 잡음항모형

시계열데이터 $y_1, y_2, \dots, y_{T-1}, y_T, y_{T+1}, \dots, y_n$ 이 다음과 같은 간섭시계열모형을 만족한다고 하자.

$$y_t = f(\delta, \omega, I_t^{(T)}) + n_t \quad (t = 1, 2, \dots)$$

이 간섭시계열모형에서 잡음항을 제거한 동태모형 $f(\delta, \omega, I_t^{(T)})$ 는 간섭시점 T , 간섭의 효과를 나타내는 모수벡터들 δ, ω 와 간섭변수(intervention variable) $\{I_t^{(T)}\}$ 의 함수이다. 여기서 $\{I_t^{(T)}\}$ 는 0이나 1의 값을 갖는 지시변수(indicator variable)이다. 따라서 간섭시계열데이터모형은 입력시계열데이터가 간섭변수 $\{I_t^{(T)}\}$ 인 전이함수모형이라고 할 수 있다. 간섭시계열분석에서는 일반적으로 잡음항 $\{n_t\}$ 가 승법계절ARIMA모형을 따른다고 가정한다.

동태모형의 형태에 따라 다양한 간섭시계열모형을 생각할 수가 있다. 하나의 간섭변수를 갖는 간섭시계열데이터의 동태모형을 다음과 같이 나타내자.

$$f_t = f(\delta, \omega, I_t^{(T)}) = \frac{\omega(B)}{\delta(B)} I_t^{(T)} \quad (t = 1, 2, \dots)$$

$$\begin{aligned} \text{단, } \omega(z) &= \omega_0 - \omega_1 z - \dots - \omega_s z^s \\ \delta(z) &= 1 - \delta_1 z - \dots - \delta_r z^r \end{aligned}$$

여기서 $\delta(z)=0$ 의 모든 근들이 단위원 밖에 있을 때 동태모형의 출력반응값이 f_t 로 수렴하기 때문에, 일반적인 간섭시계열분석에서는 $\delta(z)=0$ 의 모든 근들이 단위원 밖에 있다고 가정을 한다. 김동원(1993)은 $\delta(z)=0$ 의 모든 근들이 단위원 위, 또는 밖에 있다고 가정해서 좀더 다양한 간섭시계열모형들을 고찰하였다. 특히 동태모형이 간섭변수의 유리함수형태이고 잡음항이 ARMA모형을 갖는 간섭시계열모형은 다음과 같다.

$$y_t = \frac{\omega(B)}{\delta(B)} I_t^{(T)} + \frac{\theta(B)}{\phi(B)} v_t$$

간섭변수 $I_t^{(T)}$ 는 간섭시점 $t=T$ 에서의 간섭을 나타내는 지시변수로서 다음과 같은 세 가지 형태 중의 하나를 갖는다고 가정하자.

$$I_t^{(T)} = S_t^{(T)} = \begin{cases} 0 & (t < T) \\ 1 & (t \geq T) \end{cases}$$

$$I_t^{(n)} = P_t^{(n)} = \begin{cases} 0 & (t \neq T) \\ 1 & (t = T) \end{cases}$$

$$I_t^{(n)} = \eta_t^{(T_1, T_2)} = \begin{cases} 1 & (T_1 \leq t \leq T_2) \\ 0 & (\text{otherwise}) \end{cases}$$

여기서 $S_t^{(T)}$ 는 계단입력함수(step input function)라고 하며 간섭시점 T 에서 시작되는 수준변화를 나타내고, $P_t^{(n)}$ 는 순간과동입력함수(pulse input function)라고 하며 간섭시점 T 에서만 유효한 간섭을 나타낸다. 또한 $\eta_t^{(T_1, T_2)}$ 는 확장된 순간과동입력함수(extended pulse input function)라고 하는데 이 확장된 순간과동입력함수는 어떤 정책을 실행하는 동안 발생하는 간섭효과를 나타내는데 유용하다. 계단입력함수 $S_t^{(T)}$ 와 순간과동입력함수 $P_t^{(n)}$ 는 다음과 같은 관계를 만족한다.

$$(1-B) P_t^{(n)} = S_t^{(T)}$$

5.2.2 Box-Tiao의 간섭시계열모형

Box & Tiao(1975)는 여섯 가지 기본적인 동태모형(dynamic model)들을 바탕으로 간섭시계열데이터를 분석할 것을 제안하였다. 이후 편의상 동태모형을 $f_t = f(\omega, \delta, I_t^{(T)})$ 로 표시하자. 이 여섯 가지 동태모형들은 간섭시점 T 까지는 0이고, 간섭시점에서 계단입력 또는 순간과동입력을 받아서 시점 $(T+1)$ 에서부터 반응을 나타낸다.

【모형 1】 $f_t = \omega B S_t^{(T)}$

어느 강의 상류에 댐이 생기면 이후 강하류의 어느 지점에서의 수위는 급격히 줄어들 것이다. 이 동태모형은 이러한 현상을 나타내는 것이다. 이 동태모형은 간섭시점 T 에서 계단입력이 일어나서 시점 $(T+1)$ 부터 ω 만큼의 가파른(abrupt) 계단변화가 일어난다. 이 계단변화는 영구적(permanent)이다. 이 동태모형을 다시 쓰면 다음과 같다.

$$f_t = \begin{cases} 0 & (t = 1, \dots, T) \\ \omega & (t = T+1, \dots, n) \end{cases}$$

이 동태모형의 예를 살펴보기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM5-4.PRG에 수록되어 있다).

```

/*                                     */
/*                                     EXAM5-4.PRG                             */
/*                                     */
/*                                     BOX-TIAO INTERVENTION MODEL 1          */
/*                                     */

```

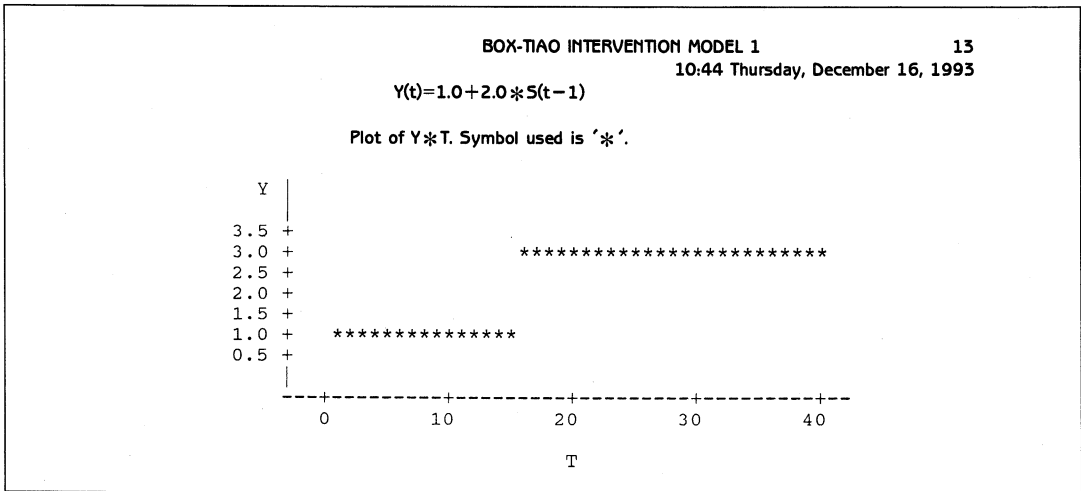



그림 5.4 동태모형 $f_t = 2BS^{(15)} + 1$

```

DATA EXAM54;
  START: T+1;
  IF STP=. THEN STP1=0;
  STP=(T>=15);
  Y=1.0+2.0*STP1;
  OUTPUT;
  STP1=STP;
  IF T<40 THEN GO TO START;

RUN;

PROC PLOT;
  PLOT Y*T='*';
  /VAXIS = 0.5 TO 3.5 BY 0.5;
  TITLE 'BOX-TIAO INTERVENTION MODEL 1';
  TITLE3 'Y(t)=1.0+2.0*S(t-1)';
RUN;
    
```

이 SAS프로그램을 수행한 결과가 그림 5.4이다. 그림 5.4는 간섭시점 15에서 계단입력이 있고 이 간섭이 시점 16에서부터 점프의 크기 2로 동태모형의 출력값에 반영되고 있다.

【모형 2】 $f_t = \frac{\omega B}{1-\delta B} S_t^{(T)}$ ($0 < \delta < 1$)

어떤 상품의 텔레비전광고를 시작하면 이 상품의 매출액이 증가할 것이다. 매출액의 증가가 지수적이라고 가정하면 매출액을 이 동태모형으로 나타낼 수 있다. 이 동태모형은 간섭시점 T 에서 계단입력이 일어나서, 자연수 k 에 대해 시점 $(T+k)$ 에서 동태모형의 출력값에 $\omega\delta^k$ 를 더한다. 만일, 시점 t 가 무한히 커지면 동태함수는 $\omega/(1-\delta)$ 에 수렴한다. 이 값을 안정상태이득(steady state gain)이라 부른다. 차분방정식을 풀어서 이 동태모형을 다

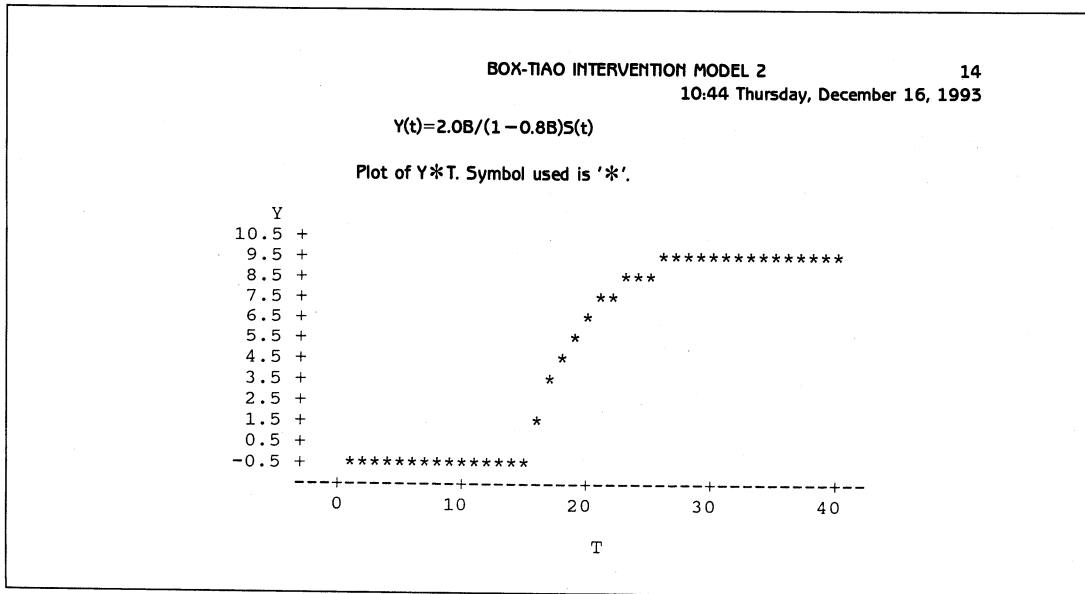


그림 5.5 동태모형 $f_t = \frac{2.0B}{1-0.8B} S_t^{(15)}$

시 쓰면 다음과 같다.

$$f_t = \begin{cases} 0 & (t = 1, \dots, T) \\ \frac{\omega(1-\delta^{t-T})}{(1-\delta)} & (t = T+1, \dots, n) \end{cases}$$

차분방정식을 푸는 자세한 내용은 부록 A를 참조하라.

이 동태모형의 예를 살펴보기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM5-5.PRG에 수록되어 있다).

```

/* */
/*          EXAM5-5.PRG          */
/* */
/*          BOX-TIAO INTERVENTION MODEL 2          */
/* */
DATA EXAM55;
  START: T+1;
  IF STP1=. THEN STP1=0;
  STP=(T>=15);
  IF Y1=. THEN Y1=0;
  Y=0.8*Y1+2.0*STP1;
  OUTPUT;
  Y1=Y;

```

```

STP1=STP;
IF T<40 THEN GO TO START;
RUN;
PROC PLOT;
PLOT Y*T='*';
/VAXIS = -0.5 TO 10.5 BY 1;
TITLE 'BOX-TIAO INTERVENTION MODEL 2';
TITLE3 'Y(t)=2.0B/ (1-0.8B)S(t)';
RUN;

```

이 SAS프로그램을 수행한 결과가 그림 5.5이다. 그림 5.5는 간섭시점 15에서 계단입력이 있고 이 간섭이 시점 16에서부터 동태함수에 영향을 미친다. 즉, 시점 $(15+k)$ 에서 동태모형의 출력값은 $2(0.8)^k$ 만큼 증가한다.

$$\text{【모형 3】 } f_t = \frac{\omega B}{1-B} S_t^{(T)}$$

어느 생산라인에 새로운 기술이 도입된 후 제품의 생산량이 산술적으로 증가한다고 가정해 보자. 이 모형은 이러한 현상을 나타내는 동태모형이다. 이 동태모형은 간섭시점 T 에서 계단입력이 일어나서, 자연수 k 에 대해 시점 $(T+k)$ 에서 동태모형의 출력값에 ω 를 더한다. 즉, 간섭시점 T 에서부터 동태함수는 기울기가 ω 인 일차함수이다. 따라서 t 가 커짐에 따라 동태함수는 무한히 커진다. 차분방정식을 풀어서 이 동태모형을 다시 쓰면 다음과 같다.

$$f_t = \begin{cases} 0 & (t = 1, \dots, T) \\ \omega(t-T) & (t = T+1, \dots, n) \end{cases}$$

이 동태모형의 예를 살펴보기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM5-6.PRG에 수록되어 있다).

```

/* */
/* EXAM5-6.PRG */
/* */
/* BOX-TIAO INTERVENTION MODEL 3 */
/* */
DATA EXAM56;
START: T+1;
IF STP1=. THEN STP1=0;
STP=(T>=15);
IF Y1=. THEN Y1=0;

```

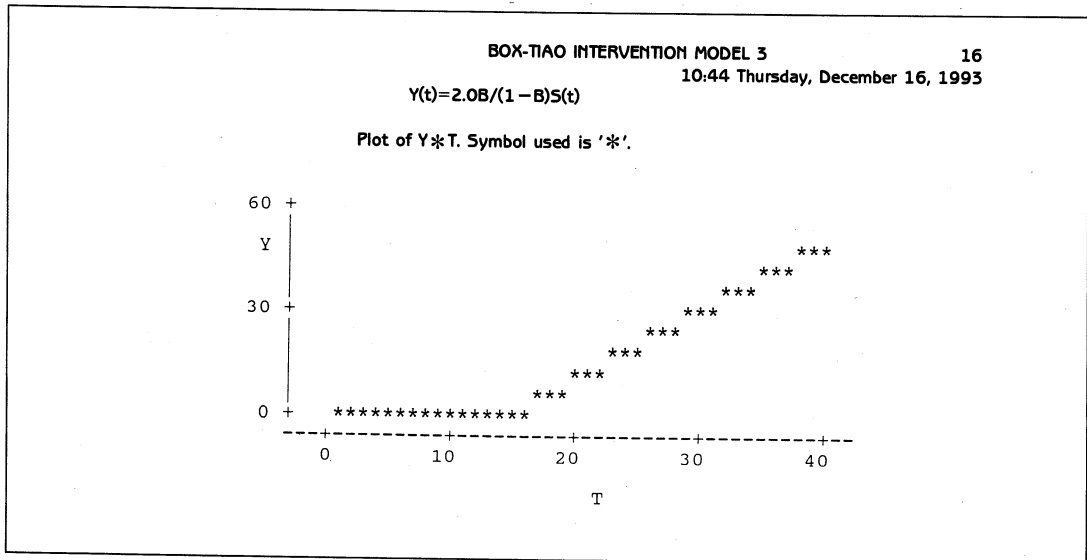


그림 5.6 동태모형 $f_t = \frac{2.0B}{1-B} S_t^{(15)}$

```

Y=Y1+2.0*STP1;
OUTPUT;
Y1=Y;
STP1=STP;
IF T<40 THEN GO TO START;

RUN;
PROC PLOT;
PLOT Y*T='*';
TITLE 'BOX-TIAO INTERVENTION MODEL 3';
TITLE3 'Y(t)=2.0B/ (1-B)S(t)';
RUN;
    
```

이 SAS프로그램을 수행한 결과가 그림 5.6이다. 그림 5.6은 간섭시점 15에서 계단입력이 있고 이 간섭이 시점 16에서부터 동태함수에 영향을 미친다. 즉, 시점 $(15+k)$ 에서 동태모형의 출력값은 $2.0k$ 이다.

【모형 4】 $f_t = \frac{\omega B}{1-\delta B} P_t^{(T)}$ ($0 < |\delta| \leq 1$)

어느 공업도시의 한 화학공장에서 큰 화재가 나서 그 도시의 공기 중 이황화탄소의 함유량이 급격히 높아졌다고 하자. 자연의 정화능력에 의해서 시간이 흐름에 따라 공기 중 이황화탄소의 함유량이 지속적으로 감소해서 오랜 시간이 흐른 뒤에는 원래의 상태로 수렴할 것이다. 이러한 현상을 위와 같은 동태모형으로 나타낼 수 있다. 이 동태모형은 간섭시점 T 에서 순간과동입력이 일어나서, 동태모형의 출력값들이 시점 $(T+1)$ 에서 ω , 시점

($T+2$)에서 $\omega\delta$, 시점 ($T+3$)에서 $\omega\delta^2$ 등 지수적으로 감소한다. 이 동태모형을 다시 쓰면 다음과 같다.

$$\begin{aligned} f_t &= 0 & (t=1, \dots, T-1, T) \\ f_{T+1} &= \omega \\ f_t &= \delta f_{t-1} & (t=T+2, T+3, \dots) \end{aligned}$$

이 차분방정식을 풀면 다음과 같다.

$$f_t = \begin{cases} 0 & (t = 1, \dots, T) \\ \omega\delta^{t-(T+1)} & (t = T+1, \dots, n) \end{cases}$$

이 동태모형의 예를 살펴보기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM5-7.PRG에 수록되어 있다).

```

/*                                     */
/*                                     EXAM5-7.PRG                             */
/*                                     */
/*                                     BOX-TIAO INTERVENTION MODEL 4           */
/*                                     */
DATA EXAM57;
  START: T+1;
  PUL=(T=15);
  IF PUL1=. THEN PUL1=0;
  IF Y1=. THEN Y1=0;
  Y=0.8*Y1+2.0*PUL1;
  OUTPUT;
  Y1=Y;
  PUL1=PUL;
  IF T<40 THEN GO TO START;

RUN;

PROC PLOT;
  PLOT Y*T='*';
  TITLE 'BOX-TIAO INTERVENTION MODEL 4';
  TITLE3 'Y(t)=2.0B/(1-0.8B)P(t)';
RUN;

```

이 SAS프로그램을 수행한 결과가 그림 5.7이다. 그림 5.7은 간섭시점 15에서 순간파동입력이 있고 이 간섭이 시점 16에서부터 동태함수에 영향을 미친다. 즉, 시점 ($15+k$)에서 동태모형의 출력값은 $2(0.8)^k$ 이다.

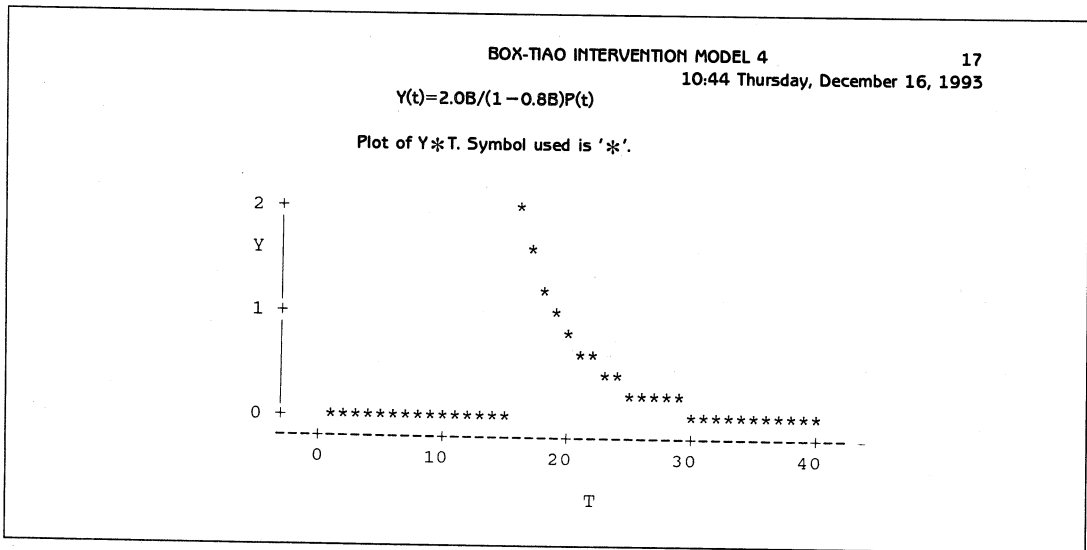


그림 5.7 동태모형 $f_t = \frac{2.0B}{1-0.8B} P_t^{(15)}$

【모형 5】 $f_t = \left(\frac{\omega_1}{1-\delta B} + \frac{\omega_2}{1-B} \right) B P_t^{(T)} \quad (0 < |\delta| \leq 1)$

어느 공업도시의 한 화학공장에서 큰 화재가 나서 그 도시의 공기 중 이황화탄소의 함유량이 급격히 높아졌다고 하자. 자연의 정화능력에 의해서 시간이 흐름에 따라 공기 중 이황화탄소의 함유량이 지수적으로 감소하나 이 화재로 인해서 자연생태계가 파괴되어 오랜 시간이 흐른 뒤에 이황화탄소의 함유량이 원래의 상태보다는 큰 값으로 수렴한다고 하자. 이러한 현상을 위와 같은 동태모형으로 나타낼 수 있다. 이 동태모형은 간섭시점 T 에서 순간과동입력이 일어나서, 동태모형의 출력값들이 시점 $(T+1)$ 에서 $(\omega_2 + \omega_1)$, 시점 $(T+2)$ 에서 $(\omega_2 + \delta\omega_1)$, 시점 $(T+3)$ 에서 $(\omega_2 + \delta^2\omega_1)$ 등 지수적으로 감소하여 오랜 시간이 흐른 뒤에는 ω_2 로 수렴한다. 이 동태모형을 다시 쓰면 다음과 같다.

$$f_t = 0 \quad (t=1, \dots, T-1, T)$$

$$(1-\delta B)(1-B) f_t = \begin{cases} \omega_1 + \omega_2 & (t=T+1) \\ -(\omega_1 + \delta\omega_2) & (t=T+2) \\ 0 & (t=T+3, T+4, \dots) \end{cases}$$

따라서, f_t 는 다음 차분방정식을 만족한다.

$$f_t = (\delta + 1)f_{t-1} - \delta f_{t-2} \quad (t=T+3, T+4, \dots)$$

이 차분방정식의 초기값들은 다음과 같다.

$$f_{T+1} = \omega_1 + \omega_2$$

$$f_{T+2} = \delta\omega_1 + \omega_2$$

$$f_{T+3} = \delta^2\omega_1 + \omega_2$$

차분방정식의 이론에 의해서 다음 식을 유도할 수 있다.

$$f_t = a_1 + a_2(t-T-1) + a_3\delta^{t-T-1} \quad (t=T+1, T+2, \dots)$$

여기서 a_1, a_2, a_3 는 초기값들에 의해 정해지는 상수들로서 다음 식을 만족한다.

$$f_{T+1} = a_1 + a_3 = \omega_1 + \omega_2$$

$$f_{T+2} = a_1 + a_2 + a_3\delta = \delta\omega_1 + \omega_2$$

$$f_{T+3} = a_1 + 2a_2 + a_3\delta^2 = \delta^2\omega_1 + \omega_2$$

즉, 다음 식들이 성립한다.

$$a_1 = \omega_2, \quad a_2 = 0, \quad a_3 = \omega_1$$

따라서, 이 동태모형을 다시 쓰면 다음과 같다.

$$f_t = \begin{cases} 0 & (t = 1, \dots, T) \\ \omega_2 + \omega_1 \delta^{t-(T+1)} & (t = T+1, \dots, n) \end{cases}$$

이 동태모형의 예를 살펴보기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM5-8.PRG에 수록되어 있다).

```

/* */
/*          EXAM5-8.PRG          */
/* */
/*          BOX-TIAO INTERVENTION MODEL 5          */
/* */
DATA EXAM58;
  START: T+1;
  PUL=(T=15);
  IF PUL1=. THEN PUL1=0;
  IF Y1=. THEN Y1=0;
  IF Y2=. THEN Y2=0;
  Y=1.8*Y1-0.8*Y2+5.0*PUL-4.4*PUL1;
  OUTPUT;
  Y2=Y1;
  Y1=Y;
  PUL1=PUL;
  IF T<40 THEN GO TO START;

RUN;
PROC PLOT;
  PLOT Y*T='*';

```

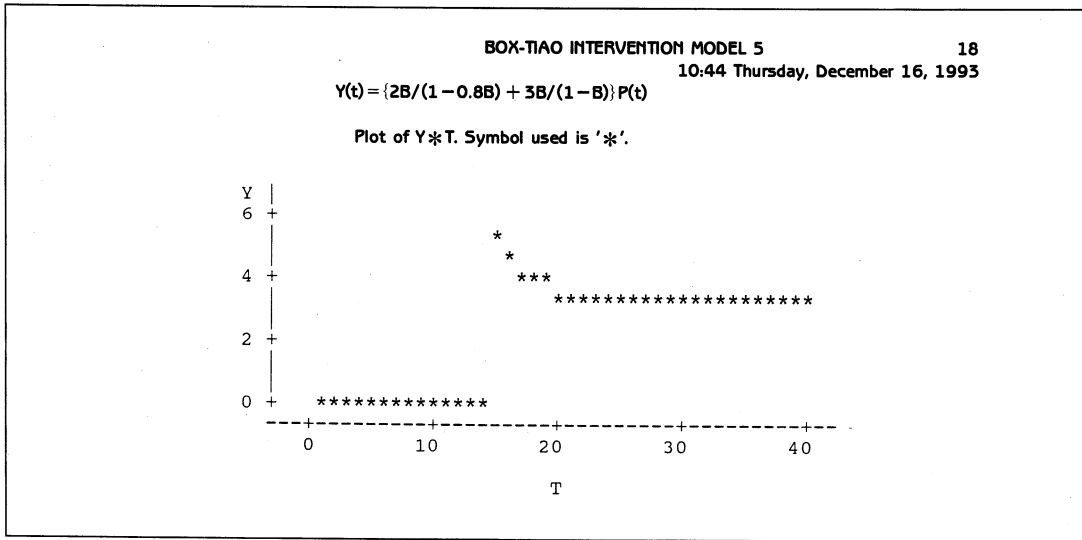


그림 5.8 동태모형 $f_t = \left(\frac{2}{1-0.8B} + \frac{3}{1-B} \right) BP_t^{(15)}$

```
TITLE 'BOX-TIAO INTERVENTION MODEL 5';
TITLE3 'Y(t)={2B/(1-0.8B) + 3B/(1-B)}P(t)';
RUN;
```

이 SAS프로그램을 수행한 결과가 그림 5.8이다. 그림 5.8은 간섭시점 15에서 순간과동입력이 있고 이 간섭이 시점 16에서부터 동태함수에 영향을 미친다. 즉, 시점 $(15+k)$ 에서 동태모형의 출력값은 $3+2(0.8)^{k-1}$ 이다.

【모형 6】 $f_t = \left(\omega_0 + \frac{\omega_1 B}{1-\delta B} + \frac{\omega_2 B}{1-B} \right) P_t^{(T)}$ ($0 < |\delta| \leq 1$)

어느 달에 전쟁이 난다는 소문이 돌아서 라면의 수요가 급증하였으나, 다음 달에 이 소문이 사실이 아닌 것으로 알려지면서 라면의 수요가 격감한 뒤, 시간이 흐르면서 다시 라면의 수요가 증가하는 현상을 모형화하기로 하자. 이러한 현상은 위와 같은 동태모형으로 나타낼 수 있다. 이 동태모형은 간섭시점 T 에서 순간과동입력이 일어나서, 동태모형의 출력값들이 간섭시점 T 에서 ω_0 로 급격히 변화하고, 시점 $(T+1)$ 에서 $(\omega_2 + \omega_1)$, 시점 $(T+2)$ 에서 $(\omega_2 + \delta\omega_1)$, 시점 $(T+3)$ 에서 $(\omega_2 + \delta^2\omega_1)$ 등 지수적으로 감소하여 오랜 시간이 흐른 뒤에는 ω_2 로 수렴한다. 이 동태모형을 다시 쓰면 다음과 같다.

$$f_t = 0 \quad (t=1, \dots, T-2, T-1)$$

$$f_T = \omega_0$$

$$(1-\delta B)(1-B) f_t = \begin{cases} -(1+\delta)\omega_0 + \omega_1 + \omega_2 & (t=T+1) \\ \delta\omega_0 - (\omega_1 + \delta\omega_2) & (t=T+2) \\ 0 & (t=T+3, T+4, \dots) \end{cases}$$

따라서, f_t 는 다음 차분방정식을 만족한다.

$$f_t = (\delta + 1)f_{t-1} - \delta f_{t-2} \quad (t=T+3, T+4, \dots)$$

이 차분방정식의 초기값들은 다음과 같다.

$$\begin{aligned} f_{T+1} &= \omega_1 + \omega_2 \\ f_{T+2} &= \delta\omega_1 + \omega_2 \\ f_{T+3} &= \delta^2\omega_1 + \omega_2 \end{aligned}$$

차분방정식의 이론에 의해서 이 식으로부터 다음 식을 유도할 수 있다.

$$f_t = a_1 + a_2(t-T-1) + a_3\delta^{t-T-1} \quad (t=T+1, T+2, \dots)$$

여기서 a_1, a_2, a_3 는 초기값들에 의해 정해지는 상수들로서 다음 식을 만족한다.

$$\begin{aligned} f_{T+1} &= a_1 + a_3 = \omega_1 + \omega_2 \\ f_{T+2} &= a_1 + a_2 + a_3\delta = \delta\omega_1 + \omega_2 \\ f_{T+3} &= a_1 + 2a_2 + a_3\delta^2 = \delta^2\omega_1 + \omega_2 \end{aligned}$$

즉, 다음 식들이 성립한다.

$$a_1 = \omega_2, \quad a_2 = 0, \quad a_3 = \omega_1$$

따라서, 이 동태모형을 다시 쓰면 다음과 같다.

$$f_t = \begin{cases} 0 & (t = 1, \dots, T-1) \\ \omega_0 & (t = T) \\ \omega_2 + \omega_1 \delta^{t-(T+1)} & (t = T+1, \dots, n) \end{cases}$$

이 동태모형의 예를 살펴보기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM5-9.PRG에 수록되어 있다).

```

/ * * /
/ *          EXAM5-9.PRG          * /
/ * * /
/ *          BOX-TIAO INTERVENTION MODEL 6          * /
/ * * /
DATA EXAM59;
  START: T+1;
  PUL=(T=15);
  IF PUL1=. THEN PUL1=0;
  IF PUL2=. THEN PUL2=0;
  IF Y1=. THEN Y1=0;
  IF Y2=. THEN Y2=0;

```

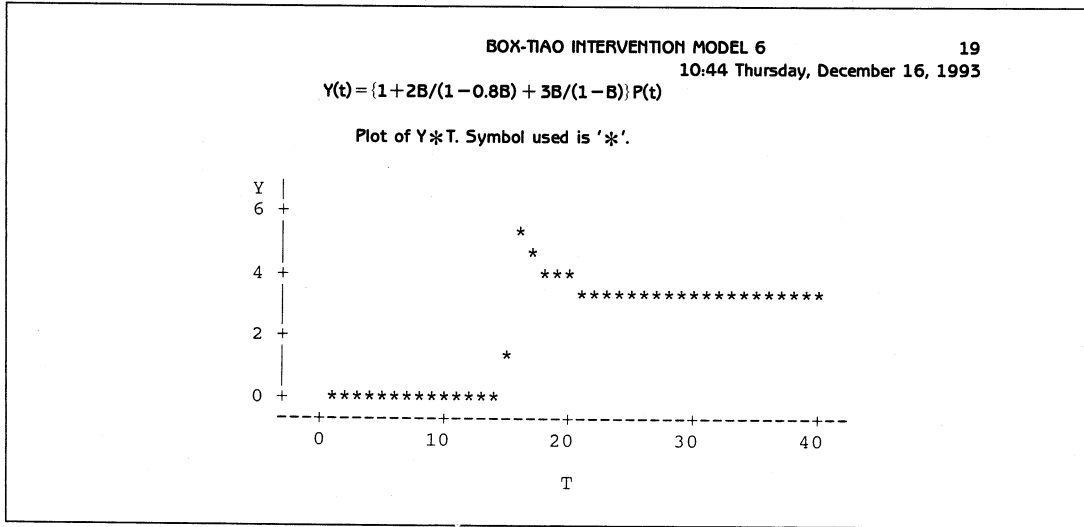


그림 5.9 동태모형 $f_t = \left(1 + \frac{2B}{1-0.8B} + \frac{3B}{1-B} \right) P_t^{(15)}$

```

Y=1.8*Y1-0.8*Y2+PUL+3.2*PUL1-3.6*PUL2;
OUTPUT;
Y2=Y1;
Y1=Y;
PUL2=PUL1;
PUL1=PUL;
IF T<40 THEN GO TO START;

RUN;
PROC PLOT;
PLOT Y*T='*';
TITLE 'BOX-TIAO INTERVENTION MODEL 6';
TITLE3 'Y(t)={1+2B/(1-0.8B) + 3B/(1-B)}P(t)';
RUN;
    
```

이 SAS프로그램을 수행한 결과가 그림 5.9이다. 그림 5.9는 간섭시점 15에서 순간파동입력이 있고 이 간섭이 시점 15에서부터 동태함수에 영향을 미친다. 즉, 동태모형의 출력값들이 시점 15에서 1이고, 시점 16에서 (3+2)이며, $k = 2, 3, \dots$ 에 대해서 시점 $(15+k)$ 에서 동태함모형의 출력값은 $3+2(0.8)^{k-1}$ 이다.

5.2.3 간섭시계열모형의 추정

간섭시계열모형을 추정하기 위해서는 제1단계로 간섭시점 이전의 시계열데이터를 이용해서 잡음항모형을 식별하고, 제2단계로 간섭시점 이전과 이후 모든 관찰점들로 구성된

전체 시계열데이터를 이용해서 잡음항모형과 동태모형을 동시에 추정한다.

간섭시점 이전의 시계열데이터를 이용해서 잡음항 $\{n_t\}$ 의 모형을 식별하기 위해 우선 차분과 계절주기가 s 인 계절차분을 적용해서 다음과 같은 정상시계열데이터 $\{w_t\}$ 를 만든다.

$$w_t = (1-B)^d (1-B^s)^D n_t$$

즉, 차분차수가 d 이고 계절주기가 s 인 계절차분차수가 D 인 비정상시계열데이터 $y_{-d-Ds+1}, y_{-d-Ds+2s}, \dots, y^0, y^1, \dots, y^n$ 에 대해서 다음 식이 성립한다.

$$w_t = (1-B)^d (1-B^s)^D \{y_t - f(\omega, \delta, I_t^{(T)})\} \quad (t = 1, 2, \dots, n)$$

이 정상시계열데이터 $\{w_t\}$ 는 다음과 같은 승법계절ARMA모형을 만족한다고 가정해 보자.

$$\phi(B) \Phi(B^s) w_t = \theta(B) \Theta(B^s) v_t$$

여기서 $\{v_t\}$ 는 분산이 σ_v^2 인 백색잡음과정이며 특성식들은 다음과 같다.

$$\phi(x) = 1 - \phi_1 x - \phi_2 x^2 - \dots - \phi_p x^p$$

$$\theta(x) = 1 - \theta_1 x - \theta_2 x^2 - \dots - \theta_q x^q$$

$$\Phi(x) = 1 - \Phi_1 x^1 - \Phi_2 x^2 - \dots - \Phi_p x^p$$

$$\Theta(x) = 1 - \Theta_1 x^1 - \Theta_2 x^2 - \dots - \Theta_q x^q$$

Box-Tiao의 동태모형들 중에서 처음 다섯 개의 모형들은 시점 $S=T$ 이전의 관찰점들에서 동태모형의 출력값들이 0이고, 여섯 번째 모형에서는 시점 $S=T-1$ 이전의 관찰점들에서 동태모형의 출력값들이 0이다. 따라서 다음 식이 성립한다.

$$w_t = (1-B)^d (1-B^s)^D y_t \quad (t = 1, 2, \dots, S)$$

이 관찰점들 $\{w_t | t = 1, 2, \dots, S\}$ 를 이용해서 정상시계열데이터 $\{w_t\}$ 를 식별한다. 즉, 승법계절ARMA모형의 차수들 p, q, P, Q 를 결정한다.

일단 잡음항모형이 식별되면 간섭시점 이전과 이후의 모든 관찰점들을 이용해서 잡음항모형과 동태모형을 다음과 같이 추정한다. 잡음항모형의 모수들과 동태함수에 포함된 계수들을 모수벡터 β 로 나타내면, 시계열데이터 $\{w_t | t = 1, 2, \dots, n\}$ 의 우도함수는 다음과 같다.

$$L(\beta, \sigma_v^2 | y) = (2\pi\sigma_v^2)^{-\frac{n}{2}} \{\det(M)\}^{\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma_v^2} S(\beta) \right\}$$

단, $\sigma_v^2 M^{-1}$ 은 벡터 $w = (w_1, \dots, w_n)'$ 의 공분산행렬

$$S(\beta) = w' M w = \sum_{t=-\infty}^n [v_t | y, \beta]^2$$

$[v_t | y, \beta]$ 는 β 와 y 가 주어졌을 때 v_t 의 조건부기대값

잡음항모형의 AR특성방정식과 동태함수의 분자방정식 $\delta(x) = 0$ 의 모든 근들이 단위원

밖에 있으며 또한 이 근들이 단위원 가까이에 있지 않을 때, 관찰점들의 개수 n 이 적당히 큰 수이면 우도함수는 지수부분에 의해 좌우된다. 즉, 지수부분의 오차제곱합 $S(\beta)$ 를 최소화하는 최소제곱추정값들은 최우추정값들과 근사하다. 시계열데이터 $\{w_t | t=1, 2, \dots, n\}$ 이 정상적이므로 적당히 큰 자연수 Q 에 대해서 만일 시점 t 가 $-Q$ 보다 작거나 같으면 오차항의 조건부기대값 $[v_t | y, \beta]$ 의 절대값은 무시해도 될 정도로 작다. 따라서 $S(\beta)$ 를 최소화하는 대신에 다음과 같은 β 와 y 가 주어졌을 때 유한개의 v_t 의 조건부기대값들의 제곱합을 최소화한다.

$$\sum_{t=-Q}^n [v_t | y, \beta]^2$$

Box & Jenkins(1970, 제7장)는 후향예측기법(back forecasting, 또는 backcasting)을 사용해서 조건부기대값들의 초기값들 $[v_0 | y, \beta], [v_{-1} | y, \beta] \dots, [v_{-Q} | y, \beta]$ 를 구하는 방법을 제시했다. 일반적으로 이 유한개의 v_t 의 조건부기대값들의 제곱합을 최소화하기 위해서는 Gauss-Marquadt 알고리즘을 사용한다.

5.2.4 PROC ARIMA를 사용한 간접시계열모형화

Woodfield(1987)는 일반적인 시계열데이터를 표현하는 데 사용할 수 있는 간접시계열 모형들을 SAS/ETS의 PROC ARIMA를 사용해서 추정하는 방법을 제시하였다. 이 소절에서는 이를 바탕으로 간접시계열모형들을 추정하는 방법을 살펴보자.

[예제 1]

시계열데이터의 간섭효과가 갑자기 나타나서 곧 사라지고, 잡음항이 ARIMA(1, 1, 2)과정인 다음과 같은 간접시계열모형을 살펴보자.

$$(1-B)\{y_t - \omega BP_t^{(T)}\} = \theta_0 + \frac{1 - \theta_1 B - \theta_2 B^2}{1 - \phi_1 B} v_t$$

이 모형을 추정하기 위해서는 다음과 같은 스테이트먼트들을 사용한다.

```
IDENTIFY VAR=Y(1) CROSS=( X(1) );
ESTIMATE P=(1) Q=(2) INPUT=(X) METHOD=ML;
```

이 스테이트먼트들의 앞에 있는 DATA스텝에서 순간충격입력 $P_{t-1}^{(T)}$ 를 변수 X로 정의해야 한다.

[예제 2]

시계열데이터에 간섭효과가 갑자기 나타나서 영구적으로 지속되며 잡음항이 ARIMA(2, 1, 2)과정인 다음과 같은 간섭시계열모형을 살펴보자.

$$(1-B)\{y_t - \omega BS_t^{(T)}\} = \theta_0 + \frac{1 - \theta_1 B - \theta_2 B^2}{1 - \phi_1 B - \phi_2 B^2} v_t$$

이 모형을 추정하기 위해서는 다음과 같은 스테이트먼트들을 사용한다.

IDENTIFY VAR=Y(1) CROSS=(X(1));
ESTIMATE P=2 Q=2 INPUT=(X) METHOD=ML;

이 스테이트먼트들의 앞에 있는 DATA스텝에서 계단입력 $S_{t-1}^{(T)}$ 를 변수 X로 정의해야 한다.

[예제 3]

간섭효과에 의해 시계열데이터가 서서히 증가(또는 감소)하나 시간이 흐름에 따라 안정된 값으로 수렴하며, 잡음항이 승법계절ARIMA(1, 1, 1) × (1, 1, 0)₁₂과정인 다음과 같은 간섭시계열모형을 살펴보자.

$$(1-B)(1-B^{12})\{y_t - \frac{\omega}{1-\delta_1 B} BS_t^{(T)}\} = \theta_0 + \frac{1-\theta_1 B}{(1-\phi_1 B)(1-\phi_{12} B^{12})} v_t$$

이 모형을 추정하기 위해서는 다음과 같은 스테이트먼트들을 사용한다.

I VAR=Y(1, 12) CROSS=(X(1, 12));
E P=(1, 12) Q=(1) INPUT=(/ (1) X) METHOD=ML;

이 스테이트먼트들의 앞에 있는 DATA스텝에서 계단입력 $S_{t-1}^{(T)}$ 를 변수 X로 정의해야 한다.

[예제 4]

간섭효과가 반영되어 시계열데이터가 갑자기 증가(또는 감소)하였다가 서서히 감소(또는 증가)하면서 시계열데이터가 시간이 흐름에 따라 간섭이 있기 이전의 값으로 수렴하며, 잡음항이 승법계절ARIMA(1, 2, 2) × (0, 2, 1)₁₂과정인 다음과 같은 간섭시계열모형을 살펴보자.

$$(1-B)^2(1-B^{12})^2 \left\{ y_t - \frac{\omega}{1-\phi_1 B} BS_t^{(T)} \right\} = \theta_0 + \frac{(1-\theta_1 B - \theta_2 B^2)(1-\phi_{12} B^{12})}{(1-\phi_1 B)} v_t$$

이 모형을 추정하기 위해서는 다음과 같은 스테이트먼트들을 사용한다.

```
I VAR=Y(1, 1, 12, 12) CROSS=( X(1, 1, 12, 12) );
E P=(1) Q=(1, 2)(12) INPUT=( / (1) X ) METHOD=ML;
```

이 스테이트먼트들의 앞에 있는 DATA스텝에서 순간충격입력 $P_{t-1}^{(T)}$ 를 변수 X로 정의해야 한다.

[예제 5] ████████████████████

간섭효과가 반영되어 시계열데이터가 시간이 흐름에 따라 산술적으로 계속 증가하며, 잡음항이 승법계절ARIMA(1, 0, 1) × (1, 0, 0)₁₂과정인 다음과 같은 간섭시계열모형을 살펴보자.

$$y_t - \frac{\omega}{1-B} BS_t^{(T)} = \theta_0 + \frac{1-\theta_1 B}{(1-\phi_1 B)(1-\phi_{12} B^{12})} v_t$$

이 모형을 추정하기 위해서는 우선 최소제곱추정법을 적용해서 다음 모형을 추정한다.

$$y_t = \theta_0 + \frac{\omega}{1-B} BS_t^{(T)} + n_t$$

여기서 $\{n_t\}$ 는 잡음항이다. 이 추정된 모형에서 잔차 $\{\hat{n}_t\}$ 를 계산한 뒤 이 잔차에 다음 승법계절ARIMA(1, 0, 1) × (1, 0, 0)₁₂모형을 적합시킨다.

$$\hat{n}_t = \frac{1-\theta_1 B}{(1-\phi_1 B)(1-\phi_{12} B^{12})} v_t$$

김동원(1993)은 좀더 다양한 형태의 동태모형들을 제시하고 이러한 동태모형들을 갖는 간섭시계열데이터들을 PROC ARIMA로 추정하는 SAS프로그램을 제공한다.

5.2.5 예 제 들

이 소절에서는 앞에서 살펴본 여러 간섭시계열모형들을 사용하여 실제 시계열데이터들을 분석하고자 한다. 모든 통계분석에서 그러하겠지만 간섭시계열분석에서는 시계열데이터의 그림, 즉 시계열산포도를 살펴보고 나서 다른 분석기법들을 적용하는 것이 좋다. 이

러한 이유로 이 소절에서 분석하는 시계열데이터들과 이들의 시계열산포도들을 제 2장에 수록하지 않고 이 소절에 수록한다.

첫 번째 예제는 [BD86] (pp. 165~171)에서 다룬 모의시계열데이터를 발생시키는 간섭시계열모형을 변형해서 간섭시계열데이터 TURNARND를 발생시킨 뒤 이를 분석한 것이다. 우선 간섭시계열데이터 TURNARND를 발생시키기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM5-10.PRG에 수록되어 있다).

```

/* */
/*          EXAM5-10.PRG          */
/* */
/*          MAKING DATASET TURNARND          */
/* */

OPTIONS LINESIZE=72 PAGESIZE=60;
LIBNAME TSA4DATA 'A:\TSA4DATA';
DATA TSA4DATA.TURNARND;
    ONE: T+1;
        XI=(T>80);
        IF T=1 THEN ERR1=1.4*RANNOR(7654321);
        ERR=0.7*ERR1+RANNOR(7654321);
        Y=15.0-10.0*XI+ERR;
        Y=FLOOR(100*Y+0.5)/100;
    OUTPUT;
    RETAIN;
    ERR1=ERR;
    IF T<120 THEN GO TO ONE;
RUN;
PROC GPLOT;
    PLOT Y*T=XI / HMINOR=0;;
    TITLE 'TURNAROUND TIMES';
    SYMBOL1 L=1 C=BLACK V=0 I=JOIN;
    SYMBOL2 L=1 C=BLACK V=1 I=JOIN;
RUN;
TITLE1:

```

이 SAS프로그램을 수행한 결과가 그림 5.10이다. 그림 5.10의 시계열산포도로부터 $\{y_t\}$ 가 전체 관찰점들의 수는 120개이고 간섭시점 T 가 80인 간섭시계열데이터임을 알 수 있다.

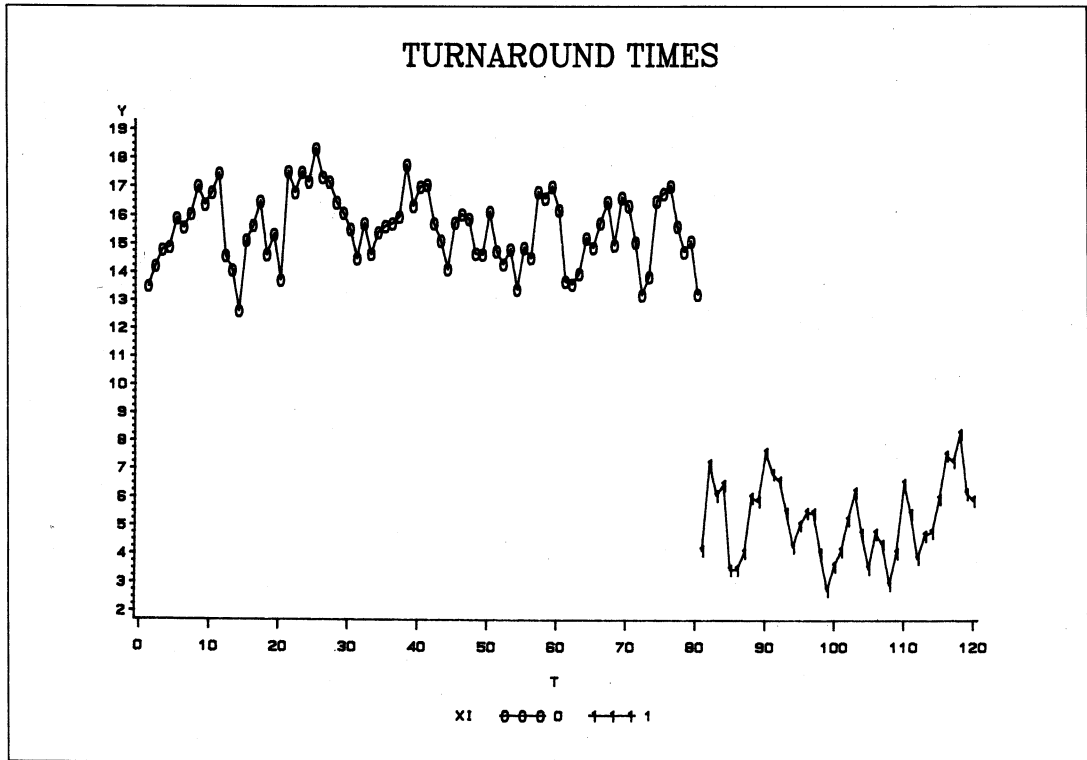


그림 5.10 시계열데이터 TURNARND의 시계열산포도

이 간접시계열데이터 TURNARND를 모형화하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM5-11.PRG에 수록되어 있다).

```

/ *                               */
/ *           EXAM5-11.PRG        */
/ *                               */
/ *           INTERVENTION ANALYSIS */
/ *           DATASET = TURNARND   */
/ *                               */

OPTIONS LINESIZE=72 PAGESIZE=60;
LIBNAME TSA4DATA 'A:\TSA4DATA';
DATA EXAM511;
    SET TSA4DATA.TURNARND;
    RUN;
TITLE 'TURNAROUND TIMES DATA';
PROC ARIMA DATA=EXAM511(OBS=80);
    I VAR=Y;
    RUN;
PROC ARIMA DATA=EXAM511;

```



```

I VAR=Y CROSS=(XI) NOPRINT;
E P=(1, 3) INPUT=(XI) PLOT;
RUN;
DATA EXAM511B;
SET EXAM511 END=EOF;
OUTPUT;
IF EOF THEN DO I=1 TO 10;
            Y=. ;
            T+1;
            XI=1;
            OUTPUT;
            END;
PROC ARIMA DATA=EXAM511B;
I VAR=Y CROSS=(XI) NOPRINT;
E P=(1, 3) INPUT=(XI) NOPRINT;
F OUT=FORE511 LEAD=10 ID=T;
RUN;
PROC PLOT DATA=FORE511(FIRSTOBS=91);
PLOT Y*T='Y' FORECAST*T='P' L95*T='L'
      U95*T='U' / OVERLAY VPOS=27;
RUN;
TITLE1;

```

이 시계열데이터에서 시점 80까지의 관찰점들은 간섭의 영향을 받지 않지만, 시점 81부터의 관찰점들은 간섭의 영향을 받아서 수준의 변화가 생겼다. 따라서 간섭시점인 80까지의 관찰점들을 사용해서 잡음항의 모형을 식별하기 위해 다음과 같은 첫 번째 PROC스텝을 수행한다.

```

PROC ARIMA DATA=EXAM511(OBS=80);
I VAR=Y;
RUN;

```

이 PROC스텝을 수행한 결과가 그림 5.11.1이다. 그림 5.11.1에서 간섭시점 80까지의 관찰점들로 구성된 시계열데이터의 표본자기상관함수는 시나브로형태를 보이고, 표본역자기상관함수와 표본부분자기상관함수는 시차 4 이후 절단형태를 보인다. 특히 시차 2에서 표본역자기상관계수와 표본부분자기상관계수가 유의적으로 0과 다르지 않다. 따라서 이 시계열데이터는 시차 2의 AR계수가 0인 부분AR(3)모형을 만족하는 것을 알 수 있다. 즉, 이 간섭시계열데이터의 잡음항모형은 다음과 같다.

$$n_t = \phi_1 n_{t-1} + \phi_3 n_{t-3} + v_t$$

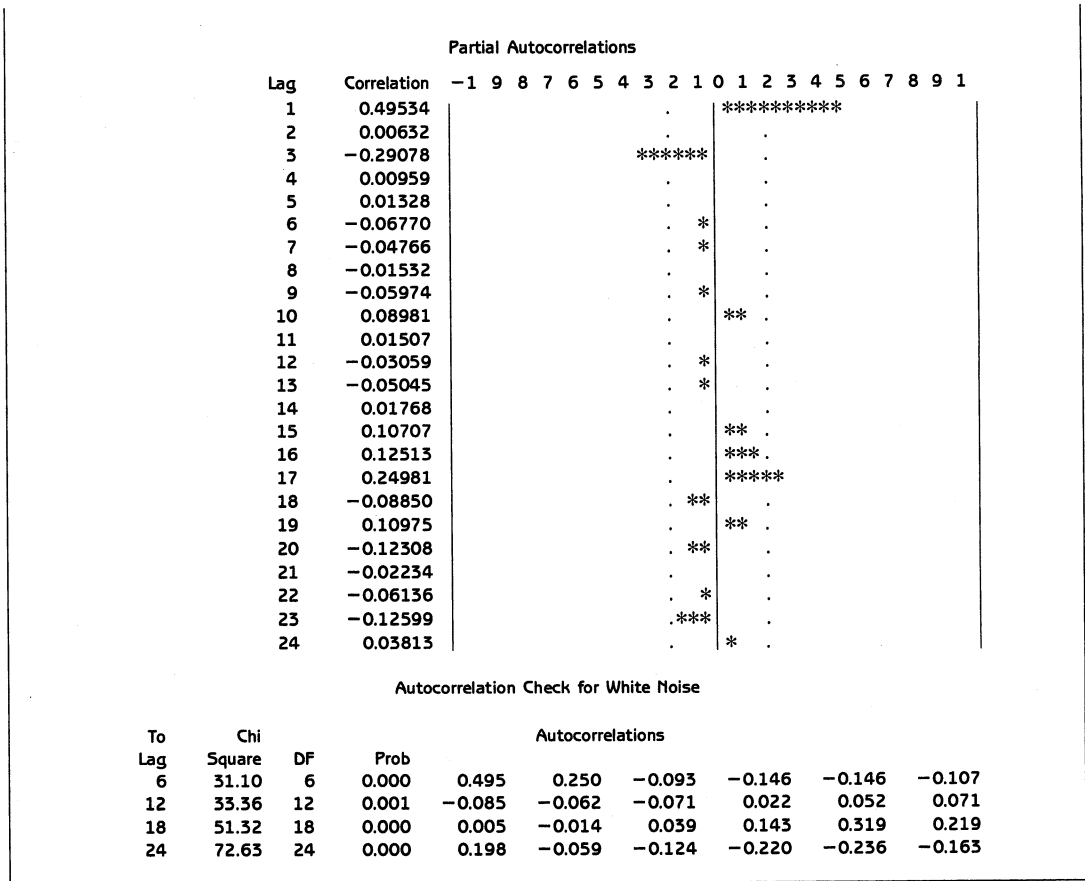


그림 5.11.1 시계열데이터 TURNARND의 잡음항모형의 식별

앞에서도 언급했듯이 이 간접시계열데이터를 모형화하기 위해서 이 잡음항모형을 추정할 필요는 없다. 제1단계에서는 단지 잡음항모형의 식별만이 필요할 뿐이다.

그림 5.10에서 시계열데이터 TURNARND에는 간접시점 80 다음에 계단변화가 있음을 알 수 있다. 따라서 이 시계열데이터는 다음과 같은 간접시계열모형을 갖는다고 할 수 있다.

$$y_t = \mu + \omega BS_t^{(80)} + \frac{1}{1 - \phi_1 B - \phi_3 B^3} v_t$$

이 간접시계열모형을 추정하기 위해서 다음과 같은 두 번째 PROC ARIMA를 수행한다.

```
PROC ARIMA DATA=EXAM511;
  I VAR=Y CROSS=(XI) NOPRINT;
  E P=(1, 3) INPUT=(XI) PLOT;
  RUN;
```

이 PROC ARIMA를 수행한 결과가 그림 5.11.2이다. 이 PROC ARIMA의 IDENTIFY스태이트먼트는 옵션 CROSSCOR=에 입력변수 XI를 지정하는 것 이외에 별다른 역

TURNAROUND TIMES DATA

25

21:46 Wednesday, December 22, 1993

ARIMA Procedure

(A)

Conditional Least Squares Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag	Variable	Shift
MU	15.55508	0.18365	84.70	0	Y	0
AR1,1	0.57485	0.07800	7.37	1	Y	0
AR1,2	-0.22516	0.08022	-2.81	3	Y	0
NUM1	-10.32085	0.31842	-32.41	0	XI	0

Constant Estimate = 10.1155687

Variance Estimate = 1.14702791

Std Error Estimate = 1.07099389

AIC = 360.937962*

SBC = 372.087929*

Number of Residuals = 120

* Does not include log determinant.

Correlations of the Estimates

Variable	Parameter	Y MU	Y AR1,1	Y AR1,2	XI NUM1
Y	MU	1.000	-0.045	-0.027	-0.575
Y	AR1,1	-0.045	1.000	-0.199	0.046
Y	AR1,2	-0.027	-0.199	1.000	0.073
XI	NUM1	-0.575	0.046	0.073	1.000

(B)

Autocorrelation Check of Residuals

To Lag	Chi Square	DF	Prob	Autocorrelations					
6	0.22	4	0.994	-0.011	0.034	-0.008	0.017	-0.012	0.000
12	3.23	10	0.976	-0.010	0.044	-0.095	-0.048	-0.093	-0.023
18	6.29	16	0.985	0.088	0.059	-0.001	-0.058	0.068	-0.052
24	11.70	22	0.963	0.111	-0.101	0.006	-0.057	-0.104	0.004

(C)

Autocorrelation Plot of Residuals

Lag	Covariance	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
0	1.147028	1.00000												*****										
1	-0.012542	-0.01093												.	*	.								
2	0.039068	0.03406												.	.	.								
3	-0.009594	-0.00836												.	.	.								
4	0.019941	0.01738												.	.	.								
5	-0.013457	-0.01173												.	.	.								
6	0.00018789	0.00016												.	.	.								
7	-0.011611	-0.01012												.	.	.								
8	0.050617	0.04413												.	.	.								
9	-0.108601	-0.09468												.	.	.								
10	-0.055582	-0.04846												.	.	.								
11	-0.106953	-0.09324												.	.	.								
12	-0.026408	-0.02302												.	.	.								
13	0.100827	0.08790												.	.	.								
14	0.068213	0.05947												.	.	.								
15	-0.0012640	-0.00110												.	.	.								
16	-0.066882	-0.05831												.	.	.								
17	0.077697	0.06774												.	.	.								
18	-0.059652	-0.05201												.	.	.								
19	0.126988	0.11071												.	.	.								
20	-0.116233	-0.10133												.	.	.								
21	0.0068664	0.00599												.	.	.								
22	-0.065440	-0.05705												.	.	.								
23	-0.119506	-0.10419												.	.	.								
24	0.0049212	0.00429												.	.	.								

"," marks two standard errors

Inverse Autocorrelations

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1		
1	-0.01090																							
2	-0.00665																							
3	-0.04391												*											
4	-0.03734												*											
5	0.03760												*											
6	0.00239																							
7	0.00261																							
8	-0.09298											**												
9	0.09116											**												
10	0.03250											**												
11	0.14482											**	*											
12	0.03147											*												
13	-0.08131										**													
14	-0.05670										*													
15	-0.00379																							
16	0.06336												*											
17	-0.05946										*													
18	0.04217												*											
19	-0.10844										**													
20	0.10188										**													
21	0.02362												*											
22	0.04986												*											
23	0.08783											**	*											
24	-0.04272										*													

Partial Autocorrelations

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1		
1	-0.01093																							
2	0.03394												*											
3	-0.00764																							
4	0.01608																							
5	-0.01086																							
6	-0.00126																							
7	-0.00913																							
8	0.04360												*											
9	-0.09315										**													
10	-0.05371										**													
11	-0.08831										**													
12	-0.02542										*													
13	0.09804												**											
14	0.06454												*											
15	-0.00261																							
16	-0.06927										*													
17	0.07075												*											
18	-0.05343										*													
19	0.11032												**											
20	-0.11611										**													
21	-0.02837										*													
22	-0.05316										*													
23	-0.09656										**													
24	0.04689										*													

(D)

Model for variable Y

Estimated Intercept = 15.5550836

Autoregressive Factors

Factor 1: $1 - 0.57485 B^{**}(1) + 0.22516 B^{**}(3)$

Input Number 1 is XI.

Overall Regression Factor = -10.3208

그림 5.11.2 간접시계열모형의 추정

할을 못 하므로 이 스테이트먼트를 수행한 결과를 화면에 출력시킬 필요가 없다. 따라서 IDENTIFY 스테이트먼트의 옵션 NOPRINT를 사용하였다. 앞서도 언급했듯이 간접시점 이전의 관찰점들 뿐 아니라 간접시점 이후의 관찰점들을 사용해서 간접시계열모형을 추정한다. 그림 5.11.2의 부분 A에서 이 PROC ARIMA의 ESTIMATE 스테이트먼트를 수행해서 조건부최소제곱추정한 간접시계열모형이 다음과 같음을 알 수 있다.

$$y_t = 15.555 - 10.321 BS_t^{(80)} + \frac{1}{1 - 0.575B - 0.225B^3} v_t$$

그림 5.11.2의 부분 B의 퍼트맨토우통계량값들로부터 이 간접시계열모형이 시계열데이터 TURNARND와 잘 부합됨을 알 수 있다. 또한, 부분 C의 잔차들의 표본자기상관함수, 표본역자기상관함수, 그리고 표본부분자기상관함수로부터 이 사실을 확인할 수 있다.

이 시계열데이터를 예측하기 위해서는 우선 입력변수를 외삽(extrapolation)하여야 한다. 이러한 목적으로 다음 DATA스텝을 수행한다.

```
DATA EXAM511B;
  SET EXAM511 END=EOF;
  OUTPUT;
  IF EOF THEN DO I=1 TO 10;
    Y=. ;
    T+1;
    XI=1;
    OUTPUT;
  END;
RUN;
```

이 PROC ARIMA를 수행하면 SAS데이터세트 EXAM511에 제121번 관찰점부터 10개의 관찰점들이 추가되어 새로운 SAS데이터세트 EXAM511B가 만들어진다. 이 추가된 관찰점들에서 변수 XI의 관찰값은 1이고 변수 Y의 관찰값은 결측(missing)이다.

새로운 SAS데이터세트 EXAM511B를 사용해서 시계열데이터 TURNARND의 미래값들을 예측하기 위해서는 다음 PROC ARIMA를 수행한다.

```
PROC ARIMA DATA=EXAM511B;
  I VAR=Y CROSS=(XI) NOPRINT;
  E P=(1, 3) INPUT=(XI) NOPRINT;
  F OUT=FORE511 LEAD=10 ID=T;
RUN;
```

이 PROC스텝을 수행한 결과가 그림 5.11.3이다. 그림 5.11.3의 부분 B에 시계열데이터 TURNARND의 예측(Forecast), 표준오차(Std Error), 신뢰수준 95%의 예측구간의 하한(Lower 95%)과 상한(Upper 95%)이 실려 있다.

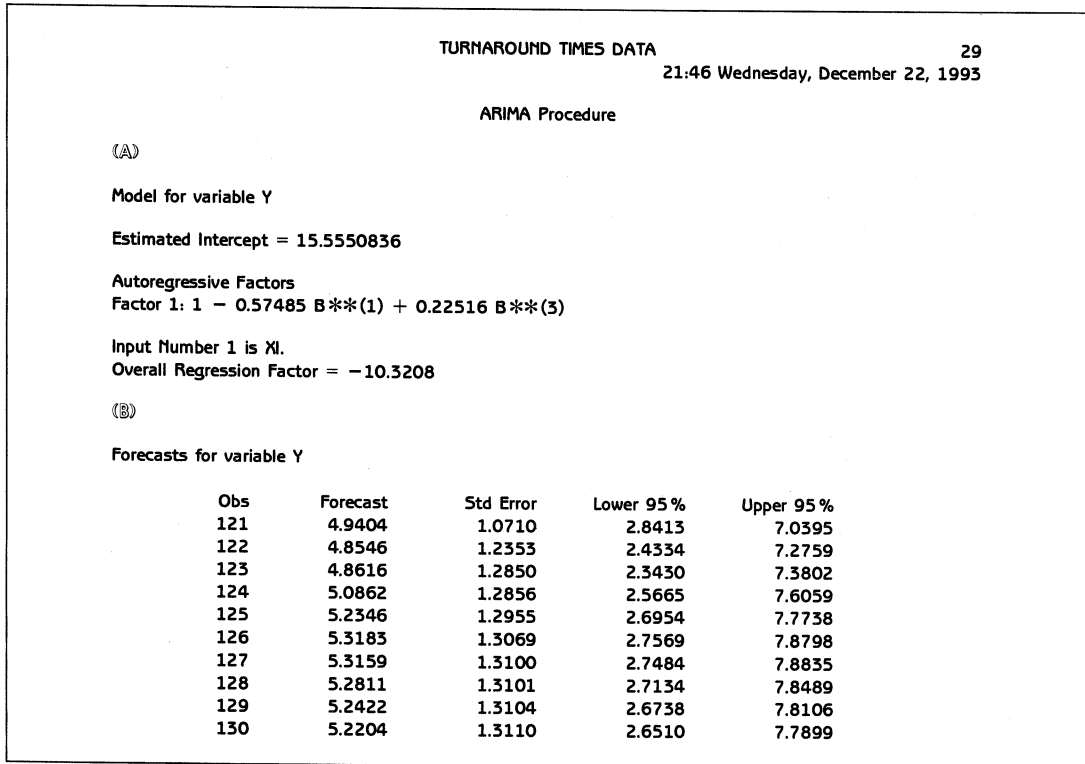


그림 5.11.3 시계열데이터 TURNARND의 예측

그림 5.11.3의 부분 B에 출력된 예측, 표준오차, 신뢰수준 95%의 예측구간의 하한과 상한의 시계열산포도를 그리기 위해서 다음 PROC PLOT를 수행해 보자.

```
PROC PLOT DATA=FORE511(FIRSTOBS=91);
  PLOT Y*T='Y' FORECAST*T='P' L95*T='L'
       U95*T='U' / OVERLAY VPOS=27;
RUN;
```

이 PROC PLOT를 수행한 결과가 그림 5.11.4이다.

두 번째 예제는 McSweeny(1978)가 제시한 1962년 1월부터 1978년 12월까지 미국 Ohio 주 Cincinnati시의 전화번호 안내원들이 하루에 받은 전화번호 안내요청 횟수의 월평균(the average daily directory assistance calls)을 기록한 시계열데이터세트 DIRCALLS이다. 이 시계열데이터는 McCleary & Hay(1980), Vandaele(1983), 그리고 Woodfield(1987)에 의해 분석되었고, 디스켓 SAS4TSA4의 SAS데이터라이브러리 TSA4DATA에 DIRCALLS.SSD로 수록되어 있다. 1974년 3월에 Cincinnati시 전화국은 전화번호 안내요청을 줄이기 위해서 20센트의 전화번호 안내요금을 받기로 하였다. 간접시계열분석을 사용해서 전화번호 안내요금을 받기로 한 이후에 전화번호 안내요청 건수의 월평균이 줄어들었는가를 분석하자. 이 시계열데이터는 1962년 1월부터 1978년 12월까지 기록되었으므로 관찰점들의 개수가 180이다.

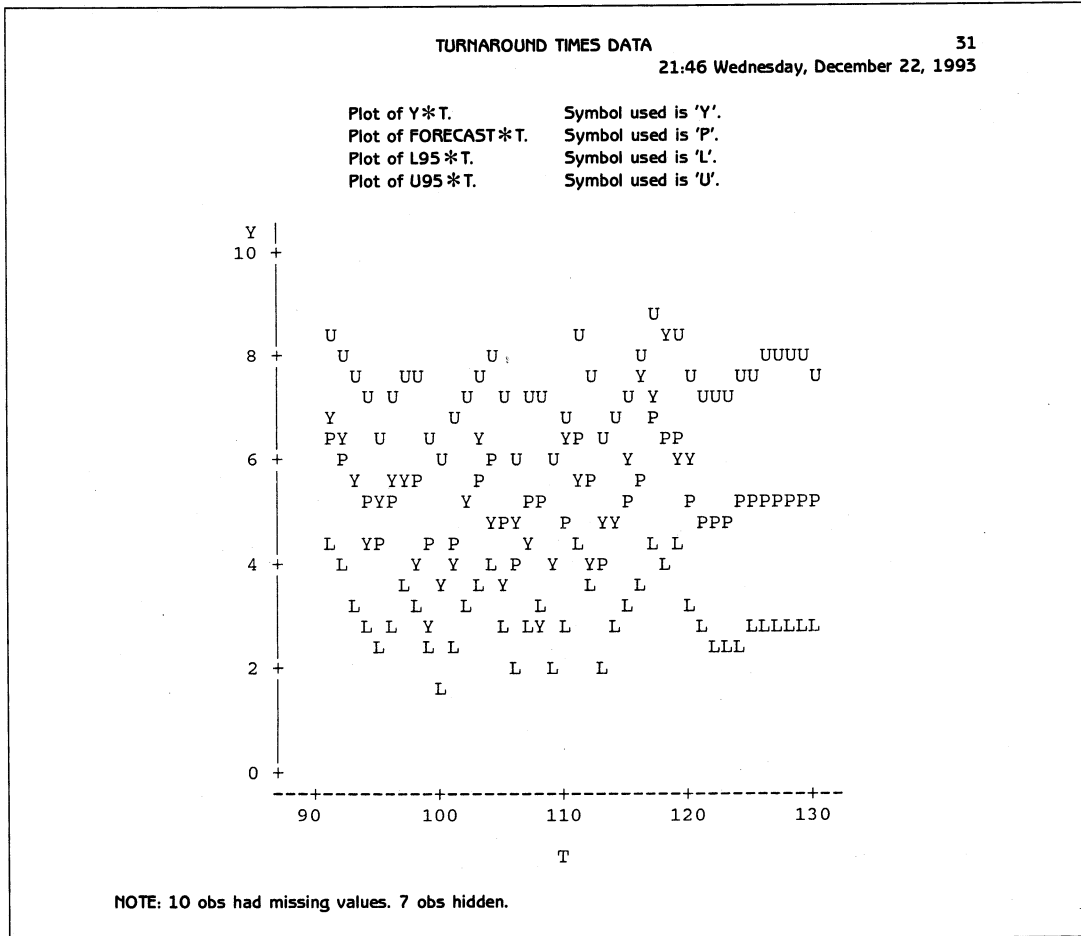


그림 5.11.4 예측된 시계열데이터 TURNARND의 산포도

시계열데이터 DIRCALLS의 시계열산포도를 그리기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM5-12.PRG에 수록되어 있다).

```

/* */
/*          EXAM5-12.PRG          */
/* */
/*          DIRECTORY ASSISTANCE CALLS AT CINCINATTI          */
/*          1962.1. - 1978.12.          */
/*          DATASET = DIRCALLS          */
/* */

OPTIONS LINESIZE=72 PAGESIZE=60;
LIBNAME TSA4DATA 'A:\TSA4DATA';
DATA TSA4DATA.DIRCALLS;
    INPUT Y @@;

```



```

Y=100*Y;
RETAIN NALJA '1DEC61'D;
NALJA=INTNX ('MONTH',NALJA,1);
FORMAT NALJA MONYY. ;
STP=(NALJA>='01MAR74'D);
LABEL Y='DAILY DIR CALLS';
DIF1Y = DIF1(Y);
DIF112Y = DIF12(DIF1Y);
CARDS:
350 339 351 364 369 331 331 340 346 341 357 398
381 367 383 375 353 361 375 371 373 366 382 429
406 403 429 425 427 409 402 409 419 404 429 463
428 449 444 467 474 463 432 453 462 456 474 514
489 475 492 525 527 533 527 522 526 513 564 599
572 587 599 601 611 620 579 582 592 581 630 663
638 631 645 682 601 595 521 521 516 496 538 575
537 534 542 538 547 540 526 548 555 545 594 643
625 616 640 625 637 634 621 641 654 649 662 699
672 704 700 711 715 718 652 664 695 704 733 772
716 712 732 755 761 748 748 750 744 731 782 810
777 816 840 868 872 811 810 762 634 626 649 697
657 549 162 177 175 162 161 165 170 172 178 186
178 178 189 205 202 185 193 200 196 204 206 227
225 217 219 236 253 213 205 210 216 218 235 241
;
TITLE1 'DIRECTORY ASSISTENTCE CALLS';
FOOTNOTE1 'n = 180, T = 01MAR74';
PROC GPLOT;
    PLOT Y*NALJA / HREF='01MAR74'D;
    SYMBOL C=BLACK V=NONE I=JOIN;
    RUN;
TITLE1;
FOOTNOTE1;

```

이 SAS프로그램을 수행한 결과가 그림 5.12이다. 그림 5.12에서 1962년 1월부터 1974년 2월까지의 간섭의 영향을 받지 않음을 알 수 있다. 즉, 관찰점들의 개수는 180개이고 간섭 시점 T 는 146이다.

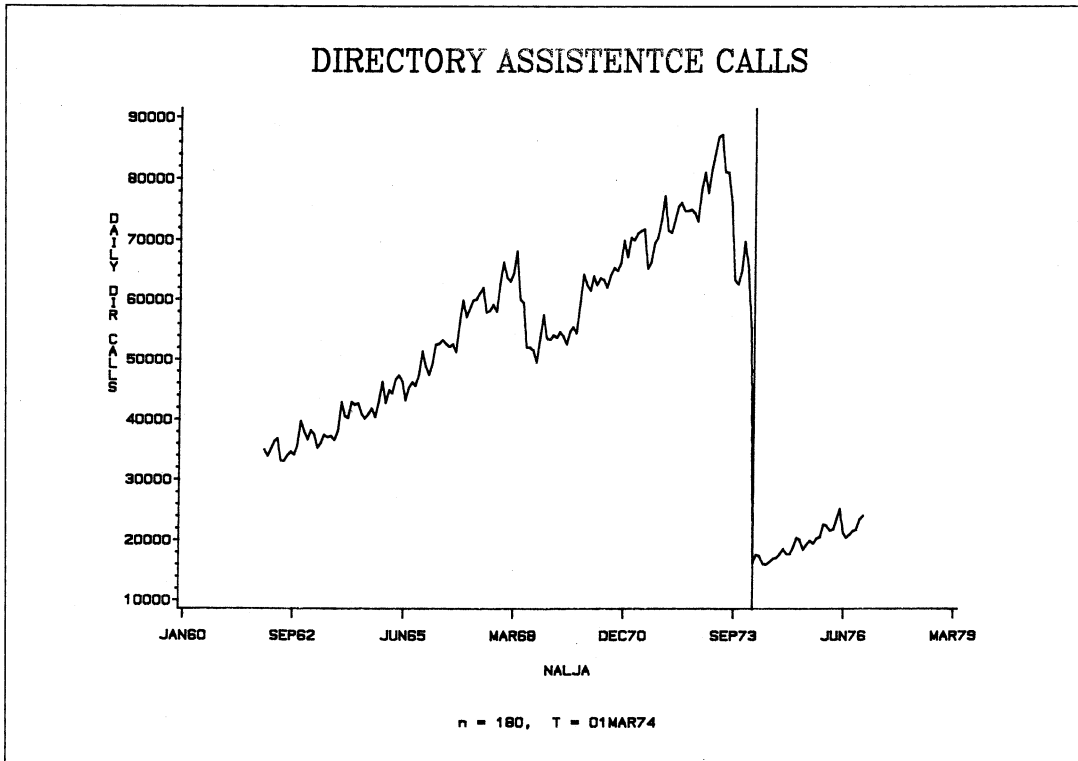


그림 5.12 시계열데이터 DIRCALLS의 시계열산포도

이 시계열데이터 DIRCALLS에 적합한 간접시계열모형을 구하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM5-13.PRG에 수록되어 있다).

```

/* */
/*          EXAM5-13.PRG          */
/* */
/*      INTERVENTION ANALYSIS      */
/*      DATASET = DIRCALLS         */
/* */
/*      INTERVENTION ANALYSIS OF TIME SERIES DIRCALLS */
/* */
OPTIONS LINESIZE=72 PAGESIZE=60;
LIBNAME TSA4DATA 'A:\TSA4DATA';
DATA EXAM513;
    SET TSA4DATA.DIRCALLS;
RUN;
PROC ARIMA DATA=EXAM513(OBS=146);
    I VAR=Y(1,12);

```

```

RUN;
PROC ARIMA DATA=EXAM513;
  I VAR=Y(1,12) CROSS=(STP(1,12)) NOPRINT;
  E Q=(12) INPUT=(STP) METHOD=ULS;
RUN;
PROC ARIMA DATA=EXAM513;
  I VAR=Y(1) CROSS=(STP(1)) NOPRIT;
  E Q=(12) INPUT=(STP) NOINT METHOD=ULS;
RUN;

```

그림 5.12에서 1962년 1월부터 1974년 2월까지의 간접의 영향을 받지 않기 때문에 이 관찰점들을 사용해서 잡음항의 모형을 식별하는 것이 타당하다. 간접시점 $T=146$ 이전까지의 시계열데이터를 식별하기 위해서 다음과 같은 첫 번째 PROC ARIMA를 수행한다.

```

PROC ARIMA DATA=EXAM513(OBS=145);
  I VAR=Y(1,12);
RUN;

```

그림 5.12에서 알 수 있듯이 시계열데이터 DIRCALLS는 간접시점 이전에는 증가하는 추세를 보인다. 이 시계열데이터의 표본자기상관함수, 표본역자기상관함수, 표본부분자기상관함수로부터 이 시계열데이터를 정상시계열데이터로 변환하기 위해서는 시차 1의 차분과 계절시차 12의 계절차분을 수행할 필요가 있음을 알 수 있다. 이 PROC ARIMA는 이렇게 차분된 시계열데이터 $\{\nabla_{12} \nabla y_t\}$ 의 잡음항모형을 식별하기 위한 것이다. 이 PROC ARIMA를 수행한 결과가 그림 5.13.1이다. 그림 5.13.1에서 차분된 시계열데이터 $\{\nabla_{12} \nabla y_t\}$ 의 시차 12에서의 표본자기상관계수, 표본역자기상관계수, 그리고 표본부분자기상관계수가 유의적으로 0과 다른 값들을 갖는다. 또한, 표본역자기상관함수와 표본부분자기상관함수의 절대값은 시차가 12, 24, ...으로 증가함에 따라 줄어든다. 따라서 차분된 시계열데이터 $\{\nabla_{12} \nabla y_t\}$ 는 계절시차가 12인 MA항이 있다고 할 수 있다. 즉, 간접시점 이전까지의 시계열데이터 $\{y_t\}$ 는 승법계절ARIMA(0, 1, 0) \times (0, 1, 1)₁₂모형에서 발생되었다고 할 수 있다. 따라서 시계열데이터 DIRCALLS의 간접시계열모형의 잡음항모형은 다음과 같다.

$$(1 - B)(1 - B^{12})n_t = (1 - \theta_{12} B^{12}) v_t$$

한 가지 유의할 점은 간접시점 이전까지의 시계열데이터가 승법계절ARIMA(0, 1, 0) \times (0, 1, 0)₁₂모형에서 발생되었다고 하였을 때, 잔차의 퍼트먼트우통계량값들로부터 간접시점 이전까지의 시계열데이터가 백색잡음성을 갖는다고 할 수 있다는 것이다.

그림 5.12의 시계열산포도에서 간접시점 1974년 2월 후에 시계열모형에 계단변화가 있었음을 알 수가 있다. 따라서 시계열데이터 DIRCALLS는 다음과 같은 간접시계열모형을 갖는다고 할 수 있다.

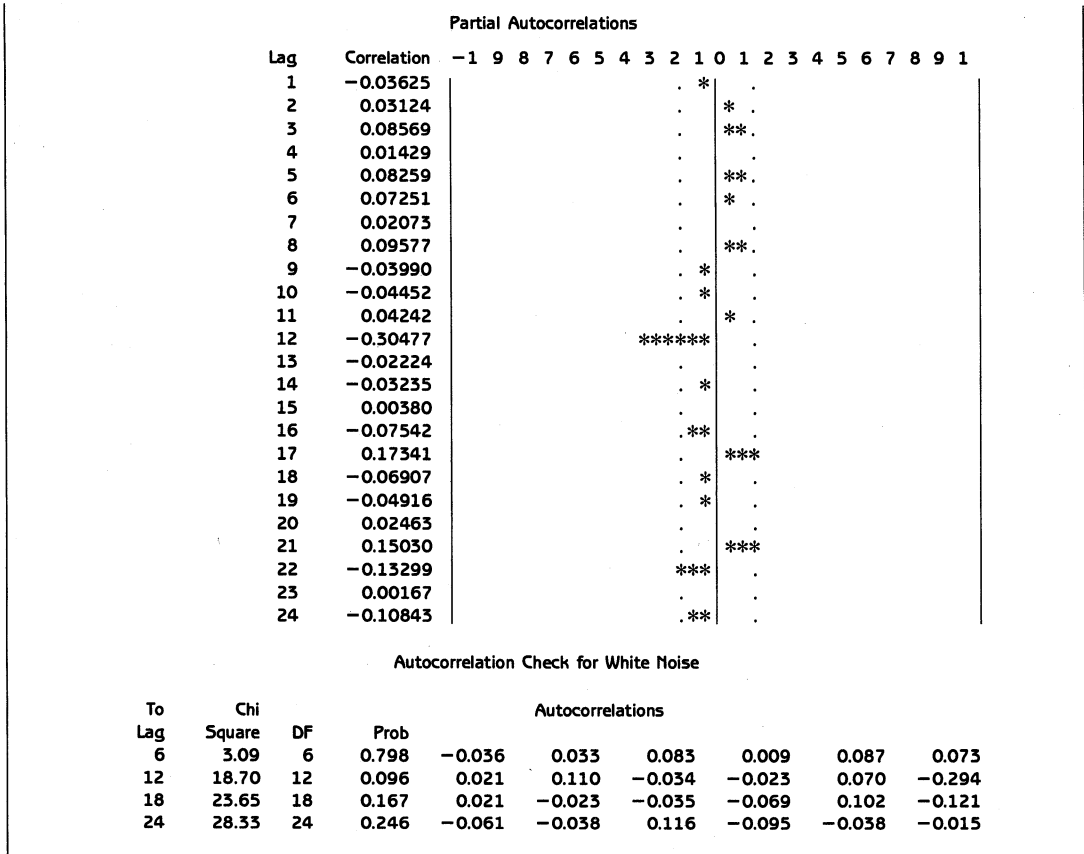


그림 5.13.1 시계열데이터 DIRCALLS의 잡음항모형의 식별

$$(1 - B)(1 - B^{12}) \{y_t - \omega BS_t^{(146)}\} = \mu + (1 - \theta_{12} B^{12}) v_t$$

이 간섭시계열모형을 추정하기 위해서 다음과 같은 PROC ARIMA를 수행한다.

```
PROC ARIMA DATA=EXAM513;
  I VAR=Y(1,12) CROSS=(STP(1,12)) NOPRINT;
  E Q=(12) INPUT=(STP) METHOD=ULS;
  RUN;
```

앞에서 언급했듯이 간섭시계열모형을 추정하기 위해서는 간섭시점 전후의 모든 관찰점들을 사용한다. 이 PROC ARIMA를 수행한 결과가 그림 5.13.2이다. 그림 5.13.2의 부분 A에서 알 수 있듯이 이 간섭시계열모형의 추정값들은 안정적이지 못하고 추정모형이 수렴하지 않을 수도 있다. 부분 B에서 추정모형이 다음과 같음을 알 수 있다.

$$(1 - B)(1 - B^{12}) \{y_t + 39698.0BS_t^{(146)}\} = -31.849 + (1 - 1.000 B^{12}) v_t$$

간섭시계열모형의 MA특성근이 단위원에 가까이 있으면 조건부최소제곱추정법보다는 비조건부최소제곱추정법이나 최우추정법을 사용하는 것이 좋다. 따라서 이 PROC ARI-

SAS

4

21:00 Thursday, December 23, 1993

ARIMA Procedure

(A)

WARNING: The model defined by the new estimates is unstable. The iteration process has been terminated.
 WARNING: Estimates may not have converged.

ARIMA Estimation Optimization Summary

Estimation Method: Unconditional Least Squares
 Parameters Estimated: 3
 Termination Criteria: Maximum Relative Change in Estimates
 Iteration Stopping Value: 0.001
 Criteria Value: 2.00014849
 Maximum Absolute Value of Gradient: 367634910
 R-Square (Relative Change in Regression SSE) from Last Iteration Step: 0.99999994
 Objective Function: Sum of Squared Residuals
 Objective Function Value: 735316712
 Marquardt's Lambda Coefficient: 0.00001
 Numerical Derivative Perturbation Delta: 0.001
 Iterations: 11
 Warning Message: Estimates may not have converged.

(B)

Unconditional Least Squares Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag	Variable	Shift
MU	-31.84941	36.97531	-0.86	0	Y	0
MA1,1	0.99998	0.15622	6.40	12	Y	0
NUM1	-39698.0	2200.7	-18.04	0	STP	0

Constant Estimate = -31.849408

Variance Estimate = 4483638.49

Std Error Estimate = 2117.46039

AIC = 3067.08522

SBC = 3076.4392

Number of Residuals = 167

Correlations of the Estimates

Variable	Parameter	Y	Y	STP
		MU	MA1,1	NUM1
Y	MU	1.000	0.020	-0.090
Y	MA1,1	0.020	1.000	0.002
STP	NUM1	-0.090	0.002	1.000

(C)

Autocorrelation Check of Residuals

To Lag	Chi Square	DF	Prob	Autocorrelations					
6	4.72	5	0.451	-0.007	-0.026	0.094	-0.056	0.104	0.061
12	8.85	11	0.636	-0.072	0.053	0.063	-0.063	-0.084	-0.014
18	12.76	17	0.752	-0.056	0.013	0.028	-0.124	-0.030	-0.027
24	19.95	23	0.645	-0.034	-0.061	0.096	-0.097	-0.112	-0.030
30	26.81	29	0.582	-0.034	0.097	0.004	-0.100	-0.063	0.096

(D)

Model for variable Y

Estimated Intercept = -31.849408

Period(s) of Differencing = 1, 12.


```
LIBNAME TSA4DATA 'A:\TSA4DATA';
```

```
DATA TSA4DATA.HONGKONG;
```

```
INPUT Y @@;
```

```
T=_N_;
```

```
XI=(T>259);
```

```
LABEL Y='DAILY STOCK INDEX';
```

```
CARDS;
```

```
1705.45 1667.36 1706.95 1708.90 1683.09 1707.57 1739.94 1734.36
1727.78 1745.72 1741.04 1747.20 1777.13 1781.13 1781.07 1779.37
1780.69 1781.62 1810.20 1777.92 1732.38 1732.94 1698.73 1698.73
1688.56 1664.81 1660.52 1699.39 1719.84 1681.65 1689.67 1718.15
1728.45 1683.82 1644.94 1688.15 1679.15 1694.75 1689.27 1706.54
1728.71 1705.73 1715.84 1717.27 1705.16 1689.78 1674.47 1672.84
1647.65 1611.57 1587.63 1597.54 1564.12 1518.90 1519.49 1515.88
1533.85 1554.92 1563.09 1502.79 1493.15 1431.68 1404.71 1331.01
1322.69 1351.01 1245.26 1218.20 1280.13 1259.16 1233.26 1113.77
1179.83 1266.52 1302.42 1335.42 1312.97 1289.24 1265.47 1249.47
1230.87 1227.72 1252.81 1236.12 1227.24 1233.06 1231.54 1270.28
1281.71 1370.90 1377.51 1380.16 1420.05 1395.92 1384.29 1435.47
1436.73 1413.43 1433.82 1443.06 1386.31 1397.80 1395.39 1401.75
1416.18 1428.68 1439.95 1451.07 1450.22 1431.10 1420.19 1399.82
1409.92 1423.82 1414.71 1403.86 1412.99 1413.19 1376.39 1381.18
1376.15 1376.15 1342.98 1334.75 1370.13 1353.11 1361.48 1375.59
1406.11 1412.78 1415.31 1405.82 1376.65 1386.78 1385.72 1409.40
1415.79 1436.57 1445.32 1423.57 1396.82 1410.00 1410.46 1386.04
1398.57 1397.08 1380.46 1405.23 1398.30 1417.42 1416.02 1390.15
1394.15 1368.38 1365.70 1322.39 1292.47 1296.40 1256.57 1270.04
1248.35 1230.62 1233.46 1257.94 1280.92 1277.01 1255.33 1267.12
1265.00 1271.60 1237.61 1231.77 1196.69 1140.58 1158.92 1125.36
1129.83 1152.69 1179.78 1196.90 1161.93 1185.13 1191.76 1188.29
1207.29 1223.19 1221.18 1223.38 1210.18 1188.18 1174.72 1167.16
1165.96 1174.30 1196.27 1200.12 1187.22 1206.86 1210.45 1203.81
1201.26 1182.48 1180.65 1193.77 1220.13 1235.84 1279.27 1306.08
1299.65 1303.09 1323.36 1320.20 1319.36 1316.36 1316.20 1351.09
1368.97 1377.55 1381.26 1387.35 1383.85 1367.53 1361.50 1334.05
1360.10 1358.37 1359.51 1353.93 1383.32 1413.47 1395.29 1392.81
1407.58 1404.27 1377.62 1371.30 1361.36 1337.48 1346.64 1327.77
1306.21 1336.98 1321.10 1306.27 1331.76 1319.08 1305.38 1266.61
1274.10 1274.55 1279.62 1268.75 1269.39 1278.56 1274.29 1266.71
```


1268.38	1271.58	1252.24	1251.86	1293.98	1294.85	1299.42	1313.37
1300.98	1295.75	1300.95	1307.74	1299.52	1283.56	1271.12	1254.47
1233.13	1185.86	1182.75	1147.28	1146.51	1142.02	1169.62	1089.87
1074.14	1066.67	966.36	994.34	937.28	978.66	1041.03	1035.55
1034.68	1067.65	1034.88	1012.00	1005.92	1048.63	1035.33	1039.03
1048.03	1076.26	1084.82	1056.29	1077.94	1105.44	1080.58	1080.37
1092.69	1090.26	1090.16	1088.28	1117.78	1113.54	1121.85	1096.12
1096.36	1012.62	988.66	986.36	927.18	862.06	816.23	881.96
865.88	888.11	883.86	864.34	857.74	867.25	894.56	880.46
878.89	886.65	873.70	833.15	832.68	753.76	774.91	803.60
771.90	771.85	822.54	864.55	851.92	859.92	851.65	843.77
851.79	875.86	860.07	824.04	831.80	810.79	787.87	820.15
801.70	792.57	789.92	773.33	747.93	732.32	704.03	710.34
676.30	698.62	743.96	752.21	751.93	771.34	754.40	777.04
803.84	802.89	772.86	757.73	754.12	753.16	748.26	758.14
764.65	780.90	776.22	783.18	783.82	765.96	761.61	769.97
772.50	798.86	845.15	851.08	859.89	870.32	866.34	874.49
889.91	906.54	908.56	871.99	879.60	888.45	880.74	884.72
899.82	887.04	895.53	894.81	888.08	895.45	891.01	896.84
906.64	915.51	931.40	944.32	974.22	990.35	983.82	990.52
992.81	1030.95	1066.30	1021.55	1032.59	1028.12	1002.74	985.54
1013.85	1014.84	1003.50	1023.81	1032.80	1032.37	1024.51	1024.07
1007.26	991.32	955.11	981.45	978.98	997.85	983.33	975.90
979.56	982.54	996.01	998.48	1014.16	1034.17	1041.86	1052.40
1056.44	1067.36	1067.14	1040.97	1043.18	1010.37	1001.48	1006.53
1041.07	1028.29	1033.76	1012.61	1019.43	997.43	1003.15	986.64
987.29	986.71	967.62	945.32	943.02	952.98	949.56	942.80
930.70	933.46	928.22	942.05	902.71	877.68	887.86	908.45
903.90	903.18	918.59	923.15	919.34	915.60	902.89	881.67
867.04	863.10	886.52	908.39	954.87	966.89	967.79	958.62
957.62	936.84	951.22	956.64	943.63	941.10	964.35	983.72
1035.97	1021.90	1016.49	1033.30	1036.42	1008.45	1027.11	1026.55
1064.73	1077.50	1071.51	1088.38	1095.22	1102.64	1086.90	1081.77
1097.96	1090.65	1074.29	1072.02	1052.84	1056.95	1059.11	1022.98
1028.25	1028.00	1020.53	1029.74	1040.58	1037.46	1047.24	1037.72
1025.13	1000.90	966.28	1000.23	993.14	982.18	981.91	981.80
971.08	965.94	955.24	950.54	916.11	842.25	836.59	849.06

TITLE1 'INDEX OF HONGKONG STOCK PRICES':

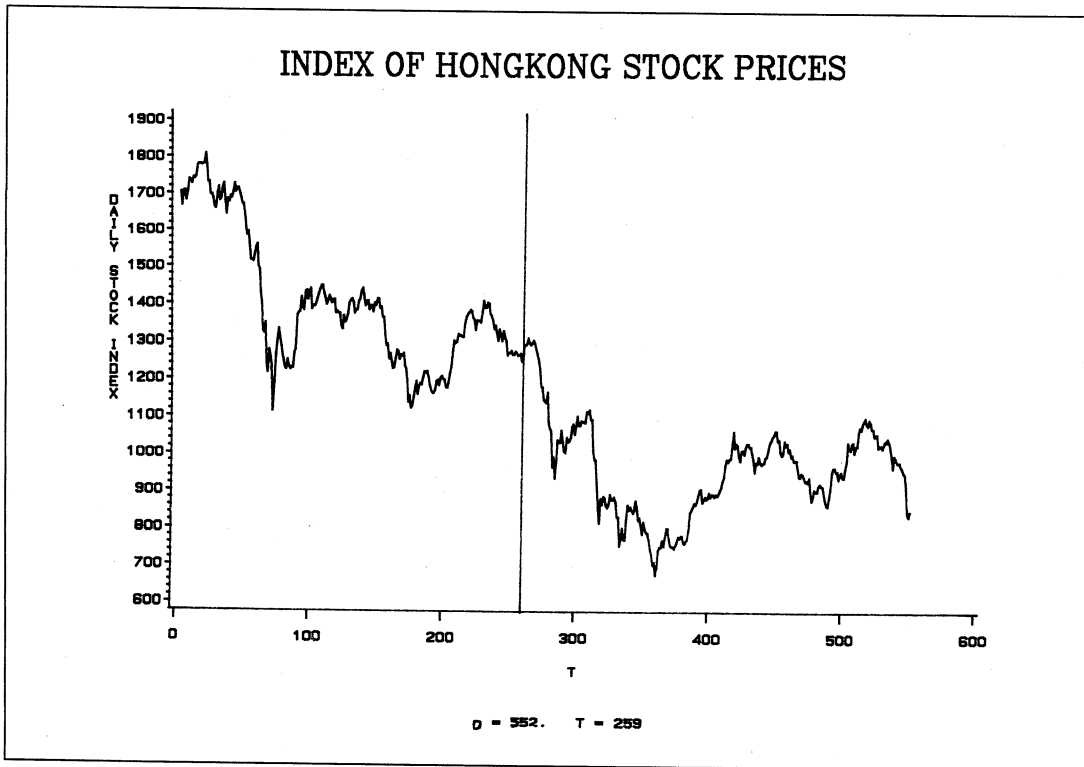


그림 5.14 시계열데이터 HONGKONG의 시계열산포도

```

FOOTNOTE1 'n = 552, T = 259';
PROC GPLOT;
    PLOT Y*T / HREF=260;
    SYMBOL C=BLACK V=NONE I=JOIN;
    RUN;
TITLE1;
FOOTNOTE1;
    
```

이 SAS프로그램을 수행한 결과가 그림 5.14이다. 그림 5.14에서 관찰점들의 개수는 552개이고 간섭시점 T 는 259임을 알 수 있다.

시계열데이터 HONGKONG에 적합한 간섭시계열모형을 구하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM5-15.PRG에 수록되어 있다).

```

/*                                     */
/*           EXAM5-15.PRG              */
/*                                     */
/*           INTERVENTION ANALYSIS     */
/*           DATASET = HONGKONG        */
/*                                     */
    
```

```

OPTIONS LINESIZE=72 PAGESIZE=60;
LIBNAME TSA4DATA 'A:\TSA4DATA';
DATA EXAM515;
    SET TSA4DATA.HONGKONG;
    RUN;
PROC ARIMA DATA=EXAM515(OBS=259);
    I VAR=Y(1);
    RUN;
PROC ARIMA DATA=EXAM515;
    I VAR=Y(1) CROSS=(XI(1)) NOPRINT;
    E INPUT=(XI) METHOD=ULS NOINT;
    RUN;
PROC ARIMA DATA=EXAM515;
    I VAR=Y(1) CROSS=(XI(1)) NOPRINT;
    E P=(6) INPUT=(XI) METHOD=ULS NOINT;
    RUN;

```

그림 5.14에서 첫 번째부터 제259번째까지의 관찰점들은 간섭의 영향을 받지 않기 때문에 이 관찰점들을 이용해서 잡음항의 모형을 식별하는 것이 타당하다. 간섭시점 $T=259$ 이전까지의 시계열데이터를 식별하기 위해서 다음과 같은 첫 번째 PROC ARIMA를 수행한다.

```

PROC ARIMA DATA=EXAM515(OBS=259);
    I VAR=Y(1);
    RUN;

```

그림 5.14에서 알 수 있듯이 시계열데이터 HONGKONG은 장기적으로 감소하는 추세를 갖는다. 시계열데이터 HONGKONG의 표본자기상관함수, 표본역자기상관함수, 표본부분자기상관함수로부터 이 시계열데이터를 정상시계열데이터로 변환하기 위해서는 시차 1의 차분을 수행할 필요가 있음을 알 수 있다. 이 PROC ARIMA는 이러한 차분된 시계열데이터 $\{\nabla y_t\}$ 를 식별하기 위한 것이다. 이 PROC ARIMA를 수행한 결과가 그림 5.15.1이다. 그림 5.15.1에서 알 수 있듯이 차분된 시계열데이터 $\{\nabla y_t\}$ 의 표본자기상관함수, 표본역자기상관함수, 그리고 표본부분자기상관함수는 백색잡음과정으로부터 나온 것들이라고 할 수 있다. 따라서 간섭시점 이전까지의 시계열데이터 $\{y_t\}$ 는 ARIMA(0, 1, 0)모형에서 발생되었다고 할 수 있다. 따라서 시계열데이터 HONGKONG의 간섭시계열모형의 잡음항모형은 다음과 같다.

$$(1-B)n_t = v_t$$

그림 5.14의 시계열산포도에서 간섭시점 $T=259$ 후에 시계열모형에 계단변화가 있었음

Partial Autocorrelations																							
Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
1	0.07678												.	**									
2	-0.07616										**	.											
3	0.12413									.		.	**										
4	0.01023									.		.	.										
5	0.06461									.		.	*										
6	-0.17436									***	.	.	.										
7	0.08724									.		.	**										
8	-0.04870									.	*	.	.										
9	0.06960									.	.	.	*										
10	0.01440																		
11	-0.08047									**	.	.	.										
12	0.04470									.	.	.	*										
13	-0.01728																		
14	0.11023									.	.	.	**										
15	0.01580																		
16	0.09881									.	.	.	**										
17	0.06415									.	.	.	*										
18	0.02592									.	.	.	*										
19	-0.02520									.	*	.	.										
20	-0.04004									.	*	.	.										
21	-0.00169																		
22	0.03519									.	.	.	*										
23	-0.06985									.	*	.	.										
24	-0.10867									**	.	.	.										

Autocorrelation Check for White Noise									
To Lag	Chi Square	DF	Prob	Autocorrelations					
6	12.93	6	0.044	0.077	-0.070	0.111	0.036	0.049	-0.149
12	17.78	12	0.123	0.047	0.011	0.009	0.044	-0.100	0.059
18	26.23	18	0.095	0.001	0.076	0.047	0.056	0.139	0.010
24	33.00	24	0.104	0.005	-0.040	-0.002	0.047	-0.099	-0.101

그림 5.15.1 시계열데이터 HONGKONG의 잡음항모형의 식별

을 알 수가 있다. 따라서 시계열데이터 HONGKONG은 다음과 같은 간접시계열모형을 갖는다고 할 수 있다.

$$(1-B)\{y_t - \omega BS_t^{(259)}\} = v_t$$

이 간접시계열모형을 추정하기 위해서 다음과 같은 두 번째 PROC ARIMA를 수행한다.

```
PROC ARIMA DATA=EXAM515;
  I VAR=Y(1) CROSS=(X1(1)) NOPRINT;
  E INPUT=(X1) METHOD=ULS NOINT;
  RUN;
```

이 PROC ARIMA를 수행한 결과가 그림 5.15.2이다. 그림 5.15.2의 부분 A에서 추정된 모형이 다음과 같음을 알 수 있다.

$$(1-B)\{y_t + 0.380 BS_t^{(259)}\} = v_t$$

부분 B의 퍼트랜트우통계량값들로부터 추정된 간접시계열모형의 잔차가 백색잡음과정에 가까움을 알 수 있다. 따라서 추정된 모형이 주어진 시계열데이터에 적합하다고 할 수 있다.

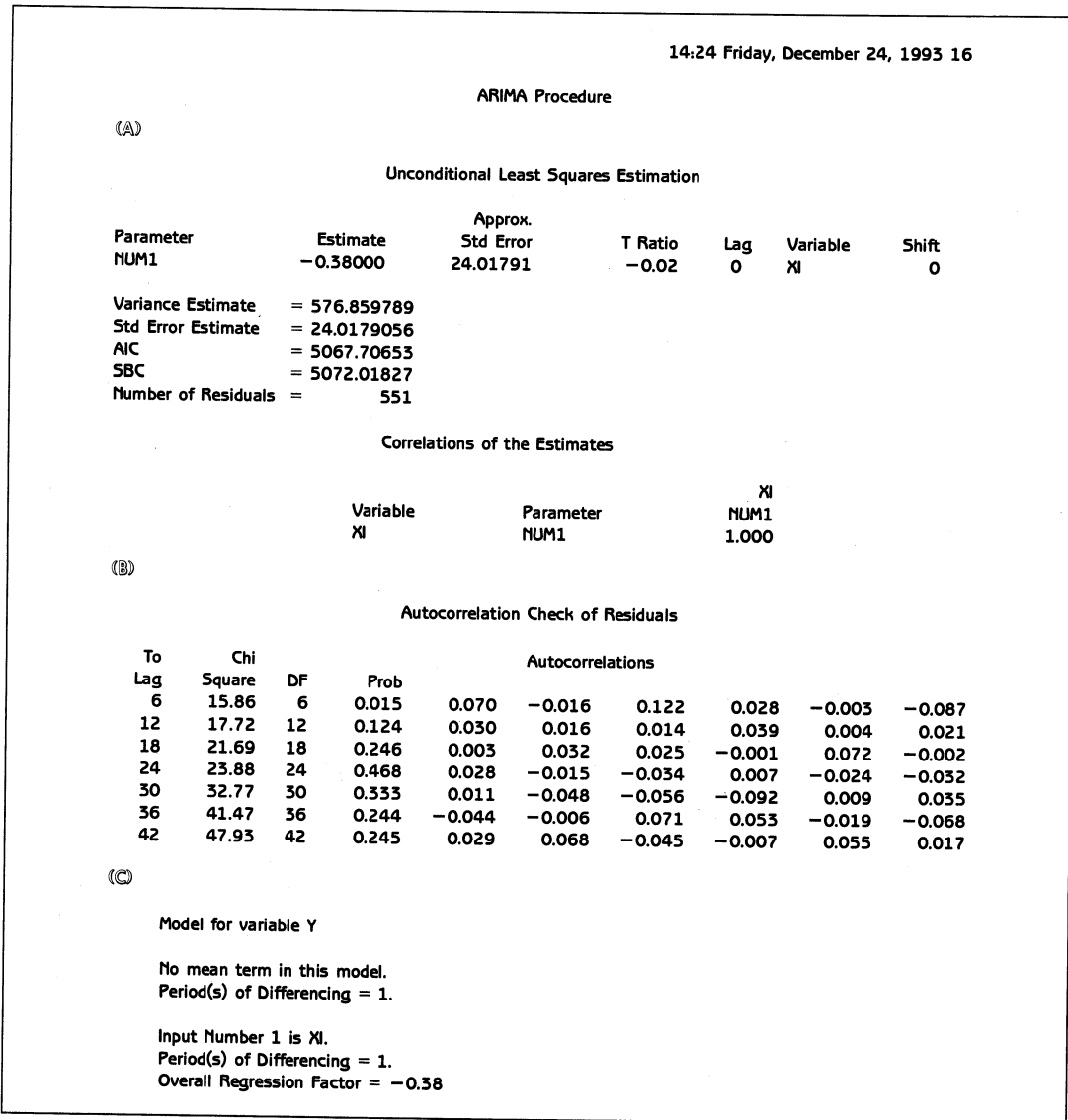


그림 5.15.2 시계열데이터 HONGKONG의 간접시계열모형추정

그림 5.15.1에서 차분된 시계열데이터 $\{V y_t\}$ 의 시차 6에서의 표본역자기상관계수와 표본 부분자기상관계수는 유의적으로 0과 다르다. 따라서 시계열데이터 HONGKONG의 간접 시계열모형의 잡음항모형은 다음과 같다고도 할 수 있다.

$$(1-B)n_t = \frac{1}{1-\phi_6 B^6} v_t$$

즉, 시계열데이터 HONGKONG은 다음과 같은 간접시계열모형을 갖는다고 할 수 있다.

$$(1-B)\{y_t - \omega BS_t^{(259)}\} = \frac{1}{1-\phi_6 B^6} v_t$$

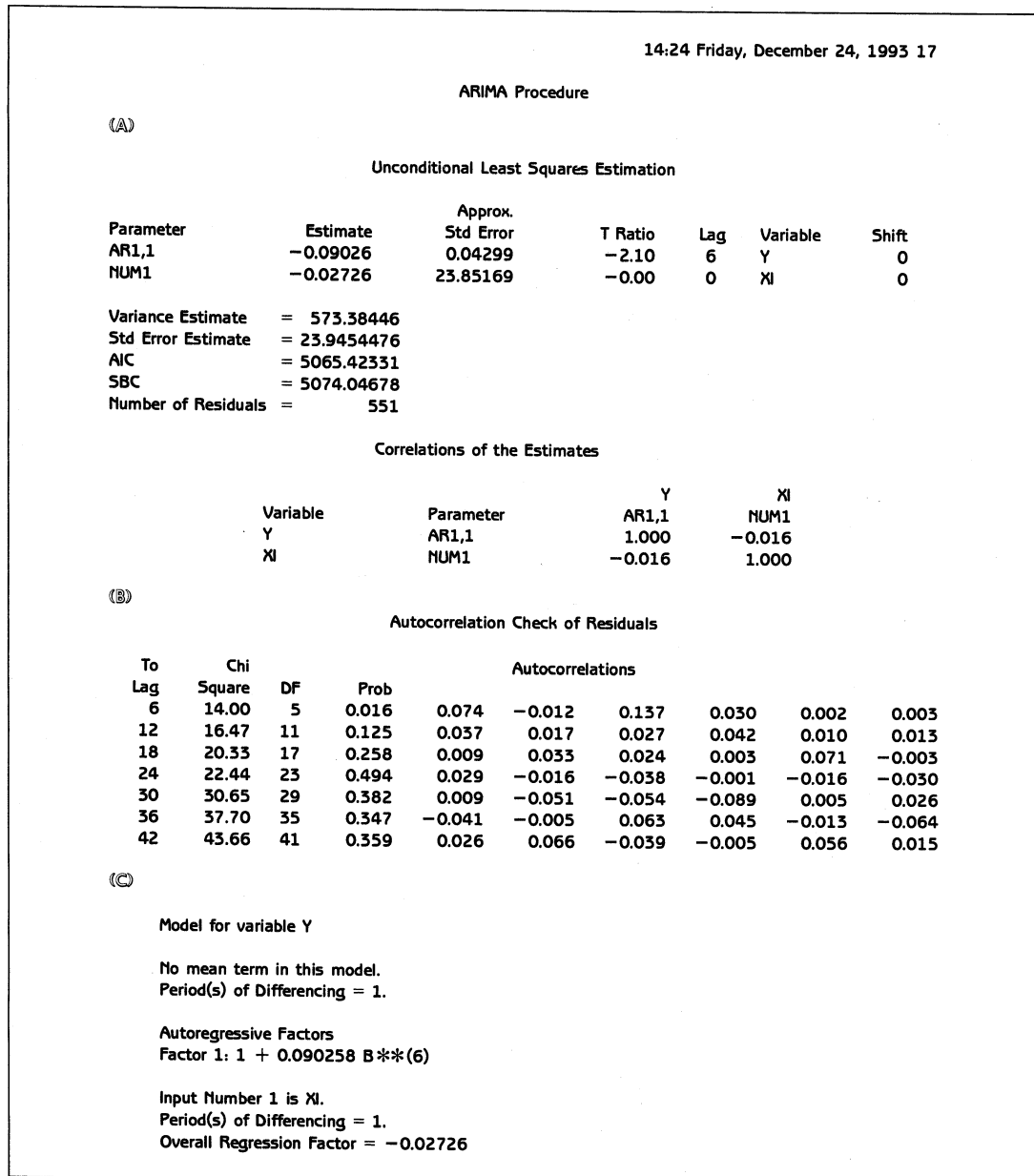


그림 5.15.3 시계열데이터 HONGKONG의 간접시계열모형추정

이 간접시계열모형을 추정하기 위해서 다음과 같은 PROC ARIMA를 수행한다.

```
PROC ARIMA DATA=EXAM515;
  I VAR=Y(1) CROSS=(XI(1)) NOPRINT;
  E P=(6) INPUT=(XI) METHOD=ULS NOINT;
  RUN;
```

이 PROC ARIMA를 수행한 결과가 그림 5.15.3이다. 그림 5.15.3의 부분 A에서 추정된 모형이 다음과 같음을 알 수 있다.


```

YEAR=55+FLOOR(( _N_ -1)/12);
MONTH=MOD( _N_ -1, 12)+1;
DATE=MDY(MONTH,1,YEAR);
X1=YEAR>=60;
SUMMER=(5<MONTH<11)*(YEAR>65);
WINTER=(YEAR>65)-SUMMER;
CARDS:
2.7 2.0 3.6 5.0 6.5 6.1 5.9 5.0 6.4 7.4 8.2 3.9
4.1 4.5 5.5 3.8 4.8 5.6 6.3 5.9 8.7 5.3 5.7 5.7
3.0 3.4 4.9 4.5 4.0 5.7 6.3 7.1 8.0 5.2 5.0 4.7
3.7 3.1 2.5 4.0 4.1 4.6 4.4 4.2 5.1 4.6 4.4 4.0
2.9 2.4 4.7 5.1 4.0 7.5 7.7 6.3 5.3 5.7 4.8 2.7
1.7 2.0 3.4 4.0 4.3 5.0 5.5 5.0 5.4 3.8 2.4 2.0
2.2 2.5 2.6 3.3 2.9 4.3 4.2 4.2 3.9 3.9 2.5 2.2
2.4 1.9 2.1 4.5 3.3 3.4 4.1 5.7 4.8 5.0 2.8 2.9
1.7 3.2 2.7 3.0 3.4 3.8 5.0 4.8 4.9 3.5 2.5 2.4
1.6 2.3 2.5 3.1 3.5 4.5 5.7 5.0 4.6 4.8 2.1 1.4
2.1 2.9 2.7 4.2 3.9 4.1 4.6 5.8 4.4 6.1 3.5 1.9
1.8 1.9 3.7 4.4 3.8 5.6 5.7 5.1 5.6 4.8 2.5 1.5
1.8 2.5 2.6 1.8 3.7 3.7 4.9 5.1 3.7 5.4 3.0 1.8
2.1 2.6 2.8 3.2 3.5 3.5 4.9 4.2 4.7 3.7 3.2 1.8
2.0 1.7 2.8 3.2 4.4 3.4 3.9 5.5 3.8 3.2 2.3 2.2
1.3 2.3 2.7 3.3 3.7 3.0 3.8 4.7 4.6 2.9 1.7 1.3
1.8 2.0 2.2 3.0 2.4 3.5 3.5 3.3 2.7 2.5 1.6 1.2
1.5 2.0 3.1 3.0 3.5 3.4 4.0 3.8 3.1 2.1 1.6 1.3
. . . . .
;
TITLE1 'OZONE IN LA AIR';
FOOTNOTE1 'n = 216, T1 = 61(60.1.), T2 = 133(66.1.)';
PROC GPLOT;
    PLOT OZONE*DATE /HREF='01JAN60'D HREF='01JAN66'D;
    SYMBOL C=BLACK V=NONE I=JOIN;
RUN;
TITLE1;
FOOTNOTE1;

```

이 SAS프로그램을 수행한 결과가 그림 5.16이다. 1960년 초 로스엔젤레스 중심가의 공기 중 오존함유량을 줄일 수 있으리라 판단되는 두 가지 일이 발생했다. 하나는 고속도로 5번(Highway No. 5, 또는 The Golden State Freeway)이 완공되어 교통량이 분산되었고,

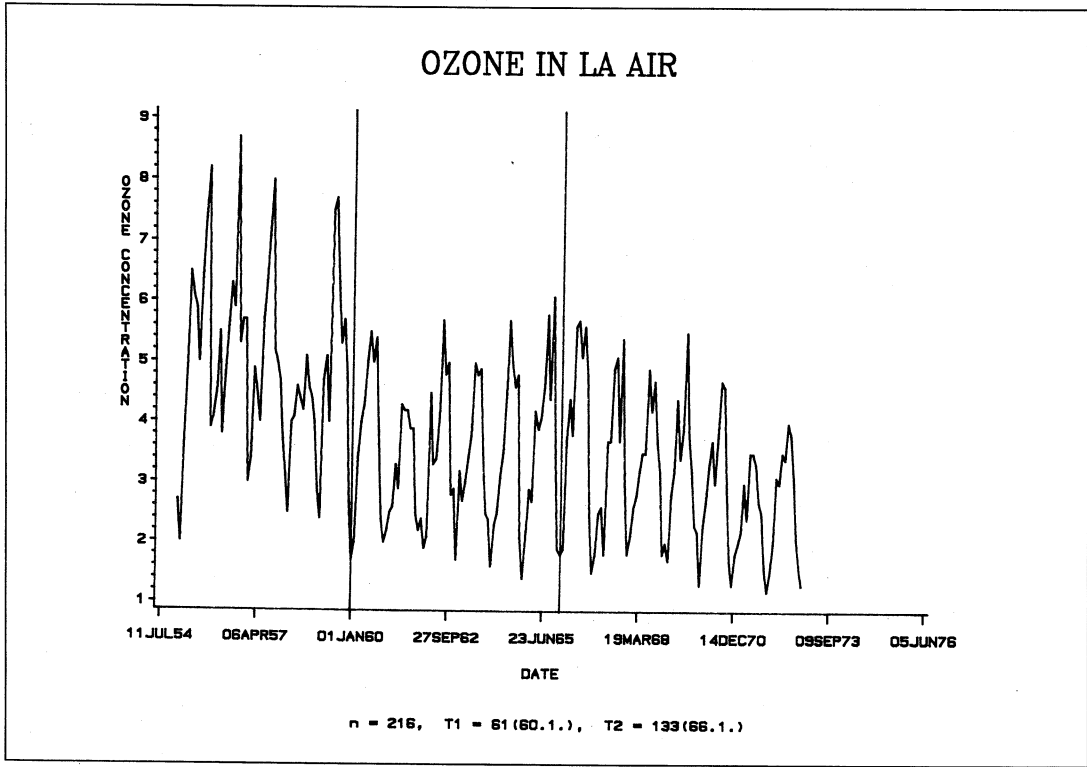


그림 5.16 시계열데이터 LAAIR의 시계열산포도

다른 하나는 이 지역에서 판매하는 가솔린에 포함된 탄화수소(hydrocarbons)의 양을 줄이도록 하는 새 법안(Rule 63)이 실행되었다. 따라서 1959년 12월이 첫 번째 간섭시점 $T1=60$ 이다. 또한, 1966년에는 오존발생량을 줄일 것으로 기대되는 새로운 자동차엔진설계법규가 실시되었다. 이 법규에 의한 엔진의 변화는 오존을 발생시키는 광화학적 반응을 감소시킨다. 이 오존발생의 감소율은 여름과 겨울의 대기온도차이와 일조량차이로 인해서 겨울보다 여름이 더 높을 것이다. 따라서 두 번째 간섭변수를 각각 여름과 겨울을 나타내는 두 개의 확장된 순간과동입력들로 정의한다. 요약하면 이 시계열데이터를 분석하기 위한 간섭시계열모형의 입력변수들은 다음과 같다.

$$S_t = \begin{cases} 0 & (t < 1960\text{년 } 1\text{월}) \\ 1 & (t \geq 1960\text{년 } 1\text{월}) \end{cases}$$

$$\eta_{1,t} = \begin{cases} 1 & (1966\text{년 } 1\text{월 이후의 } 6\text{월}, 7\text{월}, 8\text{월}, 9\text{월}, 10\text{월}) \\ 0 & (\text{다른 경우}) \end{cases}$$

$$\eta_{2,t} = \begin{cases} 1 & (1966\text{년 } 1\text{월 이후의 } 11\text{월}, 12\text{월}, 1\text{월}, 2\text{월}, 3\text{월}, 4\text{월}, 5\text{월}) \\ 0 & (\text{다른 경우}) \end{cases}$$

그러나 그림 5.16에서 이러한 간섭들의 효과를 직관적으로 발견하기는 어렵다. 지금부터 통계적인 분석을 통해서 간섭들의 효과가 있는지 여부와 효과가 있다면 크기가 얼마인가

를 조사하기로 하자.

이 시계열데이터 LAAIR에 적합한 간섭시계열모형을 구하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM5-17.PRG에 수록되어 있다).

```

/* */
/*          EXAM5-17.PRG          */
/* */
/*          INTERVENTION ANALYSIS          */
/*          DATASET = LAAIR          */
/* */
OPTIONS LINESIZE=72 PAGESIZE=60;
LIBNAME TSA4DATA 'A:\TSA4DATA';
DATA EXAM517;
  SET TSA4DATA.LAAIR;
  RUN;
PROC ARIMA DATA=EXAM517(OBS=60);
  I VAR=OZONE(12);
  RUN;
PROC ARIMA DATA=EXAM517;
  I VAR=OZONE(12) CROSS=(X1(12) SUMMER WINTER)
  NOPRINT;
  RUN;
  E Q=(1)(12) INPUT=(X1 SUMMER WINTER) NOINT
  METHOD=ML MAXIT=20;
  RUN;
  F LEAD=14 BACK=2;
  RUN;

```

1955년 1월부터 1959년 12월까지는 간섭의 영향을 받지 않기 때문에 이 기간 동안의 60개 관찰점들을 사용해서 잡음항의 모형을 식별하자. 간섭시점 $T=60$ 이전까지의 시계열데이터를 식별하기 위해서 다음과 같은 첫 번째 PROC ARIMA를 수행한다.

```

PROC ARIMA DATA=EXAM517(OBS=60);
  I VAR=OZONE(12);
  RUN;

```

그림 5.16에서 알 수 있듯이 시계열데이터 LAAIR는 계절주기 12를 갖는다. 또한, 이 시계열데이터의 표본자기상관함수, 표본역자기상관함수, 표본부분자기상관함수로부터 이 시계열데이터를 정상시계열데이터로 변환하기 위해서는 계절시차 12의 차분을 수행할 필요

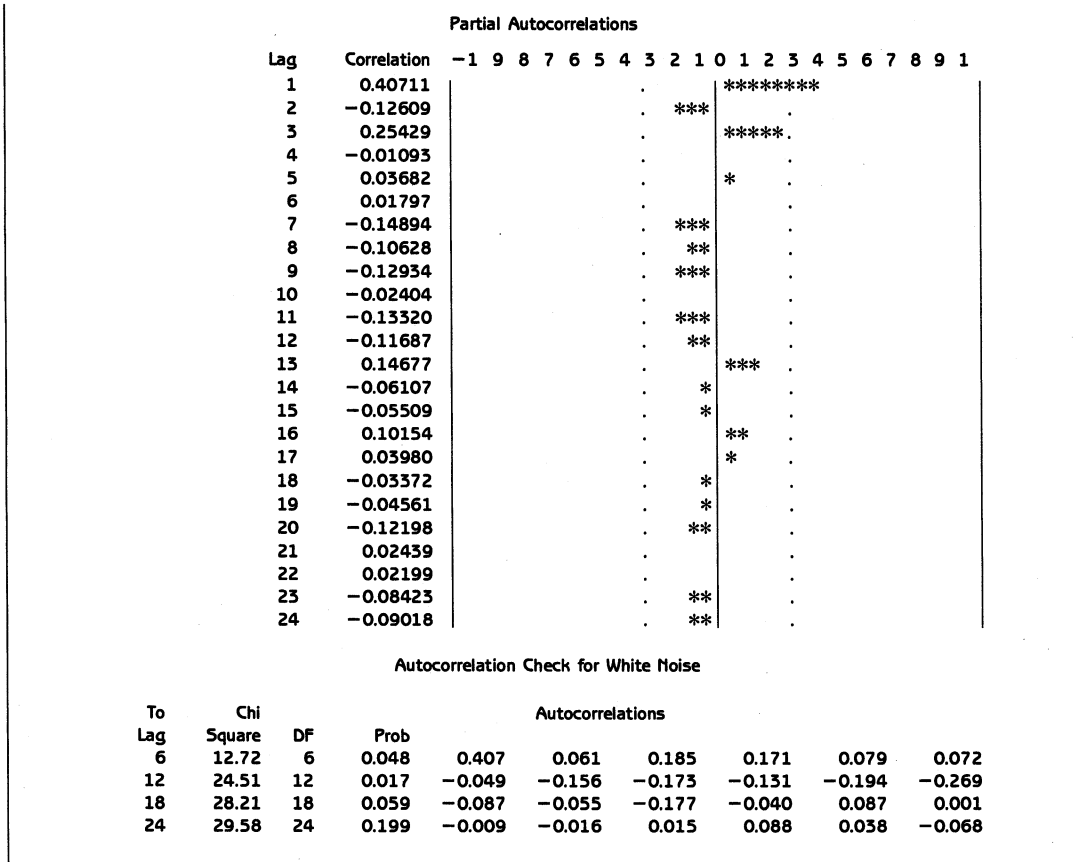


그림 5.17.1 시계열데이터 LAAIR의 잡음항모형의 식별

가 있음을 알 수 있다. 이 PROC ARIMA는 이러한 차분된 시계열데이터 $\{V_{12}y_t\}$ 를 식별하기 위한 것이다. 이 PROC ARIMA를 수행한 결과가 그림 5.17.1이다. 그림 5.17.1에서 차분된 시계열데이터 $\{V_{12}y_t\}$ 의 표본자기상관함수는 시나브로형태를 보이고, 시차 1에서 표본역자기상관계수와 표본부분자기상관계수는 유의적으로 0과 다른 값들을 가짐을 알 수 있다. 따라서 시계열데이터 LAAIR의 잡음항모형은 다음과 같다고 할 수 있다.

$$(1-B^{12})n_t = (1-\theta_1 B)v_t$$

그러나 시차 12에서 표본역자기상관계수와 표본부분자기상관계수의 절대값들은 상당히 크다. 따라서 Box & Tiao(1975, p. 73)는 시계열데이터 LAAIR의 잡음항모형을 다음과 같이 가정했다.

$$(1-B^{12})n_t = (1-\theta_1 B)(1-\theta_{12} B^{12})v_t$$

여기서는 Box & Tiao의 제안을 따르기로 한다.

앞에서 설명한 간섭들의 형태와 식별된 잡음항모형을 바탕으로 시계열데이터 LAAIR는 다음과 같은 간접시계열모형을 갖는다고 할 수 있다.

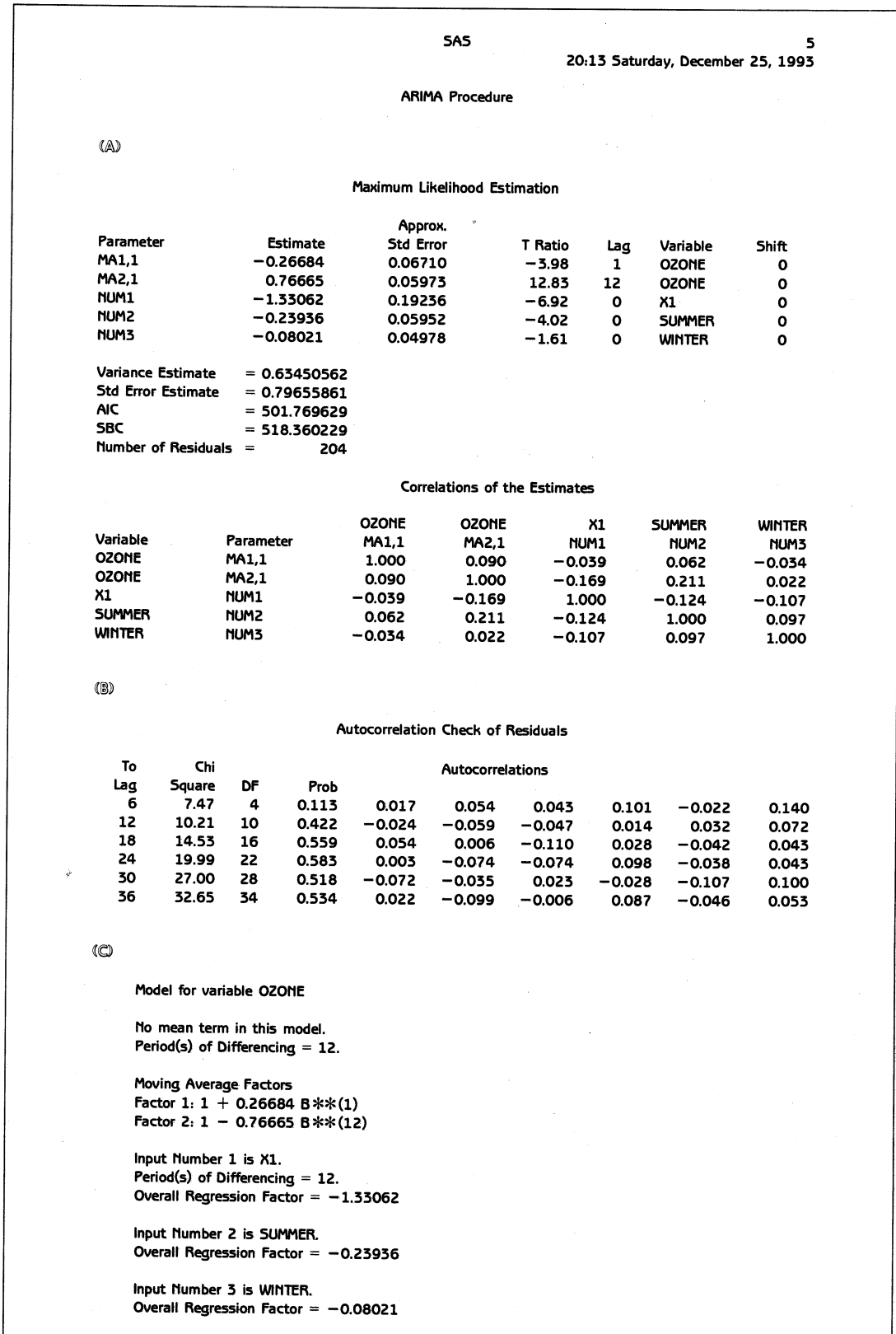


그림 5.17.2 시계열모형 LAAIR의 간섭시계열모형추정

$$y_t = \omega_0 S_t + \omega_1 \frac{\eta_{1t}}{1-B^{12}} + \omega_2 \frac{\eta_{2t}}{1-B^{12}} + \frac{(1-\theta_1 B)(1-\theta_{12} B^{12})}{1-B^{12}} v_t$$

이 간접시계열모형을 추정하기 위해서 다음과 같은 PROC ARIMA를 수행한다.

```
PROC ARIMA DATA=EXAM517;
  I VAR=OZONE(12) CROSS=(X1(12) SUMMER WINTER)
  NOPRINT;
  RUN;
  E Q=(1)(12) INPUT=(X1 SUMMER WINTER) NOINT
  METHOD=ML MAXIT=20;
  RUN;
```

이 PROC ARIMA를 수행한 결과가 그림 5.17.2이다. 그림 5.17.2의 부분 A에서 추정된 모형이 다음과 같음을 알 수 있다.

$$y_t = -1.331 S_t - 0.239 \frac{\eta_{1t}}{1-B^{12}} - 0.0802 \frac{\eta_{2t}}{1-B^{12}} + \frac{(1+0.267B)(1-0.767B^{12})}{1-B^{12}} v_t$$

간접시계열모형의 MA특성근이 단위원에 가까이 있으면 조건부최소제곱추정법보다는 비조건부최소제곱추정법이나 최우추정법을 사용하는 것이 좋다. 이 PROC ARIMA에서는 ESTIMATE스테이트먼트의 옵션 METHOD=에 ML을 지정하였다. 부분 B의 잔차의 퍼트맨토우통계량값들로부터 추정된 간접시계열모형이 주어진 시계열데이터에 적합하다고 결론지을 수 있다.

시계열데이터 LAAIR의 마지막 12개의 관찰점들의 변수 OZONE의 관찰값들은 결측이다. 이 시계열데이터를 사용해서 이 결측값들을 예측하기 위해서 다음 FORECAST스테이트먼트를 수행한다.

```
F LEAD=14 BACK=2;
  RUN;
```

이 PROC스텝을 수행한 결과가 그림 5.17.3이다. 그림 5.17.3의 부분 B에 시계열데이터 LAAIR의 표준오차, 신뢰수준 95%의 예측구간의 하한과 상한이 실려 있다.

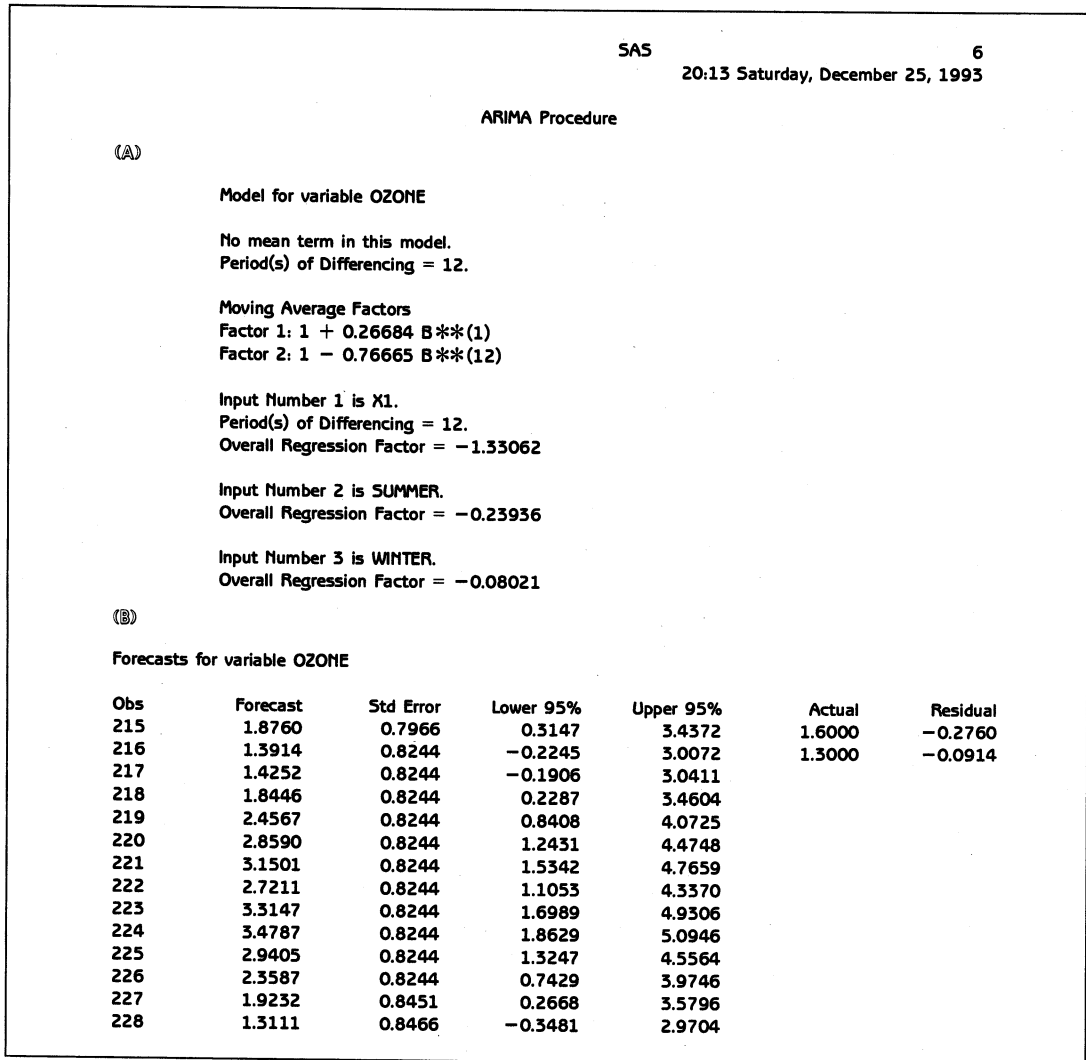


그림 5.17.3 간섭시계열모형에 의한 시계열데이터 LAAIR의 예측

5.2.6 참고사항

간섭시계열데이터에 관한 좀더 다양한 모형들과 예제들에 관해서는 Abraham(1980, 1987), Bhattacharyya & Layton(1979), Bireley(1987), Box & Tiao(1965), Capie, Mills & Wood(1986), Glass(1972), Izenman & Zabell(1981), Mills(1990, pp. 235~247), Mills & Stephenson(1987), Tiao, Box & Hamming(1975a, 1975b) 등을 참조하라.

5.3 공 누 적

5.3.1 과다차분

이 소절에서는 과다차분(overdifferencing)에 대해서 설명하기 위한 예제로써 시계열데이터 HOUSING을 잡음항이 백색잡음과정인 아닌 ARMA모형을 따르는 회귀모형으로 분석하고자 한다. 시계열데이터 HOUSING은 미국의 1974년 1사분기부터 1984년 1사분기까지의 분기별 건축자재판매액을 나타내는 변수 SALES, 분기별 착공한 주택의 수를 나타내는 변수 STARTS, 분기별 처분가능한 개인소득(disposable personal income)을 나타내는 변수 DPIC, 그리고 분기별 주택자금이자율(mortgage rates)을 나타내는 변수 MORT의 관찰값들을 기록한 것이다. 이 시계열데이터의 관찰점들의 개수는 41이고, SAS 데이터라이브러리 TSA4DATA의 영구SAS데이터세트 HOUSING.SSD에 수록되어 있다. 시계열데이터 HOUSING의 각 변수의 시계열산포도가 제 2.5절에 수록되어 있다. 이 시계열산포도들로부터 건축자재판매액은 착공주택수와 매우 비슷한 패턴을 가지고 있음을 알 수 있다.

이 시계열데이터를 [BD86](pp. 122~134)과 [CNote2](pp. 347~380)에서 잡음항이 ARMA모형을 따르는 회귀모형으로 분석하였다. 이 소절에서는 우선 과다차분의 문제점을 지적하기 위해서 [BD86]과 [CNote2]가 이 시계열데이터를 잡음항이 ARMA모형을 따르는 회귀모형으로 분석한 것을 반복해 보기로 한다.

지금부터 우리가 분석하고자 하는 것은 종속변수를 SALES(S)로, 설명변수들을 MORT(M), DPIC(D), STARTS(H)로 하는 다음과 같은 회귀모형이다.

$$S_t = \beta_0 + \beta_1 D_t + \beta_2 H_t + \beta_3 M_t + w_t$$

만일, 잡음항 $\{w_t\}$ 가 AR모형을 갖는다면 PROC AUTOREG를 사용해서 모형화할 수 있다. 그러나 잡음항 $\{w_t\}$ 가 ARMA모형을 갖는다면 PROC ARIMA를 사용해서 모형화해야 한다.

제 2.5절의 시계열산포도들로부터 변수들 DPIC, STARTS, MORT, 그리고 SALES 모두가 정상성을 갖지 않음을 알 수 있다. 이러한 경우 일반적인 시계열분석의 첫 단계는 각 변수를 정상성을 갖는 변수로 변환시키는 것이다. 시계열산포도들로부터 변수들 DPIC, STARTS, MORT, 그리고 SALES가 분산동질성을 갖는 것을 알 수 있다. 따라서 차분을 통해서 각 변수를 정상성을 갖는 변수로 변환시킬 수 있다. 각 변수의 시계열데이터를 차수 1의 차분을 취한 차분된 시계열데이터(differenced time series)의 시계열산포도는 정상성을 보인다. 각 차분된 시계열데이터의 표본자기상관함수도 정상성의 가정에서 벗어나

지 않음을 알 수 있다. 독자들은 변수 DPIC의 차분된 시계열데이터가 약간 증가하는 추세를 보임을 알 수 있을 것이다. 그러나 이러한 추세는 변수 DPIC를 종속변수로 하는 시계열모형을 찾는 경우가 아니면 별 문제가 되지 않는다. 따라서 우리가 다루고자 하는 모형은 다음과 같다.

$$\nabla S_t = 0 + \beta_1 \nabla D_t + \beta_2 \nabla H_t + \beta_3 \nabla M_t + \nabla w_t$$

이 차분된 회귀모형을 추정하기 위해서는 차분된 잡음항 $\{\nabla w_t\}$ 가 정상적이라는 가정이 필요하다. 즉, 회귀모형의 잡음항 $\{\nabla w_t\}$ 의 AR특성식이 단위근 1을 갖는다는 가정이 필요하다. 이러한 가정은 설명변수들에 관한 가정들과는 달리 시계열데이터를 모형화하는 데 매우 중요하기 때문에 이러한 가정을 하지 않으려면 차분되지 않은 변수들을 사용해서 시계열분석을 행해야 한다. 차분된 회귀모형에는 상수항이 없다. 만일 변수 SALES가 $(\beta_0 + \psi t)$ 형태의 추세를 포함한다면 차분된 회귀모형은 상수항 ψ 를 포함한다. 차분된 회귀모형을 추정하기 위해서는 PROC AUTOREG를 사용하지 못하고 PROC ARIMA를 사용해야 한다.

원래 회귀모형의 잡음항 $\{w_t\}$ 의 AR특성식이 단위근을 갖는다는 가정이 옳다는 전제하에 차분된 회귀모형을 추정하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM5-18.PRG에 수록되어 있다).

```

/* */
/* */
/* */
/* */
/* */
/* */
/* */
/* */
LIBNAME TSA4DATA 'A:TSA4DATA';
OPTIONS LINESIZE=72 PAGESIZE=60;
TITLE 'MODEL W/ THE FIRST DIFFERENCES';
PROC ARIMA DATA=TSA4DATA.HOUSING;
  I VAR=SALES(1) NOPRINT
  CROSS=(MORT(1) DPIC(1) STARTS(1));
  E INPUT=(MORT DPIC STARTS) METHOD=ML PLOT MAXIT=99;
  RUN;
  E Q=1 INPUT=(MORT DPIC STARTS) METHOD=ML MAXIT=99;
  RUN;
  E Q=1 INPUT=(MORT STARTS) METHOD=ML MAXIT=99;
  RUN;
  E Q=1 INPUT=(STARTS) METHOD=ML MAXIT=99;

```

```
RUN:
TITLE1:
```

이 PROC ARIMA의 IDENTIFY스테이트먼트의 옵션 VAR=에는 종속변수를 지정하고 옵션 CROSSCOR=에는 입력변수들을 지정한다. 여기서 종속변수는 SALES를 차분한 SALES(1)이고 입력변수들은 변수들 MORT, DPIC, STARTS를 차분한 MORT(1), DPIC(1), STARTS(1)이다. 잡음항이 ARMA과정인 회귀모형을 추정하기 위해서 PROC ARIMA를 사용할 때 IDENTIFY스테이트먼트는 단지 종속변수와 입력변수들, 그리고 이들의 차분차수들을 지정하는 데 사용한다. 따라서 이 IDENTIFY스테이트먼트에서는 옵션 NOPRINT를 사용했다.

이 PROC ARIMA의 첫 번째 ESTIMATE스테이트먼트는 다음과 같다.

```
E INPUT=(MORT DPIC STARTS)
METHOD=ML PLOT MAXIT=99;
RUN;
```

이 ESTIMATE스테이트먼트는 최우추정법을 사용해서 다음 모형을 추정하기 위한 것이다.

$$VS_t = \psi + \beta_1 \nabla M_t + \beta_2 \nabla D_t + \beta_3 \nabla H_t + v_t$$

여기서 $\{w_t\}$ 는 백색잡음과정이라고 가정한다. ESTIMATE스테이트먼트의 옵션 INPUT=에는 입력변수들을 지정한다. 앞에서도 언급했듯이 이 입력변수들은 이 ESTIMATE스테이트먼트 바로 앞의 IDENTIFY스테이트먼트의 옵션 CROSSCOR=에 지정되어 있어야 한다. 그러나 ESTIMATE스테이트먼트에서는 종속변수나 입력변수들의 차분차수나 계절차분차수를 따로 지정하지 않고 이 ESTIMATE스테이트먼트에 해당하는 IDENTIFY스테이트먼트에서 지정한 차수들을 그대로 사용한다.

이 첫 번째 ESTIMATE스테이트먼트를 수행한 결과가 그림 5.18.1이다. 독자들은 그림 5.18.1, 같은 ESTIMATE스테이트먼트를 수행한 [BD86] (pp. 126~127), 그리고 같은 ESTIMATE스테이트먼트를 수행한 [CNote2] (pp. 356~363)가 서로 다르다는 점에 유의하라. 같은 시계열데이터에 같은 통계패키지의 같은 프로그램을 적용한 결과가 서로 다르다는 것은 이상한 일이다. 이러한 현상이 일어나는 이유는 뒤에서 설명할 것이다. 부분 A에서 최우추정된 회귀모형이 다음과 같음을 알 수 있다.

$$\nabla S_t = 170.039 - 151.074 \nabla M_t + 1.002 \nabla D_t + 4.930 \nabla H_t + v_t$$

부분 B에서 알 수 있듯이 종속변수 SALES는 설명변수 STARTS와 양의 상관관계에 있고, 설명변수 MORT와는 음의 상관관계에 있으며, 설명변수 DPIC와도 음의 상관관계에 있다. 부분 A에서 설명변수들 중에서 STARTS의 추정계수의 t 통계량값만이 유의적으로 0과 다름을 알 수 있다. 그러나 올바른(correct) 모형이 아니면 t 통계량은 별 의미가 없다.

MODEL W/ THE FIRST DIFFERENCES

1

21:12 Monday, December 27, 1993

ARIMA Procedure

(A)

Maximum Likelihood Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag	Variable	Shift
MU	170.03857	181.63744	0.94	0	SALES	0
NUM1	-151.07400	112.12506	-1.35	0	MORT	0
NUM2	-1.00212	4.22489	-0.24	0	DPIC	0
NUM3	4.93009	0.39852	12.37	0	STARTS	0

Constant Estimate = 170.038573

Variance Estimate = 200686.53

Std Error Estimate = 447.980501

AIC = 605.680639

SBC = 612.436156

Number of Residuals = 40

Correlations of the Estimates

Variable	Parameter	SALES MU	MORT NUM1	DPIC NUM2	STARTS NUM3
SALES	MU	1.000	-0.028	-0.916	0.016
MORT	NUM1	-0.028	1.000	-0.071	0.349
DPIC	NUM2	-0.916	-0.071	1.000	-0.039
STARTS	NUM3	0.016	0.349	-0.039	1.000

(B)

Autocorrelation Check of Residuals

To Lag	Chi Square	DF	Prob	Autocorrelations					
6	8.90	6	0.179	-0.397	-0.028	-0.012	0.128	-0.088	0.138
12	10.27	12	0.592	-0.099	0.035	-0.033	-0.099	0.053	0.026
18	16.98	18	0.524	-0.011	-0.122	0.143	-0.189	0.014	0.156
24	23.37	24	0.498	-0.033	-0.169	0.174	-0.043	-0.039	0.093

(C)

Autocorrelations plot of Residuals

Lag	Covariance	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1		
0	200687	1.00000												*****											
1	-79669.125	-0.39698												*****											
2	-5682.473	-0.02832												*											
3	-2373.543	-0.01183																							
4	25596.562	0.12754																							
5	-17570.433	-0.08755												**											
6	27703.666	0.13804												**											
7	-19780.522	-0.09856												**											
8	7007.339	0.03492												*											
9	-6640.664	-0.03309												*											
10	-19897.518	-0.09915												**											
11	10541.993	0.05253												**											
12	5313.198	0.02648												*											
13	-2304.153	-0.01148												*											
14	-24384.039	-0.12150												**											
15	28753.116	0.14327												**											
16	-37882.935	-0.18877												****											
17	2899.354	0.01445												*											
18	31316.264	0.15605												**											
19	-6544.860	-0.03261												*											
20	-33873.612	-0.16879												**											
21	34855.776	0.17368												**											
22	-8723.817	-0.04347												*											
23	-7901.639	-0.03937												*											
24	18724.038	0.09330												**											

“.” marks two standard errors

(D)

Inverse Autocorrelations

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
1	0.50856												*****										
2	0.20827												****										
3	-0.00282																						
4	-0.17168										***												
5	-0.15771									***													
6	-0.09592									**													
7	0.06433										*												
8	0.10559											**											
9	0.18617											****											
10	0.18246											****											
11	0.07591											**											
12	0.04508											*											
13	0.05355											*											
14	0.09734											**											
15	0.06292											*											
16	0.10731											**											
17	0.03770											*											
18	-0.04631										*												
19	0.03111											*											
20	0.06078											*											

(E)

Partial Autocorrelations

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
1	-0.39698											*****											
2	-0.22069										****												
3	-0.14121										***												
4	0.07283											*											
5	-0.00035																						
6	0.16471												***										
7	0.03533											*											
8	0.02720											*											
9	-0.02400																						
10	-0.20475										****												
11	-0.10463										**												
12	-0.05685										*												
13	0.02000																						
14	-0.09278										**												
15	0.08914												**										
16	-0.12286										**												
17	-0.16250										***												
18	0.09943												**										
19	0.05263												*										
20	-0.09668										**												
21	0.07134												*										
22	0.05364												*										
23	-0.06294											*											
24	0.03711										*		*										

(F)

Model for variable SALES

Estimated Intercept = 170.038573
 Period(s) of Differencing = 1.

Input Number 1 is MORT.
 Period(s) of Differencing = 1.
 Overall Regression Factor = -151.074

Input Number 2 is DPIC.
 Period(s) of Differencing = 1.
 Overall Regression Factor = -1.00212

Input Number 3 is STARTS.
 Period(s) of Differencing = 1.
 Overall Regression Factor = 4.930094

그림 5.18.1 설명변수들이 MORT(1), DPIC(1), STARTS(1)인 최소제곱추정모형

이 ESTIMATE스테이트먼트에서는 아직 잡음항의 모형을 고려하지 않았기 때문에 설명변수 STARTS의 추정계수의 t 값이 유의적으로 0과 다르다는 것은 의미가 없다. 따라서 추정된 모형의 타당성은 t 검정이 아닌 부분 B의 퍼트맨토우통계량을 사용해서 검정해야 한다. 이 퍼트맨토우통계량값들로부터 추정된 회귀모형의 잔차들이 백색잡음과정에서 나왔다고 할 수 있다. 그러나 [SAS4TSA2] (pp. 516~3518)에서 언급되었듯이 퍼트맨토우통계량을 사용한 검진의 결과를 지나치게 믿어서는 안 되고 시차들 1, 2, 3에서의 표본자기상관계수들을 조사해 보아야 한다. 첫 번째 퍼트맨토우통계량값은 8.90이고 이에 해당하는 p 값은 0.179이다. 이 퍼트맨토우통계량값은 잔차값들의 첫 여섯 개, 즉 시차들 1, 2, 3, 4, 5, 6의 표본자기상관계수들로부터 계산된 것이다. 반면에 시차 1의 표본자기상관계수는 -0.397 로서 유의적으로 0과 다르다. 이는 나머지 다섯 개의 표본자기상관계수들이 작기 때문에 첫 번째 퍼트맨토우통계량값이 유의적으로 0과 같다고 할 수 있는 것이다. 작은 시차들의 표본자기상관계수들을 조사한 결과와 차분을 하면 잡음항이 AR모형을 가질 수도 있다는 점을 이용해서 모든 중요한 설명변수들과 적당한 잡음항모형을 포함하는 회귀모형을 사용해야 t 검정이 의미가 있다. 이는 잡음항이 백색잡음과정인 Gauss-Markov조건을 만족하는 회귀모형에서 중요한 설명변수가 빠져 있으면 t 검정통계량을 사용해서는 안 되는 것과 마찬가지로이다.

잔차들의 표본자기상관함수, 표본역자기상관함수, 그리고 표본부분자기상관함수를 사용해서 잡음항의 모형을 식별한다. 부분 C의 표본자기상관함수는 시차 2 이후에서 절단형태를 보이고 표본역자기상관함수와 표본부분자기상관함수는 시나브로형태를 보이므로 잡음항은 MA(1)과정으로부터 나왔다고 할 수 있다. 특히 시차 1에서의 표본자기상관계수는 -0.397 로 -0.5 에 가깝다. 그러므로 이 MA(1)과정의 시차 1에서의 MA계수가 1에 가깝다. 시계열데이터가 과다차분(overdifferenced)되면 잔차의 표본역자기상관함수가 시나브로형태를 보이면서 시차가 커짐에 따라 감소하는 표본역자기상관함수의 절대값의 감소정도가 작다. 따라서 이 시계열모형의 변수들은 과다차분되었음을 알 수 있다. 표본역자기상관함수의 절대값의 감소정도가 크면 시계열데이터는 적절히 차분된 것이다.

다음과 같은 간단한 회귀모형을 살펴보자.

$$y_t = \beta_0 + \beta_1 x_t + n_t$$

여기서 $\{x_t\}$ 와 $\{n_t\}$ 는 단위근을 갖는 확률과정들이다. 이 경우에는 종속변수 y 를 입력변수 x 로 회귀시킨 β_1 의 최소제곱추정량은 일치성을 갖지 못하므로 이 추정모형의 잔차값들은 잡음항을 잘 나타내지 못한다. 따라서 ESTIMATE스테이트먼트의 옵션 PLOT를 사용해서 차분이 필요한지의 여부를 결정하는 것은 불가능하다. 이 예제에서는 설명변수들이 차분을 필요로 하므로, 종속변수 SALES를 차분한 차분시계열데이터를 모형화한 다음에 ESTIMATE스테이트먼트의 옵션 PLOT를 사용해서 이 차분시계열데이터가 과다차분이 되었는지를 조사한다. 또한 과다차분된 시계열데이터의 잡음항의 시차 1에서의 MA계수가 1에 가까운 점을 이용해서 과다차분 여부를 결정한다. 이 예제에서는 부분 C에서 MA항의 계수가 1.000에 가까우므로 시계열데이터가 과다차분되었다는 것을 알 수 있다.

이 PROC ARIMA의 두 번째 ESTIMATE스테이트먼트는 다음과 같다.

```
E Q=1 INPUT=(MORT DPIC STARTS)
METHOD=ML MAXIT=99;
RUN;
```

이 ESTIMATE스테이트먼트는 최우추정법을 사용해서 다음 모형을 추정하기 위한 것이다.

$$VS_t = \psi + \beta_1 \nabla M_t + \beta_2 \nabla D_t + \beta_3 \nabla H_t + (1 - \theta_1 B) v_t$$

이 ESTIMATE스테이트먼트를 수행한 결과가 그림 5.18.2이다. 그림 5.18.2의 부분 A에서 추정된 모형이 불안정하고 추정값들이 수렴하지 않을 수도 있다는 것을 알 수 있다. 실제로 그림 5.18.2, 같은 ESTIMATE스테이트먼트를 수행한 [BD86] (pp. 129~131), 그리고 같은 ESTIMATE스테이트먼트를 수행한 [CNote2] (pp. 365~367)가 서로 다르다. 같은 시계열데이터에 같은 통계패키지의 같은 프로그램을 적용한 결과가 서로 다른 이유는 추정된 모형들이 불안정하고 추정값들이 수렴하지 않기 때문이다. 본 저자는 이러한 현상은 반복횟수를 조절하는 멈춤규칙(stopping rule)에 의한 것이 아닌가 하고 생각한다. 부분 B에서 최우추정된 회귀모형이 다음과 같음을 알 수 있다.

$$VS_t = 91.381 - 202.262 \nabla M_t + 0.896 \nabla D_t + 5.131 \nabla H_t + (1 - 1.000 B) v_t$$

부분 C의 퍼트맨트우통계량값들로부터 이 추정된 회귀모형의 잔차들이 백색잡음과정에서 나왔다고 할 수 있다. 그러나 부분 B에서 MA항의 계수가 1.000에 가까우므로 시계열데이터가 과다차분되었다는 것을 알 수 있다. 이 추정된 회귀모형에서 각 설명변수의 회귀계수의 t 통계량값은 그림 5.18.1의 최소제곱추정값의 t 통계량값과 다르다.

당연한 말이지만 종속변수와 관련이 없는 설명변수들을 포함한 회귀모형을 사용해서 예측을 해서는 안 된다. 적당하지 않은 설명변수가 모형에 포함되면 이 설명변수의 변동(fluctuation)에 의해서 예측오차분산(forecast error variance)이 불필요하게 커진다. 따라서 모형에서 불필요한 설명변수들을 제거해야 한다. 설명변수들의 회귀계수들의 t 통계량값들을 바탕으로 모형에서 불필요한 설명변수들을 제거할 때 한 번에 두 개 이상의 설명변수들을 제거해서는 안 된다. 그 이유는 모형에서 한 설명변수를 제거하면 나머지 설명변수들의 t 통계량값들이 변하기 때문이다. 그림 5.18.2에서 변수 ∇D_t 의 t 통계량값은 0.78로 유의적으로 0과 다르지 않으며, 다른 두 변수들 ∇M_t 와 ∇H_t 의 t 통계량값들보다 더 0에 가깝다. 이러한 결론을 직관적으로 설명하면 다음과 같다. 제 2.5절의 시계열산포도에서 볼 수 있듯이 차분되지 않은 변수 DPIC는 $(\alpha_0 + \alpha_1 t)$ 의 형태이다. 따라서, ∇D_t 는 상수 α_1 의 형태이고 이는 회귀모형의 상수항 ψ 에 이미 반영이 되었다. 그러므로 회귀모형에서 상수항이나 설명변수 ∇D_t 를 제거하여야 할 것이다. 그러나 상수항과 설명변수 ∇D_t 를 동시에 제거한 회귀모형은 주어진 시계열데이터와 잘 적합되지도 않을 뿐더러 예측력도 떨어질 것이다. 예측을 하기 위해서는 회귀모형에서 상수항을 제거하는 것보다 설명변수 ∇D_t 를 제거하는

MODEL W/ THE FIRST DIFFERENCES

5

21:12 Monday, December 27, 1993

ARIMA Procedure

(A)

WARNING: The model defined by the new estimates is unstable. The iteration process has been terminated.
 WARNING: Estimates may not have converged.

ARIMA Estimation Optimization Summary

Estimation Method: Maximum Likelihood
 Parameters Estimated: 5
 Termination Criteria: Maximum Relative Change in Estimates
 Iteration Stopping Value: 0.001
 Criteria Value: 27.517926
 Maximum Absolute Value of Gradient: 234254.109
 R-Square (Relative Change in Regression SSE) from Last Iteration Step: 0.02349406
 Objective Function: Log Gaussian Likelihood
 Objective Function Value: -289.06765
 Marquardt's Lambda Coefficient: 1E-8
 Numerical Derivative Perturbation Delta: 0.001
 Iterations: 13
 Warning Message: Estimates may not have converged.

(B)

Maximum Likelihood Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag	Variable	Shift
MU	91.38149	47.30628	1.93	0	SALES	0
MA1,1	0.99973	30.48149	0.03	1	SALES	0
NUM1	-202.26240	60.63966	-3.34	0	MORT	0
NUM2	0.89566	1.14910	0.78	0	DPIC	0
NUM3	5.13054	0.28083	18.27	0	STARTS	0

Constant Estimate = 91.3814878

Variance Estimate = 115435.651

Std Error Estimate = 339.758224

AIC = 588.135302

SBC = 596.579699

Number of Residuals = 40

Correlations of the Estimates

Variable	Parameter	SALES MU	SALES MA1,1	MORT NUM1	DPIC NUM2	STARTS NUM3
SALES	MU	1.000	0.360	-0.219	-0.958	-0.612
SALES	MA1,1	0.360	1.000	0.159	-0.402	-0.314
MORT	NUM1	-0.219	0.159	1.000	-0.052	0.651
DPIC	NUM2	-0.958	-0.402	-0.052	1.000	0.457
STARTS	NUM3	-0.612	-0.314	0.651	0.457	1.000

(C)

Autocorrelation Check of Residuals

To Lag	Chi Square	DF	Prob	Autocorrelations					
6	2.08	5	0.838	0.017	0.079	0.076	0.147	-0.066	0.080
12	6.50	11	0.838	-0.128	-0.079	-0.144	-0.186	-0.037	-0.044
18	11.43	17	0.834	-0.120	-0.183	0.024	-0.137	0.040	0.074
24	14.65	23	0.907	-0.073	-0.045	0.108	-0.053	-0.073	0.088

(D)

Model for variable SALES

Estimated Intercept = 91.3814878

Period(s) of Differencing = 1.


```

Moving Average Factors
Factor 1: 1 - 0.99973 B**(1)

Input Number 1 is MORT.
Period(s) of Differencing = 1.
Overall Regression Factor = -202.262

Input Number 2 is DPIC.
Period(s) of Differencing = 1.
Overall Regression Factor = 0.895659

Input Number 3 is STARTS.
Period(s) of Differencing = 1.
Overall Regression Factor = 5.130543
    
```

그림 5.18.2 설명변수들이 MORT(1), DPIC(1), STARTS(1)인 회귀모형

것이 편하다. 따라서 설명변수 ∇D_t 를 회귀모형에서 제거하기로 하고 다음과 같은 세 번째 ESTIMATE스테이트먼트를 수행한다.

```

E Q=1 INPUT=(MORT STARTS) METHOD=ML MAXIT=99;
RUN;
    
```

이 ESTIMATE스테이트먼트는 다음 회귀모형을 최우추정하기 위한 것이다.

$$\nabla S_t = \psi + \beta_1 \nabla M_t + \beta_3 \nabla H_t + (1 - \theta_1 B) v_t$$

이 ESTIMATE스테이트먼트를 수행한 결과가 그림 5.18.3이다. 그림 5.18.3의 부분 A에서도 추정된 모형이 불안정하고 추정값들이 수렴하지 않을 수도 있다는 것을 알 수 있다. 실제로 그림 5.18.3, 같은 ESTIMATE스테이트먼트를 수행한 [CNote2] (pp. 368~370)가 서로 다르다. 부분 B에서 최우추정된 회귀모형이 다음과 같음을 알 수 있다.

$$\nabla S_t = 127.426 - 208.230 \nabla M_t + 5.016 \nabla H_t + (1 - 1.000B) v_t$$

부분 B에서 설명변수 $\{\nabla D_t\}$ 를 제거함으로써 상수항의 t 통계값이 1.93에서 9.49로 증가되었음을 알 수 있다. 부분 C의 퍼트맨토우통계량값들로부터 이 추정된 회귀모형의 잔차들이 백색잡음과정에서 나왔다고 할 수 있다. 그러나 부분 B에서 MA항의 계수가 1.000에 가까우므로 시계열데이터가 과대차분되었다는 것을 알 수 있다.

```

MODEL W/ THE FIRST DIFFERENCES
21:12 Monday, December 27, 1993
7

ARIMA Procedure

(A)

WARNING: The model defined by the new estimates is unstable. The iteration process has been terminated.
WARNING: Estimates may not have converged.

ARIMA Estimation Optimization Summary

Estimation Method: Maximum Likelihood
Parameters Estimated: 4
Termination Criteria: Maximum Relative Change in Estimates
    
```

Iteration Stopping Value: 0.001
 Criteria Value: 28.1516264
 Maximum Absolute Value of Gradient: 257702.738
 R-Square (Relative Change in Regression SSE) from Last Iteration Step: 0.03079252
 Objective Function: Log Gaussian Likelihood
 Objective Function Value: -289.51856
 Marquardt's Lambda Coefficient: 1E-9
 Numerical Derivative Perturbation Delta: 0.001
 Iterations: 11
 Warning Message: Estimates may not have converged.

(B)

Maximum Likelihood Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag	Variable	Shift
MU	127.42584	13.42871	9.49	0	SALES	0
MA1,1	0.99967	27.11547	0.04	1	SALES	0
NUM1	-208.22951	59.70702	-3.49	0	MORT	0
NUM2	5.01556	0.25966	19.32	0	STARTS	0

Constant Estimate = 127.42584

Variance Estimate = 114794.527

Std Error Estimate = 338.81341

AIC = 587.057125

SBC = 593.792643

Number of Residuals = 40

Correlations of the Estimates

Variable	Parameter	SALES MU	SALES MA1,1	MORT NUM1	STARTS NUM2
SALES	MU	1.000	0.017	-0.945	-0.680
SALES	MA1,1	0.017	1.000	-0.018	-0.321
MORT	NUM1	-0.945	-0.018	1.000	0.767
STARTS	NUM2	-0.680	-0.321	0.767	1.000

(C)

Autocorrelation Check of Residuals

To Lag	Chi Square	DF	Prob	Autocorrelations					
6	3.22	5	0.666	0.032	0.124	0.107	0.172	-0.053	0.096
12	8.12	11	0.703	-0.121	-0.080	-0.147	-0.194	-0.063	-0.071
18	13.29	17	0.716	-0.133	-0.169	0.033	-0.132	0.060	0.094
24	16.09	23	0.851	-0.037	-0.024	0.131	-0.020	-0.021	0.099

(D)

Model for variable SALES

Estimated Intercept = 127.42584

Period(s) of Differencing = 1.

Moving Average Factors

Factor 1: 1 - 0.99967 B**(1)

Input Number 1 is MORT.

Period(s) of Differencing = 1.

Overall Regression Factor = -208.23

Input Number 2 is STARTS.

Period(s) of Differencing = 1.

Overall Regression Factor = 5.015564

그림 5.18.3 설명변수들이 MORT(1)과 STARTS(1)인 회귀모형

그림 5.18.3의 회귀모형에서 사용된 입력변수의 하나인 차분된 변수 ∇M_t 를 사용하지 않고 시계열분석을 하기 위해서 다음과 같은 네 번째 ESTIMATE스태이트먼트를 수행한다.

```
E Q=1 INPUT=(STARTS) METHOD=ML MAXIT=99;
RUN;
```

이 ESTIMATE스태이트먼트는 다음 회귀모형을 최우추정하기 위한 것이다.

$$\nabla S_t = \psi + \beta_3 \nabla H_t + (1 - \theta_1 B) v_t$$

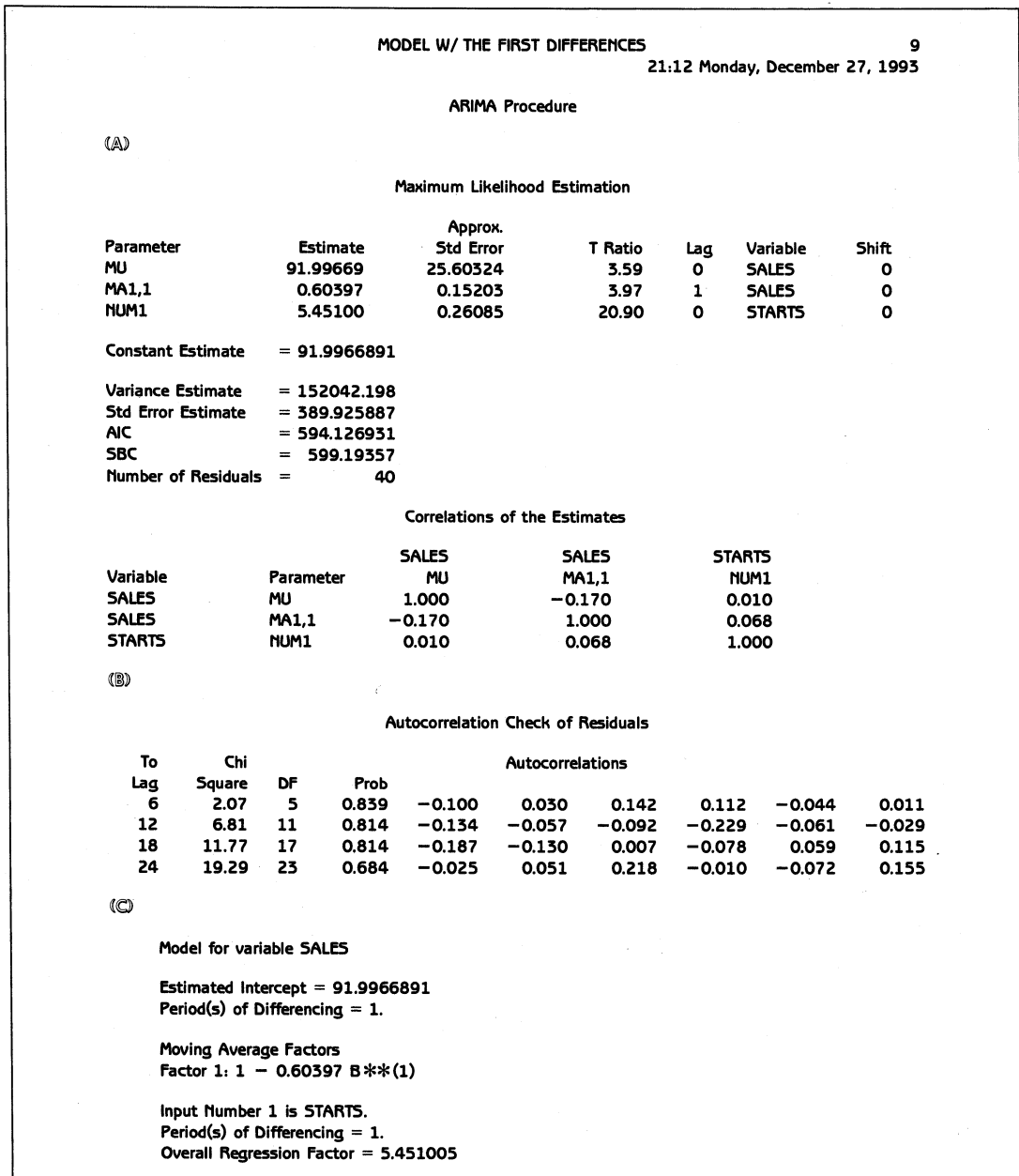


그림 5.18.4 설명변수가 STARTS(1)인 회귀모형

이 ESTIMATE스테이트먼트를 수행한 결과가 그림 5.18.4이다. 앞의 회귀모형들과 마찬가지로 그림 5.18.4, 같은 ESTIMATE스테이트먼트를 수행한 [BD86] (pp. 132~133), 같은 ESTIMATE스테이트먼트를 수행한 [CNote2] (pp. 371~373)가 서로 다르다. 부분 A에서 최우추정된 회귀모형이 다음과 같음을 알 수 있다.

$$\nabla S_t = 91.997 + 5.451 \nabla H_t + (1 - 0.604 B) v_t$$

부분 B의 퍼트맨토우통계량값들로부터 이 추정된 회귀모형의 잔차들은 백색잡음과정에서 나왔다고 할 수 있다.

지금까지의 분석은 [BD86] (pp. 122~133)과 [CNote2] (pp. 347~381)를 따른 것이다. 이 두 문헌에서는 그림 5.18.4에서 추정한 회귀모형을 시계열데이터 HOUSING에 가장 적합한 모형으로 선택하고 이 추정모형을 사용해서 예측을 할 것을 제안하였다. 그러나 그림 5.18.1에서 그림 5.18.4까지를 살펴보면 지금까지 분석한 회귀모형들에 과다차분되었음을 알 수 있다.

지금까지는 종속변수와 설명변수들을 차분을 취해서 시계열데이터 HOUSING을 분석하였고, 이러한 분석은 과다차분문제를 일으킨다는 것을 알았다. 지금부터는 과다차분문제를 피하기 위해서 차분하지 않은 변수들로 시계열모형화를 하고자 한다.

지금부터 우리가 분석하고자 하는 것은 종속변수를 SALES(S)로, 설명변수들을 MORT(M), DPIC(D), STARTS(H)로 하는 다음과 같은 회귀모형이다.

$$S_t = \beta_0 + \beta_1 D_t + \beta_2 H_t + \beta_3 M_t + w_t$$

여기서 잡음항 $\{w_t\}$ 가 ARMA모형을 갖는다고 가정하고 PROC ARIMA를 사용해서 모형화하기로 한다.

제 2.5절의 시계열산포도들로부터 변수들 DPIC, STARTS, MORT, 그리고 SALES 모두가 정상성을 갖지 않음을 알 수 있다. 앞에서 이러한 경우에 일반적인 시계열분석의 첫 단계는 각 시계열데이터를 정상성을 갖는 시계열데이터로 변환시키는 것이라고 했다. 여기서 강조하고자 하는 것은 이러한 제안이 일반적인 분석에서 통용되고 있는 것이지만 모든 경우에 그렇게 해야 한다는 것은 아니다. 그러나 어떤 시계열데이터에 분산이질성이 존재하면 변환을 사용해서 분산동질성을 갖는 시계열데이터로 변환시켜야 한다. 제 2.5절의 시계열산포도들로부터 다행히도 변수들 DPIC, STARTS, MORT, 그리고 SALES가 분산동질성을 갖는 것을 알 수 있다. 지금부터는 차분을 취하지 않고 이 시계열데이터를 분석하고자 한다.

원래 회귀모형의 잡음항 $\{w_t\}$ 가 정상적이라는 가정하에 차분되지 않은 회귀모형을 추정하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM5-19.PRG에 수록되어 있다).

```

/* */
/*          EXAM5-19.PRG          */
/* */
/*      REGRESSION WITH TIME SERIES ERROR      */
/*      DATA=HOUSEING                      */
/*      WITHOUT THE FIRST DIFFERENCES        */
/* */
LIBNAME TSA4DATA 'A:TSA4DATA';
OPTIONS LINESIZE=72 PAGESIZE=60;
TITLE 'MODEL W/ O THE FIRST DIFFERENCES';
PROC ARIMA DATA=TSA4DATA.HOUSING;
    I VAR=SALES CROSS=(MORT DPIC STARTS) NOPRINT;
    E INPUT=(MORT DPIC STARTS) METHOD=ML MAXIT=99;
    RUN;
    E INPUT=(MORT STARTS) METHOD=ML PLOT;
    E P=1 Q=1 INPUT=(MORT STARTS) METHOD=ML MAXIT=99;
    RUN;
    E INPUT=(STARTS) METHOD=ML PLOT;
    E P=2 INPUT=(STARTS) METHOD=ML MAXIT=99;
    RUN;
TITLE1;

```

이 PROC ARIMA의 IDENTIFY스테이트먼트의 옵션 VAR=에는 종속변수 SALES를 지정하고 옵션 CROSSCOR=에는 입력변수들 MORT, DPIC, STARTS 들을 지정했다.

이 PROC ARIMA의 첫 번째 ESTIMATE스테이트먼트는 다음과 같다.

```

E INPUT=(MORT DPIC STARTS) METHOD=ML MAXIT=99;
RUN;

```

이 ESTIMATE스테이트먼트는 최우추정법을 사용해서 다음 모형을 추정하기 위한 것이다.

$$S_t = \beta_0 + \beta_1 M_t + \beta_2 D_t + \beta_3 H_t + v_t$$

여기서 $\{v_t\}$ 는 백색잡음과정이라고 가정한다. 이 ESTIMATE스테이트먼트를 수행한 결과가 그림 5.19.1에 수록되어 있다. 부분 A에서 최우추정된 회귀모형이 다음과 같음을 알 수 있다.

$$S_t = 6794.8 - 166.224 M_t + 2.975 D_t + 5.497 H_t + v_t$$

이 추정된 회귀모형의 타당성을 조사하기 위해서는 부분 B의 퍼트맨투우통계량을 사용

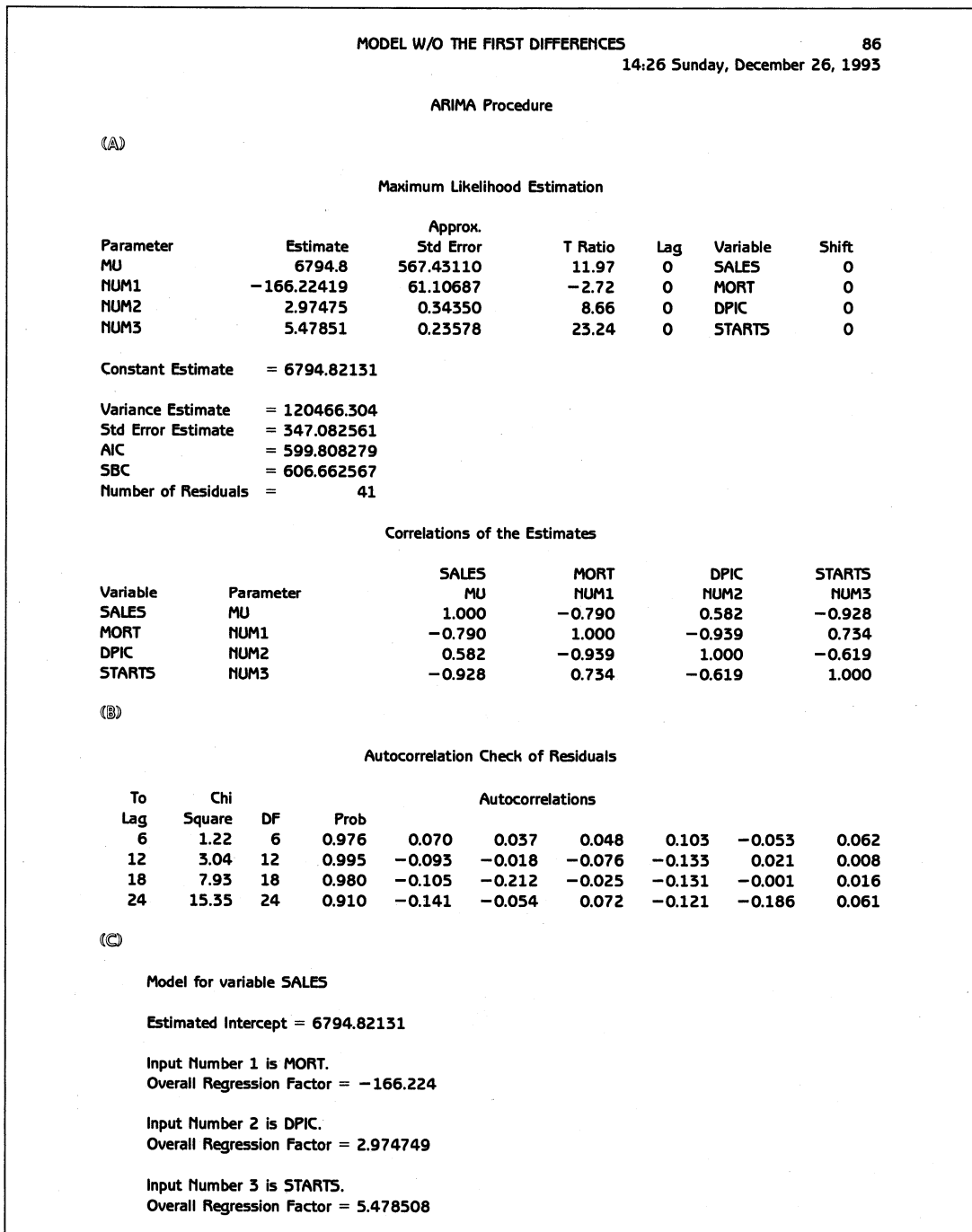


그림 5.19.1 설명변수들이 MORT, DPIC, STARTS인 최소제곱추정모형

해서 잔차들을 검정해야 한다. 이 퍼트맨토우통계량값들의 p 값들이 0.976, 0.995, 0.980, 0.910이므로 이 추정된 회귀모형의 잔차들이 백색잡음과정에서 나왔다고 할 수 있다. 또한, 잔차들의 표본자기상관함수, 표본역자기상관함수, 그리고 표본부분자기상관함수들로부터 잔차들이 백색잡음과정에서 나왔음을 재확인할 수 있다. 이로부터 잡음항 $\{v_t\}$ 가 단위근을 갖지 않는다는 사실, 즉 정상성을 갖는다는 것을 알 수 있다. 본 저자가 잡음항의 모형으로

AR(1)모형, AR(2)모형, MA(1)모형, MA(2)모형, ARMA(1, 1)모형을 갖는 회귀모형들을 이 시계열데이터에 적합시켜 보았으나, 잡음항의 모형으로 이들 중의 하나를 가정하는 회귀모형의 AIC값, BIC값, 그리고 퍼트맨토우통계량값들보다 잡음항을 백색잡음과정으로 가정하는 모형의 AIC값, BIC값, 그리고 퍼트맨토우통계량값들이 작다. 따라서 종속변수를 SALES로 하고 설명변수들을 MORT, DPIC, STARTS로 하는 회귀모형에서는 잡음항을 백색잡음과정이라고 가정하는 것이 좋다.

그림 5.18.2의 종속변수를 SALES(1)로 하고 설명변수들을 MORT(1), DPIC(1), STARTS(1)로 하고, 잡음항을 MA(1)과정이라 가정하는 추정된 회귀모형과 그림 5.19.1의 종속변수를 SALES로 하고 설명변수들을 MORT, DPIC, STARTS로 하고 잡음항을 백색잡음과정이라 가정하여 추정된 회귀모형을 비교하면 다음과 같다. 우선 전자는 수렴성이 보장되지 못하여 신빙성이 적다. 퍼트맨토우통계량값들의 관점에서 보면 후자의 잔차들이 전자의 잔차들보다 더 백색잡음과정에 가깝다. 두 모형들의 AIC값들과 BIC값들을 비교해보면 전자의 AIC값과 BIC값이 후자의 AIC값과 BIC값보다 작다. 그러나 두 모형들 사이의 AIC값들과 BIC값들의 차이는 작고 또한 이 값들을 계산하는 데 사용된 잔차들의 개수는 전자는 40이고 후자는 41이다. 따라서 AIC값과 BIC값만을 비교해서 두 모형들 사이의 우열을 가리기가 힘들다. 본 저자는 추정값들을 신뢰할 수 있는 후자, 즉 차분을 사용하지 않은 모형을 선호한다.

그림 5.19.1에서 설명변수들 MORT, DPIC, STARTS의 t 통계량값들이 모두 유의적으로 0과 다르다. 따라서 회귀모형에서 설명변수를 꼭 제거할 필요는 없다. 그러나 앞에서 수행한 차분시계열데이터의 분석결과와 비교해 보기 위해서 그림 5.19.1의 회귀모형에서 우선 설명변수 DPIC를 제외시키기로 한다. 이 PROC ARIMA의 두 번째 ESTIMATE스태이트먼트는 다음과 같다.

E INPUT=(MORT STARTS) METHOD=ML PLOT;

이 ESTIMATE스태이트먼트는 최우추정법을 사용해서 다음 모형을 추정하기 위한 것이다.

$$S_t = \beta_0 + \beta_1 M_t + \beta_3 H_t + v_t$$

여기서 $\{v_t\}$ 는 백색잡음과정이라고 가정한다. 이 첫 번째 ESTIMATE스태이트먼트를 수행한 결과가 그림 5.19.2이다. 부분 A에서 최우추정된 회귀모형이 다음과 같음을 알 수 있다.

$$S_t = 3933.4 - 330.793 M_t + 6.741 H_t + v_t$$

부분 B의 퍼트맨토우통계량값들의 p 값들을 보면 이 추정된 회귀모형의 잔차들이 백색잡음과정에서 나왔다고 할 수 없다. 부분 C의 잔차들의 표본자기상관함수는 시나브로형태를 보이고, 부분 D의 잔차들의 표본역자기상관함수는 시나브로형태 또는 시차 1 이후에서 절단형태를 보이며, 부분 E의 잔차들의 표본부분자기상관함수는 시나브로형태, 또는 시차 2 이후에서 절단형태를 보인다. 따라서 잡음항 $\{v_t\}$ 가 ARMA모형을 갖는다고 할 수 있다.

MODEL W/O THE FIRST DIFFERENCES

87

14:26 Sunday, December 26, 1993

ARIMA Procedure

(A)

Maximum Likelihood Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag	Variable	Shift
MU	3933.4	791.97229	4.97	0	SALES	0
NUM1	330.79271	36.02627	9.18	0	MORT	0
NUM2	6.74147	0.31806	21.20	0	STARTS	0

Constant Estimate = 3933.39573

Variance Estimate = 355057.496

Std Error Estimate = 595.867012

AIC = 643.218973

SBC = 648.359689

Number of Residuals = 41

Correlations of the Estimates

Variable	Parameter	SALES MU	MORT NUM1	STARTS NUM2
SALES	MU	1.000	-0.869	-0.889
MORT	NUM1	-0.869	1.000	0.568
STARTS	NUM2	-0.889	0.568	1.000

(B)

Autocorrelation Check of Residuals
Autocorrelations

To Lag	Chi Square	DF	Prob	0.361	0.293	0.188	0.011	-0.124	-0.208
6	14.19	6	0.028	0.361	0.293	0.188	0.011	-0.124	-0.208
12	34.18	12	0.001	-0.262	-0.258	-0.255	-0.344	-0.179	-0.104
18	41.73	18	0.001	-0.097	0.031	0.116	0.153	0.197	0.138
24	50.63	24	0.001	0.110	0.207	0.170	-0.022	-0.132	0.003

(C)

Autocorrelations plot of Residuals

Lag	Covariance	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
0	355057	1.00000												*****										
1	128112	0.36082												*****										
2	103953	0.29278												*****										
3	66731.884	0.18795												****										
4	4050.474	0.01141																						
5	-44099.744	-0.12420													**									
6	-73771.865	-0.20777													****									
7	-92994.040	-0.26191													****									
8	-91705.373	-0.25828													****									
9	-90570.013	-0.25509													****									
10	-122154	-0.34404													*****									
11	-63716.223	-0.17945													****									
12	-37071.352	-0.10441													**									
13	-34609.222	-0.09747													**									
14	11052.622	0.03113													*									
15	41314.519	0.11636													**									
16	54446.575	0.15335													***									
17	69959.643	0.19704													****									
18	49055.949	0.13816													***									
19	38891.057	0.10953													**									
20	73634.066	0.20739													****									
21	60525.403	0.17047													***									
22	-7876.777	-0.02218																						
23	-46996.765	-0.13236													***									
24	1220.563	0.00344																						

“.” marks two standard errors

(D)

Inverse Autocorrelations

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1		
1	-0.10296										**													
2	-0.14145										***													
3	-0.05729										*													
4	0.04772												*											
5	0.12142												**											
6	0.09117												**											
7	0.00969																							
8	0.01511																							
9	0.02383																							
10	0.14182												***											
11	-0.01798																							
12	-0.02778										*													
13	0.09308												**											
14	0.02159																							
15	-0.03158											*												
16	-0.03559										*													
17	-0.04488										*													
18	0.06905												*											
19	0.08978												**											
20	-0.10363											**												

(E)

Partial Autocorrelations

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1		
1	0.36082												*****											
2	0.18692												*****											
3	0.04123												*											
4	-0.12929										***													
5	-0.17096										***													
6	-0.14107										***													
7	-0.11290										**													
8	-0.06423										*													
9	-0.07786										**													
10	-0.23595									*****														
11	-0.01519																							
12	0.00529																							
13	-0.08769										**													
14	-0.00699																							
15	0.01710																							
16	0.00126																							
17	0.01745																							
18	-0.06963										*													
19	-0.05880										*													
20	0.12365												**											
21	0.12660												***											
22	-0.14843										***													
23	-0.25361									*****														
24	0.17425												***											

(F)

Model for variable SALES

Estimated Intercept = 3933.39573

Input Number 1 is MORT.
Overall Regression Factor = 330.7927

Input Number 2 is STARTS.
Overall Regression Factor = 6.741471

그림 5.19.2 설명변수들이 MORT와 STARTS인 최소제곱추정

본 저자가 잡음항의 모형으로 AR(1)모형, AR(2)모형, MA(1)모형, MA(2)모형, ARMA(1, 1)모형 등을 사용해서 주어진 시계열데이터에 적합시켜 본 결과 잡음항을 ARMA(1, 1)과정으로 가정하는 회귀모형의 AIC값, BIC값, 그리고 퍼트맨토우통계량값들이 잡음항의 모형으로 다른 모형을 가정하는 회귀모형의 AIC값, BIC값, 그리고 퍼트맨토우통계량값들보다 작다. 종속변수를 SALES로 하고 설명변수들을 MORT와 STARTS로 하는 회귀모형에서 잡음항이 ARMA(1, 1)과정이라는 가정하에 차분되지 않은 회귀모형을 추정하기 위해서 다음 세 번째 ESTIMATE스테이트먼트를 수행한다.

```
E P=1 Q=1 INPUT=(MORT STARTS)
METHOD=ML MAXIT=99;
RUN;
```

이 ESTIMATE스테이트먼트는 최우추정법을 사용해서 다음 모형을 추정하기 위한 것이다.

$$S_t = \beta_0 + \beta_1 M_t + \beta_3 H_t + \frac{1 - \theta_1 B}{1 - \phi_1 B} v_t$$

여기서 $\{v_t\}$ 는 백색잡음과정이라고 가정한다. 이 세 번째 ESTIMATE스테이트먼트를 수행한 결과가 그림 5.19.3이다. 부분 A에서 최우추정된 회귀모형이 다음과 같음을 알 수 있다.

$$S_t = 10980.5 - 48.636 M_t + 5.153 H_t + \frac{1 - 0.373 B}{1 - 0.990 B} v_t$$

부분 B의 퍼트맨토우통계량값들로부터 추정된 회귀모형의 잔차들이 백색잡음과정에서 나왔다고 할 수 있다. 그러나 이 퍼트맨토우통계량값들은 그림 5.19.1의 퍼트맨토우통계량값들보다 훨씬 크다. 즉, 퍼트맨토우통계량의 관점에서 본다면 MORT, DPIC, STARTS를 설명변수들로 사용하는 회귀모형이 MORT와 STARTS를 설명변수들로 사용하는 회귀모형보다 주어진 시계열데이터 HOUSING을 더 잘 반영한다. 또한 AIC와 BIC의 관점에서 같은 결론을 얻는다.

그림 5.18.3의 종속변수를 SALES(1)로 하고 설명변수들을 MORT(1)과 STARTS(1)로 하고, 잡음항을 MA(1)과정이라 가정하여 추정한 회귀모형과 그림 5.19.3의 종속변수를 SALES로, 설명변수들을 MORT와 STARTS로 하고 잡음항을 ARMA(1, 1)과정이라 가정하여 추정한 회귀모형을 비교하면 다음과 같다. 우선 전자는 수렴성이 보장되지 못하여 신빙성이 적다. 퍼트맨토우통계량값들의 관점에서 보면 전자의 잔차들이 후자의 잔차들보다 더 백색잡음과정에 가깝다. 두 모형들의 AIC값과 BIC값들을 비교해 보면 전자의 AIC값과 BIC값이 후자의 AIC값과 BIC값보다 작다. 그러나 이 값들을 계산하는 데 사용된 잔차들의 개수는 전자는 40이고 후자는 41이다.

그림 5.19.1에서 설명변수 MORT의 추정계수의 t 통계량값이 유의적으로 0과 같다. 따라서 그림 5.19.1의 회귀모형에서 우선 설명변수들 MORT와 DPIC를 제외시키기로 한다. 이 PROC ARIMA의 네 번째 ESTIMATE스테이트먼트는 다음과 같다.

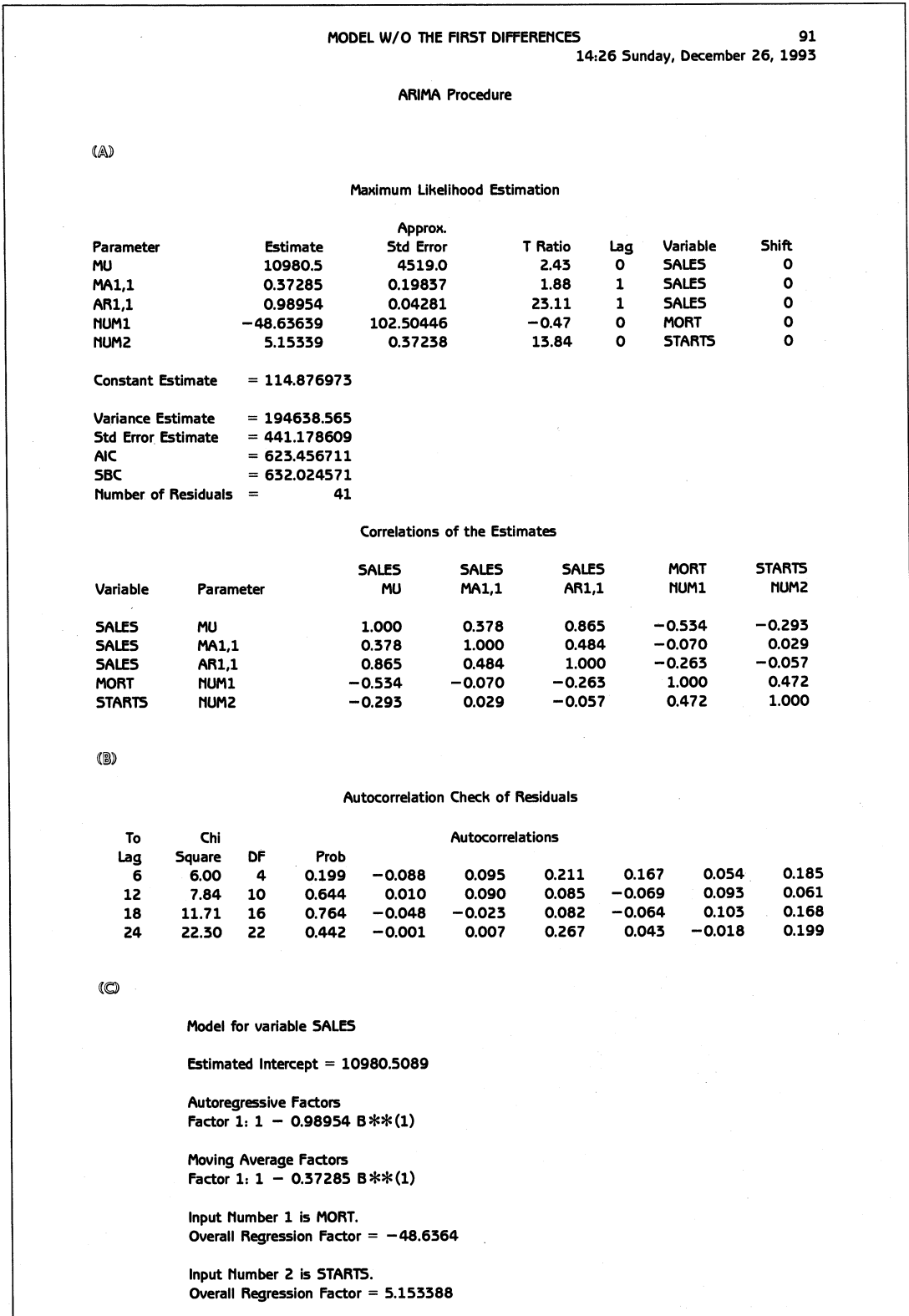


그림 5.19.3 설명변수들이 MORT, STARTS이고 잡음항이 ARMA(1, 1)과정인 모형

MODEL W/O THE FIRST DIFFERENCES

93

14:26 Sunday, December 26, 1993

ARIMA Procedure

(A)

Maximum Likelihood Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag	Variable	Shift
MU	10256.2	692.78513	14.80	0	SALES	0
NUM1	5.08381	0.46372	10.96	0	STARTS	0

Constant Estimate = 10256.1817

Variance Estimate = 1113504.61

Std Error Estimate = 1055.22728

AIC = 689.146471

SBC = 692.573616

Number of Residuals = 41

Correlations of the Estimates

Variable	Parameter	SALES MU	STARTS NUM1
SALES	MU	1.000	-0.971
STARTS	NUM1	-0.971	1.000

(B)

Autocorrelation Check of Residuals

To Lag	Chi Square	DF	Prob	Autocorrelations					
6	107.51	6	0.000	0.793	0.752	0.667	0.558	0.458	0.409
12	120.93	12	0.000	0.320	0.266	0.215	0.130	0.107	0.053
18	121.53	18	0.000	0.006	0.002	-0.018	-0.044	-0.037	-0.066
24	131.86	24	0.000	-0.091	-0.087	-0.088	-0.150	-0.177	-0.176

(C)

Autocorrelation Plot of Residuals

Lag	Covariance	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
0	1113505	1.00000												*****										
1	882982	0.79298												*****										
2	837875	0.75247												*****										
3	742801	0.66708												*****										
4	621752	0.55837												*****										
5	510039	0.45805												*****										
6	455272	0.40886												*****										
7	355898	0.31962												*****										
8	296084	0.26590												*****										
9	239482	0.21507												****										
10	145118	0.13033												***										
11	118736	0.10663												**										
12	58768.568	0.05278												*										
13	6303.113	0.00566																						
14	2658.420	0.00239																						
15	-20160.020	-0.01811																						
16	-49270.507	-0.04425											*											
17	-41158.488	-0.03696											*											
18	-73662.302	-0.06615											*											
19	-100993	-0.09070											**											
20	-96351	-0.08653											**											
21	-97999	-0.08801											**											
22	-166576	-0.14960											**											
23	-197021	-0.17694											**											
24	-196125	-0.17613											**											

"." marks two standard errors

(D)

Inverse Autocorrelations

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1		
1	-0.20624										****													
2	-0.26017									*****														
3	-0.11948									**														
4	-0.00576									.														
5	0.16139									.			***											
6	-0.04284									.	*													
7	-0.01081									.														
8	-0.01074									.														
9	-0.04865									.	*													
10	0.08671									.			**											
11	-0.06523									.	*													
12	-0.00427									.														
13	0.07605									.			**											
14	-0.02185									.														
15	-0.00348									.														
16	0.00917									.														
17	-0.06793									.	*													
18	0.01531									.			*											
19	0.04569									.			*											
20	-0.01422									.														

(E)

Partial Autocorrelations

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
1	0.79298											*****											
2	0.33314									.		*****											
3	0.01517									.													
4	-0.14452									.	***												
5	-0.11645									.	**												
6	0.08621									.		**											
7	-0.03111									.	*												
8	-0.01092									.													
9	-0.01318									.													
10	-0.12782									.	***												
11	0.05542									.		*											
12	-0.02219									.													
13	-0.03971									.	*												
14	0.06771									.		*											
15	0.00122									.													
16	-0.03272									.	*												
17	0.01045									.													
18	-0.06231									.	*												
19	-0.04675									.	*												
20	0.02167									.													
21	0.05699									.		*											
22	-0.19062									.	***												
23	-0.14618									.	***												
24	0.11142									.		**											

(F)

Model for variable SALES

Estimated Intercept = 10256.1817

Input Number 1 is STARTS.

Overall Regression Factor = 5.083806

그림 5.19.4 설명변수가 STARTS인 최소제곱추정

E INPUT=(STARTS) METHOD=ML PLOT;

이 ESTIMATE스테이트먼트는 최우추정법을 사용해서 다음 모형을 추정하기 위한 것이다.

$$S_t = \beta_0 + \beta_3 H_t + v_t$$

여기서 $\{v_t\}$ 는 백색잡음과정이라고 가정한다. 이 네 번째 ESTIMATE스테이트먼트를 수행한 결과가 그림 5.19.4이다. 부분 A에서 최우추정된 회귀모형이 다음과 같음을 알 수 있다.

$$S_t = 10256.2 + 5.084 H_t + v_t$$

부분 B의 퍼트맨토우통계량값들의 p 값들로부터 이 추정된 회귀모형의 잔차들이 백색잡음과정에서 나왔다고 할 수 없다. 부분 C의 잔차들의 표본자기상관함수는 시나브로형태를 보이고, 부분 D의 잔차들의 표본역자기상관함수는 시나브로형태 또는 시차 3 이후에서 절단형태를 보이며, 부분 E의 잔차들의 표본부분자기상관함수는 시나브로형태 또는 시차 3 이후에서 절단형태를 보인다. 따라서 잡음항 $\{v_t\}$ 가 AR(2)모형 또는 혼합ARMA모형을 갖는다고 할 수 있다.

본 저자가 잡음항의 모형으로 AR(1)모형, AR(2)모형, AR(3)모형, MA(1)모형, MA(2)모형, ARMA(1, 1)모형 등을 사용해서 주어진 시계열데이터에 적합시켜 본 결과 잡음항이 AR(2)모형을 갖는 회귀모형의 AIC값, BIC값, 그리고 퍼트맨토우통계량값들이 잡음항이 다른 모형을 갖는 회귀모형의 AIC값, BIC값, 그리고 퍼트맨토우통계량값들보다 작다. 종속변수를 SALES로 하고 설명변수를 STARTS로 하는 회귀모형에서 잡음항이 AR(2)과정이라는 가정하에 차분되지 않은 회귀모형을 추정하기 위해서 다음 다섯 번째 ESTIMATE스테이트먼트를 수행한다.

E P=2 INPUT=(STARTS) METHOD=ML MAXIT=99;
RUN;

이 ESTIMATE스테이트먼트는 최우추정법을 사용해서 다음 모형을 추정하기 위한 것이다.

$$S_t = \beta_0 + \beta_3 H_t + \frac{1}{1 - \phi_1 B - \phi_2 B^2} v_t$$

여기서 $\{v_t\}$ 는 백색잡음과정이라고 가정한다. 이 다섯 번째 ESTIMATE스테이트먼트를 수행한 결과가 그림 5.19.5이다. 부분 A에서 최우추정된 회귀모형이 다음과 같음을 알 수 있다.

$$S_t = 2651.7 + 0.309 H_t + \frac{1}{1 - 0.185 B - 0.189 B^2} v_t$$

부분 B의 퍼트맨토우통계량값들로부터 추정된 회귀모형의 잔차들이 백색잡음과정에서

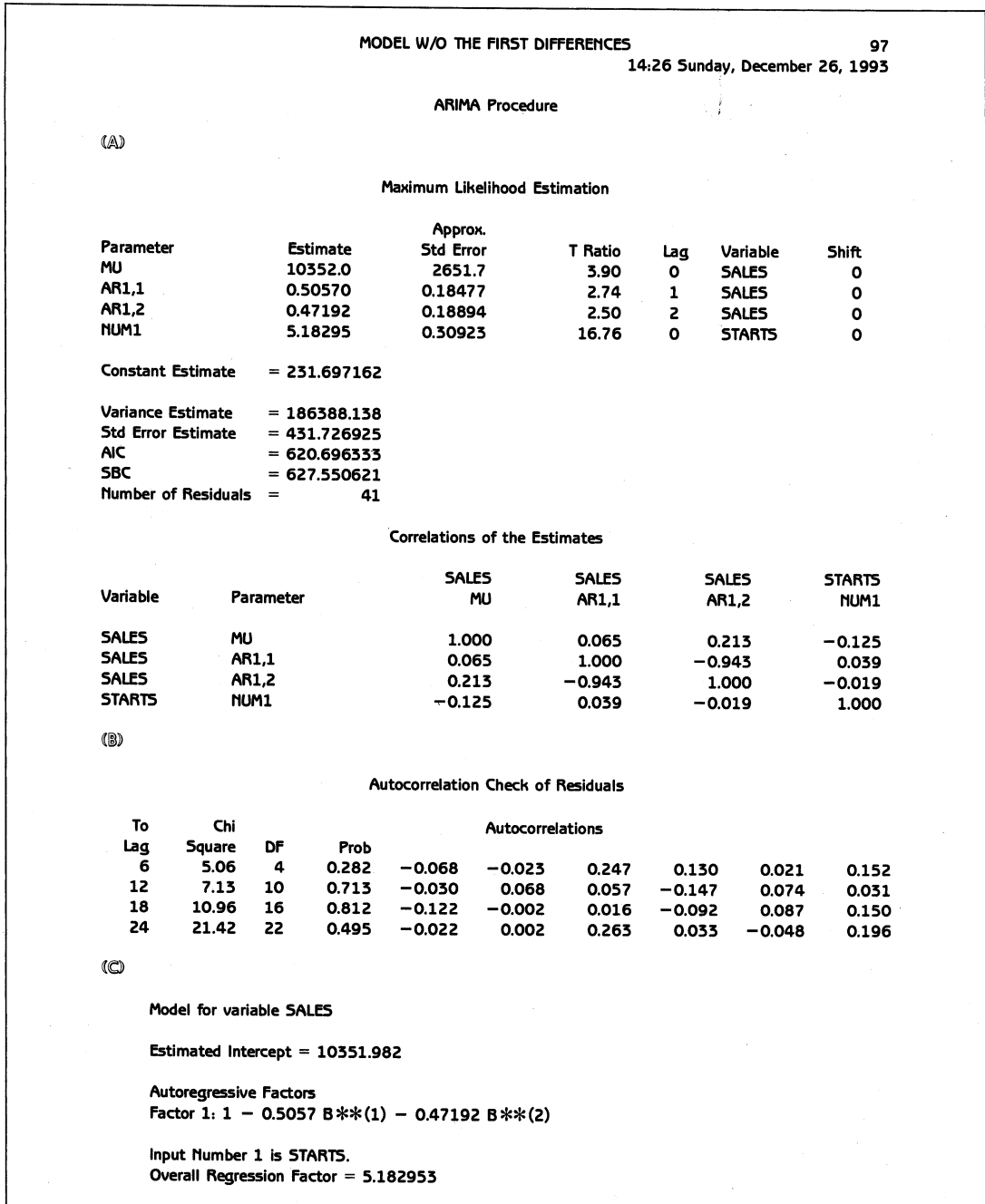


그림 5.19.5 설명변수가 STARTS이고 잡음항이 AR(2)과정인 회귀모형

나왔다고 할 수 있다. 그러나 이 퍼트맨토우통계량값들은 그림 5.19.1의 퍼트맨토우통계량 값들보다 훨씬 크다. 즉, 퍼트맨토우통계량의 관점에서 본다면 MORT, DPIC, STARTS 를 설명변수들로 사용하는 회귀모형이 STARTS를 설명변수로 사용하는 회귀모형보다 주어진 시계열데이터 HOUSING을 더 잘 반영한다. 또한 AIC와 BIC의 관점에서도 같은 결론을 얻는다.

그림 5.18.4의 종속변수를 SALES(1)로 하고, 설명변수를 STARTS(1)로 하고, 잡음항을

MA(1)과정이라 가정하여 추정된 회귀모형과 그림 5.19.5의 종속변수를 SALES로 하며, 설명변수를 STARTS로 하고, 잡음항을 AR(2)과정이라 가정하여 추정된 회귀모형을 비교하면 다음과 같다. 우선 전자는 수렴성이 반드시 보장된다고 할 수 없으므로 신빙성이 적다. 퍼트맨토우통계량값들의 관점에서 보면 전자의 잔차들이 후자의 잔차들보다 백색잡음 과정에 더 가깝다. 두 모형들의 AIC값과 BIC값들을 비교해 보면 전자의 AIC값과 BIC값이 후자의 AIC값과 BIC값보다 약간 작다. 그러나 이 값들을 계산하는 데 사용된 잔차값들의 개수는 전자는 40이고 후자는 41이다.

결론적으로 말해서 시계열데이터 HOUSING의 변수들을 차분해서 회귀분석하면 과다 차분문제가 발생하여 추정된 회귀식을 신뢰하기가 힘들다. 따라서 이 시계열데이터는 차분을 하지 않은 상태에서 분석을 하는 것이 좋다. 즉, 다음과 같은 그림 5.19.1의 추정모형이 주어진 시계열데이터에 가장 잘 부합된다.

$$S_t = 6794.8 - 166.224 M_t + 2.975 D_t + 5.497 H_t + v_t$$

단, $\{v_t\}$ 는 백색잡음과정

이 예제에서 시계열데이터의 변수들이 추세를 갖는 경우, 즉 비정상적인 경우라고 해서 반드시 차분을 취한 다음 모형화를 할 필요는 없다는 것을 알 수 있다. 이러한 경우는 지금부터 소개하고자 하는 공누적(cointegration)의 개념으로 설명할 수 있다.

한 가지 유의할 점은 이렇게 차분하지 않은 변수들 사이의 회귀식은 허위회귀(spurious regression)라는 문제를 일으킬 수도 있다는 것이다. 즉, 회귀분석에서 종속변수와 설명변수가 비정상시계열데이터들인 경우에 실제로는 이 종속변수와 설명변수가 서로 독립적임에도 불구하고 회귀분석을 하면 결정계수가 크고 회귀계수의 통계량값의 절대값이 유의적으로 큰 허위회귀문제를 발생시킬 수 있다. 앞서서도 보았듯이 이러한 경우에는 차분을 사용해서 시계열데이터들을 정상화시킨 다음에 회귀분석을 해야 한다. 물론 이 차분의 선행 조건은 시계열데이터가 비정상적인지를 검정하는 것인데, 이러한 문제를 단위근검정(unit root test)이라고 한다. 즉, 단위근검정이란 어떤 ARIMA모형의 AR특성식이 단위근 1을 갖는지 여부를 조사하는 것이다. 그러나 단위근검정에 의해서 종속변수와 설명변수로 사용되는 시계열데이터들이 단위근을 갖는 것으로 판정되더라도 항상 허위회귀가 일어나는 것은 아니다. 왜냐하면 이 시계열 데이터들이 비정상적이라 하더라도 이 시계열데이터들의 선형결합은 정상적인 시계열데이터일 수 있기 때문이다. 이러한 선형결합이 존재하면 이 시계열데이터들 사이에 공누적관계(cointegration)가 있다고 한다.

5.3.2 공누적의 개념

일반적인 경제시계열데이터는 장기기억(long-memory)확률과정과 단기기억(short memory)확률과정으로 나눌 수 있다. 단기기억시계열데이터에서는 아주 오래전에 일어난 충격(impulse, 또는 shock)이 현재의 관찰값에 미치는 영향이 거의 없으나 장기기억시계열데

이터에서는 오래된 충격이 현재값에 상당한 영향을 미친다. 시계열데이터 $\{v_t\}$ 를 다음과 같이 백색잡음과정 $\{v_t\}$ 의 선형결합으로 나타낼 수 있다고 하자.

$$x_t = \sum_{j=0}^{\infty} a_j v_{t-j}$$

단기기억시계열데이터에서는 시차 j 가 커질수록 충격반응가중값(impulse response weight) a_j 가 0으로 수렴하나 장기기억시계열데이터에서는 시차 j 가 커져도 a_j 가 0으로 수렴하지 않는다. 한 예로 단변량시계열데이터 $\{x_t\}$ 가 다음과 같은 AR(1)모형을 만족한다고 하자.

$$x_t = \phi x_{t-1} + v_t$$

여기서 $\{v_t\}$ 는 평균이 0이고 분산이 σ^2 인 백색잡음과정이다. 만일 ϕ 의 절대값이 1보다 작으면 다음 식이 성립한다.

$$x_t = \sum_{j=0}^{\infty} \phi^j v_{t-j}$$

따라서 충격반응가중값은 j 가 커짐에 따라 지수적으로 0으로 수렴한다. 따라서 정상적인 AR(1)과정은 단기기억시계열데이터이다. 그러나 만약 ϕ 가 1이라면 $\{x_t\}$ 는 다음 식을 만족하는 확률보행과정이다.

$$x_t = x_{t-1} + v_t$$

이 확률보행과정은 다음과 같이 쓸 수 있다.

$$x_t = \sum_{j=0}^{\infty} v_{t-j}$$

이 확률모형에서는 오래된 충격이나 새로운 충격이나 할 것 없이 현재의 관찰값 x_t 에 대해서 같은 충격반응가중값 1을 갖는다. 즉, 확률보행과정은 장기기억시계열데이터이다. 이 경우 x_t 의 차분은 다음 식을 만족한다.

$$\nabla x_t = v_t$$

따라서 확률보행과정의 차분시계열데이터 $\{\nabla x_t\}$ 는 단기기억시계열데이터이다.

장기기억시계열데이터 $\{x_t\}$ 를 한 번 차분하면 단기기억시계열데이터가 된다고 가정해 보자. 이 차분시계열데이터를 $\{z_t\}$ 로 표기하면 다음 식이 성립한다.

$$\nabla x_t = z_t$$

따라서 다음 식이 성립한다.

$$x_t = \sum_{j=0}^{\infty} z_{t-j}$$

여기서 x_t 는 z_t, z_{t-1}, \dots 들의 합으로 나타내어지므로 $\{x_t\}$ 를 누적시계열데이터(integrated

time series)라고 부른다. 이후 $\{x_t\}$ 를 한 번 차분한 차분시계열데이터가 단기기억시계열데이터가 되면 다음과 같이 표현한다.

$$x_t \sim I(1)$$

좀더 일반화해서 $\{x_t\}$ 를 d 번 차분해서 단기기억시계열데이터가 되면 다음과 같이 표현한다.

$$x_t \sim I(d)$$

정의에 의해서 단기기억시계열데이터는 $I(0)$ 형태이다.

시계열데이터가 $I(0)$ 형태이면 이 시계열데이터의 평균은 유한값 m 이고 분산도 유한값이다. 이 시계열데이터는 평균값과 자주 교차하려는 경향을 가지고 있으며 또한 평균에서 먼 값은 평균에 가까이 오려는 경향을 가지고 있다. 그러나 $I(1)$ 형태의 시계열데이터는 이러한 경향들을 갖지 않는다. 모든 $I(1)$ 형태의 시계열데이터가 꼭 만족하는 성질은 아니지만 일반적으로 $I(1)$ 형태의 시계열데이터는 충격이 멈춰지는 시점에서의 관찰값에 의해 결정되는 값으로 가까이 가는 성질이 있다. 즉, 시계열데이터가 $I(0)$ 형태이면 평균으로 돌아가려는 경향이 있는 반면에, $I(1)$ 형태이면 어떤 특정값으로 다시 돌아가는 경우가 거의 없다. 잘 알다시피 $I(0)$ 형태의 시계열데이터의 분산은 시간의 흐름에 따라 유한값으로 수렴한다. 한 예로 $\{x_t\}$ 가 AR계수가 ϕ 이고 백색잡음과정의 분산이 σ^2 인 AR(1)과정이라면 $\{x_t\}$ 의 분산은 다음과 같다.

$$\lim_{t \rightarrow \infty} \text{Var}(x_t) = \frac{\sigma^2}{1 - \phi^2}$$

만약, $I(1)$ 형태의 시계열데이터가 유한시간 전에 시작되었다면 기술상의 편의를 위하여 시작시점을 $t=0$ 으로 하자. 그러면 $I(1)$ 형태 시계열데이터의 분산은 시간의 흐름에 따라 선형적으로 증가한다. 예를 들어 확률보행과정 $\{x_t\}$ 의 분산은 다음과 같다.

$$\text{Var}(x_t) = t\sigma^2$$

지금까지 설명한 대로 $I(0)$ 형태의 시계열데이터와 $I(1)$ 형태의 시계열데이터는 근본적으로 다른 몇 가지 특성들을 갖는다. 많은 거시경제변수들이 $I(1)$ 형태로 나타낼 수 있는 장기기억시계열데이터이고 단지 약간의 거시경제변수들이 $I(0)$ 형태로 나타낼 수 있는 단기기억시계열데이터이기 때문에 거시경제학에서 시계열데이터들을 $I(0)$ 형태와 $I(1)$ 형태로 구분하는 것은 중요한 일이다.

5.3.3 공누적의 정의

시계열데이터들 $\{x_t\}$ 와 $\{y_t\}$ 가 $I(0)$ 형태 또는 $I(1)$ 형태인 경우 이 시계열데이터들의 선형결합들의 성질들을 살펴보자.

- ① 만일 $\{x_i\}$ 가 $I(0)$ 형태이면, $\{a+bx_i\}$ 는 $I(0)$ 형태이다. 만일 $\{x_i\}$ 가 $I(1)$ 형태이고 $b \neq 0$ 이면, $\{a+bx_i\}$ 는 $I(1)$ 형태이다.
- ② 만일 $\{x_i\}$ 와 $\{y_i\}$ 가 모두 $I(0)$ 형태이면, $\{ax_i+bx_i\}$ 는 $I(0)$ 형태이다.
- ③ 만일 $\{x_i\}$ 가 $I(1)$ 형태이고 $\{y_i\}$ 가 $I(0)$ 형태이며 $a \neq 0$ 이면, $\{ax_i+by_i\}$ 는 $I(1)$ 형태이다. 이는 $I(1)$ 형태가 $I(0)$ 형태에 비해 주도적(dominant)이라는 것을 의미한다.
- ④ 만일 $\{x_i\}$ 와 $\{y_i\}$ 가 모두 $I(1)$ 형태이면, $\{ax_i+by_i\}$ 는 일반적으로 $I(1)$ 형태이나 항상 그런 것은 아니고 $I(0)$ 형태인 경우도 있다.

네 번째 성질의 특수한 경우로 다음 정의를 할 수 있다.

정의 5.1 만일 시계열데이터들 $\{x_i\}$ 와 $\{y_i\}$ 가 모두 $I(1)$ 형태이지만 선형결합 $z_i = \mu + ax_i + by_i$ 가 존재해서 $\{z_i\}$ 가 $I(0)$ 형태이고 평균이 0이라면, $\{x_i\}$ 와 $\{y_i\}$ 는 공누적(co-integrated)되었다고 한다. ▣

공누적시계열데이터가 어떻게 발생할 수 있는지를 살펴보기 위해서 다음 시계열데이터들 $\{x_i\}$ 와 $\{y_i\}$ 를 살펴보자.

$$\begin{aligned} x_i &= \alpha w_i + \tilde{x}_i \\ y_i &= w_i + \tilde{y}_i \end{aligned}$$

만일, $\{w_i\}$ 는 $I(1)$ 형태의 시계열데이터이고 $\{\tilde{x}_i\}$ 와 $\{\tilde{y}_i\}$ 는 평균이 0인 $I(0)$ 형태의 시계열데이터들이라고 하면, $\{x_i\}$ 와 $\{y_i\}$ 는 $I(1)$ 형태의 시계열데이터들이다. 다음과 같은 시계열데이터 $\{z_i\}$ 를 살펴보자.

$$z_i = x_i - \alpha y_i = \tilde{x}_i - \alpha \tilde{y}_i$$

시계열데이터 $\{z_i\}$ 는 $I(0)$ 형태이며 평균은 0이다. 시계열데이터들 $\{x_i\}$ 와 $\{y_i\}$ 의 $I(1)$ 형태는 공통인 $I(1)$ 형태의 시계열데이터 $\{w_i\}$ 로부터 나왔고 따라서 $\{x_i\}$ 와 $\{y_i\}$ 는 공누적되었다. 이러한 성질이 항상 성립된다는 것을 증명할 수 있다. 즉, 한 쌍의 $I(1)$ 형태의 공누적 시계열데이터들은 항상 공통의 $I(1)$ 요인을 갖는다. 또한 이 성질의 역이 성립함을 쉽게 알 수 있다.

차분을 이용해서 비정상시계열데이터를 정상시계열데이터로 바꿀 수도 있다. 그러나 다변량시계열데이터인 경우에 각 변수의 차분차수를 그 변수의 성질에만 의존해서 결정하면 과다차분의 문제를 발생시킬 수 있다. 제 5.3.1 소절의 예제에서 보았듯이 이렇게 과다차분된 변수들로 구성된, 즉 단위근을 내재하는 회귀모형의 추정식은 통계적으로 열등하고 신뢰할 수가 없다. 반면에 이 다변량시계열데이터에 공누적의 개념을 도입해서 시계열모형을 구성하면 과다차분의 문제를 해결할 수 있다. 즉, 이러한 모형에 포함된 변수들은 $I(0)$ 형태를 만족하므로 정상성을 바탕으로 하는 일반적인 시계열분석방법들을 적용할 수 있다. 공누적에 관한 국내문헌으로는 최범수(1989), 정영옥(1993)과 한성신·서승환(1994)이 있다.

지금까지 공누적을 사용해서 모형화한 예제들을 살펴보았다. 그 중 하나는 제 5.1.2 소절의 첫 번째 예제로써 시계열데이터들 $\{x_t\}$ 와 $\{y_t\}$ 가 $I(1)$ 형태임에도 불구하고 차분을 이용하지 않고 공누적을 사용해서 다음 모형이 주어진 시계열데이터에 적합함을 보였다.

$$z_t = y_t - 24.482 - 0.982 x_t$$

$$(1 - 0.834 B - 0.773 B^4 - 0.618 B^5) z_t = v_t$$

단, $\{v_t\}$ 는 백색잡음과정

다른 예제는 제 5.3.1 소절의 예제로써 시계열데이터들 $\{S_t\}$, $\{M_t\}$, $\{D_t\}$ 와 $\{H_t\}$ 가 $I(1)$ 형태임에도 불구하고 차분을 이용하지 않고 공누적을 사용해서 다음 모형들이 시계열데이터 HOUSING에 적합함을 보였다.

$$S_t - 6794.8 + 166.224 M_t - 2.975 D_t - 5.497 H_t = v_{1t}$$

$$(1 - 0.990 B)(S_t - 10980.5 + 48.636 M_t - 5.153 H_t) = (1 - 0.373 B) v_{2t}$$

$$(1 - 0.185 B - 0.189 B^2)(S_t - 2651.7 - 0.309 H_t) = v_{3t}$$

단, $\{v_{1t}\}$, $\{v_{2t}\}$, $\{v_{3t}\}$ 는 백색잡음과정들

공누적을 설명하기 위해서는 오차수정모형(error-correction model)을 이해할 필요가 있다. 변수가 둘인 경우의 오차수정모형은 다음과 같다.

$$\nabla x_t = \mu_1 + \gamma_1 z_{t-1} + \sum_{j>0} (\beta_{11j} \nabla x_{t-j} + \beta_{12j} \nabla y_{t-j}) + v_{1t}$$

$$\nabla y_t = \mu_2 + \gamma_2 z_{t-1} + \sum_{j>0} (\beta_{21j} \nabla x_{t-j} + \beta_{22j} \nabla y_{t-j}) + v_{2t}$$

여기서 $z_t = x_t - \alpha y_t$, $\gamma_1 \neq 0$ 또는 $\gamma_2 \neq 0$, $(v_{1t}, v_{2t})'$ 는 이변량백색잡음과정이다. 만약 $\{x_t\}$ 와 $\{y_t\}$ 가 공누적되었고 공누적계수 α 가 올바르게 구해졌다면, $\{z_t\}$ 는 $I(0)$ 형태의 시계열데이터이다. 즉, 이 오차수정모형의 식들의 우변의 모든 항들이 $I(0)$ 형태이고 또한 좌변의 항들도 $I(0)$ 형태이다. 만일, $\{x_t\}$ 와 $\{y_t\}$ 가 $I(1)$ 형태이면서 공누적되지 않는다면, $\{z_t\}$ 가 $I(1)$ 형태이므로 이 오차수정모형은 $r_1 = r_2 = 0$ 인 경우에 한해서 성립한다. 따라서 $\{x_t\}$ 와 $\{y_t\}$ 가 공누적되어야 한다는 것은 오차수정모형이 성립하기 위한 필요조건이다. 또한 이 역도 성립함이 알려져 있다. 균형의 관점으로 이해하면 시계열데이터 $\{z_t\}$ 는 시스템이 균형에서 벗어난 정도를 나타내는 측도이므로 균형오차(equilibrium error)라고 부를 수 있다. 오차수정모형에서 x_t 와 y_t 의 다음 시점의 변화량들은 현재의 균형오차에 영향을 받을 것이다. 이것이 이 모형을 오차수정모형이라고 부르는 이유이다.

5.3.4 공누적시스템의 추정

이 소절에서는 두 개의 단변량시계열데이터들 $\{x_t\}$ 와 $\{y_t\}$ 가 공누적된다는 가정하에서 이 시계열데이터들을 2변량연립방정식시스템에 적합시키는 문제를 생각해 보자. 이 시스

템에서는 시계열데이터들 $\{x_t\}$ 와 $\{y_t\}$ 는 $I(1)$ 형태들이고 이들의 특별한 선형결합, 즉 공누적된 시계열데이터는 $I(0)$ 형태라는 특징을 갖는다고 가정한다.

Engle & Granger(1987)는 다음과 같이 한 변수 $\{y_t\}$ 를 다른 한 변수 $\{x_t\}$ 로 회귀시킨 식을 공누적회귀(cointegrating regression)라고 하였다.

$$y_t = \alpha x_t + z_t$$

여기서 $\{z_t\}$ 는 잡음항이다. 최소제곱추정법을 사용하여 이 오차항의 분산을 최소화했을 때 대표본에서 이 잡음항분산의 최소값이 유한(finite)이면 공누적시계열데이터 $\{z_t\}$ 는 $I(0)$ 형태이다. 즉, 누적된 시계열데이터는 대표본에서 잡음항의 분산이 무한(infinite)이므로, 공누적회귀모형에 최소제곱추정법을 적용하면 공누적계수를 추정할 수 있다. 또한 공누적계수 α 의 최소제곱추정량이 일치성을 갖는다는 것을 증명할 수 있다. 일반적으로 유한분산을 갖는 변수들의 선형회귀모형에서 어떤 회귀계수 β 의 최소제곱추정량 $\hat{\beta}$ 는 다음 성질을 만족한다.

$$\hat{\beta} = \beta + O_p(T^{-1/2})$$

Stock(1987)은 공누적계수 α 의 최소제곱추정량 $\hat{\alpha}$ 가 다음 성질을 만족한다는 것을 증명했다.

$$\hat{\alpha} = \alpha + O_p(T^{-1})$$

이러한 추정량은 초일치성(superconsistency)을 갖는다고 한다.

공누적시스템을 추정하는 한 가지 방법은 오차수정모형을 이용하는 것이다. 2변량벡터 (x_t, y_t) 가 차수가 $(p+1)$ 인 벡터자기회귀모형, 즉 VAR($p+1$)모형을 만족한다는 가정하에 이 확률벡터의 모형을 오차수정모형으로 표현하면 다음과 같다.

$$\begin{aligned} \nabla x_t &= \gamma_1(y_{t-1} - \alpha x_{t-1}) + \beta_{111} \nabla x_{t-1} + \beta_{112} \nabla y_{t-1} \\ &\quad + \dots + \beta_{1p1} \nabla x_{t-p} + \beta_{1p2} \nabla y_{t-p} + v_{1t} \\ \nabla y_t &= \gamma_2(y_{t-1} - \alpha x_{t-1}) + \beta_{211} \nabla x_{t-1} + \beta_{212} \nabla y_{t-1} \\ &\quad + \dots + \beta_{2p1} \nabla x_{t-p} + \beta_{2p2} \nabla y_{t-p} + v_{2t} \end{aligned}$$

여기서 (v_{1t}, v_{2t}) '는 2변량백색잡음과정이다. 가정에 의해서 이 오차수정모형의 좌변들은 $I(0)$ 형태의 차분변수들이다. 또한 이 오차수정모형의 우변들은 $I(0)$ 형태의 차분변수들과 오차수정항 $(y_{t-1} - \alpha x_{t-1})$ 이 더해져 있다. 만약 $\{x_t\}$ 와 $\{y_t\}$ 가 공누적계수 α 로 공누적된다면 $(y_{t-1} - \alpha x_{t-1})$ 역시 $I(0)$ 형태이므로 각 방정식의 우변은 $I(0)$ 형태이다. 이는 공누적되는 확률변수들을 공누적계수가 α 인 오차수정모형으로 나타낼 수 있고, 역으로 오차수정모형은 공누적표현을 할 수 있다는 것을 뜻한다. Engle & Granger(1987)는 이 성질을 증명하고 Granger의 표현정리(representation theorem)라고 명명했다.

오차수정모형은 벡터모형에 비선형의 제약이 포함된 것이므로 최우추정법을 적용하려면 반복적인 절차가 필요하다. 그러나 만약 α 의 초일치추정량(superconsistent estimate) $\hat{\alpha}$

이 주어진다면, 오차수정모형에 모수 α 대신 이 추정량 $\hat{\alpha}$ 를 대입한 뒤, 최소제곱추정법을 사용해서 나머지 계수들 β 와 γ 를 추정할 수 있다. 이것이 Engle-Granger의 2단계추정법이다. 요약하면 이 2단계추정법의 제1단계는 공누적회귀식에서 α 를 최소제곱추정한 뒤 이 회귀식의 잔차값들 $(y_t - \hat{\alpha}x_t)$ 를 구한다. 제2단계는 오차수정모형에 이 잔차값들을 대입해서 β 들과 γ 들을 추정한다. 이 Engle-Granger의 2단계추정량들은 매우 좋은 성질들을 갖는다. 왜냐하면 이 2단계추정법은 잡음항분산의 최소값을 갖는 공누적계수를 구하는 방법이고, 이 공누적계수 이외의 값들은 공누적회귀식의 잡음항의 분산이 무한대가 되도록 하기 때문이다. 또한 공누적계수의 추정량을 제외한 오차수정모형의 계수들의 추정량들은 일치추정량들이고 최우추정량들과 같은 점근분포를 갖는다.

Engle-Granger의 2단계추정법을 다음과 같은 2변량시계열데이터 $\{(x_t, y_t)'\}$ 의 예를 통해서 살펴보자.

$$\begin{aligned} x_t + \phi y_t &= n_{1t}, & n_{1t} &= n_{1,t-1} + v_{1t} \\ x_t + \pi y_t &= n_{2t}, & n_{2t} &= \rho n_{2,t-1} + v_{2t} \quad (|\rho| < 1) \end{aligned}$$

여기서 $\{(v_{1t}, v_{2t})'\}$ 는 2변량백색잡음과정으로서 v_{1t} 와 v_{2t} 는 서로 독립이 아니다. 이 연립방정식시스템을 축소형(reduced form)으로 나타내면 다음과 같다.

$$\begin{aligned} x_t &= \frac{-\pi n_{1t} + \phi n_{2t}}{\phi - \pi} \\ y_t &= \frac{n_{1t} - n_{2t}}{\phi - \pi} \end{aligned}$$

잡음항 $\{n_{1t}\}$ 가 $I(1)$ 형태이므로 $\{x_t\}$ 와 $\{y_t\}$ 도 $I(1)$ 형태이다. 반면에 $\{n_{2t}\}$ 는 $I(0)$ 형태이므로 $\{x_t + \pi y_t\}$ 는 정상적이다. 즉, $\{x_t\}$ 와 $\{y_t\}$ 는 공누적된다. 이 연립방정식의 모수들을 추정하는 것은 간단하다. 두 번째 방정식에서 y_t 를 설명변수로, x_t 를 종속변수로 하는 선형모형에 최소제곱추정법을 적용해서 π 의 일치추정량 $\hat{\pi}$ 을 구할 수 있다. 이 회귀식이 공누적회귀식이다. 즉, x_t 와 y_t 의 선형결합들 중에서 $x_t + \hat{\pi}y_t$ 이외의 선형결합의 분산은 무한대값(∞)을 갖는다. 또한 y_t 를 종속변수로, x_t 를 독립변수로 하는 선형모형에 최소제곱추정법을 적용해서 $1/\pi$ 에 대한 일치추정량을 구할 수 있다. 일단 모수 π 가 추정되면 이 추정량이 주어졌다는 조건하에 오차수정모형을 이용해서 다른 모수들을 추정할 수 있다. 이 연립방정식시스템으로부터 다음과 같은 AR모형들을 구할 수 있다.

$$\begin{aligned} \nabla x_t &= \phi \delta x_{t-1} + \pi \phi \delta y_{t-1} + a_{1t} \\ \nabla y_t &= -\delta x_{t-1} - \pi \delta y_{t-1} + a_{2t} \end{aligned}$$

단, $\delta = \frac{1 - \rho}{\pi - \phi}$

여기서 a_{1t} 와 a_{2t} 는 v_{1t} 와 v_{2t} 의 선형결합들이다. 따라서 오차수정모형은 다음과 같다.

$$\nabla x_t = \phi \delta z_{t-1} + a_{1t}$$

$$\nabla y_t = -\delta z_{t-1} + a_{2t}$$

$$\text{단, } z_t = x_t + \pi y_t$$

이 오차수정모형에 π 대신 π 의 초일치추정량을 대입하면 이 오차수정모형은 두 개의 미지수들을 갖는 두 개의 방정식들로 구성되므로 최소제곱추정법을 사용해서 ϕ 와 δ 의 일치추정량들을 쉽게 구할 수 있다. 이 추정량들이 Engle-Granger의 2단계추정량이다. 반면에 $\{x_t\}$ 와 $\{y_t\}$ 를 각각 한 번 차분한 차분시계열데이터들을 사용해서 회귀분석을 하면 모수들의 일치추정량들을 얻지 못한다.

Engle & Granger의 2단계추정법은 다음 두 가지 점에서 비판을 받고 있다. Banerjee & Dolado & Hendry & Smith(1986)에 의해서 지적되었듯이 최소제곱추정법에 의해 얻어진 공누적계수의 추정량은 유한표본(finite sample)에서는 빈약한 성질을 갖고 있다는 점이다. 또한 일부에서는 공누적을 파악하는 자체에 어려움이 존재하므로 제약이 없는 벡터자기회귀모형(vector autoregressive model, VAR모형)을 사용하는 것이 안전하다고 지적하고 있다. 이러한 비판을 검정하기 위해서 Engle & Yoo(1987)는 모의실험을 수행하였다. 이 모의실험에서는 2개의 변수들로 구성된 공누적이 있는 연립방정식시스템에서 발생시킨 모의시계열데이터들에 다음 두 가지 방법을 적용해서 20번째 미래시점들까지의 전방다단계예측값들을 구했다. 첫 번째 예측방법은 공누적제약을 가하지 않은 수준변수로 구성된 VAR모형을 사용한 것이고 두 번째 예측방법은 Engle-Granger의 2단계추정법을 사용한 것이다. 이 모의실험의 결과 단기에서는 VAR모형을 사용한 예측오차가 Engle-Granger의 2단계추정법에 의한 예측오차보다 적었으나, 장기로 갈수록 Engle-Granger의 2단계추정법에 의한 예측이 VAR모형을 사용한 예측보다 상당히 정확하였다. 변수들 사이에 공누적관계가 있음에도 불구하고 이를 무시하고 VAR모형을 추정한 경우의 추정량들의 점근분포가 Sims & Stock & Watson(1990)에 의해서 밝혀졌다.

Engle-Granger의 2단계추정법에 의한 추정량들 중에서 공누적계수의 추정량을 제외한 나머지 추정량들은 점근적으로 효율적이다. 그러나 Johansen & Juselius(1988)와 Phillips(1988)는 공누적계수의 최우추정량의 분포이론이 간단한 반면에 Engle-Granger 2단계추정량은 그렇지 못하다는 것을 밝혀냈다. 이러한 문제점을 해결하기 위해서 Ahn & Reinsel(1990), Engle & Yoo(1990), Johansen(1991), Phillips(1991), Park(1992), Stock & Watson(1993) 등이 여러 가지 방법을 제시하였다. 특히 Engle & Yoo(1990)는 Engle-Granger의 2단계추정법을 변형한 3단계추정법을 제시하였다. 이 3단계추정법을 사용하면 점근적으로 완전정보최우추정량(full information maximum likelihood estimator, FIML estimator)과 동등한 공누적계수의 추정량을 구할 수 있다.

지금부터는 Engle-Yoo의 3단계추정법으로 오차수정모형의 모수들을 추정하는 예를 살펴보자. 앞에서 2변량벡터 $x_t = (x_{1t}, x_{2t})'$ 가 차수가 $(p+1)$ 인 벡터자기회귀모형, 즉, VAR $(p+1)$ 모형을 만족한다는 가정하에 오차수정모형을 다음과 같이 정의했다.

$$\nabla x_{1,t} = \gamma_1(x_{2,t-1} - \alpha x_{1,t-1}) + \beta_{111} \nabla x_{1,t-1} + \beta_{112} \nabla x_{2,t-1}$$

$$+ \dots + \beta_{1p1} \nabla x_{1,t-p} + \beta_{1p2} \nabla x_{2,t-p} + v_{1t}$$

$$\begin{aligned} \nabla x_{2t} &= \gamma_2(x_{2,t-1} - \alpha x_{1,t-1}) + \beta_{211} \nabla x_{1,t-1} + \beta_{212} \nabla x_{2,t-1} \\ &\quad + \dots + \beta_{2p1} \nabla x_{1,t-p} + \beta_{2p2} \nabla x_{2,t-p} + v_{2t} \end{aligned}$$

여기서 $v_t = (v_{1t}, v_{2t})'$ 는 평균벡터가 영벡터이고 공분산행렬이 다음과 같은 2변량정규백색잡음과정이다.

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

이 오차수정모형을 벡터형태로 표현하면 다음과 같다.

$$A^*(B) \nabla x_t = -\gamma z_{t-1} + v_t$$

$$\text{단, } A^*(B) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \sum_{i=1}^p \begin{bmatrix} \beta_{1i1} & \beta_{1i2} \\ \beta_{2i1} & \beta_{2i2} \end{bmatrix} B^i$$

$$x_t = \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix}, \quad \nabla x_t = \begin{bmatrix} \nabla x_{1t} \\ \nabla x_{2t} \end{bmatrix}, \quad v_t = \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix}, \quad \gamma_t = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}$$

$$z_t = \alpha x_{1t} - x_{2t} = (\alpha, -1)' x_t$$

관찰점들의 개수가 T 라면 이 오차수정모형의 상수항을 생략한 근사적인 대수우도함수는 다음과 같다.

$$\begin{aligned} L &\simeq \sum_{t=1}^T L_t(\theta; \nabla x_t) \\ &= \sum_{t=1}^T L_t(\theta_1, \theta_2; \nabla x_t) \\ &= -\frac{T}{2} \det(\Sigma) - \frac{1}{2} \sum_{t=1}^T v_t' \Sigma^{-1} v_t \end{aligned}$$

여기서 $\theta = (\theta_1, \theta_2)'$ 이고, θ_1 은 모수들 β, γ, Ω 를 나타내며 θ_2 는 공누적계수 α 이다. 이 대수우도함수를 한 번 미분한 스코어함수(score function) $S(\theta)$ 를 Engle-Granger의 2단계 추정벡터 $\tilde{\theta}$ 를 중심으로 Taylor 급수로 전개하면 다음과 같은 근사식을 얻을 수 있다.

$$\frac{\partial L}{\partial \theta} = S(\theta) \simeq S(\tilde{\theta}) + \frac{\partial S(\tilde{\theta})}{\partial \theta} (\theta - \tilde{\theta})$$

이후 최우추정벡터를 $\hat{\theta}$ 로 표기하면 $S(\hat{\theta})=0$ 이므로 다음 근사식이 성립한다.

$$\hat{\theta} \simeq \tilde{\theta} - \left(\frac{\partial S(\tilde{\theta})}{\partial \theta} \right)^{-1} S(\tilde{\theta})$$

따라서 먼저 Engle-Granger의 2단계추정벡터 $\tilde{\theta}$ 를 구한 뒤, 다음과 같은 식을 사용해서

근사최우추정량 $\check{\theta}$ 를 구한다.

$$\check{\theta} = \tilde{\theta} + Q^{-1}(\theta) \frac{\partial L}{\partial \theta}$$

$$\text{단, } Q = \sum_{t=1}^T E \left[\left(\frac{\partial L_t}{\partial \theta} \right) \left(\frac{\partial L_t}{\partial \theta} \right)' \right]$$

여기서 Q 는 정보행렬(information matrix)이다. 이 정보행렬을 다음과 같이 분할하자.

$$Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}$$

여기서 Q_{11} 은 $(4p+4) \times (4p+4)$ 행렬로서 $\check{\theta}_1$ 의 분산행렬에 해당하며, Q_{22} 는 1×1 행렬, 즉 스칼라로서 $\check{\theta}_2$ 의 분산에 해당하고, Q_{12} 는 $(4p+4) \times 1$ 행렬로서 $\check{\theta}_1$ 과 $\check{\theta}_2$ 의 공분산행렬에 해당한다. 이 정보행렬 Q 는 블록대각행렬(block diagonal matrix)임이 알려져 있다. 따라서 θ_2 의 일치추정량이 주어진 상태에서 구한 근사최우추정벡터 $\check{\theta}_1$ 은 완전정보최우추정벡터 $\hat{\theta}_1$ 과 점근적으로 동등하다. 즉, 공누적계수 θ_2 의 일치추정량이 주어졌을 때 근사최우추정벡터 $\check{\theta}_1$ 은 점근적 유효성(asymptotic efficiency)을 갖는다. 이렇게 구한 근사최우추정벡터 $\check{\theta}_1$ 은 Engle-Granger의 2단계추정법에서 구한 추정량과 동일하다. 따라서 공누적계수의 추정량을 제외한 Engle-Granger의 2단계추정량들은 점근적 유효성을 지님을 알 수 있다. 이 방법은 공누적계수의 추정에도 똑같이 적용할 수 있다. 즉, 근사최우추정벡터 $\check{\theta}_1$ 을 이용해서 공누적계수의 근사최우추정량 $\check{\theta}_2$ 를 구할 수 있다. 이 추정량은 점근적 유효추정량이다. 이것이 Engle-Yoo의 3단계추정법의 제3단계이다. 즉, Engle-Yoo의 3단계추정법은 Engle-Granger 2단계추정법에서 한번 더 공누적계수를 최소제곱추정법으로 추정하는 단계를 추가한 것이다. 이 Engle-Yoo의 3단계추정법의 제3단계의 공누적계수의 추정량은 다음 식에서 구해진다.

$$\check{\theta}_2 \simeq \tilde{\theta}_2 + Q_{22}^{-1} \frac{\partial L}{\partial \theta_2}$$

$$\text{단, } Q_{22} = \sum_{t=1}^T E \left[\left(\frac{\partial L_t}{\partial \theta_2} \right)^2 \right]$$

여기서 구해진 근사최우추정량 $\check{\theta}_2$ 는 평균이 θ_2 이고 분산이 Q_{22}^{-1} 인 점근적 정규분포를 갖는다. 이와 같이 공누적계수의 근사최우추정량이 명확한 점근분포를 가지게 되므로 Sims & Stock & Watson(1990)에 의해서 지적된 Engle-Granger의 2단계추정법의 문제점을 Engle-Yoo의 3단계추정법은 가지고 있지 않다.

Engle-Yoo의 3단계추정법의 실례를 살펴보기 위해서, 앞에서 다룬 2변량시계열데이터 $\{(x_{1t}, x_{2t})\}$ 의 오차수정모형을 살펴보자.

$$A^*(B) \nabla x_t = -\gamma z_{t-1} + v_t$$

다음과 같은 행렬들과 다항식을 정의하자.

$$C := \begin{bmatrix} 1 & -\frac{\sigma_{12}}{\sigma_2^2} \\ 0 & 1 \end{bmatrix}$$

$$\gamma_c := C\gamma$$

$$v_{c_t} := Cv_t$$

$$A_c^*(B) := CA^*(B)$$

따라서 다음 식들이 성립한다.

$$A_c^*(0) = C = \begin{bmatrix} 1 & -\frac{\sigma_{12}}{\sigma_2^2} \\ 0 & 1 \end{bmatrix}$$

$$\text{Var}(v_{c_t}) = \begin{bmatrix} \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2} & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

$$\gamma_c = \begin{bmatrix} \gamma_1 - \frac{\sigma_{12}}{\sigma_2^2} \gamma_2 \\ \gamma_2 \end{bmatrix}$$

지금부터는 행렬 $A_c^*(B)$ 가 상삼각행렬(upper triangular matrix)이고, $\gamma_2 = 0$ 인 경우를 살펴보자. 즉 다음과 같은 오차수정모형을 추정하기로 하자.

$$\begin{aligned} \nabla x_{1t} &= -\gamma(x_{1,t-1} - \alpha x_{2,t-1}) + \beta_{102} \nabla x_{2t} + \beta_{111} x_{1,t-1} \\ &\quad \beta_{112} \nabla x_{2,t-1} + \dots + \beta_{1p1} \nabla x_{1,t-p} + \beta_{1p2} \nabla x_{2,t-p} + v_{1t} \end{aligned}$$

$$\nabla x_{2t} = \beta_{212} \nabla x_{2,t-1} + \beta_{222} \nabla x_{2,t-1} + \dots + \beta_{2p2} \nabla x_{2,t-p} + v_{2t}$$

$$\text{단, } \Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

공누적시계열데이터 $\{z_t\}$ 는 다음과 같이 정의되었다.

$$z_t = x_{1t} - \alpha x_{2t} = (1, \alpha)' x_t$$

Engle & Hendry & Richard(1983)는 확률벡터 w 를 주어진 조건으로 해서 모수벡터를 추정할 때 아무런 정보의 손실이 없다면 확률벡터 w 가 모수를 추정하는 데 약의생성(weak exogeneity)을 지닌다고 정의했다. 이 오차수정모형에서는 변수 x_2 는 모수 α 를 추정하는 데 있어서는 약의생변수이다. 즉, $\{\nabla x_{2t}\}$ 와 $\{v_{1t}\}$ 는 비상관이므로 다음 식들이 성립한다.

$$L_t = -\frac{1}{2} \log \sigma_1^2 - \frac{1}{2\sigma_1^2} v_{1t}^2$$

$$\frac{\partial L_t}{\partial \theta_2} = \frac{\partial L_t}{\partial \alpha} = \frac{\gamma}{\sigma_1^2} v_{1t} x_{2,t-1}$$

따라서 Engle-Yoo의 3단계추정법의 제3단계는 다음 식들을 사용해서 구한다.

$$\tilde{\alpha} = \tilde{\alpha} + \frac{\sum_t x_{2,t-1} \tilde{v}_{1t}}{\tilde{\gamma} \sum_t x_{2,t-1}^2}$$

$$Q_{22} = \frac{1}{\sigma_1^2} \left(\tilde{\gamma}^2 \sum_t x_{2,t-1}^2 \right)$$

좀더 일반적으로 변수 x_2 가 약외생성을 지니지 않은 경우를 살펴보자. 이 경우에도 공분산행렬 Σ 는 대각행렬이고 행렬 $A^*(0)$ 가 상삼각행렬이라고 가정하자. 이 오차수정모형의 대수우도함수는 다음과 같다.

$$L = -\frac{1}{2} \sum_{i=1}^2 \sum_t \frac{v_{it}^2}{\sigma_i^2} - \frac{T}{2} \sum_{i=1}^2 \log \sigma_i^2$$

대수우도함수 L 을 α 에 대해 미분하면 다음과 같다.

$$\frac{\partial L}{\partial \alpha} = \sum_{i=1}^2 \sum_t \frac{1}{\sigma_i^2} v_{it} \gamma_i x_{2,t-1}$$

다음 식처럼 가중잔차값들(weighted residual)을 쌓으면(stacking) Engle-Yoo의 3단계추정법의 제3단계를 쉽게 수행할 수 있다.

$$\tilde{v} = \left(\frac{\tilde{v}_{11}}{\tilde{\sigma}_1}, \frac{\tilde{v}_{12}}{\tilde{\sigma}_1}, \dots, \frac{\tilde{v}_{1T}}{\tilde{\sigma}_1}, \frac{\tilde{v}_{21}}{\tilde{\sigma}_2}, \frac{\tilde{v}_{22}}{\tilde{\sigma}_2}, \dots, \frac{\tilde{v}_{2T}}{\tilde{\sigma}_2} \right)'$$

$$\text{단, } \tilde{\sigma}_i^2 = \frac{1}{T} \sum_{t=1}^T \tilde{v}_{it}^2 \quad (i=1, 2)$$

벡터 \tilde{v} 에서 각 잔차항을 σ_i 의 추정량 $\tilde{\sigma}_i$ 로 나누어주는 것은 분산이질성을 고려한 것이다. 벡터 \tilde{v} 를 분산이질성이 조정된 잔차(heteroscedasticity corrected residual)라고 한다. 마찬가지로 $\tilde{\gamma}_i x_{2,t-1}$ 를 σ_i 의 추정량 $\tilde{\sigma}_i$ 로 나눈 $(2T) \times 1$ 행렬을 다음과 같이 정의한다.

$$X = \left(\frac{\tilde{\gamma}_1 x_{2,0}}{\tilde{\sigma}_1}, \dots, \frac{\tilde{\gamma}_1 x_{2,T-1}}{\tilde{\sigma}_1}, \frac{\tilde{\gamma}_2 x_{2,0}}{\tilde{\sigma}_2}, \dots, \frac{\tilde{\gamma}_2 x_{2,T-1}}{\tilde{\sigma}_2} \right)'$$

이 벡터들을 이용하면 스코어함수와 정보행렬을 다음과 같이 쓸 수 있다.

$$\frac{\partial L}{\partial \alpha} = X' v$$

$$Q_{22} = X' X$$

따라서 공누적계수 α 의 근사최우추정량 $\check{\alpha}$ 는 다음과 같다.

$$\check{\alpha} = \tilde{\alpha} + (X'X)^{-1} X'v$$

이 식에서 Engle-Yoo의 3단계추정법의 공누적계수의 근사최우추정량 $\check{\alpha}$ 는 두 추정량들의 합임을 알 수 있다. 이 추정량들 중 하나는 Engle-Granger의 2단계추정법의 공누적추정량 $\tilde{\alpha}$ 이고 다른 하나는 Engle-Granger의 2단계추정법에 의한 잔차의 분산이질성을 조정된 잔차를 분산이질성이 조정된 설명변수들에 회귀시켜서 얻은 추정량이다. 이 근사최우추정량 $\check{\alpha}$ 는 평균이 α 이고 분산이 $1/(X'X)$ 인 점근정규분포를 갖는다.

▣ 참고문헌 ▣

- 김동원(1992), 「국면조정모형의 일반화에 대한 연구」, 석사논문, 연세대학교대학원, 서울.
- 정영욱(1993), 「주성분분석과 일반화최소제곱추정법을 이용한 공누적모형의 추정」, 석사논문, 연세대학교대학원, 서울.
- 최범수(1989), 단위근과 공누적의 경제학적 의미와 그 검정법에 대한 개요, 「계간 한국개발연구」, 제11권, pp. 119~143.
- 한성신 · 서승환(1994), 「한국경제의 계량분석」, 서울종합경제연구소, 서울.
- Abraham, B. (1980), Intervention analysis and multiple time series, *Biometrika*, vol. 67, pp. 73~78.
- Abraham, B. (1987), Application of intervention analysis to road fatality series in Ontario, *J. Forecasting*, vol. 6, pp. 211~219.
- Ahn, S. K. & G. C. Reinsel(1990). Estimation for partially nonstationary multivariate autoregressive models, *J. Amer. Statist. Assoc.*, vol. 85, pp. 813~823.
- Banerjee, A., J. J. Dolado, D. F. Hendry & G. W. Smith(1986), Exploring equilibrium relationships in economics through static models: Some Monte Carlo evidence, *Oxford Bulletin of Economics and Statistics*, vol. 48, pp. 253~277.
- Bhattacharyaa, M. N. & A. P. Layton(1979), Effectiveness of seat belt legislation on the Queensland road toll-an Australian case study in intervention analysis, *J. Amer. Statist. Assoc.*, vol. 74, pp. 596~603.
- Birely, L. E. (1987), Application of intervention analysis to power plant monitoring data, *SUGI-SAS Users Group Conference Proceedings*, vol. 12, pp. 340~348.
- Box, G. E. P. & G. M. Jenkins(1976), *Time Series Analysis: Forecasting and*

- Control*, Holden-Day, New York.
- Box, G. E. P. & G. C. Tiao(1965), A change in level of a nonstationary time series, *Biometrika*, vol. 52, pp. 181~192.
- Box, G. E. P. & G. C. Tiao(1975), Intervention analysis with applications to economic and environmental problems, *J. Amer. Statist. Assoc.*, vol. 70, pp. 70~79.
- Capie, F. H., T. C. Mills & G. E. Wood(1986), Debt management and interest rates: The British stock conversion of 1932, *Applied Economics*, vol. 18, pp. 1111~1126.
- Engle, R. F. & C. Granger(1987), Cointegration and error correction: Representation, estimation, and testing, *Econometrica*, vol. 55, pp. 251~276.
- Engle, R. F., D. F. Hendry & J. F. Richard(1983), Exogeneity, *Econometrica*, vol. 51, pp. 277~304.
- Engle, R. F. & S. Yoo(1987), Forecasting and testing in cointegrated systems, *J. Econometrics*, vol. 35, pp. 143~159.
- Engle, R. F. & S. Yoo(1990), Cointegrated economic time series: an overview with new results, in *Long-Run Economic Relationships*, R. F. Engle & C. Granger, Eds., Oxford Univ. Press, New York, pp. 237~266.
- Glass, G. V. (1972), Estimating the effects of intervention into a nonstationary time series, *Amer. Educational Research J.*, vol. 9, pp. 463~477.
- Greene, W. H. (1991), *Econometric Analysis*, Macmillian Pub. Co., New York.
- Grunfeld, Y. (1958), The determination of corporate investment, Ph. D. Dissertation, Univ. of Chicago.
- Grunfeld, Y. & Z. Griliches(1960), Is aggregation necessarily bad?, *Review of Econ. and Statist.*, vol. 42, pp. 1~13.
- Izenman, A. J. & S. L. Zabell(1981), Babies and blackout: The genesis of misconception, *Social Science Research*, vol. 10, pp. 282~299.
- Johansen, S. (1991), Estimation and hypotheses testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica*, vol. 59, pp. 1551~1580.
- Johansen, S. & K. Juselius(1988), Hypothesis testing for cointegration vectors-with an application to the demand for money in Denmark and Finland, preprint, IMS, Univ. of Copenhagen.
- McCleary, R. & R. A. Hay, Jr.(1980), *Applied Time Series for the Social Sciences*, Sage Pub., Beverly Hills.
- McSweeney, A. J. (1978), Effects of response cost on the behavior of a million persons charging for directory assistance in Cincinnati, *J. Appl. Behavior Anal.*, vol. 11, pp. 47~51.

- Maddala, G. S. (1977), *Econometrics*, McGraw-Hill, New York.
- Mills, T. C. (1990), *Time Series Techniques for Economists*, Cambridge Univ. Press, Cambridge.
- Mills, T. C. & M. J. Stephenson(1987), A time series forecasting system for the UK money supply, *Economic Modelling*, vol. 4, pp. 355~369.
- Park, J. Y. (1992), Canonical cointegrating regressions, *Econometrica*, vol. 60, pp. 119~143.
- Phillips, P. C. B. (1991), Optimal inference in cointegrated system, *Econometrica*, vol. 59, pp. 283~306.
- Sims, C. A., J. H. Stock & M. W. Watson(1990), Inference in linear time series model with some unit roots, *Econometrica*, vol. 58, pp. 113~144.
- Stock, J. (1987), Asymptotic properties of least square estimators of cointegrating vectors, *Econometrica*, vol. 55, pp. 1035~1056.
- Stock, J. H. & M. W. Watson(1993). A simple estimator of cointegrating vectors in higher order integrated systems, *Econometrica*, vol 61, pp. 783~820.
- Tiao, G. C., G. E. P. Box & W. J. Hamming(1975a), Analysis of Los Angeles Photochemical smog data: A statistical overview, *J. Air Pollution Control Assoc.*, vol. 25, pp. 260~268.
- Tiao, G. C., G. E. P. Box & W. J. Hamming(1975b), A statistical analysis of Los Angeles ambient carbon monoxide data: 1955~1972, *J. Air Pollution Control Assoc.*, vol. 25, pp. 1129~1136.
- Vandaele, W. (1983), *Applied Time Series and Box-Jenkins Models*, Academic Press, New York.
- Wei, W. W. S. (1990), *Time Series Analysis: Univariate and Multivariate Methods*, Addison-Wesley Pub., Redwood City.
- Woodfield, T. J. (1987), Time series intervention analysis using SAS software, *SUGI-SAS Users Group Conference Proceedings*, vol. 12, pp. 331~339.

다변량시계열모형

- 6.1 자기공분산행렬과 자기상관행렬
- 6.2 다변량시계열의 일반모형들
- 6.3 VAR모형
- 6.4 VMA모형
- 6.5 VARMA모형

제 6 장 다변량시계열모형

일반적으로 분석해야 하는 시계열데이터는 한 개의 변수로 구성되어 있는 경우보다는 두 개 이상의 변수들로 구성되어 있는 경우가 많다. 증권시장의 주가와 주식거래량, 일별 최고온도와 최고습도, 상품의 광고비용과 판매량, 연도별 국민총생산액과 총수출액 등은 두 변수들을 동시에 분석하는 것이 필요하다. 물론 각 변수를 분석할 수도 있으나 두 변수를 같이 취급하면 좀더 효과적인 통계분석을 할 수 있고, 주어진 정보를 좀더 적극적으로 이용한다고 할 수 있다. 제 4장에서 다룬 전이함수모형은 인과관계가 분명한 경우의 입출력관계(input-output relation)를 나타낸다. 즉, 한 개의 출력변수와 여러 개의 입력변수들 사이의 관계를 나타낸다. 따라서 전이함수모형을 사용해서 여러 개의 변수들로 이루어진 시계열데이터를 분석할 수도 있다. 그러나 많은 시계열데이터에서 변수들 사이의 인과관계가 명확하지 않아서 전이함수모형을 사용할 수가 없다. 이 장에서는 이러한 시계열데이터들을 분석하는 데 유용한 다변량시계열모형(multivariate time series model)을 다루고자 한다.

6.1 자기공분산행렬과 자기상관행렬

다변량시계열데이터(multivariate time series)는 몇 개의 단변량시계열데이터(univariate time series)들로 구성되어 있다. 우선 d 개의 단변량시계열데이터들 $\{y_{1,t} | t=0, \pm 1, \pm 2, \dots\}$, $\{y_{2,t} | t=0, \pm 1, \pm 2, \dots\}$, ..., $\{y_{d,t} | t=0, \pm 1, \pm 2, \dots\}$ 이 다음 조건을 만족한다고 가정해 보자.

$$E(y_{i,t}^2) < \infty \quad (t=0, \pm 1, \pm 2, \dots, i=1, 2, \dots, d)$$

다음과 같은 벡터를 정의하자.

$$y_t := (y_{1t}, y_{2t}, \dots, y_{dt})' \quad (t = 0, \pm 1, \pm 2, \dots)$$

이 $\{y_t \mid t = \dots, -2, -1, 0, 1, 2, \dots\}$ 을 d 변량시계열데이터(d -variate time series)라고 부른다.

단변량시계열분석에서 그러했듯이 다변량시계열분석에서도 주로 1차적률과 2차적률을 분석하고 3차 이상의 적률들을 분석하는 경우는 드물다. 그 주된 이유는 계산이 복잡하기 때문이다. 또한 정규성가정하에서 3차 이상의 적률들은 큰 의미가 없기 때문이기도 하다. 다변량시계열데이터 $\{y_t\}$ 의 1차적률인 평균벡터와 2차적률인 공분산행렬은 각각 다음과 같이 정의된다.

$$\begin{aligned} \mu_t &:= E(y_t) \quad (t = 0, \pm 1, \pm 2, \dots) \\ \Gamma(s, t) &:= E\{(y_s - \mu_s)(y_t - \mu_t)'\} \quad (s, t = 0, \pm 1, \pm 2, \dots) \end{aligned}$$

단변량시계열분석에서와 마찬가지로 다변량시계열분석에서도 정상성(stationarity)은 매우 중요한 개념이다. [SAS4TSA2]의 제 9.1 절에서 기술한 단변량시계열데이터의 정상성에 관한 개념을 다변량시계열데이터에 확장할 수 있다. 즉, 어떤 확률벡터열(a sequence of random vectors) $\{\dots, y_{-2}, y_{-1}, y_0, y_1, y_2, \dots\}$ 이 엄격한 의미의 정상성(strictly stationarity)을 갖기 위한 필요충분조건은 임의의 자연수 n, k 와 임의의 시점들 t_1, t_2, \dots, t_k 에 대해 $y(t_1), y(t_2), \dots, y(t_k)$ 와 $y(t_1+n), y(t_2+n), y(t_k+n)$ 이 같은 결합확률분포를 갖는 것이다. (여기서 $y(t_i)$ 는 시점 t_i 에서의 확률벡터이다. 첨자의 첨자를 피하기 위해서 이러한 기호를 사용했다.) 이 정상성의 정의는 확률벡터열 $\{y_t\}$ 에 대한 것이나 편의상 이 확률벡터열의 실현(realization)인 다변량시계열데이터가 정상성을 갖는다고 말한다. 단변량시계열데이터의 정상성과 구별하기 위해서 d 변량시계열데이터의 정상성을 d 차원결합정상성(d -dimensional jointly stationarity)이라고도 한다.

다변량시계열데이터 $\{y_t\}$ 가 정상성을 가지면 다음 조건들이 만족된다.

- ① 평균벡터 μ_t 가 시점 t 에 의존하지 않는다.
- ② 공분산행렬 $\Gamma(s, t)$ 는 시차 $(t-s)$ 에 의존하며 시점들 t 와 s 각각에 의존하지는 않는다.

이 두 가지 조건을 만족하는 다변량시계열데이터를 넓은 의미로 정상적(stationary in wide sense)이라고 한다. 또한, 약하게 정상적(weakly stationary) 또는 2차정상적(second-order stationary)이라고도 한다. 어떤 다변량시계열데이터가 엄격한 의미의 정상성을 가지면 넓은 의미로도 정상적이다. 그러나 역은 성립하지 않는다. 예외로 다변량시계열데이터의 확률구조가 다변량정규분포를 갖는다면 이 두 정상성은 같다. 일반적으로 어떤 다변량시계열데이터가 엄격한 의미의 정상성을 갖는지를 조사하는 것은 불가능한 일이다. 따라서 다변량시계열분석에서 정상성을 가정할 때는 넓은 의미의 정상성을 의미한다. 이후 분서에서 정상성은 넓은 의미의 정상성을 뜻하기로 한다. 정상성을 갖지않은 다변량시계열데이터를 비정상다변량시계열데이터(nonstationary multivariate time series)라고

한다. 만일 d 변량시계열데이터 $\{y_t = (y_{1,t}, y_{2,t}, \dots, y_{d,t}) \mid t=0, \pm 1, \pm 2, \dots\}$ 가 d 차원결합정상성을 가지면 모든 $i (= 1, 2, \dots, d)$ 에 대해서 각 단변량시계열데이터 $\{y_{i,t} \mid t=0, \pm 1, \pm 2, \dots\}$ 는 정상성을 갖는다. 그러나 각 단변량시계열데이터들이 정상성을 갖는다고 해서 d 변량시계열데이터 $\{y_t\}$ 가 반드시 d 차원결합정상성을 갖는다는 것은 아니다.

d 변량시계열데이터 $\{y_t\}$ 가 정상적이면 평균벡터와 공분산행렬을 다음과 같이 표기한다.

$$\begin{aligned} \mu &:= E(y_t) \quad (t=0, \pm 1, \pm 2, \dots) \\ \Gamma(h) &:= E\{(y_{t+h} - \mu)(y_t - \mu)'\} \quad (t, h=0, \pm 1, \pm 2, \dots) \end{aligned}$$

이후 $\Gamma(h)$ 를 시차(lag)가 h 인 자기공분산행렬(autocovariance matrix)이라 하고, $\{\Gamma(h) \mid h=0, \pm 1, \pm 2, \dots\}$ 를 자기공분산행렬함수(autocovariance matrix function)라고 한다. 이후 시차 h 의 자기공분산행렬 $\Gamma(h)$ 의 (i, j) 원소를 $\gamma_{ij}(h)$ 로 표기하자. 즉, 다음과 같이 표기하자.

$$\Gamma(h) = \begin{bmatrix} \gamma_{11}(h) & \gamma_{12}(h) & \dots & \gamma_{1d}(h) \\ \gamma_{21}(h) & \gamma_{22}(h) & \dots & \gamma_{2d}(h) \\ \vdots & \vdots & & \vdots \\ \gamma_{d1}(h) & \gamma_{d2}(h) & \dots & \gamma_{dd}(h) \end{bmatrix}$$

정의로부터 $\gamma_{ij}(h)$ 는 $y_{i,t+h}$ 와 $y_{j,t}$ 의 공분산임을 알 수 있다. 즉, $\{\gamma_{ij}(h) \mid h=0, \pm 1, \pm 2, \dots\}$ 은 단변량시계열데이터 $\{y_{i,t} \mid t=0, \pm 1, \pm 2, \dots\}$ 의 자기공분산함수이고, $\{\gamma_{ij}(h) \mid h=0, \pm 1, \pm 2, \dots\}$ 은 단변량시계열데이터들 $\{y_{i,t} \mid t=0, \pm 1, \pm 2, \dots\}$ 와 $\{y_{j,t} \mid t=0, \pm 1, \pm 2, \dots\}$ 의 교차공분산함수(cross-covariance function)이다.

자기공분산행렬함수 $\{\Gamma(h)\}$ 는 다음 성질들을 만족한다.

- ① $\Gamma(-h) = \Gamma(h)'$
- ② $|\gamma_{ij}(h)| \leq \{\gamma_{ii}(0)\gamma_{jj}(0)\}^{\frac{1}{2}}, (i, j=1, 2, \dots, d)$
- ③ 임의의 자연수 n , 임의의 정수들 t_1, t_2, \dots, t_n , 그리고 임의의 d 차원 열벡터들 a_1, a_2, \dots, a_n 에 대해서 다음 부등식이 성립한다.

$$\sum_{i=1}^n \sum_{j=1}^n a_i' \Gamma(t_i - t_j) a_j \geq 0$$

첫 번째의 성질은 정상성의 정의에 의해 성립하고, 두 번째 성질은 Cauchy-Schwarz부등식에 의해 성립한다. 세 번째 성질은 다음 식에 의해 성립한다.

$$E\left[\sum_{i=1}^n a_i' \{y(t_i) - \mu\}\right]^2 \geq 0$$

제 4장에서 정의했듯이 $\{y_{i,t} \mid t=0, \pm 1, \pm 2, \dots\}$ 과 $\{y_{j,t} \mid t=0, \pm 1, \pm 2, \dots\}$ 의 교차상관계수(cross-correlation)는 다음과 같다.

$$\rho_{ij}(h) := \frac{\gamma_{ij}(h)}{\sqrt{\gamma_{ii}(0)\gamma_{jj}(0)}}$$

다변량시계열데이터 $\{y_t\}$ 의 시차 h 에서의 자기상관행렬(autocorrelation matrix)은 다음과 같이 정의된다.

$$P(h) := \begin{bmatrix} \rho_{11}(h) & \rho_{12}(h) & \cdots & \rho_{1d}(h) \\ \rho_{21}(h) & \rho_{22}(h) & \cdots & \rho_{2d}(h) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{d1}(h) & \rho_{d2}(h) & \cdots & \rho_{dd}(h) \end{bmatrix}$$

또한, $\{P(h) \mid h=0, \pm 1, \pm 2, \dots\}$ 을 자기상관행렬함수(autocorrelation matrix function)라 한다. 여기서 유의할 점은 $\rho_{ij}(0)$ 은 $y_{i,t}$ 와 $y_{j,t}$ 의 상관계수라는 것이다. 따라서 $\rho_{ii}(0) = 1$ 이지만, $i \neq j$ 인 경우는 $\rho_{ij}(0)$ 이 1이 아닐 수도 있다. 또한, $i \neq j$ 인 경우 어떤 $h(\neq 0)$ 에 대해 다음 부등식이 성립할 수도 있다.

$$|\rho_{ij}(h)| > |\rho_{ij}(0)|$$

[SAS4TSA2]의 제 9장에서 강조했듯이 시계열분석이란 관찰점들(observations) 사이의 관계를 규명하는 것이므로 관찰점들 사이의 관계를 나타내는 2차적률을 주로 분석한다. 따라서 다변량시계열분석에서도 자기공분산행렬함수와 자기상관행렬함수는 중요한 역할을 한다.

자기공분산행렬함수 $\{\Gamma(h)\}$ 로부터 자기상관행렬함수 $\{P(h)\}$ 를 구하기 위해서 d 차원대각행렬 Q 를 다음과 같이 정의하자.

$$Q = \text{diag}\left(\frac{1}{\sqrt{\gamma_{11}(0)}}, \frac{1}{\sqrt{\gamma_{22}(0)}}, \dots, \frac{1}{\sqrt{\gamma_{dd}(0)}}\right)$$

이 대각행렬을 사용해서 다음 식을 얻을 수 있다.

$$P(h) = Q \Gamma(h) Q \quad (h=0, \pm 1, \pm 2, \dots)$$

한 예로 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-1.PRG에 수록되어 있다).

```

/*                                                    */
/*                                                    EXAM6-1.PRG          */
/*                                                    */
/*          CALCULATING ACRMAF FROM ACVMF          */
/*                                                    */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOLOG NOCENTER;
/* AUTOCOVARIANCE MATRICES */

```

```

GAMMA0 = { 1.3333333 -0.143007 -0.016824,
           -0.143007 1.2088164 0.4825525,
           -0.016824 0.4825525 1.3145703 };
GAMMA1 = { 0.6666667 -0.071504 -0.008412,
           0.1139853 0.2513466 0.4409439,
           -0.033649 0.386529 0.4908816 };
GAMMA2 = { 0.3333333 -0.035752 -0.004206,
           0.0679706 0.133943 0.1905176,
           0.0127024 0.166228 0.2354533 };
GAMMA3 = { 0.1666667 -0.017876 -0.002103,
           0.0439411 0.0596875 0.0892671,
           0.0174048 0.076657 0.1087395 };
GAMMA4 = { 0.0833333 -0.008938 -0.001052,
           0.0262822 0.0271783 0.0413383,
           0.0140097 0.0349346 0.0504753 };
GAMMA5 = { 0.0416667 -0.004469 -0.000526,
           0.0151645 0.0123044 0.0191713,
           0.0094593 0.015916 0.0234102 };
/* CALCULATE AUTOCOVARIANCE MATRICES */
Q = INV(SQRT(DIAG(GAMMA0)));
RHO0 = Q*GAMMA0*Q;
RHO1 = Q*GAMMA1*Q;
RHO2 = Q*GAMMA2*Q;
RHO3 = Q*GAMMA3*Q;
RHO4 = Q*GAMMA4*Q;
RHO5 = Q*GAMMA5*Q;
/* PRINT AUTOCORRELATION MATRICES */
PRINT Q,
      RHO0 "" RHO1,
      RHO2 "" RHO3,
      RHO4 "" RHO5;
QUIT;

```

이 SAS프로그램은 SAS/IML로 작성되었다. SAS/IML을 모르더라도 어느 한 종류의 컴퓨터프로그래밍언어를 아는 독자들은 이 SAS프로그램을 쉽게 이해하리라 생각한다. SAS/IML의 사용법은 제 3장에서 설명하였다. 이 SAS/IML프로그램에서는 자기공분산행렬들 GAMMA0, GAMMA1, ..., GAMMA5로부터 대각행렬 Q와 자기상관행렬들 RHO0, RHO1, ..., RHO5를 구한다. 이 SAS프로그램을 수행한 결과가 그림 6.1이다.

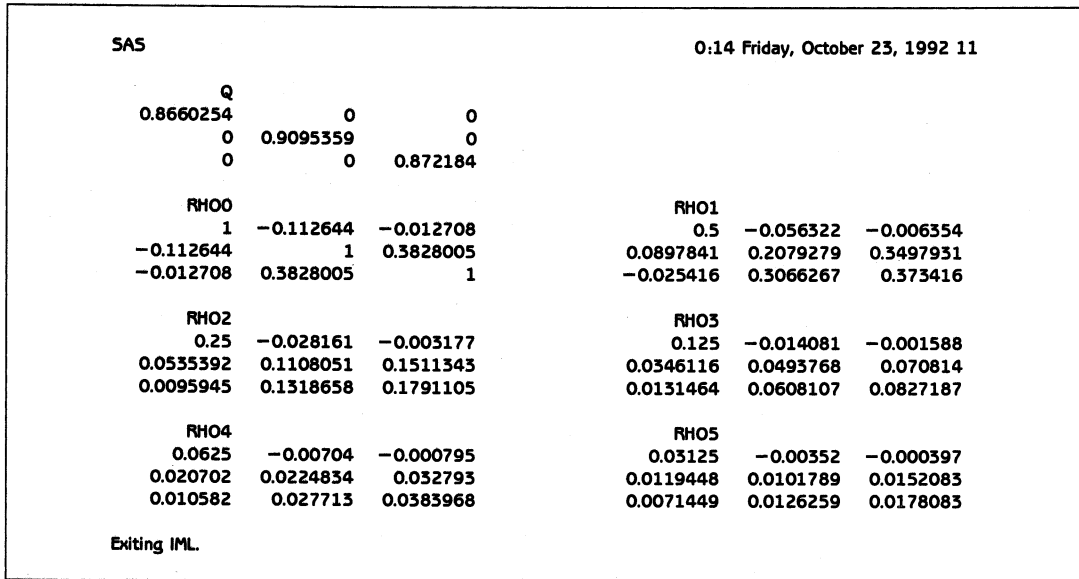


그림 6.1 자기상관행렬들

6.2 다변량시계열의 일반모형들

다변량시계열분석에서와 마찬가지로 다변량시계열분석에서도 1차적률인 평균 그 자체는 그리 중요하지 않다. 이 절에서는 평균벡터가 영벡터인 다변량시계열데이터를 다룬다.

6.2.1 다변량백색잡음과정

다변량시계열분석에서의 백색잡음과정과 마찬가지로 다변량시계열분석에서도 다변량백색잡음과정(multivariate white noise process)이 중요한 역할을 한다. 다변량시계열데이터 $\{v_t | t=0, \pm 1, \pm 2, \dots\}$ 이 다변량백색잡음과정이기 위한 필요충분조건은 $\{v_t\}$ 가 정상적이며 평균벡터는 영벡터이고 다음 식을 만족하는 자기공분산행렬함수 $\{\Gamma(h)\}$ 를 갖는 것이다.

$$\Gamma(h) = \begin{cases} \Sigma & (h=0) \\ 0 & (h \neq 0) \end{cases}$$

여기서 Σ 는 양정치행렬(positive definite matrix)이다. 수리통계학적인 이론을 전개하기 위해서 어떤 다변량시계열데이터 $\{v_t\}$ 가 다변량백색잡음과정이라는 가정 대신에 $\{v_t\}$ 가 서로 독립이고 같은 분포를 갖는다는 가정을 하기도 한다. 그러나 실제로 다변량시계열데이터를 분석하는 데 있어서는 두 가정의 차이는 없다. 다변량백색잡음과정의 한 예로써 다

음과 같은 모형을 살펴보자.

$$\begin{bmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -0.2 & 1 & 0 \\ 0 & 0.3 & 1 \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix}$$

여기서 $\{u_{1t}\}$, $\{u_{2t}\}$, $\{u_{3t}\}$ 는 서로 독립이고 분산이 1인 단변량백색잡음과정들이다. 따라서 $\{v_t = (v_{1t}, v_{2t}, v_{3t})\}$ 는 평균벡터가 영벡터이고 공분산행렬이 다음과 같은 3변량백색잡음과정이다.

$$\begin{bmatrix} 1 & 0 & 0 \\ -0.2 & 1 & 0 \\ 0 & 0.3 & 1 \end{bmatrix} \begin{bmatrix} 1 & -0.2 & 0 \\ 0 & 1 & 0.3 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -0.2 & 0 \\ -0.2 & 1.04 & 0.3 \\ 0 & 0.3 & 1.09 \end{bmatrix}$$

이 3변량백색잡음과정에서 50개의 관찰점들을 구하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA2의 파일 EXAM6-2.PRG에 수록되어 있다).

```

/*                                                    */
/*                EXAM6-2.PRG                        */
/*                                                    */
/*                GENERATE 3-VARIATE WHITE NOISE PROCESS */
/*                DATASET = WHITENOI                */
LIBNAME TSA4DATA 'A:\TSA4DATA';
DATA EXAM62:
  DO T= 1 TO 250;
    U1=RANNOR(7654321);
    U2=RANNOR(1234567);
    U3=RANNOR(1357531);
    V1 = U1;
    V2 = U2-0.2*U1;
    V3 = U3+0.3*U2;
    OUTPUT;
  END;
  KEEP T V1 V2 V3;
RUN;
DATA TSA4DATA.WHITENOI;
  SET EXAM62;
  T=T-200;
  IF T < 1 THEN DELETE;
RUN;
OPTIONS PAGESIZE=60;

```

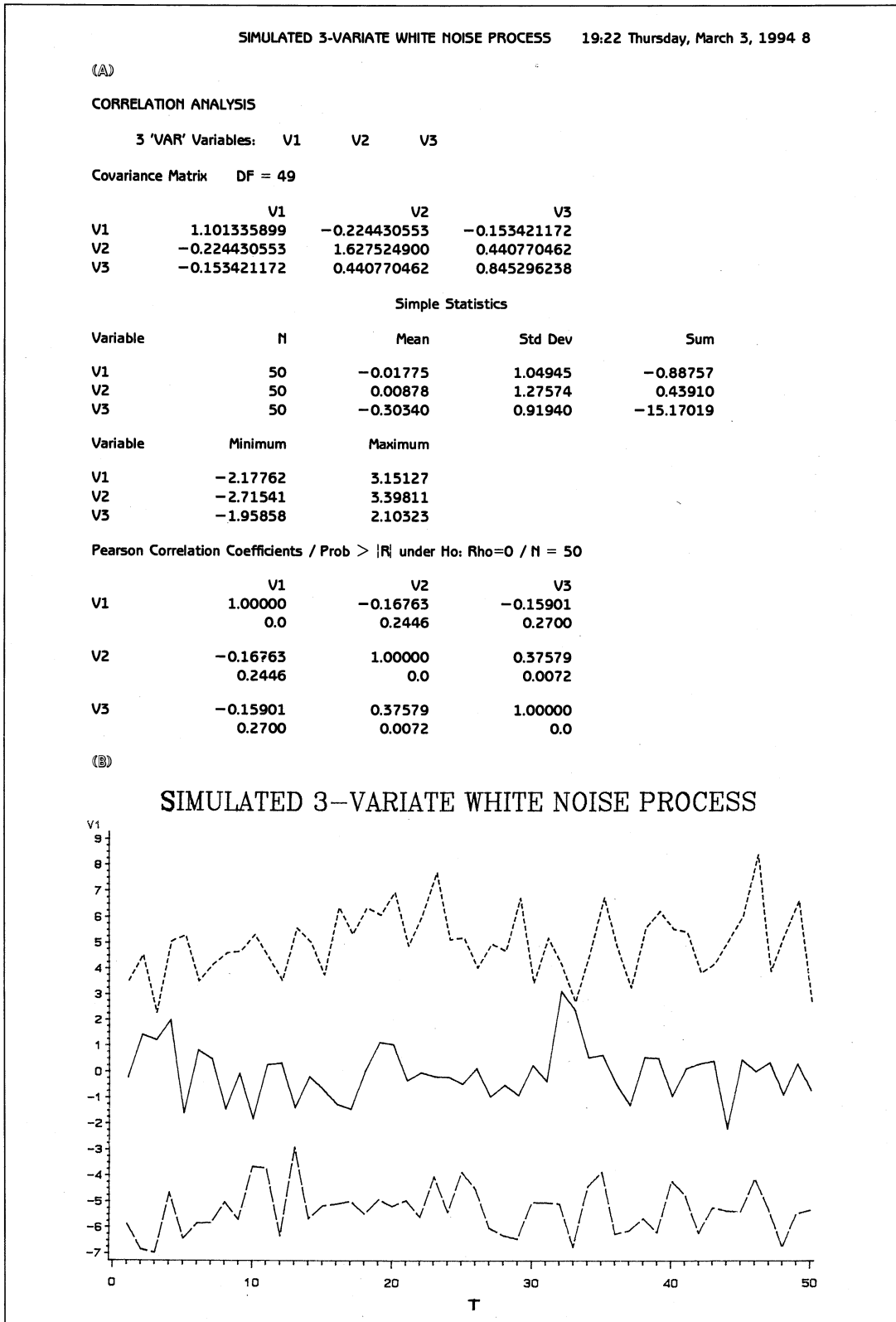


그림 6.2 3변량백색잡음과정


```

TITLE1 'SIMULATED 3-VARIATE WHITE NOISE PROCESS';
PROC CORR COV;
  VAR V1 V2 V3;
  RUN;
DATA EXAM62B;
  SET TSA4DATA.WHITENOI;
  V2 = V2 + 5.0;
  V3 = V3 - 5.0;
  RUN;
PROC GPLOT;
  PLOT (V1 V2 V3)*T / OVERLAY;
  SYMBOL1 C=BLACK V=NONE I=JOIN LINE=1;
  SYMBOL2 C=BLACK V=NONE I=JOIN LINE=2;
  SYMBOL3 C=BLACK V=NONE I=JOIN LINE=4;
  RUN;

```

이 SAS프로그램을 실행한 결과가 그림 6.2이다. 그림 6.2의 부분 A는 PROC CORR를 실행한 결과이고 부분 B는 PROC GPLOT를 실행한 결과이다.

6.2.2 VMA(∞)모형

다변량일반선형과정(multivariate general linear process) $\{y_t\}$ 는 다음 식과 같이 d 변량 백색잡음과정의 현재항과 과거항들 $\{u_t, u_{t-1}, u_{t-2}, \dots\}$ 의 선형결합으로 정의된다.

$$y_t := u_t + \Psi_1 u_{t-1} + \Psi_2 u_{t-2} + \dots = \sum_{k=0}^{\infty} \Psi_k u_{t-k}$$

여기서 Ψ_1, Ψ_2, \dots 는 $d \times d$ 행렬들이고, $\Psi_0 = I_d$ 이며, $\{u_t\}$ 는 평균벡터가 영벡터이고 공분산행렬이 Σ 인 백색잡음과정이다. 다변량일반선형모형을 무한차수의 VMA모형(vector moving-average model, 벡터이동평균모형)이라 하고 VMA(∞)모형이라고 쓴다.

다변량일반선형과정의 특성행렬(characteristic matrix)은 다음과 같이 정의된다.

$$\Psi(x) := \sum_{k=0}^{\infty} \Psi_k x^k \quad (\Psi_0 = I_d)$$

이 특성행렬과 후진연산자 B 를 사용하면 다변량일반선형모형을 다음과 같이 표기할 수 있다.

$$y_t = \Psi(B)v_t$$

VMA계수행렬(vector moving-average coefficient matrix) Ψ_k 의 (i, j) 원소를 Ψ_{kij} 라고 하자. 따라서 특성행렬 $\Psi(x)$ 의 (i, j) 원소인 $\Psi_{ij}(x)$ 는 다음과 같다.

$$\Psi_{ij}(x) = \sum_{k=0}^{\infty} \Psi_{k,ij} x^k$$

다변량일반선형과정은 다음 식으로 정의되었다.

$$y_t = \sum_{k=0}^{\infty} \Psi_k v_{t-k}$$

이 등호의 양변은 확률벡터들이다. 두 확률벡터가 같다는 것은 무엇을 뜻할까? 또한, 우변의 무한급수는 어떠한 의미에서 수렴하는 것일까? 지금부터는 이에 대한 해답을 생각해 보자. 다변량시계열데이터 $\{y_t\}$ 의 자기공분산행렬함수가 존재하기 위해서는 다음 조건이 가정되어야 한다.

$$\sum_{k=0}^{\infty} \Psi_{k,ij}^2 < \infty \quad (i, j = 1, 2, \dots, d)$$

이 조건이 만족되면 다음 식이 성립한다.

$$\lim_{n \rightarrow \infty} E \left[\left(y_t - \sum_{k=0}^n \Psi_k v_{t-k} \right)' \left(y_t - \sum_{k=0}^n \Psi_k v_{t-k} \right) \right] = 0$$

Hilbert공간에 대해서 아는 독자들은 $\sum_{k=0}^{\infty} \Psi_k v_{t-k}$ 는 평균제곱수렴(mean square convergence)을 하고, 두 확률벡터 y_t 와 $\sum_{k=0}^{\infty} \Psi_k v_{t-k}$ 는 L^2 공간에서 동일하다는 것을 알 것이다. Hilbert공간에 대한 내용은 [BD90]의 제 2장을 참조하라.

6.2.3 VAR(∞)모형

다변량시계열데이터를 모형화하는 방법 중의 하나는 다음과 같은 무한차수의 VAR모형(vector autoregressive model, 벡터자기회귀모형), 즉 VAR(∞)모형을 사용하는 것이다.

$$y_t = \sum_{k=1}^{\infty} \Pi_k y_{t-k} + v_t$$

여기서 Π^1, Π^2, \dots 은 $d \times d$ 행렬들이다. 이 VAR(∞)모형의 특성행렬은 다음과 같다.

$$\Pi(x) = - \sum_{k=1}^{\infty} \Pi_k x^k \quad (\Pi_0 = -I_d)$$

VAR계수행렬(vector autoregressive coefficient matrix) Π_k 의 (i, j) 원소를 $\Pi_{k,ij}$ 라고 하면 특성행렬 $\Pi(x)$ 의 (i, j) 원소인 $\pi_{ij}(x)$ 는 다음과 같다.

$$\pi_{ij}(x) = - \sum_{k=1}^{\infty} \pi_{k,ij} x^k$$

모든 (i, j) 에 대해서, $|x| \leq 1$ 에서 이 등식의 우변이 절대수렴(absolutely converge)하면 d 변량시계열데이터 $\{y_t\}$ 는 가역적(invertible)이라고 한다. 즉, 다음 조건이 만족되면 $\{y_t\}$ 는 가역적이다.

$$\sum_{k=0}^{\infty} |\pi_{k,ij}| < \infty \quad (i, j = 1, 2, \dots, d)$$

6.3 VAR모형

6.3.1 작인적 VAR모형

정상성을 갖는 다변량시계열데이터를 분석하는 데 사용되는 모형 중의 하나가 차수가 p 인 VAR모형(vector autoregressive model), 즉 VAR(p)모형이다. 이 d 변량VAR(p)과정 $\{z_t\}$ 는 다음 식을 만족한다.

$$z_t = \nu + \Phi_1 z_{t-1} + \Phi_2 z_{t-2} + \dots + \Phi_p z_{t-p} + v_t \quad (t=0, \pm 1, \pm 2, \dots)$$

여기서 VAR계수행렬들 $\Phi_1, \Phi_2, \dots, \Phi_p$ 는 $d \times d$ 행렬들이며, ν 는 d 차원열벡터이고, $\{v_t\}$ 는 평균벡터가 영벡터이며 공분산행렬이 Σ 인 d 변량백색잡음과정이다. 특별한 언급이 없는 한 Σ 는 양정치행렬이라고 가정한다.

행렬 $(I - \Phi_1 - \Phi_2 - \dots - \Phi_p)$ 가 정칙(nonsingular)이면 다음 식을 만족하는 벡터 μ 가 존재한다.

$$\mu := (I - \Phi_1 - \Phi_2 - \dots - \Phi_p)^{-1} \nu$$

다음과 같은 새로운 d 변량시계열데이터 $\{y_t\}$ 를 정의하자.

$$y_t := z_t - \mu$$

이 d 변량시계열데이터 $\{y_t\}$ 는 다음 d 변량VAR(p)모형을 만족한다.

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \dots + \Phi_p y_{t-p} + v_t \quad (t=0, \pm 1, \pm 2, \dots)$$

벡터 μ 가 존재하기 위한 조건, 즉 행렬 $(I - \Phi_1 - \Phi_2 - \dots - \Phi_p)$ 가 정칙이기 위한 조건은 뒤에서 다루게 될 작인성조건(作因性條件, causality condition)의 특수한 경우이다. 이후 분석하게 될 다변량시계열데이터는 이 작인성조건을 만족한다고 가정하기 때문에 벡터 μ 가 존재한다. 따라서 지금부터는 d 변량시계열데이터 $\{y_t\}$ 가 만족하는 d 변량VAR(p)모형을 분석하기로 하자. 이 다변량시계열데이터의 평균벡터는 영벡터이다. 즉, 다음 식들이 만족된다.

$$E(y_t) = 0$$

$$E(z_t) = \mu = (I - \Phi_1 - \Phi_2 - \dots - \Phi_p)^{-1} \nu$$

이 식들의 증명은 간단하다. 만일 $E(y_t) = a$ 라 놓으면 $\{y_t\}$ 의 정상성에 의해서 다음 식이 성립한다.

$$a = \Phi_1 a + \Phi_2 a + \dots + \Phi_p a$$

따라서 다음 식이 성립한다.

$$(I - \Phi_1 - \Phi_2 - \dots - \Phi_p) a = 0$$

행렬 $(I - \Phi_1 - \Phi_2 - \dots - \Phi_p)$ 가 정칙이므로 $a = 0$ 이다.

우선 VAR(1)모형에 대해서 살펴보자. 정상성을 가지며 평균벡터가 영벡터인 d 변량시계열데이터 $\{y_t\}$ 가 d 변량VAR(1)과정이면 다음 식을 만족한다.

$$y_t = \Phi_1 y_{t-1} + v_t$$

이 d 변량VAR(1)과정이 다음 식을 만족함을 쉽게 알 수 있다.

$$y_t = v_t + \Phi_1 v_{t-1} + \Phi_1^2 v_{t-2} + \dots + \Phi_1^n v_{t-n} + \Phi_1^{n+1} y_{t-n-1}$$

따라서 다음 식이 성립한다.

$$\begin{aligned} E\left\{ \left(y_t - \sum_{k=0}^n \Phi_1^k v_{t-k} \right) \left(y_t - \sum_{k=0}^n \Phi_1^k v_{t-k} \right)' \right\} \\ = E\left\{ \left(\Phi_1^{n+1} y_{t-n-1} \right) \left(\Phi_1^{n+1} y_{t-n-1} \right)' \right\} \\ = \Phi_1^{n+1} \Gamma(0) \left(\Phi_1^{n+1} \right)' \end{aligned}$$

만일, $n \rightarrow \infty$ 일 때 $\Phi_1^{n+1} \rightarrow 0$ 이면, 다음 식이 성립한다.

$$y_t = \sum_{k=0}^{\infty} \Phi_1^k v_{t-k}$$

이 식의 우변의 무한급수는 평균제곱수렴을 한다. 이렇게 다변량시계열데이터 $\{y_t\}$ 를 VMA(∞)과정으로 나타낼 수 있으면, $\{y_t\}$ 는 작인적(causal)이라고 한다. 행렬 Φ_1 의 모든 특성값들의 절대값들이 1보다 작다면 다음 식이 성립한다.

$$\lim_{n \rightarrow \infty} \Phi_1^n = 0$$

행렬 Φ_1 의 모든 특성값들의 절대값들이 1보다 작기 위한 필요충분조건은 다음 방정식의 모든 근들이 단위원 밖에 있는 것이다.

$$\det(I - \Phi_1 x) = 0$$

즉, $|x| \leq 1$ 에서는 $\det(I - \Phi_1 x) \neq 0$ 인 것이다. 이 조건을 VAR(1)모형의 작인성조건이라 한다. 이 작인성조건은 안정성조건(stability condition)이라고도 하는데, 단변량시계열 분석에서 Box-Jenkins의 정상성조건에 해당하는 것이다. 본서에서는 확률과정에서의 정상성조건과 시계열분석에서의 정상성조건을 구별하기 위해서 작인성조건이라는 용어를 사용하기로 한다. 이 용어는 [BD90] (p. 83, pp. 125~130, p. 418)에서 사용한 것이다. 결론적으로 VAR(1)과정 $\{y_t\}$ 가 작인성조건을 만족하면 VMA(∞)모형으로 나타낼 수 있다.

한 예로 다음과 같은 3변량VAR(1)과정을 살펴보자.

$$y_t = \Phi_1 y_{t-1} + v_t$$

$$\text{단, } \Phi_1 = \begin{bmatrix} 0.5 & 0.0 & 0.0 \\ 0.1 & 0.1 & 0.3 \\ 0.0 & 0.2 & 0.3 \end{bmatrix}$$

이 3변량VAR(1)과정을 VMA(∞)모형으로 나타내기 위해 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-3.PRG에 수록되어 있다).

```

/ *                               */
/ *                               EXAM6-3.PRG                               */
/ *                               */
/ *          CALCULATE MA COEFFICIENT MATRICES                          */
/ *          OF 3-VARIATE VAR(1) MODEL                                   */
OPTIONS PAGESIZE = 60;
PROC IML;
RESET NOLOG NOCENTER;
PHI1 = { 0.5  0.0  0.0,
         0.1  0.1  0.3,
         0.0  0.2  0.3 };
PSI0 = PHI1**0;
PSI1 = PSI0*PHI1;
PSI2 = PSI1*PHI1;
PSI3 = PSI2*PHI1;
PSI4 = PSI3*PHI1;
PSI5 = PSI4*PHI1;
PSI6 = PSI5*PHI1;
PSI7 = PSI6*PHI1;
PRINT PSI0 " " PSI1, PSI2 " " PSI3,
       PSI4 " " PSI5, PSI6 " " PSI7, ;
QUIT;

```

다변량시계열데이터 $\{y_t\}$ 를 VMA(∞)모형으로 표현하면 다음과 같다.

$$y_t = \sum_{k=0}^{\infty} \Psi_k v_{t-k} \quad (\Psi_0 = I)$$

따라서 VAR(1)과정을 VMA(∞)모형으로 나타내면 VMA계수행렬들은 다음 식을 만족한다.

$$\Psi_k = \Phi_1^k \quad (k=0, 1, 2, \dots)$$

이 SAS프로그램에서 PSI k 는 Ψ_k 이다. 이 SAS프로그램을 수행한 결과가 그림 6.3이다.

SAS			10:03 Thursday, October 29, 1992 1		
PSI0			PSI1		
1	0	0	0.5	0	0
0	1	0	0.1	0.1	0.3
0	0	1	0	0.2	0.3
PSI2			PSI3		
0.25	0	0	0.125	0	0
0.06	0.07	0.12	0.037	0.031	0.057
0.02	0.08	0.15	0.018	0.038	0.069
PSI4			PSI5		
0.0625	0	0	0.03125	0	0
0.0216	0.0145	0.0264	0.01225	0.00673	0.01227
0.0128	0.0176	0.0321	0.00816	0.00818	0.01491
PSI6			PSI7		
0.015625	0	0	0.0078125	0	0
0.006798	0.003127	0.0057	0.0037117	0.0014527	0.0026481
0.004898	0.0038	0.006927	0.002829	0.0017654	0.0032181
Exiting IML.					

그림 6.3 3변량 VAR(1)과정의 VMA(∞)표현

VAR(1)과정 $\{y_t\}$ 의 VMA(∞)표현을 사용해서 다음과 같이 시차 h 에서의 자기공분산행렬을 유도할 수 있다.

$$\begin{aligned}
 & \Gamma(h) \\
 &= E(y_{t+h} y_t') \\
 &= E\left\{ \left(\sum_{k=0}^{\infty} \Phi_1^k v_{t+h-k} \right) \left(\sum_{l=0}^{\infty} \Phi_1^l v_{t-l} \right)' \right\} \\
 &= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \Phi_1^k E(v_{t+h-k} v_{t-l}') (\Phi_1^l)' \\
 &= \Phi_1^h \left\{ \sum_{k=0}^{\infty} \Phi_1^k \Sigma (\Phi_1^k)'\right\}
 \end{aligned}$$

지금까지 전개한 VAR(1)모형에 관한 이론을 확장해서 VAR(p)모형에 적용할 수 있다. d 변량시계열데이터 $\{y_t\}$ 가 다음 식을 만족하는 d 변량VAR(p)과정이라고 하자.

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \dots + \Phi_p y_{t-p} + u_t$$

이 d 변량VAR(p)과정은 다음과 같이 dp 변량VAR(1)모형으로 나타낼 수 있다.

$$w_t = \Phi w_{t-1} + u_t$$

$$\text{단, } w_t := \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}, \quad \Phi = \begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_{p-1} & \Phi_p \\ I & O & \dots & O & O \\ O & I & \dots & O & O \\ \vdots & \vdots & & \vdots & \vdots \\ O & O & \dots & I & O \end{bmatrix}, \quad u_t := \begin{bmatrix} u_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

다변량시계열데이터 $\{y_t\}$ 가 정상적이기 때문에 $\{w_t\}$ 도 정상적이다. 이 dp 변량VAR(1)과정 $\{w_t\}$ 가 작인적이기 위해서는 VAR계수행렬 Φ 의 모든 특성값들의 절대값들이 1보다 작아야 한다. 즉, $|x| \leq 1$ 에서 $\det(I - \Phi x) \neq 0$ 이어야 한다. 선형대수학의 기초적인 이론을 적용해서 다음 식을 쉽게 유도할 수 있다.

$$\begin{aligned} & \det(I - \Phi x) \\ &= \det \begin{bmatrix} I - \Phi_1 x & -x\Phi_2 & \cdots & -x\Phi_{p-1} & -x\Phi_p \\ -xI & I & \cdots & O & O \\ O & -xI & \cdots & O & O \\ \vdots & \vdots & & \vdots & \vdots \\ O & O & \cdots & -xI & I \end{bmatrix} \\ &= \det \begin{bmatrix} I - \sum_{k=1}^p \Phi_k x^k & -x\Phi_1 & \cdots & -x\Phi_{p-1} & -x\Phi_p \\ O & I & \cdots & O & O \\ O & -xI & \cdots & O & O \\ \vdots & \vdots & & \vdots & \vdots \\ O & O & \cdots & -xI & I \end{bmatrix} \\ &= \det(I - \sum_{k=1}^p \Phi_k x^k) \end{aligned}$$

두 번째 행렬식의 제 p 블록열에 x^{-1} 을 곱해서 제 1블록열에 더하고, 제 $(p-1)$ 열에 x^{-2} 을 곱해서 제 1블록열에 더하고, ..., 제 2블록열에 x 를 곱해서 제 1블록열에 더하면 세 번째 행렬식이 된다. 이 식으로부터 d 변량VAR(p)과정 $\{y_t\}$ 가 작인적이기 위한 필요충분조건은 $|x| \leq 1$ 에서 다음 조건이 성립하는 것임을 알 수 있다.

$$\det(I - \sum_{k=1}^p \Phi_k x^k) \neq 0$$

다음 식을 이 d 변량VAR(p)과정의 특성행렬이라 정의한다.

$$\Phi(x) := - \sum_{k=0}^p \Phi_k x^k \quad (\Phi_0 = -I_d)$$

또한 이 특성행렬의 행렬식을 특성식이라 한다. 여기서 유의할 점은 행렬 Φ 의 특성식은 $\det[\Phi(x)]$ 가 아니라 $\det[(x)^p \Phi(x^{-1})]$ 이라는 것이다. 따라서 Lütkepohl(1991, p. 12)같은 학자들은 $\det[\Phi(x)]$ 를 VAR(p)과정의 역특성식(reverse characteristic polynomial)이라고도 한다. 또한 행렬 Φ 는 $(-x)^p \Phi(x^{-1})$ 의 친구행렬(companion matrix)형태이다. 이에 대한 자세한 내용은 Wilkinson(1965, pp. 12~13)을 참조하라.

작인성을 갖는 d 변량VAR(p)과정도 VMA(∞)모형으로 표현할 수 있다. 우선 다음과 같은 $d \times dp$ 행렬을 정의하자.

$$J := [I_d \ O \ O \ \cdots \ O]$$

행렬 J 를 사용하면 다음 식을 유도할 수 있다.

$$\begin{aligned}
 & y_t \\
 &= Jw_t \\
 &= J\left(\sum_{k=0}^{\infty} \Phi^k u_{t-k}\right) \\
 &= J\left(\sum_{k=0}^{\infty} \Phi^k J' J u_{t-k}\right) \\
 &= \sum_{k=0}^{\infty} (J \Phi^k J') v_{t-k} \\
 &= \sum_{k=0}^{\infty} \Psi_k v_{t-k}
 \end{aligned}$$

여기서 $\Psi_k = J \Phi^k J'$ 이다. 행렬 Φ 의 모든 특성값들의 절대값들이 1보다 작으므로 $\sum_{k=0}^{\infty} \Psi_k v_{t-k}$ 는 평균제곱수렴을 하고, 두 확률벡터 y_t 와 $\sum_{k=0}^{\infty} \Psi_k v_{t-k}$ 는 L^2 공간에서 동일하다. 결론적으로 작인성을 갖는 d 변량VAR(p)과정 $\{y_t\}$ 는 다음과 같은 VMA(∞)모형을 만족한다.

$$y_t = \Psi(B) v_t$$

여기서 $\Psi(x) = \sum_{k=0}^{\infty} \Psi_k x^k$ ($\Psi_0 = I_d$)이다. 정의에 의해서 VAR(p)과정은 다음 식을 만족한다.

$$\Phi(B) y_t = v_t$$

여기서 $\Phi(x) = -\sum_{k=0}^p \Phi_k x^k$ ($\Phi_0 = -I_d$)이다. 따라서 다음 식이 성립한다.

$$\Phi^{-1}(x) = \Psi(x)$$

작인성조건에 의해서 $|x| \leq 1$ 에서 $\det[\Phi(x)] \neq 0$ 이므로, $|x| \leq 1$ 인 모든 x 에 대해 다음 식을 만족한다.

$$\Phi(x) \Psi(x) = I_d$$

따라서 항등식의 원리에 의해 다음 식들을 얻는다.

$$\begin{aligned}
 \Psi_0 &= I_d \\
 \Psi_1 &= \Psi_0 \Phi_1 \\
 \Psi_2 &= \Psi_1 \Phi_1 + \Psi_0 \Phi_2 \\
 &\vdots \\
 \Psi_p &= \Psi_{p-1} \Phi_1 + \Psi_{p-2} \Phi_2 + \dots + \Psi_0 \Phi_p \\
 \Psi_i &= \Psi_{i-1} \Phi_1 + \Psi_{i-2} \Phi_2 + \dots + \Psi_{i-p} \Phi_p \quad (i=p+1, p+2, \dots)
 \end{aligned}$$

이 식들을 이용하면 VAR(p)과정을 VMA(∞)모형으로 나타낼 때의 VMA계수행렬들 $\{\Psi_i\}$ 를 쉽게 구할 수 있다.

SAS/IML을 사용해서 VAR(p)과정을 VMA(∞)모형으로 나타낼 때의 VMA계수행렬들 $\{\Psi_i\}$ 를 구하는 방법의 하나는 함수 RATIO를 사용하는 것이다. 함수 RATIO를 사용하면 행렬역함수들(matrix polynomials)의 나누기를 수행할 수 있다. 함수 RATIO를 사용

해서 다음과 같은 행렬역함수들 $A(x)$ 와 $B(x)$ 의 나누기를 수행하기로 하자.

$$C(x) = [A(x)]^{-1}B(x)$$

$$\text{단, } A(x) = A_0 + A_1x + A_2x^2 + \dots + A_{s-1}x^{s-1}$$

$$B(x) = B_0 + B_1x + B_2x^2 + \dots + B_{t-1}x^{t-1}$$

여기서 A_0, A_1, \dots, A_{s-1} 은 $n \times n$ 행렬들이고 B_0, B_1, \dots, B_{t-1} 은 $n \times m$ 행렬들이다. 나누어진 행렬역함수 $C(x)$ 를 다음과 같이 표기하자.

$$C(x) = \sum_{k=0}^{\infty} C_k x^k$$

함수 RATIO를 사용해서 $C(x)$ 를 구하기 위해서는 우선 $n \times ns$ 행렬 A 와 $n \times mt$ 행렬 B 를 다음과 같이 정의한다.

$$A = [A_0 \ A_1 \ \dots \ A_{s-1}]$$

$$B = [B_0 \ B_1 \ \dots \ B_{t-1}]$$

함수 RATIO를 계산하면 다음과 같은 $n \times mr$ 행렬 C 가 출력된다.

$$C = [C_0 \ C_1 \ \dots \ C_{r-1}]$$

이 행렬 C 를 계산하기 위해서는 다음과 같은 스테이트먼트를 수행한다.

$$C = \text{RATIO}(A, B, r, m);$$

여기서 r 은 행렬 C 를 구성하는 부행렬들의 개수이고, m 은 행렬 B_i 들의 열의 개수로서 디폴트는 1이다. 함수 COVLAG에 관한 자세한 내용은 [IML6.03] (pp. 292~293)을 참조하라.

한 예로 다음 3변량 VAR(2)과정을 살펴보자.

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + v_t$$

이 3변량 VAR(2)과정을 VMA(∞)모형으로 나타내기 위해 다음 SAS프로그램을 수행해보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-4.PRG에 수록되어 있다).

```

/*                                                    */
/*                EXAM6-4.PRG                        */
/*                                                    */
/*                CALCULATE MA COEFFICIENT MATRICES */
/*                OF 3-VARIATE VAR(2) MODEL         */
/*                                                    */
OPTIONS PAGESIZE = 60;
PROC IML;
RESET NOLOG NOCENTER;
/* PART (A) MAKING A PROGRAM */

```

```

PHI1 = { 0.5  0.1  0.0,
         0.4  0.3  0.0,
         0.0  0.0  0.3 };
PHI2 = { 0.0  0.0  0.0,
         0.25 0.0  0.0,
         0.0  0.0  0.28 };
PSI0 = PHI1**0;
PSI1 = PHI1*PSI0;
PSI2 = PHI1*PSI1 + PHI2*PSI0;
PSI3 = PHI1*PSI2 + PHI2*PSI1;
PSI4 = PHI1*PSI3 + PHI2*PSI2;
PSI5 = PHI1*PSI4 + PHI2*PSI3;
PSI6 = PHI1*PSI5 + PHI2*PSI4;
PSI7 = PHI1*PSI6 + PHI2*PSI5;
PRINT "(A) MAKING A PROGRAM", ;
PRINT PSI0 "" PSI1, PSI2 "" PSI3,
      PSI4 "" PSI5, PSI6 "" PSI7, ;
FREE / PHI1 PHI2;
/* USING SAS/ IML FUNCTION RATIO */
PHI0 = PHI1**0;
THETA0 = PHI0;
PHI = PHI0 || -PHI1 || -PHI2;
THETA = THETA0;
PSI = RATIO(PHI, THETA, 8, 3);
/* GET PSI MATRICES */
PSI0 = PSI[ , {1 2 3}];
PSI1 = PSI[ , {4 5 6}];
PSI2 = PSI[ , {7 8 9}];
PSI3 = PSI[ , {10 11 12}];
PSI4 = PSI[ , {13 14 15}];
PSI5 = PSI[ , {16 17 18}];
PSI6 = PSI[ , {19 20 21}];
PSI7 = PSI[ , {22 23 24}];
PRINT, "(B) USING SAS/ IML FUNCTION RATIO",;
PRINT PSI0 "" PSI1, PSI2 "" PSI3,
      PSI4 "" PSI5, PSI6 "" PSI7, ;
QUIT;

```

이 SAS프로그램을 수행한 결과가 그림 6.4이다. 그림 6.4의 부분 A에는 SAS/IML프로

```

19:22 Thursday, March 3, 1994 12

(A) MAKING A PROGRAM

      PSIO          PSI1
      1            0.5
      0            0.1
      0            1   0
                   0.4
                   0.3
                   0   0.3

      PSI2          PSI3
      0.29          0.202
      0.57          0.053
      0            0.13  0
                   0.412
                   0.096
                   0   0.195

      PSI4          PSI5
      0.1422        0.0361
      0.2769        0.07   0
      0            0.1621
                   0.09879
                   0.02505
                   0.19045
                   0.04869
                   0   0.10323

      PSI6          PSI7
      0.06844       0.017394
      0.132201     0.033652
      0            0   0
                   0.0474401
                   0.0120622
                   0.0917338
                   0.0233157
                   0   0.0518115

(B) USING SAS/IML FUNCTION RATIO

      PSIO          PSI1
      1            0.5
      0            0.1
      0            1   0
                   0.4
                   0.3
                   0   0.3

      PSI2          PSI3
      0.29          0.202
      0.57          0.053
      0            0.13  0
                   0.412
                   0.096
                   0   0.195

      PSI4          PSI5
      0.1422        0.0361
      0.2769        0.07   0
      0            0.1621
                   0.09879
                   0.02505
                   0.19045
                   0.04869
                   0   0.10323

      PSI6          PSI7
      0.06844       0.017394
      0.132201     0.033652
      0            0   0
                   0.0474401
                   0.0120622
                   0.0917338
                   0.0233157
                   0   0.0518115

Exiting IML.

```

그림 6.4 3변량 VAR(2)과정의 VMA(∞)표현

그램을 만들어서 VMA계수행렬들을 구한 결과가 수록되어 있고, 부분 B에는 SAS/IML 함수 RATIO를 사용해서 VMA계수행렬들을 구한 결과가 수록되어 있다.

VAR(p)과정을 VMA모형으로 나타낼 때 한 가지 유의할 점은 이 VMA모형의 차수가 반드시 무한대는 아니라는 것이다. 즉, 어떤 자연수 K 가 존재해서 K 보다 큰 k 에 대해서 Ψ_k 가 영행렬일 수 있다는 것이다. 즉, 다음 식이 성립할 수 있다.

$$y_t = \sum_{k=0}^K \Psi_k v_{t-k}$$

한 예로 다음과 같은 4변량VAR(1)모형을 살펴보자.

$$y_t = \Phi_1 y_{t-1} + v_t$$

여기서 $\{v_t\}$ 는 4변량백색잡음과정이고, Φ_1 은 4×4 행렬로서 다음과 같다.

$$\Phi_1 = \begin{bmatrix} 0 & 0.5 & 0.1 & 0.4 \\ 0 & 0 & 0.3 & 0.1 \\ 0 & 0 & 0 & 0.3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

다음 식이 성립한다.

$$\Phi^k = 0 \quad (k=4, 5, \dots)$$

이 식이 성립함을 보이기 위해 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-5.PRG에 수록되어 있다).

```

/*                                     */
/*                                     EXAM6-5.PRG                             */
/*                                     */
/*          CALCULATE MA COEFFICIENT MATRICES                             */
/*          OF 4-VARIATE VAR(1) MODEL                                     */
/*                                     */
OPTIONS PAGESIZE = 60;
PROC IML;
RESET NOLOG NOCENTER;
PHT1 = { 0.0 0.5 0.1 0.4,
         0.0 0.0 0.3 0.1,
         0.0 0.0 0.0 0.3,
         0.0 0.0 0.0 0.0 };
PSI0 = PHI1**0;
PSI1 = PSI0*PHI1;
PSI2 = PSI1*PHI1;
PSI3 = PSI2*PHI1;
PSI4 = PSI3*PHI1;
PSI5 = PSI4*PHI1;
PSI6 = PSI5*PHI1;
PSI7 = PSI6*PHI1;
PRINT PSI0 " " PSI1, PSI2 " " PSI3,
      PSI4 " " PSI5, PSI6 " " PSI7,;
QUIT;

```

이 SAS프로그램을 수행한 결과가 그림 6.5이다.

```

SAS                                     19:22 Thursday, March 3, 1994 14

PSI0
  1      0      0      0
  0      1      0      0
  0      0      1      0
  0      0      0      1

PSI1
  0      0.5    0.1    0.4
  0      0      0.3    0.1
  0      0      0      0.3
  0      0      0      0

PSI2
  0      0      0.15   0.08
  0      0      0      0.09
  0      0      0      0
  0      0      0      0

PSI3
  0      0      0      0.045
  0      0      0      0
  0      0      0      0
  0      0      0      0

PSI4
  0      0      0      0
  0      0      0      0
  0      0      0      0
  0      0      0      0

PSI5
  0      0      0      0
  0      0      0      0
  0      0      0      0
  0      0      0      0

PSI6
  0      0      0      0
  0      0      0      0
  0      0      0      0
  0      0      0      0

PSI7
  0      0      0      0
  0      0      0      0
  0      0      0      0
  0      0      0      0

Exiting IML.
    
```

그림 6.5 4변량 VAR(1)모형의 VMA(∞)표현

따라서 이 4변량 VAR(1)모형의 VMA표현은 다음과 같다.

$$y_t = \sum_{k=0}^{\infty} \Phi_1^k v_{t-k} = \sum_{k=0}^3 \Phi_1^k v_{t-k}$$

즉, 다음 두 다변량시계열모형들은 같은 자기공분산행렬함수를 갖는다.

$$y_t = \Phi_1 y_{t-1} + v_t$$

$$y_t = v_t + \Phi_1 v_{t-1} + \Phi_1^2 v_{t-2} + \Phi_1^3 v_{t-3}$$

다변량시계열분석에서는 자기공분산함수가 주어지면 유일한(unique) 착인적이고 가역

적인 ARMA모형을 얻을 수 있다. 반면에 다변량시계열분석에서는 여러 개의 VARMA모형들이 같은 자기공분산행렬함수를 가질 수 있다. 이 VARMA모형의 비유일성에 대해서는 제 6.5.7소절에서 자세히 다룰 것이다.

6.3.2 예제 들

[예제 1]

다음과 같은 3변량VAR(1)모형을 살펴보자.

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \begin{bmatrix} -0.6 & 0 & 0 \\ 0 & 0.3 & 0.5 \\ 0 & 0.4 & -0.5 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} + \begin{bmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \end{bmatrix}$$

여기서 $\{v_t = (v_{1t}, v_{2t}, v_{3t})'\}$ 는 3변량백색잡음과정으로서 평균벡터는 영벡터이고 공분산행렬은 다음과 같다.

$$\begin{bmatrix} 1 & -0.2 & 0 \\ -0.2 & 1.04 & 0.3 \\ 0 & 0.3 & 1.09 \end{bmatrix}$$

이 VAR(1)모형의 특성식은 다음과 같다.

$$\begin{aligned} \det \begin{bmatrix} 1+0.6x & 0 & 0 \\ 0 & 1-0.3x & -0.5x \\ 0 & -0.4x & 1+0.5x \end{bmatrix} \\ = (1+0.6x)\{(1-0.3x)(1+0.5x) - (-0.5x)(-0.4x)\} \\ = (1+0.6x)(1-0.5x)(1+0.7x) \end{aligned}$$

이 VAR(1)모형의 특성근들의 절대값들이 모두 1보다 크므로 이 VAR(1)모형은 stationary이다. 이 VAR(1)모형에서 150개의 관찰점들을 생성하기 위해 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-6.PRG에 수록되어 있다).

```

/*                               */
/*                               EXAM6-6.PRG                               */
/*                               */
/* GENERATE 3-VARIATE VAR(1) PROCESS                               */
/* DATASET = VAR31A                                                */
LIBNAME TSA4DATA 'A:\TSA4DATA';
DATA EXAM66;
    
```

```

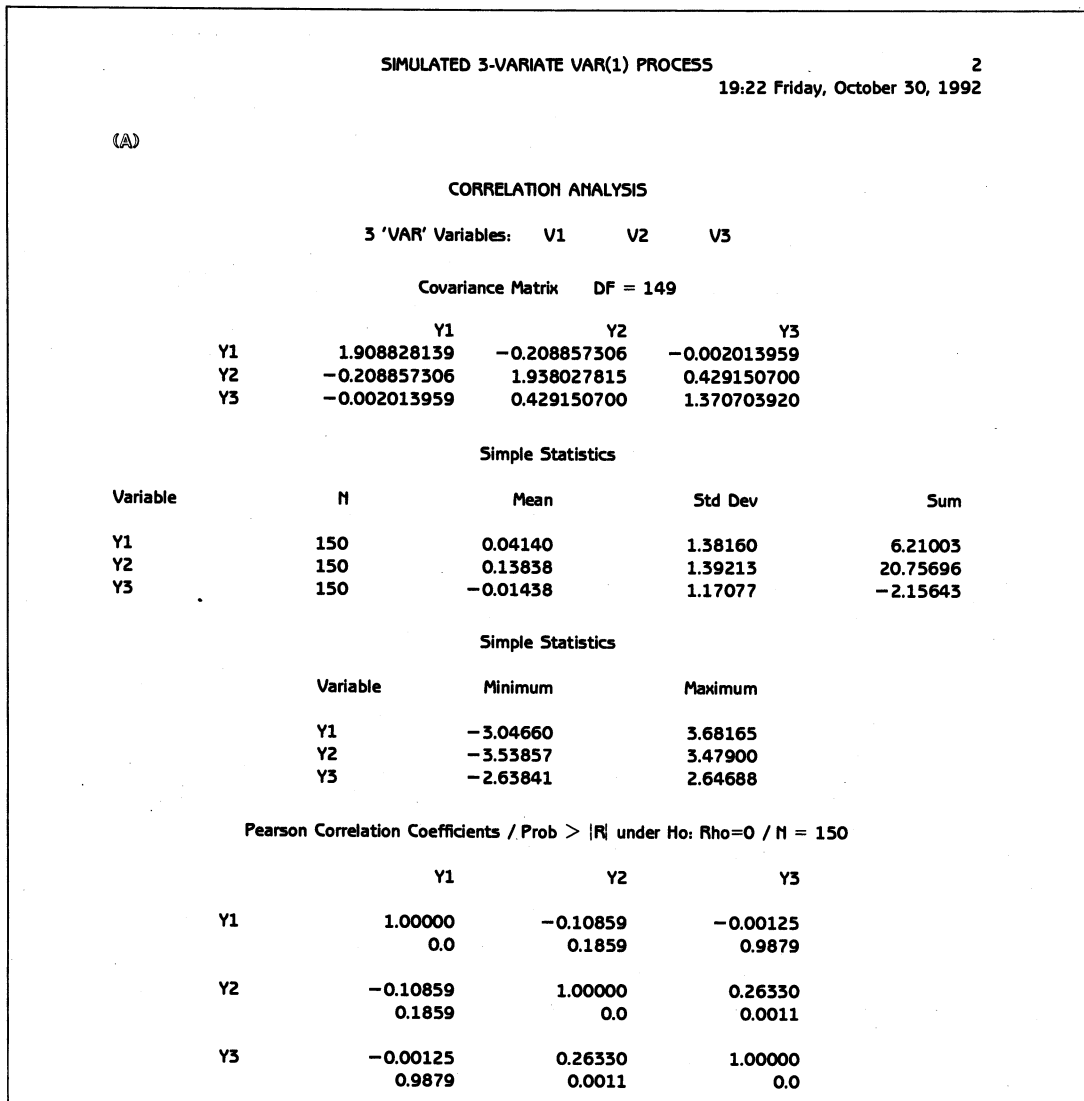
DO T = 1 TO 350;
  U1 = RANNOR(7654321);
  U2 = RANNOR(1234567);
  U3 = RANNOR(1357531);
  V1 = U1;
  V2 = U2-0.2*U1;
  V3 = U3+0.3*U2;
  IF Y1LAG1 = . THEN Y1LAG1 = 0.0;
  IF Y2LAG1 = . THEN Y2LAG1 = 0.0;
  IF Y3LAG1 = . THEN Y3LAG1 = 0.0;
  Y1 = -0.6*Y1LAG1 + V1;
  Y2 = 0.3*Y2LAG1 + 0.5*Y3LAG1 + V2;
  Y3 = 0.4*Y2LAG1 - 0.5*Y3LAG1 + V3;
  OUTPUT;
  Y1LAG1 = Y1;
  Y2LAG1 = Y2;
  Y3LAG1 = Y3;
  END;
KEEP T Y1 Y2 Y3 V1 V2 V3;
RUN;
DATA TSA4DATA.VAR31A;
  SET EXAM66;
  T = T-200;
  IF T < 1 THEN DELETE;
  RUN;
OPTIONS PAGESIZE=60;
TITLE1 'SIMULATED 3-VARIATE VAR(1) PROCESS';
PROC CORR COV;
  VAR Y1 Y2 Y3;
  RUN;
DATA EXAM66B;
  SET TSA4DATA.VAR31A;
  MY1 = Y1;
  MY2 = Y2 + 5.0;
  MY3 = Y3 - 5.0;
  KEEP T MY1 MY2 MY3;
  RUN;
PROC GPLOT;
  PLOT (MY1 MY2 MY3)*T / OVERLAY;

```

```

SYMBOL1 C=BLACK V=NONE I=JOIN LINE=1;
SYMBOL2 C=BLACK V=NONE I=JOIN LINE=2;
SYMBOL3 C=BLACK V=NONE I=JOIN LINE=4;
FOOTNOTE 'MY1=Y1 MY2=Y2+5 MY3=Y3-5';
RUN;
TITLE1;
FOOTNOTE1;
    
```

이 SAS프로그램을 실행한 결과가 그림 6.6이다. 그림 6.6의 부분 A는 PROC CORR를 실행한 결과이다. 부분 B는 PROC GPLOT를 실행한 결과인데 단변량시계열데이터들 Y1, Y2, Y3가 겹치지 않도록 MY1=Y1, MY2=Y2+5, MY3=Y3-5의 시계열산포도들을 겹쳐 그렸다. 그림 6.6에서 이 3변량시계열데이터 (Y1, Y2, Y3)'가 정상적임을 알 수 있다.



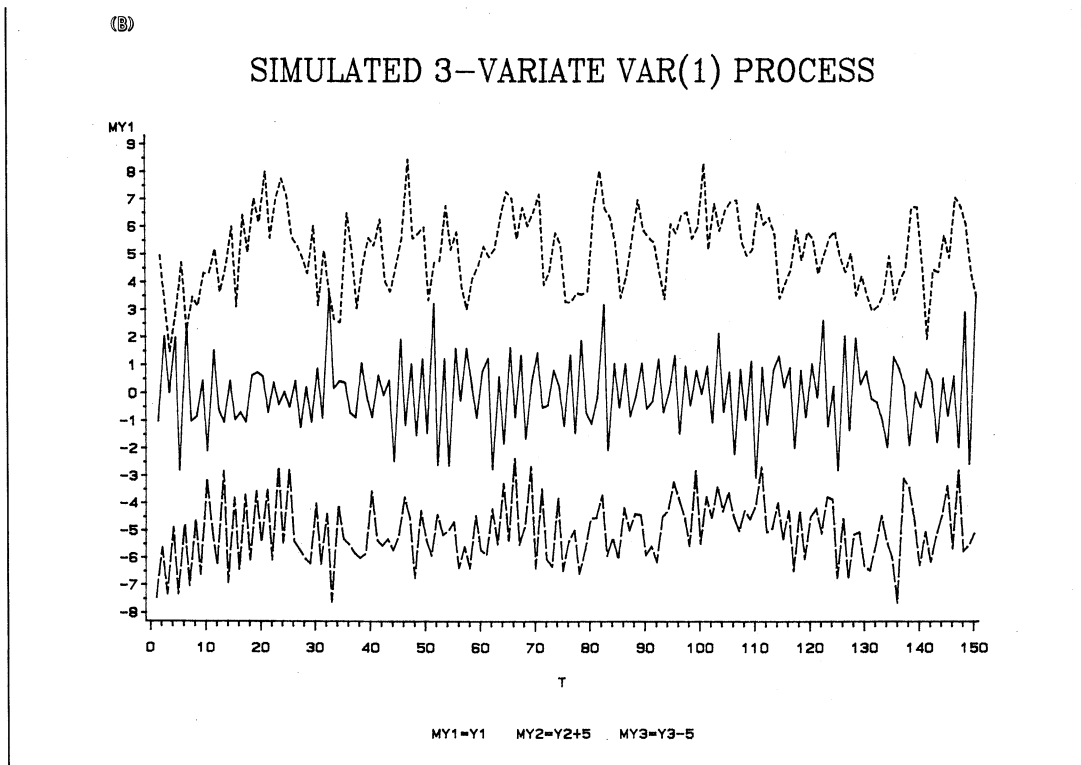


그림 6.6 3변량 VAR(1)모형에서 생성된 다변량시계열데이터 VAR31A

[예제 2] ████████████████████

다음과 같은 3변량VAR(1)모형을 살펴보자.

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \begin{bmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix}$$

여기서 $\{u_t = (u_{1t}, u_{2t}, u_{3t})\}$ 는 첫 번째 예제의 3변량백색잡음과정과 동일하다. 이 VAR(1)모형의 특성식은 다음과 같다.

$$\det \begin{bmatrix} 1-0.5x & 0 & 0 \\ -0.1x & 1-0.1x & -0.3x \\ 0 & -0.2x & 1-0.3x \end{bmatrix} = (1-0.5x)(1-0.4x-0.03x^2)$$

따라서 이 VAR(1)모형의 특성근들은 다음과 같다.

$$x_1 = 2, \quad x_2 = 2.1525, \quad x_3 = -15.4858$$

모든 특성근들의 절대값들이 1보다 크므로 이 3변량VAR(1)과정은 작인적이다. 이 VAR(1)모형에서 100개의 관찰점들을 생성하기 위해 다음 SAS프로그램을 수행해 보자

(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-7.PRG에 수록되어 있다).

```

/* */
/*          EXAM6-7.PRG          */
/* */
/*          GENERATE 3-VARIATE VAR(1) PROCESS          */
/*          DATASET = VAR22A          */
LIBNAME TSA4DATA 'A:\TSA4DATA';
DATA EXAM67;
  DO T = 1 TO 300;
    U1 = RANNOR(7654321);
    U2 = RANNOR(1234567);
    U3 = RANNOR(1357531);
    V1 = U1;
    V2 = U2 - 0.2 * U1;
    V3 = U3 + 0.3 * U2;
    IF Y1LAG1 = . THEN Y1LAG1 = 0.0;
    IF Y2LAG1 = . THEN Y2LAG1 = 0.0;
    IF Y3LAG1 = . THEN Y3LAG1 = 0.0;
    Y1 = 0.5 * Y1LAG1 + V1;
    Y2 = 0.1 * Y1LAG1 + 0.1 * Y2LAG1 + 0.3 * Y3LAG1 + V2;
    Y3 = 0.2 * Y2LAG1 + 0.3 * Y3LAG1 + V3;
    OUTPUT;
    Y1LAG1 = Y1;
    Y2LAG1 = Y2;
    Y3LAG1 = Y3;
  END;
  KEEP T Y1 Y2 Y3 V1 V2 V3;
RUN;
DATA TSA4DATA.VAR31B;
  SET EXAM67;
  T = T - 200;
  IF T < 1 THEN DELETE;
RUN;
OPTIONS PAGESIZE=60;
TITLE1 'SIMULATED 3-VARIATE VAR(1) PROCESS';
PROC CORR COV;
  VAR Y1 Y2 Y3;
RUN;

```

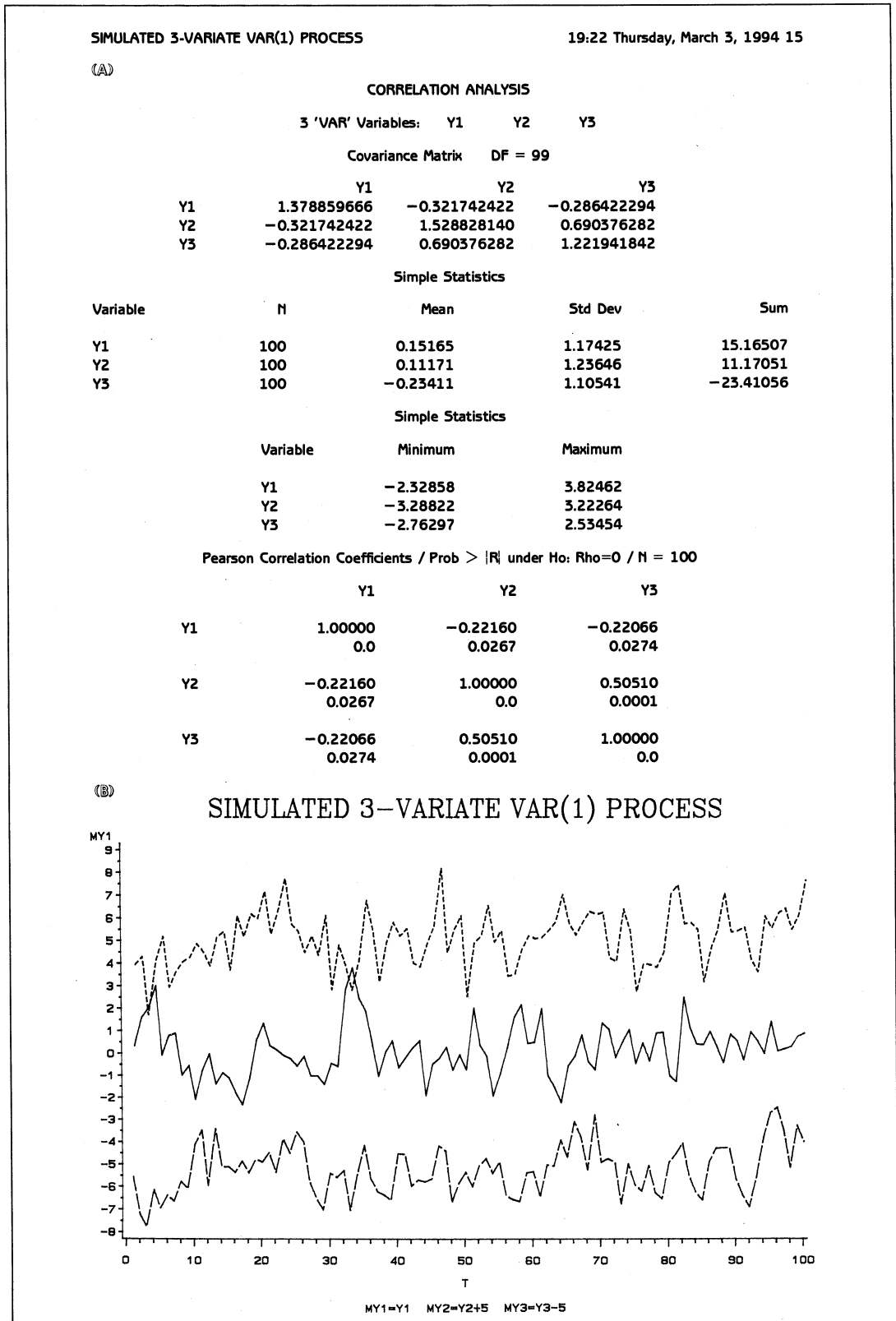


그림 6.7 3변량VAR(1)모형에서 생성된 다변량시계열데이터 VAR31B

```

DATA EXAM67B;
  SET TSA4DATA.VAR31B;
  MY1 = Y1;
  MY2 = Y2 + 5.0;
  MY3 = Y3 - 5.0;
  KEEP T MY1 MY2 MY3;
RUN;

PROC GPLOT;
  PLOT (MY1 MY2 MY3)*T / OVERLAY;
  SYMBOL1 C=BLACK V=NONE I=JOIN LINE=1;
  SYMBOL2 C=BLACK V=NONE I=JOIN LINE=2;
  SYMBOL3 C=BLACK V=NONE I=JOIN LINE=4;
  FOOTNOTE 'MY1=Y1 MY2=Y2+5 MY3=Y3-5';
RUN;

TITLE1;
FOOTNOTE1;

```

이 SAS 프로그램을 실행한 결과가 그림 6.7이다. 그림 6.7의 부분 A는 PROC CORR를 실행한 결과이다. 부분 B는 PROC GPLOT를 실행한 결과인데 단변량 시계열데이터들 Y1, Y2, Y3가 겹치지 않도록 MY1=Y1, MY2=Y2+5, MY3=Y3-5의 시계열산포도들을 겹쳐 그렸다. 부분 B의 산포도에서 이 3변량 시계열데이터 (Y1, Y2, Y3)가 정상적임을 알 수 있다.

[예제 3]

다음과 같은 2변량VAR(2)모형을 살펴보자.

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} -0.26 & 0.54 \\ 0 & 0.48 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} -0.51 & 0 \\ -0.27 & 0 \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} + \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix}$$

여기서 $\{v_t = (v_{1,t}, v_{2,t})'\}$ 는 2변량백색잡음과정으로서 평균벡터는 영벡터이고 공분산행렬은 다음과 같다.

$$\begin{bmatrix} 1 & -0.2 \\ -0.2 & 1.04 \end{bmatrix}$$

이 VAR(2)모형의 특성식은 다음과 같다.

$$\det \begin{bmatrix} 1 + 0.26x + 0.51x^2 & -0.54x \\ 0.27x^2 & 1 - 0.48x \end{bmatrix}$$

$$\begin{aligned}
 &= (1 + 0.26x + 0.51x^2)(1 - 0.48x) - (-0.54x)(0.27x^2) \\
 &= 1 - 0.22x + 0.3852x^2 - 0.099x^3
 \end{aligned}$$

따라서 이 2변량VAR(2)모형의 특성근들은 다음과 같다.

$$x_1 = -4.79397, \quad x_2 = 0.45153 + 1.37954i, \quad x_3 = 0.45153 - 1.37954i$$

모든 특성근들의 절대값들이 1보다 크므로 이 VAR(2)모형은 작인적이다. 이 2변량 VAR(2)모형에서 150개의 관찰점들을 생성하기 위해 다음 SAS프로그램을 수행해 보자 (이 SAS 프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-8.PRG에 수록되어 있다).

```

/*                                     */
/*                                     EXAM6-8.PRG                             */
/*                                     */
/* GENERATE 2-VARIATE VAR(2) PROCESS                                     */
/* DATASET = VAR22B                                                    */
LIBNAME TSA4DATA 'A:\TSA4DATA';
DATA EXAM68;
  DO T = 1 TO 350;
    U1 = RANNOR(7654321);
    U2 = RANNOR(1234567);
    V1 = U1;
    V2 = U2 - 0.2 * U1;
    IF Y1LAG1 = . THEN Y1LAG1 = 0.0;
    IF Y2LAG1 = . THEN Y2LAG1 = 0.0;
    IF Y1LAG2 = . THEN Y1LAG2 = 0.0;
    IF Y2LAG2 = . THEN Y2LAG2 = 0.0;
    Y1 = -0.26 * Y1LAG1 + 0.54 * Y2LAG1 - 0.51 * Y1LAG2 + V1;
    Y2 = 0.48 * Y2LAG1 - 0.27 * Y1LAG2 + V2;
    OUTPUT;
    Y1LAG2 = Y1LAG1;
    Y2LAG2 = Y2LAG1;
    Y1LAG1 = Y1;
    Y2LAG1 = Y2;
  END;
  KEEP T Y1 Y2 V1 V2;
RUN;
DATA TSA4DATA.VAR22A;
  SET EXAM68;
  T = T - 200;

```

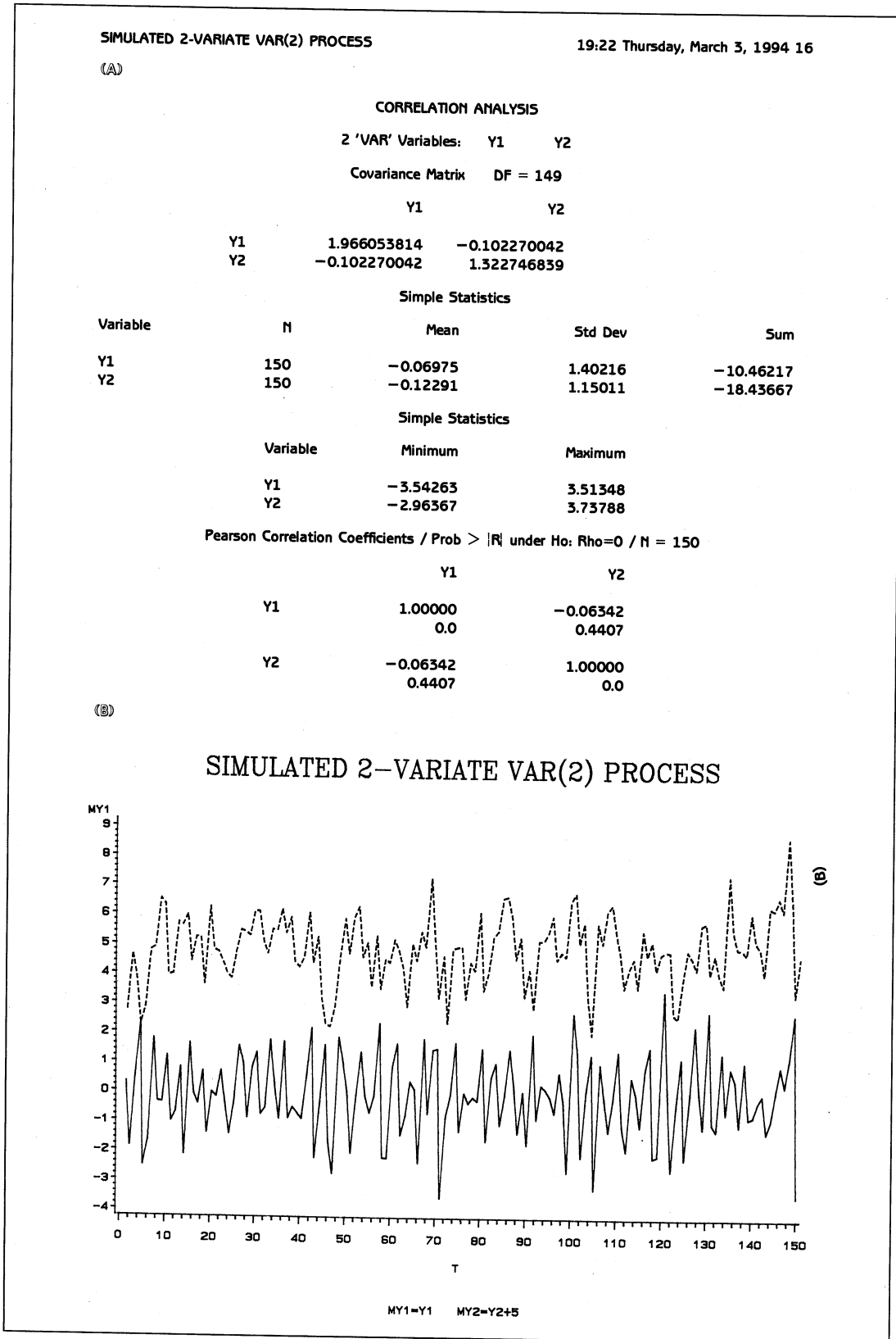


그림 6.8 2변량VAR(2)모형에서 생성된 다변량시계열데이터 VAR22A

```

IF T < 1 THEN DELETE;
RUN;
OPTIONS PAGESIZE=60;
TITLE1 'SIMULATED 2-VARIATE VAR(2) PROCESS';
PROC CORR COV;
VAR Y1 Y2;
RUN;
DATA EXAM68B;
SET TSA4DATA.VAR22A;
MY1 = Y1;
MY2 = Y2 + 5.0;
KEEP T MY1 MY2;
RUN;
PROC GPLOT;
PLOT (MY1 MY2)*T / OVERLAY;
SYMBOL1 C=BLACK V=NONE I=JOIN LINE=1;
SYMBOL2 C=BLACK V=NONE I=JOIN LINE=2;
FOOTNOTE 'MY1=Y1 MY2=Y2+5';
RUN;
TITLE1;
FOOTNOTE1;

```

이 SAS 프로그램을 실행한 결과가 그림 6.8이다. 그림 6.8의 부분 A는 PROC CORR를 실행한 결과이고, 부분 B는 PROC GPLOT를 실행한 결과이다. 부분 B의 산포도에서 이 2변량시계열데이터 (Y1, Y2)'가 정상적임을 알 수 있다.

[예제 4]

다음과 같은 2변량VAR(2)모형을 살펴보자.

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0.25 & 0 \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} + \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix}$$

여기서 $\{v_t = (v_{1,t}, v_{2,t})'\}$ 는 세 번째 예제의 2변량백색잡음과정과 동일하다. 이 2변량 VAR(2)모형의 특성식은 다음과 같다.

$$\begin{aligned} \det \begin{bmatrix} 1 - 0.5x & -0.1x \\ -0.4x - 0.25x^2 & 1 - 0.5x \end{bmatrix} \\ = 1 - x + 0.21x^2 - 0.025x^3 \end{aligned}$$

따라서 이 2변량VAR(2)모형의 특성근들은 다음과 같다.

$$x_1 = 1.3, \quad x_2 = 3.55 + 4.26i, \quad x_3 = 3.55 - 4.26i$$

모든 특성근들의 절대값들이 1보다 크므로 이 2변량VAR(2)과정은 작인적이다. 이 VAR(2)모형에서 100개의 관찰점들을 생성하기 위해 다음 SAS프로그램을 수행해 보자 (이 SAS 프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-9.PRG에 수록되어 있다).

```

/* */
/*          EXAM6-9.PRG          */
/* */
/*          GENERATE 2-VARIATE VAR(2) PROCESS          */
/* */
LIBNAME TSA4DATA 'A:\TSA4DATA';
DATA EXAM69;
  DO T = 1 TO 300;
    U1 = RANNOR(7654321);
    U2 = RANNOR(1234567);
    V1 = U1;
    V2 = U2 - 0.2*U1;
    IF Y1LAG1 = . THEN Y1LAG1 = 0.0;
    IF Y2LAG1 = . THEN Y2LAG1 = 0.0;
    IF Y1LAG2 = . THEN Y1LAG2 = 0.0;
    IF Y2LAG2 = . THEN Y2LAG2 = 0.0;
    Y1 = 0.5*Y1LAG1 + 0.1*Y2LAG1 + V1;
    Y2 = 0.4*Y2LAG1 + 0.5*Y1LAG1 + 0.25*Y1LAG2 + V2;
    OUTPUT;
    Y1LAG2 = Y1LAG1;
    Y2LAG2 = Y2LAG1;
    Y1LAG1 = Y1;
    Y2LAG1 = Y2;
  END;
  KEEP T Y1 Y2 V1 V2;
RUN;
DATA TSA4DATA.VAR22B;
  SET EXAM69;
  T = T - 200;
  IF T < 1 THEN DELETE;
RUN;

```

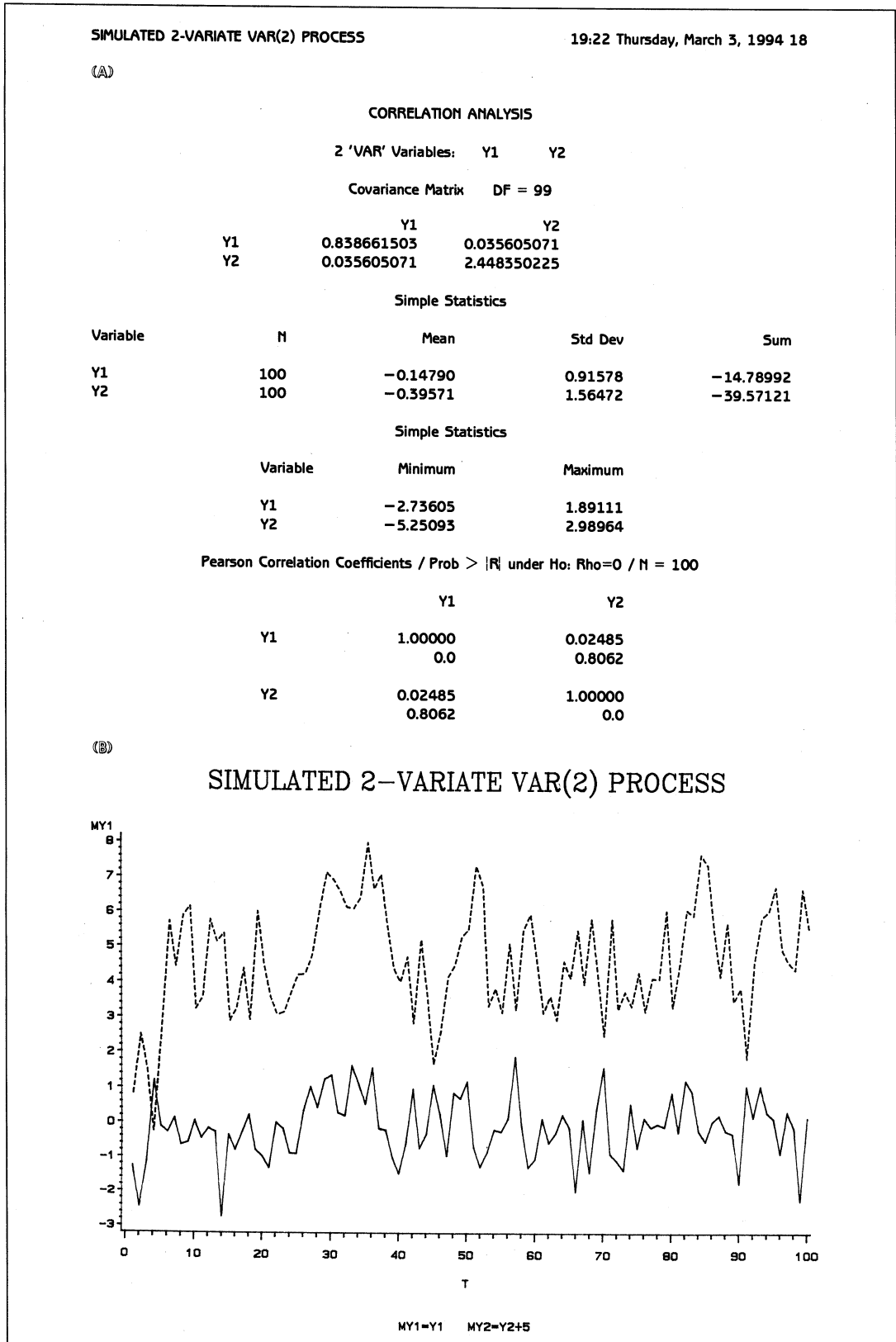



그림 6.9 2변량VAR(2)모형에서 생성된 다변량시계열데이터 VAR22B

```

OPTIONS PAGESIZE=60;
TITLE1 'SIMULATED 2-VARIATE VAR(2) PROCESS';
PROC CORR COV;
    VAR Y1 Y2;
    RUN;
DATA EXAM69B;
    SET TSA4DATA.VAR22B;
    MY1 = Y1;
    MY2 = Y2 + 5.0;
    KEEP T MY1 MY2;
    RUN;
PROC GPLOT;
    PLOT (MY1 MY2)*T / OVERLAY;
    SYMBOL1 C=BLACK V=NONE I=JOIN LINE=1;
    SYMBOL2 C=BLACK V=NONE I=JOIN LINE=2;
    FOOTNOTE 'MY1=Y1 MY2=Y2+5';
    RUN;
TITLE1;
FOOTNOTE1;

```

이 SAS 프로그램을 수행한 결과가 그림 6.9이다. 그림 6.9의 부분 A는 PROC CORR를 실행한 결과이고, 부분 B는 PROC GPLOT를 실행한 결과이다. 부분 B의 산포도에서 이 2변량시계열데이터 (Y1, Y2)'가 정상적임을 알 수 있다.

6.3.3 VAR과정의 자기공분산행렬

작인적 VAR(p)과정의 VMA표현을 사용해서 VAR(p)과정의 자기공분산행렬을 구하면 다음과 같다.

$$\begin{aligned}
 \Gamma(h) &= E(y_{t+h}y_t') \\
 &= E\left\{\left(\sum_{k=0}^{\infty} \Psi_k v_{t+h-k}\right)\left(\sum_{l=0}^{\infty} \Psi_l v_{t-l}\right)'\right\} \\
 &= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \Psi_k E(v_{t+h-k}v_{t-l}') \Psi_l' \\
 &= \sum_{k=0}^{\infty} \Psi_{k+h} \Sigma \Psi_k'
 \end{aligned}$$

그러나 VAR(p)과정의 자기공분산행렬들을 구하는 데 이 식을 사용하지는 않는다.

단변량시계열분석에서와 마찬가지로 작인적 VAR(p)과정은 다음과 같은 Yule-Walker 방정식들을 만족한다.

$$\begin{aligned}\Gamma(0) &= \Phi_1\Gamma(-1) + \Phi_2\Gamma(-2) + \dots + \Phi_p\Gamma(-p) + \Sigma \\ \Gamma(i) &= \Phi_1\Gamma(i-1) + \Phi_2\Gamma(i-2) + \dots + \Phi_p\Gamma(i-p) \quad (i = 1, 2, \dots)\end{aligned}$$

이 식들을 다음과 같이 유도할 수 있다.

$$y_t - \sum_{k=1}^p \Phi_k y_{t-k} = v_t$$

이 식의 양변에 y_{t-i}' 를 곱한 뒤 기대값을 취하면 다음과 같다.

$$\begin{aligned}\Gamma(i) - \sum_{k=1}^p \Phi_k \Gamma(i-k) \\ &= E(v_t y_{t-i}') \\ &= \sum_{h=0}^{\infty} E(v_t v_{t-i-h}') \Psi_k' \\ &= \begin{cases} \Sigma & (i=0) \\ 0 & (i>0) \end{cases}\end{aligned}$$

만일 $\Phi_1, \Phi_2, \dots, \Phi_p$ 와 초기자기공분산행렬들 $\Gamma(1-p), \Gamma(2-p), \dots, \Gamma(0)$ 이 주어지면 Yule-Walker방정식들을 사용해서 $\Gamma(1), \Gamma(2), \dots$ 을 축차적으로 구할 수 있다.

지금부터는 $\Phi_1, \Phi_2, \dots, \Phi_p$ 와 Σ 가 주어졌을 때 초기자기공분산행렬들 $\Gamma(1-p), \Gamma(2-p), \dots, \Gamma(0)$ 을 구하는 방법에 대해 다루기로 하자. 우선 VAR(1)모형의 경우를 생각해 보면 Yule-Walker방정식들로부터 다음 식들을 얻는다.

$$\begin{aligned}\Gamma(0) &= \Phi_1\Gamma(1)' + \Sigma \\ \Gamma(1) &= \Phi_1\Gamma(0)\end{aligned}$$

따라서 다음 식이 성립한다.

$$\Gamma(0) = \Phi_1\Gamma(0)\Phi_1' + \Sigma$$

이 식으로부터 다음 식을 얻는다.

$$\begin{aligned}\text{vec}\Gamma(0) &= \text{vec}(\Phi_1\Gamma(0)\Phi_1') + \text{vec}\Sigma \\ &= (\Phi_1 \otimes \Phi_1) \text{vec}\Gamma(0) + \text{vec}\Sigma\end{aligned}$$

따라서 다음 식이 성립한다.

$$\text{vec}\Gamma(0) = (I_m - \Phi_1 \otimes \Phi_1)^{-1} \text{vec}\Sigma$$

이 식에서 역행렬이 존재하는 이유는 다음과 같다. Kronecker 곱 $\Phi_1 \otimes \Phi_1$ 의 특성값들은 행렬 Φ_1 의 특성값들의 곱으로 이루어져 있으므로, 작인성조건에 의해서 $\Phi_1 \otimes \Phi_1$ 의 모든 특

성값들의 절대값들은 1보다 작다. 따라서 $\det(I - \Phi_1 \otimes \Phi_1) \neq 0$ 이다. 행렬을 벡터화하는 vec 연산자에 대해서는 Muirhead(1982)의 제 2.2절을 참조하라.

한 예로 다음 3변량VAR(1)모형에서 자기공분산행렬들을 구해 보자.

$$y_t = \Phi_1 y_{t-1} + v_t$$

$$\text{단, } \Phi_1 = \begin{bmatrix} 0.5 & 0.0 & 0.0 \\ 0.1 & 0.1 & 0.3 \\ 0.0 & 0.2 & 0.3 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1.0 & -0.2 & 0.0 \\ -0.2 & 1.04 & 0.3 \\ 0.0 & 0.3 & 1.09 \end{bmatrix}$$

이 3변량VAR(1)과정의 자기공분산행렬들을 구하기 위해 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-9.PRG에 수록되어 있다).

```

/*
/*
/*          EXAM6-10.PRG          */
/*
/*          CALCULATE AUTOCOVARANCE MATRICES          */
/*          OF 3-VARIATE VAR(1) MODEL          */
/*
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOLOG NOCENTER FW=7;
/* INPUT MATRICES */
PHI1 = { 0.5 0.0 0.0,
        0.1 0.1 0.3,
        0.0 0.2 0.3 };
SIGMA = { 1.0 -0.2 0.0,
         -0.2 1.04 0.3,
         0.0 0.3 1.09 };
PRINT PHI1 " " SIGMA;
/* MAKE VEC(SIGMA) */
SIG1 = SIGMA[,1];
SIG2 = SIGMA[,2];
SIG3 = SIGMA[,3];
VECSIG = SIG1 // SIG2 // SIG3;
/* CALCULATE VEC(GAMMA(0)) */
I9 = I(9);
VECGAM = INV(I9 - PHI1@PHI1)*VECSIG;
PRINT "(I9 - PHI1@PHI1)*VEC(GAMMA(0)) = VEC(SIGMA)",;
PRINT "(I9 - PHI1@PHI1)";
PRINT (I9 - PHI1@PHI1),;

```

```

RESET FW=9;
PRINT, VECGAM " " VECSIG,;
/* CALCULATE AUTOCOVARANCE MATRICES */
GAM1 = VECGAM[{1 2 3},1];
GAM2 = VECGAM[{4 5 6},1];
GAM3 = VECGAM[{7 8 9},1];
GAMMA0 = GAM1 || GAM2 || GAM3;
GAMMA1 = PHI1 * GAMMA0;
GAMMA2 = PHI1 * GAMMA1;
GAMMA3 = PHI1 * GAMMA2;
GAMMA4 = PHI1 * GAMMA3;
GAMMA5 = PHI1 * GAMMA4;
PRINT GAMMA0 " " GAMMA1,
      GAMMA2 " " GAMMA3,
      GAMMA4 " " GAMMA5,;
QUIT;

```

이 SAS프로그램을 수행한 결과가 그림 6.10이다.

d 변량VAR(p)과정 $\{y_t\}$ 의 초기자기공분산행렬들 $\Gamma(1-p), \Gamma(2-p), \dots, \Gamma(0)$ 을 구하기 위해서, 제 6.3.1소절에서와 같이 VAR(p)과정을 다음과 같은 VAR(1)형태로 나타내자.

$$w_t = \Phi w_{t-1} + u_t$$

이 VAR(1)과정 $\{w_t\}$ 의 자기공분산행렬함수를 $\{\Gamma_w(h)\}$ 라고 하고 u_t 의 공분산행렬을 Σ_u 라고 하면 다음 식들이 성립한다.

$$\begin{aligned}
 & \Gamma_w(0) \\
 &= E \left\{ \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix} [y_t', y_{t-1}', \dots, y_{t-p+1}'] \right\} \\
 &= \begin{bmatrix} \Gamma(0) & \Gamma(1) & \dots & \Gamma(p-1) \\ \Gamma(-1) & \Gamma(0) & \dots & \Gamma(p-2) \\ \vdots & \vdots & & \vdots \\ \Gamma(1-p) & \Gamma(2-p) & \dots & \Gamma(0) \end{bmatrix} \\
 & \Sigma_u \\
 &= E(u_t u_t') \\
 &= E(J' v_t v_t' J) \\
 &= J' \Sigma J
 \end{aligned}$$

```

19:22 Thursday, March 3, 1994 19

PHI1                      SIGMA
0.5      0      0          1      -0.2     0
0.1      0.1    0.3        -0.2    1.04    0.3
0         0.2    0.3         0      0.5     1.09

(19 - PHI1@PHI1)*VEC(GAMMA(0)) = VEC(SIGMA)

(19 - PHI1@PHI1)
0.75      0      0      0      0      0      0      0      0
-0.05     0.95  -0.15   0      0      0      0      0      0
0         -0.1   0.85   0      0      0      0      0      0
-0.05     0      0      0.95  0      0      -0.15  0      0
-0.01    -0.01  -0.03  -0.01  0.99  -0.03  -0.03  -0.03  -0.09
0         -0.02  -0.03   0     -0.02  0.97   0     -0.06  -0.09
0         0      0     -0.1   0      0      0.85   0      0
0         0      0     -0.02  -0.02  -0.06  -0.03  0.97  -0.09
0         0      0      0     -0.04  -0.06   0     -0.06  0.91

VECGAM                      VECSIG
1.3333333                    1
-0.143007                    -0.2
-0.016824                     0
-0.143007                    -0.2
1.2088164                     1.04
0.4825525                     0.3
-0.016824                     0
0.4825525                     0.3
1.3145703                     1.09

GAMMA0                      GAMMA1
1.3333333  -0.143007  -0.016824  0.6666667  -0.071504  -0.008412
-0.143007  1.2088164  0.4825525  0.1139853  0.2513466  0.4409439
-0.016824  0.4825525  1.3145703  -0.033649  0.386529  0.4908816

GAMMA2                      GAMMA3
0.3333333  -0.035752  -0.004206  0.1666667  -0.017876  -0.002103
0.0679706  0.133943  0.1905176  0.0439411  0.0596875  0.0892671
0.0127024  0.166228  0.2354533  0.0174048  0.076657  0.1087395

GAMMA4                      GAMMA5
0.0833333  -0.008938  -0.001052  0.0416667  -0.004469  -0.000526
0.0262822  0.0271783  0.0413383  0.0151645  0.0123044  0.0191713
0.0140097  0.0349346  0.0504753  0.0094593  0.015916  0.0234102

Exiting IML.

```

그림 6.10 3변량VAR(1)과정의 자기공분산행렬들

여기서 $J = [I_d \ 0 \ 0 \ \dots \ 0]$ 이다. 따라서 $\Gamma(-i) = \Gamma(i)'$ 라는 점을 유의하면 행렬 $\Gamma_w(0)$ 은 초기자기공분산행렬들 $\Gamma(1-p), \Gamma(2-p), \dots, \Gamma(0)$ 으로 이루어졌음을 알 수 있다. VAR(1)모형에서와 같은 방법으로 다음 식을 유도할 수 있다.

$$\Gamma_w(0) = \Phi \Gamma_w(0) \Phi' + \Sigma_w$$

이 식으로부터 다음 식을 유도할 수 있다.

$$\begin{aligned}
 & \text{vec} \Gamma_w(0) \\
 &= (I_{d^2} - \Phi \otimes \Phi)^{-1} \text{vec} \Sigma_w \\
 &= (I_{d^2} - \Phi \otimes \Phi)^{-1} \text{vec}(J' \Sigma J)
 \end{aligned}$$

한 예로 다음 2변량VAR(2)모형을 살펴보자.

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + v_t$$

$$\text{단, } \Phi_1 = \begin{bmatrix} -0.26 & 0.54 \\ 0.0 & 0.48 \end{bmatrix}, \Phi_2 = \begin{bmatrix} -0.51 & 0.0 \\ -0.27 & 0.0 \end{bmatrix}, \Sigma = \begin{bmatrix} 1.0 & -0.2 \\ -0.2 & 1.04 \end{bmatrix}$$

이 2변량VAR(2)과정의 자기공분산행렬들을 구하기 위해 다음 SAS프로그램을 수행해보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-11.PRG에 수록되어 있다).

```

/*                                                    */
/*                EXAM6-11.PRG                */
/*                                                    */
/*                CALCULATE AUTOCOVARIANCE MATRICES                */
/*                OF 2-VARIATE VAR(2) MODEL                */
/*                                                    */
OPTIONS PAGESIZE=90;
PROC IML;
RESET NOLOG NOCENTER FW=7;
/* INPUT DATA */
PHI1 = { -0.26 0.54,
         0.0 0.48 };
PHI2 = { -0.51 0.0,
         -0.27 0.0 };
SIGMA = { 1.0 -0.2,
          -0.2 1.04 };
PRINT PHI1 " " PHI2 " " SIGMA,;
/* MAKE SPECIAL MATRICES */
I2 = I(2);
I4 = I(4);
I16 = I(16);
J2 = J(2, 2, 0);
/* MAKE PHI-MATRIX */
R1PHI = PHI1 || PHI2;
R2PHI = I2 || J2;
PHI = R1PHI // R2PHI;
PRINT PHI,;
/* MAKE SIGMA_U MATRIX */
R1SIGU = SIGMA || J2;
R2SIGU = J(2, 4, 0);
SIGMAU = R1SIGU // R2SIGU;
PRINT SIGMAU,;

```

```

/* MAKE VEC(SIGMA_U) */
SIGU1 = SIGMAU[,1];
SIGU2 = SIGMAU[,2];
SIGU3 = SIGMAU[,3];
SIGU4 = SIGMAU[,4];
VECSIGU = SIGU1 // SIGU2 // SIGU3 // SIGU4;
/* CALCULATE VEC(GAMMA_W(0)) */
VECGAM = INV(116 - PHI@PHI)*VECSIGU;
PRINT "CALCULATE VEC(GAMMA_{W}(0)), i. e.,"
      " (116-PHI@PHI)*VECGAM = VECSIGU ";
PRINT "(116-PHI@PHI)";
PRINT (116-PHI@PHI),;
RESET FW=9;
PRINT VECGAM " " VECSIGU,;
/* MAKE GAMMA_W(0) */
GAMW1 = VECGAM[1 2 3 4,1];
GAMW2 = VECGAM[5 6 7 8,1];
GAMW3 = VECGAM[9 10 11 12,1];
GAMW4 = VECGAM[13 14 15 16,1];
GAMMAW0 = GAMW1 || GAMW2 || GAMW3 || GAMW4;
PRINT GAMMAW0,;
/* FIND AUTOCOVARANCE FUNCTION OF Y */
GAMMA0 = GAMMAW0[1 2, 1 2];
GAMMA1 = GAMMAW0[1 2, 3 4];
GAMMA2 = PHI1*GAMMA1 + PHI2*GAMMA0;
GAMMA3 = PHI1*GAMMA2 + PHI2*GAMMA1;
GAMMA4 = PHI1*GAMMA3 + PHI2*GAMMA2;
GAMMA5 = PHI1*GAMMA4 + PHI2*GAMMA3;
/* PRINT AUTOCOVARANCE FUNCTION OF Y */
PRINT GAMMA0 " " GAMMA1,
      GAMMA2 " " GAMMA3,
      GAMMA4 " " GAMMA5,;
QUIT;

```

이 SAS프로그램을 수행한 결과가 그림 6.11이다.

지금부터는 $(I - \Phi \otimes \Phi)$ 의 역행렬을 구하지 않고 자기공분산행렬들을 구하는 방법에 대해 살펴보자. Lyapunov-Kalman-Stein의 정리를 사용하면 다음 행렬방정식을 축차적으로 풀 수 있다.

$$X = AXA' + B$$

19:22 Thursday, March 3, 1994 23

PHI1			PHI2		SIGMA	
-0.26	0.54		-0.51	0	1	-0.2
0	0.48		-0.27	0	-0.2	1.04

PHI			
-0.26	0.54	-0.51	0
0	0.48	-0.27	0
1	0	0	0
0	1	0	0

SIGMAU			
1	-0.2	0	0
-0.2	1.04	0	0
0	0	0	0
0	0	0	0

CALCULATE VEC(GAMMA - (W)(0)), i. e.,

$(I6 - PHI@PHI)*VECGAM = VECSIGU$

$(I6 - PHI@PHI)$

	0.9324	0.1404	-0.1326	0	0.1404	-0.2916	0.2754	0	-0.1326
:	0.2754	-0.2601	0	0	0	0	0	0	0
	0	1.1248	-0.0702	0	0	-0.2592	0.1458	0	0
:	0.2448	-0.1377	0	0	0	0	0	0	0
	0.26	0	1	0	-0.54	0	0	0	0.51
:	0	0	0	0	0	0	0	0	0
	0	0.26	0	1	0	-0.54	0	0	0
:	0.51	0	0	0	0	0	0	0	0
	0	0	0	0	1.1248	-0.2592	0.2448	0	-0.0702
:	0.1458	-0.1377	0	0	0	0	0	0	0
	0	0	0	0	0	0.7696	0.1296	0	0
:	0.1296	-0.0729	0	0	0	0	0	0	0
	0	0	0	0	-0.48	0	1	0	0.27
:	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	-0.48	0	1	0
:	0.27	0	0	0	0	0	0	0	0
	0.26	-0.54	0.51	0	0	0	0	0	1
:	0	0	0	0	0	0	0	0	0
	0	-0.48	0.27	0	0	0	0	0	0
:	1	0	0	0	0	0	0	0	0
	-1	0	0	0	0	0	0	0	0
:	0	1	0	0	0	0	0	0	0
	0	-1	0	0	0	0	0	0	0
:	0	0	1	0	0	0	0	0	0
	0	0	0	0	0.26	-0.54	0.51	0	0
:	0	0	0	1	0	0	0	0	0
	0	0	0	0	0	-0.48	0.27	0	0
:	0	0	0	0	1	0	0	0	0
	0	0	0	0	-1	0	0	0	0
:	0	0	0	0	0	1	0	0	0
	0	0	0	0	0	-1	0	0	0
:	0	0	0	0	0	0	1	0	0

VECGAM	VECSIGU
1.7582943	1
0.2924838	-0.2
-0.198156	0
0.609473	0
0.2924838	-0.2
1.4526017	1.04
0.1938943	0
0.6448973	0
-0.198156	0
0.1938943	0
1.7582943	0
0.2924838	0
0.609473	0
0.6448973	0

0.2924838			0
1.4526017			0
GAMMAW0			
1.7582943	0.2924838	-0.198156	0.609473
0.2924838	1.4526017	0.1938943	0.6448973
-0.198156	0.1938943	1.7582943	0.2924838
0.609473	0.6448973	0.2924838	1.4526017
GAMMA0		GAMMA1	
1.7582943	0.2924838	-0.198156	0.609473
0.2924838	1.4526017	0.1938943	0.6448973
GAMMA2		GAMMA3	
-0.740507	0.0406148	0.0874893	-0.196878
-0.38167	0.2305801	-0.1297	-0.053879
GAMMA4		GAMMA5	
0.2848734	0.0013799	-0.044339	0.0801618
0.137681	-0.036828	0.0424648	0.0354796
Exiting IML.			

그림 6.11 2변량VAR(2)과정의 자기공분산행렬들

알고리즘 6.1의 제2단계에서의 멈춤규칙(stopping rule)으로는 행렬 $(B_{k+1} - B_k)$ 의 노름(norm)이 주어진 양수 ϵ 보다 작은 경우에 알고리즘의 수행을 멈추는 것으로 한다. 행렬의 노름에 관한 자세한 내용은 Wilkinson(1965, pp. 55~61)을 참조하고, 알고리즘 6.1에 대한 자세한 내용은 Anderson & Moore(1979, pp. 64~68)와 Barone (1987)를 참조하라.

VAR(1)모형에서 Lyapunov-Kalman-Stein정리를 이용하기 위해서는 다음과 같이 놓는다.

$$A = \Phi_1, \quad B = \Sigma, \quad X = \Gamma(0)$$

알고리즘 6.1 : Lyapunov-Kalman-Stein 알고리즘

다음 알고리즘을 사용하면 방정식 $X = AXA' + B$ 를 풀 수 있다.

(1단계) 초기값으로 다음 식들을 사용한다.

$$A_1 = A$$

$$B_1 = B$$

(2단계) $k = 1, 2, \dots$ 에 대해 다음 식들을 축차적으로 적용한다.

$$A_{k+1} = A_k^2$$

$$B_{k+1} = A_k B_k A_k' + B_k$$

(3단계) 이렇게 얻어진 $\{B_k \mid k=1, 2, \dots\}$ 는 다음 식을 만족한다.

$$\lim_{k \rightarrow \infty} B_k = X$$

알고리즘 6.1을 사용해서 그림 6.10의 3변량VAR(1)과정의 자기공분산행렬들을 구하기 위해 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-12.PRG에 수록되어 있다).

```

/* */
/*          EXAM6-12.PRG          */
/* */
/*          CALCULATE AUTOCOVARANCE MATRICES          */
/*          OF 3-VARIATE VAR(1) MODEL          */
/*          USING LYAPUNOV-KALMAN-STEIN THEORM          */
/* */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOLOG NOCENTER;
/* INPUT MATRICES */
PHI1 = { 0.5 0.0 0.0,
        0.1 0.1 0.3,
        0.0 0.2 0.3 };
SIGMA = { 1.0 -0.2 0.0,
         -0.2 1.04 0.3,
         0.0 0.3 1.09 };
PRINT PHI1 " " SIGMA;
/* INITIAL STEP */
A = PHI1;
B = SIGMA;
OLDB = B*1.0E+10;
/* ITERATIVE STEP */
DO K = 1 TO 30
    UNTIL (MAX(ABS(OLDB - B)) < 1.0E-8);
    EPSILON = MAX(ABS(OLDB - B));
    PRINT , "NO. OF ITERATION = " K
          " ERROR = " EPSILON;
    OLDA = A;
    OLDB = B;
    A = A*A;
    B = OLDA*B*OLDA' + B;
END;
/* CALCULATE AUTOCOVARANCE MATRICES */
GAMMA0 = B;
GAMMA1 = PHI1*GAMMA0;

```

SAS				14:19 Sunday, October 18, 1992 6			
PHI1				SIGMA			
0.5	0	0		1	-0.2	0	
0.1	0.1	0.3		-0.2	1.04	0.3	
0	0.2	0.3		0	0.3	1.09	
NO. OF ITERATION =		K	ERROR =	EPSILON			
		1		1.09E10			
NO. OF ITERATION =		K	ERROR =	EPSILON			
		2		0.25			
NO. OF ITERATION =		K	ERROR =	EPSILON			
		3		0.078125			
NO. OF ITERATION =		K	ERROR =	EPSILON			
		4		0.005188			
NO. OF ITERATION =		K	ERROR =	EPSILON			
		5		0.0000203			
GAMMA0				GAMMA1			
1.3333333	-0.143007	-0.016824		0.6666667	-0.071504	-0.008412	
-0.143007	1.2088164	0.4825525		0.1139853	0.2513466	0.4409439	
-0.016824	0.4825525	1.3145703		-0.033649	0.386529	0.4908816	
GAMMA2				GAMMA3			
0.3333333	-0.035752	-0.004206		0.1666667	-0.017876	-0.002103	
0.0679706	0.133943	0.1905176		0.0439411	0.0596875	0.0892671	
0.0127024	0.166228	0.2354533		0.0174048	0.076657	0.1087395	
GAMMA4				GAMMA5			
0.0833333	-0.008938	-0.001052		0.0416667	-0.004469	-0.000526	
0.0262822	0.0271783	0.0413383		0.0151645	0.0123044	0.0191713	
0.0140097	0.0349346	0.0504753		0.0094593	0.015916	0.0234102	
Exiting IML.							

그림 6.12 3변량VAR(1)과정의 자기공분산행렬들

```

GAMMA2 = PHI1 * GAMMA1;
GAMMA3 = PHI1 * GAMMA2;
GAMMA4 = PHI1 * GAMMA3;
GAMMA5 = PHI1 * GAMMA4;
PRINT , , GAMMA0 " " GAMMA1,
      GAMMA2 " " GAMMA3,
      GAMMA4 " " GAMMA5;
QUIT;

```

이 SAS프로그램에서 멈춤규칙으로 다음 조건을 사용하였다.

$$\max_{i,j} |B_{k+1}(i,j) - B_k(i,j)| \leq \epsilon = 10^{-8}$$

여기서 $B(i,j)$ 는 행렬 B 의 (i,j) 원소를 뜻한다. 또한 비록 이 멈춤규칙이 만족되지 않더라도 31번 이상의 반복을 하지 않도록 프로그램하였다. 이 SAS프로그램을 수행한 결과가 그림 6.12이다. 그림 6.12에서 볼 수 있듯이 이 알고리즘은 수렴성이 크고 빠르다. 다섯 번째 반복단계에서 얻은 결과가 그림 6.10의 결과와 같음을 알 수 있다.

알고리즘 6.1을 VAR(p)모형에 적용하기 위해서 다음과 같이 놓는다.

$$A = \Phi, \quad B = J'\Sigma J, \quad X = \Gamma_w(0)$$

알고리즘 6.1을 사용해서 그림 6.11의 2변량VAR(2)과정의 자기공분산행렬들을 구하기 위해 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-13.PRG에 수록되어 있다).

```

/ * * /
/ *          EXAM6-13.PRG          * /
/ * * /
/ *          CALCULATE AUTOCOVARIANCE MATRICES          * /
/ *          OF 2-VARIATE VAR(2) MODEL          * /
/ *          BY LYAPUNOV-KALMAN-STEIN THEORM          * /
OPTIONS PAGESIZE = 60;
PROC IML;
RESET NOLOG NOCENTER;
/ * INPUT DATA * /
PHI1 = { -0.26 0.54,
         0.0 0.48 };
PHI2 = { -0.51 0.0,
         -0.27 0.0 };
SIGMA = { 1.0 -0.2,
          -0.2 1.04 };
PRINT PHI1 " " PHI2 " " SIGMA,;
/ * MAKE SPECIAL MATRICES * /
I2 = I(2);
I4 = I(4);
I16 = I(16);
J2 = J(2, 2, 0);
/ * MAKE PHI-MATRIX * /
R1PHI = PHI1 || PHI2;
R2PHI = I2 || J2;
PHI = R1PHI // R2PHI;
PRINT PHI,;
/ * MAKE SIGMA_U MATRIX * /
R1SIGU = SIGMA || J2;
R2SIGU = J(2, 4, 0);
SIGMAU = R1SIGU // R2SIGU;
PRINT SIGMAU,;

```

```

/* INITIAL STEP */
A = PHI;
B = SIGMAU;
OLDB = B*1.0E+10;
/* ITERATIVE STEP */
DO K = 1 TO 30
  UNTIL ( MAX(ABS(OLDB - B)) < 1.0E-8);
  EPSILON = MAX(ABS(OLDB - B));
  PRINT , "NO. OF ITERATION = " K
        " ERROR = " EPSILON;
  OLDA = A;
  OLDB = B;
  A = A*A;
  B = OLDA*B*OLDA' + B;
END;
/* MAKE GAMMA_W(0) */
GAMMAW0 = B;
PRINT , GAMMAW0 ,;
/* FIND AUTOCOVARANCE FUNCTION OF Y */
GAMMA0 = GAMMAW0[{1 2}, {1 2}];
GAMMA1 = GAMMAW0[{1 2}, {3 4}];
GAMMA2 = PHI1*GAMMA1 + PHI2*GAMMA0;
GAMMA3 = PHI1*GAMMA2 + PHI2*GAMMA1;
GAMMA4 = PHI1*GAMMA3 + PHI2*GAMMA2;
GAMMA5 = PHI1*GAMMA4 + PHI2*GAMMA3;
/* PRINT AUTOCOVARANCE FUNCTION OF Y */
PRINT GAMMA0 " " GAMMA1,
      GAMMA2 " " GAMMA3,
      GAMMA4 " " GAMMA5;
QUIT;

```

이 SAS 프로그램을 수행한 결과가 그림 6.13이다. 그림 6.13의 자기공분산함수들은 그림 6.11의 자기공분산함수들과 동일함을 알 수 있다.

6.3.4 Whittle알고리즘

만일 $VAR(p)$ 과정의 자기공분산행렬함수 $\{\Gamma(h)\}$ 가 주어지면 Yule-Walker 방정식들을 풀어서 VAR 계수행렬들을 구할 수 있다. 이러한 과정을 단변량시계열분석에서는 Le-

```

SAS
14:19 Sunday, October 18, 1992 7

      PHI1          PHI2
    -0.26      0.54    -0.51      0
       0       0.48    -0.27      0

      SIGMA
       1      -0.2
    -0.2      1.04

      PHI
    -0.26      0.54    -0.51      0
       0       0.48    -0.27      0
       1         0         0         0
       0         1         0         0

      SIGMAU
       1      -0.2      0         0
    -0.2      1.04      0         0
       0         0         0         0
       0         0         0         0

      NO. OF ITERATION =          K      ERROR =      EPSILON
                                1      1.04E10

      NO. OF ITERATION =          K      ERROR =      EPSILON
                                2      1.04

      NO. OF ITERATION =          K      ERROR =      EPSILON
                                3      0.6584426

      NO. OF ITERATION =          K      ERROR =      EPSILON
                                4      0.0972208

      NO. OF ITERATION =          K      ERROR =      EPSILON
                                5      0.0026292

      NO. OF ITERATION =          K      ERROR =      EPSILON
                                6      1.71E-6

      GAMMA0
    1.7582943  0.2924838  -0.198156  0.609473
    0.2924838  1.4526017  0.1938943  0.6448973
    -0.198156  0.1938943  1.7582943  0.2924838
    0.609473  0.6448973  0.2924838  1.4526017

      GAMMA0          GAMMA1
    1.7582943  0.2924838  -0.198156  0.609473
    0.2924838  1.4526017  0.1938943  0.6448973

      GAMMA2          GAMMA3
    -0.740507  0.0406148  0.0874893  -0.196878
    -0.38167  0.2305801  -0.1297  -0.053879

      GAMMA4          GAMMA5
    0.2848734  0.0013799  -0.044339  0.0801618
    0.137681  -0.036828  0.0424648  0.0354796

      Exiting IML.
  
```

그림 6.13 2변량VAR(2)과정의 자기공분산행렬들

vinson-Durbin알고리즘을 사용하는데 다변량시계열분석에서는 이 알고리즘을 확장한 Whittle알고리즘(1963)을 사용한다. Whittle알고리즘을 사용해야 하는 이유는 전향VAR 모형(forward VAR model)과 후향VAR모형(backward VAR model)이 다른 형태의 Yule-Walker방정식들을 만족하기 때문이다.

다변량시계열데이터 $\{y_t\}$ 가 다음과 같은 전향 d 변량VAR(p)모형을 만족한다고 하자.

$$y_t = \Phi_1^p y_{t-1} + \Phi_2^p y_{t-2} + \dots + \Phi_p^p y_{t-p} + v_t$$

여기서 $\{v_t\}$ 는 공분산행렬이 Σ_p 인 d 변량백색잡음과정이다. 이 전향VAR(p)모형의 자기공분산행렬함수 $\{\Gamma(h)\}$ 는 다음과 같은 전향Yule-Walker방정식들을 만족한다.

$$\begin{aligned} \Gamma(i) &= \Phi_1^p \Gamma(i-1) + \Phi_2^p \Gamma(i-2) + \dots + \Phi_p^p \Gamma(i-p) \quad (i = 1, 2, \dots) \\ \Gamma(0) &= \Phi_1^p \Gamma(-1) + \Phi_2^p \Gamma(-2) + \dots + \Phi_p^p \Gamma(-p) + \Sigma_p \end{aligned}$$

따라서 초기자기공분산행렬들 $\Gamma(0), \Gamma(1), \dots, \Gamma(p)$ 가 주어지면 VAR계수행렬들 $\Phi_1^p, \Phi_2^p, \dots, \Phi_p^p$ 와 다변량백색잡음과정의 공분산행렬 Σ_p 를 구할 수 있다. 즉, $i = 1, 2, \dots, p$ 에 대해서 전향Yule-Walker방정식들을 풀어서 $\Phi_1^p, \Phi_2^p, \dots, \Phi_p^p$ 를 구하고 이 VAR계수행렬들을 다음 식에 대입해서 Σ_p 를 구한다.

$$\Sigma_p = \Gamma(0) - \Phi_1^p \Gamma(-1) - \Phi_2^p \Gamma(-2) - \dots - \Phi_p^p \Gamma(-p)$$

이 전향VAR(p)모형은 다음과 같이 나타낼 수 있다.

$$\begin{aligned} \Phi^p(B)y_t &= v_t \\ \text{단, } \Phi^p(x) &= -\Phi_0^p - \Phi_1^p x - \Phi_2^p x^2 - \dots - \Phi_p^p x^p \quad (\Phi_0^p = -I_d) \end{aligned}$$

이 전향VAR(p)모형에 연관된 다음과 같은 후향VAR(p)모형을 살펴보자.

$$\Pi^p(B^{-1})z_t = u_t$$

여기서 $\Pi^p(x) = -\Pi_0^p - \Pi_1^p x - \dots - \Pi_p^p x^p$ ($\Pi_0^p = -I_d$)이며, $\{u_t\}$ 는 d 변량백색잡음과정으로서 공분산행렬이 Ω_p 이다.

다음 등식이 만족된다고 가정하자.

$$\Pi^p(x) = (\Phi_p^p)^{-1} \Phi^p \left(\frac{1}{x} \right) x^p$$

이 가정하에서 $\{y_t\}$ 와 $\{z_t\}$ 가 같은 공분산구조(covariance structure)를 갖는다는 것을 Wiener-Khinchine관계를 사용해서 증명할 수 있다(이에 대한 자세한 내용은 [Choi92] p. 105를 참조하라). 전향VAR(p)모형의 특성방정식 $\det[\Phi^p(x)] = 0$ 의 모든 근들의 절대값들이 1보다 크므로 후향VAR(p)모형의 특성방정식 $\det[\Phi^p(x)] = 0$ 의 모든 근들의 절대값들이 1보다 작다. 따라서 $\{z_t\}$ 는 현재와 미래의 다변량백색잡음항들의 선형결합으로 나타낼 수 있다. 이 선형결합을 이용하면 다음 식이 성립함을 알 수 있다.

$$\begin{aligned} E(u_t z_{t-k}') &= 0 \quad (k = -1, -2, \dots) \\ E(u_t z_t') &= \Omega_p \end{aligned}$$

이 식들을 이용하면 후향VAR(p)과정 $\{z_t\}$ 가 다음 식을 만족한다.

$$E[\{\Pi^p(B^{-1})z_t\}z_{t-k}'] = E(u_t z_{t-k}')$$

따라서 후향VAR(p)과정 $\{z_t\}$ 는 다음 후향Yule-Walker방정식들을 만족한다.

$$\Gamma(k) = \Pi_1^p \Gamma(k+1) + \Pi_2^p \Gamma(k+2) + \dots + \Pi_p^p \Gamma(k+p) \quad (k = -1, -2, \dots)$$

또한 후향VAR(p)모형의 다변량백색잡음과정 $\{u_t\}$ 의 공분산행렬 Ω_p 는 다음 식을 만족한다.

$$\Omega_p = \Gamma(0) - \Pi_1^p \Gamma(1) - \Pi_2^p \Gamma(2) - \dots - \Pi_p^p \Gamma(p)$$

VAR(p)과정의 초기자기공분산행렬들 $\Gamma(0), \Gamma(1), \dots, \Gamma(p)$ 가 주어졌을 때 VAR계수행렬들 $\Phi_1^p, \Phi_2^p, \dots, \Phi_p^p$ 와 다변량백색잡음과정의 공분산행렬 Σ_p 를 구하기 위해서는 Whittle 알고리즘을 적용하면 된다. Levinson-Durbin알고리즘과는 달리 Whittle알고리즘은 상응하는 후향VAR과정의 VAR계수행렬들 $\Pi_1^p, \Pi_2^p, \dots, \Pi_p^p$ 와 다변량백색잡음과정의 공분산행렬 Ω_p 를 동시에 계산한다. Whittle알고리즘은 [SAS4TSA2] (p. 410)에서 기술한 단순Trench-Zohar알고리즘의 확장이라고 할 수 있다.

자기공분산행렬함수 $\{\Gamma(h)\}$ 가 주어졌을 때 Whittle알고리즘을 적용해서 VAR계수행렬들을 구하는 예로써 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4 TSA4의 파일 EXAM6-14.PRG에 수록되어 있다).

알고리즘 6.2 : Whittle알고리즘

(제 1단계) 초기행렬들로 다음 식들을 사용한다.

$$\Phi_0^0 = \Pi_0^0 = -I_d$$

$$G_0 = \Gamma(1)$$

$$\Sigma_0 = \Omega_0 = \Gamma(0)$$

(제 2단계) $k = 0, 1, \dots$ 에 대해서 다음 식들을 축차적으로 적용한다.

$$\Phi_{k+1}^{k+1} = G_k \Omega_k^{-1}$$

$$\Phi_j^{k+1} = \Phi_j^k - \Phi_{k+1}^{k+1} (\Pi^{k+1})_{k+1-j}^k \quad (j = 1, 2, \dots, k)$$

$$\Pi_{k+1}^{k+1} = G_k' \Sigma_k^{-1}$$

$$\Pi_j^{k+1} = \Pi_j^k - \Pi_{k+1}^{k+1} \Phi_{k+1-j}^k \quad (j = 1, 2, \dots, k)$$

$$G_{k+1} = \Gamma(k+2) - \sum_{j=1}^{k+1} \Phi_{k+2-j}^{k+1} \Gamma(j)$$

$$\Sigma_{k+1} = (I - \Phi_{k+1}^{k+1} \Pi_{k+1}^{k+1}) \Sigma_k$$

$$\Omega_{k+1} = (I - \Pi_{k+1}^{k+1} \Phi_{k+1}^{k+1}) \Omega_k$$

```

/* */
/*          EXAM6-14.PRG          */
/* */
/*          WHITTLE'S ALGORITHM    */
/* */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOLOG NOCENTER;
/* INPUT DATA */
C0 = { 1.3333333 -0.143007 -0.016824,
      -0.143007 1.2088164 0.4825525,
      -0.016824 0.4825525 1.3145703 };
C1 = { 0.6666667 -0.071504 -0.008412,
      0.1139853 0.2513466 0.4409439,
      -0.033649 0.386529 0.4908816 };
C2 = { 0.3333333 -0.035752 -0.004206,
      0.0679706 0.133943 0.1905176,
      0.0127024 0.166228 0.2354533 };
C3 = { 0.1666667 -0.017876 -0.002103,
      0.0439411 0.0596875 0.0892671,
      0.0174048 0.076657 0.1087395 };
C4 = { 0.0833333 -0.008938 -0.001052,
      0.0262822 0.0271783 0.0413383,
      0.0140097 0.0349346 0.0504753 };
C5 = { 0.0416667 -0.004469 -0.000526,
      0.0151645 0.0123044 0.0191713,
      0.0094593 0.015916 0.0234102 };
C = C0 || C1 || C2 || C3 || C4 || C5;
D = 3;          /* DIMENSION OF SUBMATRICES */
MAXORDER = 4;  /* MAX ORDR OF VAR MODEL */
/* CONSTANTS */
IDEN = I(D);
AROW = D;
ACOL = D*(MAXORDER+1);
L = J(AROW, ACOL, 0);
M = L;
G = J(D, D, 0);
LAMBDA = G;
V = G;

```

```

AROWS = 1:D;
ACOLS = 1:D;
/* INITIAL STEP K=0 */
L[AROWS, ACOLS] = -IDEN;
M[AROWS, ACOLS] = -IDEN;
ACOLSNXT = (D+1):(2*D);
G = C[AROWS, ACOLSNXT];
LAMBDA = C[AROWS, ACOLS];
V = LAMBDA;
/* ITERATIVE STEP */
DO K = 1 TO MAXORDER;
  LORD = L;
  MOLD = M;
  L[AROWS, ACOLS] = -IDEN;
  M[AROWS, ACOLS] = -IDEN;
  ACOLSLST = (K*D+1):(K+1)*D;
  L[AROWS, ACOLSLST] = G*INV(V);
  M[AROWS, ACOLSLST] = G'*INV(LAMBDA);
  DO J = 1 TO K-1;
    ACOLSJ = (J*D+1):(J+1)*D;
    ACOLSKMJ = ((K-J)*D+1):((K-J+1)*D);
    L[AROWS, ACOLSJ] = LORD[AROWS, ACOLSJ]
      - L[AROWS, ACOLSLST]*MOLD[AROWS, ACOLSKMJ];
    M[AROWS, ACOLSJ] = MOLD[AROWS, ACOLSJ]
      - M[AROWS, ACOLSLST]*LORD[AROWS, ACOLSKMJ];
  END;
  ACOLSNXT = ((K+1)*D+1):(K+2)*D;
  G = C[AROWS, ACOLSNXT];
  DO J = 1 TO K;
    ACOLSJ = (J*D+1):((J+1)*D);
    ACOLSKMJ = ((K-J+1)*D+1):((K-J+2)*D);
    G = G - L[AROWS, ACOLSKMJ]*C[AROWS, ACOLSJ];
  END;
  LAMBDA = (IDEN -
    L[AROWS, ACOLSLST]*M[AROWS, ACOLSLST])*LAMBDA;
  V = (IDEN -
    M[AROWS, ACOLSLST]*L[AROWS, ACOLSLST])*V;
/* PRINT THE SOLUTION */
PRINT,, "ORDER = " K,,

```

```

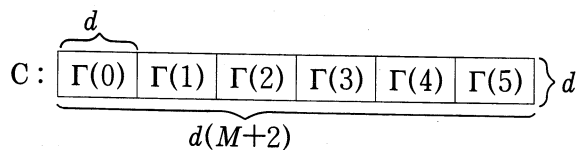
LAMBDA " " V,;
DO J = 1 TO K;
  ACOLSJ = (J*D+1):(J+1)*D;
  LKJ = L[AROWS, ACOLSJ];
  MKJ = M[AROWS, ACOLSJ];
  PRINT "LAG = " J,
        LKJ " " MKJ, ;
END;
END;
QUIT;

```

이 SAS프로그램들의 변수이름들은 알고리즘 6.2의 표기와 다르다. 이렇게 다른 변수이름들을 사용하는 이유는 다음 절에서 다루게 될 알고리즘 6.2의 확장을 고려해서이다. 이 SAS프로그램의 각 단계를 설명하면 다음과 같다.

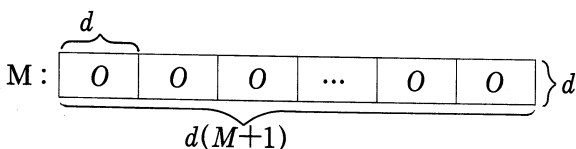
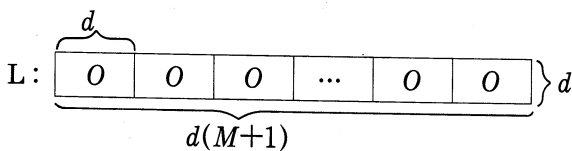
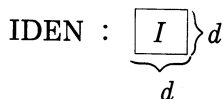
(1) INPUT DATA단계

자기공분산함수들 C_0, C_1, \dots, C_5 와 다변량시계열데이터의 차원 D , 그리고 구하고자 하는 VAR모형들의 최고차수 MAXORDER를 입력한다. 즉, $\Gamma(h)$ 를 변수 Ch 로, d 를 변수 D 로, 최고차수 M 을 변수 MAXORDER로 표시한다. 이 단계에서 유의할 점은 입력되는 자기공분산함수들의 최고시차(이 SAS프로그램에서는 5)가 최고차수 M (이 SAS프로그램에서는 4)보다 커야 한다. 입력된 자기공분산행렬들을 가로로 연결한 것이 행렬 C 이다.



(2) CONSTANTS단계

이 단계에서는 행렬의 차원과 초기값들을 지정한다.



$$G : \underbrace{\boxed{O}}_d \} d$$

$$\text{LAMBDA} : \underbrace{\boxed{O}}_d \} d$$

$$V : \underbrace{\boxed{O}}_d \} d$$

$$\text{AROWS} = \{1\ 2\ \dots\ d\}$$

$$\text{ACOLS} = \{1\ 2\ \dots\ d\}$$

(3) INITIAL STEP단계

이 단계에서는 $k=0$ 에 해당하는 값들을 할당한다.

$$L : \boxed{-I} \ \boxed{O} \ \boxed{O} \ \dots \ \boxed{O} \ \boxed{O}$$

$$M : \boxed{-I} \ \boxed{O} \ \boxed{O} \ \dots \ \boxed{O} \ \boxed{O}$$

$$\text{ACOLSNXT} = \{d+1\ d+2\ \dots\ 2d\}$$

$$G : \boxed{\Gamma(1)}$$

$$\text{LAMBDA} : \boxed{\Sigma_0}$$

$$V : \boxed{\Omega_0}$$

(4) ITERATIVE STEP단계

이 단계에서는 DO루프를 사용해서 $k=1, 2, \dots, M$ 에 대해 다음 계산을 수행한다.

$$\text{LOLD} : \boxed{\Phi_0^{k-1}} \ \boxed{\Phi_1^{k-1}} \ \dots \ \boxed{\Phi_{k-1}^{k-1}} \ \boxed{O} \ \dots \ \boxed{O}$$

$$\text{MOLD} : \boxed{\Pi_0^{k-1}} \ \boxed{\Pi_1^{k-1}} \ \dots \ \boxed{\Pi_{k-1}^{k-1}} \ \boxed{O} \ \dots \ \boxed{O}$$

$$\text{ACOLSLST} = \{kd+1\ kd+2\ \dots\ (k+1)d\}$$

$$L : \boxed{-I} \ \boxed{\Phi_1^k} \ \dots \ \boxed{\Phi_{k-1}^k} \ \boxed{\Phi_k^k} \ \dots \ \boxed{O}$$

$$M : \boxed{-I} \ \boxed{\Pi_1^k} \ \dots \ \boxed{\Pi_{k-1}^k} \ \boxed{\Pi_k^k} \ \dots \ \boxed{O}$$

① 이 DO루프 안의 첫 번째 중첩 DO루프를 $j=1, 2, \dots, k-1$ 에 대해서 수행한 결과

로 다음과 같은 행렬들을 얻는다.

$$L \quad : \quad \begin{array}{|c|c|c|c|c|c|c|} \hline -I & \Phi_1^k & \cdots & \Phi_{k-1}^k & \Phi_k^k & \cdots & O \\ \hline \end{array}$$

$$M \quad : \quad \begin{array}{|c|c|c|c|c|c|c|} \hline -I & \Pi_1^k & \cdots & \Pi_{k-1}^k & \Pi_k^k & \cdots & O \\ \hline \end{array}$$

② 두 번째 중첩DO루프 전까지의 스테이트먼트들을 수행하면 다음과 같은 결과를 얻는다.

$$ACOLSNXT = \{(k+1)d+1 \ (k+1)d+2 \ \dots \ (k+2)d\}$$

$$G \quad : \quad \boxed{\Gamma(k+1)}$$

③ 두 번째 중첩DO루프를 $j = 1, 2, \dots, k$ 에 대해서 수행한 결과로 다음 행렬을 얻는다.

$$G \quad : \quad \boxed{G_{k+1}}$$

④ 두 번째 중첩DO루프 다음의 두 스테이트먼트들을 수행하면 다음과 같은 결과를 얻는다.

$$LAMBDA: \quad \boxed{\Sigma_k}$$

$$V \quad : \quad \boxed{\Omega_k}$$

⑤ 다음의 PRINT스테이트먼트를 수행하면 다음과 같은 결과를 출력시킨다.

$$K \quad : \quad \boxed{k}$$

$$LAMBDA \quad : \quad \boxed{\Sigma_k}$$

$$V \quad : \quad \boxed{\Omega_k}$$

⑥ 세 번째 중첩DO루프는 $j = 1, 2, \dots, k$ 에 대해 다음 행렬들을 출력시킨다.

$$LKJ \quad : \quad \boxed{\Phi_j^k}$$

$$MKJ \quad : \quad \boxed{\Pi_j^k}$$

이 SAS프로그램을 수행한 결과가 그림 6.14이다. 그림 6.12에서 확인할 수 있듯이 C0, C1, ..., C5는 다음 3변량VAR(1)모형의 자기공분산행렬들이다.

$$y_t = \Phi_1 y_{t-1} + v_t$$

$$\text{단, } \Phi_1 = \begin{bmatrix} 0.5 & 0.0 & 0.0 \\ 0.1 & 0.1 & 0.3 \\ 0.0 & 0.2 & 0.3 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1.0 & -0.2 & 0.0 \\ -0.2 & 1.04 & 0.3 \\ 0.0 & 0.3 & 1.09 \end{bmatrix}$$

그림 6.14에서 볼 수 있듯이 Whittle알고리즘을 적용한 결과 다음 행렬들을 얻었다.

① VAR(1)모형을 적용

$$\Sigma_1 = \begin{bmatrix} 1.0 & -0.2 & 0.0 \\ -0.2 & 1.04 & 0.3 \\ 0.0 & 0.3 & 1.09 \end{bmatrix}, \quad \Phi_1^1 = \begin{bmatrix} 0.5 & 0.0 & 0.0 \\ 0.1 & 0.1 & 0.3 \\ 0.0 & 0.2 & 0.3 \end{bmatrix}$$

② VAR(2)모형을 적용

$$\Sigma_2 = \begin{bmatrix} 1.0 & -0.2 & 0.0 \\ -0.2 & 1.04 & 0.3 \\ 0.0 & 0.3 & 1.09 \end{bmatrix}, \quad \Phi_1^2 = \begin{bmatrix} 0.5 & 0.0 & 0.0 \\ 0.1 & 0.1 & 0.3 \\ 0.0 & 0.2 & 0.3 \end{bmatrix}, \quad \Phi_2^2 = \begin{bmatrix} 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 \end{bmatrix}$$

③ VAR(3)모형을 적용

$$\Sigma_3 = \begin{bmatrix} 1.0 & -0.2 & 0.0 \\ -0.2 & 1.04 & 0.3 \\ 0.0 & 0.3 & 1.09 \end{bmatrix}, \quad \Phi_1^3 = \begin{bmatrix} 0.5 & 0.0 & 0.0 \\ 0.1 & 0.1 & 0.3 \\ 0.0 & 0.2 & 0.3 \end{bmatrix}$$

$$\Phi_2^3 = \begin{bmatrix} 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 \end{bmatrix}, \quad \Phi_3^3 = \begin{bmatrix} 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 \end{bmatrix}$$

④ VAR(4)모형을 적용

$$\Sigma_4 = \begin{bmatrix} 1.0 & -0.2 & 0.0 \\ -0.2 & 1.04 & 0.3 \\ 0.0 & 0.3 & 1.09 \end{bmatrix}, \quad \Phi_1^4 = \begin{bmatrix} 0.5 & 0.0 & 0.0 \\ 0.1 & 0.1 & 0.3 \\ 0.0 & 0.2 & 0.3 \end{bmatrix}$$

$$\Phi_2^4 = \begin{bmatrix} 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 \end{bmatrix}, \quad \Phi_3^4 = \begin{bmatrix} 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 \end{bmatrix}, \quad \Phi_4^4 = \begin{bmatrix} 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 \end{bmatrix}$$

그림 6.14의 결과로부터 주어진 자기공분산행렬함수가 3변량VAR(1)모형의 자기공분산행렬함수임을 확인할 수 있다.

```

SAS
22:12 Friday, October 23, 1992 13

ORDER = K
1

LAMBDA
0.9999999 -0.2 6E-7
-0.2 1.0400001 0.3
6E-7 0.3 1.09

V
0.9622567 -0.119488 -0.053046
-0.119488 1.0814906 0.3118355
-0.053046 0.3118355 1.0645631

LAG = J
1

LKJ
0.5 -4.863E-7 1.7835E-7
0.0999999 0.0999999 0.3
-3.075E-7 0.1999999 0.3

MKJ
0.5193974 0.191345 -0.089189
-0.039575 0.1008494 0.2565082
0.0246801 0.2560172 0.2797531

ORDER = K
2

LAMBDA
0.9999999 -0.2 6E-7
-0.2 1.0400001 0.3
6E-7 0.3 1.09

V
0.9622567 -0.119488 -0.053046
-0.119488 1.0814906 0.3118355
-0.053046 0.3118355 1.0645631

LAG = J
1

LKJ
0.5 -5.232E-7 1.44E-7
0.0999998 0.0999999 0.3
-4.48E-7 0.1999999 0.3

MKJ
0.5193974 0.1913449 -0.089189
-0.039575 0.1008495 0.2565082
0.0246801 0.2560172 0.2797531

LAG = J
2

LKJ
2.9116E-8 1.8477E-8 1.1512E-7
1.9631E-7 8.4974E-8 -4.569E-8
2.6533E-7 -2.406E-8 7.1705E-8

MKJ
4.4794E-8 1.2546E-7 1.9886E-7
6.8624E-8 8.1104E-8 -5.477E-8
1.2192E-7 -2.428E-8 5.6916E-8

ORDER = K
3

LAMBDA
0.9999999 -0.2 6E-7
-0.2 1.0400001 0.3
6E-7 0.3 1.09

V
0.9622567 -0.119488 -0.053046
-0.119488 1.0814906 0.3118355
-0.053046 0.3118355 1.0645631

LAG = J
1

LKJ
0.5 -5.232E-7 1.44E-7
0.0999998 0.0999999 0.3
-4.48E-7 0.1999999 0.3

MKJ
0.5193974 0.1913449 -0.089189
-0.039575 0.1008495 0.2565082
0.0246801 0.2560172 0.2797531

LAG = J
2

LKJ
-1.222E-8 1.72E-9 1.1976E-7
2.3941E-7 1.1085E-7 -4.423E-8
2.9523E-7 -8.058E-9 7.3216E-8

MKJ
1.6826E-8 1.3751E-7 2.2444E-7
6.5403E-8 8.138E-8 -5.752E-8
1.2963E-7 -1.928E-8 7.0661E-8

LAG = J
3
    
```



```

      LKJ
      7.9884E-8 6.0231E-9 3.3754E-9
      -7.883E-8 2.254E-8 -5.104E-8
      -5.735E-8 -7.701E-9 -1.662E-8

      MKJ
      6.5953E-8 -5.009E-8 -3.519E-8
      2.2323E-9 2.1055E-8 -1.191E-8
      -7.09E-9 -4.162E-8 -4.193E-9

ORDER =          K
              4

      LAMBDA
      0.9999999 -0.2 6E-7
      -0.2 1.0400001 0.3
      6E-7 0.3 1.09

              V
      0.9622567 -0.119488 -0.053046
      -0.119488 1.0814906 0.3118355
      -0.053046 0.3118355 1.0645631

LAG =          J
              1

      LKJ
      0.5 -5.232E-7 1.44E-7
      0.0999998 0.0999999 0.3
      -4.48E-7 0.1999999 0.3

      MKJ
      0.5193974 0.1913449 -0.089189
      -0.039575 0.1008495 0.2565082
      0.0246801 0.2560172 0.2797531

              J
LAG =          2

      LKJ
      -1.222E-8 1.72E-9 1.1976E-7
      2.3941E-7 1.1085E-7 -4.423E-8
      2.9523E-7 -8.058E-9 7.3216E-8

      MKJ
      1.6826E-8 1.3751E-7 2.2444E-7
      6.5403E-8 8.138E-8 -5.752E-8
      1.2963E-7 -1.928E-8 7.0661E-8

              J
LAG =          3

      LKJ
      1.4492E-7 1.4128E-7 1.0374E-7
      -7.154E-8 1.0827E-8 -7.42E-8
      -9.151E-8 -2.881E-8 -2.029E-8

      MKJ
      1.1504E-7 -5.858E-8 -3.96E-8
      -1.239E-8 1.781E-8 -2.7E-8
      2.5558E-7 -4.511E-8 -3.83E-10

              J
LAG =          4

      LKJ
      -8.994E-8 1.4084E-7 -5.166E-7
      -1.231E-8 4.5714E-8 3.6941E-8
      6.391E-8 -2.293E-9 3.5592E-8

      MKJ
      -8.708E-8 -5.553E-8 7.0222E-8
      1.5616E-8 6.8185E-8 -1.787E-8
      -5.133E-7 -6.029E-8 4.7589E-8

Exiting IML.

```

그림 6.14 Whittle알고리즘

Whittle알고리즘을 SAS/IML에서 서브루틴으로 만들어 사용하기로 하자. 한 예로 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM 6-15.PRG에 수록되어 있다).

```

/* */
/*          EXAM6-15.PRG          */
/* */
/*          MODULE OF WHITTLE'S ALGORITHM          */
/*          i.e., MODULE WHITTLE          */
/* */

OPTIONS PAGESIZE=60;
PROC IML;
RESET NOLOG NOCENTER;

```

```

/ *                               */
/ *   MODULE OF WHITTLE ALGORITHM */
/ *                               */
START WHITTLE(C, D, MAXORDER);
/ *   CONSTANTS           */
IDEN = I(D);
AROW = D;
ACOL = D*(MAXORDER+1);
L = J(AROW, ACOL, 0);
M = L;
G = J(D, D, 0);
LAMBDA = G;
V = G;
AROWS = 1:D;
ACOLS = 1:D;
/ *   INITIAL STEP K=0       */
L[AROWS, ACOLS] = -IDEN;
M[AROWS, ACOLS] = -IDEN;
ACOLSNXT = (D+1):(2*D);
G = C[AROWS, ACOLSNXT];
LAMBDA = C[AROWS, ACOLS];
V = LAMBDA;
/ *   ITERATIVE STEP       */
DO K = 1 TO MAXORDER;
  LOD = L;
  MOLD = M;
  L[AROWS, ACOLS] = -IDEN;
  M[AROWS, ACOLS] = -IDEN;
  ACOLSLST = (K*D+1):(K+1)*D;
  L[AROWS, ACOLSLST] = G*INV(V);
  M[AROWS, ACOLSLST] = G*INV(LAMBDA);
  DO J = 1 TO K-1;
    ACOLSJ = (J*D+1):(J+1)*D;
    ACOLSKMJ = ((K-J)*D+1):((K-J+1)*D);
    L[AROWS, ACOLSJ] = LOD[AROWS, ACOLSJ]
      - L[AROWS, ACOLSLST]*MOLD[AROWS, ACOLSKMJ];
    M[AROWS, ACOLSJ] = MOLD[AROWS, ACOLSJ]
      - M[AROWS, ACOLSLST]*LOLD[AROWS, ACOLSKMJ];
  END;
END;

```

```

ACOLSNXT = ((K+1)*D+1):(K+2)*D;
G = C[AROWS, ACOLSNXT];
DO J = 1 TO K;
    ACOLSJ = (J*D+1):((J+1)*D);
    ACOLSKMJ = ((K-J+1)*D+1):((K-J+2)*D);
    G = G-L[AROWS, ACOLSKMJ]*C[AROWS, ACOLSJ];
    END;
LAMBDA = (IDEN -
    L[AROWS, ACOLSLST]*M[AROWS, ACOLSLST])*LAMBDA;
V = (IDEN -
    M[AROWS, ACOLSLST]*L[AROWS, ACOLSLST])*V;
/*      PRINT THE SOLUTION      */
PRINT,, "ORDER = " K,,
    LAMBDA " " V,;
DO J = 1 TO K;
    ACOLSJ = (J*D+1):(J+1)*D;
    LKJ = L[AROWS, ACOLSJ];
    MKJ = M[AROWS, ACOLSJ];
    PRINT "LAG = " J,
        LKJ " " MKJ, ;
    END;
END;
FINISH;
/*      INPUT DATA      */
GAMMA0 = { 1.7582943 0.2924838,
           0.2924838 1.4526017 };
GAMMA1 = { -0.198156 0.609473,
           0.1938943 0.6448973 };
GAMMA2 = { -0.740507 0.0406148,
           -0.38167 0.2305801 };
GAMMA3 = { 0.0874893 -0.196878,
           -0.1297 -0.053879 };
GAMMA4 = { 0.2848734 0.0013799,
           0.137681 -0.036828 };
GAMMA5 = { -0.044339 0.0801618,
           0.0424648 0.0354796 };
GAMMA = GAMMA0 || GAMMA1 || GAMMA2
        || GAMMA3 || GAMMA4 || GAMMA5;

```

```
M = 2;          /* DIMENSION OF SUBMATRICES */
MAXORDER = 4;  /* MAX ORDR OF VAR MODEL   */
RUN WHITTLE(GAMMA, M, MAXORDER);
QUIT;
```

이 SAS프로그램의 START스테이트먼트부터 FINISH스테이트먼트까지가 WHITTLE이라는 이름을 가진 SAS/IML모듈(module)이다. 이 모듈을 실행하기 위해서는 RUN스테이트먼트를 사용하는데, START스테이트먼트의 매개변수들(arguments)과 RUN스테이트먼트의 매개변수값들은 일대일 대응을 해야 한다. 이 SAS프로그램에서는 다음과 같이 매개변수들과 매개변수값들이 일대일 대응을 한다.

```
START WHITTLE( C,   D, MAXORDER) ;
                ↓   ↓     ↓
RUN  WHITTLE(GAMMA, M, MAXORDER) ;
```

이 예제에서는 M=2, MAXORDER=4이므로 다음과 같은 RUN스테이트먼트를 사용해도 된다.

```
RUN WHITTLE(GAMMA, 2, 4) ;
```

SAS/IML모듈에 대한 자세한 내용은 제3.4.4소절을 참조하라.

SAS		22:12 Friday, October 23, 1992 16			
ORDER =	K				
	1				
LAMBDA		V			
1.4419894	0.0339944	1.699306	0.2733465		
0.0339944	1.1638795	0.2733465	1.0309294		
LAG=	J				
	1				
LKJ		MKJ			
-0.188816	0.4575918	-0.139577	0.1615847		
0.0376857	0.4363721	0.2822297	0.3871327		
ORDER=	K				
	2				
LAMBDA		V			
0.9999995	-0.2	1.0132124	0.1629831		
-0.2	1.0400002	0.1629831	1.0131767		
LAG =	J				
	1				
LKJ		MKJ			
-0.26	0.54	-0.237175	0.59701		
-1.025E-8	0.4799999	0.2665304	0.4571742		
LAG =	J				
	2				

```

      LKJ              MKJ
      -0.51 1.4338E-7  -0.592122 -0.376915
      -0.27 2.3891E-8  -0.095247 -0.06063

ORDER =              K
                   3

      LAMBDA          V
      0.9999995      -0.2      1.0132124 0.1629831
      -0.2      1.0400002      0.1629831 1.0131767

LAG =              J
                   1

      LKJ              MKJ
      -0.260001 0.5399998  -0.237175 0.59701
      -3.218E-7 0.4799997  0.2665304 0.4571742

LAG =              J
                   2

      LKJ              MKJ
      -0.51 4.6646E-7  -0.592122 -0.376914
      -0.27 2.3646E-7  -0.095247 -0.06063

LAG =              J
                   3

      LKJ              MKJ
      -5.382E-7 -3.821E-9  -6.744E-7 -6.424E-7
      -5.715E-7 2.8135E-7  -5.688E-8 1.7359E-7

ORDER =              K
                   4

      LAMBDA          V
      0.9999995      -0.2      1.0132124 0.1629831
      -0.2      1.0400002      0.1629831 1.0131767

LAG =              J
                   1

      LKJ              MKJ
      -0.260001 0.5399998  -0.237175 0.59701
      -3.218E-7 0.4799997  0.2665304 0.4571742

LAG =              J
                   2

      LKJ              MKJ
      -0.51 4.0965E-7  -0.592122 -0.376914
      -0.27 2.1979E-7  -0.095247 -0.06063

LAG =              J
                   3

      LKJ              MKJ
      -5.645E-7 9.7452E-8  -7.181E-7 -5.156E-7
      -5.703E-7 3.2161E-7  -7.42E-8 2.3654E-7

LAG =              J
                   4

      LKJ              MKJ
      -1.457E-7 -3.126E-8  -1.678E-7 -7.535E-8
      -3.805E-8 -3.838E-8  -6.665E-8 -5.617E-8

Exiting IML.

```

그림 6.15 모듈화한 Whittle알고리즘

이 SAS프로그램을 수행한 결과가 그림 6.15이다. 이 SAS프로그램의 입력데이터인 GAMMA0, GAMMA1, ..., GAMMA5는 그림 6.13에서 볼 수 있듯이 다음 2변량VAR(2) 모형의 자기공분산행렬들이다.

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + v_t$$

$$\text{단, } \Phi_1 = \begin{bmatrix} -0.26 & 0.54 \\ 0.0 & 0.48 \end{bmatrix}, \Phi_2 = \begin{bmatrix} -0.51 & 0.0 \\ -0.27 & 0.0 \end{bmatrix}, \Sigma = \begin{bmatrix} 1.0 & -0.2 \\ -0.2 & 1.04 \end{bmatrix}$$

WHITTLE알고리즘을 사용해서 얻은 결과가 이 모형과 적합한가를 확인해 보라.

6.3.5 블럭토우플리츠행렬*

기본차원(basic dimension)이 $d \times e$ 인 $p \times q$ 블럭행렬은 $d \times e$ 부행렬들(submatrices)로 구성된 $pd \times qe$ 행렬을 의미한다. 이 절에서는 각 블럭행렬은 $d \times d$ 부행렬들로 이루어져 있다고 가정한다. 다음과 같은 $k \times k$ 블럭토우플리츠행렬(block Toeplitzmatrix) C_k 를 살펴보자.

$$C_k = \begin{bmatrix} C_0 & C_{-1} & \cdots & C_{-k+1} \\ C_1 & C_0 & \cdots & C_{-k+2} \\ \vdots & \vdots & & \vdots \\ C_{k-1} & C_{k-2} & \cdots & C_0 \end{bmatrix}$$

여기서 $C_i, (i=1-k, \dots, k-1)$ 은 $d \times d$ 행렬이다. 지금부터 모든 k 에 대해 C_k 가 정칙행렬이라고 가정한다. 이 소절의 목적은 블럭토우플리츠행렬 C_k 의 역행렬을 블럭LU분해(block LU decomposition)하고, 이 분해를 위한 알고리즘을 제공하는 것이다. Whittle알고리즘은 이 알고리즘의 특수한 경우이다.

만일 A 가 $p \times q$ 블럭행렬이고, 이 행렬의 (i, j) 블럭이 A_{ij} 라면 블럭전치행렬(block transposed matrix) A^T 는 블럭행렬로서 (i, j) 블럭이 A_{ji} 이다. 또한 블럭교환행렬(block exchange matrix) E_n 은 $n \times n$ 블럭행렬로서 만일 $i+j=n+1$ 이면 (i, j) 블럭이 단위행렬 I_d 이고 그렇지 않으면 (i, j) 블럭은 영행렬이다. 또한 $\tilde{A} = AE_q$ 라 정의한다.

$d \times d$ 행렬들 $L_{k,j}, U_{k,j}, M_{k,j}, N_{k,j} (k=1, 2, \dots, j=1, 2, \dots, k)$ 를 다음 연립방정식들의 해(solution)로 정의한다.

$$\begin{aligned} C_j &= L_{k,1}C_{j-1} + L_{k,2}C_{j-2} + \cdots + L_{k,k}C_{j-k} & (j=1, \dots, k) \\ C_j &= C_{j+1}U_{k,1} + C_{j+2}U_{k,2} + \cdots + C_{j+k}U_{k,k} & (j=-1, \dots, -k) \\ C_j &= M_{k,1}C_{j+1} + M_{k,2}C_{j+2} + \cdots + M_{k,k}C_{j+k} & (j=-1, \dots, -k) \\ C_j &= C_{j-1}N_{k,1} + C_{j-2}N_{k,2} + \cdots + C_{j-k}N_{k,k} & (j=1, \dots, k) \end{aligned}$$

행렬 C_k 가 정칙이므로 이 연립방정식들의 해는 유일하게 정의된다. 만일 $\{C_j | j=0, 1,$

...}이 다변량시계열데이터의 자기공분산행렬함수라면 첫 번째와 네 번째 연립방정식들은 전향Yule-Walker방정식들에 해당되고, 두 번째와 세 번째 연립방정식들은 후향Yule-Walker방정식들에 해당된다.

다음과 같은 $d \times d$ 행렬들을 정의하자.

$$L_{k,0} = U_{k,0} = M_{k,0} = N_{k,0} = -I_d \quad (k = 0, 1, \dots)$$

또한 블록삼각행렬(block triangular matrix)들 L_k , M_k , U_k , N_k 와 블록대각행렬(block diagonal matrix)들 Δ_k 와 ∇_k 를 다음과 같이 정의하자.

$$L_k = \begin{bmatrix} L_{0,0} & O & O & \cdots & O \\ L_{1,1} & L_{1,0} & O & \cdots & O \\ L_{2,2} & L_{2,1} & L_{2,0} & \cdots & O \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ L_{k-1,k-1} & L_{k-1,k-2} & L_{k-1,k-3} & \cdots & L_{k-1,0} \end{bmatrix}$$

$$M_k = \begin{bmatrix} M_{0,0} & O & O & \cdots & O \\ M_{1,1} & M_{1,0} & O & \cdots & O \\ M_{2,2} & M_{2,1} & M_{2,0} & \cdots & O \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ M_{k-1,k-1} & M_{k-1,k-2} & M_{k-1,k-3} & \cdots & M_{k-1,0} \end{bmatrix}$$

$$U_k = \begin{bmatrix} U_{0,0} & U_{1,1} & U_{2,2} & \cdots & U_{k-1,k-1} \\ O & U_{1,0} & U_{2,1} & \cdots & U_{k-1,k-2} \\ O & O & U_{2,0} & \cdots & U_{k-1,k-3} \\ O & O & O & \cdots & U_{k-1,0} \end{bmatrix}$$

$$N_k = \begin{bmatrix} N_{0,0} & N_{1,1} & N_{2,2} & \cdots & N_{k-1,k-1} \\ O & N_{1,0} & N_{2,1} & \cdots & N_{k-1,k-2} \\ O & O & N_{2,0} & \cdots & N_{k-1,k-3} \\ O & O & O & \cdots & N_{k-1,0} \end{bmatrix}$$

$$\Delta_k = \begin{bmatrix} \Lambda_0^{-1} & O & O & \cdots & O \\ O & \Lambda_1^{-1} & O & \cdots & O \\ O & O & \Lambda_2^{-1} & \cdots & O \\ O & O & O & \cdots & \Lambda_{k-1}^{-1} \end{bmatrix}$$

$$\nabla_k = \begin{bmatrix} V_0^{-1} & O & O & \cdots & O \\ O & V_1^{-1} & O & \cdots & O \\ O & O & V_2^{-1} & \cdots & O \\ O & O & O & \cdots & V_{k-1}^{-1} \end{bmatrix}$$

$$\begin{aligned} \text{단, } \Lambda_0 &= V_0 = C_0 \\ \Lambda_i &= -(L_{i,0}C_0 + L_{i,1}C_{-1} + \cdots + L_{i,i}C_{-i}) \quad (i = 1, \dots, k-1) \\ V_i &= -(M_{i,0}C_0 + M_{i,1}C_1 + \cdots + M_{i,i}C_i) \quad (i = 1, \dots, k-1) \end{aligned}$$

정리 6.7 블록LU분해

$$\begin{aligned} C_k^{-1} &= U_k \Lambda_k L_k \\ (C_k^T)^{-1} &= N_k \nabla_k M_k \end{aligned}$$



정리 6.7의 블록LU분해들을 $k = 1, 2, \dots$ 에 대해서 축차적으로 수행하는 알고리즘을 설명하기 위해서 우선 다음과 같은 행렬들을 정의하자.

$$\begin{aligned} L_k &= (L_{k,1}, L_{k,2}, \dots, L_{k,k})^T \\ U_k &= (U_{k,1}, U_{k,2}, \dots, U_{k,k})^T \end{aligned}$$

알고리즘 6.3 : 블록트우플리츠행렬의 분해를 위한 알고리즘

(제1단계) 초기행렬들로 다음 식들을 사용한다.

$$\begin{aligned} L_{0,0} &= U_{0,0} = M_{0,0} = N_{0,0} = -I_d \\ G_0 &= C_1 \\ H_0 &= C_{-1} \\ \Lambda_0 &= V_0 = C_0 \end{aligned}$$

(제2단계) $k = 0, 1, \dots$ 에 대해 다음 식들을 축차적으로 적용한다.

$$\begin{aligned} L_{k+1} &= \begin{pmatrix} L_k - L_{k+1,k+1} \tilde{M}_k \\ G_k V_k^{-1} \end{pmatrix} \\ U_{k+1} &= \begin{pmatrix} U_k - \tilde{N}_k U_{k+1,k+1} \\ V_k^{-1} H_k \end{pmatrix} \\ M_{k+1} &= \begin{pmatrix} M_k - M_{k+1,k+1} \tilde{L}_k \\ H_k \Lambda_k^{-1} \end{pmatrix} \\ N_{k+1} &= \begin{pmatrix} N_k - \tilde{U}_k N_{k+1,k+1} \\ \Lambda_k^{-1} G_k \end{pmatrix} \\ G_{k+1} &= C_{k+2} - \tilde{L}_{k+1}^T S_{k+1} \\ H_{k+1} &= C_{-k-2} - M_{k+1}^T T_{k+1} \\ \Lambda_{k+1} &= (I_d - L_{k+1,k+1} M_{k+1,k+1}) \Lambda_k \\ V_{k+1} &= V_k (I_d - U_{k+1,k+1} N_{k+1,k+1}) \end{aligned}$$

$$\begin{aligned}
M_k &= (M_{k,1}, M_{k,2}, \dots, M_{k,k})^T \\
N_k &= (N_{k,1}, N_{k,2}, \dots, N_{k,k})^T \\
S_k &= (C_1, C_2, \dots, C_k)^T \\
T_k &= (C_{-k}, C_{1-k}, \dots, C_{-1})^T \\
G_k &= C_{k+1} - \tilde{S}_k^T C_k^{-1} S_k \\
H_k &= C_{-k-1} - \tilde{T}_k^T C_k^{-1} T_k
\end{aligned}$$

이 행렬들은 다음 식들을 만족한다.

$$\begin{aligned}
\tilde{L}_k^T &= \tilde{S}_k^T C_k^{-1} \\
\tilde{U}_k &= C_k^{-1} T_k \\
M_k^T &= \tilde{T}_k^T C_k^{-1} \\
N_k &= C_k^{-1} S_k \\
G_k &= C_{k+1} - \tilde{L}_k^T S_k = C_{k+1} - \tilde{S}_k^T N_k \\
H_k &= C_{-k-1} - \tilde{T}_k^T \tilde{U}_k = C_{-k-1} - M_k^T T_k \\
\Lambda_k &= C_0 - \tilde{L}_k^T T_k = C_0 - \tilde{S}_k^T \tilde{U}_k \\
V_k &= C_0 - M_k^T S_k = C_0 - \tilde{T}_k^T N_k
\end{aligned}$$

이 행렬들을 사용해서 알고리즘 6.3을 유도할 수 있다.

알고리즘 6.4 : 대칭인 블럭토우폴리츠행렬의 LU분해를 위한 알고리즘

(제1단계) 초기행렬들로 다음 식들을 사용한다.

$$\begin{aligned}
L_{0,0} &= M_{0,0} = -I_d \\
G_0 &= C_1 \\
\Lambda_0 &= V_0 = C_0
\end{aligned}$$

(제2단계) $k = 0, 1, \dots$ 에 대해 다음 식들을 축차적으로 적용한다.

$$\begin{aligned}
L_{k+1} &= \begin{pmatrix} L_k - L_{k+1,k+1} \tilde{M}_k \\ G_k V_k^{-1} \end{pmatrix} \\
M_{k+1} &= \begin{pmatrix} M_k - M_{k+1,k+1} \tilde{L}_k \\ G_k^t \Lambda_k^{-1} \end{pmatrix} \\
G_{k+1} &= C_{k+2} - \tilde{L}_{k+1}^T S_{k+1} \\
\Lambda_{k+1} &= (I_d - L_{k+1,k+1} M_{k+1,k+1}) \Lambda_k \\
V_{k+1} &= (I_d - M_{k+1,k+1} L_{k+1,k+1}) V_k
\end{aligned}$$

만일 블럭토우플리츠행렬 C_k 가 대칭이면 모든 (i, j) 에 대해 다음 식들이 성립한다.

$$U_{i,j} = L_{i,j}'$$

$$N_{i,j} = M_{i,j}'$$

$$H_{i,j} = G_{i,j}'$$

$$\Lambda_j = \Lambda_j'$$

$$V_j = V_j'$$

따라서 이 경우에는 알고리즘 6.3을 알고리즘 6.4와 같이 간단히 할 수 있다.

알고리즘 6.4는 Whittle알고리즘을 좀더 다듬은 알고리즘이다. 정리 6.7, 알고리즘 6.3과 알고리즘 6.4에 대한 자세한 내용은 [Choi92] (pp. 79~84)와 [Choi] (1990)를 참조하라.

6.4 VMA모형

6.4.1 가역적 VMA모형

정상성을 갖는 다변량시계열데이터를 분석하는 데 사용되는 모형 중의 하나가 VMA모형 (vector moving-average model, 벡터이동평균모형)이다. 차수가 q 인 d 변량VMA과정 $\{z_t\}$ 는 다음 d 변량VMA(q)모형을 만족한다.

$$z_t = \mu + v_t - \Theta_1 v_{t-1} - \dots - \Theta_q v_{t-q} \quad (t = 0, \pm 1, \pm 2, \dots)$$

여기서 Θ 는 $d \times d$ 행렬로서 VMA계수행렬이라 부르며, μ 는 d 차원열벡터이고, $\{v_t\}$ 는 평균벡터가 영벡터이고 공분산행렬이 Σ 인 d 변량백색잡음과정이다. 특별한 언급이 없는 한 Σ 는 양정치행렬(positive definiteness matrix)이라고 가정한다.

편의상 다음과 같은 새로운 다변량시계열데이터 $\{y_t\}$ 를 정의한다.

$$y_t = z_t - \mu$$

이 다변량시계열데이터 $\{y_t\}$ 는 다음과 같은 d 변량VMA(q)모형을 만족한다.

$$y_t = v_t - \Theta_1 v_{t-1} - \Theta_2 v_{t-2} - \dots - \Theta_q v_{t-q} \quad (t = 0, \pm 1, \pm 2, \dots)$$

이 다변량시계열데이터 $\{y_t\}$ 의 평균벡터는 영벡터이다. 이후 특별한 언급이 없는 한 다변량시계열데이터 $\{z_t\}$ 대신 평균벡터가 영벡터인 다변량시계열데이터 $\{y_t\}$ 를 다루기로 한다.

다음과 같은 d 변량VMA(1)모형을 살펴보자.

$$y_t = v_t - \Theta_1 v_{t-1}$$

이 식으로부터 다음 식을 얻을 수 있다.

$$v_t = y_t + \Theta_1 y_{t-1} + \Theta_1^2 y_{t-2} + \dots + \Theta_1^n y_{t-n} + \Theta_1^{n+1} v_{t-n-1}$$

따라서 다음 등식이 성립한다.

$$\begin{aligned} & E\left\{ \left(v_t - \sum_{k=0}^n \Theta_1^k y_{t-k} \right) \left(v_t - \sum_{k=0}^n \Theta_1^k y_{t-k} \right)' \right\} \\ &= E\left\{ \left(\Theta_1^{n+1} v_{t-n-1} \right) \left(\Theta_1^{n+1} v_{t-n-1} \right)' \right\} \\ &= \Theta_1^{n+1} \Sigma \left(\Theta_1^{n+1} \right)' \end{aligned}$$

만일 $n \rightarrow \infty$ 일 때 $\Theta_1^{n+1} \rightarrow 0$ 이면, 다음 식이 성립한다.

$$y_t = - \sum_{k=1}^{\infty} \Theta_1^k y_{t-k} + v_t$$

이 식의 우변의 무한급수는 평균제곱수렴한다.

VMA과정 $\{y_t\}$ 를 VAR(∞)과정으로 나타낼 수 있으면 이 다변량시계열데이터는 가역적 (invertible)이라고 한다. 만일, 행렬 Θ_1 의 모든 특성값들의 절대값들이 1보다 작다면, 다음 식이 성립한다.

$$\lim_{n \rightarrow \infty} \Theta_1^{n+1} = 0$$

행렬의 특성값의 정의로부터 행렬 Θ_1 의 모든 특성값들의 절대값들이 1보다 작기 위한 필요충분조건은 다음 방정식의 모든 근이 단위원 밖에 있는 것이다.

$$\det(I - \Theta_1 x) = 0$$

즉, $|x| \leq 1$ 에서 $\det(I - \Theta_1) \neq 0$ 이다. 이 조건을 VMA(1)모형의 가역조건이라고 한다.

VMA(1)과정을 VAR(∞)모형으로 나타내는 예제를 살펴보기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-16.PRG에 수록되어 있다).

```

/* */
/* EXAM6-16.PRG */
/* */
/* CALCULATE AR COEFFICIENT MATRICES */
/* OF 3-VARIATE VMA(1) MODEL */
PROC IML;
RESET NOLOG NOCENTER;
THETA1 = { 0.5 0.5 -0.4,
           0.4 -0.7 0.3,
           0.0 0.2 0.5 };
PIO = THETA1**0;

```

SAS			8:30 Monday, October 26, 1992 3		
PI0			PI1		
1	0	0	0.5	0.5	-0.4
0	1	0	0.4	-0.7	0.3
0	0	1	0	0.2	0.5
PI2			PI3		
0.45	-0.18	-0.25	0.153	0.301	-0.359
-0.08	0.75	-0.22	0.26	-0.609	0.147
0.08	-0.04	0.31	0.024	0.13	0.111
PI4			PI5		
0.1969	-0.206	-0.1504	0.01605	0.21257	-0.21576
-0.1136	0.5857	-0.2132	0.17748	-0.50943	0.11455
0.064	-0.0568	0.0849	0.00928	0.08874	-0.00019
PI6			PI7		
0.093053	-0.183926	-0.050529	-0.027044	0.1651689	-0.117664
-0.115032	0.468251	-0.166546	0.1297844	-0.418601	0.1032151
0.040136	-0.057516	0.022815	-0.002938	0.0648922	-0.021902
Exiting IML.					

그림 6.16 VMA과정의 VAR(∞)표현

```

PI1 = PI0*THETA1;
PI2 = PI1*THETA1;
PI3 = PI2*THETA1;
PI4 = PI3*THETA1;
PI5 = PI4*THETA1;
PI6 = PI5*THETA1;
PI7 = PI6*THETA1;
PRINT PI0 " " PI1, PI2 " " PI3,
      PI4 " " PI5, PI6 " " PI7, ;
QUIT;
    
```

이 SAS프로그램을 수행한 결과가 그림 6.16이다.

다음과 같은 d 변량VMA(q)과정의 가역조건을 살펴보자.

$$y_t = u_t - \Theta_1 v_{t-1} - \Theta_2 v_{t-2} - \dots - \Theta_q v_{t-q}$$

이 d 변량VMA(q)과정은 다음과 같이 dq 변량VMA(1)모형으로 나타낼 수 있다.

$$w_t = u_t - \Theta u_{t-1}$$

$$\text{단, } w_t := \begin{bmatrix} y_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \Theta = \begin{bmatrix} \Theta_1 & \Theta_2 & \dots & \Theta_{q-1} & \Theta_q \\ I & 0 & \dots & 0 & 0 \\ 0 & I & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & I & 0 \end{bmatrix}, u_t = \begin{bmatrix} v_t \\ v_{t-1} \\ v_{t-2} \\ \vdots \\ v_{t-q+1} \end{bmatrix}$$

이 dq 변량VMA(1)과정 $\{w_t\}$ 가 가역적이기 위해서는 다음 방정식의 모든 근들의 절대값들이 1보다 커야 한다.

$$\det(I - \Theta x) = 0$$

제6.3.1소절에서와 같은 방법을 사용해서 다음 등식을 얻을 수 있다.

$$\det(I - \Theta x) = \det\left(I - \sum_{k=1}^q \Theta_k x^k\right)$$

따라서 dq 변량VMA(1)과정 $\{w_t\}$ 의 가역조건은 $|x| \leq 1$ 에서 다음 조건이 성립하는 것이다.

$$\det\left(I - \sum_{k=1}^q \Theta_k x^k\right) \neq 0$$

다음 식을 d 변량VMA(q)과정 $\{y_t\}$ 의 VMA특성행렬이라 정의한다.

$$\Theta(x) = - \sum_{k=0}^q \Theta_k x^k \quad (\Theta_0 = -I)$$

또한 $\det[\Theta(x)]$ 를 VMA특성식이라 한다. 여기서 유의할 점은 행렬 Θ 의 특성식은 $\det[\Theta(x)]$ 가 아니라 $\det[(-x)^q \Theta(x^{-1})]$ 라는 것이다. 따라서 일부 학자들은 $\det[\Theta(x)]$ 를 VMA(q)과정의 역특성식이라고 부르기도 한다.

가역성을 갖는 d 변량VMA(q)과정을 VAR(∞)모형으로 나타내기 위해서 다음과 같은 $d \times dq$ 행렬을 정의하자.

$$J := [I \ 0 \ 0 \ \dots \ 0]$$

행렬 J 를 사용하면 다음 식이 성립함을 알 수 있다.

$$\begin{aligned} & v_t \\ &= Ju_t \\ &= J \left(\sum_{k=0}^{\infty} \Theta^k w_{t-k} \right) \\ &= J \left(\sum_{k=0}^{\infty} \Theta^k J' J w_{t-k} \right) \\ &= \sum_k (J \Theta^k J') y_{t-k} \\ &= - \sum_{k=0}^{\infty} \Pi_k y_{t-k} \end{aligned}$$

여기서 $\Pi_k = -J \Theta^k J'$, ($k=0, 1, 2, \dots$)이다. 행렬 Θ 의 모든 특성값들의 절대값들이 1보다 작으므로 $\sum_{k=1}^{\infty} \Pi_k y_{t-k}$ 는 평균제곱수렴을 하고, 두 확률벡터들 y_t 와 $(\sum_{k=1}^{\infty} \Pi_k y_{t-k} + v_t)$ 는 L^2 공간에서 동일하다. 결론적으로 가역성을 갖는 d 변량VMA(q)과정 $\{y_t\}$ 는 다음과 같은 d 변량VAR(∞)모형을 만족한다.

$$\Pi(B)y_t = v_t$$

$$\text{단, } \Pi(x) = -\sum_{k=1}^{\infty} \Pi_k x^k \quad (\Pi_0 = -I_d)$$

또한 이 d 변량VMA(q)과정은 다음 식을 만족한다.

$$y_t = \Theta(B)v_t$$

따라서 다음 식이 성립한다.

$$\Theta^{-1}(x) = \Pi(x)$$

가역조건에 의해서 $|x| \leq 1$ 에서 $\det[\Theta(x)] \neq 0$ 이므로, $|x| \leq 1$ 인 모든 x 에 대해 다음 항등식이 성립한다.

$$\Pi(x)\Theta(x) = I_d$$

항등식의 원리에 의해 다음 식들이 성립한다.

$$\Pi_0 = -I_d$$

$$\Pi_1 = \Pi_0 \Theta_1$$

$$\Pi_2 = \Pi_1 \Theta_1 + \Pi_0 \Theta_2$$

⋮

$$\Pi_q = \Pi_{q-1} \Theta_1 + \Pi_{q-2} \Theta_2 + \dots + \Pi_0 \Theta_q$$

$$\Pi_i = \Pi_{i-1} \Theta_1 + \Pi_{i-2} \Theta_2 + \dots + \Pi_{i-q} \Theta_q \quad (i = q+1, q+2, \dots)$$

이 식들을 이용하면 d 변량VMA(q)과정을 d 변량VAR(∞)모형으로 나타낼 때의 VAR계수행렬들 $\{\Pi_i\}$ 를 구할 수 있다.

한 예로 다음과 같은 3변량VMA(3)과정을 살펴보자.

$$y_t = v_t - \Theta_1 v_{t-1} - \Theta_2 v_{t-2} - \Theta_3 v_{t-3}$$

$$\text{단, } \Theta_1 = \begin{bmatrix} 0.5 & 0.1 & 0.0 \\ 0.4 & 0.3 & 0.0 \\ 0.0 & 0.0 & 0.3 \end{bmatrix}, \quad \Theta_2 = \begin{bmatrix} 0.0 & 0.0 & 0.0 \\ 0.25 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.28 \end{bmatrix}, \quad \Theta_3 = \begin{bmatrix} 0.0 & 0.12 & -0.34 \\ 0.0 & 0.0 & 0.20 \\ 0.22 & 0.31 & -0.01 \end{bmatrix}$$

이 3변량VMA(3)과정의 VAR(∞)표현을 구하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-17.PRG에 수록되어 있다).

```

/* */
/*          EXAM6-17.PRG          */
/* */
/*          CALCULATE AR COEFFICIENT MATRICES          */
/*          OF 3-VARIATE VMA(3) MODEL          */
/* */
OPTIONS PAGESIZE = 60;
PROC IML;

```

```

RESET NOLOG NOCENTER;
THETA1 = { 0.5 0.1 0.0,
           0.4 0.3 0.0,
           0.0 0.0 0.3 };
THETA2 = { 0.0 0.0 0.0,
           0.25 0.0 0.0,
           0.0 0.0 0.28 };
THETA3 = { 0.0 0.12 -0.34,
           0.0 0.0 0.20,
           0.22 0.31 -0.01 };

PI0 = -THETA1**0;
PI1 = PI0*THETA1;
PI2 = PI1*THETA1 + PI0*THETA2;
PI3 = PI2*THETA1 + PI1*THETA2 + PI0*THETA3;
PI4 = PI3*THETA1 + PI2*THETA2 + PI1*THETA3;
PI5 = PI4*THETA1 + PI3*THETA2 + PI2*THETA3;
PI6 = PI5*THETA1 + PI4*THETA2 + PI3*THETA3;
PI7 = PI6*THETA1 + PI5*THETA2 + PI4*THETA3;
PRINT , "(A) PI(X)*THETA(X) = I",;
PRINT PI0 " " PI1, PI2 " " PI3,
       PI4 " " PI5, PI6 " " PI7, ;

/*      COMMUTING THETA & PI      */
PI0 = -THETA1**0;
PI1 = THETA1*PI0;
PI2 = THETA1*PI1 + THETA2*PI0;
PI3 = THETA1*PI2 + THETA2*PI1 + THETA3*PI0;
PI4 = THETA1*PI3 + THETA2*PI2 + THETA3*PI1;
PI5 = THETA1*PI4 + THETA2*PI3 + THETA3*PI2;
PI6 = THETA1*PI5 + THETA2*PI4 + THETA3*PI3;
PI7 = THETA1*PI6 + THETA2*PI5 + THETA3*PI4;
PRINT , "(B) THETA(X)*PI(X) = I",;
PRINT PI0 " " PI1, PI2 " " PI3,
       PI4 " " PI5, PI6 " " PI7, ;

QUIT;

```

이 SAS프로그램을 수행한 결과가 그림 6.17이다. 그림 6.17의 부분 A는 다음 항등식에서 얻은 것이다.

$$\Pi(x)\Theta(x) = I_d$$

```

SAS
23:18 Monday, March 7, 1994 1
(A)
PI(X)*THETA(X) = I
      P10          P11
      -1           0           0          -0.5         -0.1          0
      0           -1           0          -0.4         -0.3          0
      0            0          -1           0            0         -0.3

      P12          P13
      -0.29        -0.08         0         -0.202        -0.173         0.34
      -0.57        -0.13         0         -0.412        -0.096         -0.2
      0             0         -0.37        -0.22         -0.31         -0.185

      P14          P15
      -0.1902      -0.1321        0.252      -0.19119      -0.09345        0.2534
      -0.2769      -0.118         0.016      -0.20965      -0.13149        0.1166
      -0.3         -0.208         -0.1561    -0.3921      -0.2071        -0.09493

      P16          P17
      -0.0912      0.034006        0.17726    0.0000799    0.0563778       0.159858
      -0.230921   -0.171852        0.16234   -0.213554   -0.102916       0.151736
      -0.37159    -0.18509        -0.057537 -0.345948   -0.177077       0.0181195

(B)
THETA(X)*PI(X) = I
      P10          P11
      -1           0           0          -0.5         -0.1          0
      0           -1           0          -0.4         -0.3          0
      0            0          -1           0            0         -0.3

      P12          P13
      -0.29        -0.08         0         -0.202        -0.173         0.34
      -0.57        -0.13         0         -0.412        -0.096         -0.2
      0             0         -0.37        -0.22         -0.31         -0.185

      P14          P15
      -0.1902      -0.1321        0.252      -0.19119      -0.09345        0.2534
      -0.2769      -0.118         0.016      -0.20965      -0.13149        0.1166
      -0.3         -0.208         -0.1561    -0.3921      -0.2071        -0.09493

      P16          P17
      -0.0912      0.034006        0.17726    0.0000799    0.0563778       0.159858
      -0.230921   -0.171852        0.16234   -0.213554   -0.102916       0.151736
      -0.37159    -0.18509        -0.057537 -0.345948   -0.177077       0.0181195

Exiting IML.
    
```

그림 6.17 3변량VMA(3)과정의 VAR(∞)표현

부분 B는 다음 항등식에서 얻은 것이다.

$$\Theta(x)\Pi(x) = I_d$$

VMA(q)과정을 VAR(∞)모형으로 나타낼 때의 VAR계수행렬들 $\{\Pi_i\}$ 를 구하는 다른 방법들 중의 하나는 SAS/IML의 함수 RATIO를 사용하는 것이다. 제6.3.1소절에서 설명하였듯이 함수 RATIO를 사용하면 행렬역함수들(matrix polynomials)의 나누기를 수행할 수 있다. 앞의 예제의 3변량VMA(3)과정의 VAR(∞)표현을 SAS/IML함수 RATIO를 사용해서 구하기 위해 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4 TSA4의 파일 EXAM6-18.PRG에 수록되어 있다).


```

/ *                               */
/ *           EXAM6-18.PRC        */
/ *                               */
/ *           CALCULATE AR COEFFICIENT MATRICES */
/ *           OF 3-VARIATE VMA(3) MODEL        */
/ *           USING SAS/IML FUNCTION RATIO     */
OPTIONS PAGESIZE = 60;
PROC IML;
RESET NOLOG NOCENTER;
THETA1 = { 0.5 0.1 0.0,
          0.4 0.3 0.0,
          0.0 0.0 0.3 };
THETA2 = { 0.0 0.0 0.0,
          0.25 0.0 0.0,
          0.0 0.0 0.28 };
THETA3 = { 0.0 0.12 -0.34,
          0.0 0.0 0.20,
          0.22 0.31 -0.01 };
/ *           USING SAS/IML FUNCTION RATIO     */
THETA0 = THETA1**0;
PHI0 = THETA0;
PHI = PHI0;
THETA = THETA0 || -THETA1 || -THETA2 || -THETA3;
PI = -RATIO(THETA, PHI, 8, 3);
/ *           GET PI MATRICES                 */
PI0 = PI[ , {1 2 3}];
PI1 = PI[ , {4 5 6}];
PI2 = PI[ , {7 8 9}];
PI3 = PI[ , {10 11 12}];
PI4 = PI[ , {13 14 15}];
PI5 = PI[ , {16 17 18}];
PI6 = PI[ , {19 20 21}];
PI7 = PI[ , {22 23 24}];
PRINT , "USING SAS/ IML FUNCTION RATIO",;
PRINT PI0 " " PI1, PI2 " " PI3,
      PI4 " " PI5, PI6 " " PI7, ;
QUIT;

```

이 SAS프로그램을 수행한 결과가 그림 6.18이다. 그림 6.18과 그림 6.17이 동일한 결과를

SAS			23:18 Monday, March 7, 1994 4		
USING SAS/ IML FUNCTION RATIO					
PI0			PI1		
-1	0	0	-0.5	-0.1	0
0	-1	0	-0.4	-0.3	0
0	0	-1	0	0	-0.3
PI2			PI3		
-0.29	-0.08	0	-0.202	-0.173	0.34
-0.57	-0.13	0	-0.412	-0.096	-0.2
0	0	-0.37	-0.22	-0.31	-0.185
PI4			PI5		
-0.1902	-0.1321	0.252	-0.19119	-0.09345	0.2534
-0.2769	-0.118	0.016	-0.20965	-0.13149	0.1166
-0.3	-0.208	-0.1561	-0.3921	-0.2071	-0.09493
PI6			PI7		
-0.0912	0.034006	0.17726	0.0000799	0.0563778	0.159858
-0.230921	-0.171852	0.16234	-0.213554	-0.102916	0.151736
-0.37159	-0.18509	-0.057537	-0.345948	-0.177077	0.0181195
Exiting IML.					

그림 6.18 함수 RATIO를 사용한 3변량VMA(3)과정의 VAR(∞)표현

보임을 알 수 있다.

한 가지 유의할 점은 VMA과정이 가역적이라고 해서 이에 해당하는 VAR모형의 차수가 반드시 무한대일 필요는 없다는 것이다. 한 예로 다음과 같은 VMA계수행렬을 갖는 5변량VMA(1)모형을 살펴보자.

$$\Theta_1 = \begin{bmatrix} 0.0 & 0.5 & -0.4 & 0.6 & 0.3 \\ 0.0 & 0.0 & 0.3 & -0.8 & 0.2 \\ 0.0 & 0.0 & 0.0 & 0.6 & -0.1 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.8 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix}$$

다음 식이 성립함을 쉽게 알 수 있다.

$$\Theta_1^k = 0 \quad (k = 5, 6, \dots)$$

따라서 이 5변량VMA(1)과정은 5변량VAR(4)모형을 만족한다. 이 결과는 다음 SAS프로그램을 수행해서 확인할 수 있다(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM 6-19.PRG에 수록되어 있다).

```

/*                                     */
/*                                     EXAM6-19.PRG                             */
/*                                     */
/*                                     CALCULATE AR COEFFICIENT MATRICES        */
/*                                     OF 5-VARIATE VMA(1) MODEL                 */
/*                                     */

```

SAS 8:30 Monday, October 26, 1992 8

PI0				
1	0	0	0	0
0	1	0	0	0
0	0	1	0	0
0	0	0	1	0
0	0	0	0	1
PI1				
0	0.5	-0.4	0.6	0.3
0	0	0.3	-0.8	0.2
0	0	0	0.6	-0.1
0	0	0	0	0.8
0	0	0	0	0
PI2				
0	0	0.15	-0.64	0.62
0	0	0	0.18	-0.67
0	0	0	0	0.48
0	0	0	0	0
0	0	0	0	0
PI3				
0	0	0	0.09	-0.527
0	0	0	0	0.144
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
PI4				
0	0	0	0	0.072
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
PI5				
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
PI6				
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
PI7				
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

Exiting IML.

그림 6.19 VMA(1)과정의 VAR(4)표현

```

OPTIONS PAGESIZE=60;
PROC IML;
RESET NOLOG NOCENTER;
THETA1 = { 0.0 0.5 -0.4 0.6 0.3,

```

```

0.0 0.0 0.3 -0.8 0.2,
0.0 0.0 0.0 0.6 -0.1,
0.0 0.0 0.0 0.0 0.8,
0.0 0.0 0.0 0.0 0.0 };
PI0 = THETA1**0;
PI1 = PI0*THETA1;
PI2 = PI1*THETA1;
PI3 = PI2*THETA1;
PI4 = PI3*THETA1;
PI5 = PI4*THETA1;
PI6 = PI5*THETA1;
PI7 = PI6*THETA1;
PRINT PI0 " " PI1, PI2 " " PI3,
      PI4 " " PI5, PI6 " " PI7, ;
QUIT;

```

이 SAS프로그램을 수행한 결과가 그림 6.19이다.

임의의 d 변량VMA(q)과정은 정상적이며 자기공분산행렬함수는 다음과 같다.

$$\Gamma(h) = \begin{cases} \sum_{i=0}^{q-h} \Theta_{h+i} \Sigma \Theta_i' & (h=0, 1, \dots, q) \\ 0 & (h=q+1, q+2, \dots) \end{cases}$$

다음 3변량VMA(1)과정의 자기공분산행렬함수를 구해 보자.

$$y_i = v_i - \Theta_1 v_{i-1}$$

$$\text{단, } \Theta_1 = \begin{bmatrix} 0.32 & 0.03 & -0.10 \\ 0.01 & -0.46 & 0.00 \\ -0.02 & -0.12 & 0.23 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1.00 & -0.20 & 0.00 \\ -0.20 & 1.04 & 0.30 \\ 0.00 & 0.30 & 1.09 \end{bmatrix}$$

이 3변량VMA(1)과정의 자기공분산행렬함수를 구하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-20.PRG에 수록되어 있다).

```

/* */
/*          EXAM6-20.PRG          */
/* */
/*          CALCULATE AUTOCOVARANCE MATRICES          */
/*          OF 3-VARIATE VMA(1) MODEL          */
/* */
OPTIONS PAGESIZE=60;
PROC IML;

```

SAS			22:04 Tuesday, November 10, 1992 10		
GAMMA0			GAMMA1		
1.108596	-0.167972	-0.021744	-0.314	0.0628	0.1
-0.167972	1.262004	0.323868	-0.102	0.4804	0.138
-0.021744	0.323868	1.145517	-0.004	0.0518	-0.2147
GAMMA2			GAMMA3		
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
GAMMA4			GAMMA5		
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
Exiting IML.					

그림 6.20 3변량VMA(2)과정의 자기공분산행렬함수

```

RESET NOLOG NOCENTER;

/* INPUT DATA */

THETA0 = - I(3);
THETA1 = { 0.32 0.03 -0.10,
           0.01 -0.46 0.00,
           -0.02 -0.12 0.23 };
THETA2 = J(3, 3, 0);
SIGMA = { 1.0 -0.2 0.0,
          -0.2 1.04 0.3,
          0.0 0.3 1.09 };
GAMMA0 = THETA0* SIGMA*THETA0'
          + THETA1*SIGMA*THETA1';
GAMMA1 = THETA1*SIGMA*THETA0';
GAMMA2 = THETA2;
GAMMA3 = THETA2;
GAMMA4 = THETA2;
GAMMA5 = THETA2;

/* PRINT AUTOCOVARANCE FUNCTION OF Y */

PRINT GAMMA0 " " GAMMA1,
      GAMMA2 " " GAMMA3,
      GAMMA4 " " GAMMA5;

QUIT;

```

이 SAS프로그램을 수행한 결과가 그림 6.20이다.

6.4.2 가역성

[SAS4TSA2]의 제 9.2.5 소절에서 언급했듯이 어떤 단변량시계열데이터의 자기공분산구조 (autocovariance structure)가 주어졌을 때, 이 자기공분산구조를 만족하는 MA(q)모형은 최대로 2 ^{q} 개 존재한다. 한 예로 다음 4개의 MA(2)모형들은 같은 자기공분산구조를 갖는다.

$$y_t = (1 - 0.2B)(1 - 0.5B)u_t$$

$$y_t = (1 - 0.2B)(1 - 2B)v_t$$

$$y_t = (1 - 5B)(1 - 0.5B)w_t$$

$$y_t = (1 - 5B)(1 - 2B)z_t$$

여기서 $\{u_t\}$, $\{v_t\}$, $\{w_t\}$, $\{z_t\}$ 는 분산이 각각 σ_u^2 , σ_v^2 , σ_w^2 , σ_z^2 인 백색잡음과정들이다. 이 4개의 MA(2)모형들이 같은 자기공분산구조를 갖는 이유는 다음과 같다. 우선 공분산발생함수 (covariance generating function) $g(z)$ 를 다음과 같이 정의하자.

$$g(z) = \sum_{k=-\infty}^{\infty} \sigma(k) z^k$$

다음과 같은 MA(2)모형을 살펴보자.

$$y_t = \theta(B)v_t$$

여기서 $\{v_t\}$ 는 분산이 σ_v^2 인 백색잡음과정이고, $\theta(x)$ 는 다음과 같이 정의된다.

$$\theta(x) = 1 - \theta_1 x - \theta_2 x^2$$

이 MA(2)과정의 공분산발생함수는 다음과 같다.

$$g(z) = \sigma_v^2 \theta(z) \theta(z^{-1})$$

만일 $\theta(z)=0$ 의 두 근들을 각각 $1/\alpha$, $1/\beta$ 이라고 하면 다음 식이 성립한다.

$$\begin{aligned} g(z) &= \sigma_v^2 (1 - \alpha z)(1 - \beta z) \left(1 - \frac{\alpha}{z}\right) \left(1 - \frac{\beta}{z}\right) \\ &= \sigma_v^2 \left\{1 - \alpha \left(z + \frac{1}{z}\right) + \alpha^2\right\} \left\{1 - \beta \left(z + \frac{1}{z}\right) + \beta^2\right\} \end{aligned}$$

따라서 앞에서 정의한 4개의 MA(2)모형들의 공분산발생함수들은 각각 다음과 같다.

$$g_1(z) = \sigma_u^2 \left\{1 - 5 \left(z + \frac{1}{z}\right) + 25\right\} \left\{1 - 2 \left(z + \frac{1}{z}\right) + 4\right\}$$

$$g_2(z) = \sigma_v^2 \left\{1 - 5 \left(z + \frac{1}{z}\right) + 25\right\} \left\{1 - \frac{1}{2} \left(z + \frac{1}{z}\right) + \frac{1}{4}\right\}$$

$$g_3(z) = \sigma_w^2 \left\{1 - \frac{1}{5} \left(z + \frac{1}{z}\right) + \frac{1}{25}\right\} \left\{1 - 2 \left(z + \frac{1}{z}\right) + 4\right\}$$

$$g_1(z) = \sigma_v^2 \left\{ 1 - \frac{1}{5} \left(z + \frac{1}{z} \right) + \frac{1}{25} \right\} \left\{ 1 - \frac{1}{2} \left(z + \frac{1}{z} \right) + \frac{1}{4} \right\}$$

즉, 다음 식들이 성립한다.

$$g_2(z) = \left(\frac{1}{4} \frac{\sigma_v^2}{\sigma_u^2} \right) g_1(z)$$

$$g_3(z) = \left(\frac{1}{25} \frac{\sigma_w^2}{\sigma_u^2} \right) g_1(z)$$

$$g_4(z) = \left(\frac{1}{100} \frac{\sigma_z^2}{\sigma_u^2} \right) g_1(z)$$

따라서 이 4개의 MA(2)모형들은 같은 자기공분산구조를 갖는다.

이러한 결과는 다음과 같은 MA(q)모형에서도 성립한다.

$$y_t = \theta(B) v_t$$

여기서 $\{v_t\}$ 는 분산이 σ_v^2 인 백색잡음과정이고 $\theta(x)$ 는 다음과 같이 정의된다.

$$\theta(x) = 1 - \theta_1 x - \theta_2 x^2 - \dots - \theta_q x^q$$

MA특성방정식 $\theta(x) = 0$ 의 특성근들의 역수들을 z_1, z_2, \dots, z_q 라고 하면 이 MA(q)과정의 공분산발생함수는 다음과 같다.

$$\begin{aligned} g(z) &= \sigma_v^2 \prod_{i=1}^q (1 - z_i z) \left(1 - \frac{z_i}{z} \right) \\ &= \sigma_v^2 \prod_{i=1}^q \left\{ 1 - z_i \left(z + \frac{1}{z} \right) + z_i^2 \right\} \end{aligned}$$

따라서 어떤 특성근 z_j 대신 z_j 의 역수를 대입하면 공분산발생함수는 백색잡음과정의 분산인 상수 σ_v^2 만 변한다. 즉, 하나의 공분산발생함수로부터 같은 자기상관함수를 갖는 최대 2^q 개의 MA(q)모형들을 만들 수 있다. 특히 MA(q)과정의 특성방정식 $\theta(x) = 0$ 의 어떤 근도 단위원 위에 존재하지 않는다면, 이 MA(q)모형들 중에서 모든 특성근들이 단위원 밖에 존재하는 MA(q)모형은 단 하나 존재한다. 즉, 가역조건을 만족하는 MA(q)모형은 단 하나 존재한다. 그러나 어떤 특성근이 단위원 위에 존재하면 문제는 복잡해진다. 이러한 이유로 가역조건이 만족되는 시계열모형을 다루는 것이 일반적이다. 이 가역조건에 관한 자세한 내용은 [Choi92] (p. 9)를 참조하라.

다변량시계열분석도 같은 결과를 증명할 수 있다. 즉, VMA특성방정식 $\det[\Theta(x)] = 0$ 의 어느 근도 단위원 위에 존재하지 않으면 주어진 공분산구조를 갖는 가역적 VMA(q)모형은 유일하게 존재한다. 그러나 VMA(q)과정의 특성근들이 단위원 위에 존재하면 이 VMA 과정이 반드시 VAR(∞)표현을 갖는 것은 아니다. 따라서 모든 VMA과정을 가역조건을 만족하는 VMA모형으로 표현할 수 있는 것은 아니다. 그러나 특성근들이 단위원 위에 존재하는 경우는 드물고 이러한 VMA모형을 반드시 사용할 특별한 이유도 없으므로 이후 본서에서는 가역조건을 만족하는 VMA모형만을 다루기로 한다.

6.4.3 예 제 들

[예제 1]

다음과 같은 3변량VMA(1)모형을 살펴보자.

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \begin{bmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \end{bmatrix} - \begin{bmatrix} 0.5 & 0.5 & -0.4 \\ 0.4 & -0.7 & 0.3 \\ 0.0 & 0.2 & 0.5 \end{bmatrix} \begin{bmatrix} v_{1t-1} \\ v_{2t-1} \\ v_{3t-1} \end{bmatrix}$$

여기서 $\{v_t = (v_{1t}, v_{2t}, v_{3t})'\}$ 는 3변량백색잡음과정으로서 평균벡터는 영벡터이고 공분산행렬은 다음과 같다.

$$\begin{bmatrix} 1 & -0.2 & 0.0 \\ -0.2 & 1.04 & 0.3 \\ 0.0 & 0.3 & 1.09 \end{bmatrix}$$

이 3변량VMA(1)모형으로부터 150개의 관찰점들을 생성하기 위해 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-21.PRG에 수록되어 있다).

```

/* */
/*          EXAM6-21.PRG          */
/* */
/*          GENERATE 3-VARIATE VMA(1) PROCESS          */
/*          DATASET = VMA31A          */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=60;
DATA EXAM621;
  DO T = 1 TO 350;
    U1 = RANNOR(7654321);
    U2 = RANNOR(1234567);
    U3 = RANNOR(1357531);
    V1 = U1;
    V2 = U2 - 0.2 * U1;
  
```

SIMULATED 3-VARIATE VMA(1) PROCESS

1

23:33 Tuesday, November 10, 1992

(A)

CORRELATION ANALYSIS

3 'VAR' Variables: Y1

Y2

Y3

Covariance Matrix		DF = 149		
	Y1	Y2	Y3	
Y1	1.283447991	-0.239450409	-0.154675379	
Y2	-0.239450409	1.505311014	0.448395033	
Y3	-0.154675379	0.448395033	1.000795144	

Simple Statistics				
Variable	N	Mean	Std Dev	Sum
Y1	150	0.03118	1.13289	4.67757
Y2	150	0.13857	1.22691	20.78509
Y3	150	-0.04292	1.00040	-6.43801

Simple Statistics		
Variable	Minimum	Maximum
Y1	-2.68219	3.31377
Y2	-2.94782	3.88698
Y3	-2.38274	2.65294

Pearson Correlation Coefficients / Prob > R under Ho: Rho=0 / N = 150			
	Y1	Y2	Y3
Y1	1.00000 0.0	-0.17227 0.0350	-0.13648 0.0959
Y2	-0.17227 0.0350	1.00000 0.0	0.36532 0.0001
Y3	-0.13648 0.0959	0.36532 0.0001	1.00000 0.0

(B)

SIMULATED 3-VARIATE VMA(1) PROCESS

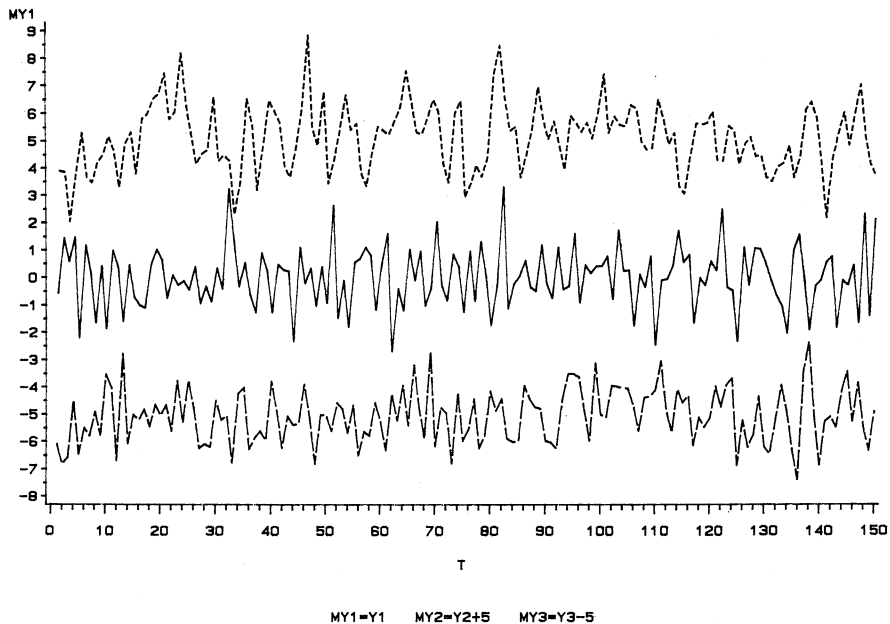


그림 6.21 3변량VMA(1)모형에서 생성된 다변량시계열데이터 VMA31A

```

V3 = U3+0.3*U2;
V1LAG1 = LAG1(V1);
V2LAG1 = LAG1(V2);
V3LAG1 = LAG1(V3);
IF V1LAG1 = . THEN V1LAG1 = 0.0;
IF V2LAG1 = . THEN V2LAG1 = 0.0;
IF V3LAG1 = . THEN V3LAG1 = 0.0;
Y1 = V1-0.32*V1LAG1-0.03*V2LAG1+0.10*V3LAG1;
Y2 = V2-0.01*V1LAG1+0.46*V2LAG1-0.00*V3LAG1;
Y3 = V3+0.02*V1LAG1+0.12*V2LAG1-0.23*V3LAG1;
OUTPUT;
END;
KEEP T Y1 Y2 Y3;
RUN;
DATA TSA4DATA.VMA31A;
SET EXAM621;
T = T-200;
IF T < 1 THEN DELETE;
RUN;
TITLE1 'SIMULATED 3-VARIATE VMA(1) PROCESS';
PROC CORR COV;
VAR Y1 Y2 Y3;
RUN;
DATA EXAM621B;
SET TSA4DATA.VMA31A;
MY1 = Y1;
MY2 = Y2 + 5.0;
MY3 = Y3 - 5.0;
KEEP T MY1 MY2 MY3;
RUN;
PROC GPLOT;
PLOT (MY1 MY2 MY3)*T / OVERLAY;
SYMBOL1 C=BLACK V=NONE I=JOIN LINE=1;
SYMBOL2 C=BLACK V=NONE I=JOIN LINE=2;
SYMBOL3 C=BLACK V=NONE I=JOIN LINE=4;
FOOTNOTE 'MY1=Y1 MY2=Y2+5 MY3=Y3-5';
RUN;

```

이 SAS프로그램을 실행한 결과가 그림 6.21이다. 그림 6.21의 부분 A는 PROC CORR

를 실행한 결과이다. 부분 ㉔는 PROC GPLOT를 실행한 결과인데 단변량시계열들 Y1, Y2, Y3가 겹쳐지지 않도록 MY1 = Y1, MY2 = Y2 + 5, MY3 = Y3 - 5의 시계열산포도를 겹쳐 그렸다. 이 산포도에서 3변량시계열데이터 (Y1, Y2, Y3)'가 정상적임을 알 수 있다.

[예제 2]

다음과 같은 2변량VMA(2)모형을 살펴보자.

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} - \begin{bmatrix} 0.4 & 0.3 \\ -0.4 & 0.5 \end{bmatrix} \begin{bmatrix} v_{1,t-1} \\ v_{2,t-1} \end{bmatrix} - \begin{bmatrix} 0.2 & -0.1 \\ -0.3 & -0.5 \end{bmatrix} \begin{bmatrix} v_{1,t-2} \\ v_{2,t-2} \end{bmatrix}$$

여기서 $\{v_t = (v_{1,t}, v_{2,t})\}$ 는 2변량백색잡음과정으로서 평균벡터는 영벡터이고 공분산행렬은 다음과 같다.

$$\begin{bmatrix} 1 & -0.2 \\ -0.2 & 1.04 \end{bmatrix}$$

이 2변량VMA(2)모형에서 120개의 관찰점들을 생성하기 위해 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-22.PRG에 수록되어 있다).

```

/*                                                    */
/*                EXAM6-22.PRG                        */
/*                                                    */
/*                GENERATE 2-VARIATE VMA(2) PROCESS  */
/*                DATASET = VMA22A                   */
/*                                                    */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=60;
DATA EXAM622;
  DO T = 1 TO 320;
    U1 = RANNOR(7654321);
    U2 = RANNOR(1234567);
    V1 = U1;
    V2 = U2 - 0.2 * U1;
    V1LAG1 = LAG1(V1);
    V2LAG1 = LAG1(V2);
    V1LAG2 = LAG2(V1);
    V2LAG2 = LAG2(V2);
    IF V1LAG1 = . THEN V1LAG1 = 0.0;
    IF V2LAG1 = . THEN V2LAG1 = 0.0;
    IF V1LAG2 = . THEN V1LAG2 = 0.0;
  
```

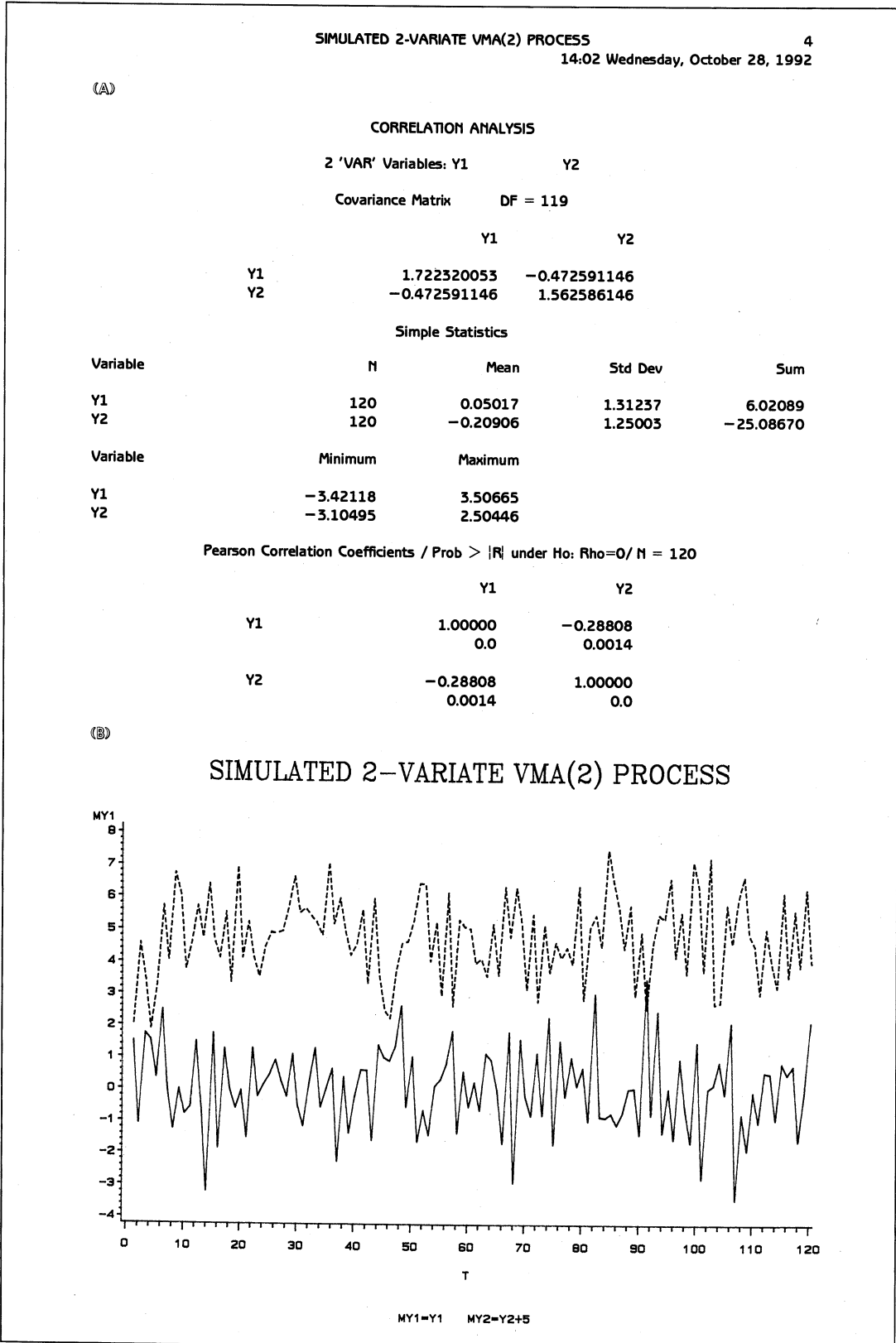


그림 6.22 2변량VMA(2)모형에서 생성된 다변량시계열데이터 VMA22A

```

IF V2LAG2 = . THEN V2LAG2 = 0.0;
Y1 = V1 - 0.4*V1LAG1 - 0.3*V2LAG1
      + 0.4*V1LAG2 - 0.5*V2LAG2;
Y2 = V2 - 0.2*V1LAG1 + 0.1*V2LAG1
      + 0.3*V1LAG2 + 0.5*V2LAG2;
OUTPUT;
END;
KEEP T Y1 Y2;
RUN;
DATA TSA4DATA.VMA22A;
SET EXAM622;
T = T-200;
IF T < 1 THEN DELETE;
RUN;
TITLE1 'SIMULATED 2-VARIATE VMA(2) PROCESS';
PROC CORR COV;
VAR Y1 Y2;
RUN;
DATA EXAM622B;
SET TSA4DATA.VMA22A;
MY1 = Y1;
MY2 = Y2 + 5.0;
KEEP T MY1 MY2 MY3;
RUN;
PROC GPLOT;
PLOT (MY1 MY2)*T / OVERLAY;
SYMBOL1 C=BLACK V=NONE I=JOIN LINE=1;
SYMBOL2 C=BLACK V=NONE I=JOIN LINE=2;
FOOTNOTE 'MY1=Y1 MY2=Y2+5';
RUN;

```

이 SAS프로그램을 실행한 결과가 그림 6.22이다. 그림 6.22의 부분 A는 PROC CORR를 실행한 결과이고 부분 B는 PROC GPLOT를 실행한 결과이다. 이 부분의 산포도에서 2변량시계열데이터 (Y1, Y2)가 정상적임을 알 수 있다.

[예제 3]

다음과 같은 3변량VMA(2)모형을 살펴보자.

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \begin{bmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \end{bmatrix} - \begin{bmatrix} 0.4 & 0.3 & -0.2 \\ 0.5 & -0.2 & -0.1 \\ 0.3 & 0.6 & -0.4 \end{bmatrix} \begin{bmatrix} v_{1,t-1} \\ v_{2,t-1} \\ v_{3,t-1} \end{bmatrix} - \begin{bmatrix} -0.4 & 0.5 & -0.7 \\ 0.4 & 0.3 & -0.6 \\ -0.5 & 0.3 & -0.4 \end{bmatrix} \begin{bmatrix} v_{1,t-2} \\ v_{2,t-2} \\ v_{3,t-2} \end{bmatrix}$$

여기서 $\{v_t\}$ 는 [예제 1]의 3변량백색잡음과정과 같다. 이 3변량VMA(2)모형에서 120개의 관찰점들을 생성하기 위해 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-23.PRG에 수록되어 있다).

```

/ * * /
/ *          EXAM6-23.PRG          * /
/ * * /
/ *          GENERATE 3-VARIATE VMA(2) PROCESS          * /
/ *          DATASET = VMA32A          * /
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=60;
DATA EXAM623;
  DO T = 1 TO 320;
    U1 = RANNOR(7654321);
    U2 = RANNOR(1234567);
    U3 = RANNOR(1357531);
    V1 = U1;
    V2 = U2 - 0.2 * U1;
    V3 = U3 + 0.3 * U2;
    V1LAG1 = LAG1(V1);
    V2LAG1 = LAG1(V2);
    V3LAG1 = LAG1(V3);
    V1LAG2 = LAG2(V1);
    V2LAG2 = LAG2(V2);
    V3LAG2 = LAG2(V3);
    IF V1LAG1 = . THEN V1LAG1 = 0.0;
    IF V2LAG1 = . THEN V2LAG1 = 0.0;
    IF V3LAG1 = . THEN V3LAG1 = 0.0;
    IF V1LAG2 = . THEN V1LAG2 = 0.0;
    IF V2LAG2 = . THEN V2LAG2 = 0.0;
    IF V3LAG2 = . THEN V3LAG2 = 0.0;
    Y1 = V1 - 0.4 * V1LAG1 - 0.3 * V2LAG1 + 0.2 * V3LAG1
          + 0.4 * V1LAG2 - 0.5 * V2LAG2 + 0.7 * V3LAG2;
    Y2 = V2 - 0.5 * V1LAG1 + 0.2 * V2LAG1 + 0.1 * V3LAG1
          - 0.4 * V1LAG2 - 0.3 * V2LAG2 + 0.6 * V3LAG2;
  END;

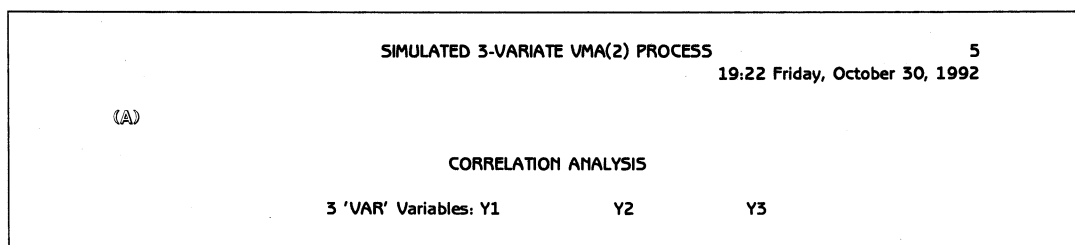
```

```

Y3 = V3 - 0.3*V1LAG1 - 0.6*V2LAG1 + 0.4*V3LAG1
      + 0.5*V1LAG2 - 0.3*V2LAG2 + 0.4*V3LAG2;
OUTPUT;
END;
KEEP T Y1 Y2 Y3;
RUN;
DATA TSA4DATA.VMA32A;
SET EXAM623;
T = T-200;
IF T < 1 THEN DELETE;
RUN;
TITLE1 'SIMULATED 3-VARIATE VMA(2) PROCESS';
PROC CORR COV;
VAR Y1 Y2 Y3;
RUN;
DATA EXAM623B;
SET TSA4DATA.VMA32A;
MY1 = Y1;
MY2 = Y2 + 5.0;
MY3 = Y3 - 5.0;
KEEP T MY1 MY2 MY3;
RUN;
PROC GPLOT;
PLOT (MY1 MY2 MY3)*T / OVERLAY;
SYMBOL1 C=BLACK V=NONE I=JOIN LINE=1;
SYMBOL2 C=BLACK V=NONE I=JOIN LINE=2;
SYMBOL3 C=BLACK V=NONE I=JOIN LINE=4;
FOOTNOTE 'MY1=Y1 MY2=Y2+5 MY3=Y3-5';
RUN;

```

이 SAS프로그램을 실행한 결과가 그림 6.23이다. 그림 6.23의 부분 A는 PROC CORR를 실행한 결과이다. 부분 B는 PROC GPLOT를 실행한 결과이다. 이 산포도에서 3변량시계열데이터 (Y1, Y2, Y3)가 정상적임을 알 수 있다.



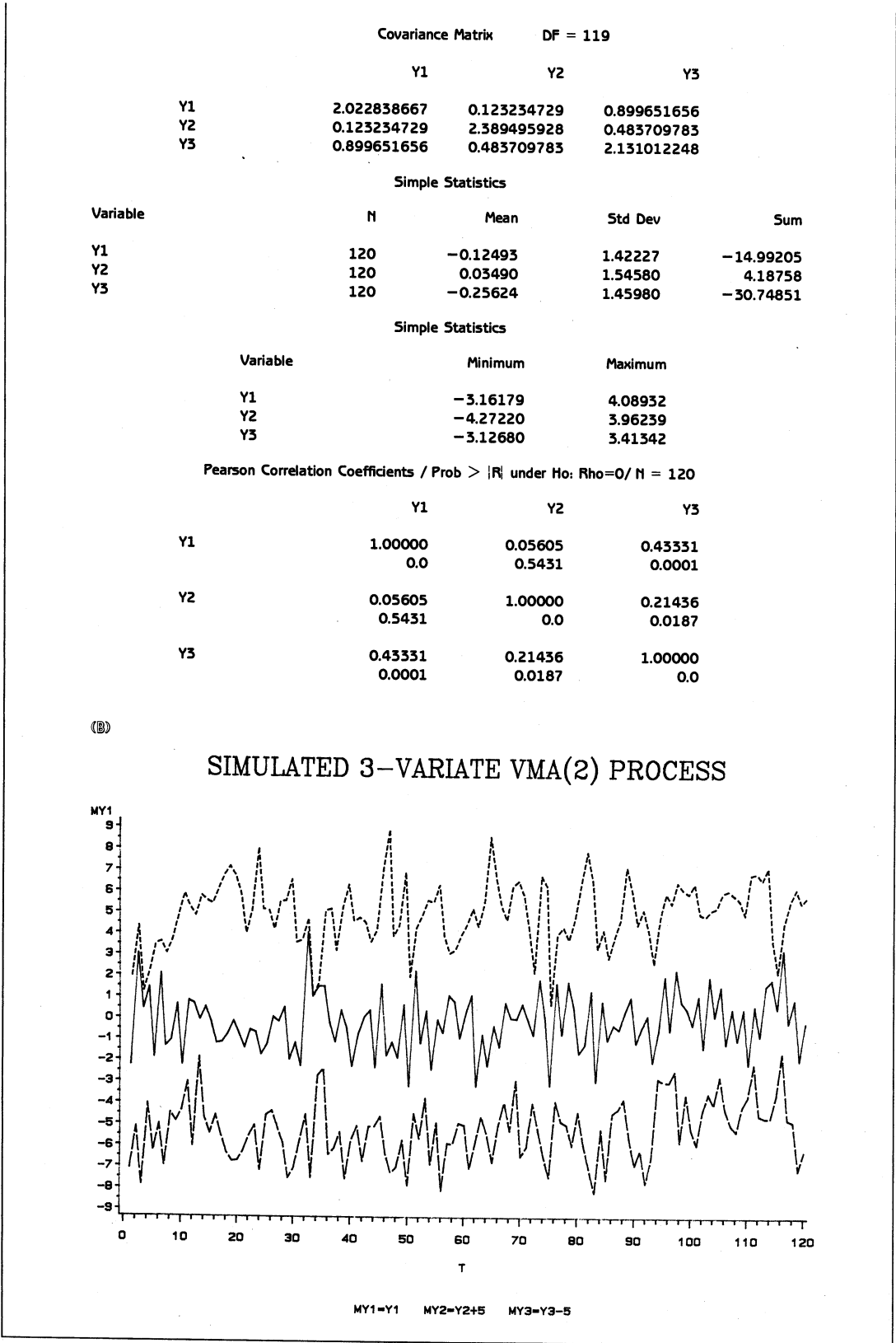


그림 6.23 3변량VMA(2)모형에서 생성된 다변량시계열데이터 VMA32A

6.4.4 VMA계수행렬

VMA(q)과정의 자기공분산행렬함수 $\{\Gamma(h)\}$ 가 주어지면 VMA계수행렬들을 구할 수 있다. 다음과 같은 VMA(1)과정을 살펴보자.

$$y_t = v_t - \Theta_1 v_{t-1}$$

이 VMA(1)과정의 자기공분산행렬들은 다음과 같다.

$$\Gamma(0) = \Sigma + \Theta_1 \Sigma \Theta_1'$$

$$\Gamma(1) = -\Theta_1 \Sigma$$

이 식들로부터 다음 식을 얻을 수 있다.

$$\Theta_1^2 \Gamma(-1) + \Theta_1 \Gamma(0) + \Gamma(1) = 0$$

이 식은 다음 식과 동일하다.

$$\Theta_1 = -\{\Theta_1 \Gamma(0) + 2\Gamma(1)\} \{2\Theta_1 \Gamma(-1) + \Gamma(0)\}^{-1}$$

본 저자는 이 식을 바탕으로 VMA계수행렬들을 구하기 위해서 알고리즘 6.5와 알고리즘 6.6을 사용할 것을 제안한다.

알고리즘 6.5 : VMA계수행렬을 구하는 알고리즘

$$\Theta_1^{(n+1)} = \frac{1}{(a+1)} [a\Theta_1^{(n)} - \{\Theta_1^{(n)} \Gamma(0) + 2\Gamma(1)\} \{2\Theta_1^{(n)} \Gamma(-1) + \Gamma(0)\}^{-1}]$$

여기서 위첨자 (n)은 n 번째 반복을 뜻하며, $a(>0)$ 는 정해진 상수이다.

한 예로 그림 6.20의 자기공분산함수로부터 VMA계수행렬을 구하기 위해 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-24.PRG에 수록되어 있다).

```

/* */
/* EXAM6-24.PRG */
/* */
/* CALCULATE VMA COEFFICIENT MATRICES */
/* OF 3-VARIATE VMA(1) MODEL */
/*
OPTIONS PAGESIZE=60;
PROC IML;
```

```

RESET NOLOG NOCENTER;
/*                                                    */
/*          FIND VMA PARAMETER OF VMA(1) MODEL      */
/*                                                    */
START VMA1COEF(M, GAMMA0, GAMMA1, THETA);
THETA = 0.5 * GAMMA0;
OLDTHETA = 1.0E+9 * I(M);
DO K = 1 TO 30
  UNTIL ( MAX(ABS(OLDTHETA - THETA)) < 1.0E-10 );
  EPSILON = MAX(ABS(OLDTHETA - THETA));
  PRINT "NO. OF ITERATION = " K " " EPSILON;
  OLDTHETA = THETA;
  THETA = (1/2.4) * (1.4 * THETA -
                    (THETA * GAMMA0 + 2.0 * GAMMA1)
                    * INV(2.0 * THETA * GAMMA1' + GAMMA0));
END;
FINISH;
/*          INPUT DATA          */
GAMMA0 = { 1.108596 -0.167972 -0.021744,
           -0.167972 1.262004 0.323868,
           -0.021744 0.323868 1.145517 };
GAMMA1 = { -0.314 0.0628 0.1,
           -0.102 0.4804 0.138,
           -0.004 0.0518 -0.2147 };
RUN VMA1COEF(3, GAMMA0, GAMMA1, THETA1);
PRINT , , THETA1;
QUIT;

```

이 SAS 프로그램을 수행한 결과가 그림 6.24이다. SAS 프로그램 EXAM6-20.PRG에서 볼 수 있듯이, 원래의 Θ_1 은 다음과 같다.

$$\Theta_1 = \begin{bmatrix} 0.32 & 0.03 & -0.10 \\ 0.01 & -0.46 & 0.00 \\ -0.02 & -0.12 & 0.23 \end{bmatrix}$$

따라서 알고리즘 6.5는 올바른 결과를 보여줄 수 있다.

알고리즘 6.5에서 상수 a 는 양수이면 된다. 본 저자가 여러 개의 a 값을 대입해 본 결과 이 예제에서는 $a = 1.4$ 가 가장 적합했다. 그러나 알고리즘 6.5의 수렴성과 초기조건에 대해서는 좀더 연구해야 할 것이다. VMA(1)과정의 자기공분산행렬함수로부터 VMA계수행렬들을 구하기 위해서는 알고리즘 6.5보다는 일반적인 다음 알고리즘을 사용할 수 있다.

```

SAS                                     11:45 Wednesday, November 11, 1992   1

NO. OF ITERATION =                      K          EPSILON
1                                     999999999
NO. OF ITERATION =                      K          EPSILON
2                                     0.5426671
NO. OF ITERATION =                      K          EPSILON
3                                     0.413945
NO. OF ITERATION =                      K          EPSILON
4                                     0.0975279
NO. OF ITERATION =                      K          EPSILON
5                                     0.0319433
NO. OF ITERATION =                      K          EPSILON
6                                     0.0038116
NO. OF ITERATION =                      K          EPSILON
7                                     0.0009687
NO. OF ITERATION =                      K          EPSILON
8                                     0.0001078
NO. OF ITERATION =                      K          EPSILON
9                                     0.0000269
NO. OF ITERATION =                      K          EPSILON
10                                    3.0126E-6
NO. OF ITERATION =                      K          EPSILON
11                                    7.4354E-7
NO. OF ITERATION =                      K          EPSILON
12                                    8.3926E-8
NO. OF ITERATION =                      K          EPSILON
13                                    2.066E-8
NO. OF ITERATION =                      K          EPSILON
14                                    3.6602E-9
NO. OF ITERATION =                      K          EPSILON
15                                    1.0605E-9
NO. OF ITERATION =                      K          EPSILON
16                                    3.171E-10

      THETA1
      0.32      0.03      -0.1
      0.01      -0.46     2.414E-13
      -0.02     -0.12      0.25

Exiting IML.
    
```

그림 6.24 3변량VMA(1)과정의 VMA계수행렬

알고리즘 6.6 : 계수행렬들을 구하는 알고리즘

$$\Theta_1^{(n+1)} = (I + A^{(n)})^{-1} [A^{(n)} \Theta_1^{(n)} - \{\Theta_1^{(n)} \Gamma(0) + 2\Gamma(1)\} \{2\Theta_1^{(n)} \Gamma(-1) + \Gamma(0)\}^{-1}]$$

여기서 위첨자 (n)은 n번째 반복을 뜻하며, A⁽ⁿ⁾은 I+A⁽ⁿ⁾이 정칙행렬이 되도록 하는 주어진 행렬이다.

알고리즘 6.5와 알고리즘 6.6을 사용하는 데 있어서 유의할 점은 두 알고리즘들의 수렴성이 아직 증명되지 않았고, 초기값에 따라 결과가 다를 수도 있다는 점이다. 특히 알고리즘 6.6에서는 계수행렬들 {A⁽ⁿ⁾ | n = 1, 2, ...}을 어떻게 선택하느냐에 따라 이 알고리즘의 수렴성질이 변할 것이다. 이에 대한 자세한 내용을 누군가가 좀더 연구해야 할 것이다. 또한 알고리즘 6.5나 알고리즘 6.6 대신 행렬역함수방정식(matrix polynomial equation)을 푸는 방법을 적용해서 Θ_1 을 구할 수도 있다. 자세한 내용은 Gantmacher(1959, pp. 227~231)

를 참조하라.

다음과 같은 VMA(q)과정의 자기공분산행렬함수로부터 VMA계수행렬들을 구하는 문제를 살펴보자.

$$y_t = -\sum_{k=0}^q \Theta_k v_{t-k} \quad (\Theta_0 = -I_d)$$

이 VMA(q)과정의 자기공분산행렬함수와 VMA계수행렬들 사이에는 다음 식들이 성립한다.

$$\sum_{i=0}^{q-h} \Theta_{h+i} \Sigma \Theta_i' = \Gamma(h) \quad (h=0, 1, \dots, q)$$

다음과 같은 대칭행렬 A_0 를 정의하자.

$$A_0 := -\Sigma^{-\frac{1}{2}}$$

이 행렬 A_0 를 이용해서 다음과 같이 행렬 A_i 를 정의한다.

$$A_i := \Theta_i \Sigma^{-\frac{1}{2}} = -\Theta_i A_0 \quad (i=1, 2, \dots, q)$$

또한 다음과 같은 행렬들을 정의하자.

$$C = \begin{bmatrix} A_0 & A_1 & \dots & A_{q-1} & A_q \\ A_1 & A_2 & \dots & A_q & O \\ A_2 & A_3 & \dots & O & O \\ \vdots & \vdots & & \vdots & \vdots \\ A_q & O & \dots & O & O \end{bmatrix}, \quad E = \begin{bmatrix} \Gamma(0) \\ \Gamma(1) \\ \Gamma(2) \\ \vdots \\ \Gamma(q) \end{bmatrix}, \quad X = \begin{bmatrix} A_0' \\ A_1' \\ A_2' \\ \vdots \\ A_q' \end{bmatrix}$$

$$D = \begin{bmatrix} A_0 & A_1 & \dots & A_{q-1} & A_q \\ O & A_0 & \dots & A_{q-2} & A_{q-1} \\ O & O & \dots & A_{q-3} & A_{q-2} \\ \vdots & \vdots & & \vdots & \vdots \\ O & O & \dots & O & A_0 \end{bmatrix}, \quad F = \begin{bmatrix} \Gamma(0)' \\ \Gamma(1)' \\ \Gamma(2)' \\ \vdots \\ \Gamma(q)' \end{bmatrix}$$

행렬들 A_0, A_1, \dots, A_q 를 사용해서 자기공분산행렬들을 다음과 같이 나타낼 수 있다.

$$CX = E$$

$$DX = F$$

본 저자는 이 식들을 바탕으로 다음 알고리즘을 사용할 것을 제안한다.

한 예로 알고리즘 6.7을 사용해서 SAS프로그램 EXAM6-24.PRG에서 다룬 VMA(1)과정의 VMA계수행렬을 구하기 위해 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-25.PRG에 수록되어 있다).


```

SAS
20:09 Monday, March 14, 1994 1

NO. OF ITERATION =      K      EPSILON
1                      0.4081389
NO. OF ITERATION =      K      EPSILON
2                      0.0934804
NO. OF ITERATION =      K      EPSILON
3                      0.0091991
NO. OF ITERATION =      K      EPSILON
4                      0.0001466
NO. OF ITERATION =      K      EPSILON
5                      4.072E-8
NO. OF ITERATION =      K      EPSILON
6                      3.22E-15
NO. OF ITERATION =      K      EPSILON
7                      2.22E-16

      SIGMA
1.0008279 -0.141264 0.0164095
-0.141264 1.0362662 0.2731346
0.0164095 0.2731346 1.0761677

      THETA1
0.3266615 0.0832507 -0.071396
-0.049326 -0.477708 0.0280605
-0.073645 -0.159173 0.2410258

Exiting IML.
    
```

그림 6.25 알고리즘 6.7을 사용해서 구한 VMA계수행렬

```

A1 = AMAT[{4 5 6},,]';
EPSILON = MAX(ABS(OLDAMAT - AMAT));
PRINT "NO. OF ITERATION = " K " " EPSILON;
OLDAMAT = AMAT;
END;

/*      FIND VMA MATRICES      */
SIGMA = A0*A0';
THETA = -A1*INV(A0);
FINISH;

/*
/*      INPUT DATA      */
GAMMA0 = { 1.108596 -0.167972 -0.021744,
          -0.167972 1.262004 0.323868,
          -0.021744 0.323868 1.145517 };
GAMMA1 = { -0.314 0.0628 0.1,
          -0.102 0.4804 0.138,
          -0.004 0.0518 -0.2147 };
RUN VMA1COE2(3, GAMMA0, GAMMA1, THETA1, SIGMA);
PRINT , SIGMA " " THETA1,;
QUIT;
    
```

이 SAS프로그램을 수행한 결과가 그림 6.25이다. SAS프로그램 EXAM6-25.PRG에서 볼 수 있듯이, 알고리즘 6.7은 2차수렴(second-order convergence)하는 것처럼 보인다. 즉, 수렴속도가 매우 빠르다. 이러한 현상은 알고리즘 6.7이 Wilson(1969)의 알고리즘을 확장

했기 때문에, 즉, Newton-Raphson해를 바탕으로 했기 때문에 생기는 현상인 것 같다. 그러나 수렴된 SIGMA나 THETA1이 원래의 행렬들(true matrices)과는 정확하게 일치하지 않으며, 이 SIGMA와 THETA1을 SAS프로그램 EXAM6-20.PRG에 대입해서 얻은 자기공분산행렬들도 원래의 자기공분산행렬들과 일치하지 않는다. 따라서 알고리즘 6.7은 좀 더 연구되어야 할 것이다.

6.5 VARMA모형

6.5.1 VARMA모형의 평균벡터

지금까지 살펴본 VAR모형이나 VMA모형과는 달리, 다변량시계열데이터가 한편으로는 VAR형태를 갖고 다른 한편으로는 VMA형태를 갖는다면 이 다변량시계열데이터는 좀더 일반적인 시계열모형을 갖게 될 것이다. 이러한 다변량시계열모형을 VARMA모형(vector autoregressive moving-average model, 벡터자기회귀이동평균모형)이라고 한다. VAR차수가 p 이고 VMA차수가 q 인 d 변량VARMA(p, q)모형은 다음과 같다.

$$z_t = \nu + \Phi_1 z_{t-1} + \Phi_2 z_{t-2} + \dots + \Phi_p z_{t-p} + v_t - \Theta_1 v_{t-1} - \Theta_2 v_{t-2} - \dots - \Theta_q v_{t-q} \\ (t = 0, \pm 1, \pm 2, \dots)$$

여기서 Φ_i 와 Θ_j 는 $d \times d$ 행렬들이며, ν 는 d 차원열벡터이고, $\{v_t\}$ 는 평균벡터가 영벡터이고 공분산행렬이 Σ 인 d 변량백색잡음과정이다. 특별한 언급이 없는 한 Σ 는 양정치행렬이라고 가정한다.

행렬 $(I - \Phi_1 - \Phi_2 - \dots - \Phi_p)$ 가 정칙이라고 가정하면 다음 식을 만족하는 벡터 μ 가 존재한다.

$$\mu := (I - \Phi_1 - \Phi_2 - \dots - \Phi_p)^{-1} \nu$$

다음과 같은 새로운 d 변량시계열데이터 $\{y_t\}$ 를 정의한다.

$$y_t = z_t - \mu \quad (t = 0, \pm 1, \pm 2, \dots)$$

이 d 변량시계열데이터 $\{y_t\}$ 는 다음 d 변량VARMA(p, q)모형을 만족한다.

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \dots + \Phi_p y_{t-p} + v_t - \Theta_1 v_{t-1} - \Theta_2 v_{t-2} - \dots - \Theta_q v_{t-q} \\ (t = 0, \pm 1, \pm 2, \dots)$$

만일, 행렬 $(I - \Phi_1 - \Phi_2 - \dots - \Phi_p)$ 가 정칙이라면 제 6.3.1소절에서와 같은 방법으로 $E(y_t) = 0$ 을 증명할 수 있다. VAR모형에서와 같이 VARMA모형이 작인성을 가지며 행렬

$(I - \Phi_1 - \Phi_2 - \dots - \Phi_p)$ 가 정칙이다. 우리가 분석하게 될 VARMA과정은 작인성조건을 만족한다고 가정하기 때문에 평균벡터 μ 가 존재한다. 따라서 지금부터는 다변량시계열데이터 $\{y_t\}$ 가 만족하는 d 변량VARMA(p, q)모형을 분석하기로 한다. 이 d 변량VARMA(p, q)모형은 다음과 같이 나타낼 수 있다.

$$\begin{aligned} \Phi(B)y_t &= \Theta(B)v_t \\ \text{단, } \Phi(x) &= -\Phi_0 - \Phi_1 x - \Phi_2 x^2 - \dots - \Phi_p x^p \quad (\Phi_0 = -I_d) \\ \Theta(x) &= -\Theta_0 - \Theta_1 x - \Theta_2 x^2 - \dots - \Theta_q x^q \quad (\Theta_0 = -I_d) \end{aligned}$$

이 $\Phi(x)$ 를 $\{y_t\}$ 의 VAR특성행렬이라 하고, $\Theta(x)$ 를 VMA특성행렬이라고 한다. 또한 $\det[\Phi(x)]$ 를 VAR특성식이라 하고, $\det[\Theta(x)]$ 는 VMA특성식이라 한다.

6.5.2 VARMA모형의 작인성

우선 d 변량VARMA(p, q)과정 $\{y_t\}$ 를 다음과 같이 $d(p+q)$ 변량VAR(1)모형으로 나타내 보자.

$$w_t = \Phi w_{t-1} + u_t$$

$$\text{단, } w_t := \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \\ v_t \\ v_{t-1} \\ \vdots \\ v_{t-q+1} \end{bmatrix}, \quad u_t := \begin{bmatrix} v_t \\ 0 \\ \vdots \\ 0 \\ v_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\Phi := \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{bmatrix}$$

$$\Phi_{11} := \begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_{p-1} & \Phi_p \\ I & O & \dots & O & O \\ O & I & \dots & O & O \\ \vdots & \vdots & & \vdots & \vdots \\ O & O & \dots & I & O \end{bmatrix}$$

$$\Phi_{12} := \begin{bmatrix} -\Theta_1 & -\Theta_2 & \dots & -\Theta_{q-1} & \Theta_q \\ O & O & \dots & O & O \\ O & O & \dots & O & O \\ \vdots & \vdots & & \vdots & \vdots \\ O & O & \dots & O & O \end{bmatrix}$$

$$\Phi_{21} := O$$

$$\Phi_{22} := \begin{bmatrix} O & O & \cdots & O & O \\ I & O & \cdots & O & O \\ O & I & \cdots & O & O \\ \vdots & \vdots & & \vdots & \vdots \\ O & O & \cdots & I & O \end{bmatrix}$$

다음 식들을 쉽게 증명할 수 있다.

$$\begin{aligned} & \det(I - \Phi x) \\ &= \det \begin{bmatrix} I - \Phi_{11} x & -\Phi_{12} x \\ O & I - \Phi_{22} x \end{bmatrix} \\ &= \det(I - \Phi_{11} x) \det(I - \Phi_{22} x) \\ &= \det(I - \Phi_{11} x) \\ &= \det\left(I - \sum_{k=1}^p \Phi_k x^k\right) \end{aligned}$$

마지막 등호는 제6.3.1소절의 방법을 사용해서 증명할 수 있다. 따라서 $\{w_i\}$ 가 작인적이기 위해서는 $|x| \leq 1$ 에서 $\det[\Phi(x)] \neq 0$ 이어야 한다. 즉, d 변량VARMA(p, q)과정 $\{y_i\}$ 가 작인적이기 위해서는 이 다변량시계열데이터의 VAR특성방정식 $\det[\Phi(x)] = 0$ 의 모든 근들이 단위원 밖에 존재하거나, 아예 근이 하나도 존재하지 않아야 한다. 즉, VAR특성식이 상수이어야 한다.

작인성을 갖는 VARMA(p, q)과정을 VMA(∞)모형으로 나타내기 위해서 다음과 같은 $d \times d(p+q)$ 행렬들 J 와 H 를 정의하자.

$$J := [I \ O \ \cdots \ O \ O \ O \ \cdots \ O]$$

$$H := [I \ O \ \cdots \ O \ I \ O \ \cdots \ O]$$

제 6.3.1소절과 같은 방법으로 다음 식들을 유도할 수 있다.

$$\begin{aligned} & y_t \\ &= J w_t \\ &= J \left(\sum_{k=0}^{\infty} \Phi^k u_{t-k} \right) \\ &= J \left(\sum_{k=0}^{\infty} \Phi^k H' J u_{t-k} \right) \\ &= \sum_{k=0}^{\infty} (J \Phi^k H') v_{t-k} \\ &= \sum_{k=0}^{\infty} \Psi_k v_{t-k} \end{aligned}$$

여기서 $\Psi_k = J \Phi^k H'$ 이다. 행렬 Φ 의 모든 특성값들의 절대값이 1보다 작으므로 $\sum_{k=0}^{\infty} \Psi_k v_{t-k}$

는 평균제곱수렴하고, 두 확률벡터들 y_t 와 $(\sum_{k=1}^{\infty} \Psi_k v_{t-k})$ 는 L^2 공간에서 동일하다. 결론적으로 작인적 VARMA(p, q)과정 $\{y_t\}$ 는 다음과 같은 VMA(∞)모형을 만족한다.

$$y_t = \Psi(B)v_t$$

여기서 $\Psi(x) = \sum_{k=0}^{\infty} \Psi_k x^k$ ($\Psi_0 = I_d$)이다. 이 VARMA(p, q)과정은 다음 식을 만족한다.

$$\Phi(B)y_t = \Theta(B)v_t$$

따라서 다음 식이 성립한다.

$$\Phi(x)\Psi(x) = \Theta(x)$$

항등식의 원리에 의해서 다음 식들을 얻는다.

$$\Psi_0 = I_d$$

$$\Psi_1 = \Phi_1 \Psi_0 - \Theta_1$$

$$\Psi_2 = \Phi_1 \Psi_1 + \Phi_2 \Psi_0 - \Theta_2$$

⋮

$$\Psi_p = \Phi_1 \Psi_{p-1} + \Phi_2 \Psi_{p-2} + \dots + \Phi_p \Psi_0 - \Theta_p$$

$$\Psi_i = \Phi_1 \Psi_{i-1} + \Phi_2 \Psi_{i-2} + \dots + \Phi_p \Psi_{i-p} - \Theta_i \quad (i = p+1, p+2, \dots)$$

$$\text{단, } \Theta_i = 0 \quad (i = q+1, q+2, \dots)$$

이 식들을 이용하면 VARMA(p, q)과정을 VMA(∞)모형으로 나타낼 때의 VMA계수행렬들 $\{\Psi_i\}$ 를 쉽게 구할 수 있다. 한 예로 다음과 같은 2변량VARMA(2, 2)과정을 살펴보자.

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + v_t - \Theta_1 v_{t-1} - \Theta_2 v_{t-2}$$

$$\text{단, } \Phi_1 = \begin{bmatrix} 0.45 & 0.21 \\ -0.34 & 0.23 \end{bmatrix}, \quad \Phi_2 = \begin{bmatrix} 0.51 & 0.29 \\ 0.25 & 0.49 \end{bmatrix}$$

$$\Theta_1 = \begin{bmatrix} 0.40 & 0.12 \\ 0.21 & -0.31 \end{bmatrix}, \quad \Theta_2 = \begin{bmatrix} -0.80 & 0.23 \\ 0.34 & -0.50 \end{bmatrix}$$

이 2변량VARMA(2, 2)과정을 VMA(∞)모형으로 나타내기 위해 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-26.PRG에 수록되어 있다).

```

/*
/*
/*          EXAM6-26.PRG          */
/*
/*          CALCULATE MA COEFFICIENT MATRICES          */
/*          OF 2-VARIATE VARMA(2, 2) MODEL          */
/*

```

PSI0		PSI1	
1	0	0.05	0.09
0	1	-0.55	0.54
PSI2		PSI3	
1.217	0.2139	0.364615	0.526311
-0.2335	1.0836	-0.724485	0.463602
PSI4		PSI5	
0.5648899	0.7575294	0.2088927	0.851297
-0.100766	0.5121217	-0.479083	0.2189707
PSI6		PSI7	
0.2522662	0.9639228	0.0623544	0.9736895
-0.089365	0.2012443	-0.288852	0.0386723

Exiting IML.

그림 6.26 2변량VARMA(2, 2)과정의 VMA(∞)표현

```

OPTIONS PAGESIZE = 60;
PROC IML;
RESET NOLOG NOCENTER;
PHI1 = { 0.45 0.21,
         -0.34 0.23 };
PHI2 = { 0.51 0.29,
         0.25 0.49 };
THETA1 = { 0.4 0.12,
           0.21 -0.31 };
THETA2 = { -0.8 0.23,
           0.34 -0.5 };
PSI0 = PHI1**0;
PSI1 = PHI1*PSI0 - THETA1;
PSI2 = PHI1*PSI1 + PHI2*PSI0 - THETA2;
PSI3 = PHI1*PSI2 + PHI2*PSI1;
PSI4 = PHI1*PSI3 + PHI2*PSI2;
PSI5 = PHI1*PSI4 + PHI2*PSI3;
PSI6 = PHI1*PSI5 + PHI2*PSI4;
PSI7 = PHI1*PSI6 + PHI2*PSI5;
PRINT PSI0 " " PSI1, PSI2 " " PSI3,
       PSI4 " " PSI5, PSI6 " " PSI7, ;
QUIT;

```

이 SAS프로그램을 수행한 결과가 그림 6.26이다.

6.5.3 VARMA모형의 가역성

우선 d 변량VARMA(p, q)과정 $\{y_t\}$ 를 다음과 같이 $d(p+q)$ 변량VMA(1)모형으로 나타내보자.

$$w_t = u_t - \Theta u_{t-1}$$

$$\text{단, } w_t := \begin{bmatrix} y_t \\ 0 \\ \vdots \\ 0 \\ y_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad u_t := \begin{bmatrix} v_t \\ u_{t-1} \\ \vdots \\ v_{t-q+1} \\ y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}$$

$$\Theta := \begin{bmatrix} \Theta_{11} & \Theta_{12} \\ \Theta_{21} & \Theta_{22} \end{bmatrix}$$

$$\Theta_{11} := \begin{bmatrix} \Theta_1 & \Theta_2 & \cdots & \Theta_{q-1} & \Theta_q \\ I & O & \cdots & O & O \\ O & I & \cdots & O & O \\ \vdots & \vdots & & \vdots & \vdots \\ O & O & \cdots & I & O \end{bmatrix}$$

$$\Theta_{12} := \begin{bmatrix} -\Phi_1 & -\Phi_2 & \cdots & -\Phi_{p-1} & \Phi_p \\ O & O & \cdots & O & O \\ O & O & \cdots & O & O \\ \vdots & \vdots & & \vdots & \vdots \\ O & O & \cdots & O & O \end{bmatrix}$$

$$\Theta_{21} := O$$

$$\Theta_{22} := \begin{bmatrix} O & O & \cdots & O & O \\ I & O & \cdots & O & O \\ O & I & \cdots & O & O \\ \vdots & \vdots & & \vdots & \vdots \\ O & O & \cdots & I & O \end{bmatrix}$$

이 $d(p+q)$ 변량VMA(1)과정 $\{w_t\}$ 가 가역적이기 위한 필요충분조건은 $|x| \leq 1$ 에서 $\det(I - \Theta x) \neq 0$ 이어야 한다. 제 6.5.2소절에서 사용한 방법을 이용해서 다음 식들을 쉽게 유도할 수 있다.

$$\begin{aligned}
& \det(I - \Theta x) \\
&= \det(I - \Theta_{11} x) \det(I - \Theta_{22} x) \\
&= \det\left(I - \sum_{k=0}^q \Theta_k x^k\right) \\
&= \det[\Theta(x)]
\end{aligned}$$

따라서 $d(p+q)$ 변량VAR(1)과정 $\{w_t\}$ 가 가역적이기 위한 필요충분조건은 $|x| \leq 1$ 에서 $\det[\Theta(x)] \neq 0$ 이어야 한다. 즉, d 변량VARMA(p, q)과정 $\{y_t\}$ 가 작인적이기 위해서는 VMA특성방정식 $\det[\Theta(x)] = 0$ 의 모든 근들이 단위원 밖에 존재해야 한다.

가역적인 VARMA(p, q)과정을 VMA(∞)모형으로 나타내기 위해 다음과 같은 $d \times d$ ($p+q$)행렬 J 와 H 를 정의하자.

$$\begin{aligned}
J &:= [I \ O \ \cdots \ 0 \ 0 \ 0 \ \cdots \ 0] \\
H &:= [I \ O \ \cdots \ 0 \ I \ O \ \cdots \ 0]
\end{aligned}$$

제 6.4.1 소절과 같은 방법으로 다음 식들을 유도할 수 있다.

$$\begin{aligned}
& v_t \\
&= J u_t \\
&= J \left(\sum_{k=0}^{\infty} \Theta^k w_{t-k} \right) \\
&= J \left(\sum_{k=0}^{\infty} \Theta^k H' J w_{t-k} \right) \\
&= \sum_{k=0}^{\infty} (J \Theta^k H') y_{t-k} \\
&= - \sum_{k=0}^{\infty} \Pi_k y_{t-k}
\end{aligned}$$

여기서 $\Pi_k = J \Theta^k H'$ ($k=0, 1, 2, \dots$)이다. 행렬 Θ 의 모든 특성값들의 절대값들이 1보다 작으므로 $\sum_{k=1}^{\infty} \Pi_k y_{t-k}$ 는 평균제곱수렴하고, 두 확률벡터들 y_t 와 $(\sum_{k=1}^{\infty} \Pi_k y_{t-k} + v_t)$ 는 L^2 공간에서 동일하다. 따라서 d 변량VARMA(p, q)과정 $\{y_t\}$ 는 다음과 같은 d 변량VAR(∞)모형을 만족한다.

$$\Pi(B) y_t = v_t$$

여기서 $\Pi(x) = -\sum_{k=0}^{\infty} \Pi_k x^k$, ($\Pi_0 = -I_d$)이다. 또한 이 d 변량 VARMA(p, q)과정은 다음 식을 만족한다.

$$\Phi(B) y_t = \Theta(B) v_t$$

따라서 다음 식이 성립한다.

$$\Theta(x) \Pi(x) = \Phi(x)$$

SAS		23:54 Thursday, October 29, 1992 2	
PI0		PI1	
-1	0	0.05	0.09
0	-1	-0.55	0.54
PI2		PI3	
1.264	0.1608	0.35002	0.2175
0.091	0.8415	0.52923	-0.466497
PI4		PI5	
-0.786754	0.0959254	-0.437751	-0.264019
0.2937029	-0.175789	-0.401875	0.3818374
PI6		PI7	
0.47363	-0.176959	0.401418	0.2218576
-0.381694	-0.053304	0.2698897	-0.301322
Exiting IML.			

그림 6.27 2변량VARMA(2, 2)과정의 VAR표현

```
PRINT PI0 " " PI1, PI2 " " PI3,
      PI4 " " PI5, PI6 " " PI7, ;
QUIT;
```

이 SAS프로그램을 수행한 결과가 그림 6.27이다.

6.5.4 예제 들

[예제 1]

다음과 같은 2변량VARMA(2, 2)과정을 살펴보자.

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 0.45 & -0.11 \\ -0.14 & -0.23 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} 0.31 & 0.19 \\ 0.05 & 0.39 \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} \\ + \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix} - \begin{bmatrix} 0.40 & 0.12 \\ 0.21 & -0.31 \end{bmatrix} \begin{bmatrix} v_{1,t-1} \\ v_{2,t-1} \end{bmatrix} - \begin{bmatrix} -0.70 & 0.23 \\ -0.24 & 0.50 \end{bmatrix} \begin{bmatrix} v_{1,t-2} \\ v_{2,t-2} \end{bmatrix}$$

여기서 $\{v_t = (v_{1t}, v_{2t})'\}$ 는 2변량백색잡음과정으로서 평균벡터가 영벡터이고 공분산행렬은 다음과 같다.

$$\begin{bmatrix} 1 & -0.2 \\ -0.2 & 1.04 \end{bmatrix}$$

이 2변량VARMA(2, 2)모형의 VAR특성식은 다음과 같다.

$$\det \begin{bmatrix} 1 - 0.45x - 0.31x^2 & 0.11x - 0.19x^2 \\ 0.14x - 0.05x^2 & 1 + 0.23x - 0.39x^2 \end{bmatrix}$$

$$= (1 - 0.45x - 0.31x^2)(1 + 0.23x - 0.39x^2) - (0.14x - 0.05x^2)(0.11x - 0.19x^2)$$

따라서 이 2변량VARMA(2, 2)모형의 VAR특성근들은 다음과 같다.

$$\begin{aligned} x_1 &= -3.09104, & x_2 &= -1.25077 \\ x_3 &= 1.22886, & x_4 &= 1.88943 \end{aligned}$$

모든 VAR특성근들의 절대값들이 1보다 크므로 이 VARMA(2, 2)모형은 작인적이다. 이 2변량VARMA(2, 2)모형의 VMA특성식은 다음과 같다.

$$\begin{aligned} & \det \begin{bmatrix} 1 - 0.4x + 0.70x^2 & -0.12x - 0.23x^2 \\ -0.21x + 0.24x^2 & 1 + 0.31x - 0.50x^2 \end{bmatrix} \\ &= (1 - 0.4x - 0.70x^2)(1 + 0.31x - 0.50x^2) - (-0.21x + 0.24x^2)(-0.12x - 0.23x^2) \end{aligned}$$

따라서 이 2변량VARMA(2, 2)모형의 VMA특성근들은 다음과 같다.

$$\begin{aligned} x_1 &= -1.24967 + 0.158789i, & x_2 &= -1.24976 - 0.158789i \\ x_3 &= 0.94677, & x_4 &= 1.64264 \end{aligned}$$

VMA특성근들 중에서 x_3 의 절대값이 1보다 작으므로 이 VARMA(2, 2)모형은 가역적이지 않다.

이 2변량VARMA(2, 2)모형에서 120개의 관찰점들을 생성하기 위해 다음 SAS프로그램을 실행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-28.PRG에 수록되어 있다).

```

/*                                                                    */
/*                                                                    */
/*          EXAM6-28.PRG                                                                    */
/*                                                                    */
/*                                                                    */
/*          GENERATE 2-VARIATE VARMA(2, 2) PROCESS                                                                    */
/*                                                                    */
/*          DATASET = VARMA22A                                                                    */
/*                                                                    */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=60;
DATA EXAM628;
  DO T = 1 TO 320;
    U1 = RANNOR(7654321);
    U2 = RANNOR(1234567);
    V1 = U1;
    V2 = U2 - 0.2*U1;
    V1LAG1 = LAG1(V1);
    V2LAG1 = LAG1(V2);
    V1LAG2 = LAG2(V1);
  
```



```

V2LAG2 = LAG2(V2);
IF V1LAG1 = . THEN V1LAG1 = 0.0;
IF V2LAG1 = . THEN V2LAG1 = 0.0;
IF V1LAG2 = . THEN V1LAG2 = 0.0;
IF V2LAG2 = . THEN V2LAG2 = 0.0;
IF Y1LAG1 = . THEN Y1LAG1 = 0.0;
IF Y2LAG1 = . THEN Y2LAG1 = 0.0;
IF Y1LAG2 = . THEN Y1LAG2 = 0.0;
IF Y2LAG2 = . THEN Y2LAG2 = 0.0;
Y1 = 0.45*Y1LAG1 - 0.11*Y2LAG1 + 0.31*Y1LAG2
      + 0.19*Y2LAG2 + V1 - 0.40*V1LAG1
      - 0.12*V2LAG1 + 0.70*V1LAG2 - 0.23*V2LAG2;
Y2 = -0.14*Y1LAG1 - 0.23*Y2LAG1 + 0.05*Y1LAG2
      + 0.39*Y2LAG2 + V2 - 0.21*V1LAG1
      + 0.31*V2LAG1 + 0.24*V1LAG2 - 0.50*V2LAG2;
OUTPUT;
Y1LAG2 = Y1LAG1;
Y2LAG2 = Y2LAG1;
Y1LAG1 = Y1;
Y2LAG1 = Y2;
END;
KEEP T Y1 Y2;
RUN;
DATA TSA4DATA.VARMA22A;
SET EXAM628;
T = T-200;
IF T < 1 THEN DELETE;
RUN;
TITLE1 'SIMULATED 2-VARIATE VARMA(2, 2) PROCESS';
PROC CORR COV;
VAR Y1 Y2;
RUN;
DATA EXAM628B;
SET TSA4DATA.VARMA22A;
MY1 = Y1;
MY2 = Y2 + 5.0;
KEEP T MY1 MY2;
RUN;
PROC GPLOT;

```

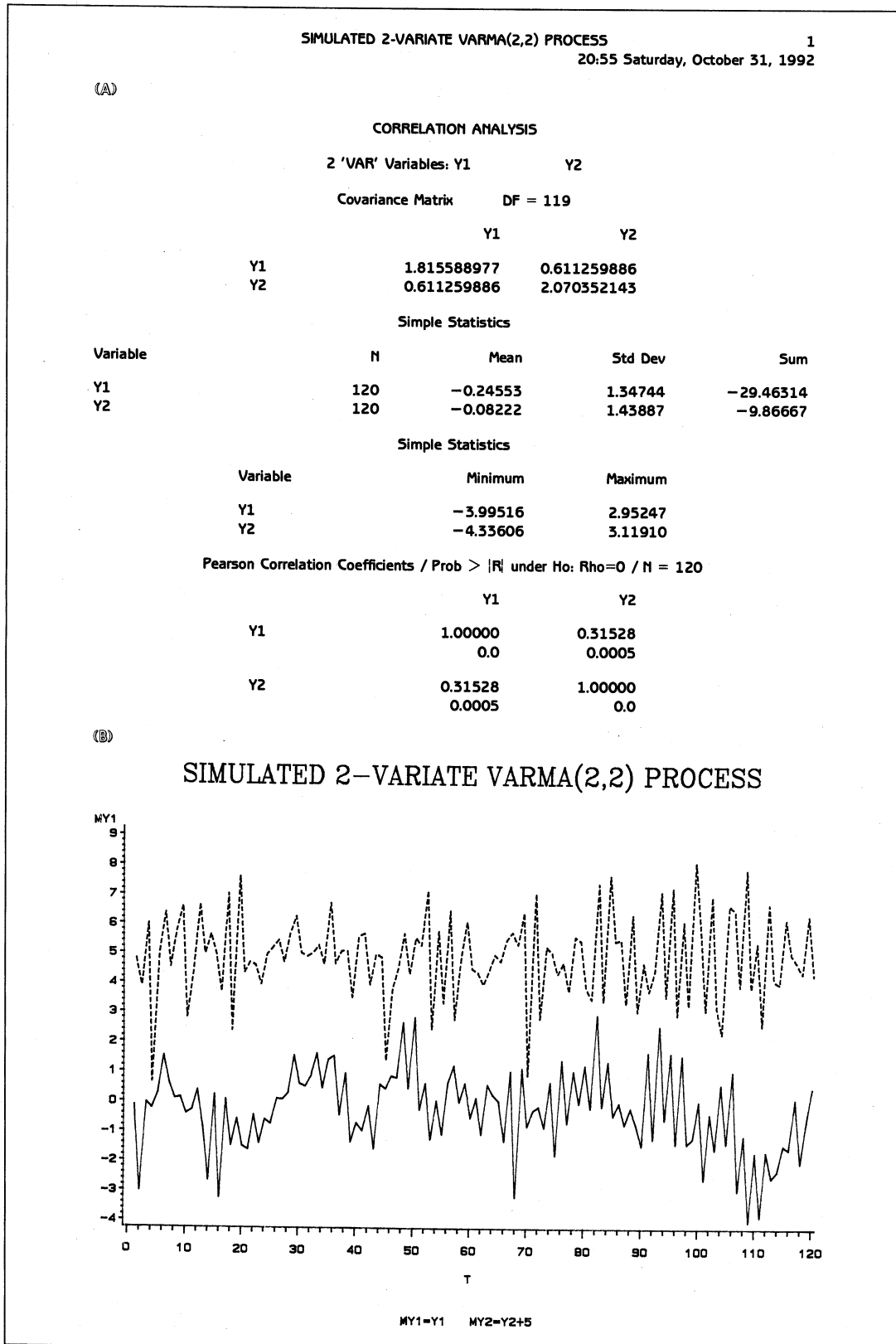


그림 6.28 2변량VARMA(2, 2)모형에서 생성된 다변량시계열데이터 VARMA2A

```

PLOT (MY1 MY2)*T / OVERLAY;
SYMBOL1 C=BLACK V=NONE I=JOIN LINE=1;
SYMBOL2 C=BLACK V=NONE I=JOIN LINE=2;
FOOTNOTE 'MY1=Y1 MY2=Y2+5';
RUN;

```

이 SAS프로그램을 수행한 결과가 그림 6.28이다. 그림 6.28의 부분 A는 PROC CORR를 실행한 결과이고, 부분 B는 PROC GPLOT를 실행한 결과이다. 부분 B의 산포도에서 이 2변량시계열데이터 (Y1, Y2)'가 정상적임을 알 수 있다.

[예제 2]

다음과 같은 3변량VARMA(1, 2)과정을 살펴보자.

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix} = \begin{bmatrix} 0.48 & -0.02 & 0.15 \\ -0.14 & 0.57 & -0.21 \\ 0.11 & -0.01 & 0.37 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} + \begin{bmatrix} v_{1,t} \\ v_{2,t} \\ v_{3,t} \end{bmatrix} \\
 - \begin{bmatrix} 0.48 & -0.21 & -0.11 \\ 0.18 & -0.49 & -0.17 \\ -0.22 & 0.23 & -0.37 \end{bmatrix} \begin{bmatrix} v_{1,t-1} \\ v_{2,t-1} \\ v_{3,t-1} \end{bmatrix} - \begin{bmatrix} -0.31 & 0.14 & 0.26 \\ -0.17 & 0.40 & 0.22 \\ 0.15 & -0.21 & 0.44 \end{bmatrix} \begin{bmatrix} v_{1,t-2} \\ v_{2,t-2} \\ v_{3,t-2} \end{bmatrix}$$

여기서 $\{v_t = (v_{1t}, v_{2t}, v_{3t})'\}$ 는 3변량백색잡음과정으로서 평균벡터가 영벡터이고 공분산행렬은 다음과 같다.

$$\begin{bmatrix} 1 & -0.2 & 0 \\ -0.2 & 1.04 & 0.3 \\ 0 & 0.3 & 1.09 \end{bmatrix}$$

이 3변량VARMA(1, 2)모형의 VAR특성식은 다음과 같다.

$$\det \begin{bmatrix} 1-0.48t & 0.02t & -0.15t \\ 0.14t & -0.57t & 0.21t \\ -0.11t & 0.01t & 1-0.37t \end{bmatrix} \\
 = 1 - 1.42t + 0.6407t^2 - 0.90455t^3$$

따라서 이 3변량VARMA(1, 2)모형의 VAR특성근들은 다음과 같다.

$$x_1 = 1.56756, \quad x_2 = 2.0143, \quad x_3 = 3.50122$$

모든 VAR특성근들의 절대값들이 1보다 크므로 이 3변량VARMA(1, 2)모형은 작인적이다. 이 3변량VARMA(1, 2)모형의 VAR특성식은 다음과 같다.

$$\det \begin{bmatrix} 1 - 0.48t + 0.31t^2 & 0.21t - 0.14t^2 & 0.11t - 0.26t^2 \\ -0.18t + 0.17t^2 & 1 + 0.49t - 0.40t^2 & 0.17t - 0.22t^2 \\ 0.22t - 0.15t^2 & -0.23t + 0.21t^2 & 1 + 0.37t - 0.44t^2 \end{bmatrix}$$

$$= 1 + 0.38t - 0.7088t^2 + 0.141444t^3 + 0.197526t^4 - 0.217916t^5$$

따라서 이 3변량VARMA(1, 2)모형의 VMA특성근들은 다음과 같다.

$$\begin{aligned} x_1 &= -1.04344 + 0.267851i, & x_2 &= -1.04344 - 0.267851i \\ x_3 &= 0.762378 + 1.63308i, & x_4 &= 0.762378 - 1.63308i \\ x_5 &= 2.09377 + 0.171955i, & x_6 &= 2.09377 - 0.171955i \end{aligned}$$

이 3변량VARMA(1, 2)모형에서 130개의 관찰점들을 생성하기 위해 다음 SAS프로그램을 실행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-29.PRG에 수록되어 있다).

```

/* */
/*          EXAM6-29.PRG          */
/* */
/*          GENERATE 3-VARIATE VARMA(1, 2) PROCESS          */
/*          DATASET = VARMA12A          */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=60;
DATA EXAM629;
  DO T = 1 TO 330;
    U1 = RANNOR(7654321);
    U2 = RANNOR(1234567);
    U3 = RANNOR(1357531);
    V1 = U1;
    V2 = U2 - 0.2*U1;
    V3 = U3 + 0.3*U2;
    V1LAG1 = LAG1(V1);
    V2LAG1 = LAG1(V2);
    V3LAG1 = LAG1(V3);
    V1LAG2 = LAG2(V1);
    V2LAG2 = LAG2(V2);
    V3LAG2 = LAG2(V3);
    IF V1LAG1 = . THEN V1LAG1 = 0.0;
    IF V2LAG1 = . THEN V2LAG1 = 0.0;
    IF V3LAG1 = . THEN V3LAG1 = 0.0;
    IF V1LAG2 = . THEN V1LAG2 = 0.0;

```

```

IF V2LAG2 = . THEN V2LAG2 = 0.0;
IF V3LAG2 = . THEN V3LAG2 = 0.0;
IF Y1LAG1 = . THEN Y1LAG1 = 0.0;
IF Y2LAG1 = . THEN Y2LAG1 = 0.0;
IF Y3LAG1 = . THEN Y3LAG1 = 0.0;
Y1 = 0.48*Y1LAG1 - 0.02*Y2LAG1 + 0.15*Y3LAG1
      + V1 - 0.48*V1LAG1 + 0.21*V2LAG1
      + 0.11*V3LAG1 + 0.31*V1LAG2
      - 0.14*V2LAG2 - 0.26*V3LAG2;
Y2 = -0.14*Y1LAG1 + 0.57*Y2LAG1 - 0.21*Y3LAG1
      + V2 - 0.18*V1LAG1 + 0.49*V2LAG1
      + 0.17*V3LAG1 + 0.17*V1LAG2
      - 0.40*V2LAG2 - 0.22*V3LAG2;
Y3 = 0.11*Y1LAG1 - 0.01*Y2LAG1 + 0.37*Y3LAG1
      + V2 + 0.22*V1LAG1 - 0.23*V2LAG1
      + 0.37*V3LAG1 - 0.15*V1LAG2
      + 0.21*V2LAG2 - 0.44*V3LAG2;

OUTPUT;
Y1LAG1 = Y1;
Y2LAG1 = Y2;
Y3LAG1 = Y3;
END;

KEEP T Y1 Y2 Y3;
RUN;

DATA TSA4DATA.VARMA12A;
SET EXAM629;
T = T-200;
IF T < 1 THEN DELETE;
RUN;

TITLE1 'SIMULATED 3-VARIATE VARMA(1, 2) PROCESS';
PROC CORR COV;
VAR Y1 Y2 Y3;
RUN;

```

SIMULATED 3-VARIATE VARMA(1, 2) PROCESS

2

20:55 Saturday, October 31, 1992

(A)

CORRELATION ANALYSIS

3 'VAR' Variables: Y1 Y2 Y3

Covariance Matrix DF = 129

	Y1	Y2	Y3
Y1	1.413787127	0.063919580	-0.006786015
Y2	0.063919580	2.730473520	1.441914574
Y3	-0.006786015	1.441914574	1.489073648

Simple Statistics

Variable	N	Mean	Std Dev	Sum
Y1	130	0.25358	1.18903	32.96594
Y2	130	0.13311	1.65241	17.30470
Y3	130	0.27246	1.22028	35.41931

Simple Statistics

Variable	Minimum	Maximum
Y1	-2.72917	3.60831
Y2	-3.98468	5.28021
Y3	-3.07592	3.74468

Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0 / N = 130

	Y1	Y2	Y3
Y1	1.00000 0.0	0.03253 0.7133	-0.00468 0.9579
Y2	0.03253 0.7133	1.00000 0.0	0.71509 0.0001
Y3	-0.00468 0.9579	0.71509 0.0001	1.00000 0.0

(B)

SIMULATED 3-VARIATE VARMA(1,2) PROCESS

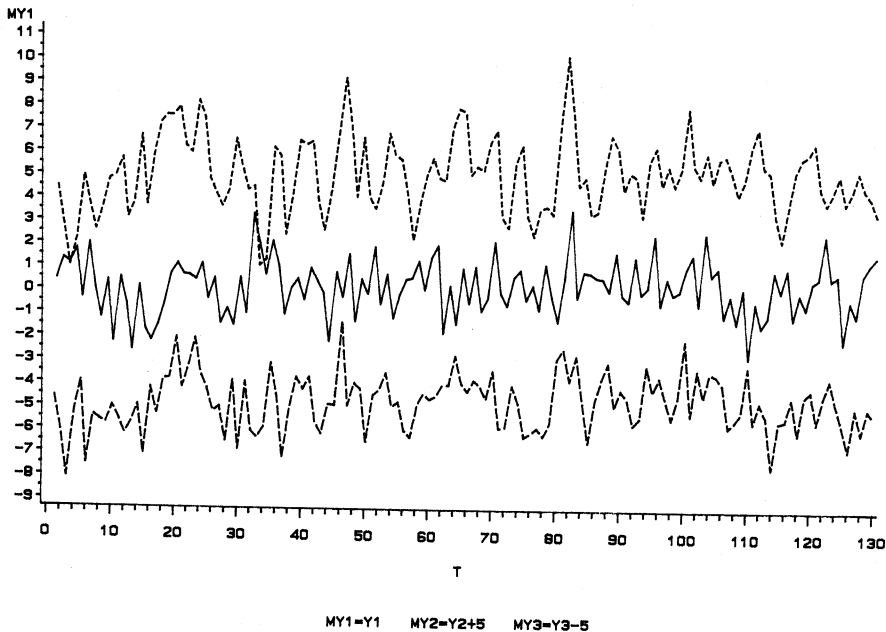


그림 6.29 3변량VARMA(1, 2)모형에서 생성된 다변량시계열데이터 VARMA12A

```

DATA EXAM629B;
  SET TSA4DATA.VARMA12A;
  MY1 = Y1;
  MY2 = Y2 + 5.0;
  MY3 = Y3 - 5.0;
  KEEP T MY1 MY2 MY3;
RUN;
PROC Gplot;
  PLOT (MY1 MY2 MY3)*T / OVERLAY;
  SYMBOL1 C=BLACK V=NONE I=JOIN LINE=1;
  SYMBOL2 C=BLACK V=NONE I=JOIN LINE=2;
  SYMBOL3 C=BLACK V=NONE I=JOIN LINE=3;
  FOOTNOTE 'MY1=Y1 MY2=Y2+5 MY3=Y3-5';
RUN;

```

이 SAS프로그램을 수행한 결과가 그림 6.29이다. 그림 6.29의 부분 A는 PROC CORR를 실행한 결과이고, 부분 B는 PROC Gplot를 실행한 결과인데, 단변량시계열들 Y1, Y2, Y3가 겹쳐지지 않도록 MY1=Y1, MY2=Y2+5, MY3=Y3-5의 시계열산포도를 그렸다. 이 산포도에서 이 3변량시계열데이터 (Y1, Y2, Y3)'가 정상적임을 알 수 있다.

[예제 3]

다음과 같은 2변량VARMA(2, 1)과정을 살펴보자.

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 0.52 & -0.02 \\ -0.03 & -0.28 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} -0.21 & -0.29 \\ -0.08 & 0.49 \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} \\
 + \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix} - \begin{bmatrix} -0.31 & 0.18 \\ 0.30 & -0.29 \end{bmatrix} \begin{bmatrix} v_{1,t-1} \\ v_{2,t-1} \end{bmatrix}$$

여기서 $\{v_t = (v_{1t}, v_{2t})\}$ 는 첫 번째 예제에서 정의된 2변량백색잡음과정이다. 이 2변량VARMA(2, 1)모형의 VAR특성식은 다음과 같다.

$$\det \begin{bmatrix} 1 - 0.52x + 0.21x^2 & 0.02x + 0.29x^2 \\ 0.03x + 0.08x^2 & 1 + 0.28x - 0.49x^2 \end{bmatrix} \\
 = 1 - 0.24x - 0.4262x^2 + 0.3033x^3 - 0.1261x^4$$

따라서 이 2변량VARMA(2, 1)모형의 VAR특성근들은 다음과 같다.

$$\begin{aligned}
 x_1 &= -1.16114, & x_2 &= 1.5545 \\
 x_3 &= 1.00594 + 1.8389i, & x_4 &= 1.00594 - 1.8389i
 \end{aligned}$$

모든 VAR특성근들의 절대값들이 1보다 크므로 이 2변량VARMA(2, 1)모형은 작인적이다. 이 2변량VARMA(2, 1)모형의 VMA특성식은 다음과 같다.

$$\det \begin{bmatrix} 1+0.31x & -0.18x \\ -0.30x & 1+0.29x \end{bmatrix} \\ = 1+0.6x+0.0359x^2$$

이 2변량VARMA(2, 1)모형의 VMA특성근들은 다음과 같다.

$$x_1 = -1.8776, \quad x_2 = -14.8355$$

이 2변량VARMA(2, 1)모형에서 120개의 관찰점들을 생성하기 위해 다음 SAS프로그램을 실행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-30.PRG에 수록되어 있다).

```

/ * * /
/ *          EXAM6-30.PRG          * /
/ * * /
/ *          GENERATE 2-VARIATE VARMA(2, 1) PROCESS          * /
/ *          DATASET = VARMA21A          * /
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=60;
DATA EXAM630;
  DO T = 1 TO 320;
    U1 = RANNOR(7654321);
    U2 = RANNOR(1234567);
    V1 = U1;
    V2 = U2 - 0.2 * U1;
    V1LAG1 = LAG1(V1);
    V2LAG1 = LAG1(V2);
    IF V1LAG1 = . THEN V1LAG1 = 0.0;
    IF V2LAG1 = . THEN V2LAG1 = 0.0;
    IF Y1LAG1 = . THEN Y1LAG1 = 0.0;
    IF Y2LAG1 = . THEN Y2LAG1 = 0.0;
    IF Y1LAG2 = . THEN Y1LAG2 = 0.0;
    IF Y2LAG2 = . THEN Y2LAG2 = 0.0;
    Y1 = 0.52 * Y1LAG1 - 0.02 * Y2LAG1 - 0.21 * Y1LAG2
        - 0.29 * Y2LAG2 + V1 + 0.31 * V1LAG1
        - 0.18 * V2LAG1;
    Y2 = - 0.03 * Y1LAG1 - 0.28 * Y2LAG1 - 0.08 * Y1LAG2

```

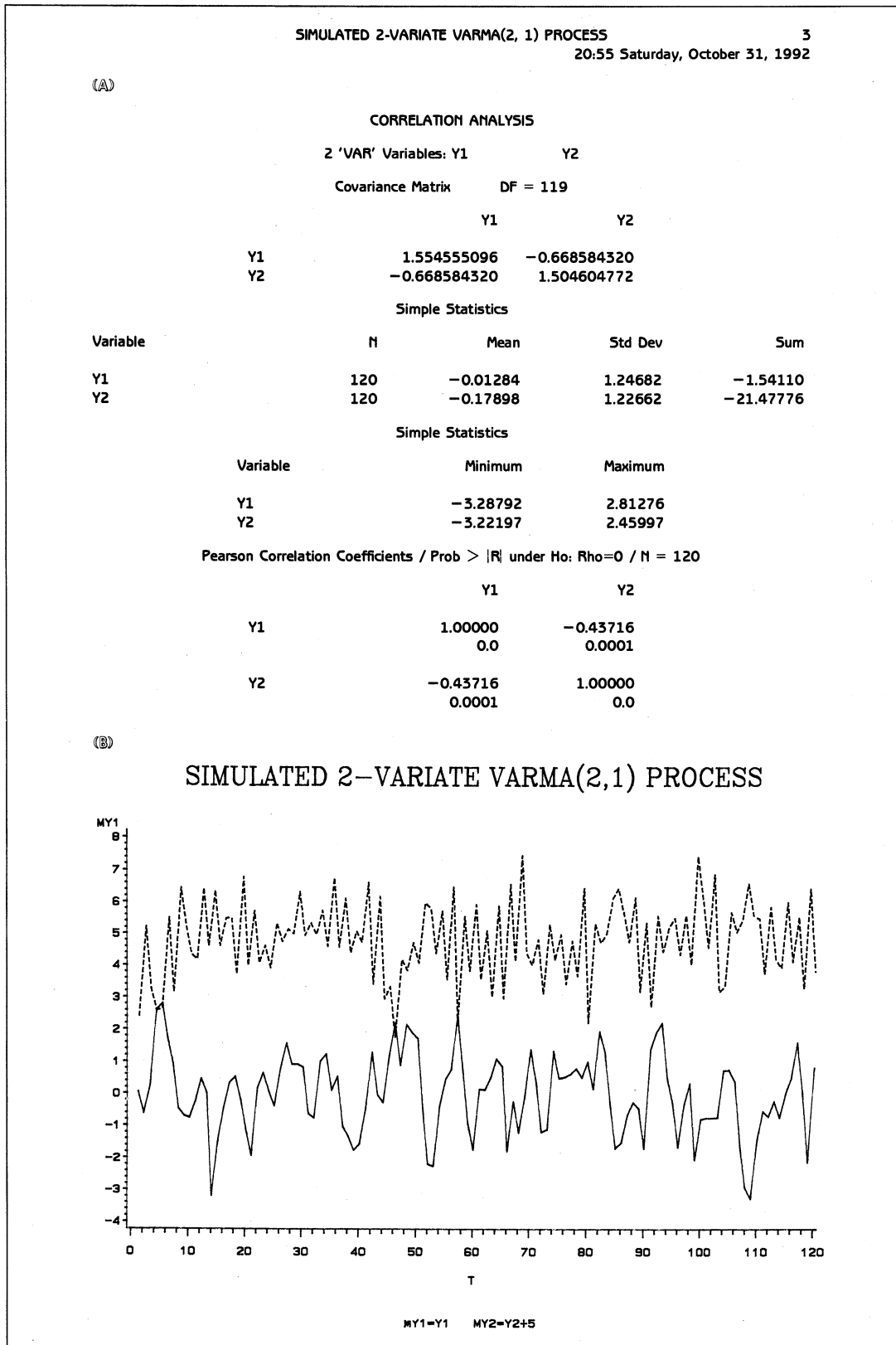



그림 6.30 2변량VARMA(2,1)모형에서 생성된 다변량시계열데이터 VARMA21A

```

      + 0.49*Y2LAG2 + V2 - 0.30*V1LAG1
      + 0.29*V2LAG1;
OUTPUT;
Y1LAG2 = Y1LAG1;
Y2LAG2 = Y2LAG1;
Y1LAG1 = Y1;
Y2LAG1 = Y2;
END;
KEEP T Y1 Y2;
RUN;
DATA TSA4DATA.VARMA21A;
SET EXAM630;
T = T-200;
IF T < 1 THEN DELETE;
RUN;
TITLE1 'SIMULATED 2-VARIATE VARMA(2, 1) PROCESS';
PROC CORR COV;
VAR Y1 Y2;
RUN;
DATA EXAM630B;
SET TSA4DATA.VARMA21A;
MY1 = Y1;
MY2 = Y2 + 5.0;
KEEP T MY1 MY2;
RUN;
PROC GPLOT;
PLOT (MY1 MY2)*T / OVERLAY;
SYMBOL1 C=BLACK V=NONE I=JOIN LINE=1;
SYMBOL2 C=BLACK V=NONE I=JOIN LINE=2;
FOOTNOTE 'MY1=Y1 MY2=Y2+5';
RUN;

```

이 SAS 프로그램을 수행한 결과가 그림 6.30이다. 그림 6.30의 부분 A는 PROC CORR를 실행한 결과이고, 부분 B는 PROC GPLOT를 실행한 결과인데, 단변량시계열들 Y1, Y2가 겹쳐지지 않도록 MY1=Y1, MY2=Y2+5의 시계열산포도를 그렸다. 부분 B로부터 2변량시계열데이터 (Y1, Y2)가 정상적임을 알 수 있다.

6.5.5 자기공분산행렬함수

다음과 같은 d 변량VARMA(p, q)과정의 자기공분산행렬함수를 구해 보자.

$$\Phi(B)y_t = \Theta(B)v_t$$

만일, 이 d 변량VARMA(p, q)과정이 작인적이면, 다음과 같은 VMA(∞)표현을 갖는다.

$$y_t = \Psi(B)v_t$$

$$\text{단, } \Psi(x) = \sum_{k=0}^{\infty} \Psi_k x^k = \Phi^{-1}(x)\Theta(x) \quad (\Psi_0 = I_d)$$

따라서 이 d 변량시계열데이터 $\{y_t\}$ 의 자기공분산행렬들은 다음과 같다.

$$\Gamma(i) = \sum_{j=0}^{\infty} \Psi_{i+j} \Sigma \Psi_j' \quad (i = 0, 1, 2, \dots)$$

그러나 자기공분산행렬들을 구할 때는 이 식보다는 확장Yule-Walker방정식들(extended Yule-Walker equations)을 사용하는 것이 편리하다. 확장Yule-Walker방정식들은 다음과 같다.

$$\Gamma(i) = \Phi_1 \Gamma(i-1) + \Phi_2 \Gamma(i-2) + \dots + \Phi_p \Gamma(i-p)$$

$$- \Theta_0 \Gamma_{yy}(i) - \Theta_1 \Gamma_{yy}(i-1) - \Theta_2 \Gamma_{yy}(i-2) - \dots - \Theta_q \Gamma_{yy}(i-q)$$

$$\text{단, } \Gamma_{yy}(j) = E(v_t y_{t-j}') \quad (\Theta_0 = -I_d)$$

작인성조건에 의해서 다음 식이 성립한다.

$$\Gamma_{yy}(0) = \Sigma$$

$$\Gamma_{yy}(j) = 0 \quad (j = 1, 2, \dots)$$

따라서 다음과 같은 확장Yule-Walker방정식들이 성립한다.

$$\Gamma(i) = \Phi_1 \Gamma(i-1) + \Phi_2 \Gamma(i-2) + \dots + \Phi_p \Gamma(i-p)$$

$$- \Theta_i \Gamma_{yy}(0) - \Theta_{i+1} \Gamma_{yy}(-1) - \dots - \Theta_q \Gamma_{yy}(i-q) \quad (i = 0, 1, \dots, q)$$

$$\Gamma(i) = \Phi_1 \Gamma(i-1) + \Phi_2 \Gamma(i-2) + \dots + \Phi_p \Gamma(i-p) \quad (i = q+1, q+2, \dots)$$

교차공분산행렬은 다음 식을 만족한다.

$$\Gamma_{yy}(-i) = E\left\{v_t \left(\sum_{j=0}^{\infty} \Psi_j v_{t+i-j}\right)'\right\} = \Sigma \Psi_i' \quad (i=0, 1, 2, \dots)$$

이 식은 다음과 같이 쓸 수 있다.

$$\Gamma_{yy}(i) = \Gamma_{yy}(-i)' = \Psi_i \Sigma \quad (i=0, 1, 2, \dots)$$

따라서 초기행렬들 $\Gamma(0), \Gamma(1), \dots, \Gamma(p-1), \Gamma_{yy}(0), \Gamma_{yy}(1), \dots, \Gamma_{yy}(q-1)$ 을 구하면, 확

$$\Sigma = \begin{bmatrix} 1.0 & -0.2 \\ -0.2 & 1.04 \end{bmatrix}$$

이 2변량VARMA(2, 2)과정의 자기공분산행렬함수를 구하기 위해 다음 SAS프로그램을 실행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-31.PRG에 수록되어 있다).

```

/* */
/*          EXAM6-31.PRG          */
/* */
/*          CALCULATE AUTOCOVARIANCE MATRICES          */
/*          OF 2-VARIATE VARMA(2, 2) MODEL          */
/*          BY LYAPUNOV - KALMAN - STEIN THEORM          */
/* */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOLOG NOCENTER;
/*      INPUT DATA      */
PHI1 = { 0.45 -0.11,
        -0.14 -0.23 };
PHI2 = { 0.31 0.19,
        0.05 0.39 };
THETA1 = { 0.40 0.12,
          0.21 -0.31 };
THETA2 = {-0.70 0.23,
          -0.24 0.50 };
SIGMA = { 1.0 -0.2,
          -0.2 1.04 };
PRINT PHI1 " " PHI2 ,
      THETA1 " " THETA2 ,
      SIGMA;
/*      MAKE SPECIAL MATRICES      */
I2 = I(2);
J2 = J(2, 2, 0);
/*      MAKE PHI-MATRIX      */
R1PHI = PHI1 || PHI2 || -THETA1 || -THETA2;
R2PHI = I2 || J2 || J2 || J2;
R3PHI = J2 || J2 || J2 || J2;
R4PHI = J2 || J2 || I2 || J2;
PHI = R1PHI // R2PHI // R3PHI // R4PHI;

```

```

PRINT PHI,;
/*      MAKE SIGMA_U MATRIX      */
R1SIGU = SIGMA || J2 || SIGMA || J2;
R2SIGU = J2 || J2 || J2 || J2;
SIGMAU = R1SIGU // R2SIGU // R1SIGU // R2SIGU;
PRINT SIGMAU,;
/*      INITIAL STEP      */
A = PHI;
B = SIGMAU;
OLDB = B*1.0E+10;
/*      ITERATIVE STEP      */
DO K = 1 TO 30
  UNTIL ( MAX(ABS(OLDB - B)) < 1.0E-15);
  EPSILON = MAX(ABS(OLDB - B));
  PRINT , "NO. OF ITERATION = " K
        " ERROR = " EPSILON;
  OLDA = A;
  OLDB = B;
  A = A*A;
  B = OLDA*B*OLDA' + B;
END;
/*      MAKE GAMMA_W(0)      */
GAMMAW0 = B;
PRINT , GAMMAW0 ,;
/*      FIND AUTOCOVARANCE FUNCTION OF Y      */
GAMYV0 = GAMMAW0[{1 2}, {5 6}];
GAMYV1 = GAMMAW0[{1 2}, {7 8}];
GAMMA0 = GAMMAW0[{1 2}, {1 2}];
GAMMA1 = GAMMAW0[{1 2}, {3 4}];
GAMMA2 = PHI1*GAMMA1 + PHI2*GAMMA0
        - THETA2*GAMYV0';
GAMMA3 = PHI1*GAMMA2 + PHI2*GAMMA1;
GAMMA4 = PHI1*GAMMA3 + PHI2*GAMMA2;
GAMMA5 = PHI1*GAMMA4 + PHI2*GAMMA3;
GAMMA6 = PHI1*GAMMA5 + PHI2*GAMMA4;
GAMMA7 = PHI1*GAMMA6 + PHI2*GAMMA5;
/*      PRINT AUTOCOVARANCE FUNCTION OF Y      */
PRINT , "SIGMA = GAMMYV0 = " GAMMYV0,
        GAMMYV1 " " GAMMYV2,,

```

SAS

13:28 Sunday, November 1, 1992 1

PHI1			PHI2	
0.45	-0.11		0.31	0.19
-0.14	-0.23		0.05	0.39
THETA1			THETA2	
0.4	0.12		-0.7	0.23
0.21	-0.31		-0.24	0.5
SIGMA				
1	-0.2			
-0.2	1.04			

PHI						
0.45	-0.11	0.31	0.19	-0.4	-0.12	0.7
-0.23						
-0.14	-0.23	0.05	0.39	-0.21	0.31	0.24
-0.5						
1	0	0	0	0	0	0
0						
0	1	0	0	0	0	0
0						
0	0	0	0	0	0	0
0						
0	0	0	0	0	0	0
0						
0	0	0	0	1	0	0
0						
0	0	0	0	0	1	0
0						
SIGMA						
1	-0.2	0	0	1	-0.2	0
0						
-0.2	1.04	0	0	-0.2	1.04	0
0						
0	0	0	0	0	0	0
0						
0	0	0	0	0	0	0
0						
0	1	-0.2	0	0	1	-0.2
0						
-0.2	1.04	0	0	-0.2	1.04	0
0						
0	0	0	0	0	0	0
0						
0	0	0	0	0	0	0
0						

NO. OF ITERATION =	K	ERROR =	EPSILON
	1		1.04E10
NO. OF ITERATION =	K	ERROR =	EPSILON
	2		1.04
NO. OF ITERATION =	K	ERROR =	EPSILON
	3		1.4206278
NO. OF ITERATION =	K	ERROR =	EPSILON
	4		0.9005827
NO. OF ITERATION =	K	ERROR =	EPSILON
	5		0.2228884
NO. OF ITERATION =	K	ERROR =	EPSILON
	6		0.0086354
NO. OF ITERATION =	K	ERROR =	EPSILON
	7		0.0000119
NO. OF ITERATION =	K	ERROR =	EPSILON
	8		2.252E-11

GAMMAWO							
3.4306437	0.1130646	1.628699	-0.821401	1	-0.2	0.096	
-0.2492							
0.1130646	1.7082674	-0.869567	-0.413523	-0.2	1.04	-0.366	
0.1532							
1.628699	-0.869567	3.4306437	0.1130646	0	0	1	
-0.2							
-0.821401	-0.413523	0.1130646	1.7082674	0	0	-0.2	
1.04							
1	-0.2	0	0	1	-0.2	0	
0							
-0.2	1.04	0	0	-0.2	1.04	0	
0							
0.096	-0.366	1	-0.2	0	0	1	
-0.2							
-0.2492	0.1532	-0.2	1.04	0	0	-0.2	
1.04							
SIGMA = GAMMYO =		GAMYVO					
		1	-0.2				
		-0.2	1.04				
GAMMY1		GAMMY2					
0.096	-0.2492						
-0.366	0.1532						
GAMMA0		GAMMA1					
3.4306437	0.1130646	1.628699	-0.821401				
0.1130646	1.7082674	-0.869567	-0.413523				
GAMMA2		GAMMA3					
2.6595487	-0.343722	1.4784389	-0.522417				
0.5276098	0.5139839	-0.751383	-0.226439				
GAMMA4		GAMMA5					
1.6726556	-0.257076	1.0347443	-0.346011				
0.3045819	0.230487	-0.523343	-0.131453				
GAMMA6		GAMMA7					
1.0995965	-0.177146	0.6965823	-0.229083				
0.1779244	0.1557119	-0.347233	-0.079581				
DOUBLE CHECK							
GAMMA1							
1.628699	-0.821401						
-0.869567	-0.413523						

그림 6.31 2변량VARMA(2, 2)과정의 자기공분산함수들

```

GAMMA0 " " GAMMA1,
GAMMA2 " " GAMMA3,
GAMMA4 " " GAMMA5,
GAMMA6 " " GAMMA7,;
/* DOUBLE-CHECK */
GAMMA1 = PHI1*GAMMA0' + PHI2*GAMMA1'
          - THETA1*GAMYV0' - THETA2*GAMYV1';
PRINT , "DOUBLE CHECK", GAMMA1,
QUIT;

```

이 SAS프로그램에서 $\Gamma(2)$ 는 다음 식을 사용해서 얻었다.

$$\Gamma(2) = \Phi_1\Gamma(1) + \Phi_2\Gamma(0) - \Theta_2\Gamma_{\text{yy}}(0)$$

또한 구해진 $\Gamma(1)$ 이 다음 식을 만족하는가를 조사하였다.

$$\Gamma(1) = \Phi_1\Gamma(0) + \Phi_2\Gamma(-1) - \Theta_1\Gamma_{\text{yy}}(0) - \Theta_2\Gamma_{\text{yy}}(-1)$$

이 SAS프로그램을 수행한 결과가 그림 6.31이다.

만일 VAR차수 p 가 VMA차수 q 보다 크다면, 교차공분산행렬들 $\Gamma_{\text{yy}}(0), \Gamma_{\text{yy}}(-1), \dots, \Gamma_{\text{yy}}(1-q)$ 를 사용하지 않고 초기자기공분산행렬들 $\Gamma(0), \Gamma(1), \dots, \Gamma(q)$ 와 다음과 같은 확장 Yule-Walker방정식들을 사용해서 자기공분산행렬함수를 구할 수 있다.

$$\Gamma(i) = \Phi_1\Gamma(i-1) + \Phi_2\Gamma(i-2) + \dots + \Phi_p\Gamma(i-p) \quad (i = q+1, q+2, \dots)$$

만일 $p \leq q$ 이면 VARMA(p, q)모형 대신 다음 가정을 만족하는 VARMA($q+1, q$)모형의 자기공분산행렬함수를 구한다.

$$\Phi_{p+1} = 0, \Phi_{p+2} = 0, \dots, \Phi_{q+1} = 0$$

이 VARMA($q+1, q$)모형의 자기공분산행렬함수는 교차공분산행렬들을 사용하지 않고 구할 수 있다. 이 자기공분산행렬함수는 원래의 VARMA(p, q)모형의 자기공분산행렬과 같다.

앞 예제의 2변량VARMA(2, 2)모형의 자기공분산행렬함수를 이 방법으로 구하기 위해 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-32.PRG에 수록되어 있다).

```

/*                                                    */
/*                                                    EXAM6-32.PRG          */
/*                                                    */
/*      CALCULATE AUTOCOVARIANCE MATRICES          */
/*      OF 2-VARIATE VARMA(2, 2) MODEL              */
/*      BY LYAPUNOV-KALMAN-STEIN THEOREM           */
/*      WITH ADDING PHI=0                           */
/*      IF NOT P > Q                                 */
/*                                                    */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOLOG NOCENTER;
/*      INPUT DATA      */
PHI1 = { 0.45 -0.11,
        -0.14 -0.23 };
PHI2 = { 0.31 0.19,
        0.05 0.39 };

```

```

PHI3 = J(2, 2, 0);
THETA1 = { 0.40  0.12,
           0.21 -0.31 };
THETA2 = {-0.70 0.23,
           -0.24 0.50 };
SIGMA = { 1.0 -0.2,
          -0.2  1.04 };
PRINT PHI1 " " PHI2 ,
        PHI3,
        THETA1 " " THETA2 ,
        SIGMA,;
/*      MAKE SPECIAL MATRICES      */
I2 = I(2);
J2 = J(2, 2, 0);
/*      MAKE PHI-MATRIX      */
R1PHI = PHI1 || PHI2 || PHI3 || -THETA1 || -THETA2;
R2PHI = I2 || J2 || J2 || J2 || J2;
R3PHI = J2 || I2 || J2 || J2 || J2;
R4PHI = J2 || J2 || J2 || J2 || J2;
R5PHI = J2 || J2 || J2 || I2 || J2;
PHI = R1PHI // R2PHI // R3PHI // R4PHI // R5PHI;
PRINT PHI,;
/*      MAKE SIGMA_U MATRIX      */
R1SIGU = SIGMA || J2 || J2 || SIGMA || J2;
R2SIGU = J2 || J2 || J2 || J2 || J2;
SIGMAU = R1SIGU // R2SIGU // R2SIGU // R1SIGU // R2SIGU;
PRINT SIGMAU,;
/*      INITIAL STEP      */
A = PHI;
B = SIGMAU;
OLDB = B*1.0E+10;
/*      ITERATIVE STEP      */
DO K = 1 TO 30
  UNTIL ( MAX(ABS(OLDB - B)) < 1.0E-15 );
  EPSILON = MAX(ABS(OLDB - B));
  PRINT , "NO. OF ITERATION = " K
        " ERROR = " EPSILON;
  OLDA = A;
  OLDB = B;

```



```

NO. OF ITERATION =          K          EPSILON
                          4          ERROR = 1.9558189

NO. OF ITERATION =          K          EPSILON
                          5          ERROR = 0.3976103

NO. OF ITERATION =          K          EPSILON
                          6          ERROR = 0.0150784

NO. OF ITERATION =          K          EPSILON
                          7          ERROR = 0.0000201

NO. OF ITERATION =          K          EPSILON
                          8          ERROR = 3.623E-11

  GAMMAWO
  3.4306437  0.1130646  1.628699  -0.821401  2.6595487  -0.343722      1
:    -0.2    0.096    -0.2492
  0.1130646  1.7082674  -0.869567  -0.413523  0.5276098  0.3139839  -0.2
:    1.04    -0.366    0.1532
  1.628699  -0.869567  3.4306437  0.1130646  1.628699  -0.821401      0
:    0        1        -0.2
  -0.821401 -0.413523  0.1130646  1.7082674  -0.869567  -0.413523      0
:    0        -0.2    1.04
  2.6595487  0.5276098  1.628699  -0.869567  3.4306437  0.1130646      0
:    0        0        0
  -0.343722  0.3139839  -0.821401  -0.413523  0.1130646  1.7082674      0
:    0        0        0
  1        -0.2    0        0        0        0        1
:   -0.2    0        0        0        0        0
  -0.2    1.04    0        0        0        0        -0.2
:   1.04    0        0        0        0        0
  0.096    -0.366    1        -0.2    0        0        0
:    0        1        -0.2
  -0.2492  0.1532    -0.2    1.04    0        0        0
:    0        -0.2    1.04

SIGMA = GAMMYO =          GAMMYO
                          1          -0.2
                          -0.2    1.04

  GAMMAO          GAMMA1
  3.4306437  0.1130646  1.628699  -0.821401
  0.1130646  1.7082674  -0.869567  -0.413523

  GAMMA2          GAMMA3
  2.6595487  -0.343722  1.4784389  -0.522417
  0.5276098  0.3139839  -0.751383  -0.226439

  GAMMA4          GAMMA5
  1.6726556  -0.257076  1.0347443  -0.346011
  0.3045819  0.230487  -0.523343  -0.131453

  GAMMA6          GAMMA7
  1.0995965  -0.177146  0.6965823  -0.229083
  0.1779244  0.1557119  -0.347233  -0.079581

Exiting IML.

```

그림 6.32 3변량VARMA(2, 2)과정의 자기공분산행렬함수

```

A = A*A;
B = OLDA*B*OLDA' + B;
END;
/*      MAKE GAMMA_W(0)      */
GAMMAWO = B;

```

```

PRINT , GAMMAW0 ,;
/ *      FIND AUTOCOVARIANCE FUNCTION OF Y      */
GAMYV0 = GAMMAW0[{{1 2}, {7 8}}];
GAMYV1 = GAMMAW0[{{1 2}, {9 10}}];
GAMMA0 = GAMMAW0[{{1 2}, {1 2}}];
GAMMA1 = GAMMAW0[{{1 2}, {3 4}}];
GAMMA2 = GAMMAW0[{{1 2}, {5 6}}];
GAMMA3 = PHI1*GAMMA2 + PHI2*GAMMA1 + PHI3*GAMMA0;
GAMMA4 = PHI1*GAMMA3 + PHI2*GAMMA2 + PHI3*GAMMA1;
GAMMA5 = PHI1*GAMMA4 + PHI2*GAMMA3 + PHI3*GAMMA2;
GAMMA6 = PHI1*GAMMA5 + PHI2*GAMMA4 + PHI3*GAMMA3;
GAMMA7 = PHI1*GAMMA6 + PHI2*GAMMA5 + PHI3*GAMMA4;
/ *      PRINT AUTOCOVARIANCE FUNCTION OF Y      */
PRINT , "SIGMA = GAMYV0 = " GAMYV0,,
      GAMMA0 " " GAMMA1,
      GAMMA2 " " GAMMA3,
      GAMMA4 " " GAMMA5,
      GAMMA6 " " GAMMA7,;
QUIT;

```

이 SAS프로그램을 수행한 결과가 그림 6.32이다. 그림 6.31의 자기공분산행렬함수와 그림 6.32의 자기공분산행렬함수는 같다.

[예제 2]

다음 3변량VARMA(1, 2)모형을 살펴보자.

$$y_t = \Phi_1 y_{t-1} + v_t - \Theta_1 v_{t-1} - \Theta_2 v_{t-2}$$

$$\text{단, } \Phi_1 = \begin{bmatrix} 0.48 & -0.02 & 0.15 \\ -0.14 & 0.57 & -0.21 \\ 0.11 & -0.01 & 0.37 \end{bmatrix}, \quad \Theta_1 = \begin{bmatrix} 0.48 & -0.21 & -0.11 \\ 0.18 & -0.49 & -0.17 \\ -0.22 & 0.23 & -0.37 \end{bmatrix}$$

$$\Theta_2 = \begin{bmatrix} -0.31 & 0.14 & 0.26 \\ -0.17 & 0.40 & 0.22 \\ 0.15 & -0.21 & 0.44 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1.0 & -0.2 & 0.0 \\ -0.2 & 1.04 & 0.3 \\ 0.0 & 0.3 & 1.09 \end{bmatrix}$$

이 3변량VARMA(1, 2)과정의 자기공분산행렬함수를 구하기 위해 다음 SAS프로그램을 실행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-33.PRG에 수록되어 있다).

```

/*
/*          EXAM6-33.PRG          */
/*
/*          CALCULATE AUTOCOVARANCE MATRICES          */
/*          OF 3-VARIATE VARMA(1, 2) MODEL          */
/*          BY LYAPUNOV - KALMAN - STEIN THEORM          */
/*

OPTIONS PAGESIZE=60;
PROC IML;
RESET NOLOG NOCENTER;
/*      INPUT DATA      */
PHI1 = { 0.48 -0.02  0.15,
        -0.14  0.57 -0.21,
        0.11 -0.01  0.37 };
THETA1 = { 0.48 -0.21 -0.11,
          0.18 -0.49 -0.17,
          -0.22  0.23 -0.37 };
THETA2 = {-0.31  0.14  0.26,
          -0.17  0.40  0.22,
          0.15 -0.21  0.44 };
SIGMA = { 1.0 -0.2  0.0 ,
          -0.2  1.04  0.3 ,
          0.0  0.3  1.09 };

PRINT PHI1,
      THETA1 " " THETA2 ,
      SIGMA;

/*      MAKE SPECIAL MATRICES      */
I3 = I(3);
J3 = J(3, 3, 0);
/*      MAKE PHI-MATRIX      */
R1PHI = PHI1 || -THETA1 || -THETA2;
R2PHI = J3 || J3 || J3;
R3PHI = J3 || I3 || J3;
PHI = R1PHI // R2PHI // R3PHI;
/*      MAKE SIGMA_U MATRIX      */
R1SIGU = SIGMA || SIGMA || J3;
R3SIGU = J3 || J3 || J3;
SIGMAU = R1SIGU // R1SIGU // R3SIGU;
/*      INITIAL STEP      */

```

```

A = PHI;
B = SIGMAU;
OLDB = B*1.0E+10;
/*      ITERATIVE STEP      */
DO K = 1 TO 30
  UNTIL ( MAX(ABS(OLDB - B)) < 1.0E-15);
  EPSILON = MAX(ABS(OLDB - B));
  OLDA = A;
  OLDB = B;
  A = A*A;
  B = OLDA*B*OLDA' + B;
  END;
/*      MAKE GAMMA_W(0)      */
GAMMAW0 = B;
/*      FIND AUTOCOVARANCE FUNCTION OF Y      */
GAMYV0 = GAMMAW0[{1 2 3}, {4 5 6}];
GAMYV1 = GAMMAW0[{1 2 3}, {7 8 9}];
GAMMA0 = GAMMAW0[{1 2 3}, {1 2 3}];
GAMMA1 = PHI1*GAMMA0 - THETA1*GAMYV0'
        - THETA2*GAMYV1';
GAMMA2 = PHI1*GAMMA1 - THETA2*GAMYV0';
GAMMA3 = PHI1*GAMMA2;
GAMMA4 = PHI1*GAMMA3;
GAMMA5 = PHI1*GAMMA4;
GAMMA6 = PHI1*GAMMA5;
GAMMA7 = PHI1*GAMMA6;
/*      PRINT AUTOCOVARANCE FUNCTION OF Y      */
PRINT , "SIGMA = GAMMYV0 = " GAMMYV0,
        GAMMYV1 ,,
        GAMMA0 " " GAMMA1,
        GAMMA2 " " GAMMA3,
        GAMMA4 " " GAMMA5,
        GAMMA6 " " GAMMA7,;
QUIT;

```

이 SAS프로그램을 수행한 결과가 그림 6.33이다.

이 3변량VARMA(1, 2)모형의 자기공분산행렬함수는 다음 조건을 만족하는 3변량VARMA(3, 2)모형의 자기공분산행렬함수이다.

$$\Phi_2 = 0, \quad \Phi_3 = 0$$

SAS			21:30 Sunday, November 1, 1992 1		
PHI1					
0.48	-0.02	0.15			
-0.14	0.57	-0.21			
0.11	-0.01	0.37			
THETA1			THETA2		
0.48	-0.21	-0.11	-0.31	0.14	0.26
0.18	-0.49	-0.17	-0.17	0.4	0.22
-0.22	0.23	-0.37	0.15	-0.21	0.44
SIGMA					
1	-0.2	0			
-0.2	1.04	0.3			
0	0.3	1.09			
SIGMA = GAMMYVO =			GAMYVO		
			1	-0.2	0
			-0.2	1.04	0.3
			0	0.3	1.09
GAMYV1					
-0.038	0.2756	0.3404			
-0.532	1.1544	0.2744			
0.378	-0.0936	0.7346			
GAMMA0			GAMMA1		
1.3523341	0.0080402	0.1562186	0.0115243	-0.081802	0.4669613
0.0080402	2.7388785	0.1306451	-0.687138	1.5166317	-0.056377
0.1562186	0.1306451	1.8160085	0.3862772	0.0511223	0.6225207
GAMMA2			GAMMA3		
0.415216	-0.347529	-0.006753	0.1976509	-0.157125	-0.011631
-0.2244	0.3491966	-0.588039	-0.177441	0.2243544	-0.306026
-0.040938	0.1111508	-0.134338	0.0327706	-0.000594	-0.044567
GAMMA4			GAMMA5		
0.1033369	-0.079996	-0.006148	0.0576618	-0.04436	-0.001888
-0.135694	0.1500044	-0.163447	-0.099298	0.1008489	-0.089215
0.0356411	-0.019747	-0.014709	0.0259112	-0.017606	-0.004484
GAMMA6			GAMMA7		
0.0335503	-0.025951	0.0002053	0.0200448	-0.015665	0.0009453
-0.070114	0.0673916	-0.049647	-0.048216	0.0446509	-0.028123
0.0169229	-0.012402	-0.000975	0.0106531	-0.008117	0.0001584
Exiting IML.					

그림 6.33 3변량VARMA(1, 2)과정의 자기공분산행렬함수

따라서 다음 프로그램을 수행해서 이 3변량VARMA(1, 2)과정의 자기공분산행렬함수를 구할 수 있다(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-34.PRG에 수록되어 있다).

```

/* */
/*          EXAM6-34.PRG          */
/* */
/*          CALCULATE AUTOCOVARIANCE MATRICES          */
/*          OF 3-VARIATE VARMA(1, 2) MODEL          */
/*          BY LYAPUNOV-KALMAN-STEIN THEORM          */
/*          WITH ADDING PHI=0          */
/* */

```



```

/*      IF NOT P > Q      */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOLOG NOCENTER;
/*      INPUT DATA      */
PHI1 = { 0.48 -0.02  0.15,
        -0.14  0.57 -0.21,
        0.11 -0.01  0.37 };
PHI2 = J(3, 3, 0);
PHI3 = J(3, 3, 0);
THETA1 = { 0.48 -0.21 -0.11,
          0.18 -0.49 -0.17,
          -0.22  0.23 -0.37 };
THETA2 = {-0.31  0.14  0.26,
          -0.17  0.40  0.22,
          0.15 -0.21  0.44 };
SIGMA = { 1.0 -0.2  0.0 ,
          -0.2  1.04  0.3 ,
          0.0  0.3  1.09 };
PRINT PHI1,
      PHI2 " " PHI3,
      THETA1 " " THETA2 ,
      SIGMA;
/*      MAKE SPECIAL MATRICES      */
I3 = I(3);
J3 = J(3, 3, 0);
/*      MAKE PHI-MATRIX      */
R1PHI = PHI1 || PHI2 || PHI3 || -THETA1 || -THETA2;
R2PHI = I3 || J3 || J3 || J3 || J3;
R3PHI = J3 || I3 || J3 || J3 || J3;
R4PHI = J3 || J3 || J3 || J3 || J3;
R5PHI = J3 || J3 || J3 || I3 || J3;
PHI = R1PHI // R2PHI // R3PHI // R4PHI // R5PHI;
/*      MAKE SIGMA_U MATRIX      */
R1SIGU = SIGMA || J3 || J3 || SIGMA || J3;
R2SIGU = J3 || J3 || J3 || J3 || J3;
SIGMAU = R1SIGU // R2SIGU // R2SIGU // R1SIGU // R2SIGU;
/*      INITIAL STEP      */
A = PHI;

```

SAS			21:30 Sunday, November 1, 1992 2		
PHI1			PHI3		
0.48	-0.02	0.15	0	0	0
-0.14	0.57	-0.21	0	0	0
0.11	-0.01	0.37	0	0	0
PHI2			PHI5		
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
THETA1			THETA2		
0.48	-0.21	-0.11	-0.31	0.14	0.26
0.18	-0.49	-0.17	-0.17	0.4	0.22
-0.22	0.23	-0.37	0.15	-0.21	0.44
SIGMA			GAMYVO		
1	-0.2	0	1	-0.2	0
-0.2	1.04	0.3	-0.2	1.04	0.3
0	0.3	1.09	0	0.3	1.09
SIGMA = GAMMYO =					
GAMMA0			GAMMA1		
1.3523341	0.0080402	0.1562186	0.0115243	-0.081802	0.4669613
0.0080402	2.7388785	0.1306451	-0.687138	1.5166317	-0.056377
0.1562186	0.1306451	1.8160085	0.3862772	0.0511223	0.6225207
GAMMA2			GAMMA3		
0.415216	-0.347529	-0.006753	0.1976509	-0.157125	-0.011631
-0.2244	0.3491966	-0.588039	-0.177441	0.2243544	-0.306026
-0.040938	0.1111508	-0.134338	0.0327706	-0.000594	-0.044567
GAMMA4			GAMMA5		
0.1033369	-0.079996	-0.006148	0.0576618	-0.04436	-0.001888
-0.135694	0.1500044	-0.163447	-0.099298	0.1008489	-0.089215
0.0356411	-0.019747	-0.014709	0.0259112	-0.017606	-0.004484
GAMMA6			GAMMA7		
0.0335503	-0.025951	0.0002053	0.0200448	-0.015665	0.0009453
-0.070114	0.0673916	-0.049647	-0.048216	0.0446509	-0.028123
0.0169229	-0.012402	-0.000975	0.0106531	-0.008117	0.0001584
8 workspace compresses.					
Exiting IML.					

그림 6.34 3변량VARMA(1, 2)과정의 자기공분산행렬함수

```

B = SIGMAU;
OLDB = B*1.0E+10;
/* ITERATIVE STEP */
DO K = 1 TO 30
  UNTIL (MAX(ABS(OLDB - B)) < 1.0E-15);
  EPSILON = MAX(ABS(OLDB - B));
  OLDA = A;
  OLDB = B;
  A = A*A;
  B = OLDA*B*OLDA' + B;

```

```

END:
/*      MAKE GAMMA_W(0)      */
GAMMAW0 = B;
/*      FIND AUTOCOVARIANCE FUNCTION OF Y      */
GAMYV0 = GAMMAW0[{1 2 3}, {10 11 12}];
GAMYV1 = GAMMAW0[{1 2 3}, {13 14 15}];
GAMMA0 = GAMMAW0[{1 2 3}, {1 2 3}];
GAMMA1 = GAMMAW0[{1 2 3}, {4 5 6}];
GAMMA2 = GAMMAW0[{1 2 3}, {7 8 9}];
GAMMA3 = PHI1*GAMMA2 + PHI2*GAMMA1 + PHI3*GAMMA0;
GAMMA4 = PHI1*GAMMA3 + PHI2*GAMMA2 + PHI3*GAMMA1;
GAMMA5 = PHI1*GAMMA4 + PHI2*GAMMA3 + PHI3*GAMMA2;
GAMMA6 = PHI1*GAMMA5 + PHI2*GAMMA4 + PHI3*GAMMA3;
GAMMA7 = PHI1*GAMMA6 + PHI2*GAMMA5 + PHI3*GAMMA4;
/*      PRINT AUTOCOVARIANCE FUNCTION OF Y      */
PRINT , "SIGMA = GAMYV0 = " GAMYV0,,
      GAMMA0 " " GAMMA1,
      GAMMA2 " " GAMMA3,
      GAMMA4 " " GAMMA5,
      GAMMA6 " " GAMMA7,;
QUIT:

```

이 SAS프로그램을 수행한 결과가 그림 6.34이다. 그림 6.34의 자기공분산행렬함수와 그림 6.33의 자기공분산행렬함수가 동일함을 알 수 있다.

[예제 3]

다음 2변량VARMA(2, 1)모형을 살펴보자.

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + v_t - \Theta_1 v_{t-1}$$

$$\text{단, } \Phi_1 = \begin{bmatrix} 0.52 & -0.02 \\ -0.03 & -0.28 \end{bmatrix}, \quad \Phi_2 = \begin{bmatrix} -0.21 & -0.29 \\ -0.08 & 0.49 \end{bmatrix}$$

$$\Theta_1 = \begin{bmatrix} -0.31 & 0.18 \\ 0.21 & -0.29 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1.0 & -0.2 \\ -0.2 & 1.04 \end{bmatrix}$$

이 2변량VARMA(2, 1)과정의 자기공분산행렬함수를 구하기 위해 다음 SAS프로그램을 실행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-35.PRG에 수록되어 있다).

```

/*
/*          EXAM6-35.PRG
/*
/*          CALCULATE AUTOCOVARIANCE MATRICES
/*          OF 2-VARIATE VARMA(2, 1) MODEL
/*          BY LYAPUNOV - KALMAN - STEIN THEORM
/*
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOLOG NOCENTER;
/*      INPUT DATA      */
PHI1 = { 0.52 -0.02,
        -0.03 -0.28 };
PHI2 = {-0.21 -0.29,
        -0.08  0.49 };
THETA1 = {-0.31  0.18,
          0.21 -0.29 };
SIGMA = { 1.0 -0.2,
          -0.2  1.04 };
PRINT PHI1 " " PHI2 ,
      THETA1,
      SIGMA;
/*      MAKE SPECIAL MATRICES      */
I2 = I(2);
J2 = J(2, 2, 0);
/*      MAKE PHI-MATRIX      */
R1PHI = PHI1 || PHI2 || -THETA1;
R2PHI = I2 || J2 || J2;
R3PHI = J2 || J2 || J2;
PHI = R1PHI // R2PHI // R3PHI;
/*      MAKE SIGMA_U MATRIX      */
R1SIGU = SIGMA || J2 || SIGMA;
R2SIGU = J2 || J2 || J2;
SIGMAU = R1SIGU // R2SIGU // R1SIGU;
/*      INITIAL STEP      */
A = PHI;
B = SIGMAU;
OLDB = B* 1.0E+10;
/*      ITERATIVE STEP      */

```

```

SAS                                21:30 Sunday, November 1, 1992  3

      PHI1          PHI2
      0.52          -0.21
     -0.03         -0.08
     -0.02         -0.29
     -0.28         0.49

      THETA1
     -0.31          0.18
      0.21         -0.29

      SIGMA
      1            -0.2
     -0.2         1.04

      SIGMA = GAMMYVO =
      1            -0.2
     -0.2         1.04

      GAMMA0          GAMMA1
     2.1124484      -0.711374
     -0.711374      1.6215748
     -0.493267      -0.519734
     -0.09733

      GAMMA2          GAMMA3
     0.4642026      -0.589183
     -0.419357      0.894326
     -0.244615      -0.186892
     -0.238849

      GAMMA4          GAMMA5
     0.0880425      -0.228033
     -0.177534      0.5578388
     0.0964365      -0.02122
     -0.081873      -0.251438

      GAMMA6          GAMMA7
     0.0847804      -0.119892
     -0.074004      0.362623
     0.0490573      0.007777
     -0.029655      -0.219445

DOUBLE CHECK

      GAMMA1
     1.330102      -0.519734
     -0.493267      -0.09733

Exiting IML.

```

그림 6.35 2변량VARMA(2, 1)과정의 자기공분산행렬함수

```

DO K = 1 TO 30
  UNTIL (MAX(ABS(OLDB - B)) < 1.0E-15);
  EPSILON = MAX(ABS(OLDB - B));
  OLDA = A;
  OLDB = B;
  A = A*A;
  B = OLDA*B*OLDA' + B;
END;

/*      MAKE GAMMA_W(0)      */
GAMMAW0 = B;

/*      FIND AUTOCOVARANCE FUNCTION OF Y      */
GAMYVO = GAMMAW0[1 2, 5 6];
GAMMA0 = GAMMAW0[1 2, 1 2];
GAMMA1 = GAMMAW0[1 2, 3 4];

```

```

GAMMA2 = PHI1 * GAMMA1 + PHI2 * GAMMA0;
GAMMA3 = PHI1 * GAMMA2 + PHI2 * GAMMA1;
GAMMA4 = PHI1 * GAMMA3 + PHI2 * GAMMA2;
GAMMA5 = PHI1 * GAMMA4 + PHI2 * GAMMA3;
GAMMA6 = PHI1 * GAMMA5 + PHI2 * GAMMA4;
GAMMA7 = PHI1 * GAMMA6 + PHI2 * GAMMA5;
/*      PRINT AUTOCOVARIANCE FUNCTION OF Y      */
PRINT , "SIGMA = GAMMYV0 = " GAMMYV0,,
      GAMMA0 " " GAMMA1,
      GAMMA2 " " GAMMA3,
      GAMMA4 " " GAMMA5,
      GAMMA6 " " GAMMA7,:
/*      DOUBLE-CHECK      */
GAMMA1 = PHI1 * GAMMA0' + PHI2 * GAMMA1'
      - THETA1 * GAMMYV0';
PRINT , "DOUBLE CHECK", GAMMA1,:
QUIT:

```

이 SAS프로그램을 수행한 결과가 그림 6.35이다. 이 SAS프로그램에서는 다음 등식이 만족되는지를 조사한다.

$$\Gamma(1) = \Phi_1 \Gamma(0) + \Phi_2 \Gamma(-1) - \Theta_1 \Gamma_{yy}(0)$$

지금까지 언급된 3개의 모형에서 자기공분산행렬함수를 구하는 데는 작인성조건이 사용되었다. 즉, 확장Yule-Walker방정식들은 작인성을 전제로 한다. 그러나 자기공분산행렬함수를 구하는 데는 가역성조건을 필요로 하지 않는다. 즉, 가역성조건의 성립 여부에 상관없이 작인성을 갖는 다변량시계열데이터의 자기공분산행렬함수는 확장Yule-Walker방정식들을 사용해서 구할 수 있다.

6.5.6 확장Whittle알고리즘

만일 VARMA(p, q)과정의 자기공분산행렬함수 $\{\Gamma(h)\}$ 가 주어지면 다음 확장Yule-Walker방정식들을 풀어서 VAR계수행렬들을 구할 수 있다.

$$\Gamma(i) = \Phi_1 \Gamma(i-1) + \Phi_2 \Gamma(i-2) + \dots + \Phi_p \Gamma(i-p) \quad (i=q+1, q+2, \dots)$$

제 6.3.4소절에서 보았듯이 이러한 문제를 해결하기 위해서는 Whittle알고리즘을 사용한다. 우리는 VAR차수 p 와 VMA차수 q 를 알지 못하므로 여러 쌍의 (p, q) 에 대해서 확장Yule-Walker방정식들을 풀어야 한다. 따라서 다음과 같은 VAR계수행렬들 $\{\Phi_i^{p,q} \mid i=1, 2,$

..., p }를 정의하자.

$$\Gamma(i) = \Phi_1^{p,q} \Gamma(i-1) + \Phi_2^{p,q} \Gamma(i-2) + \dots + \Phi_p^{p,q} \Gamma(i-p) \quad (i=q+1, q+2, \dots)$$

여기서 유의할 점은 이 VAR계수행렬들 $\{\Phi_i^{p,q}\}$ 은 VARMA(p, q)과정에서 뿐만 아니라 모든 정상적 다변량시계열데이터에서 정의된다는 것이다. 즉, 단지 $i=q+1, q+2, \dots, q+p$ 에서 이 확장Yule-Walker방정식들을 만족하는 계수행렬들을 VAR계수행렬들 $\{\Phi_i^{p,q}\}$ 라 부른다.

이 확장Yule-Walker방정식들을 풀기 위해서는 제 6.3.5소절의 확장Whittle알고리즘을 사용한다. 즉 알고리즘 6.3에 다음과 같이 할당한다.

$$C_i = \Gamma(q+i) \quad (i=0, \pm 1, \pm 2, \dots)$$

알고리즘 6.3을 적용한 결과로 다음과 같이 VAR계수행렬들을 구할 수 있다.

$$\Phi_i^{p,q} = L_{p,i} \quad (i=1, 2, \dots, p)$$

[예제 1]

그림 6.31의 자기공분산행렬함수로부터 VAR계수행렬들을 구하기 위해서 다음 SAS프로그램을 실행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-36.PRG에 수록되어 있다).

```

/*                                     */
/*                                     EXAM6-36.PRG                             */
/*                                     */
/*                                     THE EXATENDED WHITTLE ALGORITHM         */
/*                                     */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOLOG NOCENTER;
/*      INPUT DATA      */
GAMMA0 = { 3.4306437  0.1130646 ,
           0.1130646  1.7082674 };
GAMMA1 = { 1.628699  -0.821401 ,
           -0.869567  -0.413523 };
GAMMA2 = { 2.6595487  -0.343722 ,
           0.5276098   0.3139839 };
GAMMA3 = { 1.4784389  -0.522417 ,
           -0.751383  -0.226439 };
GAMMA4 = { 1.6726556  -0.257076 ,
           0.3045819   0.230487 };

```

```

GAMMA5 = { 1.0347443 -0.346011 ,
           -0.523343 -0.131453 };
GAMMA6 = { 1.0995965 -0.177146 ,
           0.1779244 0.1557119 };
GAMMA7 = { 0.6965823 -0.229083 ,
           -0.347233 -0.079581 };
C = GAMMA2 || GAMMA3 || GAMMA4 || GAMMA5
   || GAMMA6 || GAMMA7;
CM = GAMMA2 || GAMMA1 || GAMMA0 || GAMMA1'
     || GAMMA2' || GAMMA3';
D = 2;          /* DIMENSION OF SUBMATRICES */
MAXORDER = 4;  /* MAXIMUM OF VAR ORDERS */
/*   CONSTANTS   */
IDEN = I(D);
AROW = D;
ACOL = D*(MAXORDER+1);
L = J(AROW, ACOL, 0);
U = L;
M = L;
N = L;
G = J(D, D, 0);
H = G;
LAMBDA = G;
V = G;
AROWS = 1:D;
ACOLS = 1:D;
/*   INITIAL STEP K=0   */
L[AROWS, ACOLS] = -IDEN;
U[AROWS, ACOLS] = -IDEN;
M[AROWS, ACOLS] = -IDEN;
N[AROWS, ACOLS] = -IDEN;
ACOLSNXT = (D+1):(2*D);
G = C[AROWS, ACOLSNXT];
H = CM[AROWS, ACOLSNXT];
LAMBDA = C[AROWS, ACOLS];
V = LAMBDA;
/*   ITERATIVE STEP   */
DO K = 1 TO MAXORDER;
  LORD = L;

```



```

UOLD = U;
MOLD = M;
NOLD = N;
L[AROWS, ACOLS] = -IDEN;
U[AROWS, ACOLS] = -IDEN;
M[AROWS, ACOLS] = -IDEN;
N[AROWS, ACOLS] = -IDEN;
ACOLSLST = (K*D+1):(K+1)*D;
L[AROWS, ACOLSLST] = G*INV(V);
U[AROWS, ACOLSLST] = INV(V)*H;
M[AROWS, ACOLSLST] = H*INV(LAMBDA);
N[AROWS, ACOLSLST] = INV(LAMBDA)*G;
DO J = 1 TO K-1;
  ACOLSJ = (J*D+1):(J+1)*D;
  ACOLSKMJ = ((K-J)*D+1):((K-J+1)*D);
  L[AROWS, ACOLSJ] = LOLD[AROWS, ACOLSJ]
    - L[AROWS, ACOLSLST]*MOLD[AROWS, ACOLSKMJ];
  U[AROWS, ACOLSJ] = UOLD[AROWS, ACOLSJ]
    - NOLD[AROWS, ACOLSKMJ]*U[AROWS, ACOLSLST];
  M[AROWS, ACOLSJ] = MOLD[AROWS, ACOLSJ]
    - M[AROWS, ACOLSLST]*LOLD[AROWS, ACOLSKMJ];
  N[AROWS, ACOLSJ] = NOLD[AROWS, ACOLSJ]
    - UOLD[AROWS, ACOLSKMJ]*N[AROWS, ACOLSLST];
  END;
ACOLSNXT = ((K+1)*D+1):(K+2)*D;
G = C[AROWS, ACOLSNXT];
H = CM[AROWS, ACOLSNXT];
DO J = 1 TO K;
  ACOLSJ = (J*D+1):((J+1)*D);
  ACOLSKMJ = ((K-J+1)*D+1):((K-J+2)*D);
  G = G-L[AROWS, ACOLSKMJ]*C[AROWS, ACOLSJ];
  H = H-M[AROWS, ACOLSKMJ]*CM[AROWS, ACOLSJ];
  END;
LAMBDA = (IDEN -
  L[AROWS, ACOLSLST]*M[AROWS, ACOLSLST])*LAMBDA;
V = V*(IDEN -
  U[AROWS, ACOLSLST]*N[AROWS, ACOLSLST]);
/* PRINT THE SOLUTION */
PRINT,, "ORDER = " K,,

```

```

LAMBDA " " V,;
DO J = 1 TO K;
  ACOLSJ = (J*D+1):(J+1)*D;
  LKJ = L[AROWS, ACOLSJ];
  UKJ = U[AROWS, ACOLSJ];
  MKJ = M[AROWS, ACOLSJ];
  NKJ = N[AROWS, ACOLSJ];
  PRINT "LAG = " J,
        LKJ " " UKJ, MKJ " " NKJ, ;
END;
END;
QUIT;

```

이 SAS프로그램에서 유의할 점은 $\{\Gamma(h)\}$ 로부터 다음과 같은 행렬들을 만들어야 한다는 것이다.

$$C = [\Gamma(q), \Gamma(q+1), \dots, \Gamma(q+M+1)]$$

$$CM = [\Gamma(q), \Gamma(q-1), \dots, \Gamma(q-M-1)]$$

여기서 M 은 VAR계수들을 구하고자 하는 모형의 VAR차수 p 의 최대값이다. 이 SAS프로그램을 수행한 결과가 그림 6.36이다. 그림 6.31에서 볼 수 있듯이 GAMMA0, GAMMA1, ..., GAMMA7은 다음 2변량VARMA(2, 2)과정의 자기공분산행렬들이다.

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + v_t - \Theta_1 v_{t-1} - \Theta_2 v_{t-2}$$

$$\text{단, } \Phi_1 = \begin{bmatrix} 0.45 & -0.11 \\ -0.14 & -0.23 \end{bmatrix}, \quad \Phi_2 = \begin{bmatrix} 0.31 & 0.19 \\ 0.05 & 0.39 \end{bmatrix}$$

$$\Theta_1 = \begin{bmatrix} 0.40 & 0.12 \\ 0.21 & -0.31 \end{bmatrix}, \quad \Theta_2 = \begin{bmatrix} -0.70 & 0.23 \\ -0.24 & 0.50 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 1.0 & 0.2 \\ -0.2 & 1.04 \end{bmatrix}$$

그림 6.36에서 볼 수 있듯이 확장Whittle알고리즘을 적용한 결과로 다음 VAR계수행렬들을 얻었다.

$$\Phi_1^{1,2} = \begin{bmatrix} 0.728 & -0.867 \\ -0.115 & -0.847 \end{bmatrix}$$

$$\Phi_1^{2,2} = \begin{bmatrix} 0.450 & -0.110 \\ -0.140 & -0.230 \end{bmatrix}, \quad \Phi_2^{2,2} = \begin{bmatrix} 0.310 & 0.190 \\ 0.050 & 0.390 \end{bmatrix}$$

$$\Phi_1^{3,2} = \begin{bmatrix} 0.450 & -0.110 \\ -0.140 & -0.230 \end{bmatrix}, \quad \Phi_2^{3,2} = \begin{bmatrix} 0.310 & 0.190 \\ 0.050 & 0.390 \end{bmatrix}, \quad \Phi_3^{3,2} = \begin{bmatrix} 0.000 & 0.000 \\ 0.000 & 0.000 \end{bmatrix}$$

```

SAS                                     21:59 Tuesday, November 3, 1992  1

ORDER =                                K
                                      1

      LAMBDA
0.7201131  -0.104351      0.0842251  -0.220093
-0.021966  -0.130214      -0.426581  -0.025811

      J
      LAG =                             1

      LKJ
0.7278961  -0.866997      0.2090659  -0.393586
-0.114571  -0.846603      -3.120772  -0.655648

      MKJ
0.9295143  -1.59851      0.2026149  -0.237959
-0.053966  -1.376097      -2.733531  -0.321321

ORDER =                                K
                                      2

      LAMBDA
0.7460009  -0.379201      -0.234052  -0.217584
0.3399994  -0.567998      -1.670028  -0.292977

      J
      LAG =                             1

      LKJ
0.4500012  -0.110001      -0.630686  -0.353251
-0.14      -0.230002      3.030751  -7.970657

      MKJ
0.1949871  -2.874584      0.6647366  -0.153268
0.2942187  -8.79633      -1.611007  -0.444738

      J
      LAG =                             2

      LKJ
0.309999  0.1900011      2.4227319  -2.45067
0.0499994  0.3899985      -1.466098  -1.917171

      MKJ
0.6647158  -2.188016      0.1016727  -0.076245
-1.600005  -7.12617      1.2281375  0.1746771

ORDER =                                K
                                      3

      LAMBDA
0.7460024  -0.379204      -0.234054  -0.21758
0.3399971  -0.567994      -1.670035  -0.292967

      J
      LAG =                             1

      LKJ
0.449998      -0.11      -0.324577  -0.905047
-0.139997  -0.230004      0.0246586  -0.893802

      MKJ
0.255775  -0.294959      0.6647362  -0.153268
0.2700125  -1.474154      -1.611002  -0.444743

      J
      LAG =                             2

```

```

      LKJ
0.3099986 0.1900026
0.0499998 0.3899954

      UKJ
2.3118781 -1.728427
3.7962916 -13.90441

      MKJ
-0.751011 -3.711294
-5.975441 -11.45136

      NKJ
0.1016741 -0.076246
1.2281431 0.174669

      J
LAG = 3

      LKJ
2.8763E-6 -7.655E-7
-2.604E-6 4.9871E-7

      UKJ
1.5774719 -3.978381
6.1183739 -12.54232

      MKJ
0.9450029 -7.074838
3.3711646 -20.41726

      NKJ
1.4532E-6 -1.315E-6
1.263E-6 -1.527E-6

      K
ORDER = 4

      LAMBDA
0.746004 -0.379212
0.3399943 -0.567986

      V
-0.234048 -0.217578
-1.670003 -0.292958

      J
LAG = 1

      LKJ
0.4499959 -0.109987
-0.139995 -0.230015

      UKJ
-0.32456 -0.905059
0.0246779 -0.89382

      MKJ
0.2557504 -0.294954
0.2698994 -1.47413

      NKJ
0.6647295 -0.15327
-1.61102 -0.444749

      J
LAG = 2

      LKJ
0.3100027 0.1900097
0.0499976 0.3899897

      UKJ
3.280912 -2.375613
1.7739019 -3.902243

      MKJ
-0.447685 -0.311839
-7.489663 -0.78348

      NKJ
0.101676 -0.076246
1.2281144 0.1746607

      J
LAG = 3

      LKJ
2.7712E-6 2.1048E-7
-2.35E-6 -1.089E-7

      UKJ
3.5876337 -2.594984
11.472752 -29.16939

      MKJ
-0.515693 -9.080661
-5.686622 -26.72952

      NKJ
-2.126E-6 -2.173E-6
-1.238E-6 -2.181E-6

      J
LAG = 4

      LKJ
-3.652E-7 7.3512E-7
-7.07E-7 -2.708E-7

      UKJ
-0.135191 -5.83107
12.528978 -16.2638

      MKJ
0.4638904 -8.942659
10.089334 -32.26929

      NKJ
-2.995E-6 -5.617E-7
-2.88E-6 -7.467E-7

Exiting IML.

```

그림 6.36 확장Whittle알고리즘

$$\Phi_1^{4,2} = \begin{bmatrix} 0.450 & -0.110 \\ -0.140 & -0.230 \end{bmatrix}, \quad \Phi_2^{4,2} = \begin{bmatrix} 0.310 & 0.190 \\ 0.050 & 0.390 \end{bmatrix}, \quad \Phi_3^{4,2} = \begin{bmatrix} 0.000 & 0.000 \\ 0.000 & 0.000 \end{bmatrix}$$

$$\Phi_4^{4,2} = \begin{bmatrix} 0.000 & 0.000 \\ 0.000 & 0.000 \end{bmatrix}$$

이 결과로부터 주어진 자기공분산행렬함수가 2변량VARMA(2, 2)모형의 자기공분산행렬함수라고 확인할 수 있다(그러나 이러한 해석이 반드시 옳은 것은 아니라는 것을 제6.5.7소절에서 다루게 될 것이다).

확장Whittle알고리즘을 SAS/IML에서 모듈로 만들어 사용하기로 하자. 한 예로 그림 6.33의 자기공분산행렬함수로부터 VAR계수행렬들을 구하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-37.PRG에 수록되어 있다).

```

/* */
/*          EXAM6-37.PRG          */
/* */
/*          THE EXTENDED WHITTLE ALGORITHM          */
/*          MODULE BLOCKTOEP          */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOLOG NOCENTER;
/* */
/*          MODULE OF THE EXTENDED WHITTLE ALGORITHM          */
/* */
START BLKTOEP(C, CM, D, MAXORDER):
/*          CONSTANTS          */
/*          D          : DIMENSION OF SUBMATRICES          */
/*          MAXORDER : MAXIMUM OF VAR ORDERS          */
IDEN = I(D);
AROW = D;
ACOL = D*(MAXORDER+1);
L = J(AROW, ACOL, 0);
U = L;
M = L;
N = L;
G = J(D, D, 0);
H = G;
LAMBDA = G;
V = G;

```

```

AROWS = 1:D;
ACOLS = 1:D;
/*      INITIAL STEP K=0      */
L[AROWS, ACOLS] = -IDEN;
U[AROWS, ACOLS] = -IDEN;
M[AROWS, ACOLS] = -IDEN;
N[AROWS, ACOLS] = -IDEN;
ACOLSNXT = (D+1):(2*D);
G = C[AROWS, ACOLSNXT];
H = CM[AROWS, ACOLSNXT];
LAMBDA = C[AROWS, ACOLS];
V = LAMBDA;
/*      ITERATIVE STEP      */
DO K = 1 TO MAXORDER;
  LORD = L;
  UOLD = U;
  MOLD = M;
  NOLD = N;
  L[AROWS, ACOLS] = -IDEN;
  U[AROWS, ACOLS] = -IDEN;
  M[AROWS, ACOLS] = -IDEN;
  N[AROWS, ACOLS] = -IDEN;
  ACOLSLST = (K*D+1):(K+1)*D;
  L[AROWS, ACOLSLST] = G*INV(V);
  U[AROWS, ACOLSLST] = INV(V)*H;
  M[AROWS, ACOLSLST] = H*INV(LAMBDA);
  N[AROWS, ACOLSLST] = INV(LAMBDA)*G;
  DO J = 1 TO K-1;
    ACOLSJ = (J*D+1):(J+1)*D;
    ACOLSKMJ = ((K-J)*D+1):((K-J+1)*D);
    L[AROWS, ACOLSJ] = LORD[AROWS, ACOLSJ]
      - L[AROWS, ACOLSLST]*MOLD[AROWS, ACOLSKMJ];
    U[AROWS, ACOLSJ] = UOLD[AROWS, ACOLSJ]
      - NOLD[AROWS, ACOLSKMJ]*U[AROWS, ACOLSLST];
    M[AROWS, ACOLSJ] = MOLD[AROWS, ACOLSJ]
      - M[AROWS, ACOLSLST]*LORD[AROWS, ACOLSKMJ];
    N[AROWS, ACOLSJ] = NOLD[AROWS, ACOLSJ]
      - UOLD[AROWS, ACOLSKMJ]*N[AROWS, ACOLSLST];
  END;
END;

```

```

ACOLSNXT = ((K+1)*D+1):(K+2)*D;
G = C[AROWS, ACOLSNXT];
H = CM[AROWS, ACOLSNXT];
DO J = 1 TO K;
  ACOLSJ = (J*D+1):((J+1)*D);
  ACOLSKMJ = ((K-J+1)*D+1):((K-J+2)*D);
  G = G - L[AROWS, ACOLSKMJ] * C[AROWS, ACOLSJ];
  H = H - M[AROWS, ACOLSKMJ] * CM[AROWS, ACOLSJ];
  END;
LAMBDA = (IDEN -
  L[AROWS, ACOLSLST] * M[AROWS, ACOLSLST]) * LAMBDA;
V = V * (IDEN -
  U[AROWS, ACOLSLST] * N[AROWS, ACOLSLST]);
/*      PRINT THE SOLUTION      */
PRINT,, "ORDER = " K,,
      LAMBDA " " V,;
DO J = 1 TO K;
  ACOLSJ = (J*D+1):(J+1)*D;
  LKJ = L[AROWS, ACOLSJ];
  UKJ = U[AROWS, ACOLSJ];
  MKJ = M[AROWS, ACOLSJ];
  NKJ = N[AROWS, ACOLSJ];
  PRINT "LAG = " J,
      LKJ " " UKJ, MKJ " " NKJ, ;
  END;
END;
FINISH;
/*      */
/*      INPUT DATA      */
GAMMA0 = { 1.3523341  0.0080402  0.1562186,
           0.0080402  2.7388785  0.1306451,
           0.1562186  0.1306451  1.8160085 };
GAMMA1 = { 0.0115243 -0.081802  0.4669613,
           -0.687138  1.5166317 -0.056377,
           0.3862772  0.0511223  0.6225207 };
GAMMA2 = { 0.415216 -0.347529 -0.006753,
           -0.2244  0.3491966 -0.588039,
           -0.040938  0.1111508 -0.134338 };
GAMMA3 = { 0.1976509 -0.157125 -0.011631,

```

```

-0.177441 0.2243544 -0.306026,
0.0327706 -0.000594 -0.044567 };
GAMMA4 = { 0.1033369 -0.079996 -0.006148,
-0.135694 0.1500044 -0.163447,
0.0356411 -0.019747 -0.014709 };
GAMMA5 = { 0.0576618 -0.04436 -0.001888,
-0.099298 0.1008489 -0.089215,
0.0259112 -0.017606 -0.004484 };
GAMMA6 = { 0.0335503 -0.025951 0.0002053,
-0.070114 0.0673916 -0.049647,
0.0169229 -0.012402 -0.000975 };
GAMMA7 = { 0.0200448 -0.015665 0.0009453,
-0.048216 0.0446509 -0.028123,
0.0106531 -0.008117 0.0001584 };
GAMMA = GAMMA2 || GAMMA3 || GAMMA4 || GAMMA5
|| GAMMA6 || GAMMA7;
GAMMAM = GAMMA2 || GAMMA1 || GAMMA0 || GAMMA1'
|| GAMMA2' || GAMMA3';
RUN BLKTOEP(GAMMA, GAMMAM, 3, 4);
QUIT;

```

이 SAS프로그램에서는 모듈 BLKTOEP의 매개변수들 C, CM, D, MAXORDER 각각에 매개변수값들 GAMMA, GAMMAM, 3, 4를 대응시킨다. 이 SAS프로그램을 수행한 결과가 그림 6.37이다. 그림 6.34에서 볼 수 있듯이 GAMMA0, GAMMA1, ..., GAMMA7은 다음 3변량VARMA(1, 2)과정의 자기공분산행렬들이다.

$$y_t = \Phi_1 y_{t-1} + v_t - \Theta_1 v_{t-1} - \Theta_2 v_{t-2}$$

$$\text{단, } \Phi_1 = \begin{bmatrix} 0.48 & -0.02 & 0.15 \\ -0.14 & 0.57 & -0.21 \\ 0.11 & -0.01 & 0.37 \end{bmatrix}, \quad \Theta_1 = \begin{bmatrix} 0.48 & -0.21 & -0.11 \\ 0.18 & -0.49 & -0.17 \\ -0.22 & 0.23 & -0.37 \end{bmatrix}$$

$$\Theta_2 = \begin{bmatrix} -0.31 & 0.14 & 0.26 \\ -0.17 & 0.40 & 0.22 \\ 0.15 & -0.21 & 0.44 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1.0 & -0.2 & 0.0 \\ -0.2 & 1.04 & 0.3 \\ 0.0 & 0.3 & 1.09 \end{bmatrix}$$

그림 6.37에서 볼 수 있듯이 확장Whittle알고리즘을 적용한 결과로 다음 VAR계수행렬들을 얻었다.

$$\Phi_1^{1,2} = \begin{bmatrix} 0.480 & -0.020 & 0.150 \\ -0.140 & 0.570 & -0.210 \\ 0.110 & -0.010 & 0.370 \end{bmatrix}$$

$$\Phi_1^{2,2} = \begin{bmatrix} 0.480 & -0.020 & 0.150 \\ -0.140 & 0.570 & -0.210 \\ 0.110 & -0.010 & 0.370 \end{bmatrix}, \quad \Phi_2^{2,2} = \begin{bmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \end{bmatrix}$$

$$\Phi_1^{3,2} = \begin{bmatrix} 0.480 & -0.020 & 0.150 \\ -0.140 & 0.570 & -0.210 \\ 0.110 & -0.010 & 0.370 \end{bmatrix}, \quad \Phi_2^{3,2} = \begin{bmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \end{bmatrix}$$

$$\Phi_3^{3,2} = \begin{bmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \end{bmatrix}$$

$$\Phi_1^{4,2} = \begin{bmatrix} 0.480 & -0.020 & 0.150 \\ -0.140 & 0.570 & -0.210 \\ 0.110 & -0.010 & 0.370 \end{bmatrix}, \quad \Phi_2^{4,2} = \begin{bmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \end{bmatrix}$$

$$\Phi_3^{4,2} = \begin{bmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \end{bmatrix}, \quad \Phi_4^{4,2} = \begin{bmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \end{bmatrix}$$

이 결과로부터 주어진 자기공분산행렬함수가 3변량VARMA(1, 2)과정의 자기공분산행렬함수라고 확인할 수 있다(그러나 앞에서 언급한 대로 이러한 해석이 반드시 옳은 것은 아니라는 것을 제6.5.7소절에서 다룰 것이다).

```

SAS                                     22:11 Tuesday, November 3, 1992  2

ORDER =                                K
                                         1

      LAMBDA                               V
0.337996  -0.285597  -0.325404          0.2596816  -0.16702  -0.288236
0.2499989  -0.516  -0.359802           -1.46449  1.1768583  -1.160186
-0.192006  0.1164043  -0.416606         -1.231496  1.0582031  -0.803503

      LAG =                                J
                                         1

      LKJ                               UKJ
0.4799993  -0.020002  0.1500066          11.445375  -6.388751  14.220659
-0.139999  0.5699996  -0.209997          13.547231  -7.310584  15.570622
0.109999  -0.010003  0.3700103           4.8456699  -4.482402  3.9155043

      MKJ                               NKJ
-0.43302  -1.104183  1.3790981           1.9138435  -1.388221  0.5170583
-2.162645  -5.142676  23.039462           1.7065006  -1.195564  0.637596
0.0191764  -4.171727  13.62599            0.584789  -0.561739  0.7017298

ORDER =                                K
                                         2

      LAMBDA                               V
0.3379983  -0.285602  -0.325404          0.2596853  -0.167022  -0.288227
0.2499887  -0.515993  -0.359821         -1.464462  1.1768375  -1.16017
-0.192  0.1164001  -0.416601           -1.231481  1.0581928  -0.803497
    
```

LAG = J
1

LKJ			UKJ		
0.4799975	-0.019997	0.1500019	-1.793389	-0.205771	-0.117961
-0.140008	0.5700005	-0.209974	-0.196035	-0.071848	-0.612828
0.1099993	-0.01	0.3700007	0.3401824	0.0044417	0.0398734

MKJ			NKJ		
-4.068852	2.2451391	-1.350295	1.9138184	-1.388201	0.5170494
-1.974916	-3.690877	8.4005377	1.7064736	-1.195542	0.6375894
1.7986182	-3.853525	5.9343652	0.584779	-0.561732	0.7017225

LAG = J
2

LKJ			UKJ		
2.0642E-6	-1.238E-6	2.2291E-6	20.212893	-8.607658	34.669399
8.3732E-7	-3.919E-6	4.8241E-6	20.584526	-10.1485	40.98284
1.9146E-7	6.8917E-8	5.6447E-7	6.0541008	-7.344694	9.4381215

MKJ			NKJ		
5.4803965	-5.649499	1.9483765	-9.22E-7	7.566E-7	-4.538E-6
-10.81419	-2.176925	42.712261	-1.693E-6	8.4349E-7	-4.582E-6
-9.420815	-0.461599	24.344929	1.7433E-6	-1.665E-6	2.2244E-6

ORDER = K
3

LAMBDA			V		
0.3379938	-0.285598	-0.325433	0.2596675	-0.167014	-0.28824
0.2500079	-0.515993	-0.359762	-1.464463	1.1768389	-1.160185
-0.192001	0.1164015	-0.416602	-1.231481	1.0581946	-0.803502

LAG = J
1

LKJ			UKJ		
0.4800008	-0.020004	0.1500369	-1.793361	-0.205807	-0.117856
-0.139983	0.5699875	-0.210042	-0.195996	-0.07188	-0.612671
0.110001	-0.010003	0.3700022	0.3401612	0.0044635	0.0398344

MKJ			NKJ		
-4.068868	2.2451122	-1.350271	1.9137905	-1.38819	0.5170419
-1.974936	-3.69089	8.4004957	1.7064549	-1.195555	0.6375932
1.7986191	-3.85354	5.9343511	0.5847977	-0.561742	0.7017354

LAG = J
2

LKJ			UKJ		
-9.655E-6	1.8009E-6	5.0493E-6	1.8497406	-3.645915	-3.272444
2.754E-6	2.8284E-6	-9.692E-6	2.9832529	-4.768862	-0.327379
-1.158E-6	7.2483E-7	2.7556E-7	-0.815276	-0.328227	-2.388852

MKJ			NKJ		
-1.970968	0.6586684	-2.815269	8.0912E-6	-2.728E-6	2.9409E-6
-21.83035	0.1766184	9.6525996	-6.246E-7	3.3866E-7	-4.307E-6
-13.19521	0.6531121	4.6321179	2.3757E-7	-1.095E-6	9.4461E-7

LAG = J
3

LKJ			UKJ		
-5.943E-7	-2.3E-6	2.6454E-6	15.159778	6.4180919	75.6636
-3.593E-6	4.3053E-6	-4.466E-6	9.4215209	9.554993	85.165673
-4.059E-7	4.9355E-8	-1.135E-7	4.697911	-7.698602	21.975649

```

      MKJ
11.970177 -10.61193 1.9996479
2.2782609 -2.522827 86.994934
-5.000157 -1.172436 54.639522

      NKJ
4.3904E-6 -1.642E-6 3.8566E-6
5.7278E-6 -2.626E-6 3.4145E-6
-3.322E-7 9.3829E-9 -1.186E-6

ORDER =          K
              4

      LAMBDA
0.3380008 -0.285585 -0.325335
0.2500123 -0.516008 -0.359775
-0.191995 0.1163806 -0.416636

      V
0.2596919 -0.167033 -0.288234
-1.464421 1.1767764 -1.160089
-1.231456 1.0581567 -0.803446

      J
LAG =          1

      LKJ
0.4799876 -0.020013 0.1499417
-0.140006 0.5700056 -0.210029
0.1099989 -0.009984 0.3700232

      UKJ
-1.793404 -0.20588 -0.118432
-0.196031 -0.071978 -0.613268
0.3401753 0.0044606 0.0399356

      MKJ
-4.068932 2.2452223 -1.350387
-1.974846 -3.690804 8.4004038
1.7986642 -3.853482 5.9342902

      NKJ
1.9138093 -1.388161 0.5169595
1.7064484 -1.195493 0.6374919
0.584758 -0.561705 0.7016999

      J
LAG =          2

      LKJ
0.0000144 4.6187E-6 -0.000015
2.3005E-6 1.8676E-6 -3.412E-6
-5.21E-6 -2.296E-6 0.0000133

      UKJ
1.8497129 -3.646129 -3.273387
2.9833043 -4.768915 -0.327161
-0.815286 -0.328173 -2.388733

      MKJ
-1.970807 0.6587023 -2.815519
-21.82977 0.1764324 9.6523566
-13.19487 0.6530096 4.6319595

      NKJ
3.6545E-6 5.2486E-7 6.0429E-6
-8.249E-6 7.6209E-6 -5.321E-6
-2.688E-6 -8.816E-7 2.883E-6

      J
LAG =          3

      LKJ
8.6844E-6 -1.479E-6 -5.86E-6
3.9902E-6 -7.217E-7 -3.76E-7
-2.833E-6 -4.46E-7 2.8245E-6

      UKJ
11.546258 -12.04621 -6.287002
4.0179016 -6.151832 -2.79959
-2.845055 1.5047739 -0.776412

      MKJ
3.413333 1.4096216 -7.431869
-33.9996 -1.217517 28.553931
-26.27328 0.7294335 19.819638

      NKJ
-5.638E-6 2.8495E-6 1.076E-6
5.1044E-6 -2.637E-6 3.9468E-6
1.3806E-6 -7.2E-7 -7.279E-7

      J
LAG =          4

      LKJ
-1.269E-6 3.6496E-6 -4.022E-6
1.7416E-6 -6.65E-9 -2.835E-7
1.9791E-6 -2.336E-6 3.2617E-6

      UKJ
-0.62514 47.73759 162.14181
1.0439813 46.746663 179.87827
12.105694 -15.47684 41.293434

      MKJ
9.5476945 -20.58147 9.9351908
43.723509 1.7222216 141.19539
24.905721 -0.998327 83.441172

      NKJ
-5.11E-6 2.2319E-6 -1.505E-6
-4.877E-6 2.9977E-6 -1.257E-6
1.1897E-6 -1.084E-6 1.4976E-6

```

5 workspace compresses.
Exiting IML.

그림 6.37 모듈화된 확장Whittle알고리즘의 이용

6.5.7 VARMA모형의 비유일성

작인성과 가역성을 갖는 단변량ARMA모형의 자기공분산함수가 주어지면 이 ARMA모형의 차수들 p 와 q 가 유일하게 정해진다. 즉, 자기공분산함수가 주어지면 ARMA모형을 식별할 수가 있다. 다음과 같은 작인적이고 가역적인 ARMA(p, q)모형을 살펴보자.

$$\phi(B)y_t = \theta(B)v_t$$

이 단변량시계열데이터가 작인적이므로 다음과 같이 표현할 수 있다.

$$y_t = \phi^{-1}(B)\theta(B)v_t$$

여기서 유의할 점은 $\phi^{-1}(x)\theta(x)$ 가 x 의 무한차수의 멱함수(polynomial with infinite order)라는 것이다. 따라서 이 단변량시계열데이터는 MA(∞)모형으로 나타낼 수 있다. 또한 이 단변량시계열데이터는 가역적이므로 다음과 같이 표현할 수 있다.

$$v_t = \theta^{-1}(B)\phi(B)y_t$$

여기서도 $\theta^{-1}(x)\phi(x)$ 는 x 의 무한차수의 멱함수이다. 따라서 이 단변량시계열데이터는 AR(∞)모형으로도 나타낼 수 있다.

이에 반해서 이러한 성질들이 VARMA모형에서 항상 성립되는 것은 아니라는 것, 즉 VARMA모형의 경우에는 자기공분산행렬함수가 주어졌다고 해서 차수들 p 와 q 가 유일하게 정해지지 않는다는 것을 제 6.3.1소절과 제 6.4.1소절에서 알 수 있었다. 이 비유일성문제를 좀더 자세히 이해하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM6-38.PRG에 수록되어 있다).

```

/* */
/*          EXAM6-38.PRG          */
/* */
/*          CALCULATE AUTOCOVARANCE MATRICES          */
/*          OF THREE 2-VARIATE VARMA MODELS          */
/*          BY LYAPUNOV - KALMAN - STEIN THEORM          */
/*          WITH ADDING PHI=0          */
/*          IF NOT P > Q          */
/* */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOLOG NOCENTER;
/* */
/*          MODULE OF LYAPUNOV ALGORITHM          */
/* */

```

```

/*
START LYAPUNOV(PHI, SIGMAU, GAMMAW0):
/* INITIAL STEP */
A = PHI;
B = SIGMAU;
OLDB = B*1.0E+10;
/* ITERATIVE STEP */
DO K = 1 TO 30
  UNTIL ( MAX(ABS(OLDB - B)) < 1.0E-15 );
  EPSILON = MAX(ABS(OLDB - B));
  OLDA = A;
  OLDB = B;
  A = A*A;
  B = OLDA*B*OLDA' + B;
END;
/* MAKE GAMMA_W(0) */
GAMMAW0 = B;
FINISH;
/* INPUT DATA */
TRPHI1 = { 0.0 0.5,
           0.0 0.0 };
TRTHETA1 = { 0.0 0.0,
             0.3 0.0 };
SIGMA = { 1.0 -0.2,
          -0.2 1.04 };
/* MAKE SPECIAL MATRICES */
I2 = I(2);
J2 = J(2, 2, 0);
/*
/* VARMA(1, 1) MODEL */
/*
/* MAKE PHI-MATRIX */
PHI1 = TRPHI1;
PHI2 = J2;
THETA1 = TRTHETA1;
R1PHI = PHI1 || PHI2 || -THETA1;
R2PHI = I2 || J2 || J2;
R3PHI = J2 || J2 || J2;
PHI = R1PHI // R2PHI // R3PHI;

```

```

/*      MAKE SIGMA_U MATRIX      */
R1SIGU = SIGMA || J2 || SIGMA;
R2SIGU = J2 || J2 || J2;
SIGMAU = R1SIGU // R2SIGU // R1SIGU;
/*      FIND AUTOCOVARANCE FUNCTION OF Y      */
RUN LYAPUNOV(PHI, SIGMAU, GAMMAW0);
GAMMA0 = GAMMAW0[{1 2}, {1 2}];
GAMMA1 = GAMMAW0[{1 2}, {3 4}];
GAMMA2 = PHI1 * GAMMA1 + PHI2 * GAMMA0;
GAMMA3 = PHI1 * GAMMA2 + PHI2 * GAMMA1;
GAMMA4 = PHI1 * GAMMA3 + PHI2 * GAMMA2;
GAMMA5 = PHI1 * GAMMA4 + PHI2 * GAMMA3;
/*      PRINT AUTOCOVARANCE FUNCTION OF Y      */
PRINT , "THE ACVMF OF THE VARMA(1, 1) MODEL",;
PRINT , GAMMA0 " " GAMMA1,
        GAMMA2 " " GAMMA3,
        GAMMA4 " " GAMMA5,;

/*      */
/*      VAR(2) MODEL      */
/*      */
PHI1 = TRPHI - TRTHETA1;
PHI2 = TRTHETA1 * TRPHI;
/*      MAKE PHI-MATRIX      */
R1PHI = PHI1 || PHI2;
R2PHI = I2 || J2;
PHI = R1PHI // R2PHI;
/*      MAKE SIGMA_U MATRIX      */
R1SIGU = SIGMA || J2;
R2SIGU = J2 || J2;
SIGMAU = R1SIGU // R2SIGU;
/*      FIND AUTOCOVARANCE FUNCTION OF Y      */
RUN LYAPUNOV(PHI, SIGMAU, GAMMAW0);
GAMMA0 = GAMMAW0[{1 2}, {1 2}];
GAMMA1 = GAMMAW0[{1 2}, {3 4}];
GAMMA2 = PHI1 * GAMMA1 + PHI2 * GAMMA0;
GAMMA3 = PHI1 * GAMMA2 + PHI2 * GAMMA1;
GAMMA4 = PHI1 * GAMMA3 + PHI2 * GAMMA2;
GAMMA5 = PHI1 * GAMMA4 + PHI2 * GAMMA3;
/*      PRINT AUTOCOVARANCE FUNCTION OF Y      */

```

```

PRINT , "THE ACVMF OF THE VAR(2) MODEL",;
PRINT , GAMMA0 " " GAMMA1,
      GAMMA2 " " GAMMA3,
      GAMMA4 " " GAMMA5,;

/ *                               */
/ *   VMA(2) MODEL                 */
/ *                               */
/ *   MAKE PHI-MATRIX              */
PHI1 = J2;
PHI2 = J2;
PHI3 = J2;
THETA1 = TRTHETA1 - TRPHI1;
THETA2 = TRPHI1*TRTHETA1;
R1PHI = PHI1 || PHI2 || PHI1 || -THETA1 || -THETA2;
R2PHI = I2 || J2 || J2 || J2 || J2;
R3PHI = J2 || I2 || J2 || J2 || J2;
R4PHI = J2 || J2 || J2 || J2 || J2;
R5PHI = J2 || J2 || J2 || I2 || J2;
PHI = R1PHI // R2PHI // R3PHI // R4PHI // R5PHI;
/ *   MAKE SIGMA_U MATRIX          */
R1SIGU = SIGMA || J2 || J2 || SIGMA || J2;
R2SIGU = J2 || J2 || J2 || J2 || J2;
SIGMAU = R1SIGU // R2SIGU // R2SIGU // R1SIGU // R2SIGU;
/ *   FIND AUTOCOVARIANCE FUNCTION OF Y   */
RUN LYAPUNOV(PHI, SIGMAU, GAMMAW0);
GAMMA0 = GAMMAW0[{1 2}, {1 2}];
GAMMA1 = GAMMAW0[{1 2}, {3 4}];
GAMMA2 = GAMMAW0[{1 2}, {5 6}];
GAMMA3 = PHI1*GAMMA2 + PHI2*GAMMA1 + PHI3*GAMMA0;
GAMMA4 = PHI1*GAMMA3 + PHI2*GAMMA2 + PHI3*GAMMA1;
GAMMA5 = PHI1*GAMMA4 + PHI2*GAMMA3 + PHI3*GAMMA2;
/ *   PRINT AUTOCOVARIANCE FUNCTION OF Y   */
PRINT , "THE ACVMF OF THE VMA(2) MODEL",;
PRINT , GAMMA0 " " GAMMA1,
      GAMMA2 " " GAMMA3,
      GAMMA4 " " GAMMA5,;
QUIT;

```

이 SAS프로그램에서는 다음 2변량VARMA모형들의 자기공분산행렬함수들을 구하였다.

$$y_t = \Phi^* y_{t-1} + v_t - \Theta^* v_{t-1}$$

$$y_t = (\Phi^* - \Theta^*) y_{t-1} + \Phi^* \Theta^* y_{t-2} + v_t$$

$$y_t = v_t + (\Phi^* - \Theta^*) v_{t-1} + \Phi^* \Theta^* v_{t-2}$$

여기서 $\{v_t\}$ 는 평균벡터가 영벡터이고 공분산행렬이 Σ 인 2변량백색잡음과정이며 행렬들은 다음과 같이 정의된다.

$$\Phi^* = \begin{bmatrix} 0.0 & 0.5 \\ 0.0 & 0.0 \end{bmatrix}, \quad \Theta^* = \begin{bmatrix} 0.0 & 0.0 \\ 0.3 & 0.0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1.0 & -0.2 \\ -0.2 & 1.04 \end{bmatrix}$$

다음 등식들에 의해서 이 VARMA(1, 1)모형, VAR(2)모형, VMA(2)모형은 같은 자기공분산행렬함수를 갖는다.

THE ACVMF OF THE VARMA(1, 1) MODEL				
GAMMA0			GAMMA1	
1.2825	-0.17		-0.085	0.565
-0.17	1.13		-0.3	0.06
GAMMA2			GAMMA3	
-0.15	0.03		0	0
0	0		0	0
GAMMA4			GAMMA5	
0	0		0	0
0	0		0	0
THE ACVMF OF THE VAR(2) MODEL				
GAMMA0			GAMMA1	
1.2825	-0.17		-0.085	0.565
-0.17	1.13		-0.3	0.06
GAMMA2			GAMMA3	
-0.15	0.03		0	0
0	0		0	0
GAMMA4			GAMMA5	
0	0		0	0
0	0		0	0
THE ACVMF OF THE VMA(2) MODEL				
GAMMA0			GAMMA1	
1.2825	-0.17		-0.085	0.565
-0.17	1.13		-0.3	0.06
GAMMA2			GAMMA3	
-0.15	0.03		0	0
0	0		0	0
GAMMA4			GAMMA5	
0	0		0	0
0	0		0	0

Exiting IML.

그림 6.38 세 VARMA모형들의 자기공분산행렬함수들

$$(I - \Phi^*)^{-1} = I + \Phi^*$$

$$(I - \Theta^*)^{-1} = I + \Theta^*$$

이 SAS프로그램을 수행한 결과가 그림 6.38이다. 그림 6.38에서 세 VARMA모형들이 동일한 자기공분산행렬함수들을 가짐을 확인할 수 있다.

작인성조건을 만족하는 VAR특성행렬 $\Phi(x)$ 와 가역성조건을 만족하는 VMA특성행렬 $\Theta(x)$ 에 대해서 다음 세 VARMA모형들은 같은 자기공분산행렬함수를 가짐을 같은 방법으로 설명할 수 있다.

$$(I - \Phi^*B)\Phi(B)y_t = (I - \Theta^*B)\Theta(B)v_t$$

$$\Phi(B)y_t = \{I + (\Phi^* - \Theta^*)B - \Phi^*\Theta^*B^2\}\Theta(B)v_t$$

$$\{I - (\Phi^* - \Theta^*)B - \Phi^*\Theta^*B^2\}y_t = \Theta(B)v_t$$

다음의 작인적이고 가역적인 VARMA(p, q)모형을 살펴보자.

$$\Phi(B)y_t = \Theta(B)v_t$$

이 VARMA는 작인적이고 가역적이기 때문에 다음과 같이 표현할 수 있다.

$$y_t = \Phi^{-1}(B)\Theta(B)v_t$$

$$v_t = \Theta^{-1}(B)\Phi(B)y_t$$

Cramer법칙에 의해 다음 식들이 성립한다.

$$\Phi^{-1}(x) = \frac{1}{\det[\Phi(x)]} \text{Adj}[\Phi(x)]$$

$$\Theta^{-1}(x) = \frac{1}{\det[\Theta(x)]} \text{Adj}[\Theta(x)]$$

여기서 $\text{Adj}[A]$ 는 행렬 A 의 수반행렬(adjoint matrix)이다. 만일 $\det[\Phi(x)]$ 가 x 에 의존하지 않는 상수(constant)라면 $\Phi^{-1}(x)$ 는 x 의 무한차원의 멱함수가 아닌 유한차원의 멱함수이다. 따라서 작인적인 VARMA과정 중에는 유한차원의 VMA모형으로 나타낼 수 있는 것이 있다. 같은 이유로 $\det[\Theta(x)]$ 가 상수이면 이 VARMA과정을 유한차원의 VAR모형으로 나타낼 수 있다. 즉, 자기공분산행렬함수가 주어졌다고 해서 반드시 VARMA모형의 차수들 p 와 q 를 정할 수 있는 것은 아니다. Hannan(1969, 1976)은 이러한 VARMA모형의 식별문제에 대해 자세히 다루었고, Hannan & Deistler(1988, pp. 74~82)에서는 VARMA모형에 외생변수를 첨가한 VARMAX모형의 식별문제를 다루었다. Hannan은 다음 두 조건들이 만족되는 VARMA과정들만을 대상으로 식별문제를 다루었다.

(조건 A) VAR특성행렬 $\Phi(x)$ 와 VMA특성행렬 $\Theta(x)$ 의 공통좌공약항(common left divisor)의 모듈은 1이다. 즉, 만일 $\Phi(x) = C(x)A(x)$ 이고 $\Theta(x) = C(x)B(x)$ 이면, $\det[C(x)]$ 는 x 에 의존하지 않는 상수이다.

(조건 B) VAR특성방정식 $\det[\Phi(x)] = 0$ 의 모든 근은 단위원 밖에 있고, VMA특성 방

정식 $\det[\Theta(x)]=0$ 의 모든 근은 단위원 위 또는 단위원 밖에 있다.

이러한 두 조건들이 만족될 경우 다음 세 방법들 중의 한 방법을 사용하면 VARMA모형을 식별할 수 있다. 즉, 자기공분산행렬함수가 주어졌을 때 차수들 p 와 q 가 유일하게 정해진다.

- ① 주어진 자기공분산행렬함수를 갖는 모든 VARMA모형들 중에서 최소의 VMA차수 q 를 갖는 모형들을 선택한다. 이 VMA차수가 최소인 모형들 중에서 최소의 VAR차수 p 를 갖는 모형을 선택한다. 만일 Φ_p, Θ_q 가 완전랭크(full rank)를 가지면 이러한 VARMA모형은 유일하다.
- ② VAR특성행렬 $\Phi(x)$ 가 하삼각행렬형태(lower triangular matrix form)를 갖는 VARMA모형들을 대상으로 한다. 이 VAR특성행렬 $\Phi(x)$ 의 (i, j) 원소를 $\phi_{ij}(x)$ 로 표시하자. 모든 (i, j) 에 대해서 $\phi_{ij}(x)$ 의 차수가 $\phi_{ii}(x)$ 의 차수보다 작거나 같으면, 이러한 VARMA모형은 유일하다.
- ③ VAR특성행렬 $\Phi(x)$ 가 다음 형태를 갖는 VARMA모형을 대상으로 한다.

$$\Phi(x) = \phi(x)I_d$$

여기서 $\phi(x) = 1 - \phi_1 x - \phi_2 x^2 - \dots - \phi_p x^p$ 이다. 즉, $\phi(x)$ 는 단변량AR(p)과정의 특성식이다. 만일 $\phi_p \neq 0$ 이면 이러한 VARMA모형은 유일하다.

이 세 가지 방법들 중에서 첫 번째 방법이 가장 간단하고 가장 널리 쓰인다. 그러나 이 방법들은 $\det[\Phi(x)]$ 나 $\det[\Theta(x)]$ 가 (x 에 의존하지 않는) 상수인 경우에 대해서는 적용이 되지 않는다. 따라서 Hannan의 방법들을 사용해서 자기공분산행렬함수로부터 VARMA모형을 반드시 식별할 수 있는 것은 아니다. 이러한 이유로 VARMA모형의 작인성조건과 가역조건을 좀더 명확하게 기술할 필요가 있다. VARMA모형의 작인성조건은 다음과 같다. VAR특성식 $\det[\Phi(x)]$ 가 $|x| \leq 1$ 에서 0이 아니다. 즉, VAR특성방정식 $\det[\Phi(x)]=0$ 의 모든 근들의 절대값들이 1보다 크거나 VAR특성식 $\det[\Phi(x)]$ 가 x 에 의존하지 않는 상수이다. VARMA모형의 가역조건은 다음과 같다. VMA특성식 $\det[\Theta(x)]$ 가 $|x| \leq 1$ 에서 0이 아니다. 즉, VMA특성방정식 $\det[\Theta(x)]=0$ 의 모든 근의 절대값들이 1보다 크거나 VMA특성식 $\det[\Theta(x)]$ 가 x 에 의존하지 않는 상수이다. VARMA모형의 식별문제는 제 8.2절에서 다시 한 번 언급하게 될 것이다.

6.5.8 VMA계수행렬

원칙적으로 VMA모형과 마찬가지로 VARMA모형의 자기공분산행렬함수가 주어지면 VMA계수행렬들을 구할 수 있다. 그러나 현실적으로 이러한 문제를 만족스럽게 해결할 수 있는 알고리즘이 없다. 따라서 이 소절은 문제를 제기하는 수준이 될 것이다.

다음과 같은 VARMA(1, 1)모형을 살펴보자.


```

/ *          UNIDENTIFIABILITY OF VARMA MODELS          */
/ *          */
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOLOG NOCENTER;
/ *          */
/ *          SUBROUTINE FOR COEF MATRICES          */
/ *          */
START VMA1COEF(M, GAMMA0, GAMMA1, THETA);
THETA = 0.5*GAMMA0;
OLDTHETA = 1.0E+9*I(M);
DO K = 1 TO 50
  UNTIL (MAX(ABS(OLDTHETA - THETA)) < 1.0E-10 );
  EPSILON = MAX(ABS(OLDTHETA-THETA));
  OLDTHETA = THETA;
  THETA = (1/ 2.4)*(1.4*THETA-
                (THETA*GAMMA0+2.0*GAMMA1)
                *INV(2.0*THETA*GAMMA1'+GAMMA0));
END;
FINISH;
/ *          */
/ *          SUBROUTINE FOR ACVMF          */
/ *          */
START ACVMF(PHI1, THETA1, SIGMA,
            GAMMA0, GAMMA1, GAMMA2);
/ *          MAKE SPECIAL MATRICES          */
I2 = I(2);
J2 = J(2, 2, 0);
PHI2 = J2;
/ *          MAKE PHI-MATRIX          */
R1PHI = PHI1 || PHI2 || -THETA1 ;
R2PHI = I2 || J2 || J2 ;
R3PHI = J2 || J2 || J2 ;
PHI = R1PHI // R2PHI // R3PHI ;
/ *          MAKE SIGMA_U MATRIX          */
R1SIGU = SIGMA || J2 || SIGMA ;
R2SIGU = J2 || J2 || J2 ;
SIGMAU = R1SIGU // R2SIGU // R1SIGU ;
/ *          INITIAL STEP          */

```

```

A = PHI;
B = SIGMAU;
EPSILON = 100.0;
/*      ITERATIVE STEP      */
DO K = 1 TO 30
  UNTIL (EPSILON < 1.0E-10);
  OLDA = A;
  OLDB = B;
  A = A*A;
  B = OLDA*B*OLDA' + B;
  EPSILON = MAX(ABS(OLDB-B));
END;
/*      FIND AUTOCOVARIANCE FUNCTION OF Y      */
SIGMA = B[{1 2}, {5 6}];
GAMMA0 = B[{1 2}, {1 2}];
GAMMA1 = B[{1 2}, {3 4}];
GAMMA2 = PHI1*GAMMA1;
FINISH;
/*      */
/*      1. CALCULATE AUTOCOVARIANCE MATRICES      */
/*      */
/*      INPUT DATA      */
PHI1 = { 0.71 -0.19,
        0.05 0.59 };
THETA1 = { -0.60 0.02,
           0.11 -0.31 };
SIGMA = { 1.0 -0.2,
          -0.2 1.04 };
PRINT "(A) PRINT THE TRUE COEF MATRICES ";
PRINT , PHI1 " " THETA1 ,
      SIGMA;
RUN ACVMF(PHI1, THETA1, SIGMA,
          GAMMA0, GAMMA1, GAMMA2);
/*      PRINT AUTOCOVARIANCE FUNCTION OF Y      */
PRINT , "(B) ACVMF FROM COEF MATRICES ";
PRINT , SIGMA, GAMMA0 " " GAMMA1;
/*      */
/*      2. CALCULATE PHI1 & THETA1      */
/*      */

```

```

SAS
8:37 Saturday, November 21, 1992 1

(A) PRINT THE TRUE COEF MATRICES

      PHI1          THETA1
      0.71         -0.19         -0.6         0.02
      0.05          0.59          0.11        -0.51

      SIGMA
      1           -0.2
      -0.2        1.04

(B) ACVMF FROM COEF MATRICES

      SIGMA
      1           -0.2
      -0.2        1.04

      GAMMA0          GAMMA1
      5.33007        -1.050303        4.5879072        -1.325515
      -1.050303      2.3105282        -0.525175         1.6550965

(C) COEF MATRICES FROM ACVMF

      PHI1          THETA1
      0.71         -0.19        -0.596145        0.0441714
      0.05          0.59         0.0747794        -0.313855

(D) AGAIN, ACVMF FROM COEF MATRICES

      SIGMA
      1           -0.2
      -0.2        1.04

      GAMMA0          GAMMA1
      5.3335059       -1.008219        4.58333         -1.318282
      -1.008219      2.3014724        -0.465724         1.6488229

Exiting IML.

```

그림 6.39 VARMA(1, 1) 모형들의 모수행렬들과 자기공분산행렬들

```

PHI1 = GAMMA2*INV(GAMMA1);
A1 = GAMMA1' - GAMMA0*PHI1';
A2 = GAMMA0 - GAMMA1*PHI1'
      - PHI1*GAMMA1' + PHI1*GAMMA0*PHI1';
RUN VMA1COEF(2, A2, A1, THETA1);
PRINT , "(C) COEF MATRICES FROM ACVMF ";
PRINT , PHI1 " " THETA1,;

/*
/*      3. CALCULATE AUTOCOVARANCE MATRICES AGAIN
/*
/*
RUN ACVMF(PHI1, THETA1, SIGMA,
          GAMMA0, GAMMA1, GAMMA2);
/*      PRINT AUTOCOVARANCE FUNCTION OF Y
/*
PRINT , "(D) AGAIN, ACVMF FROM COEF MATRICES ";
PRINT , SIGMA, GAMMA0 " " GAMMA1,;
QUIT;

```

이 SAS프로그램을 수행한 결과가 그림 6.39이다. 그림 6.39의 부분 A에는 원래의 VAR계수행렬, VMA계수행렬, 다변량백색잡음과정의 공분산행렬이 출력되었다. 부분 B에는 이 모수행렬들로부터 구한 자기공분산행렬함수가 수록되었다. 부분 C는 이 자기공분산행렬함수들로부터 구한 VAR계수행렬과 VMA계수행렬이 수록되어 있다. 부분 C의 행렬들이 부분 A의 행렬들과 약간 차이가 있음을 알 수 있다. 이러한 차이가 단지 반올림오차(round-off error)에 의한 것인지를 확인하기 위해서 부분 C의 모수행렬들로부터 자기공분산행렬함수를 구한 것이 부분 D이다. 부분 B의 행렬들과 부분 D의 행렬들 사이에는 큰 차이가 없음을 알 수 있다.

VARMA(p, q)과정의 자기공분산행렬함수로부터 VMA계수행렬들을 구하는 방법에는 여러 가지가 있을 수 있다. 이 방법들은 본 저자의 생각일 뿐 아직 연구결과가 나온 것이 아니다. 따라서 관심있는 독자들을 위해 간단히 기술하고자 한다. 이후 이 소절에서는 자기공분산행렬함수와 VAR계수행렬들을 알고 있다고 가정한다.

VARMA과정의 VMA계수행렬들을 구하기 위해서 ARMA과정의 MA계수들을 구하는 Choi알고리즘([SAS4TSA2], p. 413)을 확장 적용할 수도 있다. 작인적 VARMA(p, q)과정은 다음과 같은 VMA(∞)모형을 만족한다.

$$y_t = \sum_{k=0}^{\infty} \Psi_k u_{t-k} \quad (\Psi_0 = I_d)$$

여기서 $\{\Psi_i\}$ 는 다음 식을 만족한다.

$$\Theta_i = \sum_{j=0}^p \Phi_j \Psi_{i-j} \quad (i = 0, 1, 2, \dots)$$

여기서 $\Psi_j = 0, (j < 0)$ 이다. 따라서 다음 등호가 만족된다.

$$\begin{aligned} [\Theta_0, \Theta_1, \dots, \Theta_q] &= G \\ [\Theta_0, \Theta_1, \dots, \Theta_q] &= [\Phi_0, \Phi_1, \dots, \Phi_p] G \end{aligned}$$

여기서 G 는 부행렬들이 $m \times m$ 행렬인 $(p+1) \times (q+1)$ 블록행렬로서 (i, j) 블록은 다음과 같다.

$$[G]_{ij} = \begin{cases} \Psi_{j-i} & (j \geq i) \\ 0 & (j < i) \end{cases}$$

또한, 확장Yule-Walker방정식들로부터 다음 식을 얻는다.

$$\Theta_i \Sigma \Psi_0' + \Theta_{i+1} \Sigma \Psi_1' + \dots + \Theta_q \Psi_{q-i}' = \sum_{j=0}^p \Phi_j \Gamma(i-j) \quad (i = 0, 1, \dots, q)$$

이 식을 정리하면 다음과 같다.

$$[\Theta_0, \Theta_1, \dots, \Theta_q] F_q = H_q$$

여기서 F_q 는 부행렬들이 $m \times m$ 행렬인 $(q+1) \times (q+1)$ 블록행렬로서 (i, j) 블록은 다음과 같다.

$$[F_q]_{ij} = \begin{cases} \sum \Psi_{i-j}' & (i \geq j) \\ 0 & (i < j) \end{cases}$$

또한, H_q 는 부행렬들이 $m \times m$ 행렬인 $(q+1) \times 1$ 블록행렬로서 $(i, 1)$ 블록은 다음과 같다.

$$[H_q]_{i,1} = \sum_{j=0}^p \Phi_j \Gamma(i-j)$$

이 식들을 정리해서 $\Sigma, \Theta_1, \dots, \Theta_q$ 에 관한 방정식들을 만드는 것보다는 $\Sigma, \Psi_1, \dots, \Psi_q$ 에 관한 방정식들을 만드는 것이 편리하다. 즉, 다음과 같은 식을 얻는다.

$$[\Phi_0, \Phi_1, \dots, \Phi_p] GF_q = H_q$$

다음과 같은 행렬들을 정의하자.

$$A_0 = \Psi_0 \Sigma^{\frac{1}{2}} = -\Sigma^{\frac{1}{2}}$$

$$A_i = \Psi_i \Sigma^{\frac{1}{2}} \quad (i=1, 2, \dots, q)$$

또한, B 와 B_q 는 부행렬들이 $m \times m$ 행렬인 $(p+1) \times (q+1)$ 블록행렬과 $(q+1) \times (q+1)$ 블록행렬로서 (i, j) 블록은 각각 다음과 같다.

$$[B]_{ij} = \begin{cases} A_{j-i} & (j \geq i) \\ 0 & (j < i) \end{cases}$$

$$[B_q]_{ij} = \begin{cases} A_{i-j}' & (i \geq j) \\ 0 & (i < j) \end{cases}$$

이 행렬들을 사용하면 다음 식을 얻는다.

$$[\Phi_0, \Phi_1, \dots, \Phi_p] BB_q = H_q$$

이 식으로부터 A_0, A_1, \dots, A_q 를 얻은 뒤 다음 식들을 이용해서 $\Psi_1, \Psi_2, \dots, \Psi_q, \Sigma$ 를 구한다.

$$\Sigma = A_0 A_0'$$

$$\Psi_i = -A_i A_0^{-1} \quad (i=1, 2, \dots, q)$$

다음 식을 사용해서 $\Psi_1, \Psi_2, \dots, \Psi_q$ 로부터 $\Theta_1, \Theta_2, \dots, \Theta_q$ 를 얻는다.

$$\Theta_i = \sum_{j=0}^p \Phi_j \Psi_{i-j} \quad (i=1, 2, \dots, q)$$

가장 어려운 문제는 A_0, A_1, \dots, A_q 를 구하는 알고리즘을 얻는 것인데 만족할 만한 알고리즘이 아직 제시되지 않았다. 다만 Newton-Raphson식을 바탕으로 한 Choi 알고리즘을 이용하는 것도 한 방법일 것이다.

VARMA(p, q)모형의 VMA계수행렬들을 구하는 방법으로는 VARMA(p, q)과정을 VMA(1)모형으로 나타내거나 VARMA(1, 1)모형으로 나타내 보는 것도 생각할 수는 있

다. 그러나 이러한 방법은 차원이 지나치게 큰 행렬의 역행렬을 구해야 하며, 제약조건하에서 행렬방정식을 풀어야 하는 문제점들이 있다.

▣ 참고문헌 ▣

- Anderson, B. D. O. & J. B. Moore(1979), *Optimal Filtering*, Prentice-Hall, Englewood Cliffs.
- Barron, P. (1987), A method for generating independent realization of a multivariate normal stationary and invertible ARMA(p, q) process, *J. Time Series Anal.*, vol. 8, pp. 125~130.
- Choi, B. S. (1990), On the block LU decomposition of block Toeplitz matrices using the vector Yule-Walker equations, *Commun. Statist. - Theory Meth.*, vol. 19, pp. 2815~2817.
- Gantmacher, F. R. (1959), *The Theory of Matrices*, vols. 1~2, Chelsea, New York.
- Hannan, E. J. (1969), The identification of vector mixed autoregressive-moving average systems, *Biometrika*, vol. 56, pp. 223~225.
- Hannan, E. J. (1976), The identification and parameterization of ARMAX and State Space Forms, *Econometrica*, vol. 44, pp. 713~723.
- Hannan, E. J. & M. Deistler(1988), *The Statistical Theory of Linear Systems*, Wiley, New York.
- Lütkepohl, H. (1991), *Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.
- Whittle, P. (1963), On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix, *Biometrika*, vol. 50, pp. 129~134.
- Wilkinson, J. H. (1965), *The Algebraic Eigen Value Problem*, Oxford Univ. Press, London.
- Wilson, G. (1969), Factorization of the covariance generating function of a pure moving average process, *SIAM J. Numer. Anal.*, vol. 6, pp. 1~7.

VAR모형의 추정과 식별

- 7.1 자기공분산행렬과 자기상관행렬의 추정
- 7.2 표본자기공분산행렬과 표본자기상관행렬의 성질
- 7.3 VAR계수행렬의 추정
- 7.4 VAR계수추정행렬들의 성질
- 7.5 VAR모형의 식별
- 7.6 VAR모형의 검진

제 7 장 VAR모형의 추정과 식별

7.1 자기공분산행렬과 자기상관행렬의 추정

정상성을 갖는 d 변량확률과정에서 T 개의 다변량시계열데이터 y_1, y_2, \dots, y_T 가 관찰되었다고 하면, 평균벡터 μ 와 자기공분산행렬 $\Gamma(h)$ 를 각각 다음과 같이 추정한다.

$$\hat{\mu} := \bar{y} = \frac{1}{T} \sum_{t=1}^T y_t$$

$$\hat{\Gamma}(h) := \begin{cases} \frac{1}{T} \sum_{t=1}^{T-h} (y_{t+h} - \bar{y})(y_t - \bar{y})' & (h = 0, 1, \dots, T-1) \\ \frac{1}{T} \sum_{t=1-h}^T (y_{t+h} - \bar{y})(y_t - \bar{y})' & (h = 1-T, 2-T, \dots, 1) \end{cases}$$

이 $\hat{\mu}$ 와 $\hat{\Gamma}(h)$ 를 각각 표본평균벡터와 시차 h 에서의 표본자기공분산행렬이라고 부른다. 표본자기공분산행렬 $\hat{\Gamma}(h)$ 의 (i, j) 원소를 $\hat{\gamma}_{ij}(h)$ 라 표기하고 $\gamma_{ij}(h)$ 의 추정량으로 간주한다.

SAS/IML의 내장함수 COVLAG를 사용하면 표본자기공분산행렬들을 구할 수 있다. 한 예로 다변량시계열데이터 VAR31A의 세 변수들 Y1, Y2, Y3의 표본자기공분산행렬들을 구하기 위해 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM7-1.PRG에 수록되어 있다).

```

/* */
/*          EXAM7-1.PRG          */
/* */
/*          COMPUTING SAMPLE ACVMF          */
/*          DATASET = VAR31A          */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=60;
PROC IML;

```

```

RESET NOLOG NOCENTER;
/* CALL TSA4DATA.VAR31A */
USE TSA4DATA.VAR31A;
READ ALL VAR{Y1 Y2 Y3} INTO Y;
CLOSE TSA4DATA.VAR31A;
/* MAKE ACVMF-MATRIX */
ACVMF = COVLAG(Y, 6);
/* MAKE A FILE TO KEEP ACVMF */
CREATE TSA4DATA.ACVMF31A FROM ACVMF;
APPEND FROM ACVMF;
CLOSE TSA4DATA.ACVMF31A;
/* MAKE ACVMF */
GAMMA0 = ACVMF[ , {1 2 3}];
GAMMA1 = ACVMF[ , {4 5 6}];
GAMMA2 = ACVMF[ , {7 8 9}];
GAMMA3 = ACVMF[ , {10 11 12}];
GAMMA4 = ACVMF[ , {13 14 15}];
GAMMA5 = ACVMF[ , {16 17 18}];
/* PRINT ACVMF */
PRINT , GAMMA0 " " GAMMA1,
      GAMMA2 " " GAMMA3,
      GAMMA4 " " GAMMA5;
QUIT;

```

SAS/IML의 함수 COVLAG를 사용해서 표본자기공분산행렬들을 구하기 위해서는 우선 주어진 다변량시계열데이터를 하나의 행렬로 만들어야 한다. 영구SAS데이터세트 TSA4DATA.VAR31A에는 3개의 변수들 각각에 대해 150개의 관찰값들이 수록되어 있다. 즉, 변수들이 3개이고 관찰점들이 150개이다. 이 관찰값들을 하나의 150×3 행렬로 읽어들이어야 SAS/IML의 함수 COVLAG를 적용할 수 있다. 이러한 과정을 수행하기 위해서 USE스테이트먼트, READ-INTO스테이트먼트, CLOSE스테이트먼트를 사용한다. 즉, 기존의 SAS 데이터세트를 SAS/IML에서 행렬형태의 관찰값들로 읽어들이기 위해서는 이 세 스테이트먼트들을 사용한다. 이 스테이트먼트들을 설명하면 다음과 같다.

첫째, USE스테이트먼트는 읽어들이는 SAS데이터세트를 여는(open)기능을 하는데 사용양식은 다음과 같다.

USE SAS데이터세트 ;

이 SAS프로그램에서는 영구SAS데이터세트 TSA4DATA.VAR31A에서 관찰값들을 읽어들이는 것을 지시한다. USE스테이트먼트에 대한 자세한 내용은 제 3.7.2 소절과 [IML6.03] (p.

315)을 참조하라.

둘째, READ-INTO스테이트먼트는 USE스테이트먼트에 지정한 SAS데이터세트의 관찰값들을 하나의 행렬로 읽어들인다. READ-INTO스테이트먼트의 사용양식은 다음과 같다.

```
READ 범위 VAR {변수들} INTO 행렬 ;
```

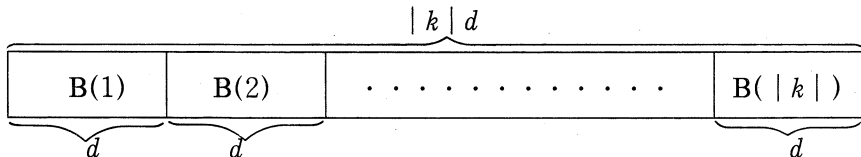
여기서 범위에 사용가능한 SAS키워드로는 ALL, CURRENT, NEXT, AFTER, POINT 등이 있다. 이 중에서 SAS키워드 ALL은 모든 관찰점들의 관찰값들을 읽어들이도록 한다. 이 범위에 대한 자세한 내용은 제3.7.5소절과 [IML6.03] (pp. 91~92)을 참조하라. SAS는 키워드 VAR는 읽어들이 변수들을 지정할 것임을 나타내며, 이 SAS키워드 VAR 다음에 읽어들이 변수들을 중괄호 안에 나열한다. SAS키워드 INTO는 읽어들이 관찰값들을 하나의 행렬로 만들 것임을 나타내며, 이 SAS키워드 INTO다음에 이 행렬의 이름을 지정한다. 이 SAS프로그램에서는 변수들 Y1, Y2, Y3의 모든 관찰점들을 읽어 행렬 Y를 만든다.

셋째, CLOSE스테이트먼트는 USE스테이트먼트에 의해서 열린 SAS데이터세트를 닫는다(close). SAS/IML의 세션(session)이 끝나면 모든 열린 파일들이 자동적으로 닫히므로 일반적으로는 CLOSE스테이트먼트를 사용하지 않아도 큰 문제가 발생하지는 않는다. 그러나 열린 파일들의 개수가 많거나 열린 파일이 큰 경우에는 필요없는 파일을 닫아 놓아야 할 것이다. SAS/IML에서 SAS데이터세트들의 입출력에 관한 자세한 내용은 제3.7절과 [IML6.03] (pp. 89~99)을 참조하라.

SAS/IML에서 표본자기공분산행렬들을 구하기 위해서는 이렇게 만들어진 행렬 Y에 함수 COVLAG를 적용한다. 함수 COVLAG의 사용양식은 다음과 같다.

```
COVLAG (A, k)
```

함수 COVLAG는 표본자기공분산행렬들(a sequence of lagged crossproduct matrices)을 계산하는데 사용된다. 입력행렬 A는 $T \times d$ 행렬이다. 여기서 T는 관찰점들의 개수이고 d는 확률벡터의 차원, 즉 A를 구성하는 확률변수들의 개수이다. 이 SAS프로그램에서는 $A=Y$, $T=150$, $d=3$ 이다. 입력상수 k의 절대값은 구해야 할 표본자기공분산행렬들의 개수이다. 만일 k가 양수이면 평균벡터를 뺀 표본자기공분산행렬들을 계산하고, k가 음수이면 평균벡터를 빼지 않은 표본자기공분산행렬들을 계산한다. 이 함수 COVLAG를 수행하면 다음과 같은 $d \times (|k| d)$ 행렬을 출력한다.



다변량시계열데이터 y_1, y_2, \dots, y_T 가 주어졌을 때 $B(1), B(2), \dots, B(|k|)$ 는 다음과 같다.

① $k < 0$ 인 경우

$$B(i) = \frac{1}{T} \sum_{t=1}^{T-i+1} y_{t+i-1} y_t' \quad (i = 1, 2, \dots, -k)$$

② $k > 0$ 인 경우

$$B(i) = \frac{1}{T} \sum_{t=1}^{T-i+1} (y_{t+i-1} - \bar{y})(y_t - \bar{y})' \quad (i = 1, 2, \dots, k)$$

즉, k 가 양수인 경우에는 제 i 번째 블록 $B(i)$ 는 $\hat{\Gamma}(i-1)$ 이다. 따라서 이 SAS프로그램의 ACVMF는 3×18 행렬로서 $\hat{\Gamma}(0), \hat{\Gamma}(1), \dots, \hat{\Gamma}(5)$ 를 옆으로 배열한 형태이다. 함수 COVLAG에 관한 자세한 내용은 [IML6.03] (p. 232)을 참조하라.

SAS/IML에서 생성된 행렬을 SAS데이터세트에 저장하기 위해서는 CREATE-FROM 스테이트먼트, APPEND 스테이트먼트, CLOSE 스테이트먼트를 사용한다. 이 SAS프로그램에서는 이 세 스테이트먼트들을 사용해서 행렬 ACVMF를 새로운 영구SAS데이터세트 TSA4DATA.ACVMF31A에 저장한다. 이 스테이트먼트들을 설명하면 다음과 같다.

첫째, CREATE 스테이트먼트는 새로이 만들고자 하는 SAS데이터세트를 여는(open) 역할을 한다. 앞에서 설명한 USE 스테이트먼트는 SAS/IML로 읽어들이고자 하는 관찰값들을 가지고 있는 기존의 SAS데이터세트를 여는 반면에, CREATE 스테이트먼트는 SAS/IML이 가지고 있는 관찰값들을 출력시킬 새로운 SAS데이터세트를 연다. 즉, USE 스테이트먼트를 사용해서 연 SAS데이터세트에서는 관찰값들을 읽을 수만 있고 쓸 수가 없으며, CREATE 스테이트먼트를 사용해서 연 SAS데이터세트에서는 쓸 수는 있으나 읽을 수가 없다. CREATE 스테이트먼트의 하나인 CREATE-FROM 스테이트먼트는 SAS/IML에서 만들어진 행렬을 새로운 SAS데이터세트에 출력하기 위해서 이 새로운 SAS데이터세트를 여는 데 사용한다. 이 CREATE-FROM 스테이트먼트의 사용양식은 다음과 같다.

CREATE SAS데이터세트 FROM 행렬 ;

이 CREATE-FROM 스테이트먼트는 SAS키워드 FROM 다음에 지정한 행렬을 SAS키워드 CREATE 다음에 지정한 SAS데이터세트에 쓰겠다는 선언을 하는 것이다. 이 선언의 결과로 지정된 SAS데이터세트가 열릴 것이다. 이 SAS프로그램에서는 영구SAS데이터세트 TSA4DATA.ACVMF31A에 행렬 ACVMF를 쓰기 위해서 이 영구SAS데이터세트를 연다. CREATE 스테이트먼트에 대한 자세한 내용은 제 3.7.2 소절과 [IML6.03] (pp. 232~233)을 참조하라.

둘째, APPEND 스테이트먼트는 CREATE 스테이트먼트를 사용해서 연 SAS데이터세트에 관찰값들을 쓰는 역할을 한다. SAS/IML에서 만들어진 행렬을 CREATE 스테이트먼트를 사용해서 연 SAS데이터세트에 기록하기 위해서는 다음과 같은 양식의 APPEND 스테이트먼트를 사용한다.

APPEND FROM 행렬 ;

이 SAS프로그램에서는 행렬 ACVMF를 영구SAS데이터세트 TSA4DATA.ACVMF31A에 쓴다. APPEND스테이트먼트에 대한 자세한 내용은 제3.7.3소절과 [IML6.03] (p. 223)을 참조하라.

셋째, CREATE스테이트먼트에 의해 열린 SAS데이터세트를 닫는 데도 CLOSE스테이트먼트를 사용한다. 앞서도 언급했듯이 SAS/IML에서의 SAS데이터세트들의 입출력에 대한 자세한 내용은 제3.7절과 [IML6.03] (pp. 89~99)을 참조하라.

이 SAS프로그램을 수행한 결과가 그림 7.1이다.

<p>SAS</p> <p style="text-align: center;">GAMMA0</p> <table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 33%;">1.8961026</td><td style="width: 33%;">-0.207465</td><td style="width: 33%;">-0.002001</td></tr> <tr><td>-0.207465</td><td>1.9251076</td><td>0.4262897</td></tr> <tr><td>-0.002001</td><td>0.4262897</td><td>1.3615659</td></tr> </table> <p style="text-align: center;">GAMMA2</p> <table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 33%;">0.7782979</td><td style="width: 33%;">0.1394627</td><td style="width: 33%;">0.0376015</td></tr> <tr><td>-0.093658</td><td>0.5317743</td><td>0.1417052</td></tr> <tr><td>0.2967557</td><td>0.0305294</td><td>0.4682391</td></tr> </table> <p style="text-align: center;">GAMMA4</p> <table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 33%;">0.5148447</td><td style="width: 33%;">0.0821527</td><td style="width: 33%;">0.0632458</td></tr> <tr><td>-0.09495</td><td>0.2368816</td><td>0.0402712</td></tr> <tr><td>0.2658899</td><td>0.0741352</td><td>0.1294852</td></tr> </table> <p>Exiting IML.</p>	1.8961026	-0.207465	-0.002001	-0.207465	1.9251076	0.4262897	-0.002001	0.4262897	1.3615659	0.7782979	0.1394627	0.0376015	-0.093658	0.5317743	0.1417052	0.2967557	0.0305294	0.4682391	0.5148447	0.0821527	0.0632458	-0.09495	0.2368816	0.0402712	0.2658899	0.0741352	0.1294852	<p style="text-align: right;">21:57 Saturday, November 21, 1992 1</p> <p style="text-align: center;">GAMMA1</p> <table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 33%;">-1.121361</td><td style="width: 33%;">0.0094321</td><td style="width: 33%;">-0.089396</td></tr> <tr><td>-0.198754</td><td>0.9131947</td><td>0.9059942</td></tr> <tr><td>-0.213856</td><td>0.6478581</td><td>-0.285762</td></tr> </table> <p style="text-align: center;">GAMMA3</p> <table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 33%;">-0.741671</td><td style="width: 33%;">0.1545137</td><td style="width: 33%;">-0.011048</td></tr> <tr><td>0.1599501</td><td>0.2366182</td><td>0.270772</td></tr> <tr><td>-0.332816</td><td>0.2136621</td><td>-0.110952</td></tr> </table> <p style="text-align: center;">GAMMA5</p> <table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 33%;">-0.510843</td><td style="width: 33%;">0.0046542</td><td style="width: 33%;">0.0066177</td></tr> <tr><td>-0.000626</td><td>0.166183</td><td>0.1845961</td></tr> <tr><td>-0.29206</td><td>0.2866318</td><td>0.0995832</td></tr> </table>	-1.121361	0.0094321	-0.089396	-0.198754	0.9131947	0.9059942	-0.213856	0.6478581	-0.285762	-0.741671	0.1545137	-0.011048	0.1599501	0.2366182	0.270772	-0.332816	0.2136621	-0.110952	-0.510843	0.0046542	0.0066177	-0.000626	0.166183	0.1845961	-0.29206	0.2866318	0.0995832
1.8961026	-0.207465	-0.002001																																																					
-0.207465	1.9251076	0.4262897																																																					
-0.002001	0.4262897	1.3615659																																																					
0.7782979	0.1394627	0.0376015																																																					
-0.093658	0.5317743	0.1417052																																																					
0.2967557	0.0305294	0.4682391																																																					
0.5148447	0.0821527	0.0632458																																																					
-0.09495	0.2368816	0.0402712																																																					
0.2658899	0.0741352	0.1294852																																																					
-1.121361	0.0094321	-0.089396																																																					
-0.198754	0.9131947	0.9059942																																																					
-0.213856	0.6478581	-0.285762																																																					
-0.741671	0.1545137	-0.011048																																																					
0.1599501	0.2366182	0.270772																																																					
-0.332816	0.2136621	-0.110952																																																					
-0.510843	0.0046542	0.0066177																																																					
-0.000626	0.166183	0.1845961																																																					
-0.29206	0.2866318	0.0995832																																																					

그림 7.1 함수 COVLAG를 사용해서 구한 표본자기공분산행렬들

교차상관계수 $\rho_{ij}(h)$ 와 자기상관행렬 $P(h)$ 를 다음과 같이 추정하고 $\hat{\rho}_{ij}(h)$ 와 $\hat{P}(h)$ 를 각각 표본교차상관계수와 표본자기상관행렬이라 부른다.

$$\hat{\rho}_{ij}(h) := \frac{\hat{\gamma}_{ij}(h)}{\sqrt{\hat{\gamma}_{ii}(0) \hat{\gamma}_{jj}(0)}}$$

$$\hat{P}(h) := [\hat{\rho}_{ij}(h)]$$

SAS/IML을 사용해서 표본자기상관행렬을 구하는 예를 살펴보기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM7-2.PRG에 수록되어 있다).

```

/*                               */
/*                               EXAM7-2.PRG                               */
/*                               */
/*   COMPUTING THE SAMPLE ACVMF   */
/*   DATASET = VAR31A              */

```

```

LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOLOG NOCENTER;
/*      CALL TSA4DATA.ACVMF31A      */
USE TSA4DATA.ACVMF31A;
READ ALL INTO ACVMF;;
CLOSE TSA4DATA.ACVMF31A;
/*      CALCULATE ACRMF      */
RHO0 = ACVMF[, {1 2 3}];
Q = INV(SQRT(DIAG(RHO0)));
RHO0 = Q*RHO0*Q;
RHO1 = Q*ACVMF[, {4 5 6}]*Q;
RHO2 = Q*ACVMF[, {7 8 9}]*Q;
RHO3 = Q*ACVMF[, {10 11 12}]*Q;
RHO4 = Q*ACVMF[, {13 14 15}]*Q;
RHO5 = Q*ACVMF[, {16 17 18}]*Q;
/*      PRINT ACRMF      */
PRINT , RHO0 " " RHO1,
RHO2 " " RHO3,
RHO4 " " RHO5,;
QUIT;

```

이 SAS프로그램의 READ스테이트먼트에서는 SAS키워드 VAR가 사용되지 않았고 변수들도 지정되지 않았다. 이는 모든 변수들의 관찰값들을 읽어들이기 위한 것이다. 이 SAS 프로그램을 수행한 결과가 그림 7.2이다.

SAS			21:37 Saturday, November 21, 1992 2		
RHO0			RHO1		
1	-0.108589	-0.001245	-0.591403	0.0049368	-0.055638
-0.108589	1	0.2633043	-0.10403	0.4743603	0.5596011
-0.001245	0.2633043	1	-0.133098	0.4001594	-0.209877
RHO2			RHO3		
0.4104725	0.0729961	0.0234021	-0.391156	0.0808739	-0.006876
-0.049021	0.2762309	0.0875252	0.0837194	0.1229117	0.1672464
0.1846923	0.018857	0.3438975	-0.207135	0.1319716	-0.081489
RHO4			RHO5		
0.2715279	0.0429995	0.0393624	-0.269418	0.0024361	0.0041187
-0.049698	0.1230485	0.0248741	-0.000327	0.086324	0.1140186
0.1654823	0.0457908	0.0951002	-0.18177	0.1770425	0.0731387
Exiting IML.					

그림 7.2 SAS/IML을 사용해서 구한 표본자기상관행렬들

SAS/ETS의 PROC STATESPACE를 사용해서 표본자기상관행렬을 구할 수 있다.

다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM7-3.PRG에 수록되어 있다).

```

/* */
/*          EXAM7-3.PRG          */
/* */
/*          ESTIMATE ACVMF      */
/*          DATASET = VAR31A     */
LIBNAME TSA4DATA 'A:\TSA4DATA';
DATA EXAM73;
    SET TSA4DATA.VAR31A;
    RUN;
OPTIONS PAGESIZE=60;
PROC STATESPACE DATA=EXAM73 PRINTOUT=LONG;
    VAR Y1 Y2 Y3;
    RUN;

```

이 SAS프로그램의 PROC STATESPACE의 옵션 PRINTOUT=에 LONG을 지정한 것은 표본자기공분산행렬들을 출력시키기 위한 것이다. 이 옵션이 사용되지 않으면 표본자기상관행렬들은 출력되나 표본자기공분산행렬들은 출력되지 않는다. 이에 대한 자세한 내용은 제8.4.1소절을 참조하라. 이 SAS프로그램을 수행한 결과의 일부가 그림 7.3이다.

그림 7.3의 부분 A에는 요약통계량들과 AIC값들이 수록되어 있다. 부분 B에는 표본자기공분산행렬들이 수록되어 있고, 부분 C에는 표본자기상관행렬들이 수록되어 있다. 여기서 유의할 점은 그림 7.3의 부분 C의 표본자기상관행렬들은 그림 7.2의 표본자기공분산행렬들과 일치하나, 그림 7.3의 부분 B의 표본자기공분산행렬들은 그림 7.1의 표본자기공분산행렬들과 일치하지 않는다는 점이다. 그러나 [ETS6.1](p. 470)에 의하면 PROC STATESPACE에서도 $\gamma(h)$ 는 다음 통계량으로 추정된다.

$$C_h = \frac{1}{T} \sum_{t=1}^{T-h} (y_{t+h} - \bar{y})(y_t - \bar{y})' \quad (h = 0, 1, \dots)$$

즉, $B(h+1) = C_h$ 이다. 따라서 SAS/IML의 함수 COVLAG 또는 PROC STATESPACE에 문제가 있는 것 같다. 이를 조사하기 위해서 $B^{-1}(h+1)C_h$ 를 h 가 0, 1, 2인 경우에 대해서 계산해 보기로 하고 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM7-4.PRG에 수록되어 있다).

```

/* */
/*          EXAM7-4.PRG          */
/* */
/*          COMPARING ACVMFS     */
/*          DATASET = ACVMF31A   */

```

SAS

1:23 Monday, November 23, 1992 1

STATESPACE Procedure

(A)

Nobs = 150

Variable	Mean	Std
Y1	0.0414	1.381603
Y2	0.13838	1.392131
Y3	-0.01438	1.170771

Information Criterion for Autoregressive Models

Lag=0	Lag=1	Lag=2	Lag=3	Lag=4
230.844003	38.387028	41.304655	48.610736	63.324667
Lag=5	Lag=6	Lag=7	Lag=8	Lag=9
69.981677	82.287743	94.248807	104.206349	115.278133
Lag=10				
118.112739				

(B)

Var/Cov Matrices of the Input Series

	Lag=0		
	Y1	Y2	Y3
Y1	1.908828	-0.208857	-0.002014
Y2	-0.208857	1.938028	0.429151
Y3	-0.002014	0.429151	1.370704
	Lag=1		
	Y1	Y2	Y3
Y1	-1.128887	0.009495	-0.089996
Y2	-0.200088	0.919324	0.912075
Y3	-0.215291	0.652206	-0.287680
	Lag=2		
	Y1	Y2	Y3
Y1	0.783521	0.140399	0.037854
Y2	-0.094287	0.535343	0.142654
Y3	0.298747	0.030734	0.471382
	Lag=3		
	Y1	Y2	Y3
Y1	-0.746649	0.155551	-0.011122
Y2	0.161024	0.238206	0.272589
Y3	-0.335049	0.215096	-0.111697
	Lag=4		
	Y1	Y2	Y3
Y1	0.518300	0.082704	0.063670
Y2	-0.095587	0.238471	0.040542
Y3	0.267674	0.074633	0.130354
	Lag=5		
	Y1	Y2	Y3
Y1	-0.514272	0.004685	0.006662
Y2	-0.000630	0.167298	0.185835
Y3	-0.294020	0.288555	0.100252

Lag=6			
	Y1	Y2	Y3
Y1	0.411958	0.050712	-0.034000
Y2	-0.142726	0.256957	0.237103
Y3	0.088304	0.124834	0.162207

Lag=7			
	Y1	Y2	Y3
Y1	-0.337451	-0.060081	0.144403
Y2	0.072264	0.185161	0.035326
Y3	-0.040686	0.049545	0.016743

Lag=8			
	Y1	Y2	Y3
Y1	0.254340	0.011567	0.055294
Y2	0.047073	-0.007243	-0.000626
Y3	0.065240	-0.002936	-0.005594

Lag=9			
	Y1	Y2	Y3
Y1	-0.102022	0.126139	0.034131
Y2	-0.066114	-0.039066	-0.031791
Y3	0.075321	-0.092326	0.032088

Lag=10			
	Y1	Y2	Y3
Y1	0.087031	0.111465	-0.049911
Y2	0.043572	-0.205981	0.046128
Y3	-0.147359	-0.188564	0.001469

(C)

Correlation Matrices of the Input Series

Lag=0				Lag=1		
	Y1	Y2	Y3	Y1	Y2	Y3
Y1	1.00000	-0.10859	-0.00125	-0.59140	0.00494	-0.05564
Y2	-0.10859	1.00000	0.26330	-0.10403	0.47436	0.55960
Y3	-0.00125	0.26330	1.00000	-0.13310	0.40016	-0.20988

Lag=2			Lag=3			
	Y1	Y2	Y3	Y1	Y2	Y3
Y1	0.41047	0.07300	0.02340	-0.39116	0.08087	-0.00688
Y2	-0.04902	0.27623	0.08753	0.08372	0.12291	0.16725
Y3	0.18469	0.01886	0.34390	-0.20713	0.13197	-0.08149

Lag=4			Lag=5			
	Y1	Y2	Y3	Y1	Y2	Y3
Y1	0.27153	0.04300	0.03936	-0.26942	0.00244	0.00412
Y2	-0.04970	0.12305	0.02487	-0.00033	0.08632	0.11402
Y3	0.16548	0.04579	0.09510	-0.18177	0.17704	0.07314

Lag=6			Lag=7			
	Y1	Y2	Y3	Y1	Y2	Y3
Y1	0.21582	0.02637	-0.02102	-0.17678	-0.03124	0.08927
Y2	-0.07421	0.13259	0.14547	0.03757	0.09554	0.02167
Y3	0.05459	0.07659	0.11834	-0.02515	0.03040	0.01221

Lag=8			Lag=9			
	Y1	Y2	Y3	Y1	Y2	Y3
Y1	0.13324	0.00601	0.03418	-0.05345	0.06558	0.02110
Y2	0.02447	-0.00374	-0.00038	-0.03437	-0.02016	-0.01951
Y3	0.04033	-0.00180	-0.00408	0.04657	-0.05665	0.02341

Lag=10			
	Y1	Y2	Y3
Y1	0.04559	0.05795	-0.03086
Y2	0.02265	-0.10628	0.02830
Y3	-0.09110	-0.11569	0.00107

그림 7.3 PROC STATESPACE를 사용해서 구한 표본자기공분산행렬들

```

LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=60;
PROC IML;
RESET NOLOG NOCENTER;
/*      CALL TSA4DATA.ACVMF31A      */
USE TSA4DATA.ACVMF31A;
READ ALL INTO ACVMF;;
CLOSE TSA4DATA.ACVMF31A;
/*      ACVMF USING COVLAGE FUNCTION      */
B1 = ACVMF[ , {1 2 3}];
B2 = ACVMF[ , {4 5 6}];
B3 = ACVMF[ , {7 8 9}];
/*      ACVMF USING PROC STAESPACE      */
C0 = { 1.908828 -0.208857 -0.002014,
      -0.208857 1.938028 0.429151,
      -0.002014 0.429151 1.370704 };
C1 = { -1.128887 0.009495 -0.089996,
      -0.200088 0.919324 0.912075,
      -0.215291 0.652206 -0.287680 };
C2 = { 0.783521 0.140399 0.037854,
      -0.094287 0.535343 0.142654,
      0.298747 0.030734 0.471382 };
/*      CALCULATE TEST MATRICES      */
TEST0 = INV(B1)*C0;
TEST1 = INV(B2)*C1;
TEST2 = INV(B3)*C2;
/*      PRINT TEST MATRICES      */
PRINT , TEST0, TEST1, TEST2, ;
QUIT;

```

이 SAS프로그램을 수행한 결과의 일부가 그림 7.4이다. 이 그림에서 볼 수 있듯이 다음 식들이 성립한다.

$$\begin{aligned}
 C_h &= 1.00671 B(h+1) \\
 &= \frac{150}{149} B(h+1) \quad (h=0, 1, 2)
 \end{aligned}$$

이 예제에서 관찰점들의 개수 T 가 150이므로 PROC STATESPACE에서 표본자기공분산함수들을 계산할 때 다음 통계량을 사용한 것이 아닌가 의심된다. 관심있는 독자들은 좀 더 여러 경우에 대해서 살펴보도록 하라.

SAS		11:49 Tuesday, February 9, 1993 3	
TEST0			
1.0067114	1.6936E-7	-4.768E-9	
1.7173E-7	1.0067115	1.531E-7	
-8.383E-8	1.9884E-7	1.0067114	
TEST1			
1.0067118	2.8014E-7	-7.18E-8	
8.3626E-8	1.0067115	-3.068E-7	
1.6393E-7	4.9072E-7	1.006712	
TEST2			
1.006711	4.2287E-7	3.108E-7	
-8.661E-7	1.0067113	-5.756E-7	
-3.797E-7	-9.031E-7	1.006712	
Exiting IML.			

그림 7.4 표본자기공분산함수들을 비교하기 위한 행렬들

$$C_h^* = \frac{1}{T-1} \sum_{t=1}^{T-h} (y_{t+h} - \bar{y})(y_t - \bar{y})' \quad (h = 0, 1, \dots)$$

그림 7.4에서 유추한 결론을 확인하기 위해서 다변량시계열데이터 VAR31A를 구성하는 단변량시계열데이터들의 표본자기공분산함수들을 구해보기로 하고 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM7-5.PRG에 수록되어 있다).

```

/* */
/*          EXAM7-5.PRG          */
/* */
/*          CALCULATING ACRMF'S  */
/*          DATASET = VAR31A     */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=60;
DATA EXAM75;
  SET TSA4DATA.VAR31A;
RUN;
PROC ARIMA DATA=EXAM75;
  IDENTIFY VAR=Y1 NLAG=5;
  IDENTIFY VAR=Y2 NLAG=5;
  IDENTIFY VAR=Y3 NLAG=5;
RUN;

```

이 SAS프로그램을 수행한 결과의 일부가 그림 7.5이다. 그림 7.5의 표본자기공분산함수들은 함수 COVLAG를 사용해서 얻은 그림 7.1의 표본자기공분산함수들과 일치함을 확인할 수 있다. 독자들은 함수 COVLAG를 사용해서 얻은 표본자기상관행렬들과 PROC STATESPACE에서 얻은 표본자기상관행렬들이 같음을 기억하라.

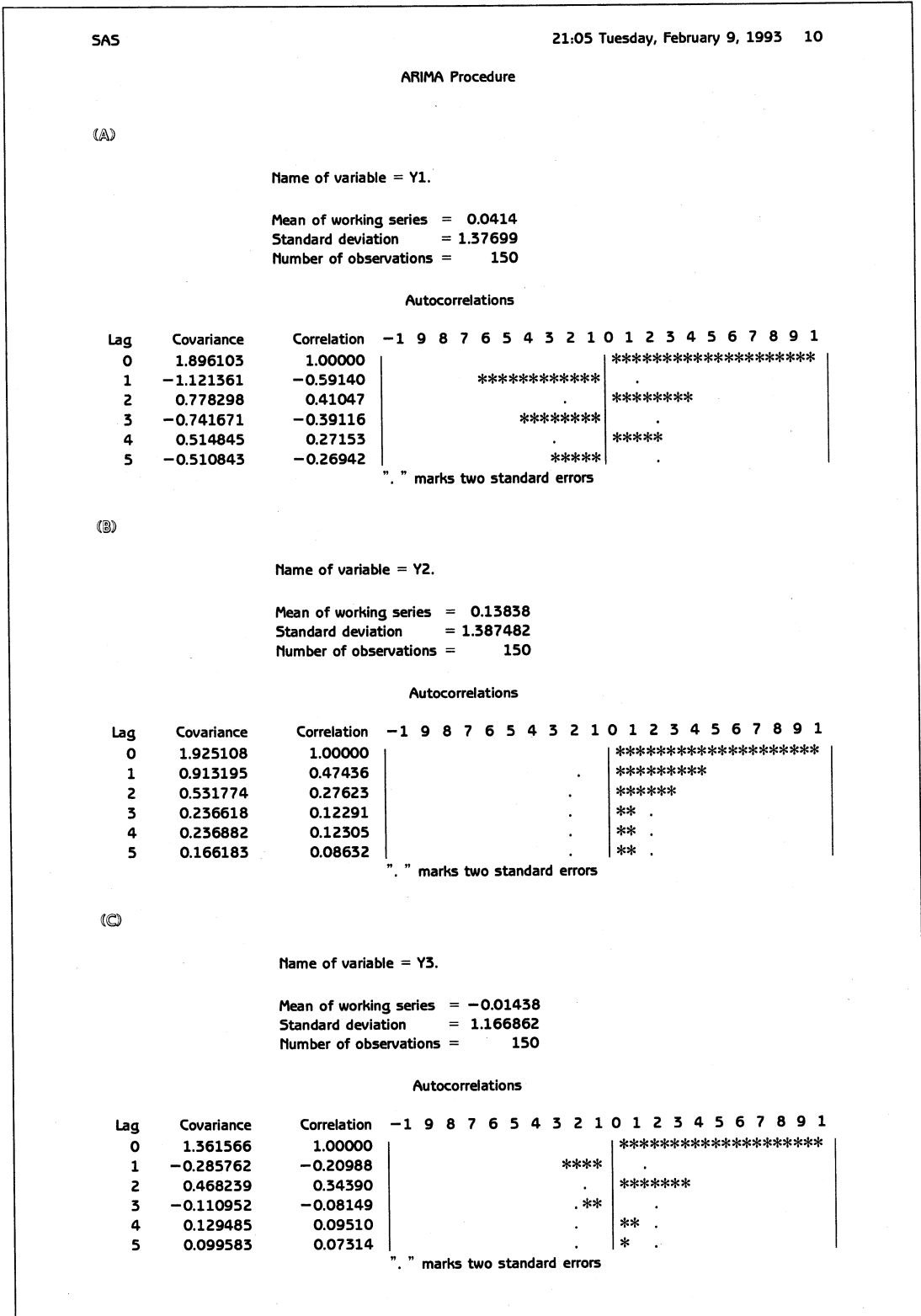


그림 7.5 PROC ARIMA를 사용해서 얻은 표본자기공분산함수들

7.2 표본자기공분산행렬과 표본자기상관행렬의 성질

이 절에서는 표본평균벡터, 표본자기공분산행렬들, 표본자기상관행렬들의 일치성(consistency)과 점근적 정규성(asymptotic normality)에 대해서 알아보자. 이러한 성질들을 유도하는 것은 매우 번잡스러운 일이다. 따라서 본서에서는 결과만을 기술하기로 한다. 이 점근성질들은 다변량시계열데이터가 정상성을 갖는다는 가정하에서 성립하는 것을 유념하라.

우선 표본평균벡터의 일치성을 살펴보자.

정리 7.1 표본평균벡터는 다음 성질들을 만족한다.

① 만일 $\lim_{T \rightarrow \infty} \gamma_{ii}(T) = 0, (i=1, 2, \dots, d)$ 이면 다음 식이 성립한다.

$$\lim_{T \rightarrow \infty} E\{(\bar{y} - \mu)'(\bar{y} - \mu)\} = 0$$

② 만일 $\sum_{h=-\infty}^{\infty} |\gamma_{ii}(h)| < \infty, (i=1, 2, \dots, d)$ 이면 다음 식이 성립한다.

$$\lim_{T \rightarrow \infty} E\{(\bar{y} - \mu)'(\bar{y} - \mu)\} = \sum_{i=1}^d \sum_{h=-\infty}^{\infty} \gamma_{ii}(h) \quad \square$$

다음 정리는 표본평균벡터의 점근적 정규성에 관한 것이다.

정리 7.2 다변량시계열데이터 $\{y_t\}$ 를 다음과 같이 쓸 수 있다고 가정하자.

$$y_t = \mu + \sum_{k=-\infty}^{\infty} C(k) v_{t-k}$$

여기서 $\{v_t | t=0, \pm 1, \pm 2, \dots\}$ 는 서로 독립이고 같은 결합확률분포를 갖는 확률벡터들로서 $E(v_t) = 0$ 이고 $E(v_t v_t') = \Sigma$ 이며, $\{C(k) | k=0, \pm 1, \pm 2, \dots\}$ 는 $m \times m$ 행렬들로서 다음 조건을 만족한다고 하자.

$$\sum_{k=-\infty}^{\infty} |c_{ij}(k)| < \infty \quad (i, j = 1, 2, \dots, d)$$

여기서 $c_{ij}(k)$ 는 행렬 $C(k)$ 의 (i, j) 원소이다. 이러한 조건하에서 $\sqrt{T}(\bar{y} - \mu)$ 는 점근적으로 다변량정규분포를 갖는다. 이 점근분포의 평균벡터는 영벡터이고 공분산행렬은 다음과 같다.

$$\left\{ \sum_{k=-\infty}^{\infty} C(k) \right\} \Sigma \left\{ \sum_{l=-\infty}^{\infty} C(l) \right\} \quad \square$$

지금부터는 $d=2$ 인 경우, 즉 2변량시계열데이터에 대해서 다루기로 한다. 다음 정리는 2변량시계열데이터의 표본교차공분산과 표본교차상관계수의 일치성에 관한 것이다.

정리 7.3 2변량시계열데이터 $\{v_t\}$ 를 다음과 같이 쓸 수 있다고 하자.

$$y_t = \sum_{k=-\infty}^{\infty} C(k)v_{t-k}$$

여기서 $\{v_t\}$ 는 서로 독립이고 같은 결합확률분포를 갖는 확률벡터들로서 평균벡터가 0이고 공분산행렬이 Σ 이며, $\{C(k)\}$ 는 2×2 행렬들로서 다음 조건을 만족한다고 하자.

$$\sum_{k=-\infty}^{\infty} |c_{ij}(k)| < \infty \quad (i, j = 1, 2)$$

여기서 $c_{ij}(k)$ 는 행렬 $C(k)$ 의 (i, j) 원소이다. 이러한 조건이 만족되면 고정된 정수 h 에 대해서 다음 식들이 성립한다.

$$\begin{aligned} \text{plim}_{T \rightarrow \infty} \hat{\gamma}_{ij}(h) &= \gamma_{ij}(h) & (i, j = 1, 2) \\ \text{plim}_{T \rightarrow \infty} \hat{\rho}_{ij}(h) &= \rho_{ij}(h) & (i, j = 1, 2) \end{aligned}$$

여기서 plim은 확률수렴 (converge in probability)을 뜻한다. ▣

표본자기공분산과 표본상관계수의 점근적 정규성에 관해서는 [SAS4TSA2]의 제10.2절을 참조하라. 다음 정리들에 표본교차공분산과 표본교차상관계수의 점근분포들을 기술한다. 이 정리들은 Bartlett(1946)의 연구결과를 확장 발전시킨 것이다.

정리 7.4 2변량시계열데이터 $\{y_t = (y_{1,t}, y_{2,t})'\}$ 가 다변량정규분포를 가지며 또한 다음 부등식을 만족한다고 하자.

$$\sum_{h=-\infty}^{\infty} |\gamma_{ij}(h)| < \infty \quad (i, j = 1, 2)$$

이 조건하에서 $\sqrt{T}\{\hat{\gamma}_{12}(0) - \gamma_{12}(0)\}, \sqrt{T}\{\hat{\gamma}_{12}(1) - \gamma_{12}(1)\}, \dots, \sqrt{T}\{\hat{\gamma}_{12}(n) - \gamma_{12}(n)\}$ 은 점근적 다변량정규분포를 가지며 점근평균들은 0이고 점근공분산은 다음과 같다.

$$\begin{aligned} \lim_{T \rightarrow \infty} T \text{Cov}\{\hat{\gamma}_{12}(r), \hat{\gamma}_{12}(s)\} \\ = \sum_{k=-\infty}^{\infty} \{\gamma_{11}(h)\gamma_{22}(h+s-r) + \gamma_{12}(h+s)\gamma_{21}(h-r)\} \end{aligned} \quad \text{▣}$$

정리 7.5 2변량시계열데이터 $\{y_t = (y_{1,t}, y_{2,t})'\}$ 의 자기공분산행렬함수가 다음 부등식을 만족한다고 하자.

$$\sum_{h=-\infty}^{\infty} |\gamma_{ij}(h)| < \infty \quad (i, j = 1, 2)$$

이 조건하에서 $\sqrt{T}\{\hat{\rho}_{12}(0) - \rho_{12}(0)\}, \sqrt{T}\{\hat{\rho}_{12}(1) - \rho_{12}(1)\}, \dots, \sqrt{T}\{\hat{\rho}_{12}(n) - \rho_{12}(n)\}$ 은 점근적으로 다변량정규분포를 가지며 점근평균들은 0이고 점근공분산은 다음과 같다.

$$\begin{aligned} & \lim_{T \rightarrow \infty} TCov\{\hat{\rho}_{12}(r), \hat{\rho}_{12}(s)\} \\ &= \sum_{k=-\infty}^{\infty} [\rho_{11}(k) \rho_{22}(k+s-r) + \rho_{12}(k+s) \rho_{21}(k-r) \\ & \quad - \rho_{12}(r) \{\rho_{11}(k) \rho_{21}(k+s) + \rho_{22}(k) \rho_{21}(k-s)\} \\ & \quad - \rho_{12}(s) \{\rho_{11}(k) \rho_{21}(k+r) + \rho_{22}(k) \rho_{21}(k-s)\} \\ & \quad + \frac{1}{2} \rho_{12}(r) \rho_{12}(s) \{\rho_{11}(k)^2 + 2\rho_{12}(k)^2 + \rho_{22}(k)^2\}] \quad \square \end{aligned}$$

따름정리 7.1 정리 7.4의 조건하에서 $\{y_{1,t}\}$ 또는 $\{y_{2,t}\}$ 가 백색잡음과정이며, 또한 어떤 구간 $[a, b]$ 에 대해서 $h \in [a, b]$ 이면 $\rho_{12}(h) = 0$ 이라고 가정하자. 이러한 조건들이 만족되면 다음 식이 성립된다.

$$\lim_{T \rightarrow \infty} TVar\{\hat{\rho}_{12}(h)\} = 1, \quad h \in [a, b] \quad \square$$

다음 정리는 두 시계열데이터들 사이의 상관계수에 관한 가설을 검정하는 데 매우 유효하다.

정리 7.6 2변량시계열데이터 $\{y_t = (y_{1,t}, y_{2,t})'\}$ 가 다음 식들을 만족한다고 하자.

$$\begin{aligned} y_{1,t} &= \sum_{k=-\infty}^{\infty} \alpha_k v_{1,t-k} \\ y_{2,t} &= \sum_{k=-\infty}^{\infty} \beta_k v_{2,t-k} \end{aligned}$$

여기서 $\{v_{1,t}\}$ 는 서로 독립이고 같은 결합확률분포를 갖는 확률변수들로서 평균이 0이고 분산이 σ_1^2 이며, $\{v_{2,t}\}$ 는 서로 독립이고 같은 결합확률분포를 갖는 확률변수들로서 평균이 0이고 분산이 σ_2^2 이며, $\{v_{1,t}\}$ 와 $\{v_{2,t}\}$ 는 서로 독립이다. 또한 다음 부등식들이 만족된다고 가정하자.

$$\sum_{k=-\infty}^{\infty} |\alpha_k| < \infty, \quad \sum_{k=-\infty}^{\infty} |\beta_k| < \infty$$

이 조건이 만족되면 $\sqrt{T}\hat{\rho}_{12}(0), \sqrt{T}\hat{\rho}_{12}(1), \dots, \sqrt{T}\hat{\rho}_{12}(n)$ 은 점근적으로 다변량정규분포를 가지며 점근평균들은 0이고 점근공분산은 다음과 같다.

$$\lim_{T \rightarrow \infty} TCov\{\hat{\rho}_{12}(r), \hat{\rho}_{12}(s)\} = \sum_{k=-\infty}^{\infty} \rho_{11}(k) \rho_{22}(h+r-s) \quad \square$$

표본자기공분산행렬함수와 표본자기상관행렬함수의 일치성과 점근적 정규성에 관한 자세한 내용은 Hannan(1970, 제4장), Fuller(1976, 제6장), [BD90](pp. 405~417)을 참조하라.

7.3 VAR계수행렬의 추정

7.3.1 Yule-Walker추정법

VAR계수행렬들을 추정하는 방법의 하나는 제 6.4.4 소절에서 소개한 Whittle알고리즘을 표본자기공분산행렬들에 적용하는 것이다. 이는 단변량시계열분석에서 Levinson-Durbin 알고리즘을 적용해서 Yule-Walker추정량을 구하는 것과 같은 방법이므로 이 추정행렬들을 Yule-Walker추정행렬들이라고 부른다. 즉, 전향VAR(p)모형의 VAR계수행렬들의 Yule-Walker추정행렬들은 다음 식을 만족한다.

$$\hat{\Gamma}(i) = \hat{\Phi}_1^i \hat{\Gamma}(i-1) + \hat{\Phi}_2^i \hat{\Gamma}(i-2) + \dots + \hat{\Phi}_p^i \hat{\Gamma}(i-p) \quad (i=1, 2, \dots, p)$$

후향VAR(p)모형의 VAR계수행렬들의 Yule-Walker추정행렬들은 다음 식을 만족한다.

$$\hat{\Gamma}(i) = \hat{\Pi}_1^i \hat{\Gamma}(i+1) + \hat{\Pi}_2^i \hat{\Gamma}(i+2) + \dots + \hat{\Pi}_p^i \hat{\Gamma}(i+p) \quad (i = -1, -2, \dots, -p)$$

또한 전향VAR(p)모형의 백색잡음과정의 공분산행렬의 추정행렬은 다음과 같다.

$$\hat{\Sigma}_p = \hat{\Gamma}(0) - \hat{\Phi}_1^p \hat{\Gamma}(-1) - \hat{\Phi}_2^p \hat{\Gamma}(-2) - \dots - \hat{\Phi}_p^p \hat{\Gamma}(-p)$$

후향VAR(p)모형의 백색잡음과정의 공분산행렬의 추정행렬은 다음과 같다.

$$\hat{\Omega}_p = \hat{\Gamma}(0) - \hat{\Phi}_1^p \hat{\Gamma}(1) - \hat{\Phi}_2^p \hat{\Gamma}(2) - \dots - \hat{\Phi}_p^p \hat{\Gamma}(p)$$

SAS/ETS의 PROC STATESPACE를 사용해서 Yule-Walker추정행렬들을 구할 수 있다. 한 예로 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM7-6.PRG에 수록되어 있다).

```

/* */
/* EXAM7-6.PRG */
/* */
/* APPLYING WHITTLE'S ALGORITHM */
/* USING PROC STATESPACE */
/* DATASET = VAR31A */
LIBNAME TSA4DATA 'A:\TSA4DATA';
DATA EXAM76;
    SET TSA4DATA.VAR31A;
RUN;
OPTIONS PAGESIZE=60;

```

```

PROC STATESPACE DATA=EXAM76 OUTAR=EXAM76B;
  VAR Y1 Y2 Y3;
  RUN;
TITLE1 'RESULT FROM WHITTLE' 'S ALGORITHM';
PROC PRINT DATA=EXAM76B;
  RUN;

```

PROC STATESPACE에서 Whittle알고리즘을 적용해서 얻은 Yule-Walker추정행렬들을 OUTPUT윈도우에 출력하기 위해서는 옵션 OUTAR=에 Yule-Walker추정행렬들을 저장할 새로운 SAS데이터세트를 지정한다. 옵션 OUTAR=에 지정된 이 SAS데이터세트가 포함하고 있는 변수들은 다음과 같다.

- ① ORDER : VAR모형의 차수
- ② AIC : 추정된 VAR모형의 AIC값이다. 제7.5.4소절에서 자세히 설명할 것이다.
- ③ SIGFi : 전향VAR(p)모형의 백색잡음과정의 공분산행렬의 추정행렬 $\hat{\Sigma}_p$ 의 제 i 번째 열
- ④ SIGBi : 후향VAR(p)모형의 백색잡음과정의 공분산행렬의 추정행렬 $\hat{\Omega}_p$ 의 제 i 번째 열
- ⑤ FORi_j : 전향VAR(p)모형의 제 i 번째 VAR계수행렬의 추정행렬 $\hat{\Phi}_p^i$ 의 제 j 번째 열
- ⑥ BACi_j : 후향VAR(p)모형의 제 i 번째 VAR계수행렬의 추정행렬 $\hat{\Pi}_p^i$ 의 제 j 번째 열

이 SAS프로그램을 수행하면 SAS데이터세트 EXAM76에 Whittle알고리즘을 적용한 결과가 SAS데이터세트 EXAM76B에 저장된다. PROC PRINT를 사용해서 SAS데이터세트 EXAM76B를 OUTPUT윈도우에 출력한 결과가 그림 7.6이다. 그림 7.6을 설명하면 다음과 같다. 변수 OBS의 값이 1, 2, 3인 행들은 차수가 0인 VAR모형을 추정한 것이다. 이는 주어진 다변량시계열데이터가 다변량백색잡음과정이라고 가정하는 것이다. 이 다변량백색잡음과정의 AIC값과 공분산행렬의 추정행렬은 다음과 같다.

$$\begin{aligned}
 AIC(0) &= 230.844 \\
 \hat{\Sigma}_0 = \hat{\Omega}_0 &= \begin{bmatrix} 1.09883 & -0.20886 & -0.00201 \\ -0.20886 & 1.93803 & 0.42915 \\ -0.00201 & 0.42915 & 1.37070 \end{bmatrix}
 \end{aligned}$$

변수 OBS의 값이 4, 5, 6인 행들은 차수가 1인 VAR모형을 추정한 것이다. 추정된 VAR(1)모형의 AIC값과 추정행렬들은 다음과 같다.

$$\begin{aligned}
 AIC(1) &= 38.387 \\
 \hat{\Sigma}_1 &= \begin{bmatrix} 1.23103 & -0.23715 & -0.11399 \\ -0.23715 & 1.10007 & 0.35109 \\ -0.11399 & 0.35109 & 0.99588 \end{bmatrix}
 \end{aligned}$$

$$\hat{\Omega}_1 = \begin{bmatrix} 1.17025 & 0.00267 & 0.04187 \\ 0.00267 & 1.33957 & 0.16745 \\ 0.04187 & 0.16745 & 0.75318 \end{bmatrix}$$

$$\hat{\Phi}_1 = \begin{bmatrix} -0.59671 & -0.04800 & -0.05150 \\ -0.06663 & 0.34368 & 0.55771 \\ -0.06898 & 0.40357 & -0.33633 \end{bmatrix}$$

$$\hat{\Pi}_1 = \begin{bmatrix} -0.60724 & -0.14367 & -0.11298 \\ 0.04935 & 0.40218 & 0.34997 \\ -0.01341 & 0.55716 & -0.38430 \end{bmatrix}$$

변수 OBS의 값이 7, 8, 9인 행들은 차수가 2인 VAR모형을 추정한 것이다. 추정된 VAR(2)모형의 AIC값과 추정행렬들은 다음과 같다.

$$\text{AIC}(2) = 41.305$$

$$\hat{\Sigma}_2 = \begin{bmatrix} 1.18647 & -0.21460 & -0.11113 \\ -0.21460 & 1.08432 & 0.33434 \\ -0.11113 & 0.33434 & 0.94003 \end{bmatrix}$$

$$\hat{\Omega}_2 = \begin{bmatrix} 1.10461 & 0.00739 & 0.03990 \\ 0.00739 & 1.28403 & 0.16420 \\ 0.03990 & 0.16420 & 0.75282 \end{bmatrix}$$

$$\hat{\Phi}_1^2 = \begin{bmatrix} -0.55818 & -0.09161 & -0.11180 \\ -0.05119 & 0.38862 & 0.59768 \\ 0.05424 & 0.46295 & -0.27844 \end{bmatrix}$$

$$\hat{\Phi}_2^2 = \begin{bmatrix} 0.07673 & 0.16999 & -0.02464 \\ 0.01672 & -0.10748 & 0.00124 \\ 0.19572 & -0.09140 & 0.00988 \end{bmatrix}$$

$$\hat{\Pi}_1^2 = \begin{bmatrix} -0.54041 & -0.22727 & 0.00296 \\ 0.13528 & 0.46557 & 0.37207 \\ 0.01783 & 0.56011 & -0.37316 \end{bmatrix}$$

$$\hat{\Pi}_2^2 = \begin{bmatrix} 0.08746 & -0.04532 & 0.25615 \\ 0.16090 & -0.07132 & -0.07720 \\ 0.00834 & -0.01520 & 0.00664 \end{bmatrix}$$

그림 7.6의 나머지 부분도 같은 방법으로 해석한다. 추정될 VAR모형들의 최대 차수를 지정하기 위해서는 PROC STATESPACE스태이트먼트의 옵션 ARMAX=을 사용한다. 한 예로 옵션 ARMAX에 5를 지정하면 VAR(0)모형, VAR(1)모형, ..., VAR(4)모형, VAR(5)모형의 VAR계수행렬들의 Yule-Walker추정행렬들을 계산할 것이다.

SAS		1:28 Monday, November 23, 1992 1						
RESULT FROM WHITTLE'S ALGORITHM								
1:28 Monday, November 23, 1992 5								
OBS	ORDER	AIC	SIGF1	SIGF2	SIGF3	SIGB1	SIGB2	SIGB3
1	0	230.844	1.90883	-0.20886	-0.00201	1.90883	-0.20886	-0.00201
2	0	230.844	-0.20886	1.93803	0.42915	-0.20886	1.93803	0.42915
3	0	230.844	-0.00201	0.42915	1.37070	-0.00201	0.42915	1.37070
4	1	38.387	1.23103	-0.23715	-0.11399	1.17025	0.00267	0.04187
5	1	38.387	-0.23715	1.10007	0.35109	0.00267	1.33957	0.16745
6	1	38.387	-0.11399	0.35109	0.99588	0.04187	0.16745	0.75318
7	2	41.305	1.18647	-0.21460	-0.11113	1.10461	0.00739	0.03990
8	2	41.305	-0.21460	1.08432	0.33434	0.00739	1.28403	0.16420
9	2	41.305	-0.11113	0.33434	0.94003	0.03990	0.16420	0.75282
10	3	48.611	1.13221	-0.21663	-0.11675	1.07401	0.02576	0.03141
11	3	48.611	-0.21663	1.07027	0.33126	0.02576	1.24634	0.17010
12	3	48.611	-0.11675	0.33126	0.93299	0.03141	0.17010	0.74499
13	4	63.325	1.11747	-0.22057	-0.12026	1.06989	0.02655	0.03093
14	4	63.325	-0.22057	1.06689	0.33085	0.02655	1.23725	0.17635
15	4	63.325	-0.12026	0.33085	0.93043	0.03093	0.17635	0.73920
16	5	69.982	1.10445	-0.22379	-0.12389	1.04967	0.03851	0.03871
17	5	69.982	-0.22379	1.05152	0.31310	0.03851	1.18933	0.15453
18	5	69.982	-0.12389	0.31310	0.88378	0.03871	0.15453	0.72294
19	6	82.288	1.09184	-0.21478	-0.12609	1.04077	0.04392	0.04077
20	6	82.288	-0.21478	1.02617	0.30621	0.04392	1.18291	0.15928
21	6	82.288	-0.12609	0.30621	0.87926	0.04077	0.15928	0.70800
22	7	94.249	1.07656	-0.19590	-0.11298	1.02969	0.04816	0.05342
23	7	94.249	-0.19590	1.00204	0.29045	0.04816	1.17875	0.15639
24	7	94.249	-0.11298	0.29045	0.86505	0.05342	0.15639	0.69089
25	8	104.206	1.02987	-0.17680	-0.11239	1.02428	0.04873	0.06235
26	8	104.206	-0.17680	0.99029	0.28964	0.04873	1.17580	0.15917
27	8	104.206	-0.11239	0.28964	0.86422	0.06235	0.15917	0.66269
28	9	115.278	1.00616	-0.17222	-0.11605	1.00507	0.04705	0.05325
29	9	115.278	-0.17222	0.98434	0.29187	0.04705	1.16183	0.15138
30	9	115.278	-0.11605	0.29187	0.85315	0.05325	0.15138	0.65030
31	10	118.113	0.98628	-0.14813	-0.08077	0.99792	0.02482	0.06144
32	10	118.113	-0.14813	0.95236	0.24606	0.02482	1.09136	0.17664
33	10	118.113	-0.08077	0.24606	0.78702	0.06144	0.17664	0.64003
OBS	FOR1_1	FOR1_2	FOR1_3	FOR2_1	FOR2_2	FOR2_3	FOR3_1	FOR3_2
1
2
3
4	-0.59671	-0.04800	-0.05150
5	-0.06663	0.34368	0.55771
6	-0.06898	0.40357	-0.33633
7	-0.55818	-0.09161	-0.11180	0.07673	0.16999	-0.02464	.	.
8	-0.05119	0.38862	0.59768	0.01672	-0.10748	0.00124	.	.
9	0.05424	0.46295	-0.27844	0.19572	-0.09140	0.00988	.	.
10	-0.56501	-0.09102	-0.06260	-0.01945	0.14594	-0.11970	-0.14952	0.13027
11	-0.06980	0.39752	0.59145	0.03139	-0.15229	-0.02095	0.04988	0.08713
12	0.05343	0.46383	-0.26559	0.16709	-0.15267	0.02673	-0.04429	0.02549
13	-0.58644	-0.10177	-0.06056	-0.04491	0.15170	-0.10393	-0.16601	0.14136
14	-0.07048	0.39468	0.59673	0.02624	-0.15220	-0.02487	0.06589	0.12085
15	0.04490	0.46491	-0.26674	0.16468	-0.15502	0.03473	-0.06291	0.03759
16	-0.59135	-0.09886	-0.06790	-0.06641	0.15712	-0.11477	-0.16809	0.13709
17	-0.06381	0.40280	0.59591	0.02884	-0.15247	-0.04092	0.06975	0.12455
18	0.03788	0.46903	-0.26795	0.13834	-0.17445	0.01849	-0.09601	0.05983
19	-0.60088	-0.09601	-0.08218	-0.08292	0.16002	-0.12658	-0.19684	0.13246
20	-0.06530	0.38862	0.58739	0.04264	-0.13277	-0.03208	0.09446	0.12406
21	0.03143	0.46306	-0.27596	0.13614	-0.16466	0.01718	-0.09938	0.06007
22	-0.60312	-0.11308	-0.08455	-0.09099	0.15696	-0.13347	-0.19212	0.14921
23	-0.06331	0.41032	0.59043	0.05302	-0.12756	-0.02908	0.08751	0.10434
24	0.03863	0.47441	-0.27034	0.15093	-0.16263	0.03393	-0.09553	0.04263
25	-0.62040	-0.08800	-0.07583	-0.09622	0.10848	-0.14622	-0.19839	0.15052
26	-0.05445	0.39939	0.58096	0.06246	-0.10493	-0.02081	0.10064	0.10208
27	0.03922	0.47477	-0.27135	0.15190	-0.16187	0.03327	-0.09592	0.04112
28	-0.63599	-0.08221	-0.07526	-0.11323	0.11038	-0.14473	-0.20457	0.13515
29	-0.06144	0.40400	0.57687	0.05530	-0.09932	-0.02153	0.09624	0.09313

30	0.03005	0.47078	-0.27138	0.14538	-0.15801	0.02768	-0.08644	0.03703
31	-0.63453	-0.07716	-0.06740	-0.08882	0.10015	-0.13861	-0.18459	0.12313
32	-0.05561	0.39873	0.56755	0.03630	-0.09015	-0.02835	0.07941	0.10585
33	0.03646	0.46162	-0.28337	0.11400	-0.14294	0.01826	-0.11413	0.05665
OBS	FOR3_3	FOR4_1	FOR4_2	FOR4_3	FOR5_1	FOR5_2	FOR5_3	FOR6_1
1
2
3
4
5
6
7
8
9
10	-0.12602
11	0.02782
12	0.07022
13	-0.19694	-0.013319	0.09203	-0.09694
14	0.00196	0.032542	0.01224	-0.05665
15	0.05260	-0.035502	0.00742	-0.03957
16	-0.17555	-0.069000	0.08878	-0.11812	-0.10371	0.01432	-0.04171	.
17	0.01405	-0.001503	-0.07509	-0.01030	-0.06363	0.00254	0.12427	.
18	0.08128	-0.084583	-0.13972	-0.05208	-0.05562	0.14930	0.11845	.
19	-0.18036	-0.090945	0.09433	-0.09816	-0.14601	0.00871	-0.09833	-0.061172
20	-0.00150	0.015969	-0.07155	-0.01022	-0.07512	-0.09065	0.21178	-0.032928
21	0.07077	-0.084492	-0.13739	-0.04190	-0.08570	0.11108	0.13200	-0.055937
22	-0.17964	-0.082775	0.09286	-0.11406	-0.14334	0.01423	-0.08870	-0.092861
23	-0.00424	0.003689	-0.07401	0.01069	-0.08474	-0.09407	0.20092	0.008871
24	0.07624	-0.076237	-0.13445	-0.02429	-0.07700	0.10325	0.11415	-0.005456
25	-0.17625	-0.068633	0.11483	-0.10768	-0.10933	0.01975	-0.11094	-0.071550
26	0.00848	0.005314	-0.08932	0.01543	-0.08415	-0.09436	0.21674	0.012231
27	0.08162	-0.075831	-0.13609	-0.02264	-0.07615	0.10744	0.11333	0.000223
28	-0.17838	-0.069514	0.12367	-0.13112	-0.10597	0.03493	-0.11659	-0.062381
29	0.00146	-0.006754	-0.09063	0.01747	-0.08671	-0.08871	0.21557	0.009462
30	0.08342	-0.060956	-0.13563	-0.00514	-0.06261	0.10309	0.12408	0.027651
31	-0.17167	-0.070253	0.14424	-0.12123	-0.10000	0.03663	-0.13388	-0.071192
32	-0.00671	-0.000133	-0.11007	0.00899	-0.08915	-0.09458	0.24807	0.021980
33	0.07188	-0.054149	-0.16672	-0.01968	-0.07010	0.09582	0.16629	0.044083
OBS	FOR6_2	FOR6_3	FOR7_1	FOR7_2	FOR7_3	FOR8_1	FOR8_2	FOR8_3
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19	0.076626	-0.06798
20	-0.048632	0.18415
21	0.004416	0.04325
22	-0.012499	-0.01704	-0.05609	0.00082	0.13333	.	.	.
23	0.052682	0.11280	0.08068	0.02194	-0.16313	.	.	.
24	0.086889	0.02497	0.08217	-0.03730	-0.08701	.	.	.
25	-0.006988	-0.02278	-0.07673	-0.13408	0.24890	-0.048543	-0.05487	0.25873
26	0.045982	0.10426	0.11837	0.10186	-0.19109	0.063059	-0.02097	-0.10266
27	0.083477	0.02421	0.07938	-0.02459	-0.08232	-0.011343	-0.02126	-0.00951
28	-0.023750	-0.02703	-0.09494	-0.11788	0.24684	-0.035820	-0.15241	0.28235
29	0.056445	0.09187	0.12641	0.09824	-0.18656	0.030455	-0.03657	-0.08133
30	0.084063	0.03165	0.09749	-0.02794	-0.10013	0.034336	-0.01613	0.03986

31	-0.031606	-0.03693	-0.11959	-0.13432	0.26008	-0.065994	-0.14558	0.29443
32	0.058670	0.10822	0.15739	0.12279	-0.19791	0.075350	-0.04982	-0.09997
33	0.090613	0.05312	0.14004	0.00713	-0.12016	0.096323	-0.03349	0.01567
OBS	FOR9_1	FOR9_2	FOR9_3	FOR10_1	FOR10_2	FOR10_3	BAC1_1	BAC1_2
1
2
3
4	-0.60724	-0.14367
5	0.04955	0.40218
6	0.01341	0.55716
7	-0.54041	-0.22727
8	0.13528	0.46557
9	0.01783	0.56011
10	-0.51407	-0.20547
11	0.11831	0.45615
12	0.01280	0.57574
13	-0.52067	-0.20676
14	0.13105	0.44212
15	0.00470	0.58787
16	-0.52159	-0.19461
17	0.14140	0.43879
18	0.00651	0.58714
19	-0.53399	-0.18370
20	0.14875	0.43353
21	0.01213	0.58853
22	-0.53034	-0.17927
23	0.14673	0.43398
24	0.01139	0.57977
25	-0.53381	-0.18218
26	0.14976	0.43397
27	0.02173	0.58159
28	0.049091	0.07117	0.12998	.	.	.	-0.52357	-0.17844
29	-0.068131	-0.02908	0.02717	.	.	.	0.15375	0.43809
30	0.070265	-0.06333	0.06044	.	.	.	0.02653	0.58869
31	0.055295	0.10786	0.04040	0.035375	0.10214	-0.12760	-0.52028	-0.18538
32	-0.068767	-0.03321	0.12786	-0.040881	-0.15392	0.10917	0.16031	0.41715
33	0.060172	-0.08624	0.20859	-0.072628	-0.21179	0.17451	0.02454	0.59792
OBS	BAC1_3	BAC2_1	BAC2_2	BAC2_3	BAC3_1	BAC3_2	BAC3_3	BAC4_1
1
2
3
4	-0.11298
5	0.34997
6	-0.38430
7	0.00296	0.08746	-0.04532	0.25615
8	0.37207	0.16090	-0.07132	-0.07720
9	-0.37316	0.00834	-0.01520	0.00664
10	0.00023	0.01564	-0.03916	0.18713	-0.14136	0.05018	-0.08342	.
11	0.37538	0.24934	-0.11698	-0.13450	0.14805	0.13167	0.01741	.
12	-0.37516	-0.02805	-0.04627	0.00921	-0.06108	0.00713	0.04905	.
13	0.00184	0.02401	-0.03957	0.18851	-0.14062	0.05817	-0.12838	-0.010288
14	0.38591	0.24956	-0.12605	-0.12322	0.20101	0.12853	0.01086	0.091356
15	-0.37995	-0.02386	-0.04742	0.00173	-0.09698	0.02932	0.06319	-0.060741
16	-0.01551	0.00600	-0.01250	0.16764	-0.13652	0.05835	-0.14090	-0.082955
17	0.39493	0.27159	-0.13258	-0.13115	0.16389	0.15488	0.00377	0.090859
18	-0.37370	-0.02208	-0.05649	-0.00509	-0.11532	0.05397	0.06023	-0.065785
19	-0.00774	-0.00444	-0.01725	0.15571	-0.15300	0.07477	-0.14749	-0.077808
20	0.39725	0.27880	-0.13643	-0.12167	0.17995	0.14649	0.01394	0.092130
21	-0.38777	-0.02366	-0.04784	-0.00594	-0.12675	0.04161	0.05709	-0.068671
22	-0.02188	0.00074	-0.02030	0.13308	-0.15067	0.09078	-0.14708	-0.081541
23	0.39627	0.27718	-0.12642	-0.11680	0.17508	0.13749	0.01284	0.085728
24	-0.36678	-0.02345	-0.05192	0.02239	-0.12257	0.02484	0.06109	-0.051778
25	-0.01108	-0.00270	-0.02052	0.12633	-0.15243	0.10145	-0.15793	-0.087245
26	0.39265	0.27424	-0.12285	-0.11257	0.16525	0.15682	0.02002	0.081294
27	-0.39252	-0.00902	-0.04955	0.02980	-0.10344	0.01573	0.08136	-0.037438
28	-0.03454	0.00324	-0.00003	0.10394	-0.14720	0.09594	-0.14960	-0.079680
29	0.36699	0.28625	-0.11134	-0.14145	0.17222	0.14236	0.02355	0.087562
30	-0.41789	-0.00645	-0.03563	0.01024	-0.09609	0.01176	0.08114	-0.021051
31	-0.03220	0.00718	0.00140	0.09962	-0.13550	0.09792	-0.16577	-0.076027

696 제 7 장 VAR모형의 추정과 식별

32	0.37519	0.29877	-0.10921	-0.15667	0.21097	0.15202	-0.03023	0.098265
33	-0.41604	-0.01150	-0.04207	0.02338	-0.11059	0.00630	0.10484	-0.026720
OBS	BAC4_2	BAC4_3	BAC5_1	BAC5_2	BAC5_3	BAC6_1	BAC6_2	BAC6_3
1
2
3
4
5
6
7
8
9
10
11
12
13	0.04786	-0.06028
14	0.02298	0.00499
15	-0.04128	-0.02438
16	0.08498	-0.03774	-0.11898	-0.070444	-0.04609	.	.	.
17	-0.06715	0.09857	0.02318	-0.047508	0.23929	.	.	.
18	-0.10951	-0.02874	-0.00708	0.051851	0.10121	.	.	.
19	0.08094	-0.04549	-0.15931	-0.037108	-0.05474	-0.069931	-0.02311	-0.065964
20	-0.07652	0.10507	0.06169	-0.044960	0.26546	0.069460	-0.02299	0.028949
21	-0.09116	-0.02520	-0.00685	0.007555	0.02701	-0.013048	0.11848	-0.010098
22	0.07473	-0.05591	-0.17346	-0.014174	-0.05826	-0.087400	-0.07747	-0.077984
23	-0.07780	0.11279	0.07122	-0.055194	0.26948	0.082708	-0.00010	-0.004883
24	-0.08739	-0.01124	0.00677	-0.019417	0.03313	0.021335	0.17282	0.033604
25	0.07669	-0.06085	-0.18848	-0.014909	-0.06002	-0.087283	-0.07137	-0.078541
26	-0.08055	0.10876	0.06578	-0.045223	0.26499	0.085105	-0.00333	-0.009720
27	-0.10166	0.00689	0.04195	-0.037664	0.05795	0.033926	0.14705	0.051331
28	0.05320	-0.04769	-0.17467	-0.013215	-0.04959	-0.054417	-0.07522	-0.077768
29	-0.07606	0.12600	0.06801	-0.064312	0.27414	0.098692	-0.01569	0.012241
30	-0.10769	0.01006	0.05280	-0.041882	0.07018	0.058119	0.12779	0.065788
31	0.06117	-0.04314	-0.17867	-0.007396	-0.03389	-0.057770	-0.08993	-0.075052
32	-0.05084	0.14295	0.05251	-0.051480	0.32816	0.088640	-0.06139	0.020448
33	-0.11637	0.00487	0.05321	-0.046958	0.05122	0.059284	0.14638	0.058901
OBS	BAC7_1	BAC7_2	BAC7_3	BAC8_1	BAC8_2	BAC8_3	BAC9_1	BAC9_2
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22	-0.03074	0.048998	0.069892
23	0.01547	0.027286	-0.069122
24	0.06429	-0.081964	-0.035244
25	-0.04232	0.038586	0.018387	-0.02857	0.064521	-0.040668	.	.
26	-0.00648	0.049848	-0.056049	-0.03393	-0.037591	-0.023133	.	.
27	0.15039	-0.056439	0.012301	0.14934	-0.048334	0.023596	.	.
28	-0.04962	0.042330	0.020611	-0.00327	0.049218	0.053114	0.05705	-0.09619
29	0.01175	0.029878	-0.039825	0.02888	-0.001546	-0.032309	0.09744	0.00107
30	0.15224	-0.058515	0.026359	0.21350	-0.067742	0.036858	0.10530	0.01373
31	-0.05035	0.043817	0.030977	0.01139	0.032915	0.058040	0.07212	-0.05118
32	0.01335	0.038538	-0.010569	0.07293	-0.052223	-0.019526	0.13624	0.14722
33	0.14677	-0.055413	0.010147	0.19518	-0.047986	0.027169	0.06952	-0.03330

OBS	BAC9_3	BAC10_1	BAC10_2	BAC10_3
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28	0.12372	.	.	.
29	-0.05876	.	.	.
30	0.04884	.	.	.
31	0.11366	0.021586	-0.016342	-0.07782
32	-0.06245	0.059245	-0.090514	-0.22244
33	0.05851	-0.053828	0.013395	0.07900

그림 7.6 PROC STATESPACE를 이용해서 구한 Yule-Walker추정량들

SAS/IML을 사용해서 Whittle알고리즘을 적용하기 위해서는 제6.4.4소절의 알고리즘 6.2를 이용한다. 이 알고리즘은 다변량시계열데이터에 적합한 VAR모형의 차수를 결정하는데 유용하게 이용될 것이다. 앞의 예제를 SAS/IML을 사용해서 해결하기 위해 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM 7-7.PRG에 수록되어 있다).

```

/* */
/*          EXAM7-7.PRG          */
/* */
/*          APPLYING WHITTLE'S ALGORITHM          */
/*          USING ALGORITHM 6.2          */
/*          DATASET = VAR31A          */
PROC IML;
RESET NOLOG NOCENTER;
/* ***** */
/*          PUT MODULE WHITTLE          */
/*          IN SUBSECTION 6.4.4 HERE          */
/* ***** */

```

```

/ *                               * /
/ *   ESTIMATE THE SAMPLE ACVMF   * /
/ *                               * /
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=60;
USE TSA4DATA.VAR31A;
READ ALL VAR{Y1 Y2 Y3} INTO Y;
CLOSE TSA4DATA.VAR31A;
SACVMF = COVLAGE(Y, 6);
/ *                               * /
/ *   APPLY WHITTLE'S ALGORITHM   * /
/ *                               * /
RUN WHITTLE(SACVMF, 3, 4);
QUIT;

```

SAS			21:56 Saturday, November 21, 1992 2		
ORDER =			K 1		
LAMBDA			V		
1.2228246	-0.235566	-0.11323	1.1624504	0.0026521	0.0415911
-0.235566	1.0927374	0.3487534	0.0026521	1.3306393	0.1663358
-0.11323	0.3487534	0.9892443	0.0415911	0.1663358	0.7481604
LAG =			J 1		
LKJ			MKJ		
-0.596709	-0.048002	-0.051505	-0.607242	-0.143667	-0.112978
-0.066629	0.3436834	0.5577051	0.049349	0.4021817	0.3499727
-0.068984	0.4035729	-0.336332	0.0134101	0.557163	-0.384299
ORDER =			K 2		
LAMBDA			V		
1.1785598	-0.213172	-0.110384	1.0972494	0.0073415	0.0396312
-0.213172	1.0770931	0.3321064	0.0073415	1.2754677	0.1631052
-0.110384	0.3321064	0.9337587	0.0396312	0.1631052	0.7478024
LAG =			J 1		
LKJ			MKJ		
-0.558176	-0.091614	-0.111797	-0.540406	-0.227267	0.0029559
-0.051189	0.388622	0.5976821	0.1352812	0.4655738	0.3720689
0.0542441	0.4629495	-0.278435	0.0178334	0.5601063	-0.373157
LAG =			J 2		
LKJ			MKJ		
0.0767263	0.1699852	-0.024643	0.0874559	-0.045325	0.2561495
0.0167207	-0.107475	0.001235	0.1608991	-0.07132	-0.077204
0.1957212	-0.091405	0.0098777	0.008342	-0.0152	0.0066435

```

ORDER =          K
          3
      LAMBDA
1.1246572 -0.215183 -0.115971      V
0.0255867 0.0311997
-0.215183 1.0631363 0.3290527      1.0668465 0.0255867 0.0311997
0.0255867 1.2380355 0.1689633
-0.115971 0.3290527 0.9267695      0.0311997 0.1689633 0.7400249

LAG =          J
          1
      LKJ
-0.565009 -0.091015 -0.062604      MKJ
-0.514071 -0.205469 0.0002343
-0.069803 0.39752 0.5914471      0.1183133 0.4561489 0.3753828
0.0534307 0.4638272 -0.265589      0.0127999 0.5757385 -0.375156

LAG =          J
          2
      LKJ
-0.019449 0.1459423 -0.119697      MKJ
0.0156431 -0.039155 0.1871277
0.0313937 -0.152285 -0.020953      0.249335 -0.116983 -0.134501
0.1670858 -0.152668 0.0267282      -0.028047 -0.046275 0.0092123

LAG =          J
          3
      LKJ
-0.149516 0.1302689 -0.126024      MKJ
-0.141365 0.050177 -0.083423
0.0498821 0.0871335 0.0278152      0.1480539 0.1316687 0.0174057
-0.044291 0.0254886 0.0702199      -0.061079 0.0071287 0.0490524

ORDER =          K
          4
      LAMBDA
1.1100247 -0.219096 -0.11946      V
1.0627556 0.0263756 0.0307203
-0.219096 1.0597749 0.3286492      0.0263756 1.2289985 0.1751724
-0.11946 0.3286492 0.9242233      0.0307203 0.1751724 0.7342731

LAG =          J
          1
      LKJ
-0.586438 -0.101774 -0.060562      MKJ
-0.520666 -0.206763 0.0018391
-0.070476 0.3946789 0.5967276      0.1310469 0.4421182 0.3859062
0.044897 0.4649137 -0.266739      0.0046976 0.5878697 -0.379951

LAG =          J
          2
      LKJ
-0.044906 0.1517012 -0.103933      MKJ
0.0240116 -0.039567 0.1885101
0.0262429 -0.1522 -0.024873      0.2495567 -0.126054 -0.123218
0.1646816 -0.155021 0.034734      -0.023859 -0.047418 0.0017284

LAG =          J
          3
      LKJ
-0.166011 0.1413634 -0.196936      MKJ
-0.140616 0.0581722 -0.128383
0.0658874 0.1208515 0.0019581      0.2010086 0.1285328 0.0108557
-0.062913 0.0375893 0.0526      -0.096977 0.0293184 0.063191

LAG =          J
          4
      LKJ
-0.013319 0.0920325 -0.096939      MKJ
-0.010288 0.0478611 -0.060275
0.0325418 0.0122439 -0.056652      0.0913558 0.0229845 0.0049885
-0.035502 0.0074192 -0.039566      -0.060741 -0.041282 -0.024379

Exiting IML.

```

그림 7.7 SAS/IML의 모듈 WHITTLE을 이용해서 구한 Yule-Walker 추정행렬들

이 SAS프로그램을 수행한 결과가 그림 7.7이다. 그림 7.6과 그림 7.7을 비교해 보면 VAR계수행렬들의 추정행렬들은 같으나 공분산행렬들의 추정행렬들은 같지 않음을 알 수 있다. 이는 함수 COVLAG와 PROC STATESPACE에 의한 표본자기상관행렬들은 동일하나 표본공분산행렬들이 동일하지 않기 때문에 생기는 현상이다.

7.3.2 최소제곱추정법

다변량시계열분석에서도 통계학에서 널리 사용되는 최소제곱추정법을 사용해서 VAR모형을 추정할 수 있다. 다음과 같은 d 변량VAR(p)모형에서 시계열데이터 y_1, y_2, \dots, y_T 가 추출되었을 때 이 VAR모형을 최소제곱추정법으로 추정해 보자.

$$y_t = \nu + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \dots + \Phi_p y_{t-p} + u_t$$

이후 초기 관찰값들 y_1, y_2, \dots, y_p 는 고정되었다고 가정하자. 편의상 다음과 같은 행렬들을 정의하자.

$$Y := (y_{p+1}, y_{p+2}, \dots, y_T)$$

$$\Phi^* := (\nu, \Phi_1, \Phi_2, \dots, \Phi_p)$$

$$x_t := \begin{bmatrix} 1 \\ y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}$$

$$X := (x_p, x_{p+1}, \dots, x_{T-1})$$

$$V := (v_{p+1}, v_{p+2}, \dots, v_T)$$

따라서 Y 는 $d \times (T-p)$ 행렬, Φ^* 는 $d \times (dp+1)$ 행렬, x_t 는 $(dp+1) \times 1$ 행렬, X 는 $(dp+1) \times (T-p)$ 행렬, V 는 $d \times (T-p)$ 행렬이다. 이 행렬들을 이용해서 주어진 VAR(p)모형을 다음과 같이 쓸 수 있다.

$$Y = \Phi^* X + V$$

이 식을 다음과 같이 쓸 수 있다.

$$\begin{aligned} & \text{vec}(Y) \\ &= \text{vec}(\Phi^* X) + \text{vec}(V) \\ &= (X' \otimes I_d) \text{vec}(\Phi^*) + \text{vec}(V) \end{aligned}$$

또한, $\text{vec}(V)$ 의 공분산행렬은 $I_{T-p} \otimes \Sigma$ 임을 쉽게 알 수 있다. 따라서 다음과 같이 정의되는 제곱합 $S = S(\Phi^*)$ 를 최소로 하는 추정량이 최소제곱추정량이다.

$$\begin{aligned}
& S(\Phi) \\
&= S(\Phi^*) \\
&= \text{vec}(V)' (I_{T-p} \otimes \Sigma)^{-1} \text{vec}(V) \\
&= \text{vec}(V)' (I_{T-p} \otimes \Sigma^{-1}) \text{vec}(V) \\
&= \text{vec}(Y - \Phi^* X)' (I_{T-p} \otimes \Sigma^{-1}) \text{vec}(Y - \Phi^* X) \\
&= \text{trace}\{(Y - \Phi^* X)' \Sigma^{-1} (Y - \Phi^* X)\}
\end{aligned}$$

여기서 유의할 점은 S 는 스칼라이고 Φ^* 는 행렬이라는 것이다. 따라서 S 를 최소로 하는 Φ^* 를 구하는 것보다는, S 를 최소로 하는 $\text{vec}(\Phi^*)$ 를 구하는 것이 더 쉬운 방법일 것이다. 이 점에 유의하여 S 를 다음과 같이 전개하자.

$$\begin{aligned}
S &= \text{vec}(Y)' (I_{T-p} \otimes \Sigma^{-1}) \text{vec}(Y) - 2 \text{vec}(\Phi^*)' (X \otimes I_d) (I_{T-p} \otimes \Sigma^{-1}) \text{vec}(Y) \\
&\quad + \text{vec}(\Phi^*)' (X \otimes I_d) (I_{T-p} \otimes \Sigma^{-1}) (X' \otimes I_d) \text{vec}(\Phi^*)
\end{aligned}$$

이 제곱합 S 를 $\text{vec}(\Phi^*)$ 로 편미분하면 다음과 같다.

$$\frac{\partial S}{\partial \text{vec}(\Phi^*)} = 2(X' X \otimes \Sigma^{-1}) \text{vec}(\Phi^*) - 2(X \otimes \Sigma^{-1}) \text{vec}(Y)$$

따라서 정규방정식은 다음과 같다.

$$(X' X \otimes \Sigma^{-1}) \text{vec}(\Phi^*) = (X \otimes \Sigma^{-1}) \text{vec}(Y)$$

이 식을 풀면 다음과 같은 $\text{vec}(\Phi^*)$ 의 최소제곱추정량을 얻는다.

$$\begin{aligned}
& \text{vec}(\tilde{\Phi}^*) \\
&= (X' X \otimes \Sigma^{-1})^{-1} (X \otimes \Sigma^{-1}) \text{vec}(Y) \\
&= \{(X' X)^{-1} \otimes \Sigma\} (X \otimes \Sigma^{-1}) \text{vec}(Y) \\
&= \{(X' X)^{-1} X \otimes I_d\} \text{vec}(Y) \\
&= \text{vec}\{Y X' (X X')^{-1}\}
\end{aligned}$$

즉, Φ^* 의 최소제곱추정량은 다음과 같다.

$$\tilde{\Phi}^* = Y X' (X X')^{-1}$$

좀더 엄격하게 말하면 이 추정량이 최소제곱추정량이기 위해서는 제곱합 S 의 헤시안(Hessian)행렬이 비음치행렬(nonnegative-definiteness matrix)임을 보여야 한다. 제곱합 S 의 헤시안행렬은 다음과 같다.

$$\frac{\partial^2 S}{\partial \text{vec}(\Phi^*) \partial \text{vec}(\Phi^*)'} = 2(X' X \otimes \Sigma^{-1})$$

이 헤시안행렬은 비음치행렬이다. 특히 다변량시계열데이터 $\{y_t\}$ 가 연속확률분포를 가지면 확률 1로(with probability 1) $X X'$ 가 정칙(nonsingular)이다. 이 경우 최소제곱추정행렬 $\tilde{\Phi}^*$ 는 유일하다.

SAS/ETS에서 VAR모형을 최소제곱추정하기 위해서 SAS매크로 %AR를 사용할 수 있다. 한 예로써 다변량시계열데이터 VAR31A를 최소제곱추정법으로 3변량VAR(2)모형에 적합시키기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM7-8.PRG에 수록되어 있다).

```

/* */
/*          EXAM7-8.PRG          */
/* */
/*      USING SAS MACRO %AR      */
/*      FOR LSE OF VAR MODELS    */
/*      W/  CONSTANT TERM        */
/*      DATASET = VAR31A         */
/* */

/* ***** */
/* SAS MACAUTOS -- Macro Library */
/* */
/* NAME      : AR                */
/* TITLE     : Vector Autoregressive Process for PROC MODEL */
/* PRODUCT   : ETS              */
/* KEYS      : ARIMA, Vector Autoregression */
/* */
/* UPDATE: 06SEP88 */
/* ***** */

options nosource2;

/* -----
*
*      Macro %AR to generate PROC MODEL code for autoregressive
*      and vector autoregressive error processes.
*
* USAGE:
*
*      There are two cases:
*
*      %AR( name, nlag [, endlolist [, laglist] ] [,M=] [,TYPE=] )
*
*      name      : prefix for %AR to use in constructing names
*                  of variables needed to define the AR process.

```


*
 * If the 'endolist' is not specified then the
 * endogenous list defaults to 'name' and 'name'
 * must be the name of the equation to which the
 * AR error process is to be applied.
 *
 * 'name' can not exceed 8 characters.
 *
 * nlag : the order of the AR process.
 *
 * enodlist : the list of equations to which the AR process
 * is to be applied.
 *
 * If more than one name is given an unrestricted
 * vector process is created with the structural
 * residuals of all the equations included as
 * regressors in each of the equations.
 *
 * If not specified, 'endolist' defaults to 'name'.
 *
 * laglist : the list of lags at which the AR terms are to be
 * added. The coefficients of the terms at lags not
 * listed are set to zero.
 *
 * All of the listed lags must be less than or equal
 * to 'nlag', and there must be no duplicates.
 *
 * If not specified, the 'laglist' defaults to
 * all lags 1 through 'nlag'.
 *
 * M=method : specifies the estimation method to implement.
 * The values of M= are: M=CLS, M=ULS, ML=ML.
 * M=CLS: conditional least squares estimates.
 * M=ULS: unconditional least squares estimates.
 * ML=ML: maximum likelihood estimates.
 *
 * M=CLS is the default.
 *
 * Only M=CLS is allowed when more than one equation

```

*           is specified. The ULS and ML methods are not
*           supported for vector AR models by %AR.
*
*   TYPE=V   : specifies that the AR process is to be applied
*             to the endogenous variables themselves instead
*             of to the structural residuals of the equations.
*
*   An alternative usage of %AR is allowed to impose
*   restrictions on a vector AR process by calling %AR
*   several times to specify different AR terms and lags
*   for different equations.
*
*   The first call has the form:
*
*   %AR( name, nlag, endolist, DEFER )
*
*   name      : prefix for %AR to use in constructing names
*             of variables needed to define the VAR process.
*
*   nlag      : the order of the AR process.
*
*   endolist  : the list of equations to which the AR process
*             is to be applied.
*
*   DEFER    : a special keyword that tells %AR to not generate
*             the AR process but to wait for further information
*             specified in later %AR calls for the same 'name'.
*
*   The later calls have the form:
*
*   %AR( name, eqlist [, varlist] [, laglist] [,TYPE=] )
*
*   name      : same as in the first call.
*
*   eqlist    : the list of equations to which the specifications
*             in this %AR call are to be applied.
*
*           Only names specified in the 'endolist' of the
*           first call for the 'name' can appear in 'eqlist'.

```

```

*
*   varlist      : the list of equations whose lagged structural
*                 residuals are to be included as regressors in
*                 the 'eqlist' equations.
*
*                 Only names specified in the 'endolist' of the
*                 first call for 'name' can appear in 'varlist'.
*
*                 If not specified 'varlist' defaults to 'endolist'.
*
*   laglist     : the list of lags at which the AR terms are to be
*                 added. The coefficients of the terms at lags not
*                 listed are set to zero.
*
*                 All of the listed lags must be less than or equal
*                 to 'nlag', and there must be no duplicates.
*
*                 If not specified, the 'laglist' defaults to
*                 all lags 1 through 'nlag'.

```

* NOTES:

- ```

*
* 1. The calls to %AR must come AFTER all of the
* equations that the process applies to.
*
* 2. Use a short 'name' for the process if parameter
* estimates are to be written to an output data set.
* %AR tries to construct parameter names <= 8 characters,
* but this is limited by the length of 'name', which
* is used as a prefix for the ar parameter names.
*
* 3. See the SAS/ETS Guide, Version 6 Edition.

```

\* EXAMPLES:

```

* PROC MODEL DATA=A;
* PARS A B C;

```

```

* Y = A + B*X1 + C*X2:
* %AR(Y, 3)
* FIT Y;
*
* PROC MODEL DATA=A;
* PARMS Y1_INT Y1_X1 Y2_INT Y2_X2;
* Y1 = Y1_INT + Y1_X1*X1;
* Y2 = Y2_INT + Y2_X2*X2;
* %AR(VEC_AR, 2, Y1 Y2)
* FIT Y1 Y2;
*
*
*
* -----*/
%MACRO AR(ARG1, ARG2, ARG3, ARG4, M=, TYPE=);

%LOCAL PNAME NLAG ENDOLIST NENDO METHOD NLAGZLAG LAGZLAG TYPECODE;
%LOCAL EQLIST NEQLIST VARLIST NVARLIST LAGLIST NLAGLIST HATNAME;
%LOCAL CALLTYPE VAR VAR2 PARMNAME PARMLIST;
%LOCAL I J K II JJ KK L L2 L3 N S1 S2;

%LET PNAME = %UPCASE(%STR(&ARG1));

%IF %LENGTH(&PNAME) <= 0 | %LENGTH(&PNAME) > 8 %THEN %DO;
 %PUT ERROR: Invalid AR process name "&PNAME". ;
 %GOTO EXIT;
%END;

%GOTO DEFSUBS;
%XRETURN: %DO; %END;

%GLOBAL &PNAME;

%IF %INDEX(0123456789, %SUBSTR(%QUOTE(&ARG2), 1, 1))>0 %THEN %DO;
 %LET CALLTYPE=D;

%LET NLAG = %EVAL(&ARG2);
%IF &NLAG<=0 | &NLAG>99 %THEN %DO;
 %PUT ERROR: Invalid lag length "&NLAG" for AR "&PNAME". ;

```

```

%GOTO EXIT;
%END;
%LET ENDOLIST = %UPCASE(&ARG3);
%IF %LENGTH(%QUOTE(&ENDOLIST))=0 %THEN %DO;
 %LET ENDOLIST = &PNAME; %LET NENDO = 1; %END;
%ELSE %LET NENDO = %_COUNT_(&ENDOLIST);
%DO I = 1 %TO &NENDO;
 %LET VAR = %SCAN(%QUOTE(&ENDOLIST), &I, %QUOTE());
 %IF %LENGTH(%QUOTE(&VAR))>40 %THEN %DO;
 %PUT ERROR: Invalid variable name "&VAR" for AR "&PNAME". ;
 %LET NENDO = -999;
 %END;
%END;
%IF &NENDO<=0 | &NENDO>99 %THEN %DO;
 %PUT ERROR: Invalid variable list for AR "&PNAME".;
 %GOTO EXIT;
%END;

%IF %UPCASE(%QUOTE(&ARG4))^=DEFER %THEN %DO;
 %LET CALLTYPE=B;
 %LET EQLIST = &ENDOLIST;
 %LET VARLIST = &ENDOLIST;
 %LET NEQLIST = &NENDO;
 %LET NVARLIST = &NENDO;
 %_LLIST_ %IF &NLAGLIST<=0 %THEN %GOTO EXIT;
%END;

%ELSE
%LET &PNAME = &NLAG, &NENDO, &ENDOLIST;

%END;

%ELSE %DO;
 %LET CALLTYPE=U;
 %LET L = %LENGTH(%QUOTE(&&&PNAME));
 %IF &L<=0 %THEN %DO;
 %PUT ERROR: The AR "&PNAME" is not defined. ;
 %GOTO EXIT;
 %END;

```

```

%LET NENDO = %SCAN(%QUOTE(&&PNAME), 2, %QUOTE(,));
%LET ENDOLIST = %SCAN(%QUOTE(&&PNAME), 3, %QUOTE(,));
%LET EQLIST = %UPCASE(&ARG2);
%IF %LENGTH(%QUOTE(&EQLIST))=0 %THEN %DO;
 %PUT ERROR: No equation list for AR " &PNAME". ;
 %GOTO EXIT;
%END;

%LET NEQLIST = % _COUNT_(&EQLIST);
%IF &NEQLIST> &NENDO %THEN %DO;
 %PUT ERROR: Too many items in equation list for AR " &PNAME". ;
 %GOTO EXIT;
%END;

%IF &NEQLIST< &NENDO %THEN
%DO I = 1 %TO &NEQLIST;
 %LET VAR = %SCAN(%QUOTE(&EQLIST), &I);
 %IF %INDEX(&ENDOLIST, &VAR) = 0 %THEN %DO;
 %PUT ERROR: Variable " &VAR" not in list for AR " &PNAME". ;
 %LET NEQLIST = -1;
 %END;
%END;

%IF &NEQLIST<=0 %THEN %GOTO EXIT;

%LET VARLIST = %UPCASE(&ARG3);
%IF %LENGTH(%QUOTE(&VARLIST))=0 %THEN %DO;
 %LET VARLIST = &ENDOLIST;
 %LET NVARLIST = &NENDO;
%END;
%ELSE %DO;
 %LET NVARLIST = % _COUNT_(&VARLIST);
 %IF &NVARLIST> &NENDO %THEN %DO;
 %PUT ERROR: Too many items in variable list for AR " &PNAME". ;
 %GOTO EXIT;
 %END;
 %DO I = 1 %TO &NVARLIST;
 %LET VAR = %SCAN(%QUOTE(&VARLIST), &I);
 %IF %INDEX(&ENDOLIST, &VAR) = 0 %THEN %DO;
 %PUT ERROR: " &VAR" not in list for AR " &PNAME". ;
 %LET NVARLIST = -1;
 %END;
 %END;
%END;

```

```

 %END;
 %END;
 %IF &NVARLIST<=0 %THEN %GOTO EXIT;
 %END;

 %LET NLAG = %SCAN(%QUOTE(&&&PNAME), 1, %QUOTE(,));
 % _LLIST_ %IF &NLAGLIST<=0 %THEN %GOTO EXIT;

 %END;

/ ***--Get the TYPE and METHOD options -----; */
%IF %LENGTH(%QUOTE(&TYPE))>0 %THEN %DO;
 %LET TYPECODE = %SUBSTR(%UPCASE(&TYPE), 1, 1);
 %IF &TYPECODE^=V & &TYPECODE^=E %THEN %DO;
 %PUT ERROR: Invalid TYPE=&TYPE for AR "&PNAME". ;
 %GOTO EXIT;
 %END;
%END;
%ELSE %LET TYPECODE=E;

%IF &TYPECODE=E %THEN %LET METHOD=CLS;
%ELSE
 %LET METHOD=;

%IF %LENGTH(%QUOTE(&M))>0 %THEN %DO;
 %IF &TYPECODE=V %THEN
 %PUT WARNING: M= option invalid for TYPE=V in AR "&PNAME". ;
 %ELSE %DO;
 %IF %LENGTH(%QUOTE(&M))<=3 %THEN %DO;
 %LET METHOD = %UPCASE(&M);
 %LET NLAGZLAG = 0;
 %END;
 %ELSE %DO;
 %LET METHOD = %SUBSTR(%UPCASE(&M), 1, 3);
 %LET NLAGZLAG = %SUBSTR(%UPCASE(&M), 4);
 %IF &NLAGZLAG=P %THEN %LET LAGZLAG = &NLAG;
 %IF &NLAGZLAG < 0 | &NLAGZLAG > %EVAL(2*&NLAG+20)
 %THEN %DO;
 %PUT ERROR: Invalid numeric value in M=&M option;
 %LET NLAGZLAG=0;

```

```

 %END;
 %IF &METHOD^=CLS %THEN %DO;
 %PUT ERROR: Invalid numeric value in M=&M option;
 %LET NLAGZLAG=0;
 %END;
%END;
%IF &METHOD^=ULS & &METHOD^=CLS & &METHOD^=ML
%THEN %DO;
%PUT ERROR: Invalid method M=&M for AR "&PNAME". M=CLS used. ;
%LET METHOD=CLS;
%END;
%END;
%END;

%IF &NENDO>1 %THEN
%IF &METHOD=ULS | &METHOD=ML %THEN %DO;
%PUT ERROR: Method M=&M not supported for vector process. M=CLS used.;
%LET METHOD=CLS;
%END;

/* ---Determin best from for construction of parameter names-----*/
%LET L = %LENGTH(&PNAME);
%LET L2 = %LENGTH(&NLAG);
%LET L3 = %LENGTH(&NENDO);
%LET S2 =;
%IF &NENDO>1 %THEN
%IF &L2=1 & &L3=1 %THEN %DO;
%IF &L=1 %THEN %LET S1 = _L;
%ELSE %IF &L=2 %THEN %LET S1 = _;
%ELSE %IF &L=3 %THEN %LET S1 =;
%ELSE
%LET S1 = _L;
%END;
%ELSE %IF &L2=2 & &L3=1 %THEN %DO;
%IF &L=1 %THEN %LET S1 = _;
%ELSE %IF &L=2 %THEN %LET S1 =;
%ELSE
%LET S1 = _L;
%END;
%ELSE %IF &L2=1 & &L3=2 %THEN %DO;
%IF &L=1 %THEN %LET S1 =;

```



```

%ELSE %LET S1 = _L;
%END;
%ELSE %DO;
%IF &L<5 %THEN %DO; %LET S1 = _L; %LET S2 = -; %END;
%ELSE %IF &L=5 %THEN %DO; %LET S1 = L; %LET S2 = -; %END;
%ELSE %IF &L=6 %THEN %DO; %LET S1 = ; %LET S2 = -; %END;
%ELSE %DO;
%PUT ERROR: The name "&PNAME" is too long for AR.;
%GOTO EXIT;
%END;
%END;
%ELSE
%IF &L2=1 %THEN %DO;
%IF &L<6 %THEN %LET S1 = _L;
%ELSE %IF &L=6 %THEN %LET S1 = _;
%ELSE
%LET S1 = _L;
%END;
%ELSE %DO;
%IF &L<5 %THEN %LET S1 = _L;
%ELSE %IF &L=5 %THEN %LET S1 = _;
%ELSE
%LET S1 = -L;
%END;

/ * ----- */

%IF &NENDO > 1 %THEN %LET HATNAME = &PNAME;
%ELSE
%LET HATNAME = _PRED_;

/ *** --Generate needed code-----; */

%IF &CALLTYPE=D | &CALLTYPE=B %THEN %DO;

%* --insert assingment statement for temporary PRED. vars---;
%IF &TYPECODE=E %THEN
%DO I = 1 %TO &NENDO;
%LET VAR = %SCAN(%QUOTE(&ENDOLIST), &I);
/ * &HATNAME. _&VAR = ZLAG0(PRED.&VAR); */
&HATNAME. _&VAR = PRED. &VAR;
%END;

```

```

%END;

%IF &CALLTYPE=U | &CALLTYPE=B %THEN %DO;

%DO I = 1 %TO &NEQLIST;

/ *** -- insert assingment statement for endogenous variable --- */
%LET VAR = %SCAN(%QUOTE(&EQLIST), &I);
%IF &NEQLIST=&NENDO %THEN %LET II = &I;
%ELSE %DO JJ = 1 %TO &NENDO;
%IF &VAR=%SCAN(%QUOTE(&ENDOLIST), &JJ) %THEN %LET II=&JJ;
%END;
%LET PARMLIST = ;
%LET N = 0;

%IF &METHOD=ULS | &METHOD=ML %THEN %IF &NLAG>1 %THEN %DO;
IF _OBS_ > &NLAG THEN
%END;

&VAR = PRED. &VAR

%DO J = 1 %TO &NLAGLIST;
%LET L = %SCAN(%QUOTE(&LAGLIST), &J);
%IF &L < &NLAGZLAG %THEN %LET LAGZLAG = LAG;
%ELSE %LET LAGZLAG = ZLAG;
%DO K = 1 %TO &NVARLIST;
%LET VAR2 = %SCAN(%QUOTE(&VARLIST), &K);
%IF &NVARLIST=&NENDO %THEN %LET KK = &K;
%ELSE %DO JJ = 1 %TO &NENDO;
%IF &VAR2= %SCAN(%QUOTE(&ENDOLIST), &JJ) %THEN
%LET KK=&JJ;
%END;
%LET N = %EVAL(&N+1);
%IF &NENDO>1 %THEN
%LET PARMNAME = &PNAME. &S1. &L. &S2. _&II. _&KK;
%ELSE
%LET PARMNAME = &PNAME. &S1. &L;
%LET PARMLIST = &PARMLIST &PARMNAME;

```

```

%IF &TYPECODE=E %THEN
%STR(+&PARMNAME*&LAGZLAG*&L(&VAR2-&HATNAME._&VAR2));
%ELSE
%STR(+&PARMNAME*&LAGZLAG&L(&VAR2));
%END;
%END;
%IF &NLAGZLAG > &NLAG %THEN %DO;
%IF &TYPECODE=E %THEN
%STR(+0*LAG&NLAGZLAG(&VAR2-&HATNAME._&VAR2));
%ELSE
%STR(+0*LAG&NLAGZLAG(&VAR2));
%END;
%STR(:)

/ *** --insert startup code for method=ULS or methdo=ML ---: */
%IF &METHOD=ULS | &METHOD=ML %THEN %DO;
%IF &NLAG>1 %THEN %_ULS_;
%ELSE %DO;
IF _OBS_=1 THEN
RESID.&VAR = RESID.&VAR*SQRT(1-&PARMNAME*&PARMNAME);
%IF &METHOD=ML %THEN %DO;
IF _OBS_=1 THEN
&PNAME._WGT_ =
(1-&PARMNAME*&PARMNAME)**(-1/(2*_NUSED_));
ELSE &PNAME._WGT_ = ZLAG1(&PNAME._WGT_);
RESID.&VAR = RESID.&VAR*&PNAME._WGT_;
%END;
%END;
%END;

/ *** --insert parms statements ---: */
PARMS
%LET N = 0;
%DO J = 1 %TO &NLAGLIST;
%DO K = 1 %TO &NVARLIST;
%LET N = %EVAL(&N+1);
%LET PARMNAME = %SCAN(%QUOTE(&PARMLIST),&n);
&PARMNAME 0
%END;

```

```
%END;
```

```
;
```

```
/ *** --construct label statements--- */
```

```
%LET N = 0;
```

```
%DO J = 1 %TO &NLAGLIST;
```

```
%LET L = %SCAN(%QUOTE(&LAGLIST), &J);
```

```
%DO K = 1 %TO &NVARLIST;
```

```
%LET N = %EVAL(&N+1);
```

```
%LET PARMNAME = %SCAN(%QUOTE(&PARMLIST), &N);
```

```
%LET VAR2 = %SCAN(%QUOTE(&VARLIST), &K);
```

```
LABEL &PARMNAME =
```

```
%IF &NENDO>1 %THEN %DO;
```

```
%IF %LENGTH(&PNAME)+ %LENGTH(&VAR)
+ %LENGTH(&VAR2)+%LENGTH(&K) > 20
```

```
%THEN
```

```
%STR("AR(&PNAME) &VAR: L&L,&VAR2");
```

```
%ELSE
```

```
%STR("AR(&PNAME) &VAR: LAG&L PARM FOR &VAR2");
```

```
%END;
```

```
%ELSE %STR("AR(&PNAME) &VAR LAG&L PARAMETER");
```

```
%END;
```

```
%END;
```

```
%END;
```

```
%END;
```

```
%GOTO ENDIT;
```

```
/ *** ----- */
```

```
%DEFSUBS: %DO; %END;
```

```
%MACRO _COUNT_(ARG, DEL);
```

```
%LOCAL I N D;
```

```
%IF %LENGTH(&DEL)=0 %THEN %LET D=%QUOTE(); %ELSE %LET D=&DEL;
```

```
%LET N=0;
```

```
%DO I = 1 %TO 999;
```

```
%IF %LENGTH(%SCAN(%QUOTE(&ARG), &I, &D))=0 %THEN %LET I=9999;
```

```

 %ELSE %LET N = %EVAL(&N+1);
 %END;
 &N
 %MEND _COUNT_;

%MACRO _LLIST_ ; %LOCAL I J;

 %LET LAGLIST = %UPCASE(&ARG4);
 %IF %LENGTH(&LAGLIST)>0 & %QUOTE(&LAGLIST)^=DEFER %THEN %DO;
 %LET NLAGLIST = %_COUNT_(&LAGLIST);
 %DO I = 1 %TO &NLAGLIST;
 %LET J = %SCAN(%QUOTE(&LAGLIST), &I);
 %IF &J<=0 | &J>&NLAG %THEN %DO;
 %PUT ERROR: Invalid lag value "&J" for "&PNAME";
 %LET NLAGLIST = -999;
 %END;
 %END;
 %END;
 %ELSE %DO;
 %LET NLAGLIST = &NLAG;
 %LET LAGLIST = ;
 %DO I = 1 %TO &NLAGLIST;
 %LET LAGLIST = &LAGLIST &I;
 %END;
 %END;

 %IF &NLAGLIST<=0 | &NLAGLIST>&NLAG %THEN %DO;
 %PUT ERROR: Invalid lag list for AR "&PNAME";
 %LET NLAGLIST = -1;
 %END;

%MEND _LLIST_;

/ *** -----; */

/ *** --MACRO TO GENERATE CODE FOR ULS AND ML METHODS-----; */
%MACRO _ULS_ ;

%LOCAL P K P_1 K_1 P_K J N PARMS PARM;

```

```
%MACRO _GAMMA_;
```

```
 ** GET PHI_J_J FROM REVERSED LEVINSON RECURRENCE;
```

```
 %DO K = 1 %TO &NLAG;
```

```
 &PNAME._PHI_&K = -%SCAN(%QUOTE(&PARMS), &K, ?);
```

```
 %END;
```

```
 %IF &NLAG=2 %THEN %DO;
```

```
 &PNAME._PHI_1=(&PNAME._PHI_1_&PNAME._PHI_2*&PNAME._PHI_1)
 / (1 - &PNAME._PHI_2*&PNAME._PHI_2);
```

```
 %END;
```

```
 %ELSE
```

```
 %DO P = &NLAG %TO 2 %BY -1;
```

```
 %LET P_1 = %EVAL(&P-1);
```

```
 &PNAME._TEMP_1 = 1/ (1 - &PNAME._PHI_&P*&PNAME._PHI_&P);
```

```
 %DO K = 1 %TO &P_1;
```

```
 %LET P_K = %EVAL(&P-&K);
```

```
 &PNAME._L_&K = &PNAME._TEMP_1
```

```
 (&PNAME._PHI_&K - &PNAME._PHI_&P&PNAME._PHI_&P_K);
```

```
 %END;
```

```
 %DO K = 1 %TO &P_1; &PNAME._PHI_&K = &PNAME._L_&K; %END;
```

```
 %END;
```

```
IF _ITER_>0 THEN
```

```
IF ABS(&PNAME._PHI_1)>=1
```

```
%DO K=2 %TO &NLAG; | ABS(&PNAME._PHI_&K)>=1 %END;
```

```
THEN PUT 'WARNING: AT' _ITER_ =
```

```
 ' THE %AR PROCESS ' &PNAME ' IS UNSTABLE.'
```

```
 /' %DO K=1 %TO &NLAG; &PNAME._PHI_&K = %END; ;
```

```
&PNAME._GAM_0 = 1
```

```
%DO K=1 %TO &NLAG; / (1 - &PNAME._PHI_&K*&PNAME._PHI_&K) %END; ;
```

```
%IF &METHOD=ML %THEN %DO;
```

```
 &PNAME._WGT_ = &PNAME._GAM_0;
```

```
%END;
```

```

%DO P = 1 %TO &NLAG; %LET P_1 = %EVAL(&P-1);

%** GET PHI_P_K FROM FORWARD LEVINSON RECURRENCE:
 %IF &P=2 %THEN %DO;
 &PNAME._PHI_1 = &PNAME._PHI_1
 + &PNAME._PHI_2*&PNAME._PHI_1;
 %END;
 %ELSE %IF &P>2 %THEN %DO;
 %DO K = 1 %TO &P_1;
 %LET P_K = %EVAL(&P-&K);
 &PNAME._L_&K = &PNAME._PHI_&K
 + &PNAME._PHI_&P*&PNAME._PHI_&P_K;
 %END;
 %DO K=1 %TO &P_1; &PNAME._PHI_&K=&PNAME._L_&K; %END;
 %END;

%** COMPUTE GAMMA_P;
 &PNAME._GAM_&P =
 %DO K = 1 %TO &P; %LET P_K = %EVAL(&P-&K);
 - &PNAME._PHI_&K*&PNAME._GAM_&P_K
 %END; ;

%** COMPUTE determinant-----;
 %IF &METHOD=ML %THEN %DO;
 &PNAME._WGT_ = &PNAME._WGT_*(&PNAME._GAM_0
 %DO K = 1 %TO &P;
 + &PNAME._PHI_&K*&PNAME._GAM_&K
 %END;);
 %END;
%END;

%IF &METHOD=ML %THEN %DO;
 &PNAME._WGT_ = &PNAME._WGT_ ** (1/ (2*_NUSED_));
%END;
%MEND _GAMMA_;

/ ** *-----; */

%LET PARMS =;

```

```

%LET N = 1;
%DO P = 1 %TO &NLAGLIST;
 %LET K = %SCAN(%QUOTE(&LAGLIST), &P);
 %DO J = 1 %TO %EVAL(&K-&N);
 %LET PARMS =&PARMS?0; %LET N = %EVAL(&N+1); %END;
 %LET PARMS =&PARMS? %SCAN(%QUOTE(&PARMLIST), &P);
 %LET N = %EVAL(&N+1);
 %END;
%DO J = &N %TO &NLAG;
 %LET PARMS =&PARMS?0; %LET N = %EVAL(&N+1); %END;
%IF &N^= %EVAL(&NLAG+1) %THEN
%PUT ERROR: AR &PNAME ULS BUG IN SET UP OF PARMS. N=&N NLAG=&NLAG;

IF _OBS_ <= &NLAG THEN DO;

 IF _OBS_ = 1 THEN DO; / *-- Initialize the Kalman filter --*/
 % _GAMMA_
 &PNAME._R_ = &PNAME._GAM_0;
 &PNAME._L_1 = &PNAME._GAM_1;
 &PNAME._K_1 = &PNAME._GAM_1;
 %DO K=1 %TO &NLAG; &PNAME._Z_&K = 0; %END;
 %DO K=2 %TO &NLAG;
 %LET K_2 = %EVAL(&K-2);
 &PNAME._L_&K = &PNAME._GAM_&K_2;
 &PNAME._K_&K = &PNAME._GAM_&K_2;
 %END;
 END;

 ELSE DO; / *-- Update the Kalman filter --*/

 %* UPDATE R:
 &PNAME._R_ = ZLAG1(&PNAME._R_
 - &PNAME._L_1* &PNAME._L_1/ &PNAME._R_);

 %* UPDATE THE ZS:
 %DO K=&NLAG %TO 2 %BY -1; %LET K_1 = %EVAL(&K-1);
 &PNAME._Z_&K = ZLAG1(
 &PNAME._Z_&K_1+&PNAME._K_&K* &PNAME._TEMP_1);
 %END;

```



```

&PNAME._Z_1 = ZLAG1(
 %DO K=1 %TO &NLAG;
 %LET PARM = %SCAN(%QUOTE(&PARMS), &K, ?);
 %IF %QUOTE(&PARM)^=0 %THEN
 %DO: + &PARM * &PNAME._Z_ &K %END;
 %END;
 + &PNAME._K_1 * &PNAME._TEMP_1);

%* UPDATE THE KS;
%DO K=&NLAG %TO 2 %BY -1; %LET K_1 = %EVAL(&K-1);
&PNAME._K_ &K = ZLAG1(&PNAME._K_ &K
 - &PNAME._TEMP_2 * &PNAME._L_ &K_1);
%END;
&PNAME._K_1 = ZLAG1(
 &PNAME._K_1 - &PNAME._TEMP_2 * &PNAME._TEMP_3);

%* UPDATE THE LS;
%DO K=&NLAG %TO 2 %BY -1; %LET K_1 = %EVAL(&K-1);
&PNAME._L_ &K = ZLAG1(&PNAME._L_ &K_1
 - &PNAME._TEMP_2 * &PNAME._K_ &K);
%END;
&PNAME._L_1 = ZLAG1(&PNAME._TEMP_3
 - &PNAME._TEMP_2 * &PNAME._K_1);
END;

&VAR = PRED.&VAR + &PNAME._Z_1;
&PNAME._TEMP_1 = -RESID.&VAR / &PNAME._R_;
RESID.&VAR = RESID.&VAR / SQRT(&PNAME._R_);

&PNAME._TEMP_2 = &PNAME._L_1 / &PNAME._R_;
&PNAME._TEMP_3 =
%DO K=1 %TO &NLAG;
 %LET PARM = %SCAN(%QUOTE(&PARMS), &K, ?);
 %IF %QUOTE(&PARM)^=0 %THEN %DO: + &PARM * &PNAME._L_ &K %END;
%END; ;
END;

%IF &METHOD=ML %THEN %DO;
IF _OBS_ > 1 THEN &PNAME._WGT_ = ZLAG1(&PNAME._WGT_);

```

```

RESID.&VAR = RESID.&VAR* &PNAME._WGT_ ;
%END;

%MEND _ULS_ ;
%GOTO XRETURN;

/ ***-----; */

%EXIT::
:ABORT: *'%AR' ERROR INSERTED ABORT STATEMENT;
%ENDIT: %MEND AR;

/ * * /
/ * APPLYING %AR TO A VAR(2) MODEL * /
/ * * /
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=60;
DATA EXAM78;
SET TSA4DATA.VAR31A;
RUN;
PROC MODEL DATA=EXAM78;
PARMS A1-A3;
Y1 = A1;
Y2 = A2;
Y3 = A3;
%AR(Y, 2, Y1 Y2 Y3, TYPE=V)
FIT Y1 Y2 Y3;
RUN;

```

이 SAS프로그램의 앞부분은 SAS매크로 %AR를 설명하는 주석문들과 다음과 같은 OPTIONS스테이트먼트로 구성되었다.

```
OPTIONS NOSOURCES2 ;
```

이 OPTIONS스테이트먼트의 옵션 NOSOURCE2는 %INCLUDE스테이트먼트에 의해  
포함된 2차출처스테이트먼트들(secondary source statements)을 LOG윈도우에 나타  
내지 않도록 한다. 반대의 기능을 하는 옵션은 SOURCE2이다. 이와 마찬가지로 SAS출처  
스테이트먼트들(SAS source statements)을 LOG윈도우에 나타내거나, 반대로 나타내지  
않기 위해서는 각각 옵션 SOURCE와 옵션 NOSOURCE를 사용한다. 이 옵션들에 대해서

는 [Lang6.03] (p. 518)을 참조하라.

SAS매크로 %AR는 다음 스테이트먼트부터 시작한다.

```
% MACRO AR(ARG1, ARG2, ARG3, ARG4, ARG5, M=, TYPE=);
```

또한, SAS매크로 %AR는 다음 스테이트먼트로 끝난다.

```
% ENDIT: % MEND AR;
```

이 두 스테이트먼트들 사이의 스테이트먼트들이 SAS매크로 %AR를 구성한다. 물론 이 SAS매크로를 이해하거나 작성할 수 있다면 좋을 것이다. 그러나 시계열분석을 하기 위해서 SAS매크로에 관한 내용을 반드시 알아야 하는 것은 아니다. SAS매크로에 관한 기본적인 내용은 『SAS Guide to Macro Processing』(1990, pp. 3~22)을 참조하라. SAS매크로 %AR를 사용하기 위해서는 이 SAS매크로 %AR가 적용되는 PROC스텝 앞에 SAS매크로 %AR를 놓아야 한다.

SAS매크로 %AR는 SAS/ETS의 프로시저인 PROC MODEL과 함께 사용되어 AR모형과 VAR모형을 분석하는 데 이용된다. 따라서 SAS매크로 %AR를 이용하기 위해서는 PROC MODEL에 대해서 알고 있어야 한다. PROC MODEL은 하나 이상의 비선형확률방정식들(nonlinear stochastic equations)로 구성되어 있는 연립방정식시스템의 구조를 이해하고, 모수들을 추정하며, 미래값들을 예측하는 데 사용되는 매우 다양한 기능을 가진 방대한 프로시저이다. 이 PROC MODEL은 계량경제모형, 재무관리모형(financial planning model), 다변량비선형회귀모형, 확률연립방정식모형(simultaneous equations model), 이산형동적시뮬레이션(discrete-time dynamic simulation) 등을 분석하는 데 사용된다. PROC MODEL의 기능들을 자세히 설명하려면 많은 지면을 필요로 할 뿐 아니라 본서의 목적과도 어울리지 않는다. 따라서 이 절에서는 SAS매크로 %AR를 사용하는 데 불편이 없을 정도만을 설명하고자 한다. PROC MODEL에 관한 자세한 내용은 [ETS6.]의 제13장을 참조하라.

SAS매크로 %AR는 PROC MODEL과 함께 사용되어 AR모형들에 관한 프로그래밍 스테이트먼트들을 생성하는 데 사용된다. 확률방정식들의 오차항들, 또는 내생변수(endogenous variables)들이 AR모형, 또는 VAR모형을 갖는 경우에 SAS매크로 %AR를 적용한다. 현재 우리들의 문제는 VAR모형을 최소제곱추정하는 것이므로 내생변수들이 VAR모형을 갖는 경우이다. 이 경우의 SAS매크로 %AR의 사용양식은 다음과 같다.

```
%AR(다변량변수, 차수, 단변량변수들, TYPE=V)
```

여기서 다변량변수는 다변량시계열데이터의 이름, 차수는 VAR모형의 차수, 단변량변수들은 다변량시계열데이터를 구성하는 단변량시계열데이터들의 이름들, 'TYPE=V'는 확률방정식들의 오차항들이 아닌 내생변수들이 VAR모형을 가짐을 나타낸다. 이 SAS프로그램에서는 3개의 단변량시계열데이터들  $\{y_{1t}\}, \{y_{2t}\}, \{y_{3t}\}$ 를 하나의 3변량시계열데이터  $\{y_t =$

$(y_{1t}, y_{2t}, y_{3t})'$ 로 간주해서 3변량VAR(2)모형을 최소제곱추정한다.

이 SAS프로그램의 PROC MODEL스태이트먼트는 SAS데이터세트 EXAM78에 PROC MODEL을 적용하기 위한 것이고, PARM스태이트먼트는 A1, A2, A3가 모수들임을 선언하는 것이다. 또한 'Y1=A1', 'Y2=A2', 'Y3=A3'는 이 3변량VAR(2)모형이 상수항  $\nu = (A1, A2, A3)'$ 을 포함하고 있음을 나타낸다. 따라서 이 SAS프로그램은 다음과 같은 3변량VAR(2)모형을 최소제곱추정하기 위한 것이다.

$$y_t = \nu + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + v_t$$

단,  $\nu = (A1, A2, A3)'$

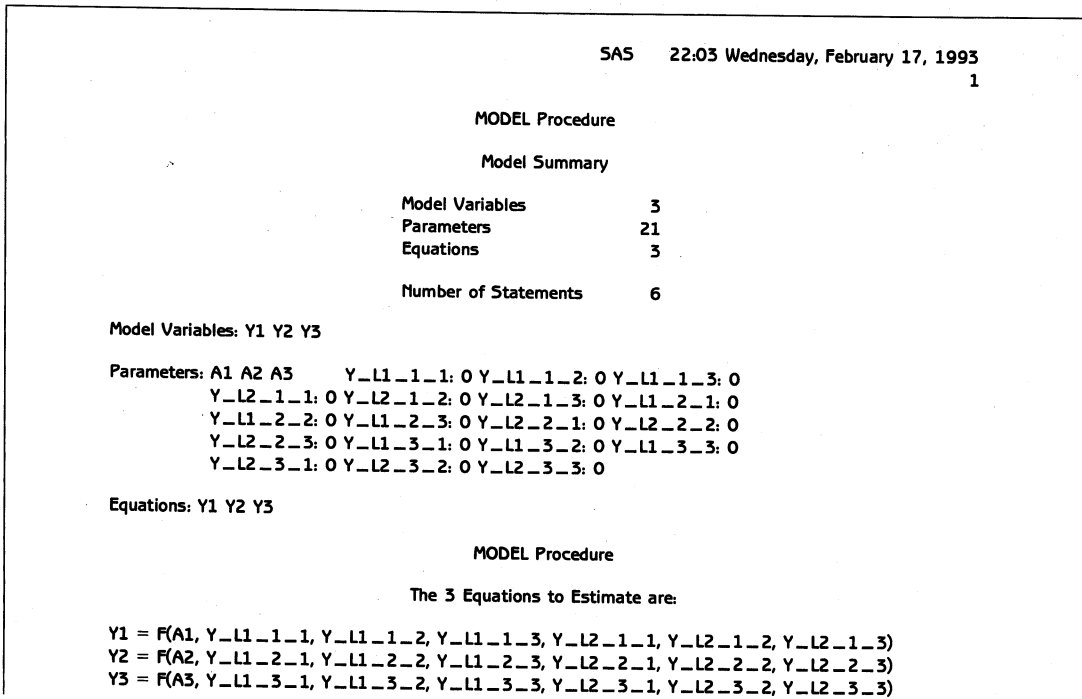
이 SAS프로그램을 수행한 결과가 그림 7.8이다. 그림 7.8로부터 최소제곱추정벡터와 최소제곱추정행렬들이 다음과 같음을 알 수 있다.

$$\tilde{\nu} = (0.0315, 0.1044, -0.0831)'$$

$$\tilde{\Phi}_1 = \begin{bmatrix} -0.5791 & -0.0764 & -0.1181 \\ -0.0558 & 0.3957 & 0.5969 \\ 0.0576 & 0.4773 & -0.2784 \end{bmatrix}$$

$$\tilde{\Phi}_2 = \begin{bmatrix} 0.0865 & 0.1739 & -0.0329 \\ 0.0258 & -0.1077 & -0.0018 \\ 0.2203 & -0.0949 & 0.0037 \end{bmatrix}$$

이 최소제곱추정행렬들이 제7.3.1소절의 Yule-Walker추정행렬들과 크게 다르지 않음을 알 수 있다.



MODEL Procedure  
OLS Estimation

OLS Estimation Summary

Dataset Option            Dataset  
DATA=                    EXAM78

Parameters Estimated        21

Minimization Summary

Method                    GAUSS  
Iterations                 1

Final Convergence Criteria

R                            0  
PPC                         4.27E-14  
RPC(Y\_L1\_2\_3)            596902  
Object                     0.39683326  
Trace(S)                  3.29241222  
Objective Value          3.13876632

Observations Processed

Read                        150  
Solved                      150

MODEL Procedure  
OLS Estimation

Nonlinear OLS Summary of Residual Errors

| Equation | DF Model | DF Error | SSE       | MSE     | Root MSE | R-Square | Adj R-Sq |
|----------|----------|----------|-----------|---------|----------|----------|----------|
| Y1       | 7        | 143      | 172.04729 | 1.20313 | 1.09687  | 0.3951   | 0.3697   |
| Y2       | 7        | 143      | 160.79911 | 1.12447 | 1.06041  | 0.4432   | 0.4198   |
| Y3       | 7        | 143      | 137.96855 | 0.96482 | 0.98225  | 0.3245   | 0.2961   |

Nonlinear OLS Parameter Estimates

| Parameter | Estimate   | Approx. Std Err | T Ratio | Approx. Prob> T | Label                      |
|-----------|------------|-----------------|---------|-----------------|----------------------------|
| A1        | 0.031480   | 0.09143         | 0.34    | 0.7311          |                            |
| A2        | 0.104445   | 0.08859         | 1.18    | 0.2393          |                            |
| A3        | -0.083079  | 0.08188         | -1.01   | 0.3120          |                            |
| Y_L1_1_1  | -0.579055  | 0.08470         | -6.84   | 0.0001          | AR(Y) Y1: LAG1 PARM FOR Y1 |
| Y_L1_1_2  | -0.076353  | 0.09299         | -0.82   | 0.4130          | AR(Y) Y1: LAG1 PARM FOR Y2 |
| Y_L1_1_3  | -0.118136  | 0.09594         | -1.23   | 0.2202          | AR(Y) Y1: LAG1 PARM FOR Y3 |
| Y_L2_1_1  | 0.086527   | 0.08645         | 1.00    | 0.3186          | AR(Y) Y1: LAG2 PARM FOR Y1 |
| Y_L2_1_2  | 0.173888   | 0.07937         | 2.19    | 0.0301          | AR(Y) Y1: LAG2 PARM FOR Y2 |
| Y_L2_1_3  | -0.032922  | 0.10520         | -0.31   | 0.7548          | AR(Y) Y1: LAG2 PARM FOR Y3 |
| Y_L1_2_1  | -0.055827  | 0.08189         | -0.68   | 0.4965          | AR(Y) Y2: LAG1 PARM FOR Y1 |
| Y_L1_2_2  | 0.395652   | 0.08990         | 4.40    | 0.0001          | AR(Y) Y2: LAG1 PARM FOR Y2 |
| Y_L1_2_3  | 0.596902   | 0.09275         | 6.44    | 0.0001          | AR(Y) Y2: LAG1 PARM FOR Y3 |
| Y_L2_2_1  | 0.025819   | 0.08357         | 0.31    | 0.7578          | AR(Y) Y2: LAG2 PARM FOR Y1 |
| Y_L2_2_2  | -0.107715  | 0.07674         | -1.40   | 0.1626          | AR(Y) Y2: LAG2 PARM FOR Y2 |
| Y_L2_2_3  | -0.0017518 | 0.10171         | -0.02   | 0.9863          | AR(Y) Y2: LAG2 PARM FOR Y3 |
| Y_L1_3_1  | 0.057559   | 0.07585         | 0.76    | 0.4492          | AR(Y) Y3: LAG1 PARM FOR Y1 |
| Y_L1_3_2  | 0.477304   | 0.08328         | 5.73    | 0.0001          | AR(Y) Y3: LAG1 PARM FOR Y2 |
| Y_L1_3_3  | -0.278401  | 0.08592         | -3.24   | 0.0015          | AR(Y) Y3: LAG1 PARM FOR Y3 |
| Y_L2_3_1  | 0.220285   | 0.07741         | 2.85    | 0.0051          | AR(Y) Y3: LAG2 PARM FOR Y1 |
| Y_L2_3_2  | -0.094873  | 0.07108         | -1.33   | 0.1841          | AR(Y) Y3: LAG2 PARM FOR Y2 |
| Y_L2_3_3  | 0.0037048  | 0.09421         | 0.04    | 0.9687          | AR(Y) Y3: LAG2 PARM FOR Y3 |

Number of Observations        Statistics for System  
Used                            150            Objective            3.1388  
Missing                         0            Objective \* N      470.8149

그림 7.8 SAS매크로 %AR를 사용한 VAR모형의 최소제곱추정

만일 VAR(p)과정  $\{y_t\}$ 의 평균벡터  $\mu$ 를 알고 있는 경우에는 SAS매크로 %AR를 사용해서 최소제곱추정을 하기 위해서는 우선 다음과 같은 평균조정된(mean-adjusted) 다변량 시계열데이터  $\{y_t^0\}$ 를 구한다.

$$y_t^0 = y_t - \mu$$

이 평균조정된 다변량시계열데이터는 다음과 같은 상수항을 포함하지 않는 VAR(p)모형을 만족한다.

$$y_t^0 = \Phi_1 y_{t-1}^0 + \Phi_2 y_{t-2}^0 + \dots + \Phi_p y_{t-p}^0 + v_t$$

이 VAR모형을 최소제곱추정하기 위해서 우선 다음과 같은 행렬들을 정의하자.

$$Y^0 := (y_{p+1}^0, y_{p+2}^0, \dots, y_T^0)$$

$$\Phi := (\Phi_1, \Phi_2, \dots, \Phi_p)$$

$$z_t := \begin{bmatrix} y_t^0 \\ y_{t-1}^0 \\ \vdots \\ y_{t-p+1}^0 \end{bmatrix}$$

$$Z := (z_p, z_{p+1}, \dots, z_{T-1})$$

앞에서와 같은 방법으로  $\Phi$ 의 최소제곱추정량을 구하면 다음과 같다.

$$\tilde{\Phi} = Y_0 Z' (ZZ')^{-1}$$

상수항을 포함하지 않는 VAR모형을 최소제곱추정하는 예로써 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM7-9.PRG에 수록되어 있다).

```

/* */
/* EXAM7-9.PRG */
/* */
/* USING SAS MACRO %AR */
/* FOR LSE OF VAR MODELS */
/* W/ O CONSTANT TERM */
/* DATASET = VAR31A */
/* */

/ ***** /
/* PUT SAS MACRO %AR HERE */
/ ***** /

```

SAS 22:23 Wednesday, February 17, 1993

1

MODEL Procedure

Model Summary

Model Variables 3  
 Parameters 18  
 Equations 3  
  
 Number of Statements 6

Model Variables: Y1 Y2 Y3

Parameters: Y\_L1\_1\_1: 0 Y\_L1\_1\_2: 0 Y\_L1\_1\_3: 0 Y\_L2\_1\_1: 0  
 Y\_L2\_1\_2: 0 Y\_L2\_1\_3: 0 Y\_L1\_2\_1: 0 Y\_L1\_2\_2: 0  
 Y\_L1\_2\_3: 0 Y\_L2\_2\_1: 0 Y\_L2\_2\_2: 0 Y\_L2\_2\_3: 0  
 Y\_L1\_3\_1: 0 Y\_L1\_3\_2: 0 Y\_L1\_3\_3: 0 Y\_L2\_3\_1: 0  
 Y\_L2\_3\_2: 0 Y\_L2\_3\_3: 0

Equations: Y1 Y2 Y3

MODEL Procedure

The 3 Equations to Estimate are:

$$Y1 = F(Y_{L1_1_1}, Y_{L1_1_2}, Y_{L1_1_3}, Y_{L2_1_1}, Y_{L2_1_2}, Y_{L2_1_3})$$

$$Y2 = F(Y_{L1_2_1}, Y_{L1_2_2}, Y_{L1_2_3}, Y_{L2_2_1}, Y_{L2_2_2}, Y_{L2_2_3})$$

$$Y3 = F(Y_{L1_3_1}, Y_{L1_3_2}, Y_{L1_3_3}, Y_{L2_3_1}, Y_{L2_3_2}, Y_{L2_3_3})$$

MODEL Procedure

OLS Estimation

OLS Estimation Summary

Dataset Option Dataset  
 DATA= EXAM79

Parameters Estimated 18

Minimization Summary

Method GAUSS  
 Iterations 1

Final Convergence Criteria

R 0  
 PPC 2.15E-14  
 RPC(Y\_L1\_2\_3) 584614.3  
 Object 0.39337047  
 Trace(S) 3.28833967  
 Objective Value 3.15680608

Observations Processed

Read 150  
 Solved 150

MODEL Procedure

OLS Estimation

Nonlinear OLS Summary of Residual Errors

| Equation | DF Model | DF Error | SSE       | MSE     | Root MSE | R-Square | Adj R-Sq |
|----------|----------|----------|-----------|---------|----------|----------|----------|
| Y1       | 6        | 144      | 172.18991 | 1.19576 | 1.09351  | 0.3946   | 0.3736   |
| Y2       | 6        | 144      | 162.36910 | 1.12756 | 1.06187  | 0.4377   | 0.4182   |
| Y3       | 6        | 144      | 138.96190 | 0.96501 | 0.98235  | 0.3196   | 0.2960   |

| Nonlinear OLS Parameter Estimates |           |                       |          |                 |                            |
|-----------------------------------|-----------|-----------------------|----------|-----------------|----------------------------|
| Parameter                         | Estimate  | Approx. Std Err       | T Ratio  | Approx. Prob> T | Label                      |
| Y_L1_1_1                          | -0.577459 | 0.08431               | -6.85    | 0.0001          | AR(Y) Y1: LAG1 PARM FOR Y1 |
| Y_L1_1_2                          | -0.071905 | 0.09181               | -0.78    | 0.4348          | AR(Y) Y1: LAG1 PARM FOR Y2 |
| Y_L1_1_3                          | -0.121840 | 0.09504               | -1.28    | 0.2019          | AR(Y) Y1: LAG1 PARM FOR Y3 |
| Y_L2_1_1                          | 0.088461  | 0.08600               | 1.03     | 0.3054          | AR(Y) Y1: LAG2 PARM FOR Y1 |
| Y_L2_1_2                          | 0.176663  | 0.07872               | 2.24     | 0.0263          | AR(Y) Y1: LAG2 PARM FOR Y2 |
| Y_L2_1_3                          | -0.037631 | 0.10399               | -0.36    | 0.7180          | AR(Y) Y1: LAG2 PARM FOR Y3 |
| Y_L1_2_1                          | -0.050531 | 0.08188               | -0.62    | 0.5381          | AR(Y) Y2: LAG1 PARM FOR Y1 |
| Y_L1_2_2                          | 0.410412  | 0.08915               | 4.60     | 0.0001          | AR(Y) Y2: LAG1 PARM FOR Y2 |
| Y_L1_2_3                          | 0.584614  | 0.09229               | 6.33     | 0.0001          | AR(Y) Y2: LAG1 PARM FOR Y3 |
| Y_L2_2_1                          | 0.032235  | 0.08351               | 0.39     | 0.7001          | AR(Y) Y2: LAG2 PARM FOR Y1 |
| Y_L2_2_2                          | -0.098509 | 0.07644               | -1.29    | 0.1996          | AR(Y) Y2: LAG2 PARM FOR Y2 |
| Y_L2_2_3                          | -0.017378 | 0.10098               | -0.17    | 0.8636          | AR(Y) Y2: LAG2 PARM FOR Y3 |
| Y_L1_3_1                          | 0.053345  | 0.07574               | 0.70     | 0.4824          | AR(Y) Y3: LAG1 PARM FOR Y1 |
| Y_L1_3_2                          | 0.465563  | 0.08248               | 5.64     | 0.0001          | AR(Y) Y3: LAG1 PARM FOR Y2 |
| Y_L1_3_3                          | -0.268627 | 0.08538               | -3.15    | 0.0020          | AR(Y) Y3: LAG1 PARM FOR Y3 |
| Y_L2_3_1                          | 0.215181  | 0.07726               | 2.79     | 0.0061          | AR(Y) Y3: LAG2 PARM FOR Y1 |
| Y_L2_3_2                          | -0.102195 | 0.07072               | -1.45    | 0.1506          | AR(Y) Y3: LAG2 PARM FOR Y2 |
| Y_L2_3_3                          | 0.016134  | 0.09342               | 0.17     | 0.8631          | AR(Y) Y3: LAG2 PARM FOR Y3 |
| Number of Observations            |           | Statistics for System |          |                 |                            |
| Used                              | 150       | Objective             | 3.1568   |                 |                            |
| Missing                           | 0         | Objective * N         | 473.5209 |                 |                            |

그림 7.9 상수항을 포함하지 않은 VAR모형의 최소제곱추정

```

/* */
/* APPLYING %AR TO A VAR(3) MODEL */
/* */

LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=60;
DATA EXAM79;
 SET TSA4DATA.VAR31A;
RUN;
PROC MODEL DATA=EXAM79;
 Y1 = 0;
 Y2 = 0;
 Y3 = 0;
 %AR(Y, 2, Y1 Y2 Y3, TYPE=V)
FIT Y1 Y2 Y3;
RUN;

```

이 SAS프로그램은 다음 3변량VAR(1)모형을 최소제곱추정하는 것이다.

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + v_t$$

이 PROC MODEL에서는 PARMS스테이트먼트가 필요하지 않다. 세 개의 식들 'Y1=0', 'Y2=0', 'Y3=0'은 이 VAR(3)모형에 상수항이 포함되지 않음을 나타내는데, 이 식들은 PROC MODEL 안에 반드시 포함되어야 한다. 이 SAS프로그램을 수행한 결과가 그림 7.9이다. 그림



7.9로부터 최소제곱추정행렬들이 다음과 같음을 알 수 있다.

$$\tilde{\Phi}_1 = \begin{bmatrix} -0.5775 & -0.0719 & -0.1218 \\ -0.0505 & 0.4014 & 0.5846 \\ 0.0533 & 0.4656 & -0.2686 \end{bmatrix}$$

$$\tilde{\Phi}_2 = \begin{bmatrix} 0.0885 & 0.1767 & -0.0376 \\ 0.0322 & -0.0985 & -0.0174 \\ 0.2152 & -0.1022 & 0.0161 \end{bmatrix}$$

SAS매크로 %AR에 관한 자세한 내용은 [ETS6.1] (pp. 68~73)을 참조하라. 여기서 유의할 점은 본절의 최소제곱추정은 초기값들이 주어졌다고 가정하는 조건부최소제곱추정이라는 것이다. SAS매크로 %AR에서는 VAR모형을 추정하는데 조건부최소제곱추정법만을 사용할 수 있고, 비조건부최소제곱추정법이나 최우추정법을 사용할 수는 없다.

SAS매크로 %AR를 사용하는 대신 SAS/IML로 프로그램을 만들어 VAR모형을 최소제곱추정할 수 있다. 최소제곱추정법을 사용해서 다변량시계열데이터 VAR31A를 VAR(2)모형에 적합시키기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM7-10.PRG에 수록되어 있다).

```

/* */
/* EXAM7-10.PRG */
/* */
/* SAS/ IML PROGRAM */
/* FOR LSE OF VAR MODELS */
/* W/ CONSTANT TERM */
/* DATASET = VAR31A */
/* */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=60;
DATA EXAM710;
 SET TSA4DATA.VAR31A;
 S1L0 = Y1;
 S1L1 = LAG(S1L0);
 S1L2 = LAG(S1L1);
 S2L0 = Y2;
 S2L1 = LAG(S2L0);
 S2L2 = LAG(S2L1);
 S3L0 = Y3;
 S3L1 = LAG(S3L0);
 S3L2 = LAG(S3L1);

```

| SAS          |           |           | 22:23 Wednesday, February 17, 1993 1 |           |           |
|--------------|-----------|-----------|--------------------------------------|-----------|-----------|
| PHI1HAT      |           |           | PHI2HAT                              |           |           |
| -0.570773    | -0.081252 | -0.09567  | 0.0929259                            | 0.1667442 | -0.022222 |
| -0.058918    | 0.396859  | 0.5894444 | 0.0235511                            | -0.105527 | -0.004964 |
| 0.0478586    | 0.4789246 | -0.298581 | 0.2122562                            | -0.089674 | -0.003657 |
| NUHAT        |           | SIGMAHAT  |                                      |           |           |
| 0.0323583    |           | 1.1460813 | -0.226805                            | -0.134972 |           |
| 0.1077593    |           | -0.226805 | 1.0852119                            | 0.3233024 |           |
| -0.059951    |           | -0.134972 | 0.3233024                            | 0.8846221 |           |
| Exiting IML. |           |           |                                      |           |           |

그림 7.10 SAS/IML을 사용한 VAR모형의 최소제곱추정

```

X0 = 1.0;
IF _N_ < 3 THEN DELETE;
DROP Y1-Y3 V1-V3;
RUN;

/* */
PROC IML;
RESET NOLOG NOCENTER;
USE EXAM710;
READ ALL VAR{S1L0 S2L0 S3L0} INTO YTR;
READ ALL VAR{X0 S1L1 S2L1 S3L1
 S1L2 S2L2 S3L2} INTO XTR;
CLOSE EXAM710;
TMP = NROW(YTR); /* TMP = T - P */
PHISTHAT = YTR' * XTR * INV(XTR' * XTR);
V = YTR' - PHISTHAT * XTR';
SIGMAHAT = V * V' / TMP;
FREE / PHISTHAT SIGMAHAT;
NUHAT = PHISTHAT[, {1}];
PHI1HAT = PHISTHAT[, {2 3 4}];
PHI2HAT = PHISTHAT[, {5 6 7}];
FREE PHISTHAT;
PRINT, PHI1HAT " " PHI2HAT,
 NUHAT " " SIGMAHAT, ;
QUIT;

```

이 SAS프로그램을 수행한 결과가 그림 7.10이다. 그림 7.10과 그림 7.8이 비슷한 결과를 보여줌을 알 수 있다.

### 7.3.3 최우추정법

물론 최우추정법을 사용해서 VAR모형을 추정할 수도 있다. 우선 VAR(p)모형을 다음과 같이 정리하자.

$$y_t - \mu = \Phi_1(y_{t-1} - \mu) + \Phi_2(y_{t-2} - \mu) + \dots + \Phi_p(y_{t-p} - \mu) + v_t$$

이후 문제를 간단히 하기 위해서 초기관찰벡터들  $y_1, y_2, \dots, y_p$ 가 고정되었다고 가정하고, 표기법은 제7.3.2소절을 따른다. 또한 다음과 같은 행렬들을 정의하자.

$$\begin{aligned} y &:= \text{vec}(Y) \\ v &:= \text{vec}(V) \\ f &:= \text{vec}(\Phi) = \text{vec}(\Phi_1, \Phi_2, \dots, \Phi_p) \end{aligned}$$

$$y_t := \begin{bmatrix} y_{tp} \\ y_{t(p-1)} \\ \vdots \\ y_{t1} \end{bmatrix}, \quad m_t := \begin{bmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{bmatrix}, \quad \mu := \begin{bmatrix} \mu \\ \mu \\ \vdots \\ \mu \\ \vdots \\ \mu \end{bmatrix}$$

여기서  $m_t$ 는  $dp \times 1$ 행렬이고,  $m$ 은  $(T-p)d \times 1$ 행렬이다. 또한 오차항  $\{v_t\}$ 는 평균벡터가 영벡터이고 공분산행렬이  $\Sigma$ 인 다변량정규분포를 갖는다고 가정한다.

오차벡터  $v$ 의 공분산행렬이  $I_{T-p} \otimes \Sigma$ 이므로  $v$ 의 결합확률밀도함수는 다음과 같다.

$$f_v(v) = (2\pi)^{-d(T-p)/2} |I_{T-p} \otimes \Sigma|^{-1/2} \exp\left\{-\frac{1}{2}v'(I_{T-p} \otimes \Sigma)^{-1}v\right\}$$

또한, 벡터  $v$ 는 다음과 같이 쓸 수 있다.

$$v = J_1(y - m) + J_2(y_t - m_t)$$

여기서  $J_1$ 은 차원이  $d \times d$ 인 부행렬들로 이루어진  $(T-p) \times (T-p)$ 블록행렬로서  $(i, j)$ 블록의 부행렬은 각각  $I_d$  ( $i=j$ 인 경우),  $-\Phi_{i-j}$  ( $1 \leq i-j \leq p$ 인 경우),  $O$  (그 이외의 경우)이다. 또한,  $J_2$ 는 차원이  $d \times d$ 인 부행렬들로 이루어진  $(T-p) \times p$ 블록행렬로서  $(i, j)$ 블록의 부행렬은 각각  $-\Phi_{p+i-j}$  ( $1-p \leq i-j \leq 0$ 인 경우),  $O$  (그 이외의 경우)이다. 따라서 다음 식이 성립한다.

$$\det\left(\frac{\partial v}{\partial y}\right) = 1$$

주어진 VAR(p)모형을 다음과 같이 쓸 수 있다.

$$Y^0 = \Phi Z + V$$

이 식을 다음과 같이 쓸 수 있다.

$$\text{vec}(Y^0) = \text{vec}(\Phi Z) + \text{vec}(V)$$

즉, 다음 식이 성립한다.

$$y = m + (Z' \otimes I_d) f + v$$

따라서  $y$ 의 결합확률밀도함수는 다음과 같다.

$$f_y(y) = (2\pi)^{-d(T-p)/2} |I_{T-p} \otimes \Sigma|^{-1/2} \exp \left[ -\frac{1}{2} \{y - m - (Z' \otimes I_d) f\}' (I_{T-p} \otimes \Sigma^{-1}) \{y - m - (Z' \otimes I_d) f\} \right]$$

따라서 대수우도함수는 다음과 같다.

$$\begin{aligned} & l \\ &= \ln f_y(y) \\ &= -\frac{d(T-p)}{2} \ln(2\pi) - \frac{T-p}{2} \ln |\Sigma| \\ &\quad - \frac{1}{2} \{y - m - (Z' \otimes I_m) f\}' (I_{T-p} \otimes \Sigma^{-1}) \{y - m - (Z' \otimes I_d) f\} \\ &= -\frac{d(T-p)}{2} \ln(2\pi) - \frac{T-p}{2} \ln |\Sigma| \\ &\quad - \frac{1}{2} \sum_{t=p+1}^T \left\{ (y_t - \mu) - \sum_{i=1}^p \Phi_i (y_{t-i} - \mu) \right\}' \Sigma^{-1} \left\{ (y_t - \mu) - \sum_{i=1}^p \Phi_i (y_{t-i} - \mu) \right\} \\ &= -\frac{d(T-p)}{2} \ln(2\pi) - \frac{T-p}{2} \ln |\Sigma| \\ &\quad - \frac{1}{2} \sum_{t=p+1}^T (y_t - \sum_{i=1}^p \Phi_i y_{t-i})' \Sigma^{-1} (y_t - \sum_{i=1}^p \Phi_i y_{t-i}) \\ &\quad + \mu' (I_d - \sum_{i=1}^p \Phi_i)' \Sigma^{-1} \sum_{i=1}^p \Phi_i (y_t - \sum_{i=1}^p \Phi_i y_{t-i}) \\ &\quad - \frac{T-p}{2} \mu' (I_d - \sum_{i=1}^p \Phi_i)' \Sigma^{-1} (I_d - \sum_{i=1}^p \Phi_i) \mu \end{aligned}$$

다음 식을 상기해 보자.

$$v = \text{vec}(V) = \text{vec}(Y^0 - \Phi Z)$$

이 식을 사용해서 우도함수와 대수우도함수를 다음과 같이 나타낼 수 있다.

$$\begin{aligned}
& f_y(y) \\
&= \{(2\pi)^d |\Sigma|\}^{-(T-p)/2} \exp\left\{-\frac{1}{2} \text{vec}(Y^0 - \Phi Z)' (I_{T-p} \otimes \Sigma^{-1}) \text{vec}(Y^0 - \Phi Z)\right\} \\
&= \{(2\pi)^d |\Sigma|\}^{-(T-p)/2} \exp\left[-\frac{1}{2} \text{trace}\{(Y^0 - \Phi Z)' \Sigma^{-1} (Y^0 - \Phi Z)\}\right] \\
l &= -\frac{T-p}{2} \ln\{(2\pi)^d |\Sigma|\} - \frac{1}{2} \text{trace}\{(Y^0 - \Phi Z)' \Sigma^{-1} (Y^0 - \Phi Z)\}
\end{aligned}$$

최우추정량을 구하기 위해서 대수우도함수의 1차 도함수들을 구하면 다음과 같다.

$$\begin{aligned}
\frac{\partial l}{\partial \mu} &= (I_d - \sum_{i=1}^p \Phi_i)' \Sigma^{-1} \sum_{t=p+1}^T \{(y_t - \sum_{i=1}^p \Phi_i y_{t-i}) - (T-p)(I_d - \sum_{i=1}^p \Phi_i) \mu\} \\
\frac{\partial l}{\partial f} &= (Z \otimes I_d) (I_{T-p} \otimes \Sigma^{-1}) \{y - m - (Z \otimes I_d) f\} \\
&= (Z \otimes \Sigma^{-1}) (y - m) - (ZZ' \otimes \Sigma^{-1}) f \\
\frac{\partial l}{\partial \Sigma} &= -\frac{T-p}{2} \Sigma + \frac{1}{2} \Sigma^{-1} (Y^0 - \Phi Z) (Y^0 - \Phi Z)' \Sigma^{-1}
\end{aligned}$$

따라서 다음 방정식들의 해가 최우추정량  $\check{\mu}$ ,  $\check{f}$ ,  $\check{\Sigma}$ 이다.

$$\begin{aligned}
\mu &= \frac{1}{T-p} (I_d - \sum_{i=1}^p \Phi_i)^{-1} \sum_{t=p+1}^T (y_t - \sum_{i=1}^p \Phi_i y_{t-i}) \\
f &= \{(ZZ')^{-1} Z \otimes I_m\} (y - m) \\
\Sigma &= \frac{1}{T-p} (Y^0 - \Phi Z) (Y^0 - \Phi Z)'
\end{aligned}$$

여기서 유의할 점은 이 세 식들 중에서 어느 한 식을 먼저 구할 수는 없다는 것이다. 첫 번째 식은  $\check{\Phi}_i$ 를, 두 번째 식은  $\check{\mu}$ 를 필요로 하며, 세 번째 식은  $\check{\Phi}$ 와  $\check{\mu}$ 를 필요로 한다. 따라서 최우추정을 하기 위해서는 이 세 식들을 반복해서(iteratively) 푸는 방법을 적용하거나 우도함수 자체를 최소화하는 알고리즘을 사용해야 할 것이다.

유감스럽게도 SAS에서는 VAR모형을 최우추정하는 프로시저를 제공하고 있지 않다. 그러나 다음 절에서 알 수 있듯이 초기관찰벡터들이 고정되었다고 가정하는 경우에는 조건부최소제곱추정의 결과가 최우추정의 결과와 크게 다르지 않기 때문에 독자가 많은 노력을 들여 최우추정량을 구하기 위한 SAS프로그램을 만들 필요는 없다고 생각한다.

## 7.4 VAR계수추정행렬들의 성질

다변량시계열분석에서 VAR계수추정량들의 유한표본 성질들(small sample properties)을 구하는 것은 매우 어려운 일이다. 따라서 이 절에서는 VAR계수추정량들의 접근성질들만을 조사하기로 한다.

### 7.4.1 최소제곱추정량의 성질

우선 제 7.3.2소절에서 다룬 최소제곱추정량의 성질을 살펴보자. 다음 조건들은 VAR계수행렬의 최소제곱추정행렬이 일치성과 점근적 정규성을 갖기 위한 충분조건들이다.

(조건 A)  $\text{plim}_{T \rightarrow \infty} \frac{1}{T-p} XX'$ 가 존재하며, 이 극한행렬을  $G$ 라 했을 때  $G$ 는 정칙(non-singular)이다.

(조건 B)  $\frac{1}{\sqrt{T}} \text{vec}(VX)'$ 는 점근적으로 다변량정규분포를 가지며, 이 점근분포의 평균벡터는 영벡터이고 공분산행렬은  $G \otimes \Sigma$ 이다.

이러한 조건들이 만족되는 다변량시계열모형들을 가정하기 위해서 다음과 같은 표준다변량백색잡음과정(standard multivariate white noise process)을 정의하자.

**정의 7.1** 다변량확률과정  $\{v_t = (v_{1t}, v_{2t}, \dots, v_{mt})'\}$ 이 표준다변량백색잡음과정이기 위해서는  $v_t$ 가 연속형 확률벡터이고,  $E(v_t) = 0$ ,  $\Sigma = E(v_t v_t')$ 는 정칙,  $s \neq t$ 이면  $v_s$ 와  $v_t$ 는 서로 독립이고, 모든  $t$ 에 대해서 다음 부등식을 만족하는 양수  $C (< \infty)$ 가 존재해야 한다.

$$E | v_{it} v_{jt} v_{kt} v_{lt} | < C \quad (i, j, k, l = 1, 2, \dots, m) \quad \blacksquare$$

한 예로  $\{v_t\}$ 가 다변량정규분포를 갖는 다변량백색잡음과정이면  $\{v_t\}$ 는 표준다변량백색잡음과정이다.

**보조정리 7.1** 작인성을 갖는  $d$ 변량시계열데이터  $\{y_t\}$ 가 다음 VAR( $p$ )모형을 만족한다고 하자.

$$y_t = \nu + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \dots + \Phi_p y_{t-p} + v_t$$

만일  $\{v_t\}$ 가 표준다변량백색잡음과정이면 이 VAR( $p$ )과정은 (조건 A)와 (조건 B)를 만족한다. ▣

보조정리 7.1의 조건이 만족되면 제 7.3.2소절에서 구한 최소제곱추정량이 일치성과 점근정규분포를 갖는다.

**정리 7.7** 작인성을 갖는  $d$ 변량시계열데이터  $\{y_t\}$ 가 보조정리 7.1의 조건을 만족하면 최소제곱추정량  $\tilde{\Phi}^* = YX'(XX')^{-1}$ 은 다음 성질들을 만족한다.

$$\begin{aligned} \text{plim}_{T \rightarrow \infty} \tilde{\Phi}^* &= \Phi^* \\ \sqrt{T} \text{vec}(\tilde{\Phi}^* - \Phi^*) &\overset{d}{\rightarrow} N(0, G^{-1} \otimes \Sigma) \end{aligned}$$

여기서  $\overset{d}{\rightarrow}$ 는 분포수렴(converge in distribution)을 뜻한다. ▣

정리 7.7을 적용해서  $\Phi^*$ 에 대한 통계적 추론을 하기 위해서는  $\Sigma$ 를 추정해야 한다. 우선 다음과 같은 추정량을 정의하자.

$$\begin{aligned} \tilde{\Sigma}_p &= \frac{1}{T-p} \sum_{t=p+1}^T \tilde{v}_t \tilde{v}_t' \\ \text{단, } \tilde{v}_t &= y_t - \tilde{\Phi}_1 y_{t-1} - \tilde{\Phi}_2 y_{t-2} - \dots - \tilde{\Phi}_p y_{t-p} \end{aligned}$$

이 공분산추정행렬은 다음과 같이 쓸 수 있다.

$$\begin{aligned} \tilde{\Sigma}_p &= \frac{1}{T-p} \tilde{V} \tilde{V}' \\ &= \frac{1}{T-p} (Y - \tilde{\Phi}^* X)(Y - \tilde{\Phi}^* X)' \\ &= \frac{1}{T-p} \{Y - YX'(XX')^{-1}X\} \{Y - YX'(XX')^{-1}X\}' \\ &= \frac{1}{T-p} Y \{I - X'(XX')^{-1}X\}' \end{aligned}$$

이 공분산추정행렬의 자유도를 조사하기 위해서 다음 식을 살펴보자.

$$\begin{aligned} &\text{trace}\{I - X'(XX')^{-1}X\} \\ &= \text{trace}\{I\} \text{trace}\{(XX')^{-1}(XX')\} \\ &= T-p - (dp+1) \end{aligned}$$

따라서 다음에 정의되는  $\check{\Sigma}_p$ 는 공분산추정행렬  $\tilde{\Sigma}_p$ 의 불편추정행렬이다.

$$\check{\Sigma}_p = \frac{T-p}{T-(d+1)p-1} \tilde{\Sigma}_p$$

**정리 7.8** 작인성을 갖는  $d$ 변량시계열데이터  $\{y_t\}$ 가 보조정리 7.1의 조건을 만족하면 다음 식들이 성립한다.

$$\text{plim}_{T \rightarrow \infty} \check{\Sigma} = \text{plim}_{T \rightarrow \infty} \tilde{\Sigma}_p = \text{plim}_{T \rightarrow \infty} \frac{1}{T-p} VV' = \Sigma \quad \text{▣}$$

표본평균벡터는 다음과 같은 성질을 갖는다.

**정리 7.9** 작인성을 갖는  $d$ 변량시계열데이터  $\{y_t\}$ 가 보조정리 7.1의 조건을 만족하면 표본평균벡터  $\bar{y}$ 는 다음 식들을 만족한다.

$$\begin{aligned} \text{plim}_{T \rightarrow \infty} \bar{y} &= \mu \\ \sqrt{T}(\bar{y} - \mu) &\xrightarrow{d} N(0, H) \\ \text{단, } H &= (I_d - \Phi_1 - \dots - \Phi_p)^{-1} \Sigma (I_d - \Phi_1' - \dots - \Phi_p')^{-1} \quad \square \end{aligned}$$

제 7.3.2 소절의 후반부에서 평균벡터  $\mu$ 를 알고 있는 경우에 VAR계수행렬들  $\Phi_1, \Phi_2, \dots, \Phi_p$ 의 최소제곱추정행렬들을 구하였다. 이 추정행렬들에서 평균벡터  $\mu$ 대신 표본평균벡터  $\bar{y}$ 를 대입한 추정행렬들을  $\hat{\Phi}_1, \hat{\Phi}_2, \dots, \hat{\Phi}_p$ 라고 하자. 즉, 다음과 같은 식들을 정의하자.

$$\begin{aligned} \hat{y}_t^0 &:= y_t - \bar{y} \\ \hat{Y} &:= (\hat{y}_{p+1}^0, \hat{y}_{p+2}^0, \dots, \hat{y}_T^0) \\ \hat{z}_t &:= \begin{bmatrix} \hat{y}_t^0 \\ \hat{y}_{t-1}^0 \\ \vdots \\ \hat{y}_{t-p+1}^0 \end{bmatrix} \\ \hat{Z} &:= (\hat{z}_p, \hat{z}_{p+1}, \dots, \hat{z}_{T-1}) \\ \hat{\Phi} &:= \hat{Y}^0 \hat{Z}' (Z Z')^{-1} \end{aligned}$$

이 추정행렬들  $\hat{\Phi}_1, \hat{\Phi}_2, \dots, \hat{\Phi}_p$ 는 다음과 같은 점근성질들을 갖는다.

**정리 7.10** 작인성을 갖는  $d$ 변량시계열데이터  $\{y_t\}$ 가 보조정리 7.1의 조건을 만족하면 다음 성질들이 만족된다.

$$\begin{aligned} \text{plim}_{T \rightarrow \infty} \hat{\Phi} &= \Phi \\ \sqrt{T} \text{vec}(\hat{\Phi} - \Phi) &\xrightarrow{d} N(0, \Gamma_z(0)^{-1} \otimes \Sigma) \\ \text{단, } \Gamma_z(0) &= E(z_t z_t') \quad \square \end{aligned}$$

### 7.4.2 Yule-Walker추정량의 성질

제 7.3.1 소절에서 다룬 Yule-Walker추정량  $\hat{\Phi}^p = (\hat{\Phi}_1^p, \hat{\Phi}_2^p, \dots, \hat{\Phi}_p^p)$ 는 다음과 같은 점근성질들을 갖는다.

**정리 7.11** 작인성을 갖는  $d$ 변량시계열데이터  $\{y_t\}$ 가 보조정리 7.1의 조건을 만족하면 다음 성질들이 만족된다.

$$\text{plim}_{T \rightarrow \infty} \hat{\Phi}^p = \Phi$$



$$\sqrt{T} \text{vec}(\hat{\Phi}^p - \Phi) \xrightarrow{d} N(0, \Gamma_s(0)^{-1} \otimes \Sigma) \quad \square$$

대표본인 경우에는 Yule-Walker추정량과 최소제곱추정량이 같은 점근성질들을 가지나, 소표본인 경우에는 Yule-Walker추정량의 편의가 최소제곱추정량의 편의보다 클 것이다. 그러나 계산량을 고려할 때 특히 VAR모형의 차수를 모르는 경우에는 Yule-Walker추정량을 사용하는 것이 유리할 수도 있다.

### 7.4.3 최우추정량의 성질

평균벡터  $\mu$ 의 최우추정벡터  $\check{\mu}$ 는 양쪽 끝의 가중행렬을 제외하고는 표본평균벡터  $\bar{y}$ 와 같다. 즉, 다음과 같은 근사식이 성립한다.

$$\begin{aligned} \check{\mu} &= (I_d - \sum_{i=1}^p \check{\Phi}_i)^{-1} \frac{1}{T-p} \sum_{t=p+1}^T (y_t - \sum_{i=1}^p \check{\Phi}_i y_{t-1}) \\ &\simeq (I_d - \sum_{i=1}^p \check{\Phi}_i)^{-1} (I_d - \sum_{i=1}^p \check{\Phi}_i) \bar{y} \\ &= \bar{y} \end{aligned}$$

따라서  $\check{\mu}$ 는 다음과 같은 성질을 갖는다.

**정리 7.12** 작인성을 갖는  $d$ 변량시계열데이터  $\{y_t\}$ 가 보조정리 7.1의 조건을 만족하면 평균벡터  $\mu$ 의 최우추정벡터  $\check{\mu}$ 는 다음 식들을 만족한다.

$$\begin{aligned} \text{plim}_{T \rightarrow \infty} \check{\mu} &= \mu \\ \sqrt{T} (\check{\mu} - \mu) &\xrightarrow{d} N(0, H) \quad \square \end{aligned}$$

최우추정행렬들  $\check{\Phi}_1, \check{\Phi}_2, \dots, \check{\Phi}_p$ 에서  $\check{\mu}$ 대신  $\bar{y}$ 를 대입하면 이 추정행렬들은  $\hat{\Phi}_1, \hat{\Phi}_2, \dots, \hat{\Phi}_p$ 와 같다. 따라서 다음과 같은 점근성질을 갖는다.

**정리 7.13** 작인성을 갖는  $d$ 변량시계열데이터  $\{y_t\}$ 가 보조정리 7.1의 조건을 만족하면 다음 성질들을 만족한다.

$$\begin{aligned} \text{plim}_{T \rightarrow \infty} \check{\Phi} &= \Phi \\ \sqrt{T} \text{vec}(\check{\Phi} - \Phi) &\xrightarrow{d} N(0, \Gamma_s(0)^{-1} \otimes \Sigma) \quad \square \end{aligned}$$

다변량백색잡음과정의 공분산행렬  $\Sigma$ 의 최우추정행렬  $\check{\Sigma}$ 는 다음과 같은 점근성질을 갖는다.

**정리 7.14** 작인성을 갖는  $d$ 변량시계열데이터  $\{y_t\}$ 가 보조정리 7.1의 조건을 만족하면 다음 성질들이 만족된다.

$$\begin{aligned} \text{plim}_{T \rightarrow \infty} \text{vec } h(\check{\Sigma}) &= \text{vec } h(\Sigma) \\ \sqrt{T} \text{vec } h(\check{\Sigma} - \Sigma) &\xrightarrow{d} N(0, 2D_m^+(\Sigma \otimes \Sigma)D_m^+) \end{aligned}$$

여기서  $D_m$ 은  $d$ 차원 복사행렬이다. 즉, 임의의  $d \times d$ 행렬  $A$ 에 대해 다음 식을 만족시키는 행렬이다.

$$\text{vec}(A) = D_m \text{vec } h(A)$$

또한,  $D_m^+$ 는  $D_m$ 의 Moore-Penrose의 일반화역행렬이다. ▣

최우추정량들은 다음 성질을 만족한다.

**정리 7.15** 작인성을 갖는  $d$ 변량시계열데이터  $\{y_t\}$ 가 보조정리 7.1의 조건을 만족하면 최우추정량들  $\check{\mu}, \check{\Phi}, \check{\Sigma}$ 는 점근적으로 독립이다. ▣

제 7.4.1소절에서 정의한 최소제곱추정량을 이용해서 다음과 같은 잔차제곱교차곱합행렬  $S(p)$ 를 정의하자.

$$S(p) = \sum_{t=p+1}^T \tilde{v}_t \tilde{v}_t'$$

따라서 다음 식이 성립한다.

$$\check{\Sigma}_p = \frac{1}{T - (d+1)p - 1} S(p)$$

만일, 다변량백색잡음과정이 다변량정규분포를 가지면 다음 성질을 만족한다.

**정리 7.16** 작인성을 갖는  $d$ 변량VAR( $p$ )과정  $\{y_t\}$ 의 다변량백색잡음과정  $\{v_t\}$ 가 다변량정규분포를 가지면  $S(p)$ 는 자유도가  $(T-p-1)$ 인 Wishart분포  $W_{T-p-1}(\Sigma)$ 를 가지며,  $\check{\Phi}^*$ 와  $S(p)$ 는 서로 독립이다. ▣

이 절에서 다룬 추정량들의 성질에 관한 자세한 내용은 Wei(1990, 제14장), Lütkepohl(1991, 제3장), Hannan(1970, 제6장)을 참조하라.

## 7.5 VAR모형의 식별

VAR모형의 차수를 정하는 것은 중요하면서도 까다로운 문제이다. 다변량시계열분석의 VAR모형의 식별은 단변량시계열분석의 AR모형의 식별보다 더 복잡하지만, VAR모형의 식별문제는 기본적으로 AR모형의 식별에 사용되는 방법들을 확장해서 해결한다. 우선 VAR모형의 차수를 좀더 명확히 정의하기 위해서 다음 VAR모형을 살펴보자.

$$y_t = \nu + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \dots + \Phi_p y_{t-p} + v_t$$

만일,  $\Phi_p \neq 0$ 이고  $\Phi_i = 0 (i > p)$ 이면 이 VAR모형의 차수를  $p$ 라 한다.

실제 문제에 있어서는 주어진 유한개의 관찰벡터들  $y_1, y_2, \dots, y_T$ 를 바탕으로 VAR모형의 차수를 결정해야 한다. 따라서 대상이 되는 가능한 VAR모형의 차수도 유한이라 가정한다. 즉, 차수의 상한  $K$ 를 사전에 정해 놓아야 한다. 물론  $K$ 는 진짜차수(true order)  $p$ 보다는 커야 하지만 지나치게 커서도 안 된다. 이 차수의 상한  $K$ 에 관한 자세한 내용은 [Choi92] (pp. 75~76)를 참조하라.

### 7.5.1 공분산추정행렬의 편의

VAR모형의 차수를 정하는 과정에서 다변량백색잡음과정의 공분산행렬  $\Sigma$ 의 추정행렬을 자주 사용하게 된다. 제 7.3절에서 정의한 공분산행렬의 추정행렬들을 다시 살펴보자. 최소제곱추정법을 사용해서 주어진 다변량시계열데이터를 VAR( $k$ )모형에 적합시켜 얻은 잔차벡터를 다음과 같이 표기하자.

$$\tilde{v}_i(k) = y_i - \tilde{\Phi}_{k,1} y_{i-1} - \tilde{\Phi}_{k,2} y_{i-2} - \dots - \tilde{\Phi}_{k,k} y_{i-k} \quad (i = k+1, k+2, \dots, T)$$

이 잔차벡터는 최우추정법에 의한 잔차들과 동일하다. 따라서 다변량백색잡음과정의 공분산행렬  $\Sigma_k$ 의 최우추정행렬은 다음과 같다.

$$\check{\Sigma}_k = \frac{1}{T-k} \sum_{i=k+1}^T \tilde{v}_i(k) \tilde{v}_i(k)'$$

이 최우추정행렬의 편의를 줄이기 위해서 다음과 같은 추정행렬을 정의하자.

$$\tilde{\Sigma}_k = \frac{1}{T-(d+1)k-1} \sum_{i=k+1}^T \tilde{v}_i(k) \tilde{v}_i(k)'$$

이 추정행렬들  $\check{\Sigma}_k$ 나  $\tilde{\Sigma}_k$ 는 최우추정법 또는 최소제곱추정법에 의한 것이므로 여러 차수에 대해서 이 추정행렬들을 구하는 데는 많은 계산량을 필요로 한다. 따라서 Yule-Walker

추정행렬  $\hat{\Sigma}_k$ 를 변형해서 쓰는 방법이 필요하다.

Choi(1994)는 다음 식을 제시했다.

$$E(\hat{\Sigma}_p) = \left(1 - \frac{dp}{T}\right) \Sigma - \frac{1}{T} \sum_{i=0}^p \sum_{j=0}^p |j-i| \Phi_i \Gamma(j-i) \Phi_j' + O\left(\frac{1}{T^2}\right)$$

따라서 다음과 같이 정의되는 공분산행렬의 추정행렬의 편의가 Yule-Walker추정량의 편의보다 작을 것이다.

$$\bar{\Sigma}_k = \left(1 - \frac{dk}{T}\right)^{-1} \left(\Sigma_k + \frac{1}{T} \sum_{i=0}^k \sum_{j=0}^k |j-i| \hat{\Phi}_{k,i} \hat{\Gamma}(j-i) \hat{\Phi}_{k,j}'\right)$$

이 식에 의해서 다음 식이 성립한다.

$$E\left(\frac{T-(d+1)p-1}{T-p} \bar{\Sigma}_p\right) = E(\check{\Sigma}_p) + O\left(\frac{1}{T^2}\right)$$

즉, 다음과 같이 정의되는 추정행렬은 최우추정행렬  $\check{\Sigma}_k$ 와 같은 점근성질들을 갖는다.

$$\Sigma_k^* := \left(1 - \frac{1}{T}\right) \left(\hat{\Sigma}_k + \frac{1}{T} \sum_{i=0}^k \sum_{j=0}^k |j-i| \hat{\Phi}_{k,i} \hat{\Gamma}(j-i) \hat{\Phi}_{k,j}'\right)$$

또한,  $\sqrt{T} \text{vec}(\check{\Sigma}_p - \Sigma)$ 와  $\sqrt{T} \text{vec}(\Sigma_p^* - \Sigma)$ 는 같은 점근분포를 갖는다. 따라서 최우추정행렬  $\check{\Sigma}_k$  대신에  $\Sigma_k^*$ 를 사용하는 것이 타당성을 갖는다.

### 7.5.2 부분자기상관계수의 확장

단변량시계열데이터의 부분자기상관계수를 다변량시계열데이터에 확장시킨 것들 중의 하나가 Tiao & Box(1981)의 부분자기회귀행렬(partial autoregression matrix)이다. 시차  $s$ 인 부분자기회귀행렬은 정상적인 다변량시계열데이터의 자기공분산행렬들을 VAR( $s$ ) 모형의 Yule-Walker방정식들에 적용시켰을 때 시차  $s$ 에서의 VAR계수행렬이다. 즉, 제 6.4.4소절에서 정의한  $\Phi_s^*$ 이다. 여기서 주의할 것은 부분자기회귀행렬은 VAR과정에서만 정의되는 것이 아니라 모든 정상적 다변량시계열데이터에서 정의된다는 점이다. 자기공분산행렬들이 주어졌을 때는 Whittle알고리즘을 사용해서 부분자기회귀행렬들  $\Phi_1^*, \Phi_2^*, \dots$ 을 쉽게 구할 수 있다. 여기서 유의할 점은 단변량시계열데이터에서와는 달리  $\Phi_s^*$ 의 각 원소가 상관계수의 성질들을 반드시 만족시키지는 않는다는 것이다.

단변량시계열데이터에서 정상시계열데이터  $\{y_t\}$ 의 시차  $s$ 인 부분자기상관계수는  $y_{t+1}, y_{t+2}, \dots, y_{t+s-1}$ 이 주어진 조건하에서  $y_t$ 와  $y_{t+s}$ 의 자기상관계수이다. 즉,  $y_t$ 와  $y_{t+s}$  각각에서  $y_{t+1}, y_{t+2}, \dots, y_{t+s-1}$ 의 선형종속성(linear dependency)을 제거한 나머지들의 상관계수이다. Heyse & Wei(1985)는 단변량시계열데이터에서 부분자기상관계수가 갖는 이러한 성질을 확장해서 다변량시계열데이터의 부분시차상관행렬(partial lag correlation matrix)을 정의했다. 정상성을 갖는 다변량시계열데이터  $\{y_t\}$ 의 부분시차상관행렬을 정의하기 위해 다음과 같은 잔차벡터들을 정의하자.

$$u_{t+s}^{s-1} := y_{t+s} - \sum_{k=1}^{s-1} A_k^{s-1} y_{t+s-k}$$

$$v_t^{s-1} := y_t - \sum_{k=1}^{s-1} B_k^{s-1} y_{t+k}$$

여기서 다변량선형회귀계수행렬들  $A_1^{s-1}, A_2^{s-1}, \dots, A_{s-1}^{s-1}$ 과  $B_1^{s-1}, B_2^{s-1}, \dots, B_{s-1}^{s-1}$ 은 각각  $E(\|u_{t+s}^{s-1}\|^2)$ 과  $E(\|v_t^{s-1}\|^2)$ 을 최소화하는 행렬들이다. 이 잔차벡터들  $u_{t+s}^{s-1}$ 과  $v_t^{s-1}$ 의 상관행렬을 시차  $s$ 의 부분시차자기상관행렬이라 하고  $P_H(s)$ 라 쓴다.

잔차벡터들  $u_{t+s}^{s-1}$ 과  $v_t^{s-1}$ 의 공분산행렬들을 다음과 같이 표기하자.

$$V_u(s) = \text{Var}(u_{t+s}^{s-1})$$

$$V_v(s) = \text{Var}(v_t^{s-1})$$

$$V_w(s) = \text{Cov}(u_{t+s}^{s-1}, v_t^{s-1})$$

시차  $s$ 의 부분자기회귀행렬과 부분시차상관행렬은 각각 다음과 같이 쓸 수 있다.

$$\Phi_s^s = V_w(s) V_w^{-1}(s)$$

$$P_H(s) = D_w^{-1}(s) V_w^{-1}(s) D_w^{-1}(s)$$

여기서  $D_u(s)$ 와  $D_w(s)$ 는 각각  $V_u(s)$ 와  $V_w(s)$ 의 대각원소들의 제곱근들을 대각원소로 하는 대각행렬들이다.

Quenouille(1957, p. 40)와 Ansley & Newbold(1979)는 단변량시계열데이터의 부분자기상관계수를 다변량시계열데이터에 확장시켜 다음과 같은 시차  $s$ 의 다변량부분자기상관행렬(multivariate partial autocorrelation matrix)  $P_Q(s)$ 를 정의했다.

$$P_Q(s) := W_u^{-1}(s) V_w(s) W_w^{-1}(s)$$

여기서  $W_u(s)$ 와  $W_w(s)$ 는 대각원소들이 양수들이 대칭행렬들로서 다음 식들을 만족한다.

$$V_u(s) = W_u^2(s)$$

$$V_w(s) = W_w^2(s)$$

단변량시계열데이터의 경우에는 부분자기회귀행렬, 부분시차상관행렬, 다변량부분자기상관행렬은 동일하다. 즉, 다음 식이 성립한다.

$$\Phi_s^s = P_H(s) = P_Q(s)$$

그러나 다변량시계열데이터인 경우에는 이 세 행렬들이 동일하지 않다. 특히 부분자기회귀행렬이나 다변량부분자기상관행렬은 상관계수의 행렬이 아니다.

VAR( $p$ )과정에서는 이 세 행렬들 모두가 절단성질(cut-off property)을 갖는다. 즉, 다음 식들이 성립한다.

$$\Phi_s^s = 0 \quad (s=p+1, p+2, \dots)$$

$$P_H(s) = 0 \quad (s=p+1, p+2, \dots)$$



```

/* INPUT DATA */
C0 = { 1.3333333 -0.143007 -0.016824,
 -0.143007 1.2088164 0.4825525,
 -0.016824 0.4825525 1.3145703 };
C1 = { 0.6666667 -0.071504 -0.008412,
 0.1139853 0.2513466 0.4409439,
 -0.033649 0.386529 0.4908816 };
C2 = { 0.3333333 -0.035752 -0.004206,
 0.0679706 0.133943 0.1905176,
 0.0127024 0.166228 0.2354533 };
C3 = { 0.1666667 -0.017876 -0.002103,
 0.0439411 0.0596875 0.0892671,
 0.0174048 0.076657 0.1087395 };
C4 = { 0.0833333 -0.008938 -0.001052,
 0.0262822 0.0271783 0.0413383,
 0.0140097 0.0349346 0.0504753 };
C5 = { 0.0416667 -0.004469 -0.000526,
 0.0151645 0.0123044 0.0191713,
 0.0094593 0.015916 0.0234102 };

C = C0 || C1 || C2 || C3 || C4 || C5;

D = 3; /* DIMENSION OF SUBMATRICES */
MAXORDER = 4; /* MAX ORDR OF VAR MODEL */
/* CONSTANTS */
IDEN = I(D);
AROW = D;
ACOL = D*(MAXORDER+1);
L = J(AROW, ACOL, 0);
M = L;
G = J(D, D, 0);
LAMBDA = G;
V = G;
AROWS = 1:D;
ACOLS = 1:D;
/* INITIAL STEP K=0 */
L[AROWS, ACOLS] = -IDEN;
M[AROWS, ACOLS] = -IDEN;
ACOLSNXT = (D+1):(2*D);
G = C[AROWS, ACOLSNXT];
LAMBDA = C[AROWS, ACOLS];

```

```

V = LAMBDA;
/* ITERATIVE STEP */
DO K = 1 TO MAXORDER;
 LOD = L;
 MOLD = M;
 L[AROWS, ACOLS] = -IDEN;
 M[AROWS, ACOLS] = -IDEN;
 ACOLSLST = (K*D+1):(K+1)*D;
 L[AROWS, ACOLSLST] = G*INV(V);
 M[AROWS, ACOLSLST] = G'*INV(LAMBDA);
/* CALCULATE SQUARE ROOT MATRICES */
WUU = SQRT(DIAG(LAMBDA));
OLDWUU = WUU*1.0E+10;
DO J = 1 TO 30
 UNTIL (MAX(ABS(OLDWUU - WUU)) < 1.0E-10);
 OLDWUU = WUU;
 WUUVIN = INV(WUU);
 WUU = (2.0*WUU + LAMBDA*WUUVIN + WUUVIN*LAMBDA)/ 4.0;
 UEPSIL = MAX(ABS(WUU - OLDWUU));
END;
WV = SQRT(DIAG(V));
OLDWV = WV*1.0E+10;
DO J = 1 TO 30
 UNTIL (MAX(ABS(OLDWV - WV)) < 1.0E-10);
 OLDWV = WV;
 WVVIN = INV(WV);
 WV = (2.0*WV + V*WVVIN + WVVIN*V)/ 4.0;
 VEPSIL = MAX(ABS(WV - OLDWV));
END;
/* CALCULATE PARTIAL ACR MATRICES */
TPART = G*INV(V);
HPART = INV(SQRT(DIAG(V)))*G'
 *INV(SQRT(DIAG(LAMBDA)));
QPART = INV(WUU)*G*INV(WV);
FREE WUU WUUVIN OLDWUU;
FREE WV WVVIN OLDWV;
/* PRINT THE SOLUTION */
PRINT, "ORDER = " K,
 "ERROR TO CALCULATE WUU = " UEPSIL,

```



```

SAS
15:47 Tuesday, March 2, 1993 1

ORDER = K
 1

ERROR TO CALCULATE WUU = UEPSIL
5.033E-12

ERROR TO CALCULATE WWV = VEPSIL
5.033E-12

TIAO PAR MATRIX = TPART
 0.5 -4.863E-7 1.7835E-7
0.09999999 0.0999999 0.3
-3.075E-7 0.1999999 0.3

HEYSE PLACR MATRIX = HPART
 0.5 0.0897841 -0.025416
-0.056322 0.2079279 0.3066267
-0.006354 0.3497931 0.373416

QUENOUILLE MPACR MATRIX = QPART
0.5076784 -0.021576 0.0151621
0.1389318 0.1073042 0.277
-0.038221 0.2286419 0.2850174

ORDER = K
 2

ERROR TO CALCULATE WUU = UEPSIL
8.333E-11

ERROR TO CALCULATE WWV = VEPSIL
3.597E-11

TIAO PAR MATRIX = TPART
2.9116E-8 1.8477E-8 1.1512E-7
1.9631E-7 8.4974E-8 -4.569E-8
2.6533E-7 -2.406E-8 7.1705E-8

HEYSE PLACR MATRIX = HPART
2.0085E-8 1.8111E-7 2.484E-7
5.039E-8 5.1098E-8 -3.258E-8
1.2287E-7 -3.094E-8 5.0831E-8

QUENOUILLE MPACR MATRIX = QPART
3.9299E-8 4.3204E-8 1.1611E-7
1.579E-7 7.8896E-8 -3.598E-8
2.2832E-7 -3.983E-8 6.589E-8

ORDER = K
 3

ERROR TO CALCULATE WUU = UEPSIL
8.333E-11

ERROR TO CALCULATE WWV = VEPSIL
3.597E-11

TIAO PAR MATRIX = TPART
7.9884E-8 6.0231E-9 3.3754E-9
-7.883E-8 2.254E-8 -5.104E-8
-5.735E-8 -7.701E-9 -1.662E-8

HEYSE PLACR MATRIX = HPART
7.7446E-8 -7.581E-8 -5.212E-8
-1.903E-9 1.6858E-8 -6.134E-9
1.196E-9 -4.099E-8 -1.583E-8

QUENOUILLE MPACR MATRIX = QPART
7.2123E-8 4.3638E-9 -1.924E-9
-6.319E-8 2.1805E-8 -4.548E-8
-4.502E-8 -1.003E-8 -9.806E-9

```

```

ORDER = K
 4

 UEPSIL
ERROR TO CALCULATE WUU = 8.333E-11

 VEPSIL
ERROR TO CALCULATE WVW = 3.597E-11

 TPART
TIAO PAR MATRIX = -8.994E-8 1.4084E-7 -5.166E-7
 -1.231E-8 4.5714E-8 3.6941E-8
 6.391E-8 -2.293E-9 3.5592E-8

 HPART
HEYSE PLACR MATRIX = -7.745E-8 -1.926E-8 5.8473E-8
 1.9027E-9 5.8866E-8 9.045E-10
 -4.858E-7 5.1543E-8 3.1363E-8

 QPART
QUENOUILLE MPACR MATRIX = -8.912E-8 7.8276E-8 -5.078E-7
 -3.399E-8 6.2386E-8 -1.185E-8
 6.539E-8 -1.01E-8 3.8721E-8

7 workspace compresses.
Exiting IML.

```

그림 7.11 부분자기회귀행렬, 부분시차상관행렬, 그리고 다변량부분자기상관행렬

```

"ERROR TO CALCULATE WVW = " VEPSIL,
" TIAO PAR MATRIX = " TPART,
" HEYSE PLACR MATRIX = " HPART,
" QUENOUILLE MPACR MATRIX = " QPART,;
/ * END OF PRINTING */
DO J = 1 TO K-1;
 ACOLSJ = (J*D+1):(J+1)*D;
 ACOLSKMJ = ((K-J)*D+1):((K-J+1)*D);
 L[AROWS, ACOLSJ] = LOD[AROWS, ACOLSJ]
 - L[AROWS, ACOLSLST]*MOLD[AROWS, ACOLSKMJ];
 M[AROWS, ACOLSJ] = MOLD[AROWS, ACOLSJ]
 - M[AROWS, ACOLSLST]*LOLD[AROWS, ACOLSKMJ];
END;
ACOLSNXT = ((K+1)*D+1):(K+2)*D;
G = C[AROWS, ACOLSNXT];
DO J = 1 TO K;
 ACOLSJ = (J*D+1):((J+1)*D);
 ACOLSKMJ = ((K-J+1)*D+1):((K-J+2)*D);
 G = G-L[AROWS, ACOLSKMJ]*C[AROWS, ACOLSJ];
END;
LAMBDA = (IDEN -
 L[AROWS, ACOLSLST]*M[AROWS, ACOLSLST])*LAMBDA;

```

```
V = (IDEN -
 M[AROWS, ACOLSLST]*L[AROWS, ACOLSLST])*V;
END;
QUIT;
```

이 SAS프로그램을 실행한 결과가 그림 7.11이다. 그림 7.11에서 주어진 자기공분산행렬 함수가 VAR(1)모형에서 나온 것을 알 수 있다. 실제로 그림 6.2에서 이 자기공분산행렬 함수가 3변량VAR(1)모형에서 나온 것임을 확인할 수 있다.

부분자기회귀행렬, 부분시차상관행렬과 다변량부분자기상관행렬을 추정하기 위해서는 각 행렬들의 정의식에 자기공분산행렬들 대신에 Yule-Walker추정행렬들 또는 최소제곱추정행렬들을 대입한다.

### 7.5.3 우도비검정방법

가장 고전적인 방법으로  $d$ 변량VAR모형의 차수를 정하기 위해서는 우도비검정(likelihood ratio test)을 이용하는 것이다. 우선 다음 귀무가설을 검정하는 문제를 생각해 보기로 하자.

$$H_0 : \Phi_{k+1} = \Phi_{k+2} = \dots = \Phi_K = 0$$

제약조건이 없는 경우에는 우도함수의 최대값은 다음과 같다.

$$L(\check{\Phi}_1^K, \dots, \check{\Phi}_K^K, \check{\Sigma}_K) = X |\check{\Sigma}_K|^{-(T-K)/2}$$

귀무가설하에서 우도함수의 최대값은 다음과 같다.

$$L(\check{\Phi}_1^k, \dots, \check{\Phi}_k^k, \check{\Sigma}_k) = X |\check{\Sigma}_k|^{-(T-k)/2}$$

따라서 대수우도비검정통계량은 다음과 같다.

$$\lambda(k) = (T-k) \ln |\check{\Sigma}_k| - (T-K) \ln |\check{\Sigma}_K|$$

이 대수우도비검정통계량은 다음 성질을 만족한다.

**정리 7.17** 귀무가설  $H_0$ 하에서  $\lambda(k)$ 는 점근적으로 자유도가  $d^2(K-k)$ 인 카이제곱분포를 갖는다. 즉, 다음 성질을 만족한다.

$$\lambda(k) \xrightarrow{d} \chi_{d^2(K-k)}^2 \quad \blacksquare$$

이 귀무가설  $H_0$ 를 택한다는 것은 VAR모형이 가질 수 있는 차수의 최대값이  $k$ 라는 것이

지 차수가  $k$ 라는 것을 의미하지는 않는다. 따라서 축차검정법(sequential testing procedure)을 사용하는 것이 타당할 것이다. Anderson(1971, pp. 34~46, pp. 260~276)은 다음 귀무가설들을 축차적으로 검정할 것을 제안하였다.

$$\begin{aligned} H_0^{(K)} &: \Phi_K = 0 \\ H_0^{(K-1)} &: \Phi_K = \Phi_{K-1} = 0 \\ &\vdots \\ H_0^{(1)} &: \Phi_K = \Phi_{K-1} = \dots = \Phi_1 = 0 \end{aligned}$$

각 귀무가설을 검정하는 데는 정리 7.17의 우도비검정통계량을 사용한다. 만약 어떤 귀무가설이 기각되면 뒤따르는 모든 귀무가설들이 기각된다. 즉,  $H_0^{(k+1)}$ 이 채택되고  $H_0^{(k)}$ 가 기각되면 이 VAR모형의 차수는  $k$ 라고 결론짓는다. 따라서 이 축차우도비검정법의 제1종오차는 각각의 귀무가설을 검정하는데 적용되는 제1종오차와 다르다. 특히 유의할 점은 귀무가설  $H_0^{(k)}$ 을 앞의 귀무가설들  $H_0^{(K)}, \dots, H_0^{(k+1)}$ 이 참(true)이라는 조건하에서 검정하는 것이 아니라 사전에 가정된 최대차수의 VAR( $K$ )모형에 대해서 검정하는 것이다. 따라서 각 귀무가설을 검정하는 우도비검정통계량은 서로(점근적으로도) 독립이 아니다. 따라서 이 축차우도비검정법의 제1종오차를 구하는 것은 쉬운 일이 아니다.

Jenkins & Alavi(1981)와 Tiao & Box(1981)는 다음 귀무가설들을 축차적으로 검정할 것을 제안하였다.

$$\begin{aligned} H_0^K &: \Phi_K = 0 \\ H_0^{K-1} &: \Phi_{K-1} = 0 \\ &\vdots \\ H_0^1 &: \Phi_1 = 0 \end{aligned}$$

귀무가설  $H_0^k$ 와 이에 대립하는 대립가설  $H_A^k$ 를 좀더 명확하게 쓰면 다음과 같다.

$$\begin{aligned} H_0^k &: \Phi_k = 0 \mid \Phi_K = \Phi_{K-1} = \dots = 0 \\ \text{대 } H_A^k &: \Phi_k \neq 0 \mid \Phi_K = \Phi_{K-1} = \dots = 0 \end{aligned}$$

이 귀무가설  $H_0^k$ 를 검정하는 데는 정리 7.16을 사용한다. 즉, 귀무가설  $H_0^{k+1}$ 가 채택되었을 때 귀무가설  $H_0^k$ 를 검정하기 위해서는 다음과 같은 우도비검정통계량을 사용한다.

$$U_k := \frac{|S(k)|}{|S(k-1)|}$$

이러한 축차검정법을 축차부분자기회귀검정법(sequential partial autoregression test)이라고 한다.

정리 7.16과 Bartlett(1938)의 근사식을 사용하면 다음 정리를 얻는다.

**정리 7.18** 귀무가설  $H_0^{k+1}$ 이 채택된 상황에서  $H_0^k$ 를 검정할 때, 귀무가설  $H_0^k$ 하에서 다음 점근성질이 성립한다.

$$-\left(T - \frac{3}{2} - km\right) \ln U_k \xrightarrow{d} \chi_m^2 \quad \square$$

정리 7.18로부터 다음과 같은 따름정리를 유도할 수 있다. Tiao & Box(1981)는 이 따름정리를 사용해서 VAR모형의 차수를 정할 것을 제안하였다.

**따름정리 7.2** 귀무가설  $H_0^{k+1}$ 이 채택된 상황에서  $H_0^k$ 를 검정할 때 귀무가설  $H_0^k$ 하에서 다음 점근성질이 성립한다.

$$T(\ln |\hat{\Sigma}_{k-1}| - \ln |\check{\Sigma}_k|) \xrightarrow{d} \chi_m^2 \quad \square$$

축차우도비검정에서와 마찬가지로 축차부분자기회귀검정이나 Tiao-Box검정법에 의해서 선택되는 VAR모형의 차수는 검정에 사용된 유의수준에 의존한다. 더구나 축차부분자기회귀검정이나 Tiao-Box검정법에서 귀무가설  $H_0^k$ 를 기각한다는 것은 귀무가설들  $H_0^{k-1}$ ,  $H_0^{k-2}$ , ...,  $H_0^0$ 을 자동적으로 기각한다는 것을 포함한다. 따라서 각각의 귀무가설의 유의수준은 축차부분자기회귀검정이나 Tiao-Box검정의 유의수준과 구별되어야 한다. 만일 제j번째 단일검정에서  $H_0^{k+1-j}$ 가 참인데도 불구하고  $H_0^{k+1-j}$ 를 기각하는 기각역을  $R_j$ 라 하고, 이 제j번째 단일검정의 유의수준을  $\gamma_j = \Pr(R_j)$ 라 하자. 축차부분자기회귀검정과정이나 Tiao-Box검정과정의 제j번째 검정의 제1종오차는 다음과 같다.

$$\alpha_j = \Pr(R_1 \cup R_2 \cup \dots \cup R_j)$$

Paulsen & Tjøstheim(1985)은  $H_0^k, H_0^{k-1}, \dots, H_0^{k+1-k}$ 가 참일 때 우도비검정통계량들  $U_i$ 와  $U_j$ (단,  $i < j \leq k$ )는 점근적으로 독립임을 보였다. 따라서 다음 식이 점근적으로 성립한다.

$$\begin{aligned} \alpha_k &= \Pr\left(\bigcup_{i=1}^k R_i\right) \\ &= 1 - \Pr\left\{\left(\bigcap_{i=1}^k R_i^c\right)\right\} \\ &= 1 - \prod_{i=1}^k \Pr(R_i^c) \\ &= 1 - (1 - \gamma_1)(1 - \gamma_2) \dots (1 - \gamma_k) \end{aligned}$$

따라서 축차부분자기회귀검정이나 Tiao-Box검정에서 귀무가설들의 개수가 커질수록

제1종오차가 커짐을 알 수 있다.

축차우도비검정법(LRSTAT), 축차부분자기회귀검정법(PARSTAT), Tiao-Box검정법(TBSTAT)을 이용해서 VAR모형의 차수를 정하기 위해서 다음 SAS프로그램을 수행해보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM7-12.PRG에 수록되어 있다).

```

/ * */
/ * EXAM7-12.PRG */
/ * */
/ * DETERMINE VAR ORDER */
/ * THROUGH THE LR TESTS */
/ * DATASET = VAR22A */
/ * */
PROC IML;
RESET NOLOG NOCENTER;
/ * */
/ * MODULE FOR LR TESTS OF VAR IDENTIFICATION */
/ * */
START LRTEST(C, T, D, MAXORDER):
/ * CONSTANTS */
IDEN = I(D);
AROW = D;
ACOL = D*(MAXORDER+1);
L = J(AROW, ACOL, 0);
M = L;
G = J(D, D, 0);
LAMBDA = G;
V = G;
AROWS = 1:D;
ACOLS = 1:D;
LNDETSIG = J(1, MAXORDER+2, 0);
/ * INITIAL STEP K=0 */
L[AROWS, ACOLS] = -IDEN;
M[AROWS, ACOLS] = -IDEN;
ACOLSNXT = (D + 1):(2*D);
G = C[AROWS, ACOLSNXT];
LAMBDA = C[AROWS, ACOLS];
V = LAMBDA;
LNDETSIG[1, 1] = LOG(DET(LAMBDA));
/ * ITERATIVE STEP */

```

```

DO K = 1 TO MAXORDER:
 LORD = L:
 MORD = M:
 L[AROWS, ACOLS] = -IDEN:
 M[AROWS, ACOLS] = -IDEN:
 ACOLSLST = (K*D+1):(K+1)*D:
 L[AROWS, ACOLSLST] = G*INV(V):
 M[AROWS, ACOLSLST] = G'*INV(LAMBDA):
 DO J = 1 TO K-1:
 ACOLSJ = (J*D+1):(J+1)*D:
 ACOLSKMJ = ((K-J)*D+1):((K-J+1)*D):
 L[AROWS, ACOLSJ] = LORD[AROWS, ACOLSJ]
 - L[AROWS, ACOLSLST]*MORD[AROWS, ACOLSKMJ]:
 M[AROWS, ACOLSJ] = MORD[AROWS, ACOLSJ]
 - M[AROWS, ACOLSLST]*LORD[AROWS, ACOLSKMJ]:
 END:
 ACOLSNXT = ((K+1)*D+1):(K+2)*D:
 G = C[AROWS, ACOLSNXT]:
 DO J = 1 TO K:
 ACOLSJ = (J*D+1):((J+1)*D):
 ACOLSKMJ = ((K-J+1)*D+1):((K-J+2)*D):
 G = G-L[AROWS, ACOLSKMJ]*C[AROWS, ACOLSJ]:
 END:
 LAMBDA = (IDEN -
 L[AROWS, ACOLSLST]*M[AROWS, ACOLSLST])*LAMBDA:
 V = (IDEN -
 M[AROWS, ACOLSLST]*L[AROWS, ACOLSLST])*V:
 LNDETSIG[1, K+1] = LOG(DET(LAMBDA)):
 END:
FREE / T D MAXORDER LNDETSIG:
/* VAR MODEL IDENTIFICATION USING LR TESTS */
PRINT "[SEQUENTIAL LR TEST FOR IDENTIFICATION]":
DO J = 1 TO MAXORDER:
 MORDMJ = MAXORDER - J:
 TESTORD = MORDMJ + 1:
 LRSTAT = T*(LNDETSIG[1, MORDMJ+1] -
 LNDETSIG[1, MAXORDER+1]):
 DF = D*D*J:
 PRINT, "ORDER = " TESTORD " " "DF = " DF,

```

```

"LR STAT = " LRSTAT;
END;
PRINT ,, "[SEQUENTIAL PARTIAL AUTOREGRESSION TEST]", ;
DO J = 1 TO MAXORDER;
MORDMJ = MAXORDER - J ;
TESTORD = MORDMJ +1;
PARSTAT = (T-1.5-(MORDMJ+1)*D)*(LNDETSIG[1, MORDMJ+1]
- LNDETSIG[1, MORDMJ+2] +
D*(LOG(T-D*MORDMJ-1) - LOG(T-D*(MORDMJ+1)-1)));
TBSTAT = T*(LNDETSIG[1, MORDMJ+1]
- LNDETSIG[1, MORDMJ+2]);
DF = D*D;
PRINT "ORDER = " TESTORD " " "DF = " DF,
"PAR STAT = " PARSTAT " "
"TIAO & BOX STAT = " TBSTAT,;
END;
FINISH;
/* */
/* ESTIMATE THE SAMPLE ACVMF */
/* */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=60;
MAXORDER = 6;
USE TSA4DATA.VAR22A;
READ ALL VAR{Y1 Y2} INTO Y;
CLOSE TSA4DATA.VAR22A;
T = NROW(Y);
DIMEN = NCOL(Y);
SACVMF = COVLAG(Y, MAXORDER+2);
/* */
/* APPLY THE MODULE LRTEST */
/* */
RUN LRTEST(SACVMF, T, DIMEN, MAXORDER);
QUIT;

```

이 SAS프로그램은 다변량시계열데이터 VAR22A의 VAR차수를 정하기 위한 것이다. 이 SAS프로그램을 다른 시계열데이터에 적용하기 위해서는 이 SAS프로그램의 끝에서 12번째 줄부터 9번째 줄까지를 변형시키면 된다. 즉, 'MAXORDER='에 대상으로 하는 VAR차수의 최대값을 할당하고, USE스테이트먼트에는 분석하고자 하는 SAS데이터셋을 지정



하며, READ스테이트먼트에서는 이 SAS데이터세트를 행렬형태로 바꾸며, CLOSE스테이트먼트에서는 이 SAS데이터세트를 닫는다. 이 SAS프로그램에서 유의할 점은 각 검정통계량에서 백색잡음과정의 공분산행렬의 최우추정행렬  $\hat{\Sigma}_k$  대신 Yule-Walker추정행렬  $\hat{\Sigma}_k^*$ 를 사용하였다. 이는 최우추정행렬들을 구하는 데 많은 시간과 컴퓨터작업공간을 필요로 하는 데 비해서, Yule-Walker추정행렬들을 사용한 결과와 크게 다르지 않기 때문이다. 물론 Yule-Walker추정행렬을 변형한  $\Sigma_k^*$ 를 사용하는 것이 최우추정행렬을 사용한 결과에 더 가까울 것이다.

이 SAS프로그램을 수행한 결과가 그림 7.12이다. 각 귀무가설의 유의수준을 0.05로 하는 축차우도비검정법을 적용하면 다음 귀무가설들을 채택한다.

$$\begin{aligned} H_0^{(6)} : \Phi_6 &= 0 \\ H_0^{(5)} : \Phi_6 = \Phi_5 &= 0 \\ H_0^{(4)} : \Phi_6 = \Phi_5 = \Phi_4 &= 0 \\ H_0^{(3)} : \Phi_6 = \Phi_5 = \Phi_4 = \Phi_3 &= 0 \end{aligned}$$

그러나 다음 귀무가설을 기각한다.

$$H_0^{(2)} : \Phi_6 = \Phi_5 = \Phi_4 = \Phi_3 = \Phi_2 = 0$$

즉, 축차우도비검정법을 사용하면 VAR모형의 차수가 2라고 결론지을 수 있다. 각 귀무가설의 유의수준을 0.05로 하는 축차부분자기회귀검정법을 적용하면 다음 귀무가설들을 채택한다.

$$\begin{aligned} H_0^6 : \Phi_6 &= 0 \\ H_0^5 : \Phi_5 &= 0 \\ H_0^4 : \Phi_4 &= 0 \\ H_0^3 : \Phi_3 &= 0 \end{aligned}$$

반면에 다음 귀무가설을 기각한다.

$$H_0^2 : \Phi_2 = 0$$

따라서 축차부분자기회귀검정법을 사용해도 VAR모형의 차수가 2라고 결론지을 수 있다. 각 귀무가설의 유의수준을 0.05로 했으므로 축차부분자기회귀검정과정의 제5번째 검정의 점근적 제1종오차는 다음과 같다.

$$\begin{aligned} \alpha_5 &= 1 - (1 - 0.05)^5 \\ &\approx 0.226 \end{aligned}$$

또한, 그림 7.12에서 Tiao-Box검정법도 같은 결과를 보여줄 수 있다.

| SAS                                        |                      | 21:13 Saturday, June 19, 1993 8 |                     |
|--------------------------------------------|----------------------|---------------------------------|---------------------|
| [ SEQUENTIAL LR TEST FOR IDENTIFICATION ]  |                      |                                 |                     |
| ORDER =                                    | TESTORD<br>6         | DF =                            | DF<br>4             |
| LR STAT =                                  | LRSTAT<br>1.4162887  |                                 |                     |
| ORDER =                                    | TESTORD<br>5         | DF =                            | DF<br>8             |
| LR STAT =                                  | LRSTAT<br>6.6445686  |                                 |                     |
| ORDER =                                    | TESTORD<br>4         | DF =                            | DF<br>12            |
| LR STAT =                                  | LRSTAT<br>10.173887  |                                 |                     |
| ORDER =                                    | TESTORD<br>3         | DF =                            | DF<br>16            |
| LR STAT =                                  | LRSTAT<br>14.212393  |                                 |                     |
| ORDER =                                    | TESTORD<br>2         | DF =                            | DF<br>20            |
| LR STAT =                                  | LRSTAT<br>115.31135  |                                 |                     |
| ORDER =                                    | TESTORD<br>1         | DF =                            | DF<br>24            |
| LR STAT =                                  | LRSTAT<br>183.05827  |                                 |                     |
| [ SEQUENTIAL PARTIAL AUTOREGRESSION TEST ] |                      |                                 |                     |
| ORDER =                                    | TESTORD<br>6         | DF =                            | DF<br>4             |
| PAR STAT =                                 | PARSTAT<br>5.2454137 | TIAO & BOX STAT =               | TBSTAT<br>1.4162887 |
| ORDER =                                    | TESTORD<br>5         | DF =                            | DF<br>4             |
| PAR STAT =                                 | PARSTAT<br>8.7846552 | TIAO & BOX STAT =               | TBSTAT<br>5.2282799 |
| ORDER =                                    | TESTORD<br>4         | DF =                            | DF<br>4             |
| PAR STAT =                                 | PARSTAT<br>7.2636064 | TIAO & BOX STAT =               | TBSTAT<br>3.5293179 |
| ORDER =                                    | TESTORD<br>3         | DF =                            | DF<br>4             |
| PAR STAT =                                 | PARSTAT<br>7.7949781 | TIAO & BOX STAT =               | TBSTAT<br>4.0385065 |
| ORDER =                                    | TESTORD<br>2         | DF =                            | DF<br>4             |
| PAR STAT =                                 | PARSTAT<br>101.35097 | TIAO & BOX STAT =               | TBSTAT<br>101.09896 |

|                                         |                      |                   |                     |
|-----------------------------------------|----------------------|-------------------|---------------------|
| ORDER =                                 | TESTORD<br>1         | DF =              | DF<br>4             |
| PAR STAT =                              | PARSTAT<br>70.125675 | TIAO & BOX STAT = | TBSTAT<br>67.746917 |
| 5 workspace compresses.<br>Exiting IML. |                      |                   |                     |

그림 7.12 우도비방법을 사용한 VAR모형의 식별

### 7.5.4 벌칙함수식별방법

다변량시계열데이터를 차수가 높은 VAR모형에 적합시키면 다변량백색잡음과정의 공분산행렬  $\Sigma$ 의 추정행렬인 평균오차제공행렬의 대각원소들이 작아진다. 즉, 다변량시계열데이터를 구성하는 각 단변량시계열데이터의 평균오차제공이 작아진다. 그러나 이 평균오차제공들을 아주 작게 하는 차수를 택하면 이 차수는 지나치게 커서 모수절약원칙에 위배된다. 더구나 차수가 커지면 추정모형의 편의는 줄지만 분산이 늘어나는 Grenander현상이 나타난다. 이러한 문제점을 해결하기 위해서 높은 차수를 택하는 데 따르는 벌칙(penalty)을 평균오차제공행렬의 행렬식에 더한 통계량을 최소화하는 차수를 택한다. 이러한 식별방법을 벌칙함수식별방법(penalty function identification method)이라고 한다. 다른 다변량시계열분석기법들처럼 이 벌칙함수식별방법들도 원래는 단변량시계열분석을 위해서 제안되었다가 후에 다변량시계열분석을 위해 변형되었다. 벌칙함수식별방법들의 자세한 배경과 이론에 대해서는 [Choi92]의 제 3장과 [SAS4TSA2] (pp. 511~514)를 참조하라.

가장 먼저 제시된 벌칙함수식별통계량은 Akaike(1969)의 FPE(final prediction error)이다. VAR모형의 식별을 위한 FPE는 다음과 같이 정의된다.

$$FPE(k) := \left( \frac{T+kd+1}{T-kd-1} \right)^d \det(\check{\Sigma}_k)$$

이 FPE(k)를 최소로 하는 차수를 최소FPE추정량이라 부르고  $\hat{k}(FPE)$ 라 표기한다.

Akaike(1974)는 Kullback-Leibler정보수와 최우추정량의 성질을 이용해서 AIC(Akaike's information criterion)를 제시했다. VAR모형의 식별을 위한 AIC는 다음과 같이 정의된다.

$$AIC(k) := \ln \det(\check{\Sigma}_k) + \frac{2d^2k}{T}$$

이 AIC(k)를 최소로 하는 차수를 최소AIC추정량이라 부르고  $\hat{k}(AIC)$ 라 표기한다.

두 통계량 FPE와 AIC는 다음 식을 만족한다.

$$\ln FPE(k) = AIC(k) + \frac{2d}{T} + O\left(\frac{1}{T^2}\right)$$

따라서 관찰점들의 개수  $T$ 가 크면 최소FPE추정량과 최소AIC추정량은 거의 같은 성질을 가질 것이다.

회귀분석이나 단변량시계열분석에서 최소FPE추정량과 최소AIC추정량이 일치성을 갖지 못하는 잘 알려져 있다. Paulsen & Tjøstheim(1985)은 VAR모형의 식별에서도 이 두 추정량은 일치성을 갖지 못함을 보였다. 즉, 다변량시계열데이터가 VAR(p)모형에서 생성되었을 때 최소AIC추정량은 다음 성질들을 만족한다.

$$\lim_{T \rightarrow \infty} \Pr(\hat{k}(\text{AIC}) \geq p) = 1$$

$$\lim_{T \rightarrow \infty} \Pr(\hat{k}(\text{AIC}) > p) > 0$$

물론 최소FPE추정량도 최소AIC추정량과 같은 점근성질들을 갖는다.

최소FPE추정량과 같은 점근성질을 갖는 통계량으로 Parzen(1977)이 제시한 최소CAT 추정량이 있다. VAR모형의 식별을 위한 CAT(criterion of autoregressive transfer function)는 다음과 같이 정의된다.

$$\text{CAT}(k) = \text{trace} \left[ \frac{d}{T} \sum_{j=0}^k \left( \frac{T}{T-jd-1} \check{\Sigma}_j \right)^{-1} - \left( \frac{T}{T-kd-1} \check{\Sigma}_j \right)^{-1} \right]$$

이 CAT(k)를 최소로 하는 차수를 최소CAT추정량이라 하고  $\hat{k}(\text{CAT})$ 라 표기한다.

앞에서 언급했듯이 최소FPE추정량, 최소AIC추정량과 최소CAT추정량은 같은 점근성질들을 갖는다. 특히 이 추정량들은 일치성을 갖지 못한다. 여기서 일치성의 정의를 복습해 보자. VAR(p)과정의 차수의 추정량  $\hat{k}$ 가 일치성을 갖기 위한 필요충분조건은 다음과 같다.

$$\lim_{T \rightarrow \infty} \Pr(\hat{k} = p) = 1$$

차수의 추정량  $\hat{k}$ 가 강일치성(strong consistency)을 갖기 위한 조건은 다음과 같다.

$$\Pr(\lim_{T \rightarrow \infty} \hat{k} = p) = 1$$

어떤 벌칙함수식별방법이 일치성(또는 강일치성)을 갖는다는 것은 이 식별방법에 의한 추정량이 일치성(또는 강일치성)을 갖는 것을 뜻한다.

VAR모형의 벌칙함수식별방법들의 일치성을 조사하기 위해서 Quinn(1980)은 다음과 같은 정리를 제시하였다.

**정리 7.19** 다변량시계열데이터  $\{y_t\}$ 가 표준다변량백색잡음과정을 오차항으로 하는 d변량VAR(p)과정에서 생성되었다고 하자. 가능한 VAR차수의 상한을 K라고 하고,  $p \leq K$ 라고 가정하자. 또한  $\hat{k}$ 를 다음 벌칙함수를 최소로 하는 차수의 추정량이라고 하자.

$$P(k) = \ln \det(\check{\Sigma}_k) + \frac{kC(T)}{T} \quad (0 \leq k \leq K)$$

여기서 C(T)는 관찰점들의 개수 T의 비감소함수이다. 이러한 조건하에서 차수의 추정량  $\hat{k}$ 가 일치성을 갖기 위한 필요충분조건은 다음과 같다.

$$\lim_{T \rightarrow \infty} C(T) = \infty, \quad \lim_{T \rightarrow \infty} \frac{C(T)}{T} = 0$$

또한 차수의 추정량  $\hat{k}$ 가 강일치성을 갖기 위한 필요충분조건은 다음과 같다.

$$\lim_{T \rightarrow \infty} \frac{C(T)}{2 \ln \ln T} > 1 \quad \square$$

Paulsen(1984)은 VAR특성방정식  $\det(\Phi(z))=0$ 가 1을 근으로 가질 때도 정리 7.19의 조건 아래서 일치성을 가짐을 보였다. 그러나 강일치성은 다루지 않았다.

정리 7.19에 의해서 FPE, AIC, 그리고 CAT는 일치성을 갖는 식별방법들이 아님을 알 수 있다. Paulsen & Tjøstheim(1985)는 최소AIC추정량  $\hat{k}$ (AIC)의 점근분포를 제시했다. 이 점근분포에 의하면 다변량시계열데이터의 차원  $d$ 가 증가할수록 과대식별을 하는 확률  $\Pr(\hat{k}(AIC) > p)$ 의 극한값이 감소한다. 더구나 다변량시계열데이터의 차원이 1, 즉 단변량시계열데이터에서보다는 차원이 2, 즉 2변량시계열데이터에서 과대식별의 점근확률이 VAR모형의 차수에 덜 의존한다. 이러한 성질들은 차원  $d$ 가 큰 경우에 더욱 명백하게 나타난다. 실제로 다변량시계열데이터의 차원  $d$ 가 5 이상이면 AIC에 의해서 과대식별하는 점근확률은 무시해도 될 정도이다. 이 결론은 모의실험에서도 예증되었다. 결론적으로 말해서 차원  $d$ 가 큰 경우에는 최소FPE추정량, 최소AIC추정량, 그리고 최소CAT추정량이 진짜차수  $p$ 일 점근확률은 거의 1이다. Koreisha & Pukkila(1993)는 차원  $d$ 가 큰 경우에 대해서 벌칙함수식별방법들을 모의실험을 통해서 비교하였다.

일치성을 갖는 벌칙함수식별방법들에는 Hannan & Quinn(1979)이 제시한 HQC(Hannan-Quinn criterion), Akaike(1977), Schwarz(1978), Rissanen(1978), 그리고 Kashyap(1977)이 제시한 BIC(Bayesian information criterion), Quinn(1980)이 제시한 QC(Quinn's criterion)가 있다. 이 벌칙함수식별방법들의 통계량들은 다음과 같다.

$$HQC(k) = \ln \det(\check{\Sigma}_k) + \frac{2kd^2 \ln \ln T}{T}$$

$$BIC(k) = \ln \det(\check{\Sigma}_k) + \frac{kd^2 \ln T}{T}$$

$$QC(k) = \ln \det(\check{\Sigma}_k) + \frac{2k \ln \ln T}{T}$$

이 HQC( $k$ ), BIC( $k$ )와 QC( $k$ )를 최소로 하는 차수들을 각각 최소HQC추정량, 최소BIC추정량과 최소QC추정량이라 부르고, 각각  $\hat{k}$ (HQC),  $\hat{k}$ (BIC)와  $\hat{k}$ (QC)라 표기한다.

정리 7.19에 의해서 최소BIC추정량은 강일치성을 가짐을 알 수 있다. 반면에 최소HQC추정량은 차원  $d$ 가 1인 경우에 일치성을 갖고 차원  $d$ 가 2 이상인 경우에는 강일치성을 갖는다. 또한 최소QC추정량은 일치성을 갖는다.

원래 Hannan & Quinn(1979)과 Quinn(1980)은 벌칙함수들을 다음과 같이 정의했다.

$$HQC_\alpha(k) = \ln \det(\check{\Sigma}_k) + \frac{2\alpha kd^2 \ln \ln T}{T} \quad (\alpha > 1)$$

$$QC_\alpha(k) = \ln \det(\check{\Sigma}_k) + \frac{2\alpha k \ln \ln T}{T} \quad (\alpha > 1)$$

이 두 벌칙함수들을  $\alpha=1$ 에서 정의하지 않는 이유는 선택된 차수의 추정량들이 강일치성을 갖지 않기 때문이다. 그러나 이 벌칙함수식별방법들은 잘 사용되지 않고 대신 HQC



```

RESET NOLOG NOCENTER;
/* */
/* MODULE FOR VAR IDENTIFICATION */
/* BY PENALTY FUNCTION METHODS */
/* */
/* C = AUTOCOVARIANCE MATRICES */
/* T = NO. OF OBSERVATIONS */
/* D = DIMENSION OF VECTOR */
/* MAXORDER = MAX OF VAR ORDERS */
START PFMIDEN(C, T, D, MAXORDER):
/* CONSTANTS */
IDEN = I(D);
AROW = D;
ACOL = D*(MAXORDER+1);
L = J(AROW, ACOL, 0);
M = L;
G = J(D, D, 0);
LAMBDA = G;
V = G;
AROWS = 1:D;
ACOLS = 1:D;
LNDETSIG = J(1, MAXORDER+1, 0);
/* INITIAL STEP K=0 */
L[AROWS, ACOLS] = -IDEN;
M[AROWS, ACOLS] = -IDEN;
ACOLSNXT = (D + 1):(2*D);
G = C[AROWS, ACOLSNXT];
LAMBDA = C[AROWS, ACOLS];
V = LAMBDA;
LNDETSIG[1, 1] = LOG(DET(LAMBDA));
/* THE IDENTIFICATION CRITERIA */
LAMBSTR = (T-1)/ T*LAMBDA;
LNSIG = LOG(DET(LAMBDA));
LNSIGSTR = LOG(DET(LAMBSTR));
AIC = LNSIG;
AICSTR = LNSIGSTR;
BIC = LNSIG;
BICSTR = LNSIGSTR;
HQC = LNSIG;

```

```

HQCSTR = LNSIGSTR;
PRINT , "ORDER = 0",
 AIC " " AICSTR,
 BIC " " BICSTR,
 HQC " " HQCSTR,;
FREE LAMBSTR AIC AICSTR BIC BICSTR HQC HQCSTR;
/* ITERATIVE STEP */
DO K = 1 TO MAXORDER;
 LORD = L;
 MOLD = M;
 L[AROWS, ACOLS] = -IDEN;
 M[AROWS, ACOLS] = -IDEN;
 ACOLSLST = (K*D+1):(K+1)*D;
 L[AROWS, ACOLSLST] = G*INV(V);
 M[AROWS, ACOLSLST] = G*INV(LAMBDA);
 DO J = 1 TO K-1;
 ACOLSJ = (J*D+1):(J+1)*D;
 ACOLSKMJ = ((K-J)*D+1):((K-J+1)*D);
 L[AROWS, ACOLSJ] = LORD[AROWS, ACOLSJ]
 - L[AROWS, ACOLSLST]*MOLD[AROWS, ACOLSKMJ];
 M[AROWS, ACOLSJ] = MOLD[AROWS, ACOLSJ]
 - M[AROWS, ACOLSLST]*LORD[AROWS, ACOLSKMJ];
 END;
 ACOLSNXT = ((K+1)*D+1):(K+2)*D;
 G = C[AROWS, ACOLSNXT];
 DO J = 1 TO K;
 ACOLSJ = (J*D+1):(J+1)*D;
 ACOLSKMJ = ((K-J+1)*D+1):((K-J+2)*D);
 G = G-L[AROWS, ACOLSKMJ]*C[AROWS, ACOLSJ];
 END;
 LAMBDA = (IDEN -
 L[AROWS, ACOLSLST]*M[AROWS, ACOLSLST])*LAMBDA;
 V = (IDEN -
 M[AROWS, ACOLSLST]*L[AROWS, ACOLSLST])*V;
 LNDETSIG[1, K+1] = LOG(DET(LAMBDA));
/* THE IDENTIFICATION CRITERIA */
 LAMBSTR = LAMBDA;
 DO II = 1 TO K;
 ACOLSII = (II*D+1):(II+1)*D;

```



```

DO JJ = (II+1) TO K;
 ACOLSJJ = (JJ*D+1):(JJ+1)*D;
 JMI = JJ - II;
 ACOLSJMI = (JMI*D+1):(JMI+1)*D;
 LAMBSTR = LAMBSTR + 2.0/ T*JMI
 *L[AROWS, ACOLSII]*C[AROWS, ACOLSJMI]
 *L[AROWS, ACOLSJJ]4;
 END;
END;
LAMBSTR = (T-1)/ T*LAMBSTR;
LNSIG = LOG(DET(LAMBDA));
LNSIGSTR = LOG(DET(LAMBSTR));
AIC = LNSIG + 2.0*D*D*K/ T;
AICSTR = LNSIGSTR + 2.0*D*D*K/ T;
BIC = LNSIG + LOG(T)*D*D*K/ T;
BICSTR = LNSIGSTR + LOG(T)*D*D*K/ T;
HQC = LNSIG + 2.0*LOG(LOG(T))*D*D*K/ T;
HQCSTR = LNSIGSTR + 2.0*LOG(LOG(T))*D*D*K/ T;
PRINT , "ORDER = " K,
 AIC " " AICSTR,
 BIC " " BICSTR,
 HQC " " HQCSTR ,;
FREE LAMBSTR AIC AICSTR BIC BICSTR HQC HQCSTR;
END;
FREE / T D MAXORDER;
FINISH;
/ * * /
/ * ESTIMATE THE SAMPLE ACVMF * /
/ * * /
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=60;
MAXORDER = 6;
USE TSA4DATA.VAR22A;
READ ALL VAR{Y1 Y2} INTO Y;
CLOSE TSA4DATA.VAR22A;
T = NROW(Y);
DIMEN = NCOL(Y);
SACVMF = COVLAG(Y, MAXORDER+2);
/ * * /

```

```

/ * APPLY THE MODULE PFMIDEN * /
/ * * /
RUN PFMIDEN(SACVMF, T, DIMEN, MAXORDER);
QUIT;

```

이 SAS프로그램을 수행하기 위해서는 끝에서 12번째 줄에서 식별하고자 하는 차수의 가능한 최대값을 지정한다. 이 예제에서는 'MAXORDER=6'이라 했으므로 가능한 차수

| SAS                                      |           | 16:43 Sunday, June 27, 1993 1 |           |
|------------------------------------------|-----------|-------------------------------|-----------|
| ORDER = 0                                |           | ORDER =                       | K<br>4    |
| AIC                                      | AICSTR    | AIC                           | AICSTR    |
| 0.938331                                 | 0.924953  | -0.024427                     | -0.042407 |
| BIC                                      | BICSTR    | BIC                           | BICSTR    |
| 0.938331                                 | 0.924953  | 0.2967074                     | 0.2787276 |
| HQC                                      | HQCSTR    | HQC                           | HQCSTR    |
| 0.938331                                 | 0.924953  | 0.1060397                     | 0.0880599 |
| ORDER =                                  | K<br>1    | ORDER =                       | K<br>5    |
| AIC                                      | AICSTR    | AIC                           | AICSTR    |
| 0.5400182                                | 0.5266402 | -0.005949                     | -0.022915 |
| BIC                                      | BICSTR    | BIC                           | BICSTR    |
| 0.6203018                                | 0.6069238 | 0.3954691                     | 0.3785035 |
| HQC                                      | HQCSTR    | HQC                           | HQCSTR    |
| 0.5726349                                | 0.5592569 | 0.1571345                     | 0.1401689 |
| ORDER =                                  | K<br>2    | ORDER =                       | K<br>6    |
| AIC                                      | AICSTR    | AIC                           | AICSTR    |
| -0.080642                                | -0.098606 | 0.0379425                     | 0.0205578 |
| BIC                                      | BICSTR    | BIC                           | BICSTR    |
| 0.0799257                                | 0.0619609 | 0.5196442                     | 0.5022595 |
| HQC                                      | HQCSTR    | HQC                           | HQCSTR    |
| -0.015408                                | -0.033373 | 0.2336426                     | 0.2162579 |
| ORDER =                                  | K<br>3    |                               |           |
| AIC                                      | AICSTR    |                               |           |
| -0.054232                                | -0.072399 |                               |           |
| BIC                                      | BICSTR    |                               |           |
| 0.1866192                                | 0.1684519 |                               |           |
| HQC                                      | HQCSTR    |                               |           |
| 0.0436185                                | 0.0254511 |                               |           |
| 28 workspace compresses.<br>Exiting IML. |           |                               |           |

그림 7.13 벌칙함수식별방법들에 의한 VAR모형의 식별

의 최대값을 6으로 했다. 또한 끝에서 11번째 줄의 USE스테이트먼트에 분석하고자 하는 SAS데이터세트를 지정하며, 10번째 줄의 READ스테이트먼트에서 이 SAS데이터세트를 행렬형태로 바꾸며, 9번째 줄의 CLOSE스테이트먼트에서 이 SAS스테이트먼트를 닫는다. 이 SAS프로그램에서는 AIC, BIC, HQC만을 계산하고 FPE, CAT, QC 등은 계산하지 않았다. 이는 PC SAS를 사용하기 때문에 발생하는 컴퓨터의 작업마당(workspace)의 부족을 고려한 것이다. WINDOWS용 SAS, 또는 워크스테이션용 SAS를 쓰는 독자라면 후자들도 쉽게 계산할 수 있으리라 생각한다. 또한 이 SAS프로그램에서는 최우추정량  $\hat{\Sigma}$  대신 Yule-Walker추정량  $\hat{\Sigma}_k$ 를 사용했다. 이것도 컴퓨터의 작업마당의 부족을 고려한 것이기는 하나, VAR모형의 식별과정에서 최우추정량 대신 Yule-Walker추정량을 쓰는 것이 일반적이기 때문이기도 하다. 이 SAS프로그램에서는 최우추정량  $\hat{\Sigma}$ 에 가까운 제 7.5.1소절에서 정의한 변형된 Yule-Walker추정량  $\Sigma_k^*$ 를 사용해서 AIC, BIC, HQC를 계산하고 이를 각각 AICSTR, BICSTR, HQCSTR로 표기했다. 이 SAS프로그램을 수행한 결과가 그림 7.13이다. 그림 7.13으로부터 AIC, BIC, HQC, AICSTR, BICSTR, HQCSTR 모두 차수 2를 선택함을 알 수 있다.

### 7.5.5 VAR모형 식별방법들의 비교

Lütkephol(1985)은 시뮬레이션을 통해서 VAR모형의 식별방법들을 비교하였다. 이 소절은 이 논문의 내용을 요약한 것이다.

전반적으로 BIC와 HQC가 가장 작은 전방일단계예측오차의 평균(the smallest average squared forecasting errors)을 가지며, 또한 가장 자주 진짜(true) VAR차수를 식별한다. 이 두 벌칙함수식별방법들은 일치성을 갖는다. 일치성은 관찰점들의 개수  $T$ 가 100 또는 200인 경우에서의 식별방법들의 행태를 조사하는 좋은 기준이 된다. 또한 FPE, AIC, 그리고 CAT는 앞의 통계량보다는 약간 큰 전방일단계예측오차의 평균을 갖는다. 반면에 나머지 식별방법들은 앞에서 언급된 식별방법들에 비해서 열등한 예측능력을 갖는다. 관찰점들의 개수  $T$ 가 증가함에 따라 BIC, HQC, AIC, FPE, 그리고 CAT의 예측능력 차이는 점차 줄어든다. 특히 유의할 점은 비록 식별방법들의 식별능력 차이는 VAR차수, VAR 특성근들의 절대값의 크기와 형태, 그리고 다변량백색잡음과정의 공분산행렬의 구조 등에 의존하기는 하지만 상당히 넓은 범위의 VAR모형들을 시뮬레이션분석한 결과에서 식별방법의 우열은 일정하게 나타났다는 것이다.

축차우도비검정법이나 축차부분자기회귀검정법은 모형을 잘 식별하지 못하는데, 이는 소표본에서 이 검정법들의 통계량들의 진짜 확률분포들이 정리 7.17이나 정리 7.18의 점근분포들과 잘 부합되지 않는다는 것을 의미한다. 이 점은 관찰점들의 개수가 아주 작을 때 귀무가설을 지나치게 자주 기각한다는 것에서도 확인할 수 있다. 이후 축차검정법들에서 가정된 최고차수를  $K$ 라 하고,  $K$ 는 진짜차수  $p$ 보다 크다고 가정하며, 각 귀무가설을 검정하는 유의수준을  $\alpha$ 라고 하자. 비록 소표본에서 검정통계량들의 분포가 점근카이제곱분포

들과 정확히 같다고 가정한다 하더라도, 첫 번째 귀무가설을 기각할 확률은  $\alpha$ 이다. 따라서 관찰점들의 개수가 크고 유의수준  $\alpha$ 가 작다하더라도 축차검정법들에 의해서 차수를 결정할 때 차수를 과대식별할 확률이 0이 아닌 양수이다. 따라서 각 귀무가설을 검정하기 위한 유의수준이 5%이고 가정된 최고차수  $K$ 와 진짜차수  $p$ 의 차이가 크다면, 축차우도비검정법이나 축차부분자기회귀검정법은 식별능력이나 예측능력에서 열등할 것이다. 그렇다고 해서 VAR모형을 식별하는데 축차검정법들을 절대 사용하지 말라는 뜻은 아니다. 벌칙함수 식별방법들을 사용하면 단순히 통계패키지가 선택해 준 차수들을 이용해야 하는 반면에 축차검정법들을 사용하면 각 귀무가설을 검정하면서 주어진 다변량시계열데이터를 파악할 수도 있다. 전체적으로 보아 지금까지 언급된 식별방법들 중에서 어느 것도 차수를 과소식별하는 경향을 보이지 않는다. 따라서 정리 7.20에 의하면 BIC나 HQC가 진짜차수를 잘 식별해내는 것은 당연하다. 결과적으로 다변량시계열데이터에 주어진 특별한 정보가 없고 VAR모형의 유한차수  $p$ 를 선택하거나 예측하는 것이 목적이라면 BIC나 HQC를 사용하는 것이 타당할 것이다. 그러나 다변량시계열데이터와 그에 따르는 특별한 정보가 있다면 다른 식별방법을 사용하는 것이 좋을 수도 있다. 예를 들어, 사전정보가 있다면 베이지안 이론을 이용하는 Smith & Spiegelhalter(1980), Atkinson(1978, 1980), Poskitt & Tremayne(1983), 그리고 Chow(1981) 등의 기법들을 이용할 수 있을 것이다.

## 7.6 VAR모형의 검진

VAR모형화를 끝내기 위해서는 VAR모형의 식별과정과 추정과정을 거쳐서 선택된 모형이 주어진 다변량시계열데이터와 잘 부합되는지를 검진해야 한다. 시계열분석에서의 검진의 기초에 관해서는 [SAS4TSA2] (pp. 514~1518)를 참조하라. 이 절에서 다루는 방법들은 VAR모형의 검진뿐만 아니라 VARMA모형의 검진에도 적용할 수 있다.

### 7.6.1 다변량백색잡음과정의 표본자기상관행렬함수

VAR모형의 다변량백색잡음과정의 독립성(또는 비상관성)을 검정하기 위해서는 잔차들의 표본자기상관행렬함수를 분석해 볼 필요가 있다. 이 목적을 위해서 먼저 다변량백색잡음과정의 자기상관행렬함수의 성질들을 살펴보자.

우선  $v_1, v_2, \dots, v_T$ 를 평균벡터가 0이고 공분산행렬이  $\Sigma$ 인 표준 $d$ 변량백색잡음과정의 표본벡터들이라고 가정하자. 이  $d$ 변량백색잡음과정의 표본자기공분산행렬은 다음과 같다.

$$C_h := \frac{1}{T} \sum_{t=1}^{T-h} v_{t+h} v_t' \quad (h = 0, 1, \dots, T-1)$$

이 표본자기공분산행렬함수의  $(i, j)$ 원소를  $c_{ij,h}$ 라 하자. 또한 시차 0에서의 표본자기공분산행렬  $C_0$ 의 대각원소들의 제곱근들로 이루어진 대각행렬  $D$ 를 다음과 같이 정의하자.

$$D := \text{diag}(\sqrt{c_{11,0}}, \sqrt{c_{22,0}}, \dots, \sqrt{c_{dd,0}})$$

이  $d$ 변량백색잡음과정의 표본자기상관행렬은 다음과 같다.

$$R_h := D^{-1}C_h D^{-1} \quad (h=0, 1, \dots, T-1)$$

이 표본자기상관행렬  $R_h$ 의  $(i, j)$ 원소를  $r_{ij,h}$ 라 하자.

**정리 7.21** 만일  $v_1, v_2, \dots, v_T$ 가 표준 $d$ 변량백색잡음과정에서 발생되었으면,  $\sqrt{T}\text{vec}(C_1), \sqrt{T}\text{vec}(C_2), \dots, \sqrt{T}\text{vec}(C_m)$ 은 점근적으로 독립이며 같은 점근정규분포를 갖는다. 이 점근정규분포의 평균벡터는 영벡터이고 공분산행렬은  $\Sigma \otimes \Sigma$ 이다.  $\square$

**정리 7.22** 정리 7.21의 조건하에서  $\sqrt{T}\text{vec}(R_1), \sqrt{T}\text{vec}(R_2), \dots, \sqrt{T}\text{vec}(R_m)$ 은 점근적으로 독립이며 같은 점근정규분포를 갖는다. 이 점근분포의 평균벡터는 영벡터이고 공분산행렬은  $\Delta \otimes \Delta$ 이다. 여기서  $\Delta$ 는 이  $d$ 변량백색잡음과정의 시차 0에서의 자기상관행렬이다. 즉,  $\Sigma$ 의  $(i, j)$ 원소를  $\sigma_{ij}$ 라 할 때  $\Delta$ 의  $(i, j)$ 원소  $\delta_{ij}$ 는 다음과 같다.

$$\delta_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}} \sqrt{\sigma_{jj}}} \quad \square$$

정리 7.21과 정리 7.22는 정리 7.4와 정리 7.5로부터 유도할 수 있다. 자세한 내용은 Hannan(1970, pp. 220~224)을 참조하라. 표본자기공분산행렬  $C_h$ 대신 평균이 조정된 다음과 같은 표본자기공분산행렬을 사용해도 정리 7.21과 정리 7.22는 성립한다.

$$C_h := \frac{1}{T} \sum_{t=1}^{T-h} (v_{t+h} - \bar{v})(v_t - \bar{v})' \quad (h=0, 1, \dots, T-1)$$

$$\text{단, } \bar{v} = \frac{1}{T} \sum_{t=1}^T v_t$$

다변량백색잡음과정에서 시차 0의 자기상관행렬은  $\Delta$ 이고 그 이외의 시차의 자기상관행렬은 영행렬이다. 시차가  $h (\neq 0)$ 인 자기상관행렬의  $(i, j)$ 원소를  $\rho_{ij}(h)$ 라고 하자. 정리 7.22를 사용해서 다음과 같은 귀무가설을 검정할 수가 있다.

$$H_0 : \rho_{ij}(h) = 0 \quad \text{대} \quad H_1 : \rho_{ij}(h) \neq 0$$

정리 7.22에 의하면 다음과 같은 부등식이 만족되면 유의수준 5%에서 귀무가설을 기각한다.

$$|r_{ij,h}| > \frac{2}{\sqrt{T}}$$

이 검정법을 이용하면 주어진 다변량시계열데이터가 다변량백색잡음과정에서 발생되었

는가를 검정할 수 있다. 그러나 이 검정법을 사용할 때 유의할 점이 있다. 만일  $g \neq h$ 이면  $\sqrt{TR_h}$ 와  $\sqrt{TR_g}$ 는 점근적으로 서로 독립이지만,  $\sqrt{TR_h}$ 의 원소들이 점근적으로 서로 독립인 것은 아니라는 점이다. 따라서 제 7.5.2 소절에서 언급되었듯이 이 검정법은 제1종오차를 조정해야 하는 문제를 발생시킨다. 또한 이 검정법은 정리 7.22의 점근이론을 바탕으로 하므로, 대표본인 경우에 한해서 적용할 수 있다. Dufour & Roy(1985)는 소표본인 경우에는  $\sqrt{TR_{g,h}} (h > 0)$ 의 표본분산이 1과 다른 경우가 많음을 보였다. 특히 표본분산이 1보다 작은 경우가 많기 때문에 이 검정법은 주어진 유의수준에서 기대되는 귀무가설의 기각횟수보다 실제로 귀무가설을 기각하는 횟수가 적다. 즉, 이 검정법은 보수적(conservative)이다. 따라서 이 검정법을 지나치게 믿어서는 안 된다. 그러나 이 검정법은 간단하기 때문에 널리 사용되고 있다.

### 7.6.2 잔차벡터들의 표본자기상관행렬함수

앞에서는 다변량백색잡음과정  $\{v_t\}$ 의 표본자기공분산행렬함수와 이 함수의 점근분포에 대해서 살펴보았다. 지금부터는 다변량시계열데이터  $\{y_t\}$ 에 VAR( $p$ )모형을 적합시켜서 얻은 잔차벡터들의 표본자기공분산행렬함수와 표본자기상관행렬함수의 점근분포들에 대해 살펴보자.

이 절에서는  $\hat{\Phi}_1, \hat{\Phi}_2, \dots, \hat{\Phi}_p, \hat{\Sigma}$ 를 각각  $\Phi_1, \Phi_2, \dots, \Phi_p, \Sigma$ 의 최소제곱추정량 또는 점근적으로 이와 같은 성질들을 갖는 추정량이라 하자. 이 추정량은 일치성을 갖고, 점근적으로 다변량정규분포를 가지며, 또한 점근적 효율성(asymptotic efficiency)을 갖는다. 이 추정모형의 잔차벡터  $\hat{v}_t$ 를 다음과 같이 정의하자.

$$\hat{v}_t := y_t - \hat{\Phi}_1 y_{t-1} - \dots - \hat{\Phi}_p y_{t-p}$$

이 잔차벡터들의 표본자기공분산행렬과 표본자기상관행렬은 다음과 같이 정의된다.

$$\hat{C}_h := \frac{1}{T} \sum_{t=1}^{T-h} \hat{v}_{t+h} \hat{v}_t' \quad (h=0, 1, \dots)$$

$$\hat{R}_h := \hat{D}^{-1} \hat{C}_h \hat{D}^{-1} \quad (h=0, 1, \dots)$$

단,  $\hat{D}$ 는  $\hat{C}_0$ 의 대각원소들의 제곱근들로 이루어진 대각행렬이다.

Chitturi(1974, 1976)는 잔차벡터들의 표본자기공분산행렬함수와 표본자기상관행렬함수의 점근분포를 다음과 같이 제시했다.

**정리 7.23** 만일  $\{y_t\}$ 가  $d$ 변량VAR( $p$ )과정이라면 잔차벡터들의 표본자기공분산행렬함수는 다음과 같은 점근분포를 갖는다.

$$\sqrt{T}(\text{vec}(\hat{C}_1)', \dots, \text{vec}(\hat{C}_m)')' \xrightarrow{d} N(0, A_m)$$

여기서  $A_m = I_m \otimes \Sigma \otimes \Sigma - \bar{G} (\Gamma_y^{-1}(0) \otimes \Sigma) G'$ 이고,  $\Gamma_y(0)$ 은  $(y_t', y_{t-1}', \dots, y_{t-p+1}')$ 의 공분산행렬이며,  $G$ 는  $d^2 m \times d^2 m$  행렬로서 다음과 같이 정의된다.

$$\bar{G} = G \otimes I_d$$

$$\text{단, } G := \begin{bmatrix} \Sigma & \Phi_1 \Sigma & \Phi_2 \Sigma & \cdots & \Phi_{m-1} \Sigma \\ 0 & \Sigma & \Phi_1 \Sigma & \cdots & \Phi_{m-2} \Sigma \\ 0 & 0 & \Sigma & \cdots & \Phi_{m-3} \Sigma \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \Phi_{m-p} \Sigma \end{bmatrix}$$

▣

**정리 7.24** 정리 7.23의 조건하에서 잔차벡터들의 표본자기상관행렬함수는 다음과 같은 점근분포를 갖는다.

$$\sqrt{T}(\text{vec}(\hat{R}_1)', \dots, \text{vec}(\hat{R}_m)')' \xrightarrow{d} N(0, B_m)$$

여기서  $B_m = (I_m \otimes D_0^{-1} \otimes D_0^{-1}) A_m (I_m \otimes D_0^{-1} \otimes D_0^{-1})'$ 이고  $D_0$ 는 다음과 같이 정의되는  $D$ 의 확률극한행렬이다.

$$D_0 = \text{plim}_{T \rightarrow \infty} D$$

▣

정리 7.24의  $B_m$ 은 다음과 같이 쓸 수 있다.

$$B_m = I_m \otimes \Delta \otimes \Delta - H_m$$

여기서  $H_m$ 은 비음치행렬(nonnegative definite matrix)이다. 따라서 다변량백색잡음과정의 표본자기상관행렬의 원소들의 점근분산은 이에 해당하는 잔차벡터들의 표본자기상관행렬의 원소들의 점근분산보다 크거나 같다. 특히 공분산행렬  $B_m$ 의 대각원소는 1보다 작거나 같다. 이는 표본자기상관행렬  $\hat{R}_h$ 의  $(i, j)$ 원소  $\hat{r}_{i,j,h}$  ( $h > 0$ )의 분산이  $1/T$ 보다 작거나 같은 것을 의미한다. 따라서 제 7.6.1소절의 검정법을 사용해서 잔차벡터들의 백색잡음성을 조사하면, 이 검정법은 보수적인 결과를 보여 줄 것이다. 즉, 이 검정법은 주어진 유의수준에서 기대되는 귀무가설의 기각횟수보다 실제로 귀무가설을 기각하는 횟수가 적을 것이다. 특히 시차  $h(>0)$ 가 작으면  $\sqrt{T}\hat{r}_{i,j,h}$ 의 분산이 1보다 작은 경향이 있다. 반면에 시차  $h$ 가 커지면  $\sqrt{T}\hat{r}_{i,j,h}$ 의 분산은 1에 가까워진다. 이러한 성질은  $j$ 가 무한대로 커짐에 따라  $\Phi_{j-1}, \Phi_{j-2}, \dots, \Phi_{j-p}$ 는 각각 영행렬에 수렴하고 따라서 행렬  $G$ 에서 이 부분에 해당하는 부분행렬들이 영행렬로 수렴하기 때문이다.

### 7.6.3 퍼트맨토우검정통계량

잔차벡터들의 백색잡음성을 조사하기 위해서 어떤 특정시차  $h(>0)$ 의 표본자기공분산 행렬  $\hat{R}_h$ 가 유의적으로 영행렬과 다른가를 조사할 것이 아니라  $(\hat{R}_1, \hat{R}_2, \dots, \hat{R}_m)$ 이 영행렬과 유의적으로 다른가를 조사해야 할 것이다. 이러한 목적으로 사용하는 통계량이 퍼트맨토우검정통계량이다.

정리 7.23의  $A_m$ 은 다음과 같이 쓸 수 있다.

$$A_m = (I_{d^2m} - P_m)(I_m \otimes \Sigma \otimes \Sigma)$$

여기서  $P_m$ 은  $d^2m \times d^2m$  행렬이다. 만일 임의의  $i(>m-p)$ 에 대해서  $\Phi_i$ 가 영행렬에 가깝다면, 즉  $m$ 이 충분히 큰 수라면,  $P_m$ 은 멱등행렬(idempotent matrix)이며 대칭행렬이다. 따라서  $m$ 이 충분히 크면  $P_m$ 은 사영행렬(projection matrix)에 가깝다.

직적(Kronecker product)의 성질들을 이용해서 다음 식을 증명할 수 있다.

$$\begin{aligned} & T(\text{vec}(\hat{C}_1)', \dots, \text{vec}(\hat{C}_m)')(I_m \otimes \hat{\Sigma} \otimes \hat{\Sigma})^{-1}(\text{vec}(\hat{C}_1)', \dots, \text{vec}(\hat{C}_m)')' \\ &= T \sum_{h=1}^m \text{vec}(\hat{C}_h)' (\hat{\Sigma} \otimes \hat{\Sigma})^{-1} \text{vec}(\hat{C}_h) \\ &= T \sum_{h=1}^m \text{vec}(\hat{C}_h)' (\hat{C}_0^{-1} \otimes \hat{C}_0^{-1}) \text{vec}(\hat{C}_h) \\ &= T \sum_{h=1}^m \text{vec}(\hat{C}_h)' (\hat{C}_0^{-1} \otimes I_d) (I_d \otimes \hat{C}_0^{-1}) \text{vec}(\hat{C}_h) \\ &= T \sum_{h=1}^m \text{vec}(\hat{C}_h)' (\hat{C}_0^{-1} \otimes I_d)' (I_d \otimes \hat{C}_0^{-1}) \text{vec}(\hat{C}_h) \\ &= T \sum_{h=1}^m [(\hat{C}_0^{-1} \otimes I_d) \text{vec}(\hat{C}_h)] (I_d \otimes \hat{C}_0^{-1}) \text{vec}(\hat{C}_h) \\ &= T \sum_{h=1}^m \text{vec}(\hat{C}_h \hat{C}_0^{-1})' (I_d \otimes \hat{C}_0^{-1}) \text{vec}(\hat{C}_h) \\ &= T \sum_{h=1}^m [\text{vec}(\hat{C}_0^{-1} \hat{C}_h)']' (I_d \otimes \hat{C}_0^{-1}) \text{vec}(\hat{C}_h) \\ &= T \sum_{h=1}^m \text{trace}(\hat{C}_0^{-1} \hat{C}_h') \hat{C}_0^{-1} \hat{C}_h \\ &= T \sum_{h=1}^m \text{trace}(\hat{C}_h' \hat{C}_0^{-1} \hat{C}_h \hat{C}_0^{-1}) \end{aligned}$$

행렬  $(I - P_m)$ 이 점근적 사영행렬이라는 성질을 이용하면 다음 정리가 성립한다.

**정리 7.25** 만일 잔차벡터  $\{\hat{v}_t\}$ 가  $d$ 변량VAR( $p$ )과정에 최소제곱추정법 또는 점근적으로 이와 동일한 추정법을 적용해서 얻어졌다면 충분히 큰  $m$ 에 대해 다음 성질이 성립한다.

$$Q_m = T \sum_{h=1}^m \text{trace}(\hat{C}_h' \hat{C}_0^{-1} \hat{C}_h \hat{C}_0^{-1}) \xrightarrow{d} \chi^2_{d^2(m-p)} \quad \blacksquare$$



이후  $Q_m$ 을 퍼트맨토우검정통계량이라 부른다.

**따름정리 7.3** 정리 7.25의 조건하에서 충분히 큰  $m$ 에 대해 다음 성질이 성립한다.

$$T \sum_{h=1}^m \text{trace}(\hat{R}_h \hat{R}_0^{-1} \hat{R}_h \hat{R}_0^{-1}) \xrightarrow{d} \chi_{d^2(m-p)}^2 \quad \square$$

정리 7.25에서 유의할 점은 관찰점들의 개수  $T$ 가 커짐에 따라  $m$ 도 커져야 한다는 것이다. 자세한 내용은 Hosking(1980, p. 607)을 참조하라. 따라서 지나치게 작은  $m$ 에 대한 퍼트맨토우검정통계량을 사용해서는 안 된다.

[SAS4TSA2] (pp. 516~518)에 기술되어 있듯이 단변량시계열분석에 사용되는 퍼트맨토우검정통계량은 보수적(conservative)인 경향을 보인다. 즉, 모의실험을 해서 얻은 단변량시계열데이터의 퍼트맨토우검정통계량의 경험분포(empirical distribution)는 점근카이제곱분포에 비해서 왼쪽으로 질량(mass)이 모여 있다. 이와 같은 현상은 다변량시계열데이터의 퍼트맨토우검정통계량에서도 마찬가지이다. 이러한 현상을 치유하기 위해서 Hosking(1980)은 다음과 같은 변형된 퍼트맨토우검정통계량을 사용할 것을 제안하였다.

$$Q_m^* = T^2 \sum_{h=1}^m \frac{1}{T-h} \text{trace}(\hat{C}_h \hat{C}_0^{-1} \hat{C}_h \hat{C}_0^{-1})$$

**따름정리 7.4** 정리 7.25의 조건하에서 충분히 큰 수  $m$ 에 대해 다음 성질이 성립한다.

$$Q_m^* \xrightarrow{d} \chi_{d^2(m-p)}^2 \quad \square$$

오늘날 퍼트맨토우검정통계량  $Q_m$ 보다는 변형된 퍼트맨토우검정통계량  $Q_m^*$ 이 더 널리 사용되기 때문에 이후 퍼트맨토우검정통계량은  $Q_m^*$ 를 뜻하기로 한다.

Hosking(1980)은 VARMA( $p, q$ )모형의 잔차벡터들로부터 퍼트맨토우검정통계량들  $Q_m$ 과  $Q_m^*$ 을 정의했다. 이들도 정리 7.25, 따름정리 7.3과 따름정리 7.4를 만족하는데, VAR( $p$ )모형의 경우와 다른 점은 점근분포인 카이제곱분포의 자유도가  $d^2(m-p)$ 가 아닌  $d^2(m-p-q)$ 라는 것이다.

퍼트맨토우검정통계량은 여러 학자에 의해서 제시되어 왔다. Chitturi(1974)는 VAR모형의 퍼트맨토우검정통계량, Chitturi(1976)는 다변량백색잡음과정의 퍼트맨토우검정통계량, Hosking(1980)은 VARMA모형의 퍼트맨토우검정통계량, Li & McLead(1981)는 VARMA모형의 퍼트맨토우검정통계량을 제시하였다. Hosking(1981a)은 Chitturi, Hosking, 그리고 Li & McLead의 퍼트맨토우검정통계량들이 동일한 것임을 보였다. Hosking(1981b)은 이 퍼트맨토우검정통계량을 Lagrange승수검정으로 유도하였다. Li(1985)는 특수한 VAR모형에 대한 퍼트맨토우검정통계량을 유도하였다.

퍼트맨토우검정통계량을 사용해서 추정된 VAR모형을 검진하는 예를 들기 위해서, 우선 다변량시계열데이터 VAR31A를 조건부최소제곱추정법으로 3변량VAR(2)모형에 적합시

켜 잔차벡터들을 구하자. 이러한 목적으로 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM7-14.PRG에 수록되어 있다).

```

/* */
/* EXAM7-14.PRG */
/* */
/* PREPARING DATASET EXAM714B */
/* FOR PORTMANTEAU TEST */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=60;
DATA EXAM714;
 SET TSA4DATA.VAR31A;
 S1L0 = Y1;
 S1L1 = LAG(S1L0);
 S1L2 = LAG(S1L1);
 S2L0 = Y2;
 S2L1 = LAG(S2L0);
 S2L2 = LAG(S2L1);
 S3L0 = Y3;
 S3L1 = LAG(S3L0);
 S3L2 = LAG(S3L1);
 X0 = 1.0;
 IF _N_ < 3 THEN DELETE;
 DROP Y1-Y3 V1-V3;
 RUN;
/* */
/* CALCULATE THE OLS RESIDUALS */
/* */
PROC IML;
 RESET NOLOG NOCENTER NONAME;
 USE EXAM714;
 READ ALL VAR{S1L0 S2L0 S3L0} INTO YTR;
 READ ALL VAR{X0 S1L1 S2L1 S3L1
 S1L2 S2L2 S3L2} INTO XTR;
 CLOSE EXAM714;
 VARORDER = 2;
 PHISTHAT = YTR' * XTR * INV(XTR' * XTR);
 RESIDTR = YTR' - PHISTHAT * XTR';
 RESID = RESIDTR';

```



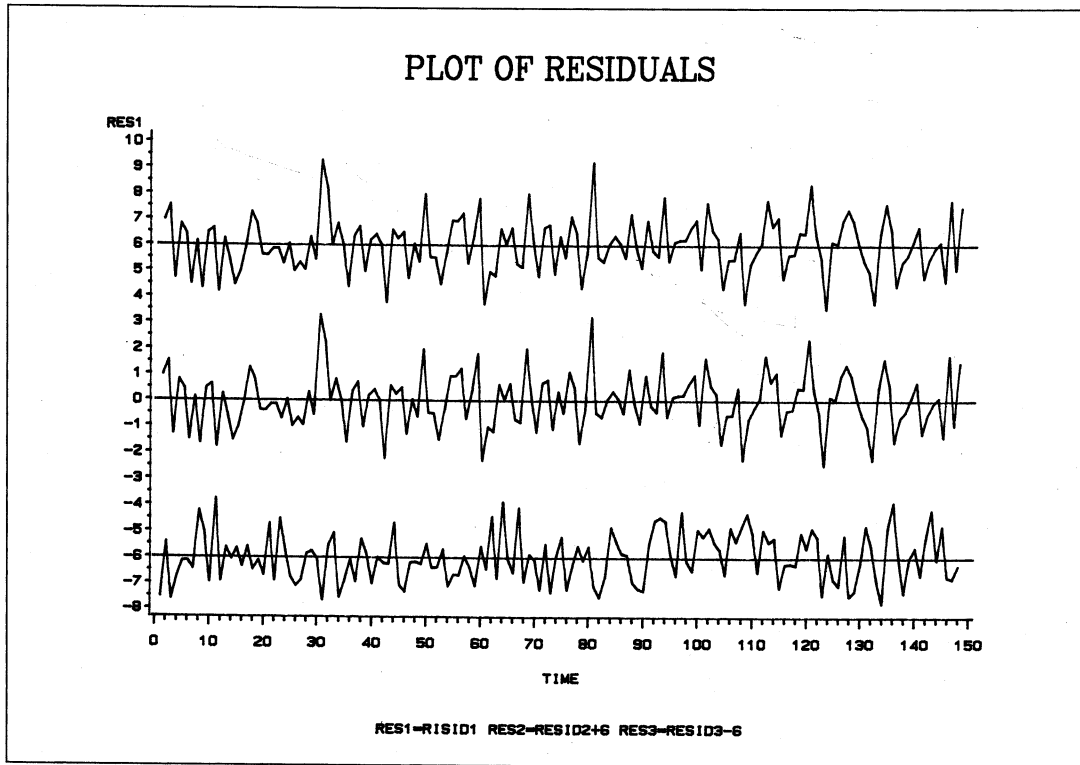


그림 7.14 잔차벡터들의 시계열산포도

```

/* DATASET = EXAM714B */
LIBNAME TSA4DATA 'A:\TSA4DATA';

/* */
/* CALCULATE THE PORTMANTEAU STATISTIC */
/* */
PROC IML;
RESET NOLOG NOCENTER NONAME;

/* */
/* MODULE FOR PORTMANTEAU STATISTIC */
/* */

START PORTMANT(D, T, RESID1, RESID2, RESID3, VARORDER);
/* MAKE THE MATRIX OF RESIDUALS */
RESIDMAT = RESID1 || RESID2 || RESID3;
MAXLAG = INT(T/ 3); /* MAX ORDER OF COV TERMS */
COVRESID = COVLAG(RESIDMAT, MAXLAG+1);
FREE / D T MAXLAG VARORDER COVRESID;
COV0 = COVRESID[, 1:D];
IC0 = INV(COV0);
FREE COV0;

```

```

PMAT = J(D, D, 0);
MD6 = INT(MAXLAG/ 6);
DO K = 0 TO MD6-1;
 K6 = 6*K;
 DO J = (K6+1) TO (K6+6);
 COLSJ = (J*D+1):(J+1)*D;
 COVJ = COVRESID[, COLSJ];
 PMAT = PMAT + COVJ' * IC0 * COVJ * IC0 / (T-J);
 END;
PORTMANT = T*T*TRACE(PMAT);
LAGORDER = 6*(K+1);
DOF = D*D*(LAGORDER - VARORDER);
PRINT "SUM UP TO LAG = " LAGORDER
 " PORTMANTEAU = " PORTMANT " D. F. = " DOF;
END;
FINISH;
/ * * /
/ * APPLY THE MODULE PORTMANT * /
/ * * /
USE TSA4DATA.EXAM714B;
READ ALL;
CLOSE TSA4DATA.EXAM714B;
D = 3; / * DIM OF VAR TIME SERIES * /
T = 148; / * NO OF RESIDUALS * /
VARORDER = 2; / * OREDER OF VAR MODEL * /
RUN PORTMANT(D, T, RESID1, RESID2, RESID3, VARORDER);
QUIT;

```

이 SAS프로그램의 밑에서 5번째 줄에서는 변수 D에 다변량시계열의 차원을 3으로 지정했고, 4번째 줄에서는 변수 T에 잔차벡터들의 개수를 148개로 지정했으며, 3번째 줄에서는 변수 VARORDER에 VAR모형의 차수를 2로 지정했다. 이 다변량시계열데이터의 차원이 3이므로 RUN스태이트먼트에는 3개의 잔차변수들 RESID1, RESID2, 그리고 RESID3가 사용되었다. 만일 다변량시계열의 차원이 3이 아니면 모듈 PORTMANT의 첫째 줄과 셋째 줄을 그리고 RUN스태이트먼트를 수정해야 한다. 만일 다변량시계열의 차원이 2이면 다음과 같이 수정한다.

```

START PORTMANT (D, T, RESID1, RESID2, VARORDER);
RESIDMANT = RESID1 || RESID2;
RUN PORTMANT(D, T, RESID1, RESID2, VARORDER);

```

| SAS             |    | 9:07 Saturday, July 3, 1993 1 |                      |
|-----------------|----|-------------------------------|----------------------|
| SUM UP TO LAG = | 6  | PORTMANTEAU =                 | 31.045188 D.F. = 36  |
| SUM UP TO LAG = | 12 | PORTMANTEAU =                 | 92.913635 D.F. = 90  |
| SUM UP TO LAG = | 18 | PORTMANTEAU =                 | 134.10445 D.F. = 144 |
| SUM UP TO LAG = | 24 | PORTMANTEAU =                 | 179.30364 D.F. = 198 |
| SUM UP TO LAG = | 30 | PORTMANTEAU =                 | 227.03566 D.F. = 252 |
| SUM UP TO LAG = | 36 | PORTMANTEAU =                 | 276.88054 D.F. = 306 |
| SUM UP TO LAG = | 42 | PORTMANTEAU =                 | 331.44727 D.F. = 360 |
| SUM UP TO LAG = | 48 | PORTMANTEAU =                 | 376.79603 D.F. = 414 |
| Exiting IML.    |    |                               |                      |

그림 7.15 퍼트맨투우검정통계량을 사용한 VAR모형의 검진

이 SAS프로그램을 수행한 결과가 그림 7.15이다. 그림 7.15로부터 추정된 VAR(2)모형의 잔차벡터들이 백색잡음성을 갖는다고 결론지을 수 있다.

### ▣ 참고문헌 ▣

- Akaike, H. (1969), Fitting autoregressive models for prediction, *Ann. Inst. Statist. Math.*, vol. 21, pp. 407~419.
- Akaike, H. (1974), A new look at the statistical model identification, *IEEE Trans. Automatic Control*, vol. AC~19, pp. 716~723.
- Akaike, H. (1977), An objective use of Bayesian models, *Ann. Inst. Statist. Math.*, vol. 29, pp. 9~20.
- Ansley, C. F. & P. Newbold(1979), Multivariate partial autocorrelations, *ASA Proceedings of Business and Economic Statistics Section*, pp. 349~353.
- Atkinson, A. C. (1978), Posterior probabilities for choosing a regression model, *Biometrika*, vol. 65, pp. 39~48.
- Atkinson, A. C. (1980), A note on the generalized information criterion for choice of a model, *Biometrika*, vol. 67, pp. 413~418.
- Bartlett, M. S. (1938), Further aspects of the theory of multiple regression, *Proceeding of the Cambridge Philosophical Society*, vol. 34, pp. 180~185.
- Bartlett, M. S. (1946), On the theoretical specification and sampling properties of autocorrelated time series, *J. Roy. Statist. Soc., Suppl.*, vol. 8, pp. 27~41.

- Chitturi, R. V. (1974), Distribution of residual autocorrelations in multiple autoregressive schemes, *J. Amer. Statist. Assoc.*, vol. 69, pp. 928~934.
- Chitturi, R. V. (1976), Distribution of multivariate white noise autocorrelations, *J. Amer. Statist. Assoc.*, vol. 71, pp. 223~226.
- Choi, B. S. (1994), On the covariance matrix estimators of the white noise process of a vector autoregressive model, *Commun. Statist. - Theory Meth.*, vol. 23, pp. 249~256.
- Chow, G. C. (1981), A comparison of the information and posterior probability criteria for model selection, *J. Econometrics*, vol. 16, pp. 21~33.
- Dufour, J. M. & R. Roy(1985), Some robust exact results on sample autoregressions and tests of randomness, *J. Econometrics*, vol. 29, pp. 257~273.
- Fuller, W. A. (1976), *Introduction to Statistical Time Series*, Wiley, New York.
- Hannan, E. J. (1970), *Multiple Time Series*, Wiley, New York.
- Hannan, E. J. & B. G. Quinn(1979), The determination of the order of an autoregression, *J. Roy. Statist. Soc., Ser. B*, vol. 41, pp. 190~195.
- Heyse, J. & W. W. S. Wei(1985), The partial lag autocorrelation function, *Technical Report* No. 32, Department of Statistics, Temple University.
- Hosking, J. R. M. (1980), The multivariate portmanteau statistic, *J. Amer. Statist. Assoc.*, vol. 75, pp. 602~607.
- Hosking, J. R. M. (1981a), Equivalent forms of the multivariate portmanteau statistic, *J. Roy. Statist. Soc., Ser. B*, vol. 43, pp. 261~262. Correction: vol. 51, (1989), p. 303.
- Hosking, J. R. M. (1981b), Lagrange-multiplier tests of multivariate time-series models, *J. Roy. Statist. Soc., Ser. B*, vol. 43, pp. 219~230.
- Jenkins, G. M. & A. S. Alavi(1981), Some aspects of modelling and multivariate time series, *J. Time Series Anal.*, vol. 2, pp. 1~47.
- Kashyap, R. L. (1977), A bayesian comparison of different classes of dynamic models using empirical data, *IEEE Trans. Automatic Control*, vol. AC-22, pp. 715~727.
- Koreisha, S. G. & T. Pukkila(1983), Determining the order of a vector autoregression when the number of component series is large, *J. Time Series Anal.*, vol. 14, pp. 47~69.
- Li, W. K. (1985), Distribution of residual autocorrelations in multivariate autoregressive index models, *Biometrika*, vol. 72, pp. 686~688.
- Li, W. K. & A. I. McLead(1981), Distribution of the residual autocorrelations in multivariate ARMA time series models, *J. Roy. Statist. Soc., Ser. B*, vol. 43, pp. 231~239.

- Lütkepohl, H. (1985), Comparison of criteria for estimating the order of a vector autoregressive process, *J. Time Series Anal.*, vol. 6, pp. 35~52.
- Lütkepohl, H. (1991), *Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.
- Parzen, E. (1977), Multiple time series: Determining the order of approximating autoregressive schemes, in *Multivariate Analysis IV*, P. R. Krishnaiah, Ed., North-Holland, Amsterdam, pp. 283~295.
- Paulsen, J. (1984), Order determination of multivariate autoregressive time series with unit roots, *J. Time Series Anal.*, vol. 5, pp. 115~127.
- Paulsen, J. & D. Tjøstheim(1985), On the estimation of residual variance and order in autoregressive time series, *J. Roy. Statist. Soc., Ser. B*, vol. 47, pp. 216~228.
- Poskitt, D. S. & A. R. Tremayne(1983), On the posterior odds of time series models, *Biometrika*, vol. 70, pp. 157~162.
- Quenouille, M. H. (1957), *The Analysis of Multiple Time Series*, Griffin, London.
- Quinn, B. G. (1980), Order determination for a multivariate autoregression, *J. Roy. Statist. Soc., Ser. B*, vol. 42, pp. 182~185.
- Quinn, B. G.(1988), A note on AIC order determination for multivariate autoregressions, *J. Time Series Anal.*, vol. 9, pp. 241~245.
- Rissanen, J. (1978), Modeling by shortest data description, *Automatica*, vol. 14, pp. 465~471.
- Schwarz, G. (1978), Estimating the dimension of a model, *Ann. Statist.*, vol. 6, pp. 461~464.
- Smith, A. F. M. & D. J. Spiegelhalter(1980), Bayes factors and choice criteria for linear models, *J. Roy. Statist. Soc., Ser. B*, vol. 42, pp. 213~220.
- Tiao, G. C. & G. E. P. Box(1981), Modelling multivariate time series with applications, *J. Amer. Statist. Assoc.*, vol. 76, pp. 802~816.
- Wei, W. S. (1990), *Time Series Analysis: Univariate and Multivariate Methods*, Addison-Wesley, Redwood City.



## 상태공간모형

- 8.1 상태공간모형
- 8.2 상태공간모형화의 기초
- 8.3 상태공간모형화
- 8.4 PROC STATESPACE의 자세한 사용법
- 8.5 수리적 배경
- 8.6 상태공간모형화방법의 예제들



## 제 8 장 상태공간모형

### 8.1 상태공간모형

상태공간모형화(state space modeling)는 Kalman(1960)과 Kalman & Buch(1961)에 의해 처음으로 소개되었고 칼만필터링(Kalman filtering)으로도 알려져 있다. 상태공간모형은 원래 항공학 분야에서 개발된 현대제어이론(modern control theory)의 기본적인 모형 중의 하나이나 오늘날에는 경제학, 의학, 공학 등 여러 분야에서 널리 이용되는 예측모형으로서 단변량시계열분석과 다변량시계열분석 모두에 사용된다. Akaike(1974a)는 상태공간모형을 VARMA모형과 연결시켜서 VARMA모형화에서 발생하는 식별의 비유일성 문제를 해결하는 방법을 제시하였다. 이 장에서는 Akaike의 상태공간모형을 이용한 VARMA모형의 모형화(modeling)와 예측(prediction)에 대해서 다루고자 한다. 독자들이 유의할 것은 이 Akaike의 상태공간모형화방법에서는 상태공간모형을 사용하지만 칼만필터링을 직접 사용하는 것은 아니라는 점이다.

상태공간모형을 이용한 VARMA모형의 모형화와 예측은 이해하기가 쉬운 내용이 아니다. 따라서 본서에서는 Akaike의 상태공간모형화방법을 세 단계로 나누어 설명하고자 한다. 제 8.1절에서는 제1단계로 간단한 예제를 가지고 SAS/ETS의 PROC STATESPACE의 기초적인 사용법을 설명한다. 제 8.2절에서는 제2단계로 상태공간모형의 이론에 대해서 설명한다. 제 8.3절에서는 PROC STATESPACE의 자세한 사용법을 설명한다. 제 8.4절에서는 제3단계로 Akaike의 상태공간모형화방법의 수리적 측면을 자세하게 설명한다. 제 8.5절에서는 여러 예제들을 분석해서 상태공간모형을 이용한 VARMA모형화의 유용성을 보인다.

## 8.2 상태공간모형화의 기초

### 8.2.1 PROC STATESPACE의 기초적 사용법

시스템의 상태(state)는 미래를 예측하기 위해서 필요한 현재와 과거의 정보의 최소집합(minimal set)으로 정의된다. 따라서 시스템의 미래현상(future behavior)을 현재의 상태와 미래의 입력(input)을 사용해서 완전히 설명할 수 있다. 즉, 어느 시스템의 상태벡터(state vector)는 미래의 모든 값(또는 벡터)을 예측하기 위해서 필요한 현재와 과거의 정보의 최소량으로 정의된다. 따라서 시스템의 미래 출력벡터(output vector)는 현재의 상태벡터(state vector)와 미래의 입력벡터(input vector)를 사용해서 계산할 수 있다. 상태공간모형은 현재의 상태가 주어졌을 때 시스템의 미래는 현재의 상태에 의존한다는 마아코프성질(Markov property)을 바탕으로 하므로 시스템의 상태공간표현(state space representation)을 마아코비안표현(Markovian representation)이라 부르기도 한다.

상태공간모형화방법이 단변량시계열분석에서도 쓰이기는 하지만 이 방법의 주된 사용목적은 다변량시계열분석에 있다. ARMA모형은 VARMA모형의 특수한 경우이다. 따라서 이후 이 장에서는 특별한 언급이 없는 한 ARMA모형은 VARMA모형에 포함되는 것으로 한다.

Akaike(1974a)는 정상적  $d$ 변량시계열데이터  $\{y_t\}$ 가 VARMA과정이면 다음과 같은 상태공간모형으로 나타낼 수 있음을 보였다.

$$z_{t+1} = Fz_t + Gu_{t+1}$$

$$y_t = [I_d \ O] z_t$$

단,  $k \geq d$

$z_t$ 는  $k \times 1$ 인 상태벡터(state vector)

$F$ 는  $k \times k$ 인 변환행렬(transition matrix)

$G$ 는  $k \times d$ 인 입력행렬(input matrix)

$u_t$ 는  $d \times 1$ 인 쇄신벡터(innovation vector), 또는

충격벡터(shock vector)로서 평균벡터가

영벡터이고 공분산행렬이  $\Sigma$ 임

$y_t$ 는  $d \times 1$ 인 출력벡터(output vector)

이 상태공간모형의 첫 번째 방정식을 상태방정식(state equation)이라 하고 두 번째 방정식을 출력방정식(output equation)이라고 한다. 시점  $t$ 까지의 시계열데이터가 주어졌을 때 시점  $(t+k)$ 의 최량예측(best predictor)은 다음과 같다.

$$y_{t+k|t} = E(y_{t+k} | y_t, y_{t-1}, \dots, y_1)$$

물론  $y_{t|t} = y_t$ 이다. 상태벡터  $z_t$ 의 첫  $d$ 개 원소들은  $y_t$ 의 원소들이고 나머지 원소들은  $y_{t+1|t}$ ,  $y_{t+2|t}$ , ... 들의 원소들 중에서 해당되는 것들로 구성할 수 있다.

VARMA모형을 상태공간모형으로 나타내는 몇 가지 예제들을 살펴보자.

[예제 1] 

---

다음과 같은 AR(1)모형을 살펴보자.

$$y_t = 0.7y_{t-1} + v_t$$

이 AR(1)모형은 다음과 같이 상태공간모형으로 나타낼 수 있다.

$$z_{t+1} = Fz_t + Gv_{t+1}$$

$$y_t = z_t$$

여기서  $F = [0.7]$ ,  $G = [1]$ 이다.

[예제 2] 

---

다음과 같은 AR(2)모형을 살펴보자.

$$y_t = 0.6y_{t-1} + 0.3y_{t-2} + v_t$$

이 AR(2)모형은 다음과 같이 상태공간모형으로 나타낼 수 있다.

$$z_{t+1} = Fz_t + Gv_{t+1}$$

$$y_t = [1 \ 0] z_t$$

$$\text{단, } z_t = (y_t, y_{t+1|t})'$$

$$F = \begin{bmatrix} 0 & 1 \\ 0.3 & 0.6 \end{bmatrix}, \quad G = \begin{bmatrix} 1 \\ 0.6 \end{bmatrix}$$

[예제 3] 

---

다음과 같은 MA(1)모형을 살펴보자.

$$y_t = v_t + 0.7v_{t-1}$$

이 MA(1)모형은 다음과 같이 상태공간모형으로 나타낼 수 있다.

$$z_{t+1} = Fz_t + Gv_{t+1}$$

$$y_t = [1 \ 0] z_t$$

$$\text{단, } z_t = (y_t, y_{t+1|t})'$$

$$F = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad G = \begin{bmatrix} 1 \\ 0.7 \end{bmatrix}$$

## [예제 4]

다음과 같은 2변량VARMA(1, 1)모형을 살펴보자.

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 0.45 & -0.11 \\ -0.14 & -0.23 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} - \begin{bmatrix} -0.40 & 0.12 \\ 0.00 & 0.00 \end{bmatrix} \begin{bmatrix} v_{1,t-1} \\ v_{2,t-1} \end{bmatrix}$$

이 VARMA(1, 1)모형은 다음과 같은 상태공간모형으로 나타낼 수 있다.

$$z_{t+1} = Fz_t + Gv_{t+1}$$

$$y_t = [I_2 \ 0] z_t$$

$$\text{단, } z_t = (y_{1,t}, y_{2,t}, y_{1,t+1|t})'$$

$$F = \begin{bmatrix} 0 & 0 & 1 \\ -0.14 & -0.23 & 0 \\ 0.0154 & 0.0253 & 0.45 \end{bmatrix}, \quad G = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0.05 & -0.23 \end{bmatrix}$$

상태벡터  $z_t$ 의 첫  $d$ 개 원소들은  $y_t$ 의 원소들이므로  $z_t$ 의 나머지 원소들을 상태공간모형에서 제거함으로써 VARMA모형을 유도할 수 있다. 따라서 Akaike의 상태공간모형화는 VARMA모형화의 한 방법으로 사용된다고 할 수 있다. 그러나 일반적으로는 추정된 상태공간모형을 VARMA모형으로 변환시키지는 않고 추정된 상태공간모형 자체를 예측등의 목적에 직접 사용한다.

SAS/ETS의 PROC STATESPACE를 사용해서 단변량시계열데이터를 분석하는 예를 살펴보기 위해서 우선 다음과 같은 AR(2)모형에서 관찰점들이 200개인 모의실험데이터를 생성하자.

$$y_t = 0.6y_{t-1} + 0.3y_{t-2} + v_t$$

이러한 목적으로 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4 TSA4의 파일 EXAM8-1.PRG에 수록되어 있다).

```

/* */
/* EXAM8-1.PRG */
/* */
/* GENERATE AR(2) DATA VAR12A */
/* WITH COEFFICIENTS=0.6, 0.3 & T=200 */
/* */
LIBNAME TSA4DATA 'A:\TSA4DATA';
DATA EXAM81;
 START:T+1;
 IF Y1=. THEN Y1=0;
 IF Y2=. THEN Y2=0;

```

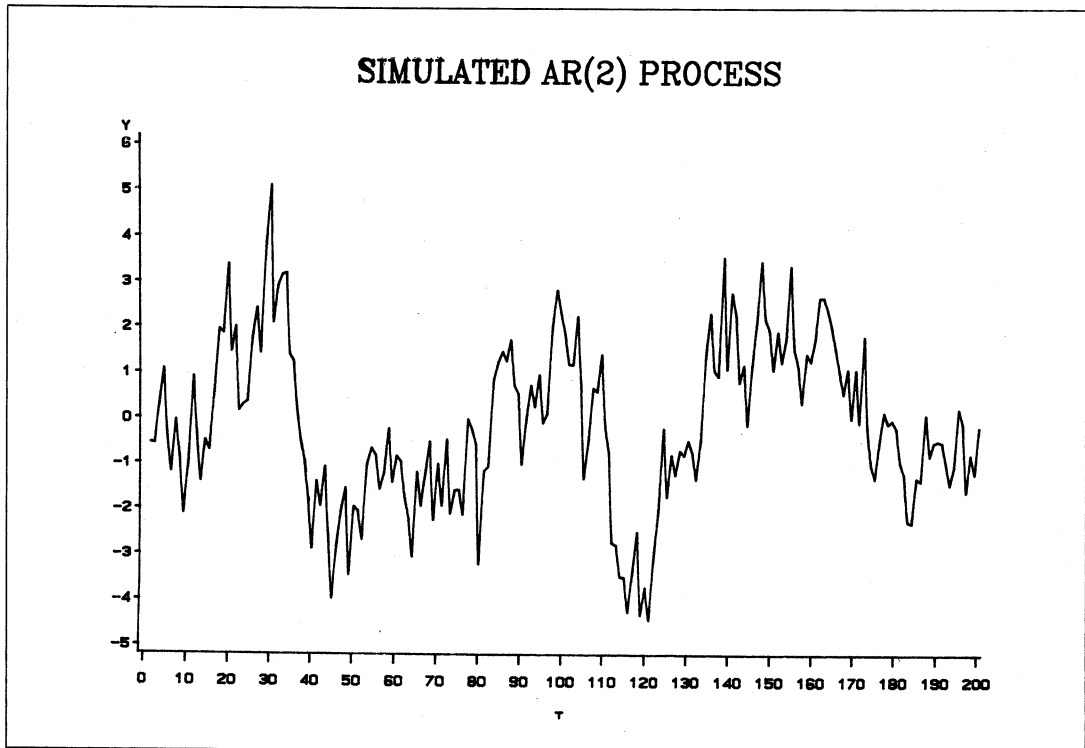


그림 8.1 단변량시계열데이터 VAR12의 시계열산포도

```

Y=0.6*Y1 + 0.3*Y2 + RANNOR(7654321);
OUTPUT;
Y2=Y1;
Y1=Y;
IF T < 300 THEN GO TO START;
RUN;
DATA TSA4DATA.VAR12A;
SET EXAM81;
T=T-100;
IF T < 1 THEN DELETE;
KEEP Y;
RUN;
PROC GPLOT;
PLOT Y*T / HMINOR=0 VMINOR=0;
SYMBOL1 C=BLACK V=NONE I=JOIN LINE=1;
TITLE1 'SIMULATED AR(2) PROCESS';
RUN;

```

이 SAS프로그램의 DATA스텝들이 수행되면 SAS데이터라이브러리 TSA4DATA에 영구SAS데이터세트 VAR12A.SSD가 만들어진다. 또한, PROC GPLOT가 수행되면 그림 8.1

과 같은 시계열산포도가 그려진다. 이 시계열산포도로부터 이 단변량시계열데이터가 정상적임을 알 수 있다. 독자들은 정상성조건이 상태공간모형화의 필요조건임을 반드시 기억해야 한다.

이 영구SAS데이터세트의 변수 Y를 상태공간모형화하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM8-2.PRG에 수록되어 있다).

```

/* */
/* EXAM8-2.PRG */
/* */
/* STATE SPACE MODELING OF VAR12A */
/* */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=80 LINESIZE=72;
PROC STATESPACE DATA=TSA4DATA.VAR12A;
 VAR Y;
RUN;

```

이 SAS프로그램을 수행한 결과가 그림 8.2이다. 그림 8.2의 각 부분을 설명하면 다음과 같다.

- (A) 관찰점들의 개수  $T$
- (B) 각 변수의 변수이름, 표본평균과 표본표준편차, 단일 변수에 차분을 취하면 차분의 차수
- (C) 시차 0에서 10까지의 VAR모형들의 AIC값들
- (D) 표본자기상관행렬함수의 도표적 표현(schematic representation)
- (E) 표본부분자기상관행렬함수의 도표적 표현
- (F) 최소AIC추정값을 갖는 VAR모형에서 VAR계수행렬들의 Yule-Walker추정행렬들
- (G) 선택된 상태공간모형의 예비추정값들
- (G1) 상태벡터  $z_t$
- (G2) 변환행렬  $F$ 의 예비추정행렬
- (G3) 입력행렬  $G$ 의 예비추정행렬
- (G4) 쇄신벡터의 공분산행렬  $\Sigma$ 의 예비추정행렬
- (H) 선택된 상태공간모형의 최종추정값들
- (H1) 상태벡터  $z_t$
- (H2) 변환행렬  $F$ 의 최종추정행렬
- (H3) 입력행렬  $G$ 의 최종추정행렬
- (H4) 쇄신벡터의 공분산행렬  $\Sigma$ 의 최종추정행렬
- (I) 최종적으로 선택된 상태공간모형에서 행렬들  $F$ 와  $G$ 의 원소들, 이들의 추정값들, 표준오차들, 그리고  $t$ 값들



그림 8.2를 다음과 같이 해석할 수 있다.

- ① 부분 A와 부분 B에서 분석대상이 되는 변수 Y의 관찰점들의 개수는 200개이고 표본평균은 0, 표본표준편차는 1.791이라는 것을 알 수 있다.
- ② 부분 C에서 변수 Y를 차수가 0에서 10까지의 VAR모형들에 적합시킬 때 AIC값을 최소로 하는 차수가 2라는 것을 알 수 있다.
- ③ 부분 D의 표본자기상관함수의 도표적 표현으로부터 이 표본자기상관함수가 시나브로형태를 가짐을 알 수 있다.
- ④ 부분 E의 표본부분자기상관함수의 도표적 표현으로부터 이 표본부분자기상관함수가 시차 3 이후부터 절단형태를 가짐을 알 수 있다. 이 절단형태와 부분 D의 표본자기상관함수의 시나브로형태를 바탕으로 이 시계열데이터가 AR(2)모형을 갖는 것을 확인할 수 있다.
- ⑤ 부분 F로부터 최소AIC값을 갖는 AR(2)모형의 Yule-Walker 추정모형은 다음과 같음을 알 수 있다.

$$y_t = Y(t) + 0.119$$

$$y_t = 0.668y_{t-1} + 0.181y_{t-2} + v_t$$

- ⑥ 부분 G으로부터 초기에 선택된 상태벡터는 다음과 같음을 알 수 있다.

$$z_t = (y_t, y_{t+1|t})'$$

- ⑦ 부분 G로부터 초기에 선택된 상태공간모형의 추정행렬들은 다음과 같음을 알 수 있다.

$$\hat{F} = \begin{bmatrix} 0 & 1 \\ 0.063 & 0.812 \end{bmatrix}, \quad \hat{G} = \begin{bmatrix} 1 \\ 0.668 \end{bmatrix}, \quad \hat{\Sigma} = [1.039]$$

- ⑧ 부분 H으로부터 최종적으로 선택된 상태벡터는 다음과 같음을 알 수 있다.

$$z_t = (y_t, y_{t+1|t})'$$

따라서 시점  $t$ 에서  $y_t$ 와  $y_{t+1|t}$ 의 값들만 있으면 모든 미래시점들에서의 값들을 예측할 수 있다.

- ⑨ 부분 H로부터 최종적으로 선택된 상태공간모형의 추정행렬들은 다음과 같음을 알 수 있다.

$$\hat{F} = \begin{bmatrix} 0 & 1 \\ -0.104 & 1.014 \end{bmatrix}, \quad \hat{G} = \begin{bmatrix} 1 \\ 0.662 \end{bmatrix}, \quad \hat{\Sigma} = [1.038]$$

STATESPACE Procedure

(A)

Nobs = 200

(B)

| Variable | Mean     | Std      |
|----------|----------|----------|
| Y        | -0.11884 | 1.791011 |

(C)

Information Criterion for Autoregressive Models

| Lag=0      | Lag=1     | Lag=2     | Lag=3     | Lag=4     |
|------------|-----------|-----------|-----------|-----------|
| 233.112018 | 16.334461 | 11.671213 | 13.630674 | 14.081670 |
| Lag=5      | Lag=6     | Lag=7     | Lag=8     | Lag=9     |
| 16.036176  | 17.977597 | 19.060161 | 17.062326 | 19.048223 |
| Lag=10     |           |           |           |           |
| 21.025320  |           |           |           |           |

(D)

Schematic Representation of Correlations

| Name/Lag | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|----------|---|---|---|---|---|---|---|---|---|---|----|
| Y        | + | + | + | + | + | + | + | + | + | + | +  |

+ is > 2\*std error, - is < -2\*std error, . is between

(E)

Schematic Representation of Partial Autocorrelations

| Name/Lag | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|----------|---|---|---|---|---|---|---|---|---|----|
| Y        | + | + | . | . | . | . | . | . | . | .  |

+ is > 2\*std error, - is < -2\*std error, . is between

(F)

Yule-Walker Estimates for the Min AIC

|   | Lag=1   | Lag=2   |
|---|---------|---------|
| Y | 0.66790 | 0.18102 |

(G)

Selected Statespace Form and Preliminary Estimates

(G1)

State Vector

Y(T;T)                  Y(T+1;T)

(G2)

Estimate of the Transition Matrix

|  |       |       |
|--|-------|-------|
|  | 0     | 1     |
|  | 0.065 | 0.812 |

(G3)

Input Matrix for the Innovation

|  |       |
|--|-------|
|  | 1     |
|  | 0.668 |

(G4)

Variance Matrix for the Innovation

1.03910115

(H)

Selected Statespace Form and Fitted Model

(H1)

State Vector

Y(T;T)                  Y(T+1;T)

(H2)

Estimate of the Transition Matrix

|  |        |       |
|--|--------|-------|
|  | 0      | 1     |
|  | -0.104 | 1.014 |

|      |                                    |            |           |          |
|------|------------------------------------|------------|-----------|----------|
| (H3) | Input Matrix for the Innovation    |            |           |          |
|      |                                    | 1          |           |          |
|      |                                    | 0.662      |           |          |
| (H4) | Variance Matrix for the Innovation |            |           |          |
|      |                                    | 1.03767381 |           |          |
| (I)  | Parameter Estimates                |            |           |          |
|      | Parameter                          | Estimate   | Std. Err. | T value  |
|      | F(2,1)                             | -0.10372   | 0.295959  | -0.35046 |
|      | F(2,2)                             | 1.01387    | 0.350407  | 2.893406 |
|      | G(2,1)                             | 0.66223    | 0.070664  | 9.37158  |

그림 8.2 단변량시계열데이터 VAR12A의 상태공간모형화

⑩ 부분 1로부터 선택된 상태공간모형의 최종추정행렬들의 원소들이 0과 유의적으로 다른지 여부를 조사할 수 있다. 이 예제에서는 원소들  $F(2, 2)$ 와  $G(2, 1)$ 의 추정값들이 유의적으로 0과 다르다고 할 수 있다. 원소  $F(2, 1)$ 의 추정값은 유의적으로 0과 다르지 않다. 이러한 경우에는 이 원소를 0으로 고정하고 상태공간모형을 추정할 수도 있다. 이에 대한 자세한 내용은 제 8.1.2 소절을 참조하라.

상태공간모형이 구해지면 이 상태공간모형으로부터 VARMA모형을 유도할 수 있다. 그림 8.2에서 최종적으로 구한 상태공간모형은 다음과 같다.

$$\begin{aligned}
 y_t &= Y(t) + 0.119 \\
 z_{t+1} &= Fz_t + Gv_{t+1} \\
 y_t &= [1 \ 0] z_t \\
 \text{단, } z_t &= (y_{1t}, y_{t+1|t})' \\
 \hat{F} &= \begin{bmatrix} 0 & 1 \\ -0.104 & 1.014 \end{bmatrix}, \quad \hat{G} = \begin{bmatrix} 1 \\ 0.662 \end{bmatrix}, \quad \hat{\Sigma} = [1.038]
 \end{aligned}$$

이 상태공간모형식을 전개하면 다음과 같다.

$$\begin{aligned}
 y_{t+1} &= y_{t+1|t} + v_{t+1} \\
 y_{t+2|t+1} &= -0.104y_t + 1.014y_{t+1|t} + 0.662v_{t+1}
 \end{aligned}$$

첫 번째 식을 두 번째 식에 대입하면 다음 식을 얻는다.

$$y_{t+2} - v_{t+2} = -0.104y_t + 1.014(y_{t+1} - v_{t+1}) + 0.662v_{t+1}$$

이 식을 정리하면 다음과 같은 ARMA(2, 1)모형을 얻는다.

$$y_t = -1.014y_{t-1} - 0.104y_{t-2} + v_t - 0.352v_{t-1}$$

이 ARMA(2, 1)모형의 AR특성근의 하나는 1에 가깝다. 즉, 이 ARMA(2, 1)모형은 정상성조건을 만족하지 못한다.

SAS/ETS의 PROC STATESPACE를 사용한 다변량시계열데이터를 분석하는 예를 살



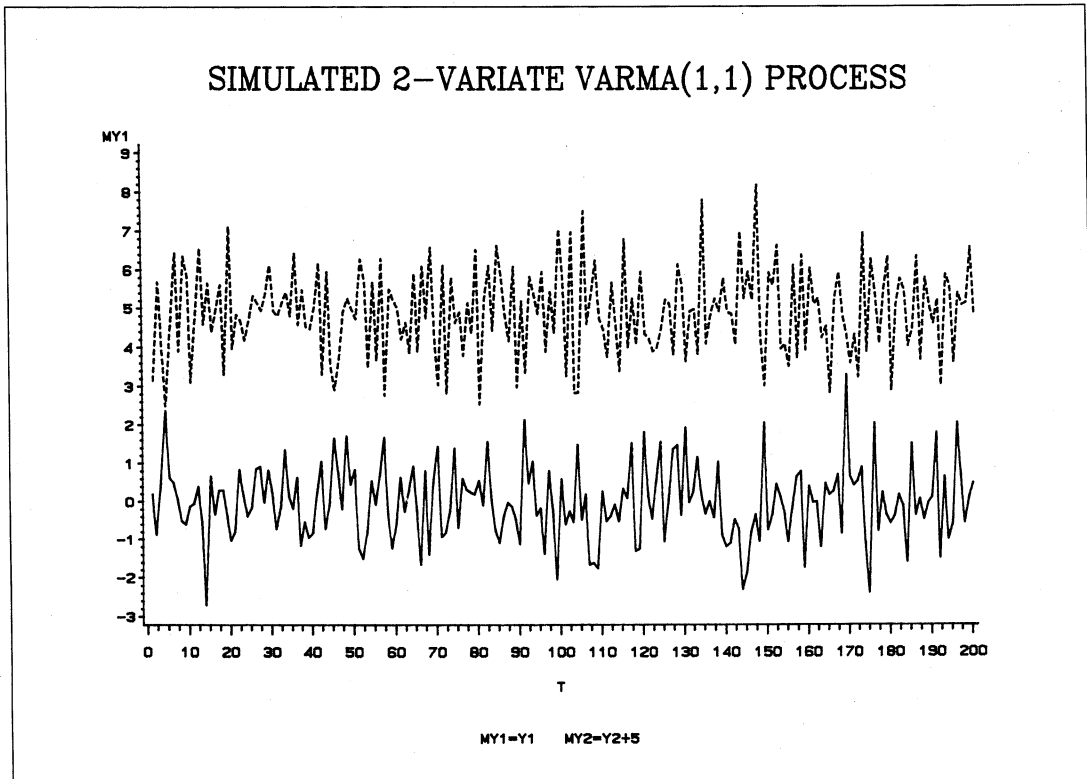


그림 8.3 다변량시계열데이터 VARMA211의 시계열산포도

```

RUN;
TITLE1 'SIMULATED 2-VARIATE VARMA(1,1) PROCESS';
DATA EXAM83B;
 SET TSA4DATA.VARMA211;
 MY1 = Y1;
 MY2 = Y2 + 5.0;
 KEEP T MY1 MY2;
RUN;
PROC GPLOT;
 PLOT (MY1 MY2)*T / OVERLAY;
 SYMBOL1 C=BLACK V=NONE I=JOIN LINE=1;
 SYMBOL2 C=BLACK V=NONE I=JOIN LINE=2;
 FOOTNOTE 'MY1=Y1 MY2=Y2+5';
RUN;

```

이 SAS프로그램의 DATA스텝들을 수행하면 SAS데이터라이브러리 TSA4DATA에 영구SAS데이터세트 VARMA211.SSD가 만들어진다. PROC GPLOT가 수행되면 그림 8.3과 같은 시계열산포도가 그려진다. 이 시계열산포도로부터 이 다변량시계열데이터가 정상성 조건을 만족한다는 것을 알 수 있다.

이 영구SAS데이터세트의 변수들 Y1과 Y2를 상태공간모형화하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM8-4.PRG에 수록되어 있다).

```

/* */
/* EXAM8-4.PRG */
/* */
/* STATE SPACE MODELING OF VARMA211 */
/* */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=80 LINESIZE=72;
PROC STATESPACE DATA=TSA4DATA.VARMA211;
 VAR Y1 Y2;
 RUN;

```

이 SAS프로그램을 수행한 결과가 그림 8.4이다. 그림 8.4의 각 부분을 해석하면 다음과 같다.

- ① 부분 A와 부분 B에서 분석대상이 되는 변수들 Y1과 Y2의 관찰점들의 개수는 200이고 변수 Y1의 표본평균은  $-0.0071$ , 표본표준편차는  $0.9694$ 이며, 변수 Y2의 표본평균은  $-0.0643$ , 표본표준편차는  $1.1238$ 인 것을 알 수 있다.
- ② 부분 C에서 이 2변량시계열데이터를 차수가 0에서 10까지의 VAR모형들에 적합시킬 때 AIC값을 최소로 하는 차수가 2라는 것을 알 수 있다.
- ③ 부분 D의 표본자기상관행렬함수의 도표적 표현으로부터 이 다변량시계열데이터가 정상성을 가짐을 알 수 있다. 또한 이 표본자기상관행렬함수가 시나브로형태, 또는 시차 3 이후에서 절단형태를 갖는다고 할 수 있다. 그러나 시차 10에서의 표본자기상관행렬이 유의적으로 영행렬과 다르므로 시나브로형태라고 하는 것이 좋겠다.
- ④ 부분 E의 표본부분자기상관행렬함수의 도표적 표현으로부터 이 표본부분자기상관행렬함수가 시나브로형태 또는 시차 3 이후부터 절단형태를 가짐을 알 수 있다. 이 절단형태와 부분 D의 표본자기상관함수의 시나브로형태를 바탕으로 이 2변량시계열데이터가 VAR(2)모형을 갖는다는 결론을 확인할 수 있다.
- ⑤ 부분 F로부터 최소AIC값을 갖는 VAR(2)모형의 Yule-Walker추정모형이 다음과 같음을 알 수 있다.

$$y_{1,t} = Y1(t) + 0.007$$

$$y_{2,t} = Y2(t) + 0.064$$

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} -0.072 & -0.156 \\ -0.118 & -0.372 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} -0.071 & -0.158 \\ 0.087 & -0.071 \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix}$$

- ⑥ 부분 G으로부터 초기에 선택된 상태벡터는 다음과 같음을 알 수 있다.

$$z_t = (y_{1,t}, y_{2,t}, y_{1,t+1}, y_{2,t+1})'$$

SAS 12:49 Thursday, July 8, 1993 4

STATESPACE Procedure

(A) Nobs = 200

(B)

| Variable | Mean     | Std      |
|----------|----------|----------|
| Y1       | -0.00713 | 0.96935  |
| Y2       | -0.06427 | 1.123824 |

(C) Information Criterion for Autoregressive Models

| Lag=0     | Lag=1     | Lag=2     | Lag=3     | Lag=4     |
|-----------|-----------|-----------|-----------|-----------|
| 26.670015 | 2.267159  | 0.056566  | 5.393204  | 6.107841  |
| Lag=5     | Lag=6     | Lag=7     | Lag=8     | Lag=9     |
| 11.096724 | 16.763063 | 23.838702 | 26.616654 | 31.673740 |
| Lag=10    |           |           |           |           |
| 33.422365 |           |           |           |           |

(D) Schematic Representation of Correlations

| Name/Lag | 0  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|----------|----|---|---|---|---|---|---|---|---|---|----|
| Y1       | +- | . | . | . | . | . | . | . | . | . | .  |
| Y2       | -+ | . | . | . | . | . | . | . | . | . | .  |

+ is > 2\*std error, - is < -2\*std error, . is between

(E) Schematic Representation of Partial Autocorrelations

| Name/Lag | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|----------|---|---|---|---|---|---|---|---|---|----|
| Y1       | . | . | . | . | . | . | . | . | . | .  |
| Y2       | . | . | . | . | . | . | . | . | . | .  |

+ is > 2\*std error, - is < -2\*std error, . is between

(F) Yule-Walker Estimates for the Min AIC

|    | Lag=1    |          | Lag=2    |          |
|----|----------|----------|----------|----------|
|    | Y1       | Y2       | Y1       | Y2       |
| Y1 | -0.07165 | -0.15590 | -0.07116 | -0.15845 |
| Y2 | -0.11808 | -0.37168 | 0.08734  | -0.07119 |

(G) Selected Statespace Form and Preliminary Estimates

(G1) State Vector

| Y1(T;T) | Y2(T;T) | Y1(T+1;T) |
|---------|---------|-----------|
|---------|---------|-----------|

(G2) Estimate of the Transition Matrix

|    | 0      | 0      | 1     |
|----|--------|--------|-------|
| Y1 | -0.095 | -0.322 | 0.272 |
| Y2 | -0.024 | -0.027 | 0.730 |

(G3) Input Matrix for the Innovation

|    | 1      | 0      |
|----|--------|--------|
| Y1 | 0      | 1      |
| Y2 | -0.072 | -0.156 |

(G4) Variance Matrix for the Innovation

|            |            |
|------------|------------|
| 0.90101357 | -0.2596382 |
| -0.2596382 | 1.09963879 |

(H) Selected Statespace Form and Fitted Model

(H1) State Vector

| Y1(T;T) | Y2(T;T) | Y1(T+1;T) |
|---------|---------|-----------|
|---------|---------|-----------|

| (H2)      | Estimate of the Transition Matrix                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |           |          |           |         |        |          |          |          |        |          |         |          |        |          |         |          |        |         |          |          |        |          |        |         |        |          |          |          |        |          |          |          |        |         |          |          |
|-----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|----------|-----------|---------|--------|----------|----------|----------|--------|----------|---------|----------|--------|----------|---------|----------|--------|---------|----------|----------|--------|----------|--------|---------|--------|----------|----------|----------|--------|----------|----------|----------|--------|---------|----------|----------|
|           | $\begin{bmatrix} 0 & 0 & 1 \\ -0.092 & -0.309 & 0.414 \\ 0.060 & -0.016 & 0.603 \end{bmatrix}$                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |           |          |           |         |        |          |          |          |        |          |         |          |        |          |         |          |        |         |          |          |        |          |        |         |        |          |          |          |        |          |          |          |        |         |          |          |
| (H3)      | Input Matrix for the Innovation                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |           |          |           |         |        |          |          |          |        |          |         |          |        |          |         |          |        |         |          |          |        |          |        |         |        |          |          |          |        |          |          |          |        |         |          |          |
|           | $\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -0.104 & -0.164 \end{bmatrix}$                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |           |          |           |         |        |          |          |          |        |          |         |          |        |          |         |          |        |         |          |          |        |          |        |         |        |          |          |          |        |          |          |          |        |         |          |          |
| (H4)      | Variance Matrix for the Innovation                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |           |          |           |         |        |          |          |          |        |          |         |          |        |          |         |          |        |         |          |          |        |          |        |         |        |          |          |          |        |          |          |          |        |         |          |          |
|           | $\begin{bmatrix} 0.88337679 & -0.2691965 \\ -0.2691965 & 1.10788592 \end{bmatrix}$                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |           |          |           |         |        |          |          |          |        |          |         |          |        |          |         |          |        |         |          |          |        |          |        |         |        |          |          |          |        |          |          |          |        |         |          |          |
| (D)       | Parameter Estimates                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |           |          |           |         |        |          |          |          |        |          |         |          |        |          |         |          |        |         |          |          |        |          |        |         |        |          |          |          |        |          |          |          |        |         |          |          |
|           | <table border="1"> <thead> <tr> <th>Parameter</th> <th>Estimate</th> <th>Std. Err.</th> <th>T value</th> </tr> </thead> <tbody> <tr> <td>F(2,1)</td> <td>-0.09241</td> <td>0.091633</td> <td>-1.00843</td> </tr> <tr> <td>F(2,2)</td> <td>-0.30947</td> <td>0.08721</td> <td>-3.54856</td> </tr> <tr> <td>F(2,3)</td> <td>0.414431</td> <td>0.40767</td> <td>1.016584</td> </tr> <tr> <td>F(3,1)</td> <td>0.05966</td> <td>0.061594</td> <td>0.968608</td> </tr> <tr> <td>F(3,2)</td> <td>-0.01627</td> <td>0.0764</td> <td>-0.2129</td> </tr> <tr> <td>F(3,3)</td> <td>0.603492</td> <td>0.149804</td> <td>4.028552</td> </tr> <tr> <td>G(3,1)</td> <td>-0.10397</td> <td>0.073366</td> <td>-1.41717</td> </tr> <tr> <td>G(3,2)</td> <td>-0.1637</td> <td>0.064601</td> <td>-2.53405</td> </tr> </tbody> </table> | Parameter | Estimate | Std. Err. | T value | F(2,1) | -0.09241 | 0.091633 | -1.00843 | F(2,2) | -0.30947 | 0.08721 | -3.54856 | F(2,3) | 0.414431 | 0.40767 | 1.016584 | F(3,1) | 0.05966 | 0.061594 | 0.968608 | F(3,2) | -0.01627 | 0.0764 | -0.2129 | F(3,3) | 0.603492 | 0.149804 | 4.028552 | G(3,1) | -0.10397 | 0.073366 | -1.41717 | G(3,2) | -0.1637 | 0.064601 | -2.53405 |
| Parameter | Estimate                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | Std. Err. | T value  |           |         |        |          |          |          |        |          |         |          |        |          |         |          |        |         |          |          |        |          |        |         |        |          |          |          |        |          |          |          |        |         |          |          |
| F(2,1)    | -0.09241                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 0.091633  | -1.00843 |           |         |        |          |          |          |        |          |         |          |        |          |         |          |        |         |          |          |        |          |        |         |        |          |          |          |        |          |          |          |        |         |          |          |
| F(2,2)    | -0.30947                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 0.08721   | -3.54856 |           |         |        |          |          |          |        |          |         |          |        |          |         |          |        |         |          |          |        |          |        |         |        |          |          |          |        |          |          |          |        |         |          |          |
| F(2,3)    | 0.414431                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 0.40767   | 1.016584 |           |         |        |          |          |          |        |          |         |          |        |          |         |          |        |         |          |          |        |          |        |         |        |          |          |          |        |          |          |          |        |         |          |          |
| F(3,1)    | 0.05966                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | 0.061594  | 0.968608 |           |         |        |          |          |          |        |          |         |          |        |          |         |          |        |         |          |          |        |          |        |         |        |          |          |          |        |          |          |          |        |         |          |          |
| F(3,2)    | -0.01627                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 0.0764    | -0.2129  |           |         |        |          |          |          |        |          |         |          |        |          |         |          |        |         |          |          |        |          |        |         |        |          |          |          |        |          |          |          |        |         |          |          |
| F(3,3)    | 0.603492                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 0.149804  | 4.028552 |           |         |        |          |          |          |        |          |         |          |        |          |         |          |        |         |          |          |        |          |        |         |        |          |          |          |        |          |          |          |        |         |          |          |
| G(3,1)    | -0.10397                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 0.073366  | -1.41717 |           |         |        |          |          |          |        |          |         |          |        |          |         |          |        |         |          |          |        |          |        |         |        |          |          |          |        |          |          |          |        |         |          |          |
| G(3,2)    | -0.1637                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | 0.064601  | -2.53405 |           |         |        |          |          |          |        |          |         |          |        |          |         |          |        |         |          |          |        |          |        |         |        |          |          |          |        |          |          |          |        |         |          |          |

그림 8.4 다변량시계열데이터 VARMA211의 상태공간모형화

⑦ 부분 ㉔로부터 초기에 선택된 상태공간모형의 추정행렬들은 다음과 같음을 알 수 있다.

$$\hat{F} = \begin{bmatrix} 0 & 0 & 1 \\ -0.095 & -0.322 & 0.272 \\ -0.024 & -0.027 & 0.730 \end{bmatrix}, \hat{G} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -0.072 & -0.156 \end{bmatrix}, \hat{\Sigma} = \begin{bmatrix} 0.901 & -0.260 \\ -0.260 & 1.100 \end{bmatrix}$$

⑧ 부분 ㉕으로부터 최종적으로 선택된 상태벡터는 다음과 같음을 알 수 있다.

$$z_t = (y_{1,t}, y_{2,t}, y_{1,t+1|t})'$$

따라서 시점  $t$ 에서  $y_{1,t}$ ,  $y_{2,t}$ 와  $y_{1,t+1|t}$ 의 값만 있으면 모든 미래시점들의 값들을 예측할 수 있다.

⑨ 부분 ㉖로부터 선택된 상태공간모형의 최종추정행렬들이 다음과 같음을 알 수 있다.

$$\hat{F} = \begin{bmatrix} 0 & 0 & 1 \\ -0.092 & -0.309 & 0.414 \\ 0.060 & -0.016 & 0.603 \end{bmatrix}, \hat{G} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -0.104 & -0.164 \end{bmatrix}, \hat{\Sigma} = \begin{bmatrix} 0.883 & -0.269 \\ -0.269 & 1.108 \end{bmatrix}$$

⑩ 부분 ㉗로부터 선택된 상태공간모형의 최종추정행렬들의 원소들이 0과 유의적으로 다른지 여부를 조사할 수 있다. 이 예제에서는 원소들  $F(2, 2)$ ,  $F(3, 3)$ 과  $G(3, 2)$ 의 추정값들이 유의적으로 0과 다르다고 할 수 있다.



그림 8.4에서 최종적으로 구한 상태공간모형은 다음과 같다.

$$y_{1,t} = Y1(t) + 0.007$$

$$y_{2,t} = Y2(t) + 0.064$$

$$z_{t+1} = F z_t + G v_{t+1}$$

$$y_t = [I \ 0] z_t$$

$$\text{단, } z_t = (y_{1,t}, y_{2,t}, y_{1,t+1|t})'$$

$$\hat{F} = \begin{bmatrix} 0 & 0 & 1 \\ -0.092 & -0.309 & 0.414 \\ 0.060 & -0.016 & 0.603 \end{bmatrix}, \quad \hat{G} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -0.104 & -0.164 \end{bmatrix},$$

$$\hat{\Sigma} = \begin{bmatrix} 0.883 & -0.269 \\ -0.269 & 1.108 \end{bmatrix}$$

이 상태공간모형식을 전개하면 다음과 같다.

$$y_{1,t+1} = y_{1,t+1|t} + v_{t+1}$$

$$y_{2,t+1} = -0.092y_{1,t} - 0.309y_{2,t} + 0.414y_{1,t+1|t} + v_{2,t+1}$$

$$y_{1,t+2|t+1} = 0.060y_{1,t} - 0.016y_{2,t} + 0.603y_{1,t+1|t} - 0.104v_{1,t+1} - 0.164v_{2,t+1}$$

첫 번째 식을 세 번째 식에 대입하면 다음 식을 얻는다.

$$y_{1,t+2} = 0.603y_{1,t+1} + 0.060y_{1,t} - 0.016y_{2,t} + v_{1,t+2} - 0.707v_{1,t+1} - 0.164v_{2,t+1}$$

첫 번째 식을 두 번째 식에 대입하면 다음 식을 얻는다.

$$y_{2,t+1} = -0.092y_{1,t} - 0.309y_{2,t} - 0.414(y_{1,t+1} - v_{1,t+1}) + v_{2,t+1}$$

이 식을 정리된 세 번째 식에 대입하면 다음 식을 얻는다.

$$y_{2,t+1} = 0.158y_{1,t} - 0.309y_{2,t} + 0.025y_{1,t-1} - 0.007y_{2,t-1} \\ + v_{2,t+1} - 0.293v_{1,t} - 0.068v_{2,t}$$

이 식을 정리하면 다음과 같은 2변량VARMA(2, 1)모형을 얻는다.

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 0.603 & 0 \\ 0.158 & -0.309 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} 0.060 & -0.016 \\ 0.025 & -0.007 \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} \\ + \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} - \begin{bmatrix} 0.707 & 0.164 \\ 0.293 & 0.068 \end{bmatrix} \begin{bmatrix} v_{1,t-1} \\ v_{2,t-1} \end{bmatrix}$$

## 8.2.2 RESTRICT스태이먼트

그림 8.2와 그림 8.4의 부분 ㉠에서 볼 수 있듯이 변환행렬  $F$ 와 입력행렬  $G$ 의 원소들의 최

종추정값들 중에서 0과 유의적으로 다르지 않은 것이 있다. 이러한 경우에는 이에 해당하는 원소를 0으로 고정하고서 상태공간모형을 다시 추정할 수 있다. 이러한 과정은 PROC STATESPACE의 RESTRICT스태이트먼트를 사용해서 수행한다.

제 8.1.1소절의 첫 번째 예제에서 모수  $F(2, 1)$ 의 추정값이 유의적으로 0과 다르지 않았다. 따라서 이 원소를 0으로 고정시키고 상태공간모형화하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM8-5.PRG에 수록되어 있다).

```

/* */
/* EXAM8-5.PRG */
/* */
/* STATE SPACE MODELING OF VAR12A */
/* WITH RESTRICT OPTION */
/* */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=80 LINESIZE=72;
PROC STATESPACE DATA=TSA4DATA.VAR12A;
 VAR Y;
 RESTRICT F(2,1)=0;
RUN;

```

이 SAS프로그램의 RESTRICT스태이트먼트에서 원소  $F(2, 1)$ 을 0으로 고정시켰다. 이 SAS프로그램을 수행한 결과가 그림 8.5이다. 그림 8.5의 부분 A는 RESTRICT스태이트먼트를 사용하지 않은 그림 8.2의 부분 A에서 부분 F까지와 동일하다. 그림 8.5의 부분 B는 예비추정단계의 결과이다. 부분 B는 원소  $F(2, 1)$ 의 값이 0으로 고정된 것을 제외하고는 그림 8.2의 부분 G와 동일하다. 즉, RESTRICT스태이트먼트를 사용하는 경우의 예비추정단계의 결과는 지정된 원소들이 주어진 값들로 고정되는 것을 제외하고는 RESTRICT스태이트먼트를 사용하지 않는 경우의 예비추정단계의 결과와 동일하다. 그림 8.5의 부분 C는 RESTRICT스태이트먼트를 사용한 경우의 최종추정단계의 결과이다. 예상대로 이 부분은 RESTRICT스태이트먼트를 사용하지 않는 경우의 최종추정단계의 결과와 다르다. 그림 8.5의 부분 D는 RESTRICT스태이트먼트를 사용해서 구한 최종추정값들, 표준오차들, 그리고  $t$ 통계량값들이다. 부분 D에서 모수들  $F(2, 2)$ 와  $G(2, 1)$ 의 추정값들이 0과 유의적으로 차이를 알 수 있다. RESTRICT스태이트먼트를 사용해서 추정한 상태공간모형은 다음과 같다.

$$\begin{aligned}
 y_t &= Y_t + 0.119 \\
 z_{t+1} &= \hat{F}z_t + \hat{G}v_{t+1} \\
 y_t &= [1 \ 0]z_t \\
 \text{단, } z_t &= (y_t, y_{t+1})'
 \end{aligned}$$

SAS 19:41 Thursday, July 8, 1993 5

(A)

STATESPACE Procedure

Nobs = 200

| Variable | Mean     | Std      |
|----------|----------|----------|
| Y        | -0.11884 | 1.791011 |

Information Criterion for Autoregressive Models

| Lag=0      | Lag=1     | Lag=2     | Lag=3     | Lag=4     |
|------------|-----------|-----------|-----------|-----------|
| 233.112018 | 16.334461 | 11.671213 | 13.630674 | 14.081670 |
| Lag=5      | Lag=6     | Lag=7     | Lag=8     | Lag=9     |
| 16.036176  | 17.977597 | 19.060161 | 17.062326 | 19.048223 |
| Lag=10     |           |           |           |           |
| 21.025320  |           |           |           |           |

Schematic Representation of Correlations

| Name/Lag | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|----------|---|---|---|---|---|---|---|---|---|---|----|
| Y        | + | + | + | + | + | + | + | + | + | + | +  |

+ is > 2\*std error, - is < -2\*std error, . is between

Schematic Representation of Partial Autocorrelations

| Name/Lag | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|----------|---|---|---|---|---|---|---|---|---|----|
| Y        | + | + | . | . | . | . | . | . | . | .  |

+ is > 2\*std error, - is < -2\*std error, . is between

Yule-Walker Estimates for the Min AIC

|   | Lag=1   | Lag=2   |
|---|---------|---------|
| Y | 0.66790 | 0.18102 |

(B)

Selected Statespace Form and Preliminary Estimates

State Vector

| Y(T,T) | Y(T+1,T) |
|--------|----------|
|--------|----------|

Estimate of the Transition Matrix

|       |       |
|-------|-------|
| 0     | 1     |
| 0.000 | 0.812 |

Input Matrix for the Innovation

|       |
|-------|
| 1     |
| 0.668 |

Variance Matrix for the Innovation

1.03910115

(C)

Selected Statespace Form and Fitted Model

State Vector

| Y(T,T) | Y(T+1,T) |
|--------|----------|
|--------|----------|

Estimate of the Transition Matrix

|       |       |
|-------|-------|
| 0     | 1     |
| 0.000 | 0.890 |

|                                    |                     |          |           |
|------------------------------------|---------------------|----------|-----------|
| Input Matrix for the Innovation    |                     |          |           |
| 1                                  |                     |          |           |
| 0.660                              |                     |          |           |
| Variance Matrix for the Innovation |                     |          |           |
| 1.03796418                         |                     |          |           |
| (D)                                | Parameter Estimates |          |           |
|                                    | Parameter           | Estimate | Std. Err. |
|                                    | F(2,2)              | 0.889875 | 0.038896  |
|                                    | F(2,1)              | 0.659588 | 0.069165  |
|                                    |                     |          | T value   |
|                                    |                     |          | 22.87857  |
|                                    |                     |          | 9.536475  |

그림 8.5 RESTRICT스태이트먼트를 사용한 상태공간모형화

$$\hat{F} = \begin{bmatrix} 0 & 1 \\ 0.000 & 0.890 \end{bmatrix}, \quad \hat{G} = \begin{bmatrix} 1 \\ 0.660 \end{bmatrix}, \quad \hat{\Sigma} = [1.038]$$

이 상태공간모형식을 전개하면 다음과 같다.

$$y_{t+1} = y_{t+1|t} + v_{t+1}$$

$$y_{t+2|t+1} = 0.890y_{t+1|t} + 0.660v_{t+1}$$

첫 번째 식을 두 번째 식에 대입하면 다음 식을 얻는다.

$$y_{t+2} - v_{t+2} = 0.890(y_{t+1} - v_{t+1}) + 0.660v_{t+1}$$

이 식을 정리하면 다음과 같은 ARMA(1, 1)모형을 얻는다.

$$y_t = 0.890y_{t-1} + v_t - 0.230v_{t-1}$$

따라서 최종상태공간모형으로부터 다변량시계열데이터 VAR12A는 ARMA(1, 1)모형을 갖는다고 결론지을 수 있다.

제 8.1.1 소절의 두 번째 예제에서 모수들  $F(2, 1)$ ,  $F(2, 3)$ ,  $F(3, 1)$ ,  $F(3, 2)$ , 그리고  $G(3, 1)$ 의 추정값들이 0과 유의적으로 다르지 않았다. 따라서 이 원소들을 동시에 0으로 고정시키고 상태공간모형화를 할 수도 있다. 그러나 이 방법은 타당하지가 않다. 그 이유는 모수가 0과 확률적으로 동일하다는 검정을 하는 데 카이제곱검정법이나  $F$ 검정법이 아닌  $t$ 검정법을 사용하였기 때문이다. 즉, 어떤 모수들이 0이라는 동시검정(simultaneous test)을 한 것이 아니라 개개의 모수가 0이라는 단일검정(single test)들을 사용해서 유의성검정을 했기 때문이다. 따라서 PROC STATESPACE의 RESTRICT스태이트먼트를 사용해서 모수들을 제약하기 위해서  $t$ 검정법을 사용하는 경우에는 제약조건을 하나씩만 추가시켜야 한다.

그림 8.4의 부분 Ⅱ에서  $t$ 통계량값의 절대값이 가장 작은 원소는  $F(3, 2)$ 이다. 원소  $F(3, 2)$ 를 0으로 고정시키고 상태공간모형화하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM8-6.PRG에 수록되어 있다).

SAS 19:41 Thursday, July 8, 1993 8

**STATESPACE Procedure**

Selected Statespace Form and Fitted Model

State Vector

Y1(T;T)      Y2(T;T)      Y1(T+1;T)

Estimate of the Transition Matrix

|  |        |        |       |
|--|--------|--------|-------|
|  | 0      | 0      | 1     |
|  | -0.091 | -0.303 | 0.430 |
|  | 0.062  | 0.000  | 0.611 |

Input Matrix for the Innovation

|  |        |        |
|--|--------|--------|
|  | 1      | 0      |
|  | 0      | 1      |
|  | -0.103 | -0.170 |

Variance Matrix for the Innovation

|  |            |            |
|--|------------|------------|
|  | 0.88356732 | -0.2690761 |
|  | -0.2690761 | 1.10799497 |

Parameter Estimates

| Parameter | Estimate | Std. Err. | T value  |
|-----------|----------|-----------|----------|
| F(2,1)    | -0.09149 | 0.092046  | -0.994   |
| F(2,2)    | -0.30334 | 0.084167  | -3.604   |
| F(2,3)    | 0.430195 | 0.416485  | 1.032918 |
| F(3,1)    | 0.06186  | 0.060142  | 1.028555 |
| F(3,3)    | 0.611368 | 0.149053  | 4.10169  |
| G(3,1)    | -0.10312 | 0.073361  | -1.40562 |
| G(3,2)    | -0.16993 | 0.054737  | -3.10453 |

그림 8.6 한 개의 제약조건식을 사용한 상태공간모형화

```

/* */
/* EXAM8-6.PRG */
/* */
/* STATE SPACE MODELING OF VARMA211 */
/* WITH 1 RESTRICTION */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=80 LINESIZE=72;
PROC STATESPACE DATA = TSA4DATA.VARMA211;
 RESTRICT F(3,2)=0;
 VAR Y1 Y2;
RUN;

```

이 SAS프로그램의 RESTRICT스태이트먼트에서 원소  $F(3, 2)$ 를 0으로 고정시켰다. 이 SAS프로그램을 수행한 결과 중에서 최종추정단계의 결과가 그림 8.6이다.

그림 8.6에서 모수들  $F(2, 1)$ ,  $F(2, 3)$ ,  $F(3, 1)$ , 그리고  $G(3, 1)$ 의 추정값들이 0과 유의적

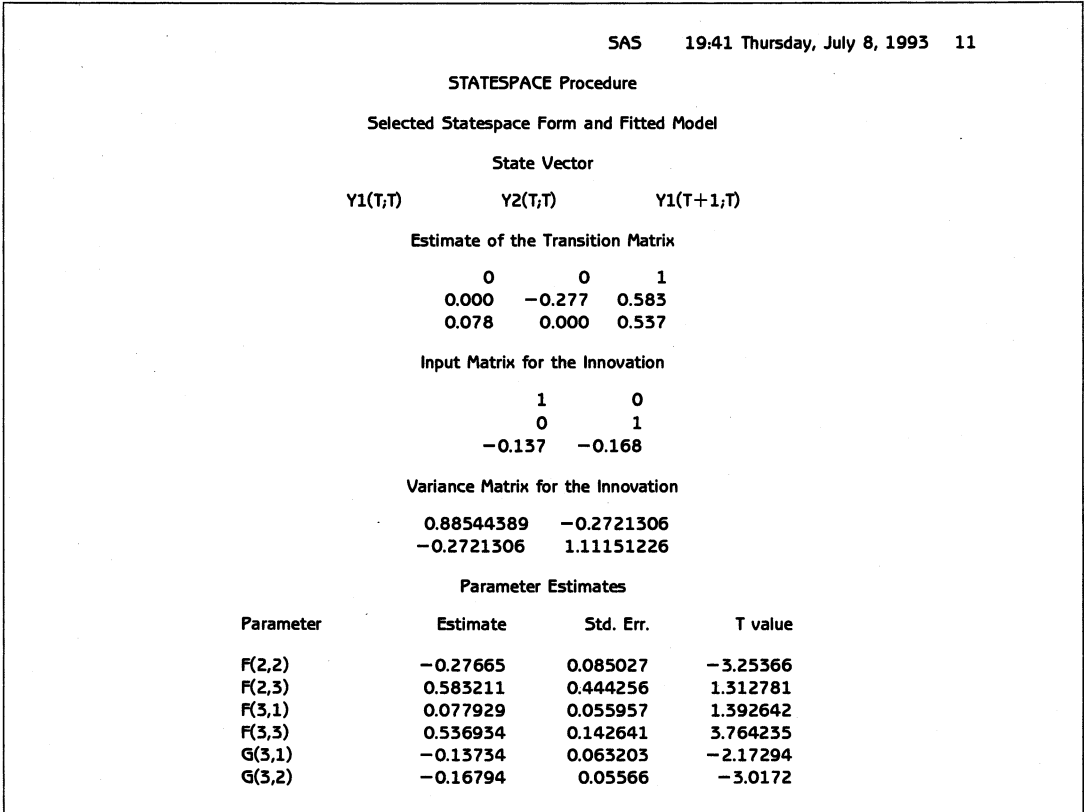


그림 8.7 두 개의 제약조건식들을 사용한 상태공간모형화

으로 다르지 않음을 알 수 있다. 이 중에서  $t$ 통계량값의 절대값이 가장 작은 원소는  $F(2, 1)$ 이다. 원소들  $F(3, 2)$ 와  $F(2, 1)$ 을 0으로 고정시키고 상태공간모형화하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM8-7.PRG에 수록되어 있다).

```

/* */
/* EXAM8-7.PRG */
/* */
/* STATE SPACE MODELING OF VARMA211 */
/* WITH 2 RESTRICTIONS */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=80 LINESIZE=72;
PROC STATESPACE DATA=TSA4DATA.VARMA211;
 RESTRICT F(3,2)=0 F(2,1)=0;
 VAR Y1 Y2;
RUN;

```

이 SAS프로그램의 RESTRICT스태이트먼트에서 원소들  $F(3, 2)$ 와  $F(2, 1)$ 을 0으로 고정시켰다. 이 SAS프로그램을 수행한 결과 중에서 최종추정단계의 결과가 그림 8.7이다.

SAS 19:41 Thursday, July 8, 1993 14

**STATESPACE Procedure**

Selected Statespace Form and Fitted Model

State Vector

Y1(T;T)      Y2(T;T)      Y1(T+1;T)

Estimate of the Transition Matrix

|       |        |       |
|-------|--------|-------|
| 0     | 0      | 1     |
| 0.000 | -0.331 | 0.000 |
| 0.053 | 0.000  | 0.692 |

Input Matrix for the Innovation

|        |        |
|--------|--------|
| 1      | 0      |
| 0      | 1      |
| -0.136 | -0.174 |

Variance Matrix for the Innovation

|            |            |
|------------|------------|
| 0.88293202 | -0.2687164 |
| -0.2687164 | 1.1251655  |

Parameter Estimates

| Parameter | Estimate | Std. Err. | T value  |
|-----------|----------|-----------|----------|
| F(2,2)    | -0.33075 | 0.066494  | -4.97404 |
| F(3,1)    | 0.0526   | 0.065978  | 0.797239 |
| F(3,3)    | 0.692342 | 0.113464  | 6.101894 |
| G(3,1)    | -0.13635 | 0.070623  | -1.93066 |
| G(3,2)    | -0.17437 | 0.052767  | -3.30458 |

그림 8.8 세 개의 제약조건식들을 사용한 상태공간모형화

그림 8.7에서 모수들  $F(2, 3)$ ,  $F(3, 1)$ , 그리고  $G(3, 1)$ 의 추정값들이 0과 유의적으로 다르지 않음을 알 수 있다. 이 중에서  $t$ 통계량값의 절대값이 가장 작은 원소는  $F(2, 3)$ 이다. 원소들  $F(3, 2)$ ,  $F(2, 1)$ 과  $F(2, 3)$ 을 0으로 제약하는 조건하에서 상태공간모형화하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM8-8.PRG에 수록되어 있다).

```

/* */
/* EXAM8-8.PRG */
/* */
/* STATE SPACE MODELING OF VARMA211 */
/* WITH 3 RESTRICTIONS */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=80 LINESIZE=72;
PROC STATESPACE DATA=TSA4DATA.VARMA211;
 RESTRICT F(3,2)=0 F(2,1)=0 F(2,3)=0;
 VAR Y1 Y2;
RUN;

```

| SAS 19:41 Thursday, July 8, 1993 17       |            |           |          |
|-------------------------------------------|------------|-----------|----------|
| STATESPACE Procedure                      |            |           |          |
| Selected Statespace Form and Fitted Model |            |           |          |
| State Vector                              |            |           |          |
| Y1(T;T)                                   | Y2(T;T)    | Y1(T+1;T) |          |
| Estimate of the Transition Matrix         |            |           |          |
| 0                                         | 0          | 1         |          |
| 0.000                                     | -0.001     | 0.000     |          |
| 0.000                                     | 0.000      | 0.741     |          |
| Input Matrix for the Innovation           |            |           |          |
|                                           | 1          | 0         |          |
|                                           | 0          | 1         |          |
|                                           | -0.091     | -0.171    |          |
| Variance Matrix for the Innovation        |            |           |          |
| 0.88276761                                | -0.2612745 |           |          |
| -0.2612745                                | 1.12516851 |           |          |
| Parameter Estimates                       |            |           |          |
| Parameter                                 | Estimate   | Std. Err. | T value  |
| F(2,2)                                    | -0.00064   | 0.070614  | -0.00909 |
| F(3,3)                                    | 0.741104   | 0.117612  | 6.30129  |
| G(3,1)                                    | -0.09094   | 0.042329  | -2.14835 |
| G(3,2)                                    | -0.17106   | 0.061662  | -2.77415 |

그림 8.9 네 개의 제약조건식들을 사용한 상태공간모형화

이 SAS프로그램의 RESTRICT스태이트먼트에서 원소들  $F(3, 2)$ ,  $F(2, 1)$ 와  $F(2, 3)$ 을 0으로 고정시켰다. 이 SAS프로그램을 수행한 결과 중에서 최종추정단계의 결과가 그림 8.8이다.

그림 8.8에서 모수들  $F(3, 1)$ 과  $G(3, 1)$ 의 추정값들이 0과 유의적으로 다르지 않음을 알 수 있다. 이 중에서  $t$ 통계량값의 절대값이 가장 작은 원소는  $F(3, 1)$ 이다. 원소들  $F(3, 2)$ ,  $F(2, 1)$ ,  $F(2, 3)$ 과  $F(3, 1)$ 을 0으로 제약하는 조건하에서 상태공간모형화하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM 8-9.PRG에 수록되어 있다).

```

/* */
/* EXAM8-9.PRG */
/* */
/* STATE SPACE MODELING OF VARMA211 */
/* WITH 4 RESTRICTIONS */
/* */
LIBNAME TSA4DATA 'A:TSA4DATA';
OPTIONS PAGESIZE=80 LINESIZE=72;
PROC STATESPACE DATA=TSA4DATA.VARMA211;

```



| SAS 19:41 Thursday, July 8, 1993 20       |            |           |          |
|-------------------------------------------|------------|-----------|----------|
| STATESPACE Procedure                      |            |           |          |
| Selected Statespace Form and Fitted Model |            |           |          |
| State Vector                              |            |           |          |
| Y1(T;T)                                   | Y2(T;T)    | Y1(T+1;T) |          |
| Estimate of the Transition Matrix         |            |           |          |
| 0                                         | 0          | 1         |          |
| 0.000                                     | 0.000      | 0.000     |          |
| 0.000                                     | 0.000      | 0.758     |          |
| Input Matrix for the Innovation           |            |           |          |
|                                           | 1          | 0         |          |
|                                           | 0          | 1         |          |
|                                           | -0.087     | -0.231    |          |
| Variance Matrix for the Innovation        |            |           |          |
| 0.88934551                                | -0.2898091 |           |          |
| -0.2898091                                | 1.26298136 |           |          |
| Parameter Estimates                       |            |           |          |
| Parameter                                 | Estimate   | Std. Err. | T value  |
| F(3,3)                                    | 0.757801   | 0.079865  | 9.488549 |
| G(3,1)                                    | -0.08666   | 0.040864  | -2.12063 |
| G(3,2)                                    | -0.23093   | 0.057968  | -3.98371 |

그림 8.10 다섯 개의 제약조건식들을 사용한 상태공간모형화

```

RESTRICT F(3,2)=0 F(2,1)=0 F(2,3)=0
 F(3,1)=0;
VAR Y1 Y2;
RUN;

```

이 SAS프로그램의 RESTRICT스태이트먼트에서 원소들  $F(3, 2)$ ,  $F(2, 1)$ ,  $F(2, 3)$ 과  $F(3, 1)$ 을 0으로 고정시켰다. 이 SAS프로그램을 수행한 결과 중에서 최종추정단계의 결과가 그림 8.9이다.

그림 8.9에서 모수  $F(2, 2)$ 의 추정값이 0과 유의적으로 다르지 않음을 알 수 있다. 원소들  $F(3, 2)$ ,  $F(2, 1)$ ,  $F(2, 3)$ ,  $F(3, 1)$ 과  $F(2, 2)$ 를 0으로 고정시키고 상태공간모형화하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM8-10.PRG에 수록되어 있다).

```

/* */
/* EXAM8-10.PRG */
/* */
/* STATE SPACE MODELING OF VARMA211 */
/* WITH 5 RESTRICTIONS */
LIBNAME TSA4DATA 'A:\TSA4DATA';

```

```

OPTIONS PAGESIZE=80 LINESIZE=72;
PROC STATESPACE DATA=TSA4DATA.VARMA211;
 RESTRICT F(3,2)=0 F(2,1)=0 F(2,3)=0
 F(3,1)=0 F(2,2)=0;
VAR Y1 Y2;
RUN;

```

이 SAS프로그램의 RESTRICT스태이트먼트에서 원소들  $F(3, 2)$ ,  $F(2, 1)$ ,  $F(2, 3)$ ,  $F(3, 1)$ , 그리고  $F(2, 2)$ 를 0으로 고정시켰다. 이 SAS프로그램을 수행한 결과 중에서 최종 추정단계의 결과가 그림 8.10이다.

그림 8.10에서 모수들  $F(3, 3)$ ,  $G(3, 1)$ 과  $G(2, 1)$ 의 추정값들이 0과 유의적으로 다를 수 있다. 따라서 최종적으로 추정된 상태공간모형은 다음과 같다.

$$y_{1,t} = Y1(t) + 0.007$$

$$y_{2,t} = Y2(t) + 0.064$$

$$z_{t+1} = \hat{F}z_t + \hat{G}v_{t+1}$$

$$y_{2,t} = [I_2 \ 0]z_t$$

$$\text{단, } z_t = (y_{1,t}, y_{2,t}, y_{1,t+1|t})'$$

$$\hat{F} = \begin{bmatrix} 0 & 0 & 1 \\ 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.758 \end{bmatrix}, \quad \hat{G} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -0.087 & -0.231 \end{bmatrix},$$

$$\hat{\Sigma} = \begin{bmatrix} 0.889 & -0.290 \\ -0.290 & 1.263 \end{bmatrix}$$

이 상태공간모형식을 전개하면 다음과 같다.

$$y_{1,t+1} = y_{1,t+1|t} + v_{1,t+1}$$

$$y_{2,t+1} = v_{2,t+1}$$

$$y_{1,t+2|t+1} = 0.758y_{1,t+1|t} - 0.087v_{1,t+1} - 0.231v_{2,t+1}$$

첫 번째 식을 세 번째 식에 대입하면 다음 식을 얻는다.

$$y_{1,t+2} = 0.758y_{1,t+1} + v_{1,t+2} - 0.845v_{1,t+1} - 0.231v_{2,t+1}$$

따라서 그림 8.10의 상태공간모형으로부터 이 2변량시계열데이터는 다음과 같은 2변량 VARMA(1, 1)모형을 갖는다고 결론지을 수 있다.

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 0.758 & 0.000 \\ 0.000 & 0.000 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} - \begin{bmatrix} 0.845 & 0.231 \\ 0.000 & 0.000 \end{bmatrix} \begin{bmatrix} v_{1,t-1} \\ v_{2,t-1} \end{bmatrix}$$

RESTRICT스태이트먼트를 사용할 때 제약조건을 하나씩 더해가는 방식은 회귀분석에서 타당한 모형을 선택하기 위해서 사용하는 후향제거방법(backward elimination met-

hod)과 비슷하다. 따라서 전향추가방법(forward selection method)이나 단계별방법(step-wise method)을 적용하는 것도 생각해 볼 수 있다. 회귀분석에서 이러한 방법들이 어떠한 단점들을 갖고 있는가를 독자들은 잘 알고 있으리라 생각한다. 상태공간모형화에서 PROC STATESPACE의 RESTRICT스태이트먼트를 사용해서 제약조건을 가할 때도 같은 문제점들이 발생한다. 특히  $t$ 검정법을 바탕으로 하는 후향제거법을 사용하면 제약조건을 추가할 때마다 제1종오차가 증가되어 전체적인 제1종오차를 조절할 수 없는 것이 가장 큰 문제이다. 더구나 추가하는 제약조건들의 순서에 의해서도 최종적으로 선택한 상태공간모형이 달라질 것이다. 객관적으로 제약조건을 정하기 위해서 벌칙함수식별방법들인 AIC, BIC, HQC, 그리고 퍼트맨토우통계량 등을 사용할 수 있다. 이 중에서 퍼트맨토우통계량은 보수적(conservative)인 성질을 가지므로 그리 유용하지가 않을 것이다. 그림 8.4, 그림 8.6부터 그림 8.10까지의 상태공간모형들 중에서 타당한 모형을 고르기 위해서 AIC, BIC, 그리고 HQC를 적용하기로 하자. 이러한 목적으로 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM8-11.PRG에 수록되어 있다).

```

/* */
/* EXAM8-11.PRG */
/* */
/* CHOOSING RESTRICT CONDITIONS */
/* USING THE PENALTY FUNTION METHODS */
/* */
OPTIONS PAGESIZE=60 LINESIZE=72;
PROC IML;
RESET NOLOG NOCENTER NONAME;
T = 200;
K = 8;
/* FROM EXAM8-4.OUT */
COV84 = { 0.88337679 -0.2691965,
 -0.2691965 1.10788592 };
NORES = 0;
AIC = LOG(DET(COV84)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COV84)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COV84)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT , "NO OF RESTRICTION = " NORES,
 " AIC = " AIC,
 " BIC = " BIC,
 " HQC = " HQC;
/* FROM EXAM8-6.OUT */
COV86 = { 0.88356732 -0.2690761,

```

```

-0.2690761 1.10799497 };
NORES = 1;
AIC = LOG(DET(COV86)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COV86)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COV86)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT, "NO OF RESTRICTION = " NORES,
 " AIC = " AIC,
 " BIC = " BIC,
 " HQC = " HQC;
/* FROM EXAM8-7.OUT */
COV87 = { 0.88544389 -0.2721306,
 -0.2721306 1.11151226 };
NORES = 2;
AIC = LOG(DET(COV87)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COV87)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COV87)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT, "NO OF RESTRICTION = " NORES,
 " AIC = " AIC,
 " BIC = " BIC,
 " HQC = " HQC;
/* FROM EXAM8-8.OUT */
COV88 = { 0.88293202 -0.2687164,
 -0.2687164 1.1251655 };
NORES = 3;
AIC = LOG(DET(COV88)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COV88)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COV88)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT, "NO OF RESTRICTION = " NORES,
 " AIC = " AIC,
 " BIC = " BIC,
 " HQC = " HQC;
/* FROM EXAM8-9.OUT */
COV89 = { 0.88276761 -0.2612745,
 -0.2612745 1.12516851 };
NORES = 4;
AIC = LOG(DET(COV89)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COV89)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COV89)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT, "NO OF RESTRICTION = " NORES,

```

```

" AIC = " AIC,
" BIC = " BIC,
" HQC = " HQC;
/* FROM EXAM8-10.OUT */
COV810 = { 0.88934551 -0.2898091,
 -0.2898091 1.26298136 };
NORES = 5;
AIC = LOG(DET(COV810)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COV810)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COV810)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT, "NO OF RESTRICTION = " NORES,
" AIC = " AIC,
" BIC = " BIC,
" HQC = " HQC,;
QUIT;

```

이 SAS프로그램을 수행한 결과가 그림 8.11이다. 그림 8.11에서 볼 수 있듯이 AIC, BIC, 그리고 HQC가 제한조건이 4개인 상태공간모형, 즉 원소들  $F(3, 2)$ ,  $F(2, 1)$ ,  $F(2, 3)$ , 그리고  $F(3, 1)$ 을 0으로 고정시킨 상태공간모형에서 최소값을 갖는다. 이 상태공간모형은 다음과 같다.

$$y_{1,t} = Y1(t) + 0.007$$

$$y_{2,t} = Y2(t) + 0.064$$

$$z_{t+1} = \hat{F}z_t + \hat{G}v_{t+1}$$

$$y_t = [I_2 \ 0]z_t$$

$$\text{단, } z_t = (y_{1,t}, y_{2,t}, y_{1,t+1|t})'$$

$$\hat{F} = \begin{bmatrix} 0 & 0 & 1 \\ 0.000 & -0.001 & 0.000 \\ 0.000 & 0.000 & 0.741 \end{bmatrix}, \quad \hat{G} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -0.091 & -0.171 \end{bmatrix},$$

$$\hat{\Sigma} = \begin{bmatrix} 0.883 & -0.261 \\ -0.261 & 1.125 \end{bmatrix}$$

이 상태공간모형식을 전개하면 다음과 같다.

$$y_{1,t+1} = y_{1,t+1|t} + v_{1,t+1}$$

$$y_{2,t+1} = -0.001y_{2,t} + v_{2,t+1}$$

$$y_{1,t+2|t+1} = 0.741y_{1,t+1|t} - 0.091v_{1,t+1} - 0.171v_{2,t+1}$$

첫 번째 식을 세 번째 식에 대입하면 다음 식을 얻는다.

$$y_{1,t+2} = 0.741y_{1,t+1} + v_{1,t+2} - 0.832v_{1,t+1} - 0.171v_{2,t+1}$$

|                     |           |                                |   |
|---------------------|-----------|--------------------------------|---|
| SAS                 |           | 22:17 Sunday, August 1, 1993 1 |   |
| NO OF RESTRICTION = |           |                                | 0 |
| AIC =               | -0.01848  |                                |   |
| BIC =               | 0.1134528 |                                |   |
| HQC =               | 0.0349113 |                                |   |
| NO OF RESTRICTION = |           |                                | 1 |
| AIC =               | -0.028069 |                                |   |
| BIC =               | 0.0873719 |                                |   |
| HQC =               | 0.0186481 |                                |   |
| NO OF RESTRICTION = |           |                                | 2 |
| AIC =               | -0.034171 |                                |   |
| BIC =               | 0.064778  |                                |   |
| HQC =               | 0.0058719 |                                |   |
| NO OF RESTRICTION = |           |                                | 3 |
| AIC =               | -0.032039 |                                |   |
| BIC =               | 0.0504191 |                                |   |
| HQC =               | 0.0013306 |                                |   |
| NO OF RESTRICTION = |           |                                | 4 |
| AIC =               | -0.037964 |                                |   |
| BIC =               | 0.0280026 |                                |   |
| HQC =               | -0.011268 |                                |   |
| NO OF RESTRICTION = |           |                                | 5 |
| AIC =               | 0.0684873 |                                |   |
| BIC =               | 0.117962  |                                |   |
| HQC =               | 0.0885089 |                                |   |
| Exiting IML.        |           |                                |   |

그림 8.11 벌칙함수식별방법들에 의한 제약조건의 선택

따라서  $\{y_t\}$ 는 다음과 같은 2변량VARMA(1, 1)모형을 갖는다고 결론지을 수 있다.

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 0.741 & 0.000 \\ 0.000 & -0.000 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} - \begin{bmatrix} 0.832 & 0.171 \\ 0.000 & 0.000 \end{bmatrix} \begin{bmatrix} v_{1,t-1} \\ v_{2,t-1} \end{bmatrix}$$

지금까지 벌칙함수식별방법들에 의해서 선택한 상태공간모형이 최적의 모형이라고 결론지을 수는 없다. 이는  $t$ 점정법에 의해 선택된 상태공간모형들 중에서 벌칙함수를 최소화하는 상태공간모형을 택했기 때문이다. 따라서 전향추가방법으로 선택한 상태공간모형들에 벌칙함수식별방법들을 적용하거나, 모든 가능한(all possible) 상태공간모형들에 벌칙함수식별방법들을 적용하거나, 또는 출력벡터내의 원소들의 순서를 달리하면, 후향제거방법으로 선택한 상태공간모형들에 벌칙함수식별방법들을 적용해서 선택한 최적의 상태공간모형과 다른 상태공간모형이 최적의 상태공간모형으로 선택될 수도 있다. 상태공간모형화에서 PROC STATESPACE의 RESTRICT스태이트먼트를 사용해서 제약조건을 가할 때,

이 제약조건들을 정하기 위한 좀더 타당한 방법은 역시 동시검정을 하는 것이라 판단된다. 제약조건을 갖는 VAR모형의 다변량최우추정량들도 점근다변량정규분포를 갖는다. 따라서 본 저자의 생각으로는 상태공간모형의 어떤 모수들이 0인지를 동시검정하는 카이제곱통계량이나  $F$ 통계량을 쉽게 구할 수 있으리라 생각된다. 관심있는 독자들은 이 문제를 풀어 보고 좋은 결과를 얻으면 본 저자에게도 알려 주기를 바란다.

### 8.2.3 상태공간모형을 사용한 예측

시계열분석의 주된 목적 중의 하나는 예측(forecast, 또는 prediction)을 하는 것이다. PROC STATESPACE를 사용하면 단변량시계열데이터뿐만 아니라 다변량시계열데이터도 예측할 수 있다. 이러한 목적을 위해서는 PROC STATESPACE스태이트먼트의 옵션 LEAD=을 사용한다. 그러나 옵션 LEAD=을 사용해도 OUTPUT윈도우에 예측의 결과가 나타나는 것은 아니다. 예측의 결과를 출력시키기 위해서는 PROC STATESPACE스태이트먼트의 옵션 PRINT를 사용하거나 PROC STATESPACE스태이트먼트의 옵션 OUT=을 옵션 LEAD=과 같이 사용해야 한다. 옵션 OUT=에 지정한 SAS데이터세트에 예측결과가 저장된다. 이 SAS데이터세트를 PROC PRINT를 사용해서 OUTPUT윈도우에 출력시킨다.

상태공간모형을 사용해서 제 8.1.1소절의 첫 번째 예제의 단변량시계열데이터 VAR12A를 예측하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM8-12.PRG에 수록되어 있다).

```

/* */
/* EXAM8-12.PRG */
/* */
/* STATE SPACE MODEL FORECASTING OF VAR12A */
/* */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=80 LINESIZE=72;
TITLE 'FORECAST BY STATE SPACE MODEL';
PROC STATESPACE DATA=TSA4DATA.VAR12A
 LEAD=15 OUT=TSA4DATA.FORE812;
VAR Y;
RESTRICT F(2,1)=0;
RUN;
PROC PRINT DATA=TSA4DATA.FORE812(FIRSTOBS=191);
RUN;

```

이 SAS프로그램에서 PROC STATESPACE스태이트먼트의 옵션 LEAD=에 15가 지

| FORECAST BY STATE SPACE MODEL |          |          |          |         | 11                            |
|-------------------------------|----------|----------|----------|---------|-------------------------------|
|                               |          |          |          |         | 21:41 Saturday, July 10, 1993 |
| OBS                           | Y        | FOR1     | RES1     | STD1    |                               |
| 191                           | -0.50244 | -0.43756 | -0.06489 | 1.01881 |                               |
| 192                           | -0.98774 | -0.44526 | -0.54249 | 1.01881 |                               |
| 193                           | -1.44808 | -0.76713 | -0.68095 | 1.01881 |                               |
| 194                           | -1.04172 | -1.14489 | 0.10317  | 1.01881 |                               |
| 195                           | 0.23544  | -0.96384 | 1.19928  | 1.01881 |                               |
| 196                           | -0.08669 | -0.07975 | -0.00694 | 1.01881 |                               |
| 197                           | -1.61653 | -0.08864 | -1.52790 | 1.01881 |                               |
| 198                           | -0.78480 | -1.09974 | 0.31494  | 1.01881 |                               |
| 199                           | -1.20985 | -0.78399 | -0.42586 | 1.01881 |                               |
| 200                           | -0.16745 | -0.99163 | 0.82418  | 1.01881 |                               |
| 201                           | .        | -0.35190 | .        | 1.01881 |                               |
| 202                           | .        | -0.32623 | .        | 1.22047 |                               |
| 203                           | .        | -0.30339 | .        | 1.35909 |                               |
| 204                           | .        | -0.28307 | .        | 1.45955 |                               |
| 205                           | .        | -0.26498 | .        | 1.53445 |                               |
| 206                           | .        | -0.24889 | .        | 1.59126 |                               |
| 207                           | .        | -0.23457 | .        | 1.63484 |                               |
| 208                           | .        | -0.22182 | .        | 1.66855 |                               |
| 209                           | .        | -0.21048 | .        | 1.69477 |                               |
| 210                           | .        | -0.20039 | .        | 1.71524 |                               |
| 211                           | .        | -0.19141 | .        | 1.73128 |                               |
| 212                           | .        | -0.18342 | .        | 1.74388 |                               |
| 213                           | .        | -0.17631 | .        | 1.75380 |                               |
| 214                           | .        | -0.16998 | .        | 1.76161 |                               |
| 215                           | .        | -0.16435 | .        | 1.76777 |                               |

그림 8.12 PROC STATESPACE를 사용한 단변량시계열의 예측

정되었으므로 주어진 200개의 관찰점들에서의 예측, 잔차, 그리고 표준편차들을 계산하고, 마지막 관찰점 이후의 15개 선행시점(lead)에서 예측과 표준편차들을 계산해서 이 값들을 옵션 OUT=에 지정한 영구SAS데이터세트 FORE812에 저장한다. 이 SAS스테이트먼트 FORE812를 OUTPUT윈도우에 출력시키기 위해서 PROC PRINT를 수행한다. 이 PROC PRINT스테이트먼트의 옵션 FIRSTOBS=에 191이 지정되었으므로 제191번째 관찰점부터의 결과물이 출력될 것이다. 이 SAS프로그램을 수행한 결과가 그림 8.12이다. 그림 8.12에서 변수 Y는 분석대상의 단변량시계열데이터이고, 변수 FOR1은 상태공간모형에 의한 예측(또는 적합), 변수 RES1은 잔차, 변수 STD1은 표준편차이다.

그림 8.12의 PROC STATESPACE에 의한 예측과 PROC ARIMA에 의한 예측을 비교하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM8-13.PRG에 수록되어 있다).

```

/* */
/* EXAM8-13.PRG */
/* */
/* ARIMA MODEL FORECASTING OF VAR12A */
/* */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=80 LINESIZE=76;
TITLE 'FORECAST BY ARIMA MODEL';
PROC ARIMA DATA=TSA4DATA.VAR12A;

```



| FORECAST BY ARIMA MODEL |          |          |                             |
|-------------------------|----------|----------|-----------------------------|
|                         |          |          | 35                          |
|                         |          |          | 19:15 Sunday, July 11, 1993 |
| OBS                     | Y        | FORECAST | STD                         |
| 191                     | -0.50244 | -0.44894 | 1.02375                     |
| 192                     | -0.98774 | -0.45650 | 1.02375                     |
| 193                     | -1.44808 | -0.77787 | 1.02375                     |
| 194                     | -1.04172 | -1.15562 | 1.02375                     |
| 195                     | 0.23544  | -0.97567 | 1.02375                     |
| 196                     | -0.08669 | -0.09281 | 1.02375                     |
| 197                     | -1.61653 | -0.10005 | 1.02375                     |
| 198                     | -0.78480 | -1.10924 | 1.02375                     |
| 199                     | -1.20985 | -0.79571 | 1.02375                     |
| 200                     | -0.16745 | -1.00287 | 1.02375                     |
| 201                     | .        | -0.36445 | 1.02375                     |
| 202                     | .        | -0.34602 | 1.22581                     |
| 203                     | .        | -0.32960 | 1.36499                     |
| 204                     | .        | -0.31498 | 1.46603                     |
| 205                     | .        | -0.30196 | 1.54149                     |
| 206                     | .        | -0.29036 | 1.59881                     |
| 207                     | .        | -0.28002 | 1.64286                     |
| 208                     | .        | -0.27082 | 1.67698                     |
| 209                     | .        | -0.26263 | 1.70357                     |
| 210                     | .        | -0.25533 | 1.72436                     |
| 211                     | .        | -0.24882 | 1.74068                     |
| 212                     | .        | -0.24303 | 1.75352                     |
| 213                     | .        | -0.23788 | 1.76364                     |
| 214                     | .        | -0.23328 | 1.77162                     |
| 215                     | .        | -0.22919 | 1.77793                     |

그림 8.13 PROC ARIMA에 의한 단변량시계열데이터의 예측

```

I VAR=Y NOPRINT;
E P=1 Q=1 NOPRINT;
F LEAD=15 OUT=TSA4DATA.FORE813 NOPRINT;
RUN;
PROC PRINT DATA=TSA4DATA.FORE813(FIRSTOBS=191);
VAR Y FORECAST STD;
RUN;

```

제 8.1.2 소절에서 알 수 있듯이 이 단변량시계열데이터에 알맞는 ARMA모형의 차수는  $(p, q) = (1, 1)$  이다. 즉, 이 단변량시계열데이터를 ARMA(1, 1)모형으로 식별한다. 따라서 PROC ARIMA의 ESTIMATE스태이트먼트의 옵션 P=에 1을, 옵션 Q=에 1을 지정하였다. PROC ARIMA의 FORECAST스태이트먼트의 옵션 NOPRINT를 사용했으므로 예측의 결과가 OUTPUT윈도우에 나타나지 않는다. 이 예측의 결과를 출력시키기 위해서 PROC PRINT를 수행한다. 이 SAS프로그램을 수행한 결과가 그림 8.13이다.

그림 8.13을 그림 8.12와 비교해 보면 PROC ARIMA에 의한 예측이 PROC STATESPACE에 의한 예측보다 전반적으로 약간 작다. 반면에 PROC ARIMA에 의한 표준편차가 PROC STATESPACE에 의한 표준편차보다 약간 크다. PROC STATESPACE에 의한 잔차와 PROC ARIMA에 의한 잔차를 비교하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM8-14.PRG에 수록되어 있다).

```

/* */
/* */ EXAM8-14.PRG */
/* */ */
/* */ COMPARISON OF FORECASTING RESIDUALS */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=80 LINESIZE=76;
TITLE 'COMPARISON OF RESIDUALS';
DATA EXAM814;
 MERGE TSA4DATA.FORE812 TSA4DATA.FORE813;
 T=_N_;
 RESSTATE = RES1;
 RESARIMA = Y - FORECAST;
 LABEL RESSTATE='RESIDUAL BY PROC STATESPACE';
 LABEL RESARIMA='RESIDUAL BY PROC ARIMA';
 LABEL T = 'TIME';
 IF T > 200 THEN DELETE;
 KEEP T RESSTATE RESARIMA;
 RUN;
PROC PLOT DATA=EXAM814(FIRSTOBS=151);
 PLOT RESSTATE*T='S' RESARIMA*T='A'
 / VPOS=25 VREF=0 OVERLAY
 HAXIS = 150 TO 200 BY 5;
 RUN;

```

이 SAS프로그램을 수행한 결과가 그림 8.14이다. 그림 8.14에서 볼 수 있듯이 PROC ARIMA에 의한 잔차와 PROC STATESPACE에 의한 잔차가 거의 일치한다. 즉, 이 예제에서는 두 프로시저가 비슷한 결과를 보인다.

상태공간모형을 사용해서 제 8.1.1 소절의 두 번째 예제의 다변량시계열데이터 VARMA 211을 예측하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4 TSA4의 파일 EXAM8-15.PRG에 수록되어 있다).

```

/* */ */
/* */ EXAM8-15.PRG */
/* */ */
/* */ STATE SPACE FORECASTING OF VARMA211 */
/* */ */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=80 LINESIZE=76;
TITLE 'FORECAST BY STATE SPACE MODEL';

```

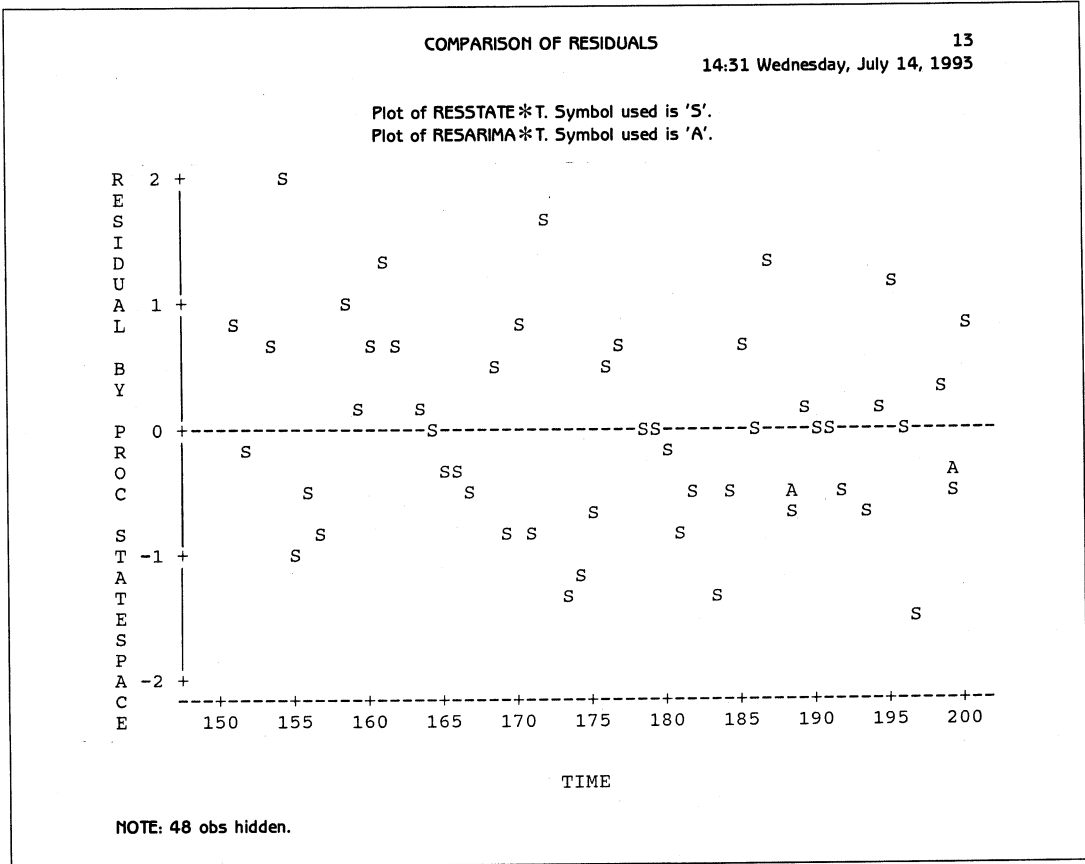


그림 8.14 PROC STATESPACE와 PROC ARIMA에 의한 잔차들의 비교

```

PROC STATESPACE DATA=TSA4DATA.VARMA211
 LEAD=15 OUT=TSA4DATA.FORE815;
 RESTRICT F(3,2)=0 F(2,1)=0 F(2,3)=0 F(3,1)=0;
 VAR Y1 Y2;
 RUN;
PROC PRINT DATA=TSA4DATA.FORE815(FIRSTOBS=186);
 RUN;

```

이 SAS프로그램의 PROC STATESPACE 스테이트먼트의 옵션 LEAD=에 15가 지정되었으므로 주어진 200개의 관찰점들에서의 예측과 표준편차를 계산하고, 마지막 관찰점 이후의 15개 선행시점들에서 예측, 잔차 그리고 표준편차들을 계산해서 이 값들을 옵션 OUT=에 지정한 영구SAS데이터세트 FORE815에 저장한다. 이 SAS스테이트먼트 FORE815를 OUTPUT 윈도우에 출력시키기 위해서 PROC PRINT를 수행한다. PROC PRINT 스테이트먼트의 옵션 FIRSTOBS=에 186이 지정되었으므로 제186번째 관찰점부터의 결과물이 출력될 것이다. 이 SAS프로그램을 수행한 결과가 그림 8.15이다. 그림 8.15에서 변수 Y1은 분석대상의 다변량시계열데이터의 첫 번째 원소이고, 변수 FOR1, 변수 RES1, 그리고 변수 STD1은 각각 변수 Y1의 상태공간모형에 의한 예측(또는 적합), 잔차 그리

| FORECAST BY STATE SPACE MODEL |          |          |          |         |          |           |          |         |
|-------------------------------|----------|----------|----------|---------|----------|-----------|----------|---------|
| 21:07 Saturday, July 10, 1993 |          |          |          |         |          |           |          |         |
| OBS                           | Y1       | FOR1     | RES1     | STD1    | Y2       | FOR2      | RES2     | STD2    |
| 186                           | -0.35968 | 0.08424  | -0.44392 | 0.93956 | 1.38208  | -0.063990 | 1.44607  | 1.06074 |
| 187                           | 0.07490  | -0.14641 | 0.22132  | 0.93956 | -1.30414 | -0.065202 | -1.23894 | 1.06074 |
| 188                           | -0.45997 | 0.08145  | -0.54142 | 0.93956 | 0.82814  | -0.063477 | 0.89162  | 1.06074 |
| 189                           | -0.02564 | -0.04477 | 0.01912  | 0.93956 | 0.13664  | -0.064846 | 0.20148  | 1.06074 |
| 190                           | 0.11598  | -0.07123 | 0.18721  | 0.93956 | -0.38686 | -0.064402 | -0.32246 | 1.06074 |
| 191                           | 1.80152  | -0.01650 | 1.81802  | 0.93956 | 0.25818  | -0.064066 | 0.32224  | 1.06074 |
| 192                           | -1.47352 | -0.23452 | -1.23900 | 0.93956 | -1.98487 | -0.064480 | -1.92039 | 1.06074 |
| 193                           | 0.66057  | 0.26552  | 0.39504  | 0.93956 | 0.90553  | -0.063040 | 0.96857  | 1.06074 |
| 194                           | -0.98978 | -0.00667 | -0.98311 | 0.93956 | 0.60919  | -0.064896 | 0.67409  | 1.06074 |
| 195                           | -0.58119 | -0.03270 | -0.54849 | 0.93956 | -1.38068 | -0.064705 | -1.31597 | 1.06074 |
| 196                           | 2.05518  | 0.24891  | 1.80627  | 0.93956 | 0.42269  | -0.063428 | 0.48612  | 1.06074 |
| 197                           | 0.69242  | -0.06479 | 0.75721  | 0.93956 | 0.09506  | -0.064586 | 0.15964  | 1.06074 |
| 198                           | -0.56261 | -0.14603 | -0.41657 | 0.93956 | 0.15057  | -0.064375 | 0.21494  | 1.06074 |
| 199                           | 0.09747  | -0.10896 | 0.20643  | 0.93956 | 1.61319  | -0.064411 | 1.67760  | 1.06074 |
| 200                           | 0.49213  | -0.38834 | 0.88047  | 0.93956 | -0.16908 | -0.065350 | -0.10373 | 1.06074 |
| 201                           | .        | -0.35197 | .        | 0.93956 | .        | -0.064206 | .        | 1.06074 |
| 202                           | .        | -0.26269 | .        | 0.95648 | .        | -0.064273 | .        | 1.06074 |
| 203                           | .        | -0.19653 | .        | 0.96566 | .        | -0.064273 | .        | 1.06074 |
| 204                           | .        | -0.14749 | .        | 0.97066 | .        | -0.064273 | .        | 1.06074 |
| 205                           | .        | -0.11116 | .        | 0.97339 | .        | -0.064273 | .        | 1.06074 |
| 206                           | .        | -0.08422 | .        | 0.97489 | .        | -0.064273 | .        | 1.06074 |
| 207                           | .        | -0.06427 | .        | 0.97571 | .        | -0.064273 | .        | 1.06074 |
| 208                           | .        | -0.04947 | .        | 0.97616 | .        | -0.064273 | .        | 1.06074 |
| 209                           | .        | -0.03851 | .        | 0.97641 | .        | -0.064273 | .        | 1.06074 |
| 210                           | .        | -0.03039 | .        | 0.97655 | .        | -0.064273 | .        | 1.06074 |
| 211                           | .        | -0.02437 | .        | 0.97662 | .        | -0.064273 | .        | 1.06074 |
| 212                           | .        | -0.01990 | .        | 0.97666 | .        | -0.064273 | .        | 1.06074 |
| 213                           | .        | -0.01660 | .        | 0.97669 | .        | -0.064273 | .        | 1.06074 |
| 214                           | .        | -0.01415 | .        | 0.97670 | .        | -0.064273 | .        | 1.06074 |
| 215                           | .        | -0.01233 | .        | 0.97671 | .        | -0.064273 | .        | 1.06074 |

그림 8.15 PROC STATESPACE를 사용한 다변량시계열데이터의 예측

고 표준편차들이다. 변수 Y2는 분석대상의 다변량시계열데이터의 두 번째 원소이고, 변수 FOR2, 변수 RES2, 그리고 변수 STD2는 각각 변수 Y2의 상태공간모형에 의한 예측(또는 적합), 잔차 그리고 표준편차들이다.

다변량시계열데이터 VARMA211의 PROC STATESPACE에 의한 예측과 PROC ARIMA에 의한 예측을 비교하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM8-16.PRG에 수록되어 있다).

```

/*
/*
/* EXAM8-16.PRG
/*
/* ARIMA FORECASTING OF VARMA211
/*
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=80 LINESIZE=76;
TITLE 'FORECAST BY ARIMA MODEL';
PROC ARIMA DATA=TSA4DATA.VARMA211;
 I VAR=Y1 NOPRINT;
 E P=1 NOPRINT;

```

```

F LEAD=15 OUT=TSA4DATA.FORE816A NOPRINT;
RUN;
I VAR=Y2 NOPRINT;
E Q=1 NOPRINT;
F LEAD=15 OUT=TSA4DATA.FORE816B NOPRINT;
RUN;
PROC PRINT DATA=TSA4DATA.FORE816A(FIRSTOBS=186);
VAR Y1 FORECAST STD;
RUN;
PROC PRINT DATA=TSA4DATA.FORE816B(FIRSTOBS=186);
VAR Y2 FORECAST STD;
RUN;

```

독자들이 잘 알다시피 PROC ARIMA는 단변량시계열분석을 위한 것이다. 따라서 이 다변량시계열데이터의 각 성분인 단변량시계열데이터에 PROC ARIMA를 적용한다. 변수 Y1에 알맞는 ARIMA모형의 차수를 찾기 위해서 Box-Jenkins방법을 적용한 결과, 이 단변량시계열데이터를 백색잡음과정으로 식별하였다. 백색잡음모형은 예측에 크게 도움이 되지 않으므로 벌칙함수식별방법들을 적용해서 이 단변량시계열데이터를 식별하기로 한다. 이 중에서 AIC값을 최소로 하는 차수는  $(p, q) = (1, 0)$ 이고, BIC값을 최소로 하는 차수도  $(p, q) = (1, 0)$ 이다. 즉, 이 단변량시계열데이터 Y1을 ARMA(1, 0)모형으로 식별한다. 따라서 PROC ARIMA의 첫 번째 ESTIMATE스테이트먼트의 옵션 P=에 1을 할당하였다. 이 PROC ARIMA의 첫 번째 IDENTIFY스테이트먼트와 첫 번째 ESTIMATE스테이트먼트 모두에 옵션 NOPRINT를 사용했으므로 이 스테이트먼트들을 수행한 결과가 출력되지 않는다. 첫 번째 FORECAST스테이트먼트의 옵션 LEAD=에 15를 지정하였으므로 주어진 200개의 관찰점들에서의 변수 Y1의 예측, 잔차, 그리고 표준편차를 계산하고, 마지막 관찰점 이후의 15개 선행시점들에서 예측과 표준편차들을 계산해서 이 값들을 옵션 OUT=에 지정한 영구SAS데이터세트 FORE816A에 저장한다. 이 FORECAST스테이트먼트의 옵션 NOPRINT를 사용했으므로 이 스테이트먼트를 수행한 결과가 출력되지 않는다. 변수 Y2에 알맞는 ARIMA모형의 차수를 찾기 위해서 Box-Jenkins방법을 적용한 결과, 이 단변량시계열데이터를 AR(1)모형 또는 MA(1)모형으로 식별하였다. 벌칙함수식별방법들을 적용하면, AIC를 최소로 하는 차수는  $(p, q) = (0, 1)$ 이고, BIC를 최소로 하는 차수도  $(p, q) = (0, 1)$ 이다. 즉, 이 단변량시계열데이터 Y2를 MA(1)모형으로 식별한다. 따라서 PROC ARIMA의 두 번째 ESTIMATE스테이트먼트의 옵션 Q=에 1을 할당하였다. 이 두 번째 IDENTIFY스테이트먼트와 두 번째 ESTIMATE스테이트먼트 모두에 옵션 NOPRINT를 사용했으므로 이 스테이트먼트들을 수행한 결과들이 출력되지 않는다. 두 번째 FORECAST스테이트먼트의 옵션 LEAD=에 15를 지정하였으므로 주어진 200개의 관찰점들에서의 변수 Y2의 예측, 잔차, 그리고 표준편차를 계산하고, 마지막 관찰점 이후의 15개 선행시점들에서 예측과 표준편차들을 계산해서 이 값들을 옵션 OUT=에 지정한

| FORECAST BY ARIMA MODEL     |          |           |         |
|-----------------------------|----------|-----------|---------|
|                             |          |           | 2       |
| 20:13 Sunday, July 11, 1993 |          |           |         |
| OBS                         | Y1       | FORECAST  | STD     |
| 186                         | -0.35968 | -0.032391 | 0.97166 |
| 187                         | 0.07490  | -0.001388 | 0.97166 |
| 188                         | -0.45997 | -0.008540 | 0.97166 |
| 189                         | -0.02564 | 0.000263  | 0.97166 |
| 190                         | 0.11598  | -0.006885 | 0.97166 |
| 191                         | 1.80152  | -0.009216 | 0.97166 |
| 192                         | -1.47352 | -0.036958 | 0.97166 |
| 193                         | 0.66057  | 0.016945  | 0.97166 |
| 194                         | -0.98978 | -0.018179 | 0.97166 |
| 195                         | -0.58119 | 0.008983  | 0.97166 |
| 196                         | 2.05518  | 0.002258  | 0.97166 |
| 197                         | 0.69242  | -0.041133 | 0.97166 |
| 198                         | -0.56261 | -0.018704 | 0.97166 |
| 199                         | 0.09747  | 0.001952  | 0.97166 |
| 200                         | 0.49213  | -0.008912 | 0.97166 |
| 201                         | .        | -0.015407 | 0.97166 |
| 202                         | .        | -0.007054 | 0.97179 |
| 203                         | .        | -0.007191 | 0.97179 |
| 204                         | .        | -0.007189 | 0.97179 |
| 205                         | .        | -0.007189 | 0.97179 |
| 206                         | .        | -0.007189 | 0.97179 |
| 207                         | .        | -0.007189 | 0.97179 |
| 208                         | .        | -0.007189 | 0.97179 |
| 209                         | .        | -0.007189 | 0.97179 |
| 210                         | .        | -0.007189 | 0.97179 |
| 211                         | .        | -0.007189 | 0.97179 |
| 212                         | .        | -0.007189 | 0.97179 |
| 213                         | .        | -0.007189 | 0.97179 |
| 214                         | .        | -0.007189 | 0.97179 |
| 215                         | .        | -0.007189 | 0.97179 |
| OBS                         | Y2       | FORECAST  | STD     |
| 186                         | 1.38208  | 0.15143   | 1.06307 |
| 187                         | -1.30414 | -0.47393  | 1.06307 |
| 188                         | 0.82814  | 0.21431   | 1.06307 |
| 189                         | 0.13664  | -0.26793  | 1.06307 |
| 190                         | -0.38686 | -0.19805  | 1.06307 |
| 191                         | 0.25818  | 0.00011   | 1.06307 |
| 192                         | -1.98487 | -0.14912  | 1.06307 |
| 193                         | 0.90553  | 0.55012   | 1.06307 |
| 194                         | 0.60919  | -0.18163  | 1.06307 |
| 195                         | -1.38068 | -0.32704  | 1.06307 |
| 196                         | 0.42269  | 0.28893   | 1.06307 |
| 197                         | 0.09506  | -0.10761  | 1.06307 |
| 198                         | 0.15057  | -0.13062  | 1.06307 |
| 199                         | 1.61319  | -0.15685  | 1.06307 |
| 200                         | -0.16908 | -0.65406  | 1.06307 |
| 201                         | .        | -0.22490  | 1.06307 |
| 202                         | .        | -0.06294  | 1.12078 |
| 203                         | .        | -0.06294  | 1.12078 |
| 204                         | .        | -0.06294  | 1.12078 |
| 205                         | .        | -0.06294  | 1.12078 |
| 206                         | .        | -0.06294  | 1.12078 |
| 207                         | .        | -0.06294  | 1.12078 |
| 208                         | .        | -0.06294  | 1.12078 |
| 209                         | .        | -0.06294  | 1.12078 |
| 210                         | .        | -0.06294  | 1.12078 |
| 211                         | .        | -0.06294  | 1.12078 |
| 212                         | .        | -0.06294  | 1.12078 |
| 213                         | .        | -0.06294  | 1.12078 |
| 214                         | .        | -0.06294  | 1.12078 |
| 215                         | .        | -0.06294  | 1.12078 |

그림 8.16 PROC ARIMA를 사용한 다변량시계열데이터의 예측

영구SAS데이터세트 FORE816B에 저장한다. 이 FORECAST스테이트먼트의 옵션 NOPRINT를 사용했으므로 이 스테이트먼트를 수행한 결과가 출력되지 않는다. 이 예측결과들을 OUTPUT윈도우에 출력시키기 위해서 두 개의 PROC PRINT들을 수행하였다. 이 SAS프로그램을 수행한 결과가 그림 8.16이다.

다변량시계열데이터 VARMA211의 PROC STATESPACE에 의한 잔차와 PROC ARIMA에 의한 잔차를 비교하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM8-17.PRG에 수록되어 있다).

```

/* */
/* EXAM8-17.PRG */
/* */
/* COMPARISON OF RESIDUALS */
/* */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=80 LINESIZE=76;
TITLE 'COMPARISON OF RESIDUALS';
DATA EXAM817A;
 SET TSA4DATA.FORE815;
 RSTATE1 = RES1;
 RSTATE2 = RES2;
 KEEP RSTATE1 RSTATE2;
 RUN;
DATA EXAM817B;
 SET TSA4DATA.FORE816A;
 RARIMA1 = Y1 - FORECAST;
 KEEP RARIMA1;
 RUN;
DATA EXAM817C;
 SET TSA4DATA.FORE816B;
 RARIMA2 = Y2 - FORECAST;
 KEEP RARIMA2;
 RUN;
DATA EXAM817D;
 MERGE EXAM817A EXAM817B EXAM817C;
 T = _N_;
 IF T > 200 THEN DELETE;
 LABEL RSTATE1='RESIDUAL OF Y1 BY PROC STATESPACE';
 LABEL RARIMA1='RESIDUAL OF Y1 BY PROC ARIMA';
 LABEL RSTATE2='RESIDUAL OF Y2 BY PROC STATESPACE';
 LABEL RARIMA2='RESIDUAL OF Y2 BY PROC ARIMA';

```

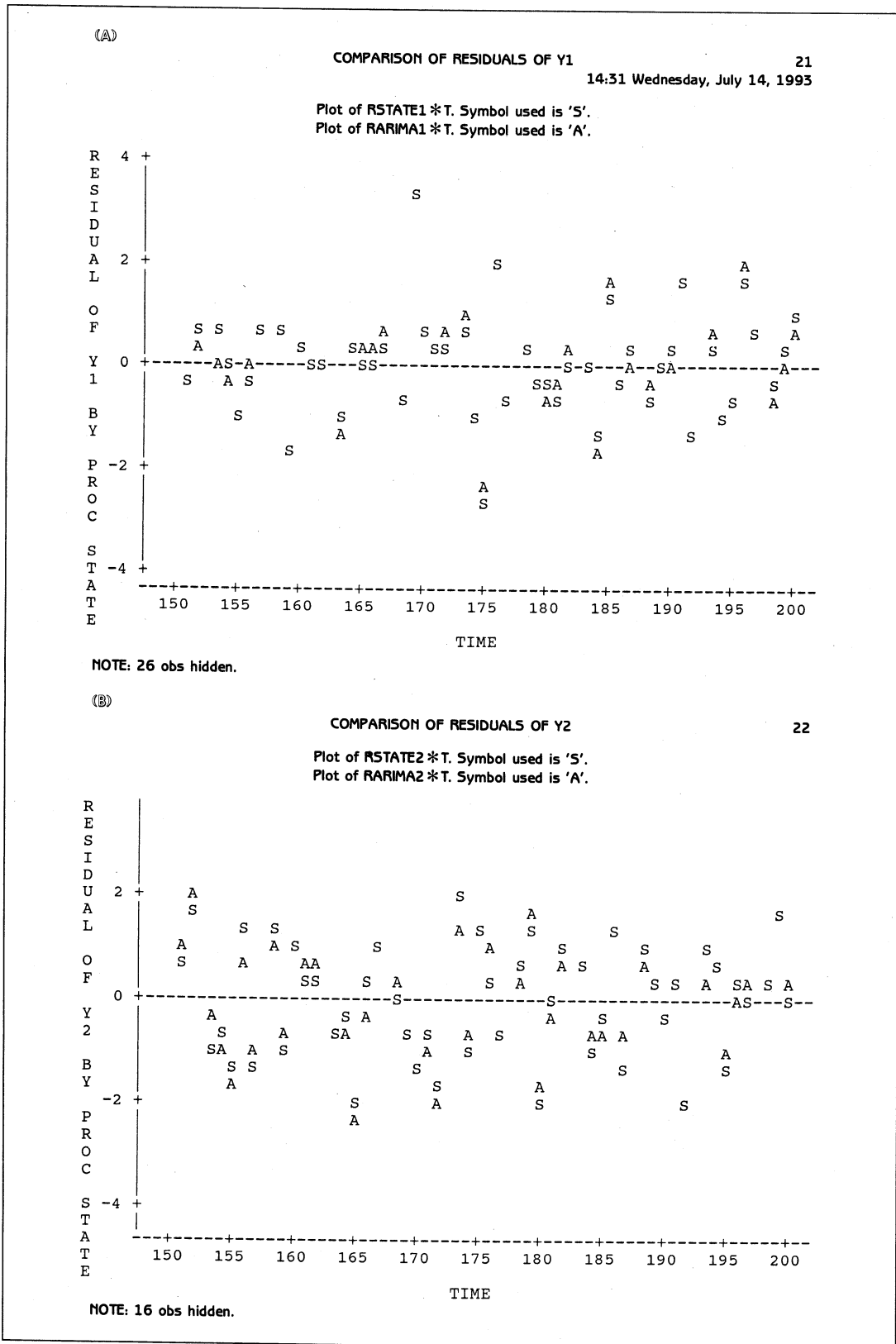


그림 8.17 PROC STATESPACE와 PROC ARIMA에 의한 잔차들의 비교



```

LABEL T = 'TIME';
KEEP T RSTATE1 RARIMA1 RSTATE2 RARIMA2;
RUN;
PROC PLOT DATA=EXAM817D(FIRSTOBS=151);
TITLE 'COMPARISON OF RESIDUALS OF Y1';
PLOT RSTATE1*T='S' RARIMA1*T='A'
 / VPOS=25 VREF=0 OVERLAY
 HAXIS = 150 TO 200 BY 5;
TITLE 'COMPARISON OF RESIDUALS OF Y2';
PLOT RSTATE2*T='S' RARIMA2*T='A'
 / VPOS=25 VREF=0 OVERLAY
 HAXIS = 150 TO 200 BY 5;
RUN;

```

이 SAS프로그램을 수행한 결과가 그림 8.17이다. 그림 8.17에서 볼 수 있듯이 PROC ARIMA에 의한 예측과 PROC STATESPACE에 의한 예측이 거의 일치한다. 즉, 이 예제에서는 두 프로시저들이 비슷한 결과를 낳는다. 상식적으로 보아 이 결과는 그렇게 만족스러운 것은 아니다. 그 이유는 다음과 같다. PROC ARIMA를 사용해서 예측을 하면 다변량시계열데이터를 구성하는 각 단변량시계열데이터의 정보만을 가지고 예측을 하는 반면에 PROC STATESPACE를 사용하면 다변량시계열데이터가 가지고 있는 모든 정보를 가지고 예측을 한다. 따라서 PROC STATESPACE를 사용한 예측이 PROC ARIMA를 사용한 예측보다 정확해야 할 것이다. 이 예제에서 두 프로시저들이 비슷한 결과를 내는 이유는 제 8.1.2소절의 결론부분에서 알 수 있듯이 두 변수들 Y1과 Y2의 상관정도가 약해서 일 것이다.

### 8.3 상태공간모형화

이 절에서는 상태공간모형을 사용해서 VARMA모형을 구축하는 방법에 대해서 논한다. VARMA모형의 상태공간모형화는 Akaike(1973, 1974a, 1974b, 1975, 1976)에 의해서 개발된 것이다. 앞서도 언급했듯이 이 상태공간모형화방법을 완전히 이해하기 위해서는 상당한 정도의 수학적 지식 및 통계학적 지식이 있어야 한다. 상태공간모형화방법에 대한 자세한 수리적 증명은 제 8.4절로 미루고 이 절에서는 상태공간모형화방법의 골격만을 설명하고자 한다. 이 상태공간모형화방법에 대한 한글로 쓰인 문헌으로는 정은이(1993)가 있다. 이 석사논문은 상태공간모형화방법에 대해서 좀더 자세한 내용을 공부하고 싶은 독자들에게 많은 도움이 되리라 생각한다.

## 8.3.1 상태공간모형과 VARMA모형

이 소절에서는 단변량시계열데이터, 또는 다변량시계열데이터에서 상태공간모형과 VARMA모형 사이의 관계를 알아보기로 하자. Akaike(1974a)는 임의의 VARMA모형을 상태공간모형으로 나타낼 수 있고 또한 임의의 상태공간모형을 VARMA모형으로 나타낼 수 있음을 보였다. 우선 몇 가지 예제를 살펴보자.

첫 번째 예제는 다음과 같은 단변량ARMA(2, 1)모형이다.

$$y_t = 1.1y_{t-1} - 0.24y_{t-2} + v_t + 0.5v_{t-1}$$

이 ARMA과정을 MA( $\infty$ )모형으로 나타내면 다음과 같다.

$$y_t = v_t + 1.6v_{t-1} + \dots$$

따라서 다음 식들이 성립한다.

$$\begin{aligned} y_{t+1} &= y_{t+1|t} + v_{t+1} \\ y_{t+2|t+1} &= y_{t+2|t} + 1.6v_{t+1} \\ &= 1.1y_{t+1|t} - 0.24y_t + 1.6v_{t+1} \end{aligned}$$

이 식들을 정리하면 다음과 같다.

$$\begin{aligned} z_{t+1} &= Fz_t + Gv_{t+1} \\ y_t &= [1 \ 0] z_t \\ \text{단, } z_t &= (y_t, y_{t+1|t})' \\ F &= \begin{bmatrix} 0 & 1 \\ -0.24 & 1.1 \end{bmatrix}, \quad G = \begin{bmatrix} 1 \\ 1.6 \end{bmatrix} \end{aligned}$$

지금까지는 단변량ARMA(2, 1)과정을 상태공간모형으로 나타내었다. 지금부터는 이 상태공간모형으로부터 ARMA모형을 유도하자. 이 상태공간모형을 전개하면 다음과 같다.

$$\begin{aligned} y_{t+1} &= y_{t+1|t} + v_{t+1} \\ y_{t+2|t+1} &= 1.1y_{t+1|t} - 0.24y_t + 1.6v_{t+1} \end{aligned}$$

첫 번째 식을 두 번째 식에 대입하면 다음과 같다.

$$y_{t+2} - v_{t+2} = 1.1(y_{t+1} - v_{t+1}) - 0.24y_t + 1.6v_{t+1}$$

이 식을 정리하면 다음과 같은 ARMA(2, 1)모형을 얻는다.

$$y_t = 1.1y_{t-1} - 0.24y_{t-2} + v_t + 0.5v_{t-1}$$

두 번째 예제는 다음과 같은 2변량VAR(2)모형이다.

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 1.40 & 0.30 \\ 0.10 & 1.10 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} - \begin{bmatrix} 0.48 & 0.12 \\ 0.06 & 0.28 \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} + \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix}$$

이 식을 전개하면 다음과 같다.

$$\begin{aligned} y_{1,t} &= 1.40y_{1,t-1} + 0.30y_{2,t-1} - 0.48y_{1,t-2} - 0.12y_{2,t-2} + v_{1,t} \\ y_{2,t} &= 0.10y_{1,t-1} + 1.10y_{2,t-1} - 0.06y_{1,t-2} - 0.28y_{2,t-2} + v_{2,t} \end{aligned}$$

첫 번째 식으로부터 다음 식들을 얻는다.

$$\begin{aligned} y_{1,t+1|t} &= 1.40y_{1,t} + 0.30y_{2,t} - 0.48y_{1,t-1} - 0.12y_{2,t-1} \\ y_{1,t+2|t} &= 1.40y_{1,t+1|t} + 0.30y_{2,t+1|t} - 0.48y_{1,t} - 0.12y_{2,t} \end{aligned}$$

두 번째 식으로부터 다음 식들을 얻는다.

$$\begin{aligned} y_{2,t+1|t} &= 0.10y_{1,t} + 1.10y_{2,t} - 0.06y_{1,t-1} - 0.28y_{2,t-1} \\ y_{2,t+2|t} &= 0.10y_{1,t+1|t} + 1.10y_{2,t+1|t} - 0.06y_{1,t} - 0.28y_{2,t} \end{aligned}$$

이 식들을 정리하면 다음과 같다.

$$z_{t+1} = Fz_t + Gv_{t+1}$$

$$y_t = [I_2 \ O] z_t$$

$$\text{단, } z_t = (y_{1,t}, y_{2,t}, y_{1,t+1|t}, y_{2,t+1|t})'$$

$$F = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -0.48 & -0.12 & 1.40 & 0.30 \\ -0.06 & -0.28 & 0.10 & 1.10 \end{bmatrix}, \quad G = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1.4 & 0.3 \\ 0.1 & 1.1 \end{bmatrix}$$

지금까지는 2변량VAR(2)과정을 상태공간모형으로 나타내었다. 이제는 이 상태공간모형으로부터 VARMA모형을 유도하자. 이 상태공간모형을 전개하면 다음과 같다.

$$\begin{aligned} y_{1,t+1} &= y_{1,t+1|t} + v_{1,t+1} \\ y_{2,t+1} &= y_{2,t+1|t} + v_{2,t+1} \\ y_{1,t+2|t+1} &= -0.48y_{1,t} - 0.12y_{2,t} + 1.40y_{1,t+1|t} + 0.30y_{2,t+1|t} \\ &\quad + 1.40v_{1,t+1} + 0.30v_{2,t+1} \\ y_{2,t+2|t+1} &= -0.06y_{1,t} - 0.28y_{2,t} + 0.10y_{1,t+1|t} + 1.10y_{2,t+1|t} \\ &\quad + 0.10v_{1,t+1} + 1.10v_{2,t+1} \end{aligned}$$

첫 번째 식과 두 번째 식을 세 번째 식과 네 번째 식에 대입해서 정리하면 다음과 같다.

$$\begin{aligned} y_{1,t+2} &= 1.40y_{1,t+1} + 0.30y_{2,t+1} - 0.48y_{1,t} - 0.12y_{2,t} + v_{1,t+2} \\ y_{2,t+2} &= 0.10y_{1,t+1} + 1.10y_{2,t+1} - 0.06y_{1,t} - 0.28y_{2,t} + v_{2,t+2} \end{aligned}$$

이 식들을 정리하면 다음과 같다.

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 1.40 & 0.30 \\ 0.10 & 1.10 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} - \begin{bmatrix} 0.48 & 0.12 \\ 0.06 & 0.28 \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} + \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix}$$

세 번째 예제는 다음과 같은 2변량VARMA(1, 1)모형이다.

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 0.45 & -0.11 \\ -0.14 & -1.23 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} - \begin{bmatrix} -0.40 & 0.12 \\ 0.00 & 0.00 \end{bmatrix} \begin{bmatrix} v_{1,t-1} \\ v_{2,t-1} \end{bmatrix}$$

이 2변량VARMA(1, 1)모형을 다음과 같은 상태공간모형으로 나타낼 수 있음을 제 8.1.1 소절에서 보았다.

$$z_{t+1} = Fz_t + Gv_{t+1}$$

$$y_t = [I_2 \ 0] z_t$$

$$\text{단, } z_t = (y_{1,t}, y_{2,t}, y_{1,t+1|t})'$$

$$F = \begin{bmatrix} 0 & 0 & 1 \\ -0.14 & -0.23 & 0 \\ 0.0154 & 0.0253 & 0.45 \end{bmatrix}, \quad G = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0.05 & -0.23 \end{bmatrix}$$

이제 이 상태공간모형으로부터 VARMA모형을 유도하자. 이 상태공간모형을 전개하면 다음과 같다.

$$y_{1,t+1} = y_{1,t+1|t} + v_{1,t+1}$$

$$y_{2,t+1} = -0.14y_{1,t} - 0.23y_{2,t} + v_{2,t+1}$$

$$y_{1,t+2|t+1} = 0.0154y_{1,t} + 0.0253y_{2,t} + 0.45y_{1,t+1|t} \\ + 0.05v_{1,t+1} - 0.23v_{2,t+1}$$

첫 번째 식을 세 번째 식에 대입해서 정리하면 다음과 같다.

$$y_{1,t+1} = 0.45y_{1,t} - 0.11y_{2,t} + v_{1,t+1} - 0.40v_{1,t} + 0.12v_{2,t}$$

이 식과 위의 두 번째 식을 정리하면 다음과 같은 2변량VARMA(1, 1)모형을 얻는다.

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 0.45 & -0.11 \\ -0.14 & -0.23 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} - \begin{bmatrix} -0.40 & 0.12 \\ 0.00 & 0.00 \end{bmatrix} \begin{bmatrix} v_{1,t-1} \\ v_{2,t-1} \end{bmatrix}$$

지금까지 설명했듯이 임의의 VARMA모형을 상태공간모형으로 나타낼 수 있고 또한 임의의 상태공간모형을 VARMA모형으로 나타낼 수 있다. 상태공간모형을 VARMA모형 형태로 전환시킬 때 유의할 점은 상태벡터의 원소들이 선형독립이 아니면 VAR특성행렬과 VMA특성행렬이 비단순(nontrivial) 공통항(common factor)을 갖는다. 따라서 VARMA모형의 차수들이 필요 이상으로 커진다. 따라서 상태공간모형화방법을 적용할 때는 최저차수의 상태벡터를 찾아야 한다. 이러한 이유로 상태벡터를 미래를 예측하기 위한 최소의

정보량으로 정의한다. 앞에서 보았듯이 상태벡터  $z_t$ 의 첫  $d$ 개 원소들은  $y_t$ 의 원소들이므로, 상태벡터  $z_t$ 의 나머지 원소들을 상태공간모형에서 제거함으로써 VARMA모형을 유도할 수 있다. 따라서 Akaike의 상태공간모형화방법은 VARMA모형화방법의 하나로 사용될 수 있다. 그러나 일반적으로는 추정된 상태공간모형을 VARMA모형으로 변환시키지는 않고, 선택된 상태공간모형 자체를 동태관계분석이나 예측 등의 목적에 직접 사용한다.

지금부터는 일반적인 VARMA과정을 상태공간모형으로 나타내보자. 제 6장에서 정의한  $d$ 차원VARMA모형은 다음과 같다.

$$\begin{aligned} \Phi(B)y_t &= \Theta(B)v_t \\ \text{단, } \Phi(z) &= -\Phi_0 - \Phi_1 z - \dots - \Phi_p z^p \quad (\Phi_0 = -I_d) \\ \Theta(z) &= -\Theta_0 - \Theta_1 z - \dots - \Theta_q z^q \quad (\Theta_0 = -I_d) \end{aligned}$$

여기서  $\{v_t\}$ 는 평균벡터가 영벡터이고 공분산행렬이  $\Sigma$ 인  $d$ 변량백색잡음과정이다. 이 VARMA( $p, q$ )모형이 작인성조건과 가역성조건을 만족한다고 가정하자. 제 6장에서 정의하였듯이 VARMA( $p, q$ )과정  $\{y_t\}$ 가 작인적이기 위해서는 VAR특성방정식  $\det[\Phi(x)] = 0$ 의 모든 근이 단위원 밖에 존재해야 하며, 또한 가역적이기 위해서는 VMA특성방정식  $\det[\Theta(x)] = 0$ 의 모든 근이 단위원 밖에 존재해야 한다. 이 다변량시계열데이터는 작인성조건이 만족된다고 가정하므로  $\{y_t\}$ 는 다음과 같은 VMA모형으로 나타낼 수 있다.

$$y_t = \sum_{j=0}^{\infty} \Psi_j B^j \quad (\Psi_0 = I_d)$$

따라서  $\Psi_j$ 는 입력시계열데이터가  $\{v_t\}$ 이고 출력시계열데이터가  $\{y_t\}$ 인 시간에 대해 불변인 선형시스템(time-invariant linear system)의 충격반응행렬(impulse response matrix)이다.

제 8.1 절에서 정의했듯이  $\{y_1, y_2, \dots, y_t\}$ 가 주어진 경우에  $y_{t+i}$ 의 최량예측  $y_{t+i|t}$ 는 다음과 같다.

$$y_{t+i|t} = E(y_{t+i} | y_s, s \leq t)$$

또한,  $w_{s|t}$ 는  $w_s$ 를 힐버트공간  $L\{y_s, y_{t-1}, \dots\}$ 으로 투영한 것으로 다음과 같다.

$$w_{s|t} = E(w_s | y_s, y_{t-1}, \dots)$$

이후  $p > q$ 라고 가정하기 위해서 다음과 같이 정의한다.

$$\Phi_{p+1} = \Phi_{p+2} = \dots = \Phi_q = 0$$

이러한 조건하에서 VARMA( $p, q$ )모형은 다음 정리에서와 같은 모형으로 나타낼 수 있다.

**정리 8.1** 작인성을 갖는 VARMA( $p, q$ )모형은 다음과 같은 모형으로 나타낼 수 있다.

$$\begin{bmatrix} y_{t+1} \\ y_{t+2|t+1} \\ \vdots \\ y_{t+p-1|t+1} \\ y_{t+p|t+1} \end{bmatrix} = \begin{bmatrix} O & I_d & O & \cdots & O & O \\ O & O & I_d & \cdots & O & O \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ O & O & O & \cdots & O & I_d \\ \Phi_p & \Phi_{p-1} & \Phi_{p-2} & \cdots & \Phi_2 & \Phi_1 \end{bmatrix} \begin{bmatrix} y_t \\ y_{t+1|t} \\ \vdots \\ y_{t+p-2|t} \\ y_{t+p-1|t} \end{bmatrix} = \begin{bmatrix} I_d \\ \Psi_1 \\ \vdots \\ \Psi_{p-2} \\ \Psi_{p-1} \end{bmatrix} u_{t+1}$$

$$y_t = [I_d \ O \ \cdots \ O \ O] \begin{bmatrix} y_t \\ y_{t+1|t} \\ \vdots \\ y_{t+p-2|t} \\ y_{t+p-1|t} \end{bmatrix}$$



정리 8.1에서 정의한  $(y_{t|t}, y_{t+1|t}, \dots, y_{t+p-1|t})'$ 는 상태벡터일 가능성을 가지고 있지만 반드시 상태벡터라고는 할 수 없다. 상태벡터는 미래시점들의 벡터들을 예측하기 위해서 필요한 현재와 과거의 정보의 최소량으로 정의되므로, 이 벡터가 상태벡터이기 위해서는 이 벡터의 원소들이 선형독립이어야 한다. 따라서 이 벡터의 원소들로 구성된 부벡터(sub-vector)들 중에서 원소들이 선형독립인 부벡터가 상태벡터이다. 정의에 의해서 벡터  $y_{t|t}$ 의 원소들은 선형독립이므로 상태벡터  $z_t$ 의 첫  $d$ 개의 원소는 벡터  $y_{t|t}$ 의 원소들이다.

정리 8.1에서 보았듯이 상태벡터의 처음  $d$ 개의 원소는  $y_t$ 의 원소들과 같도록 할 수 있다. 따라서 상태공간모형식에서 상태벡터의 이 첫  $d$ 개의 원소들을 제외한 나머지 원소들을 제거하면 상태공간모형을 VARMA모형으로 나타낼 수 있다. 즉, 앞의 여러 예제들에서 보였고 제 8.4.1소절에서 증명하듯이 임의의 유한차원 상태벡터의 상태공간모형을 VARMA모형으로 나타낼 수 있다.

지금까지 설명한 대로 유한차원 상태벡터의 상태공간모형을 VARMA모형으로 나타낼 수 있고 또한 작인성을 갖는 VARMA모형을 상태공간모형으로 나타낼 수 있지만, 상태공간모형은 VARMA모형에 비해서 다음과 같은 장점들이 있다.

- ① 상태공간모형의 정준형태(canonical form)를 쉽게 찾을 수 있기 때문에 VARMA모형이 갖는 식별문제를 쉽게 해결할 수 있다.
- ② 상태공간모형을 사용하면 Kalman필터링을 사용해서 예측을 위한 간단한 반복기법(recursive procedure)을 만들 수도 있다.

이러한 이유들로 Akaike(1974a)는 상태공간모형을 사용해서 VARMA과정을 모형화할 것을 제안하였다. Akaike에 의한 상태공간모형을 사용한 VARMA모형화방법은 세 단계로 나눌 수 있다. [ETS6.1]의 제 18장에서는 이 세 단계를 각각 VAR모형단계(autoregressive models phase), 정준상관분석단계(canonical correlations phase) 그리고 최종추정단계(final estimation phase)라 부른다. 지금부터는 단변량시계열데이터, 또는 다변량시계열데이터  $y_1, y_2, \dots, y_T$ 가 주어졌을 때 이 시계열데이터에 적합한 상태공간모형을 구하는 과정을 이 세 단계로 나누어 설명하고자 한다.

## 8.3.2 VAR모형단계

상태공간모형화를 하기 위해서 주어진  $d$ 변량시계열데이터  $y_1, y_2, \dots, y_T$ 가 정상적 VAR-MA( $p, q$ )모형에서 생성된 것이라고 가정하자. 또한, 벡터  $y_t$ 의 공분산행렬  $\Gamma(0)$ 가 정칙(nonsingular)이라고 가정하자. 만약 이 행렬이 비정칙(singular)이면  $d$ 차원벡터  $y_t$ 의 원소들  $y_{1,t}, y_{2,t}, \dots, y_{d,t}$ 는 선형종속이다. 이는 이 다변량시계열데이터를 구성하는 단변량시계열데이터들 사이에 선형종속관계가 성립함을 뜻한다. 이러한 경우에는 선형독립성을 유지하는 단변량시계열데이터들로 구성된 다변량시계열데이터를 다루기로 한다. 이 다변량시계열데이터의 시차 0에서의 자기공분산행렬  $\Gamma(0)$ 은 정칙이다.

정리 8.1에 의해서 다음에 정의되는 벡터  $z_t$ 가 상태벡터의 후보이다.

$$z_t = (y_t, y_{t+1}, \dots, y_{t+p-1})'$$

이  $z_t$ 를 상태벡터의 후보라고 한 이유는 상태벡터는 미래를 예측하기 위해서 필요한 현재와 과거의 모든 정보를 함축하는 최소량이기 때문이다. 즉,  $z_t$ 의 원소들 사이에 선형종속관계가 존재할 수 있기 때문에, 비록  $z_t$ 가 미래를 예측하기 위해서 필요한 현재와 과거의 모든 정보를 포함했다 하더라도  $z_t$ 는 미래를 예측하기 위해서 필요한 현재와 과거의 모든 정보를 함축하는 최소량은 아닐 수도 있다. 벡터  $z_t$ 의 원소들이 선형종속인 경우에는 벡터  $z_t$ 의 원소들 중에서 선형독립인 원소들의 최대집합은 벡터  $z_t$ 와 같은 정보를 가지고 있다. 선형대수학을 수강한 독자라면  $z_t$ 의 원소들로 생성된 힐버트공간이 벡터  $z_t$ 의 원소들 중에서 선형독립인 원소들의 최대집합에 의해 생성된 힐버트공간과 같다는 것을 알고 있을 것이다. 따라서 상태공간모형화를 하기 위해서는 우선 정리 8.1의 VAR차수  $p$ 를 선택한 다음 벡터  $z_t$ 의 원소들 중에서 선형독립인 원소들의 최대집합을 선택해야 한다. 지금부터는 주어진  $d$ 변량시계열데이터  $y_1, y_2, \dots, y_T$ 에 알맞는 VAR차수  $p$ 를 선택하는 방법에 대해 다루고, 제 8.2.3소절에서 이 벡터  $z_t$ 의 원소들 중에서 선형독립인 원소들의 최대집합을 선택하는 방법에 대해서 다룬다. 편의상  $m$ 을 다음과 같이 정의한다.

$$m = p-1$$

시점  $t$ 에서  $\{y_t, y_{t-1}, \dots\}$ 로 생성된 힐버트공간  $D_t = L\{y_t, y_{t-1}, \dots\}$ 를 데이터공간(data space)이라 하고  $\{y_t, y_{t+1}, \dots, y_{t+m|t}\}$ 로 생성된 힐버트공간  $F_t = L\{y_t, y_{t+1}, \dots, y_{t+m|t}\}$ 을 예측공간(predictor space)이라 하자. 힐버트공간의 중요한 성질의 하나인 직교성질(orthogonal property)에 의해서 다음 등식이 성립한다.

$$E(y_{t+j} - y_{t+j|t}) y'_{t-i} = 0 \quad (i, j = 0, 1, \dots)$$

$$\text{즉, } E y_{t+j} y'_{t-i} = E y_{t+j|t} y'_{t-i} \quad (i, j = 0, 1, \dots)$$

예측공간의 기저(basis)와 데이터공간의 기저 사이의 공분산행렬  $\Gamma_\infty$ 는 다음과 같다.

$$\Gamma_\infty = \text{Cov}((y'_t, y'_{t+1|t}, \dots, y'_{t+m|t})', (y'_t, y'_{t-1}, y'_{t-2}, \dots)')$$

따라서 힐버트공간의 직교성질에 의해서 다음 식이 성립한다.

$$\Gamma_\infty = \begin{bmatrix} \Gamma(0) & \Gamma(1) & \Gamma(2) & \dots \\ \Gamma(1) & \Gamma(2) & \Gamma(3) & \dots \\ \vdots & \vdots & \vdots & \\ \Gamma(k) & \Gamma(k+1) & \Gamma(k+2) & \dots \\ \vdots & \vdots & \vdots & \\ \Gamma(m) & \Gamma(m+1) & \Gamma(m+2) & \dots \end{bmatrix}$$

여기서  $\Gamma(j)$ 는 시차  $j$ 에서의 자기공분산행렬이다. 이러한 형태의 행렬을 블럭행렬행렬 (block Hankel matrix)이라고 한다.

**정리 8.2** 예측공간이 유한차원이면, 다음 식을 만족하는 자연수  $r (\leq m)$ 과 계수행렬  $A_1, A_2, \dots, A_r$ 이 존재한다.

$$\Gamma(r+j) = \sum_{i=1}^r A_i \Gamma(r+j-i) \quad (j = 1, 2, \dots) \quad \square$$

정리 8.2를 이용하면  $\Gamma_\infty$ 의 랭크는 다음과 같은 유한차원 블럭행렬행렬  $\Gamma_m$ 의 랭크와 같음을 알 수 있다.

$$\Gamma_m = \begin{bmatrix} \Gamma(0) & \Gamma(1) & \Gamma(2) & \dots & \Gamma(m) \\ \Gamma(1) & \Gamma(2) & \Gamma(3) & \dots & \Gamma(m+1) \\ \vdots & \vdots & \vdots & & \vdots \\ \Gamma(k) & \Gamma(k+1) & \Gamma(k+2) & \dots & \Gamma(k+m) \\ \vdots & \vdots & \vdots & & \vdots \\ \Gamma(m) & \Gamma(m+1) & \Gamma(m+2) & \dots & \Gamma(2m) \end{bmatrix}$$

따라서 예측공간이  $\{y_t, y_{t+1|t}, \dots, y_{t+m|t}\}$ 에 의해서 생성된다고 하면 차수  $m$ 을 구하기 위해서 무한차원 벡터  $(y'_t, y'_{t-1}, \dots)'$ 와 벡터  $(y'_t, y'_{t+1}, \dots, y'_{t+m})'$ 의 공분산행렬을 분석하는 대신 유한차원 벡터  $(y'_t, y'_{t-1}, \dots, y'_{t-m})'$ 와 벡터  $(y'_t, y'_{t+1}, \dots, y'_{t+m})'$ 의 공분산행렬을 분석한다. 이는 근본적으로 미래시점의 시계열데이터  $y_{t+1}, y_{t+2}, \dots, y_{t+m}$ 를 예측하기 위해서 필요한 모든 정보(essentially all the information relevant for prediction of future values of the time series)가 벡터  $(y'_t, y'_{t-1}, \dots, y'_{t-m})'$ 에 포함되어 있음을 뜻한다. 따라서 이후 분석대상이 되는 시계열데이터는  $d$ 변량VAR( $m$ )모형에서 생성되었다고 가정한다.

PROC STATESPACE에서는 시계열데이터  $y_1, y_2, \dots, y_T$ 가 주어지면 이  $d$ 변량VAR모형의 차수는 AIC방법에 의해서 결정되고 VAR계수행렬들의 추정행렬들로는 Yule-Walker 추정행렬들이 사용된다. 여기까지를 상태공간모형화의 VAR모형단계라고 부른다. AIC방법에 의한 VAR모형의 차수선택과 Yule-Walker추정법에 대해서는 본서의 제 7장을 참조



하라. PROC STATESPACE 스테이트먼트의 옵션 ARMAX=을 사용해서 추정할 Yule-Walker 추정행렬들의 최대시차를 지정할 수 있으며, PROC STATESPACE 스테이트먼트의 옵션 OUTAR=을 사용해서 Yule-Walker 추정행렬들을 지정된 SAS 데이터세트에 저장할 수 있다. PROC STATESPACE 스테이트먼트의 옵션 MINIC을 사용하면 최소AIC값에 해당하는 Yule-Walker 추정행렬들만을 PROC STATESPACE 스테이트먼트의 옵션 OUTAR=에 지정된 SAS 데이터세트에 저장할 수 있다. 이들 옵션들에 대해서는 제 8.4 절에서 자세히 설명할 것이다.

### 8.3.3 정준상관분석단계

정리 8.1에서 언급한 대로 임의의 VARMA 모형은 다음과 같은 상태공간모형으로 표현할 수 있다.

$$\begin{aligned} z_{t+1} &= Fz_t + Gv_{t+1} \\ y_t &= Hz_t \end{aligned}$$

그러나 이 VARMA 모형의 상태공간표현이 유일한(unique) 것은 아니다. 이 상태벡터  $z_t$ 의 차원을  $k$ 라 하자. 만일  $k \times k$  행렬  $M$ 이 정칙이라면 이 VARMA 모형은 다음과 같은 상태벡터  $w_t$ 의 상태공간모형으로 나타낼 수 있다.

$$\begin{aligned} w_{t+1} &= MF M^{-1} w_t + MG v_{t+1} \\ y_t &= HM^{-1} w_t \\ \text{단, } w_t &= M z_t \end{aligned}$$

즉, 상태공간모형도 VARMA 모형과 마찬가지로 식별의 문제를 지니고 있다. 따라서 상태공간모형을 사용해서 VARMA 모형화하기 위해서는 상태공간모형의 표준적(canonical) 형태를 정할 필요가 있다.

제 8.2.2 소절에서 상태공간벡터  $z_t$ 의 원소들은 벡터들  $y_t, y_{t+1}, \dots, y_{t+m|t}$ 의 원소들로 구성할 수 있다는 것을 알았다. 또한, SAS/ETS의 PROC STATESPACE에서는  $p(=m+1)$ 의 값으로 주어진 시계열데이터를 VAR 모형들에 적합시켰을 때 최소AIC값을 갖는 차수를 선택하는 것을 알았다. Akaike(1974a)는 이 상태벡터  $z_t$ 를 사용한 상태공간모형으로 주어진 VARMA 과정을 상태공간모형화할 것을 제안하였다. 지금부터는  $y_t, y_{t+1}, \dots, y_{t+m|t}$ 의 원소들 중에서 상태벡터  $z_t$ 를 구성하는 원소들을 찾는 방법을 살펴보자.

상태벡터란 미래를 예측하기 위해서 필요한 현재와 과거의 모든 정보를 포함하는 최소 집합이다. 예측벡터들  $y_t, y_{t+1}, \dots, y_{t+m|t}$ 의 구성성분들 사이에 선형종속이 존재할 수 있기 때문에, 상태벡터는 이 예측벡터들의 구성성분들 중에서 선형독립인 원소들의 최대집합으로 이루어진다. 따라서 상태벡터를 구하기 위해서는 예측벡터들  $y_t, y_{t+1}, \dots, y_{t+m|t}$ 에 의해 생성되는 예측공간  $F_t = L\{y_t, y_{t+1}, \dots, y_{t+m|t}\}$ 의 기저를 찾아야 한다. 이 예측공간의 기저

가 여러 개 존재할 수 있으므로 상태벡터도 여러 개 존재할 수 있다. 따라서 VARMA과정을 상태공간모형화하는 과정에서도 식별의 문제가 발생한다. Akaike(1974a)는 이러한 식별의 문제를 해결하기 위해서 다음 원소들의 순서로 선형독립성을 조사할 것을 제안하였다.

$$y_{1,t}, y_{2,t}, \dots, y_{d,t}, y_{1,t+1}, y_{2,t+1}, \dots, y_{d,t+1}, \dots, y_{1,t+m}, y_{2,t+m}, \dots, y_{d,t+m}$$

이 과정을 자세히 설명하면 다음과 같다. 가정에 의해서  $y_{1,t}, y_{2,t}, \dots, y_{d,t}$ 는 선형독립이다. 따라서 우선 다음 원소들이 선형독립인가를 조사한다.

$$y_{1,t}, y_{2,t}, \dots, y_{d,t}, y_{1,t+1}$$

만일 이 원소들이 선형종속이면, 이후  $y_{1,t+2}, y_{1,t+3}, \dots$ 은 분석대상에서 제외시킨다. 만일 이 원소들이 선형독립이면 다음 원소들이 선형독립인가를 조사한다.

$$y_{1,t}, y_{2,t}, \dots, y_{d,t}, y_{1,t+1}, y_{2,t+1}$$

만일 이 원소들이 선형종속이면 이후  $y_{2,t+2}, y_{2,t+3}, \dots$ 은 분석대상에서 제외시킨다. 이와 같은 과정을 반복해서 선형독립인 원소들의 최대집합을 구하면, 이 최대집합의 원소들이 상태벡터를 이룬다. 이렇게 상태벡터의 차원을 결정하는 것을 상태공간모형의 식별이라고 한다.

Akaike(1974a)는 이러한 선형독립성을 조사하기 위해서 정준상관분석(canonical correlation analysis)을 사용할 것을 제안하였다. 정준상관분석에 관한 자세한 내용은 제 8.5.3 소절을 참조하라.

**정리 8.3** 정규확률벡터  $y$ 의 차원이 정규확률벡터  $x$ 의 차원보다 크거나 같고 두 확률벡터들이 비퇴화(nondegenerate)이면, 조건부확률벡터  $x|y$ 의 차원은 확률벡터  $x$ 와 확률벡터  $y$ 의 정준상관계수들 중에서 0이 아닌 정준상관계수들의 개수와 같다. ▣

제 8.2.2 소절에서 언급되었듯이 VAR모형단계에서 선택한 VAR모형의 차수  $p(=m+1)$ 에 의해 정해지는  $\{y_t, y_{t-1}, \dots, y_{t-m}\}$ 이 미래를 예측하기 위한 모든 정보를 가지고 있다고 가정하기 때문에, 정준상관분석은 다음과 같이 정의되는 데이터공간  $D_t$ 와 예측공간  $F_t$  사이에서 행해진다.

$$D_t = L\{y_{1,t}, \dots, y_{d,t}, y_{1,t-1}, \dots, y_{d,t-1}, \dots, y_{1,t-m}, \dots, y_{d,t-m}\}$$

$$F_t = L\{y_{1,t}, \dots, y_{d,t}, y_{1,t+1}, \dots, y_{d,t+1}, \dots, y_{1,t+m}, \dots, y_{d,t+m}\}$$

상태벡터를 선택할 때 이용되는 중요한 성질의 하나는 만일  $y_{i,t+j|t}$ 가 상태벡터에서 제외되면 또한  $y_{i,t+j|t} (n=j+1, j+2, \dots)$ 도 상태벡터에서 제외된다. 왜냐하면  $y_{i,t+j|t}$ 를 전항들의 선형결합으로 나타낼 수 있으면  $y_{i,t+n|t}$  또한 전항들의 선형결합으로 나타낼 수 있기 때문이다.

**정리 8.4** 만일  $y_{i,t+j|t}$  를 전항들의 선형결합으로 나타낼 수 있으면, 임의의  $n(>j)$ 에 대해서  $y_{i,t+n|t}$ 도 전항들의 선형결합으로 나타낼 수 있다. ▣

이후  $i=1, 2, \dots, d$ , 그리고  $j=0, 1, \dots$  대해서 예측공간  $F_t$ 의  $(i, j)$ 번째 부분공간  $F_t^{(i,j)}$ 를 다음과 같이 정의하자.

$$F_t^{(i,j)} = L\{y_{1,t}, \dots, y_{d,t}, \dots, y_{i-1,t+j|t}, y_{i,t+j|t}\}$$

각 부분공간  $F_t^{(i,j)}$ 에서의 상태벡터  $z_t^{(i,j)}$ 는 부분공간  $F_t^{(i,j)}$ 와 데이터공간  $D_t$ 사이의 정준상관 분석을 통하여 결정된다. 따라서 다음과 같은 단계들을 거쳐 예측공간  $F_t$ 의 상태벡터를 찾는다.

(1단계) 예측공간  $F_t$ 의 첫 번째 부분공간  $F_t^{(d,0)} = L\{y_{1,t}, y_{2,t}, \dots, y_{d,t}\}$ 와 데이터공간  $D_t$ 사이의 정준상관계수들이 모두 1이므로 이 부분공간에서 상태벡터는 다음과 같다.

$$z_t^{(d,0)} = (y_{1,t}, y_{2,t}, \dots, y_{d,t})'$$

(2단계) 부분공간  $F_t^{(1,1)}$ 은 다음과 같다.

$$F_t^{(1,1)} = L\{y_{1,t}, y_{2,t}, \dots, y_{d,t}, y_{1,t+1|t}\}$$

① 만일  $F_t^{(1,1)}$ 과  $D_t$ 사이의 최소정준상관계수가 0이라면, 이에 해당하는  $F_t^{(1,1)}$ 의 원소들의 선형결합은 데이터공간  $D_t$ 의 원소들과 비상관관계에 있으므로 부분공간  $F_t^{(d,0)}$ 가 주는 정보 이외의 새로운 정보를 주지 못한다. 따라서 정리 8.3에 의해서  $y_{1,t+1|t}$ 는 상태벡터에서 제외된다. 또한 정리 8.4에 의해서  $y_{i,t+j|t}$  ( $j=2, 3, \dots$ )는 상태벡터에서 제외되므로 앞으로의 고려대상에서 제외한다. 따라서 새로운 상태벡터는 부분공간  $F_t^{(d,0)}$ 에서의 상태벡터와 같다. 즉, 다음 등식이 성립한다.

$$z_t^{(1,1)} = z_t^{(d,0)}$$

② 만일  $F_t^{(1,1)}$ 과  $D_t$ 사이의 최소정준상관계수가 0이 아니라면,  $y_{1,t+1|t}$ 를 현재의 상태벡터에 추가한다. 따라서 부분공간  $F_t^{(1,1)}$ 에서의 상태벡터는 다음과 같다.

$$z_t^{(1,1)} = (y_{1,t}, y_{2,t}, \dots, y_{d,t}, y_{1,t+1|t})'$$

(3단계) 제2단계의 과정을  $i=2, 3, \dots, d$ 에 대해서 수행한다. 즉, 부분공간  $F_t^{(i,1)}$ 을 다음과 같이 정의한다.

$$F_t^{(i,1)} = L\{y_{1,t}, y_{2,t}, \dots, y_{d,t}, y_{1,t+1|t}, \dots, y_{i,t+1|t}\}$$

① 만일  $F_t^{(i,1)}$ 과  $D_t$ 사이의 최소정준상관계수가 0이면, 정리 8.3에 의해서  $y_{i,t+1|t}$ 는 상태벡터에서 제외된다. 또한 정리 8.4에 의해서  $y_{i,t+j|t}$  ( $j=2, 3, \dots$ )은 상

상태벡터에서 제외된다. 따라서 부분공간  $F_t^{(i,1)}$ 에서의 상태벡터는 부분공간  $F_t^{(i-1,1)}$ 에서의 상태벡터와 같다. 즉, 다음 등식이 성립한다.

$$z_t^{(i,1)} = z_t^{(i-1,1)}$$

- ② 만일  $F_t^{(i,1)}$ 과  $D_t$  사이의 최소정준상관계수가 0이 아니라면,  $y_{i,t+1|t}$ 를 현재의 상태벡터에 추가한다. 즉, 부분공간  $F_t^{(i,1)}$ 에서의 상태벡터는 다음과 같다.

$$z_t^{(i,1)} = \begin{bmatrix} z_t^{(i-1,1)} \\ y_{i,t+1|t} \end{bmatrix}$$

(4단계) 제 2단계와 제 3단계의 과정을  $j=2, 3, \dots, m$  그리고  $i=1, 2, \dots, d$ 에 대해서 적용한다. 즉,  $y_{i,t+j-1|t}$ 가 상태벡터에 포함되지 않으면,  $y_{i,t+j|t}$ 도 상태벡터에 포함시키지 않는다. 그러나  $y_{i,t+j-1|t}$ 가 상태벡터에 포함되면 부분공간  $F_t^{(i,j)}$ 를 다음과 같이 정의한다.

$$F_t^{(i,j)} = L\{y_{1,t}, y_{2,t}, \dots, y_{d,t}, \dots, y_{i-1,t+j|t}, y_{i,t+j|t}\}$$

- ① 만일  $F_t^{(i,j)}$ 과  $D_t$  사이의 최소정준상관계수가 0이라면  $y_{i,t+j|t}$ 를 상태벡터에서 제외한다. 따라서 부분공간  $F_t^{(i,j)}$ 에서의 상태벡터는 부분공간  $F_t^{(i-1,j)}$ 의 상태벡터와 같다. 즉, 다음 식이 성립한다.

$$z_t^{(i,j)} = z_t^{(i-1,j)}$$

- ② 만일  $F_t^{(i,j)}$ 과  $D_t$  사이의 최소정준상관계수가 0이 아니라면,  $y_{i,t+j|t}$ 가 현재의 상태벡터에 추가된다. 즉,  $i=2, 3, \dots, d$ 인 경우의 부분공간  $F_t^{(i,j)}$ 에서의 상태벡터는 다음과 같다.

$$z_t^{(i,j)} = \begin{bmatrix} z_t^{(i-1,j)} \\ y_{i,t+j|t} \end{bmatrix}$$

또한  $i=1$ 인 경우의 부분공간  $F_t^{(1,j)}$ 에서 새로운 상태벡터는 다음과 같다.

$$z_t^{(1,j)} = \begin{bmatrix} z_t^{(d,j-1)} \\ y_{1,t+j|t} \end{bmatrix}$$

(5단계) 이와 같은 방법으로  $F_t^{(i,j)}$ 과  $D_t$  사이의 최소정준상관계수가 0이라는 귀무가설을 검정해서  $y_{i,t+j|t}$ 를 상태벡터에 추가할지 여부를 결정한다. 예측공간  $F_t$ 에 더 이상 상태벡터로 추가할 수 있는 원소가 없을 때 상태벡터의 선택을 끝낸다. 즉, 다음 등식들을 만족하는 정수들  $p_1, p_2, \dots, p_d$ 가 존재하면 상태벡터는  $z_t^{(d,p_d-1)}$ 이다.

$$z_t^{(1,p_1)} = z_t^{(1,p_1-1)}, \dots, z_t^{(d,p_d)} = z_t^{(d,p_d-1)}$$

이후 상태벡터는 다음과 같이 표기한다.

$$z_t = (z_{1,t}, z_{2,t}, \dots, z_{k,t})'$$

주어진 시계열데이터  $y_1, y_2, \dots, y_T$ 에 적합한 상태공간모형을 구하기 위해서는  $i=1, 2, \dots, d$ , 그리고  $j=1, 2, \dots, m$ 인 각  $(i, j)$ 단계에서 부분공간  $F_t^{(i,j)}$ 과  $D_t$  사이의 표본정준상관계수가 유의적으로 0과 다른가를 검정해야 한다. Bartlett (1939)은 표본정준상관계수의 유의성을 검정하기 위하여 다음과 같은 접근카이제곱검정을 사용할 것을 제안하였다.

**정리 8.5** 최소정준상관계수  $\rho_{min}$ 가 0이라는 귀무가설하에서 다음 통계량은 자유도가  $\{d(m+1)-q+1\}$ 인 접근카이제곱분포를 갖는다.

$$\chi^2 = - \left[ T - \frac{1}{2} \{d(m+1) + (q+1)\} \right] \ln(1 - \hat{\rho}_{min}^2)$$

여기서  $q$ 는 최소정준상관계수가 0인가를 검정하고자 하는 부분공간  $F_t^{(i,j)}$ 의 차원이다. ▣

Robinson(1973)은 MA모형에 정리 8.5의 Bartlett검정을 적용했을 때 문제가 발생할 수 있음을 보였다. Akaike(1976)는 최소정준상관계수가 0이라는 귀무가설을 검정하기 위해서 AIC를 변형한 다음과 같은 정보판단기준을 사용할 것을 제안하였다.

$$DIC = -t \ln(1 - \hat{\rho}_{min}^2) - 2\{d(m+1) - q + 1\}$$

만일,  $DIC < 0$ 이면 최소정준상관계수를 0으로 판단하고, 그렇지 않으면 0보다 큰 것으로 판단한다.

SAS/ETS의 PROC STATESPACE에서는 디플트로 DIC검정법을 사용해서 상태벡터를 정한다. 그러나 PROC STATESPACE스테이트먼트의 옵션 CANCELL을 사용하면 Bartlett검정의 결과도 출력시킬 수 있다. 또한 PROC STATESPACE스테이트먼트의 옵션 SIGCORR를 사용하면 DIC대신 다른 벌칙함수식별방법들을 사용할 수도 있고, PROC STATESPACE스테이트먼트의 옵션 DIMMAX=을 사용하면 상태벡터의 최대차원을 지정할 수 있다. PROC STATESPACE스테이트먼트의 옵션 PASTMIN=을 사용하면 정준상관분석에 사용하는 과거 데이터들의 시차의 최소값을 지정할 수 있다. 이 옵션들에 대한 자세한 내용은 제 8.4절을 참조하라. 특히 기억해 둘 것은 PROC STATESPACE의 FORM스테이트먼트를 사용해서 상태벡터를 임의로 지정할 수도 있다는 점이다. 이 FORM스테이트먼트에 대한 자세한 내용은 제 8.4절을 참조하라.

지금까지는 예측변수들  $y_t, y_{t-1}, \dots, y_{t-m}$ 의 원소들 중에서 선형독립인 원소들의 최대집합을 상태벡터로 선택하는 과정을 살펴보았다. 그러나 상태벡터가 정해졌다고 해서 변환행렬  $F$ 의 0이 아닌 원소들의 개수가 정해진 것은 아니다. 지금부터는 선택된 상태벡터에 해당하는 변환행렬  $F$ 의 형태를 정하는 과정을 살펴보자. 상태벡터  $z_t$ 의 차원이  $k$ 이므로 변환행렬  $F$ 는  $k \times k$ 행렬이다. 우선 다음과 같은  $d(m+1)$ 차원 벡터  $s$ 를 정의하자.

$$s = \begin{bmatrix} y_{t|t} \\ y_{t+1|t} \\ \vdots \\ y_{t+m-1|t} \\ y_{t+m|t} \end{bmatrix}$$

만일  $n = jd + i$  라고 하면,  $s$ 의 제  $n$ 번째 원소  $s_n$ 은 다음과 같다.

$$s_n = y_{i,t+j|t}$$

상태벡터  $z_t$ 는  $s_n = z_{i,t}$ 일 때  $h_i = n$ 으로 정의되는  $h_1, h_2, \dots, h_k$ 에 의해 결정된다. 벡터  $(h_1, h_2, \dots, h_k)'$ 를  $z_t$ 의 구조특성벡터(structural characteristic vector)라고 한다. 만일  $s_n = y_{i,t+j|t}$ 를 그 전항들의 선형결합으로 나타낼 수 있으면,  $t$ 를  $t+q$  ( $q=1, 2, \dots$ )로 대치시킨  $s_{n+qd} = y_{i,t+q+j|t}$  또한 전항들의 선형결합으로 표현할 수 있다. 만일  $i=1, 2, \dots, d$ 에 대해  $y_{i,t+j|t}$ 가  $s$ 의 전항들과 선형종속인  $j$ 들 중에서 가장 작은 값을  $j_i$ 라 하면  $d$ 차원벡터  $(j_1, j_2, \dots, j_d)'$  또한 상태벡터의 구조를 결정한다. 즉  $j=0, 1, \dots, j_i-1, i=1, 2, \dots, d$ 에 대해  $y_{i,t+j|t}$ 들을  $(i+jd)$ 의 오름차순으로 배열한 것이 상태벡터  $z_t$ 이다.

변환행렬  $F$ 를 결정하기 위해서는 다음 정리가 필요하다.

**정리 8.6** 만일  $z_{n,t} = y_{i,t+j-1|t}$ 라 하면  $z_{n,t+1|t}$ 는 다음과 같다.

$$z_{n,t+1|t} = y_{i,t+j|t} \quad \square$$

우선  $j=j_i$ 인 경우를 살펴보자. 만일  $z_{n,t} = y_{i,t+j-1|t}$ 라고 하면 정리 8.6에 의해서 다음 식이 성립한다.

$$z_{n,t+1|t} = y_{i,t+j|t}$$

즉,  $z_{n,t+1|t}$ 는 벡터  $s$ 에서  $y_{i,t+j|t}$ 의 전항들의 선형결합으로 표현된다. 만일  $h_1, h_2, \dots, h_k$  중에서  $(h_n + d)$ 보다 작은 수들의 개수를  $c(n)$ 이라 하면,  $m=c(n)+1, c(n)+2, \dots, k$ 에 대해서 변환행렬  $F$ 의  $(n, m)$ 번째 원소는  $F(n, m)=0$ 이다. 지금부터는  $m=1, 2, \dots, c(n)$ 에서  $F(n, m)$ 의 원소를 구해보자. 부분공간  $F_t^{(i,j)}$ 은 다음과 같다.

$$F_t^{(i,j)} = L\{z_{1,t}, z_{2,t}, \dots, z_{n,t}, y_{i,t+j|t}\}$$

이 부분공간  $F_t^{(i,j)}$ 와 데이터공간  $D_t$  사이의 최소의 표본정준상관계수가 유의적으로 0과 다르지 않다. 따라서 다음 식을 만족하는 상수들  $a_1, a_2, \dots, a_{n+1}$ 이 존재한다.

$$a_{n+1}y_{i,t+j|t} = \sum_{m=1}^n a_m z_{m,t}$$

이 식에서  $a_{n+1}=0$ 이면,  $z_{1,t}, z_{2,t}, \dots, z_{n,t}$ 가 선형독립이라는 가정에 어긋난다. 따라서 일반성을 잃지 않고  $a_{n+1}$ 을 1로 놓을 수 있다. 따라서 다음 식들이 성립한다.

$$F(n, m) = a_m \quad (m = 1, 2, \dots, c(n))$$

따라서 다음 식이 성립한다.

$$z_{n,t+1|t} = \sum_{m=1}^k F(n, m) z_{m,t} = \sum_{m=1}^{c(n)} a_m z_{m,t}$$

정의에 의해서  $z_{1,t}, z_{2,t}, \dots, z_{k,t}$ 가 선형독립이므로 이 식의 표현은 유일(unique)하다.

지금부터  $j < j_i$ 인 경우를 살펴보자.  $z_{n,t}$ 가  $y_{i,t+j-1|t}$ 와 같다면  $y_{i,t+j|t}$ 는  $z_i$ 의 원소로  $z_{d(n),t}$ 이다. 여기서  $d(n)$ 은  $h_1, h_2, \dots, h_k$  중에서  $(h_n + d)$ 보다 작거나 같은 수들의 개수이다. 이 경우에  $z_{n,t+1} = y_{i,t+j|t+1}$ 이므로 정리 8.6에 의해  $z_{n,t+1|t}$ 는  $y_{i,t+j|t}$ 와 같다. 이 값이  $z_i$ 의  $d(n)$ 번째 원소이다. 따라서  $m=d(n)$ 인 경우  $F(n, m)=1$ , 그렇지 않은 경우  $F(n, m)=0$  이라고 하면 다음 식이 성립한다.

$$z_{n,t+1|t} = \sum_{m=1}^k F(n, m) z_{m,t} = z_{d(i),t} z_{m,t}$$

정의에 의해서  $z_{1,t}, z_{2,t}, \dots, z_{k,t}$ 가 선형독립이므로 이 식의 표현은 유일하다.

마지막 두 식들과 다음 식에 의해 변환행렬  $F$ 는 원소들이 성긴(sparse) 행렬이라는 것을 알 수 있다.

$$z_{t+1|t} = Fz_t$$

실제 시계열데이터를 분석할 때 모형이 어떤 특별한 형태를 갖는다는 것을 알고 있으면 변환행렬  $F$ 나 입력행렬  $G$ 의 특정 원소들을 0으로 고정한다. 이러한 목적을 위해서 PROC STATESPACE의 RESTRICT스태이트먼트를 사용한다.

### 8.3.4 최종추정단계

VAR모형단계와 정준상관분석단계에서 얻은 추정값들을 초기값들로 이용해서 행렬들  $F, G$ , 그리고  $\Sigma$ 의 원소들의 효율적인 추정량들을 얻기 위한 최종추정단계에 대해서 살펴보자. Akaike(1973, 1974a, 1976)는 스펙트럴밀도함수를 바탕으로 한 최우추정법을 사용할 것을 제안했으나 이 방법이 널리 사용되지는 않는다. 이 소절에서는 PROC STATESPACE에서 사용하는 최우추정법과 조건부최소제곱추정법을 설명하고자 한다.

우선 최우추정법에 대해서 설명하면 다음과 같다. 상태방정식에서 다음 식을 얻는다.

$$z_t = (I - FB)^{-1} Gv_t$$

이 식으로부터 다음 식들을 얻는다.

$$y_t = H(I - FB)^{-1} Gv_t$$

$$v_t = [H(I - FB)^{-1} G]^{-1} y_t$$

따라서 대수우도함수는 다음 식을 만족한다.

$$\ln L(F, G, \Sigma | y_1, \dots, y_T) \propto -\frac{T}{2} \{d \ln 2\pi + \ln \det(\Sigma) + \text{trace}(\Sigma^{-1} S_0)\}$$

$$\text{단, } S_0 = \frac{1}{T} \sum_{i=1}^T v_i v_i'$$

이 대수우도함수를  $\Sigma$ 에 대해서 미분한 정규방정식을 풀면 다음과 같은  $\Sigma$ 의 추정행렬을 얻는다.

$$\hat{\Sigma} = \hat{S}_0 = \frac{1}{T} \sum_{i=1}^T \hat{v}_i \hat{v}_i'$$

이 추정행렬을 대수우도함수에 대입하면 다음과 같다.

$$\ln L(F, G, S_0 | y_1, \dots, y_T) \propto -\frac{T}{2} \{d \ln 2\pi + \ln \det(S_0) + \text{trace}(T I_d)\}$$

따라서 우도함수를 최대화하는 것은 단지  $\ln \det(\hat{S}_0)$ 를 최소화하는 것과 같다. 이  $\hat{S}_0$ 는 모수들의 비선형함수이므로  $\ln \det(\hat{S}_0)$ 는 수치적 절차(numerical procedure)를 사용해서 최소화한다.

PROC STATESPACE에서는 다음과 같은 방법으로  $\ln \det(\hat{S}_0)$ 를 최소화한다. 주어진 VARMA과정이 가역적이라고 가정했으므로 다음 식이 성립한다.

$$v_t = [H(I - FB)^{-1}G]^{-1} y_t \equiv \sum_{i=0}^{\infty} \Pi_i y_{t-i}$$

따라서 쇄신벡터의 공분산행렬  $S_0$ 는 다음과 같다.

$$S_0 = \sum_{i=0}^n \sum_{j=0}^n \Pi_i \Gamma(j-i) \Pi_j'$$

여기서  $\Gamma(j)$ 는 다변량시계열데이터  $\{y_t\}$ 의 시차  $j$ 에서의 자기공분산행렬이다. 시계열데이터의 관찰점들의 개수는 한정되어 있으므로 무한이 많은 표본자기공분산행렬들을 사용해서  $S_0$ 를 추정할 수는 없다. 따라서  $\hat{S}_0$  대신 다음과 같이 정의되는 절단된 형태의 근사식을 사용한다.

$$S_0^{(n)} = \sum_{i=0}^n \sum_{j=0}^n \Pi_i \Gamma(j-i) \Pi_j'$$

이 근사식에 사용되는 자기공분산행렬들의 최고차수  $n$ 을 PROC STATESPACE 스테이트먼트의 옵션 KLAG=에 지정한다. 물론  $n$ 의 값을 충분히 크게 해서 이 값보다 큰 시차에서는 행렬들  $\Pi_{n+1}, \Pi_{n+2}, \dots$ 이 영행렬에 가까워야 할 것이다. 특히 VMA부분이 있는 시계열데이터에서 좋은 추정량들을 얻기 위해서는  $n$ 의 값을 크게 해야 한다.

Akaike(1973b)는 이 대수우도함수에  $(-1)$ 을 곱한 함수를 두 번 미분해서 얻은 행렬, 즉 헤시안행렬(Hessian matrix)은 근사적으로 비음정치(nonnegative definiteness)임을 증명하였다. 따라서 추정행렬들은 대수함수 최대화의 제2차조건을 만족하는 것을 알 수 있다. PROC STATESPACE에서는 이 헤시안행렬의 근사행렬의 역행렬을 변환행렬  $F$ 와 입력행렬  $G$ 의 모수들의 추정량들의 공분산행렬의 추정행렬로 사용한다.



만약 시계열데이터의 VAR특성근이 단위원에 가깝다면, 즉 시계열데이터가 비정상성에 가까운 행태를 보인다면, 최종추정단계에서 PROC STATESPACE스태이트먼트의 옵션 RESIDEST를 사용해야 한다. 이 옵션을 사용하면 조건부최소제곱추정법을 적용해서 모형을 추정한다. 이 방법을 좀더 설명하면 다음과 같다. 상태공간모형을 사용해서 변수들의 표본평균들을 시작값(start-up value)들로 하는 잔차들을 계산한다. 주어진 시계열데이터를 구성하는 어떤 변수가 상태벡터에 나타나는 횟수의 최대값을  $i$ 라 하면, 제  $i$ 번째 이후의 잔차들을 사용해서 잔차벡터의 공분산행렬  $S_0$ 를 추정한다. 이 공분산행렬의 행렬식을 최소화하는 데는 Gauss-Marquardt알고리즘을 사용한다. 이 알고리즘에 관한 자세한 내용은 Kennedy & Gentle(pp. 483~485)을 참조하라. 조건부추정법은 앞의 방법에 비해 안정적인 결과를 주나, 계산적인 측면에서 보면 비싼(computationally expensive) 방법이다.

상태공간모형화의 예비분석단계(preliminary analysis stage)에서는 계산적으로 비싼 최종추정단계를 꼭 수행할 필요는 없다. 이러한 경우에는 PROC STATESPACE스태이트먼트의 옵션 NOEST를 사용해서 최종추정단계의 수행을 억제시킨다. 최종추정단계에서 사용할 수 있는 PROC STATESPACE스태이트먼트의 옵션들은 다음과 같다. 옵션 COVB는 앞에서 정의한 모수들의 추정량들의 공분산행렬의 추정행렬을 OUTPUT윈도우에 출력시킨다. 옵션 OUTMODEL=에는 모수들의 추정값들과 표준오차들을 저장할 SAS데이터세트를 지정한다. 옵션 MAXIT=에는 최우추정법이나 조건부최소제곱추정법을 적용할 때 반복횟수(number of iteration)의 상한선을 지정한다. 옵션 ITPRINT는 추정의 각 반복 단계에서 나타나는 결과물을 OUTPUT윈도우에 출력시킨다. 옵션 SINGULAR=에는 어떤 행렬이 비정칙행렬(nonsingula)인지를 판단하는 기준이 되는 값을 지정한다. 옵션 DETTIL=과 옵션 PARMTOL=은 추정과정에서 수렴성을 조사하는 데 사용한다. 이 옵션들에 대한 자세한 내용은 제 8.4 절을 참조하라.

## 8.4 PROC STATESPACE의 자세한 사용법

제 8.1 절에서 PROC STATESPACE의 간단한 사용법을 소개하였다. 이 절에서는 PROC STATESPACE의 자세한 사용법을 소개하고자 한다. 이 절의 내용은 [ETS6.1] (pp. 463~494)에 수록되어 있다.

### 8.4.1 PROC STATESPACE의 사용법

PROC STATESPACE의 사용양식은 다음과 같다.

```

PROC STATESPACE 옵션들;
 VAR 변수들;
 BY 변수들;
 FORM 변수 정수 ... 변수 정수;
 ID 변수;
 INITIAL F(행,열)=값 ... F(행,열)=값
 G(행,열)=값 ... G(행,열)=값;
 RESTRICT F(행,열)=값 ... F(행,열)=값
 G(행,열)=값 ... G(행,열)=값;

```

이 PROC STATESPACE의 스테이트먼트들 중에서 PROC STATESPACE 스테이트먼트와 VAR 스테이트먼트는 필수적이고 나머지 스테이트먼트들은 옵션이다. 지금부터는 각 스테이트먼트에 대해서 설명하고자 한다.

#### (1) PROC STATESPACE 스테이트먼트

PROC STATESPACE 스테이트먼트의 사용양식은 다음과 같다.

```

PROC STATESPACE 옵션들;

```

PROC STATESPACE 스테이트먼트에는 다섯 가지 종류의 옵션이 있다. 첫째는 데이터의 입력과 자기상관행렬함수(data input and autocorrelation)에 관한 옵션들, 둘째는 Yule-Walker 추정(Yule-Walker autoregressive estimate)에 관한 옵션들, 셋째는 정준상관분석단계(canonical correlation phase)에 관한 옵션들, 넷째는 최종추정단계(final estimation phase)에 관한 옵션들, 그리고 마지막으로 예측(forecasting)에 관한 옵션들이다.

##### (1.1) 데이터의 입력과 자기상관행렬함수에 관한 옵션들

###### ① DATA = SAS데이터세트

이 옵션에는 PROC STATESPACE에서 분석할 SAS데이터세트를 지정한다. 만일 이 옵션을 사용하지 않으면 가장 최근에 만들어진 SAS데이터세트를 분석한다.

###### ② LAGMAX = 정수

이 옵션에는 계산할 표본자기공분산행렬들의 시차들의 개수를 지정한다. 만일 'LAGMAX=3'이면 시차 0, 1, 2, 3에서의 표본자기공분산행렬들이 계산된다. 옵션 NOCENTER가 사용되지 않으면 시차  $i$ 의 표본자기공분산행렬은 다음과 같이 계산된다고 한다.

$$\hat{\Gamma}(i) = \frac{1}{T} \sum_{t=i+1}^T (y_{t-i} - \bar{y})(y_t - \bar{y})'$$

옵션 NOCENTER가 사용되면 시차  $i$ 의 표본자기공분산행렬은 다음과 같이 계산된다

고 한다.

$$\hat{\Gamma}(i) = \frac{1}{T} \sum_{t=i+1}^T y_{t-i} y_t'$$

SAS/ETS의 공식지침서인 [ETS6.1] (p. 465)에서는 이 식들의 합이 't=i'에서 시작하나 본 저자의 생각으로는 't=i+1'에서 시작해야 한다. 다변량시계열데이터  $\{y_t\}$ 의 원소들은 VAR스테이트먼트에 지정한 변수들로서 VAR스테이트먼트에 지정한 순서대로 구성된다. 이 옵션의 디폴트는 10이다.

### ③ NOCENTER

PROC STATESPACE에서는 주어진 시계열데이터를 분석하기 전에 시계열데이터에서 표본평균을 빼서 분석할 시계열데이터의 표본평균을 0으로 만든다. 그러나 옵션 NOCENTER를 사용하면 주어진 시계열데이터를 분석하기 전에 표본평균을 빼는 과정을 생략한다.

### ④ PRINTOUT = LONG|NONE|SHORT

이 옵션은 OUTPUT윈도우에 출력되는 기술통계량의 양을 조정한다. 만일 'PRINTOUT=NONE'을 지정하면 관찰점들의 개수, 각 변수의 변수이름, 표본평균과 표본표준편차, 예비추정단계의 결과, 그리고 최종추정단계의 결과를 출력한다. 만일 'PRINTOUT=SHORT'를 지정하면, 'PRINTOUT=NONE'을 지정했을 때의 결과물 이외에도 각 VAR모형에 해당하는 AIC값, 표본자기공분산행렬들의 구조적 표현, 표본부분자기공분산행렬들의 구조적 표현, 그리고 최소AIC값을 갖는 VAR모형의 계수행렬들의 Yule-Walker추정행렬들이 출력된다. 만일 'PRINTOUT = LONG'을 지정하면, 'PRINTOUT = SHORT'를 지정했을 때의 결과물 이외에도 주어진 시계열데이터의 표본자기공분산행렬들, 표본자기상관행렬들, 표본부분자기공분산행렬들, 그리고 각 VAR모형에 의해 생성되는 잔차들의 공분산행렬과 상관행렬을 출력한다. 이 옵션의 디폴트는 SHORT이다.

## (1.2) Yule-Walker추정에 관한 옵션들

### ① ARMAX = 정수

이 옵션에는 계산할 Yule-Walker추정행렬들의 최대시차를 지정한다. 이 옵션의 디폴트는 10이다.

### ② MINIC

이 옵션을 사용하면 최소AIC값을 갖는 VAR모형의 Yule-Walker추정행렬들을 옵션 OUTAR=에 지정한 SAS데이터세트에 출력한다.

### ③ OUTAR = SAS데이터세트

이 옵션에는 Yule-Walker추정단계의 결과물을 저장할 SAS데이터세트를 지정한다.

## (1.3) 정준상관분석단계의 옵션들

## ① CANGCORR

VAR모형단계에서 정준상관계수들, DIC값들, Bartlett의 카이제곱통계량값들과 이에 해당하는 자유도들을 출력한다.

## ② DIMMAX = 정수

이 옵션에는 상태벡터  $z_t$ 의 차원의 상한을 지정한다. 이 옵션의 디폴트는 10이다.

## ③ NOEST

이 옵션을 사용하면 최종추정단계를 수행하지 않는다. 최종추정단계는 계산적으로 비싸므로 (computationally expensive) 예비분석단계에서는 수행하지 않는 것이 좋을 수도 있다.

## ④ PASTMIN = 정수

이 옵션에는 정준상관분석단계에서 사용하는 데이터공간을 구성하는 과거항들의 시차들의 최대값의 하한을 지정한다. 이 옵션의 디폴트는 0이다.

## ⑤ SIGCORR = 숫자

이 옵션에는 상태공간모형을 식별하는데 사용하는 벌칙함수식별방법에서 자유도에 곱하는 숫자를 지정한다. 이 숫자는 정리 7.19의  $C(T)$ 에 해당하는 것으로 이 옵션의 디폴트는 2이다. 이는 AIC의  $C(T)$ 에 해당하는 것이다. 만일 BIC나 HQC를 사용하려면  $\ln(T)$ 나  $\ln \ln(T)$ 에 해당하는 숫자들을 지정해야 한다. 여기서  $T$ 는 관찰점들의 개수이다. 이론적으로 이 옵션에 1.0보다 작은 수를 지정하는 것은 의미가 없다. 이 옵션에 큰 값을 지정하면 선택되는 상태벡터의 차원은 작다. 따라서 이 옵션에 지정하는 숫자가 클수록 좀더 간단한 모형이 선택될 것이다.

## (1.4) 최종추정단계의 옵션들

## ① COVB

이 옵션을 사용하면 Fisher정보행렬 (information matrix)의 역행렬의 추정행렬을 출력한다. 대수우도함수를 모수벡터로 2차미분한 통계량에 기대값을 취한 뒤  $(-1)$ 을 곱한 행렬을 Fisher정보행렬이라 부른다. 이 Fisher정보행렬의 역행렬이 최우추정벡터의 공분산행렬임이 알려져 있다. 따라서 옵션 COVB를 사용해서 얻은 행렬은 추정벡터의 공분산행렬의 근사행렬로 사용할 수 있다.

## ② DETTOL = 숫자, 또는 PARMTOL = 숫자

이 옵션들은 반복적인 추정단계에서 수렴성을 검정하는 데 사용한다. 만일 추정값들의 상대적인 변화가 옵션 PARMTOL=에 지정한 값보다 작으며 쉐신벡터의 공분산행렬의 추정행렬의 행렬식의 상대적 변화가 옵션 DETTOL=에 지정한 값보다 작으면, 추정의 반복단계가 멈춰지고 현재의 추정값들이 최종추정값들로 사용된다. 옵션들 DETTOL=의

디폴트는 0.001이고 옵션 PARMTOL=의 디폴트는  $1.0E-5$ 이다.

③ ITPRINT

이 옵션을 사용하면 반복적인 추정과정의 각 단계의 추정값들을 OUTPUT윈도우에 출력한다.

④ KLAG = 정수

이 옵션에는 최종추정단계의 최우추정법에서 사용되는 자기공분산행렬들의 최고차수를 지정한다. 이 옵션의 디폴트는 15이다. 만일 VMA부분이 있는 경우에 좋은 추정값들을 얻기 위해서는 이 옵션에 큰 값을 할당해야 한다.

⑤ MAXIT = 정수

이 옵션에는 최우추정법이나 조건부최소제곱추정법을 적용할 때 허용하는 반복횟수의 상한선을 지정한다.

⑥ OUTMODEL = SAS데이터세트

이 옵션에는 최종추정값들과 이들의 표준오차들을 저장할 SAS데이터세트를 지정한다.

⑦ RESIDEST

이 옵션을 사용하면 조건부최소제곱추정법을 사용해서 상태공간모형을 최종추정한다. 제 8.3.4 소절에서 언급했듯이 이 조건부최소제곱추정법의 결과는 최우추정법의 결과에 비해 안정적이지만 계산적 측면에서 비싸다.

⑧ SINGULAR = 숫자

이 옵션에는 어느 행렬이 비정칙(singular)인지 여부를 결정하는 판단기준이 되는 숫자를 지정한다. 어떤 행렬에 소과법(掃過法, sweeping method)을 적용할 때 주축값(pivot)이 이 옵션에 지정된 값보다 작으면 이 행렬을 특이행렬이라고 판정한다. 이 옵션의 디폴트는  $1.0E-07$ 이다.

(1.5) 예측에 사용되는 옵션들

① BACK = 정수

이 옵션에는 시계열데이터의 마지막 관찰점으로부터 역순으로 몇 번째(the number of steps from the end of the given time series data) 관찰점으로부터 전방일단계예측(one-step ahead prediction)을 할 것인가를 지정한다. 이 옵션의 디폴트는 0이다.

② INTERVAL = 시간구간

이 옵션에는 관찰점들 사이의 시간구간(time interval)을 지정한다. 사용가능한 시간구간으로는 YEAR, QTR, MONTH, WEEK, 그리고 DAY가 있다. 일반적으로 이 옵션은 ID변수와 함께 사용되어, 입력된 시계열데이터가 올바른 순서로 배열되었는지를 조사하거나 결측점이 없는가를 조사하는 데 이용된다. 또한 이 옵션은 마지막 관찰점 다음의 ID변

수의 값들을 외삽(extrapolation)하는 데 사용한다. 시간구간에 대한 자세한 내용은 [SAS-4TSA2]의 제 2.5 절을 참조하라.

③ INTPER = 정수

이 옵션이 옵션 INTERVAL=과 함께 사용되면 옵션 INTERVAL=에 지정된 시간구간의 배수를 시간구간으로 지정한다. 한 예로 두 달마다 관찰된 시계열데이터에는 'INTERVAL=MONTH'와 'INTPER=2'를 사용해야 한다. 이 옵션이 옵션 INTERVAL=과 함께 사용되는 경우에는 디폴트가 1이다. 이 옵션이 옵션 INTERVAL=과 함께 사용되지 않으면 정확한 숫자구간(exact numerical interval)을 지정하는 데 사용한다.

④ LEAD = 정수

이 옵션에는 예측할 선시차(lead)들의 개수를 지정한다. 예측이 시작되는 시점은 옵션 BACK=에 지정한다. 이 옵션의 디폴트는 0이다.

⑤ OUT = SAS데이터세트

이 옵션에는 예측과 표준편차 들을 저장할 SAS데이터세트를 지정한다.

⑥ PRINT

이 옵션을 사용하면 예측과 이들의 표준편차들을 OUTPUT윈도우에 출력시킨다.

(2) BY스테이트먼트

BY스테이트먼트의 사용양식은 다음과 같다.

BY 변수들;

다른 SAS프로시저의 BY스테이트먼트와 마찬가지로 PROC STATESPACE에서 BY스테이트먼트를 사용하면 BY변수들에 의해서 구분되어진 그룹별로 상태공간모형화를 한다. BY스테이트먼트를 사용하기 위해서는 SAS데이터세트가 BY변수의 데이터값들 순으로 관찰점들이 배열되어 있어야 한다. BY스테이트먼트에 대한 자세한 내용은 [SAS4TSA1] (p. 91)을 참조하라.

(3) FORM스테이트먼트

FORM스테이트먼트의 사용양식은 다음과 같다.

FORM 변수 정수 ... 변수 정수;

FORM스테이트먼트는 어떤 변수와 그 변수의 시차변수들이 몇 번 상태벡터에 나타나지를 지정하는 데 사용한다. 한 예로 다음 FORM스테이트먼트를 살펴보자.

FORM X 3 Y 2;

이 FORM스테이트먼트에 의해서 상태공간벡터에 다음 항들이 포함된다.

$$x_{t|t}, x_{t+1|t}, x_{t+2|t}, y_{t|t}, y_{t+1|t}$$

#### (4) ID스테이트먼트

ID스테이트먼트의 사용양식은 다음과 같다.

ID 변수;

ID스테이트먼트에 의해 지정된 ID변수는 옵션 OUT=에 지정된 SAS데이터세트나 예측결과와 출력물에 포함된다. 만일 옵션 INTERVAL=이나 옵션 INTPER=가 사용되는 경우에는 예측시점들에 대해서 ID변수가 외삽된다.

#### (5) INITIAL스테이트먼트

INITIAL스테이트먼트의 사용양식은 다음과 같다.

INITIAL F(행,열)=값 ... F(행,열)=값  
G(행,열)=값 ... G(행,열)=값;

INITIAL스테이트먼트에는 변환행렬  $F$ 와 입력행렬  $G$ 의 원소들의 초기값들을 할당한다. 만일 할당된 원소가 모수가 아니라 상태공간모형의 구조상 사전에 정해지는 구조적 원소(structural element)라면 해당되는 초기값은 무시된다. 다음은 INITIAL스테이트먼트의 한 예이다.

$$\text{INITIAL F}(3, 2) = 0.4 \text{ F}(3, 4) = -0.2 \text{ G}(5, 1) = 0;$$

#### (6) RESTRICT스테이트먼트

RESTRICT스테이트먼트의 사용양식은 다음과 같다.

RESTRICT F(행,열)=값 ... F(행,열)=값  
G(행,열)=값 ... G(행,열)=값;

RESTRICT스테이트먼트에는 변환행렬  $F$ 와 입력행렬  $G$ 의 원소들의 값들을 할당한다. 만일 할당된 원소가 모수가 아니라 상태공간모형의 구조상 사전에 정해지는 구조적 원소라면 해당되는 값은 무시된다. 이 RESTRICT스테이트먼트는 상태공간모형을 사용해서

전이함수모형을 분석하는 데 유용하다. 다음은 RESTRICT스태이트먼트의 한 예이다.

RESTRICT F(3, 2) = 0 F(3, 4) = 0 F(5, 1) = 0 G(5, 1) = 0;

### (7) VAR스태이트먼트

VAR스태이트먼트의 사용양식은 다음과 같다.

VAR 변수(차분, ..., 차분) ... 변수(차분, ..., 차분);

VAR스태이트먼트에는 입력시계열데이터를 구성하는 변수들을 지정한다. 다변량시계열 데이터의 제  $i$ 번째 성분은 VAR스태이트먼트의 제  $i$ 번째 변수이다. 만일 VAR스태이트먼트의 첫 번째 변수를  $Y_1$ , 두 번째 변수를  $Y_2$ , ..., 제  $d$ 번째 변수를  $Y_d$ 라고 하고 VAR모형단계에서 차수  $p(=m+1)$ 이 선택된다면 다음 변수들이 상태벡터에 포함되는지 여부를 이 변수들이 배열된 순서대로 조사한다.

$$Y_{1,t}, Y_{2,t}, \dots, Y_{d,t}, Y_{1,t+1|t}, Y_{2,t+1|t}, \dots, Y_{d,t+1|t}, \dots, Y_{1,t+m|t}, Y_{2,t+m|t}, \dots, Y_{d,t+m|t}$$

전이함수모형의 관점에서 볼 때 입력시계열데이터들에 해당하는 변수들을 앞에 놓고 출력시계열데이터들에 해당하는 변수들을 뒤에 놓는다.

여러 번 강조했듯이 상태공간모형화는 정상시계열데이터에만 적용한다. 따라서 비정상시계열데이터는 정상시계열데이터로 바꾼 다음에 상태공간모형화를 해야 한다. 우선 분산의 이질성문제는 대수변환등 변환을 통해서 해결하고, 시계열데이터의 평균의 변화문제는 차분, 또는 계절차분을 사용해서 해결한다. PROC ARIMA의 IDENTIFICATION스태이트먼트의 옵션 VAR= 또는 옵션 CROSSCOR=에서와 마찬가지로 PROC STATESPACE의 VAR스태이트먼트에서도 각 변수의 차분과 계절차분을 수행할 수 있다. 한 예로 변수 Y에 시차 1의 차분을 취한 뒤, 계절시차 12인 계절차분을 취하기 위해서는 다음 VAR스태이트먼트를 사용한다.

VAR Y(1, 12);

이 VAR스태이트먼트의 사용양식은 PROC ARIMA의 IDENTIFICATION스태이트먼트의 옵션 VAR=, 또는 옵션 CROSSCOR=과 마찬가지로이다. 이에 대한 자세한 내용은 [SAS4TSA2] (p. 467), 또는 본서의 제 5.1.3소절을 참조하라. 한 가지 유의할 점은 변수 Y가 이렇게 차분되었어도 예측단계에서는 원래의 변수 Y에 대한 예측의 결과를 출력한다는 것이다.



## 8.4.2 단변량시계열데이터 BJ\_A

제 8.1.1 소절에서는 예제들을 사용해서 PROC STATESPACE의 간단한 사용법을 설명했다. 지금부터는 두 예제들의 분석을 통해서 PROC STATESPACE의 자세한 사용법을 설명하고자 한다. 이 예제들은 [ETS6.1] (pp. 479~493)에서 PROC STATESPACE를 설명하는 데 사용한 것이다.

이 소절에서는 Box & Jenkins(1976, p. 525)의 Series A인 단변량시계열데이터 BJ\_A를 상태공간모형화하고자 한다. 이 단변량시계열데이터의 변수 X는 어느 화학공정의 매 두 시간마다 측정된 농도(chemical process concentration readings: every two hours)를 기록한 것이다. 이 단변량시계열데이터의 관찰점들의 개수는 197이다. 이 단변량시계열데이터는 제 2.1 절에 수록되어 있고, 변수 X의 시계열산포도가 그림 2.1에 수록되어 있다. 이 다변량시계열데이터는 디스켓 SAS4TSA4의 서브디렉토리 TSA4DATA에 영구SAS데이터세트 BJA.SSD로 저장되어 있다.

단변량시계열데이터 BJ\_A를 상태공간모형화하기 위해서 다음 SAS프로그램을 수행해보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM8-18.PRG에 수록되어 있다).

```

/* */
/* EXAM8-18.PRG */
/* */
/* STATE SPACE MODELING */
/* DATASET = BJ_A */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=72;
/* ----- */
/* USE PROC ARIMA TO FIT MODEL AND FORECAST */
/* ----- */
TITLE 'ARIMA MODELING';
DATA EXAM818;
 SET TSA4DATA.BJA;
 RUN;
PROC ARIMA;
 I VAR=X NOPRINT;
 E P=1 Q=1;
 F LEAD=10 OUT=EXAM818B NOPRINT;
 RUN;

```

```

/ * ----- */
/ * USE PROC STATESPACE TO SELECT AND FIT */
/ * ----- */
TITLE 'STATESPACE MODELING';
PROC STATESPACE CANCELL PRINTOUT=LONG;
 VAR X;
 RUN;
/ * ----- */
/ * USE PROC STATESPACE TO SELECT AND FIT */
/ * ----- */
PROC STATESPACE RESIDEST LEAD=10
 OUT=EXAM818C OUTMODEL=MODEL1;
 VAR X;
 RESTRICT F(2,1)=0;
 RUN;
/ * ----- */
/ * PRINT MODEL */
/ * ----- */
TITLE 'MODEL FROM PROC STATESPACE';
PROC PRINT DATA=MODEL1;
 RUN;
/ * ----- */
/ * COMPARE FORECASTS */
/ * ----- */
TITLE 'COMPARING FORECASTS';
DATA;
 MERGE EXAM818B EXAM818C;
 RUN;
PROC PRINT DATA=_LAST_(FIRSTOBS=180);
 RUN;

```

이 SAS프로그램의 첫 번째 프로시저는 다음과 같다.

```

PROC ARIMA;
 I VAR=X NOPRINT;
 E P=1 Q=1;
 F LEAD=10 OUT=EXAM818B NOPRINT;
 RUN;

```

이 프로시저를 수행하는 목적은 PROC STATESPACE에 의한 예측의 결과를 PROC

```

ARIMA MODELING 15:52 Saturday, July 24, 1993
33

ARIMA Procedure

Name of variable = X.

Mean of working series = 17.06244
Standard deviation = 0.398232
Number of observations = 197

Conditional Least Squares Estimation

Parameter Estimate Approx.
 Std Error T Ratio Lag
MU 17.03883 0.09510 179.17 0
MA1,1 0.55797 0.08828 6.32 1
AR1,1 0.90365 0.04670 19.35 1

Constant Estimate = 1.64168984

Variance Estimate = 0.09965051
Std Error Estimate = 0.31567469
AIC = 107.73974*
SBC = 117.589351*
Number of Residuals = 197
*Does not include log determinant.

Correlations of the Estimates

Parameter MU MA1,1 AR1,1
MU 1.000 0.004 -0.025
MA1,1 0.004 1.000 0.738
AR1,1 -0.025 0.738 1.000

Autocorrelation Check of Residuals

To Chi Autocorrelations
Lag Square DF Prob
6 4.90 4 0.297 0.041 0.004 -0.090 -0.082 -0.085 0.021
12 14.75 10 0.141 0.168 0.039 0.065 0.021 -0.071 -0.089
18 22.43 16 0.130 0.000 0.110 -0.107 0.014 0.061 0.091
24 27.25 22 0.202 -0.036 0.094 -0.085 -0.009 -0.046 0.046
30 36.99 28 0.119 0.042 0.043 0.118 -0.090 0.039 -0.121
36 50.37 34 0.035 0.030 0.122 0.119 0.068 -0.145 -0.016

Model for variable X

Estimated Mean = 17.0388318

Autoregressive Factors
Factor 1: 1 - 0.90365 B**(1)

Moving Average Factors
Factor 1: 1 - 0.55797 B**(1)

```

그림 8.18.1 PROC ARIMA의 수행결과

ARIMA에 의한 예측의 결과와 비교하기 위한 준비를 하는 것이다. 즉, 이 프로시저는 단변량시계열데이터 BJ\_A에 ARIMA모형을 적용해서 예측을 하는 것이다. 이 단변량시계열데이터를 식별하기 위해서 IDENTIFY스테이트먼트의 옵션 VAR=에 변수 X를 할당하였다. 그러나 IDENTIFY스테이트먼트의 옵션 NOPRINT를 사용했으므로 이 스테이트먼트를 수행한 결과가 OUTPUT윈도우에 출력되지 않는다. 이는 단지 본서의 지면을 아끼기 위한 것이다. 이 식별단계에서 변수 X를 ARMA(1,1)모형으로 식별하였다. ESTI-

MATE스테이트먼트에서는 ARMA(1, 1)모형을 조건부최소제곱추정법으로 추정하였다. FORECAST스테이트먼트에서 옵션 LEAD=에 10을 할당하여 10개의 선시차들에 대해서 예측을 하였다. FORECAST스테이트먼트에서 옵션 NOPRINT를 사용하였으므로 이 예측의 결과가 OUTPUT윈도우에 출력되지 않는다. FORECAST스테이트먼트의 옵션 OUT=을 사용해서 이 예측의 결과를 SAS데이터세트 EXAM818B에 수록하였다. 이 프로시저를 수행한 결과가 그림 8.18.1이다. 그림 8.18.1에서 볼 수 있듯이 추정된 ARMA모형은 다음과 같다.

$$y_t = X(t) - 17.062$$

$$y_t = 0.904y_{t-1} + v_t - 0.558v_{t-1}$$

이 SAS프로그램의 두 번째 프로시저는 다음과 같다. 이 프로시저는 단변량시계열데이터 BJ\_A를 상태공간모형화하기 위한 것이다.

```
PROC STATESPACE CANCORR PRINTOUT=LONG;
VAR X;
RUN;
```

이 프로시저를 수행한 결과가 그림 8.18.2이다. PROC STATESPACE스테이트먼트의 옵션 PRINTOUT=에 LONG을 할당함으로써 디폴트인 SHORT를 할당했을 때보다 부분 D의 주어진 시계열데이터의 표본자기공분산행렬들, 부분 E의 표본자기상관행렬들, 부분 G의 표본부분자기상관행렬들, 부분 J의 각 VAR모형에 이 시계열데이터를 적합시켰을 때 백색잡음과정의 공분산행렬의 Yule-Walker추정행렬, 그리고 부분 K의 각 VAR모형에 시계열데이터를 적합시켰을 때 백색잡음과정의 상관행렬의 Yule-Walker추정행렬들이 추가적으로 출력되었다. 또한 PROC STATESPACE스테이트먼트의 옵션 CANCORR를 사용함으로써 부분 L의 DIC값들과 Bartlett통계량값들이 출력되었다.

그림 8.18.2의 각 부분을 설명하면 다음과 같다. 부분 A에서 이 SAS데이터세트의 관찰점들의 개수가 197, 부분 B에서 표본평균이 17.062, 표본표준편차가 0.399임을 알 수 있다. 부분 C에는 VAR(0)모형, VAR(1)모형, ..., VAR(10)모형에 해당하는 AIC값들이 수록되어 있다. 이 중에서 VAR(7)모형의 AIC값이 최소이다. 즉, VAR모형단계에서 VAR(7)모형을 선택한다. 부분 D에는 VAR스테이트먼트에 지정된 변수들의 표본자기공분산행렬들이 수록되어 있다. 이 예제에서는 VAR스테이트먼트에 하나의 변수가 지정되었으므로 표본자기공분산행렬들이 아닌 표본자기공분산들이 실려 있다. 부분 E에는 VAR스테이트먼트에 지정된 변수들의 표본자기상관행렬들이 수록되어 있다. 즉, 부분 D에 해당하는 표본자기상관행렬이 수록되어 있다. 이 예제에서는 VAR스테이트먼트에 하나의 변수가 지정되었으므로 표본자기상관행렬들이 아닌 표본자기상관계수들이 실려 있다. 부분 F에는 VAR스테이트먼트에 지정된 변수들의 표본자기상관행렬들의 도표적 표현(schematic representation)이 수록되어 있다. 이 도표적 표현에서 '+'는 표본자기상관계수가 표준오차의 2배보다 큰 것을, '-'는 표본자기상관계수가 표준오차의 (-2)배보다 작은 것을, '.'은 표본자기상관

STATESPACE MODELING

36

15:52 Saturday, July 24, 1993

STATESPACE Procedure

(A) Nobs = 197

(B)

| Variable | Mean     | Std      |
|----------|----------|----------|
| X        | 17.06244 | 0.399247 |

(C) Information Criterion for Autoregressive Models

| Lag=0       | Lag=1       | Lag=2       | Lag=3       | Lag=4       |
|-------------|-------------|-------------|-------------|-------------|
| -361.761017 | -437.216070 | -448.124845 | -447.046264 | -445.994023 |
| Lag=5       | Lag=6       | Lag=7       | Lag=8       | Lag=9       |
| -444.848471 | -445.884051 | -448.755121 | -446.953416 | -444.972197 |
| Lag=10      |             |             |             |             |
| -443.025084 |             |             |             |             |

(D) Var/Cov Matrices of the Input Series

| Lag=0    | Lag=1    | Lag=2    | Lag=3    | Lag=4    |
|----------|----------|----------|----------|----------|
| X        | X        | X        | X        | X        |
| 0.159398 | 0.090883 | 0.078912 | 0.063433 | 0.056697 |
| Lag=5    | Lag=6    | Lag=7    | Lag=8    | Lag=9    |
| X        | X        | X        | X        | X        |
| 0.052105 | 0.055751 | 0.062466 | 0.051247 | 0.048478 |
| Lag=10   |          |          |          |          |
| X        |          |          |          |          |
| 0.040626 |          |          |          |          |

(E) Correlation Matrices of the Input Series

| Lag=0   | Lag=1   | Lag=2   | Lag=3   | Lag=4   | Lag=5   |
|---------|---------|---------|---------|---------|---------|
| X       | X       | X       | X       | X       | X       |
| 1.00000 | 0.57016 | 0.49506 | 0.39795 | 0.35570 | 0.32688 |
| Lag=6   | Lag=7   | Lag=8   | Lag=9   | Lag=10  |         |
| X       | X       | X       | X       | X       |         |
| 0.34976 | 0.39189 | 0.32150 | 0.30413 | 0.25487 |         |

(F) Schematic Representation of Correlations

| Name/Lag | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|----------|---|---|---|---|---|---|---|---|---|---|----|
| X        | + | + | + | + | + | + | + | + | + | + | +  |

+ is > 2\*std error, - is < -2\*std error, . is between

(G) Partial Autoregressive Matrices

| Lag=1     | Lag=2     | Lag=3      | Lag=4     | Lag=5      |
|-----------|-----------|------------|-----------|------------|
| X         | X         | X          | X         | X          |
| 0.5701648 | 0.2518453 | 0.0683106  | 0.0692777 | 0.0657869  |
| Lag=6     | Lag=7     | Lag=8      | Lag=9     | Lag=10     |
| X         | X         | X          | X         | X          |
| 0.1236565 | 0.1562788 | -0.0317185 | 0.0097637 | -0.0163838 |

(H) Schematic Representation of Partial Autocorrelations

| Name/Lag | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|----------|---|---|---|---|---|---|---|---|---|----|
| X        | + | + | . | . | . | . | + | . | . | .  |

+ is > 2\*std error, - is < -2\*std error, . is between

|      |                                                    |              |         |          |          |          |
|------|----------------------------------------------------|--------------|---------|----------|----------|----------|
| (I)  | Yule-Walker Estimates for the Min AIC              |              |         |          |          |          |
|      | Lag=1                                              | Lag=2        | Lag=3   | Lag=4    | Lag=5    |          |
|      | X                                                  | X            | X       | X        | X        |          |
| X    | 0.37262                                            | 0.19682      | 0.02035 | 0.01375  | -0.01484 |          |
|      | Lag=6                                              | Lag=7        |         |          |          |          |
|      | X                                                  | X            |         |          |          |          |
| X    | 0.06240                                            | 0.15628      |         |          |          |          |
| (J)  | Variance/Covariance of the Residuals               |              |         |          |          |          |
|      | Lag=0                                              | Lag=1        | Lag=2   | Lag=3    | Lag=4    |          |
|      | X                                                  | X            | X       | X        | X        |          |
| X    | 0.15940                                            | 0.10758      | 0.10076 | 0.10029  | 0.09980  |          |
|      | Lag=5                                              | Lag=6        | Lag=7   | Lag=8    | Lag=9    |          |
|      | X                                                  | X            | X       | X        | X        |          |
| X    | 0.09937                                            | 0.09785      | 0.09546 | 0.09537  | 0.09536  |          |
|      | Lag=10                                             |              |         |          |          |          |
|      | X                                                  |              |         |          |          |          |
| X    | 0.09533                                            |              |         |          |          |          |
| (K)  | Residual Correlations from AR Models               |              |         |          |          |          |
|      | Lag=0                                              | Lag=1        | Lag=2   | Lag=3    | Lag=4    | Lag=5    |
|      | X                                                  | X            | X       | X        | X        | X        |
| X    | 1.00000                                            | 1.00000      | 1.00000 | 1.00000  | 1.00000  | 1.00000  |
|      | Lag=6                                              | Lag=7        | Lag=8   | Lag=9    | Lag=10   |          |
|      | X                                                  | X            | X       | X        | X        |          |
| X    | 1.00000                                            | 1.00000      | 1.00000 | 1.00000  | 1.00000  |          |
| (L)  | Canonical Correlations Analysis                    |              |         |          |          |          |
|      | State vector                                       | Correlations |         | Infor.   | Chisq    | D.F.     |
|      | X(T;T),X(T+1;T)                                    | 1.0000       | 0.3369  | 9.737346 | 23.31562 | 7        |
|      | X(T;T),X(T+1;T),X(T+2;T)                           | 1.0000       | 0.3699  | 0.1675   | -6.39369 | 5.520937 |
| (M)  | Selected Statespace Form and Preliminary Estimates |              |         |          |          |          |
| (M1) | State Vector                                       |              |         |          |          |          |
|      | X(T;T)                                             | X(T+1;T)     |         |          |          |          |
| (M2) | Estimate of the Transition Matrix                  |              |         |          |          |          |
|      |                                                    | 0            | 1       |          |          |          |
|      |                                                    | -0.036       | 0.932   |          |          |          |
| (M3) | Input Matrix for the Innovation                    |              |         |          |          |          |
|      |                                                    | 1            |         |          |          |          |
|      |                                                    | 0.373        |         |          |          |          |
| (M4) | Variance Matrix for the Innovation                 |              |         |          |          |          |
|      |                                                    | 0.09546355   |         |          |          |          |
| (N)  | Selected Statespace Form and Fitted Model          |              |         |          |          |          |
| (N1) | State Vector                                       |              |         |          |          |          |
|      | X(T;T)                                             | X(T+1;T)     |         |          |          |          |
| (N2) | Estimate of the Transition Matrix                  |              |         |          |          |          |
|      |                                                    | 0            | 1       |          |          |          |
|      |                                                    | -0.166       | 1.109   |          |          |          |
| (N3) | Input Matrix for the Innovation                    |              |         |          |          |          |
|      |                                                    | 1            |         |          |          |          |
|      |                                                    | 0.392        |         |          |          |          |

|      |                                    |          |           |          |
|------|------------------------------------|----------|-----------|----------|
| (N4) | Variance Matrix for the Innovation |          |           |          |
|      | 0.09866764                         |          |           |          |
| (O)  | Parameter Estimates                |          |           |          |
|      | Parameter                          | Estimate | Std. Err. | T value  |
|      | F(2,1)                             | -0.16564 | 0.117181  | -1.41352 |
|      | F(2,2)                             | 1.108647 | 0.146643  | 7.560155 |
|      | G(2,1)                             | 0.391536 | 0.069791  | 5.610135 |

그림 8.18.2 PROC STATESPACE의 수행결과

계수가 표준오차의 2배보다 작으며 표준오차의 (-2)배보다 큰 것을 나타낸다. 이러한 표본자기상관행렬의 도표적 표현은 주어진 시계열데이터의 성질을 파악하는 데 도움이 된다. 부분 ㉔에는 VAR스테이트먼트에 지정된 변수들의 표본부분자기상관행렬들이 수록되어 있다. 즉, 부분 ㉔에 해당하는 표본부분자기상관행렬이 수록되어 있다. 이 예제에서는 VAR스테이트먼트에 하나의 변수가 지정되었으므로 표본부분자기상관행렬들이 아닌 표본부분자기상관계수들이 실려 있다. 부분 ㉕에는 VAR스테이트먼트에 지정된 변수들의 표본부분자기상관행렬들의 도표적 표현이 수록되어 있다. 부분 ㉕에서와 마찬가지로 이 도표적 표현에서 ‘+’는 표본부분자기상관계수가 표준오차의 2배보다 큰 것을, ‘-’는 표본부분자기상관계수가 표준오차의 (-2)배보다 작은 것을, ‘.’은 표본부분자기상관계수가 표준오차의 2배보다 작으며 표준오차의 (-2)배보다 큰 것을 나타낸다. 이러한 표본부분자기상관행렬의 도표적 표현도 주어진 시계열데이터의 성질을 파악하는 데 도움이 된다. 부분 ㉖에는 최소AIC값을 갖는 VAR모형의 VAR계수행렬들의 Yule-Walker추정행렬들이 수록되어 있다. 이 예제에서는 VAR(7)모형에서 최소의 AIC값을 가지므로 VAR(7)모형의 VAR계수행렬들이 추정되었다. 부분 ㉖에는  $k = 0, 1, \dots, 10$  각각에 대하여 VAR( $k$ )모형에 시계열데이터를 적합시켰을 때 백색잡음과정의 공분산행렬의 Yule-Walker추정행렬이 수록되어 있다. 부분 ㉗에는 VAR( $k$ )모형에 시계열데이터를 적합시켰을 때 백색잡음과정의 상관행렬의 Yule-Walker추정행렬이 수록되었다. 부분 ㉘는 PROC STATESPACE스테이트먼트의 옵션 CANCORR를 사용함으로써 출력된 것이다. 부분 ㉘에는 상태벡터(State vector)들, 표본정준상관계수들(Correlations), AIC값(Infor.)들, Bartlett통계량값(Chisq)들, 그리고 해당되는 자유도(D.F.)가 출력되었다. 부분 ㉙은 예비추정단계의 결과물이다. 부분 ㉚는 예비추정단계의 상태벡터이다. 부분 ㉛2는 변환행렬  $F$ 의 예비추정행렬이다. 이 추정행렬은 제 8.2.3소절의 정리 8.6다음에 설명했듯이 정준상관분석을 사용해서 얻는다. 부분 ㉛3는 입력행렬  $G$ 의 예비추정행렬이다. 이 추정행렬은 다음 성질을 이용해서 VAR모형단계에서 얻는다. 입력행렬  $G$ 의 제( $r, s$ )원소를  $G(r, s)$ 라 하고  $v_{n,t}$ 를  $v_t$ 의 제  $n$ 번째 원소라 하자. 만일 상태벡터의 제  $i$ 번째 원소가  $y_{i,t+k|t}$  라면  $G(i, n)$ 은 시차  $k$ 에서  $v_{n,t}$ 에 대한  $y_{i,t}$ 의 충격반응값으로부터 얻어진다. 이 충격반응값들의 추정값들은 VAR모형의 Yule-Walker추정모형으로부터 얻는다. 부분 ㉛4는 쇠신과정의 공분산행렬의 예비추정행렬이다. 이 추정행렬은 VAR모형단계에서 백색잡음과정의 공분산행렬의 Yule-Walker추정행렬들 중에서

최소AIC값을 갖는 VAR모형에 해당하는 것이다. 부분 N은 최종추정단계의 결과이다. 부분 NI은 최종추정단계의 상태벡터이다. 부분 N2는 변환행렬  $F$ 의 최종추정행렬이고, 부분 N3는 입력행렬  $G$ 의 최종추정행렬이며, 부분 N4는 쇄신과정의 공분산행렬의 최종추정행렬이다. 이 추정행렬들은 제 8.2.4 소절에서 설명한 최우추정법에 의해서 구한다. 부분 O는 부분 N의 행렬들  $F$ 와  $G$ 의 원소들 중에서 구조적 원소들이나 RESTRICT스테이트먼트에 의해서 사전에 할당된 원소들이 아니고 최종상태공간모형에서 추정된 원소들의 최종추정값들, 표준오차들, 그리고 이에 해당하는  $t$ 값들이 수록되어 있다.

상태공간모형의 최종추정식은 다음과 같다.

$$y_t = X(t) - 17.062$$

$$z_{t+1} = Fz_t + Gv_{t+1}$$

$$yz_t = [1 \ 0] z_t$$

$$\text{단, } z_t = (y_t, y_{t+1|t})'$$

$$\hat{F} = \begin{bmatrix} 0 & 1 \\ -0.036 & 0.932 \end{bmatrix}, \quad \hat{G} = \begin{bmatrix} 1 \\ 0.373 \end{bmatrix}, \quad \hat{\Sigma} = [0.0955]$$

이 상태공간모형을 전개하면 다음과 같다.

$$y_{t+1} = y_t + v_{t+1}$$

$$y_{t+2|t} = -0.166y_t + 1.109y_{t+1|t} + 0.392v_{t+1}$$

첫 번째 식을 두 번째 식에 대입하면 다음 식을 얻는다.

$$y_t = 1.109y_{t-1} - 0.166y_{t-2} + v_t - 0.717v_{t-1}$$

따라서  $\{y_t\}$ 는 ARMA(2, 1)모형을 이룬다.

이 SAS프로그램의 세 번째 프로시저는 다음과 같다. 이 프로시저는 단변량시계열데이터 BJ\_A를 상태공간모형화하고 이 모형을 사용해서 예측을 하기 위한 것이다.

```
PROC STATESPACE RESIDEST LEAD=10
 OUT=EXAM818C OUTMODEL=MODEL1;
VAR X;
RESTRICT F(2,1)=0;
RUN;
```

이 프로시저에서는 PROC STATESPACE스테이트먼트의 옵션 LEAD=에 10을 할당함으로써 열 개의 선시차들에 대해서 예측을 한다. 또한 PROC STATESPACE스테이트먼트의 옵션 OUT=에 EXAM818C를 할당함으로써 예측의 결과를 SAS데이터세트 EXAM818C에 출력한다. PROC STATESPACE스테이트먼트의 옵션 OUT=에 지정한 SAS데이터세트의 내용은 제 8.2.3 소절에서 간단히 소개하였으나 여기서 좀더 자세히 설명하면 다음과 같다.



- ① 변수 OBS는 관찰점들의 순서를 나타낸다.
- ② 다음 열에는 VAR스테이트먼트에 지정한 순서대로 변수 중의 하나가 나타난다. 이 변수를 제 $x$ 번째 변수라고 하자.
- ③ 다음 열의 변수 FOR $x$ 는 제 $x$ 번째 변수의 예측이다. 시계열데이터의 마지막 관찰점으로부터 옵션 BACK=에 지정한 숫자만큼의 역순의 관찰점으로부터 전방일단계 예측을 한다.
- ④ 다음 열의 변수 RES $x$ 는 제 $x$ 번째 변수의 잔차이다.
- ⑤ 다음 열의 변수 STD $x$ 는 제 $x$ 번째 변수의 표본표준편차이다. 이 열 다음에는 제3단계로 돌아가 새로운 변수 FOR $x$ 가 나타난다.
- ⑥ 만일 BY변수가 존재하면 이 변수가 출력된다.
- ⑦ 만일 ID변수가 존재하면 이 변수가 출력된다.

| STATESPACE Procedure |                                                    |          |           |
|----------------------|----------------------------------------------------|----------|-----------|
| (A)                  | Selected Statespace Form and Preliminary Estimates |          |           |
| (A1)                 | State Vector                                       |          |           |
|                      | X(T;T)                                             | X(T+1;T) |           |
| (A2)                 | Estimate of the Transition Matrix                  |          |           |
|                      | 0                                                  | 1        |           |
|                      | 0.000                                              | 0.932    |           |
| (A3)                 | Input Matrix for the Innovation                    |          |           |
|                      |                                                    | 1        |           |
|                      |                                                    | 0.373    |           |
| (A4)                 | Variance Matrix for the Innovation                 |          |           |
|                      | 0.09546355                                         |          |           |
| (B)                  | Selected Statespace Form and Fitted Model          |          |           |
| (B1)                 | State Vector                                       |          |           |
|                      | X(T;T)                                             | X(T+1;T) |           |
| (B2)                 | Estimate of the Transition Matrix                  |          |           |
|                      | 0                                                  | 1        |           |
|                      | 0.000                                              | 0.908    |           |
| (B3)                 | Input Matrix for the Innovation                    |          |           |
|                      |                                                    | 1        |           |
|                      |                                                    | 0.328    |           |
| (B4)                 | Variance Matrix for the Innovation                 |          |           |
|                      | 0.09835728                                         |          |           |
| (C)                  | Parameter Estimates                                |          |           |
|                      | Parameter                                          | Estimate | Std. Err. |
|                      | F(2,2)                                             | 0.908008 | 0.044666  |
|                      | G(2,1)                                             | 0.328294 | 0.060494  |
|                      |                                                    |          | T value   |
|                      |                                                    |          | 20.32874  |
|                      |                                                    |          | 5.426915  |

그림 8.18.3 RESTRICT스테이트먼트를 사용한 상태공간모형화

또한 이 프로시저에서는 PROC STATESPACE 스테이트먼트의 옵션 RESIDEST를 사용하여 최종추정단계에서 최우추정법이 아닌 조건부최소제곱추정법을 사용한다. 또한 PROC STATESPACE 스테이트먼트의 옵션 OUTMODEL=에 SAS 데이터세트 MODEL1을 지정했으므로 최종적으로 추정된 상태공간모형이 이 SAS 데이터세트에 저장된다. 또한 RESTRICT 스테이트먼트를 사용해서 변환행렬  $F$ 의 원소  $F(2, 1)$ 을 0으로 고정시켰다. 이 프로시저를 수행한 결과 중에서 그림 8.18.2에 나타나지 않은 부분이 그림 8.18.3이다.

그림 8.18.3의 각 부분을 설명하면 다음과 같다. 부분 A는 예비추정단계의 결과이다. 부분 A1은 예비추정단계의 상태벡터, 부분 A2는 변환행렬  $F$ 의 예비추정행렬, 부분 A3는 입력행렬  $G$ 의 예비추정행렬, 그리고 부분 A4는 쇄신과정의 공분산행렬의 예비추정행렬이다. 부분 B는 최종추정단계의 결과이다. 여기서 유의할 점은 부분 A3의 원소  $F(2, 1)$ 이 0.000인 것을 제외하면 부분 A는 그림 8.18.2의 부분 M과 동일하다. 부분 B1은 최종추정단계의 상태벡터, 부분 B2는 변환행렬  $F$ 의 최종추정행렬이고, 부분 B3는 입력행렬  $G$ 의 최종추정행렬, 그리고 부분 B4는 쇄신과정의 공분산행렬의 최종추정행렬이다. 이 추정행렬들은 제 8.3.4 소절에서 설명한 조건부최소제곱추정법에 의한 것이다. 부분 C에는 부분 B의 행렬들  $F$ 와  $G$ 의 원소들 중에서 구조적 원소들이나 RESTRICT 스테이트먼트에 의해 정해지지 않고 추정된 원소들의 최종추정값들, 표준오차들, 그리고 이에 해당하는  $t$ 통계량값들이 수록되어 있다.

상태공간모형의 최종추정식은 다음과 같다.

$$y_t = X(t) - 17.062$$

$$z_{t+1} = \hat{F}z_t + \hat{G}v_{t+1}$$

$$y_t = [1 \ 0] z_t$$

$$\text{단, } z_t = (y_t, y_{t+1|t})'$$

$$\hat{F} = \begin{bmatrix} 0 & 1 \\ 0.000 & 0.908 \end{bmatrix}, \quad \hat{G} = \begin{bmatrix} 1 \\ 0.328 \end{bmatrix}, \quad \hat{\Sigma} = [0.0984]$$

이 상태공간모형을 전개하면 다음과 같다.

$$y_{t+1} = y_{t+1|t} + v_{t+1}$$

$$y_{t+2|t+1} = 0.908y_{t+1|t} + 0.328v_{t+1}$$

첫 번째 식을 두 번째 식에 대입하면 다음 식을 얻는다.

$$y_t = 0.908y_{t-1} + v_t - 0.580v_{t-1}$$

따라서 이 단변량시계열데이터는 ARMA(1, 1)모형에서 나왔다고 할 수 있다. 이 ARMA(1, 1)모형은 그림 8.18.2의 ARMA(2, 1)모형과 별로 다르지 않다는 것을 다음 근사식에서 알 수 있다.

$$1 - 1.109z + 0.166z^2 \approx 1 - 0.908z \quad (z \approx 0)$$

| MODEL FROM PROC STATESPACE |          |     |         |     |     |     |     |     |     |     |      |         | 46                            |
|----------------------------|----------|-----|---------|-----|-----|-----|-----|-----|-----|-----|------|---------|-------------------------------|
|                            |          |     |         |     |     |     |     |     |     |     |      |         | 15:52 Saturday, July 24, 1995 |
| OBS                        | STATEVEC | F_1 | F_2     | F_3 | F_4 | F_5 | F_6 | F_7 | F_8 | F_9 | F_10 | G_1     | SIG_1                         |
| 1                          | X(T;T)   | 0   | 1.00000 | .   | .   | .   | .   | .   | .   | .   | .    | 1.00000 | 0.098357                      |
| 2                          | Std      | .   | .       | .   | .   | .   | .   | .   | .   | .   | .    | .       | .                             |
| 3                          | X(T+1;T) | 0   | 0.90801 | .   | .   | .   | .   | .   | .   | .   | .    | 0.32829 | .                             |
| 4                          | Std      | .   | 0.04467 | .   | .   | .   | .   | .   | .   | .   | .    | 0.06049 | .                             |

그림 8.18.4 옵션 OUTMODEL=에 저장된 결과물

이 SAS프로그램의 네 번째 프로시저는 다음과 같다. 이 프로시저는 두 번째 PROC STATESPACE의 옵션 OUTMODEL=에 지정한 SAS데이터세트를 OUTPUT윈도우에 출력시키기 위한 것이다.

```
PROC PRINT DATA=MODEL1;
RUN;
```

이 프로시저를 수행한 결과가 그림 8.18.4이다. 그림 8.18.4의 각 부분을 설명하면 다음과 같다.

- ① 변수 STATEVEC에는 상태공간벡터의 원소들의 이름들과 이들의 표본표준편차를 의미하는 Std가 관찰값들로 수록되어 있다.
- ② 변수 F\_x는 변환행렬 F의 추정행렬의 제x번째 행이 수록되어 있다. 이 변수들의 개수는 옵션 DIMMAX=에 지정된 값이다. 상태벡터의 차원이 이보다 작으면 남은 변수들에는 결측(missing)으로 기록된다. 이 예제에서는 변수들 F\_3, F\_4, ..., F\_10이 결측으로 기록되었다.
- ③ 변수 G\_x는 입력행렬 G의 추정행렬 제x번째 행이 수록되어 있다. 이 변수들의 개수는 다변량시계열데이터 {y\_t}의 차원 d이다. 이 예제에서는 d가 1이므로 변수 G\_1만이 출력되었다.
- ④ 변수 SIG\_x는 백색잡음과정의 공분산행렬의 추정행렬의 제x번째 행이 수록되어 있다. 이 변수들의 개수는 다변량시계열데이터 {y\_t}의 차원 d이다. 이 예제에서는 d가 1이므로 변수 SIG\_1만이 출력되었다.
- ⑤ 만일 BY변수가 존재하면 이 변수가 출력된다.

이 SAS프로그램의 두 번째 DATA스텝과 다섯 번째 PROC스텝은 다음과 같다.

```
DATA:
MERGE EXAM818B EXAM818C;
RUN;
PROC PRINT DATA = _LAST _ (FIRSTOBS = 180);
RUN;
```

이 DATA스텝은 PROC ARIMA의 FORECAST스테이트먼트의 옵션 OUT=에 지정한

| COMPARING FORECASTS           |      |          |         |         |         |          |         |          |         |
|-------------------------------|------|----------|---------|---------|---------|----------|---------|----------|---------|
| 47                            |      |          |         |         |         |          |         |          |         |
| 15:52 Saturday, July 24, 1993 |      |          |         |         |         |          |         |          |         |
| OBS                           | X    | FORECAST | STD     | L95     | U95     | RESIDUAL | FOR1    | RES1     | STD1    |
| 180                           | 17.4 | 17.1276  | 0.31567 | 16.5089 | 17.7463 | 0.27241  | 17.1351 | 0.26489  | 0.31362 |
| 181                           | 17.5 | 17.2132  | 0.31567 | 16.5945 | 17.8319 | 0.28680  | 17.2154 | 0.28462  | 0.31362 |
| 182                           | 17.9 | 17.2955  | 0.31567 | 16.6768 | 17.9143 | 0.60446  | 17.2948 | 0.60525  | 0.31362 |
| 183                           | 17.0 | 17.4798  | 0.31567 | 16.8611 | 18.0985 | -0.47976 | 17.4721 | -0.47208 | 0.31362 |
| 184                           | 17.0 | 17.2714  | 0.31567 | 16.6527 | 17.8901 | -0.27143 | 17.2794 | -0.27942 | 0.31362 |
| 185                           | 17.0 | 17.1552  | 0.31567 | 16.5365 | 17.7739 | -0.15519 | 17.1677 | -0.16772 | 0.31362 |
| 186                           | 17.2 | 17.0903  | 0.31567 | 16.4716 | 17.7090 | 0.10967  | 17.1030 | 0.09702  | 0.31362 |
| 187                           | 17.3 | 17.1233  | 0.31567 | 16.5046 | 17.7420 | 0.17672  | 17.1311 | 0.16890  | 0.31362 |
| 188                           | 17.4 | 17.1762  | 0.31567 | 16.5575 | 17.7949 | 0.22377  | 17.1802 | 0.21977  | 0.31362 |
| 189                           | 17.4 | 17.2403  | 0.31567 | 16.6216 | 17.8591 | 0.15965  | 17.2415 | 0.15846  | 0.31362 |
| 190                           | 17.0 | 17.2761  | 0.31567 | 16.6574 | 17.8948 | -0.27612 | 17.2771 | -0.27709 | 0.31362 |
| 191                           | 18.0 | 17.1578  | 0.31567 | 16.5391 | 17.7765 | 0.84219  | 17.1664 | 0.83362  | 0.31362 |
| 192                           | 18.2 | 17.4375  | 0.31567 | 16.8188 | 18.0562 | 0.76252  | 17.4305 | 0.76951  | 0.31362 |
| 193                           | 17.6 | 17.6627  | 0.31567 | 17.0440 | 18.2814 | -0.06266 | 17.6493 | -0.04926 | 0.31362 |
| 194                           | 17.8 | 17.5809  | 0.31567 | 16.9622 | 18.1996 | 0.21911  | 17.5791 | 0.22090  | 0.31362 |
| 195                           | 17.7 | 17.6044  | 0.31567 | 16.9857 | 18.2231 | 0.09559  | 17.6041 | 0.09591  | 0.31362 |
| 196                           | 17.2 | 17.5830  | 0.31567 | 16.9642 | 18.2017 | -0.38296 | 17.5858 | -0.38575 | 0.31362 |
| 197                           | 17.4 | 17.3981  | 0.31567 | 16.7794 | 18.0169 | 0.00185  | 17.4110 | -0.01097 | 0.31362 |
| 198                           | .    | 17.3642  | 0.31567 | 16.7455 | 17.9829 | .        | 17.3753 | .        | 0.31362 |
| 199                           | .    | 17.3328  | 0.33400 | 16.6782 | 17.9875 | .        | 17.3465 | .        | 0.33009 |
| 200                           | .    | 17.3045  | 0.34826 | 16.6219 | 17.9871 | .        | 17.3204 | .        | 0.34307 |
| 201                           | .    | 17.2789  | 0.35948 | 16.5743 | 17.9835 | .        | 17.2967 | .        | 0.35342 |
| 202                           | .    | 17.2558  | 0.36838 | 16.5338 | 17.9778 | .        | 17.2751 | .        | 0.36173 |
| 203                           | .    | 17.2349  | 0.37550 | 16.4989 | 17.9708 | .        | 17.2556 | .        | 0.36843 |
| 204                           | .    | 17.2160  | 0.38122 | 16.4688 | 17.9631 | .        | 17.2378 | .        | 0.37387 |
| 205                           | .    | 17.1989  | 0.38582 | 16.4427 | 17.9551 | .        | 17.2217 | .        | 0.37830 |
| 206                           | .    | 17.1835  | 0.38954 | 16.4200 | 17.9470 | .        | 17.2070 | .        | 0.38191 |
| 207                           | .    | 17.1696  | 0.39255 | 16.4002 | 17.9389 | .        | 17.1937 | .        | 0.38486 |

그림 8.18.5 PROC ARIMA와 PROC STATESPACE에 의한 예측들의 비교

SAS데이터세트 EXAM818B와 두 번째 PROC STATESPACE의 PROC STATESPACE 스테이트먼트의 옵션 OUT=에 지정한 SAS데이터세트 EXAM818C를 가로결합시키는 것이다. PROC PRINT는 이렇게 만들어진 SAS데이터세트의 제 180번째 관찰점부터 OUTPUT윈도우에 출력시키기 위한 것이다. 이 PROC PRINT를 수행한 결과가 그림 8.18.5이다. 그림 8.18.5의 각 부분을 설명하면 다음과 같다.

- ① 변수 OBS는 관찰점의 순서를 나타낸다.
- ② 변수 X는 분석대상의 시계열데이터를 나타낸다.
- ③ 변수들 FORECAST, STD, L95, U95, 그리고 RESIDUAL은 PROC ARIMA의 FORECAST스테이트먼트에 의한 것으로 각각 예측, 표준오차, 신뢰도 95%의 예측구간의 상한과 하한, 그리고 잔차를 나타낸다.
- ④ 변수들 FOR1, RES1, 그리고 STD1은 PROC STATESPACE에 의한 것으로 각각 변수 X에 대한 예측, 잔차, 그리고 표본표준편차를 나타낸다.

그림 8.18.5에서 볼 수 있듯이 이 예제에서는 PROC ARIMA에 의한 예측과 PROC STATESPACE에 의한 예측이 큰 차이가 없음을 알 수 있다.

## 8.4.3 다변량시계열데이터 BJ\_J

이 소절에서는 Box & Jenkins(1976, pp. 532~533)의 Series J인 2변량시계열데이터 BJ\_J를 상태공간모형화하기로 한다. 이 2변량시계열데이터는 296개의 관찰점들로 구성되어 있으며, 변수 X는 입력개스비율(input gas rate)을 나타내고 변수 Y는 출력개스내의 이산화탄소의 퍼센트(CO<sub>2</sub> percent in output gas)를 나타낸다. 이 다변량시계열데이터는 제 2.2절에 수록되어 있으며 이 다변량시계열데이터의 시계열산포도가 그림 2.2에 수록되어 있다. 또한 이 다변량시계열데이터는 디스켓 SAS4TSA4의 서브디렉토리 TSA4DAT-A에 영구SAS데이터세트 BJJ.SSD로 저장되어 있다.

다변량시계열데이터 BJ\_J를 상태공간모형화하기 위한 준비단계의 하나로 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM8-19.PRG에 수록되어 있다).

```

/* */
/* EXAM8-19.PRG */
/* */
/* STATE SPACE MODELING */
/* DATASET = BJ_J */
LIBNAME TSA4DATA 'A:\TSA4DATA';
/* EXAM8-19.PRG */
/* */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=60;
/* */
/* FIT STATESPACE MODELS */
/* TO GAS FURNACE DATA */
/* */
DATA EXAM819;
 SET TSA4DATA.BJJ;
 RUN;
TITLE 'STATESPACE MODELING';
PROC STATESPACE PRINTOUT=NONE;
 VAR X Y;
 RUN;
PROC STATESPACE PRINTOUT=NONE;
 VAR X Y;

```

```

RESTRICT G(3,2)=0;
RUN;
PROC STATESPACE PRINTOUT=NONE;
VAR X Y;
RESTRICT G(3,2)=0 G(4,1)=0;
RUN;
PROC STATESPACE PRINTOUT=NONE;
VAR X Y;
RESTRICT G(3,2)=0 G(4,1)=0 F(3,1)=0;
RUN;
PROC STATESPACE PRINTOUT=NONE;
VAR X Y;
RESTRICT G(3,2)=0 G(4,1)=0 F(3,1)=0 F(3,2)=0;
RUN;
PROC STATESPACE PRINTOUT=NONE;
VAR X Y;
RESTRICT G(3,2)=0 G(4,1)=0 F(3,1)=0 F(3,2)=0
 F(3,3)=0;
RUN;
PROC STATESPACE PRINTOUT=NONE;
VAR X Y;
RESTRICT G(3,2)=0 G(4,1)=0 F(3,1)=0 F(3,2)=0
 F(3,3)=0 F(3,4)=0;
RUN;

```

이 SAS프로그램을 수행한 결과의 일부가 그림 8.19이다. 첫 번째 PROC스텝을 수행한 결과의 일부가 그림 8.19의 부분 A와 부분 B이다. 부분 A는 최종추정단계에서의 상태벡터이다. 부분 B에서 모수  $G(3, 2)$ 의 최종추정값이 유의적으로 0과 다르지 않으며,  $t$ 통계량값들의 절대값들 중에서 모수  $G(3, 2)$ 의  $t$ 통계량값의 절대값이 가장 작은 것을 알 수 있다. 따라서 두 번째 PROC STATESPACE의 RESTRICT스테이트먼트에 ' $G(3, 2)=0$ '을 지정하였다. 이 PROC스텝을 수행한 결과 중에서 최종추정단계의 결과물이 그림 8.19의 부분 C이다. 부분 C에서 모수  $G(4, 1)$ 의 최종추정값이 유의적으로 0과 다르지 않으며  $t$ 통계량값들의 절대값들 중에서 모수  $G(4, 1)$ 의  $t$ 통계량값의 절대값이 가장 작은 것을 알 수 있다. 따라서 세 번째 PROC STATESPACE의 RESTRICT스테이트먼트에 ' $G(4, 1)=0$ '을 지정하였다. 이 PROC스텝을 수행한 결과 중에서 최종추정단계의 결과물이 그림 8.19의 부분 D이다. 부분 D에서 모수  $F(3, 1)$ 의 최종추정값이 유의적으로 0과 다르지 않으며  $t$ 통계량값들의 절대값들 중에서 모수  $F(3, 1)$ 의  $t$ 통계량값의 절대값이 가장 작은 것을 알 수 있다. 모수  $F(3, 1)$ 의 추정값이 유의적으로 0과 다르지 않음을 알 수 있다. 따라서 네 번째 PROC STATESPACE의 RESTRICT스테이트먼트에 ' $F(3, 1)=0$ '을 지정하였다. 이 PROC스텝

STATESPACE MODELING

1

13:57 Monday, August 16, 1993

STATESPACE Procedure

(A)

Selected Statespace Form and Fitted Model

State Vector

X(T;T)                      Y(T;T)                      X(T+1;T)                      Y(T+1;T)  
 Y(T+2;T)

(B)

Variance Matrix for the Innovation

0.03557876                      -0.007285  
 -0.007285                      0.09557739

Parameter Estimates

| Parameter | Estimate | Std. Err. | T value  |
|-----------|----------|-----------|----------|
| F(3,1)    | -0.86192 | 0.072961  | -11.8135 |
| F(3,2)    | 0.030609 | 0.026167  | 1.16975  |
| F(3,3)    | 1.724235 | 0.061599  | 27.99136 |
| F(3,4)    | -0.05483 | 0.030169  | -1.81726 |
| F(5,1)    | -0.34839 | 0.135253  | -2.57582 |
| F(5,2)    | 0.292124 | 0.046299  | 6.309469 |
| F(5,3)    | -0.09435 | 0.096527  | -0.9774  |
| F(5,4)    | -1.09823 | 0.109525  | -10.0272 |
| F(5,5)    | 1.671418 | 0.083737  | 19.9604  |
| G(3,1)    | 1.92442  | 0.058162  | 33.08744 |
| G(3,2)    | -0.00416 | 0.035255  | -0.11812 |
| G(4,1)    | 0.015621 | 0.095771  | 0.163107 |
| G(4,2)    | 1.258495 | 0.055742  | 22.57733 |
| G(5,1)    | 0.08058  | 0.151622  | 0.531452 |
| G(5,2)    | 1.353204 | 0.091388  | 14.80719 |

(C)

Variance Matrix for the Innovation

0.03557878                      -0.0072856  
 -0.0072856                      0.09558192

Parameter Estimates

| Parameter | Estimate | Std. Err. | T value  |
|-----------|----------|-----------|----------|
| F(3,1)    | -0.86225 | 0.073018  | -11.8088 |
| F(3,2)    | 0.030625 | 0.026165  | 1.170445 |
| F(3,3)    | 1.724564 | 0.061654  | 27.9716  |
| F(3,4)    | -0.05484 | 0.030167  | -1.818   |
| F(5,1)    | -0.35236 | 0.135877  | -2.59324 |
| F(5,2)    | 0.290997 | 0.046362  | 6.27669  |
| F(5,3)    | -0.0918  | 0.096935  | -0.94698 |
| F(5,4)    | -1.09489 | 0.109742  | -9.97692 |
| F(5,5)    | 1.668703 | 0.083984  | 19.86934 |
| G(3,1)    | 1.924843 | 0.057739  | 33.33707 |
| G(4,1)    | 0.015223 | 0.095751  | 0.158986 |
| G(4,2)    | 1.256743 | 0.05508   | 22.81671 |
| G(5,1)    | 0.079889 | 0.151532  | 0.527207 |
| G(5,2)    | 1.350866 | 0.090916  | 14.85833 |

(D)

Variance Matrix for the Innovation

0.03557464                      -0.007285  
 -0.007285                      0.09560056

Parameter Estimates

| Parameter | Estimate | Std. Err. | T value  |
|-----------|----------|-----------|----------|
| F(3,1)    | -0.00168 | 0.077162  | -0.02183 |
| F(3,2)    | 0.030851 | 0.027882  | 1.106497 |
| F(3,3)    | 1.724639 | 0.064672  | 26.66757 |
| F(3,4)    | -0.0551  | 0.032225  | -1.70993 |
| F(5,1)    | -0.35242 | 0.240225  | -1.46704 |
| F(5,2)    | 0.289241 | 0.098965  | 2.92265  |
| F(5,3)    | -0.09309 | 0.147112  | -0.63279 |
| F(5,4)    | -1.09059 | 0.263284  | -4.14224 |
| F(5,5)    | 1.665801 | 0.200215  | 8.320041 |
| G(3,1)    | 1.926006 | 0.057377  | 33.56728 |
| G(4,2)    | 1.2552   | 0.056204  | 22.33306 |
| G(5,1)    | 0.059541 | 0.088841  | 0.670194 |
| G(5,2)    | 1.349563 | 0.091152  | 14.80568 |

(E)

Variance Matrix for the Innovation

|            |            |
|------------|------------|
| 0.04556675 | -0.0081029 |
| -0.0081029 | 0.09615995 |

Parameter Estimates

| Parameter | Estimate | Std. Err. | T value  |
|-----------|----------|-----------|----------|
| F(3,2)    | -0.00036 | 0.032134  | -0.01133 |
| F(3,3)    | 1.031448 | 0.024644  | 41.85365 |
| F(3,4)    | 0.218426 | 0.033886  | 6.445898 |
| F(5,1)    | -0.35048 | 0.258649  | -1.35503 |
| F(5,2)    | 0.205362 | 0.083658  | 2.454783 |
| F(5,3)    | -0.16341 | 0.191439  | -0.85361 |
| F(5,4)    | -0.87377 | 0.201354  | -4.33944 |
| F(5,5)    | 1.515873 | 0.14918   | 10.16138 |
| G(3,1)    | 1.690075 | 0.048208  | 35.05831 |
| G(4,2)    | 1.209341 | 0.052233  | 23.15304 |
| G(5,1)    | 0.071251 | 0.089306  | 0.79783  |
| G(5,2)    | 1.274005 | 0.082605  | 15.42283 |

(F)

Variance Matrix for the Innovation

|            |            |
|------------|------------|
| 0.0508894  | -0.0091991 |
| -0.0091991 | 0.09570294 |

Parameter Estimates

| Parameter | Estimate | Std. Err. | T value  |
|-----------|----------|-----------|----------|
| F(3,3)    | 0.001889 | 0.023823  | 0.079277 |
| F(3,4)    | 0.026227 | 0.007943  | 3.30205  |
| F(5,1)    | -0.21836 | 0.301109  | -0.7252  |
| F(5,2)    | 0.278593 | 0.091636  | 3.040208 |
| F(5,3)    | -0.25203 | 0.229298  | -1.09914 |
| F(5,4)    | -1.04608 | 0.225979  | -4.62911 |
| F(5,5)    | 1.629707 | 0.166314  | 9.798971 |
| G(3,1)    | 1.692025 | 0.042675  | 39.64948 |
| G(4,2)    | 1.253884 | 0.055512  | 22.58756 |
| G(5,1)    | 0.149804 | 0.093478  | 1.602561 |
| G(5,2)    | 1.334305 | 0.089985  | 14.82811 |

(G)

Variance Matrix for the Innovation

|           |            |
|-----------|------------|
| 0.6413151 | -0.186015  |
| -0.186015 | 0.15744988 |

Parameter Estimates

| Parameter | Estimate | Std. Err. | T value  |
|-----------|----------|-----------|----------|
| F(3,4)    | 0.000085 | 0.019046  | 0.004445 |
| F(5,1)    | -0.39168 | 0.107332  | -3.64923 |



|        |          |          |          |
|--------|----------|----------|----------|
| F(5,2) | 0.091949 | 0.120212 | 0.764885 |
| F(5,3) | -0.29513 | 0.108586 | -2.71791 |
| F(5,4) | -0.68301 | 0.268476 | -2.54402 |
| F(5,5) | 1.428819 | 0.176449 | 8.097624 |
| G(3,1) | 0.982796 | 0.025303 | 38.84126 |
| G(4,2) | 1.340337 | 0.057364 | 23.36546 |
| G(5,1) | 0.293572 | 0.063974 | 4.5889   |
| G(5,2) | 1.468016 | 0.099007 | 14.82742 |

(H)

Variance Matrix for the Innovation

|           |           |
|-----------|-----------|
| 177115564 | -11663831 |
| -11663831 | 768114.2  |

Parameter Estimates

| Parameter | Estimate | Std. Err. | T value  |
|-----------|----------|-----------|----------|
| F(5,1)    | -0.00039 | 2.398E-6  | -161.162 |
| F(5,2)    | 0.334274 | 1.643E-6  | 999      |
| F(5,3)    | -0.18174 | 1.962E-9  | -999     |
| F(5,4)    | -1.23557 | 2.44E-9   | -999     |
| F(5,5)    | 1.787475 | 4.759E-8  | 999      |
| G(3,1)    | 1.925887 | .         | .        |
| G(4,2)    | 1.299793 | .         | .        |
| G(5,1)    | 0.142421 | .         | .        |
| G(5,2)    | 1.361696 | .         | .        |

그림 8.19 RESTRICT스테이트먼트를 사용한 여러 상태공간모형들

을 수행한 결과 중에서 최종추정단계의 결과물이 그림 8.19의 부분 H이다. 부분 H에서 모수  $F(3, 2)$ 의 최종추정값이 유의적으로 0과 다르지 않으며  $t$ 통계량값들의 절대값들 중에서 모수  $F(3, 2)$ 의  $t$ 통계량값의 절대값이 가장 작은 것을 알 수 있다. 따라서 다섯 번째 PROC STATESPACE의 RESTRICT스테이트먼트에 ' $F(3, 2)=0$ '을 지정하였다. 이 PROC스텝을 수행한 결과 중에서 최종추정단계의 결과물이 그림 8.19의 부분 I이다. 부분 I에서 모수  $F(3, 3)$ 의 최종추정값이 유의적으로 0과 다르지 않으며  $t$ 통계량값들의 절대값들 중에서 모수  $F(3, 3)$ 의  $t$ 통계량값의 절대값이 가장 작은 것을 알 수 있다. 따라서 여섯 번째 PROC STATESPACE의 RESTRICT스테이트먼트에 ' $F(3, 3)=0$ '을 지정하였다. 이 PROC스텝을 수행한 결과 중에서 최종추정단계의 결과물이 그림 8.19의 부분 G이다. 부분 G에서 모수  $F(3, 4)$ 의 최종추정값이 유의적으로 0과 다르지 않으며  $t$ 통계량값들의 절대값들 중에서 모수  $F(3, 4)$ 의  $t$ 통계량값의 절대값이 가장 작은 것을 알 수 있다. 따라서 일곱 번째 PROC STATESPACE의 RESTRICT스테이트먼트에 ' $F(3, 4)=0$ '을 지정하였다. 이 PROC스텝을 수행한 결과 중에서 최종추정단계의 결과물이 그림 8.19의 부분 H이다. 부분 H에서 모든 모수들의 추정값들이 유의적으로 0과 다르다. 따라서 여기서 후향제거방법을 멈추도록 한다.

좀더 객관적으로 상태공간모형을 식별하기 위해서 벌칙함수식별방법들을 적용하기로 하자. 이러한 목적으로 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4 TSA4의 파일 EXAM8-20.PRG에 수록되어 있다).

```

/*
/*
EXAM8-20.PRG
*/
*/

```

```

/ * CHOOSING RESTRICT CONDITIONS */
/ * USING THE PENALTY FUNTION METHODS */
OPTIONS LINESIZE=72 PAGESIZE=60;
PROC IML;
RESET NOLOG NOCENTER NONAME;
T = 296;
K = 15;
/ * WITHOUT ANY RESTRICTION */
COV0 = { 0.03557876 -0.007285,
 -0.007285 0.09557739 };
NORES = 0;
AIC = LOG(DET(COV0)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COV0)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COV0)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT , "NO OF RESTRICTION = " NORES,
 " AIC = " AIC,
 " BIC = " BIC,
 " HQC = " HQC;
/ * WITH 1 RESTRICTION */
COV1 = { 0.03557878 -0.0072856,
 -0.0072856 0.09558192 };
NORES = 1;
AIC = LOG(DET(COV1)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COV1)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COV1)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT , "NO OF RESTRICTION = " NORES,
 " AIC = " AIC,
 " BIC = " BIC,
 " HQC = " HQC;
/ * WITH 2 RESTRICTIONS */
COV2 = { 0.03557464 -0.007285,
 -0.007285 0.09560056 };
NORES = 2;
AIC = LOG(DET(COV2)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COV2)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COV2)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT , "NO OF RESTRICTION = " NORES,
 " AIC = " AIC,
 " BIC = " BIC,

```

```

 " HQC = " HQC;
/* WITH 3 RESTRICTIONS */
COV3 = { 0.04556675 -0.0081029,
 -0.0081029 0.09615995 };
NORES = 3;
AIC = LOG(DET(COV3)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COV3)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COV3)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT, "NO OF RESTRICTION = " NORES,
 " AIC = " AIC,
 " BIC = " BIC,
 " HQC = " HQC;
/* WITH 4 RESTRICTIONS */
COV4 = { 0.0508894 -0.0091991,
 -0.0091991 0.09570294 };
NORES = 4;
AIC = LOG(DET(COV4)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COV4)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COV4)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT, "NO OF RESTRICTION = " NORES,
 " AIC = " AIC,
 " BIC = " BIC,
 " HQC = " HQC;
/* WITH 5 RESTRICTIONS */
COV5 = { 0.6413151 -0.186015,
 -0.186015 0.15744988 };
NORES = 5;
AIC = LOG(DET(COV5)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COV5)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COV5)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT, "NO OF RESTRICTION = " NORES,
 " AIC = " AIC,
 " BIC = " BIC,
 " HQC = " HQC;
/* WITH 6 RESTRICTIONS */
COV6 = { 177115564 -11663831,
 -11663831 768114.2 };
NORES = 6;
AIC = LOG(DET(COV6)) + (K-NORES)*2.0/ T;

```

```

SAS
22:17 Sunday, August 1, 1993 2
NO OF RESTRICTION = 0
 AIC = -5.598204
 BIC = -5.411192
 HQC = -5.523328
NO OF RESTRICTION = 1
 AIC = -5.604915
 BIC = -5.430371
 HQC = -5.535031
NO OF RESTRICTION = 2
 AIC = -5.611589
 BIC = -5.449512
 HQC = -5.546697
NO OF RESTRICTION = 3
 AIC = -5.364336
 BIC = -5.214727
 HQC = -5.304435
NO OF RESTRICTION = 4
 AIC = -5.267811
 BIC = -5.130669
 HQC = -5.212902
NO OF RESTRICTION = 5
 AIC = -2.644891
 BIC = -2.520217
 HQC = -2.594974
NO OF RESTRICTION = 6
 AIC = 17.140279
 BIC = 17.252486
 HQC = 17.185205
Exiting IML.

```

그림 8.20 벌칙함수식별방법들에 의한 상태공간모형식별

```

BIC = LOG(DET(COV6)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COV6)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT , "NO OF RESTRICTION = " NORES,
 " AIC = " AIC,
 " BIC = " BIC,
 " HQC = " HQC,;
QUIT;

```

이 SAS프로그램을 수행한 결과가 그림 8.20이다. 그림 8.20에서 AIC, BIC, 그리고 HQC 모두 제한조건이 두 개인, 즉 RESTRICT스태이트먼트에 'G(3, 2)=0 G(4, 1)=0'을 할당한 상태공간모형을 선택한다.

다변량시계열데이터 BJ\_J를 상태공간모형화하기 위해서 다음 SAS프로그램을 수행해보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM8-21.PRG에 수록되어 있다).

```

/ * */
/ * EXAM8-21.PRG */
/ * */
/ * STATE SPACE MODELING */
/ * DATASET = BJ_J */

LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=60;

/ * ----- */
/ * FIT A TRANSFER MODEL */
/ * ----- */

TITLE 'ARIMA MODELING';
DATA EXAM821;
 SET TSA4DATA.BJJ;
 RUN;
PROC ARIMA DATA=EXAM821;
 I VAR = Y CROSSCOR=(X) NOPRINT;
 E P=2 INPUT=(3$ (1,2)/ (1)X) MAXIT=30;
 F LEAD=5 OUT=EXAM821B NOPRINT;
 RUN;
PROC PRINT DATA=EXAM821B(FIRSTOBS=285);
 RUN;

/ * ----- */
/ * DEFAULT MODEL AND DESCRIPTIVE STATISTICS */
/ * ----- */

TITLE 'STATESPACE MODELING';
PROC STATESPACE DATA=EXAM821 CANCORR LEAD=5
 OUT=EXAM821C MAXIT=30;

 VAR X Y;
 RESTRICT G(3,2)=0 G(4,1)=0;
 RUN;
PROC PRINT DATA=EXAM821C(FIRSTOBS=285);
 RUN;

```

```

/ * ----- */
/ * TRANSFER FUNTION MODEL */
/ * ----- */
TITLE 'TRANSFER FUNCTION MODEL BY PROC STATESPACE';
PROC STATESPACE DATA=EXAM821 LEAD=5
 OUT=EXAM821D MAXIT=30;
 VAR X Y;
 RESTRICT F(3,2)=0 F(3,4)=0 G(3,2)=0 G(4,1)=0
 G(5,1)=0;
 RUN;
PROC PRINT DATA=EXAM821D(FIRSTOBS=285);
 RUN;
/ * ----- */
/ * COMPLICATED TRANSFER FUNCTION MODEL */
/ * WITH FORM STATEMENT MODEL */
/ * ----- */
TITLE 'COMPLICATED TRANSFER FUNCTION MODEL';
PROC STATESPACE DATA=EXAM821 CANCORR MAXIT=30
 LEAD=5 OUT=EXAM821E;
 VAR X Y;
 FORM X 3 Y 3;
 RESTRICT F(5,2)=0 F(5,4)=0 F(5,6)=0
 G(3,2)=0 G(4,1)=0 G(5,2)=0 G(6,1)=0;
 RUN;
PROC PRINT DATA=EXAM821E(FIRSTOBS=285);
 RUN;

```

이 SAS프로그램의 첫 번째 프로시저는 다음과 같다.

```

PROC ARIMA DATA=EXAM821;
 I VAR=Y CROSSCOR=(X) NOPRINT;
 E P=2 INPUT=(3$ (1,2)/ (1)X) MAXIT=30;
 F LEAD=5 OUT=EXAM821B NOPRINT;
 RUN;

```

이 프로시저를 수행하는 목적은 PROC STATESPACE에 의한 예측결과를 PROC ARIMA에 의한 예측결과와 비교하기 위해서이다. 이 프로시저는 다변량시계열데이터 BJ-J에 PROC ARIMA를 적용해서 전이함수모형화를 하는 것이다. 이 PROC스텝은 [ETS6.1] (pp. 153~162)에서 PROC ARIMA의 사용법을 설명하기 위한 세 번째 예제에서 인용한 것이다. 자세한 내용은 제 4.7.1소절을 참조하라. FORECAST스테이트먼트에서 읍

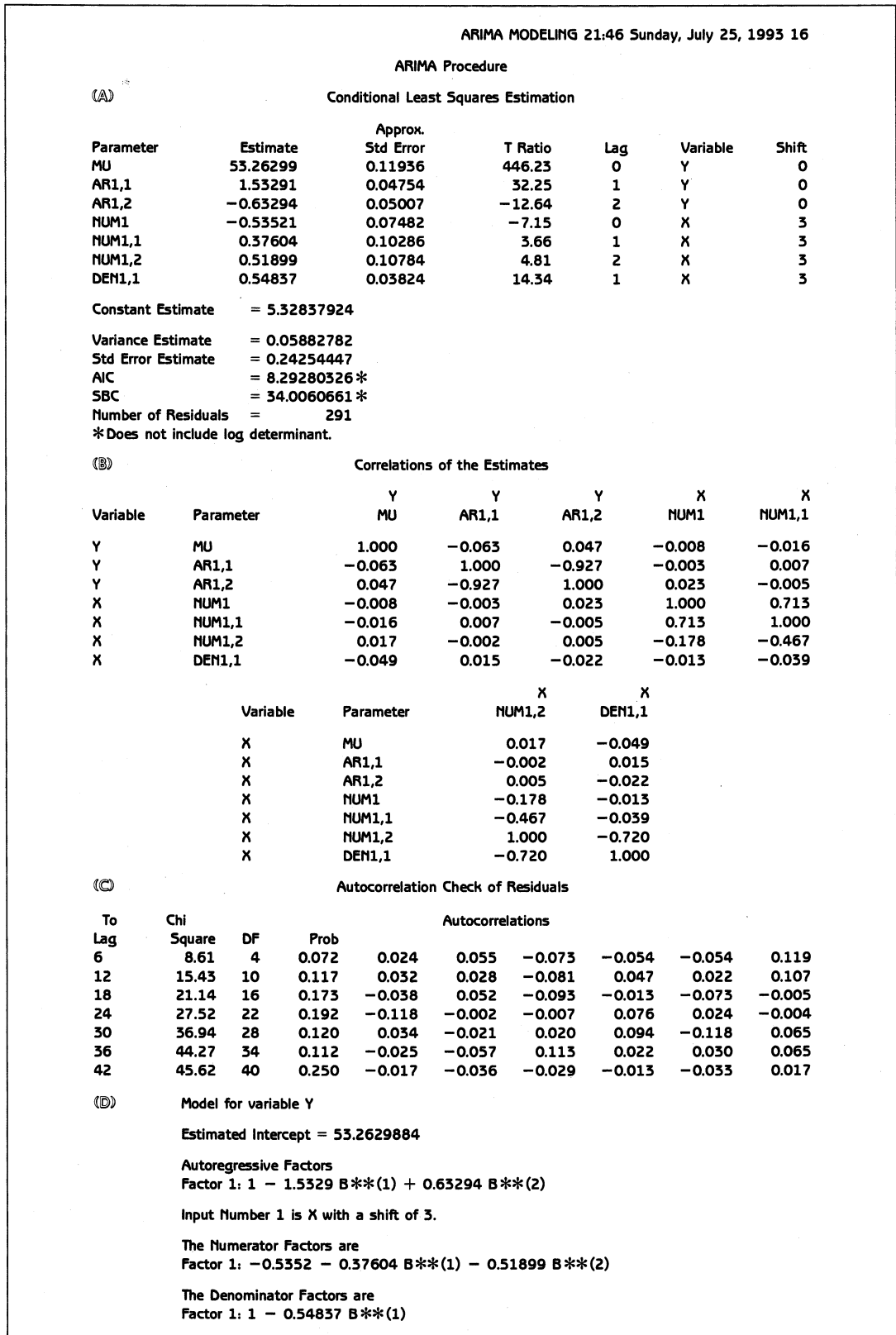


그림 8.21.1 PROC ARIMA의 수행결과

| ARIMA MODELING 21:46 Sunday, July 25, 1993 18 |      |          |         |         |         |          |
|-----------------------------------------------|------|----------|---------|---------|---------|----------|
| OBS                                           | Y    | FORECAST | STD     | L95     | U95     | RESIDUAL |
| 285                                           | 52.6 | 52.6129  | 0.24254 | 52.1376 | 53.0883 | -0.01294 |
| 286                                           | 53.0 | 52.9980  | 0.24254 | 52.5226 | 53.4734 | 0.00200  |
| 287                                           | 54.3 | 53.7222  | 0.24254 | 53.2468 | 54.1975 | 0.57783  |
| 288                                           | 56.0 | 55.4123  | 0.24254 | 54.9369 | 55.8877 | 0.58772  |
| 289                                           | 57.0 | 56.9863  | 0.24254 | 56.5110 | 57.4617 | 0.01366  |
| 290                                           | 58.0 | 57.2844  | 0.24254 | 56.8090 | 57.7598 | 0.71561  |
| 291                                           | 58.6 | 58.0456  | 0.24254 | 57.5702 | 58.5209 | 0.55445  |
| 292                                           | 58.5 | 58.2814  | 0.24254 | 57.8060 | 58.7568 | 0.21861  |
| 293                                           | 58.3 | 57.7346  | 0.24254 | 57.2592 | 58.2100 | 0.56538  |
| 294                                           | 57.8 | 57.5772  | 0.24254 | 57.1018 | 58.0526 | 0.22281  |
| 295                                           | 57.3 | 57.0157  | 0.24254 | 56.5403 | 57.4910 | 0.28433  |
| 296                                           | 57.0 | 56.6199  | 0.24254 | 56.1446 | 57.0953 | 0.38006  |
| 297                                           | .    | 56.5267  | 0.24254 | 56.0513 | 57.0020 | .        |
| 298                                           | .    | 56.0439  | 0.44392 | 55.1738 | 56.9139 | .        |
| 299                                           | .    | 55.6022  | 0.60866 | 54.4093 | 56.7952 | .        |

그림 8.21.2 PROC ARIMA에 의한 예측

선 LEAD=에 5를 할당하여 다섯 개의 선시차들에 대해서 예측을 하였다. 이 예측의 결과를 SAS데이터세트 EXAM821B에 수록하였다. 이 프로시저를 수행한 결과가 그림 8.21.1이다. 그림 8.21.1에서 추정한 전이함수모형은 출력시계열데이터가 Y, 입력시계열데이터가 X, 지연모수가 3, 충격반응함수는 분자가 2차이고 분모가 1차인 유리함수, 그리고 잡음시계열데이터는 AR(2)과정으로 구성되어 있다. 그림 8.21.1의 부분 A에는 이 전이함수모형의 추정결과가 수록되어 있다. 부분 B에는 추정값들 사이의 상관계수들이 수록되어 있다. 부분 C에는 잔차들의 난수성을 조사하기 위한 퍼트맨토우통계량값들이 수록되어 있다. 이 퍼트맨토우통계량값들로부터 추정된 전이함수모형이 주어진 다변량시계열데이터와 잘 부합된다고 할 수 있다. 부분 D에서 확인할 수 있듯이 추정된 전이함수모형은 다음과 같다.

$$y_t = 53.263 - \frac{0.535 + 0.376B + 0.519B^2}{1 - 0.548B} B^3 x_t + \frac{1}{1 - 1.533B + 0.633B^2} v_t$$

이 SAS프로그램의 두 번째 프로시저는 다음과 같다.

```
PROC PRINT DATA=EXAM821B(FIRSTOBS=285);
RUN;
```

이 프로시저는 PROC ARIMA의 FORECAST스태이트먼트의 옵션 OUT=에 지정한 예측들의 SAS데이터세트 EXAM821B를 OUTPUT윈도우에 출력하기 위한 것이다. 이 PROC PRINT를 수행한 결과가 그림 8.21.2이다. 앞의 PROC ARIMA의 FORECAST스태이트먼트의 옵션 LEAD=에 5가 지정되었음에도 불구하고 그림 8.21.2에 세 개의 선시차들에 대해서만 예측이 된 것은 전이함수모형의 지연모수가 3이기 때문이다. 만일 다섯 개의 선시차들에 대해서 예측을 하려면 사전에 입력변수 X의 미래값들을 예측해 두어야 한다.



이 SAS프로그램의 세 번째 프로시저는 다음과 같다. 이 프로시저는 다변량시계열데이터 BJ-J를 상태공간모형화하기 위한 것이다.

```
PROC STATESPACE DATA=EXAM821 CANCORR LEAD=5
 OUT=EXAM821C MAXIT=30;
VAR X Y;
RESTRICT G(3,2)=0 G(4,1)=0;
RUN;
```

이 프로시저에서는 그림 8.20에서 식별한 최적의 상태공간모형을 추정하기 위해서 RESTRICT스테이트먼트에 'G(3, 2)=0 G(4, 1)=0'를 지정했다. 또한 PROC STATESPACE 스테이트먼트의 옵션 LEAD=에 5를 할당함으로써 다섯 개의 선시차들에 대해서 예측을 하고, PROC STATESPACE스테이트먼트의 옵션 OUT=에 EXAM821C를 할당함으로써 예측의 결과를 SAS데이터세트 EXAM821C에 출력한다. 이 프로시저를 수행한 결과가 그림 8.21.3이다. 그림 8.21.3의 각 부분을 자세히 설명하면 다음과 같다. 부분 A에서 이 SAS 데이터세트의 관찰점들의 개수가 296, 부분 B에서 변수들 X와 Y의 표본평균이 각각 -0.057, 53.509, 표본표준편차가 각각 1.073, 3.202임을 알 수 있다. 부분 C에는 VAR(0)모형, VAR(1)모형, ..., VAR(10)모형에 해당하는 AIC값들이 수록되어 있다. 이 중에서 VAR(4)모형의 AIC값이 최소이다. 즉, 상태공간모형화의 VAR모형단계에서 VAR(4)모형을 선택한다. 부분 D에는 VAR스테이트먼트에 지정된 변수들의 표본자기상관행렬들의 도표적 표현이 수록되어 있다. 부분 E에는 VAR스테이트먼트에 지정된 변수들의 표본부분자기상관행렬들의 도표적 표현이 수록되어 있다. 부분 F에는 최소AIC값을 갖는 VAR모형의 VAR계수행렬들의 Yule-Walker추정행렬들이 수록되어 있다. 이 예제에서는 VAR(4)모형에서 최소의 AIC값을 가지므로 VAR(4)모형의 VAR계수행렬들이 추정되었다. 부분 G는 PROC STATESPACE스테이트먼트의 옵션 CANCORR를 사용함으로써 출력된 것으로, 상태벡터들, 표본정준상관계수들, AIC값들, Bartlett통계량값들, 그리고 이에 해당되는 자유도들이 출력되었다. 부분 H는 예비추정단계의 결과물이다. 부분 H1은 예비추정단계의 상태벡터이고, 부분 H2는 변환행렬 F의 예비추정행렬이며, 부분 H3는 입력행렬 G의 예비추정행렬이며, 부분 H4는 쇄신과정의 공분산행렬의 예비추정행렬이다. 부분 I는 최종추정단계의 결과물이다. 부분 II는 최종추정단계의 상태벡터이다. 부분 I2는 변환행렬 F의 최종추정행렬이고, 부분 I3는 입력행렬 G의 최종추정행렬이며, 부분 I4는 쇄신과정의 공분산행렬의 최종추정행렬이다. 부분 J는 행렬들 F와 G의 최종적으로 추정된 원소들의 최종추정값들, 표준오차들, 그리고 이에 해당하는 t통계량값들이 수록되어 있다.

상태공간모형의 최종추정식은 다음과 같다.

$$y_{1,t} = X(t) + 0.057$$

$$y_{2,t} = Y(t) - 53.509$$

$$z_{t+1} = \hat{F}z_t + \hat{G}v_{t+1}$$

$$y_t = [I_2 \ O] z_t$$

$$\text{단, } z_t = (y_{1,t}, y_{2,t}, y_{1,t+1|t}, y_{2,t+1|t}, y_{2,t+2|t})'$$

$$\hat{F} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ -0.002 & 0.031 & 1.725 & -0.055 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -0.352 & 0.289 & -0.093 & -1.091 & 1.666 \end{bmatrix}, \quad \hat{G} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1.926 & 0.000 \\ 0.000 & 1.255 \\ 0.060 & 1.350 \end{bmatrix}$$

$$\hat{\Sigma} = \begin{bmatrix} 0.0356 & -0.0073 \\ -0.0073 & 0.0956 \end{bmatrix}$$

이 상태공간모형식을 전개하면 다음과 같다.

$$y_{1,t+1} = X(t) + 0.057$$

$$y_{2,t+1} = Y(t) - 53.509$$

$$y_{1,t+1} = y_{1,t+1|t} + v_{1,t+1}$$

$$y_{2,t+1} = y_{2,t+1|t} + v_{2,t+1}$$

$$y_{1,t+2|t+1} = -0.002y_{1,t} + 0.031y_{2,t} + 1.725y_{1,t+1|t} - 0.055y_{2,t+1|t} + 1.926v_{1,t+1}$$

$$y_{2,t+2|t+1} = y_{2,t+2|t} + 1.255v_{2,t+1}$$

$$y_{2,t+3|t+1} = -0.352y_{1,t} + 0.289y_{2,t} - 0.093y_{1,t+1|t} - 1.091y_{2,t+1|t} + 1.666y_{2,t+2|t} + 0.060v_{1,t+1} + 1.350v_{2,t+1}$$

첫 번째 식과 두 번째 식을 세 번째 식과 네 번째 식에 대입하면 다음 식들을 얻는다.

$$y_{1,t+2} = 1.725y_{1,t+1} - 0.055y_{2,t+1} - 0.002y_{1,t} + 0.031y_{2,t} + v_{1,t+2} + 0.201v_{1,t+1} + 0.055v_{2,t+1}$$

$$y_{2,t+2|t+1} = y_{2,t+2|t} - v_{2,t+2} - 1.255v_{2,t+1}$$

이 식들을 정리된 다섯 번째 식에 대입하면 다음 식을 얻는다.

$$y_{2,t+3} = 1.666y_{2,t+2} - 0.093y_{1,t+1} - 1.091y_{2,t+1} - 0.352y_{1,t} + 0.298y_{2,t} + v_{2,t+3} - 0.411v_{2,t+2} + 0.153v_{1,t+1} + 0.350v_{2,t+1}$$

이 식을 정리하면 다음과 같은 2변량VARMA(3, 2)모형을 얻는다.

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 1.725 & -0.055 \\ 0 & 1.666 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} -0.002 & 0.031 \\ -0.093 & -1.091 \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ -0.352 & 0.298 \end{bmatrix} \begin{bmatrix} y_{1,t-3} \\ y_{2,t-3} \end{bmatrix} + \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} - \begin{bmatrix} -0.201 & -0.055 \\ 0 & 0.411 \end{bmatrix} \begin{bmatrix} v_{1,t-1} \\ v_{2,t-1} \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ -0.153 & -0.350 \end{bmatrix} \begin{bmatrix} v_{1,t-2} \\ v_{2,t-2} \end{bmatrix}$$

STATESPACE MODELING 21:46 Sunday,24, 1993  
19

STATESPACE Procedure

(A)

Nobs = 296

(B)

| Variable | Mean     | Std      |
|----------|----------|----------|
| X        | -0.05683 | 1.072766 |
| Y        | 53.50912 | 3.202121 |

(C)

Information Criterion for Autoregressive Models

| Lag=0       | Lag=1       | Lag=2       | Lag=3       | Lag=4       |
|-------------|-------------|-------------|-------------|-------------|
| 651.386156  | -1033.57211 | -1632.95659 | -1645.12498 | -1651.52305 |
| Lag=5       | Lag=6       | Lag=7       | Lag=8       | Lag=9       |
| -1648.91248 | -1649.34335 | -1643.15323 | -1638.55591 | -1634.79898 |
| Lag=10      |             |             |             |             |
| -1633.58501 |             |             |             |             |

(D)

Schematic Representation of Correlations

| Name/Lag | 0  | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 |
|----------|----|----|----|----|----|----|----|----|----|----|----|
| X        | +- | +- | +- | +- | +- | +- | +- | +- | +- | +- | +- |
| Y        | -+ | -+ | -+ | -+ | -+ | -+ | -+ | -+ | -+ | -+ | -+ |

+ is > 2\*std error, - is < -2\*std error, . is between

(E)

Schematic Representation of Partial Autocorrelations

| Name/Lag | 1  | 2  | 3 | 4  | 5  | 6  | 7  | 8  | 9  | 10 |
|----------|----|----|---|----|----|----|----|----|----|----|
| X        | +  | -  | + | .. | .. | -  | .. | .. | .. | .. |
| Y        | -+ | -- | - | +  | .. | .. | .. | .. | .. | +  |

+ is > 2\*std error, - is < -2\*std error, . is between

(F)

Yule-Walker Estimates for the Min AIC

|   | Lag=1    |          | Lag=2    |          |
|---|----------|----------|----------|----------|
|   | X        | Y        | X        | Y        |
| X | 1.92589  | -0.00124 | -1.20166 | 0.00422  |
| Y | 0.05050  | 1.29979  | -0.02046 | -0.32770 |
|   | Lag=3    |          | Lag=4    |          |
|   | X        | Y        | X        | Y        |
| X | 0.11692  | -0.00867 | 0.10424  | 0.00327  |
| Y | -0.71182 | -0.25701 | 0.19541  | 0.13342  |

(G)

Canonical Correlations Analysis

| State vector                                          | Correlations |        | Infor. | Chisq    | D.F.     |   |
|-------------------------------------------------------|--------------|--------|--------|----------|----------|---|
| X(T;T),Y(T;T),X(T+1;T)                                | 1.0000       | 1.0000 | 0.8049 | 292.9228 | 304.7481 | 8 |
| X(T;T),Y(T;T),X(T+1;T),<br>Y(T+1;T)                   | 1.0000       | 1.0000 | 0.9067 | 122.3358 | 134.7237 | 7 |
| X(T;T),Y(T;T),X(T+1;T),<br>Y(T+1;T),X(T+2;T)          | 1.0000       | 1.0000 | 0.9094 | -1.54701 | 10.34705 | 6 |
| X(T;T),Y(T;T),X(T+1;T),<br>Y(T+1;T),Y(T+2;T)          | 1.0000       | 1.0000 | 0.9101 | 0.940392 | 12.80924 | 6 |
| X(T;T),Y(T;T),X(T+1;T),<br>Y(T+1;T),Y(T+2;T),Y(T+3;T) | 1.0000       | 1.0000 | 0.9130 | -7.94103 | 2.041584 | 5 |
|                                                       | 0.6288       | 0.2266 | 0.0833 |          |          |   |

(H)

Selected Statespace Form and Preliminary Estimates

(HI)

State Vector

|          |        |          |          |
|----------|--------|----------|----------|
| X(T;T)   | Y(T;T) | X(T+1;T) | Y(T+1;T) |
| Y(T+2;T) |        |          |          |

(H2) Estimate of the Transition Matrix

|        |       |        |        |       |
|--------|-------|--------|--------|-------|
| 0      | 0     | 1      | 0      | 0     |
| 0      | 0     | 0      | 1      | 0     |
| -0.847 | 0.027 | 1.712  | -0.050 | 0     |
| 0      | 0     | 0      | 0      | 1     |
| -0.198 | 0.334 | -0.182 | -1.236 | 1.787 |

(H3) Input Matrix for the Innovation

|  |       |       |
|--|-------|-------|
|  | 1     | 0     |
|  | 0     | 1     |
|  | 1.926 | 0.000 |
|  | 0.000 | 1.300 |
|  | 0.142 | 1.362 |

(H4) Variance Matrix for the Innovation

|            |            |
|------------|------------|
| 0.0352738  | -0.0073378 |
| -0.0073378 | 0.09756908 |

Estimates give unstable system.  
 No improvement after 10 step halvings.  
 Convergence has been assumed.

(I) Selected Statespace Form and Preliminary Estimates

(II) State Vector

|          |        |          |          |
|----------|--------|----------|----------|
| X(T;T)   | Y(T;T) | X(T+1;T) | Y(T+1;T) |
| Y(T+2;T) |        |          |          |

(I2) Estimate of the Transition Matrix

|        |       |        |        |       |
|--------|-------|--------|--------|-------|
| 0      | 0     | 1      | 0      | 0     |
| 0      | 0     | 0      | 1      | 0     |
| -0.002 | 0.031 | 1.725  | -0.055 | 0     |
| 0      | 0     | 0      | 0      | 1     |
| -0.352 | 0.289 | -0.093 | -1.091 | 1.666 |

(I3) Input Matrix for the Innovation

|  |       |       |
|--|-------|-------|
|  | 1     | 0     |
|  | 0     | 1     |
|  | 1.926 | 0.000 |
|  | 0.000 | 1.255 |
|  | 0.060 | 1.350 |

(I4) Variance Matrix for the Innovation

|            |            |
|------------|------------|
| 0.03557464 | -0.007285  |
| -0.007285  | 0.09560056 |

(J) Parameter Estimates

| Parameter | Estimate | Std. Err. | T value  |
|-----------|----------|-----------|----------|
| F(3,1)    | -0.00168 | 0.077162  | -0.02183 |
| F(3,2)    | 0.030851 | 0.027882  | 1.106497 |
| F(3,3)    | 1.724639 | 0.064672  | 26.66757 |
| F(3,4)    | -0.0551  | 0.032225  | -1.70993 |
| F(5,1)    | -0.35242 | 0.240225  | -1.46704 |
| F(5,2)    | 0.289241 | 0.098965  | 2.92265  |
| F(5,3)    | -0.09309 | 0.147112  | -0.63279 |
| F(5,4)    | -1.09059 | 0.263284  | -4.14224 |
| F(5,5)    | 1.665801 | 0.200215  | 8.320041 |
| G(3,1)    | 1.926006 | 0.057377  | 33.56728 |
| G(4,2)    | 1.2552   | 0.056204  | 22.33306 |
| G(5,1)    | 0.059541 | 0.088841  | 0.670194 |
| G(5,2)    | 1.349563 | 0.091152  | 14.80568 |

그림 8.21.3 RESTRICT스태이트먼트를 사용한 상태공간모형화

| STATESPACE MODELING 21.46 Sunday, July 25, 1993 |        |          |          |         |      |         |          |         |
|-------------------------------------------------|--------|----------|----------|---------|------|---------|----------|---------|
| 25                                              |        |          |          |         |      |         |          |         |
| OBS                                             | X      | FOR1     | RES1     | STD1    | Y    | FOR2    | RES2     | STD2    |
| 285                                             | -0.824 | -1.20861 | 0.38461  | 0.18861 | 52.6 | 52.5722 | 0.02784  | 0.30919 |
| 286                                             | -0.740 | -1.27771 | 0.53771  | 0.18861 | 53.0 | 53.0516 | -0.05164 | 0.30919 |
| 287                                             | -0.528 | -1.12832 | 0.60032  | 0.18861 | 54.3 | 53.7980 | 0.50195  | 0.30919 |
| 288                                             | -0.204 | -0.77902 | 0.57502  | 0.18861 | 56.0 | 55.3107 | 0.68925  | 0.30919 |
| 289                                             | 0.034  | -0.26893 | 0.30293  | 0.18861 | 57.0 | 56.9172 | 0.08281  | 0.30919 |
| 290                                             | 0.204  | 0.05013  | 0.15387  | 0.18861 | 58.0 | 57.3107 | 0.68931  | 0.30919 |
| 291                                             | 0.253  | 0.32207  | -0.06907 | 0.18861 | 58.6 | 57.7389 | 0.86107  | 0.30919 |
| 292                                             | 0.195  | 0.36865  | -0.17365 | 0.18861 | 58.5 | 57.9559 | 0.54410  | 0.30919 |
| 293                                             | 0.131  | 0.25403  | -0.12303 | 0.18861 | 58.3 | 57.5158 | 0.78416  | 0.30919 |
| 294                                             | 0.017  | 0.17511  | -0.15811 | 0.18861 | 57.8 | 57.2282 | 0.57181  | 0.30919 |
| 295                                             | -0.182 | -0.01877 | -0.16323 | 0.18861 | 57.3 | 56.7896 | 0.51037  | 0.30919 |
| 296                                             | -0.262 | -0.35408 | 0.09208  | 0.18861 | 57.0 | 56.4231 | 0.57694  | 0.30919 |
| 297                                             | .      | -0.43553 | .        | 0.18861 | .    | 56.3331 | .        | 0.30919 |
| 298                                             | .      | -0.75751 | .        | 0.40931 | .    | 55.7815 | .        | 0.49621 |
| 299                                             | .      | -1.30270 | .        | 0.74943 | .    | 55.3318 | .        | 0.64753 |
| 300                                             | .      | -2.23465 | .        | 1.31710 | .    | 55.0827 | .        | 0.75089 |
| 301                                             | .      | -3.84117 | .        | 2.28959 | .    | 55.1628 | .        | 0.87469 |

그림 8.21.4 PROC STATESPACE에 의한 예측

네 번째 PROC스텝은 다음과 같다.

```
PROC PRINT DATA= EXAM821C(FIRSTOBS=285);
```

```
RUN;
```

이 프로시저는 앞의 PROC STATESPACE의 PROC STATESPACE스태이트먼트의 옵션 OUT=에 지정한 SAS데이터세트 EXAM821C를 OUTPUT윈도우에 출력시키기 위한 것이다. 이 프로시저를 수행한 결과가 그림 8.21.4이다.

이 SAS프로그램의 다섯 번째 프로시저는 다음과 같다. 이 프로시저는 다변량시계열데이터 BJ\_J를 PROC STATESPACE를 사용해서 전이함수모형화하기 위한 것으로 [ETS6.1] (pp. 485~493)의 예제를 그대로 사용한 것이다.

```
PROC STATESPACE DATA=EXAM821 LEAD=5
```

```
OUT=EXAM821D MAXIT=30;
```

```
VAR X Y;
```

```
RESTRICT F(3,2)=0 F(3,4)=0 G(3,2)=0 G(4,1)=0
```

```
G(5,1)=0;
```

```
RUN;
```

이 프로시저에서는 PROC STATESPACE를 사용해서 전이함수모형을 추정하기 위해서 RESTRICT스태이트먼트에 'F(3, 2)=0 F(3, 4)=0 G(3, 2)=0 G(4, 1)=0 G(5, 1)=0'를 지정하였다. 또한 PROC STATESPACE스태이트먼트의 옵션 LEAD=에 5를 할당함으로써 다섯 개의 선시차들에 대해서 예측을 하고, PROC STATESPACE스태이트먼트의 옵션 OUT=에 EXAM821D를 할당함으로써 예측의 결과를 SAS데이터세트 EXAM821D에 출력한다. 이 프로시저를 수행한 결과의 일부분이 그림 8.21.5이다. 부분 A는 최종추정단계

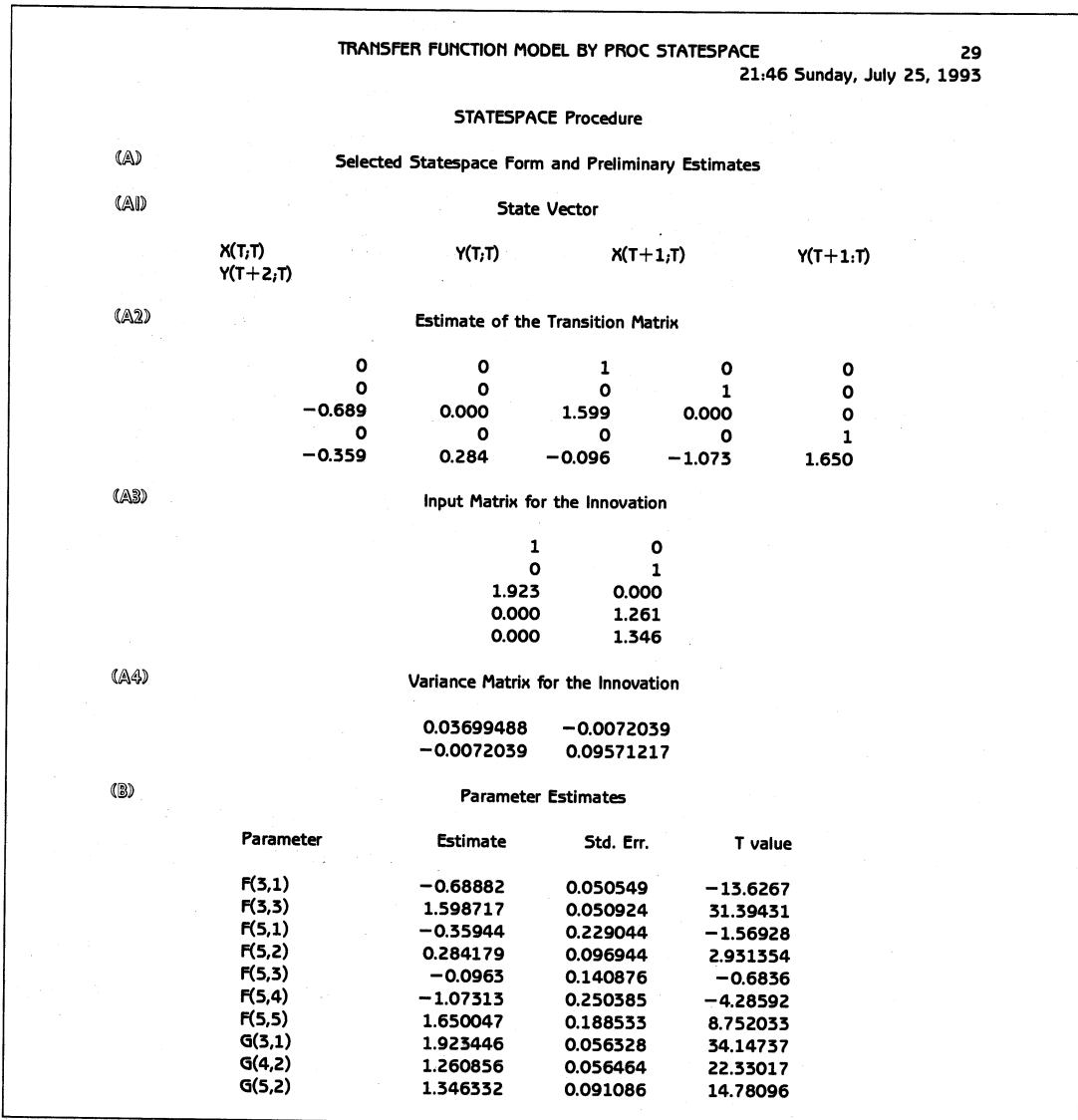


그림 8.21.5 PROC STATESPACE에 의한 전이함수모형화

를 수행한 결과이다. 부분 A1은 최종추정단계의 상태벡터이고, 부분 A2는 변환행렬  $F$ 의 최종추정행렬이고, 부분 A3는 입력행렬  $G$ 의 최종추정행렬이며, 부분 A4는 쇄신과정의 공분산행렬의 최종추정행렬이다. 부분 B에는 행렬들  $F$ 와  $G$ 의 원소들의 최종추정값들, 표준오차들, 그리고 이에 해당하는  $t$ 통계량값들이 수록되어 있다.

상태공간모형의 최종추정식은 다음과 같다.

$$y_{1,t} = X(t) + 0.057$$

$$y_{2,t} = Y(t) - 53.509$$

$$z_{t+1} = \hat{F}z_t + \hat{G}v_{t+1}$$

$$y_t = [I_2 \ 0] z_t$$

$$\text{단, } z_t = (y_{1,t}, y_{2,t}, y_{1,t+1|t}, y_{2,t+1|t}, y_{2,t+2|t})'$$

$$\hat{F} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ -0.689 & 0.000 & 1.599 & 0.000 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -0.359 & 0.284 & -0.096 & -1.073 & 1.650 \end{bmatrix}, \quad \hat{G} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1.923 & 0.000 \\ 0.000 & 1.261 \\ 0.000 & 1.346 \end{bmatrix}$$

$$\hat{\Sigma} = \begin{bmatrix} 0.0370 & -0.0072 \\ -0.0072 & 0.0957 \end{bmatrix}$$

이 상태공간모형식을 전개하면 다음과 같다.

$$\begin{aligned} y_{1,t+1} &= y_{1,t+1|t} + v_{1,t+1} \\ y_{2,t+1} &= y_{2,t+1|t} + v_{2,t+1} \\ y_{1,t+2|t+1} &= -0.689y_{1,t} + 1.599y_{1,t+1|t} + 1.923v_{1,t+1} \\ y_{2,t+2|t+1} &= y_{2,t+2|t} + 1.261v_{2,t+1} \\ y_{2,t+3|t+1} &= -0.359y_{1,t} + 0.284y_{2,t} - 0.096y_{1,t+1|t} - 1.073y_{2,t+1|t} \\ &\quad + 1.650y_{2,t+2|t} + 1.346v_{2,t+1} \end{aligned}$$

첫 번째 식과 두 번째 식을 세 번째 식과 네 번째 식에 대입하면 다음 식들을 얻는다.

$$\begin{aligned} y_{1,t} &= 1.599y_{1,t+1} - 0.689y_{1,t} + v_{1,t+2} + 0.324v_{1,t+1} \\ y_{2,t+2|t} &= y_{2,t+2} - v_{2,t+2} - 1.261v_{2,t+1} \end{aligned}$$

이 식들을 정리된 다섯 번째 식에 대입하면 다음 식을 얻는다.

$$\begin{aligned} y_{2,t+3} &= 1.650y_{2,t+2} - 0.096y_{1,t+1} - 1.073y_{2,t+1} - 0.359y_{1,t} + 0.284y_{2,t} \\ &\quad + v_{2,t+3} - 0.389v_{2,t+2} + 0.096v_{1,t+1} + 0.338v_{2,t+1} \end{aligned}$$

이 식을 정리하면 다음과 같은 2변량VARMA(3, 2)모형을 얻는다.

$$\begin{aligned} \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} &= \begin{bmatrix} 1.599 & 0 \\ 0 & 1.650 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} -0.689 & 0 \\ -0.096 & -1.073 \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ -0.359 & 0.284 \end{bmatrix} \begin{bmatrix} y_{1,t-3} \\ y_{2,t-3} \end{bmatrix} \\ &\quad + \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} - \begin{bmatrix} -0.324 & 0 \\ 0 & 0.389 \end{bmatrix} \begin{bmatrix} v_{1,t-1} \\ v_{2,t-1} \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ -0.096 & -0.338 \end{bmatrix} \begin{bmatrix} v_{1,t-2} \\ v_{2,t-2} \end{bmatrix} \end{aligned}$$

다섯 번째 PROC스텝은 다음과 같다.

```
PROC PRINT DATA = EXAM821D(FIRSTOBS = 285);
RUN;
```

이 프로시저는 앞의 PROC스텝의 PROC STATESPACE스테이트먼트의 옵션 OUT=에 지정한 SAS데이터세트 EXAM821D를 OUTPUT윈도우에 출력시키는 것이다. 이 프로시저를 수행한 결과가 그림 8.21.6이다.

| TRANSFER FUNCTION MODEL BY PROC STATESPACE |        |          |          |         |      |         |          | 30                          |
|--------------------------------------------|--------|----------|----------|---------|------|---------|----------|-----------------------------|
|                                            |        |          |          |         |      |         |          | 21:46 Sunday, July 25, 1993 |
| OBS                                        | X      | FOR1     | RES1     | STD1    | Y    | FOR2    | RES2     | STD2                        |
| 285                                        | -0.824 | -0.84523 | 0.02123  | 0.19234 | 52.6 | 52.6435 | -0.04348 | 0.30937                     |
| 286                                        | -0.740 | -0.79276 | 0.05276  | 0.19234 | 53.0 | 53.0570 | -0.05703 | 0.30937                     |
| 287                                        | -0.528 | -0.60346 | 0.07546  | 0.19234 | 54.3 | 53.7221 | 0.57793  | 0.30937                     |
| 288                                        | -0.204 | -0.31502 | 0.11102  | 0.19234 | 56.0 | 55.2045 | 0.79553  | 0.30937                     |
| 289                                        | 0.034  | 0.06849  | -0.03449 | 0.19234 | 57.0 | 56.8103 | 0.18971  | 0.30937                     |
| 290                                        | 0.204  | 0.17856  | 0.02544  | 0.19234 | 58.0 | 57.2110 | 0.78899  | 0.30937                     |
| 291                                        | 0.253  | 0.30586  | -0.05286 | 0.19234 | 58.6 | 57.6790 | 0.92097  | 0.30937                     |
| 292                                        | 0.195  | 0.24167  | -0.04667 | 0.19234 | 58.5 | 57.9358 | 0.56425  | 0.30937                     |
| 293                                        | 0.131  | 0.11720  | 0.01380  | 0.19234 | 58.3 | 57.5213 | 0.77875  | 0.30937                     |
| 294                                        | 0.017  | 0.07447  | -0.05747 | 0.19234 | 57.8 | 57.2532 | 0.54677  | 0.30937                     |
| 295                                        | -0.182 | -0.08684 | -0.09516 | 0.19234 | 57.3 | 56.8102 | 0.48976  | 0.30937                     |
| 296                                        | -0.262 | -0.33870 | 0.07670  | 0.19234 | 57.0 | 56.4356 | 0.56438  | 0.30937                     |
| 297                                        | .      | -0.27371 | .        | 0.19234 | .    | 56.3432 | .        | 0.30937                     |
| 298                                        | .      | -0.26224 | .        | 0.41697 | .    | 55.7801 | .        | 0.49787                     |
| 299                                        | .      | -0.23583 | .        | 0.62009 | .    | 55.3016 | .        | 0.64912                     |
| 300                                        | .      | -0.20152 | .        | 0.78351 | .    | 54.9329 | .        | 0.75401                     |
| 301                                        | .      | -0.16484 | .        | 0.90331 | .    | 54.6713 | .        | 0.88095                     |

그림 8.21.6 PROC STATESPACE를 사용해 구축한 전이함수모형의 예측

이 SAS프로그램의 여섯 번째 프로시저는 다음과 같다. 이 프로시저는 다변량시계열데이터 BJ\_J를 PROC STATESPACE를 사용해서 좀더 복잡한 전이함수모형에 적합시키기 위한 것으로 [ETS6.1] (pp. 485~493)의 예제를 그대로 사용한 것이다.

```

PROC STATESPACE DATA=EXAM821 CANCERR MAXIT=30
 LEAD=5 OUT=EXAM821E;
VAR X Y;
FORM X 3 Y 3;
RESTRICT F(5,2)=0 F(5,4)=0 F(5,6)=0
 G(3,2)=0 G(4,1)=0 G(5,2)=0 G(6,1)=0;
RUN;

```

이 프로시저는 PROC STATESPACE를 사용해서 좀더 복잡한 전이함수모형을 추정하기 위한 것이다. 우선 FORM스테이트먼트를 사용해서 다음 변수들이 상태벡터를 구성하도록 하였다.

$$X(T), Y(T), X(T+1;T), Y(T+1;T), X(T+2;T), Y(T+2;T)$$

PROC STATESPACE를 사용해서 전이함수모형을 추정하기 위해서 RESTRICT스테이트먼트에 'F(5, 2)=0 F(5, 4)=0 F(5, 6)=0 G(3, 2)=0 G(4, 1)=0 G(5, 2)=0 G(6, 1)=0'을 지정하였다. 또한 PROC STATESPACE스테이트먼트의 옵션 LEAD=에 5를 할당함으로써 다섯 개의 선시차들에 대해서 예측을 하고, PROC STATESPACE스테이트먼트의 옵션 OUT=에 EXAM821E를 할당함으로써 예측의 결과를 SAS데이터세트 EXAM821E에 출력한다. 이 프로시저를 수행한 결과의 일부분이 그림 8.21.7이다. 부분 A는 PROC STATESPACE스테이트먼트의 옵션 CANCERR를 사용함으로써 출력된 것으로, 상태백



COMPLICATED TRANSFER FUNCTION MODEL

33

21:46 Sunday, July 25, 1993

STATESPACE Procedure

(A)

Canonical Correlations Analysis

| State vector                                                       | Correlations               |                  |                  | Infor.   | Chisq    | D.F. |
|--------------------------------------------------------------------|----------------------------|------------------|------------------|----------|----------|------|
| X(T;T),Y(T;T),X(T+1;T)                                             | 1.0000                     | 1.0000           | 0.8049           | 292.9228 | 304.7481 | 8    |
| X(T;T),Y(T;T),X(T+1;T),<br>Y(T+1;T)                                | 1.0000<br>0.6075           | 1.0000           | 0.9067           | 122.3358 | 134.7237 | 7    |
| X(T;T),Y(T;T),X(T+1;T),<br>Y(T+1;T),X(T+2;T)                       | 1.0000<br>0.6103           | 1.0000<br>0.1863 | 0.9094           | -1.54701 | 10.34705 | 6    |
| X(T;T),Y(T;T),X(T+1;T),<br>Y(T+1;T),X(T+2;T),Y(T+2;T)              | 1.0000<br>0.6200           | 1.0000<br>0.2208 | 0.9121<br>0.1770 | -0.57992 | 9.340518 | 5    |
| X(T;T),Y(T;T),X(T+1;T),<br>Y(T+1;T),X(T+2;T),Y(T+2;T),<br>X(T+3;T) | 1.0000<br>0.6204<br>0.1613 | 1.0000<br>0.2214 | 0.9127<br>0.1821 | -0.1949  | 7.752361 | 4    |
| X(T;T),Y(T;T),X(T+1;T),<br>Y(T+1;T),X(T+2;T),Y(T+2;T),<br>Y(T+3;T) | 1.0000<br>0.6299<br>0.0478 | 1.0000<br>0.2461 | 0.9148<br>0.1774 | -7.32366 | 0.671773 | 4    |

(B)

Selected Statespace Form and Preliminary Estimates

(B1)

State Vector

| X(T;T)<br>X(T+2;T) | Y(T;T)<br>Y(T+2;T) | X(T+1;T) | Y(T+1;T) |
|--------------------|--------------------|----------|----------|
|--------------------|--------------------|----------|----------|

(B2)

Estimate of the Transition Matrix

|       |       |        |        |       |       |
|-------|-------|--------|--------|-------|-------|
| 0     | 0     | 1      | 0      | 0     | 0     |
| 0     | 0     | 0      | 1      | 0     | 0     |
| 0     | 0     | 0      | 0      | 1     | 0     |
| 0     | 0     | 0      | 0      | 0     | 1     |
| 0.432 | 0.000 | -1.913 | 0.000  | 2.360 | 0.000 |
| 0.331 | 0.281 | -1.350 | -1.108 | 0.702 | 1.719 |

(B3)

Input Matrix for the Innovation

|       |       |
|-------|-------|
| 1     | 0     |
| 0     | 1     |
| 1.926 | 0.000 |
| 0.000 | 1.300 |
| 2.507 | 0.000 |
| 0.000 | 1.362 |

(B4)

Variance Matrix for the Innovation

|            |            |
|------------|------------|
| 0.0352738  | -0.0073378 |
| -0.0073378 | 0.09756908 |

Estimates give unstable system.  
No improvement after 10 step halvings.  
Convergence has been assumed.

(C)

Selected Statespace Form and Fitted Model

(C1)

State Vector

| X(T;T)<br>X(T+2;T) | Y(T;T)<br>Y(T+2;T) | X(T+1;T) | Y(T+1;T) |
|--------------------|--------------------|----------|----------|
|--------------------|--------------------|----------|----------|

(C2)

Estimate of the Transition Matrix

|   |   |   |   |   |   |
|---|---|---|---|---|---|
| 0 | 0 | 1 | 0 | 0 | 0 |
|---|---|---|---|---|---|

|      |                                    |            |           |          |       |       |
|------|------------------------------------|------------|-----------|----------|-------|-------|
|      | 0                                  | 0          | 0         | 1        | 0     | 0     |
|      | 0                                  | 0          | 0         | 0        | 1     | 0     |
|      | 0                                  | 0          | 0         | 0        | 0     | 1     |
|      | 0.515                              | 0.000      | -0.003    | 0.000    | 2.156 | 0.000 |
|      | 0.085                              | 0.255      | -1.006    | -1.022   | 0.540 | 1.641 |
| (C3) | Input Matrix for the Innovation    |            |           |          |       |       |
|      | 1                                  |            |           |          |       |       |
|      | 0                                  | 0          |           |          |       |       |
|      | 1.919                              | 0.000      |           |          |       |       |
|      | 0.000                              | 1.266      |           |          |       |       |
|      | 2.523                              | 0.000      |           |          |       |       |
|      | 0.000                              | 1.355      |           |          |       |       |
| (C4) | Variance Matrix for the Innovation |            |           |          |       |       |
|      | 0.03522668                         | -0.007253  |           |          |       |       |
|      | -0.007253                          | 0.09546338 |           |          |       |       |
| (D)  | Parameter Estimates                |            |           |          |       |       |
|      | Parameter                          | Estimate   | Std. Err. | T value  |       |       |
|      | F(5,1)                             | 0.514664   | 0.12418   | 4.144489 |       |       |
|      | F(5,3)                             | -0.00335   | 0.281175  | -0.0119  |       |       |
|      | F(5,5)                             | 2.156293   | 0.175757  | 12.26859 |       |       |
|      | F(6,1)                             | 0.085297   | 0.583185  | 0.14626  |       |       |
|      | F(6,2)                             | 0.254607   | 0.118326  | 2.151738 |       |       |
|      | F(6,3)                             | -1.00594   | 1.108475  | -0.9075  |       |       |
|      | F(6,4)                             | -1.02228   | 0.304248  | -3.36002 |       |       |
|      | F(6,5)                             | 0.540282   | 0.644432  | 0.838385 |       |       |
|      | F(6,6)                             | 1.64112    | 0.22113   | 7.421528 |       |       |
|      | G(3,1)                             | 1.919238   | 0.057602  | 33.3192  |       |       |
|      | G(4,2)                             | 1.266427   | 0.057145  | 22.16148 |       |       |
|      | G(5,1)                             | 2.523116   | 0.12456   | 20.25619 |       |       |
|      | G(6,2)                             | 1.35515    | 0.09247   | 14.65506 |       |       |

그림 8.21.7 PROC STATESPACE를 사용한 복잡한 전이함수모형화

터들, 표본정준상관계수들, DIC값들, Bartlett통계량값들, 그리고 이에 해당되는 자유도들이 출력되었다. FORM스테이트먼트에 지정한 여섯 개의 변수들의 모형에서는 DIC값이 음수이다. 따라서 이 상태벡터는 적당한 것이 아니다. 또한 유의수준 0.05에서 자유도가 5인 카이제곱분포의 임계값은 11.07이다. 따라서 Bartlett검정도 이 결론을 지지한다. 그러나 FORM스테이트먼트에서 상태벡터를 지정하였으므로 이 상태벡터가 이후 분석에서 사용될 것이다. 부분 B는 예비추정단계의 결과물이다. 부분 D는 예비추정단계의 상태벡터이고, 부분 B2는 변환행렬 F의 예비추정행렬이며, 부분 B3는 입력행렬 G의 예비추정행렬이며, 부분 B4는 쇄신과정의 공분산행렬의 예비추정행렬이다. 부분 C는 최종추정단계의 결과물이다. 부분 C1은 최종추정단계의 상태벡터이다. 부분 C2는 변환행렬 F의 최종추정행렬이고, 부분 C3는 입력행렬 G의 최종추정행렬이며, 부분 C4는 쇄신과정의 공분산행렬의 최종추정행렬이다. 부분 J에는 부분 D의 행렬들 F와 G의 추정된 원소들의 최종추정값들, 표준오차들, 이에 해당하는 t통계량값들이 수록되어 있다.

상태공간모형의 최종추정식은 다음과 같다.

$$y_{1,t} = X(t) + 0.057$$

$$y_{2,t} = Y(t) - 53.509$$

| COMPLICATED TRANSFER FUNCTION MODEL |        |           |            |         |      |           |            | 37                          |
|-------------------------------------|--------|-----------|------------|---------|------|-----------|------------|-----------------------------|
|                                     |        |           |            |         |      |           |            | 21:46 Sunday, July 25, 1993 |
| OBS                                 | X      | FOR1      | RES1       | STD1    | Y    | FOR2      | RES2       | STD2                        |
| 285                                 | -0.824 | 1.6831E40 | -1.6831E40 | 0.18769 | 52.6 | 7.6751E39 | -7.6751E39 | 0.30897                     |
| 286                                 | -0.740 | 2.3458E40 | -2.3458E40 | 0.18769 | 53.0 | 1.0697E40 | -1.0697E40 | 0.30897                     |
| 287                                 | -0.528 | 3.2693E40 | -3.2693E40 | 0.18769 | 54.3 | 1.4908E40 | -1.4908E40 | 0.30897                     |
| 288                                 | -0.204 | 4.5563E40 | -4.5563E40 | 0.18769 | 56.0 | 2.0777E40 | -2.0777E40 | 0.30897                     |
| 289                                 | 0.034  | 6.3501E40 | -6.3501E40 | 0.18769 | 57.0 | 2.8956E40 | -2.8956E40 | 0.30897                     |
| 290                                 | 0.204  | 8.85E40   | -8.85E40   | 0.18769 | 58.0 | 4.0356E40 | -4.0356E40 | 0.30897                     |
| 291                                 | 0.253  | 1.2334E41 | -1.2334E41 | 0.18769 | 58.6 | 5.6244E40 | -5.6244E40 | 0.30897                     |
| 292                                 | 0.195  | 1.719E41  | -1.719E41  | 0.18769 | 58.5 | 7.8386E40 | -7.8386E40 | 0.30897                     |
| 293                                 | 0.131  | 2.3957E41 | -2.3957E41 | 0.18769 | 58.3 | 1.0925E41 | -1.0925E41 | 0.30897                     |
| 294                                 | 0.017  | 3.3389E41 | -3.3389E41 | 0.18769 | 57.8 | 1.5225E41 | -1.5225E41 | 0.30897                     |
| 295                                 | -0.182 | 4.6533E41 | -4.6533E41 | 0.18769 | 57.3 | 2.1219E41 | -2.1219E41 | 0.30897                     |
| 296                                 | -0.262 | 6.4853E41 | -6.4853E41 | 0.18769 | 57.0 | 2.9573E41 | -2.9573E41 | 0.30897                     |
| 297                                 | .      | 9.0384E41 | .          | 0.18769 | .    | 4.1216E41 | .          | 0.30897                     |
| 298                                 | .      | 2.9944E42 | .          | 0.40618 | .    | 1.0964E42 | .          | 0.49857                     |
| 299                                 | .      | 6.4537E42 | .          | 0.62389 | .    | 2.0865E42 | .          | 0.65106                     |
| 300                                 | .      | 1.4371E43 | .          | 1.27901 | .    | 2.9602E42 | .          | 0.75773                     |
| 301                                 | .      | 3.2508E43 | .          | 2.88980 | .    | 4.532E42  | .          | 0.80471                     |

그림 8.21.8 복잡한 전이함수모형에 의한 예측

$$z_{t+1} = \hat{F}z_t + \hat{G}v_{t+1}$$

$$y_t = [I_2 \ 0]z_t$$

$$\text{단, } z_t = (y_{1,t}, y_{2,t}, y_{1,t+1|t}, y_{2,t+1|t}, y_{1,t+2|t}, y_{2,t+2|t})'$$

$$\hat{F} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0.432 & 0.000 & -1.913 & 0.000 & 2.360 & 0.000 \\ 0.331 & 0.281 & -1.350 & -1.108 & 0.702 & 1.719 \end{bmatrix}, \quad \hat{G} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1.926 & 0.000 \\ 0.000 & 1.300 \\ 2.507 & 0.000 \\ 0.000 & 1.362 \end{bmatrix}$$

$$\hat{\Sigma} = \begin{bmatrix} 0.0353 & -0.0073 \\ -0.0073 & 0.0976 \end{bmatrix}$$

이 상태공간모형식으로부터 2변량VARMA모형을 유도할 수 있다. 그러나 이 VARMA모형을 유도할 필요가 있을 정도로 이 상태공간모형이 주어진 시계열데이터를 잘 나타내고 있는 것은 아니다. 따라서 이 유도를 독자들에게 숙제로 남겨둔다.

여덟 번째 PROC스텝은 다음과 같다.

```
PROC PRINT DATA= EXAM821E(FIRSTOBS=285);
```

```
RUN;
```

이 프로시저는 앞의 PROC스텝의 PROC STATESPACE스테이트먼트의 옵션 OUT=에 지정한 SAS데이터세트 EXAM821E를 OUTPUT윈도우에 출력시키기 위한 것이다. 이 프로시저를 수행한 결과가 그림 8.21.8이다. 그림 8.21.8에서 잔차들의 절대값이 지나치게 큰 것을 알 수 있다. 따라서 이 전이함수모형은 결코 잘 선택된 것이 아니다. 이 예제에서 지침서(manual)에 나온 결과물이라고 해서 반드시 옳은 것은 아니라는 것을 알 수 있다.

SAS프로그램 EXAM8-21.PRG의 세 개의 상태공간모형들 중에서 최적의 모형을 식별하기 위해서 벌칙함수식별방법들을 적용하기로 하자. 이러한 목적으로 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM8-22.PRG에 수록되어 있다).

```

/ * * /
/ * EXAM8-22.PRG * /
/ * * /
/ * CHOOSING RESTRICT CONDITIONS * /
/ * USING THE PENALTY FUNTION METHODS * /
/ * * /
OPTIONS LINESIZE=72 PAGESIZE=60;
PROC IML;
RESET NOLOG NOCENTER NONAME;
T = 296;
/ * WITH 2 RESTRICTIONS * /
K = 15;
COV2 = { 0.03557464 -0.007285,
 -0.007285 0.09560056 };
NORES = 2;
AIC = LOG(DET(COV2)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COV2)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COV2)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT , "NO OF RESTRICTION = " NORES,
 " AIC = " AIC,
 " 2BIC = " BIC,
 " HQC = " HQC;
/ * TRANSFER FUNCTION MODEL WITH 5 RESTRICTIONS * /
K = 15;
COV5 = { 0.03699488 -0.0072039,
 -0.0072039 0.09571217 };
NORES = 5;
AIC = LOG(DET(COV5)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COV5)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COV5)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT , "NO OF RESTRICTION = " NORES,
 " AIC = " AIC,
 " BIC = " BIC,

```

```

SAS
22:17 Sunday, August 1, 1993 3

NO OF RESTRICTION = 2
AIC = -5.611589
BIC = -5.449512
HQC = -5.546697

NO OF RESTRICTION = 5
AIC = -5.590583
BIC = -5.465909
HQC = -5.540666

NO OF RESTRICTION = 7
AIC = -5.622893
BIC = -5.460817
HQC = -5.558001

Exiting IML.

```

그림 8.22 벌칙함수식별방법들에 의한 전이함수모형들의 비교

```

" HQC = " HQC;
/* TRANSFER FUNCTION MODEL WITH 7 RESTRICTIONS */
K = 20;
COV7 = { 0.03522668 -0.007253,
 -0.007253 0.09546338 };
NORES = 7;
AIC = LOG(DET(COV7)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COV7)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COV7)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT, "NO OF RESTRICTION = " NORES,
" AIC = " AIC,
" BIC = " BIC,
" HQC = " HQC,;
QUIT;

```

이 SAS프로그램을 수행한 결과가 그림 8.22이다. 그림 8.22에서 AIC방법과 HQC방법은 제한조건이 일곱 개인 복잡한 형태의 전이함수모형을 선택하고, BIC방법은 제한조건이 다섯 개인 전이함수모형을 선택한다. 그림 8.21.8에서 볼 수 있듯이 제한조건이 일곱 개인 복잡한 형태의 전이함수모형은 예측의 관점에서 볼 때 결코 좋은 것은 아니다. 따라서 모형의 선택은 어느 한 가지 기준에 전적으로 의존할 것이 아니라는 것을 알 수 있다.

## 8.5 수리적 배경

이 절에서는 지금까지 설명한 상태공간모형화방법의 수학적 배경 및 통계학적 배경을 설명하고자 한다. 따라서 이 절을 이해하기 위해서는 어느 정도의 수학적 지식 및 통계학적 지식이 있어야 한다. 단순히 상태공간모형을 사용해서 다변량시계열데이터를 분석하고자 한다면 이 절을 꼭 읽을 필요는 없다. 이 절의 목적은 상태공간모형화방법을 잘 이해해서 상태공간모형을 사용한 VARMA모형분석을 좀더 정확하게 하도록 하는 것이다.

앞에서도 여러 번 언급했듯이 VARMA모형의 상태공간모형화방법은 Akaike에 의해서 개발되었다. 따라서 이 방법을 잘 이해하기 위해서는 Akaike의 논문들을 잘 이해해야 한다. Akaike(1973)는 정규성가정하에서 VARMA모형의 최우추정법을 제시하였다. 이 방법은 스펙트럴밀도함수를 바탕으로 한 것으로써 VARMA과정의 상태공간모형화의 제3단계인 최종추정단계에서 상태공간모형의 우도함수를 최대화하는 데 적용된다. Akaike(1974a)는 VARMA과정의 상태공간모형화방법을 체계적으로 제시한 최초의 논문이다. SAS/ETS의 PROC STATESPACE는 단지 이 논문의 내용을 프로그램한 것에 불과하다고 해도 과언이 아니다. 따라서 이 논문은 제 8.3절의 내용의 대부분을 포함하고 있다. Akaike(1974b)는 정규성가정하에서 가장 자연스러운(most natural) 상태공간이 예측공간임을 보였다. 이는 정규성가정하에서 정리 8.1의 상태공간모형이 가장 자연스러운 형태임을 보인 것이다. Akaike(1976)는 VARMA과정의 상태공간모형화방법을 체계적으로 제시한 Akaike(1974a)를 수리적으로 보강해서 자세히 설명한 논문이다. [ETS6.1]의 제 18장에서도 PROC STATESPACE에 관한 자세한 이론은 이 논문을 참조할 것을 권하고 있다. Akaike(1978)는 VARMA과정의 상태공간모형화방법을 포함한 제어이론(control theory)에서 사용하는 상태공간모형의 식별을 간단히 설명하고 있다.

지금부터는 제 8.3절에서 제시된 정리들을 증명하고 그 외에 상태공간모형화방법을 이해하는 데 필요한 수리적인 내용을 설명하고자 한다.

### 8.5.1 VARMA모형의 상태공간모형화

앞에서 언급하였듯이 여기서 다루어지는 VARMA모형은 작인성조건과 가역성조건을 만족한다고 가정한다. 다시 설명하면 VARMA과정  $\{y_t\}$ 가 작인적이기 위해서는 이 다변량시계열데이터의 VAR특성방정식  $\det[\Phi(x)] = 0$ 의 모든 근이 단위원 밖에 존재해야 한다. 또한 이 VARMA과정이 가역적이기 위해서는 이 다변량시계열데이터의 VMA특성방정식  $\det[\Theta(x)] = 0$ 의 모든 근이 단위원 밖에 존재해야 한다.

다음 정리는 정리 8.1을 확장한 것이다.

**정리 8.7** Akaike(1974a)

작인성조건과 가역성조건을 만족하는 VARMA모형은 상태공간표현(state space representation)으로 나타낼 수 있고, 임의의 상태공간모형을 VARMA모형으로 나타낼 수 있다.  $\square$

(증명) 우선  $d$ 변량VARMA과정  $\{y_t\}$ 를 상태공간모형으로 나타내보자. 제 6.6.2소절에서 보았듯이 작인성가정에 의하여  $y_t$ 는 다음과 같은 VMA( $\infty$ )모형으로 나타낼 수 있다.

$$y_t = \sum_{j=0}^{\infty} \Psi_j v_{t-j} \quad (\Psi_0 = I_d)$$

여기서  $\Psi_j$ 는 입력시계열데이터가  $d$ 변량백색잡음과정  $\{v_t\}$ 이고 출력시계열데이터가  $\{y_t\}$ 인 선형시스템에서 시차  $j$ 의 충격반응가중행렬(impulse response weight matrix)이다. 이 식으로부터 다음 식이 성립한다.

$$y_{t+i|t} = \sum_{j=i}^{\infty} \Psi_j v_{t+i-j}$$

이 식으로부터 다음과 같은 식을 얻는다.

$$\begin{aligned} y_{t+i|t+1} &= \sum_{j=i-1}^{\infty} \Psi_j v_{t+i-j} \\ &= \sum_{j=i}^{\infty} \Psi_j v_{t+i-j} + \Psi_{i-1} v_{t+1} \\ &= y_{t+i|t} + \Psi_{i-1} v_{t+1} \end{aligned}$$

최량예측의 정의로부터 다음 식들이 성립함을 알 수 있다.

$$\begin{aligned} y_{t+h|t} &= y_t \quad (h=0, -1, -2, \dots) \\ v_{t+h|t} &= 0 \quad (h=1, 2, \dots) \end{aligned}$$

이후 임의의 VARMA( $p, q$ )과정에서  $p < q$ 이면 다음 등식이 성립한다고 가정하자.

$$\Psi_{p+1} = \Psi_{p+2} = \dots = \Psi_q = 0$$

즉,  $p > q$ 라 가정하자. 이러한 가정하에서 다음 식이 성립한다.

$$y_{t+p|t} = \Phi_1 y_{t+p-1|t} + \Phi_2 y_{t+p-2|t} + \dots + \Phi_p y_{t|t}$$

따라서 미래의 예측  $y_{t+p+i|t}$  ( $i=0, 1, 2, \dots$ )는  $y_{t|t}, y_{t+1|t}, \dots, y_{t+p-1|t}$ 의 선형결합으로 나타낼 수 있으므로 상태벡터의 후보로 다음과 같이 벡터  $w_t$ 를 정의하자.

$$w_t = (y_{t|t}', y_{t+1|t}', \dots, y_{t+p-1|t}')'$$

앞의 식들을 정리하면 이 VARMA( $p, q$ )과정은 다음과 같은 마아코프표현으로 나타낼 수 있다.

$$w_{t+1} = Fw_t + Gv_{t+1}$$

$$y_t = [I_d \ 0 \ \cdots \ 0 \ 0] w_t$$

$$\text{단, } F = \begin{bmatrix} 0 & I_d & 0 & \cdots & 0 & 0 \\ 0 & 0 & I_d & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & I_d \\ \Phi_p & \Phi_{p-1} & \Phi_{p-2} & \cdots & \Phi_2 & \Phi_1 \end{bmatrix}, \quad G = \begin{bmatrix} I_d \\ \Psi_1 \\ \vdots \\ \Psi_{p-2} \\ \Psi_{p-1} \end{bmatrix}$$

여기서 유의할 점은  $w_t$ 가 상태벡터가 아닐 수도 있다는 것이다. 그 이유는 비록  $w_t$ 가 미래의 벡터들을 예측할 수 있는 모든 정보를 갖기는 했지만 이 정보량이 미래의 모든 벡터들을 예측하기 위한 최소량이라고 할 수는 없기 때문이다. 이  $w_t$ 가 상태벡터이기 위해서는  $w_t$ 의 원소들이 선형독립이어야 한다. 따라서 상태벡터  $z_t$ 를 다음과 같이 정의하자. 우선  $w_t$ 의 성분들로 이루어진 집합의 부분집합들 중에서  $y_t$ 의  $d$ 개의 성분들을 포함하면서 원소들이 선형독립인 부분집합들 중에서 원소의 개수가 가장 작은 부분집합들 중의 하나를  $Z$ 라 하자. 상태벡터  $z_t$ 의 성분들은 집합  $Z$ 의 원소들을 나열한 것으로서 첫  $d$ 개의 성분들은  $y_t$ 의  $d$ 개의 성분들을 나열한 것이다. 만일  $w_t$ 의 차원이  $a$ 이고  $z_t$ 의 차원이  $b$ 라면 위의 마아코프표현에서  $w_t$ 의 성분들 중에서  $z_t$ 의 성분이 아닌  $(b-a)$ 개의 성분들에 해당하는 열들을 제거하면 상태공간표현이 된다.

지금부터는 상태공간모형을 VARMA모형으로 나타내자. 다변량시계열데이터  $\{y_t\}$ 가 다음과 같은 상태공간모형을 갖는다고 하자.

$$z_{t+1} = Fz_t + Gv_{t+1}$$

$$y_t = Hz_t$$

상태벡터  $z_t$ 의 차원을  $n$ 이라 하고 변환행렬  $F$ 의 특성방정식을 다음과 같이 표현하자.

$$\det[\lambda I - F] = \sum_{i=0}^n \phi_i \lambda^{n-i} \quad (\phi_0 = 1)$$

여기서  $\phi_0 (=1), \phi_1, \dots, \phi_n$ 은 스칼라들임에 유의하라. Cayley-Hamilton정리에 의해서 다음 식이 성립한다.

$$\sum_{i=0}^p \phi_i F^{n-i} = 0$$

상태방정식을 축차적으로 대입하면 다음 식을 얻을 수 있다.

$$z_{t+i} = F^i z_t + F^{i-1} G v_{t+1} + \cdots + F G v_{t+i-1} + G v_{t+i}$$

이 식과 출력방정식으로부터 다음 식들을 얻을 수 있다.



$$\begin{aligned}
y_{t+n} &= H(F^n z_t + F^{n-1} G v_{t+1} + \dots + G v_{t+n}) \\
\phi_1 y_{t+n-1} &= H \phi_1 (F^{n-1} z_t + F^{n-2} G v_{t+1} + \dots + G v_{t+n-1}) \\
&\vdots \\
\phi_{n-1} y_{t+1} &= H \phi_{n-1} (F z_t + G v_{t+1}) \\
\phi_n y_t &= H \phi_n z_t
\end{aligned}$$

이 식들로부터 다음 식이 성립한다.

$$\begin{aligned}
& y_{t+n} + \phi_1 y_{t+n-1} + \dots + \phi_{n-1} y_{t+1} + \phi_n y_t \\
&= H(F^n + \phi_1 F^{n-1} + \dots + \phi_{n-1} F + \phi_n I) z_t \\
&\quad + H(F^{n-1} + \phi_1 F^{n-2} + \dots + \phi_{n-1} I) G v_{t+1} \\
&\quad \vdots \\
&\quad + H(F + \phi_1 I) G v_{t+n-1} \\
&\quad + H G v_{t+n} \\
&= H(F^{n-1} + \phi_1 F^{n-2} + \dots + \phi_{n-1} I) G v_{t+1} \\
&\quad \vdots \\
&\quad + H(F + \phi_1 I) G v_{t+n-1} \\
&\quad + H G v_{t+n}
\end{aligned}$$

다음과 같은 행렬들을 정의하자.

$$\theta_i = H(F^i + \phi_1 F^{i-1} + \dots + \phi_i I) G \quad (i = 0, 1, \dots, n-1)$$

이 정의로부터 다음 식이 성립한다.

$$\begin{aligned}
& y_{t+n} + \phi_1 y_{t+n-1} + \dots + \phi_{n-1} y_{t+1} + \phi_n y_t \\
&= \theta_0 v_{t+n} + \theta_1 v_{t+n-1} + \dots + \theta_{n-1} v_{t+1}
\end{aligned}$$

이 식으로부터 상태공간표현을 갖는 임의의 다변량확률과정을 VARMA모형으로 표현할 수 있음을 알 수 있다. Q.E.D.

## 8.5.2 VAR모형단계의 이론적 배경

내적공간  $L_t(y)$ 는 확률과정  $\{y_{t+h|h} = \dots, -1, 0, 1, \dots\}$ 에 의해 생성되는 힐버트공간이고,  $L_t(\text{past})$ 는 현재와 과거의 원소들  $y_t, y_{t-1}, y_{t-2}, \dots$ 으로 생성되는  $L_t(y)$ 의 부분공간이며,  $L_t(\text{future})$ 는 현재와 미래의 원소들  $y_t, y_{t+1}, y_{t+2}, \dots$ 으로 생성되는  $L_t(y)$ 의 부분공간이다. 또한  $y_t | \text{future}$ 와  $y_t | \text{past}$ 는 각각  $y_t$ 의  $L_t(\text{future})$ 와  $L_t(\text{past})$ 로의 사영(projection)들이고,  $L_t(\text{past} | \text{future})$ 는  $y_t' | \text{future}$ ,  $y_{t-1}' | \text{future}$ ,  $y_{t-2}' | \text{future}$ ,  $\dots$ 으로 생성된  $L_t(\text{future})$ 의 부분공간이며,  $L_t(\text{future} | \text{past})$ 는  $y_t' | \text{past}$ ,  $y_{t+1}' | \text{past}$ ,  $y_{t+2}' | \text{past}$ ,  $\dots$ 으로 생성된  $L_t$

(past)의 부분공간이다.

Akaike(1974b)는 예측공간이 유한차원이기 위한 필요충분조건은 자기공분산행렬함수가 유한차원 인자분해(finite dimensional factorization)되는 것과 동일함을 보였다.

**정리 8.8** Akaike(1974b)

자기공분산행렬함수가 다음과 같이 유한차원 인자분해되는 것은 예측공간이 유한차원이기 위한 필요충분조건이다.

$$\Gamma(j) = HF^jD \quad (j=0, 1, 2, \dots)$$

$$\text{단, } P = E(z_t z_t') \quad (D=PH')$$

(증명) 우선 자기공분산행렬함수가 유한차원 인자분해되면 예측공간이 유한차원임을 보이자. 자기공분산행렬함수가 다음과 같이 유한차원 인자분해된다고 가정하자.

$$\Gamma(j) = HF^jD \quad (j=0, 1, 2, \dots)$$

변환행렬  $F$ 의 최소다항방정식(minimal polynomial equation)이 다음과 같다고 하자.

$$\lambda^r + \sum_{i=1}^r A_i \lambda^{r-i} = 0 \quad (r \leq n)$$

Cayley-Hamilton정리에 의해 모든  $j(\geq 0)$ 에 대해 다음 식이 성립한다.

$$H(F^r + \sum_{i=1}^r A_i F^{r-i})D = 0$$

이 식의 양변에  $F^j$ 를 곱하면 다음 식이 성립한다.

$$HF^{r+j}D + \sum_{i=1}^r A_i HF^{r+j-i}D = 0 \quad (j=0, 1, 2, \dots)$$

이 식으로부터 다음 식이 성립한다.

$$\Gamma(r+j) + \sum_{i=1}^r A_i \Gamma(r+j-i) = 0 \quad (j=0, 1, 2, \dots)$$

따라서 다음 식들이 성립한다.

$$E[y_{t+n}(y_{t-r-m} - \sum_{i=1}^r A_i y_{t-r-m+i})'] = 0 \quad (m, n = 0, 1, 2, \dots)$$

$$E[(y_{t+r+m} - \sum_{i=1}^r A_i y_{t+r+m+i})' y_{t-n}] = 0 \quad (m, n = 0, 1, 2, \dots)$$

이 식들을 정리하면 다음과 같다.

$$y_{t-r-m} | \text{future}_t = \sum_{i=1}^r A_i y_{t-r-m+i} | \text{future}_t \quad (m = 0, 1, 2, \dots)$$

$$y_{t+r+m} | \text{past}_t = \sum_{i=1}^r A_i y_{t+r+m-i} | \text{past}_t \quad (m = 0, 1, 2, \dots)$$

첫 번째 식으로부터  $L_t(\text{past} | \text{future})$ 는 유한차원이며 벡터  $(y_t' | \text{future}, y_{t-1}' | \text{future},$

...  $y_{t-r+1} | \text{future}_t$ '의 원소들로부터 생성됨을 알 수 있다. 즉, 이 벡터의 원소들은  $L_t(\text{past} | \text{future})$ 의 기저를 이룬다. 두 번째 식으로부터  $L_t(\text{future} | \text{past})$ 가 유한차원이며 벡터  $(y_t' | \text{past}, y_{t+1}' | \text{past}, \dots, y_{t+r-1}' | \text{past}_t)$ '의 원소들로부터 생성됨을 알 수 있다. 즉, 이 벡터의 원소들은  $L_t(\text{future} | \text{past})$ 의 기저를 이룬다. 이 두 부분공간들  $L_t(\text{past} | \text{future})$ 와  $L_t(\text{future} | \text{past})$  중에서  $L_t(\text{future} | \text{past})$ 를 예측공간이라 부른다. 따라서 예측공간  $F_t = L(y_t, y_{t+1}, y_{t+2}, \dots)$ 는 유한차원공간이다.

지금부터는 예측공간이 유한차원이면 자기공분산행렬함수의 유한차원 인자분해가 존재함을 보이자. 예측공간  $L_t(\text{future} | \text{past})$ 의 차원을  $k$ 라고 가정하고 하나의 기저를 선택하여  $z_{t1}, z_{t2}, \dots, z_{tk}$ 라고 하자. 만일  $z_t = (z_{t1}, z_{t2}, \dots, z_{tk})'$ 라고 하면  $y_t | \text{past}_t$ 는  $z_t$ 의 선형변환으로 나타낼 수 있다. 즉, 다음 관계식을 만족하는  $d \times k$ 행렬  $H$ 가 존재한다.

$$y_t | \text{past}_t = Hz_t$$

따라서 다음 출력방정식이 성립한다.

$$y_t = Hz_t + \zeta_t$$

여기서  $\zeta_t$ 는 측정오차벡터(measurement error vector), 또는 교란벡터(disturbance vector)라고 부른다. 이 측정오차벡터는  $z_t$ 와 독립인 다변량백색잡음과정이다. 만일  $v_{t+1}$ 을 시점  $(t+1)$ 에서  $y_{t+1}$ 의 쇠신벡터라고 하면 다음 식이 성립한다.

$$v_{t+1} = y_{t+1} - y_{t+1} | \text{past}_t$$

정의에 의해서  $z_{t+1}$ 은  $z_t$ 로 설명할 수 있는 부분과  $z_t$ 로 설명할 수 없는 부분이 있다. 따라서 다음 식을 만족하는 행렬들  $F$ 와  $G$ 가 존재한다.

$$z_{t+1} = Fz_t + Gv_{t+1}$$

이 식이 상태방정식이다. 따라서  $y_t$ 의 상태공간표현이 결정되었다. 이 상태공간표현을 이용하여 다음 식들을 유도할 수 있다.

$$\begin{aligned} & \Gamma_z(j) \\ &= E(z_{t+j} z_t') \\ &= E(Fz_{t+j-1} + Gv_{t+j}) z_t' \\ &= F\Gamma_z(j) + GE(v_{t+j} z_t') \\ &= F\Gamma_z(j) \quad (j=1, 2, 3, \dots) \end{aligned}$$

따라서 다음 식이 성립한다.

$$\Gamma_z(j) = F^j \Gamma_z(0) \quad (j=0, 1, 2, \dots)$$

이 식을 이용하면 다음 식들을 유도할 수 있다.

$$\begin{aligned}
 & \Gamma(j) \\
 &= E(y_{t+j}y_t') \\
 &= E(Hz_{t+j} + \xi_{t+j})(Hz_t + \xi_t)' \\
 &= H\Gamma_z(j)H' \\
 &= HF^j\Gamma_z(0)H' \\
 &= HF^jPH' \\
 &= HF^jD \qquad (j=0, 1, 2, \dots)
 \end{aligned}$$

이 식이 자기공분산행렬함수의 유한차원 인자분해이다. Q. E. D.

앞에서 언급했듯이 두 부분공간들  $L_t(\text{past} | \text{future})$ 와  $L_t(\text{future} | \text{past})$  중에서  $L_t(\text{future} | \text{past})$ 를 예측공간이라 부른다. 정리 8.7로부터 상태공간표현은 예측공간으로부터 결정되며 상태공간의 기저로서 최소 마야코비안표현을 할 수 있음을 알았다. 만일 시스템의 차원이  $p(=m+1)$ 이라면 예측공간의 기저를 찾는 것은  $y_{t+1}, y_{t+2}, \dots, y_{t+m}$ 에 의해 생성되는 예측공간의 범위내로 제한할 수 있다. 제 8.2.2 소절에서 기술했듯이 PROC STATESPACE에서는 AIC방법을 사용해서 차수  $m$ 을 결정한다.

정리 8.7로부터 다음 정리를 쉽게 유도할 수 있다.

**정리 8.9** 예측공간이 유한차원이면 다음 식을 만족하는 자연수  $r(r \leq m)$ 과 계수행렬들  $A_1, A_2, \dots, A_r$ 이 존재한다.

$$\Gamma(r+j) = \sum_{i=1}^r A_i \Gamma(r+j-i) \qquad (j=0, 1, 2, \dots) \quad \blacksquare$$

(증명) 정리 8.7에 의해서 예측공간이 유한차원이면 자기공분산행렬함수는 다음과 같이 유한차원 인자분해된다.

$$\Gamma(j) = HF^jD \qquad (j=0, 1, 2, \dots)$$

따라서 정리 8.7의 증명에서와 같이 Cayley-Hamilton정리를 사용해서 다음 식을 유도할 수 있다.

$$\Gamma(r+j) = \sum_{i=1}^r A_i \Gamma(r+j-i) \qquad (j=0, 1, 2, \dots) \qquad \text{Q. E. D.}$$

### 8.5.3 정준상관분석단계의 이론적 배경

상태공간모형화의 정준상관분석단계를 이해하기 위해서는 우선 다변량통계분석기법의 하나인 정준상관분석을 이해하고 있어야 한다. 정준상관분석(canonical correlation analysis)은 한 개 이상의 변수들로 구성된 변수 집단들 사이의 상관관계를 분석하는 다변량통계분석기법이다. 즉, 정준상관분석이란 기준변수들  $y_1, y_2, \dots, y_r$ 의 선형결합들과 설명변수

들  $x_1, x_2, \dots, x_p$ 의 선형결합들 사이의 상관계수들 중에서 절대값이 최대인 선형결합들을 찾아내어, 주어진 두 변수집단들 사이에 어떠한 통계적 관계가 있는지를 알아내는 다변량통계분석기법이다.

확률벡터  $(x_1, x_2, \dots, x_p, y_1, y_2, \dots, y_q)'$ 가 다음과 같은 다변량정규분포를 갖는다고 가정하자.

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}\right)$$

여기서  $x = (x_1, x_2, \dots, x_p)'$ ,  $y = (y_1, y_2, \dots, y_q)'$ 이다. 또한 공분산행렬을 다음과 같이 표기하자.

$$\Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}$$

이 공분산행렬  $\Sigma$ 는 양정치행렬(positive definite matrix)이다. 이후  $n$ 을 다음과 같이 정의하자.

$$n = p + q$$

확률벡터들  $x$ 와  $y$ 의 임의의 선형결합을 다음과 같이 정의하자.

$$\begin{aligned} u &= \alpha'x \\ v &= \gamma'y \end{aligned}$$

여기서  $u$ 와  $v$ 는 각각 기준변수와 설명변수의 정준상관득점(canonical scores)을,  $\alpha$ 는 설명변수집단의 정준가중벡터(canonical weight vector)를, 그리고  $\gamma$ 는 기준변수집단의 정준가중벡터를 의미한다. 우리의 목적은 변수집단들의 속성을 가장 잘 나타내는 선형결합들을 찾고자 하는 것이고, 이는 정준가중벡터들  $\alpha$ 와  $\gamma$ 를 어떻게 결정하느냐는 문제로 귀착된다. 이 문제에 대한 해답으로 정준상관분석에서는  $u$ 와  $v$ 의 단순상관계수(Pearson correlation coefficient)의 절대값을 가장 크게 하는  $\alpha$ 와  $\gamma$ 를 선택한다.

지금부터는 최대상관계수를 갖는 선형결합들을 찾아보기로 하자. 여기서 확률변수  $u$ 에 상수를 곱한 확률변수와 확률변수  $v$ 에 상수를 곱한 확률변수의 상관계수의 절대값은 확률변수들  $u$ 와  $v$ 의 상관계수의 절대값과 같으므로  $\alpha$ 와  $\gamma$ 를 정규화시킬 필요가 있다. 따라서 확률변수들  $u$ 와  $v$ 의 분산이 1이 되게 하는  $\alpha$ 와  $\gamma$ 를 선택한다. 즉,  $\alpha$ 와  $\gamma$ 는 다음 식들을 만족한다고 가정한다.

$$\begin{aligned} E(u^2) &= E(\alpha'xx'\alpha) = \alpha'\Sigma_{xx}\alpha = 1 \\ E(v^2) &= E(\gamma'yy'\gamma) = \gamma'\Sigma_{yy}\gamma = 1 \end{aligned}$$

여기서  $E(u)=0$ ,  $E(v)=0$ 이다. 이 가정하에서 확률변수들  $u$ 와  $v$ 의 상관계수는 다음과 같다.

$$E(uv) = E(\alpha'xy'\gamma) = \alpha'\Sigma_{xy}\gamma$$

확률변수들  $u$ 와  $v$ 의 상관계수를 최대화하는  $\alpha$ 와  $\gamma$ 를 찾아내기 위해서 Lagrange승수법을 사용하기로 하고 다음 식을 정의하자.

$$\varphi = \alpha' \Sigma_{xy} \gamma - \frac{1}{2} \lambda (\alpha' \Sigma_{xx} \alpha - 1) - \frac{1}{2} \mu (\gamma' \Sigma_{yy} \gamma - 1)$$

여기서  $\lambda$ 와  $\mu$ 는 Lagrange승수들이다. 함수  $\varphi$ 를  $\alpha$ 와  $\gamma$ 로 미분하여 0으로 놓은 정규방정식을 풀면 다음과 같은 최대화의 제1차조건식들을 얻는다.

$$\begin{aligned} \Sigma_{xy} \gamma - \lambda \Sigma_{xx} \alpha &= 0 \\ \Sigma_{xy}' \alpha - \mu \Sigma_{yy} \gamma &= 0 \end{aligned}$$

이 식들에 확률변수들  $u$ 와  $v$ 의 분산들이 1인 것을 적용하면 다음 식이 성립한다.

$$\lambda = \mu = \alpha' \Sigma_{xy} \gamma$$

따라서 제1차조건식들을 다음과 같은 행렬방정식으로 표현할 수 있다.

$$\begin{bmatrix} -\lambda \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & -\lambda \Sigma_{yy} \end{bmatrix} \begin{bmatrix} \alpha \\ \gamma \end{bmatrix} = 0$$

이 행렬방정식이 영벡터가 아닌 해(nontrivial solution)를 갖기 위해서는 다음 식이 성립해야 한다.

$$\det \begin{bmatrix} -\lambda \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & -\lambda \Sigma_{yy} \end{bmatrix} = 0$$

편의상 이 방정식을 정준상관분석의 특성방정식이라 하자. 이 방정식이  $n$ 차다항식이므로  $n$ 개의 근, 즉  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ 을 갖는다. 여기서 유의할 점은 만일  $\lambda$ 가 정준상관분석의 특성방정식의 근이면 다음 식들이 성립한다.

$$\begin{aligned} \Sigma_{xy} \gamma - (-\lambda) \Sigma_{xx} (-\alpha) &= 0 \\ \Sigma_{xy}' (-\alpha) - (-\mu) \Sigma_{yy} \gamma &= 0 \\ \lambda = \mu = (-\alpha)' \Sigma_{xx} \gamma &= 0 \end{aligned}$$

즉, 다음 행렬방정식이 성립한다.

$$\begin{bmatrix} -\lambda \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & -\lambda \Sigma_{yy} \end{bmatrix} \begin{bmatrix} -\alpha \\ \gamma \end{bmatrix} = 0$$

따라서  $(-\lambda)$ 도 정준상관분석의 특성방정식의 근이다. 즉,  $d = \min(p, q)$ 라고 정의하면 정준상관분석의 특성방정식의 근들은 다음과 같다.

$$\lambda_1, \lambda_2, \dots, \lambda_d, 0, \dots, 0, -\lambda_d, -\lambda_{d-1}, \dots, -\lambda_1$$

따라서 이후 정준상관분석의 특성방정식의 근의 절대값을 최대로 하는 선형결합을 찾는

대신 정준상관분석의 특성방정식의 근을 최대로 하는 선형결합을 찾기로 한다.

물론 정준상관분석의 특성방정식의 근  $\lambda$ 는 확률변수들  $u=\alpha'x$ 와  $v=\gamma'y$ 의 상관계수이다. 따라서 최대상관계수를 갖는 선형결합들을 찾기 위해서  $\lambda=\lambda_1$ 으로 선택한다. 이 상관계수  $\lambda_1$ 을 확률벡터들  $x$ 와  $y$ 의 첫 번째 상관계수라고 한다. 또한  $\lambda=\lambda_1$ 에 대응하는 해를  $\alpha^{(1)}$ 과  $\gamma^{(1)}$ 이라 하면,  $u_1=\alpha^{(1)'x}$ ,  $v_1=\gamma^{(1)'y}$ 의 관계가 성립한다. 따라서 확률변수들  $u_1$ 과  $v_1$ 은 최대상관계수를 갖는 확률벡터들  $x$ 와  $y$ 의 선형결합들이다.

지금부터는 확률벡터  $x$ 의 두 번째 선형결합  $u=\alpha'x$ 와 확률벡터  $y$ 의 두 번째 선형결합  $v=\gamma'y$ 를 찾아보자. 이 두 번째 선형결합들  $u_2$  및  $v_2$ 는 첫 번째 선형결합들  $u_1$  및  $v_1$ 과의 상관계수가 0인 모든 선형결합들 중에서 확률변수들  $u_2$ 와  $v_2$ 의 상관계수가 최대가 되는 선형결합들이다. 이러한 과정을 반복해서 제  $r$ 번째 단계에서 선형결합들  $u_1=\alpha^{(1)'x}$ ,  $v_1=\gamma^{(1)'y}$ , ...,  $u_r=\alpha^{(r)'x}$ ,  $v_r=\gamma^{(r)'y}$ 를 얻었다고 하자. 이 선형결합들에 대응하는 정준상관분석의 특성방정식의 근들은  $\lambda^{(1)}=\lambda_1$ ,  $\lambda^{(2)}=\lambda_2$ , ...,  $\lambda^{(r)}=\lambda_r$ 이다. 이 근들을 정준상관계수들이라 부른다. 이 절차를 정준상관분석의 특성방정식의 음이 아닌 근에 대해서 제 1차조건식을 만족하는 해를 모두 찾을 때까지 계속한다. 즉, 확률벡터  $(x', y')$ 의 공분산행렬에서 첫 선형결합들로 설명되지 못한 부분으로부터 두 번째 선형결합들을 도출하고 선택된 선형결합들의 효과를 제외한 것을 대상으로 선형결합들을 계속 도출하는데 도출가능한  $u_i$ 들(또는,  $v_i$ 들)의 최대 개수는 설명변수들의 개수와 기준변수들의 개수 중에서 작은 값인  $d$ 이다. 여기서  $(u_r, v_r) = (\alpha^{(r)'x}, \gamma^{(r)'y})$ 를 제  $r$ 번째 정준변량(canonical variate)의 쌍이라고 한다. 이 제  $r$ 번째 정준변량의 쌍은 처음  $(r-1)$ 개 정준변량들의 쌍과 상관되지 않은 변량들 중에서 최대상관계수를 가지며, 제  $r$ 번째 정준상관계수는 정준상관분석의 특성방정식의 근들 중에서 제  $r$ 번째로 큰 근이다.

정준상관분석의 여러 정리들 중에서 상태공간모형화에 사용하는 중요한 정리는 다음과 같다.

**정리 8.10** 정규확률벡터(normal random vector)  $x$ 의 차원이 정규확률벡터  $y$ 의 차원보다 크거나 같고 두 확률벡터들이 비퇴화(nondegenerate)하면, 조건부확률벡터  $y|x$ 의 차원은 확률벡터  $y$ 와 확률벡터  $x$ 의 정준상관계수들 중에서 0이 아닌 정준상관계수의 개수와 같다. ▣

(증명) 다음 정준상관분석의 특성방정식을 살펴보자.

$$\det \begin{bmatrix} -\lambda \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & -\lambda \Sigma_{yy} \end{bmatrix} = 0$$

가정에 의해서 행렬  $\Sigma_{xx}$ 는 정칙이므로, 만일  $\lambda \neq 0$ 인 경우에는 다음 식이 성립한다.

$$\det [\lambda \Sigma_{xx}] \det [\lambda \Sigma_{yy} - \Sigma_{yx} (\lambda \Sigma_{xx})^{-1} \Sigma_{xy}] = 0$$

따라서 다음 식이 성립한다.

$$\det [\lambda \Sigma_{yy} - \Sigma_{yx} (\lambda \Sigma_{xx})^{-1} \Sigma_{xy}] = 0$$

가정에 의해서 행렬  $\Sigma_{yy}$ 가 정칙(nonsingular)이므로 다음 식을 얻는다.

$$\det [\lambda^2 I - \Sigma_{yy}^{-1/2} \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yy}^{-1/2}] = 0$$

이 식과  $\text{rank}(A) = \text{rank}(AA')$ 라는 성질을 이용하면 다음 식들이 성립함을 알 수 있다.

$$\begin{aligned} & \text{number of nonzero canonical correlations of } y \text{ and } x \\ &= \text{number of nonzero } \lambda \text{ (or } \lambda^2) \\ &= \text{number of nonzero eigenvalues of } \Sigma_{yy}^{-1/2} \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yy}^{-1/2} \\ &= \text{rank}(\Sigma_{yy}^{-1/2} \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yy}^{-1/2}) \\ &= \text{rank}(\Sigma_{yy}^{-1/2} \Sigma_{yx} \Sigma_{xx}^{-1/2}) \\ &= \text{rank}(\Sigma_{yx}) \end{aligned}$$

만일  $x$ 를 설명벡터,  $y$ 를 기준벡터라고 하면 다변량정규분포이론에 의해서 다음 식이 성립한다.

$$\text{var}(y|x) = \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy}$$

따라서 다음 등식이 성립한다.

$$\begin{aligned} & \text{dimension of } y|x \\ &= \text{rank}(\Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy}) \\ &= \text{rank}(\Sigma_{yx} \Sigma_{xx}^{-1/2}) \\ &= \text{rank}(\Sigma_{yx}) \end{aligned}$$

이 식으로부터 조건부확률벡터  $y|x$ 의 차원은  $\Sigma_{yx}$ 의 랭크와 같다는 것을 알 수 있다. 따라서 조건부확률벡터  $y|x$ 의 차원은 확률벡터  $y$ 와 확률벡터  $x$ 의 정준상관계수들 중에서 0이 아닌 정준상관계수들의 개수와 같다. Q.E.D.

상태벡터를 선택할 때 사용하는 중요한 성질의 하나는 만일  $y_{i,t+j|t}$ 가 상태벡터에서 제외되면  $y_{i,t+n|t}$  ( $n=j+1, j+2, \dots$ )도 앞으로의 상태벡터의 원소로서의 고려대상에서 제외된다는 것이다. 왜냐하면  $y_{i,t+j|t}$ 를 전항들의 선형결합으로 나타낼 수 있으면  $y_{i,t+n|t}$  또한 전항들의 선형결합으로 나타낼 수 있기 때문이다.

**정리 8.11** 만일  $y_{i,t+j|t}$ 가 전항들의 선형결합으로 나타낼 수 있으면, 임의의  $n(>j)$ 에 대해서  $y_{i,t+n|t}$ 도 전항들의 선형결합으로 나타낼 수 있다. ▣

(증명) 가정에 의해서 다음 식을 만족하는 상수열  $\{a_{j,m}\}$ 이 존재한다.

$$y_{i,t+j|t} = \sum_{n=0}^{i-1} \sum_{m=1}^d a_{n,m} y_{m,t+n|t} + \sum_{m=1}^{i-1} a_{j,m} y_{m,t+j|t}$$



따라서 다음 식이 성립한다.

$$y_{i,t+j} = \sum_{n=0}^{i-1} \sum_{m=1}^d a_{n,m} y_{m,t+n|t} + \sum_{m=1}^{i-1} a_{j,m} y_{m,t+j|t} + v_{1,t+j} + \sum_{n=0}^{i-1} \sum_{m=1}^d b_{n,m} v_{m,t+n}$$

이 식으로부터 다음 식이 성립한다.

$$y_{i,t+j+1} = \sum_{n=0}^{i-1} \sum_{m=1}^d a_{n,m} y_{m,t+n+1|t} + \sum_{m=1}^{i-1} a_{j,m} y_{m,t+j+1|t} + v_{i,t+j+1} + \sum_{n=0}^{i-1} \sum_{m=1}^d b_{n,m} v_{m,t+n+1}$$

따라서 다음 식이 성립한다.

$$y_{i,t+j+1|t} = \sum_{n=0}^{i-1} \sum_{m=1}^d a_{n,m} y_{m,t+n+1|t} + \sum_{m=1}^{i-1} a_{j,m} y_{m,t+j+1|t}$$

즉,  $y_{i,t+j+1|t}$ 는 전항들의 선형결합으로 나타낼 수 있다. 따라서 수학적귀납법에 의해서 임의의  $n(>j)$ 에 대해서  $y_{i,t+n|t}$ 도 전항들의 선형결합으로 나타낼 수 있다. Q.E.D.

Bartlett검정에 대한 자세한 내용이나 증명은 Anderson(1984, pp. 497~498)을 참조하라.

**정리 8.12** 최소정준상관계수  $\rho$ 가 0이라는 귀무가설하에서 다음 통계량은 점근카이제곱분포를 갖고 자유도는  $d(m+1) - q + 1$ 이다.

$$\chi^2 = - \left[ T - \frac{1}{2} \{d(m+1) + (q+1)\} \right] \ln(1 - \hat{\rho}^2)$$

단,  $\hat{\rho}$ 는 최소정준상관계수의 추정량이고,  $q$ 는 현재단계에서의 상태벡터  $F^{(ij)}$ 의 차원이다. ▣

**정리 8.13** 만일  $z_{n,t} = y_{i,t+j-1|t}$ 라 하면  $z_{n,t+1|t}$ 는 다음과 같다. ▣

$$z_{n,t+1|t} = y_{i,t+j|t}$$

(증명) 정의에 의해서 다음 식이 성립한다.

$$z_{n,t} = y_{i,t+j-1|t}$$

이 식에 의해 다음 식이 성립한다.

$$y_{i,t+j-1} = z_{n,t} + v_{i,t+j-1}$$

따라서 다음 식이 성립한다.

$$y_{i,t+j} = z_{n,t+1} + v_{i,t+j}$$

이 식에 의해서 다음 식이 성립한다.

$$y_{i,t+j|t} = z_{n,t+1|t}$$

Q.E.D

다변량정규분포에서 발생한 서로 독립인 관찰점들로 구성된 데이터세트를 분석하는 다변량통계분석에서 정준상관분석은 두 개의 확률벡터들 사이의 공분산구조(covariance structure)를 나타내는 선형모형의 최우추정과정(maximum likelihood estimation procedure)으로 간주할 수 있다. 이에 대한 자세한 내용은 Anderson(1984, pp. 502~508)을 참조하라. Akaike(1976, pp. 66~68)는 이 이론을 바탕으로 정준상관분석단계에서 상대벡터를 결정하는데 AIC의 변형형태인 DIC를 사용할 것을 제안하였다. 차원이 각각  $s$ 와  $t(\leq s)$ 인 확률벡터들  $x$ 와  $y$ 가 다음과 같은 선형모형을 만족한다고 하자.

$$y = Ax + v$$

여기서  $A$ 는 종속확률벡터를  $y$ 로 하고 독립확률벡터를  $x$ 로 하는 다변량회귀분석의 회귀계수행렬이고, 확률벡터  $v$ 는 확률벡터  $x$ 와 비상관관계(uncorrelated)에 있다. 이 선형모형의 자유모수(free parameter)들의 개수는 확률벡터  $x$ 의 공분산행렬의 자유모수들의 개수, 확률벡터  $v$ 의 공분산행렬의 자유모수들의 개수, 그리고 회귀계수행렬  $A$ 의 자유모수들의 개수의 합이다. 만일 회귀계수행렬  $A$ 의 랭크가  $q(\leq t)$ 라고 가정하면 이 개수들은 각각  $s(s+1)/2$ ,  $t(t+1)/2$ , 그리고  $q(s+t-q)$ 이다. 따라서 자유모수들의 개수들의 합은 다음과 같다.

$$F(q) = \frac{1}{2}s(s+1) + \frac{1}{2}t(t+1) + q(s+t-q)$$

만일  $T$ 개의 관찰점들이 주어지면 이 모형의 AIC는 다음과 같다.

$$AIC(q) = T \log \prod_{i=1}^q (1 - c_i^2) + 2F(q)$$

여기서  $c_i$ 는 이 데이터세트의 제  $i$ 번째 표본정준상관계수이다. Akaike(1974)는 다음과 같이 DIC를 정의했다.

$$DIC(q) = AIC(q) - AIC(t)$$

여기서  $AIC(t)$ 는 회귀계수행렬  $A$ 에 어떠한 제약조건도 가하지 않았을 때의 AIC값이다. 따라서 이 모형의 DIC는 다음과 같이 쓸 수 있다.

$$DIC(q) = -T \log \prod_{i=q+1}^t (1 - c_i^2) - 2(t-q)(s-q)$$

만일 진짜(true) 정준상관계수들 중에서 가장 큰  $q$ 개를 제외한 나머지 값들이 0, 또는 음수라면, 일반적 조건하에서(under regularity conditions) 이 식의 우변의 첫째 항의 기대값은 점근적으로  $(t-q)$ 에 어떤 양수를 곱한 것과 같다. 지금 다루고 있는 서로 독립인 관찰점들로 구성된 데이터세트인 경우에는 이 양수의 값은 1이고, 시계열데이터인 경우에는 이 양수가 꼭 1이라고 할 수는 없다. 이후 음이 아닌 진짜 정준상관계수들 중에서 0이 아닌 개수를  $q_0$ 라 하자. 만일  $q < q_0$ 면, 관찰점들의 개수  $T$ 가 증가함에 따라  $DIC(q)$ 는 한없이 증가할 것이다. 따라서 만일  $DIC(q)$ 를 최소로 하는  $q$ 를 차수로 선택한다면, 적어도 관찰점들의 개수  $T$ 가 아주 큰 경우에는 이 선택된 차수가 진짜 차수  $q_0$ 보다 작지는 않을 것

이다. 반면에 만일  $(s-q)$ 가 크고 또한  $q \geq q_0$ 라면,  $DIC(q)$ 는 음수일 것이다. 따라서 이 선택된 차수가 진짜 차수  $q$ 보다 클 확률이 양수이다. 그러나  $(s-q)$ 가 큰 경우에는 이 확률은 아주 작을 것이다. 이  $DIC(q)$ 를 일반화 시켜서 다음과 같이 정의하자.

$$DIC(q) = -T \log \prod_{i=q+1}^T (1 - c_i^2) + 2T^b(t-q)(s-q) \quad (0 < b < 1)$$

관찰점들의 개수  $T$ 를 크게 함으로써 이 일반화된  $DIC$ 를 최소화하는 차수가 진짜 차수보다 클 확률을 원하는 대로 작게 만들 수 있다. 그러나 지수  $b$ 를 주관적으로 선택할 수밖에 없음을 유의해라. 비록 이  $DIC$ 가 서로 독립인 관찰점들로 구성된 데이터세트인 경우에 대해서 유도하였지만, Akaike는 모의실험과 실제 시계열데이터분석을 통해서  $DIC(q)$ 를 최소화하는 기준이 유용함을 보이고 이 벌칙함수기준을 상태공간모형의 식별에 사용할 것을 권하였다. 정준상관분석의 경우에는  $t=s=(m+1)d$ 이다.

#### 8.5.4 Akaike의 최종추정단계

Akaike(1973, 1974a)는 스펙트럴밀도함수를 사용해서 상태공간모형을 최우추정하였으나, 이 방법이 널리 쓰이지는 않는다. 이 소절에서는 이 최우추정법을 간략히 소개하고자 한다.

정상다변량시계열데이터  $y_1, y_2, \dots, y_T$ 가 주어지고 이 다변량시계열데이터에 적합한 상태공간모형을 식별하였다고 가정하자. 지금부터는 식별된 행렬들  $F, G$ 와  $\Sigma$ 의 효율적인 추정량들을 얻기 위한 최우추정법에 대해서 살펴보자. 상태방정식에서 다음 식을 얻는다.

$$z_i = (I - FB)^{-1} Gv_i$$

이 식과 상태공간모형으로부터 다음 식들을 얻는다.

$$y_i = H(I - FB)^{-1} Gv_i$$

$$v_i = [H(I - FB)^{-1} G]^{-1} y_i$$

따라서 대수우도함수는 다음과 같은 근사식을 만족한다.

$$\ln L(F, G, \Sigma | y_1, \dots, y_T) \propto -\frac{T}{2} \{d \ln 2\pi + \ln \det(\Sigma) + \text{trace}(\Sigma^{-1} S_0)\}$$

$$\text{단, } S_0 = \frac{1}{T} \sum_{i=1}^T v_i v_i'$$

정상다변량시계열데이터들  $\{y_i\}$ ,  $\{z_i\}$ , 그리고  $\{v_i\}$ 의 Fourier 변환(transform)들은 다음과 같이 정의된다.

$$Y(f) = \frac{1}{\sqrt{T}} \sum_{i=1}^T \exp(-i2\pi ft) y_i$$

$$Z(f) = \frac{1}{\sqrt{T}} \sum_{i=1}^T \exp(-i2\pi ft) z_i$$

$$V(f) = \frac{1}{\sqrt{T}} \sum_{i=1}^T \exp(-i2\pi ft) v_i$$

Fourier분석의 이론으로부터 다음 식들이 성립함을 알 수 있다.

$$V(f) = [H\{I - \exp(-i2\pi f)F\}^{-1}G]^{-1}Y(f)$$

$$S_0 = \int_{-\frac{1}{2}}^{\frac{1}{2}} V(f)V^*(f)df$$

여기서 \*는 공액이항(conjugate transpose)을 뜻한다. 이 식을 이용해서 다음과 같은 대수우도함수의 근사식을 구할 수 있다.

$$\ln L \propto -\frac{T}{2} \left\{ d \ln 2\pi + \ln \det(\Sigma) + \int_{-\frac{1}{2}}^{\frac{1}{2}} \text{trace}[Y(f)Y^*(f)P^{-1}(f)]df \right\}$$

여기서  $P(f)$ 는 다음과 같이 정의되는 스펙트럴밀도함수행렬(spectral density matrix)이다.

$$P(f) = [H\{I - \exp(-i2\pi f)F\}^{-1}G] \Sigma [H\{I - \exp(-i2\pi f)F\}^{-1}G]^*$$

다음과 같은 식을 정의하자.

$$P_T(f) = E\{Y(f)Y^*(f)\}$$

또한 관찰점들의 개수  $T$ 가 무한히 커질 때 다음 식이 성립한다고 가정하자.

$$\lim_{T \rightarrow \infty} P_T(f) = P_\infty(f)$$

대수우도함수의 근사식을  $(-T)$ 로 나눈 식은 관찰점들의 개수  $T$ 가 무한히 커지면 다음 식에 수렴한다.

$$\frac{1}{2} \left\{ d \ln 2\pi + \ln \det(\Sigma) + \int_{-\frac{1}{2}}^{\frac{1}{2}} \text{trace}[P_\infty(f)P^{-1}(f)]df \right\}$$

이 식을 최소화하는 행렬들이 근사적 최우추정행렬들이다. 관찰점들의 개수  $T$ 가 큰 경우에는 이 식의 적분항은 무시할 수 있을 정도로 작다. 또한,  $S_0$ 는 근사적으로  $\Sigma$ 의 최우추정행렬이다. 따라서 우도함수를 최대화하는 것은 단지  $\det(S_0)$ 를 최소화하는 것과 같다. 이 행렬식을 최소화하는 과정은 제 8.3.4 소절에서 다룬 것과 같다.

### 8.5.5 Kalman필터링

이 소절에서는 상태공간모형의 일반형을 소개하고, 상태공간모형을 추정하는 데 가장 널리 쓰이는 Kalman필터링에 대해서 살펴보자.

시스템이 선형이고 시간에 불변(time invariant)일 때, 이 시스템은 다음과 같은 상태방정식과 출력방정식으로 이루어진 상태공간모형으로 표현할 수 있다.

$$z_{t+1} = Fz_t + Gx_{t+1}$$

$$y_t = Hz_t + \zeta_t$$

단,  $z_t$ 는  $k \times 1$ 인 상태벡터(state vector)

$F$ 는  $k \times k$ 인 변환행렬(transition matrix)

$G$ 는  $k \times n$ 인 입력행렬(input matrix)

$x_t$ 는  $n \times 1$ 인 시스템의 입력벡터(input vector)

$y_t$ 는  $d \times 1$ 인 출력벡터(output vector)

$H$ 는  $d \times k$ 인 출력행렬(output matrix), 또는  
관찰행렬(observation matrix)

$\zeta_t$ 는  $d \times 1$ 인 측정오차(disturbance, measurement error)

상태방정식은 시스템방정식(system equation)이나 변환방정식(transition equation)이라 하고, 출력방정식은 측정방정식(measurement equation), 또는 관찰방정식(observation equation)이라고 부른다.

첫 번째 예제로 다음과 같은 다변량시스템을 생각해 보자.

$$Y_t = \nu(B)x_t$$

$$y_t = Y_t + \zeta_t$$

$$\text{단, } \nu(z) = \sum_{j=0}^{\infty} \nu_j z^j$$

$B$ 는 후진연산자(backshift operator)

여기서  $Y_t$ 는 출력시계열데이터,  $y_t$ 는 관찰된 출력시계열데이터,  $x_t$ 는 입력시계열데이터,  $\zeta_t$ 는 잡음항(또는 오차항)이다. 또한  $\nu(z)$ 를 전이함수라 하며 계수  $\{\nu_j\}$ 는 충격반응가중값(impulse response weight)이라고 부른다. 따라서 이 모형을 전이함수모형이라 부르기도 한다. 만일  $j = -1, -2, \dots$ 에 대해  $\nu_j = 0$ 이면 이 전이함수모형은 작인적(causal)이라고 정의한다. 이 전이함수모형이 작인적인 경우, 전이함수  $\nu(z)$ 가 다음과 같이 유리함수형태를 갖는다고 가정하자.

$$\nu(z) = \frac{\Theta(z)}{\Phi(z)}$$

$$\text{단, } \Phi(z) = 1 + \Phi_1 z + \dots + \Phi_p z^p$$

$$\Theta(z) = 1 + \Theta_1 z + \dots + \Theta_q z^q$$

이러한 조건하에서 이 다변량시스템은 다음과 같이 쓸 수 있다.

$$Y_t + \Phi_1 Y_{t-1} + \dots + \Phi_p Y_{t-p} = \Theta_0 x_t + \Theta_1 x_{t-1} + \dots + \Theta_q x_{t-q}$$

$$y_t = Y_t + \zeta_t$$

두 번째 예제로 입력시계열데이터  $\{x_t\}$ 를 다음과 같은 유한차수의 MA모형으로 나타낼 수 있다고 하자.

$$x_t = \Theta(B)v_t$$

단,  $\{v_t\}$ 는 백색잡음과정

$$\Theta(z) = 1 - \Theta_1 z - \dots - \Theta_q z^q$$

이 경우에는  $E(v_t | x_i, i \leq t-1) = v_t$ 를 사용해서  $v_t$ 를 다음과 같이 표현할 수 있다.

$$v_t = x_t - E(x_t | x_i, i \leq t-1)$$

즉,  $v_t$ 는  $x_t$ 의 전방일단계예측오차이다. 따라서 상태공간모형은 다음과 같이 나타낼 수 있다.

$$z_{t+1} = Fz_t + Gv_{t+1}$$

$$y_t = Hz_t + \zeta_t$$

여기서  $v_t$ 와  $\zeta_t$ 는 서로 독립이다. 물론 이 식의 행렬들  $F$ 와  $G$ 는 앞에서 정의한 행렬들  $F$ 나  $G$ 와는 다르지만 편의상 같은 기호를 사용하기로 한다.

세 번째 예제로 만일 측정오차가 없고( $\zeta_t \equiv 0$ ),  $\{x_t\}$ 가 백색잡음과정( $x_t \equiv v_t$ )이라면, 다음과 같은 상태공간모형을 얻을 수 있다.

$$z_{t+1} = Fz_t + Gv_{t+1}$$

$$y_t = Hz_t$$

이 경우  $\{y_t\}$ 는 백색잡음과정인 입력시계열데이터  $\{v_t\}$ 에 대해 선형이고 시간에 불변인 시스템의 출력시계열데이터이다. 이 상태방정식은 출력시계열데이터와 입력시계열데이터를 동일시( $y_t \equiv x_t$ )함으로써 두 번째 예제의 상태방정식으로부터 유도할 수 있다. 이 경우  $v_t$ 는  $y_t$ 의 쇄신항(innovation process)이다.

필터링의 문제는 다음과 같이 설명할 수 있다. 관찰된 출력시계열데이터  $y_t$ 는 신호(signal)  $Y_t$ 와 잡음항(noise)  $\zeta_t$ 로 구성되며 다음과 같이 나타낼 수 있다.

$$y_t = Y_t + \zeta_t$$

지금부터는 시점  $t$ 까지의 관찰벡터들  $y_1, y_2, \dots, y_t$ 가 주어졌을 때  $y_{t+m}$ 의 추정에 대해서 살펴보자. 여기서  $m$ 은 양수일 수도, 음수일 수도, 0일 수도 있다. 만일  $m$ 이 양수이면 예측문제(prediction problem)이고,  $m$ 이 0이면 필터문제(filter problem)이며,  $m$ 이 음수일 경우 평활문제(smoothing problem)로 분류한다. 그러나 만일  $\zeta_t \equiv 0$ 이면, 이 문제는  $m$ 이 양수인 경우에만 의미가 있다. 단변량시계열분석에서의 필터링문제는 Wiener(1949)에 의해서 다음과 같은 가정하에 처음으로 연구되었다.

- ①  $Y_{t+m}$ 의 추정량  $\hat{Y}_{t+m}$ 은  $\{Y_i | i \leq t\}$ 의 선형함수를 바탕으로 한다.
- ② 선형예측함수의 적합성 판단기준은 평균제곱오차이다. 즉,  $E(\hat{Y}_{t+m} - Y_{t+m})^2$ 이다.
- ③  $Y_t$ 와  $\zeta_t$ 는 결합정상적(jointly stationary)이다.

만일  $Y_t$ 가 정규오차항을 갖는 유한차수의 VARMA모형을 따르고  $\zeta_t$ 도 다변량정규분포를 갖는다고 가정하면, Kalman & Buch(1961)에 의해서 제안되고 Kalman필터라고 알려진 반복기법을 사용해서 선형최소제곱오차필터  $\hat{Y}_{t+m}$ 을 구할 수 있다. 이 Kalman필터는 널리 알려져 있으며 인자분석(factor analysis)을 이용하여 간단히 유도할 수도 있다(Priestley & Subba Rao [1975]).

다변량시계열데이터  $\{y_t\}$ 가 VARMA모형을 만족한다고 가정할 때, 정리 8.1에 의해서 다음과 같은 상태공간모형으로 나타낼 수 있다.

$$\begin{aligned} z_{t+1} &= Fz_t + Gu_{t+1} \\ y_t &= Hz_t + \zeta_t \end{aligned}$$

여기서  $\{z_t\}$ 와  $\{y_t\}$ 는 결합정상적 정규확률과정이고,  $\{u_t\}$ 와  $\{\zeta_t\}$ 도 평균벡터들이 0인 다변량정규백색잡음과정이며,  $E(u_t \zeta_t') = 0$ 라고 가정한다. 상태공간모형으로부터 상태벡터  $z_t$ 는 다음과 같이 표시된다.

$$z_t = F^{t-j} z_j + \sum_{i=j+1}^t F^{t-i} G u_{i-1} \quad (0 \leq j \leq t)$$

따라서 다음과 같은 성질이 성립한다.

$$\begin{aligned} E(z_t) &= 0 & (t=1, 2, \dots) \\ E(z_t v_s') &= 0 & (s \geq t) \\ E(z_t \zeta_s') &= 0 & (t=1, 2, \dots, s=1, 2, \dots) \\ E(y_t v_s') &= 0 & (s \geq t) \end{aligned}$$

다변량시계열데이터  $\{y_t\}$ 의 전방일단계오차는 다음과 같다.

$$e_t = y_t - E(y_t | y_i, i \leq t-1)$$

정상성가정하에서,  $E(y_t | y_i, i \leq t-1)$ 은 선형최소제곱오차예측벡터이다. 출력방정식의 양변에 조건부기대값을 취하면 다음과 같다.

$$E(y_t | y_i, i \leq t-1) = HE(z_t | y_i, i \leq t-1) = H\hat{z}_{t|t-1}$$

여기서  $\hat{z}_{t|t-1}$ 은  $y_{t-1}, y_{t-2}, \dots$ 이 주어진 상태에서  $z_t$ 의 선형최소제곱오차예측벡터를 나타낸다. 따라서 전방일단계예측오차는 다음과 같다.

$$\begin{aligned} e_t &= Hf_t + \zeta_t & (t=1, 2, \dots) \\ \text{단, } f_t &= z_t - \hat{z}_{t|t-1} \end{aligned}$$

여기서  $f_t$ 는 필터오차라 하며  $f_t$ 의 평균벡터는 0이다. 또한  $f_t$ 의 공분산행렬을  $C_{t|t-1} = E[f_t f_t']$ 라고 하면 다음 식이 성립한다.

$$C_{t|t-1} = F C_{t-1|t-1} F' + G Q G'$$

여기서  $Q$ 는  $\{v_t\}$ 조건부확률밀도함수의 공분산행렬이다.

만약 최적필터추정량  $\hat{z}_{t|t}$ 와 필터오차  $f_t$ 의 공분산행렬  $C_t$ 가  $t=0, 1, \dots$ 에 대해 주어지면 최적의 전방일단계예측벡터는 다음과 같다.

$$\hat{z}_{t|t-1} = F \hat{z}_{t-1|t-1}$$

또한,  $\hat{z}_{t|t-1}$ 은  $\{y_i, i \leq t-1\}$ 의 선형함수이기 때문에,  $f_t$ 와  $\xi_t$ 는 서로 독립이며  $\{e_t\}$ 가 서로 독립이다. 따라서 다음 식은 인자분석에서 사용되는 식과 같은 형태이다.

$$e_t = H f_t + \xi_t \quad (t=1, 2, \dots)$$

단,  $f_t = z_t - \hat{z}_{t|t-1}$

다변량시계열데이터  $\{y_t, y_{t-1}, \dots\}$ 이 주어진 상태에서  $f_t$ 의 추정에 대한 문제를 생각해 보자. 이  $f_t$ 의 선형최소제곱오차추정벡터는 다음과 같다.

$$\hat{f}_t = E(f_t | y_t, i \leq t)$$

이 식의 우변은  $\{y_t\}$ 의 선형함수이다. 또한  $\{y_t\}$ 와  $\{e_t\}$  사이에 선형관계가 있으므로 다음 식이 성립한다.

$$\begin{aligned} \hat{f}_t &= E(f_t | e_t, i \leq t) \\ &= E(z_t - \hat{z}_{t|t-1} | e_t, i \leq t) \\ &= \hat{z}_{t|t} - \hat{z}_{t|t-1} \end{aligned}$$

또한  $\{z_t\}$ 와  $\{e_t\}$ 는 결합정상적이므로 선형최소제곱오차예측벡터  $\hat{z}_{t|t}$ 는  $e_t, e_{t-1}, \dots$ 의 선형결합이다.

$$\hat{z}_{t|t} = A_0 e_t + A_1 e_{t-1} + \dots$$

여기서  $\{e_t\}$ 는 직교적이므로  $e_t$ 를 제거해도 이 방정식의 계수들은 변하지 않는다. 따라서 다음 식이 성립한다.

$$\hat{z}_{t|t-1} = A_1 e_{t-1} + A_2 e_{t-2} + \dots$$

따라서  $\hat{f}_t$ 는 다음과 같이  $e_t$ 의 함수가 된다.

$$\begin{aligned} \hat{f}_t &= \hat{z}_{t|t} - \hat{z}_{t|t-1} = A_0 e_t \\ \text{단, } \hat{f}_t &= [\hat{f}_{1,t}, \hat{f}_{2,t}, \dots, \hat{f}_{k,t}] \end{aligned}$$



$$A_0 = [a_1, a_2, \dots, a_k]$$

즉, 다음 식이 성립한다.

$$\hat{f}_{it} = a_i' e_t \quad (i=1, 2, \dots, k)$$

조건부기대값의 성질에 의해서 벡터  $a_i$ 가 다음 식을 최소화하는 벡터임을 알 수 있다.

$$Q_i = E(\hat{f}_{it} - f_{it})^2$$

Lawley & Maxwell(1971)은 이러한 문제유형은 인자분석을 통하여 인자점수(factor score)를 추정하는 것과 동일함을 보이고, 또한  $f_t$ 의 최적추정량은 다음 식을 만족함을 보였다.

$$\hat{f}_t = C_{t|t-1} H' \Sigma_e^{-1} e_t = C_{t|t-1} H' \Sigma_e^{-1} [y_t - H \hat{z}_{t|t-1}]$$

단,  $\Sigma_e$ 는  $e_t$ 의 공분산행렬

이 식으로부터  $\hat{z}_{t|t}$ 는 다음 식을 만족함을 알 수 있다.

$$\hat{z}_{t|t} = \hat{z}_{t|t-1} + K_t [y_t - H \hat{z}_{t|t-1}]$$

단,  $K_t = C_{t|t-1} H' \Sigma_e^{-1}$

이  $K_t$ 를 Kalman이득행렬(Kalman gain matrix)이라 부른다.

필터오차  $\hat{f}_t$ 의 공분산행렬의 식으로부터  $t=0, 1, \dots$ 에 대하여 다음 식이 성립함을 알 수 있다.

$$K_t = C_{t|t-1} H' [H C_{t|t-1} H' + R]^{-1}$$

$$C_{t|t-1} = F C_{t-1|t-1} F' + G Q G'$$

$$C_{t|t} = [I - K_t H] C_{t|t-1}$$

또한, 다음 식이 성립한다.

$$\begin{aligned} & \hat{z}_{t|t-1} \\ &= E[z_t | y_i, i \leq t-1] \\ &= F E[z_{t-1} | y_i, i \leq t-1] \\ &= F \hat{z}_{t-1|t-1} \end{aligned}$$

이 식은 바로 전 시점에서의 상태벡터추정량  $\hat{z}_{t-1|t-1}$ 을 최신화시켜서 현재 상태벡터의 추정량을 계산하는 반복식이다.

위 식들을 이용하여 제  $m$  단계의 반복필터(recursive filter)를 다음과 같이 유도할 수 있다.

$$z_{t+m} = F^m z_t + \sum_{i=1}^m F^{m-i} G v_{t+i-1}$$

또한  $i \geq 1$ 에 대하여  $E(v_{t+i-1} | y_j, j \leq t) = 0$ 이므로 다음 식이 성립한다.

$$\hat{z}_{t+m} = E[z_{t+m} | y_i, i \leq t] = F^m \hat{z}_{t|t}$$

Kalman의 일단계반복법은 다음과 같다. 시점  $(t-1)$ 까지의 관찰값  $\{y_i | i \leq t-1\}$ 을 가지고 시점  $(t-1)$ 에서의 상태벡터의 추정량  $\hat{z}_{t-1|t-1}$ 을 계산한다. 단지 이 관찰값들을 사용한  $z_t$ 의 추정량은  $\hat{z}_{t|t-1} = F\hat{z}_{t-1|t-1}$ 이고,  $y_t$ 의 예측값은  $\hat{y}_{t|t-1} = H\hat{z}_{t|t-1}$ 이다. 시점  $t$ 에서의 관찰값  $y_t$ 를 이용할 수 있을 때, 과거의 예측벡터  $\hat{z}_{t|t-1}$ 과 예측오차인  $(y_t - H\hat{z}_{t|t-1})$  사이의 선형결합으로  $z_t$ 의 추정값을 최신화시킬 수 있다. Kalman이득행렬인  $K_t$ 는 단순히 가중인자(weighting factor)로 간주할 수 있다. 이러한 의미에서 Kalman필터링은 지수가중평활법(exponential moving average method)에 의한 예측의 일반화이다. 필터링오차의 공분산행렬  $D_t$ 는 다음과 같다.

$$D_t = E[(z_t - \hat{z}_{t|t})(z_t - \hat{z}_{t|t})'] = C_{t|t-1} - K_t \Sigma_e K_t'$$

단,  $\hat{z}_{t|t} = F\hat{z}_{t-1|t-1} + K_t e_t$

여기서  $\hat{z}_{t|t}$ 와  $D_t$ 를 다음과 같은 반복식을 사용해서 구할 수 있다.

$$\begin{aligned} \hat{Y}_{t|t-1} &= HF\hat{z}_{t-1|t-1} \\ e_t &= Y_t - \hat{Y}_{t|t-1} \\ C_{t|t-1} &= FD_{t-1}F' + G\Sigma_vG' \\ \Sigma_e &= HC_{t|t-1}H' + \Sigma_\zeta \\ K_t &= C_{t|t-1}H'\Sigma_e^{-1} \end{aligned}$$

따라서 다음 식이 성립한다.

$$\begin{aligned} \hat{z}_{t|t} &= F\hat{z}_{t-1|t-1} + Ke_t \\ D_t &= C_{t|t-1} - K\Sigma_eK' \end{aligned}$$

위에서  $\Sigma_v$ 와  $\Sigma_\zeta$ 은 각각  $v_t$ 와  $\zeta_t$ 의 공분산행렬이다. 이 반복식을 적용하기 위해서는  $H, F, G, \Sigma_v, \Sigma_\zeta$ , 그리고 초기시점  $t_0$ 에서의  $\hat{z}_{t_0|t_0}, D_{t_0}$ 를 알아야 한다.

반복최신화식인 Kalman필터는 다른 방법으로 유도할 수 있다. 예를 들면,  $v_t$ 와  $\zeta_t$ 가 다변량정규분포를 갖는다는 가정하에, 상태벡터의 사전분포와 사후분포를 이용하여 Kalman필터를 유도할 수 있다.

이 Kalman필터는 컴퓨터의 발전으로 알고리즘을 수행하기가 쉬워진 덕분에 많은 분야에서 널리 사용되고 있다. 이 방법은 시계열데이터의 예측, 베이지안예측, 시점에 따라 변하는 계수들(time varying coefficients)을 갖는 시계열모형의 분석, 결측값분석, 품질관리 등에 적용되고 있다.

## 8.6 상태공간모형화방법의 예제들

이 절에서는 지금까지 다룬 상태공간모형화방법을 여러 시계열데이터들에 적용해서 이 방법의 유용성을 보이고자 한다. PROC STATESPACE를 사용해서 시계열데이터를 분석한 국내문헌으로는 정은이(1993) 이외에도 박준경·이호창(1985), 신호중(1986), 양현조(1992) 등이 있다.

### 8.6.1 시뮬레이션데이터 MA1A

단변량시계열데이터 MA1A는 다음과 같은 MA(1)모형에서 발생시킨 시뮬레이션데이터이다.

$$y_t = v_t + 0.8v_{t-1}$$

여기서  $\{v_t\}$ 는 분산이 1인 백색잡음과정이다. 단변량시계열데이터 MA1A의 관찰점들의 개수는 100이다. 이 단변량시계열데이터를 발생시키는 SAS프로그램이 제 2.7절에 수록되어 있고, 또한 디스켓 SAS4TSA4의 서브디렉토리 TSA4DATA, 즉, 영구SAS데이터라이브러리 TSA4DATA의 파일 MA1A.SSD에 수록되어 있다. 이 단변량시계열데이터의 시계열산포도가 그림 2.7이다. 이 시계열데이터를 상태공간모형화하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM8-23.PRG에 수록되어 있다).

```

/* */
/* EXAM8-23.PRG */
/* */
/* STATE SPACE MODELING */
/* DATASET = MA1A */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=72;
DATA EXAM823;
 SET TSA4DATA.MA1A;
RUN;
TITLE 'STATESPACE MODELING';
PROC STATESPACE ITPRINT CANCORR;

```

```

VAR Y;
RUN;
/* ----- */
/* WITH RESTRICT STATEMENT */
/* ----- */
PROC STATESPACE CANCECORR OUTAR=EXAM823B;
VAR Y;
RESTRICT F(2,2)=0;
RUN;
PROC PRINT DATA=EXAM823B;
RUN;
PROC STATESPACE CANCECORR DATA=EXAM823;
VAR Y;
RESTRICT F(2,2)=0 F(2,1)=0;
RUN;

```

이 SAS프로그램의 첫 번째 PROC스텝은 다음과 같다.

```

PROC STATESPACE ITPRINT CANCECORR;
VAR Y;
RUN;

```

이 프로시저를 수행한 결과가 그림 8.23.1이다. 그림 8.23.1의 부분 A에서 이 SAS데이터 세트의 관찰점들의 개수가 100, 부분 B에서 변수 Y의 표본평균이 0.211, 표본표준편차가 1.186임을 알 수 있다. 부분 C에는 VAR(0)모형, VAR(1)모형, ..., VAR(10)모형에 해당하는 AIC값들이 수록되어 있다. 이 중에서 VAR(3)모형의 AIC값이 최소이다. 즉, 상태공간모형화방법의 VAR모형단계에서 VAR(3)모형을 선택한다. 부분 D에는 VAR스테이트먼트에 지정된 변수들의 표본자기상관행렬들의 도표적 표현이 수록되어 있고, 부분 E에는 표본부분자기상관행렬들의 도표적 표현이 수록되어 있다. 부분 F에는 최소 AIC값을 갖는 VAR모형의 VAR계수행렬들의 Yule-Walker추정행렬들이 수록되어 있다. 이 예제에서는 VAR(3)모형이 최소의 AIC값을 가지므로 VAR(3)모형의 VAR계수행렬들이 추정되었다. 부분 G는 PROC STATESPACE스테이트먼트의 옵션 CANCECORR를 사용함으로써 출력된 것으로, 상태벡터들, 표본정준상관계수들, AIC값들, Bartlett통계량값들, 그리고 해당하는 자유도들이 출력되었다. 부분 H는 예비추정단계를 수행한 결과이다. 부분 H1은 예비추정단계의 상태벡터이고, 부분 H2는 변환행렬 F의 예비추정행렬이며, 부분 H3는 입력행렬 G의 예비추정행렬이며, 부분 H4는 쇄신과정의 공분산행렬의 예비추정행렬이다. 부분 I는 PROC STATESPACE스테이트먼트의 옵션 ITPRINT를 사용함으로써 출력된 것으로, 최종추정단계의 각 반복단계인 Iter에서의 변환행렬 F와 입력행렬 G의 원소들의 추정값들  $F(2, 1)$ ,  $F(2, 2)$ ,  $G(2, 1)$ 과 백색잡음과정의 공분산행렬의 추정행렬의 원소 SIGMA

```

STATESPACE MODELING
1
13:50 Sunday, August 1, 1993

STATESPACE Procedure

(A) Nobs = 100

(B) Variable Mean Std
Y 0.211388 1.186429

(C) Information Criterion for Autoregressive Models

Lag=0 Lag=1 Lag=2 Lag=3 Lag=4
34.189661 15.559334 0.509357 0.267194 1.168187

Lag=5 Lag=6 Lag=7 Lag=8 Lag=9
3.154724 4.703545 6.379448 8.092996 9.030008

Lag=10
10.947035

(D) Schematic Representation of Correlations

Name/Lag 0 1 2 3 4 5 6 7 8 9 10
Y + +

+ is > 2*std error, - is < -2*std error, . is between

(E) Schematic Representation of Partial Autocorrelations

Name/Lag 1 2 3 4 5 6 7 8 9 10
Y + -

+ is > 2*std error, - is < -2*std error, . is between

(F) Yule-Walker Estimates for the Min AIC

Lag=1 Lag=2 Lag=3
Y Y Y
Y 0.66165 -0.48567 0.14890

(G) Canonical Correlations Analysis

State vector Correlations Infor. Chisq D.F.
Y(T;T),Y(T+1;T) 1.0000 0.4295 14.39115 20.08528 3
Y(T;T),Y(T+1;T),Y(T+2;T) 1.0000 0.4575 0.0507 -3.74256 0.254864 2

(H) Selected Statespace Form and Preliminary Estimates

(H1) State Vector

Y(T;T) Y(T+1;T)

(H2) Estimate of the Transition Matrix

0 1
-0.249 0.263

(H3) Input Matrix for the Innovation

1
0.662

(H4) Variance Matrix for the Innovation

0.94428424

```

| ① Iterative Fitting |      |                          |        |            |            |            |
|---------------------|------|--------------------------|--------|------------|------------|------------|
| Iter                | Half | Determinant              | Lambda | F(2,1)     | F(2,2)     | G(2,1)     |
| 0                   | 0    | 0.93769133<br>0.93769133 | 0.1    | -0.2493173 | 0.26313851 | 0.66165426 |
| 1                   | 0    | 0.93749852<br>0.93749852 | 0.01   | -0.2537855 | 0.27169138 | 0.67219694 |
| 2                   | 0    | 0.93747756<br>0.93747756 | 0.001  | -0.2581047 | 0.28008544 | 0.67494126 |
| 3                   | 0    | 0.93747743<br>0.93747743 | 0.0001 | -0.2578883 | 0.27965188 | 0.67521926 |
| 4                   | 0    | 0.93747742<br>0.93747742 | 1E-5   | -0.2579689 | 0.27979466 | 0.67520004 |

| ② Selected Statespace Form and Fitted Model |          |           |          |
|---------------------------------------------|----------|-----------|----------|
| ③ State Vector                              |          |           |          |
| Y(T;T)                                      | Y(T+1;T) |           |          |
| ④ Estimate of the Transition Matrix         |          |           |          |
| 0                                           | 1        |           |          |
| -0.258                                      | 0.280    |           |          |
| ⑤ Input Matrix for the Innovation           |          |           |          |
| 1                                           |          |           |          |
| 0.675                                       |          |           |          |
| ⑥ Variance Matrix for the Innovation        |          |           |          |
| 0.93747742                                  |          |           |          |
| ⑦ Parameter Estimates                       |          |           |          |
| Parameter                                   | Estimate | Std. Err. | T value  |
| F(2,1)                                      | -0.25797 | 0.139653  | -1.84722 |
| F(2,2)                                      | 0.279795 | 0.202156  | 1.38405  |
| G(2,1)                                      | 0.6752   | 0.099465  | 6.788337 |

그림 8.23.1 단변량시계열데이터 MA1A의 상태공간모형화

(1, 1), 그리고 백색잡음과정의 공분산행렬의 추정행렬의 행렬식 Determinant 등이 출력되었다. 그 외에도 변수들 Half와 Lambda가 출력되었다. 그러나 [ETS6.1]의 어느 곳에도 이들에 대한 설명이 없다. 본 저자의 생각으로는 변수 Lambda는 백색잡음과정의 공분산행렬의 추정행렬의 행렬식의 상대적 변화량을 뜻하는 것 같다. 부분 ②는 최종추정단계를 수행한 결과이다. 부분 ③은 최종추정단계의 상태벡터이다. 부분 ④는 변환행렬 F의 최종추정행렬이고, 부분 ⑤는 입력행렬 G의 최종추정행렬이며, 부분 ⑥는 쇄신과정의 공분산행렬의 최종추정행렬이다. 부분 ⑦에는 행렬들 F와 G의 추정된 원소들의 최종추정값들, 표준오차들, 이에 해당하는 t통계량값들이 수록되어 있다.

그림 8.23.1의 부분 ⑦에서 원소 F(2, 2)의 추정값이 0과 유의적으로 다르지 않음을 알 수 있다. 따라서 이 SAS프로그램의 두 번째 프로시저에서는 RESTRICT스태이트먼트를 사용해서 이 원소를 0으로 고정시킨 뒤 상태공간모형을 추정한다. 이 프로시저는 다음과 같다.

| STATESPACE MODELING  |                                           | 9                            |           |          |
|----------------------|-------------------------------------------|------------------------------|-----------|----------|
|                      |                                           | 13:50 Sunday, August 1, 1993 |           |          |
| STATESPACE Procedure |                                           |                              |           |          |
| (A)                  | Selected Statespace Form and Fitted Model |                              |           |          |
| (A1)                 | State Vector                              |                              |           |          |
|                      | Y(T;T)                                    | Y(T+1;T)                     |           |          |
| (A2)                 | Estimate of the Transition Matrix         |                              |           |          |
|                      | 0                                         | 1                            |           |          |
|                      | -0.121                                    | 0.000                        |           |          |
| (A3)                 | Input Matrix for the Innovation           |                              |           |          |
|                      | 1                                         |                              |           |          |
|                      | 0.602                                     |                              |           |          |
| (A4)                 | Variance Matrix for the Innovation        |                              |           |          |
|                      | 0.95905053                                |                              |           |          |
| (B)                  | Parameter Estimates                       |                              |           |          |
|                      | Parameter                                 | Estimate                     | Std. Err. | T value  |
|                      | F(2,1)                                    | -0.12063                     | 0.108143  | -1.11548 |
|                      | G(2,1)                                    | 0.60198                      | 0.087     | 6.919344 |

그림 8.23.2 RESTRICT스테이트먼트를 사용한 상태공간모형화

```

PROC STATESPACE CANCORR OUTAR=EXAM823B;
VAR Y;
RESTRICT F(2,2)=0;
RUN;

```

이 프로시저에서는 RESTRICT스테이트먼트에 'F(2, 2)=0'을 지정했다. 또한 PROC STATESPACE스테이트먼트의 옵션 OUTAR=에 EXAM823B를 할당함으로써 VAR모형단계를 수행한 결과를 SAS데이터세트 EXAM823B에 출력한다. 이 프로시저를 수행한 결과의 일부분이 그림 8.23.2이다. 부분 A는 최종추정단계를 수행한 결과이다. 부분 A1은 최종추정단계의 상태벡터이다. 부분 A2는 변환행렬 F의 최종추정행렬이고, 부분 A3는 입력행렬 G의 최종추정행렬이며, 부분 A4는 쇄신과정의 공분산행렬의 최종추정행렬이다. 부분 B에는 행렬들 F와 G의 추정된 원소들의 최종추정값들, 표준오차들, 이에 해당하는 t통계량값들이 수록되어 있다.

세 번째 PROC스텝은 다음과 같다.

```

PROC PRINT DATA= EXAM823B;
RUN;

```

이 프로시저는 두 번째 PROC스텝의 PROC STATESPACE스테이트먼트의 옵션 OUTAR=에 지정한 SAS데이터세트 EXAM823B를 OUTPUT윈도우에 출력시키기 위한

| STATESPACE MODELING |          |           |           |           |           |          |           |          | 10                           |
|---------------------|----------|-----------|-----------|-----------|-----------|----------|-----------|----------|------------------------------|
|                     |          |           |           |           |           |          |           |          | 13:50 Sunday, August 1, 1993 |
| OBS ORDER           | AIC      | SIGF1     | SIGB1     | FOR1_1    | FOR2_1    | FOR3_1   | FOR4_1    | FOR5_1   |                              |
| 1                   | 0        | 34.1897   | 1.40761   | 1.40761   | .         | .        | .         | .        | .                            |
| 2                   | 1        | 15.5593   | 1.14522   | 1.14522   | 0.43176   | .        | .         | .        | .                            |
| 3                   | 2        | 0.5094    | 0.96570   | 0.96570   | 0.60270   | -0.39593 | .         | .        | .                            |
| 4                   | 3        | 0.2672    | 0.94428   | 0.94428   | 0.66165   | -0.48567 | 0.14890   | .        | .                            |
| 5                   | 4        | 1.1682    | 0.93396   | 0.93396   | 0.67722   | -0.53644 | 0.21808   | -0.10455 | .                            |
| 6                   | 5        | 3.1547    | 0.93384   | 0.93384   | 0.67843   | -0.53897 | 0.22430   | -0.11240 | 0.011602                     |
| 7                   | 6        | 4.7035    | 0.92963   | 0.92963   | 0.67766   | -0.53143 | 0.20925   | -0.07624 | -0.033917                    |
| 8                   | 7        | 6.3794    | 0.92663   | 0.92663   | 0.68147   | -0.53336 | 0.20491   | -0.06434 | -0.064146                    |
| 9                   | 8        | 8.0930    | 0.92398   | 0.92398   | 0.67843   | -0.52771 | 0.20148   | -0.06778 | -0.053187                    |
| 10                  | 9        | 9.0300    | 0.91421   | 0.91421   | 0.67293   | -0.52981 | 0.20941   | -0.07325 | -0.060156                    |
| 11                  | 10       | 10.9470   | 0.91345   | 0.91345   | 0.66997   | -0.52934 | 0.20726   | -0.07043 | -0.061889                    |
| OBS                 | FOR6_1   | FOR7_1    | FOR8_1    | FOR9_1    | FOR10_1   | BAC1_1   | BAC2_1    | BAC3_1   |                              |
| 1                   | .        | .         | .         | .         | .         | .        | .         | .        | .                            |
| 2                   | .        | .         | .         | .         | .         | 0.43176  | .         | .        | .                            |
| 3                   | .        | .         | .         | .         | .         | 0.60270  | -0.39593  | .        | .                            |
| 4                   | .        | .         | .         | .         | .         | 0.66165  | -0.48567  | 0.14890  | .                            |
| 5                   | .        | .         | .         | .         | .         | 0.67722  | -0.53644  | 0.21808  | .                            |
| 6                   | .        | .         | .         | .         | .         | 0.67843  | -0.53897  | 0.22430  | .                            |
| 7                   | 0.06709  | .         | .         | .         | .         | 0.67766  | -0.53143  | 0.20925  | .                            |
| 8                   | 0.10564  | -0.056883 | .         | .         | .         | 0.68147  | -0.53336  | 0.20491  | .                            |
| 9                   | 0.07712  | -0.020436 | -0.053483 | .         | .         | 0.67843  | -0.52771  | 0.20148  | .                            |
| 10                  | 0.09783  | -0.074700 | 0.016279  | -0.10283  | .         | 0.67293  | -0.52981  | 0.20941  | .                            |
| 11                  | 0.09572  | -0.068669 | 0.001020  | -0.08345  | -0.028799 | 0.66997  | -0.52934  | 0.20726  | .                            |
| OBS                 | BAC4_1   | BAC5_1    | BAC6_1    | BAC7_1    | BAC8_1    | BAC9_1   | BAC10_1   |          |                              |
| 1                   | .        | .         | .         | .         | .         | .        | .         | .        | .                            |
| 2                   | .        | .         | .         | .         | .         | .        | .         | .        | .                            |
| 3                   | .        | .         | .         | .         | .         | .        | .         | .        | .                            |
| 4                   | .        | .         | .         | .         | .         | .        | .         | .        | .                            |
| 5                   | -0.10455 | .         | .         | .         | .         | .        | .         | .        | .                            |
| 6                   | -0.11240 | 0.011602  | .         | .         | .         | .        | .         | .        | .                            |
| 7                   | -0.07624 | -0.033917 | 0.06709   | .         | .         | .        | .         | .        | .                            |
| 8                   | -0.06434 | -0.064146 | 0.10564   | -0.056883 | .         | .        | .         | .        | .                            |
| 9                   | -0.06778 | -0.053187 | 0.07712   | -0.020436 | -0.053483 | .        | .         | .        | .                            |
| 10                  | -0.07325 | -0.060156 | 0.09783   | -0.074700 | 0.016279  | -0.10283 | .         | .        | .                            |
| 11                  | -0.07043 | -0.061889 | 0.09572   | -0.068669 | 0.001020  | -0.08345 | -0.028799 | .        | .                            |

그림 8.23.3 VAR모형단계에서 추정된 VAR모형

것이다. 이 옵션 OUTAR=은 VAR모형단계에서 추정된 VAR모형을 지정한 SAS데이터세트에 출력시킨다. 이 프로시저를 수행한 결과가 그림 8.23.3이다. 그림 8.23.3의 각 부분을 설명하면 다음과 같다.

- ① 변수 ORDER에는 VAR모형의 차수가 수록되어 있다.
- ② 변수 SIGB<sub>k</sub>에는 후향VAR모형의 백색잡음과정의 공분산행렬의 Yule-Walker 추정행렬의 제k번째 열이 수록되어 있다.
- ③ 변수 SIGF<sub>k</sub>에는 전향VAR모형의 백색잡음과정의 공분산행렬의 Yule-Walker 추정행렬의 제k번째 열이 수록되어 있다.
- ④ 변수 AIC에는 AIC값들이 수록되어 있다.
- ⑤ 변수 BAC<sub>k-j</sub>에는 후향VAR모형의 시차 k의 VAR계수행렬의 Yule-Walker 추정행렬의 제j번째 열이 수록되어 있다.
- ⑥ 변수 FOR<sub>k-j</sub>에는 전향VAR모형의 시차 k의 VAR계수행렬의 Yule-Walker 추정행렬의 제j번째 열이 수록되어 있다.



| STATESPACE MODELING  |                                           | 14                           |           |          |
|----------------------|-------------------------------------------|------------------------------|-----------|----------|
|                      |                                           | 13:50 Sunday, August 1, 1993 |           |          |
| STATESPACE Procedure |                                           |                              |           |          |
| (A)                  | Selected Statespace Form and Fitted Model |                              |           |          |
| (A1)                 | State Vector                              |                              |           |          |
|                      | Y(T,T)                                    | Y(T+1,T)                     |           |          |
| (A2)                 | Estimate of the Transition Matrix         |                              |           |          |
|                      | 0                                         | 1                            |           |          |
|                      | 0.000                                     | 0.000                        |           |          |
| (A3)                 | Input Matrix for the Innovation           |                              |           |          |
|                      | 1                                         |                              |           |          |
|                      | 0.644                                     |                              |           |          |
| (A4)                 | Variance Matrix for the Innovation        |                              |           |          |
|                      | 0.96981091                                |                              |           |          |
| (B)                  | Parameter Estimates                       |                              |           |          |
|                      | Parameter                                 | Estimate                     | Std. Err. | T value  |
|                      | G(2,1)                                    | 0.643755                     | 0.076326  | 8.434242 |

그림 8.23.4 두 개의 제약조건을 사용한 상태공간모형화

그림 8.23.2의 부분 B에서 원소  $F(2, 1)$ 의 추정값이 0과 유의적으로 다르지 않음을 알 수 있다. 따라서 이 SAS프로그램의 네 번째 프로시저에서는 RESTRICT스태이트먼트에 'F(2, 2)=0 F(2, 1)=0'을 할당한 뒤 상태공간모형을 추정한다. 이 프로시저는 다음과 같다.

```
PROC STATESPACE CANCORR DATA=EXAM823;
VAR Y;
RESTRICT F(2,2)=0 F(2,1)=0;
RUN;
```

이 프로시저를 수행한 결과의 일부분이 그림 8.23.4이다. 부분 A는 최종추정단계를 수행한 결과이다. 부분 A1은 최종추정단계의 상태벡터이다. 부분 A2는 변환행렬  $F$ 의 최종추정행렬이고, 부분 A3는 입력행렬  $G$ 의 최종추정행렬이며, 부분 A4는 쇄신과정의 공분산행렬의 최종추정행렬이다. 부분 B에는 행렬들  $F$ 와  $G$ 의 추정된 원소들의 최종추정값들, 표준오차들, 이에 해당하는  $t$ 통계량값들이 수록되어 있다. 부분 B로부터 더 이상 0으로 고정시켜야 할 원소가 없음을 알 수 있다.

SAS프로그램 EXAM8-23.PRGM의 3개의 상태공간모형들 중에서 최적의 모형을 식별하기 위해서 벌칙함수식별방법들을 적용하기로 하자. 이러한 목적으로 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM8-24.PRGM에 수록되어 있다).

```

/ * */
/ * EXAM8-24.PRG */
/ * */
/ * CHOOSING RESTRICT CONDITIONS */
/ * USING THE PENALTY FUNTION METHODS */
OPTIONS PAGESIZE=60 LINESIZE=72;
PROC IML;
RESET NOLOG NOCENTER NONAME;
T = 100;
K = 3;
/ * WITH NO RESTRICTION */
COV0 = { 0.93747742 };
NORES = 0;
AIC = LOG(DET(COV0)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COV0)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COV0)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT , "NO OF RESTRICTION = "NORES,
 " AIC = "AIC,
 " BIC = "BIC,
 " HQC = "HQC;
/ * WITH 1 RESTRICTION */
COV1 = { 0.95905053 };
NORES = 1;
AIC = LOG(DET(COV1)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COV1)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COV1)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT , "NO OF RESTRICTION = "NORES,
 " AIC = "AIC,
 " BIC = "BIC,
 " HQC = "HQC;
/ * WITH 2 RESTRICTIONS */
COV2 = { 0.96981091 };
NORES = 2;
AIC = LOG(DET(COV2)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COV2)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COV2)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT , "NO OF RESTRICTION = "NORES,
 " AIC = "AIC,
 " BIC = "BIC,

```

| SAS                 |           | 22:17 Sunday, August 1, 1993 4 |   |
|---------------------|-----------|--------------------------------|---|
| NO OF RESTRICTION = |           |                                | 0 |
| AIC =               | -0.004563 |                                |   |
| BIC =               | 0.0735925 |                                |   |
| HQC =               | 0.0270682 |                                |   |
| NO OF RESTRICTION = |           |                                | 1 |
| AIC =               | -0.001812 |                                |   |
| BIC =               | 0.0502919 |                                |   |
| HQC =               | 0.0192757 |                                |   |
| NO OF RESTRICTION = |           |                                | 2 |
| AIC =               | -0.010654 |                                |   |
| BIC =               | 0.0153975 |                                |   |
| HQC =               | -0.000111 |                                |   |
| Exiting IML.        |           |                                |   |

그림 8.24 벌칙함수식별방법들에 의한 모형들의 비교

```

 " HQC = "HQC, ;
QUIT;

```

이 SAS프로그램을 수행한 결과가 그림 8.24이다. 그림 8.24에서 AIC, BIC, 그리고 HQC 모두가 제한조건이 두 개인 상태공간모형을 최적의 상태공간모형으로 선택한다.

상태공간모형이 구해지면 이 상태공간모형으로부터 VARMA모형을 유도할 수 있다. 그림 8.23.4에서 최종적으로 구한 상태공간모형은 다음과 같다.

$$\begin{aligned}
 y_t &= Y(t) - 0.211 \\
 z_{t+1} &= \hat{F}z_t + \hat{G}v_{t+1} \\
 y_t &= [1 \ 0]z_t \\
 \text{단, } z_t &= (y_t, y_{t+1|t})' \\
 \hat{F} &= \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \hat{G} = \begin{bmatrix} 1 \\ 0.664 \end{bmatrix}, \quad \hat{\Sigma} = [0.970]
 \end{aligned}$$

이 상태공간모형식을 전개하면 다음과 같다.

$$\begin{aligned}
 y_{t+1} &= y_{t+1|t} + v_{t+1} \\
 y_{t+2|t+1} &= 0.664v_{t+1}
 \end{aligned}$$

첫 번째 식을 두 번째 식에 대입하면 다음 식을 얻는다.

$$y_{t+2} = v_{t+2} + 0.664v_{t+1}$$

즉, 단변량시계열데이터 MA1A는 다음과 같은 MA(1)모형에서 발생되었다고 할 수 있다.

$$y_t = v_t + 0.664v_{t-1}$$

## 8.6.2 단변량시계열 LEADPROD

단변량시계열데이터 LEADPROD는 1981년 1월부터 1985년 12월까지의 미국의 월별 납생산량을 기록한 것이다. 이 단변량시계열데이터의 관찰점들의 개수는 60이다. 단변량시계열데이터 LEADPROD는 제 2.6절에 수록되어 있고, 또한 SAS데이터라이브러리 TSA4DATA의 파일 LEADPROD.SSD에 수록되어 있다. 이 단변량시계열데이터의 시계열산포도가 그림 2.6이다. 이 단변량시계열데이터를 상태공간모형화하기 위해서 다음 SAS 프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM8-25.PRG에 수록되어 있다).

```

/* */
/* EXAM8-25.PRG */
/* */
/* STATE SPACE MODELING */
/* DATASET = LEADPROD */
/* */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=72;
/* ----- */
/* DEFAULT MODEL */
/* ----- */
DATA EXAM825;
 SET TSA4DATA.LEADPROD;
RUN;
TITLE 'STATESPACE MODELING';
PROC STATESPACE CANCORR DATA=EXAM825;
 VAR LEADPROD;
RUN;

```

이 SAS프로그램을 수행한 결과가 그림 8.25이다. 그림 8.25의 부분 A에서 이 SAS데이터 세트의 관찰점들의 개수가 60, 부분 B에서 변수 LEADPROD의 표본평균이 44202.25, 표본표준편차가 8900.983임을 알 수 있다. 부분 C에는 VAR(0)모형, VAR(1)모형, ..., VAR(10)모형에 해당하는 AIC값들이 수록되어 있다. 이 중에서 VAR(1)모형의 AIC값이 최소이다. 즉, 상태공간모형화의 VAR모형단계에서 VAR(1)모형을 선택한다. 부분 D에는 VAR스태이트먼트에 지정된 변수 LEADPROD의 표본자기상관계수들의 도표적 표현이

```

STATESPACE MODELING
15
13:50 Sunday, August 1, 1993

STATESPACE Procedure

(A) Nobs = 60

(B) Variable Mean Std
LEADPROD 44202.25 8900.983

(C) Information Criterion for Autoregressive Models

Lag=0 Lag=1 Lag=2 Lag=3 Lag=4
1091.270042 1067.180457 1069.014375 1071.006646 1072.407068

Lag=5 Lag=6 Lag=7 Lag=8 Lag=9
1074.264767 1075.251757 1077.240791 1078.7936 1080.745158

Lag=10
1082.542523

(D) Schematic Representation of Correlations

Name/Lag 0 1 2 3 4 5 6 7 8 9 10
LEADPROD + + +

+ is > 2*std error, - is < -2*std error, . is between

(E) Schematic Representation of Partial Autocorrelations

Name/Lag 1 2 3 4 5 6 7 8 9 10
LEADPROD +

+ is > 2*std error, - is < -2*std error, . is between

(F) Yule-Walker Estimates for the Min AIC

Lag=1
LEADPROD
LEADPROD 0.59382

(G) Canonical Correlations Analysis

State vector Correlations Infor. Chisq D.F.
LEADPROD(T;T), 1.0000 0.0526 -1.83392 0.164698 1
LEADPROD(T+1;T)

(H) Selected Statespace Form and Preliminary Estimates

(H1) State Vector
LEADPROD(T;T)

(H2) Estimate of the Transition Matrix
0.594

(H3) Input Matrix for the Innovation
1

(H4) Variance Matrix for the Innovation
51290064.4

(I) Selected Statespace Form and Fitted Model

(I1) State Vector
LEADPROD(T;T)

```

|      |                                    |          |           |
|------|------------------------------------|----------|-----------|
| (I2) | Estimate of the Transition Matrix  |          |           |
|      | 0.594                              |          |           |
| (I3) | Input Matrix for the Innovation    |          |           |
|      | 1                                  |          |           |
| (I4) | Variance Matrix for the Innovation |          |           |
|      | 51290064.4                         |          |           |
| (J)  | Parameter Estimates                |          |           |
|      | Parameter                          | Estimate | Std. Err. |
|      | F(1,1)                             | 0.593821 | 0.103788  |
|      |                                    |          | T value   |
|      |                                    |          | 5.721498  |

그림 8.25 PROC STATESPACE를 사용한 LEADPROD.DAT의 상태공간모형화

수록되어 있고, 부분 E에는 이 변수의 표본부분자기상관계수들의 도표적 표현이 수록되어 있다. 이 도표적 표현들에 의해서 변수 LEADPROD가 AR(1)모형을 가짐을 확인할 수 있다. 부분 F에는 최소AIC값을 갖는 AR(1)모형의 AR계수들이 Yule-Walker방법으로 추정되었다. 부분 G는 PROC STATESPACE스테이트먼트의 옵션 CANCELL을 사용함으로써 출력된 것으로, 상태벡터들, 표본정준상관계수들, AIC값들, Bartlett통계량값들, 그리고 해당되는 자유도들이 출력되었다. 부분 H는 예비추정단계를 수행한 결과이다. 부분 H1은 예비추정단계의 상태벡터이고, 부분 H2는 변환행렬 F의 예비추정행렬이며, 부분 H3는 입력행렬 G의 예비추정행렬이며, 부분 H4는 쇠신과정의 공분산행렬의 예비추정행렬이다. 부분 I는 최종추정단계를 수행한 결과이다. 부분 I1은 최종추정단계의 상태벡터이다. 부분 I2는 변환행렬 F의 최종추정행렬이고, 부분 I3는 입력행렬 G의 최종추정행렬이며, 부분 I4는 쇠신과정의 공분산행렬의 최종추정행렬이다. 부분 J에는 최종추정된 원소 F(1, 1)의 최종추정값, 표준오차, 이에 해당하는 t통계량값이 수록되어 있다. 원소 F(1, 1)의 추정값이 0과 유의적으로 다르므로 다음과 같은 상태공간모형을 최종추정식으로 한다.

$$\begin{aligned}
 y_t &= \text{LEADPROD}(t) - 44202.25 \\
 z_{t+1} &= \hat{F}z_t + \hat{G}v_{t+1} \\
 y_t &= z_t \\
 \text{단, } \hat{F} &= [0.594], \quad \hat{G} = [1], \quad \hat{\Sigma} = [51290064.4]
 \end{aligned}$$

이 상태공간모형식을 전개하면 다음과 같다.

$$y_t = 0.594y_{t-1} + v_t$$

따라서 단변량시계열데이터 LEADPROD는 AR(1)모형을 이룬다고 할 수 있다.

## 8.6.3 다변량시계열데이터 MINKMUSK

다변량시계열데이터 MINKMUSK는 Canada의 Hudson's Bay Company가 1842년부터 1890년까지의 밍크(mink)와 사향취(muskkrat)의 털가죽(pelt)을 유럽에 수출한 양을 기록한 것이다. 이 다변량시계열데이터의 관찰점들의 개수는 49이고, 변수 MINK는 밍크가죽의 수를, 변수 MUSKRAT는 사향취가죽의 수를 기록한 것이다. 다변량시계열데이터 MINKMUSK는 제 2.8절에 수록되어 있으며 또한 SAS데이터라이브러리 TSA4DATA의 파일 MINKMUSK.SSD에 수록되어 있다. 이 다변량시계열데이터의 시계열산포도가 그림 2.8이다.

이 다변량시계열데이터를 상태공간모형화하기 위해서는 이 다변량시계열데이터를 정상 시계열데이터로 바꾸어야 한다. 제 2.8절의 그림 2.8에서 실선은 변수 MINK의 시계열산포도를 그린 것이고 점선은 변수 MUSKRAT의 시계열산포도를 그린 것이다. 두 변수의 분산들은 모두가 시간이 흐름에 따라 증가함을 알 수 있다. 이 분산의 이질성문제를 해결하기 위해서 이 변수들을 대수변환하기 위해 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM8-26.PRG에 수록되어 있다).

```

/ * */
/ * EXAM8-26.PRG */
/ * */
/ * LOG-TRANSFORMATION OF */
/ * DATASET = MINKMUSK */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS LINESIZE=72 PAGESIZE=60;
TITLE 'HUDSON' 'S BAY FUR TRAPPING RECORDS 1842-1890';
DATA EXAM826;
 SET TSA4DATA.MINKMUSK;
 LMINK = LOG(MINK);
 LMUSKRAT = LOG(MUSKRAT);
RUN;
SYMBOL1 C=BLACK V=NONE L=1 I=JOIN;
SYMBOL2 C=BLACK V=NONE L=4 I=JOIN;
AXIS1 ORDER=(1842 TO 1892 BY 5);
AXIS2 ORDER=(9.0 TO 14 BY 1);
PROC GPLOT DATA=EXAM826;
 PLOT LMINK*YEAR / HAXIS=AXIS1 VAXIS=AXIS2;
 PLOT2 LMUSKRAT*YEAR / HAXIS=AXIS1 VAXIS=AXIS2;
RUN;

```

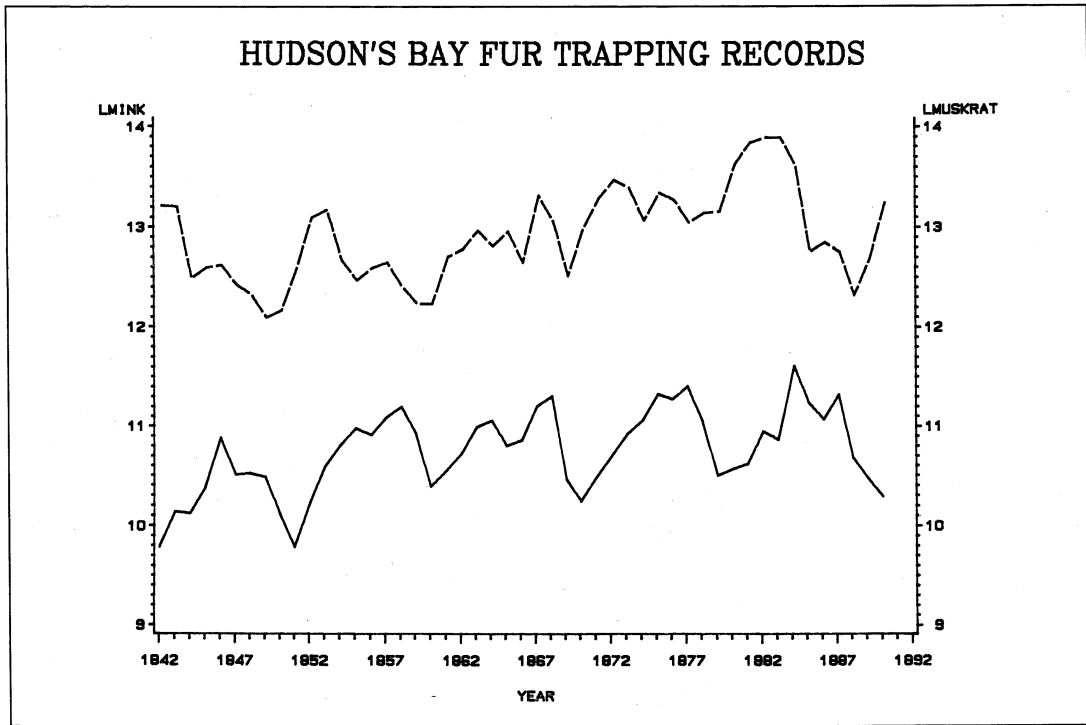


그림 8.26 대수변환된 변수들 LMINK와 LMUSKRAT의 시계열산포도들

이 SAS프로그램을 수행한 결과가 그림 8.26이다. 그림 8.26에서 실선은 변수 MINK를 대수변환한 변수 LMINK의 시계열산포도이고 점선은 변수 MUSKRAT를 대수변환한 변수 LMUSKRAT의 시계열산포도이다. 이 시계열산포도들의 분산들은 시간이 흘러도 거의 변화가 없음을 알 수 있다. 반면에 대수변환된 변수들 LMINK와 LMUSKRAT의 평균들은 시간이 흐름에 따라 증가함을 알 수 있다.

PROC STATESPACE를 사용하여 시계열데이터를 상태공간모형화하기 위해서는 시계열데이터가 정상적이어야 한다. 따라서 대수변환된 변수들 LMINK와 LMUSKRAT가 가지고 있는 추세를 제거해야 한다. 추세를 제거하기 위해서는 PROC REG를 사용하거나 PROC STATESPACE의 VAR스테이트먼트를 사용할 수 있다.

[BD86] (pp. 186~195)은 PROC REG를 적용해서 시계열데이터의 추세를 제거했다. 이 방법으로 추세를 제거한 뒤 상태공간모형화하기 위해서 다음 SAS프로그램을 수행해 보자 (이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM8-27.PRG에 수록되어 있다).

```

/* */
/* EXAM8-27.PRG */
/* */
/* STATE SPACE MODELING */
/* WITH REGRESSION DETREND */
/* DATASET = MINKMUSK */
LIBNAME TSA4DATA 'A:\TSA4DATA';

```



```

OPTIONS PAGESIZE=72;
TITLE 'HUDSON' 'S BAY FUR TRAPPING RECORDS 1842-1890';
DATA EXAM827;
 SET TSA4DATA.MINKMUSK;
 LMINK = LOG(MINK);
 LMUSKRAT = LOG(MUSKRAT);
 T+1;
 RUN;

/* ----- */
/* DETREND THE LOG-TRANSFORMED DATA */
/* AND APPLY THE STATESPACE MODEL */
/* ----- */
PROC REG DATA=EXAM827;
 MODEL LMINK LMUSKRAT = T;
 OUTPUT OUT=EXAM827B R=RLMINK RLMUSK;
 RUN;
PROC STATESPACE DATA=EXAM827B LEAD=10
 PRINT NOCENTER CANCELL;
 VAR RLMINK RLMUSK;
 RUN;

```

이 SAS프로그램의 첫 번째 PROC스텝은 다음과 같다.

```

PROC REG DATA=EXAM827;
 MODEL LMINK LMUSKRAT = T;
 OUTPUT OUT=EXAM827B R=RLMINK RLMUSK;
 RUN;

```

이 PROC스텝은 PROC REG를 사용하여 추세를 제거하여 정상시계열데이터를 만드는 것이다. 이 프로시저를 수행한 결과가 그림 8.27.1이다. 그림 8.27.1의 부분 A와 부분 B로부터 추세식들이 다음과 같음을 알 수 있다.

$$t = \text{YEAR} - 1841$$

$$\text{LMINK}(t) = 10.383 + 0.0139*t + \text{RLMINK}(t)$$

$$\text{LMUSKRAT}(t) = 12.499 + 0.0159*t + \text{RLMUSK}(t)$$

이 SAS프로그램의 두 번째 PROC스텝은 다음과 같다.

```

PROC STATESPACE DATA=EXAM827B LEAD=10
 PRINT NOCENTER CANCELL;
 VAR RLMINK RLMUSK;
 RUN;

```

| HUDSON'S BAY FUR TRAPPING RECORDS 1842-1890 |          |                    |                |                       | 9                            |
|---------------------------------------------|----------|--------------------|----------------|-----------------------|------------------------------|
|                                             |          |                    |                |                       | 19:05 Sunday, August 1, 1993 |
| Model: MODEL1                               |          |                    |                |                       |                              |
| (A) Dependent Variable: LMINK               |          |                    |                |                       |                              |
| Analysis of Variance                        |          |                    |                |                       |                              |
| Source                                      | DF       | Sum of Squares     | Mean Square    | F Value               |                              |
| Model                                       | 1        | 1.90597            | 1.90597        | 13.519                |                              |
| Error                                       | 47       | 6.62645            | 0.14099        |                       |                              |
| C Total                                     | 48       | 8.53242            |                |                       |                              |
|                                             | Root MSE | 0.37548            | R-square       | 0.2234                |                              |
|                                             | Dep Mean | 10.74528           | Adj R-sq       | 0.2069                |                              |
|                                             | C.V.     | 3.49441            |                |                       |                              |
| Parameter Estimates                         |          |                    |                |                       |                              |
| Variable                                    | DF       | Parameter Estimate | Standard Error | T for H0: Parameter=0 | Prob >  T                    |
| INTERCEP                                    | 1        | 10.382692          | 0.11226140     | 92.487                | 0.0001                       |
| T                                           | 1        | 0.013946           | 0.00379296     | 3.677                 | 0.0006                       |
| (B) Dependent Variable: LMUSKRAT            |          |                    |                |                       |                              |
| Analysis of Variance                        |          |                    |                |                       |                              |
| Source                                      | DF       | Sum of Squares     | Mean Square    | F Value               |                              |
| Model                                       | 1        | 2.49188            | 2.49188        | 15.164                |                              |
| Error                                       | 47       | 7.72320            | 0.16432        |                       |                              |
| C Total                                     | 48       | 10.21508           |                |                       |                              |
|                                             | Root MSE | 0.40537            | R-square       | 0.2439                |                              |
|                                             | Dep Mean | 12.91312           | Adj R-sq       | 0.2279                |                              |
|                                             | C.V.     | 3.13920            |                |                       |                              |
| Parameter Estimates                         |          |                    |                |                       |                              |
| Variable                                    | DF       | Parameter Estimate | Standard Error | T for H0: Parameter=0 | Prob >  T                    |
| INTERCEP                                    | 1        | 12.498523          | 0.12119615     | 103.126               | 0.0001                       |
| T                                           | 1        | 0.015946           | 0.00409484     | 3.894                 | 0.0003                       |

그림 8.27.1 PROC REG를 사용한 추세의 제거

이 프로시저는 그림 8.27.1의 잔차들 RLMINK와 RLMUSK를 상태공간모형화하기 위한 것이다. 이 프로시저를 수행한 결과가 그림 8.27.2이다. 그림 8.27.2의 부분 A에서 이 SAS데이터세트의 관찰점들의 개수가 49, 변수들 RLMINK와 RLMUSK의 표본표준편차들이 각각 0.368과 0.397임을 알 수 있다. PROC STATESPACE스태이트먼트의 옵션 NOCENTER를 사용했으므로 표본평균들은 0으로 간주된다. VAR모형들 중에서 VAR(1)모형이 최소의 AIC값을 갖는다. 즉, 상태공간모형화의 VAR모형단계에서 VAR(1)모

HUDSON'S BAY FUR TRAPPING RECORDS 1842-1890

11

19:05 Sunday, August 1, 1993

(A)

STATESPACE Procedure

Nobs = 49

| Variable | Std      |
|----------|----------|
| RLMINK   | 0.367741 |
| RLMUSK   | 0.397009 |

Information Criterion for Autoregressive Models

| Lag=0       | Lag=1       | Lag=2       | Lag=3       | Lag=4       |
|-------------|-------------|-------------|-------------|-------------|
| -188.624096 | -251.778386 | -245.704605 | -242.548102 | -236.654144 |
| Lag=5       | Lag=6       | Lag=7       | Lag=8       | Lag=9       |
| -230.490470 | -229.249919 | -227.126909 | -225.203450 | -221.890966 |
| Lag=10      |             |             |             |             |
| -221.111675 |             |             |             |             |

Schematic Representation of Correlations

| Name/Lag | 0 | 1 | 2 | 3 | 4  | 5  | 6 | 7 | 8 | 9  | 10 |
|----------|---|---|---|---|----|----|---|---|---|----|----|
| RLMINK   | + | + | . | . | -  | -  | - | - | - | +  | +  |
| RLMUSK   | . | + | - | - | .. | .. | + | + | + | .. | .. |

+ is > 2\*std error, - is < -2\*std error, . is between

Schematic Representation of Partial Autocorrelations

| Name/Lag | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 |
|----------|----|----|----|----|----|----|----|----|----|----|
| RLMINK   | +  | .. | .. | .. | .. | .. | .. | .. | .. | .. |
| RLMUSK   | -+ | .. | .. | .. | .. | .. | .. | .. | .. | .. |

+ is > 2\*std error, - is < -2\*std error, . is between

Yule-Walker Estimates for the Min AIC

|        | Lag=1    |         |
|--------|----------|---------|
|        | RLMINK   | RLMUSK  |
| RLMINK | 0.56875  | 0.29826 |
| RLMUSK | -0.46839 | 0.62716 |

(B)

Canonical Correlations Analysis

| State vector                              | Correlations |        |        | Infor.   | Chisq    | D.F. |
|-------------------------------------------|--------------|--------|--------|----------|----------|------|
| RLMINK(T;T),RLMUSK(T;T),<br>RLMINK(T+1;T) | 1.0000       | 1.0000 | 0.1080 | -3.42514 | 0.56313  | 2    |
| RLMINK(T;T),RLMUSK(T;T),<br>RLMUSK(T+1;T) | 1.0000       | 1.0000 | 0.1627 | -2.68591 | 1.287271 | 2    |

(C)

Selected Statespace Form and Preliminary Estimates

State Vector

| RLMINK(T;T) | RLMUSK(T;T) |
|-------------|-------------|
|-------------|-------------|

Estimate of the Transition Matrix

|        |       |
|--------|-------|
| 0.569  | 0.298 |
| -0.468 | 0.627 |

Input Matrix for the Innovation

|   |   |
|---|---|
| 1 | 0 |
| 0 | 1 |

|     |                                           |            |             |          |
|-----|-------------------------------------------|------------|-------------|----------|
|     | Variance Matrix for the Innovation        |            |             |          |
|     |                                           | 0.07913099 | 0.00270282  |          |
|     |                                           | 0.00270282 | 0.06307163  |          |
| (D) | Selected Statespace Form and Fitted Model |            |             |          |
|     | State Vector                              |            |             |          |
|     | RLMINK(T;T)                               |            | RLMUSK(T;T) |          |
|     | Estimate of the Transition Matrix         |            |             |          |
|     |                                           | 0.569      | 0.298       |          |
|     |                                           | -0.468     | 0.627       |          |
|     | Input Matrix for the Innovation           |            |             |          |
|     |                                           | 1          | 0           |          |
|     |                                           | 0          | 1           |          |
|     | Variance Matrix for the Innovation        |            |             |          |
|     |                                           | 0.07913099 | 0.00270282  |          |
|     |                                           | 0.00270282 | 0.06307163  |          |
| (E) | Parameter Estimates                       |            |             |          |
|     | Parameter                                 | Estimate   | Std. Err.   | T value  |
|     | F(1,1)                                    | 0.568749   | 0.109253    | 5.205807 |
|     | F(1,2)                                    | 0.298262   | 0.101203    | 2.947173 |
|     | F(2,1)                                    | -0.468359  | 0.097544    | -4.80183 |
|     | F(2,2)                                    | 0.627158   | 0.090356    | 6.940953 |

그림 8.27.2 PROC STATESPACE를 사용한 잔차들의 상태공간모형화

형을 선택한다. 이러한 결론의 타당성은 VAR스태이트먼트에 지정된 변수들 RLMINK와 RLMUSK의 표본자기상관행렬들의 도표적 표현과 표본부분자기상관행렬들의 도표적 표현에서 확인할 수 있다. 또한 이 VAR(1)모형의 VAR계수행렬들의 Yule-Walker추정행렬들이 출력되었다. 부분 B에는 PROC STATESPACE스태이트먼트의 옵션 CANCELL을 사용함으로써 출력된 것으로, 상태벡터들, 표본정준상관계수들, AIC값들, Bartlett통계량 값들, 그리고 해당되는 자유도들이 출력되었다. 부분 C에서 상태벡터가 다음과 같다는 것을 알 수 있다.

$$(RLMINK(t;t), RLMUSK(t;t))'$$

부분 C에는 예비추정단계를 수행한 결과가 실려 있다. 즉, 예비추정단계의 상태벡터, 변환행렬  $F$ 의 예비추정행렬, 입력행렬  $G$ 의 예비추정행렬, 쇄신과정의 공분산행렬의 예비추정행렬이 실려 있다. 부분 D에는 최종추정단계를 수행한 결과가 실려 있다. 즉, 최종추정단계의 상태벡터, 변환행렬  $F$ 의 최종추정행렬, 입력행렬  $G$ 의 최종추정행렬, 쇄신과정의 공분산행렬의 최종추정행렬이 수록되어 있다. 부분 E는 행렬들  $F$ 와  $G$ 의 원소들의 최종추정값들, 표준오차들, 이에 해당하는  $t$ 통계량값들이 수록되어 있다. 이 추정된 원소들의 최종추정값들의  $t$ 통계량값들이 유의적으로 0과 다르므로 이 상태공간모형에 RESTRICT스태이트먼트를 적용할 필요가 없다.

그림 8.27.2에서 최종적으로 구한 상태공간모형은 다음과 같다.

$$y_{1,t} = \text{RLMINK}(t)$$

$$y_{2,t} = \text{RLMUSK}(t)$$

$$z_{t+1} = \hat{F}z_t + \hat{G}v_{t+1}$$

$$y_t = z_t$$

$$\text{단, } z_t = (y_{1,t}, y_{2,t})'$$

$$\hat{F} = \begin{bmatrix} 0.569 & 0.298 \\ -0.468 & 0.627 \end{bmatrix}, \quad \hat{G} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \hat{\Sigma} = \begin{bmatrix} 0.079 & 0.003 \\ 0.003 & 0.063 \end{bmatrix}$$

이 식을 정리하면 다음과 같은 2변량VAR(1)모형을 얻는다.

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 0.569 & 0.298 \\ -0.468 & 0.627 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix}$$

지금부터는 추세를 제거하기 위해서 PROC REG를 사용하는 대신에 PROC STATESPACE의 VAR스테이트먼트를 사용하기로 하자. 이러한 목적으로 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM8-28.PRG에 수록되어 있다).

```

/* */
/* EXAM8-28.PRG */
/* */
/* STATE SPACE MODELING */
/* WITH VAR DETREND */
/* DATASET = MINKMUSK */
/* */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=60 LINESIZE=72;
TITLE 'HUDSON' 'S BAY FUR TRAPPING RECORDS 1842-1890';
DATA EXAM827;
 SET TSA4DATA.MINKMUSK;
 LMINK = LOG(MINK);
 LMUSKRAT = LOG(MUSKRAT);
 T+1;
 RUN;
PROC STATESPACE NOCENTER NOEST CANCORR;
 VAR LMINK(1) LMUSKRAT(1);
 RUN;
PROC STATESPACE NOCENTER NOEST CANCORR PASTMIN=1;
 VAR LMINK(1) LMUSKRAT(1);

```

```

RUN;
PROC STATESPACE NOCENTER NOEST CANCECORR PASTMIN=2;
VAR LMINK(1) LMUSKRAT(1);
RUN;
PROC STATESPACE NOCENTER NOEST CANCECORR PASTMIN=3;
VAR LMINK(1) LMUSKRAT(1);
RUN;
PROC STATESPACE NOCENTER NOEST CANCECORR PASTMIN=4;
VAR LMINK(1) LMUSKRAT(1);
RUN;

```

이 SAS프로그램은 상태공간모형화를 사용하여 대수변환된 변수들 LMINK와 LMUSKRAT에 적당한 모형을 선택하기 위한 준비단계이다. 앞에서 언급한 대로 변수들 LMINK와 LMUSKRAT의 추세를 제거하기 위해서 VAR스테이트먼트에서 차분을 취한다. 이 SAS프로그램이 상태공간모형화를 위한 준비단계이고 PROC STATESPACE의 최종추정 단계는 계산적으로 비싸기 때문에 최종추정단계를 수행하지 않도록 하기 위해서 PROC STATESPACE들에서 옵션 NOEST를 사용하였다. 이 프로시저들을 수행한 결과의 일부가 그림 8.28이다.

첫 번째 PROC STATESPACE를 수행한 결과의 일부가 그림 8.28의 부분 A와 부분 B에 수록되어 있다. 그림 8.28의 부분 A에는 관찰점들의 개수, 대수변환된 변수들의 표본평균들, 표본표준편차들, AIC값들, 표본자기상관행렬들과 표본부분자기상관행렬들의 도표적 표현들, 그리고 최소AIC값을 갖는 VAR모형인 VAR(1)모형의 VAR계수행렬의 Yule-Walker추정행렬이 수록되어 있다. 부분 B의 앞부분은 PROC STATESPACE스테이트먼트의 옵션 CANCECORR를 사용함으로써 출력된 것으로, 상태벡터들, 표본정준상관계수들, AIC값들, Bartlett통계량값들, 그리고 해당되는 자유도들이 출력되었다. 여기서 유의할 점은 상태벡터 '(LMINK( $t:t$ ), LMUSKRAT( $t:t$ ), LMUSKRAT( $t+1:t$ ))'의 최소의 표본정준상관계수가 0.3409이고 이에 해당하는 DIC, 즉 변수 Infor.의 값이 1.928로 양수임에도 불구하고, 상태벡터가(LMINK( $t:t$ ), LMUSKRAT( $t:t$ ), LMUSKRAT( $t+1:t$ ), LMUSKRAT( $t+2:t$ ))일 가능성을 조사하지 않았다는 것이다. 이는 부분 A에서 볼 수 있듯이 여러 VAR모형들 중에서 VAR(1)모형이 최소AIC값을 가지므로 정준상관분석단계에서 사용하는 데이터공간을 구성하는 과거항들의 시차들이 0과 1이기 때문이다. 부분 B의 뒷부분에는 예비추정단계를 수행한 결과가 실려 있다. 즉, 예비추정단계의 상태벡터, 변환행렬  $F$ 의 예비추정행렬, 입력행렬  $G$ 의 예비추정행렬, 쇄신과정의 공분산행렬의 예비추정행렬이 실려 있다. 여기서 다시 한 번 주의할 점은 상태벡터가 '(LMINK( $t:t$ ), LMUSKRAT( $t:t$ ))'라는 것이다.

두 번째 PROC STATESPACE에는 PROC STATESPACE스테이트먼트의 옵션 PASTMIN=에 1이 할당되어 있다. 이 옵션 PASTMIN=은 데이터공간을 구성하는 과거항들의 시차들의 최대값의 하한을 지정하는 것이다. 따라서 이 PROC스텝에서는 데이터공간

HUDSON'S BAY FUR TRAPPING RECORDS 1842-1890

13

15:43 Wednesday, August 4, 1993

(A)

STATESPACE Procedure

Nobs = 48

Variable Std  
 LMINK 0.317705  
 LMINK has been differenced.  
 The periods are 1.  
 LMUSKRAT 0.324383  
 LMUSKRAT has been differenced.  
 The periods are 1.

Information Criterion for Autoregressive Models

|                       |                      |                      |                      |                      |
|-----------------------|----------------------|----------------------|----------------------|----------------------|
| Lag=0<br>-220.187082  | Lag=1<br>-241.667489 | Lag=2<br>-240.757380 | Lag=3<br>-239.496722 | Lag=4<br>-232.413712 |
| Lag=5<br>-227.937115  | Lag=6<br>-225.578113 | Lag=7<br>-221.503922 | Lag=8<br>-218.812578 | Lag=9<br>-224.681773 |
| Lag=10<br>-222.133207 |                      |                      |                      |                      |

Schematic Representation of Correlations

| Name/Lag | 0 | 1 | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 |
|----------|---|---|----|----|----|----|----|----|----|----|----|
| LMINK    | + | + | .. | .. | .. | .. | .. | .. | .. | +  | +  |
| LMUSKRAT | + | - | -  | .. | .. | .. | .. | .. | .. | .. | .. |

+ is > 2\*std error, - is < -2\*std error, . is between

Schematic Representation of Partial Autocorrelations

| Name/Lag | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 |
|----------|----|----|----|----|----|----|----|----|----|----|
| RLMINK   | .. | .. | .. | .. | .. | .. | .. | .. | +  | .. |
| LMUSKRAT | -  | -  | .. | .. | .. | .. | .. | .. | .. | .. |

+ is > 2\*std error, - is < -2\*std error, . is between

Yule-Walker Estimates for the Min AIC

|          |          |          |
|----------|----------|----------|
|          | Lag=1    |          |
|          | LMINK    | LMUSKRAT |
| LMINK    | 0.00621  | 0.28874  |
| LMUSKRAT | -0.64978 | 0.17579  |

(B)

STATESPACE Procedure

Canonical Correlations Analysis

| State vector                                 | Correlations  | Infor. | Chisq    | D.F.     |   |
|----------------------------------------------|---------------|--------|----------|----------|---|
| LMINK(T;T),LMUSKRAT(T;T),<br>LMINK(T+1;T)    | 1.0000 1.0000 | 0.0546 | -3.85663 | 0.140386 | 2 |
| LMINK(T;T),LMUSKRAT(T;T),<br>LMUSKRAT(T+1;T) | 1.0000 1.0000 | 0.3409 | 1.929268 | 5.805742 | 2 |

Selected Statespace Form and Preliminary Estimates

State Vector

LMINK(T;T) LMUSKRAT(T;T)

Estimate of the Transition Matrix

|        |       |
|--------|-------|
| 0.006  | 0.289 |
| -0.650 | 0.176 |

Input Matrix for the Innovation

|   |   |
|---|---|
| 1 | 0 |
| 0 | 1 |

Variance Matrix for the Innovation

|            |            |
|------------|------------|
| 0.09208463 | 0.01995379 |
| 0.01995379 | 0.06414716 |

(C)

STATESPACE Procedure

Canonical Correlations Analysis

| State vector                                 | Correlations |        |        | Infor.   | Chisq    | D.F. |
|----------------------------------------------|--------------|--------|--------|----------|----------|------|
| LMINK(T;T),LMUSKRAT(T;T),<br>LMINK(T+1;T)    | 1.0000       | 1.0000 | 0.0546 | -3.85663 | 0.140386 | 2    |
| LMINK(T;T),LMUSKRAT(T;T),<br>LMUSKRAT(T+1;T) | 1.0000       | 1.0000 | 0.3409 | 1.929268 | 5.805742 | 2    |

Selected Statespace Form and Preliminary Estimates

State Vector

|            |               |
|------------|---------------|
| LMINK(T;T) | LMUSKRAT(T;T) |
|------------|---------------|

Estimate of the Transition Matrix

|        |       |
|--------|-------|
| 0.006  | 0.289 |
| -0.650 | 0.176 |

Input Matrix for the Innovation

|   |   |
|---|---|
| 1 | 0 |
| 0 | 1 |

Variance Matrix for the Innovation

|            |            |
|------------|------------|
| 0.09208463 | 0.01995379 |
| 0.01995379 | 0.06414716 |

(D)

STATESPACE Procedure

Canonical Correlations Analysis

| State vector                                                     | Correlations |        |                  | Infor.   | Chisq    | D.F. |
|------------------------------------------------------------------|--------------|--------|------------------|----------|----------|------|
| LMINK(T;T),LMUSKRAT(T;T),<br>LMINK(T+1;T)                        | 1.0000       | 1.0000 | 0.2489           | -4.9298  | 2.942273 | 4    |
| LMINK(T;T),LMUSKRAT(T;T),<br>LMUSKRAT(T+1;T)                     | 1.0000       | 1.0000 | 0.4241           | 1.515432 | 9.118956 | 4    |
| LMINK(T;T),LMUSKRAT(T;T),<br>LMUSKRAT(T+1;T),<br>LMUSKRAT(T+2;T) | 1.0000       | 1.0000 | 0.5079<br>0.1488 | -4.92467 | 1.041726 | 3    |

Selected Statespace Form and Preliminary Estimates

State Vector

|            |               |                 |
|------------|---------------|-----------------|
| LMINK(T;T) | LMUSKRAT(T;T) | LMUSKRAT(T+1;T) |
|------------|---------------|-----------------|

Estimate of the Transition Matrix

|       |        |       |
|-------|--------|-------|
| 0.006 | 0.289  | 0     |
| 0     | 0      | 1     |
| 0.052 | -0.255 | 0.586 |



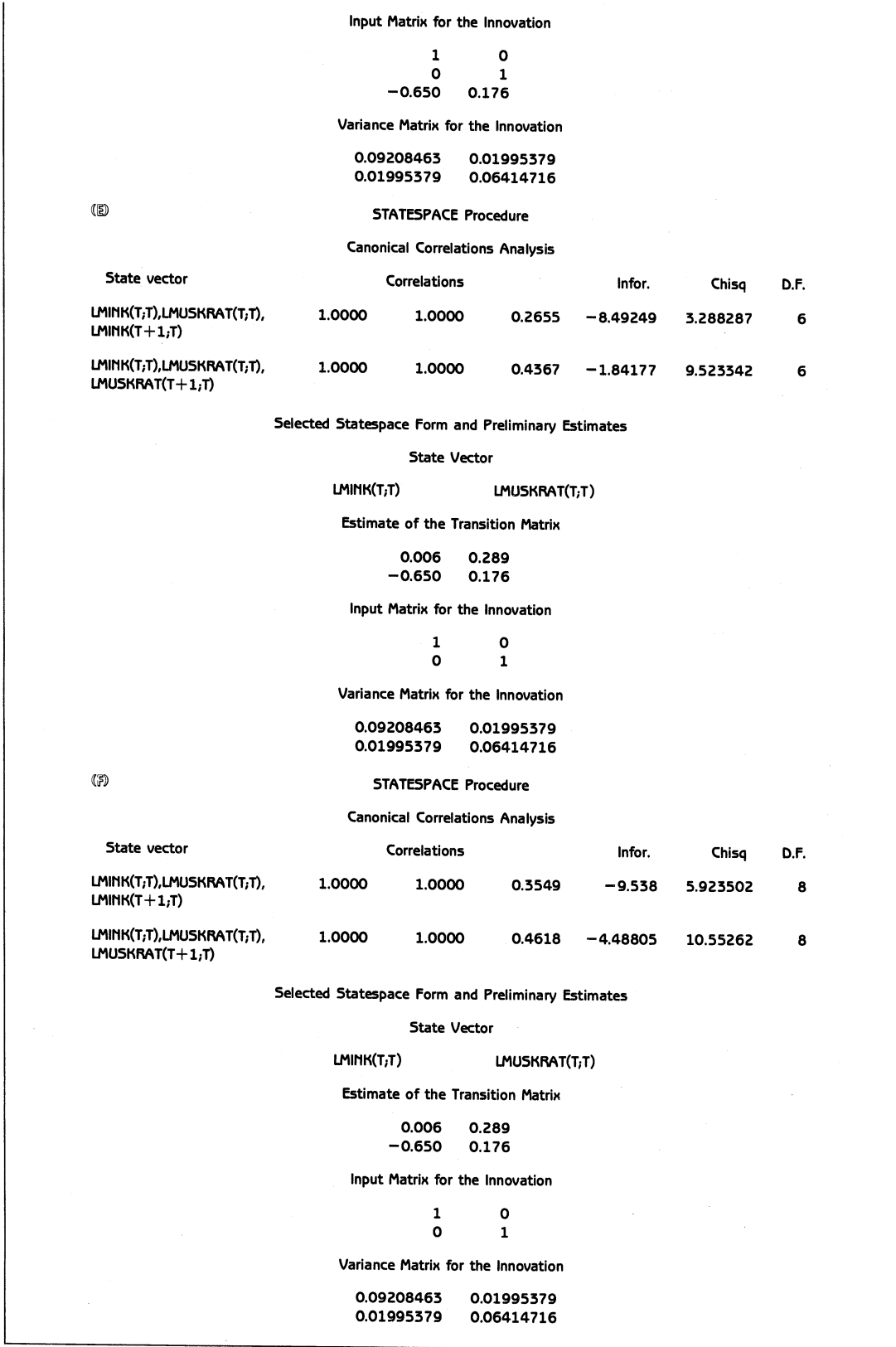


그림 8.28 옵션 PASTMIN =을 사용한 PROC STATESPACE들의 결과

에  $y_t$ 와  $y_{t+1}$ 가 포함되어야 한다. 이 PROC스텝을 수행한 결과의 일부가 그림 8.28의 부분 C에 실려 있다. 그러나 부분 C는 부분 B와 다르지 않다. 그 이유는 부분 B의 데이터공간에도  $y_t$ 와  $y_{t+1}$ 가 포함되었기 때문이다.

세 번째 PROC STATESPACE에는 PROC STATESPACE스테이트먼트의 옵션 PASTMIN=에 2가 할당되어 있다. 따라서 이 PROC스텝에서는 데이터공간에  $y_t, y_{t+1}, y_{t+2}$ 가 포함되어야 한다. 이 PROC스텝을 수행한 결과의 일부가 그림 8.28의 부분 D에 실려 있다. 부분 D의 앞부분은 PROC STATESPACE스테이트먼트의 옵션 CANCELL을 사용함으로써 출력된 것으로서, 상태벡터들, 표본정준상관계수들, AIC값들, Bartlett통계량값들, 그리고 해당되는 자유도들이 출력되었다. 이 부분에서 상태벡터가 '(LMINK( $t:t$ ), LMUSKRAT( $t:t$ ), LMUSKRAT( $t+1:t$ ))'임을 알 수 있다. 부분 D의 뒷부분에는 예비추정단계를 수행한 결과가 실려 있다. 즉, 예비추정단계의 상태벡터, 변환행렬  $F$ 의 예비추정행렬, 입력행렬  $G$ 의 예비추정행렬, 쇄신과정의 공분산행렬의 예비추정행렬이 실려 있다.

네 번째 PROC STATESPACE에는 PROC STATESPACE스테이트먼트의 옵션 PASTMIN=에 3이 할당되어 있다. 따라서 이 PROC스텝에서는 데이터공간에  $y_t, y_{t+1}, y_{t+2}, y_{t+3}$ 가 포함되어야 한다. 이 PROC스텝을 수행한 결과의 일부가 그림 8.28의 부분 E에 실려 있다. 부분 E에서 상태벡터는 '(LMINK( $t:t$ ), LMUSKRAT( $t:t$ ))'이다. 이 상태벡터에 다른 원소를 추가하면 새로운 상태벡터의 최소의 정준상관계수에 해당하는 DIC, 즉 변수 Infor.의 값들이 모두 음수이다. 따라서 이 단계에서 상태공간벡터의 원소들의 추가선택을 멈춘다.

다섯 번째 PROC STATESPACE에는 PROC STATESPACE스테이트먼트의 옵션 PASTMIN=에 4가 할당되어 있다. 따라서 이 PROC스텝에서는 데이터공간에는  $y_t, y_{t+1}, y_{t+2}, y_{t+3}, y_{t+4}$ 가 포함되어야 한다. 이 PROC스텝을 수행한 결과의 일부가 그림 8.28의 부분 F에 실려 있다. 부분 F의 상태벡터는 '(LMINK( $t:t$ ), LMUSKRAT( $t:t$ ))'이다. 이 상태벡터에 다른 원소를 추가하면 새로운 상태벡터의 최소의 정준상관계수에 해당하는 DIC, 즉 변수 Infor.의 값들이 모두 음수이다. 따라서 이 단계에서 상태공간벡터의 원소들의 추가선택을 멈춘다.

그림 8.28에서 이 예제에서는 옵션 PASTMIN=에 3 또는 2를 할당하는 것이 타당함을 알 수 있다. 물론 옵션 PASTMIN=에 4를 할당할 수도 있으나 최종적인 상태공간모형이 3을 할당했을 때와 같아진다. 타당한 상태공간모형을 선택하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM8-29.PRG에 수록되어 있다).

```

/* */
/* EXAM8-29.PRG */
/* */
/* STATE SPACE MODELING */
/* DATASET = MINKMUSK */

```

```

LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=60 LINESIZE=72;
TITLE 'HUDSON' 'S BAY FUR TRAPPING RECORDS 1842-1890';
DATA EXAM829;
 SET TSA4DATA.MINKMUSK;
 LMINK = LOG(MINK);
 LMUSKRAT = LOG(MUSKRAT);
 T+1;
 RUN;
TITLE 'MODEL 3-0';
PROC STATESPACE NOCENTER PASTMIN=3;
 VAR LMINK(1) LMUSKRAT(1);
 RUN;
TITLE 'MODEL 3-1';
PROC STATESPACE NOCENTER PASTMIN=3;
 VAR LMINK(1) LMUSKRAT(1);
 RESTRICT F(1,1)=0;
 RUN;
TITLE 'MODEL 2-0';
PROC STATESPACE NOCENTER PASTMIN=2;
 VAR LMINK(1) LMUSKRAT(1);
 RUN;
TITLE 'MODEL 2-1';
PROC STATESPACE NOCENTER PASTMIN=2;
 VAR LMINK(1) LMUSKRAT(1);
 RESTRICT F(1,1)=0;
 RUN;
TITLE 'MODEL 2-2';
PROC STATESPACE NOCENTER PASTMIN=2;
 VAR LMINK(1) LMUSKRAT(1);
 RESTRICT F(1,1)=0 F(3,1)=0;
 RUN;
TITLE 'MODEL 2-3';
PROC STATESPACE NOCENTER PASTMIN=2;
 VAR LMINK(1) LMUSKRAT(1);
 RESTRICT F(1,1)=0 F(3,1)=0 G(3,2)=0;
 RUN;

```

이 SAS프로그램을 수행한 결과의 일부가 그림 8.29이다. 첫 번째 PROC스텝에서 PROC

(A)

MODEL 3-0

3

17:37 Wednesday, August 4, 1993

Variance Matrix for the Innovation

0.09208463 0.01995379  
0.01995379 0.06414716

Parameter Estimates

| Parameter | Estimate | Std. Err. | T value  |
|-----------|----------|-----------|----------|
| F(1,1)    | 0.006214 | 0.14065   | 0.04418  |
| F(1,2)    | 0.288741 | 0.137758  | 2.096006 |
| F(2,1)    | -0.64978 | 0.117408  | -5.53441 |
| F(2,2)    | 0.175794 | 0.114992  | 1.528745 |

(B)

MODEL 3-1

7

17:37 Wednesday, August 4, 1993

Variance Matrix for the Innovation

0.09208837 0.0199546  
0.0199546 0.06414733

Parameter Estimates

| Parameter | Estimate | Std. Err. | T value  |
|-----------|----------|-----------|----------|
| F(1,2)    | 0.289975 | 0.133659  | 2.169511 |
| F(2,1)    | -0.65113 | 0.112527  | -5.78638 |
| F(2,2)    | 0.176059 | 0.113847  | 1.546447 |

(C)

MODEL 2-0

9

17:37 Wednesday, August 4, 1993

Variance Matrix for the Innovation

0.09208463 0.02122177  
0.02122177 0.05295349

Parameter Estimates

| Parameter | Estimate | Std. Err. | T value  |
|-----------|----------|-----------|----------|
| F(1,1)    | 0.006215 | 0.140807  | 0.044136 |
| F(1,2)    | 0.288741 | 0.137908  | 2.093712 |
| F(3,1)    | -0.14121 | 0.251504  | -0.56145 |
| F(3,2)    | -0.1898  | 0.14993   | -1.26594 |
| F(3,3)    | 0.331558 | 0.275076  | 1.205332 |
| G(3,1)    | -0.55769 | 0.115005  | -4.8493  |
| G(3,2)    | -0.17561 | 0.14744   | -1.19108 |

(D)

MODEL 2-1

13

17:37 Wednesday, August 4, 1993

Variance Matrix for the Innovation

0.09208837 0.02122263  
0.02122263 0.05295369

Parameter Estimates

| Parameter | Estimate | Std. Err. | T value  |
|-----------|----------|-----------|----------|
| F(1,2)    | 0.289979 | 0.135025  | 2.147597 |
| F(3,1)    | -0.13701 | 0.233467  | -0.58687 |
| F(3,2)    | -0.19105 | 0.147395  | -1.2962  |
| F(3,3)    | 0.331843 | 0.274996  | 1.206721 |
| G(3,1)    | -0.55912 | 0.110334  | -5.06758 |
| G(3,2)    | -0.17533 | 0.147299  | -1.19028 |

|     |                                    |                                 |           |          |
|-----|------------------------------------|---------------------------------|-----------|----------|
| (E) | MODEL 2-2                          | 17                              |           |          |
|     |                                    | 17:37 Wednesday, August 4, 1993 |           |          |
|     | Variance Matrix for the Innovation |                                 |           |          |
|     | 0.09208837                         | 0.02076825                      |           |          |
|     | 0.02076825                         | 0.05318102                      |           |          |
|     | Parameter Estimates                |                                 |           |          |
|     | Parameter                          | Estimate                        | Std. Err. | T value  |
|     | F(1,2)                             | 0.289976                        | 0.135028  | 2.147522 |
|     | F(3,2)                             | -0.22629                        | 0.139371  | -1.62369 |
|     | F(3,3)                             | 0.464935                        | 0.123319  | 3.77019  |
|     | G(3,1)                             | -0.59059                        | 0.102966  | -5.73575 |
|     | G(3,2)                             | -0.1286                         | 0.134293  | -0.95764 |
| (F) | MODEL 2-3                          | 21                              |           |          |
|     |                                    | 17:37 Wednesday, August 4, 1993 |           |          |
|     | Variance Matrix for the Innovation |                                 |           |          |
|     | 0.09219234                         | 0.02120271                      |           |          |
|     | 0.02120271                         | 0.05433557                      |           |          |
|     | Parameter Estimates                |                                 |           |          |
|     | Parameter                          | Estimate                        | Std. Err. | T value  |
|     | F(1,2)                             | 0.321414                        | 0.131895  | 2.436897 |
|     | F(3,2)                             | -0.28612                        | 0.131018  | -2.1838  |
|     | F(3,3)                             | 0.51488                         | 0.127043  | 4.052798 |
|     | G(3,1)                             | -0.60144                        | 0.104764  | -5.74085 |

그림 8.29 옵션 RESTRICT=을 사용한 최종추정단계들

STATESPACE스테이트먼트의 옵션 PASTMIN=에 3을 할당하였다. 이 PROC스텝을 수행한 결과물 중에서 최종추정단계의 결과가 그림 8.29의 부분 A이다. 부분 A에서 모수 F(1, 1)의 추정값이 유의적으로 0과 다르지 않음을 알 수 있다. 따라서 두 번째 PROC STATESPACE의 RESTRICT스테이트먼트에 'F(1, 1)=0'을 지정하였다. 이 PROC스텝을 수행한 결과물 중에서 최종추정단계의 결과가 그림 8.29의 부분 B이다. 부분 B에서 모든 모수들의 추정값들이 유의적으로 0과 다르다. 따라서 여기서 후향제거방법을 멈추도록 한다. 세 번째 PROC스텝에서 PROC STATESPACE스테이트먼트의 옵션 PASTMIN=에 2를 할당하였다. 이 PROC스텝을 수행한 결과물 중에서 최종추정단계의 결과가 그림 8.29의 부분 C이다. 부분 C에서 모수 F(1, 1)의 추정값이 유의적으로 0과 다르지 않음을 알 수 있다. 따라서 네 번째 PROC STATESPACE의 RESTRICT스테이트먼트에 'F(1, 1)=0'을 지정하였다. 이 PROC스텝을 수행한 결과 중에서 최종추정단계의 결과물이 그림 8.29의 부분 D이다. 부분 D에서 모수 F(3, 1)의 추정값이 유의적으로 0과 다르지 않음을 알 수 있다. 따라서 다섯 번째 PROC STATESPACE의 RESTRICT스테이트먼트에 'F(1, 1)=0 F(3, 1)=0'을 지정하였다. 이 PROC스텝을 수행한 결과 중에서 최종추정단계의 결과물이 그림 8.29의 부분 E이다. 부분 E에서 모수 G(3, 2)의 추정값이 유의적으로 0과 다르지 않음을 알 수 있다. 따라서 여섯 번째 PROC STATESPACE의 RESTRICT스테이트먼트에 'F(1, 1)=0 F(3, 1)=0 G(3, 2)=0'을 지정하였다. 이 PROC스텝을 수행한 결과

중에서 최종추정단계의 결과물이 그림 8.29의 부분 Ⅱ이다. 부분 Ⅱ에서 모든 모수들의 추정값들이 유의적으로 0과 다르다. 따라서 여기서 후향제거방법을 멈추도록 한다.

좀더 객관적으로 상태공간모형을 식별하기 위해서 벌칙함수식별방법들을 적용하기로 하자. 이러한 목적으로 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4 TSA4의 파일 EXAM8-30.PRG에 수록되어 있다).

```

/ * */
/ * EXAM8-30.PRG */
/ * */
/ * CHOOSING RESTRICT CONDITIONS */
/ * USING THE PENALTY FUNTION METHODS */
OPTIONS LINESIZE=72 PAGESIZE=60;
PROC IML;
RESET NOLOG NOCENTER NONAME;
T = 49;
/ * WITH PASTMIN=3 */
PSTMIN=3;
K=4;
/ * WITHOUT ANY RESTRICTION */
COVAR = { 0.09208463 0.01995379,
 0.01995379 0.06414716 };
NORES = 0;
AIC = LOG(DET(COVAR)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COVAR)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COVAR)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT, "WITH PASTMIN = " PSTMIN " "
 "NO OF RESTRICTION = " NORES,
 " AIC = " AIC,
 " BIC = " BIC,
 " HQC = " HQC;
/ * WITH 1 RESTRICTION */
COVAR = { 0.09208837 0.0199546,
 0.0199546 0.06414733 };
NORES = 1;
AIC = LOG(DET(COVAR)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COVAR)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COVAR)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT, "WITH PASTMIN = " PSTMIN " "
 "NO OF RESTRICTION = " NORES,

```

```

" AIC = " AIC,
" BIC = " BIC,
" HQC = " HQC;
/* WITH PASTMIN=2 */
PSTMIN=2;
K=7;
/* WITHOUT ANY RESTRICTION */
COVAR = { 0.09208463 0.02122177,
 0.02122177 0.05295349 };
NORES = 0;
AIC = LOG(DET(COVAR)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COVAR)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COVAR)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT , "WITH PASTMIN = " PSTMIN " "
 "NO OF RESTRICTION = " NORES,
 " AIC = " AIC,
 " BIC = " BIC,
 " HQC = " HQC;
/* WITH 1 RESTRICTION */
COVAR = { 0.09208837 0.02122263,
 0.02122263 0.05295369 };
NORES = 1;
AIC = LOG(DET(COVAR)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COVAR)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COVAR)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT , "WITH PASTMIN = " PSTMIN " "
 "NO OF RESTRICTION = " NORES,
 " AIC = " AIC,
 " BIC = " BIC,
 " HQC = " HQC;
/* WITH 2 RESTRICTIONS */
COVAR = { 0.09208837 0.02076825,
 0.02076825 0.05318102 };
NORES = 2;
AIC = LOG(DET(COVAR)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COVAR)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COVAR)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT , "WITH PASTMIN = " PSTMIN " "
 "NO OF RESTRICTION = " NORES,

```

| SAS            |           | 20:18 Wednesday, August 4, 1993 1 |   |
|----------------|-----------|-----------------------------------|---|
| WITH PASTMIN = | 3         | NO OF RESTRICTION =               | 0 |
| AIC =          | -5.038141 |                                   |   |
| BIC =          | -4.883706 |                                   |   |
| HQC =          | -4.979548 |                                   |   |
| WITH PASTMIN = | 3         | NO OF RESTRICTION =               | 1 |
| AIC =          | -5.078916 |                                   |   |
| BIC =          | -4.963091 |                                   |   |
| HQC =          | -5.034972 |                                   |   |
| WITH PASTMIN = | 2         | NO OF RESTRICTION =               | 0 |
| AIC =          | -5.134581 |                                   |   |
| BIC =          | -4.864321 |                                   |   |
| HQC =          | -5.032045 |                                   |   |
| WITH PASTMIN = | 2         | NO OF RESTRICTION =               | 1 |
| AIC =          | -5.175357 |                                   |   |
| BIC =          | -4.943705 |                                   |   |
| HQC =          | -5.087469 |                                   |   |
| WITH PASTMIN = | 2         | NO OF RESTRICTION =               | 2 |
| AIC =          | -5.207173 |                                   |   |
| BIC =          | -5.01413  |                                   |   |
| HQC =          | -5.133933 |                                   |   |
| WITH PASTMIN = | 2         | NO OF RESTRICTION =               | 3 |
| AIC =          | -5.227218 |                                   |   |
| BIC =          | -5.072784 |                                   |   |
| HQC =          | -5.168626 |                                   |   |
| Exiting IML.   |           |                                   |   |

그림 8.30 벌칙함수식별방법들에 의한 상태공간모형의 식별

```

" AIC = " AIC,
" BIC = " BIC,
" HQC = " HQC;
/ * WITH 3 RESTRICTIONS */
COVAR = { 0.09219234 0.02120271,
 0.02120271 0.05433557 };
NOES = 3;
AIC = LOG(DET(COVAR)) + (K-NOES)*2.0/ T;
BIC = LOG(DET(COVAR)) + (K-NOES)*LOG(T)/ T;
HQC = LOG(DET(COVAR)) + (K-NOES)*2*LOG(LOG(T))/ T;

```



```

PRINT, "WITH PASTMIN = " PSTMIN " "
 "NO OF RESTRICTION = " NORES,
 " AIC = " AIC,
 " BIC = " BIC,
 " HQC = " HQC, ;
QUIT:

```

이 SAS프로그램을 수행한 결과가 그림 8.30이다. 그림 8.30에서 AIC, BIC, 그리고 HQC 모두가 옵션 PASTMIN=에 2를 할당하고 RESTRICT스태이트먼트에 'F(1, 1)=0 F(3, 1)=0 G(3, 2)=0'을 할당한 상태공간모형을 최적의 상태공간모형으로 선택한다.

다변량시계열데이터 MINKMUSK를 상태공간모형화하고 이 상태공간모형을 이용해서 예측을 하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4 TSA4의 파일 EXAM8-31.PRG에 수록되어 있다).

```

/* */
/* */
/* */
/* */
/* */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=60 LINESIZE=72;
TITLE 'HUDSON' 'S BAY FUR TRAPPING RECORDS 1842-1890';
DATA EXAM831;
 SET TSA4DATA.MINKMUSK;
 LMINK = LOG(MINK);
 LMUSKRAT = LOG(MUSKRAT);
 T+1;
 RUN;
PROC STATESPACE LEAD=10 OUT=EXAM831B
 PASTMIN=2 NOCENTER;
 VAR LMINK(1) LMUSKRAT(1);
 RESTRICT F(1,1)=0 F(3,1)=0 G(3,2)=0;
 RUN;
DATA EXAM831C;
 SET EXAM831B;
 ILMINK = EXP(LMINK);
 ILMUSK = EXP(LMUSKRAT);
 FOREMINK = EXP(FOR1+STD1*STD1/2.0);
 FOREMUSK = EXP(FOR2+STD2*STD2/2.0);

```

```

RUN;
PROC PRINT DATA=EXAM831C(FIRSTOBS=40);
 VAR ILMINK FOREMINK ILMUSK FOREMUSK;
RUN;
PROC STATESPACE DATA=EXAM831 LEAD=10 NOCENTER PRINT;
 VAR LMINK(1) LMUSKRAT(1);
RUN;

```

이 SAS프로그램의 첫 번째 프로시저는 다음과 같다.

```

PROC STATESPACE LEAD=10 OUT=EXAM831B
 PASTMIN=2 NOCENTER;
 VAR LMINK(1) LMUSKRAT(1);
 RESTRICT F(1,1)=0 F(3,1)=0 G(3,2)=0;
RUN;

```

이 프로시저에서는 그림 8.30에서 식별한 최적의 상태공간모형을 추정하기 위해서 RE-STRIC트스테이트먼트에 'F(1, 1)=0 F(3, 1)=0 G(3, 2)=0'을 지정하였다. 또한 PROC STATESPACE스테이트먼트의 옵션 LEAD=에 10을 할당함으로써 10개의 선시차들에 대해서 예측을 하고, PROC STATESPACE스테이트먼트의 옵션 OUT=에 EXAM831B를 할당함으로써 예측의 결과를 SAS데이터세트 EXAM831B에 출력시킨다. 이 프로시저를 수행한 결과 중에서 최종추정단계의 결과가 그림 8.31의 부분 A에 수록되어 있다.

그림 8.31의 부분 A에서 최종적인 상태공간모형은 다음과 같다는 것을 알 수 있다.

$$\begin{aligned}
 y_{1,t} &= \text{LMINK}(t) - \text{LMINK}(t-1) \\
 y_{2,t} &= \text{LMUSKRAT}(t) - \text{LMUSKRAT}(t-1) \\
 z_{t+1} &= \hat{F}z_t + \hat{G}v_{t+1} \\
 y_{2,t} &= [I_2 \ 0]z_t
 \end{aligned}$$

단,  $z_t = (y_{1,t}, y_{2,t}, y_{2,t+1|t})'$

$$\hat{F} = \begin{bmatrix} 0 & 0.321 & 0 \\ 0 & 0 & 1 \\ 0 & -0.286 & 0.515 \end{bmatrix}, \quad \hat{G} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -0.601 & 0 \end{bmatrix}, \quad \hat{\Sigma} = \begin{bmatrix} 0.092 & 0.021 \\ 0.021 & 0.054 \end{bmatrix}$$

이 상태공간모형식을 전개하면 다음과 같다.

$$\begin{aligned}
 y_{1,t+1} &= 0.321y_{2,t} + v_{1,t+1} \\
 y_{2,t+1} &= y_{2,t+1|t} + v_{2,t+1} \\
 y_{2,t+2|t+1} &= -0.286y_{2,t} + 0.515y_{2,t+1|t} - 0.601v_{1,t+1}
 \end{aligned}$$

두 번째 식을 세 번째 식에 대입하면 다음 식을 얻는다.

$$y_{2,t+2} = 0.515y_{2,t+1} - 0.286y_{2,t} + v_{2,t+2} - 0.601v_{1,t+1} - 0.515v_{2,t+1}$$

이 식을 정리하면 다음과 같은 2변량VARMA(2, 1)모형을 얻는다.

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 0 & 0.321 \\ 0 & 0.515 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & -0.286 \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} + \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0.601 & 0.515 \end{bmatrix} \begin{bmatrix} v_{1,t-1} \\ v_{2,t-1} \end{bmatrix}$$

두 번째 DATA스텝과 두 번째 PROC스텝은 다음과 같다. 이 DATA스텝은 앞의 PROC스텝에서 생성한 예측의 결과를 역대수변환(inverse log transformation)하기 위한 것이다. 또한 이 PROC스텝은 역대수변환한 예측값들을 출력시키기 위한 것이다.

```
DATA EXAM827D;
 SET EXAM827C;
 ILMINK = EXP(LMINK);
 ILMUSK = EXP(LMUSKRAT);
 FOREMINK = EXP(FOR1+STD1*STD1/2.0);
 FOREMUSK = EXP(FOR2+STD2*STD2/2.0);
 RUN;
PROC PRINT DATA=EXAM827D(FIRSTOBS=40);
 VAR ILMINK FOREMINK ILMUSK FOREMUSK;
 RUN;
```

여기서 유의할 점은 변수들 LMINK와 LMUSKRAT가 대수변환된 변수들이라는 것이다. [SAS4TSA2] (pp. 609~610)에서 설명했듯이 대수변환된 변수의 최소평균제곱오차예측을 역대수변환한 것이 원래 변수의 최소평균제곱오차예측은 아니다. 시계열데이터  $\{y_t\}$ 를 대수변환한 시계열데이터  $\{z_t\}$ 를 생각해 보자. 만일  $z_1, z_2, \dots, z_T$ 가 주어졌을 때  $\tilde{z}_i(k)$ 를  $z_{t+k}$ 의 최소평균제곱오차예측이라 하고,  $\sigma_k^2$ 를 최소평균제곱오차예측의 분산이라고 하면 시점  $t$ 에서의  $y_{t+k}$ 의 최소평균제곱오차는 다음과 같다.

$$\exp \left\{ \tilde{z}_i(k) + \frac{1}{2} \sigma_k^2 \right\}$$

변수들 FOREMINK와 FOREMUSK는 변수들 MINK와 MUSKRAT의 최소평균제곱오차예측들이다. PROC PRINT는 이 예측들을 OUTPUT윈도우에 출력시키기 위한 것이다. 이 프로시저들을 수행한 결과가 그림 8.31의 부분 B에 수록되어 있다. 부분 B에서 변수들 ILMINK와 ILMUSK는 원래 변수들 MINK와 MUSKRAT의 관찰값들이다.

이 SAS프로그램의 세 번째 프로시저는 다음과 같다. 이 프로시저는 [BD86] (pp. 189~195)의 예제를 그대로 사용한 것이다.

```
PROC STATESPACE DATA=EXAM831 LEAD=10 NOCENTER PRINT;
 VAR LMINK(1) LMUSKRAT(1);
 RUN;
```

그림 8.30에서 알 수 있듯이 이 PROC스텝에서 구한 상태공간모형이 그림 8.31의 부분 A의 상태공간모형보다 더 좋은 것은 아니다. 그러나 SAS의 출판물인 [BD86] (pp. 191~

(A)

STATESPACE Procedure

Selected Statespace Form and Fitted Model

State Vector

LMINK(T;T)                      LMUSKRAT(T;T)                      LMUSKRAT(T+1;T)

Estimate of the Transition Matrix

|       |        |       |
|-------|--------|-------|
| 0.000 | 0.321  | 0     |
| 0     | 0      | 1     |
| 0.000 | -0.286 | 0.515 |

Input Matrix for the Innovation

|        |       |
|--------|-------|
| 1      | 0     |
| 0      | 1     |
| -0.601 | 0.000 |

Variance Matrix for the Innovation

|            |            |
|------------|------------|
| 0.09219234 | 0.02120271 |
| 0.02120271 | 0.05433557 |

Parameter Estimates

| Parameter | Estimate | Std. Err. | T value  |
|-----------|----------|-----------|----------|
| F(1,2)    | 0.321414 | 0.131895  | 2.436897 |
| F(3,2)    | -0.28612 | 0.131018  | -2.1838  |
| F(3,3)    | 0.51488  | 0.127043  | 4.052798 |
| G(3,1)    | -0.60144 | 0.104764  | -5.74085 |

(B)

| OBS | ILMINK | FOREMINK  | ILMUSK  | FOREMUSK   |
|-----|--------|-----------|---------|------------|
| 40  | 40834  | 47256.56  | 1028187 | 1011361.74 |
| 41  | 56882  | 45805.12  | 1081489 | 1071392.22 |
| 42  | 52258  | 60540.95  | 1082999 | 898971.71  |
| 43  | 110610 | 54747.82  | 817003  | 1045029.15 |
| 44  | 76393  | 105796.49 | 347050  | 517675.99  |
| 45  | 64215  | 60752.07  | 380022  | 356549.37  |
| 46  | 82941  | 69234.87  | 344818  | 469286.13  |
| 47  | 43641  | 84181.93  | 223615  | 331126.75  |
| 48  | 35400  | 39761.21  | 322360  | 329464.52  |
| 49  | 29363  | 41694.23  | 574742  | 470808.71  |
| 50  | .      | 37028.57  | .       | 765584.75  |
| 51  | .      | 42556.42  | .       | 765033.84  |
| 52  | .      | 43768.21  | .       | 726036.24  |
| 53  | .      | 43715.25  | .       | 724274.88  |
| 54  | .      | 44371.01  | .       | 750444.30  |
| 55  | .      | 45781.67  | .       | 781905.31  |
| 56  | .      | 47420.59  | .       | 808082.22  |
| 57  | .      | 48984.96  | .       | 830273.48  |
| 58  | .      | 50473.97  | .       | 852502.67  |
| 59  | .      | 51981.41  | .       | 876484.40  |

(C)

STATESPACE Procedure

Selected Statespace Form and Fitted Model

State Vector

LMINK(T;T)                      LMUSKRAT(T;T)

Estimate of the Transition Matrix

|        |       |
|--------|-------|
| 0.006  | 0.289 |
| -0.650 | 0.176 |

| Input Matrix for the Innovation    |            |            |          |
|------------------------------------|------------|------------|----------|
|                                    | 1          | 0          |          |
|                                    | 0          | 1          |          |
| Variance Matrix for the Innovation |            |            |          |
|                                    | 0.09208463 | 0.01995379 |          |
|                                    | 0.01995379 | 0.06414716 |          |
| Parameter Estimates                |            |            |          |
| Parameter                          | Estimate   | Std. Err.  | T value  |
| F(1,1)                             | 0.006214   | 0.14065    | 0.04418  |
| F(1,2)                             | 0.288741   | 0.137758   | 2.096006 |
| F(2,1)                             | -0.64978   | 0.117408   | -5.53441 |
| F(2,2)                             | 0.175794   | 0.114992   | 1.528745 |

그림 8.31 PROC STATESPACE에 의한 상태공간모형화와 예측

195)를 좇아 이 PROC스텝을 수행해 보자. 이 프로시저를 수행한 결과 중에서 최종추정단계가 그림 8.31의 부분 C에 수록되어 있다. 이 상태공간모형의 최종추정식은 다음과 같다.

$$y_{1,t} = \text{LMINK}(t) - \text{LMINK}(t-1)$$

$$y_{2,t} = \text{LMUSKRAT}(t) - \text{LMUSKRAT}(t-1)$$

$$z_{t+1} = \hat{F}z_t + \hat{G}v_{t+1}$$

$$y_t = z_t$$

$$\text{단, } z_t = (y_{1,t}, y_{2,t})'$$

$$\hat{F} = \begin{bmatrix} 0.006 & 0.289 \\ -0.650 & 0.176 \end{bmatrix}, \hat{G} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \hat{\Sigma} = \begin{bmatrix} 0.092 & 0.020 \\ 0.020 & 0.064 \end{bmatrix}$$

이 상태공간모형식을 전개하면 다음과 같은 2변량VAR(1)모형을 얻는다.

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 0.006 & 0.289 \\ -0.650 & 0.176 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix}$$

#### 8.6.4 다변량시계열데이터 HOUSECON

다변량시계열데이터 HOUSECON은 1983년 1월부터 1989년 10월까지의 월별 건설도금액(construction contracts in millions of dollars)인 변수 CONSTR와 착공한 주택의 수(private housing starts in thousands of units)인 변수 HSTARTS, 그리고 월평균 신주택자금이자율(average new home mortgage rate)인 변수 INTRATE를 기록한 것이다. 이 다변량시계열데이터의 관찰점들의 개수는 82이다. 다변량시계열데이터 HOUSECON은 제 2.4절에 수록되어 있으며, 또한 SAS데이터라이브러리 TSA4DATA의 파일 HOUSECON.SSD에 수록되어 있다. 변수 CONSTR, 변수 INTRATE, 그리고 변수 HSTARTS의 시계열산포도들이 각각 그림 2.4.1, 그림 2.4.2와 그림 2.4.3이다.

다변량시계열데이터 HOUSECON을 상태공간모형화하기 위해서는 이 다변량시계열데이터를 정상시계열로 바꾸어야 한다. 제 2.4절의 그림 2.4.1에서 변수 CONSTR의 분산이 시간이 흐름에 따라 증가함을 알 수 있다. 또한 그림 2.4.3에서 변수 HSTARTS의 분산도 시간이 흐름에 따라 증가함을 알 수 있다. 이 분산의 이질성문제를 해결하기 위해서 이 변수들을 대수변환하기로 하고 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM8-32.PRG에 수록되어 있다).

```

/ * */
/ * EXAM8-32.PRG */
/ * */
/ * LOG TRANSFORMATION OF */
/ * DATASET = HOUSECON */
LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS LINESIZE=72 PAGESIZE=60;
TITLE 'HOUSE CONSTRUCTION DATA';
DATA EXAM832;
 SET TSA4DATA.HOUSECON;
 LCONSTR = LOG(CONSTR);
 LHSTART = LOG(HSTARTS);
RUN;
SYMBOL1 C=BLACK V=NONE L=1 I=JOIN;
SYMBOL2 C=BLACK V=NONE L=4 I=JOIN;
AXIS1 ORDER=('01JAN83'D TO '01JAN90'D BY YEAR)
 LABEL=('DATE');
AXIS2 ORDER=(9.0 TO 11.5 BY 0.5)
 LABEL=('LOG(CONTR)');
AXIS3 ORDER=(3.7 TO 5.3 BY 0.2)
 LABEL=('LOG(HSTARTS)');
PROC Gplot DATA=EXAM832;
 PLOT LCONSTR*DATE
 / HAXIS=AXIS1 VAXIS=AXIS2;
 PLOT2 LHSTART*DATE
 / HAXIS=AXIS1 VAXIS=AXIS3;
RUN;

```

이 SAS프로그램을 수행한 결과가 그림 8.32이다. 그림 8.32에서 실선은 변수 CONSTR를 대수변환한 LCONSTR의 시계열산포도이고 점선은 변수 HSTARTS를 대수변환한 LHSTART의 시계열산포도이다. 이 시계열산포도들은 시간이 흘러도 분산들이 거의 변화가 없음을 알 수 있다. 반면에 대수변환된 변수들 LCONSTR와 LHSTART의 평균들은

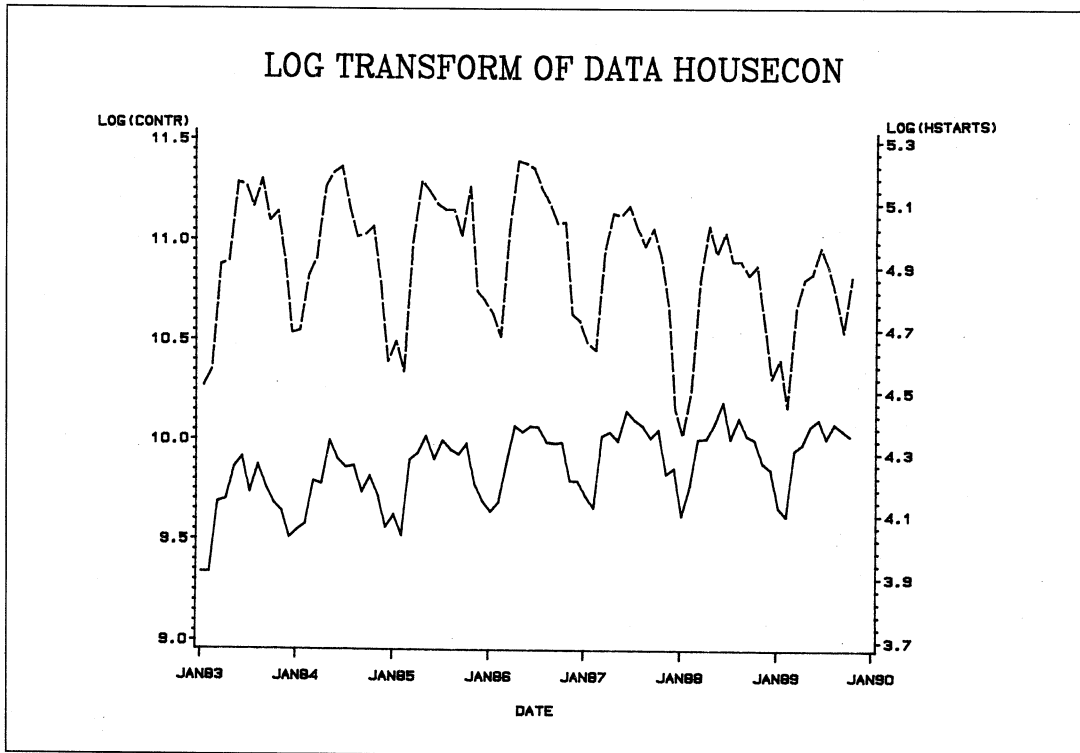


그림 8.32 대수변환된 변수들 LCONSTR와 LHSTART의 시계열산포도들

계절주기 12를 가지고 변동하고 있음을 알 수 있다.

이 대수변환된 시계열데이터를 상태공간모형화하기 위해서는 이 대수변환된 시계열데이터를 정상시계열데이터로 바꾸어야 한다. 그림 8.32에서 변수들 LCONSTR와 LHSTART가 계절주기 12를 가지고 변동함을 알 수 있다. 이 계절성의 문제를 해결하기 위해서 이 변수들을 계절차분하기로 한다. 이렇게 계절차분한 대수변환시계열데이터는 정상성을 보인다. 계절차분된 대수변환시계열데이터를 여러 상태공간모형들에 적합시키기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM 8-33.PRG에 수록되어 있다).

```

/* */
/* EXAM8-33.PRG */
/* */
/* APPLYING SEVERAL STATE SPACE MODELS */
/* DATASET = HOUSECON */
/* */

LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=60 LINESIZE=72;
TITLE 'STATESPACE MODELING';
DATA EXAM833;
 SET TSA4DATA.HOUSECON;

```

```

LCONSTR=LOG(CONSTR);
LHSTART=LOG(HSTARTS);
RUN;
PROC STATESPACE DATA=EXAM833;
VAR LHSTART(12) LCONSTR(12);
RUN;
PROC STATESPACE DATA=EXAM833;
VAR LHSTART(12) LCONSTR(12);
RESTRICT G(3,2)=0;
RUN;
PROC STATESPACE DATA=EXAM833;
VAR LHSTART(12) LCONSTR(12);
RESTRICT G(3,2)=0 F(1,2)=0;
RUN;
PROC STATESPACE DATA=EXAM833;
VAR LCONSTR(12) LHSTART(12);
RUN;
PROC STATESPACE DATA=EXAM833;
VAR LCONSTR(12) LHSTART(12);
RESTRICT F(2,1)=0;
RUN;
PROC STATESPACE DATA=EXAM833;
VAR LCONSTR(12) LHSTART(12);
RESTRICT F(2,1)=0 F(2,3)=0;
RUN;
PROC STATESPACE DATA=EXAM833;
VAR LCONSTR(12) LHSTART(12);
RESTRICT F(2,1)=0 F(2,3)=0 G(3,1)=0;
RUN;
PROC STATESPACE DATA=EXAM833;
VAR LCONSTR(12) LHSTART(12);
RESTRICT F(2,1)=0 F(2,3)=0 G(3,1)=0 F(2,2)=0;
RUN;
PROC STATESPACE DATA=EXAM833;
VAR LCONSTR(12) LHSTART(12);
RESTRICT F(2,1)=0 F(2,3)=0 G(3,1)=0
 F(2,2)=0 G(3,2)=0;
RUN;

```



이 SAS프로그램을 수행한 결과 중에서 각 PROC STATESPACE의 최종추정단계의 결과물들이 그림 8.33이다. 첫 번째 PROC스텝에서 세 번째 PROC스텝까지의 VAR스테이트먼트에는 변수 LHSTART(12)를 먼저 지정하고 변수 LCONSTR(12)를 다음에 지정하였다. 첫 번째 PROC스텝을 수행한 결과 중에서 최종추정단계의 결과가 그림 8.33의 부분 A와 부분 B이다. 부분 A에서 상태벡터가 LHSTART( $t:t$ ), LCONSTR( $t:t$ ), 그리고 LCONSTR( $t+1:t$ )로 구성되어 있음을 알 수 있다. 부분 B에서 모수  $G(3, 2)$ 의 추정값이 유의적으로 0과 다르지 않음을 알 수 있다. 따라서 두 번째 PROC STATESPACE의 RESTRICT스테이트먼트에 ' $G(3, 2)=0$ '을 지정하였다. 이 PROC스텝을 수행한 결과 중에서 최종추정단계의 결과물이 그림 8.33의 부분 C이다. 부분 C에서 모수  $F(1, 2)$ 의 추정값이 유의적으로 0과 다르지 않음을 알 수 있다. 따라서 세 번째 PROC STATESPACE의 RESTRICT스테이트먼트에 ' $G(3, 2)=0 F(1, 2)=0$ '을 지정하였다. 이 PROC스텝을 수행한 결과 중에서 최종추정단계의 결과물이 그림 8.33의 부분 D이다. 부분 D에서 모든 모수들의 추정값들이 유의적으로 0과 다르다. 따라서 여기서 후향제거방법을 멈추도록 한다.

상태벡터를 구성하는 원소들은 다변량시계열데이터를 구성하는 원소들의 순서에 의존한다. 즉, 다변량시계열데이터를 구성하는 단변량시계열데이터들의 순서에 따라 상태벡터를 구성하는 원소들이 달라질 수 있고, 비록 상태벡터를 구성하는 원소들이 같다 하더라도 순서가 다르면 전혀 다른 상태공간모형을 구성할 수도 있다. 이를 예증하기 위해서 네 번째 PROC스텝에서 아홉 번째 PROC스텝까지의 VAR스테이트먼트에는 변수 LCONSTR(12)를 먼저 지정하고 변수 LHSTART(12)를 다음에 지정하였다. 네 번째 PROC스텝을 수행한 결과 중에서 최종추정단계의 결과가 그림 8.33의 부분 E와 부분 F이다. 부분 E에서 상태벡터가 LCONSTR( $t:t$ ), LHSTART( $t:t$ ), 그리고 LCONSTR( $t+1:t$ )로 구성되어 있음을 알 수 있다. 부분 F에서 모수  $F(2, 1)$ 의 추정값이 유의적으로 0과 다르지 않음을 알 수 있다. 따라서 다섯 번째 PROC STATESPACE의 RESTRICT스테이트먼트에 ' $F(2, 1)=0$ '을 지정하였다. 이 PROC스텝을 수행한 결과 중에서 최종추정단계의 결과물이 그림 8.33의 부분 G이다. 부분 G에서 모수  $F(2, 3)$ 의 추정값이 유의적으로 0과 다르지 않음을 알 수 있다. 따라서 여섯 번째 PROC STATESPACE의 RESTRICT스테이트먼트에 ' $F(2, 1)=0 F(2, 3)=0$ '을 지정하였다. 이 PROC스텝을 수행한 결과 중에서 최종추정단계의 결과물이 그림 8.33의 부분 H이다. 부분 H에서 모수  $G(3, 1)$ 의 추정값이 유의적으로 0과 다르지 않음을 알 수 있다. 따라서 일곱 번째 PROC STATESPACE의 RESTRICT스테이트먼트에 ' $F(2, 1)=0 F(2, 3)=0 G(3, 1)=0$ '을 지정하였다. 이 PROC스텝을 수행한 결과 중에서 최종추정단계의 결과물이 그림 8.33의 부분 I이다. 부분 I에서 모수  $F(2, 2)$ 의 추정값이 유의적으로 0과 다르지 않음을 알 수 있다. 따라서 여덟 번째 PROC STATESPACE의 RESTRICT스테이트먼트에 ' $F(2, 1)=0 F(2, 3)=0 G(3, 1)=0 F(2, 2)=0$ '을 지정하였다. 이 PROC스텝을 수행한 결과 중에서 최종추정단계의 결과물이 그림 8.33의 부분 J이다. 부분 J에서 모수  $G(3, 2)$ 의 추정값이 유의적으로 0과 다르지 않음을 알 수 있다. 따라서 아홉 번째 PROC STATESPACE의 RESTRICT스테이트먼트에 ' $F(2, 1)=0 F(2, 3)=0 G(3, 1)=0 F(2, 2)=0 G(3, 2)=0$ '을 지정하였다. 이 PROC스텝을 수행한 결과 중에

|     |                                    |               |                   |
|-----|------------------------------------|---------------|-------------------|
| (A) | State Vector                       |               |                   |
|     | LHSTART(T;T)                       | LCONST(R(T;T) | LCONST(R(T+1;T)   |
| (B) | Variance Matrix for the Innovation |               |                   |
|     | 0.01121776                         | 0.00353991    |                   |
|     | 0.00353991                         | 0.00491669    |                   |
|     | Parameter Estimates                |               |                   |
|     | Parameter                          | Estimate      | Std. Err. T value |
|     | F(1,1)                             | 0.413336      | 0.12372 3.340897  |
|     | F(1,2)                             | 0.10313       | 0.186635 0.552578 |
|     | F(3,1)                             | -0.17888      | 0.058942 -3.03483 |
|     | F(3,2)                             | 0.352884      | 0.123348 2.860884 |
|     | F(3,3)                             | 0.486236      | 0.210952 2.304963 |
|     | G(3,1)                             | 0.194446      | 0.090218 2.155291 |
|     | G(3,2)                             | -0.06889      | 0.130951 -0.52609 |
| (C) | Variance Matrix for the Innovation |               |                   |
|     | 0.01122676                         | 0.00356911    |                   |
|     | 0.00356911                         | 0.00494631    |                   |
|     | Parameter Estimates                |               |                   |
|     | Parameter                          | Estimate      | Std. Err. T value |
|     | F(1,1)                             | 0.398448      | 0.120503 3.306528 |
|     | F(1,2)                             | 0.147366      | 0.166752 0.883744 |
|     | F(3,1)                             | -0.1786       | 0.059518 -3.01082 |
|     | F(3,2)                             | 0.321636      | 0.118917 2.704717 |
|     | F(3,3)                             | 0.524959      | 0.211249 2.485022 |
|     | G(3,1)                             | 0.17538       | 0.079417 2.208327 |
| (D) | Variance Matrix for the Innovation |               |                   |
|     | 0.01126666                         | 0.00354306    |                   |
|     | 0.00354306                         | 0.00495459    |                   |
|     | Parameter Estimates                |               |                   |
|     | Parameter                          | Estimate      | Std. Err. T value |
|     | F(1,1)                             | 0.448038      | 0.106855 4.192948 |
|     | F(3,1)                             | -0.16985      | 0.058644 -2.89621 |
|     | F(3,2)                             | 0.298559      | 0.117617 2.538404 |
|     | F(3,3)                             | 0.493158      | 0.221132 2.230157 |
|     | G(3,1)                             | 0.17882       | 0.07961 2.2462    |
| (E) | State Vector                       |               |                   |
|     | LCONST(R(T;T)                      | LHSTART(T;T)  | LCONST(R(T+1;T)   |
| (F) | Variance Matrix for the Innovation |               |                   |
|     | 0.00487052                         | 0.0034434     |                   |
|     | 0.0034434                          | 0.01114823    |                   |
|     | Parameter Estimates                |               |                   |
|     | Parameter                          | Estimate      | Std. Err. T value |
|     | F(2,1)                             | 0.097908      | 0.173221 0.565221 |
|     | F(2,2)                             | 0.370646      | 0.129537 2.861319 |
|     | F(2,3)                             | 0.271218      | 0.382195 0.709632 |
|     | F(3,1)                             | 0.386463      | 0.137489 2.810857 |
|     | F(3,2)                             | -0.18245      | 0.067837 -2.68961 |
|     | F(3,3)                             | 0.544735      | 0.209245 2.603334 |
|     | G(3,1)                             | -0.08101      | 0.131949 -0.61395 |
|     | G(3,2)                             | 0.181977      | 0.091154 1.996372 |

|     |                                    |            |            |          |
|-----|------------------------------------|------------|------------|----------|
| (G) | Variance Matrix for the Innovation |            |            |          |
|     |                                    | 0.0048728  | 0.00345405 |          |
|     |                                    | 0.00345405 | 0.01119791 |          |
|     | Parameter Estimates                |            |            |          |
|     | Parameter                          | Estimate   | Std. Err.  | T value  |
|     | F(2,2)                             | 0.403822   | 0.117333   | 3.441669 |
|     | F(2,3)                             | 0.269823   | 0.383328   | 0.703897 |
|     | F(3,1)                             | 0.380982   | 0.134229   | 2.838296 |
|     | F(3,2)                             | -0.1777    | 0.066568   | -2.66945 |
|     | F(3,3)                             | 0.524809   | 0.207228   | 2.532524 |
|     | G(3,1)                             | -0.10261   | 0.126297   | -0.81245 |
|     | G(3,2)                             | 0.189313   | 0.09025    | 2.097652 |
| (H) | Variance Matrix for the Innovation |            |            |          |
|     |                                    | 0.00492118 | 0.00355473 |          |
|     |                                    | 0.00355473 | 0.01126666 |          |
|     | Parameter Estimates                |            |            |          |
|     | Parameter                          | Estimate   | Std. Err.  | T value  |
|     | F(2,2)                             | 0.448041   | 0.106672   | 4.20017  |
|     | F(3,1)                             | 0.350868   | 0.11878    | 2.953926 |
|     | F(3,2)                             | -0.17317   | 0.057573   | -3.00779 |
|     | F(3,3)                             | 0.454187   | 0.201792   | 2.250763 |
|     | G(3,1)                             | -0.10154   | 0.116687   | -0.87018 |
|     | G(3,2)                             | 0.205318   | 0.087863   | 2.336791 |
| (I) | Variance Matrix for the Innovation |            |            |          |
|     |                                    | 0.00495515 | 0.00354395 |          |
|     |                                    | 0.00354395 | 0.01126666 |          |
|     | Parameter Estimates                |            |            |          |
|     | Parameter                          | Estimate   | Std. Err.  | T value  |
|     | F(2,2)                             | 0.000875   | 0.106855   | 0.008189 |
|     | F(3,1)                             | 0.298228   | 0.117582   | 2.536346 |
|     | F(3,2)                             | -0.16998   | 0.058615   | -2.89988 |
|     | F(3,3)                             | 0.493696   | 0.220945   | 2.234474 |
|     | G(3,2)                             | 0.178951   | 0.079609   | 2.247877 |
| (J) | Variance Matrix for the Innovation |            |            |          |
|     |                                    | 0.00521987 | 0.0043618  |          |
|     |                                    | 0.0043618  | 0.01409633 |          |
|     | Parameter Estimates                |            |            |          |
|     | Parameter                          | Estimate   | Std. Err.  | T value  |
|     | F(3,1)                             | 0.259334   | 0.11921    | 2.175445 |
|     | F(3,2)                             | -0.18082   | 0.078008   | -2.31796 |
|     | F(3,3)                             | 0.601467   | 0.223075   | 2.696257 |
|     | G(3,2)                             | 0.056494   | 0.072369   | 0.780642 |
| (K) | Variance Matrix for the Innovation |            |            |          |
|     |                                    | 0.00535783 | 0.00459937 |          |
|     |                                    | 0.00459937 | 0.01409633 |          |
|     | Parameter Estimates                |            |            |          |
|     | Parameter                          | Estimate   | Std. Err.  | T value  |
|     | F(3,1)                             | 0.319587   | 0.116246   | 2.749221 |
|     | F(3,2)                             | -0.16438   | 0.062385   | -2.63495 |
|     | F(3,3)                             | 0.475289   | 0.215366   | 2.206884 |

그림 8.33 옵션 RESTRICT=을 사용한 최종추정단계들

서 최종추정단계의 결과물이 그림 8.33의 부분 K이다. 부분 K의 모든 모수들의 추정값들이 유의적으로 0과 다르다. 따라서 여기서 후향제거방법을 멈추도록 한다.

좀더 객관적으로 상태공간모형을 식별하기 위해서 벌칙함수식별방법들을 적용하기로 하자. 이러한 목적으로 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4 TSA4의 파일 EXAM8-34.PRG에 수록되어 있다).

```

/* */
/* EXAM8-34.PRG */
/* */
/* CHOOSING RESTRICT CONDITIONS */
/* USING THE PENALTY FUNTION METHODS */
/* */
OPTIONS PAGESIZE=60 LINESIZE=72;
PROC IML;
RESET NOLOG NOCENTER NONAME;
T = 70;
K = 7;
/* WITH NO RESTRICTION */
COV0 = { 0.01121776 0.00353991,
 0.00353991 0.00491669 };
NORES = 0;
AIC = LOG(DET(COV0)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COV0)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COV0)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT , "NO OF RESTRICTION = " NORES,
 " AIC = " AIC,
 " BIC = " BIC,
 " HQC = " HQC;
/* WITH 1 RESTRICTION */
COV1 = { 0.01122676 0.00356911,
 0.00356911 0.00494631 };
NORES = 1;
AIC = LOG(DET(COV1)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COV1)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COV1)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT , "NO OF RESTRICTION = " NORES,
 " AIC = " AIC,
 " BIC = " BIC,
 " HQC = " HQC;
/* WITH 2 RESTRICTIONS */

```

```

COV2 = { 0.01126666 0.00354306,
 0.00354306 0.00495459 };

NORES = 2;
AIC = LOG(DET(COV2)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COV2)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COV2)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT , "NO OF RESTRICTION = " NORES,
 " AIC = " AIC,
 " BIC = " BIC,
 " HQC = " HQC;

/ * WITH NO RESTRICTION */
COV0 = { 0.00487052 0.0034434,
 0.0034434 0.01114823 };

NORES = 0;
AIC = LOG(DET(COV0)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COV0)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COV0)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT , "NO OF RESTRICTION = " NORES,
 " AIC = " AIC,
 " BIC = " BIC,
 " HQC = " HQC;

/ * WITH 1 RESTRICTION */
COV1 = { 0.0048728 0.00345405,
 0.00345405 0.01119791 };

NORES = 1;
AIC = LOG(DET(COV1)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COV1)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COV1)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT , "NO OF RESTRICTION = " NORES,
 " AIC = " AIC,
 " BIC = " BIC,
 " HQC = " HQC;

/ * WITH 2 RESTRICTIONS */
COV2 = { 0.00492118 0.00355473,
 0.00355473 0.01126666 };

NORES = 2;
AIC = LOG(DET(COV2)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COV2)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COV2)) + (K-NORES)*2*LOG(LOG(T))/ T;

```

```

PRINT , "NO OF RESTRICTION = " NORES,
 " AIC = " AIC,
 " BIC = " BIC,
 " HQC = " HQC;

/ * WITH 3 RESTRICTIONS */
COV3 = { 0.00495515 0.00354395,
 0.00354395 0.01126666 };
NORES = 3;
AIC = LOG(DET(COV3)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COV3)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COV3)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT , "NO OF RESTRICTION = " NORES,
 " AIC = " AIC,
 " BIC = " BIC,
 " HQC = " HQC;

/ * WITH 4 RESTRICTIONS */
COV4 = { 0.00521987 0.0043618,
 0.0043618 0.01409633 };
NORES = 4;
AIC = LOG(DET(COV4)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COV4)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COV4)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT , "NO OF RESTRICTION = " NORES,
 " AIC = " AIC,
 " BIC = " BIC,
 " HQC = " HQC;

/ * WITH 5 RESTRICTIONS */
COV5 = { 0.00535783 0.00459937,
 0.00459937 0.01409633 };
NORES = 5;
AIC = LOG(DET(COV5)) + (K-NORES)*2.0/ T;
BIC = LOG(DET(COV5)) + (K-NORES)*LOG(T)/ T;
HQC = LOG(DET(COV5)) + (K-NORES)*2*LOG(LOG(T))/ T;
PRINT , "NO OF RESTRICTION = " NORES,
 " AIC = " AIC,
 " BIC = " BIC,
 " HQC = " HQC, ;

QUIT;

```

| SAS                 | 16:17 Thursday, August 5, 1993 1 |
|---------------------|----------------------------------|
| NO OF RESTRICTION = | 0                                |
| AIC =               | -9.86311                         |
| BIC =               | -9.63826                         |
| HQC =               | -9.773797                        |
| NO OF RESTRICTION = | 1                                |
| AIC =               | -9.887719                        |
| BIC =               | -9.694991                        |
| HQC =               | -9.811165                        |
| NO OF RESTRICTION = | 2                                |
| AIC =               | -9.905231                        |
| BIC =               | -9.744624                        |
| HQC =               | -9.841436                        |
| NO OF RESTRICTION = | 0                                |
| AIC =               | -9.867403                        |
| BIC =               | -9.642554                        |
| HQC =               | -9.77809                         |
| NO OF RESTRICTION = | 1                                |
| AIC =               | -9.891413                        |
| BIC =               | -9.698685                        |
| HQC =               | -9.814859                        |
| NO OF RESTRICTION = | 2                                |
| AIC =               | -9.915901                        |
| BIC =               | -9.755295                        |
| HQC =               | -9.852106                        |
| NO OF RESTRICTION = | 3                                |
| AIC =               | -9.933802                        |
| BIC =               | -9.805317                        |
| HQC =               | -9.882766                        |
| NO OF RESTRICTION = | 4                                |
| AIC =               | -9.730574                        |
| BIC =               | -9.63421                         |
| HQC =               | -9.692297                        |
| NO OF RESTRICTION = | 5                                |
| AIC =               | -9.762527                        |
| BIC =               | -9.698284                        |
| HQC =               | -9.737009                        |
| Exiting IML.        |                                  |

그림 8.34 벌칙함수식별방법들에 의한 상태공간모형의 식별

이 SAS프로그램을 수행한 결과가 그림 8.34이다. 그림 8.34에서 VAR스테이트먼트에 변수 LHSTART(12)를 먼저 지정하고 다음에 변수 LCONSTR(12)를 지정한 경우에는 AIC, BIC, 그리고 HQC 모두가 RESTRICT스테이트먼트에 'G(3, 2)=0 F(1, 2)=0'을 할당한 상태공간모형을 선택하는 것을 알 수 있다. 또한 VAR스테이트먼트에 변수 LCONSTR(12)를 먼저 지정하고 다음에 변수 LHSTART(12)를 지정한 경우에는 AIC, BIC, 그리고 HQC 모두가 RESTRICT스테이트먼트에 'F(2, 1)=0 F(2, 3)=0 G(3, 1)=0'을 할당한 상태공간모형을 선택하는 것을 알 수 있다.

다변량시계열데이터 HOUSECON을 상태공간모형화하기 위해서 다음 SAS프로그램을 수행해 보자(이 SAS프로그램은 디스켓 SAS4TSA4의 파일 EXAM8-35.PRG에 수록되어 있다).

```

/* */
/* EXAM8-35.PRG */
/* */
/* STATE SPACE MODELING */
/* DATASET = HOUSECON */
/* */

LIBNAME TSA4DATA 'A:\TSA4DATA';
OPTIONS PAGESIZE=60 LINESIZE=72;
TITLE 'STATESPACE MODELING';
DATA EXAM835;
 SET TSA4DATA.HOUSECON;
 LCONSTR=LOG(CONSTR);
 LHSTART=LOG(HSTARTS);
RUN;

PROC STATESPACE DATA=EXAM835 CANCORR;
 VAR LHSTART(12) LCONSTR(12);
 RESTRICT G(3,2)=0 F(1,2)=0;
RUN;

PROC STATESPACE DATA=EXAM835 CANCORR;
 VAR LCONSTR(12) LHSTART(12);
 RESTRICT F(2,1)=0 F(2,3)=0 G(3,1)=0;
RUN;

```

이 SAS프로그램의 첫번째 프로시저는 다음과 같다.

```

PROC STATESPACE DATA=EXAM835 CANCORR;
 VAR LHSTART(12) LCONSTR(12);
 RESTRICT G(3,2)=0 F(1,2)=0;
RUN;

```



STATESPACE MODELING

1

20:44 Thursday, March 10, 1994

(A)

STATESPACE Procedure

Nobs = 70

|                               |          |          |
|-------------------------------|----------|----------|
| Variable                      | Mean     | Std      |
| LHSTART                       | -0.03004 | 0.118728 |
| LHSTART has been differenced. |          |          |
| The periods are 12.           |          |          |
| LCONSTR                       | 0.04734  | 0.078701 |
| LCONSTR has been differenced. |          |          |
| The periods are 12.           |          |          |

Information Criterion for Autoregressive Models

|             |             |             |             |             |
|-------------|-------------|-------------|-------------|-------------|
| Lag=0       | Lag=1       | Lag=2       | Lag=3       | Lag=4       |
| -675.086487 | -683.136911 | -690.722772 | -688.547156 | -684.576302 |
| Lag=5       | Lag=6       | Lag=7       | Lag=8       | Lag=9       |
| -681.919695 | -687.297687 | -680.469688 | -685.572262 | -685.197890 |
| Lag=10      |             |             |             |             |
| -684.236878 |             |             |             |             |

Schematic Representation of Correlations

|          |    |    |    |    |    |    |    |    |    |    |    |
|----------|----|----|----|----|----|----|----|----|----|----|----|
| Name/Lag | 0  | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 |
| LHSTART  | ++ | ++ | ++ | .+ | .. | .. | .. | .. | .. | .. | .. |
| LCONSTR  | ++ | +  | .+ | .+ | .. | .. | ++ | .. | .. | .+ | .. |

+ is > 2\*std error, - is < -2\*std error, . is between

Schematic Representation of Partial Autocorrelations

|          |    |    |    |    |    |    |    |    |    |    |
|----------|----|----|----|----|----|----|----|----|----|----|
| Name/Lag | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 |
| LHSTART  | +  | .. | .. | .. | .. | .. | .. | -+ | .. | -. |
| LCONSTR  | .. | .+ | .+ | .. | .. | .. | .. | .. | +  | .. |

+ is > 2\*std error, - is < -2\*std error, . is between

Yule-Walker Estimates for the Min AIC

|         |         |         |          |         |
|---------|---------|---------|----------|---------|
|         | Lag=1   |         | Lag=2    |         |
| LHSTART | 0.50747 | LCONSTR | 0.09747  | LHSTART |
| LCONSTR | 0.17599 | 0.02785 | -0.19606 | LCONSTR |
|         |         |         |          | 0.21750 |
|         |         |         |          | 0.35989 |

(B)

Canonical Correlations Analysis

|                           |              |        |        |          |          |   |
|---------------------------|--------------|--------|--------|----------|----------|---|
| State vector              | Correlations | Infor. | Chisq  | D.F.     |          |   |
| LHSTART(T;T),LCONSTR(T;T) | 1.0000       | 1.0000 | 0.2743 | -2.52532 | 5.318261 | 4 |
| LHSTART(T+1;T)            |              |        |        |          |          |   |
| LHSTART(T;T),LCONSTR(T;T) | 1.0000       | 1.0000 | 0.4276 | 6.133989 | 13.73016 | 4 |
| LCONSTR(T+1;T)            |              |        |        |          |          |   |
| LHSTART(T;T),LCONSTR(T;T) | 1.0000       | 1.0000 | 0.4805 | -1.91074 | 4.00163  | 3 |
| LCONSTR(T+1;T),           | 0.2382       |        |        |          |          |   |
| LCONSTR(T+2;T)            |              |        |        |          |          |   |

(C)

Selected Statespace Form and Preliminary Estimates

State Vector

|              |              |                |
|--------------|--------------|----------------|
| LHSTART(T;T) | LCONSTR(T;T) | LCONSTR(T+1;T) |
|--------------|--------------|----------------|

Estimate of the Transition Matrix

|        |       |       |
|--------|-------|-------|
| 0.413  | 0.000 | 0     |
| 0      | 0     | 1     |
| -0.235 | 0.364 | 0.736 |

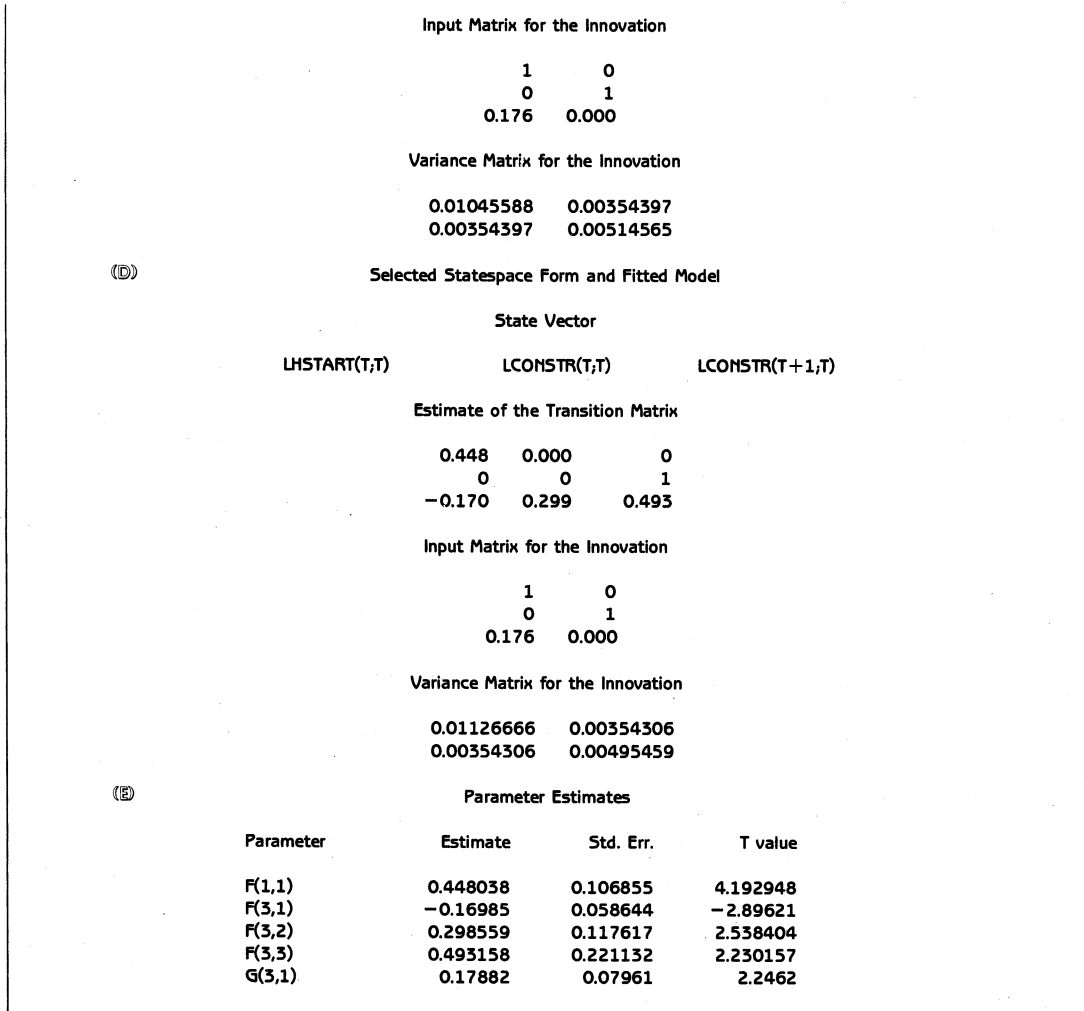


그림 8.35.1 다변량시계열데이터 HOUSECON의 상태공간모형화 1

그림 8.34에서 VAR스테이트먼트에 변수 LHSTART(12)를 먼저 지정하고 변수 LCONSTR(12)를 다음에 지정한 경우에는 AIC, BIC, 그리고 HQC 모두가 RESTRICT스테이트먼트에 'G(3, 2)=0 F(1, 2)=0'을 할당한 상태공간모형을 선택하였다. 이 PROC시스템에서는 이 상태공간모형을 추정한다. 이 PROC시스템을 수행한 결과가 그림 8.35.1이다. 그림 8.35.1의 부분 A에는 VAR모형단계를 수행한 결과가 수록되어 있다. 부분 B에는 PROC STATESPACE스테이트먼트의 옵션 CANCORR를 수행한 결과가 수록되어 있고, 부분 C에는 예비추정단계를 수행한 결과가 수록되어 있다. 부분들 D와 E에는 최종추정단계를 수행한 결과가 수록되어 있다. 이 상태공간모형의 최종추정식은 다음과 같다.

$$\begin{aligned}
 y_{1,t} &= LHSTART(t) - LHSTART(t-12) + 0.030 \\
 y_{2,t} &= LCONSTR(t) - LCONSTR(t-12) - 0.047 \\
 z_{t+1} &= \hat{F}z_t + \hat{G}v_{t+1} \\
 y_t &= [I_2 \ 0] z_t
 \end{aligned}$$

$$\text{단, } z_t = (y_{1,t}, y_{2,t}, y_{2,t+1|t})'$$

$$\hat{F} = \begin{bmatrix} 0.448 & 0.000 & 0 \\ 0 & 0 & 1 \\ -0.170 & 0.299 & 0.493 \end{bmatrix}, \hat{G} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0.176 & 0.000 \end{bmatrix}, \hat{\Sigma} = \begin{bmatrix} 0.011 & 0.003 \\ 0.003 & 0.005 \end{bmatrix}$$

이 상태공간모형식을 전개하면 다음과 같다.

$$y_{1,t+1} = 0.448y_{1,t} + v_{1,t+1}$$

$$y_{2,t+1} = y_{2,t+1|t} + v_{2,t+1}$$

$$y_{2,t+2|t+1} = -0.170y_{1,t} + 0.299y_{2,t} + 0.493y_{2,t+1|t} + 0.176v_{1,t+1}$$

두 번째 식을 세 번째 식에 대입하면 다음 식을 얻는다.

$$y_{2,t+2} = 0.493y_{2,t+1} - 0.170y_{1,t} + 0.299y_{2,t} + v_{2,t+2} - 0.493v_{2,t+1} + 0.176v_{1,t+1}$$

이 식을 정리하면 다음과 같은 2변량VARMA(2, 1)모형을 얻는다.

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 0.448 & 0 \\ 0 & 0.493 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ -0.170 & 0.299 \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} \\ + \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ -0.176 & 0.493 \end{bmatrix} \begin{bmatrix} v_{1,t-1} \\ v_{2,t-1} \end{bmatrix}$$

이 SAS프로그램의 두 번째 PROC스텝은 다음과 같다.

```
PROC STATESPACE DATA=EXAM835 CANCORR;
 VAR LCONSTR(12) LHSTART(12);
 RESTRICT F(2,1)=0 F(2,3)=0 G(3,1)=0;
 RUN;
```

그림 8.34에서 VAR스테이트먼트에 변수 LCONSTR(12)를 먼저 지정하고 다음에 변수 LHSTART(12)를 지정한 경우에는 AIC, BIC, 그리고 HQC 모두가 RESTRICT스테이트먼트에 'F(2, 1)=0 F(2, 3)=0 G(3, 1)=0'을 할당한 상태공간모형을 선택하였다. 이 PROC스텝에서는 이러한 제약조건들을 갖는 상태공간모형을 추정한다. 이 PROC스텝을 수행한 결과 중에서 최종추정단계가 그림 8.35.2이다. 이 상태공간모형의 최종추정식은 다음과 같다.

$$y_{1,t} = \text{LCONSTR}(t) - \text{LCONSTR}(t-12) - 0.047$$

$$y_{2,t} = \text{LHSTART}(t) - \text{LHSTART}(t-12) + 0.030$$

$$z_{t+1} = \hat{F}z_t + \hat{G}v_{t+1}$$

$$y_t = [I_2 \ 0] z_t$$

$$\text{단, } z_t = (y_{1,t}, y_{2,t}, y_{1,t+1|t})'$$

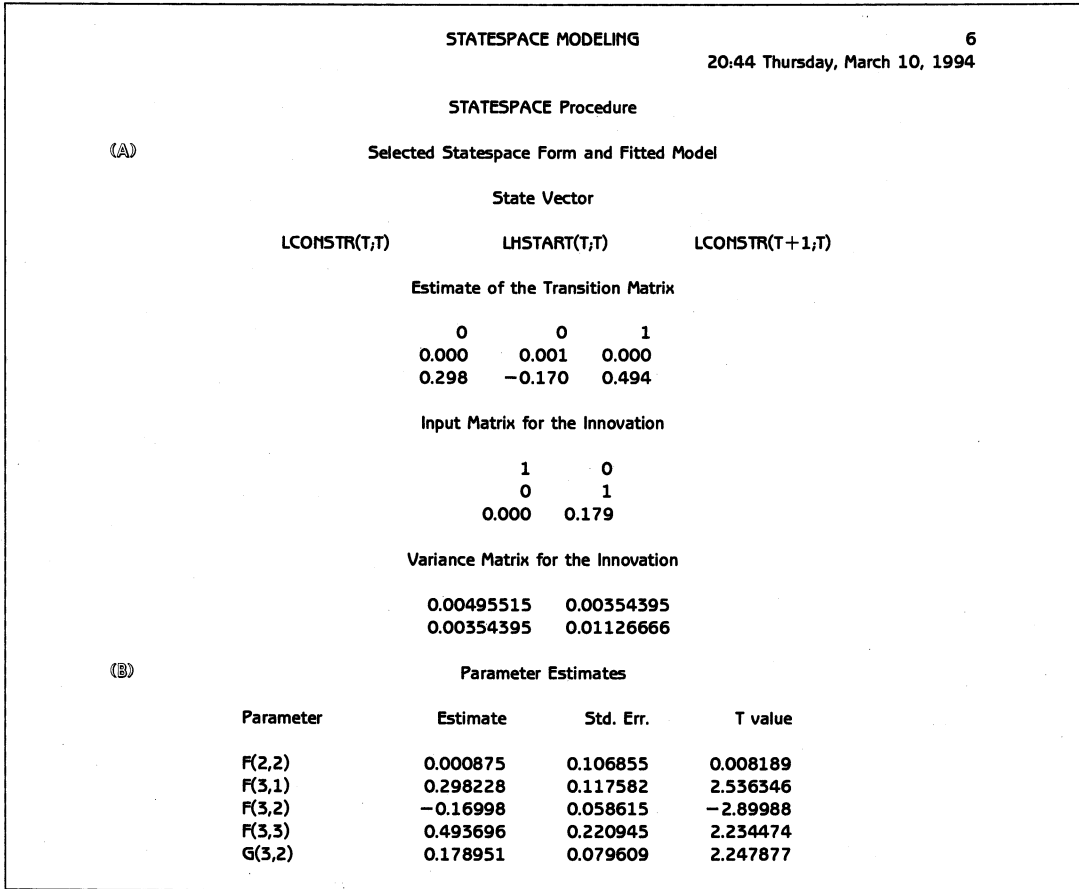


그림 8.35.2 다변량시계열데이터 HOUSECON의 상태공간모형화 2

$$\hat{F} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.001 & 0 \\ 0.298 & -0.170 & 0.494 \end{bmatrix}, \hat{G} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0.179 \end{bmatrix}, \hat{\Sigma} = \begin{bmatrix} 0.0050 & 0.0035 \\ 0.0035 & 0.0113 \end{bmatrix}$$

이 상태공간모형식을 전개하면 다음과 같다.

$$\begin{aligned} y_{1,t+1} &= y_{1,t+1|t} + v_{1,t+1} \\ y_{2,t+1} &= 0.001y_{2,t} + v_{2,t+1} \\ y_{1,t+2|t} &= 0.298y_{1,t} - 0.170y_{2,t} + 0.494y_{1,t+1|t} + 0.179v_{2,t+1} \end{aligned}$$

첫 번째 식을 두 번째 식에 대입하면 다음 식을 얻는다.

$$y_{1,t+2} = 0.494y_{1,t+1} + 0.298y_{1,t} - 0.170y_{2,t} + v_{1,t+2} - 0.494v_{1,t+1} + 0.179v_{2,t+1}$$

이 식을 정리하면 다음과 같은 2변량VARMA(2, 1)모형을 얻는다.

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 0.494 & 0 \\ 0 & 0.001 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} 0.298 & -0.170 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix}$$

$$+ \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} - \begin{bmatrix} 0.494 & -0.179 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v_{1,t-1} \\ v_{2,t-1} \end{bmatrix}$$

그림 8.35.1과 그림 8.35.2에서 상태벡터를 구성하는 원소들은 다변량시계열데이터를 구성하는 원소들의 순서에 의존하고, 또한 비록 상태벡터를 구성하는 원소들이 같다 하더라도 순서가 다르면 전혀 다른 상태공간모형을 이룰 수도 있다는 것을 보았다. 그림 8.34에서 벌칙함수식별방법들을 적용하면 그림 8.35.1의 상태공간모형보다는 그림 8.35.2의 상태공간모형이 더 타당한 것을 알 수 있다. 그러나 벌칙함수식별기준들이 아닌 다른 기준으로 비교하면 이와 다른 결론을 얻을 수도 있음에 유의하라.

### ▣ 참고문헌 ▣

- 박준경 · 이호창(1985), 「경제변수의 시계열분석」, 정책보고, vol. 85-02, 한국개발연구원, 서울.
- 신호중(1986), 「경제시계열 변수의 상태공간모형 설정」, 산경연구, 제2집, 목원대학교, 대전, pp. 89~104.
- 양현조(1992), 「전이함수모형과 상태공간모형의 비교연구」, 석사논문, 연세대학교 대학원, 서울.
- 정은이(1993), 「상태공간모형화에서 VAR모형의 선택이 미치는 영향」, 석사논문, 연세대학교 대학원, 서울.
- Akaike, H. (1973), Maximum likelihood identification of Gaussian autoregressive moving average models, *Biometrika*, vol. 60, pp. 255~265.
- Akaike, H. (1974a), Markovian representation of stochastic process and its application to the analysis of autoregressive moving average process, *Ann. Inst. Statist. Math.*, vol. 26, pp. 363~387.
- Akaike, H. (1974b), Stochastic theory of minimal realization, *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 667~674.
- Akaike, H. (1975), Markovian representation of stochastic process by canonical variables, *SIAM J. Contr.*, vol. 9, pp. 162~173.
- Akaike, H. (1976), Canonical correlation analysis of time series and the use of an information criterion, in *System Identification: Advances and Case Studies*, R. K. Mehra & D. G. Lainiotis, Eds., Academic Press, New York, pp. 27~96.
- Akaike, H. (1978), On the identification of state space models and their use in con-

- trol, in *Directions in Time Series*, D. R. Brillinger & G. C. Tiao, Eds., Institute of Mathematical Statistics, Hayward, pp. 175~187.
- Anderson, T. W. (1984), *Introduction to Multivariate Statistical Analysis*, (Second edition), Wiley, New York.
- Bartlett, M. S. (1939), A note on tests of significance in multivariate analysis, *Proc. of Cambridge Philosophical Soc.*, vol. 35, pp. 180~185.
- Kalman, R. E. (1960), A new approach to linear filtering and prediction problems, *Trans. ASME J. Basic Engineering*, Ser. D, vol. 82, pp. 35~45.
- Kalman, R. E. & R. S. Buch (1961), New results in linear filtering and prediction problems, *Trans. ASME J. Basic Engineering*, Ser. D, vol. 83, pp. 95~108.
- Kennedy, W. J. & J. E. Gentle (1980), *Statistical Computing*, Marcel Dekker, New York.
- Robinson, P. M. (1973), Generalized canonical analysis for time series, *J. Multivariate Anal.*, vol. 3, pp. 141~160.

# 부 록

---

---

- A. 점화식 연구
- ⌘ SAS / ETS® Software Reference Card
- ⌘ SAS / IML™ Software Reference Card





## 부록 A 점화식 연구

시계열분석을 공부하기 위해서는 우선 차분방정식(difference equation)에 대해서 잘 알고 있어야 한다. 차분방정식을 점화식이라고도 부르는데 이는 일본식 용어이다. 이 부록 A에는 본 저자가 고등학교 시절 점화식에 대해서 나름대로 생각해보았던 것들을 정리하여 1973년에 문지사에서 출판하였던 「점화식 연구」를 재수록하고자 한다.

### ● 머리글 ●

고등학교 2학년 재학시, 나는 건강상의 이유로 잠시 학교를 쉬고, 대전으로 내려간 적이 있었다. 그 당시, 나는 점화식에 흥미를 가지고 있었으며, 무료한 시간을 보내기 위해서 이 부분에 관해 공부해 보기로 했다.

그러나 점화식에 관한 서적은 구할 수 없었으며, 대부분의 책은 수열의 일부분으로 약간을 기술했을 뿐이었다. 또한 풀이과정에 있어서도 문제를 만드는 과정을 역으로 하는, 즉 수학적이기보다는 경험적인 방법을 사용했기 때문에, 호기심을 만족시키지 못했고 일종의 불신감까지 주어 새로운 방법을 생각하기에 이르렀다. 이 때 점화식을 체계적으로 연구하기 위해서 이 책의 골격이라 할 수 있는 ‘수열은 정의역이 자연수인 함수다’라는 정의를 내리고 이러한 사실을 바탕으로 점화식을 체계적으로 연구하였다.

다시 학교로 돌아와서, 이러한 방법을 친구들에게 가르쳐 주고 좀더 연구를 계속하던 중, 친구들이 책을 쓰기를 권했고 나 자신도 이러한 방법의 보급을 위해 책을 쓸 것을 생각했었다.

막상 대학에 들어와서 두 차례에 걸쳐 원고를 완성하였으나, 아직 사회인이 아닌 나로서는 출판이라는 것이 그리 쉬운 일은 아니었다. 여러 가지 난관에 부딪쳐서 포기하려고도 했으나, 끊임없는 친구들의 독려와 열성어린 도움이 나로 하여금 이 책을 완성시키게 했다. 사실상 이 책은 주위의 희생적인 정신과 시간적, 물질적인 도움이 없었다면 빛을

보지 못했을 것이다.

나 자신은 이 책을 참고서라기보다는 하나의 연구 보고서로 생각하고 싶다. 이 책에 사용되는 많은 문제는 창작했으며 본 저자의 편협적인 사고방식이 기술되지 않도록 하기 위해서 연습문제는 기존의 문제를 사용했다. 그러나 풀이에 있어서는 좀더 새롭고 체계적인 방법으로 풀었으며 적어도 몇몇 문제는 아직까지 일반적 해법을 구하지 못했던 것을 나 자신이 매혹될 만큼이나 멋있게 풀었음을 자부한다. 또한 책의 정도가 높은 이유로 한정된 부수를 한정된 독자층을 위해 발행했음을 말해 둔다.

마지막으로 이 책을 내는데 많은 도움을 주신 고태흠 선생님과 여러 은사님들, 편집과 교정에 수고해 준 친구들, 출판비를 대준 친구들, 그리고 출판에 여러 가지 편의를 보아 주신 경기서림 이승학 씨께 감사드린다.

1973. 11. 5

최 병 선

▶ 편집과 교정에 수고해 주신 분들 ◀

|              |                        |
|--------------|------------------------|
| 제A.2절, 제A.3절 | 박 준(법학과)               |
| 제A.4절        | 강영석(물리학과)              |
| 제A.3절, 제A.4절 | 최문기(의예과)               |
| 제A.5절, 제A.6절 | 정윤희(물리학과)              |
| 제A.7절        | 왕규창(의예과), 박찬용(약학과)     |
|              | 조경옥(미생물학과), 유영숙(수학교육과) |

## A.1 점화식이란 무엇인가?

### 1. 점화식이란

수열(또는 적분)에서 인접(隣接)한 항 사이의 관계를 규정하는 식이다.

(**보기**) :  $a_{n+1}=2a_n+3$ )

### 2. 점화식의 분류

#### 1) 항수로 본 분류

① 인접 2항간 :  $a_{n+1}=f(a_n)$  (**보기**) :  $a_{n+1}=3a_n+2$ )

② 인접 3항간 :  $a_{n+2}=g(a_{n+1}, a_n)$  (**보기**) :  $a_{n+2}=2a_{n+1}+a_n$ )

#### 2) 함수의 형태로 본 분류

1)의 ①, ②에서 각각의 점화식에 대응하는 함수  $y=f(x)$ ,  $z=g(x, y)$ 를 생각할 때  $f, g$ 의 꼴에 따라 다음과 같이 나눌 수 있다.

① 정 식 : (**보기**)  $a_{n+1}=3a_n+2$  :  $y=3x+2$   $f$ : 1차식  
 $a_{n+2}=2a_{n+1}+a_n$  :  $z=2x+y$   $g$ : 다변수 1차식

② 분 수 식 : (**보기**)  $a_{n+1}=\frac{2a_n-3}{a_n+1}$  :  $y=\frac{2x-3}{x+1}$   $f$ : 분수식

③ 무 리 식 : (**보기**)  $a_{n+1}=\sqrt{a_n+1}$  :  $y=\sqrt{x+1}$   $f$ : 무리식

④ 삼각함수 : (**보기**)  $a_{n+1}=\sin(a_n)$  :  $y=\sin x$   $f$ : 삼각함수식

⑤ 대수함수 : (**보기**)  $a_{n+1}=\log a_n$  :  $y=\log x$   $f$ : 대수함수식

그 외에 지수함수등 많은 형태가 있다.

#### 3) 초기조건(初期條件)

점화식에서 일반항을 구하려면 어떠한 조건이 필요하다. 예를 들어  $a_{n+1}=3a_n$ 은 정의에 의해 공비 3인 등비수열이므로  $a_n=a_1 \times 3^{n-1}$ (단,  $a_1$ 은 초항)이다. 이때  $a_1=1$ 이면  $a_n=3^{n-1}$ ,  $a_1=3$ 이면  $a_n=3^n$ 이 되듯이  $a_1$ 을 정함에 따라 하나의 수열이 결정된다. 따라서 수열이 확정되려면 우선  $a_1$ 이 확정되어야 한다. 이러한  $a_1$ 을 초기조건이라 한다. 제1항에서 ①의 경우는 1개, ②의 경우는 2개의 초기조건이 필요하다.

**보기** 1.  $a_{n+1}=2a_n+4$ ,  $a_1=1$ (초기조건)인 수열의 제4항을 구하여라.

**풀이**  $a_1=1, a_2=2 \times a_1+4=2 \times 1+4=6$   
 $a_3=2 \times a_2+4=2 \times 6+4=16$   
 $a_4=2 \times a_3+4=2 \times 16+4=36$

$\therefore a_4=36$

**보기 2.**  $a_{n+2}=a_{n+1}+a_n, a_1=1, a_2=1$ (초기조건)인 수열을 제10항까지 구하여라.

**풀이**

|       |   |   |   |   |   |   |    |    |    |    |
|-------|---|---|---|---|---|---|----|----|----|----|
| $n$   | 1 | 2 | 3 | 4 | 5 | 6 | 7  | 8  | 9  | 10 |
| $a_n$ | 1 | 1 | 2 | 3 | 5 | 8 | 13 | 21 | 34 | 55 |

**3. 점화식의 필요성과 수열에의 응용**

1) 나열된 수열  $a_1, a_2, a_3, \dots, a_n$ 에서 일반항  $a_n$ 을 보통은 직관적인 방법을 사용하여 구한다. 그러나 이러한 직관적인 방법이 어려울 때 점화식을 사용하여 간단히 얻을 수도 있다.

**보기 1.** 다음 수열의 일반항을 구하여라.

$$a_1=2, a_2=\frac{3}{2+2}, a_3=\frac{3}{2+\frac{3}{2+2}}, a_4=\frac{3}{2+\frac{3}{2+\frac{3}{2+2}}}, \dots$$

**풀이** 이 수열에서  $a_n$ 을 직관적으로 추정한다는 것은 매우 어렵고 번잡한 일이다. 이 때 각 항을 관찰하면 다음과 같은 규칙성을 얻을 수 있다.

$$a_1=2, a_2=\frac{3}{2+a_1}, a_3=\frac{3}{2+a_2}, a_4=\frac{3}{2+a_3}, \dots$$

결국, 이 수열은  $a_1=2$ (초기조건),  $a_{n+1}=\frac{3}{2+a_n}$ 으로 나타낼 수 있으며 이 식을 사용하여 일반항을 구할 수 있다(제A.7절을 참조하라).

**보기 2.** 다음 수열의 극한을 조사하여라.

$$a_1=\sqrt{3}, a_2=\sqrt{3+\sqrt{3}}, a_3=\sqrt{3+\sqrt{3+\sqrt{3}}},$$

$$a_4=\sqrt{3+\sqrt{3+\sqrt{3+\sqrt{3}}}}, \dots$$

**풀이** 이 수열의 극한을 이러한 상태에서 조사한다는 것은 거의 불가능한 일이다. 각 항을 관찰하여 다음과 같은 규칙을 발견할 수 있다.

$$a_1=\sqrt{3}, a_2=\sqrt{3+a_1}, a_3=\sqrt{3+a_2}, a_4=\sqrt{3+a_3}, \dots$$

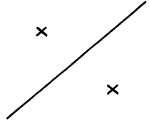
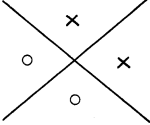
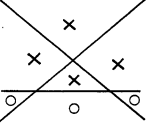
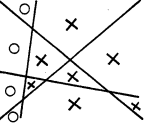
따라서 이 수열은  $a_1=\sqrt{3}$ (초기조건),  $a_{n+1}=\sqrt{3+a_n}$ 이 되므로 이 식을 이용하여 간단히 극한을 조사할 수 있다(제A.5절과 제A.6절을 참조하라).

2) 수열의 일반적인 나열이 힘든 대신에 인접한 항 사이의 관계를 쉽게 알 수 있는

경우 점화식을 사용한다.

**보기** 1. 평면 위에  $n$ 개의 직선이 있다(단, 어느 세 직선도 한 점에서 만나지 않으며, 어느 두 직선도 평행하지 않는다). 이 직선들에 의해서 평면은 몇 개로 나누어 지겠는가?

**풀이** 이러한 문제는 다음과 같이 생각할 수 있다. 직선  $r$ 개가 있을 때 평면수를  $a_r$ 로 생각하자.

| $r$   | 1                                                                                 | 2                                                                                 | 3                                                                                 | 4                                                                                 | ... | $n$                   |
|-------|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|-----|-----------------------|
|       |  |  |  |  | ... |                       |
| $a_r$ | $a_1=2$                                                                           | $a_2=2+2$<br>$=a_1+2$                                                             | $a_3=2+2+3$<br>$=a_2+3$                                                           | $a_4=2+2+3+4$<br>$=a_3+4$                                                         | ... | $a_n$<br>$=a_{n-1}+n$ |

결국, 제  $r$ 번째 직선을 한 개 더 그을 때마다  $r$ 개의 부분이 더 생기므로  $a_1=2$ (초기조건),  $a_n=a_{n-1}+n$ 이란 관계식을 유도하여  $a_n$ 을 구할 수 있다(제A.2절을 참조하라).

#### 4. 일반수열에서 점화식을 만드는 법

1) 수열  $\{a_n\}$ 의 일반항에서  $n$ 을 포함한 식을 제거한다.

**보기** 1.  $a_n=3^n+2$ 인 수열의 점화식을 구하여라.

**풀이**  $a_{n+1}=3^{n+1}+2=3 \cdot 3^n+2$  .....①  
 $a_n=3^n+2$  .....②  
 ①-②×3 :  $a_{n+1}-3a_n=-4$ , (단,  $a_1=5$ )

**보기** 2.  $a_n=3^n+2n+1$ 인 수열의 점화식을 구하여라.

**풀이**  $a_n=3^n+2n+1$  .....①  
 $a_{n+1}=3^{n+1}+2(n+1)+1=3 \cdot 3^n+2n+3$  .....②  
 $a_{n+2}=3^{n+2}+2(n+2)+1=9 \cdot 3^n+2n+5$  .....③  
 ③-② :  $a_{n+2}-a_{n+1}=6 \cdot 3^n+2$  .....④  
 ②-① :  $a_{n+1}-a_n=2 \cdot 3^n+2$  .....⑤  
 ④-⑤×3 :  $a_{n+2}-4a_{n+1}+3a_n=-4$ , (단,  $a_1=6, a_2=14$ )

2) 점화식 속에  $n$ 에 관한 식을 포함하면  $n$ 의 식을 제거하여 다른 점화식으로 유도할 수 있다.

**보기** 1. 점화식  $a_{n+1}=3a_n+2^n$ 에서  $2^n$ 을 제거한 다른 형태의 점화식으로 고쳐라.

**풀이**  $a_{n+1}=3a_n+2^n$  .....①  
 $a_{n+2}=3a_{n+1}+2^{n+1}=3a_{n+1}+2\cdot 2^n$  .....②  
 ②-① $\times 2$   $2a_{n+2}-2a_{n+1}=3a_{n+1}-6a_n$   
 $\therefore a_{n+2}=5a_{n+1}-6a_n$

**보기 2.** 점화식  $a_{n+2}+2a_{n+1}-3a_n=n+3$ 에서  $n$ 을 제거한 점화식으로 변형하여라.

**풀이**  $a_{n+2}+2a_{n+1}-3a_n=n+3$  .....①  
 $a_{n+3}+2a_{n+2}-3a_{n+1}=n+4$  .....②  
 ②-① :  $a_{n+3}+a_{n+2}-5a_{n+1}+3a_n=1$

이 책의 특징

점화식으로부터 수열의 일반항을 구하는 직관에 의존한 일반적인 방법은 축차대입법이다. 예를 들어  $a_1=1, a_{n+1}=2a_n+2^n$ 인 수열의 일반항은  $a_1=1, a_2=2a_1+2=4=2\times 2, a_3=2a_2+2^2=12=3\times 4, a_4=2a_3+2^3=32=4\times 8, a_5=2a_4+2^4=80=5\times 16$ 에 의해  $a_n=n2^{n-1}$ 로 추정한다. 이 축차대입법은 다음과 같이 항상 수학적 귀납법으로 검토하여야 한다.

$n=1$ 에서  $a_1=1\times 2^0$ 는 1  
 $n=k$ 일 때  $a_k=k\cdot 2^{k-1}$ 이라 하면  
 $n=k+1$ 에서  $a_{k+1}=2a_k+2^k=k\cdot 2^k+2^k=2^k(k+1)$   
 $n=k+1$ 에서 성립하므로 모든 자연수  $n$ 에 대해  $a_n=n\cdot 2^{n-1}$ 이다.

본서에서는 좀더 수학적 추리력에 의존한 방법을 사용하기 위해 수열을 다음과 같이 정의한다.

- \*1. 수열은 정의역이 자연수인 함수다.
- \*2. 수열  $\{a_n\}$ 과 수열  $\{b_n\}$ 이 있을 때  $\{\lambda a_n : \lambda$ 는 상수 $\}, \{a_n + b_n\}, \{a_n - b_n\}, \{a_n \times b_n\}, \{a_n \div b_n$  단,  $b_n \neq 0\}$  또한 수열이다.

이러한 생각은 보통 역사적으로 수열에서 함수로 도입되는 사실을 역이용하여 수열을 공부하는 데 있어서 함수의 개념을 도입한 것이다.

**보기 1.** 위에서 든 예를 \*1과 \*2를 이용하여 풀어라.

**풀이**  $a_1=1, a_{n+1}=2a_n+2^n$ 의 양변을  $2^n$ 으로 나누면 다음과 같다.  
 $\frac{a_{n+1}}{2^n} = \frac{a_n}{2^{n-1}} + 1$  .....①  
 만일,  $\frac{a_n}{2^{n-1}} = b_n$ 이라면,  $b_1 = \frac{a_1}{2^0} = 1$ 이고 다음 식이 성립한다. ....②  
 ①, ② :  $b_{n+1} = b_n + 1$  .....③  
 ③은 공차가 1인 등차수열이므로  $b_n = b_1 + (n-1) = n$ 이다. ....④  
 ②, ④ :  $a_n = b_n 2^{n-1} = n \cdot 2^{n-1}$

❖ 설명 : ②에서  $2^{n-1} = c_n$ 으로 생각하면  $\frac{a_n}{c_n}$ 도 수열이므로  $b_n$ 으로 둔다.

이 때  $b_{n+1} = \frac{a_{n+1}}{c_{n+1}} = \frac{a_{n+1}}{2^n}$ 이므로 ①에서  $b_{n+1} = b_n + 1 \cdots \cdots$  ③으로 둘 수 있다.

**보기 2.**  $a_1 = 1$ ,  $a_n = \left(1 - \frac{1}{n^2}\right)a_{n-1}$ 인 수열의 극한을 조사하라. ( $n \geq 2$ )

(1968년도 서울대 입시문제)

**풀이**  $a_n = \left(\frac{n^2-1}{n^2}\right)a_{n-1} = \frac{(n+1)(n-1)}{n \cdot n}a_{n-1}$ 의 양변에  $\frac{n}{n+1}$ 을 곱한다.

$$\frac{n}{n+1}a_n = \frac{n-1}{n}a_{n-1} \quad \cdots \cdots \text{①}$$

$$\frac{n}{n+1}a_n = b_n \text{이라 하면, } b_1 = \frac{1}{2}a_1 = \frac{1}{2} \text{이고 다음 식이 성립한다.} \quad \cdots \cdots \text{②}$$

$$\text{①, ② : } b_n = b_{n-1} \text{이므로 } b_n = b_{n-1} = b_{n-2} = \cdots = b_2 = b_1 = \frac{1}{2} \text{이다.} \quad \cdots \cdots \text{③}$$

$$\text{②, ③ : } a_n = \frac{n+1}{n}b_n = \frac{n+1}{2n}$$

$$\therefore \lim_{n \rightarrow \infty} a_n = \frac{1}{2}$$

❖ 설명 : ①에서  $\frac{n}{n+1} = c_n$ 으로 생각하면  $a_n \times c_n$ 도 수열이므로  $b_n$ 으로 둔다.

이 때  $b_{n+1} = a_{n+1} \times c_{n+1} = a_{n+1} \times \frac{n-1}{n}$ 이므로 ①에서  $b_n = b_{n-1} \cdots \cdots$  ③으로 둘 수 있다.

**보기 3.**  $a_1 = 1$ ,  $a_{n+1} = 2a_n + n - 1$ 일 때  $a_n$ 을 구하여라.

**풀이**  $a_{n+1} = 2a_n + n - 1 = 2a_n + 2n - (n+1)$

$$a_{n+1} + (n+1) = 2(a_n + n) \quad \cdots \cdots \text{①}$$

$$a_n + n = b_n \text{이라 하면 } b_1 = a_1 + 1 = 2 \text{이다.} \quad \cdots \cdots \text{②}$$

$$\text{①, ②에서 } b_{n+1} = 2b_n \text{은 공비 2인 등비수열이므로 } b_n = 2^{n-1}b_1 = 2^n \text{이다.} \quad \cdots \cdots \text{③}$$

$$\text{②, ③ : } a_n = b_n - n = 2^n - n$$

❖ 설명 : ②에서  $n = c_n$ 이라 하면  $a_n + c_n$ 도 수열이므로  $b_n$ 으로 둔다. 이 때

$$b_{n+1} = a_{n+1} + c_{n+1} = a_{n+1} + (n+1) \text{이므로 ①에서 } b_{n+1} = 2b_n \text{이 성립한다.}$$

결국 본서에서 모든 점화수열은 \*1과 \*2를 사용하여 다른 수열로 고친 뒤, 고친 수열의 점화식이 등비수열 또는 등차수열이 되도록 하는 방법을 썼다. 제A.2절과 제A.3절은 모든 수열을 등비수열 또는 등차수열로 고치는 방법을 연구해 놓은 것이다.

또한 제A.5절과 제A.6절의 극한의 문제는 재래의 그래프에 의존한 직관적인 방법을 지양하고 좀더 수학적으로 풀기 위해 극한의 문제를 다음과 같이 두 단계로 나누어 풀었다.

제1단계 : 수열이 수렴함을 증명한다.

제2단계 : 제1단계의 성립하에 극한값을 구한다.

제1단계의 수렴함을 증명하기 위해서 종래의 미적분에서 쓰던 '함수가 구간에서 단조함수이고 유계이면 수렴한다'는 사실을 수열에 이용하였다.

만일 수열이 단조함수가 아닌 경우는  $A < X < B$ 에서  $\lim B = A$ 이면  $\lim X = A$ 가 됨을 이용하였다.

제2단계의 극한값을 구하는 과정에 있어서는 종래의 그래프를 사용하는 방법인 특성방정식을 사용하였다.





2. 수열  $a_1, a_2, a_3, \dots, a_n, \dots$ 에서  $a_1=1, a_{n+1}=a_n+(n+1)$ 이다. 다음 물음에 답하여라.

(1)  $a_n$ 을  $n$ 의 식으로 나타내어라.

(2)  $\sum_{k=1}^n \frac{1}{a_k}$ 을 구하여라.

(3)  $\sum_{k=1}^{\infty} \frac{1}{a_k}$ 을 구하여라.

3. 수열  $\{a_n\}, \{b_n\}$ 에서  $a_0=1, b_n=a_n-a_{n-1}$ 이다. 또 수열  $\{b_n\}$ 은 공차가  $a_1+b_1$ 인 등차 수열이다. ( $n \geq 1$ )

(1) 일반항을  $a_1$ 과  $n$ 으로 나타내어라.

(2)  $\lim_{n \rightarrow \infty} \frac{1}{n^3} \sum_{k=0}^n a_k = 1$ 일 때  $a_1$ 을 구하여라.

4. 수열  $\{a_n\}$ 에서  $a_1 > 0$ 이고  $a_{i+1} - a_i = -\frac{10}{3}$ 이다. 2차방정식  $x^2 + 2a_i x + a_i^2 - \frac{9}{4} = 0$ 의 실근을 작은 순서로 나열해서 얻은 수열의 제  $n$ 항까지의 합을  $S_n$ 이라 하자.

(1)  $S_n$ 을  $a_1$ 과  $n$ 으로 표시하여라.

(2)  $\lim_{n \rightarrow \infty} \frac{S_n}{n^2}$ 을 구하여라.

**정리 2.**  $a_{n+1} = pa_n + f(n)$ 이면 다음 식이 성립한다.

$$a_n = \{a_1 - g(1)\}p^{n-1} + g(n)$$

단,  $g(n+1) = pg(n) + f(n)$

**증명**  $a_{n+1} = pa_n + f(n)$  .....①

$g(n+1) = pg(n) + f(n)$  .....②

①-② :  $a_{n+1} - g(n+1) = p\{a_n - g(n)\}$ 에서  
 $a_n - g(n) = b_n$ 이라면 다음 식이 성립한다. ....③

$b_{n+1} = pb_n$   
 $\therefore b_n = p^{n-1}b_1 = p^{n-1}\{a_1 - g(1)\}$  .....④

③, ④ :  $a_n = b_n + g(n) = \{a_1 - g(1)\}p^{n-1} + g(n)$  Q.E.D

※ 유의 : 위와 같은 방법에서는  $g(n)$ 을 구하는 것이 가장 중요하다. 이  $g(n)$ 을 구하는 방법은 다음의 각 따름정리에서 여러 가지 형태로 분류하여 기술하였다. 정리 1은 정리 2의  $p=1$ 인 경우이나 정리 2와 같은 방법을 쓰려면  $g(n)$ 을 구하는 것이 어렵다. 다음의 따름정리는  $p=1$ 에 적용되지 않음을 유의하라.

**정리 2-1.** [ $f(n)$ 이 상수함수]

$a_{n+1} = pa_n + q$  ( $p, q$ 는 상수,  $p \neq 1$ )이면 다음 식이 성립한다.

$$a_n = \frac{q}{1-p}(1-p^{n-1}) + a_1 p^{n-1}$$

**[증명]**  $f(n)$ 이 상수함수이면  $g(n)$ 도 상수함수이다.

$$a_{n+1} = pa_n + q \quad \dots\dots\dots ①$$

$$x = px + q, \left(x = \frac{q}{1-p}\right) \quad \dots\dots\dots ②$$

$$① - ② : a_{n+1} - x = p(a_n - x) \text{에서 } a_n - x = b_n \text{이라면} \quad \dots\dots\dots ③$$

$$b_{n+1} = pb_n$$

$$\therefore b_n = p^{n-1}b_1 = (a_1 - x)p^{n-1} \quad \dots\dots\dots ④$$

$$③, ④ : a_n = b_n + x = (a_1 - x)p^{n-1} + x = x(1 - p^{n-1}) + a_1p^{n-1}$$

$$= \frac{q}{1-p}(1 - p^{n-1}) + a_1p^{n-1} \quad \text{Q.E.D}$$

**[보기 1.]**  $a_1=1, a_{n+1}=2a_n+3$ 이다.  $a_n$ 을 구하여라.

**[풀이]**  $a_{n+1}=2a_n+3 \quad \dots\dots\dots ①$

$$x=2x+3, (x=-3) \quad \dots\dots\dots ②$$

$$①-② : a_{n+1}-x=2(a_n-x) \text{에서 } a_n-x=b_n \text{이라면} \quad \dots\dots\dots ③$$

$$b_{n+1}=2b_n$$

$$\therefore b_n=2^{n-1}b_1=2^{n-1}(a_1-x)=2^{n-1} \quad \dots\dots\dots ④$$

$$③, ④ : a_n = b_n + x = 2^{n-1} - 3$$

《유제》  $a_1=0, a_{n+1}=-\frac{1}{2}a_n+1$ 일 때  $a_n$ 을 구하여라. ☞  $a_n = \frac{2}{3} \left\{ 1 - \left(-\frac{1}{2}\right)^{n-1} \right\}$

**[정리 2-2.]** [ $f(n)$ 이 정함수]

$a_{n+1}=pa_n+f(n)$ 에서  $f(n)$ 이 정함수이면  $g(n)$ 도 같은 차수의 미지의 정함수로 놓아  $g(n)$ 을 구한 뒤  $a_n$ 을 구한다. (단,  $p \neq 1$ )

**[보기 2.]**  $a_1=6, a_{n+1}=3a_n-4n-4$ 에서  $a_n$ 을 구하여라.

( $f(n) = -4n - 4$ , 즉 1차식이므로  $g(n) = an + \beta$ 로 놓는다.)

**[풀이]**  $a_{n+1}=3a_n-4n-4 \quad \dots\dots\dots ①$

$$\alpha(n+1) + \beta = 3(\alpha n + \beta) - 4n - 4 \quad \dots\dots\dots ②$$

$$① - ② : a_{n+1} - \alpha(n+1) - \beta = 3(a_n - \alpha n - \beta) \text{에서}$$

$$a_n - \alpha n - \beta = b_n \text{이라 하자.} \quad \dots\dots\dots ③$$

$$②, ③ : b_{n+1} = 3b_n$$

$$\therefore b_n = 3^{n-1}b_1 \quad \dots\dots\dots ④$$

②에서 항등식의 원리를 적용  $\alpha=2, \beta=3$ 이므로  $b_1=1$ 이다.

$$③, ④ : a_n = 2n + 3 + 3^{n-1}$$

《유제》  $a_1=1, a_{n+1}=2a_n+n-1$ 일 때  $a_n$ 을 구하여라. ☞  $a_n = 2^n - n$

**정리 2-3.** [ $f(n)$ 이 지수함수]

$$a_{n+1} = pa_n + lq^n \quad (\text{단, } p \neq 1, l \neq 0)$$

$$\text{[I]} \quad p \neq q, \quad a_n = \frac{lq^n}{q-p} + p^{n-1} \left( a_1 - \frac{lq}{q-p} \right)$$

$$\text{[II]} \quad p = q, \quad a_n = \{a_1 + (n-1)l\}p^{n-1}$$

**증명** [I]  $q \neq p$ 인 경우  $f(n)$ 이  $lq^n$  꼴이면  $g(n) = xq^n$ 으로 놓는다.

$$a_{n+1} = pa_n + lq^n \quad \dots\dots\dots \text{①}$$

$$xq^{n+1} = p(xq^n) + lq^n \quad \dots\dots\dots \text{②}$$

$$\text{①} - \text{②} : a_{n+1} - xq^{n+1} = p(a_n - xq^n) \text{에서}$$

$$b_n = a_n - xq^n \text{이라면 다음 식이 성립한다.} \quad \dots\dots\dots \text{③}$$

$$b_{n+1} = pb_n$$

$$\therefore b_n = p^{n-1}b_1 = p^{n-1}(a_1 - xq) \quad \dots\dots\dots \text{④}$$

②에서 항등식의 원리를 적용하면 다음 식이 성립한다.

$$xq = px + l \quad \therefore x = \frac{l}{q-p}$$

$$\text{③, ④} : a_n = xq^n + b_n = \frac{l}{q-p}q^n + \left( a_1 - \frac{lq}{q-p} \right) p^{n-1}$$

[II]  $q = p$ 인 경우에는 위의  $xq = px + l$ 에서  $l=0$ 가 조건에 모순이므로 이러한 방법을 쓸 수 없다.

$a_{n+1} = pa_n + lp^n$ 의 양변을  $p^n$ 으로 나누어

$$\frac{a_{n+1}}{p^n} = \frac{a_n}{p^{n-1}} + l \text{에서 } \frac{a_n}{p^{n-1}} = b_n \text{이라고 하자.} \quad \dots\dots\dots \text{①}$$

$$b_{n+1} = b_n + l$$

$$\therefore b_n = b_1 + (n-1)l = a_1 + (n-1)l \quad \dots\dots\dots \text{②}$$

$$\text{①, ②} : a_n = p^{n-1}b_n = \{a_1 + (n-1)l\}p^{n-1} \quad \text{Q.E.D}$$

※ 유의: [II]에서  $g(n)$ 을 구하여  $a_n$ 을 구하려면  $g(n) = (xn+y)q^n$ 으로 놓는다(보기 4의 방법 2를 참조하라).

[I]에서도 [II]와 같이 양변을  $q^n$ 으로 나누어 구할 수도 있으나 한 단계의 과정이 늘어나므로 위의 방법이 좋다(보기 3의 방법 2를 참조하라).

**보기 3.**  $a_1=1, a_{n+1}=3a_n+2^n$ 일 때  $a_n$ 을 구하여라.

▶ **방법 1.** [ $f(n)=2^n$ 이므로  $g(n)=x2^n$ 으로 놓는다.]

$$a_{n+1} = 3a_n + 2^n \quad \dots\dots\dots \text{①}$$

$$x2^{n+1} = 3(x2^n) + 2^n, \quad (x=-1) \quad \dots\dots\dots \text{②}$$

$$\text{①} - \text{②} : a_{n+1} - x2^{n+1} = 3(a_n - x2^n) \text{에서}$$

$$a_n - x2^n = b_n \text{이라면 다음 식이 성립한다.} \quad \dots\dots\dots \text{③}$$

$$b_{n+1} = 3b_n \quad \therefore b_n = 3^{n-1}b_1 = 3^{n-1}(a_1 - 2x) = 3^n \quad \dots\dots\dots \text{④}$$

$$\text{③, ④} : a_n = 3^n - 2^n$$

▶ 방법 2.  $a_{n+1}=3a_n+2^n$ 의 양변을  $2^n$ 으로 나누면 다음과 같다. ....①

$$\frac{a_{n+1}}{2^n} = \frac{3}{2} \times \frac{a_n}{2^{n-1}} + 1$$

이 식에서  $\frac{a_n}{2^{n-1}}=b_n$ 이라 놓으면 다음 식이 성립한다. ....②

$$b_{n+1} = \frac{3}{2}b_n + 1 \quad \dots\dots③$$

$$x = \frac{3}{2}x + 1 \quad \dots\dots④$$

③-④ :  $b_{n+1}-x = \frac{3}{2}(b_n-x)$ 에서  
 $b_n-x=c_n$ 이라면 다음 식이 성립한다. ....⑤

$$c_{n+1} = \frac{3}{2}c_n$$

$$\therefore c_n = \left(\frac{3}{2}\right)^{n-1} c_1 = \frac{3^n}{2^{n-1}} \quad \dots\dots⑥$$

②, ⑤, ⑥ :  $a_n = 2^{n-1}b_n = 2^{n-1}(c_n+x) = 2^{n-1}\left(\frac{3^n}{2^{n-1}}-2\right) = 3^n - 2^n$

《유제》  $a_1=0, a_{n+1}=4a_n+2^{n-1}$ 에서  $a_n$ 을 구하여라. ☞  $a_n=2^{n-2}(2^{n-1}-1)$

☞ 4.  $a_{n+1}=2a_n-3 \cdot 2^n, a_1=4$ 인  $a_n$ 을 구하여라.

▶ 방법 1.  $a_{n+1}=2a_n-3 \cdot 2^n$ 의 양변을  $2^n$ 으로 나누면

$$\frac{a_{n+1}}{2^n} = \frac{a_n}{2^{n-1}} - 3 \text{에서}$$

$\frac{a_n}{2^{n-1}}=b_n$ 이라 놓으면 다음 식이 성립한다. ....①

$$b_{n+1} = b_n - 3 \quad \dots\dots②$$

$$\therefore b_n = b_1 - 3(n-1) = 7 - 3n \quad \dots\dots②$$

①, ② :  $a_n = 2^{n-1}b_n = (7-3n)2^{n-1}$

▶ 방법 2. [ $g(n)=(xn+y)2^n$ 으로 한다.] ....①

$$\{x(n+1)+y\}2^{n+1} = 2 \times (xn+y) \cdot 2^n - 3 \cdot 2^n \quad \dots\dots②$$

②에서 양변을  $2^n$ 으로 나눈 뒤 항등식 원리를 적용하면  
 $x = -\frac{3}{2}$ 이고  $y$ 는 부정이므로 0으로 한다.

①-② :  $a_{n+1} - \{x(n+1)+y\} \cdot 2^{n+1} = 2\{a_n - (xn+y) \cdot 2^n\}$ 에서  
 $a_n - (xn+y) \cdot 2^n = b_n$ 이라면, 다음 식이 성립한다. ....③

$$b_{n+1} = 2b_n \quad \dots\dots④$$

$$\therefore b_n = 2^{n-1}b_1 = 7 \cdot 2^{n-1}$$

③, ④ :  $a_n = (xn+y) \cdot 2^n + b_n = -3n \cdot 2^{n-1} + 7 \cdot 2^{n-1} = 2^{n-1}(7-3n)$   
 《유제》  $a_1=1, a_{n+1}=3a_n+3^n$ 인  $a_n$ 을 구하여라. ☞  $a_n=3^{n-1}n$

**정리 2-4.**  $\{f(n)\}$ 이 지수함수 $\times$ 정함수

$$a_{n+1} = pa_n + q^n h(n) \quad (\text{단, } h(n) \text{은 정함수, } p \neq 1)$$

$$[\text{I}] \quad p \neq q, \quad a_n = l(n)q^n + \{a_1 - ql(1)\}p^{n-1}$$

$$[\text{단, } ql(n+1) - pl(n) = h(n)]$$

$$[\text{II}] \quad p = q, \quad a_n = p^{n-1} \left\{ a_1 + \sum_{k=1}^{n-1} h(k) \right\}$$

**증명**  $[\text{I}] \quad p \neq q$ 인 경우

$$a_{n+1} = pa_n + q^n h(n) \quad \dots\dots\dots ①$$

$$q^{n+1}l(n+1) = pq^n l(n) + q^n h(n) \quad \dots\dots\dots ②$$

$$① - ② : a_{n+1} - q^{n+1}l(n+1) = p\{a_n - q^n l(n)\} \text{에서} \quad \dots\dots\dots ③$$

$$a_n - q^n l(n) = b_n \text{이라면 다음 식의 성립한다.} \quad \dots\dots\dots ④$$

$$\begin{aligned} ③, ④ : b_{n+1} &= pb_n \\ \therefore b_n &= p^{n-1}b_1 = p^{n-1}\{a_1 - ql(1)\} \quad \dots\dots\dots ⑤ \end{aligned}$$

$$④, ⑤ : a_n = b_n + q^n l(n) = q^n l(n) + p^{n-1}\{a_1 - ql(1)\}$$

이 때 ②의 양변을  $q^n$ 으로 나누면  $ql(n+1) - pl(n) = h(n)$ 이다. 이 때  $l(n)$ 은  $h(n)$ 과 같은 차수의 미지의 정식으로 두어 항등식의 원리에 의해 구한다.

$[\text{II}] \quad p = q$ 의 경우

$a_{n+1} = pa_n + p^n h(n)$ 의 양변을  $p^n$ 으로 나누면

$$\frac{a_{n+1}}{p^n} = \frac{a_n}{p^{n-1}} + h(n) \text{에서 } \frac{a_n}{p^{n-1}} = b_n \text{이라면}$$

다음 식이 성립한다. .....①

$$b_{n+1} = b_n + h(n)$$

$$\therefore b_n = b_1 + \sum_{k=1}^{n-1} h(k) = a_1 + \sum_{k=1}^{n-1} h(k) \quad \dots\dots\dots ②$$

$$①, ② : a_n = p^{n-1}b_n = p^{n-1} \left\{ a_1 + \sum_{k=1}^{n-1} h(k) \right\} \quad \text{Q.E.D}$$

※ 유의 :  $[\text{I}]$ 에서  $[\text{II}]$ 와 같은 방법을 써도 되나, 계산이 1단계가 늘어 복잡해진다 (보기 5의 방법 2를 참조하라).

$[\text{II}]$ 에서  $[\text{I}]$ 과 같이  $l(n)$ 을 구하려면  $l(n)$ 을  $h(n)$ 보다 1차 높은 정식으로 둔다 (보기 6의 방법 2를 참조하라).

**보기 5.**  $a_{n+1} = 2a_n + 3^n(n+2)$ ,  $a_1 = 2$ 일 때  $a_n$ 을 구하여라.

▶ **방법 1.**  $\{h(n)\}$ 이 1차식이므로  $l(n) = an + \beta$ 로 둔다.]

$$a_{n+1} = 2a_n + 3^n(n+2) \quad \dots\dots\dots ①$$

$$3^{n+1}(an+1+\beta) = 2 \times 3^n(an+\beta) + 3^n(n+2) \quad \dots\dots\dots ②$$

$$①, ② : a_{n+1} - 3^{n+1}(an+1+\beta) = 2 \times \{a_n - 3^n(an+\beta)\}$$

$$a_n - 3^n(an+\beta) = b_n \text{이라면, 다음 식이 성립한다.} \quad \dots\dots\dots ③$$

$$b_{n+1} = 2b_n$$

$$\therefore b_n = 2^{n-1}b_1 = 2^{n-1}(a_1 - 3a - 3\beta) \quad \dots\dots\dots ④$$

또한 ②의 양변을  $3^n$ 으로 나누어 항등식원리를 적용하면  $a=1, \beta=-1$ 이다.

$$\textcircled{3}, \textcircled{4} : a_n = b_n + 3^n(an + \beta) = 2^n + (n-1)3^n$$

▶ **방법 2.**  $a_{n+1} = 2a_n + 3^n(n+2)$ 의 양변을  $3^n$ 으로 나누면 다음 식이 성립한다.

$$\frac{a_{n+1}}{3^n} = \frac{2}{3} \times \frac{a_n}{3^{n-1}} + (n+2)$$

이 식에서  $\frac{a_n}{3^{n-1}} = b_n$ 이라면, 다음 식이 성립한다. .....①

$$b_{n+1} = \frac{2}{3}b_n + (n+2) \quad \dots\dots\dots ②$$

$$\alpha(n+1) + \beta = \frac{2}{3}(an + \beta) + (n+2), \quad (\alpha=3, \beta=-3) \quad \dots\dots\dots ③$$

$$\textcircled{2} - \textcircled{3} : b_{n+1}(\overline{an+1} + \beta) = \frac{2}{3}(b_n - an - \beta) \text{에서}$$

$c_n = b_n - (an + \beta)$ 라하면 다음 식이 성립한다. .....④

$$c_{n+1} = \frac{2}{3}c_n$$

$$\therefore c_n = \left(\frac{2}{3}\right)^{n-1} c_1 = \left(\frac{2}{3}\right)^{n-1} (b_1 - a - \beta) = \frac{2^n}{3^{n-1}} \quad \dots\dots\dots ⑤$$

$$\textcircled{1}, \textcircled{4}, \textcircled{5} : a_n = 3^{n-1}b_n = 3^{n-1}\{c_n + (an + \beta)\} = 3^{n-1}\left(\frac{2^n}{3^{n-1}} + 3n - 3\right) = 2^n + 3^n(n-1)$$

《유제》  $a_1=2, a_{n+1}=2a_n+3^{-n}(7-10n)$ 에서  $a_n$ 을 구하여라. 답  $a_n=2^{n-1}+3 \cdot 3^{-n}(2n-1)$

**보기 6.**  $a_{n+1} = -a_n + (-1)^n(n+1), a_1=0$ 인  $a_n$ 을 구하여라.

▶ **방법 1.**  $a_{n+1} = -a_n + (-1)^n(n+1)$ 의 양변을  $(-1)^n$ 으로 나누면 다음 식이 성립한다.

$$\frac{a_{n+1}}{(-1)^n} = \frac{a_n}{(-1)^{n-1}} + (n+1) \text{에서 } \frac{a_n}{(-1)^{n-1}} = b_n \text{이라면}$$

$a_1=b_1=0$ 이고 다음 식이 성립한다. .....①

$$b_{n+1} = b_n + (n+1)$$

$$\therefore b_n = b_1 + \sum_{k=1}^{n-1} (k+1) = \frac{1}{2}(n-1)(n+2) \quad \dots\dots\dots ②$$

$$\textcircled{1}, \textcircled{2} : a_n = (-1)^{n-1}b_n = \frac{1}{2}(-1)^{n-1}(n-1)(n+2)$$

▶ **방법 2.**  $a_{n+1} = -a_n + (-1)^n(n+1), h(n) = n+1$ 이 1차식이므로  $l(n)$ 은 2차식  $xn^2 + yn + z$ 로 둔다.

$$a_{n+1} = -a_n + (-1)^n(n+1) \quad \dots\dots\dots ①$$

$$(-1)^{n+1}\{x(n+1)^2 + y(n+1) + z\} = -(-1)^n(xn^2 + yn + z) + (-1)^n(n+1) \quad \dots\dots\dots ②$$

①-②하고  $a_n - (-1)^n(xn^2 + yn + z) = b_n$ 이라면 다음 식이 성립한다. .....③

$$b_{n+1} = -b_n$$

$$\therefore b_n = (-1)^{n-1}b_1 = (-1)^{n-1}(a_1 + x + y + z) \quad \dots\dots\dots ④$$

②에서 양변을  $(-1)^n$ 으로 나누어 계산하면  $x = -\frac{1}{2}$ ,  $y = -\frac{1}{2}$ ,  $z$ 는 부정이므로  $z=0$ 으로 한다. 따라서  $b_1 = a_1 + x + y + z = -1$ 이다.

$$\begin{aligned} \text{③, ④} : a_n &= (-1)^n(xn^2 + yn + z) + b_n \\ &= \frac{1}{2}(-1)^{n+1}(n^2 + n) + (-1)^n \\ &= \frac{1}{2}(-1)^{n+1}(n+2)(n-1) \end{aligned}$$

《유제》  $a_1=0$ ,  $2a_{n+1} = a_n + 2n(n+1)(n-1) \cdot 2^{-n}$ 에서  $a_n$ 을 구하여라.

답  $a_n = 2^{-(n+1)}(n-2)(n-1)n(n+1)$

**정리 2-5.**  $\{f(n)$ 이 기타의 함수꼴

[I] 분수함수꼴 : 부분분수분해

[II] 계승함수 :  $(n+1)! = n!(n+1)$

$\left. \begin{aligned} \frac{1}{n!} &= \frac{n+1}{(n+1)!} \end{aligned} \right\}$  을 이용하여  $f(n)$ 을 분해

**보기 7.**  $a_{n+1} = 3a_n - \frac{2n+6}{n(n+1)(n+2)}$ ,  $a_1 = \frac{3}{2}$ 인  $a_n$ 을 구하여라.

**풀이**  $\frac{2n+6}{n(n+1)(n+2)} = \frac{\alpha}{n(n+1)} + \frac{\beta}{(n+1)(n+2)}$ 이면 다음 식이 성립한다.

$$2n+6 = \alpha(n+2) + \beta n = (\alpha+\beta)n + 2\alpha \quad \therefore \alpha=3, \beta=-1$$

따라서 다음 식이 성립한다.

$$a_{n+1} = 3a_n - \frac{3}{n(n+1)} + \frac{1}{(n+1)(n+2)}$$

즉, 다음 식이 성립한다.

$$a_{n+1} - \frac{1}{(n+1)(n+2)} = 3 \left\{ a_n - \frac{1}{n(n+1)} \right\}$$

이 식에서  $b_n = a_n - \frac{1}{n(n+1)}$  이라면 다음 식이 성립한다. .....①

$$b_{n+1} = 3b_n \quad \therefore b_n = 3^{n-1}b_1 = 3^{n-1}, (\because b_1 = a_1 - \frac{1}{2} = 1) \quad \text{.....②}$$

$$\text{①, ②} : a_n = b_n + \frac{1}{n(n+1)} = 3^{n-1} + \frac{1}{n(n+1)}$$

《유제》  $a_1=1$ ,  $2a_{n+1} - a_n = \frac{n-2}{n(n+1)(n+2)}$ 일 때  $a_n$ 을 구하여라.

답  $a_n = \left(\frac{1}{2}\right)^n + \frac{1}{n(n+1)}$

**보기 8.**  $a_{n+1} = 2a_n - \frac{2n+1}{(n+1)!}$ ,  $a_1=3$ 인  $a_n$ 을 구하여라.



$$\text{[풀이]} \quad a_{n+1} = 2a_n - \frac{2n+1}{(n+1)!} = 2a_n - \frac{2(n+1)-1}{(n+1)!} = 2a_n - \frac{2}{n!} + \frac{1}{(n+1)!}$$

$a_{n+1} - \frac{1}{(n+1)!} = 2\left(a_n - \frac{1}{n!}\right)$ 에서  $a_n - \frac{1}{n!} = b_n$ 이라 놓으면 다음 식이 성립한다.

.....①

$$b_{n+1} = 2b_n$$

$$\therefore b_n = 2^{n-1}b_1 = 2^n, \quad (\because b_1 = a_1 - 1 = 2)$$

.....②

$$\text{①, ②} : a_n = b_n + \frac{1}{n!} = 2^n + \frac{1}{n!}$$

《유제》  $2a_1 = 1$ ,  $2a_{n+1} = 6a_n + n!$  ( $n \geq 2$ )인  $a_n$ 을 구하여라.

$$\text{[답]} \quad a_n = \frac{1}{2}n!$$

## ══════ 연습문제 II ══════

1. 수열  $\{a_n\}$ 은  $a_{n+1} = pa_n + q$ 의 관계를 만족한다. 이 수열의 수렴, 발산을 조사하고 수렴하는 경우 극한값을  $a_1, p, q$ 로 나타내어라.

2. 2차방정식  $a_{n-1}x^2 - a_nx + 1 = 0$ 의 2근을  $a_n, \beta_n$ 이라 할 때 다음과 같은 관계식을 만족한다.

$$3a_n - a_n\beta_n + 3\beta_n = 1, \quad (n=2, 3, \dots)$$

(1)  $a_n$ 과  $a_{n-1}$ 의 관계식을 구하여라.

(2)  $a_n$ 을  $a_1$ 과  $n$ 의 식으로 나타내어라.

(3)  $\lim_{n \rightarrow \infty} a_n$ 을 구하여라.

3.  $a_1 = 1 + i$ ,  $a_2 = 1 + \sqrt{3}i$ ,  $a_{n+1} = a_1 + ka_n$  ( $n \geq 2$ )로 정의된 복소수열이 있다(단,  $k$ 는 양수). 복소수  $a_n$ 의 편각의  $\tan$ 의 극한값을 구하여라.

4.  $x_1 = a$ ,  $\tan 2x_{n+1} \cdot \tan x_n = 1$ ,  $\frac{\pi}{3} < x_n < \frac{2}{3}\pi$ 라 하자.

(1)  $x_{n+1}$ 과  $x_n$ 과의 관계식을 삼각함수를 사용하지 않고 나타내어라.

(2)  $x_n$ 을 구하여라.

(3)  $\lim_{n \rightarrow \infty} x_n$ 을 구하여라.

5.  $a_1 = 1$ ,  $a_n = 3a_{n-1} + 2n + 2$ 에서  $a_n$ 을 구하고 극한을 조사하여라.

6.  $a, b$  (단,  $a > b$ ),  $r, s$  (단,  $r \neq s$ )는 양수이다. 수열  $\{c_n\}$ 이 다음과 같은 조건을 만족한

다고 하자.

$$c_1 = a - b, \quad c_n = rc_{n-1} - b \cdot s^{n-1}$$

- (1)  $c_n$ 을 구하여라.
  - (2)  $c_k = 0$ 에서,  $\frac{s}{r} = x$ ,  $\frac{a}{b} = y$ 라 할 때  $y$ 를  $x$ 와  $k$ 로 표시하여라.
7.  $a_{n+1} + a_n = 2n^3 + 3n^2 + 3n - 3$ ,  $a_1 = 1$ 인  $a_n$ 을 구하여라.
  8.  $a_{n+1} = 3a_n + 5 \cdot 3^n n(n+1)(n+2)(n+3)$ ,  $a_1 = 0$ 인  $a_n$ 을 구하여라.
  9.  $a_{n+1} = 2a_n + 3^{n-1} + 2n + 7$ 인  $a_n$ 을 구하여라. (단,  $a_1 = 0$ )
  10.  $a_{n+1} - 3a_n = 3^n - 2n - 1$ 인  $a_n$ 을 구하여라. (단,  $a_1 = 1$ )
  11.  $a_{n+1} = 2a_n + (6n+5)(-1)^{n+1}$ ,  $a_1 = 0$ 인  $a_n$ 을 구하여라.
  12.  $a_{n+1} = 2a_n - \frac{2n^2 + 7n + 4}{n(n+1)(n+2)} + 2^n$ ,  $a_1 = 2$ 인  $a_n$ 을 구하여라.
  13.  $a_1 = 1$ ,  $2a_{n+1} = a_n + (-1)^{n+1}$ , ( $n \geq 1$ )이 성립한다고 하자.
    - (1)  $a_n$ 을 구하여라.
    - (2)  $\lim_{n \rightarrow \infty} \sum_{k=1}^{2n} a_k$ 를 구하여라.
    - (3)  $\lim_{n \rightarrow \infty} \sum_{k=1}^{2n+1} a_k$ 를 구하여라.

**정리 3.**  $a_{n+1} = r(n)a_n$ 이면  $a_n = \frac{S(1)}{S(n)} a_1 a^{n-1}$ 이다.  
 단,  $r(n) = \frac{\alpha S(n)}{S(n+1)}$

**증명**  $a_{n+1} = r(n)a_n = \frac{\alpha S(n)}{S(n+1)} a_n$ 의 양변에  $S(n+1)$ 을 곱하면 다음 식이 성립한다.  
 $a_{n+1}S(n+1) = \alpha a_n S(n)$   
 이 식에서  $a_n S(n) = b_n$ 이라면 다음 식이 성립한다. .....①  
 $b_{n+1} = \alpha b_n$   
 $\therefore b_n = \alpha^{n-1} b_1 = \alpha^{n-1} a_1 S(1)$   
 ①에서  $a_n = \frac{b_n}{S(n)} = \frac{S(1)}{S(n)} a_1 \alpha^{n-1}$  Q.E.D

※ 유의 :  $a_n$ 의 계수가 수가 아니고 식일 경우, 위와 같은 방법을 사용한다.

**보기 1.**  $a_1=1$ ,  $a_n=\left(1-\frac{1}{n^2}\right)a_{n-1}$ 인  $a_n$ 을 구하라. (단,  $n \geq 2$ )

**풀이**  $a_n=\left(1-\frac{1}{n^2}\right)a_{n-1}=\frac{n^2-1}{n^2}a_{n-1}=\frac{(n-1)(n+1)}{n \cdot n}a_{n-1}$

$\frac{n}{n+1}a_n=\frac{n-1}{n}a_{n-1}$ 에서  $\frac{n}{n+1}a_n=b_n$ 이라면,

$b_1=\frac{1}{2}$ 이고 다음 식이 성립한다. .....①

$b_n=b_{n-1}$

$\therefore b_n=b_1=\frac{1}{2}$  .....②

①, ② :  $a_n=\frac{n+1}{n}$   $b_n=\frac{n+1}{2n}$

**보기 2.**  $a_1=1$ ,  $na_{n+1}=(n+1)a_n-2$ 인  $a_n$ 을 구하여라.

**풀이**  $na_{n+1}=(n+1)a_n-2$ 의 양변을  $n(n+1)$ 로 나누면 다음과 같다.

$\frac{a_{n+1}}{n+1}=\frac{a_n}{n}-\frac{2}{n(n+1)}=\frac{a_n}{n}-2\left(\frac{1}{n}-\frac{1}{n+1}\right)$

$\frac{a_{n+1}-2}{n+1}=\frac{a_n-2}{n}$ 에서  $\frac{a_n-2}{n}=b_n$ 이라면, .....①

$b_1=-1$ 이고 다음 식이 성립한다.

$b_{n+1}=b_n$

$\therefore b_n=b_1=-1$

①에서  $a_n=2+nb_n=2-n$

**보기 3.**  $a_{n+1}=(n+1)a_n+n$ ,  $a_1=1$ 인  $a_n$ 을 구하여라.

**풀이**  $a_{n+1}=(n+1)a_n+n$ 의 양변을  $(n+1)!$ 로 나누면 다음과 같다.

$\frac{a_{n+1}}{(n+1)!}=\frac{n+1}{(n+1)!}a_n+\frac{(n+1)-1}{(n+1)!}=\frac{a_n}{n!}+\frac{1}{n!}-\frac{1}{(n+1)!}$

$\frac{a_{n+1}+1}{(n+1)!}=\frac{a_n+1}{n!}$ 에서  $\frac{a_n+1}{n!}=b_n$ 이라면, .....①

$b_1=2$ 이고 다음 식이 성립한다.

$b_{n+1}=b_n$

$\therefore b_n=b_1=2$

①에서  $a_n=n! \cdot b_n-1=2 \cdot n! -1$

### 연습문제 III

1.  $a_1=1$ ,  $a_{n+1}=2^n a_n$ 인  $a_n$ 을 구하여라.

2.  $a_1=0$ ,  $n^2 a_n = (n-1)^2 a_{n-1} + n(n-1)$ 인  $a_n$ 을 구하여라. (단  $n \geq 2$ )
3.  $a_1=3$ ,  $a_{n+1} = \frac{-n(2n+3)}{(n+1)(2n+1)} a_n + \frac{2n+3}{n+1} (-1)^n$ 인  $a_n$ 을 구하여라.
4.  $a_1=2$ ,  $(n+1)a_{n+1} = 2a_n + n!$  ( $n^2 + 2n - 1$ )인  $a_n$ 을 구하여라.
5.  $a_1 = \frac{3}{2}$ ,  $a_{n+1} = \frac{-n(2n+3)}{(n+2)(2n+1)} a_n$ 인  $a_n$ ,  $\sum_{k=1}^n a_k$ 를 구하여라.
6.  $a_2=9$ ,  $(n-1)a_{n+1} = (n+1)a_n - 2$ 인  $a_n$ 을 구하여라. (단  $n \geq 2$ )
7.  $a_1=1$ 인 수열  $\{a_n\}$ 에서  $S_n = \sum_{k=1}^n a_k = n^2 a_n$ 이다. 이 때  $a_1$ ,  $S_n$ ,  $\lim_{n \rightarrow \infty} S_n$ 을 구하여라.

## A.3 3항간의 점화식(선형점화식)

3항간의 점화식  $a_{n+2} - pa_{n+1} + qa_n = f(n)$ 에서 일반항  $a_n$ 을 구하려면,  $p = \alpha + \beta$ ,  $q = \alpha\beta$ 인  $\alpha, \beta$ 를 구하여 원식을 다음과 같이 고친다.

$$a_{n+2} - (\alpha + \beta)a_{n+1} + \alpha\beta a_n = f(n)$$

이 식은 다음과 같다.

$$(a_{n+2} - \alpha a_{n+1}) - \beta(a_{n+1} - \alpha a_n) = f(n)$$

이 식에서  $a_{n+1} - \alpha a_n = b_n$ 이라 놓자.

.....①

따라서  $b_{n+1} - \beta b_n = f(n)$ 이 되어 2항간의 점화식으로 귀착되므로 제 A.2 절의 방법에 의해  $b_n$ 을 구할 수 있다. 또한 ①에서,  $b_n$ 이 구해지면 ①식 역시 2항간의 점화식이 되어  $a_n$ 을 구할 수 있다. 결국 3항간의 점화식은  $\alpha, \beta$ 를 구하는 과정과 2단계의 2항간의 점화식으로 풀 수 있다. 이 때  $\alpha + \beta = p$ ,  $\alpha\beta = q$ 라 했으므로,  $t^2 - pt + q = 0$ 의 2근이  $\alpha, \beta$ 이므로, 근의 공식에 의해  $\alpha, \beta$ 를 구할 수 있다.

**보기 1.**  $a_1 = a$ ,  $a_2 = b$ ,  $a_{n+2} = pa_{n+1} + qa_n$  (단,  $p + q = 1$ )인  $a_n$ 을 구하여라.

**풀이**  $a_{n+2} - pa_{n+1} - qa_n = a_{n+2} + (q-1)a_{n+1} - qa_n = 0$

$(a_{n+2} - a_{n+1}) + q(a_{n+1} - a_n) = 0$ 에서  $a_{n+1} - a_n = b_n$ 이라면 다음 식이 성립한다.

$$b_{n+1} = -qb_n \quad \text{.....①}$$

$$\therefore b_n = (-q)^{n-1} b_1 = (-q)^{n-1} (a_2 - a_1) = (-q)^{n-1} (b - a) \quad \text{.....②}$$

①, ②에서  $a_{n+1} - a_n = (-q)^{n-1} (b - a)$ 이므로 다음 식이 성립한다.

$$a_n = a + \sum_{k=1}^{n-1} (b-a)(-q)^{k-1} = \frac{b+aq + (a-b)(-q)^{n-1}}{1+q}$$

❖ 참고 : 원식  $a_{n+2} + qa_{n+1} = a_{n+1} + qa_n$ 에서  $a_{n+1} + qa_n = b_n$ 이라 놓으면 다음 식이 성립한다.

$$b_{n+1} = b_n$$

$$\therefore b_n = b_1 = a_2 + qa_1 = b + aq$$

따라서  $a_{n+1} + qa_n = b + aq$ 로 해서 구할 수 있다.

《유제》  $a_1 = 1$ ,  $a_2 = 2$ ,  $3a_{n+2} = 4a_{n+1} - a_n$ 인  $a_n$ 을 구하여라.  $\text{답 } a_n = \frac{1}{2} \left\{ 5 - \left( \frac{1}{3} \right)^{n-2} \right\}$

**보기 2.**  $a_1 = 1$ ,  $a_2 = 1$ ,  $a_{n+2} = a_{n+1} + a_n$ 에서  $a_n$ 을 구하여라.

**풀이** 방정식  $t^2 - t - 1 = 0$ 의 2근이  $\alpha = \frac{1-\sqrt{5}}{2}$ ,  $\beta = \frac{1+\sqrt{5}}{2}$ 이므로 원식은 다음과 같이 쓸 수 있다.

$$a_{n+2} - (\alpha + \beta)a_{n+1} + \alpha\beta a_n = 0$$

이 식  $(a_{n+2} - \alpha a_{n+1}) = \beta(a_{n+1} - \alpha a_n)$ 에서  $a_{n+1} - \alpha a_n = b_n$ 이라면 다음 식이 성립한다.

$$b_{n+1} = \beta b_n$$

$$\therefore b_n = \beta^{n-1} b_1 = \beta^{n-1} (a_2 - \alpha a_1) = \beta^n$$

$$\therefore a_{n+1} - \alpha a_n = \beta^n \quad \dots\dots\dots \textcircled{1}$$

이 때  $\alpha$ 와  $\beta$ 는 교환해도 상관없으므로 다음 식이 성립한다.

$$a_{n+1} - \beta a_n = \alpha^n \quad \dots\dots\dots \textcircled{2}$$

$$\textcircled{1} - \textcircled{2} : (\beta - \alpha) a_n = \beta^n - \alpha^n$$

$$\therefore a_n = \frac{\beta^n - \alpha^n}{\beta - \alpha} = \frac{1}{\sqrt{5}} \left\{ \left( \frac{1 + \sqrt{5}}{2} \right)^n - \left( \frac{1 - \sqrt{5}}{2} \right)^n \right\}$$

《유제》  $a_{n+2} = a_{n+1} + 2a_n$ ,  $a_1 = 1$ ,  $a_2 = 0$ 인  $a_n$ 을 구하여라.  $\textcircled{\square} a_n = \frac{2}{3} \{ 2^{n-2} + (-1)^{n-1} \}$

**보기 3.**  $a_{n+2} + 2a_{n+1} + a_n = 0$ ,  $a_1 = 1$ ,  $a_2 = 0$ 인  $a_n$ 을 구하여라.

**풀이** 방정식  $t^2 + 2t + 1 = 0$ 에서  $t = -1$ (중근)이므로, 원식  $(a_{n+2} + a_{n+1}) = -(a_{n+1} + a_n)$ 에서  $a_{n+1} + a_n = b_n$ 이라면 다음 식이 성립한다.  $\dots\dots\dots \textcircled{1}$

$$b_{n+1} = -b_n$$

$$\therefore b_n = (-1)^{n-1} b_1 (-1)^{n-1} (a_2 + a_1) = (-1)^{n+1} \quad \dots\dots\dots \textcircled{2}$$

$\textcircled{1}$ ,  $\textcircled{2}$ 에서  $a_{n+1} + a_n = (-1)^{n+1}$ 의 양변을  $(-1)^{n+1}$ 로 나누면 다음 식이 성립한다.

$$\frac{a_{n+1}}{(-1)^{n+1}} - \frac{a_n}{(-1)^n} = 1$$

이 식에서  $\frac{a_n}{(-1)^n} = c_n$ 이라면 다음 식이 성립한다.  $\dots\dots\dots \textcircled{3}$

$$c_{n+1} = c_n + 1$$

$$\therefore c_n = c_1 + (n-1) = n-2, \quad (\because c_1 = -a_1 = -1) \quad \dots\dots\dots \textcircled{4}$$

$$\textcircled{3}, \textcircled{4} : a_n = (-1)^n (n-2)$$

《유제》  $a_{n+2} = 4a_{n+1} - 4a_n$ ,  $a_1 = 1$ ,  $a_2 = 4$ 인  $a_n$ 을 구하여라.  $\textcircled{\square} a_n = n2^{n-1}$

**보기 4.**  $a_{n+2} = 4a_{n+1} - 3a_n + 2$ ,  $a_1 = 1$ ,  $a_2 = -1$ 인  $a_n$ 을 구하여라.

**풀이** 방정식  $t^2 - 4t + 3 = 0$ 에서  $t = 1$ ,  $t = 3$ 이므로, 원식  $(a_{n+2} - a_{n+1}) - 3(a_{n+1} - a_n) = 2$ 에서  $a_{n+1} - a_n = b_n$ 이라면 다음 식이 성립한다.  $\dots\dots\dots \textcircled{1}$

$$b_{n+1} = 3b_n + 2, \quad (b_1 = a_2 - a_1 = -2) \quad \dots\dots\dots \textcircled{2}$$

$$x = 3x + 2, \quad (x = -1) \quad \dots\dots\dots \textcircled{3}$$

$\textcircled{2} - \textcircled{3}$ 에서  $b_n - x = c_n$ 이라면,  $c_1 = b_1 - x = -1$ 이고 다음 식이 성립한다.  $\dots\dots\dots \textcircled{4}$

$$c_{n+1} = 3c_n$$

$$\therefore c_n = 3^{n-1} c_1 = -3^{n-1} \quad \dots\dots\dots \textcircled{5}$$

$$\textcircled{1}, \textcircled{4}, \textcircled{5} : a_{n+1} - a_n = b_n = c_n + x = -(3^{n-1} + 1)$$

$$\therefore a_n = 1 + \sum_{k=1}^{n-1} \{ -(3^{k-1} + 1) \} = \frac{1}{2} (5 - 2n - 3^{n-1})$$

《유제》  $a_{n+2} + a_{n+1} - 2a_n = 2$ ,  $a_1 = 0$ ,  $a_2 = 2$ 인  $a_n$ 을 구하여라.  $\textcircled{\square} \frac{1}{9} \{ 6n - 2 - (-2)^{n+1} \}$

**보기 5.**  $a_{n+2}-a_{n+1}-2a_n=2$ ,  $a_1=0$ ,  $a_2=2$ 인  $a_n$ 을 구하여라.

**풀이** 방정식  $t^2-t-2=0$ 에서  $t=2$ ,  $t=-1$ 이므로, 원식  $(a_{n+2}+a_{n+1})=2(a_{n+1}+a_n)+2$ 에서  $a_{n+1}+a_n=b_n$ 이라면 다음 식이 성립한다. ....①

$b_{n+1}=2b_n+2$ , ( $b_1=a_2+a_1=2$ ) .....②

$x=2x+2$ , ( $x=-2$ ) .....③

②-③에서  $b_n-x=c_n$ 이라면, ( $c_1=b_1-x=4$ ) 다음 식이 성립한다. ....④

$c_{n+1}=2c_n$

$\therefore c_n=2^{n-1}c_1=2^{n+1}$  .....⑤

①, ④, ⑤ :  $a_{n+1}+a_n=b_n=c_n+x=2^{n+1}-2$  .....⑥

$(\alpha \cdot 2^{n+1} + \beta) + (\alpha \cdot 2^n + \beta) = 2^{n+1} - 2$ , ( $\alpha = \frac{2}{3}$ ,  $\beta = -1$ ) .....⑦

⑥-⑦에서  $a_n - (\alpha \cdot 2^n + \beta) = d_n$ 이라면,

$d_1 = a_1 - 2\alpha - \beta = -\frac{1}{3}$ 이고 다음 식이 성립한다. ....⑧

$d_{n+1} = -d_n$

$\therefore d_n = (-1)^{n-1}d_1 = \frac{1}{3}(-1)^n$  .....⑨

⑧, ⑨ :  $a_n = \alpha \cdot 2^n + \beta + d_n = \frac{1}{3}\{2^{n+1} + (-1)^n - 3\}$

《유제》  $a_{n+2}-5a_{n+1}+6a_n=7$ ,  $a_2=3$ ,  $a_1=1$ 인  $a_n$ 을 구하여라.

☞  $a_n = \frac{1}{2}\{3^{n+1} - 7 \cdot 2^n + 7\}$

**보기 6.**  $a_{n+2}=4a_{n+1}-4a_n-4$ ,  $a_1=0$ ,  $a_2=8$ 인  $a_n$ 을 구하여라.

**풀이** 방정식  $t^2-4t+4=0$ 에서  $t=2$ (중근)이므로, 원식  $(a_{n+2}-2a_{n+1})=2(a_{n+1}-2a_n)-4$ 에서  $a_{n+1}-2a_n=b_n$ 이라면 다음 식이 성립한다. ....①

$b_{n+1}=2b_n-4$ , ( $b_1=a_2-2a_1=8$ ) .....②

$x=2x-4$ , ( $x=4$ ) .....③

②-③에서  $b_n-x=c_n$ 이라면,  $c_1=b_1-x=4$ 이고 다음 식이 성립한다. ....④

$c_{n+1}=2c_n$

$\therefore c_n=2^{n-1}c_1=2^{n+1}$  .....⑤

①, ④, ⑤ :  $a_{n+1}-2a_n=b_n=c_n+x=2^{n+1}+4$ 의 양변을  $2^{n+1}$ 로 나누면

$\frac{a_{n+1}}{2^{n+1}} = \frac{a_n}{2^n} + 1 + \frac{1}{2^{n-1}}$ 에서  $\frac{a_n}{2^n} = d_n$ 이라면,

$d_1 = \frac{a_1}{2} = 0$ 이고 다음 식이 성립한다. ....⑥

$d_{n+1} = d_n + 1 + 2^{-(n-1)}$

$\therefore d_n = 0 + \sum_{k=1}^{n-1} \{2^{-(k-1)} + 1\} = (n+1) - 2^{-(n-2)}$  .....⑦

⑥, ⑦ :  $a_n = 2^n d_n = (n+1)2^n - 4$

《유제》  $4a_{n+2}+4a_{n+1}+a_n=1$ ,  $a_1=1$ ,  $a_2=0$ 인  $a_n$ 을 구하여라.

$$\text{답 } a_n = \frac{1}{9} \left\{ (3n-7) \left( -\frac{1}{2} \right)^{n-2} + 1 \right\}$$

**보기 7.**  $a_{n+2} - 4a_{n+1} + 3a_n = 2n + 1$ ,  $a_1 = -1$ ,  $a_2 = 0$ 인  $a_n$ 을 구하여라.

- 풀이** 방정식  $t^2 - 4t + 3 = 0$ 에서  $t=1$ ,  $t=3$ 이므로, 원식  $a_{n+2} - a_{n+1} = 3(a_{n+1} - a_n) + 2n + 1$ 에서  $a_{n+1} - a_n = b_n$ 이라면 다음 식이 성립한다. .....①  
 $b_{n+1} = 3b_n + 2n + 1$ , ( $b_1 = a_2 - a_1 = 1$ ) .....②  
 $\{p(n+1) + q\} = 3\{pn + q\} + 2n + 1$ , ( $p = -1$ ,  $q = -1$ ) .....③  
 ②-③에서  $b_n = (pn + q) = c_n$ 이라면,  
 $c_1 = b_1 - p - q = 3$ 이고 다음 식이 성립한다. ....④  
 $c_{n+1} = 3c_n$   
 $\therefore c_n = 3^{n-1} c_1 = 3^n$  .....⑤  
 ①, ④, ⑤ :  $a_{n+1} - a_n = b_n = c_n + (pn + q) = 3^n - n - 1$   
 $\therefore a_n = -1 + \sum_{k=1}^{n-1} (3^k - k - 1) = \frac{1}{2} \{3^n - n(n+1) - 3\}$

《유제》  $a_{n+2} - a_n = 2n + 7$ ,  $a_1 = 0$ ,  $a_2 = 2$ 인  $a_n$ 을 구하여라.

$$\text{답 } a_n = \frac{1}{2} (n^2 + 5n - 8) + (-1)^{n-1}$$

**보기 8.**  $2a_{n+2} = 7a_{n+1} - 3a_n + 4n + 8$ ,  $a_1 = -1$ ,  $a_2 = 2$ 인  $a_n$ 을 구하여라.

- 풀이** 방정식  $t^2 - \frac{7}{2}t + \frac{3}{2} = 0$ 에서  $t = \frac{1}{2}$ ,  $t = 3$ 이므로, 원식  $a_{n+2} - 3a_{n+1} = \frac{1}{2}(a_{n+1} - 3a_n) + 2n + 4$ 에서  $a_{n+1} - 3a_n = b_n$ 이라면 다음 식이 성립한다. ....①  
 $b_{n+1} - \frac{1}{2}b_n = 2n + 4$ , ( $b_1 = a_2 - 3a_1 = 5$ ) .....②  
 $x(n+1) + y - \frac{1}{2}(xn + y) = 2n + 4$ , ( $x = 4$ ,  $y = 0$ ) .....③  
 ②-③에서  $b_n - (xn + y) = c_n$ 이라면,  
 $c_1 = b_1 - x - y = 1$ 이고 다음 식이 성립한다. ....④  
 $c_{n+1} = \frac{1}{2}c_n$   
 $\therefore c_n = \left(\frac{1}{2}\right)^{n-1}$  .....⑤  
 ①, ④, ⑤ :  $a_{n+1} - 3a_n = b_n = c_n + (xn + y) = \left(\frac{1}{2}\right)^{n-1} + 4n$  .....⑥  
 $\left\{p\left(\frac{1}{2}\right)^n + q(n+1) + r\right\} - 3\left\{p\left(\frac{1}{2}\right)^{n-1} + qn + r\right\} = \left(\frac{1}{2}\right)^{n-1} + 4n$  .....⑦  
 항등식의 원리에 의해서  $p = -\frac{2}{5}$ ,  $q = -2$ ,  $r = -1$ 이다.  
 ⑥-⑦에서  $a_n - p\left(\frac{1}{2}\right)^{n-1} - qn - r = d_n$ 이라면,  
 $d_1 = a_1 - p - q - r = \frac{12}{5}$ 이고 다음 식이 성립한다. ....⑧



$$d_{n+1}=3d_n$$

$$\therefore d_n=3^{n-1}d_1=\frac{4}{5}\cdot 3^n \quad \dots\dots\dots\textcircled{9}$$

$$\textcircled{8}, \textcircled{9} : a_n=p\left(\frac{1}{2}\right)^{n-1}+qn+r+d_n=\frac{4}{5}\times 3^n-\frac{1}{5}\times\left(\frac{1}{2}\right)^{n-2}-2n-1$$

《유제》  $a_{n+2}=a_{n+1}+6a_n+2n+5$ ,  $a_1=1$ ,  $a_2=5$ 인  $a_n$ 을 구하여라.

$$\text{답} \quad \frac{1}{45}\{11\times 3^{n+1}+(-2)^{n-1}-15n-40\}$$

**보기 9.**  $4a_{n+2}-4a_{n+1}+a_n=4n$ ,  $a_1=2$ ,  $a_2=1$ 인  $a_n$ 을 구하여라.

**풀이** 방정식  $t^2-t+\frac{1}{4}=0$ 에서  $t=\frac{1}{2}$ (중근)이므로, 원식  $(a_{n+2}-\frac{1}{2}a_{n+1})=\frac{1}{2}(a_{n+1}-\frac{1}{2}a_n)+n$

에서  $a_{n+1}-\frac{1}{2}a_n=b_n$ 이라면 다음 식이 성립한다. .....①

$$b_{n+1}=\frac{1}{2}b_n+n, \quad (b_1=a_2-\frac{1}{2}a_1=0) \quad \dots\dots\dots\textcircled{2}$$

$$p(n+1)+q=\frac{1}{2}(pn+q)+n, \quad (p=2, q=-4) \quad \dots\dots\dots\textcircled{3}$$

②-③에서  $b_n-(pn+q)=c_n$ 이라면,  
 $c_1=b_1-p-q=2$ 이고 다음 식이 성립한다. .....④

$$c_{n+1}=\frac{1}{2}c_n$$

$$\therefore c_n=\left(\frac{1}{2}\right)^{n-1}c_1=\left(\frac{1}{2}\right)^{n-2} \quad \dots\dots\dots\textcircled{5}$$

$$\textcircled{1}, \textcircled{4}, \textcircled{5} : a_{n+1}-\frac{1}{2}a_n=b_n=c_n+(pn+q)=\left(\frac{1}{2}\right)^{n-2}+2n-4 \quad \dots\dots\dots\textcircled{6}$$

$$\{x(n+1)+y\}-\frac{1}{2}(xn+y)=2n-4, \quad (x=4, y=-16) \quad \dots\dots\dots\textcircled{7}$$

⑥-⑦에서  $a_n-(xn+y)=d_n$ 이라면,  
 $d_1=a_1-x-y=14$ 이고 다음 식이 성립한다. .....⑧

$$d_{n+1}=\frac{1}{2}d_n+\left(\frac{1}{2}\right)^{n-2}$$

이 식의 양변을  $2^n$ 배하면  $2^n d_{n+1}=2^{n-1}d_n+4$ 이다. 이 식에서  $2^{n-1}d_n=e_n$ 이라면,  
 $e_1=14$ 이고 다음 식이 성립한다. .....⑨

$$e_{n+1}=e_n+4$$

$$\therefore e_n=e_1+4\times(n-1)=4n+10 \quad \dots\dots\dots\textcircled{10}$$

$$\textcircled{8}, \textcircled{9}, \textcircled{10} : a_n=xn+y+d_n=xn+y+\frac{e_n}{2^{n-1}}=4n-16+\frac{2n+5}{2^{n-2}}$$

❖참고 : ⑥의 양변  $\times 2^{n+1}$ 을 하면 다음과 같다.

$$2^{n+1}a_{n+1}-2^n a_n=8+n2^{n+2}-2^{n+3}$$

이 식에서  $2^n a_n=b'_n$ 이라면 다음 식이 성립한다. .....⑦'

$$b'_{n+1}-b'_n=8-2^{n+3}+n\cdot 2^{n+2}, \quad (\text{단}, b'_1=2a_1=4)$$

$$\therefore b_n' = b_1' + \sum_{k=1}^{n-1} (8 - 2^{k+3} + k \cdot 2^{k+2}) = n \cdot 2^{n+2} - 2^{n+4} + 8n + 20 \quad \dots\dots\dots \textcircled{8}'$$

$$\textcircled{7}', \textcircled{8}' : a_n = \frac{b_n'}{2^n} = 4n - 16 + \frac{2n+5}{2^{n-2}}$$

《유제》  $a_{n+2} = 6a_{n+1} - 9a_n + 2n + 1$ ,  $a_1 = 0$ ,  $a_2 = 1$ 인  $a_n$ 을 구하여라.

$$\text{답} \quad a_n = \frac{1}{4} \left\{ (4n-9) \cdot 3^{n-1} + 2n + 3 \right\}$$

**보기 10.**  $a_{n+2} + 2a_{n+1} = 3a_n + 2^n$ ,  $a_1 = 1$ ,  $a_2 = 2$ 인  $a_n$ 을 구하여라.

**풀이** 방정식  $t^2 + 2t - 3 = 0$ 에서  $t = 1$ ,  $t = -3$ 이므로, 원식  $(a_{n+2} - a_{n+1}) + 3(a_{n+1} - a_n) = 2^n$ 에서  $a_{n+1} - a_n = b_n$ 이라 하면 다음 식이 성립된다. .....①

$$b_{n+1} + 3b_n = 2^n, \quad (b_1 = a_2 - a_1 = 1) \quad \dots\dots\dots \textcircled{2}$$

$$x \cdot 2^n + 3x \cdot 2^{n-1} = 2^n, \quad \left(x = \frac{2}{5}\right) \quad \dots\dots\dots \textcircled{3}$$

②-③에서  $b_n - x \cdot 2^{n-1} = c_n$ 이라 하면,

$$c_1 = b_1 - x = \frac{3}{5} \text{이고 다음 식이 성립한다.} \quad \dots\dots\dots \textcircled{4}$$

$$c_{n+1} = -3c_n$$

$$\therefore c_n = (-3)^{n-1} c_1 = \frac{3}{5} (-3)^{n-1} \quad \dots\dots\dots \textcircled{5}$$

$$\textcircled{1}, \textcircled{4}, \textcircled{5} : a_{n+1} - a_n = b_n = c_n + x \cdot 2^{n-1} = \frac{1}{5} \left\{ 2^n - (-3)^n \right\}$$

$$\therefore a_n = 1 + \sum_{k=1}^{n-1} \frac{1}{5} \left\{ 2^k - (-3)^k \right\} = \frac{1}{20} \left\{ 2^{n+2} + (-3)^n + 15 \right\}$$

《유제》  $a_{n+2} - a_n = 2^n$ ,  $a_1 = 0$ ,  $a_2 = 1$ 인  $a_n$ 을 구하여라. \text{답} \quad a\_n = \frac{1}{6} \left\{ 2^{n+1} + (-1)^n - 3 \right\}

**보기 11.**  $a_{n+2} - 2a_{n+1} - 3a_n = 2^n$ ,  $a_1 = 0$ ,  $a_2 = 1$ 인  $a_n$ 을 구하여라.

**풀이** 방정식  $t^2 - 2t - 3 = 0$ 에서  $t = -1$ ,  $t = 3$ 이므로, 원식  $a_{n+2} + a_{n+1} = 3(a_{n+1} + a_n) + 2^n$ 에서  $a_{n+1} + a_n = b_n$ 이라 하면 다음 식이 성립한다. .....①

$$b_{n+1} = 3b_n + 2^n, \quad (b_1 = a_2 + a_1 = 1) \quad \dots\dots\dots \textcircled{2}$$

$$x \cdot 2^n = 3x2^{n-1} + 2^n, \quad (x = -2) \quad \dots\dots\dots \textcircled{3}$$

②-③에서  $b_n - x \cdot 2^{n-1} = c_n$ 이라 하면, .....④

$$c_1 = b_1 - x = 3 \text{이고 다음 식이 성립한다.}$$

$$c_{n+1} = 3c_n$$

$$\therefore c_n = 3^{n-1} c_1 = 3^n \quad \dots\dots\dots \textcircled{5}$$

$$\textcircled{1}, \textcircled{4}, \textcircled{5} : a_{n+1} + a_n = b_n = c_n + x \cdot 2^{n-1} = 3^n - 2^n \quad \dots\dots\dots \textcircled{6}$$

$$(p \cdot 3^n + q \cdot 2^n) + (p \cdot 3^{n-1} + q \cdot 2^{n-1}) = 3^n - 2^n, \quad \left(p = \frac{3}{4}, \quad q = -\frac{2}{3}\right) \quad \dots\dots\dots \textcircled{7}$$

⑥-⑦에서  $a_n - p \cdot 3^{n-1} - q \cdot 2^{n-1} = d_n$ 이라 하면,

$$d_1 = a_1 - p - q = -\frac{1}{12} \text{이고 다음 식이 성립한다.} \quad \dots\dots\dots \textcircled{8}$$

$$d_{n+1} = -d_n$$

$$\therefore d_n = (-1)^{n-1} d_1 = \frac{1}{12} (-1)^n \quad \dots\dots\dots ⑨$$

$$⑧, ⑨ : a_n = p \cdot 3^{n-1} + q \cdot 2^{n-1} + d_n = \frac{1}{12} \{ 3^{n+1} - 2^{n+2} + (-1)^n \}$$

《유제》  $a_{n+2} - 5a_{n+1} + 6a_n = (-1)^{n-1}$ ,  $a_1 = -1$ ,  $a_2 = 0$ 인  $a_n$ 을 구하여라.

$$\text{답} \quad a_n = \frac{1}{12} \{ 3^{n+2} - 5 \cdot 2^{n+2} - (-1)^n \}$$

**보기 12.**  $a_{n+2} = 6a_{n+1} - 9a_n + 2^n$ ,  $a_1 = 0$ ,  $a_2 = 1$ 인  $a_n$ 을 구하여라.

**풀이** 방정식  $t^2 - 6t + 9 = 0$ 에서  $t = 3$ (중근)이므로, 원식  $(a_{n+2} - 3a_{n+1}) = 3(a_{n+1} - 3a_n) + 2^n$ 에서  $a_{n+1} - 3a_n = b_n$ 이라하면 다음 식이 성립한다. .....①

$$b_{n+1} = 3b_n + 2^n, \quad (b_1 = a_2 - 3a_1 = 1) \quad \dots\dots\dots ②$$

$$x \cdot 2^{n+1} = 3 \cdot x \cdot 2^n + 2^n, \quad (x = -1) \quad \dots\dots\dots ③$$

②-③에서  $b_n - x \cdot 2^n = c_n$ 이라하면,

$$c_1 = b_1 - 2x = 3 \text{이고 다음 식이 성립한다.} \quad \dots\dots\dots ④$$

$$c_{n+1} = 3c_n$$

$$\therefore c_n = 3^{n-1} c_1 = 3^n \quad \dots\dots\dots ⑤$$

$$①, ④, ⑤ : a_{n+1} - 3a_n = b_n = c_n + x \cdot 2^n = 3^n - 2^n \quad \dots\dots\dots ⑥$$

$$y \cdot 2^{n+1} - 3y \cdot 2^n = -2^n, \quad (y = 1) \quad \dots\dots\dots ⑦$$

⑥-⑦ :  $a_{n+1} - y \cdot 2^{n+1} = 3(a_n - y \cdot 2^n) + 3^n$ 의 양변을  $3^n$ 으로 나누면 다음 식이 성립한다.

$$\frac{a_{n+1} - y \cdot 2^{n+1}}{3^n} = \frac{a_n - y \cdot 2^n}{3^{n-1}} + 1$$

이 식에서  $\frac{a_n - y \cdot 2^n}{3^{n-1}} = d_n$ 이라하면 다음 식이 성립한다. .....⑧

$$d_{n+1} = d_n + 1$$

$$\therefore d_n = d_1 + (n-1) = n-3, \quad (\because d_1 = a_1 - 2y = -2) \quad \dots\dots\dots ⑨$$

$$⑧, ⑨ : a_n = y \cdot 2^n + 3^{n-1} d_n = 2^n + (n-3) \cdot 3^{n-1}$$

❖ 참고 : ⑥의 양변을  $3^n$ 으로 나누면 다음 식이 성립한다.

$$\frac{a_{n+1}}{3^n} = \frac{a_n}{3^{n-1}} + 1 - \left(\frac{2}{3}\right)^n$$

이 식에서  $\frac{a_n}{3^{n-1}} = b'_n$ 라하면,  $b'_1 = a_1 = 0$ 이고 다음 식이 성립한다. .....⑦'

$$b'_{n+1} = b'_n + 1 - \left(\frac{2}{3}\right)^n$$

$$\therefore b'_n = b'_1 + \sum_{k=1}^{n-1} \left\{ 1 - \left(\frac{2}{3}\right)^k \right\} = n-3 + 2 \left(\frac{2}{3}\right)^{n-1} \quad \dots\dots\dots ⑧'$$

$$⑦', ⑧' : a_n = 3^{n-1} b'_n = (n-3) \cdot 3^{n-1} + 2^n$$

《유제》  $a_{n+2} + a_{n+1} - 6a_n = 2^n$ ,  $a_1 = 1$ ,  $a_2 = 3$ 인  $a_n$ 을 구하여라.

$$\text{답} \quad \frac{1}{25} \{ 2^{n-1} (5n+23) + (-3)^n \}$$

**보기 13.**  $a_{n+2}=3a_{n+1}-2a_n+3^n(n+1)$ ,  $a_1=0$ ,  $a_2=1$ 인  $a_n$ 을 구하여라.

**풀이** 방정식  $t^2-3t+2=0$ 에서  $t=1$ ,  $t=2$ 이므로, 원식  $(a_{n+2}-a_{n+1})=2(a_{n+1}-a_n)+3^n(n+1)$

에서  $a_{n+1}-a_n=b_n$ 이라면 다음 식이 성립한다. .....①

$$b_{n+1}=2b_n+3^n(n+1), (b_1=a_2-a_1=1) \quad \dots\dots\dots②$$

$$3^{n+1}\{x(n+1)+y\}=2 \cdot 3^n(xn+y)+3^n(n+1), (x=1, y=-2) \quad \dots\dots\dots③$$

②-③에서  $b_n-3^n(xn+y)=c_n$ 이라면,  
 $c_1=b_1-3x-3y=4$ 이고 다음 식이 성립한다. ....④

$$c_{n+1}=2c_n$$

$$\therefore c_n=2^{n-1}c_1=2^{n+1} \quad \dots\dots\dots⑤$$

$$①, ④, ⑤ : a_{n+1}-a_n=b_n=c_n+3^n(xn+y)=2^{n+1}+3^n(n-2)$$

따라서 다음 식이 성립한다.

$$a_n=0+\sum_{k=1}^{n-1}\left\{2^{k+1}+3^k(k-2)\right\}=2^{n+1}+\frac{1}{4} \cdot 3^n(2n-7)-\frac{1}{4}$$

$$\begin{aligned} \text{❖ 참고 : } \sum_{k=1}^{n-1}3^k(k-2) &= \sum_{k=1}^{n-1}\left\{\frac{1}{4} \cdot 3^{k+1}(2k+1-7)-\frac{1}{4} \cdot 3^k(2k-7)\right\} \\ &= \sum_{k=1}^n\left\{\frac{1}{4} \cdot 3^k(2k-7)\right\}-\sum_{k=1}^{n-1}\left\{\frac{1}{4} \cdot 3^k(2k-7)\right\} \\ &= \frac{1}{4}\left\{3^n(2n-7)-3(2-7)\right\} \end{aligned}$$

《유제》  $a_{n+2}+a_{n+1}-2a_n=(-1)^n(n+1)$ ,  $a_1=1$ ,  $a_2=2$ 인  $a_n$ 을 구하여라.

$$\text{답 } a_n=\frac{1}{12}\left\{13-(-2)^{n+3}+(-1)^{n-1}(6n+9)\right\}$$

**보기 14.**  $a_{n+2}-5a_{n+1}-6a_n=3^{n+1}(2n+1)$ ,  $a_1=0$ ,  $a_2=1$ 인  $a_n$ 을 구하여라.

**풀이** 방정식  $t^2-5t-6=0$ 에서  $t=6$ ,  $t=-1$ 이므로, 원식  $(a_{n+2}+a_{n+1})=6(a_{n+1}+a_n)+3^{n+1}$

$(2n+1)$ 에서  $a_{n+1}+a_n=b_n$ 이라면 다음 식이 성립한다. ....①

$$b_{n+1}=6b_n+3^{n+1}(2n+1), (b_1=a_2+a_1=1) \quad \dots\dots\dots②$$

$$3^{n+1}(xn+1+y)=6 \cdot 3^n(xn+y)+3^{n+1}(2n+1), (x=-2, y=-3) \quad \dots\dots\dots③$$

②-③에서  $b_n-3^n(xn+y)=c_n$ 이라면,  
 $c_1=b_1-3x-3y=16$ 이고 다음 식이 성립한다. ....④

$$c_{n+1}=6c_n$$

$$\therefore c_n=6^{n-1}c_1=16 \times 6^{n-1} \quad \dots\dots\dots⑤$$

$$①, ④, ⑤ : a_{n+1}+a_n=b_n=c_n+3^n(xn+y)=16 \times 6^{n-1}-3^n(2n+3) \quad \dots\dots\dots⑥$$

이 때  $d_n=p \cdot 6^n-3^n(qn+r)$  중에서 다음을 만족하는 것을 구해보자. ....⑦

$$d_{n+1}+d_n=16 \times 6^{n-1}-3^n(2n+3) \quad \dots\dots\dots⑧$$

⑦, ⑧에서  $p=\frac{8}{21}$ ,  $q=\frac{1}{2}$ ,  $r=\frac{3}{8}$ 이다.

$$\therefore d_1=-\frac{19}{56}$$

⑥-⑧에서  $a_n - d_n = e_n$ 이라면,  $e_1 = a_1 - d_1 = \frac{19}{56}$ 이고 다음 식이 성립한다. ……⑨

$$e_{n+1} = -e_n$$

$$\therefore e_n = (-1)^{n-1} e_1 = \frac{19}{56} (-1)^{n-1} \quad \dots\dots\dots ⑩$$

$$\text{⑦, ⑨, ⑩} : a_n = d_n + e_n = \frac{8}{21} \times 6^n - 3^n \left( \frac{n}{2} + \frac{3}{8} \right) + \frac{19}{56} \times (-1)^{n-1}$$

《유제》  $a_{n+2} + 5a_{n+1} + 4a_n = 2^{n+1}(3n+4)$ ,  $a_1 = a_2 = 1$ 인  $a_n$ 을 구하여라.

$$\text{답} \quad a_n = \frac{1}{9} \{ 5(-1)^n + 6(-4)^{n-1} + (3n+1)2^n \}$$

**㉮기 15.**  $a_{n+2} - 4a_n = 2^{n+2}(n+1)$ ,  $a_1 = 0$ ,  $a_2 = 1$ 인  $a_n$ 을 구하여라.

**풀이** 방정식  $t^2 - 4 = 0$ 에서  $t = 2$ ,  $t = -2$ 이므로, 원식  $a_{n+2} + 2a_{n+1} = 2(a_{n+1} + 2a_n) + 2^{n+2}(n+1)$ 에서  $a_{n+1} + 2a_n = b_n$ 이라면 다음 식이 성립한다. ……①

$b_{n+1} = 2b_n + 2^{n+2}(n+1)$ 의 양변을  $2^{n+1}$ 로 나누면 다음 식이 성립한다.

$$\frac{b_{n+1}}{2^{n+1}} = \frac{b_n}{2^n} + 2n + 2$$

이 식에서  $\frac{b_n}{2^n} = c_n$ 이라면,  $c_1 = \frac{b_1}{2} = \frac{a_2 + 2a_1}{2} = \frac{1}{2}$ 이고 다음 식이 성립한다. ……②

$$c_{n+1} = c_n + 2n + 2$$

$$\therefore c_n = \frac{1}{2} + \sum_{k=1}^{n-1} 2(k+1) = \frac{1}{2} (2n^2 + 2n - 3) \quad \dots\dots\dots ③$$

$$\text{①, ②, ③} : a_{n+1} + 2a_n = b_n = 2^{n-1} (2n^2 + 2n - 3) \quad \dots\dots\dots ④$$

이 때  $d_n = 2^n (pn^2 + qn + r)$  중에서 다음을 만족하는 것을 구해보자. ……⑤

$$d_{n+1} + 2d_n = 2^{n-1} (2n^2 + 2n - 3) \quad \dots\dots\dots ⑥$$

$$\text{⑤, ⑥} : p = \frac{1}{4}, \quad q = 0, \quad r = -\frac{1}{2}$$

$$\therefore d_1 = 2(p + q + r) = -\frac{1}{2} \quad \dots\dots\dots ⑦$$

④-⑤에서  $a_n - d_n = e_n$ 이라면,  $e_1 = a_1 - d_1 = \frac{1}{2}$ 이고 다음 식이 성립한다. ……⑧

$$e_{n+1} = -2e_n$$

$$\therefore e_n = (-2)^{n-1} e_1 = -(-2)^{n-2} \quad \dots\dots\dots ⑨$$

$$\text{⑧, ⑨} : a_n = e_n + d_n = 2^{n-2} (n^2 - 2) - (-2)^{n-2}$$

《유제》  $a_{n+2} - 4a_{n+1} + 4a_n = 2^{n+1}n$ ,  $a_1 = 0$ ,  $a_2 = 0$ 인  $a_n$ 을 구하여라.

$$\text{답} \quad a_n = 2^{n-1} \frac{n(n-1)(n-2)}{6}$$

**㉮기 16.**  $n^2 a_n = (n-2)^2 a_{n-2} + 2n - 1$ ,  $a_1 = 0$ ,  $a_2 = 1$ 인  $a_n$ 을 구하여라. ( $n \geq 3$ )

**풀이** 우선  $n^2 a_n = b_n$ 이라면,  $b_{n+2} = b_n + 2(n+2) - 1 = b_n + 2n + 3$ , ( $b_1 = 0$ ,  $b_2 = 4$ )이다. ……①

방정식  $t^2 - 1 = 0$ 에서  $t = 1$ ,  $t = -1$ 이므로,



수열  $\{a_n\}$ 이 있다.

(1)  $a_n$ 을 구하여라.

(2)  $\sum_{k=0}^{\infty} a_k$ 을 구하여라.

5. 수열  $\{a_n\}$ 에서  $a_1=a_2=1$ ,  $a_{r+2}=a_{r+1}+a_r$ , ( $r=1, 2, \dots$ )일 때  $\lim_{n \rightarrow \infty} \frac{a_n}{a_{n-1}}$ 을 구하여라.

6.  $a_1=i$ ,  $a_2=1$ ,  $a_{n+2}=1-a_n$ 인  $a_n$ 을 구하여라.

7.  $a_0=1$ ,  $a_1=2$ ,  $a_{n+1}=3a_n-2a_{n-1}$ 인 수열  $\{a_n\}$ 이 있다.

(1)  $a_n$ 을 구하여라.

(2)  $S_n = \sum_{k=0}^n ka_k x^k$ 를 구하여라.

8.  $0 < |a| < |b|$ 일 때 다음 세 조건을 만족하는 수열의 일반항을 구하여라.

$$ax_{n-1} + bx_{n+1} = (a+b)x_n, \quad -ax_0 + bx_1 = 0, \quad \sum_{n=0}^{\infty} x_n = 1$$

9.  $\alpha + \beta + \gamma = \pi$ , ( $\alpha, \beta, \gamma > 0$ )이고  $0 < a_n < \pi$ 이다.

$$a_1 = \alpha, \quad a_2 = \frac{\beta + \gamma}{2}, \quad a_{n+2} = \frac{1}{2}(a_{n+1} + a_n)$$
인 수열  $\{a_n\}$ 이 있다.

(1)  $a_n + 2a_{n+1} = \pi$ 임을 보여라.

(2)  $(\sin a_n + \sin a_{n+1}) \cos a_{n+2} \leq \sin a_{n+1}$ 을 증명하여라.

(3) (2)의 부등식이 모든  $n$ 에 대해 등호가 성립하면  $a_n$ 은 일정함을 말하라.

10. 수열  $\{c_n\}$ 의 일반항  $c_n$ 이 공차  $d$ 인 등차수열의 제  $n$ 항과 공비  $\gamma$  ( $\gamma \neq 1$ )인 등비수열의 제  $n$ 항의 합으로 표시되기 위한 필요충분조건은  $c_{n+2} - (1+\gamma)c_{n+1} + \gamma c_n = d(1-\gamma)$ 임을 증명하라.

11. 식  $-ax_0 + \beta x_1 = 0$ ,  $-(\alpha + n\beta)x_n + \alpha x_{n-1} + (n+1)\beta x_{n+1} = 0$ 을 만족하는 수열  $\{x_n\}$ 이 있다. 일반항  $x_n$ 을  $n$ 과  $x_0$ ,  $\alpha$ ,  $\beta$ 로 나타내어라.

12.  $a_{n+2} - 9a_n = 2 - 8n - 5 \cdot 2^n$ ,  $a_1 = -3$ ,  $a_2 = 42$ 인  $a_n$ 을 구하여라.

13.  $a_{n+2} - 4a_{n+1} + 3a_n = 15 \cdot (-2)^{n-1} - 4n + 2$ ,  $a_1 = 4$ ,  $a_2 = 5$ 인  $a_n$ 을 구하여라.

14.  $a_{n+2} - 5a_{n+1} + 6a_n = 3^n + 4n + 4$ ,  $a_1 = 4$ ,  $a_2 = 10$ 인  $a_n$ 을 구하여라.

### A.4 점화방정식

점화방정식이란 서로 다른 2개 이상의 수열들의 일반항을 일정한 관계식으로 연결시켜 놓은 방정식이다.

**보기**  $a_1=1, a_{n+1}=pa_n+qb_n$   
 $b_1=0, b_{n+1}=pb_n+qa_n$

일반적으로 점화방정식도 방정식의 일종이므로, 보통의 방정식과 같이 대입법, 가감법 또는 이를 겸용하여 풀 수 있다. 하지만 특수한 경우를 제외하고는 대부분 대입법에 의해서만 풀린다.

1) 대입법

$$a_{n+1}=f(a_n, b_n) \dots\dots\dots \textcircled{1}$$

$$b_{n+1}=g(a_n, b_n) \dots\dots\dots \textcircled{2}$$

대입법이란 위와 같은 점화식의 경우, ①에서  $b_n=h(a_{n+1}, a_n)$ 으로 고친 뒤 ②에 대입하여  $h(a_{n+2}, a_{n+1})=g(a_n, h(a_{n+1}, a_n))$ 를 구한다. 이 식은 수열  $\{a_n\}$ 에 관한 식이므로 앞에서 공부한 인접 2항간 또는 인접 3항간 등의 점화식으로 귀착하여 풀 수 있다.

**보기**  $a_{n+1}=5a_n+b_n \dots\dots \textcircled{1}$   
 $b_{n+1}=3a_n-2b_n \dots\dots \textcircled{2}$   
 이 점화방정식을  $a_n$ 에만 관한 식으로 유도하여라.

**풀이** ①에서  $b_n=a_{n+1}-5a_n$ 이다. 이 식을 ②에 대입하면 다음 식을 얻는다.

$$a_{n+2}-5a_{n+1}=3a_n-2(a_{n+1}-5a_n)$$

$$\therefore a_{n+2}=3a_{n+1}+13a_n$$

2) 가감법

점화방정식  $a_{n+1}=f(a_n, b_n), b_{n+1}=g(a_n, b_n)$ 에서 두 식을 적당히 가감하여 수열  $\{a_n+b_n\}$ 과 수열  $\{a_n-b_n\}$ 의 일반항을 구할 수 있을 때 이 방법을 쓴다.

**보기**  $a_1=1, a_{n+1}=pa_n+qb_n \dots\dots \textcircled{1}$   
 $b_1=0, b_{n+1}=pb_n+qa_n \dots\dots \textcircled{2}$   
 이 점화방정식을 만족하는 수열  $\{a_n+b_n\}, \{a_n-b_n\}$ 을 구하여라.

**풀이** ①+② :  $a_{n+1}+b_{n+1}=(p+q)(a_n+b_n)$

$$\therefore a_n+b_n=(p+q)^{n-1}(a_1+b_1)=(p+q)^{n-1}$$



$$\begin{aligned} \text{①}-\text{②} : a_{n+1} - b_{n+1} &= (p-q)(a_n - b_n) \\ \therefore a_n - b_n &= (p-q)^{n-1}(a_1 - b_1) = (p-q)^{n-1} \end{aligned}$$

※ 유의 : 미지의 수열이 3개인 점화방정식에서는 주로 가감법과 대입법을 겸용하여 풀다 (보기 4를 참조하라).

**보기 1.** 다음 수열의 일반항  $a_n$ 과  $b_n$ 을 구하여라. (단,  $a, b, p, q$ 는 상수)

$$\begin{aligned} a_1 &= a, \quad a_{n+1} = pa_n + qb_n && \dots\dots\dots \text{①} \\ b_1 &= b, \quad b_{n+1} = qa_n + pb_n && \dots\dots\dots \text{②} \end{aligned}$$

**풀이**    ①+② :  $a_{n+1} + b_{n+1} = (p+q)(a_n + b_n)$   
 $\therefore a_n + b_n = (p+q)^{n-1}(a_1 + b_1) = (a+b)(p+q)^{n-1}$  .....③

①-② :  $a_{n+1} - b_{n+1} = (p-q)(a_n - b_n)$   
 $\therefore a_n - b_n = (p-q)^{n-1}(a_1 - b_1) = (a-b)(p+q)^{n-1}$  .....④

③+④ :  $a_n = \frac{1}{2}\{(a+b)(p+q)^{n-1} + (a-b)(p-q)^{n-1}\}$

③-④ :  $b_n = \frac{1}{2}\{(a+b)(p+q)^{n-1} - (a-b)(p-q)^{n-1}\}$

《유제》 점  $p_n(x_n, y_n)$ 과 점  $p_{n+1}(x_{n+1}, y_{n+1})$ 은 다음과 같은 관계가 있다.

$$\begin{cases} 3x_{n+1} = 2x_n + y_n \\ 3y_{n+1} = 2y_n + x_n \end{cases}$$

만일  $n$ 이 한없이 커지면 점  $p_n$ 은 어느 점에 가까이 가는가? 이 점  $p(x, y)$ 를  $p_0$ 의 좌표  $(x_0, y_0)$ 로 나타내어라.       $\square p\left(\frac{x_0+y_0}{2}, \frac{x_0+y_0}{2}\right)$

**보기 2.**  $a_1=1, a_{n+1}=2a_n - b_n + 1$  .....①'  
 $b_1=3, b_{n+1}=4a_n - 3b_n - 2$  .....②'  
 ①', ②'를 만족하는  $a_n, b_n$ 을 구하여라.

**풀이**    ①'에서  $b_n = -a_{n+1} + 2a_n + 1$ 을 ②'에 대입하면 다음과 같다. .....①

$$-a_{n+2} + 2a_{n+1} + 1 = 4a_n - 3(-a_{n+1} + 2a_n + 1) - 2$$

즉,  $a_{n+2} + a_{n+1} - 2a_n = 6$ 이다. (단,  $a_2 = 2a_1 - b_1 + 1 = 0$ )

방정식  $t^2 + t - 2 = 0$ 에서  $t=1, t=-2$ 이므로,  $(a_{n+2} - a_{n+1}) + 2(a_{n+1} - a_n) = 6$ 에서

$$a_{n+1} - a_n = c_n \text{ 이라면 다음 식이 성립한다.} \dots\dots\dots \text{②}$$

$$c_{n+1} + 2c_n = 6, \quad (c_1 = a_2 - a_1 = -1) \dots\dots\dots \text{③}$$

$$x + 2x = 6, \quad (x=2) \dots\dots\dots \text{④}$$

③-④에서  $c_n - x = d_n$ 이라면,  $d_1 = c_1 - x = -3$ 이고 다음 식이 성립한다. .....⑤

$$d_{n+1} = -2d_n$$

$$\therefore d_n = (-2)^{n-1}d_1 = -3 \cdot (-2)^{n-1} \dots\dots\dots \text{⑥}$$

②, ⑤, ⑥ :  $a_{n+1} - a_n = c_n = d_n + x = 2 - 3 \cdot (-2)^{n-1}$

$$\therefore a_n = 1 + \sum_{k=1}^{n-1} \{2 - 3(-2)^{k-1}\} = (-2)^{n-1} + 2(n-1) \dots\dots\dots \text{⑦}$$

①, ⑦ :  $b_n = -a_{n+1} + 2a_n + 1 = (-2)^{n+1} + 2n - 3$

《유제》  $a_1=1, b_1=-3, a_{n+1}=2a_n+b_n+5, b_{n+1}=4a_n-b_n+2$ 인  $a_n$ 과  $b_n$ 을 구하여라.

$$\text{답 } a_n = \frac{1}{10}\{8 \cdot 3^n - 3 \cdot (-2)^n\} - 2, \quad b_n = \frac{1}{10}\{8 \cdot 3^n + 12(-2)^n\} - 3$$

**보기 3.**  $a_1=1, 2a_{2n}=a_{2n-1}, 3a_{2n+1}=-a_{2n}$ 인  $a_n$ 을 구하여라.

**풀이**  $a_{2n+1} = \left(-\frac{1}{3}\right)a_{2n} = \left(-\frac{1}{3}\right)\left(\frac{1}{2}a_{2n-1}\right) = \left(-\frac{1}{6}\right)a_{2n-1}$

$$\therefore a_{2n-1} = \left(-\frac{1}{6}\right)^{n-1} \quad \dots\dots\dots \text{①}$$

$$a_{2n} = \frac{1}{2}a_{2n-1} = \frac{1}{2}\left(-\frac{1}{6}\right)^{n-1} \quad \dots\dots\dots \text{②}$$

$$\text{①, ② : } a_n = \frac{1+(-1)^n}{2} \times \left\{ \frac{1}{2}\left(-\frac{1}{6}\right)^{\frac{n-1}{2}} \right\} + \frac{1-(-1)^n}{2} \times \left(-\frac{1}{6}\right)^{\frac{n-1}{2}}$$

《유제》  $a_1=1, a_{2n}=pa_{2n-1}, a_{2n+1}=a_{2n}+q$ 에서  $a_n$ 을 구하여라.

$$\text{답 } p=1 \text{ 일 때 } a_{2n-1}=a_{2n}=1+(n-1)q$$

$$p \neq 1 \text{ 일 때 } a_{2n-1} = p^{n-1}\left(1 - \frac{q}{1-p}\right) + \frac{q}{1-p}, \quad a_{2n} = p^n\left(1 - \frac{q}{1-p}\right) + \frac{pq}{1-p}$$

**보기 4.**  $2a_1=b+c, 2b_1=c+a, 2c_1=a+b$

$$2a_{n+1}=b_n+c_n \quad \dots\dots\dots \text{①}$$

$$2b_{n+1}=c_n+a_n \quad \dots\dots\dots \text{②}$$

$$2c_{n+1}=a_n+b_n \quad \dots\dots\dots \text{③}$$

위의 조건들을 만족하는  $a_n, b_n, c_n$ 을 구하여라.

**풀이** ①+②+③ :  $2(a_{n+1}+b_{n+1}+c_{n+1})=2(a_n+b_n+c_n)$ 에서

$$a_n+b_n+c_n=a_1+b_1+c_1=a+b+c \quad \dots\dots\dots \text{④}$$

$$\text{①, ④ : } 2a_{n+1}+a_n=a+b+c \quad \dots\dots\dots \text{⑤}$$

$$2x+x=a+b+c \quad \dots\dots\dots \text{⑥}$$

$$\text{⑤-⑥에서 } a_n-x=d_n \text{ 이라면 다음 식이 성립한다.} \quad \dots\dots\dots \text{⑦}$$

$$d_{n+1} = -\frac{1}{2}d_n$$

$$\therefore d_n = \left(-\frac{1}{2}\right)^{n-1} d_1 = \left(-\frac{1}{2}\right)^{n-1} (a_1-x) = \left(-\frac{1}{2}\right)^{n-1} \cdot \frac{b+c-2a}{6} \quad \dots\dots\dots \text{⑧}$$

$$\text{⑦, ⑧ : } a_n = x + d_n = \frac{a+b+c}{3} + \frac{1}{6} \times \left(-\frac{1}{2}\right)^{n-1} (b+c-2a)$$

같은 방법으로 다음 식들을 구할 수 있다.

$$b_n = \frac{a+b+c}{3} + \left(-\frac{1}{2}\right)^{n-1} \times \frac{a+c-2b}{6}$$

$$c_n = \frac{a+b+c}{3} + \left(-\frac{1}{2}\right)^{n-1} \times \frac{a+b-2c}{6}$$

《유제》 
$$\begin{cases} x_1=3, & x_{n+1}=3-z_n \\ y_1=6, & y_{n+1}=18-4z_n-6x_n \\ z_1=-4, & z_{n+1}=x_n-5y_n+9 \end{cases}$$

이 식들을 만족하는  $x_n, y_n, z_n$ 을 다음의 순서로 구하여라.

- 1)  $x_n + y_n + z_n = c$  (상수)이다.
- 2)  $x_n, y_n, z_n$ 을 구하여라.

$$\text{답 } x_n = 3^n - 2^n + 2, \quad y_n = 2 \cdot 3^n - 2^n + 2, \quad z_n = 2^{n+1} - 3^{n+1} + 1$$

══════ 연습문제 ══════

1. 2개의 수열  $\{a_n\}$ 과  $\{b_n\}$ 이 다음을 만족한다.

$$a_0 = 1, \quad a_{n+1} = 2pa_n + p^2b_n \quad \dots\dots\dots \textcircled{1}$$

$$b_0 = 0, \quad b_{n+1} = 2pb_n + p^2a_n \quad \dots\dots\dots \textcircled{2}$$

- (1)  $a_n, b_n$ 을  $n$ 과  $p$ 로 나타내어라.
- (2)  $a_n$ 과  $b_n$ 이 모두 수렴하도록, 실수  $p$ 의 범위를 정하여라.

2. 평면 위의 점  $P_n(x_n, y_n)$ 과  $P_{n+1}(x_{n+1}, y_{n+1})$ 은 다음의 관계가 있다.

$$x_{n+1} = 2x_n + 3y_n, \quad y_{n+1} = x_n + 2y_n$$

또한  $P_0(1, 0)$ 이다. 원점을  $O$ 라 할 때 직선  $OP_n$ 의 기울기의 극한값을 구하여라.

3. 2개의 수열의 일반항  $a_n, b_n$  사이에 다음과 같은 관계가 있다.

$$a_{n+1} = pa_n + qb_n + 1 \quad \dots\dots\dots \textcircled{1}$$

$$b_{n+1} = qa_n + pb_n \quad \dots\dots\dots \textcircled{2}$$

이 때 두 수열이 수렴하기 위한  $p, q$ 의 조건을 구하라.

4.  $a_1 = 1, a_{2n} = 2a_{2n-1}, a_{2n+1} = a_{2n} + 2^{n-1}$ 인 수열  $\{a_n\}$ 이 있다.

- (1) 일반항  $a_n$ 을 구하여라.
- (2)  $\sum_{k=1}^{2n} a_k$ 를 구하여라.

5. 수열  $\{x_n\}$ 과  $\{y_n\}$ 이 다음 조건을 만족한다.

(조건 1)  $x_{n+1} = px_n + qy_n, y_{n+1} = rx_n + sy_n$  (단,  $p, q, r, s$ 는 상수)

(조건 2) 방정식  $t^2 - (p+s)t + (ps-qr) = 0$  서로 다른 두 근은  $\alpha, \beta$ 이다.

$$(1) (\alpha - \beta)x_n = (\alpha^n - \beta^n)x_1 + A(\alpha^{n-1} - \beta^{n-1})$$

$$(\alpha - \beta)y_n = (\alpha^n - \beta^n)y_1 + B(\alpha^{n-1} - \beta^{n-1})$$

이 식들이 성립하기 위한  $A, B$ 를  $p, q, r, s, x_1, y_1$ 으로 나타내어라.

- (2)  $p = q = 1, r = 2, s = 3, x_1 = y_1 = 1$ 일 때  $\lim_{n \rightarrow \infty} \frac{x_n}{y_n}$ 을 구하여라.

$$6. \quad x_1 = -1, \quad 6x_{n+1} = x_n + 2y_n + 3z_n - 4 \quad \dots\dots\dots \textcircled{1}$$

$$y_1 = -1, \quad 6y_{n+1} = 2x_n + 3y_n + z_n - 5 \quad \dots\dots\dots \textcircled{2}$$

$$z_1 = 0, \quad 6z_{n+1} = 3x_n + y_n + 2z_n + 9 \quad \dots\dots\dots \textcircled{3}$$

이 식들을 만족하는 수열  $\{x_n\}$ ,  $\{y_n\}$ ,  $\{z_n\}$ 에서 다음 물음에 답하여라.

(1)  $x_n + y_n + z_n = -2$ 를 증명하여라.

(2)  $x_n, y_n, z_n$ 을 구하여라.

## A.5 점화식과 극한

### 1. 점화식에서 극한값을 구하는 방법

1) 수열  $\{a_n\}$ 의 일반항을 구해서  $\lim_{n \rightarrow \infty} a_n$ 을 조사한다(보기 1을 참조하라).

2) 수열  $\{a_n\}$ 의 일반항을 구할 수 없는 경우

① 수열  $\{a_n\}$ 이 단조수열이면 수학적 귀납법을 사용하여 단조성을 조사하고 유계(상계와 하계)를 정한 뒤, 특성방정식을 사용하여 극한값을 구한다(보기 2를 참조하라).

**정리 1)**  $a_1 \leq a_2 \leq a_3 \leq \dots \leq a_n \leq \dots \leq U$  ( $U$ 는 상수)이면 수열  $\{a_n\}$ 은 단조증가수열,  $U$ 를 상계(上界)라 한다.

2)  $a_1 \geq a_2 \geq a_3 \geq \dots \geq a_n \geq \dots \geq L$  ( $L$ 은 상수)이면 수열  $\{a_n\}$ 은 단조감소수열,  $L$ 을 하계(下界)라 한다.

1) 또는 2)의 수열은 극한값을 갖는다.

※ 유의: 상계와 하계를 합하여 유계라 한다. 위의 방법은 단조인 동시에 유계이어야만 극한값을 갖는다.

② 특성방정식 중 적당한 근  $\alpha$ 를 구하고,  $0 < k < 1$ 인  $k$ 를 구하여 다음과 같이 극한값을 유도할 수 있다. 우선 점화식으로부터 다음 식을 유도한다.

$$0 \leq |a_n - \alpha| \leq k|a_{n-1} - \alpha|$$

따라서 다음 식이 성립한다.

$$0 \leq |a_n - \alpha| \leq k^{n-1}|a_1 - \alpha|$$

이 때  $\lim_{n \rightarrow \infty} k^{n-1}|a_1 - \alpha| = 0$ 이다. 따라서 다음 식이 성립한다(보기 3을 참조하라).

$$\lim_{n \rightarrow \infty} |a_n - \alpha| = 0, \quad \text{즉} \quad \lim_{n \rightarrow \infty} a_n = \alpha$$

③ 특성방정식과 그래프를 사용한다. 이 경우는 1), 2)에서 어떤 수열이 극한값을 갖는다는 것을 증명한 부분을 생략한 것으로 그래프를 사용한 직관에 의한 방법이다(보기 4를 참조하라).

### 2. 특성방정식과 점화식의 극한

**정리 2** 2항간의 점화식  $a_{n+1} = f(a_n)$ 에서 수열  $\{a_n\}$ 이 극한값을 가지면, 그 극한값은 방정식  $x = f(x)$ 의 한 근이다. 이 방정식을 특성방정식이라 한다.

**[증명]** 수열  $\{a_n\}$ 이 극한값을 갖는다고 했으므로  $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} a_{n+1} = \alpha$ 이다.

또한  $f(a_n) = a_{n+1}$ 에 의해서 다음 식이 성립한다.

$$\lim_{n \rightarrow \infty} a_{n+1} = \lim_{n \rightarrow \infty} f(a_n) = f(\lim_{n \rightarrow \infty} a_n)$$

즉,  $\alpha = f(\alpha)$ 이다. 따라서  $\alpha$ 는  $x=f(x)$ 의 한 근이다.

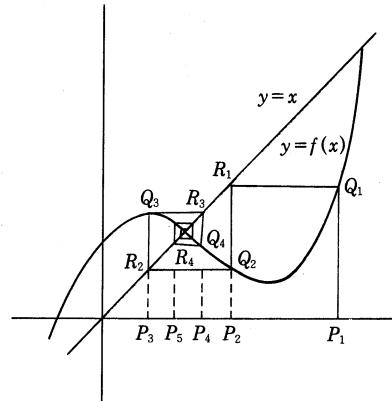
Q.E.D

이 특성방정식의 근을 특성근이라 한다.

### 3. 특성방정식과 그래프

앞에서 배운 특성방정식  $x=f(x)$ 의 근은  $y=f(x)$ 와  $y=x$ 가 만나는 점의  $x$ 좌표이다. 이러한 관점에서 생각하여 보자.

지금  $x$ 축 위의 점  $P_1(a_1, 0)$ 를 잡고, 곡선 위의  $x$ 좌표가  $a_1$ 인 점을  $Q_1$ 이라고 하면  $Q_1$ 의  $y$ 좌표는  $f(a_1) = a_2$ , 즉  $Q_1(a_1, a_2)$ 이다. 또  $Q_1$ 에서  $x$ 축에 평행선을 그어  $y=x$ 와 만나는 점을  $R_1$ 이라고 하면, ( $Q$ 의  $y$ 좌표) = ( $R_1$ 의  $y$ 좌표) = ( $R_1$ 의  $x$ 좌표)에서  $R_1(a_2, a_2)$ 이다.  $R_1$ 에서  $x$ 축에 수선을 내려 그 발을  $P_2$ 라하면  $P_2$ 는  $(a_2, 0)$ 이다.  $P_2$ 에서 차례로, 위와 같은 순으로  $P_2 \rightarrow Q_2 \rightarrow R_2 \rightarrow P_3 \rightarrow Q_3 \rightarrow \dots$ 를 구하면  $P_1(a_1, 0)$ ,  $Q_i(a_i, a_{i+1})$ ,  $R_i(a_{i+1}, a_{i+1})$ 이 된다. 이러한 경과로 직관에 의해서 수열의 수렴, 발산을 알 수 있고, 만일 수렴한다면 점  $Q_i$ 는  $y=x$ 와  $y=f(x)$ 의 교점에 한없이 가까워짐을 알 수 있다. 따라서  $\lim_{n \rightarrow \infty} a_n$ 은 교점의  $x$ 좌표가 된다.



**[보기 1.]**  $a_1=1, a_{n+1}=a_n \sin \theta + b_n \cos \theta$  .....①

$b_1=0, b_{n+1}=b_n \sin \theta + a_n \cos \theta$  .....②

수열  $\{a_n\}$ 과  $\{b_n\}$ 이 수렴할  $\theta$ 의 조건을 구하여라.

**[풀이]** ①+② :  $a_{n+1} + b_{n+1} = (\sin \theta + \cos \theta)(a_n + b_n)$

①-② :  $a_{n+1} - b_{n+1} = (\sin \theta - \cos \theta)(a_n - b_n)$

만일  $a_n + b_n = c_n (c_1=1)$ ,  $a_n - b_n = d_n (d_1=1)$ 이라면, 다음 식들이 성립한다.

$$c_{n+1} = (\sin \theta + \cos \theta) c_n, \quad d_{n+1} = (\sin \theta - \cos \theta) d_n$$

$$\therefore c_n = (\sin \theta + \cos \theta)^{n-1}, \quad d_n = (\sin \theta - \cos \theta)^{n-1}$$

$$\therefore a_n = \frac{1}{2} \{c_n + d_n\} = \frac{1}{2} \{(\sin \theta + \cos \theta)^{n-1} + (\sin \theta - \cos \theta)^{n-1}\}$$

$$b_n = \frac{1}{2} \{c_n - d_n\} = \frac{1}{2} \{(\sin \theta + \cos \theta)^{n-1} - (\sin \theta - \cos \theta)^{n-1}\}$$

공비  $r$ 인 무한등비수열에서 수렴할 조건은  $-1 < r \leq 1$ 이므로 다음 식들이 성립한다.

$$-1 < \sin\theta + \cos\theta = \sqrt{2}\sin\left(\theta + \frac{\pi}{4}\right) \leq 1 \quad \dots\dots\dots ③$$

$$-1 < \sin\theta - \cos\theta = \sqrt{2}\sin\left(\theta - \frac{\pi}{4}\right) \leq 1 \quad \dots\dots\dots ④$$

$$③ : 2n\pi - \frac{\pi}{2} < \theta \leq 2n\pi, \quad 2n\pi + \frac{\pi}{2} \leq \theta < 2n\pi + \pi \quad \dots\dots\dots ⑤$$

$$④ : 2m\pi < \theta \leq 2m\pi + \frac{\pi}{2}, \quad 2m\pi + \pi \leq \theta < 2m\pi + \frac{3}{2}\pi \quad (\text{단, } m, n \text{은 정수}) \quad \dots\dots\dots ⑥$$

⑤, ⑥의 공통범위는  $\theta = 2l\pi + \frac{\pi}{2}$ 이다. (단,  $l$ 은 정수)

**보기 2.**  $a_1 = \frac{3}{2}$ ,  $a_{n+1} = \frac{1}{3}(a_n^2 + 2)$ 인 수열  $\{a_n\}$ 의 극한값을 다음 순서로 구하여라.

- (1) 수열의 각 항은 양수임을 증명하라.
- (2) 이 수열은 단조감소수열임을 증명하라.
- (3)  $\lim_{n \rightarrow \infty} a_n$ 을 구하여라.

**풀이** (1)  $n=1$ 에서  $a_1 = \frac{3}{2} > 0$ 이다.

$n=k$ 에서  $a_k > 0$ 이라 가정하자.

$$n=k+1 \text{일 때 } a_{k+1} = \frac{1}{3}(a_k^2 + 2) > \frac{2}{3} > 0 \text{이다.}$$

따라서 모든 자연수  $n$ 에 대해  $a_n > 0$ 이다.

(2)  $a_n > a_{n+1}$ 을 수학적 귀납법으로 증명하자.

$$n=1 \text{일 때 } a_2 - a_1 = \frac{1}{3}(a_1^2 + 2) - a_1 = -\frac{1}{12} < 0 \text{이다. } \therefore a_2 < a_1$$

$n=k$ 일 때,  $a_{k+1} < a_k$ 라 가정하자.

$n=k+1$ 일 때 다음 식이 성립한다.

$$a_{k+2} - a_{k+1} = \frac{1}{3}(a_{k+1}^2 + 2) - \frac{1}{3}(a_k^2 + 2) = \frac{1}{3}(a_{k+1} - a_k)(a_{k+1} + a_k) < 0$$

$$\therefore a_{k+2} < a_{k+1}$$

따라서 모든 자연수  $n$ 에 대해  $a_{n+1} < a_n$ 이다.

(3) (1), (2)에서 유계단조수열이므로 극한값을 갖는다. 따라서  $\lim_{n \rightarrow \infty} a_{n+1} = \lim_{n \rightarrow \infty} a_n = \alpha$ 이므

로  $\alpha = \frac{1}{3}(\alpha^2 + 2)$  즉,  $\alpha = 1, 2$ 이다. 이 때  $0 < a_n \leq \frac{3}{2}$ 이므로  $0 \leq \lim_{n \rightarrow \infty} a_n = \alpha \leq \frac{3}{2}$ 이다.

$$\therefore \lim_{n \rightarrow \infty} a_n = 1$$

**보기 3.**  $x_1 = \frac{3}{4}\pi$ ,  $2x_{n+1} + \cos x_n = \pi$ 일 때 수열  $\{x_n\}$ 의 극한을 구하여라.

**풀이** 특성방정식  $2x = \pi - \cos x$ 에서  $x = \frac{\pi}{2}$ 이다. 이 근을 이용하여 원식을 고치면 다음 식이 성립한다.

$$x_{n+1} - \frac{\pi}{2} = -\frac{1}{2}\cos x_n = -\frac{1}{2}\sin\left(\frac{\pi}{2} - x_n\right) = \frac{1}{2}\sin\left(x_n - \frac{\pi}{2}\right)$$

따라서 다음 식이 성립한다.

$$0 \leq \left| x_{n+1} - \frac{\pi}{2} \right| = \frac{1}{2} \left| \sin \left( x_n - \frac{\pi}{2} \right) \right| \leq \frac{1}{2} \left| x_n - \frac{\pi}{2} \right|, \quad (\because |\sin x| \leq |x|)$$

$$\therefore 0 \leq \left| x_n - \frac{\pi}{2} \right| \leq \left( \frac{1}{2} \right)^{n-1} \left| x_1 - \frac{\pi}{2} \right|$$

$$\therefore \lim_{n \rightarrow \infty} x_n = \frac{\pi}{2}$$

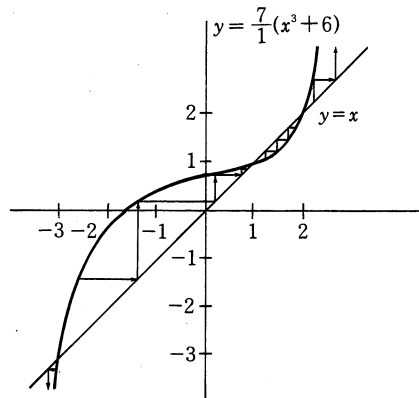
**보기 4.** 그래프를 이용하여 다음 수열의 극한값을  $p$ 에 따라 조사하여라.

$$a_1 = p, \quad a_{n+1} = \frac{1}{7}(a_n^3 + 6)$$

**풀이** 특성방정식  $x = \frac{1}{7}(x^3 + 6)$ 의 근은  $y = x$ ,  $y = \frac{1}{7}(x^3 + 6)$ 가 만나는 점이다. 특성근들은

$x = 1, 2, -3$ 이고, 그림에 따라 수렴, 발산을 조사하면 다음과 같다.

- i)  $p < -3$                        $\lim_{n \rightarrow \infty} a_n = -\infty$
- ii)  $p = -3$                          $\lim_{n \rightarrow \infty} a_n = -3$
- iii)  $-3 < p < 1$                    $\lim_{n \rightarrow \infty} a_n = 1$
- iv)  $p = 1$                            $\lim_{n \rightarrow \infty} a_n = 1$
- v)  $1 < p < 2$                        $\lim_{n \rightarrow \infty} a_n = 1$
- vi)  $p = 2$                            $\lim_{n \rightarrow \infty} a_n = 2$
- vii)  $2 < p$                           $\lim_{n \rightarrow \infty} a_n = \infty$



|     |           |    |    |    |          |
|-----|-----------|----|----|----|----------|
| $p$ |           | -3 |    | 2  |          |
| 발산  | 발산        | 수렴 | 수렴 | 수렴 | 발산       |
| 극한  | $-\infty$ | -3 | 1  | 2  | $\infty$ |

══════ 연습문제 ══════

1. 수열  $\{x_n\}$ 은  $x_n = q \sin x_{n-1} + a$  (단,  $0 < q < 1$ ,  $a$ 는 상수)이다. 이 수열이 수렴함을 증명하고 이 때의 극한값을 근으로 하는 방정식을 구하여라.

2. 수열  $\{a_n\}$ 에서  $a_{n+1} = \sin\left(\frac{\pi}{2} a_n\right)$ ,  $a_1 = \frac{1}{2}$ 이다.

- (1) 임의의 자연수  $n$ 에 대해  $0 < a_n < 1$ 을 증명하여라.
- (2) 임의의 자연수  $n$ 에 대해  $a_n < a_{n+1}$ 을 증명하여라.
- (3) 수열  $\{a_n\}$ 의 극한값을 구하여라.



3. 초항  $a_1$ 인 수열  $\{a_n\}$ 에서  $a_{n+1}$ 은 방정식  $a_n x^2 + 2x - a_n = 0$ 의 근 중에서 절대값이 1보다 작은 근이다. (단,  $a_n \neq 0$ )
- (1)  $a_{n+1}$ 을  $a_n$ 으로 나타내어라.
  - (2) 수열  $\{a_n\}$ 의 수렴 또는 발산을 조사하고 수렴하는 경우 극한값을 구하여라.
4.  $9u_{n+1} = (10 - u_n^2)u_n$ ,  $0 < u_1 < \frac{3}{2}$ 인 수열  $\{u_n\}$ 이 있다.
- (1)  $0 < u_n < \frac{3}{2}$ 을 증명하여라.
  - (2)  $u_1 > u_2$ 이면  $u_n > u_{n+1}$ 이고,  $u_1 < u_2$ 이면  $u_n < u_{n+1}$ 을 증명하여라.
  - (3)  $\lim_{n \rightarrow \infty} u_n$ 을 구하여라.
5. 그래프를 이용하여  $a_1 = a$ ,  $a_{n+1} = \sqrt[3]{3a_n - 2}$ 인 수열  $\{a_n\}$ 의 수렴 또는 발산을 조사하고 극한값을 구하여라.
6. 수열  $\{a_n\}$ 에서  $a_1 > a$ 이고  $a_{n+1} = f(a_n)$ 이다. 또한 함수  $f(x)$ 의 도함수  $f'(x)$ 는  $0 < f'(x) < 1$ 이고,  $a$ 는  $f(x) = x$ 의 한 근이다.
- (1)  $a_n > a$ 를 증명하여라.
  - (2)  $a_n > a_{n+1}$ 을 증명하여라.
  - (3)  $\lim_{n \rightarrow \infty} a_n$ 을 구하여라.

### A.6 무리식의 점화식

일반적으로 무리함수꼴과 2차 이상의 다항식꼴의 점화식에서 일반항은 구할 수 없다. 따라서 극한의 문제는 앞에서 배운 방법을 사용하기로 한다. 그러나  $(a_{n+2})^p = (a_{n+1})^q (a_n)^r$  과 같은 것은 일반항을 구할 수 있는 특수한 것으로서, 우선  $a_n > 0$  을 증명한 뒤 양변에  $\log$  를 취하여  $\log a_n = b_n$  으로 놓아  $pb_{n+2} = qb_{n+1} + rb_n$  꼴로 고친 다음 앞에서 배운 점화식으로 귀착시킨다.

**[보기] 1.**  $2a_{n+1}^3 = a_n^4$ ,  $a_1 = 2$  를 만족하는 수열의 일반항을 구하여라.

**[풀이]** 우선 일반항  $a_n$  이 양수임을 증명한다.

$n=1$ 에서  $a_1 = 2 > 0$ 이다.

$n=k$ 에서  $a_k > 0$ 이라 가정하자.

$n=k+1$ 에서  $2a_{k+1}^3 = a_k^4 > 0$ 이다.  $\therefore a_{k+1} > 0$

따라서 모든 자연수  $n$ 에 대해  $a_n > 0$ 이다. .....①

위의 결과를 사용하여 원식의 양변에 밑이 2인  $\log$  를 취하면 다음과 같다.

$$3\log_2 a_{n+1} + 1 = 4\log_2 a_n$$

이 식에서  $\log_2 a_n = b_n$  이라면 다음 식이 성립한다. .....②

$$3b_{n+1} = 4b_n - 1, (b_1 = 1) \quad \text{.....③}$$

$$3x = 4x - 1, (x = 1) \quad \text{.....④}$$

③-④ :  $3(b_{n+1} - x) = 4(b_n - x)$  에서  $b_n - x = c_n$  이라면 다음 식이 성립한다. .....⑤

$$c_{n+1} = \frac{4}{3}c_n$$

$$\therefore c_n = \left(\frac{4}{3}\right)^{n-1} c_1 = 0, (\because c_1 = b_1 - x = 0) \quad \text{.....⑥}$$

②, ⑤, ⑥ :  $a_n = 2^{b_n} = 2^{c_n+x} = 2$

《유제》  $a_1 = 1, a_2 = 2, a_{n+2} = \sqrt{a_n a_{n+1}}$  에서  $\lim_{n \rightarrow \infty} a_n$  을 구하여라. [답]  $3\sqrt{4}$

**[보기] 2.**  $a_1 = 1, b_1 = 8, a_{n+1}^2 = a_n b_{n+1}, b_{n+1}^2 = a_{n+1} b_n, a_n > 0, b_n > 0$  이라 하자.  
 (1)  $a_n, b_n$  을 구하여라.  
 (2)  $\lim_{n \rightarrow \infty} a_n, \lim_{n \rightarrow \infty} b_n$  을 구하여라.

**[풀이]** 우선  $a_n > 0, b_n > 0$  이므로 밑이 2인  $\log$  를 취하여  $\log_2 a_n = x_n, \log_2 b_n = y_n$  이라면  $x_1 = 0, y_1 = 3$  이고 다음 식이 성립한다.

$$2x_{n+1} = x_n + y_{n+1} \quad \text{.....①}$$

$$2y_{n+1} = x_{n+1} + y_n \quad \text{.....②}$$

①+② :  $x_{n+1} + y_{n+1} = x_n + y_n$

$$\therefore x_n + y_n = 3 \quad \dots\dots\dots ③$$

$$\begin{aligned} ① - ② : x_{n+1} - y_{n+1} &= \frac{1}{3}(x_n - y_n) \\ \therefore x_n - y_n &= \left(\frac{1}{3}\right)^{n-1} (x_1 - y_1) = -\left(\frac{1}{3}\right)^{n-2} \quad \dots\dots\dots ④ \end{aligned}$$

$$\frac{③ + ④}{2} : x_n = \frac{1}{2} \left\{ 3 - \left(\frac{1}{3}\right)^{n-2} \right\}$$

$$\frac{③ - ④}{2} : y_n = \frac{1}{2} \left\{ 3 + \left(\frac{1}{3}\right)^{n-2} \right\}$$

$$\therefore a_n = 2^{\frac{1}{2}} \left\{ 3 - \left(\frac{1}{3}\right)^{n-2} \right\} \quad b_n = 2^{\frac{1}{2}} \left\{ 3 + \left(\frac{1}{3}\right)^{n-2} \right\}$$

$$\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n = 2\sqrt{2}$$

**보기 3.**  $a_1 = 1$ ,  $a_{n+1} = \sqrt{a_n + 1} + 1$ 인 수열의 극한을 조사하여라.

▶ **방법 1.** 수열의 일반항이  $1 \leq a_n < 3$ 이라 가정한다.

$$n=1, 1 \leq a_1 = 1 < 3$$

$$n=2, 1 \leq a_2 = 1 + \sqrt{2} < 3$$

$n=k$ 일 때  $1 \leq a_k < 3$ 이라 가정하자.

$n=k+1$ 일 때  $y = \sqrt{x+1} + 1$ 은 구간  $(1, 3)$ 에서 단조증가함수이므로 다음 식이 성립한다.

$$1 \leq \sqrt{1+1} + 1 \leq a_{k+1} = \sqrt{a_k + 1} + 1 < \sqrt{3+1} + 1 = 3$$

따라서 모든 자연수  $n$ 에 대해  $1 \leq a_n < 3$ 이다. .....①

$\{a_n\}$ 은 유계단조증가수열이므로 극한값을 갖는다.

만일  $\lim_{n \rightarrow \infty} a_{n+1} = \lim_{n \rightarrow \infty} a_n = \alpha$ 라고 놓으면 다음 식이 성립한다.

$$\alpha = \sqrt{\alpha + 1} + 1, \therefore \alpha = 3 \quad \dots\dots\dots ②$$

①, ②에서  $\lim_{n \rightarrow \infty} a_n = 3$

▶ **방법 2.** 특성방정식  $x = \sqrt{x+1} + 1$ 의 특성근 3을 식의 양변에서 빼면 다음 식이 성립한다.

$$a_{n+1} - 3 = \sqrt{a_n + 1} - 2 = \frac{a_n - 3}{\sqrt{a_n + 1} + 2}$$

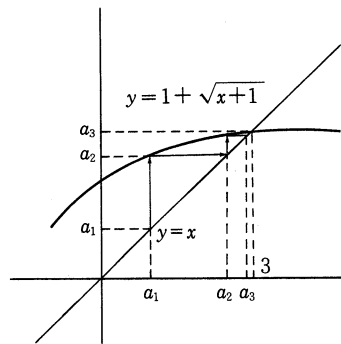
따라서 다음 식이 성립한다.

$$0 \leq |a_{n+1} - 3| = \frac{|a_n - 3|}{\sqrt{a_n + 1} + 2} \leq \frac{1}{2} |a_n - 3|,$$

$$(\because \sqrt{a_n + 1} \geq 0)$$

$$\therefore 0 \leq |a_n - 3| \leq \frac{1}{2^{n-1}} |a_1 - 3|$$

이 때  $\lim_{n \rightarrow \infty} \frac{1}{2^{n-1}} |a_1 - 3| = 0$ 이므로  $\lim_{n \rightarrow \infty} a_n = 3$ 이다.



▶ **방법 3.** 그래프에 의해서  $\lim_{n \rightarrow \infty} a_{n+1} = \lim_{n \rightarrow \infty} a_n = \alpha$ 라 하고, 준식의 양변에  $\lim$ 를 취하면 다음과 같다.

$$\alpha = \sqrt{\alpha + 1} + 1$$

즉,  $y=x$ 와  $y=\sqrt{x+1}+1$ 의 교점의  $x$ 좌표가  
수열의 극한값이므로  $a=3$ , 따라서  $\lim_{n \rightarrow \infty} a_n=3$ 이다.

《유제》  $a_1=1$ ,  $a_{n+1}=\sqrt{a_n+2}$ 일 때  $\lim_{n \rightarrow \infty} a_n$ 을 구하여라.

답 2

### ══════ 연습문제 ══════

1.  $a_1=4$ ,  $a_n=\sqrt[3]{a_{n-1}}$ 일 때  $\lim_{n \rightarrow \infty} a_1 a_2 a_3 \cdots a_n$ 을 구하여라.
2.  $y_1=x\sqrt{x}$ ,  $y_n=x\sqrt{xy_{n-1}}$ 일 때  $\lim_{n \rightarrow \infty} y_n$ 을 구하여라.
3.  $a_1=1$ ,  $a_2=10$ ,  $(a_{n+2})^2 a_n=(a_{n+1})^3$ (단,  $n \geq 1$ ),  $a_n > 0$ 인 수열의 일반항과  $\lim_{n \rightarrow \infty} a_n$ 을 구하여라.
4.  $a_1=1$ ,  $b_1=2$ ,  $c_1=5$ 인  $a_{n+1}=\sqrt{b_n c_n}$ ,  $b_{n+1}=\sqrt{c_n a_n}$ ,  $c_{n+1}=\sqrt{a_n b_n}$ 인 수열에서  $a_n$ ,  $b_n$ ,  $c_n$ 을 구하고 각각의 극한을 구하여라.
5.  $a$ 는  $1 < a < 2$ 인 상수이고  $x_1=a$ ,  $x_{n+1}=\frac{1}{3}(x_n^2+2)$ 인 수열이 있다.
  - (1)  $1 < x < a$ 를 증명하여라. (단,  $n \geq 2$ )
  - (2)  $x_n - 1 \leq \frac{1}{3}(a+1)(x_{n-1}-1)$ 을 증명하여라.
  - (3)  $\lim_{n \rightarrow \infty} x_n$ 을 구하여라.
6.  $y=1-\frac{1}{2}x^2$ 과 직선  $y=x$ 를 이용하여  $a_1=p$ ,  $a_{n+1}=1-\frac{1}{2}a_n^2$ 으로 정의된 수열의 극한을 조사하여라.
7.  $f(x)=x(2-x)$ 이고  $x_{n+1}=f(x_n)$ 인 수열  $\{x_n\}$ 이 있다.
  - (1)  $x_n$ 을  $x_1$ 과  $n$ 으로 나타내어라.
  - (2) 이 수열이 수렴하기 위한  $x_1$ 의 범위를 조사하여라.
8.  $f_1(x)=(x-a)^2$ ,  $f_{n+1}(x)=[f_n(x)-a]^2$ 으로 정의된  $x$ 의 정식들이 있다.  $f_i(x)$ 의 정수항, 일차항의 계수를  $a_i$ ,  $b_i$ 라 하자. 즉,  $f_i(x)=a_i+b_i x+\cdots$  라고 하자.

(1)  $a_{n+1}$ ,  $b_{n+1}$ 을  $a_n$ 과  $b_n$ 으로 나타내어라.

(2)  $a=2$ 일 때  $a_n$ ,  $b_n$ 을 구하여라.

9. 양수  $x$ 에 대해  $2a_1=x$ ,  $2a_{n+1}=a_n^2+1$ 인 수열  $\{a_n\}$ 이 있다.

(1)  $x \neq 2$ 일 때  $a_1 < a_2 < a_3 < \dots < a_n < \dots$ 을 증명하여라.

(2)  $x < 2$ 일 때  $a_n < 1$ 을 증명하고,  $\varepsilon$ 을  $0 < \varepsilon < 1 - \frac{x}{2}$ 로 잡을 때  $a_1, a_2, a_3, \dots, a_{n-1},$

$a_n$ 까지가  $1 - \varepsilon$  이하로 되면, 개수  $n$ 에 대해 다음 부등식이 성립함을 증명하여라.

$$2 - x < n\varepsilon^2$$

### A.7 분수식의 점화식

일반적으로 분모가 1차식형의 분수점화식은 다음과 같이 구할 수 있다.

$a_1 = a, a_{n+1} = \frac{pa_n + q}{ra_n + s}$ 인 수열  $\{a_n\}$ 의 특성방정식  $x = \frac{px + q}{rx + s}$ 의 근을  $\alpha, \beta$ 라 하자.  
 즉,  $rx^2 + (s-p)x - q = 0$ 의 2근을  $\alpha, \beta$ 라 한다. (단,  $r \neq 0, ps - qr \neq 0$ 이다.  $\therefore r = 0$   
 이면 위의 식이 1차식의 점화식이 되고  $ps - qr = 0$ 이면  $a_{n+1} = \frac{psa_n + qs}{rsa_n + ss} = \frac{qra_n + qs}{s(ra_n + s)}$   
 $= \frac{q(ra_n + s)}{s(ra_n + s)} = \frac{q}{s}$ 이므로  $a_n = c$ (상수)인 수열이다)

▶ **방법 1.** 위의 식 양변에서  $\alpha$ (특성근)를 빼면 다음 식이 성립한다.

$$a_{n+1} - \alpha = \frac{pa_n + q}{ra_n + s} - \alpha = \frac{(p - r\alpha)a_n + q - as}{ra_n + s} = \frac{(p - r\alpha)(a_n - \alpha)}{ra_n + s} \dots\dots\dots ①$$

( $\because ra^2 + (s-p)a - q = 0$ 에서  $q - as = a(r\alpha - p)$ )

이 때 ①에서  $\alpha, \beta$ 는 바꾸어도 상관없으므로 다음 식이 성립한다.

$$a_{n+1} - \beta = \frac{(p - r\beta)(a_n - \beta)}{ra_n + s} \dots\dots\dots ②$$

$$① \div ② : \frac{a_{n+1} - \alpha}{a_{n+1} - \beta} = \frac{p - r\alpha}{p - r\beta} \times \frac{a_n - \alpha}{a_n - \beta}$$

다음과 같은 식들을 정의하자.

$$\frac{p - r\alpha}{p - r\beta} = \lambda, \frac{a_n - \alpha}{a_n - \beta} = b_n \dots\dots\dots ③$$

다음 식이 성립한다.

$$b_{n+1} = \lambda b_n$$

$$\therefore b_n = (\lambda)^{n-1} b_1 \dots\dots\dots ④$$

$$③, ④ : a_n = \frac{\alpha - \beta b_n}{1 - b_n} = \frac{\alpha - \beta(\lambda)^{n-1} b_1}{1 - (\lambda)^{n-1} b_1}$$

<주의> 이 방법은  $\alpha, \beta$ 가 서로 다를 때에 한해서 쓸 수 있다.

▶ **방법 2.** ①의 양변에 역수를 취하면 다음과 같다.

$$\frac{1}{a_{n+1} - \alpha} = \frac{1}{p - r\alpha} \times \frac{\gamma(a_n - \alpha) + s + r\alpha}{a_n - \alpha} = \frac{\gamma}{p - r\alpha} + \frac{s + r\alpha}{p - r\alpha} \times \frac{1}{a_n - \alpha}$$

이 때  $\frac{1}{a_n - \alpha} = c_n, \frac{s + r\alpha}{p - r\alpha} = A, \frac{\gamma}{p - r\alpha} = B$ 라고 정의하면 다음 식이 성립한다.

$$c_{n+1} = Ac_n + B \dots\dots\dots ⑤$$

여기서  $c_n$ 을 구한 뒤 ⑤에서  $a_n = \frac{1}{c_n} + \alpha$ 에 대입한다.

<주의> 여기서  $p - r\alpha$ 가 분모에 나왔는데 이것이 만일  $p - r\alpha = 0$ 이라면 다음 식이 성립한다.

$$\gamma\left(\frac{p}{\gamma}\right)^2 + (s-p)\left(\frac{p}{\gamma}\right) - q = 0$$

즉,  $ps = qr$ 이 되어 제시된 조건에 모순이다.

$$\therefore p - \gamma a \neq 0$$

**보기 1.**  $a_{n+1} = \frac{2a_n+3}{a_n+4}$ ,  $a_1=0$ 인  $a_n$ 을 구하여라.

▶ **방법 1.** 특성방정식  $x = \frac{2x+3}{x+4}$ 에서  $x=1$ ,  $x=-3$ 을 원식의 양변에서 빼면 다음 식들이 성립한다.

$$a_{n+1}-1 = \frac{2a_n+3}{a_n+4} - 1 = \frac{a_n-1}{a_n+4} \quad \dots\dots\dots ①$$

$$a_{n+1}+3 = \frac{2a_n+3}{a_n+4} + 3 = \frac{5(a_n+3)}{a_n+4} \quad \dots\dots\dots ②$$

① ÷ ② :  $\frac{a_{n+1}-1}{a_{n+1}+3} = \frac{1}{5} \frac{a_n-1}{a_n+3}$ 에서  $\frac{a_n-1}{a_n+3} = b_n$ 이라면,  $b_1 = \frac{a_1-1}{a_1+3} = -\frac{1}{3}$ 이고 다음 식이 성립한다. \dots\dots\dots ③

$$b_{n+1} = \frac{1}{5} b_n$$

$$\therefore b_n = \left(\frac{1}{5}\right)^{n-1} b_1 = -\frac{1}{3} \times \frac{1}{5^{n-1}} \quad \dots\dots\dots ④$$

$$③, ④ : a_n = \frac{3b_n+1}{1-b_n} = \frac{3(5^{n-1}-1)}{3 \cdot 5^{n-1}+1}$$

▶ **방법 2.** ①의 양변에서 역수를 취하면 다음 식이 성립한다.

$$\frac{1}{a_{n+1}-1} = \frac{a_n+4}{a_n-1} = \frac{a_n-1+5}{a_n-1} = \frac{5}{a_n-1} + 1$$

$$\frac{1}{a_n-1} = c_n \text{이라면 다음 식이 성립한다.} \quad \dots\dots\dots ⑤$$

$$c_{n+1} = 5c_n + 1, (c_1 = -1) \quad \dots\dots\dots ⑥$$

$$x = 5x + 1, (4x = -1) \quad \dots\dots\dots ⑦$$

$$⑥ - ⑦ : c_{n+1} - x = 5(c_n - x)$$

$$\therefore c_n - x = 5^{n-1}(c_1 - x)$$

$$\text{따라서 } c_n = 5^{n-1}(c_1 - x) + x = -\frac{1}{4}(1 + 3 \cdot 5^{n-1}) \text{이다.} \quad \dots\dots\dots ⑧$$

$$⑤, ⑧ : a_n = \frac{1}{c_n} + 1 = \frac{3(5^{n-1}-1)}{3 \cdot 5^{n-1}+1}$$

《유제》  $a_1=4$ ,  $a_{n+1} = \frac{3a_n+2}{a_n+4}$ 인  $a_n$ 을 구하여라.

$$\text{답 } a_n = \frac{5^{n-1} + 2^{n-1}}{5^{n-1} - 2^{n-2}}$$

**보기 2.**  $a_1=1$ ,  $a_{n+1} = \frac{a_n-4}{a_n-3}$ 인  $a_n$ 을 구하여라.

**풀이** 특성방정식  $x = \frac{x-4}{x-3}$ 에서  $x=2$ (중근)를 원식의 양변에서 빼면 다음과 같다.

$$a_{n+1}-2 = \frac{a_n-4}{a_n-3} - 2 = \frac{-(a_n-2)}{a_n-3}$$

이 식의 양변에 역수를 취하면 다음과 같다.

$$\frac{1}{a_{n+1}-2} = \frac{-(a_n-3)}{a_n-2} = -1 + \frac{1}{a_n-2}$$

이 식에서  $\frac{1}{a_n-2} = b_n$ 이라 놓으면 다음 식이 성립한다. .....①

$$b_{n+1} = b_n - 1$$

$$\therefore b_n = b_1 + (n-1)(-1) = -n \quad (\because b_1 = -1) \quad \text{.....②}$$

$$\text{①, ②} : a_n = \frac{1}{b_n} + 2 = 2 - \frac{1}{n} = \frac{2n-1}{n}$$

《유제》  $a_{n+1} = \frac{3a_n-1}{a_n+1}$ ,  $a_1=0$ 인  $a_n$ 을 구하여라. ☞  $a_n = \frac{n-1}{n-3}$

❖ 참고 : **보기 1**과 **보기 2**에서 특성방정식과 그래프를 생각하여 보고 또한 극한값과의 관계도 조사해 보아라.

**보기 3.**  $a_1 > 1$ ,  $a_{n+1} = \frac{1}{2} \left( a_n + \frac{1}{a_n} \right)$ 인 수열의 극한을 구하여라.

**풀이** 특성방정식  $x = \frac{1}{2} \left( x + \frac{1}{x} \right)$ 의 근  $x = \pm 1$ 을 원식의 양변에서 빼면 다음과 같다.

$$a_{n+1} + 1 = \frac{1}{2} \left( a_n + \frac{1}{a_n} \right) + 1 = \frac{1}{2} \frac{(a_n+1)^2}{a_n} \quad \text{.....①}$$

$$a_{n+1} - 1 = \frac{1}{2} \left( a_n + \frac{1}{a_n} \right) - 1 = \frac{1}{2} \frac{(a_n-1)^2}{a_n} \quad \text{.....②}$$

$$\text{①} \div \text{②} : \frac{a_{n+1}+1}{a_{n+1}-1} = \left( \frac{a_n+1}{a_n-1} \right)^2 \quad \text{.....③}$$

지금부터  $a_n > 1$ 을 수학적 귀납법으로 증명하자.

$n=1$ 에서  $a_1 > 1$ 이고,  $n=k$ 에서  $a_k > 1$ 이라고 가정하자.

$$n=k+1 \text{에서 } a_{k+1} = \frac{1}{2} \left( a_k + \frac{1}{a_k} \right) \geq \sqrt{a_k \cdot \frac{1}{a_k}} = 1$$

(이 때 등호가 성립하지 않는 것은 상승, 상가평균에 있어서 등호가 성립하려면  $a_k = \frac{1}{a_k}$ , 즉  $a_k = \pm 1$ 일 때이나 여기서는  $a_k > 1$ 이기 때문이다.) 따라서 모든 자연수  $n$ 에 대해  $a_n > 1$ 이다.

$\therefore \frac{a_n+1}{a_n-1} > 0$ 이므로 ③의 양변에  $\log$ 를 취하여  $b_n = \log \left( \frac{a_n+1}{a_n-1} \right)$ 이라 놓으면 다음 식이 성립한다.

$$b_{n+1} = 2b_n$$

$$\therefore b_n = 2^{n-1}b_1$$

$$\therefore a_n = \frac{10^{b_n} + 1}{10^{b_n} - 1} = \frac{1 + \left( \frac{1}{10} \right)^{b_n}}{1 - \left( \frac{1}{10} \right)^{b_n}} \quad \left( \text{단, } b_n = 2^{n-1} \log \left( \frac{1+a_1}{1-a_1} \right) \right)$$





- (3)  $\lim_{n \rightarrow \infty} x_n$ 과  $\lim_{n \rightarrow \infty} y_n$ 을 구하라.
8.  $c_1=2$ ,  $c_{n+1}=\frac{2c_n+3}{c_n+2}$ 인  $c_n$ 을 기약분수로 표시하면  $\frac{a_n}{b_n}$ 이다. (단  $b_n > 0$ )
- (1)  $n \geq 1$ 에 대하여  $a_{n+1}$ ,  $b_{n+1}$ 을  $a_n$ 과  $b_n$ 으로 표시하여라.
  - (2)  $n > 1$ 에 대하여  $a_n + \sqrt{3}b_n$ ,  $a_n - \sqrt{3}b_n$ 을 구하여라.
  - (3)  $n > 1$ 에 대하여  $c_n$ 을 구하고  $\lim_{n \rightarrow \infty} c_n$ 을 구하여라.
9.  $x_1=4$ ,  $x_{n+1}=\frac{5x_n-9}{x_n-1}$ 이고,  $y_n=\frac{1}{x_n-3}$ 이다.
- (1)  $y_{n+1}$ 을  $y_n$ 으로 나타내어라.
  - (2)  $y_n$ 을 구하여라.
  - (3)  $x_n$ 을 구하고  $\lim_{n \rightarrow \infty} x_n$ 을 구하여라.
10.  $a_{n+1}=\frac{1-a_n}{1+a_n}$ ,  $a_1=a$  (단,  $a \neq -1$ )인 수열  $\{a_n\}$ 에서 수렴 또는 발산을 조사하고 수렴하는 경우 극한값을 구하여라.
11.  $a_1 > \sqrt{k}$  (단,  $k > 0$ ),  $a_{n+1}=\frac{1}{2}\left(a_n + \frac{k}{a_n}\right)$ 인 수열  $\{a_n\}$ 에서 다음을 증명하여라.
- (1)  $a_n > \sqrt{k}$
  - (2)  $a_n - \sqrt{k} < \frac{1}{2}(a_{n-1} - \sqrt{k})$
  - (3)  $\lim_{n \rightarrow \infty} a_n = \sqrt{k}$
12.  $x_0 > \sqrt{a}$ ,  $x_{n+1}=\frac{1}{2}\left(\frac{a}{x_n} + x_n\right)$  ( $n=0, 1, 2, \dots$ )이다.
- (1)  $u_{n+1}=a-x_{n+1}^2$ 을  $x_n$ 의 식으로 나타내고 부호를 조사하여라.
  - (2)  $n \geq 1$  일 때  $\left|\frac{u_{n+1}}{u_n}\right|$ 과  $\frac{1}{2}$ 의 대소를 비교하여라.
  - (3)  $u_n$ 의 수렴함을 증명하고  $\lim_{n \rightarrow \infty} u_n$ 을 구하여라.
  - (4)  $x_n$ 의 극한값을 구하여라.
13.  $a_1=a$  ( $a \neq 1$ ),  $a_{n+1}=\frac{n(n+1)}{2n-a_n}$  ( $n=1, 2, \dots$ )인 수열  $\{a_n\}$ 에서 일반항을 구하여라.

## 연습문제 해답

## A.2 연습문제 I

$$1. (1) a_n = \frac{1}{6}(n-1)n(2n-1)$$

$$(2) a_n = n! + 1$$

$$(3) a_n = \frac{2}{3} \left\{ 1 - \left(-\frac{1}{2}\right)^n \right\}$$

$$(4) a_n = \frac{5}{3} - \frac{2n+3}{3^n}$$

$$(5) a_n = \sqrt{n}$$

$$2. (1) a_n = 1 + \sum_{k=1}^{n-1} (k+1) = \frac{n(n+1)}{2}$$

$$(2) \sum_{k=1}^n \frac{1}{a_k} = \sum_{k=1}^n \frac{2}{k(k+1)} = \sum_{k=1}^n \left( \frac{2}{k} - \frac{2}{k+1} \right) = 2 \left( 1 - \frac{1}{n+1} \right)$$

$$(3) \sum_{k=1}^{\infty} \frac{1}{a_k} = \lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{2}{k(k+1)} = 2$$

$$3. (1) b_n = b_1 + (a_1 + b_1)(n-1) = (2n-1)a_1 - n, \quad (\because b_1 = a_1 - a_0 = a_1 - 1)$$

$$a_{k+1} - a_k = b_{k+1}$$

$$\therefore a_n = a_1 + \sum_{k=1}^{n-1} b_{k+1} = n^2 a_1 - \frac{(n+2)(n-1)}{2}$$

$$(2) \lim_{n \rightarrow \infty} \frac{1}{n^3} \sum_{k=0}^n a_k = \lim_{n \rightarrow \infty} \frac{1}{n^3} \sum_{k=0}^n \left\{ k^2 a_1 - \frac{(k+2)(k+1)}{2} \right\} = \frac{1}{3} \left( a_1 - \frac{1}{2} \right)$$

$$\therefore \frac{1}{3} \left( a_1 - \frac{1}{2} \right) = 1 \quad \therefore a_1 = \frac{7}{2}$$

$$4. x^2 + 2a_i x + a_i^2 = \frac{9}{4}, \quad \text{즉 } (x+a_i)^2 = \left(\frac{3}{2}\right)^2 \text{의 2근을 } a_i, \beta_i \text{라고 하자. (단, } a_i < \beta_i)$$

$$a_i = -a_i - \frac{3}{2}, \quad \beta_i = -a_i + \frac{3}{2}$$

$$(1) a_{i+1} - \beta_i = \left(-a_{i+1} - \frac{3}{2}\right) - \left(-a_i + \frac{3}{2}\right) = a_i - a_{i+1} - 3 = \frac{10}{3} - 3 = \frac{1}{3} > 0$$

따라서  $a_{i+1} > \beta_i$ 이고, 또한  $a_i < \beta_i$ 이므로  $a_i < \beta_i < a_{i+1}$ 이다. 즉, 문제의 수열은 다음과 같다.

$$a_1, \beta_1, a_2, \beta_2, a_3, \beta_3, \dots, a_n, \beta_n, \dots$$

또한 근과 계수와의 관계에서  $a_i + \beta_i = -2a_i$ 이고, 제의에서 수열  $\{a_n\}$ 은 초항이  $a_1$ 이고 공차가  $-\frac{10}{3}$ 인 등차수열이다. 따라서  $S_n$ 은 다음 식들을 만족한다.

$$\begin{aligned} n=2m : S_{2m} &= \sum_{i=1}^m (a_i + \beta_i) = -2 \sum_{i=1}^m a_i = -m \times \left\{ 2a_1 + (m-1) \times \left( -\frac{10}{3} \right) \right\} \\ &= -\frac{n}{2} \left\{ 2a_1 - \frac{10}{3} \times \left( \frac{n}{2} - 1 \right) \right\} = -\frac{n}{6} (5n - 6a_1 - 10) \end{aligned}$$

$$\begin{aligned} n=2m-1 : S_{2m-1} &= S_{2m} - \beta_m \\ &= -m \left\{ 2a_1 + (m-1) \times \left( -\frac{10}{3} \right) \right\} - \left( \frac{10}{3}m - a_1 - \frac{11}{6} \right) \\ &\quad \left[ \because \beta_m = -a_m + \frac{3}{2} = - \left\{ a_1 + (m-1) \times \left( -\frac{10}{3} \right) \right\} + \frac{3}{2} \right] \\ &= -\frac{n}{6} (5n - 6a_1 - 10) - \frac{2}{3} \quad (\because n=2m-1) \end{aligned}$$

$$\text{두 개를 합성하면 } S_n = \frac{n}{6} (5n - 6a_1 - 10) + \frac{1 - (-1)^n}{2} \times \left( -\frac{2}{3} \right)$$

❖ 참고 :  $a_1, b_2, a_3, b_4, a_5, b_6, a_7, b_8, \dots$ 인 수열의 일반항은 다음과 같다.

$$\frac{1 - (-1)^n}{2} a_n + \frac{1 + (-1)^n}{2} b_n$$

$$(2) \lim_{n \rightarrow \infty} \frac{S_n}{n^2} = \frac{5}{6}$$

## A.2 연습문제 II

1. (a)  $p \neq 1$ 인 경우 정리 2-1에서  $a_n = p^{n-1} \left( a_1 - \frac{a}{1-p} \right) + \frac{q}{1-p}$ 이다. 따라서  $p$ 값에 따라 다음과 같은 성질을 갖는다.

$$|p| < 1, \lim_{n \rightarrow \infty} p^{n-1} = 0 \quad \therefore \lim_{n \rightarrow \infty} a_n = \frac{q}{1-p}$$

$$p = -1, \lim_{n \rightarrow \infty} p^{n-1} \text{은 } \pm 1 \text{로 진동} \quad \therefore \lim_{n \rightarrow \infty} a_n \text{도 진동}$$

$$|p| > 1, \lim_{n \rightarrow \infty} p^{n-1} \text{은 발산} \quad \therefore \lim_{n \rightarrow \infty} a_n \text{도 발산}$$

- (b)  $p$ 값에 관계없이  $a_1 = \frac{q}{1-p}$ 이면  $\lim_{n \rightarrow \infty} a_n = \frac{q}{1-p}$ 이다.

- (c)  $p=1$ 인 경우

$$a_{n+1} = a_n + q \quad \therefore a_n = a_1 + (n-1)q$$

$$q=0 \text{이면 } \lim_{n \rightarrow \infty} a_n = a_1$$

$$q > 0 \text{ 이면 } \lim_{n \rightarrow \infty} a_n = \infty, \quad q < 0 \text{ 이면 } \lim_{n \rightarrow \infty} a_n = -\infty$$

(a), (b), (c)에서 수렴조건과 극한값은 다음과 같다.

$$|p| < 1 \text{에서 극한값 } \frac{q}{1-p}$$

$$a_1 = \frac{q}{1-p} \text{에서 극한값 } \frac{q}{1-p}$$

$$p=1 \text{이고 } q=0 \text{이면 극한값 } a_1$$

2. (1) 방정식  $a_{n-1}x^2 - ax_n + 1 = 0$ 이 두 근을 갖고 있으므로  $a_{n-1} \neq 0$ 이다. 또한 근과 계수

와의 관계에서  $a_n + \beta_n = \frac{a_n}{a_{n-1}}$ ,  $a_n\beta_n = \frac{1}{a_{n-1}}$ 이므로 다음 식이 성립한다.

$$3(a_n + \beta_n) + a_n\beta_n = \frac{3a_n}{a_{n+1}} + \frac{1}{a_{n-1}} = 1$$

$$\therefore a_n = \frac{1}{3}a_{n-1} + \frac{1}{3}$$

$$(2) \text{ 따라서 } a_n = \left(\frac{1}{3}\right)^{n-1} \left(a_1 - \frac{1}{2}\right) + \frac{1}{2}$$

$$(3) \lim_{n \rightarrow \infty} a_n = \frac{1}{2}$$

3. (가)  $k \neq 1$ 인 경우

$$a_n = \frac{1-k^{n-1}}{1-k} + \frac{1 + (\sqrt{3}-1)k^{n-2} - \sqrt{3}k^{n-1}}{1-k} i$$

따라서 편각의  $\tan$ 는  $\tan\theta_n = \frac{1 + (\sqrt{3}-1)k^{n-2} - \sqrt{3}k^{n-1}}{1-k^{n-1}}$ 이다.

$$\therefore \lim_{n \rightarrow \infty} \tan\theta_n = \begin{cases} 1 & (0 < k < 1) \\ \frac{\sqrt{3}k - \sqrt{3} + 1}{k} & (k > 1) \end{cases}$$

(나)  $k=1$ 이면  $a_{n+1} = a_n + a_1$ 이므로 다음 식이 성립한다.

$$\begin{aligned} a_n &= (n-2)a_1 + a_2 \\ &= (n-1) + (n-2 + \sqrt{3})i \end{aligned}$$

$$\therefore \tan\theta_n = \frac{n-2 + \sqrt{3}}{n-1}$$

$$\therefore \lim_{n \rightarrow \infty} \tan\theta_n = 1$$

$$4. (1) \tan(2x_{n+1}) = \frac{1}{\tan x_n} = \tan\left(\frac{1}{2}\pi - x_n\right)$$

$$\therefore 2x_{n+1} = \frac{1}{2}\pi - x_n + l\pi \quad (l \text{은 정수})$$

이 때  $\frac{\pi}{3} < x_n < \frac{2}{3}\pi$ 이므로  $2x_{n+1} = \frac{3}{2}\pi - x_n$ 이다.

(2)  $x_n = \frac{\pi}{2} + \left(-\frac{1}{2}\right)^{n-1} \left(\alpha - \frac{\pi}{2}\right)$

(3)  $\lim_{n \rightarrow \infty} x_n = \frac{\pi}{2}$

❖ 참고 :  $\tan(2x_{n+1})\tan x_n = 1$ 에서  $\sin(2x_{n+1})\sin(x_n) = \cos(2x_{n+1})\cos(x_n)$   
 $\therefore \cos(2x_{n+1} + x_n) = 0$   
 $\therefore 2x_{n+1} + x_n = m\pi + \frac{\pi}{2}$ , ( $m$ 은 정수)

5.  $a_n = 3a_{n-1} + 2n + 2$  .....①

$(pn+1+q) = 3(pn+q) + 2n + 2$ , ( $p = -1$ ,  $q = -\frac{3}{2}$ ) .....②

①-②에서  $a_n - \{p(n+1) + q\} = b_n$ 이라면

$b_1 = a_1 - 2p - q = \frac{9}{2}$ 이고 다음 식이 성립한다. ....③

$b_{n+1} = 3b_n$

$\therefore b_n = 3^{n-1}b_1 = \frac{1}{2} \cdot 3^{n+1}$  .....④

③, ④ :  $a_n = b_n + p(n+1) + q = \frac{1}{2} \cdot 3^{n+1} - (n+1) - \frac{3}{2} = \frac{1}{2}(3^{n+1} - 2n - 5)$

$\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} \frac{3^{n+1}}{2} \left(1 - \frac{2n}{3^{n+1}} - \frac{5}{3^{n+1}}\right) = \infty$ , ( $\because \lim_{n \rightarrow \infty} \frac{2n}{3^{n+1}} = 0$ )

6. (1)  $c_n = r \cdot c_{n-1} - b \cdot s^{n-1}$  .....①

$p \cdot s^{n-1} = r(p \cdot s^{n-2}) - b \cdot s^{n-1}$ , ( $p = \frac{bs}{r-s}$ ) .....②

①-②에서  $c_n - p \cdot s^{n-1} = d_n$ 이라면  $d_1 = c_1 - p$ 이고

다음 식이 성립한다. ....③

$d_n = r \cdot d_{n-1}$

$\therefore d_n = r^{n-1}d_1 = r^{n-1}(c_1 - p)$  .....④

③, ④ :  $c_n = ps^{n-1} + d_n = \frac{bs^n}{r-s} + \left(a - b - \frac{bs}{r-s}\right)r^{n-1} = ar^{n-1} - \frac{b(s^n - r^n)}{s-r}$

(2)  $c_k = 0$ 이면  $ar^{k-1} = \frac{b(s^k - r^k)}{s-r}$ 의 양변을  $br^{k-1}$ 로 나누어 다음 식을 얻는다.

$\frac{a}{b} = \frac{1}{s-r} \left\{ s \left(\frac{s}{r}\right)^{k-1} - r \right\}$

이 식에  $\frac{a}{b} = y$ ,  $\frac{s}{r} = x$ , ( $s = rx$ )를 대입하면 다음 식을 얻는다.

$y = \frac{1}{r(x-1)} \left\{ rx(x)^{k-1} - r \right\} = \frac{x^k - 1}{x-1}$

7.  $a_{n+1} = -a_n + 2n^3 + 3n^2 + 3n - 3$ , ( $a_1 = 1$ ) .....①

$g(n) = \alpha n^3 + \beta n^2 + \gamma n + \delta$  중에서 다음을 만족하는 것을 구하자. ....②

$g(n+1) + g(n) = 2n^3 + 3n^2 + 3n - 3$  .....③

②, ③ :  $\alpha = 1$ ,  $\beta = 0$ ,  $\gamma = 0$ ,  $\delta = -2$

①-③에서  $a_n - g(n) = b_n$ 이라면,  $b_1 = a_1 - g(1) = 2$ 이고 다음 식이 성립한다. ....④

$$b_{n+1} = -b_n$$

$$\therefore b_n = (-1)^{n-1} b_1 = (-1)^{n-1} \times 2 \quad \dots\dots\dots ⑤$$

$$④, ⑤ : a_n = b_n + g(n) = 2 \times (-1)^{n-1} + n^3 - 2$$

8.  $a_{n+1} = 3a_n + 5 \cdot 3^n n(n+1)(n+2)(n+3)$ 의 양변을  $3^n$ 으로 나누면 다음 식이 성립한다.

$$\frac{a_{n+1}}{3^n} = \frac{a_n}{3^{n-1}} + 5n(n+1)(n+2)(n+3)$$

이 식에서  $\frac{a_n}{3^{n-1}} = b_n$ 이라면,  $b_1 = a_1 = 0$ 이고 다음 식이 성립한다. ....①

$$b_{n+1} = b_n + 5n(n+1)(n+2)(n+3)$$

$$\begin{aligned} \therefore b_n &= 0 + \sum_{k=1}^{n-1} 5k(k+1)(k+2)(k+3) \\ &= (n-1)n(n+1)(n+2)(n+3) \quad \dots\dots\dots ② \end{aligned}$$

$$①, ② : a_n = 3^{n-1} b_n = 3^{n-1} (n-1)n(n+1)(n+2)(n+3)$$

9.  $a_{n+1} - 2a_n = 3^{n-1} + 2n + 7$ , ( $a_1 = 0$ ) ....①

$g(n) = \alpha \cdot 3^{n-1} + \beta n + \gamma$  중 다음을 만족하는 것을 구하자. ....②

$$g(n+1) - 2g(n) = 3^{n-1} + 2n + 7 \quad \dots\dots\dots ③$$

②, ③에서  $\alpha = 1$ ,  $\beta = -2$ ,  $\gamma = -9$ ,  $g(1)$ 은  $\alpha + \beta + \gamma = -10$ 이다.

①-③에서  $a_n - g(n) = b_n$ 이라면,  $b_1 = a_1 - g(1) = 10$ 이고 다음 식이 성립한다. ....④

$$b_{n+1} = 2b_n$$

$$\therefore b_n = 2^{n-1} b_1 = 5 \cdot 2^n \quad \dots\dots\dots ⑤$$

$$④, ⑤ : a_n = b_n + g(n) = 5 \cdot 2^n + 3^{n-1} - 2n - 9$$

10.  $a_{n+1} - 3a_n = 3^n - 2n - 1$  ....①

$$\{\alpha(n+1) + \beta\} - 3(\alpha n + \beta) = -2n - 1, (\alpha = 1, \beta = 1) \quad \dots\dots\dots ②$$

①-②에서  $a_n - (\alpha n + \beta) = b_n$ 이라면,  $b_1 = a_1 - \alpha - \beta = -1$ 이고 다음 식이 성립한다. ....③

$$b_{n+1} = 3b_n + 3^n$$

이 식의 양변을  $3^n$ 으로 나누면 다음 식이 성립한다.

$$\frac{b_{n+1}}{3^n} = \frac{b_n}{3^{n-1}} + 1$$

이 식에서  $\frac{b_n}{3^{n-1}} = c_n$ 이라면,  $c_1 = b_1 = -1$ 이고 다음 식이 성립한다. ....④

$$c_{n+1} = c_n + 1$$

$$\therefore c_n = c_1 + (n-1) = n-2 \quad \dots\dots\dots ⑤$$

$$③, ④, ⑤ : a_n = \alpha n + \beta + 3^{n-1} c_n = (n-2) \cdot 3^{n-1} + n + 1$$

11.  $a_{n+1} = 2a_n + (6n+5)(-1)^{n+1}$  ....①

$g(n) = (an + \beta)(-1)^n$ 인  $g(n)$  중 다음을 만족하는 것을 구하자.

$$g(n+1) = 2g(n) + (6n+5)(-1)^{n+1} \quad \dots\dots\dots ②$$

따라서  $\alpha=2, \beta=1, g(1) = -(\alpha+\beta) = -3$ 이다. .....③

①-②에서  $a_n - g(n) = b_n$ 이라면,

$$b_1 = a_1 - g(1) = 3 \text{이고 다음 식이 성립한다.} \quad \dots\dots\dots ④$$

$$b_{n+1} = 2b_n$$

$$\therefore b_n = 2^{n-1}b_1 = 3 \cdot 2^{n-1} \quad \dots\dots\dots ⑤$$

$$④, ⑤ : a_n = b_n + g(n) = 3 \cdot 2^{n-1} + (2n+1)(-1)^n$$

12.  $\frac{2n^2+7n+4}{n(n+1)(n+2)} = \frac{2(an+\beta)}{n(n+1)} - \frac{\alpha(n+1)+\beta}{(n+1)(n+2)}$  라고 하면,  $\alpha=2, \beta=1$ 이다. 따라서

$a_{n+1} = 2a_n + 2^n - \frac{2(2n+1)}{n(n+1)} + \frac{2(n+1)+1}{(n+1)(n+2)}$  에서  $b_n = a_n - \frac{2n+1}{n(n+1)}$  이라고 하면 다음 식이 성립한다.

$$b_{n+1} = 2b_n + 2^n$$

이 식에서 양변을  $2^n$ 으로 나누면 다음 식이 성립한다.

$$\frac{b_{n+1}}{2^n} = \frac{b_n}{2^{n-1}} + 1$$

이 식의  $\frac{b_n}{2^{n-1}} = c_n$ 이라 하면,  $c_1 = b_1 = a_1 - \frac{3}{2} = \frac{1}{2}$ 이고 다음 식이 성립한다.

$$c_{n+1} = c_n + 1$$

$$\therefore c_n = c_1 + (n-1) = n - \frac{1}{2}$$

$$\therefore a_n = b_n + \frac{2n+1}{n(n+1)} = 2^{n-1}c_n + \frac{2n+1}{n(n+1)} = 2^{n-2}(2n-1) + \frac{(2n+1)}{n(n+1)}$$

13. (1)  $2a_{n+1} = a_n + (-1)^{n+1}$  .....①

$$2x(-1)^{n+1} = x \cdot (-1)^n + (-1)^{n+1}, \left(x = \frac{1}{3}\right) \quad \dots\dots\dots ②$$

①-②에서  $a_n - x(-1)^n = b_n$ 이라면,  $b_1 = a_1 + x = \frac{4}{3}$ 이고 다음 식이 성립한다.

$$\dots\dots\dots ③$$

$$b_{n+1} = \frac{1}{2}b_n$$

$$\therefore b_n = \left(\frac{1}{2}\right)^{n-1} b_1 = \frac{1}{3} \left(\frac{1}{2}\right)^{n-3} \quad \dots\dots\dots ④$$

$$③, ④ : a_n = b_n + x(-1)^n = \frac{1}{3} \left(\frac{1}{2}\right)^{n-3} + \frac{1}{3}(-1)^n = \frac{1}{3} \left\{ \left(\frac{1}{2}\right)^{n-3} + (-1)^n \right\}$$

$$(2) \sum_{k=1}^{2n} a_k = \frac{8}{3} \left\{ 1 - \left(\frac{1}{2}\right)^{2n} \right\} \therefore \lim_{n \rightarrow \infty} \sum_{k=1}^{2n} a_k = \frac{8}{3}$$

$$(3) \sum_{k=1}^{2n+1} a_k = \frac{8}{3} \left\{ 1 - \left(\frac{1}{2}\right)^{2n+1} \right\} - \frac{1}{3} \therefore \lim_{n \rightarrow \infty} \sum_{k=1}^{2n+1} a_k = \frac{7}{3}$$



══════ A.2 연습문제 III ══════

1.  $a_{n+1}=2^n a_n=2^{\frac{1}{2}n(n+1)-\frac{1}{2}n(n-1)}$   $a_n$ 의 양변을  $2^{\frac{1}{2}n(n+1)}$ 로 나누면 다음 식이 성립한다.

$$\frac{a_{n+1}}{(\sqrt{2})^{n(n+1)}} = \frac{a_n}{(\sqrt{2})^{n(n-1)}}$$

이 식에서  $b_n = \frac{a_n}{(\sqrt{2})^{n(n-1)}}$ 이라 하면,  $b_1 = a_1 = 1$ 이고 다음 식이 성립한다.

$$b_{n+1} = b_n \quad \dots\dots\dots ①$$

$$\therefore b_n = b_1 = 1 \quad \dots\dots\dots ②$$

$$①, ② : a_n = b_n (\sqrt{2})^{n(n-1)} = (\sqrt{2})^{n(n-1)} = 2^{\frac{1}{2}n(n-1)}$$

❖ 참고 : 양변에 밑이 2인 log를 취하여  $\log_2 a_n = b_n$ 이라하면  $b_{n+1} = b_n + n$ 으로 해서 구할 수 있다.

2.  $n^2 a_n = (n-1)^2 a_{n-1} + n(n-1)$ 에서  $b_n = n^2 a_n$ 이라하면,  $b_1 = a_1 = 0$ 이고 다음 식이 성립한다.

$$b_n = b_{n-1} + n(n-1) \quad \dots\dots\dots ①$$

$$\therefore b_n = 0 + \sum_{k=1}^{n-1} k(k+1) = \frac{1}{3}(n-1)n(n+1) \quad \dots\dots\dots ②$$

$$①, ② : a_n = \frac{b_n}{n^2} = \frac{n^2-1}{3n}$$

3. 원식의 양변에  $\frac{n+1}{2n+3}$ 을 곱하면 다음 식을 얻는다.

$$\frac{n+1}{2n+3} a_{n+1} = -\frac{n}{2n+1} a_n + (-1)^n$$

$$\text{이 식에서 } \frac{n}{2n+1} a_n = b_n \text{이라하면 } b_{n+1} = -b_n + (-1)^n \text{이다.} \quad \dots\dots\dots ①$$

이 식에서 양변을  $(-1)^n$ 으로 나누면 다음 식을 얻는다.

$$\frac{b_{n+1}}{(-1)^n} = \frac{b_n}{(-1)^{n-1}} + 1$$

$$\text{이 식에서 } \frac{b_n}{(-1)^{n-1}} = c_n \text{이라하면, } c_1 = b_1 = \frac{1}{3} a_1 = 1 \text{이고 다음 식이 성립한다.}$$

$$\dots\dots\dots ②$$

$$c_{n+1} = c_n + 1$$

$$\therefore c_n = c_1 + (n-1) = n \quad \dots\dots\dots ③$$

$$①, ②, ③ : a_n = \frac{2n+1}{n} b_n = \frac{2n+1}{n} (-1)^{n-1} c_n = (2n+1) (-1)^{n-1}$$

4.  $(n+1)a_{n+1} = 2a_n + n!(n^2+2n-1)$ 의 양변에  $n!$ 을 곱하면 다음 식이 성립한다.

$$(n+1)! a_{n+1} = 2(n! a_n) + n! \cdot n! \{(n+1)^2 - 2\}$$

$$= 2n! a_n + \{(n+1)!\}^2 - 2(n!)^2$$

$$\therefore (n+1)! a_{n+1} - \{(n+1)!\}^2 = 2\{n! a_n - (n!)^2\}$$

이 식에서  $n! a_n - (n!)^2 = b_n$ 이라면 다음 식을 얻는다. .....①

$$b_{n+1} = 2b_n$$

$$\therefore b_n = 2^{n-1} b_1 = 2^{n-1} (a_1 - 1) = 2^{n-1} \quad \text{.....②}$$

$$\text{①, ②} : a_n = \frac{1}{n!} \{b_n + (n!)^2\} = \frac{1}{n!} \{2^{n-1} + (n!)^2\}$$

5.  $a_{n+1} = \frac{-n(2n+3)}{(n+2)(2n+1)} a_n = \frac{-n(n+1)(2n+3)}{(n+1)(n+2)(2n+1)} a_n$

이 식의 양변에  $\frac{(n+1)(n+2)}{2n+3}$ 를 곱한 뒤  $a_n \frac{n(n+1)}{2n+1} = b_n$ 이라면 다음 식이 성립한다.

$$b_{n+1} = -b_n \quad \text{.....①}$$

$$\therefore b_n = (-1)^{n-1} b_1 = (-1)^{n-1}, (\because b_1 = \frac{2}{3} a_1 = 1) \quad \text{.....②}$$

$$\text{①, ②} : a_n = \frac{2n+1}{n(n+1)} b_n = (-1)^{n-1} \frac{2n+1}{n(n+1)}$$

$$S_n = \sum_{k=1}^n a_k = \sum_{k=1}^n (-1)^{k-1} \left( \frac{1}{k} + \frac{1}{k+1} \right) = \sum_{k=1}^n \left\{ (-1)^{k-1} \frac{1}{k} - (-1)^k \frac{1}{k+1} \right\} = 1 - \frac{(-1)^n}{n+1}$$

6. 식  $(n-1)a_{n+1} = (n+1)a_n - 2$ 의 양변을  $(n-1)n(n+1)$ 로 나누면 다음 식이 성립한다.

$$\frac{a_{n+1}}{n(n+1)} = \frac{a_n}{n(n-1)} - \frac{2}{n(n+1)(n-1)} = \frac{a_n}{n(n-1)} - \left\{ \frac{1}{n(n-1)} - \frac{1}{n(n+1)} \right\}$$

$$\therefore \frac{a_{n+1}-1}{n(n+1)} = \frac{a_n-1}{n(n-1)}$$

이 식에서  $\frac{a_n-1}{n(n-1)} = b_n$ 이라면 다음 식이 성립한다. .....①

$$b_{n+1} = b_n$$

$$\therefore b_n = b_2 = 4, (\because b_2 = \frac{a_2-1}{2} = 4) \quad \text{.....②}$$

$$\text{①, ②} : a_n = n(n-1) b_n + 1 = 4n(n-1) + 1 = (2n-1)^2$$

7.  $n \geq 2$ 인 경우  $S_n - S_{n-1} = n^2 a_n - (n-1)^2 a_{n-1} = a_n$ 이다. 따라서  $(n^2-1)a_n = (n-1)$

$(n+1)a_n = (n-1)^2 a_{n-1}$ 의 양변을  $n-1$ 로 나누어 다음 식을 얻는다.

$$(n+1)a_n = (n-1)a_{n-1}$$

이 식의 양변에  $n$ 을 곱하여  $n(n+1)a_n = b_n$ 이라면 다음 식이 성립한다.

$$b_n = b_{n-1} \quad \text{.....①}$$

$$\therefore b_n = b_2 = 6a_2 = 2, (\because 4a_2 = a_1 + a_2 \text{에서 } a_2 = \frac{1}{3}) \quad \text{.....②}$$

$$\text{①, ②} : a_n = \frac{b_n}{n(n+1)} = \frac{2}{n(n+1)}$$

$n=1$ 이면  $a_1=1$ 이므로 모든 자연수  $n$ 에 대해  $a_n = \frac{2}{n(n+1)} = \frac{2}{n} - \frac{2}{n+1}$ 이다.

$$S_n = \sum_{k=1}^n \left( \frac{2}{k} - \frac{2}{k+1} \right) = 2 - \frac{2}{n+1} = \frac{2n}{n+1}$$

$$\lim_{n \rightarrow \infty} S_n = 2$$

### A.3 연습문제

#### 1. [보기 1]의 유형

(1) 방정식  $t^2 - (\cos^2 \alpha)t - \sin^2 \alpha = 0$ 에서  $t=1$ ,  $t = -\sin^2 \alpha$ 이다.

$$x_{n+2} - x_{n+1} = (-\sin^2 \alpha)(x_{n+1} - x_n)$$

$$\therefore x_{n+1} - x_n = (-\sin^2 \alpha)^{n-1}(x_2 - x_1) = (-\sin^2 \alpha)^{n-1}$$

$$\therefore x_n = 1 + \sum_{k=1}^{n-1} (-\sin^2 \alpha)^{k-1} = 1 + \frac{1 - (-\sin^2 \alpha)^{n-1}}{1 + \sin^2 \alpha}$$

(2)  $\lim_{n \rightarrow \infty} x_n = 1 + \frac{1}{1 + \sin^2 \alpha}$ , ( $\because 0 < \alpha < \frac{\pi}{2}$ 에서  $-1 < -\sin^2 \alpha < 0$ )

#### 2. [보기 1]의 유형

(1) 방정식  $t^2 - \frac{4}{3}t + \frac{1}{3} = 0$ 의 근은  $t=1$ ,  $t = \frac{1}{3}$ 이다.

$$a_{n+2} - a_{n+1} = \left(\frac{1}{3}\right)(a_{n+1} - a_n)$$

$$\therefore a_{n+1} - a_n = \left(\frac{1}{3}\right)^{n-1}(a_2 - a_1) = -\frac{4}{3^n}$$

$$\therefore a_n = 3 + \sum_{k=1}^{n-1} \left(-\frac{4}{3^k}\right) = 1 + 2 \times \left(\frac{1}{3}\right)^{n-1}$$

(2)  $a_k - 1 = 2 \times \left(\frac{1}{3}\right)^{k-1}$

$$\therefore \log(a_k - 1) = \log 2 - (k-1)\log 3$$

$$\therefore S_n = \sum_{k=1}^n \frac{1}{\{\log 2 - (k-1)\log 3\}\{\log 2 - k\log 3\}}$$

$$= \frac{1}{\log 3} \sum_{k=1}^n \left\{ \frac{1}{-\log 2 + (k-1)\log 3} - \frac{1}{-\log 2 + k\log 3} \right\}$$

$$= \frac{1}{\log 3} \left( -\frac{1}{\log 2} - \frac{1}{-\log 2 + n\log 3} \right)$$

$$\therefore \lim_{n \rightarrow \infty} S_n = \frac{1}{-\log 2 \cdot \log 3}$$

3. 주어진 식을 근의 공식에 넣어  $c_n = c_{n+1} - c_{n-1}$ 을 적용시키면 다음과 같다.

$$x = \frac{1}{2} \{ (1 \pm \sqrt{5})c_n + 2c_{n-1} \}$$

$$\therefore \alpha_n = \frac{1 + \sqrt{5}}{2} c_n + c_{n-1}, \quad \beta_n = \frac{1 - \sqrt{5}}{2} c_n + c_{n-1}$$

(1) 이 때  $\alpha_{n+1} = \frac{1 + \sqrt{5}}{2} c_{n+1} + c_n = \frac{1 + \sqrt{5}}{2} (c_n + c_{n-1}) + c_n$

$$= \frac{1+\sqrt{5}}{2} \left( \frac{1+\sqrt{5}}{2} c_n + c_{n-1} \right)$$

$$\therefore \frac{a_{n+1}}{a_n} = \frac{1+\sqrt{5}}{2}$$

같은 방법으로 다음 식을 얻을 수 있다.

$$\frac{\beta_{n+1}}{\beta_n} = \frac{1-\sqrt{5}}{2}$$

(2)  $n=1$ 이면  $x^2 - (c_2 + c_0)x + (c_2c_0 - c_1^2) = x^2 - x - 1 = 0$ 에서 특성근은 다음과 같다.

$$a_1 = \frac{1+\sqrt{5}}{2}, \quad \beta_1 = \frac{1-\sqrt{5}}{2}$$

또한 (1)에서  $a_{n+1} = \frac{1+\sqrt{5}}{2} a_n$ 이다.

$$\therefore a_n = \left( \frac{1+\sqrt{5}}{2} \right)^{n-1} a_1 = \left( \frac{1+\sqrt{5}}{2} \right)^n$$

같은 방법으로 다음 식을 얻는다.

$$\beta_{n+1} = \frac{1-\sqrt{5}}{2} \beta_n$$

$$\therefore \beta_n = \left( \frac{1-\sqrt{5}}{2} \right)^{n-1} \beta_1 = \left( \frac{1-\sqrt{5}}{2} \right)^n$$

#### 4. [보기 1]의 유형

(1)  $\theta(a_{n+1} - a_n) = \sin\theta(a_n - a_{n-1})$ 에서 다음 식이 성립한다.

$$(a_{n+1} - a_n) = \frac{\sin\theta}{\theta} (a_n - a_{n-1})$$

$$\therefore a_{n+1} - a_n = \left( \frac{\sin\theta}{\theta} \right)^n (a_1 - a_0)$$

따라서  $a_n = a_1 + \sum_{k=1}^{n-1} \left( \frac{\sin\theta}{\theta} \right)^k (a_1 - a_0)$ 이고 또한  $a_1 = \frac{\sin\theta}{\theta} a_0$ 이므로 다음 식이 성립한다.

$$a_n = \frac{\sin\theta}{\theta} a_0 + a_0 \frac{\sin\theta}{\theta} \left\{ \left( \frac{\sin\theta}{\theta} \right)^{n-1} - 1 \right\} = a_0 \left( \frac{\sin\theta}{\theta} \right)^n$$

(2) 수열  $\{a_n\}$ 은 초항  $a_0$ , 공비  $\frac{\sin\theta}{\theta}$ 인 무한등비수열이므로 다음 식이 성립한다.

$$\sum_{n=0}^{\infty} a_n = \frac{\theta a_0}{\theta - \sin\theta}$$

5. [보기 2]에서  $a_n = \frac{\beta^n - \alpha^n}{\beta - \alpha}$ , (단,  $\alpha = \frac{1-\sqrt{5}}{2}$ ,  $\beta = \frac{1+\sqrt{5}}{2}$ )이므로 다음 식이 성립한다.

$$\lim_{n \rightarrow \infty} \frac{a_n}{a_{n-1}} = \lim_{n \rightarrow \infty} \frac{\beta^n - \alpha^n}{\beta^{n-1} - \alpha^{n-1}} = \lim_{n \rightarrow \infty} \frac{\beta - \alpha \left( \frac{\alpha}{\beta} \right)^{n-1}}{1 - \left( \frac{\alpha}{\beta} \right)^{n-1}} = \beta = \frac{1+\sqrt{5}}{2}, \quad \left( \because -1 < \frac{\alpha}{\beta} < 0 \right)$$

❖ 참고 :  $a_{n+2} = a_{n+1} + a_n$ 의 양변을  $a_{n+1}$ 로 나누어  $\frac{a_{n+1}}{a_n} = b_n$ 이라면  $b_{n+1} = 1 + \frac{1}{b_n}$ 이 되어 제 A.7절의 분수점화식으로 귀착시킬 수도 있다.

### 6. [보기 5]의 유형

방정식  $t^2 + 1 = 0$ 에서  $t = i$ ,  $t = -i$ 이므로 원식  $(a_{n+2} + ia_{n+1}) - i(a_{n+1} + ia_n) = 1$ 에서  $a_{n+1} + ia_n = b_n$ 이라면,  $b_1 = 0$ 이고 다음 식이 성립한다.

$$b_n = \left(\frac{1+i}{2}\right)(1 - i^{n-1})$$

또한  $a_{n+1} + ia_n = \left(\frac{1+i}{2}\right)(1 - i^{n-1})$ 에서  $a_n$ 을 구하면 다음과 같다.

$$a_n = \frac{1}{4}(3i-1)(-i)^{n-1} + \frac{1}{4}(1+i)i^n + \frac{1}{2}$$

### 7. [보기 1]의 유형

$$(1) a_{n+1} - a_n = 2(a_n - a_{n-1})$$

$$\therefore a_{n+1} - a_n = 2^n(a_1 - a_0) = 2^n$$

$$\therefore a_n = 2 + \sum_{k=1}^{n-1} 2^k = 2^n$$

$$(2) S = \sum_{k=0}^n k \cdot x^k \cdot 2^k = \sum_{k=0}^n k(2x)^k$$

$$(가) x \neq \frac{1}{2}, S - 2xS = \frac{2x\{1 - (2x)^n\}}{1 - 2x} - n(2x)^{n+1}$$

$$\therefore S = \frac{2x\{1 - (2x)^n\}}{(1 - 2x)^2} - \frac{n(2x)^{n+1}}{1 - 2x}$$

$$(나) x = \frac{1}{2}, S = \sum_{k=0}^n k = \frac{1}{2}n(n+1)$$

### 8. [보기 1]의 유형

$x_{n+1} - x_n = \left(\frac{a}{b}\right)(x_n - x_{n-1})$ 이므로 다음 식이 성립한다.

$$x_{n+1} - x_n = \left(\frac{a}{b}\right)^n (x_1 - x_0) = x_0 \left(\frac{a}{b} - 1\right) \left(\frac{a}{b}\right)^n, (\because bx_1 = ax_0)$$

$$\therefore x_n = \left(\frac{a}{b}x_0\right) + \sum_{k=1}^{n-1} x_0 \left(\frac{a}{b} - 1\right) \left(\frac{a}{b}\right)^k = \left(\frac{a}{b}\right)^n x_0$$

이 때  $x_n$ 은 공비가  $\frac{a}{b}$ 인 등비수열이므로 다음 식이 성립한다.

$$\sum_{n=0}^{\infty} x_n = \frac{x_0}{1 - \left(\frac{a}{b}\right)} = 1$$

$$\therefore x_0 = \frac{b-a}{b}$$

$$\therefore x_n = \left(\frac{b-a}{b}\right) \left(\frac{a}{b}\right)^n$$

9. [보기] 1의 유형

(1) 방정식  $t^2 - \frac{1}{2}t - \frac{1}{2} = 0$ 에서  $t=1$ ,  $t=-\frac{1}{2}$ 이므로 원식

$a_{n+2} + \frac{1}{2}a_{n+1} = a_{n+1} + \frac{1}{2}a_n$ 에서  $a_{n+1} + \frac{1}{2}a_n = b_n$ 이라면 다음 식이 성립한다.

$$b_{n+1} = b_n$$

$$\therefore b_n = b_1 = a_2 + \frac{1}{2}a_1 = \frac{a + \beta + \gamma}{2} = \frac{\pi}{2}$$

$$\therefore 2a_{n+1} + a_n = \pi$$

(2)  $(\sin a_n + \sin a_{n+1}) \cos a_{n+2}$

$$= 2\sin\left(\frac{a_n + a_{n+1}}{2}\right) \cos\left(\frac{a_n - a_{n+1}}{2}\right) \cos\left(\frac{a_n + a_{n+1}}{2}\right)$$

$$= \sin(a_n + a_{n+1}) \cos\left(\frac{a_n - a_{n+1}}{2}\right)$$

$$= \sin(\pi - a_{n+1}) \cos\left(\frac{a_n - a_{n+1}}{2}\right)$$

$$\leq \sin(\pi - a_{n+1}) = \sin(a_{n+1})$$

$$\left(\because \cos\left(\frac{a_n - a_{n+1}}{2}\right) \leq 1, 0 < a_n < \pi \text{에서 } \sin(\pi - a_{n+1}) > 0\right)$$

(3)  $-\pi < a_n - a_{n+1} < \pi$ 이다. 또한 (2)에서 등호가 성립하려면  $\cos\left(\frac{a_n - a_{n+1}}{2}\right) = 1$ 이다.

$$\text{즉, } \frac{a_n - a_{n+1}}{2} = 2m\pi, \quad (m \text{은 정수}) \quad \dots\dots\dots \textcircled{1}$$

$$\textcircled{1}, \textcircled{2} : a_n - a_{n+1} = 0 \quad \dots\dots\dots \textcircled{2}$$

$$\therefore a_n = a_{n+1}$$

따라서  $a_n$ 은 일정하여야 한다.

10. 수열  $\{c_n\}$ 이 등차수열  $\{a_n\}$ 과 등비수열  $\{b_n\}$ 의 합으로 되었다면 다음 식들이 성립한다.

$$c_n = a_n + b_n$$

$$c_{n+1} = a_{n+1} + b_{n+1} = (a_n + d) + rb_n$$

$$c_{n+2} = a_{n+2} + b_{n+2} = (a_n + 2d) + r^2b_n$$

$$\therefore c_{n+2} - (1+r)c_{n+1} + rc_n$$

$$= \{(a_n + 2d) + r^2b_n\} - (1+r)\{(a_n + d) + rb_n\} + r(a_n + b_n)$$

$$= d(1-r)$$

역으로  $c_{n+2} - (1+r)c_{n+1} + rc_n = d(1-r)$ 이므로  $(c_{n+2} - c_{n+1}) - r(c_{n+1} - c_n) = d$

$$(1-r) \text{에서 } c_{n+1} - c_n = x_n \text{이라면 다음 식이 성립한다.} \quad \dots\dots\dots \textcircled{1}$$

$$x_{n+1} - rx_n = d(1-r), \quad (x_1 = c_2 - c_1) \quad \dots\dots\dots \textcircled{2}$$

$$p - rp = d(1-r), \quad (p = d) \quad \dots\dots\dots \textcircled{3}$$

$\textcircled{2} - \textcircled{3}$ 에서  $x_n - p = y_n$ 이라면,

$$y_1 = x_1 - p = c_2 - c_1 - d \text{이고 다음 식이 성립한다.} \quad \dots\dots\dots \textcircled{4}$$

$$y_{n+1} = ry_n$$

$$\therefore y_n = r^{n-1}y_1 \quad \dots\dots\dots ⑤$$

$$①, ④, ⑤ : c_{n+1} - c_n = x_n = y_n + p = r^{n-1}y_1 + d$$

$$\begin{aligned} \therefore c_n &= c_1 + \sum_{k=1}^{n-1} (y_1 r^{k-1} + d) \\ &= c_1 + y_1 \times \frac{r^{n-1} - 1}{r - 1} + d(n-1) \\ &= \frac{c_2 - rc_1 - d}{1 - r} + d(n-1) + \frac{c_2 - c_1 - d}{r - 1} r^{n-1} \end{aligned}$$

따라서  $c_n$ 은  $\frac{c_2 - rc_1 - d}{1 - r}$ 를 초항으로 하고 공차를  $d$ 로 하는 등차수열의 제  $n$ 항과 초항이  $\frac{c_1 - c_2 + d}{1 - r}$ 이고 공비  $r$ 인 등비수열의 제  $n$ 항의 합이다.

11. 원식이  $\beta(n+1x_{n+1} - nx_n) = \alpha(x_n - x_{n-1})$ 이므로

$$\begin{aligned} \beta\{nx_n - (n-1)x_{n-1}\} &= \alpha(x_{n-1} - x_{n-2}) \\ \beta\{(n-1)x_{n-1} - (n-2)x_{n-2}\} &= \alpha(x_{n-2} - x_{n-3}) \\ \beta\{(n-2)x_{n-2} - (n-3)x_{n-3}\} &= \alpha(x_{n-3} - x_{n-4}) \\ &\vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \\ \beta(3x_3 - 2x_2) &= \alpha(x_2 - x_1) \\ +) \beta(2x_2 - x_1) &= \alpha(x_1 - x_0) \\ \hline \beta(nx_n - x_1) &= \alpha(x_{n-1} - x_0) \end{aligned}$$

$$\therefore n\beta x_n = \alpha x_{n-1} + \beta x_1 - \alpha x_0 = \alpha x_{n-1}, \quad (\because \beta x_1 - \alpha x_0 = 0)$$

이 식의 양변  $\times (n-1)!$  한 뒤,  $n! x_n = y_n$ 이라 두면,  $y_0 = x_0$ 이고 다음 식이 성립한다. .....①

$$\beta n! x_n = \alpha(n-1)! x_{n-1}$$

$$\therefore y_n = \left(\frac{\alpha}{\beta}\right)^n y_0 = \left(\frac{\alpha}{\beta}\right)^n x_0 \quad \dots\dots\dots ②$$

$$①, ② : x_n = \frac{y_n}{n!} = \frac{1}{n!} \left(\frac{\alpha}{\beta}\right)^n x_0$$

12~14. 인접 3항간의 점화식  $a_{n+2} - pa_{n+1} + qa_n = f(n)$ 을 방정식  $t^2 - pt + q = 0$ 의 2근을 구하여 2단계의 2항간의 점화식으로 유도하여 푸는 대신에  $g(n+2) - pg(n+1) + qg(n) = f(n)$ 이 되는  $g(n)$ 을 먼저 구하여  $a_n - g(n) = b_n$ 이라 두면  $b_{n+2} - pb_{n+1} + qb_n = 0$ 이 되므로 여기서  $b_n$ 을 구하여  $a_n = b_n + g(n)$ 으로 풀 수 있다. 여기서는  $f(n) = s^n + tn + u$ 인 경우의 예를 세 개 들었다.  
\* 앞에서 풀 방법과 비교하여 보아라.

12. [ (type 1)  $(\alpha, \beta, s, 1 \neq)$ 인 경우 ]  
[  $g(n) = xn + y + zs^n$ 으로 둔다. ]

$$a_{n+2} - 9a_n = 2 - 8n - 5 \cdot 2^n \quad \dots\dots\dots ①$$

$$\{x(n+2) + y + z \cdot 2^{n+2}\} - 9(xn + y + z \cdot 2^n) = 2 - 8n - 5 \cdot 2^n \quad \dots\dots\dots ②$$

②에서 항등식의 원리를 적용하면  $x=1, y=0, z=1$ 이다.

①-②에서  $a_n - (xn + y + z \cdot 2^n) = b_n$ 이라면,

$$b_1 = -6, b_2 = 36 \text{이고 다음 식이 성립한다.} \quad \dots\dots\dots ③$$

$$b_{n+2} - 9b_n = 0$$

$$\text{즉, } b_{n+2} - 3b_{n+1} = (-3)(b_{n+1} - 3b_n)$$

$$\therefore b_{n+1} - 3b_n = (-3)^{n-1}(b_2 - 3b_1) = -2(-3)^{n+2} \quad \dots\dots\dots ④$$

$$p(-3)^{n+1} - 3p(-3)^n = -2(-3)^{n+2}, \quad (p=3) \quad \dots\dots\dots ⑤$$

④-⑤에서  $b_n - p(-3)^n = c_n$ 이라면,

$$c_1 = b_1 + 9 = 3 \text{이고 다음 식이 성립한다.} \quad \dots\dots\dots ⑥$$

$$c_{n+1} = 3c_n$$

$$\therefore c_n = 3^{n-1}c_1 = 3^n \quad \dots\dots\dots ⑦$$

$$\begin{aligned} \text{③, ⑥, ⑦ : } a_n &= b_n + xn + y + z \cdot 2^n = c_n + (-3)^n \cdot p + xn + y + z \cdot 2^n \\ &= 3^n - (-3)^{n+1} + n + 2^n \end{aligned}$$

13. [ (type 2) ( $\alpha=1$ ), ( $1, \beta, s \neq 1$ )인 경우 ]  
 $[g(n) = xn^2 + yn + z + ws^n$ 으로 둔다.]

$$a_{n+2} - 4a_{n+1} + 3a_n = 15(-2)^{n-1} - 4n + 2 \quad \dots\dots\dots ①$$

$g(n) = xn^2 + yn + z + w(-2)^{n-1}$  중에서 다음을 만족하는 것을 구하자.  $\dots\dots\dots ②$

$$g(n+2) - 4g(n+1) + 3g(n) = 15(-2)^{n-1} - 4n + 2 \quad \dots\dots\dots ③$$

②, ③에서  $x=1, y=-1, z$ 는 부정이므로 0으로 놓음,  $w=1$ 이다.

①-③에서  $a_n - g(n) = b_n$ 이라면,  $b_1 = a_1 - g(1) = 3, b_2 = a_2 - g(2) = 5$ 이고 다음 식이 성립한다.  $\dots\dots\dots ④$

$$b_{n+2} - 4b_{n+1} + 3b_n = 0$$

$$(b_{n+2} - b_{n+1}) = 3(b_{n+1} - b_n)$$

$$\therefore b_{n+1} - b_n = 3^{n-1}(b_2 - b_1) = 2 \cdot 3^{n-1}$$

$$\text{따라서 } b_n = 3 + \sum_{k=1}^{n-1} 2 \cdot 3^{k-1} = 3^{n-1} + 2 \text{이다.} \quad \dots\dots\dots ⑤$$

$$\text{④, ⑤ : } a_n = b_n + g(n) = 3^{n-1} + n^2 - n + (-2)^{n-1} + 2$$

14. [ (type 3) ( $\alpha = s \neq \beta, \alpha \neq 1, \beta \neq 1$ )인 경우 ]  
 $[g(n) = (xn + y)s^n + zn + w$ 로 둔다.]

$$a_{n+2} - 5a_{n+1} + 6a_n = 3^n + 4n + 4 \quad \dots\dots\dots ①$$

$g(n) = (xn + y) \cdot 3^{n-1} + zn + w$  중에서 다음을 만족하는 것을 구하자.  $\dots\dots\dots ②$

$$g(n+2) - 5g(n+1) + 6g(n) = 3^n + 4n + 4 \quad \dots\dots\dots ③$$

②, ③에서  $x=1, y$ 는 부정이므로 0이고  $z=2, w=5$ 이다.

①-③에서  $a_n - g(n) = b_n$ 이라 놓으면,  $b_1 = a_1 - g(1) = -4, b_2 = a_2 - g(2) = -5$ 이고 다음 식이 성립한다.  $\dots\dots\dots ④$

$$b_{n+2} - 5b_{n+1} + 6b_n = 0$$

$$\text{즉, } b_{n+2} - 2b_{n+1} = 3(b_{n+1} - 2b_n)$$



$$\therefore b_{n+1} - 2b_n = 3^{n-1}(b_2 - 2b_1) = 3^n \quad \dots\dots\dots ⑤$$

$$c \cdot 3^{n+1} - 2c \cdot 3^n = 3^n, \quad (c=1) \quad \dots\dots\dots ⑥$$

⑤-⑥에서  $b_n - c \cdot 3^n = c_n$ 이라면,

$$c_1 = b_1 - 3c = -7 \text{이고 다음 식이 성립한다.} \quad \dots\dots\dots ⑦$$

$$c_{n+1} = 2c_n$$

$$\therefore c_n = 2^{n-1} c_1 = -7 \cdot 2^{n-1} \quad \dots\dots\dots ⑧$$

$$④, ⑦, ⑧ : a_n = b_n + g(n) = c_n + c \cdot 3^n + g(n) = (n+3) \cdot 3^{n-1} - 7 \cdot 2^{n-1} + 2n + 5$$

### A.4 연습문제

1. (1) ①+② :  $a_n + b_n = (2p+p^2)(a_{n-1} + b_{n-1})$   
 $a_n + b_n = (2p+p^2)^n (a_0 + b_0) = (2p+p^2)^n \quad \dots\dots\dots ③$

①-② :  $a_n - b_n = (2p-p^2)(a_{n-1} - b_{n-1})$   
 $a_n - b_n = (2p-p^2)^n (a_0 - b_0) = (2p-p^2)^n \quad \dots\dots\dots ④$

$$\frac{③+④}{2} : a_n = \frac{1}{2} \{ (2p+p^2)^n + (2p-p^2)^n \}$$

$$\frac{③-④}{2} : b_n = \frac{1}{2} \{ (2p+p^2)^n - (2p-p^2)^n \}$$

(2) 수렴하기 위해서는 공비가  $-1 < 2p+p^2 \leq 1$ ,  $-1 < 2p-p^2 \leq 1$ 이어야 하므로 공통범위  $1 - \sqrt{2} < p \leq \sqrt{2} - 1$ 이 수렴하기 위한 조건이다.

2.  $x_{n+1} = 2x_n + y_n \quad \dots\dots\dots ①$

$$y_{n+1} = x_n + 2y_n \quad \dots\dots\dots ②$$

②를 ①에 대입하여 정리하면 다음과 같다.

$$x_{n+1} = y_{n+2} - 2y_{n+1} = 2(y_{n+1} - 2y_n) + 3y_n$$

$$\therefore y_{n+2} - 4y_{n+1} + y_n = 0 \quad \dots\dots\dots ③$$

방정식  $t^2 - 4t + 1 = 0$ 의 근을  $\alpha = 2 + \sqrt{3}$ ,  $\beta = 2 - \sqrt{3}$ 이라 하자.

③에서  $y_{n+2} - \alpha y_{n+1} = \beta(y_{n+1} - \alpha y_n)$ 이다.  
 $\therefore y_{n+1} - \alpha y_n = \beta^n (y_1 - \alpha y_0) = \beta^n, \quad (\because y_1 = x_0 + 2y_0 = 1) \quad \dots\dots\dots ④$

④에서  $\alpha, \beta$ 를 바꾸어도 상관없으므로 다음 식이 성립한다.  
 $y_{n+1} - \beta y_n = (y_1 - \beta y_0) = \alpha^n \quad \dots\dots\dots ⑤$

$$⑤-④ : (\alpha - \beta)y_n = \alpha^n - \beta^n \quad \dots\dots\dots ⑥$$

$$\overline{OP_n} \text{의 기울기} = \frac{y_n}{x_n} = \frac{y_n}{y_{n+1} - 2y_n} = \frac{\alpha^n - \beta^n}{\sqrt{3}(\alpha^n + \beta^n)}$$

$$\lim_{n \rightarrow \infty} \frac{y_n}{x_n} = \lim_{n \rightarrow \infty} \frac{1}{\sqrt{3}} \left\{ \frac{1 - \left(\frac{\beta}{\alpha}\right)^n}{1 + \left(\frac{\beta}{\alpha}\right)^n} \right\} = \frac{1}{\sqrt{3}}, \quad (\because 0 < \frac{\beta}{\alpha} < 1)$$

3. ①+② :  $a_n + b_n = (p+q)(a_{n-1} + b_{n-1}) + 1$

$$\textcircled{1}-\textcircled{2} : a_n - b_n = (p-q)(a_{n-1} - b_{n-1}) + 1$$

$a_n + b_n = x_n$ ,  $a_n - b_n = y_n$ 이라 놓으면 다음 식들이 성립한다.

$$x_n = (p+q)x_{n-1} + 1$$

$$y_n = (p-q)y_{n-1} + 1$$

이 때  $a_n$ ,  $b_n$ 이 수렴하기 위한 필요충분조건은  $x_n$ ,  $y_n$ 이 수렴하는 것이므로  $x_n$ ,  $y_n$ 이 수렴하기 위한 조건  $|p+q| < 1$ ,  $|p-q| < 1$ 이 답이다.

4. (1)  $a_{2n+1} = a_{2n} + 2^{n-1} = 2a_{2n-1} + 2^{n-1}$ 의 양변을  $2^{n-1}$ 로 나누면 다음 식이 성립한다.

$$\frac{a_{2n+1}}{2^{n-1}} = \frac{a_{2n-1}}{2^{n-2}} + 1$$

이 식에서  $b_n = \frac{a_{2n-1}}{2^{n-2}}$ 이라면,  $b_1 = 2a_1 = 2$ 이고 다음 식이 성립한다.

$$\dots\dots\dots\textcircled{1}$$

$$b_{n+1} = b_n + 1$$

$$\therefore b_n = b_1 + (n-1) = n+1 \dots\dots\dots\textcircled{2}$$

$$\textcircled{1}, \textcircled{2} : a_{2n-1} = (n+1)2^{n-2} \dots\dots\dots\textcircled{3}$$

$$\text{또한 } a_{2n} = 2a_{2n-1} = 2^{n-1}(n+1) \text{이다.} \dots\dots\dots\textcircled{4}$$

$$\textcircled{3}, \textcircled{4} : a_n = \frac{n+3}{2}(\sqrt{2})^{n-3} \times \frac{1-(-1)^n}{2} + \frac{n+2}{2} \times \frac{(\sqrt{2})^{n-2} + (-1)^n}{2}$$

$$(2) S_{2n} = \sum_{k=1}^{2n} a_k = \sum_{k=1}^n (a_{2k} + a_{2k-1}) = \sum_{k=1}^n \{2^{k-2}(k+1) + 2^{k-1}(k+1)\}$$

$$= \sum_{k=1}^{2n} 3 \cdot 2^{k-2}(k+1) = 3n2^{n-1}$$

( $2S_{2n} - S_{2n} = S_{2n}$ 의 방법으로 풀다)

5. (1) 조건 1의  $qy_n = x_{n+1} - px_n$ 을  $qy_{n+1} = rx_{n+1} + sqy_n$ 에 대입하고 정리하면 다음과 같다.

$$x_{n+2} - (p+s)x_{n+1} + (ps-qr)x_n = 0$$

이 식을 조건 2에 적용시키면  $p+s = \alpha + \beta$ ,  $ps-qr = \alpha\beta$ 에서 다음 식이 성립한다.

$$x_{n+2} - \alpha x_{n+1} = \beta(x_{n+1} - \alpha x_n)$$

$$\therefore x_{n+1} - \alpha x_n = \beta^{n-1}(x_2 - \alpha x_1) \dots\dots\dots\textcircled{1}$$

이 때  $\alpha$ ,  $\beta$ 는 교환이 가능하므로  $x_{n+1} - \beta x_n = \alpha^{n-1}(x_2 - \beta x_1)$ 이다.

$$\dots\dots\dots\textcircled{2}$$

$$\textcircled{2} - \textcircled{1} : (\alpha - \beta)x_n = \alpha^{n-1}(x_2 - \beta x_1) - \beta^{n-1}(x_2 - \alpha x_1) \dots\dots\dots\textcircled{3}$$

이 때  $x_2 = px_1 + qy_1$ ,  $p+s = \alpha + \beta$ 에서 다음 식이 성립한다.

$$x_2 - \beta x_1 = (qy_1 - sx_1) + \alpha x_1, \quad x_2 - \alpha x_1 = (qy_1 - sx_1) + \beta x_1 \dots\dots\dots\textcircled{4}$$

$$\textcircled{3}, \textcircled{4} : (\alpha - \beta)x_n = (\alpha^n - \beta^n)x_1 + (qy_1 - sx_1)(\alpha^{n-1} - \beta^{n-1})$$

$$\therefore A = qy_1 - sx_1$$

같은 방법으로  $B = rx_1 - py_1$ 이다.

(2)  $p=q=1$ ,  $r=2$ ,  $s=3$ 이면 조건 2에서  $t^2 - 4t + 1 = 0$ 이므로,  $\alpha = 2 + \sqrt{3}$ ,  $\beta = 2 - \sqrt{3}$ 이다. 조건 1의 결과에서  $A = -2$ ,  $B = 1$ 이므로 다음 식이 성립한다.

$$\begin{aligned}\lim_{n \rightarrow \infty} \frac{x_n}{y_n} &= \lim_{n \rightarrow \infty} \frac{(a^n - \beta^n) - 2(a^{n-1} - \beta^{n-1})}{(a^n - \beta^n) + (a^{n-1} - \beta^{n-1})} = \lim_{n \rightarrow \infty} \frac{a - \beta \left(\frac{\beta}{a}\right)^{n-1} - 2 + 2\left(\frac{\beta}{a}\right)^{n-1}}{a - \beta \left(\frac{\beta}{a}\right)^{n-1} + 1 - \left(\frac{\beta}{a}\right)^{n-1}} \\ &= \frac{a-2}{a+1} = \frac{\sqrt{3}-1}{2}, \quad \left(\because 0 < \frac{\beta}{a} = \frac{2-\sqrt{3}}{2+\sqrt{3}} < 1\right)\end{aligned}$$

6. ①+②+③ :  $6(x_{n+1} + y_{n+1} + z_{n+1}) = 6(x_n + y_n + z_n)$

$$\therefore x_n + y_n + z_n = x_1 + y_1 + z_1 = -2 \quad \dots\dots\dots ④$$

$$\text{①, ④} : 6x_{n+1} = -2x_n - y_n - 10 \quad \dots\dots\dots ⑤$$

$$\text{②, ④} : 6y_{n+1} = x_n + 2y_n - 7 \quad \dots\dots\dots ⑥$$

⑤에서  $y_n = -6x_{n+1} - 2x_n - 10$ 을 ⑥에 대입하면 다음 식을 얻는다.

$$12x_{n+2} - x_n + 11 = 0 \quad \dots\dots\dots ⑦$$

$$12x - x + 11 = 0, \quad (x = -1) \quad \dots\dots\dots ⑧$$

⑦-⑧에서  $x_n - x = a_n$ 이라면 다음 식이 성립한다.  $\dots\dots\dots ⑨$

$$12a_{n+2} - a_n = 0 \quad \dots\dots\dots ⑩$$

방정식  $12t^2 - 1 = 0$ 의 2근을  $\alpha = \frac{\sqrt{3}}{6}$ ,  $\beta = -\frac{\sqrt{3}}{6}$ 이라면 ⑩에서 다음 식이 성립한다.

$$a_{n+2} - \alpha a_{n+1} = \beta(a_{n+1} - \alpha a_n)$$

$$\therefore a_{n+1} - \alpha a_n = \beta^{n-1}(a_2 - \alpha a_1) \quad \dots\dots\dots ⑪$$

이 때  $\alpha$ ,  $\beta$ 를 바꾸어도 되므로  $a_{n+1} - \beta a_n = \alpha^{n-1}(a_2 - \beta a_1)$ 이다.  $\dots\dots\dots ⑫$

$$\text{①, ⑨} : a_1 = x_1 - x = 0, \quad a_2 = x_2 - x = -\frac{1}{6}$$

$$\text{⑪-⑫} : (\beta - \alpha)a_n = \beta^{n-1}(a_2 - \alpha a_1) - \alpha^{n-1}(a_2 - \beta a_1)$$

이 식에  $\alpha, \beta, a_1, a_2$ 를 대입하면 다음 식이 성립한다.

$$a_n = -\left(\frac{\sqrt{3}}{6}\right)^n - \left(-\frac{\sqrt{3}}{6}\right)^n \quad \dots\dots\dots ⑬$$

$$\text{⑨, ⑬} : x_n = a_n + x = -\left\{1 + \left(\frac{\sqrt{3}}{6}\right)^n + \left(-\frac{\sqrt{3}}{6}\right)^n\right\} \quad \dots\dots\dots ⑭$$

$$\text{⑤, ⑬} : y_n = -2 + 6 \times \left\{\left(\frac{\sqrt{3}}{6}\right)^{n+1} + \left(-\frac{\sqrt{3}}{6}\right)^{n+1}\right\} + 2 \times \left\{\left(\frac{\sqrt{3}}{6}\right)^n + \left(-\frac{\sqrt{3}}{6}\right)^n\right\} \quad \dots\dots\dots ⑮$$

$$\text{④, ⑬, ⑮} : z_n = 1 - 6 \times \left\{\left(\frac{\sqrt{3}}{6}\right)^{n+1} + \left(-\frac{\sqrt{3}}{6}\right)^{n+1}\right\} - \left\{\left(\frac{\sqrt{3}}{6}\right)^n + \left(-\frac{\sqrt{3}}{6}\right)^n\right\}$$

### A.5 연습문제

1. 수열  $x_{n+1} = q \sin(x_n) + a \dots ①$ 에서 특성근을  $\alpha$ 라 하자.

$$\text{즉, } \alpha = q \sin \alpha + a \quad \dots\dots\dots ②$$

$$\textcircled{1}-\textcircled{2} : x_{n+1}-\alpha=q(\sin x_n-\sin \alpha)=2q\cos \frac{x_n+\alpha}{2}\sin \frac{x_n-\alpha}{2}$$

$$\therefore 0 \leq |x_{n+1}-\alpha|=2q\left|\cos \frac{x_n+\alpha}{2}\right|\left|\sin \frac{x_n-\alpha}{2}\right| \leq 2q\left|\frac{x_n-\alpha}{2}\right|=q|x_n-\alpha|$$

$$\left(\because |\sin x| \leq |x|, \left|\cos \frac{x_n+\alpha}{2}\right| \leq 1\right)$$

따라서  $0 \leq |x_n-\alpha| \leq q^{n-1}|x_1-\alpha|$ 이다.

이 때  $\lim_{n \rightarrow \infty} q^{n-1}|x_1-\alpha|=0$ , ( $\because 0 < q < 1$ )

$$\therefore \lim_{n \rightarrow \infty} x_n = \alpha$$

따라서 수열  $\{x_n\}$ 은 수렴하고, 그 극한값은  $x=q\sin x + a$ 의 근이다.

2. (1)  $n=1$ 일 때  $0 < a_1 = \frac{1}{2} < 1$

$n=k$ 일 때  $0 < a_k < 1$ 이면  $0 < \frac{\pi}{2}a_k < \frac{\pi}{2}$ 이다.

$n=k+1$ 에서  $0 < a_{k+1} = \sin\left(\frac{\pi}{2}a_k\right) < 1$ 이다.

따라서 모든 자연수  $n$ 에 대해  $0 < a_n < 1$ 이다.

(2)  $n=1$ 일 때  $a_2 = \sin\left(\frac{\pi}{2}a_1\right) = \sin\frac{\pi}{4} = \frac{1}{\sqrt{2}} > \frac{1}{2} = a_1 \therefore a_2 > a_1$

$n=k$ 일 때  $a_{k+1} > a_k$ 이면  $\frac{\pi}{2} > \frac{\pi}{2}a_{k+1} > \frac{\pi}{2}a_k > 0$ 이다.

$n=k+1$ 에서  $1 > \sin\left(\frac{\pi}{2}a_{k+1}\right) > \sin\left(\frac{\pi}{2}a_k\right) > 0$ , 즉  $a_{k+2} > a_{k+1}$ 이다.

따라서 모든 자연수  $n$ 에 대해  $a_{n+1} > a_n$ 이다.

(3) (1), (2)에서 유계단조수열이므로 극한값을 갖는다.

즉,  $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} a_{n+1} = \alpha$  이면  $0 \leq \lim_{n \rightarrow \infty} a_n = \alpha \leq 1$ 이다.

또한  $\alpha = \sin\frac{\pi}{2}\alpha$ 이므로  $\alpha=1$ , 즉  $\lim_{n \rightarrow \infty} a_n = 1$ 이다.

3. (1) 근의 공식에서  $a_{n+1} = \frac{-1 \pm \sqrt{1+a_n^2}}{a_n}$ 이다. 이 때  $|a_{n+1}| < 1$ 이므로 다음 식이 성립한다.

$$a_{n+1} = \frac{-1 + \sqrt{1+a_n^2}}{a_n}$$

$$(2) a_{n+1} = \frac{-1 + \sqrt{1+a_n^2}}{a_n} = \frac{a_n}{1 + \sqrt{1+a_n^2}}$$

$$\therefore 0 \leq |a_{n+1}| = \frac{|a_n|}{1 + \sqrt{1+a_n^2}} \leq \frac{|a_n|}{2}$$

$$\therefore 0 \leq |a_n| \leq \frac{|a_1|}{2^{n-1}}$$

이 때  $\lim_{n \rightarrow \infty} \frac{a_1}{2^{n-1}} = 0$ 이므로  $\lim_{n \rightarrow \infty} a_n = 0$ 이다.

4. (1)  $n=1$ 에서  $0 < u_1 < \frac{3}{2}$

$n=k$ 일 때  $0 < u_k < \frac{3}{2}$ 이라고 하자.

$n=k+1$ 에서  $u_{k+1} = f(u_k) = \frac{1}{9}u_k(10 - u_k^2)$ 에서  $u_{k+1}$ 의 범위는  $0 < x < \frac{3}{2}$ 에서  $y = \frac{1}{9}x(10 - x^2)$ 의  $y$ 범위와 같다. 또한  $0 < x < \frac{3}{2}$ 에서  $y' = \frac{1}{9}(10 - 3x^2) > 0$ 이다. 따라서  $y$ 는  $0 < y < f(\frac{3}{2}) = \frac{31}{24} < \frac{3}{2}$ 이고, 따라서  $0 < u_{k+1} < \frac{3}{2}$ 이다. 즉, 모든  $n$ 에 대해  $0 < u_n < \frac{3}{2}$ 이다.

(2) i)  $u_1 > u_2$ 의 경우  $u_{k-1} > u_k$ 이면 평균값의 정리에서 다음 식이 성립한다.

$$u_{k+1} - u_k = f(u_k) - f(u_{k-1}) = (u_k - u_{k-1})f'(c) < 0 \quad (\text{단, } u_{k-1} > c > u_k)$$

$$(\because 0 < u_n < \frac{3}{2} \text{에서 } 0 < c < \frac{3}{2} \text{이고, 또 이 구간에서 } f'(c) > 0)$$

$$\therefore u_{k+1} < u_k$$

따라서 모든 자연수  $n$ 에 대해  $u_{n+1} < u_n$ 이다.

ii) 같은 방법으로  $u_1 < u_2$ 이면  $u_n < u_{n+1}$ 이다.

(3) i)  $\frac{3}{2} > u_1 > u_2 > u_3 > \dots > u_n > \dots > 0$

ii)  $\frac{3}{2} > \dots > u_n > u_{n-1} > \dots > u_2 > u_1 > 0$

i), ii) 모두 유계단조수열이므로 극한값을 갖는다. 즉, 특성방정식  $9a = (10 - a^2)a$ 에서  $a = 0, 1, -1$ 이다. 여기서  $0 \leq \lim_{n \rightarrow \infty} u_n \leq \frac{3}{2}$ 에 맞는 근은  $0, 1$ 이다.

이 때 그래프를 사용 조사하면  $0 < u_1 \leq 1$ 에서  $u_1 \leq u_2$ 이고  $\lim_{n \rightarrow \infty} u_n = 1$ 이다. 따라서

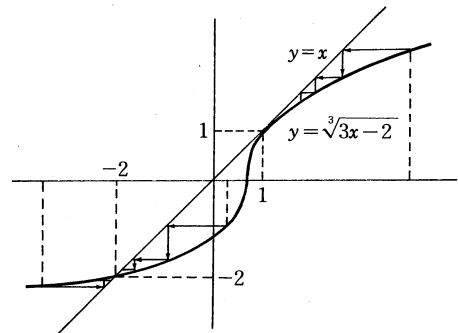
$1 < u_1 < \frac{3}{2}$ 이면  $u_1 > u_2$ 이고  $\lim_{n \rightarrow \infty} u_n = 1$ , 따라서 극한값은  $1$ 이다.

5. 특성방정식  $x = \sqrt[3]{3x-2}$ 는  $y=x$ 와  $y = \sqrt[3]{3x-2}$ 가 만나는 것으로 나타낼 수 있다. 이 때 특성방정식의 근( $x = -2, 1$ )과 위의 두 그래프를 고찰하여 다음을 얻을 수 있다.

$a \leq -2$ 일 때  $\lim_{n \rightarrow \infty} a_n = -2$ 이다.

$-2 < a < 1$ 에서  $\lim_{n \rightarrow \infty} a_n = -2$ 이다.

$1 \leq a$ 에서  $\lim_{n \rightarrow \infty} a_n = 1$ 이다.



6. (1)  $a_1 > a$ 이고,  $a_k > a$ 라 가정하면 다음 식이 성립한다.

$$a_{k+1} - a = f(a_k) - a = f(a_k) - f(a) = f'(c)(a_k - a), \quad (\text{단, } a < c < a_k)$$

이 때  $a_k > a$ 이고  $f'(c) > 0$ 이므로  $a_{k+1} > a$ 이다. 따라서 모든 자연수  $n$ 에 대해  $a_n > a$ 이다.

$$(2) \quad a_{n+1} - a_n = a_{n+1} - a - (a_n - a) = f(a_n) - f(a) - (a_n - a) \\ = f'(c)(a_n - a) - (a_n - a) = (a_n - a)(f'(c) - 1) < 0, \quad (\text{단, } a < c < a_n)$$

$$(\because a_n > a, \quad 0 < f'(c) < 1)$$

$$\therefore a_{n+1} < a_n$$

(3) (1), (2)에서 유계단조수열이므로 수렴한다. 이 때의 극한값을  $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} a_{n+1} = x$ 라고 하면  $x = f(x)$ 의 근, 즉  $a$ 이다.

### A.6 연습문제

1. 우선  $a_n > 0$ 을 증명하기로 한다.

$n=1$ 일 때  $a_1 = 4 > 0$ 이다.

$n=k$ 일 때  $a_k > 0$ 이라 가정하면,  $n=k+1$ 일 때  $a_{k+1} = \sqrt[3]{a_k} > 0$ 이다.

$\therefore$  모든 자연수  $n$ 에 대해  $a_n > 0$ 이다.

이러한 조건하에  $\log$ 를 취하면서  $\log a_n = b_n$ 이라면 다음 식이 성립한다.

$$b_n = \frac{1}{3} b_{n-1} \quad \therefore b_n = \left(\frac{1}{3}\right)^{n-1} \log 4, \quad (\because b_1 = \log a_1 = \log 4)$$

이 때  $A = a_1 a_2 a_3 a_4 \cdots a_{n-1} a_n$ 이라면 다음 식이 성립한다.

$$\log A = \sum_{k=1}^n \log a_k = \sum_{k=1}^n b_k = \sum_{k=1}^n \left(\frac{1}{3}\right)^{k-1} \log 4 = \frac{3}{2} \log 4 \left\{ 1 - \left(\frac{1}{3}\right)^n \right\}$$

$$\therefore \lim_{n \rightarrow \infty} \log A = \log 8 \quad \therefore \lim_{n \rightarrow \infty} A = 8$$

2.  $y_n = x \sqrt{x y_{n-1}} = x \sqrt{x} \sqrt{y_{n-1}} = y_1 \sqrt{y_{n-1}}$ 의 양변에  $\log$ 를 취하여  $\log y_n = z_n$ 이라면,

$z_1 = \log y_1 = \frac{3}{2} \log x$ 이고 다음 식이 성립한다.

$$z_n = \frac{1}{2} z_{n-1} + z_1$$

$$\therefore z_n = z_1 \left\{ 2 - \left(\frac{1}{2}\right)^{n-1} \right\}$$

따라서 다음 식이 성립한다.

$$\log y_n = (\log y_1) \left\{ 2 - \left(\frac{1}{2}\right)^{n-1} \right\}$$

$$\lim_{n \rightarrow \infty} \log y_n = 2 \log y_1 = 3 \log x$$

$$\therefore \lim_{n \rightarrow \infty} y_n = x^3$$

3.  $a_n > 0$ 이라 했으므로 양변에  $\log$ 를 취하여  $\log a_n = b_n$ 이라하면 다음 식이 성립한다.

$$2b_{n+2} + b_n = 3b_{n+1}, \quad (b_1 = \log a_1 = 0, \quad b_2 = \log a_2 = 1)$$

따라서 다음 식이 성립한다.

$$b_{n+2} - b_{n+1} = \frac{1}{2}(b_{n+1} - b_n)$$

$$\therefore b_{n+1} - b_n = \left(\frac{1}{2}\right)^{n-1} (b_2 - b_1) = \left(\frac{1}{2}\right)^{n-1}$$

$$\therefore b_n = 0 + \sum_{k=1}^{n-1} \left(\frac{1}{2}\right)^{k-1} = 2 - \left(\frac{1}{2}\right)^{n-2}$$

$$\therefore a_n = 10^{2 - \left(\frac{1}{2}\right)^{n-2}}$$

$$\lim_{n \rightarrow \infty} a_n = 100$$

4. 3식의 양변에  $\log$ 를 취하고  $\log a_n = x_n$ ,  $\log b_n = y_n$ ,  $\log c_n = z_n$ 이라하면 다음 식이 성립한다.

$$x_{n+1} = \frac{1}{2}(y_n + z_n), \quad (x_1 = 0) \quad \dots\dots\dots ①$$

$$y_{n+1} = \frac{1}{2}(z_n + x_n), \quad (y_1 = \log 2) \quad \dots\dots\dots ②$$

$$z_{n+1} = \frac{1}{2}(x_n + y_n), \quad (z_1 = \log 5) \quad \dots\dots\dots ③$$

①+②+③에서  $x_n + y_n + z_n = d_n$ 이라하면,  $d_1 = x_1 + y_1 + z_1 = 1$ 이고 다음 식이 성립한다.

$$d_{n+1} = d_n$$

$$\therefore d_n = d_1 = 1 \quad \text{즉,} \quad x_n + y_n + z_n = 1 \quad \dots\dots\dots ④$$

$$\text{①, ④} : x_{n+1} = \frac{1}{2}(1 - x_n)$$

$$\text{즉,} \quad x_{n+1} - \frac{1}{3} = \left(-\frac{1}{2}\right)\left(x_n - \frac{1}{3}\right)$$

$$\therefore x_n - \frac{1}{3} = \left(-\frac{1}{2}\right)^{n-1}\left(x_1 - \frac{1}{3}\right)$$

$$\text{즉,} \quad x_n = \frac{1}{3} - \frac{1}{3}\left(-\frac{1}{2}\right)^{n-1}$$

$$\therefore a_n = 10^{x_n} = 10^{\frac{1}{3}\left[1 - \left(-\frac{1}{2}\right)^{n-1}\right]}, \quad \lim_{n \rightarrow \infty} a_n = {}^3\sqrt{10}$$

같은 방법으로 ②, ④에서  $y_n = \frac{1}{3} + \left(-\frac{1}{2}\right)^{n-1}\left(\log 2 - \frac{1}{3}\right)$ 가 성립한다.

$$b_n = 10^{y_n} = 10^{\frac{1}{3} + \left(-\frac{1}{2}\right)^{n-1}\left(\log 2 - \frac{1}{3}\right)}, \quad \lim_{n \rightarrow \infty} b_n = {}^3\sqrt{10}$$

같은 방법으로 ③, ④에서  $z_n = \frac{1}{3} + \left(-\frac{1}{2}\right)^{n-1}\left(\log 5 - \frac{1}{3}\right)$ 가 성립한다.

$$\therefore c_n = 10^{z_n} = 10^{\frac{1}{3} + \left(-\frac{1}{2}\right)^{n-1}\left(\log 5 - \frac{1}{3}\right)}$$

$$\lim_{n \rightarrow \infty} c_n = {}^3\sqrt{10}$$

5. (1)  $n=2$ 에서  $x_2 = \frac{x_1^2+2}{3} = \frac{a^2+2}{3} > 1$ 이다. ( $\because 1 < a < 2$ )

$n=k$ 에서  $x_k > 1$ 이라 가정하자.

$n=k+1$ 에서  $x_{k+1} = \frac{x_k^2+2}{3} > \frac{1+2}{3} = 1$ 이다.

$\therefore$  모든 자연수  $n$ 에 대해  $x_n > 1$ 이다. .....①

$n=2$ 에서  $x_2 - a = \frac{a^2+2}{3} - a = \frac{1}{3}(a-1)(a-2) < 0$ 이다.

$n=k$ 에서  $x_k < a$ 라고 가정하자.

$n=k+1$ 에서  $x_{k+1} - a = \frac{x_k^2+2}{3} - a < \frac{a^2+2}{3} - a = \frac{1}{3}(a-1)(a-2) < 0$ 이다.

$\therefore$  모든 자연수  $n$ 에 대해  $x_n < a$ 이다. .....②

①, ② :  $1 < x_n < a$

(2)  $x_n - 1 = \frac{x_{n-1}^2+2}{3} - 1 = \frac{1}{3}(x_{n-1}-1)(x_{n-1}+1)$

이 때  $x_{n-1} - 1 > 0$ 이고,  $\frac{1}{3} < \frac{x_{n-1}+1}{3} \leq \frac{a+1}{3}$  이므로 다음 식이 성립한다.

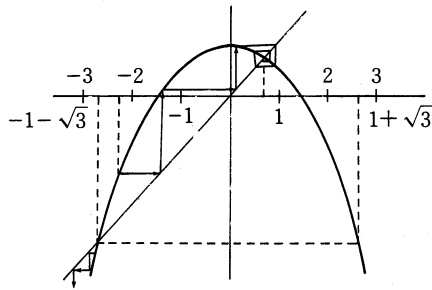
$$0 < x_n - 1 \leq \frac{a+1}{3}(x_{n-1} - 1)$$

(3) (2)에서  $0 < x_n - 1 \leq \left(\frac{a+1}{3}\right)^{n-1} (x_1 - 1)$

또  $1 < a < 2$ 에서  $0 < \frac{a+1}{3} < 1$ 이므로  $\lim_{n \rightarrow \infty} \left(\frac{a+1}{3}\right)^{n-1} = 0$ 이다.

$\therefore \lim_{n \rightarrow \infty} x_n = 1$

- |                                      |                                                   |
|--------------------------------------|---------------------------------------------------|
| 6. $p < -1 - \sqrt{3}$ ,             | 발산                                                |
| $p > 1 + \sqrt{3}$ ,                 | 발산                                                |
| $p = -1 - \sqrt{3}$ ,                | $\lim_{n \rightarrow \infty} a_n = -\sqrt{3} - 1$ |
| $p = 1 + \sqrt{3}$ ,                 | $\lim_{n \rightarrow \infty} a_n = 1 + \sqrt{3}$  |
| $-1 - \sqrt{3} < p < 1 + \sqrt{3}$ , | $\lim_{n \rightarrow \infty} a_n = \sqrt{3} - 1$  |
| (단, $ p  \neq \sqrt{2}$ )            |                                                   |
| $ p  = \sqrt{2}$ ,                   | $\lim_{n \rightarrow \infty} a_n = 0$             |



7. (1)  $x_{n+1} = x_n(2 - x_n)$ 의 특성방정식  $x = x(2 - x)$ 의 근을 양변에서 빼면 다음과 같다.

$$x_{n+1} - 1 = -(x_n - 1)^2$$

이 식에서  $1 - x_n = u_n$ 이라면 다음 식이 성립한다.

$$u_{n+1} = u_n^2 \tag{.....①}$$

(가)  $x_n \neq 1$ 인 경우  $u_n > 0$ 이므로 양변에  $\log$ 를 취하면 다음과 같다.

$$\log u_{n+1} = 2 \log u_n$$

$$\therefore \log u_n = 2^{n-1} \log u_1 = 2^{n-1} \log(1 - x_1)$$



$$\therefore u_n = (1-x_1)^{2^{n-1}}, \text{ 즉 } x_n = 1(1-x_1)^{2^{n-1}} \dots\dots\dots ②$$

(나)  $x_n=1$ 이면  $u_n=0$ 이 되어 ①을 만족한다.  $\dots\dots\dots ③$

$$②, ③ : x_n = 1 - (1-x_1)^{2^{n-1}}$$

(2) 수렴하기 위해서는  $-1 < (1-x_1)^2 \leq 1$ , 즉  $2 \geq x_1 \geq 0$ 이어야 한다.

8. (1)  $f_{n+1}(x) = \{f_n(x)\}^2 - 2af_n(x) + a^2$ 에서  
 상수항의 계수를 비교하면  $a_{n+1} = a_n^2 - 2a \cdot a_n + a^2 = (a_n - a)^2$ 이고,  
 1차항의 계수를 비교하면  $b_{n+1} = 2a_nb_n - 2ab_n = 2b_n(a_n - a)$ 이다.

(2)  $a=2$ 이면,  $a_{n+1} = (a_n - 2)^2 \dots\dots\dots ①$

$$b_{n+1} = 2b_n(a_n - 2) \dots\dots\dots ②$$

또  $f_1(x) = (x-2)^2 = x^2 - 4x + 4$ 에서  $a_1=4, b_1=-4$ 이다.

①에서  $a_2 = (a_1 - 2)^2 = 4, a_3 = (a_2 - 2)^2 = 4$ 이다.

여기서  $a_n=4$ 로 추정한다.

$n=1$ 일 때  $a_1=4$

$n=k$ 일 때  $a_k=4$ 로 가정하면,  $n=k+1$ 에서  $a_{k+1} = (a_k - 2)^2 = 4$ 이다.

$\therefore$  모두  $n$ 에 대해  $a_n=4$ 이다.

$$② : b_{n+1} = 2b_n(a_n - 2) = 4b_n$$

$$\therefore b_n = 4^{n-1}b_1 = -4^n$$

9. (1)  $a_{n+1} > a_n$ 의 수학적 귀납법으로 증명한다.

$$n=1 \text{에서 } a_2 - a_1 = \frac{1}{2}(a_1^2 + 1) - a_1 = \frac{1}{2}(a_1 - 1)^2 = \frac{1}{2}\left(\frac{x}{2} - 1\right)^2 > 0, (\because x \neq 2)$$

$n=k$ 에서  $a_{k+1} > a_k$ 라 가정하자.

$$n=k+1 \text{에서 } a_{k+2} - a_{k+1} = \frac{1}{2}(a_{k+1}^2 + 1 - a_{k+1}^2 - 1) = \frac{1}{2}(a_{k+1} + a_k)(a_{k+1} - a_k) > 0$$

따라서 모든 자연수  $n$ 에 대해  $a_n < a_{n+1}$ 이다.

❖ 참고 :  $a_n > 0$ 임을 수학적 귀납법으로 증명하는 과정은 다음과 같다.

$$n=1 \text{에서 } 2a_1 = x > 0 \therefore a_1 > 0$$

$n=k$ 일 때  $a_k > 0$ 이라면

$$n=k+1 \text{에서 } 2a_{k+1} = a_k^2 + 1 > 1 > 0, \therefore a_{k+1} > 0$$

따라서 모든 자연수  $n$ 에 대해  $a_n > 0$ 이다.

(2)  $n=1$ 에서  $a_1 = \frac{x}{2} < 1$

$x=k$ 에서  $a_k < 1$ 이라 가정하자.

$$n=k+1 \text{에서 } a_{k+1} - 1 = \frac{1}{2}(a_k^2 + 1) - 1 = \frac{1}{2}(a_k + 1)(a_k - 1) < 0 \text{이다.}$$

따라서 모든 자연수  $n$ 에 대해  $a_n < 1$ 이다. 다음 식이 성립한다.

$$a_{k+1} - a_k = \frac{1}{2}(a_k^2 + 1) - a_k = \frac{1}{2}(a_k - 1)^2 \dots\dots\dots ①$$

이 식에서  $0 < \varepsilon < 1 - \frac{x}{2}$ 이고  $a_1 < a_2 < \dots < a_{n-1} < a_n \leq 1 - \varepsilon$ 이면 다음 성질들이 성립

한다.

$$\text{만일 } k=1, 2, \dots, n-2, n-1 \text{이면, } \frac{1}{2}(a_k-1)^2 > \frac{\varepsilon^2}{2} \text{이다.} \quad \dots\dots\dots \textcircled{2}$$

$$\text{만일 } k=n \text{이면, } \frac{1}{2}(a_k-1)^2 \geq \frac{\varepsilon^2}{2} \text{이다.} \quad \dots\dots\dots \textcircled{3}$$

또한  $n \geq 2$ 에서 ①의 양변에  $\sum_{k=1}^n$ 을 붙이면, ②, ③에 의해서 다음 식이 성립한다.

$$\begin{aligned} \sum_{k=1}^n (a_{k+1} - a_k) &= a_{n+1} - a_1 > \sum_{k=1}^n \frac{\varepsilon^2}{2} = \frac{n}{2} \varepsilon^2 \\ \therefore 2-x > 2(a_{n+1} - a_1) &> n\varepsilon^2, \quad (\because 2a_{n+1} < 2, 2a_1 = x) \\ n=1 \text{에서 } 2-x > 2\varepsilon > \varepsilon^2, \quad (\because 2 > \varepsilon) \\ \text{따라서 모든 자연수 } n \text{에 대해 } 2-x > n\varepsilon^2 \text{이다.} \end{aligned}$$

**===== A.7 연습문제 =====**

1.  $\frac{1}{a_n} = b_n$ 이라면,  $b_1 = \frac{1}{a_1} = 1$ 이고 다음 식이 성립한다.

$$\begin{aligned} b_{n+1} - b_n &= (n+1)^2 \\ \therefore b_n &= 1 + \sum_{k=1}^{n-1} (k+1)^2 = \frac{1}{6}n(n+1)(2n+1) \\ \therefore a_n &= \frac{1}{b_n} = \frac{6}{n(n+1)(2n+1)} \end{aligned}$$

2.  $na_n a_{n-1} + a_n - a_{n-1} = 0$ 의 양변을  $a_n a_{n-1}$ 로 나누면 다음 식이 성립한다.

$$\begin{aligned} \frac{1}{a_n} - \frac{1}{a_{n-1}} &= n \\ \text{이 식에서 } \frac{1}{a_n} = b_n \text{이라면, } b_1 &= \frac{1}{a_1} = 1 \text{이고 다음 식이 성립한다.} \\ b_{n+1} - b_n &= n+1 \\ \therefore b_n &= 1 + \sum_{k=1}^{n-1} (k+1) = \frac{1}{2}n(n+1) \\ \therefore a_n &= \frac{1}{b_n} = \frac{2}{n(n+1)} \\ \sum_{n=1}^{\infty} a_n &= \lim_{n \rightarrow \infty} \sum_{k=1}^n a_k = \lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{2}{k(k+1)} = \lim_{n \rightarrow \infty} 2 \left( 1 - \frac{1}{n+1} \right) = 2 \end{aligned}$$

3. 원식을 역수로 취하면 다음과 같다.

$$\begin{aligned} \frac{1}{b_{n+1}} = \frac{b_n+1}{b_n} &= 1 + \frac{1}{b_n} \\ \text{이 식에서 } \frac{1}{b_n} = a_n \text{으로 놓으면, } b_1 &= \frac{1}{a_1} > 0 \text{이고 다음 식이 성립한다.} \\ a_{n+1} &= a_n + 1 \end{aligned}$$

$$\therefore a_n = a_1 + (n-1)$$

$$b_n = \frac{1}{a_n} = \frac{1}{a_1 + (n-1)}$$

$$\therefore \lim_{n \rightarrow \infty} b_n = 0$$

또한 다음 식이 성립한다.

$$b_k^2 = \left\{ \frac{1}{a_1 + (k-1)} \right\}^2 < \frac{1}{(k-1+a_1)(k-2+a_1)} = \frac{1}{k-2+a_1} - \frac{1}{k-1+a_1}$$

$$\therefore \sum_{k=2}^n b_k^2 < \sum_{k=2}^n \left( \frac{1}{k-2+a_1} - \frac{1}{k-1+a_1} \right) = \frac{1}{a_1} - \frac{1}{n-1+a_1} < \frac{1}{a_1} = b_1$$

따라서  $b_2^2 + b_3^2 + \dots + b_n^2 < b_1$ 이다.

4. (1) 원식을 역수로 취하면 다음과 같다.

$$\frac{1}{a_n} = \frac{1}{a_{n-1}} + (2n-1)$$

이 식에서  $\frac{1}{a_n} = b_n$ 이라면,  $b_1 = \frac{1}{a_1} = 1$ 이고 다음 식이 성립한다.

$$b_{n+1} = b_n + (2n+1)$$

$$\therefore b_n = 1 + \sum_{k=1}^{n-1} (2k+1) = n^2$$

$$\therefore a_n = \frac{1}{n^2}$$

$$\begin{aligned} (2) S_n &= \sum_{k=1}^n \frac{1}{\sqrt{a_k a_{n+1-k}}} = \sum_{k=1}^n \sqrt{k^2(n+1-k)^2} = \sum_{k=1}^n k(n+1-k) \\ &= (n+1) \sum_{k=1}^n k - \sum_{k=1}^n k^2 = \frac{1}{6} n(n+1)(n+2) \end{aligned}$$

$$(3) \lim_{n \rightarrow \infty} \frac{a_n S_n}{n} = \lim_{n \rightarrow \infty} \frac{n(n+1)(n+2)}{6n^3} = \frac{1}{6}$$

5. 원식의 양변을  $a_n a_{n+1} a_{n+2}$ 로 나누면 다음과 같다.

$$\frac{4}{a_n} + \frac{1}{a_{n+2}} = \frac{4}{a_{n+1}}$$

이 식에서  $\frac{1}{a_n} = b_n$ 이라면,  $b_1 = 1$ ,  $b_2 = 4$ 이고 다음 식이 성립한다. ……①

$$b_{n+2} - 4b_{n+1} + 4b_n = 0, \text{ 즉 } b_{n+2} - 2b_{n+1} = 2(b_{n+1} - 2b_n)$$

$$\therefore b_{n+1} - 2b_n = 2^{n-1}(b_2 - 2b_1) = 2^n$$

이 식의 양변을  $2^n$ 으로 나누면 다음과 같다.

$$\frac{b_{n+1}}{2^n} - \frac{b_n}{2^{n-1}} = 1$$

이 식에서  $\frac{b_n}{2^{n-1}} = c_n$ 이라면,  $c_1 = b_1 = 1$ 이고 다음 식이 성립한다. ……②

$$c_{n+1} = c_n + 1$$

$$\therefore c_n = 1 + (n-1) = n \quad \dots\dots\dots\text{③}$$

$$\textcircled{1}, \textcircled{2}, \textcircled{3} : a_n = \frac{1}{b_n} = \frac{1}{c_n 2^{n-1}} = \frac{1}{a 2^{n-1}}$$

6. 원식의 양변을  $a_n a_{n+1}$ 로 나누면 다음과 같다.

$$\frac{a_{n+2}}{a_{n+1}} = 2 + \frac{a_{n+1}}{a_n}$$

이 식에서  $\frac{a_{n+1}}{a_n} = b_n$ 이라면,  $b_1 = \frac{a_2}{a_1} = 2$ 이고 다음 식이 성립한다. ……①

$$b_{n+1} = b_n + 2$$

$$\therefore b_n = 2 + 2(n-1) = 2n \quad \text{……②}$$

①, ②에서  $a_{n+1} = 2na_n$ 의 양변을  $n!$ 로 나누면 다음과 같다.

$$\frac{a_{n+1}}{n!} = 2 \frac{a_n}{(n-1)!}$$

이 식에서  $\frac{a_n}{(n-1)!} = c_n$ 이라면,  $c_1 = 1$ 이고 다음 식이 성립한다. ……③

$$c_{n+1} = 2c_n$$

$$\therefore c_n = 2^{n-1} c_1 = 2^{n-1} \quad \text{……④}$$

$$\textcircled{3}, \textcircled{4} : a_n = (n-1)! c_n = (n-1)! 2^{n-1}$$

$$7. (1) x_{n+1} - 1 = \frac{3x_n + 1}{x_n + 3} - 1 = \frac{2(x_n - 1)}{x_n + 3}$$

$$x_{n+1} + 1 = \frac{3x_n + 1}{x_n + 3} + 1 = \frac{4(x_n + 1)}{x_n + 3}$$

이 식들에서 다음 식이 성립한다.

$$y_{n+1} = \frac{x_{n+1} - 1}{x_{n+1} + 1} = \frac{1}{2} \frac{x_n - 1}{x_n + 1} = \frac{1}{2} y_n, \quad \left( \text{단, } y_1 = \frac{x_1 - 1}{x_1 + 1} = \frac{1}{3} \right)$$

수열  $\{y_n\}$ 은 초항  $\frac{1}{3}$ , 공비  $\frac{1}{2}$ 인 등비수열이다.

$$(2) (1) \text{에서 } y_n = \left(\frac{1}{2}\right)^{n-1} y_1 = \frac{1}{3} \left(\frac{1}{2}\right)^{n-1}$$

$$y_n = \frac{x_n - 1}{x_n + 1} \text{에서 } x_n = \frac{1 + y_n}{1 - y_n} = \frac{3 \cdot 2^{n-1} + 1}{3 \cdot 2^{n-1} - 1} \text{이다.}$$

$$(3) \lim_{n \rightarrow \infty} y_n = 0 \quad \therefore \lim_{n \rightarrow \infty} x_n = 1$$

$$8. (1) c_{n+1} = \frac{a_{n+1}}{b_{n+1}} = \frac{2c_n + 3}{c_n + 2} = \frac{2a_n + 3b_n}{a_n + 2b_n} = 2 - \frac{b_n}{a_n + 2b_n}$$

이 때  $a_n, b_n$ 이 서로소이므로,  $a_n + 2b_n$ 과  $b_n$ 도 서로소이다.

$$\therefore a_{n+1} = 2a_n + 3b_n, \quad b_{n+1} = a_n + 2b_n$$

(2) (1)의 결과와  $c_1 = 2$ , 즉  $a_1 = 2, b_1 = 1$ 에서 다음 식이 성립한다.

$$a_{n+1} + \sqrt{3}b_{n+1} = 2a_n + 3b_n + \sqrt{3}a_n + 2\sqrt{3}b_n = (2 + \sqrt{3})(a_n + \sqrt{3}b_n)$$

$$\therefore a_n + \sqrt{3}b_n = (2 + \sqrt{3})^{n-1} (a_1 + \sqrt{3}b_1) = (2 + \sqrt{3})^n \quad \text{……①}$$

$$\text{같은 방법으로 } a_n - \sqrt{3}b_n = (2 - \sqrt{3})^n \text{이다.} \quad \text{……②}$$

(3)  $\frac{\textcircled{1}+\textcircled{2}}{2} : a_n = \frac{1}{2} \{ (2+\sqrt{3})^n + (2-\sqrt{3})^n \}$

$\frac{\textcircled{1}-\textcircled{2}}{2\sqrt{3}} : b_n = \frac{1}{2\sqrt{3}} \{ (2+\sqrt{3})^n - (2-\sqrt{3})^n \}$

$\therefore c_n = \frac{a_n}{b_n} = \sqrt{3} \frac{(2+\sqrt{3})^n + (2-\sqrt{3})^n}{(2+\sqrt{3})^n - (2-\sqrt{3})^n}$

이 때  $\gamma = \frac{2-\sqrt{3}}{2+\sqrt{3}}$  이라면  $\lim_{n \rightarrow \infty} c_n = \lim_{n \rightarrow \infty} \frac{\sqrt{3}(1+\gamma^n)}{1-\gamma^n} = \sqrt{3}, (\because 0 < \gamma < 1)$

9. (1)  $y_{n+1} = \frac{1}{x_{n+1}-3} = \frac{x_n-1}{2(x_n-3)} = \frac{1}{2} + \frac{1}{x_n-3} = \frac{1}{2} + y_n$

(2)  $y_{n+1} - y_n = \frac{1}{2} \therefore y_n = y_1 + \frac{1}{2}(n-1) = \frac{1}{2}(n+1), (\because y_1 = \frac{1}{x_1-3} = 1)$

(3)  $y_n = \frac{1}{x_n-3}$  에서  $x_n = \frac{1}{y_n} + 3 = \frac{2}{n+1} + 3$  이다.

$\therefore \lim_{n \rightarrow \infty} x_n = 3$

10. 특성방정식  $x = \frac{1-x}{1+x}$  에 의해 일반항을 구하면 다음과 같다.

$a_n = \frac{(a_1^2+1) + (-1)^{n+1}(a_1^2+2a_1-1)}{2(1+a_1)}$

이 식에서 수렴 또는 발산을 조사해도 되나, 이 문제는 특수한 방법을 써 보자.

$a_{n+1} + 1 = \frac{1-a_n}{1+a_n} + 1 = \frac{2}{a_n+1}$  .....①

같은 방법으로  $a_n + 1 = \frac{2}{a_{n-1}+1}$  이다. ....②

①, ② :  $a_n = a_{n-2}$

$n=2m$ 의 경우 :  $a_{2m} = a_{2m-2} = \dots = a_4 = a_2$

$n=2m+1$ 의 경우 :  $a_{2m+1} = a_{2m-1} = \dots = a_3 = a_1$

이 때 이 수열이 수렴하려면  $\lim_{m \rightarrow \infty} a_{2m} = \lim_{m \rightarrow \infty} a_{2m+1}$  이어야 하므로  $a_1 = a_2$  이다.

즉,  $a_1 = \frac{1-a_1}{1+a_1}$  일 때  $(a_1 = -1 \pm \sqrt{2})$  수렴하고 그 외에는 발산한다.

$a_1 = -1 \pm \sqrt{2}$  에서  $-1 \pm \sqrt{2}$  에 수렴(복호동순)한다.

$a_1 \neq -1 \pm \sqrt{2}$  에서 발산한다.

❖ 참고 : 앞에서 말한 경우는  $a_n = \frac{(a_1^2+1) + (-1)^{n+1}(a_1^2+2a_1-1)}{2(1+a_1)}$  에서  $a_1^2+2a_1-1=0$

인 경우 수렴함을 의미한다.

11. (1)  $a_n > \sqrt{k}$  를 수학적 귀납법으로 증명한다.

$n=1$ 에서  $a_1 > \sqrt{k}$  이고  $n=r$ 에서  $a_r > \sqrt{k}$  라 가정하자.

$n=r+1$ 에서  $a_{r+1} = \frac{1}{2} \left( a_r + \frac{k}{a_r} \right) \geq \sqrt{a_r \cdot \frac{k}{a_r}} = \sqrt{k}$  이다.

( $\because$  상승, 상가평균에서 등호가 성립하려면  $a_r = \frac{k}{a_r}$ , 즉  $a_r = \pm\sqrt{k}$ 일 경우이나  $a_r > \sqrt{k}$ 이므로 등호가 성립하는 경우는 없다)

$$(2) \quad a_n - \sqrt{k} = \frac{1}{2} \left( a_{n-1} + \frac{k}{a_{n-1}} \right) - \sqrt{k} = \frac{1}{2a_{n-1}} (a_{n-1} - \sqrt{k})^2$$

$$= \frac{a_{n-1} - \sqrt{k}}{a_{n-1}} \times \frac{a_{n-1} - \sqrt{k}}{2} < \frac{1}{2} (a_{n-1} - \sqrt{k})$$

( $\because$  (1)에서  $0 < \frac{a_{n-1} - \sqrt{k}}{a_{n-1}} = 1 - \frac{\sqrt{k}}{a_{n-1}} < 1$ )

(3) (2)에서  $0 < a_n - \sqrt{k} < \frac{1}{2} (a_{n-1} - \sqrt{k})$ , 즉  $0 < a_n - \sqrt{k} < \frac{1}{2^{n-1}} (a_1 - \sqrt{k})$ 이다.

$\therefore \lim_{n \rightarrow \infty} a_n = \sqrt{k}$

12. (1)  $u_{n+1} = a - x_{n+1}^2 = a - \frac{1}{4} \left( x_n + \frac{a}{x_n} \right)^2 = -\frac{1}{4} \left( \frac{a}{x_n} - x_n \right)^2 < 0$

(2)  $\frac{u_{n+1}}{u_n} = \left| \frac{-\frac{1}{4} \left( \frac{a}{x_n} - x_n \right)^2}{a - x_n^2} \right| = \left| \frac{a - x_n^2}{4x_n^2} \right| = \frac{x_n^2 - a}{4x_n^2}$ , ( $\because a - x_n^2 < 0$ )

$$\left| \frac{u_{n+1}}{u_n} \right| - \frac{1}{2} = \frac{x_n^2 - a}{4x_n^2} - \frac{1}{2} = -\frac{(x_n^2 + a)}{4x_n^2} < 0$$

$\therefore \left| \frac{u_{n+1}}{u_n} \right| < \frac{1}{2}$

(3) (2)에서  $0 < |u_{n+1}| < \frac{1}{2} |u_n|$ 이다.

$\therefore 0 < |u_n| < \frac{1}{2^{n-1}} |u_1|$

$\lim_{n \rightarrow \infty} \frac{|u_1|}{2^{n-1}} = 0 \quad \therefore \lim_{n \rightarrow \infty} u_n = 0$

(4)  $\lim_{n \rightarrow \infty} u_n = 0$ 이므로  $\lim_{n \rightarrow \infty} x_n^2 = a$ 이다.  $\therefore \lim_{n \rightarrow \infty} x_n = \sqrt{a}$

( $\because n=0$ 에서  $x_0 > 0$ 이고,  $n=r$ 에서  $x_r > 0$ 이라 가정하자.

$n=r+1$ 에서  $x_{r+1} = \frac{1}{2} \left( \frac{a}{x_r} + x_r \right) > \sqrt{\frac{a}{x_r} \cdot x_r} = \sqrt{a} > 0$ 이다.

따라서  $n=0, 1, 2, \dots$ 에서  $x_n > 0$ 이다)

13.  $a_{n+1} = \frac{n(n+1)}{2n-a_n}$ 의 양변을  $n+1$ 로 나누면 다음과 같다.

$$\frac{a_{n+1}}{n+1} = \frac{n}{2n-a_n} = \frac{1}{2 - \frac{a_n}{n}}$$

이 식에서  $\frac{a_n}{n} = b_n$ 이라하면,  $b_1 = a_1 = a$ 이고 다음 식이 성립한다.  $\dots\dots\dots$ ①

점화식  $b_{n+1} = \frac{1}{2-b_n}$ 에서 특성방정식  $x = \frac{1}{2-x}$ 의 근  $x=1$ 을 양변에서 빼면

다음과 같다.

$$b_{n+1}-1 = \frac{1}{2-b_n}-1 = \frac{b_n-1}{2-b_n}$$

이 식의 양변에 역수를 취하면 다음과 같다.

$$\frac{1}{b_{n+1}-1} = \frac{2-b_n}{b_n-1} = \frac{1}{b_n-1}-1$$

이 식에서  $\frac{1}{b_n-1} = c_n$ 이라면 다음 식이 성립한다. .....②

$$c_{n+1} = c_n - 1$$

$$\therefore c_n = c_1 - (n-1) = \frac{a}{a-1} - n, \quad \left( \because c_1 = \frac{1}{a-1} \right) \quad \text{.....③}$$

$$\textcircled{1}, \textcircled{2}, \textcircled{3} : a_n = nb_n = n \left( \frac{1}{c_n} + 1 \right) = \frac{n\{(2a-1) - n(a-1)\}}{a - (a-1)n}$$





# SAS/ETS® Software Reference Card

This reference card is a source of quick information. Complete descriptions are found in the *SAS/ETS User's Guide, Version 6, First Edition*.

## Statement Descriptions

The form of a SAS statement is specified with these conventions:

- KEYWORD** *parameter...[item | item | item] options;*
- where
- bold** indicates that you use exactly the same spelling and form as shown.
- italics* means that you supply your own information.
- bracketed information ([ ]) is optional. Parameters not in brackets are required.
- three periods (...) mean that more than one of the parameters preceding ... may be specified.
- vertical bar (|) separating keywords and options means to choose only one of the terms separated by the vertical bars.
- options* are keyword options specific to a particular SAS statement.

## Procedures Used in SAS/ETS Software

**PROC ARIMA** [DATA=SASdataset OUT=SASdataset];  
 The IDENTIFY, ESTIMATE, and FORECAST statements, when used, must be specified in that order.  
 [BY variables;]  
 [ESTIMATE [ALTPARM AR=numbers  
 BACKLIM=-number CONVERGE=number  
 DELTA=number GRID GRIDVAL=number  
 INITVAL=(values variable...)  
 INPUT=(form1 variable1 form2 variable2...)  
 MA=numbers MAXITER | MAXIT=number  
 METHOD=ML | ULS | CLS  
 MU=number NOCONSTANT | NOINT  
 NODF NOEST NOLS NOPRINT  
 NOSTABLE P=lag | (lag,...,lag)...(lag,...,lag)  
 PLOT PRINTALL Q=lag | (lag,...,lag)...(lag,...,lag)  
 SINGULAR=number];  
 [FORECAST [BACK=number ID=variable  
 INTERVAL=interval | number INTPER=number  
 LEAD=number NOOUTALL NOPRINT  
 OUT=SASdataset PRINTALL];  
 [IDENTIFY VAR=variable [(d<sub>1</sub>, d<sub>2</sub>,...,d<sub>k</sub>)] [CENTER CLEAR

SAS/ETS is a registered trademark of SAS Institute Inc., Cary, NC, USA.

CROSSCORR | CROSSCOR | CROSS=(variable [(d<sub>1</sub>, d<sub>2</sub>,...,d<sub>k</sub>)...variable(d<sub>1</sub>, d<sub>2</sub>,...,d<sub>k</sub>)])  
 DATA=SASdataset  
 NLAG=number NOMISS NOPRINT  
 OUTCOV=SASdataset];  
**PROC AUTOREG** [DATA=SASdataset  
 OUTEST=SASdataset];  
 [BY variables;]  
 [label:] MODEL dependent=regressors [/ ALL  
 BACKSTEP COEF CONVERGE=number CORRB  
 COVB GINV ITER ITPRINT MAXITER=number  
 METHOD=ML | ULS | YW  
 NLAG=p | (numberlist)  
 NOINT NOMISS NOPRINT PARTIAL  
 SLSTAY=value];  
 [OUTPUT [OUT=SASdataset ALPHACLI=number  
 ALPHACLM=number CONSTANT=variable  
 LCL=name LCLM=name PREDICTED | P=name  
 PREDICTEDM | PM=name RESIDUAL | R=name  
 RESIDUALM | RM=name TRANSFORM=variables  
 UCL=name UCLM=name];  
**PROC CITIBASE** [ASCII BEGINYR=n  
 DBNAME="pathname" EBCDIC ENDYR=n  
 FORMAT=OLD INFILE=fileref LENGTH=n  
 NOPRINT OUT=SASdataset  
 OUTCONT=SASdataset OUTSELECT  
 TYPE=frequency];  
 [SELECT specifications;]  
 where specifications can be a name, an alphabetic range, an order range, or a prefix.  
**PROC COMPUTAB** [CDEC=d CSPACE=n  
 CWIDTH=w DATA=SASdataset  
 FUZZ=nnnn INITMISS NOPRINT NORTR  
 NOTRANSPOSE | NOTRANS OPTIONS  
 OUT=SASdataset RTS=n SUMONLY];  
 [BY variables;]  
 [CELL names / FORMAT=format;]  
 [COLxxxx: statements for column variables;]  
 [COLUMNS | COL names [/ CHAR 'text'  
 FORMAT | F=format \_LABEL\_ LJC  
 MTITLE | M='text' +n \_NAME\_ NOPRINT  
 NOZERO \_PAGE\_ \_TITLES\_ | \_TITLE\_  
 ZERO='text'];  
 where names can be of the form name1, name2,...namen.  
 [INIT terms initialvalues;]  
 [INIT anchorname [locatorname] values  
 [locatorname values] +n n\*value;]  
 [program statements;]  
 [ROWxxxx: statements for row variables;]  
 [ROWS | ROW names [/ CHAR 'text'  
 FORMAT | F=format \_LABEL\_ LJC +n \_NAME\_  
 NOPRINT NOZERO OL | DOL  
 OVERPRINT | OVP \_PAGE\_ SKIP UL | DUL  
 ZERO='text'];  
 where names can be of the form name1, name2,...namen.  
 [SUMBY variables;]  
**PROC EXPAND** [DATA=SASdataset FACTOR=n | n:m  
 FROM=interval OUT=SASdataset  
 OUTEST=SASdataset TO=interval];  
 [BY variables;]  
 [CONVERT variables [=newname] [/ METHOD=option  
 OBSERVED= attribute | (from-attribute, to-attribute)];  
 [ID variable;]  
**PROC FORECAST** [ALPHA=n  
 AR | NLAGS=n DATA=SASdataset  
 INTERVAL=interval INTPER=n LEAD=n  
 METHOD=STEPAR | EXPO | WINTERS |  
 ADDWINTERS OUT=SASdataset  
 OUTACTUAL OUTALL OUTEST=SASdataset  
 OUTFULL OUTLIMIT OUTRESID OUTSTD  
 OUT1STEP SEASONS=interval |  
 (interval<sub>1</sub>[interval<sub>2</sub>[interval<sub>3</sub>]]) | n | (n<sub>1</sub>[n<sub>2</sub>[n<sub>3</sub>]])  
 SINGULAR=n SINTPER=m | (m<sub>1</sub>[m<sub>2</sub>[m<sub>3</sub>]])  
 SLENTRY=n SLSTAY=n START=n TREND=n  
 WEIGHT=ω | (ω<sub>1</sub>[ω<sub>2</sub>[ω<sub>3</sub>]]) ZEROMISS];

```

[BY variables;]
[ID variables;]
[VAR variables;]
PROC MODEL [BLOCK COLLIN CONVERGE=n CORR
CORRB CORRS COVB COVS DATA=SASdataset
DETAILS DW DYNAMIC FLOW FORECAST
FSRSQ GRAPH I ITALL ITDETAILS ITOLS ITPRINT
ITSUR IT2SLS IT3SLS JACOBI LIST LISTCODE
LISTDEP LISTDER MAXERRORS=n
MAXITER=n MAXSUBITER=n
METHOD=GAUSS | MARQUARDT
MODEL=modelname | (model-list)
NAHEAD=n NESTIT NEWTON NOPRINT
NOSTORE N2SLS | 2SLS N3SLS | 3SLS OLS
OUTACTUAL OUTALL OUTCOV OUTERRORS
OUTLAGS OUTMODEL=model-name OUTPREDICT
OUTRESID PRINTALL SEED SEIDEL SIMULATE
SINGLE | ONEPASS SINGULAR=n SOLVEPRINT
START=s STATIC STATS SUR THEIL TRACE
TYPE=name VARDEF XPX XREF];
[ARRAY arrayname [{dimensions}][${variables and constants}];
[BY variables;]
[CONTROL variable [value]...;]
[ENDOGENOUS | ENDO variable [initial values]...;]
[EXOGENOUS variable [initial values]...;]
FIT [equations] [DROP=(parameters)
PARMS=(parameter [values]...)
START=(parameter values...) / BLOCK COLLIN
CONVERGE=n CORR CORRB CORRS
COV COVB COVS
DATA=SASdataset DETAILS DW
ESTDATA=SASdataset FSRSQ I ITALL ITDETAILS
ITOLS ITPRINT ITSUR IT2SLS IT3SLS FLOW GRAPH
LIST LISTCODE LISTDEP LISTDER MAXERRORS=n
MAXITER=n MAXSUBITER=n
METHOD=GAUSS | MARQUARDT NESTIT
NOPRINT N2SLS | 2SLS N3SLS | 3SLS OLS
OUTACTUAL OUTALL OUTCOV | COVOUT
OUTLAGS OUT=SASdataset OUTEST=SASdataset
OUTPREDICT OUTRESID OUTS=SASdataset
OUTUSED=SASdataset PRINTALL
SDATA=SASdataset SUR TRACE TYPE=name
VARDEF=N | WGT | DF XPX XREF];
[INCLUDE modelfiles;]
[INSTRUMENTS [instruments _EXOG_
EXCLUDE=(parameters)
[/ NOINTERCEPT | NOINT
STOREINST | STORE];]
[LABEL variable='label'...;]
[OUTVARS variables;]
[PARAMETERS | PARMS variable [value]...;]
[program statements;]
[RESET [options] [PURGE];]
[RANGE variable [=first] [TO last];]
[RETAIN variables initial-values...;]
[SOLVE [variables] [SATISFY=equations]
[/ BLOCK CONVERGE=n DATA=SASdataset
DETAILS DYNAMIC ESTDATA=SASdataset FLOW
FORECAST GRAPH ITPRINT JACOBI LIST
LISTCODE LISTDEP LISTDER MAXERRORS=n
MAXITER=n NAHEAD=n NEWTON NOPRINT
OUT=SASdataset OUTACTUAL OUTALL
OUTERRORS OUTLAGS OUTPREDICT OUTRESID
PRINTALL RANDOM=n SDATA=SASdataset
SEED=n SEIDEL SIMULATE SINGLE | ONEPASS
SOLVEPRINT START=s STATIC STATS THEIL
TRACE TYPE=name XREF];]
[VAR variable [initial values]...;]
[WEIGHT variable;]
PROC MORTGAGE [AMOUNT | A=d COMPOUND=time-
unit FIRSTPAYMENT | FIRSTP=yyyy:n
INTERVAL=payment-interval LIFE | L=n
MONTHS | M=m NOPRINT NYEARS=n
OUT=SASdataset OUTSUM=SASdataset
PAYMENT | P=p PRINTLOG | PRINTL
RATE | R=r YEARLY];

```

```

PROC PDLREG [DATA=SASdataset];
[BY variables;]
[label:] MODEL dependent=effects [/ ALL
CORRB COVB I METHOD=ML | ULS | YW
NLAG=n | (numberlist) NOINT NOPRINT PARTIAL
STB XPX];
where effects can be of the form variable(p,d,l,q).
[OUTPUT [OUT=SASdataset ALPHACLI=number
ALPHACL=number CONSTANT=variable
LCL=name LCLM=name PREDICTED | P=name
PREDICTEDM | PM=name RESIDUAL | R=name
RESIDUALM | RM=name TRANSFORM=variables
UCL=name UCLM=name];]
[RESTRICT restriction1, restriction2, ...restrictionk;]
PROC SIMLIN [DATA=SASdataset EST=SASdataset
ESTPRINT INTERIM=n NOPRINT NORED
OUTEST=SASdataset START=number TOTAL
TYPE=value];
[BY variables;]
ENDOGENOUS variables;
EXOGENOUS variables;
[ID variables;]
[LAGGED lagvariable endogenousvariable number...;]
[OUTPUT [OUT=SASdataset PREDICTED | P=names
RESIDUAL | R=names];]
PROC SPECTRA [A ADJMEAN | CENTER
COEF CROSS DATA=SASdataset K
OUT=SASdataset P PH S WHITEST];
[BY variables;]
VAR variables;
[WEIGHTS constants;]
PROC STATESPACE [ARMAX=value
BACK=number CANCORR COVB
DATA=SASdataset DETTOL | PARMTOL=number
DIMMAX=number INTERVAL=interval
INTPER=number ITPRINT KLAG=number
LAGMAX=number LEAD=number MAXIT=number
MINIC NOCENTER NOEST OUT=SASdataset
OUTAR=SASdataset OUTMODEL=SASdataset
PASTMIN=number PRINT
PRINTOUT=LONG | NONE | SHORT
RESIDEST SIGCORR=number
SINGULAR=number];
[BY variables;]
[FORM variable value...;]
[ID variable;]
[INITIAL F(row,column)=value
F(row,column)=value...G(row,column)=value
G(row,column)=value...;]
[RESTRICT F(row,column)=value
F(row,column)=value...G(row,column)=value
G(row,column)=value...;]
VAR variable (difference, difference,...)...;
PROC SYSLIN [ALL ALPHA=value CONVERGE=number
COVOUT COV3OUT DATA=SASdataset
DFNO EPSILON=number FIML FIRST ITPRINT
ITSUR IT3SLS K=value LIML
MAXITER | MAXIT=number MELO NODFS
NOINCLUDE NOPRINT OLS OUT=SASdataset
OUTEST=SASdataset OUTSSCP=SASdataset
REDUCED SDIAG SIMPLE SINGULAR=number SUR
USSCP USSCP2 VARDEF=DF | N | WDF | WEIGHT |
WDF | WEIGHT | WGT 2SLS 3SLS];
[label:] MODEL response=regressors [/ ALL
ALPHA=value CORR COVB DW I K=value
NOINT NOPRINT OVERID PLOT STB UNREST XPX];
[BY variables;]
ENDOGENOUS variables;
[ID variables;]
[IDENTITY identity1, identity2, ...identityk];
[INSTRUMENTS variables;]
[OUTPUT [PREDICTED | PRED | P=variable
RESIDUAL | RESID | R=variable];]
[RESTRICT restriction1, restriction2, ...restrictionk;]
[SRESTRICT restriction1, restriction2, ...restrictionk;]
[STEST equation1, equation2, ...equationk [/ PRINT];]
[TEST equation1, equation2, ...equationk [/ PRINT];]

```

```

[VAR variables;]
[WEIGHT variable;]
PROC X11 [DATA=SASdataset YRAHEADOUT];
With PROC X11 you must specify either the
MONTHLY or QUARTERLY statement.
[BY variables;]
[ID variables;]
[MACURVES [month=specification STABLE];]
MONTHLY [ADDITIVE CHARTS=ALL | NONE | STANDARD
DATE=variable END=mmmyy EXCLUDE=n
FULLWEIGHT=n LENGTH NDEC=number
PMFACTOR=variable
PRINTOUT=FULL | LONG | NONE | STANDARD
START=mmmyy SUMMARY TDCOMPUTE=year
TDREGR=ADJUST | NONE | PRINT | TEST
TRENDADJ TRENDMA=9 | 13 | 23
ZEROWEIGHT=n];
[OUTPUT [OUT=SASdataset table=variables...];]
[PDWEIGHTS [day=n];]
QUARTERLY [ADDITIVE CHARTS=ALL | NONE | STANDARD
DATE=variable END='yyQq'
FULLWEIGHT=n NDEC=number
PRINTOUT=FULL | LONG | NONE | STANDARD
START='yyQq' SUMMARY TRENDADJ
ZEROWEIGHT=n];
[TABLES tablenames;]
[VAR variables;]

```



SAS Institute Inc.  
 SAS Circle  Box 8000  
 Cary, NC 27512-8000  
 Phone (919) 677-8000  
 Fax (919) 677-8123



## 부록 ④

# SAS/IML™ Software Reference Card

This Reference Card is a source of quick information that can be detached if desired. Complete descriptions of all statements, operators, functions, and calls are found in the *SAS/IML User's Guide, Release 6.03 Edition*.

### Operators

| Operation                | Symbol | Syntax type | Data type |
|--------------------------|--------|-------------|-----------|
| sign reverse             | -      | prefix      | num       |
| addition                 | +      | infix       | num       |
| subtraction              | -      | infix       | num       |
| index creation           | :      | infix       | num       |
| matrix multiplication    | *      | infix       | num       |
| element multiplication   | #      | infix       | num       |
| direct product           | @      | infix       | num       |
| matrix power             | **     | infix       | num       |
| element power            | ##     | infix       | num       |
| division                 | /      | infix       | num       |
| horizontal concatenation |        | infix       | both      |
| vertical concatenation   | //     | infix       | both      |
| element maximum          | <>     | infix       | both      |
| element minimum          | ><     | infix       | both      |
| and                      | &      | infix       | num       |
| or                       |        | infix       | num       |
| not                      | ^      | prefix      | num       |
| less than                | <      | infix       | both      |
| greater than             | >      | infix       | both      |
| equal to                 | =      | infix       | both      |
| less or equal            | <=     | infix       | both      |
| greater or equal         | >=     | infix       | both      |
| not equal                | ^=     | infix       | both      |
| transpose                | \      | postfix     | both      |
| subscript                | [ ]    | postfix     | both      |

### Operator Precedence

#### Group I (highest priority)

^ \ subscripts -(prefix) ## \*\*

#### Group II

\* # <> >< / @

#### Group III

+ -

#### Group IV

|| // :

#### Group V

< <= > >= = ^ =

#### Group VI

&

#### Group VII (lowest priority)

|

### Subscript Reduction Operators

|     |                  |
|-----|------------------|
| +   | addition         |
| #   | multiplication   |
| <>  | maximum          |
| ><  | minimum          |
| <:> | index of maximum |
| >:< | index of minimum |
| :   | mean             |
| ##  | sum of squares   |

### Functions

| Function                      | Example                                | Arg type |
|-------------------------------|----------------------------------------|----------|
| all elements nonzero?         | B=ALL(A);                              | num      |
| any elements nonzero?         | I=ANY(A);                              | num      |
| combine diagonally            | C=BLOCK(A, B);                         | num      |
| block transpose               | Z=BTRAN(N, NX, MX);                    | num      |
| get ASCII byte by number      | CHAR=BYTE(number);                     | char     |
| change substrings             | CALL CHANGE(matrix, old, new);         | char     |
| numeric to character          | C=CHAR(NUM, W, D);                     | num      |
| choose elements conditionally | X=CHOOSE(CONDITION, TRESULT, FRESULT); | both     |
| element concatenation         | C=CONCAT(A, B);                        | char     |

(continued)

## Functions (continued)

| Function                                       | Example                                                   | Arg type |
|------------------------------------------------|-----------------------------------------------------------|----------|
| contents of data set                           | R=CONTENTS( <i>lib, mem</i> );                            | char     |
| character reshaping                            | R=CSHAPE( <i>matrix, nrow, ncol, size&lt;, pad&gt;</i> ); | char     |
| cumulative sum                                 | C=CUSUM( <i>matrix</i> );                                 | num      |
| find SAS data sets                             | R=DATASETS( <i>libname</i> );                             | char     |
| delete a data set                              | CALL DELETE( <i>lib, mem</i> );                           |          |
| create design matrix                           | X=DESIGN(A);                                              | num      |
| create full-rank design matrix                 | X=DESIGNF(A);                                             | num      |
| determinant                                    | D=DET(A);                                                 | num      |
| diagonal                                       | D=DIAG(A);                                                | num      |
| index generation                               | I=DO( <i>i1, i2, inc</i> );                               | num      |
| reduce to row echelon form                     | E=ECHOLON(A);                                             | num      |
| eigenvalues and vectors                        | CALL EIGEN(M, E, A);                                      | num      |
| eigenvalues                                    | H=EIGVAL(A);                                              | num      |
| create eigenvector matrix                      | E=EIGVEC(A);                                              | num      |
| compute eigenvalues and eigenvectors           | CALL GENEIG(M, E, A, B);                                  | num      |
| generalized inverse                            | G=GINV(A);                                                | num      |
| compute Gram-Schmidt orthonormal factorization | CALL GSORTH(P, T, LINDEP, A);                             | num      |
| Cholesky root                                  | U=HALF(A);                                                | num      |
| horizontal direct product                      | H=HDIR(A, B);                                             | num      |
| reduce to Hermite normal form                  | H=HERMITE(A);                                             | num      |
| solve homogenous system of linear equations    | H=HOMOGEN(A);                                             | num      |
| identity matrix                                | B=I(A);                                                   | num      |
| insert a new element                           | R=INSERT(X, Y, ROW, COL);                                 | both     |
| matrix inverse                                 | I=INV(A);                                                 | num      |

(continued)

## Functions (continued)

| Function                             | Example                                                | Arg type |
|--------------------------------------|--------------------------------------------------------|----------|
| update a matrix inverse              | R=INVUPDT(A, X, W);                                    | num      |
| perform iterative proportional fit   | CALL IPF(FIT, STATUS, DM, TABLE, CONFIG, INITAB, MOD); | num      |
| matrix of identical values           | B=J(NR, NC, VAL);                                      | num      |
| solve linear complementarity problem | CALL LCP(RC, W, Z, M, Q, EPSILON);                     | num      |
| length of longest string             | B=LENGTH(A);                                           | char     |
| location of nonzeros                 | B=LOC(A);                                              | num      |
| solve linear program                 | CALL LP(RC, X, DUAL, A, B, CNTL, U, L, BASIS);         | num      |
| evaluate margin totals               | CALL MARG(LOCMAR, MARGINAL, DIM, TABLE, CONFIG);       | num      |
| maximum value                        | B=MAX(A);                                              | both     |
| minimum value                        | B=MIN(A);                                              | both     |
| name of arguments                    | N=NAME(A, ...);                                        | both     |
| number of columns                    | N=NCOL(A);                                             | both     |
| length of element                    | B=NLENG(A);                                            | both     |
| number or rows                       | K=NROW(A);                                             | both     |
| character to numeric                 | N=NUM(A);                                              | char     |
| optimal scaling                      | OPSCAL(MLEVEL, QUANTI);                                | num      |
| orthogonal polynomials               | A=ORPOL(X, MD, W);                                     | num      |
| produce scatter plot                 | CALL PGRAF(XY, ID, YLABEL, TITLE);                     |          |
| ranking values                       | R=RANK(A);                                             | num      |
| ranking with ties averaged           | R=RANKTIE(X);                                          | num      |
| remove element                       | R=REMOVE( <i>matrix, indices</i> );                    | both     |
| rename a SAS data set                | CALL RENAME( <i>lib, mem, new</i> );                   | char     |
| repeat a matrix                      | R=REPEAT( <i>matrix, rows, cols</i> );                 | num      |
| Cholesky decomposition               | U=ROOT(A);                                             | num      |
| row concatenation                    | R=ROWCATB(A, B);                                       | char     |
| row concatenation                    | R=ROWCATC(A, B);                                       | char     |
| set difference                       | D=SETDIF(A, B);                                        | both     |
| reshape                              | B=SHAPE(A, NROW, NCOL);                                | both     |

(continued)

## Functions (continued)

| Function                     | Example                 | Arg type |
|------------------------------|-------------------------|----------|
| solve linear system          | X=SOLVE(A,C);           | num      |
| produce tone                 | CALL SOUND(freq<,dur>); | num      |
| convert symmetric to square  | SQR=SQRSYM(SYM);        | num      |
| sum of squares               | S=SSQ(A);               | num      |
| find out matrices in storage | A=STORAGE();            | char     |
| substring                    | A=SUBSTR(B,START,LEN);  | char     |
| sum                          | S=SUM(A);               | num      |
| singular value decomp        | CALL SVD(U,M,V,A);      | num      |
| sweep                        | B=SWEEP(A,1:5);         | num      |
| convert square to symmetric  | SYM=SYMSQR(sqr);        | num      |
| trace                        | T=TRACE(X);             | num      |
| return type of operand       | T=TYPE(A);              | char     |
| union set operator           | U=UNION(matrix<,...>);  | both     |
| form unique set, union       | U=UNIQUE(matrix<,...>); | both     |
| value assignment             | CALL VALSET(A,B);       | char     |
| value lookup                 | V=VALUE(A);             | char     |
| diagonal to vector           | V=VECDIAG(A);           | num      |
| set intersection             | R=XSECT(A,B);           | both     |

## Scalar Functions

| Function                    | Usage                  |
|-----------------------------|------------------------|
| absolute value              | A=ABS(X);              |
| arc-cosine                  | S=ARCOS(X);            |
| arc-sine                    | S=ARSIN(X);            |
| arc-tangent                 | S=ATAN(X);             |
| cosine                      | S=COS(X);              |
| exponential                 | B=EXP(X);              |
| integer value (truncation)  | I=INT(X);              |
| natural logarithm           | Y=LOG(X);              |
| modulo (remainder)          | Y=MOD(X,D);            |
| chi-square distribution     | P=PROBCHI(X,DF,NC);    |
| F distribution              | P=PROBF(X,NDF,DDF,NC); |
| inverse normal distribution | P=PROBIT(X);           |
| normal distribution         | P=PROBNORM(X);         |
| t distribution              | P=PROBT(X,DF,NC);      |
| normal random number        | N=NORMAL(SEED);        |
| sine                        | S=SIN(X);              |
| square root                 | S=SQRT(X);             |
| tangent                     | S=TAN(X);              |
| uniform random number       | U=UNIFORM(SEED);       |

## General Purpose Commands

| Command                     | Example               |
|-----------------------------|-----------------------|
| module definition finish    | FINISH;               |
| free storage                | FREE names;           |
| storage retrieve            | LOAD names;           |
| storage remove              | REMOVE names;         |
| set options                 | RESET options;        |
| print matrices              | PRINT names messages; |
| exit from IML               | QUIT;                 |
| module execute              | RUN name arguments;   |
| show status                 | SHOW items;           |
| module definition start     | START name arguments; |
| storage store               | STORE names;          |
| associates print attributes | MATTRIB name;         |

## Control Statements

|                  |                                       |
|------------------|---------------------------------------|
| IF-THEN/ELSE     | conditional statements                |
| DO-END           | grouping                              |
| iterative DO-END | iteration                             |
| GOTO             | transfer control                      |
| LINK             | goto with return link                 |
| RETURN           | return to caller                      |
| STOP             | stop execution                        |
| ABORT            | stop execution and exit IML           |
| PAUSE            | pause execution                       |
| RESUME           | resume from pause                     |
| PUSH             | push lines into command input stream  |
| QUEUE            | queue lines into command input stream |
| EXECUTE          | execute lines immediately             |

## RESET Options

| Option    | Effect                                                                                                           |
|-----------|------------------------------------------------------------------------------------------------------------------|
| PRINT     | prints all results automatically.                                                                                |
| NAME      | controls printing of matrix name and default row and column names.                                               |
| FLOW      | traces flow of execution with messages.                                                                          |
| DETAILS   | causes more details to be shown.                                                                                 |
| FUZZ      | specifies to print very small numbers as zero.                                                                   |
| FW=       | specifies field width for printing matrices.                                                                     |
| STORAGE=  | specifies the matrix library storage.                                                                            |
| DEFLIB=   | is the default libname for SAS data sets.                                                                        |
| LOG       | routes output to log rather than output scroll.                                                                  |
| DEBUG     | is for IML debugging only.                                                                                       |
| ALL       | prints intermediate results automatically.                                                                       |
| CENTER    | centers output.                                                                                                  |
| LINESIZE= | specifies linesize.                                                                                              |
| PAGESIZE= | specifies pagesize.                                                                                              |
| SPACES=   | specifies number of spaces between output.                                                                       |
| SPILL     | specifies that a spill file be used.                                                                             |
| CASE      | specifies that character strings are converted to uppercase before comparisons in data processing are performed. |
| BREAK     | enables a break exit and escape mechanism                                                                        |

**Data Processing Commands****Opening and Closing**

|        |                                                 |
|--------|-------------------------------------------------|
| EDIT   | opens a SAS data set for read and write access. |
| CREATE | opens a new SAS data set.                       |
| CLOSE  | closes a SAS data set or external file.         |
| SETIN  | selects an open data set for input.             |
| SETOUT | selects an open data set for output.            |
| USE    | opens a SAS data set for reading.               |

**Showing and Resetting**

|               |                                           |
|---------------|-------------------------------------------|
| SHOW DATASETS | shows data sets currently active.         |
| SHOW CONTENTS | shows contents of current data set.       |
| RESET DEFLIB= | sets up the default libname.              |
| RESET CASE    | converts comparison strings to uppercase. |
| RESET SPILL   | allows spill file for large data sets.    |

**Input and Output**

|         |                                             |
|---------|---------------------------------------------|
| LIST    | displays records.                           |
| READ    | reads data.                                 |
| REPLACE | replaces data in a data set.                |
| APPEND  | adds records at the end of the data set.    |
| FIND    | finds records satisfying conditions.        |
| DELETE  | marks records as deleted.                   |
| FORCE   | forces out data to a data set.              |
| PURGE   | purges deleted records from a SAS data set. |

**External Files**

|           |                                    |
|-----------|------------------------------------|
| INFILE    | opens to an external input file.   |
| INPUT     | reads from an external input file. |
| FILE      | opens to an external output file.  |
| PUT       | writes to an external output file. |
| CLOSEFILE | closes an external file.           |

**Applications**

|         |                                                 |
|---------|-------------------------------------------------|
| INDEX   | indexes a data set.                             |
| SORT    | sorts a data set.                               |
| SUMMARY | obtains summary statistics from a SAS data set. |

**Display Commands**

|         |                              |
|---------|------------------------------|
| WINDOW  | opens a window on display.   |
| DISPLAY | displays fields in a window. |

**Call Routines and Functions**


|             |                                                     |
|-------------|-----------------------------------------------------|
| DATASETS    | is a function to obtain members in a data library.  |
| CONTENTS    | is a function to obtain variables in a member.      |
| CALL RENAME | is a call routine to rename a SAS data set.         |
| CALL DELETE | is a call routine to delete (erase) a SAS data set. |

**Graphics Commands**

|          |                                           |
|----------|-------------------------------------------|
| GBLKVP   | defines blanking viewport.                |
| GBLKVPD  | deletes the blanking viewport.            |
| GCLOSE   | closes the graphics segment.              |
| GDELETE  | deletes a graphics segment.               |
| GDRAW    | draws a polyline.                         |
| GDRAWL   | draws individual lines.                   |
| GGRID    | draws a grid.                             |
| GINCLUDE | includes a graphics segment.              |
| GOPEN    | opens a graphics segment.                 |
| GPIE     | draws pie slices.                         |
| GPIEXY   | converts from polar to world coordinates. |
| GPOINT   | marks a point.                            |
| GPOLY    | fills a polygon.                          |
| GPORT    | changes the viewport.                     |
| GPORTPOP | pops the viewport.                        |
| GPORTSTK | stacks the viewport.                      |
| GSCALE   | gets round numbers for labeling axis.     |
| GSCRIPT  | writes with fonts.                        |
| GSET     | sets attribute.                           |
| GSHOW    | shows the graph.                          |
| GSTART   | initializes graphics.                     |
| GSTOP    | stops graphics.                           |
| GSTRLEN  | finds the string length.                  |
| GTEXT    | places text horizontally on the graph.    |
| GVTEXT   | places text vertically on the graph.      |
| GWINDOW  | defines the data window.                  |
| GXAXIS   | draws a horizontal axis.                  |
| GYAXIS   | draws a vertical axis.                    |

**Time Series Functions**

|          |                                                             |
|----------|-------------------------------------------------------------|
| ARMACOV  | computes the autocovariance of an ARMA.                     |
| ARMALIK  | computes the log-likelihood function for an ARMA model.     |
| COVLAG   | computes autocovariance estimates for a vector time series. |
| FFT      | performs finite Fourier transform routine.                  |
| HANKEL   | generates Hankel and block-Hankel matrices.                 |
| IFFT     | performs inverse finite Fourier transform.                  |
| PRODUCT  | multiplies matrices of polynomials.                         |
| RATIO    | computes terms of the ratio of two matrices of polynomials. |
| TOEPLITZ | generates Toeplitz and block-Toeplitz matrices.             |

 SAS Institute Inc.  
SAS Circle □ Box 8000  
Cary, NC 27512-8000  
Phone (919) 467-8000  
Fax (919) 469-3737

Copyright © 1988 by SAS Institute Inc.



색 인

---

---



## 국 문 색 인

**ㄱ** =====  
 가능일반화최소제곱추정값 227  
 가로결합연산자 75  
 가역성조건 819, 876  
 가역적 520, 577  
 가역적 VMA모형 576  
 Gauss-Markov가정 226  
 간섭시계열모형 420  
 간소화된 사다리꼴 형태 110  
 건너뛰기 136  
 결정적 변수 349  
 결합정상적 212, 893  
 경계기법 112  
 계단입력함수 422  
 계열상관 354  
 계획행렬 115  
 공능적 494, 497  
 공능적 관계 494  
 공능적 회귀 499  
 공분산발생함수  $g(z)$  588  
 공분산추정행렬  $\Sigma$ 의 불편추정행렬 733  
 공분산추정행렬의 편의 737  
 공분산행렬 821  
 과다차분 471, 476, 497  
 교차공분산함수 212, 513  
 교차상관계수 513  
 교차상관관계 282  
 교차상관함수 212  
 구조특성벡터 828  
 국부적 환경 125  
 균형오차 498  
 Granger의 표현정리 499  
 Gram-Schmidt의 정규직교분해 119  
 Grunfeld의 시계열데이터 25

**ㄴ** =====  
 내부파일 149  
 넓은 의미로 정상적 512  
 누적시계열데이터 495  
 능형회귀추정법 242  
  
**ㄷ** =====  
 다공선성 242  
 다변량백색잡음과정의 표본자기상관행렬함수  
     762  
 다변량부분자기상관행렬 739  
 다변량시계열데이터 13, 14  
 다변량시계열모형 511  
 다변량시계열분석 14  
 다변량시스템 891  
 다변량일반선형과정 519  
 다변량정규백색잡음과정 893  
 다변량정규분포 883  
 단기기억 494  
 단변량시계열데이터 13  
 단순 Trench-Zohar알고리즘의 확장 559  
 단위근검정 494  
 단위행렬 102  
 단일검정 794  
 단일입력단일출력시스템 200  
 대수우도함수 830  
 대칭인 블록토우플리츠행렬의 LU분해를 위한  
     알고리즘 575  
 대화적 187  
 대화형 140  
 대화형시스템 57  
 더하기연산자 71  
 데이터공간 821  
 데이터처리명령어 149

도표적 표현 842, 782

동시검정 794

동태모형 421

Do함수 84

D배열 258

d변량 VAR(p)과정 547

d변량 VARMA(p, q)과정 610

d변량 VARMA(p, q)과정의 자기공분산행렬함수 625

d차원결합정상성 512

ㄹ

Levinson-Durbin알고리즘 227, 690

LOG윈도우 56

LEAD 805

READ-INTO스테이트먼트 677

RESET스테이트먼트 61

RESTRICT스테이트먼트 792

Lyapunov-Kalman-Stein의 정리 552

ㄴ

떡등행렬 766

떡함수인자 263

모듈 123, 570

모퉁이방법 251

모형의 검진 281

Moore-Penrose일반역행렬 111

무한기억예측 409

무한차수의 떡함수 658

문자조작서브루틴 118

문자조작함수 107

ㄷ

Bartlett검정 827

Box-Jenkins방법 245

반복비교적합 119

반복처리 132

반복필터 895

배치처리 59

별칭함수식별방법 753

범위피연산자 157, 186

변환행렬 778

보통최소제곱추정법 354

부분공간 825, 879

부분시차상관행렬 738

부분자기상관계수의 확장 738

부분자기회귀행렬 738

부행렬 84, 86

부호바꾸기연산자 71

분산의 이질성 838

분포수렴 733

VAR계수추정행렬들의 성질 731

VAR과정의 자기공분산행렬 544

VAR구문 150, 161

VAR모형 식별방법들의 비교 761

VAR모형의 검진 762

VAR( $\infty$ )모형 520, 521

VAR스테이트먼트 838

VAR(p)과정의 역특성식 525

VAR(p)과정의 자기공분산행렬 544

VAR특성방정식 876

VAR특성행렬 818

VARMA모형 816, 605

VARMA모형의 비유일성 658

VARMA모형의 평균벡터 605

VARMA(p, q)모형 819

VARMA(q+1, q)모형의 자기공분산행렬함수 631

VMA계수행렬 599, 664

VMA계수행렬을 구하는 알고리즘 599  
 VMA모형 576  
 VMA( $\infty$ )모형 519  
 VMA특성방정식 876  
 VMA특성행렬 818  
 블록대각행렬 101  
 블록LU분해 574  
 블록토우플리츠행렬의 분해를 위한 알고리즘 574  
 블록행켈행렬 822  
 비결정적 349  
 비교연산자 163  
 비음정치행렬 111  
 비음치행렬 765  
 비정상다변량시계열데이터 512  
 비조건부예측 409  
 비조건부최소제곱추정량 275  
 빈 행렬 84  
 빼기연산자 72

人 

---

사영 879  
 사전백색화 233  
 3배열방법 258  
 3변량 VAR(1)과정 553  
 3변량 VAR(1)모형 532, 535, 547  
 3변량 VARMA(1, 2)과정 617  
 3변량 VARMA(1, 2)모형 635  
 3변량 VMA(1)과정 586, 590  
 3변량 VMA(2)모형 595  
 상삼각행렬 112  
 상태 778  
 상태공간모형 778, 816  
 상태공간모형화 777, 815  
 상태방정식 778

상태벡터 778, 820, 823  
 서브루틴 91, 117  
 선형대수서브루틴 118  
 선형대수함수 110  
 선형최소제곱오차예측벡터 893  
 세로결합연산자 75  
 SAS데이터라이브러리 143  
 SAS데이터세트 149  
 SAS/IML 저장라이브러리 143  
 SAS/IML 함수 91  
 SAS/IML 환경 53, 57  
 소과법 835  
 쇄신벡터 778  
 수열함수 105  
 수축요인 243  
 순간파동입력함수 422  
 순서함수 106  
 스칼라함수 92  
 스코올구문 161  
 시간에 대해 불변 200  
 시계열서브루틴 123  
 시계열함수 116  
 쓸기연산자 112

○ 

---

OUTPUT윈도우 56  
 안정성 200  
 안정적 200  
 R배열 258  
 약하게 정상적 512  
 양정치행렬 111, 118  
 APPEND스테이트먼트 678  
 Engle-Granger의 2단계추정법 500  
 Engle-Yoo의 3단계추정법 501  
 S배열 258

- 역대수변환 929
- 영숫자 70
- 예측 287, 805
- 예측공간 822
- 예측벡터 823
- 예측시스템 13
- 5변량 VMA(1)모형 584
- 오염 225
- 오차수정모형 498
- 완전랭크 115
- 외부파일 149, 189
- 우도비검정방법 745
- 원소곱하기연산자 77
- 원소기준연산자 71
- 원소기준연산자들 77
- 원소나누기연산자 77
- 원소논리곱연산자 82
- 원소논리부연산자 83
- 원소논리연산자들 81
- 원소논리합연산자 82
- 원소떡연산자 78
- 원소비교연산자들 80
- 원소최대연산자 78
- 원소최소연산자 78
- Wiener-Khinchine 관계 558
- 유색잡음 226
- 유연성 187
- USE스테이트먼트 676
- 유한기억예측 409
- 유한표본성질들 731
- Yule-Walker 방정식 220, 226
- Yule-Walker 추정량의 성질 734
- Yule-Walker 추정법 228, 690
- Yule-Walker 추정행렬 822
- INITIAL스테이트먼트 837
- 2단계최소제곱추정법 226
- 2변량 VAR(2)과정 555
- 2변량 VAR(2)과정의 자기공분산행렬 549
- 2변량 VAR(2)모형 538, 541
- 2변량 VARMA(2, 1)과정 621
- 2변량 VARMA(2, 2)과정 613, 626
- 2변량 VARMA(2, 1)모형 641
- 2변량 VARMA(2, 2)모형의 자기공분산행렬합수 631
- 2변량 VMA(2)모형 593
- 이중사전백색화 273
- 2차수렴 604
- 2차정상적 512
- 2차출처스테이트먼트들 720
- 이진코드 191
- 이진파일 191
- 인자분해 882
- INTO구문 166
- 인포맷 191
- 일반목적명령어 140
- 일치성 732, 687
- 입력시계열데이터 199
- 입력행렬 778

x 

---

- 자기공분산구조 588
- 자기공분산행렬 511, 513
- 자기공분산행렬과 자기상관행렬의 추정 675
- 자기공분산행렬합수 882, 513, 625
- 자기상관관계 282
- 자기상관행렬 511
- 작인성 200
- 작인성조건 876, 521
- 작인적 201, 891
- 작인적 구조 199

작인적 조건 819  
 잔차벡터들의 표본자기상관행렬함수 764  
 잡음항 199  
 장기기억 확률과정 494  
 저장라이브러리 143  
 전방다단계예측 412  
 전방일단계예측 835, 412  
 전방일단계예측오차 892  
 전이함수 199  
 전이함수모형 199  
 전이함수모형화 860  
 전치연산 72  
 전치연산자 72  
 전향 VAR모형 557  
 전향 Yule-Walker방정식 557  
 절단성질 739  
 절대수렴 520  
 점근적 정규성 687, 732  
 점근적 효율성 764  
 점근카이제곱검정 827  
 점근명령어 154, 174  
 점축조정서브루틴 118  
 점축조정함수 110  
 정규화 883  
 정상성 782, 512  
 정준가중벡터 883  
 정준변량 885  
 정준상관계수 885  
 정준상관득점 883  
 정준상관분석 824, 882  
 정칙 821  
 제어스테이트먼트 130  
 제어이론 876  
 조건부예측 409  
 조건부최소제곱추정법 831

주축원소 112  
 지연모수 201, 263  
 직교다항식 111  
 직적 766  
 집합함수 109

---

 ㅈ

침자감측연산자 87  
 침자연산자 71, 84  
 침자연산자들 84  
 침자창조연산자 84  
 최량예측 778  
 최소정준상관계수 827  
 최소제곱 732  
 최소제곱추정량의 성질 732  
 최소제곱추정법 219, 700  
 최소CAT추정량 754  
 최소평균제곱오차 929  
 최소평균제곱오차예측 929  
 최신화 111, 151  
 Choi알고리즘[SAS4TSA2] 669, 670  
 최우추정량의 성질 735, 753  
 최우추정법 829, 889, 729  
 축소형 500  
 축약함수 98  
 축차검정법 746  
 축차부분자기회귀검정법 746  
 출력방정식 778  
 출력벡터 778  
 출력시계열데이터 199  
 충격반응가중값 200, 891  
 충격반응가중함수 201  
 충격반응행렬 819  
 친구행렬 525

ㄱ

Cayley-Hamilton정리 878  
 Kalman필터링 890  
 Cochrane-Orcutt방법 227  
 Cholesky분해 111  
 CALL스테이트먼트 91, 117  
 QUIT스테이트먼트 53  
 Kullback-Leibler정보수 753  
 Kronecker곱 73, 100  
 CREATE스테이트먼트 678  
 CLOSE스테이트먼트 677, 679  
 Cleveland 방법 264

ㄷ

텍스트파일 149, 190  
 특성방정식 878  
 특이값 120  
 특이값분해 120  
 특이행렬 835

ㅍ

파일참조이름 190  
 %AR 721  
 퍼트맨투우검정통계량 766  
 퍼트맨투우통계량 282  
 평균제곱오차 893  
 포맷 194  
 Fourier변환 889  
 포인터 192  
 FORM스테이트먼트 836  
 표본교차공분산과 표본교차상관계수의 일치성 688  
 표본교차공분산과 표본교차상관계수의 접근분포 688

표본교차상관계수 679  
 표본교차상관함수 213  
 표본자기공분산행렬들 677  
 표본자기상관행렬 679  
 표본평균벡터의 점근적 정규성 687  
 표준다변량백색잡음과정 732  
 Prais-Winsten방법 227  
 FROM구문 166  
 피드백 현상 273  
 Fisher정보행렬 834  
 PGM윈도우 51  
 필터 199  
 필터오차 894

ㅎ

Hannan의 방법들 664  
 할당스테이트먼트 70  
 함수모듈 71, 127  
 행렬계산함수 100  
 행렬곱하기연산자 72  
 행렬기준연산자 71  
 행렬역연산자 72  
 행렬역함수들 582  
 행렬역함수방정식 601  
 행렬연산자 71  
 행렬정의스테이트먼트 70  
 행렬조작함수 101  
 행렬조회함수 96  
 행렬함수 96  
 Hermite정규형태 111  
 헤시안행렬 830  
 현재관찰점 152  
 현재SAS데이터세트 149  
 현재입력SAS데이터세트 149  
 현재입력파일 191



현재입출력SAS데이터세트 150

현재출력SAS데이터세트 149

현재출력파일 193

Whittle알고리즘 557

확률 수렴 688

확률적 변수 349

확장 Yule-Walker방정식들 625, 644

확장 whittle알고리즘 644

후향VAR모형 557

후향VAR( $p$ )모형 558

후향제거법 801

WHERE구문 150, 163

힐버트공간 819, 879, 520

## 영문색인

### A

a sequence of lagged crossproduct matrices  
*677*  
absolutely converge *520*  
access command *154*  
AIC(Akaike's information criterion) *753*,  
*756, 888*  
All *65*  
alphanumeric *70*  
ARMAX *823*  
assignment statement *70*  
asymptotic efficiency *764*  
asymptotic normality *687*  
autocovariance structure *588*

### B

backward VAR model *557*  
binary code *191*  
binary fire *191*  
block diagonal matrix *101*  
bordering method *112*  
Box & Tiao(1975) *420, 467*  
BREAK *69*

### C

CANCORR *863*  
causal *201, 891*  
causality *200*  
causality condition *521*  
causal structure *199*  
CENTER *62*  
character manipulation function *96*  
character manipulation subroutine *117*

Chemical Process Readings Data *19*  
Chitturi(1974, 1976) *764, 767*  
Cointegration *494*  
colored noise *226*  
companion matrix *525*  
consistency *687*  
contaminated *225*  
control program subroutine *117*  
converge in distribution *733*  
converge in probability *688*  
covariance generating function *588*  
COVB *831*  
COVLAG *677*  
cross-correlation *513*  
cross-correlation function, CCRF *212*  
cross-covariance function, CCVF *212, 513*  
current input file *191*  
current input/output SAS dataset *150*  
current input SAS dataset *149*  
current observation *152*  
current output file *193*  
current output SAS dataset *149*  
current SAS dataset *149*  
cut-off property *739*

### D

data processing commands *149*  
d-dimensional jointly stationarity *512*  
design matrix *115*  
DETAILS *68*  
deterministic *349*  
diagnostic checking *281*  
DIC *827, 888*  
double prewhitening *273*

**E** 

---

elementwise comparison operators *80*  
 elementwise logical-and operator *82*  
 elementwise logical-not operator *83*  
 elementwise logical operators *81*  
 elementwise logical-or operator *82*  
 elementwise operators *77*  
 empty matrix *84*  
 Engle & Granger(1987) *499*  
 Engle & Hendry & Richard(1983) *504*  
 equilibrium *498*  
 error-correction model *498*  
 extended Yule-Walker equations *625*  
 external file *149, 189*

**F** 

---

feasible generalized least squares estimate,  
     FGLS estimate *227*  
 file reference name *190*  
 filter *199*  
 flexibility *187*  
 FLOW *68*  
 format *194*  
 forward VAR mode *557*  
 FPE(final prediction error) *753, 755*  
 FROM clause *166*  
 full rank *115*  
 FW *68*

**G** 

---

Gas Furnace Data *21*  
 general purpose command *140*  
 graphics subroutine *117*  
 Grunfeld(1958) *387*

**H** 

---

Hannan & Quinn(1979) *755*  
 Hermite normal form *111*  
 horizontal concatenation operator *75*  
 horizontal direct product *100*  
 Hosking(1980) *767*  
 HOUSECON Data *29*  
 HOUSING Data *33*

**I** 

---

idempotent matrix *766*  
 identity matrix *102*  
 impulse response function *201*  
 impulse response weight *200, 891*  
 index creation operator *84*  
 innovation process *892*  
 input format *191*  
 integrated *495*  
 interactive *187*  
 interactive mode *140*  
 interface function *96*  
 interface subroutine *117*  
 internal file *149*  
 INTO clause *166*  
 invertible *520, 577*  
 Iterative Proportional Fit *119*

**J** 

---

joint stationary *212*  
 jointly stationary *893*  
 jump *136*

**K** 

---

Kalman gain matrix *895*

Kronecker product 766

**L** 

---

LEADPROD Data 38

Li & McLead(1981) 767

linear algebraic function 96

linear algebraic subroutine 117

LINESIZE 68

local environment 125

long-memory 494

Lütkepohl(1985) 756, 761

**M** 

---

MA1A Data 40

manipulation and reshaping function 96

matrix arithmetic function 96

matrix inquiry function 96

matrix polynomial equation 601

matrix polynomials 582

McSweeny(1978) 445

MINKMUSK Data 41

module 123, 570

Moore-Penrose's generalized inverse matrix  
111

multicolinearity 242

multivariate partial autocorrelation matrix  
739

multivariate time series model 511

**N** 

---

NAME 67

NOALL 65

NOCENTER 62

NODETAILS 68

NOFLOW 68

noise term 199

NOLOG 61

NONAME 67

nondeterministic 349

nonnegative definite matrix 111, 765

nonsingular 821

nonstationary multivariate time series 512

NOPRINT 64

**O** 

---

ordering function 96

orthonormal decomposition 119

orthogonal polynomial 111

overdifferenced 476

overdifferencing 471

**P** 

---

Partial autoregression matrix 738

Partial lag correlation matrix 738

Parzen(1977) 754

PASTMIN 834

Paulsen(1984) 755

Paulsen & Tjøstheim(1985) 747, 754

Pearson correlation coefficient 883

pivot elements 112

Polynomial factor 263

polynomial with infinite order 658

Positive definite matrix 111, 118

prewhiten 233

PROC ARIMA 401

PROC AUTOREG 413

PROC MODEL 721

PROC STATESPACE 832

**Q** =====

Quinn(1980) 753, 755

**P** =====

range operand 157  
 reduced form 500  
 reduction function 96  
 representation theorem 499  
 reverse characteristic polynomial 525  
 ridge regression method 242  
 RIVER Data 44  
 row reduced echelon form 110

**S** =====

SAS 721  
 SAS4TSA1 14, 15  
 SAS4TSA2 13, 15  
 SAS/IML environment 53, 57  
 scope clause 161  
 secondary source statements 720  
 second-order convergence 604  
 second-order stationary 512  
 sequence function 96  
 sequential partial autoregression test 746  
 sequential testing procedure 746  
 set function 96  
 short memory 494  
 SHOW 69  
 shrinkage factor 243  
 singular value 120  
 singular value decomposition 120  
 SPACE 68  
 spectral density matrix 890  
 SPILL 68

stability 200  
 stable 200  
 standard multivariate white noise process 732  
 Stationary in wide sense 512  
 Statistical function 96  
 statistical subroutine 117  
 Subroutine 91, 117  
 Subscript operator 84  
 Subscript operators 84  
 Subscript reduction operator 87  
 sweep operator 112

**T** =====

text file 190  
 time-invariant linear system 819  
 time series function 96  
 time series subroutine 117  
 transfer function 199  
 transform 889  
 transpose operator 72

**U** =====

upper triangular matrix 112

**V** =====

VAR clause 150, 161  
 vector moving-average model 576  
 vertical concatenation operator 75

**W** =====

weakly stationary 512  
 Wei(1990) 453  
 WHERE clause 150, 163  
 Wilson(1969) 604



◆ 독자들의 의견을 듣고 싶습니다 ◆

저자는 다변량시계열분석과 같은 책은 저자와 독자가 함께 만들어가는 것이 좋다고 생각합니다. 본서 속의 오류나 본서에 대한 의견을 보내 주시면 개정판을 낼 때마다 최대한 반영하도록 하겠습니다. 이 page를 복사해서 본서에 대한 제안이나 의견을 다음 주소로 보내주십시오.

120-749

주소: 서울시 서대문구 신촌동 연세대학교 상경대학

최 병 선 교수

팩스: (02) 313-5331

※ 본서에 첨부된 디스켓에 관해서는 세경사로 문의 바랍니다.





▣ 저자약력 ▣

서울대학교 수학과 졸업(이학사)

미국 Stanford대학교 대학원 졸업(경제학 석사, 통계학 석사,

Ph. D. in Statistics <major> and Economics <minor>)

연세대학교 상경대학교수 역임

현재 서울대학교 경제학부 교수 (재무경제학 담당)

Email bschoi12@snu.ac.kr

다변량시계열분석

ISBN 89-7127-042-X

1995년 1월 23일 제1판 1쇄 인쇄

1995년 1월 25일 제1판 1쇄 발행

|     |         |
|-----|---------|
| 저자  | 최병선     |
| 발행자 | 김재은     |
| 조판  | 신홍인쇄(주) |



발행처 **신** **經** **社**

서울특별시 마포구 염리동 36-249/302호

등록: 1979. 8. 31 제 10-19호(倫)

전화: (02) 719-3996

팩스: (02) 719-3995





## 최병선 교수 Creative Commons Book List (<http://s-space.snu.ac.kr/handle/10371/314>) <sup>1</sup>

1. (제2판) SAS를 이용한 현대통계학 (이성백교수와 공저)  
<http://s-space.snu.ac.kr/handle/10371/94393>
2. 행렬의 대각화와 Jordan표준형 (이성백교수와 공저)  
<http://s-space.snu.ac.kr/handle/10371/94394>
3. Fourier 해석 입문  
<http://s-space.snu.ac.kr/handle/10371/94413>
4. Lebesgue적분 입문  
<http://s-space.snu.ac.kr/handle/10371/94414>
5. Wavelet 해석  
<http://s-space.snu.ac.kr/handle/10371/94415>
6. 회귀분석(상)  
<http://s-space.snu.ac.kr/handle/10371/94433>
7. 다변량시계열분석  
<http://s-space.snu.ac.kr/handle/10371/94434>
8. 이산형 재무모형의 수리적 배경  
<http://s-space.snu.ac.kr/handle/10371/94455>
9. 회귀분석(하)  
<http://s-space.snu.ac.kr/handle/10371/94456>
10. 단변량시계열분석  
<http://s-space.snu.ac.kr/handle/10371/94457>
11. 금융공학 IV: Monte Carlo Methods for Finance and Economics  
<http://s-space.snu.ac.kr/handle/10371/98466>
12. 금융공학 V: Introduction to Financial Engineering with R  
<http://s-space.snu.ac.kr/handle/10371/99003>

---

<sup>1</sup>현대문화사(hun1249@hanmail.net)에서 이 책들의 POD판을 구하실 수 있습니다.