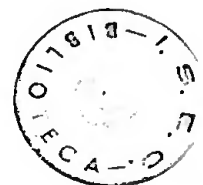


1. 2. 3. 4. 5.
E. Bibliotecas
1364-G. 36436

RESERVADO

HA31.7
T45
1990



PAULO JOÃO FIGUEIREDO CABRAL TELES

**MODELIZAÇÃO DE
SÉRIES CRONOLÓGICAS
MULTIVARIADAS**

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre em Métodos Matemáticos para Economia e Gestão de Empresas.

INSTITUTO SUPERIOR DE ECONOMIA E GESTÃO

da

UNIVERSIDADE TÉCNICA DE LISBOA

Março de 1990





À minha família

AGRADECIMENTOS

Ao Prof. Dr. Daniel Müller, pela contribuição decisiva dada para a minha formação no domínio das séries cronológicas e pela excelente orientação e extraordinária paciência, sem o que este trabalho nunca teria sido possível.

Aos meus colegas, em especial aos que se encontravam ou encontram ainda no estrangeiro, cujo apoio bibliográfico foi fundamental.

Ao Prof. Dr. Paulo Jorge Gomes, pelas facilidades concedidas e pelos valiosos conselhos.

Ao António José Castilho e ao Nuno Castilho, pelo excelente tratamento de texto.

À minha família, cujo apoio foi sempre imprescindível.

Todos os erros, lacunas e deficiências que esta dissertação apresenta são da exclusiva responsabilidade do seu autor.

RESUMO

A modelização de séries cronológicas multivariadas constitui o objecto do presente trabalho. A definição de conceitos específicos ao caso multivariado precede a análise das diferentes fases que constituem a metodologia conducente à modelização de uma qualquer série cronológica: Identificação, Estimação e Confirmação do Diagnóstico. Além disso é ainda analisada a utilização de um modelo em Previsão.

Por outro lado, reservou-se o tratamento da fase da Identificação para o fim, conferindo-lhe primordial importância, uma vez que se pode considerar esta fase como a que constitui o principal desafio ao analista.

Palavras-chave: simultaneidade; processo estocástico multivariado; série cronológica multivariada; matriz de covariância; matriz de correlação; matriz de correlação parcial; estacionaridade; modelos ARMA multivariados; estimação; verosimilhança exacta; verosimilhança aproximada; distribuição assintótica; consistência; confirmação do diagnóstico; resíduos de estimação; qualidade do ajustamento; previsão; previsor linear óptimo; identifiabilidade; redundância; identificação; modelos bivariados; matriz de correlação parcial q-condicionada estimada; matriz de correlação alargada estimada; estrutura sobreparametrizada; simplificação da estrutura identificada.

ABSTRACT

The modelling of multivariate time series is the subject of this study. The definition of the concepts specific to the multivariate case precedes the analysis of the different phases comprising the methodology conducing to the modelling of a time series : Identification, Estimation , and Diagnostic Checking. In addition, the use of the model in Prediction is analyzed.

Consideration of the Identification phase, however, was left to the last, stressing its importance, because precisely this aspect of the study is one which poses the greatest challenge to the analyst.

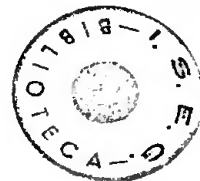
Key-words: simultaneity; multivariate stochastic process; multivariate time series; covariance matrix; correlation matrix; partial correlation matrix; stationarity; multivariate ARMA models; estimation; exact likelihood; approximate likelihood; asymptotic distribution; consistence; diagnostic checking; estimation residuals; quality of fit; prediction; best linear predictor; identifiability; redundancy; identification; bivariate models; q-conditioned sample partial correlation matrix; extended sample correlation matrix; overparameterized structure; simplification of the identified structure.

INDICE

0. INTRODUÇÃO.....	1
I CONCEITOS INTRODUTÓRIOS.....	8
1. PROCESSO ESTOCÁSTICO MULTIVARIADO	9
2. PROPRIEDADES DE SEGUNDA ORDEM DE UM PROCESSO ESTOCÁSTICO MULTIVARIADO.....	12
2.1. PROCESSO ESTOCÁSTICO MULTIVARIADO ESTACIONÁRIO	12
2.2. TRANSFORMAÇÕES DE UMA SÉRIE CRONOLÓGICA	19
II. MODELOS PROBABILÍSTICOS.....	23
1. PROCESSOS ESTOCÁSTICOS MULTIVARIADOS	24
1.1. RUÍDO BRANCO OU PROCESSO PURAMENTE ALEATÓRIO MULTIVARIADO.....	24
1.2. PROCESSO AUTO-REGRESSIVO MULTIVARIADO.....	24
1.3. PROCESSO DE MÉDIAS MÓVEIS MULTIVARIADO.....	29
1.4. PROCESSO MISTO AUTO-REGRESSIVO E DE MÉDIAS MÓVEIS MULTIVARIADO.....	35
1.5. PROCESSO MISTO AUTO-REGRESSIVO E DE MÉDIAS MÓVEIS INTEGRADO MULTIVARIADO	40
III. ESTIMAÇÃO.....	45
1. ESTIMAÇÃO DO VALOR ESPERADO, DA MATRIZ DE COVARIÂNCIA, DA MATRIZ DE CORRELAÇÃO E	
ESTIMAÇÃO PRELIMINAR DOS PARÂMETROS.....	47
1.1. ESTIMAÇÃO DO VALOR ESPERADO	47
1.2. ESTIMAÇÃO DA MATRIZ DE COVARIÂNCIA, DA MATRIZ DE CORRELAÇÃO E DA MATRIZ DE	
CORRELAÇÃO PARCIAL.....	49
1.3. ESTIMAÇÃO PRELIMINAR DOS PARÂMETROS.....	52
1.3.1. ESTIMAÇÃO PRELIMINAR DOS PARÂMETROS DE MODELOS AUTO-REGRESSIVOS.....	53
1.3.2. ESTIMAÇÃO PRELIMINAR DOS PARÂMETROS DE MODELOS DE MÉDIAS MÓVEIS.....	53
2. ESTIMAÇÃO DOS PARÂMETROS.....	55
2.1. ESTIMAÇÃO DOS PARÂMETROS DE MODELOS AUTO-REGRESSIVOS.....	55
2.2. ESTIMAÇÃO DOS PARÂMETROS DE MODELOS DE MÉDIAS MÓVEIS.....	58
2.3. ESTIMAÇÃO DOS PARÂMETROS DE MODELOS MISTOS AUTO - REGRESSIVOS E DE MÉDIAS MÓVEIS. .	64
3. ESTIMAÇÃO DA ORDEM DO MODELO.....	66
3.1. ESTIMAÇÃO DA ORDEM DE MODELOS AUTO-REGRESSIVOS.....	66
3.2. ESTIMAÇÃO DAS ORDENS DE MODELOS MISTOS AUTO-REGRESSIVOS E DE MÉDIAS MÓVEIS	70
4. DISTRIBUIÇÃO ASSINTÓTICA DOS ESTIMADORES	73
4.1. DISTRIBUIÇÃO DOS ESTIMADORES DA FUNÇÃO DE CORRELAÇÃO E DA FUNÇÃO DE CORRELAÇÃO	
PARCIAL.....	73
4.2. DISTRIBUIÇÃO DOS ESTIMADORES DOS PARÂMETROS.....	76

4.3. DISTRIBUIÇÃO DOS ESTIMADORES DA FUNÇÃO DE COVARIÂNCIA E DA FUNÇÃO DE CORRELAÇÃO RESIDUAIS.....	77
IV. CONFIRMAÇÃO DO DIAGNÓSTICO.....	81
1. TESTE DE INDEPENDÊNCIA DE DOIS PROCESSOS ESTOCÁSTICOS UNIVARIADOS	81
1.2. TESTES GLOBAIS.....	86
1.2.1. TESTE DE HAUGH.....	87
1.2.2. TESTE DE KOCH E YANG.....	88
2. CRITÉRIOS GERAIS.....	91
3. ANÁLISE DOS RESÍDUOS.....	92
4. TESTES DE QUALIDADE DO AJUSTAMENTO.....	94
4.1. SOBREPAREMETRIZAÇÃO.....	94
4.2. TESTES GLOBAIS.....	95
4.3. TESTES DE MULTIPLICADORES DE LAGRANGE.....	98
4.4. TESTE INFORMAL.....	108
V. PREVISÃO.....	112
1. EXTENSÃO DO MÉTODO DE BOX E JENKINS.....	113
2. MÉTODO DE GRANGER E NEWBOLD.....	118
VI. IDENTIFIABILIDADE E REDUNDÂNCIA.....	123
1. REGRAS DE IDENTIFIABILIDADE.....	124
VII IDENTIFICAÇÃO.....	127
1. MODELOS DE PEQUENA DIMENSÃO.....	131
1.1. MÉTODO RESIDUAL.....	131
1.2. MÉTODO COMPARATIVO.....	134
1.3. MODELOS BIVARIADOS.....	137
1.3.1. RELAÇÃO BIDIRECCIONAL.....	139
1.3.2. RELAÇÃO UNIDIRECCIONAL.....	153
1.3.2.1. MÉTODO DE SIMS.....	154
1.3.2.2. MÉTODO DE BOX E JENKINS.....	156
1.3.2.3. MÉTODO DE PRIESTLEY.....	162
1.3.2.4. MÉTODO DE GRANGER E NEWBOLD.....	168
2. EXTENSÕES DO MÉTODO DE BOX E JENKINS.....	174
2.1. UTILIZAÇÃO DE FUNÇÕES MATRICIAIS.....	175
2.2. UTILIZAÇÃO DE FUNÇÕES ESCALARES.....	196
3. IDENTIFICAÇÃO DE UMA ESTRUTURA SIMPLIFICADA.....	208

3.1. IDENTIFICAÇÃO DOS ELEMENTOS SIGNIFICATIVOS NOS OPERADORES DE UM MODELO ARMA	209
3.2. MODELO FACTORIAL.....	218
3.3. MODELO INDICE AUTO-REGRESSIVO.....	231
3.4. MODELO AUTO-REGRESSIVO DE POLINÓMIO AMORTECIDO.....	237
3.5. ANÁLISE CANÓNICA	246
3.6. CONCLUSÃO	258
NOTAS	261



0. INTRODUÇÃO

Existe frequentemente interesse na observação não apenas de uma série cronológica isoladamente, mas na observação simultânea de diferentes séries inter-relacionadas. Quando isso acontece, está-se na presença de uma série cronológica multivariada, cuja definição rigorosa será proposta mais à frente.

A necessidade de analisar uma série cronológica multivariada surge em diferentes áreas do conhecimento. Assim, por exemplo, no domínio da Engenharia pode revelar-se útil estudar as variações simultâneas da corrente e da voltagem; no que se refere à Meteorologia e Geofísica, a atenção pode recair sobre os registos relativos aos movimentos sísmicos, à temperatura e à pressão feitas em diferentes estações meteorológicas no mesmo período de tempo; em Economia, é conferida grande importância à teia de relações existentes entre a oferta de moeda, os preços, os salários e a taxa de juro. A um nível mais simples, é útil por vezes estudar isoladamente a relação entre a oferta de moeda e os preços, ou entre estes e o desemprego, ou ainda entre este e a taxa de juro; um outro exemplo vulgarmente utilizado, e amplamente discutido em Economia, é o da análise entre a procura e a oferta de um dado bem ou serviço.

Os exemplos acabados de enunciar são suficientes para ilustrar a importância do tema que vai ser estudado no presente trabalho.

Note-se que qualquer das grandezas acabadas de referir pode ser estudada isoladamente, sendo tratada como uma série cronológica univariada. Embora tal procedimento não deixasse de fornecer informações úteis, nunca conseguiria revelar o que poderá ser o mais importante nesse estudo, ou seja, a interdependência entre as variáveis. Deste modo, os principais motivos para considerar a análise e modelização conjunta das diferentes séries univariadas podem ser resumidos nas duas seguintes vertentes principais:

(i) A possibilidade da percepção e compreensão das relações dinâmicas existentes entre essas séries. Estas podem estar contemporaneamente relacionadas, uma pode conduzir outras ou pode ainda haver relações em vários sentidos entre algumas séries. Uma melhor compreensão destas relações pode, por exemplo, num estudo sobre a poluição atmosférica, permitir delinear uma estratégia apropriada de controle anti-poluição, de modo a melhorar a qualidade ambiental.

(ii) A capacidade de aumentar a precisão das previsões. Quando existe informação sobre uma série contida nos dados históricos de outra, é possível obter melhores previsões quando se procede à sua modelização conjunta.

Tal como na Teoria das Probabilidades, em que o estudo das relações entre diferentes variáveis aleatórias tem que ser feito com recurso à sua distribuição conjunta, não bastando o conhecimento das distribuições marginais, na análise de uma série cronológica multivariada é necessária a utilização de instrumentos que permitam descrever não só as propriedades de cada série em si, mas também as inter-relações entre as diferentes séries. A propósito, refira-se que grande parte da teoria relativa ao tratamento das séries cronológicas univariadas é generalizável às séries multivariadas, embora se levantem novos problemas.

Por outro lado, tal como no caso univariado, não é utilizado qualquer conhecimento "a priori" sobre possíveis relações de comportamento para a modelização de uma série cronológica multivariada, admitindo-se apenas que cada série univariada está relacionada com as outras.

Afigura-se também importante sublinhar que é possível distinguir dois tipos de situações no que respeita à relação entre as diferentes séries cronológicas univariadas que compõem uma série multivariada. Adoptando a terminologia de Jenkins e Watts (1968), num primeiro tipo

de situação as séries são encaradas como estando "em pé de igualdade", residindo o interesse em estudar a interacção entre elas; num segundo tipo de situação, existe uma relação de causalidade entre as séries, sendo umas os "inputs" e outras os "outputs", em que o objectivo consiste em estudar as propriedades do sistema (linear) de modo a que os valores dos segundos possam ser previstos a partir dos valores dos primeiros. No presente estudo, será o primeiro tipo de situação que irá ser considerado.

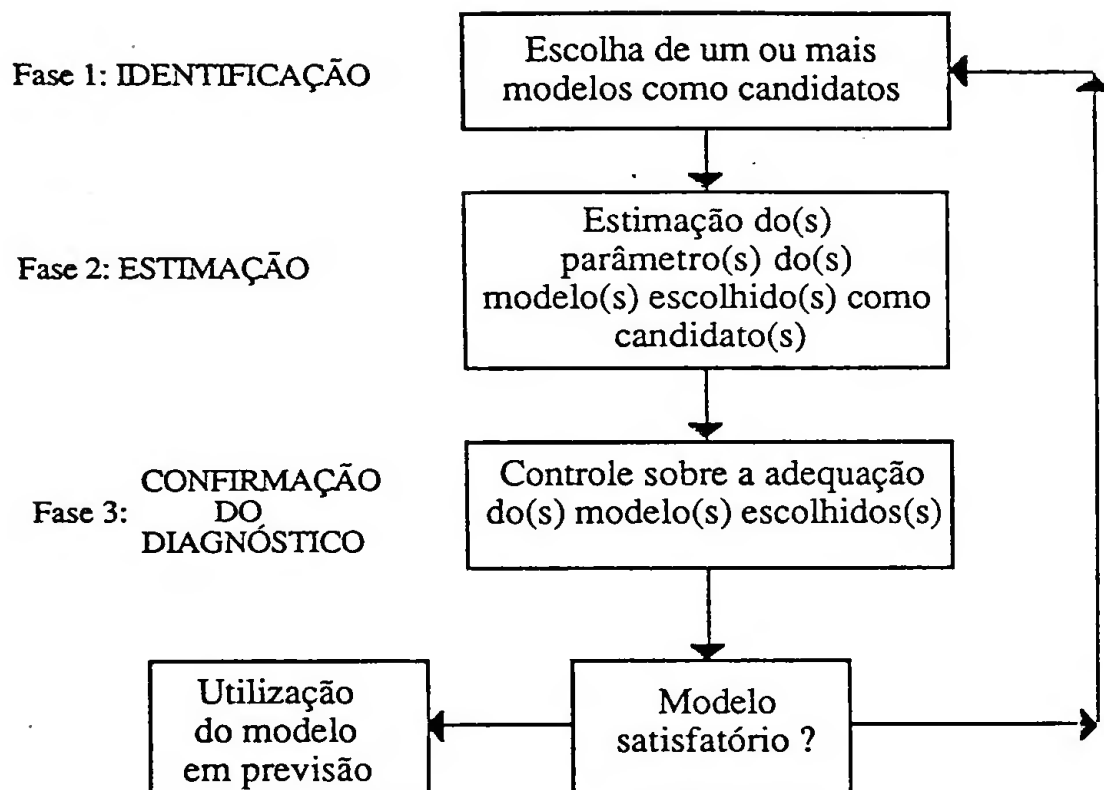
Assim, todo o processo de modelização de uma série cronológica visa atingir certos fins, de que se destacam:

1º) Descrição: pretende-se aqui formular um modelo matemático que descreva a série.

2º) Explicação: quando se observa mais do que uma série simultaneamente, é possível utilizar a evolução de cada uma para melhor descrever as outras. Isto poderá permitir um conhecimento mais profundo do mecanismo que gerou uma dada série.

3º) Predição (ou previsão): a partir do modelo estabelecido, pretende-se predizer (prever) valores futuros da série.

Estes objectivos são alcançados por etapas, dando origem ao processo de modelização de uma série cronológica, que é válido tanto no caso de uma série univariada como no de uma multivariada e que pode ser esquematizado da forma:



Assim, a modelização de uma série cronológica passa pelas fases da identificação, da estimação e da confirmação do diagnóstico, considerando-se um dado modelo como apto para a geração de previsões se for julgado satisfatório.

Deste modo, a primeira fase consiste na identificação da série, isto é, na escolha de um modelo que permita a adequada descrição da série. Na segunda fase, procede-se à estimação dos parâmetros do modelo escolhido na fase anterior e, finalmente, na terceira fase pretende-se verificar se o modelo acabado de estimar é o melhor para a série cronológica.

Nestas condições, um dado modelo só pode ser utilizado para previsão após ter passado por estas três fases e ter sido considerado adequado para a série cronológica em análise:

Tal como já foi referido, todos os métodos e instrumentos de análise de uma série cronológica univariada continuam presentes na modelização de uma série multivariada.

No contexto agora descrito, vai-se passar à análise de cada uma das fases referidas, focando-se também a utilização de um modelo para a geração de previsões. No entanto, a ordem adoptada para a análise de cada uma das fases não é a que foi acabada de referir, embora seja esse o desenvolvimento normal de um qualquer processo de modelização de uma série cronológica. deixando-se para o fim o tratamento do problema da identificação.

Assim, apresenta-se inicialmente um conjunto de resultados necessários à caracterização de um processo estocástico e que permitem a descrição de diferentes modelos probabilísticos, ou seja, de classes de processos estocásticos multivariados que se revelam externamente úteis na modelização de qualquer série cronológica, apontando-se também alguns problemas que se colocam em alguns aspectos de utilização dos referidos modelos nessa modelização. Em seguida, passa-se à análise dos métodos de estimação não só das funções necessárias à identificação de uma série cronológica, mas também dos parâmetros desconhecidos incluídos na especificação dos modelos já referidos, conferindo-se alguma atenção aos principais problemas e dificuldades encontrados na implementação desses métodos. Por outro lado, tendo em conta importantes fins inferenciais, considera-se ainda neste âmbito as propriedades e as distribuições dos estimadores obtidos.

No entanto, conforme foi referido, um modelo só pode ser utilizado para previsão após a verificação da sua qualidade. Deste modo, é-se imediatamente conduzido à fase de crítica do modelo, ou seja, de confirmação do diagnóstico, onde se irá analisar um conjunto de procedimentos e de testes que pretendem averiguar se o modelo anteriormente estimado tem uma qualidade suficiente para a sua utilização em previsão e revelar eventuais sinais de falta de adequação à série cronológica, permitindo nesse caso evidenciar quaisquer características

apresentadas pela série cronológica que possivelmente não tenham ainda sido tomadas em conta.

Admitindo que um dado modelo cumpre os requisitos necessários para ser considerado satisfatório, analisa-se então a sua utilização para a geração de previsões referentes aos valores que a série cronológica futuramente virá a assumir. Para o efeito, serão utilizados dois métodos que, apesar de se basearem em diferentes critérios de optimalidade do previsor, acabam por conduzir ao mesmo resultado, o que significa que são equivalentes. Em seguida, analisa-se rapidamente o problema da identifiabilidade de um dado modelo, ou seja, o da sua especificação de uma forma única a partir da função de covariância do processo estocástico multivariado.

Finalmente, analisa-se em último lugar o problema da identificação de uma série cronológica, ou seja, o da selecção de um modelo que lhe seja adequado. Embora a análise desta fase tenha sido relegada para último lugar, apesar de, conforme já foi referido, ela ser a primeira na modelização de uma série cronológica, tal não significa de forma nenhuma que seja uma fase de importância secundária.

Pelo contrário, pretende-se com tal opção enfatizar a sua extrema relevância e o carácter fundamental que assume a identificação de uma qualquer série cronológica. Com efeito, uma boa identificação é imprescindível para que as fase subsequentes se desenrolem sem problemas dignos de nota, uma vez que lhes fornece a matéria-prima que aí irá ser trabalhada e, conseqüentemente, para a geração de previsões de boa qualidade. Assim, a finalidade de deixar a análise da fase da identificação para o último lugar, consiste em conferir-lhe maior importância, sendo essa análise o objectivo principal do presente estudo. Estas considerações conduzem ao tratamento desta questão com grande detalhe, começando-se por abordagens que, sendo pioneiras na

identificação de séries cronológicas multivariadas, não são também suficientemente genéricas para abarcar qualquer caso. Com efeito, as críticas feitas a essas propostas evidenciam a sua incapacidade para tratar séries com um elevado número de componentes, passando-se por isso a métodos posteriores cuja formulação é suficientemente geral para lhes permitir eficácia e flexibilidade em qualquer situação. Constatada a boa qualidade destes métodos, apesar de algumas imperfeições, estuda-se por fim a introdução de simplificações nos modelos assim seleccionados.

A importância, que será amplamente sublinhada, de que se reveste a identificação de uma série cronológica multivariada, converte esta fase num enorme desafio que se coloca a qualquer analista. Este desafio é tanto maior quanto maior for o número de componentes dessa série, tornando o seu tratamento muito mais complexo e difícil do que no caso univariado. Uma vez que, conforme foi referido atrás, se constata a necessidade de considerar diversas variáveis conjuntamente, não sendo satisfatória a análise univariada de cada uma delas, afigurou-se extremamente útil e relevante aceitar o desafio de analisar cuidadosamente as metodologias mais indicadas para a identificação de uma série cronológica multivariada.

Convém no entanto referir que se admite que, para a leitura do presente estudo, se dispõe de um bom conhecimento da análise de séries cronológicas univariadas.

I. CONCEITOS INTRODUTÓRIOS

Tendo em vista o caso multivariado, irá ser agora apresentado um conjunto de definições e conceitos fundamentais a toda a análise subsequente e que se revelam extremamente úteis na prossecução do estudo.

1. Processo estocástico multivariado

Um processo estocástico multivariado pode ser definido a partir da generalização do conceito de processo estocástico univariado. Começemos então por definir este último.

Assim, seja T um conjunto não vazio e (Ω, \mathcal{F}, P) um espaço de probabilidades em que Ω é um conjunto qualquer, \mathcal{F} é uma σ -álgebra em Ω e P é uma medida de probabilidade definida em (Ω, \mathcal{F}) .

Considere-se o espaço mensurável $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, designando $\mathcal{B}_{\mathbb{R}}$ a σ -álgebra dos borelianos de \mathbb{R} .

Neste contexto, define-se, para cada $t \in T$, a variável aleatória $X(t)$, isto é, uma aplicação mensurável:

$$X(t): \Omega \rightarrow \mathbb{R} \quad (1.1.1)$$

ou seja, X é uma aplicação que permite obter $X(\omega)$, com $\omega \in \Omega$, e em que $X^{-1}(R)$ pertence a \mathcal{F} para cada intervalo de números reais limitados R .

À família de variáveis aleatórias definidas para todo o $t \in T$ $\{X(t), t \in T\}$ chama-se processo estocástico (univariado), isto é, trata-se de uma família de variáveis aleatórias definidas no mesmo espaço de probabilidades.

O espaço (Ω, \mathcal{F}, P) é o espaço de probabilidades de base do processo estocástico, sendo $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ o seu espaço dos estados e \mathbb{R} o conjunto dos estados.

Deste modo, um processo estocástico não é mais do que uma aplicação de X definida para elementos de $\Omega \times T$ e assumindo valores em \mathbb{R} :

$$X: \Omega \times T \rightarrow \mathbb{R} \quad (1.1.2)$$

tal que, se t é fixo, torna-se numa aplicação de Ω em \mathbb{R} , ou seja, converte-se numa variável aleatória.

Por outro lado, se se fixar $\omega \in \Omega$, o processo estocástico torna-se numa função real de t . Esta função designa-se de trajectória ou realização do processo estocástico.

O conjunto T é designado de conjunto dos índices do processo estocástico. Se $T = \{0, \pm 1, \pm 2, \dots\}$ ou $T = \{0, 1, 2, \dots\}$ o processo estocástico diz-se de parâmetro ou tempo discreto e se $T = \mathbb{R}$ ou $T = \mathbb{R}^+$ diz-se de parâmetro ou tempo contínuo.

Nestas condições, pode-se definir uma série cronológica univariada como uma trajectória particular de um processo estocástico univariado limitada no tempo.

Uma série cronológica discreta é uma trajectória de um processo estocástico de parâmetro ou tempo discreto, ou seja, é uma série que só é observável em certos momentos do tempo. Opostamente, uma série cronológica contínua é uma trajectória de um processo estocástico de parâmetro ou tempo contínuo, ou seja, é uma série cuja observação é feita continuamente no tempo. Como é óbvio, esta distinção nada significa sobre se os valores das observações são discretos ou contínuos. O presente trabalho ocupar-se-à apenas de séries discretas.

A notação que será utilizada para uma série cronológica (univariada) é a seguinte:

$$\{X_t, t \in T\} ; T = \{t_1, t_2, \dots, t_n\} \quad (1.1.3)$$

ou seja, tem-se a sucessão de observações $X_{t_1}, X_{t_2}, \dots, X_{t_n}$.

Esta definição é facilmente generalizável ao caso de um processo estocástico multivariado, que consiste num vector de processos estocásticos univariados[†].

[†] sempre que se tenha vectores ou matrizes, utilizar-se-à letras carregadas.

$$\{X_t^T\} = \{X_{1t}, X_{2t}, \dots, X_{mt}\} \quad (1.1.3)$$

Do mesmo modo, uma série cronológica multivariada é uma trajectória particular de um processo estocástico multivariado limitada no tempo, confinando-se também a análise às séries discretas no caso multivariado.

2. Propriedades de segunda ordem de um processo estocástico multivariado

Antes de prosseguir, é necessário apresentar um conjunto de instrumentos e resultados relativos a um processo estocástico multivariado que são indispensáveis para a sua caracterização e, portanto, para a análise de uma série cronológica multivariada. Por outro lado, apresenta-se também a definição de processo estocástico multivariado estacionário, uma vez que este tipo de processo desempenha um papel fundamental naquela análise.

2.1. Processo estocástico multivariado estacionário

Considere-se m processos estocásticos univariados:

$$\{X_{it}, t=0, \pm 1, \pm 2, \dots\} \quad i=1, \dots, m \quad (1.2.1)$$

com $E(X_{it}^2) < \infty$.

Designar-se-à o valor esperado de um processo estocástico $\{X_{it}\}$ por:

$$E(X_{it}) = \mu_{it} \quad (1.2.2)$$

e a respectiva função de autocovariância por:

$$\text{cov}(X_{it}, X_{i,t-k}) = E[(X_{it} - \mu_{it})(X_{i,t-k} - \mu_{i,t-k})] = \gamma_{ii}(t, t-k) \quad i=1, \dots, m ; k=0, \pm 1, \pm 2, \dots \quad (1.2.3)$$

Esta função descreve a estrutura da covariância de cada processo. Uma vez que é necessário considerar a estrutura da covariância entre os diferentes processos, define-se a função de covariância cruzada entre $\{X_{it}\}$ e $\{X_{jt}\}$:

$$\text{cov}(X_{it}, X_{j,t-k}) = E[(X_{it} - \mu_{it})(X_{j,t-k} - \mu_{j,t-k})] = \gamma_{ij}(t, t-k) \quad \begin{array}{l} i, j = 1, \dots, m \quad i \neq j; \\ k = 0, \pm 1, \pm 2, \dots \end{array} \quad (1.2.4)$$

Ficam assim especificadas as características de segunda ordem de um processo estocástico multivariado.

A utilização de notação matricial revela-se muito vantajosa no tratamento de processos estocásticos multivariados. Assim, representando um processo estocástico multivariado $\{X_t\}$ da forma:

$$X_t = \begin{bmatrix} X_{1t} \\ X_{2t} \\ \vdots \\ X_{mt} \end{bmatrix} \quad t = 0, \pm 1, \pm 2, \dots \quad (1.2.5)$$

as características de segunda ordem de $\{X_t\}$ são descritas pelo vector dos valores esperados:

$$\mu_t = E(X_t) = \begin{bmatrix} \mu_{1t} \\ \mu_{2t} \\ \vdots \\ \mu_{mt} \end{bmatrix} \quad (1.2.6)$$

e pela matriz de covariância:

$$\Gamma(t, t-k) = E[(X_t - \mu_t)(X_{t-k} - \mu_{t-k})^*] \quad (1.2.7)$$

designando * a transposição e conjugação. No entanto, admitir-se-à que $\{X_t\}$ é um processo real.

Tal como no caso univariado, os processos multivariados estacionários desempenham um importante papel na modelização de uma série cronológica.

Assim, um processo estocástico multivariado diz-se fortemente estacionário se a distribuição conjunta de (X_1, \dots, X_n) e de

$(X_{t_1+k}, \dots, X_{t_n+k})$ for a mesma, para todo o t_1, \dots, t_n e k inteiros. Por outro lado, um processo estocástico diz-se estacionário de segunda ordem se:

$$\begin{aligned} E(X_t) &= E(X_{t+k}) = \mu && \text{para todo o } t, k; \\ \Gamma(t_1, t_2) &= \Gamma(t_1 - t_2) && \text{para todo o } t_1, t_2. \end{aligned} \quad (1.2.8)$$

Uma vez que a definição de estacionaridade forte é muito exigente, utilizar-se-à habitualmente esta segunda definição, estabelecida em (1.2.8), para caracterizar o que é um processo estacionário. Deste modo, o vector dos valores esperados passará a ser designado por:

$$\mu = E(X_t) = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{bmatrix} \quad (1.2.9)$$

e a matriz de covariância por:

$$\Gamma(k) = E[(X_t - \mu)(X_{t+k} - \mu)^T] \quad (1.2.10)$$

Isto significa que ambas as características de segunda ordem do processo estocástico multivariado são independentes de t , ou seja, do momento considerado.

O elemento de ordem (i, j) da matriz $\Gamma(k)$ é:

$$\text{cov}(X_{it}, X_{jt+k}) = E[(X_{it} - \mu_i)(X_{jt+k} - \mu_j)] = \gamma_{ij}(k) \quad (1.2.11)$$

Se $i=j$, trata-se da função de autocovariância de cada processo $\gamma_{ii}(k)$ e, se $i \neq j$ ($i, j=1, \dots, m$), trata-se da função de covariância cruzada entre $\{X_{it}\}$ e $\{X_{jt}\}$ designada por $\gamma_{ij}(k)$. Fica assim bem evidenciado que se trata de funções apenas de k .

Deste modo, dois processos são conjuntamente estacionários se cada um deles o for, o que implica que as respectivas funções de

autocovariância sejam independentes de t , e se a respectiva função de covariância cruzada também for independente do momento considerado.

No entanto, note-se que, contrariamente à função de autocovariância, a função de covariância cruzada não é uma função par, uma vez que se verifica a relação:

$$\gamma_{ij}(k) \neq \gamma_{ij}(-k) \quad i \neq j \quad (1.2.12)$$

verificando-se antes a equação:

$$\begin{aligned} \gamma_{ij}(k) &= E[(X_{it} - \mu_i)(X_{j,t+k} - \mu_j)] = \\ &= E[(X_{i,t+k} - \mu_i)(X_{jt} - \mu_j)] = \\ &= E[(X_{jt} - \mu_j)(X_{i,t+k} - \mu_i)] = \gamma_{ji}(-k) \end{aligned} \quad (1.2.13)$$

Assim, a matriz de covariância de um processo estocástico multivariado estacionário verifica as propriedades (Brockwell e Davis, 1987, pág.393):

$$1^a) \Gamma(k) = \Gamma(-k)$$

$$2^a) |\gamma_{ij}(k)| \leq [\gamma_{ii}(0)\gamma_{jj}(0)]^{1/2} \quad i, j = 1, \dots, m$$

$$3^a) \gamma_{ii}(k) \text{ é uma função de autocovariância, } i = 1, \dots, m$$

$$4^a) \sum_{l,h=1}^N \mathbf{a}_l^T \Gamma(l-h) \mathbf{a}_h \geq 0 \quad \text{para } N \in \{1, 2, \dots\} \text{ e } \mathbf{a}_1, \dots, \mathbf{a}_N \in \mathbb{R}^m$$

A primeira propriedade resulta da equação (1.2.13), a segunda da desigualdade de Cauchy-Schwarz e a terceira do facto de $\gamma_{ii}(k)$ ser a função de autocovariância do processo estocástico estacionário $\{X_{ii}\}$.

Uma vez que estas funções estão dependentes da unidade de medida de cada processo $\{X_{ii}\}$ componente de $\{X_i\}$, revela-se útil normalizá-las,

nomeadamente para fins interpretativos. Nestas condições, define-se a função de autocorrelação de $\{X_{i_t}\}$ da forma:

$$\rho_{ii}(k) = \frac{\gamma_{ii}(k)}{\gamma_{ii}(0)} \quad i = 1, \dots, m \quad (1.2.14)$$

e a função de correlação cruzada entre $\{X_{i_t}\}$ e $\{X_{j_t}\}$ da forma:

$$\rho_{ij}(k) = \frac{\gamma_{ij}(k)}{[\gamma_{ii}(0)\gamma_{jj}(0)]^{1/2}} \quad i, j = 1, \dots, m \quad i \neq j \quad (1.2.15)$$

Note-se que $\gamma_{ii}(0) = V(X_{i_t}) = \sigma_i^2$.

A função $\rho_{ij}(k)$ mede a correlação existente entre X_{i_t} e $X_{j_{t-k}}$ e, tal como $\gamma_{ij}(k)$, também não é uma função par:

$$\rho_{ij}(k) = \rho_{ji}(-k) \quad i, j = 1, \dots, m; \quad i \neq j \quad (1.2.16)$$

Assim, obtém-se a matriz de correlação de um processo estocástico multivariado estacionário $\{X_t\}$, de elemento genérico $\rho_{ij}(k)$ e assumindo a forma:

$$\rho(k) = \Gamma_0^{-1/2} \Gamma(k) \Gamma_0^{-1/2} \quad (1.2.17)$$

sendo $\Gamma_0 = \text{diag} \{ \gamma_{11}(0) \dots \gamma_{mm}(0) \}$ a matriz diagonal das variâncias dos diferentes processos $\{X_{i_t}\}$, $i = 1, \dots, m$. A matriz $\rho(k)$ verifica as propriedades (Brockwell e Davis, 1987, pág.393):

$$1^a) \rho(k) = \rho^T(-k)$$

$$2^a) |\rho_{ij}(k)| \leq [\gamma_{ii}(0)\gamma_{jj}(0)]^{1/2} \quad i, j = 1, \dots, m$$

$$2^a) |\rho_{ij}(k)| \leq [\rho_{ii}(0)\rho_{jj}(0)]^{1/2} \quad i, j = 1, \dots, m$$

$$3^a) \rho_{ii}(k) \text{ é uma função de autocorrelação} \quad i = 1, \dots, m$$

$$4^a) \sum_{i,h=1}^N \mathbf{b}_i^T \rho(l-h) \mathbf{b}_h \geq 0 \quad \text{para } N \in \{1,2,\dots\} \text{ e } \mathbf{b}_1, \dots, \mathbf{b}_N \in \mathbb{R}^m$$

É importante notar que, para $i \neq j$, $\rho_{ij}(0)$ não é necessariamente igual a 1, sendo possível que, para $k \neq 0$, $|\gamma_{ij}(k)| > |\gamma_{ij}(0)|$, o que significa também que $\rho_{ij}(k)$ pode atingir o seu máximo para um qualquer valor de k contrariamente ao que se passa quando $i=j$, em que esse máximo é atingido para $k=0$.

Por outro lado, há ainda a considerar a matriz de correlação parcial. Para o efeito, define-se a regressão de X_t sobre X_{t-l} ($l=1,\dots,k$) que assume a forma:

$$X_t = \phi_1 X_{t-1} + \dots + \phi_k X_{t-k} + U_t \quad (1.2.18)$$

onde ϕ_l ($l=1,\dots,k$) são matrizes de parâmetros, de dimensão $(m \times m)$ e U_t é um vector de dimensão $(m \times 1)$, constituindo o termo residual da regressão. Assim, Tiao e Box (1979, 1981) e Jenkins e Alavi (1981) definem a matriz de correlação parcial, também designada pelos primeiros de matriz de auto-regressão parcial, como a última matriz de parâmetros ϕ_k na regressão acima, ou seja:

$$P(k) = \phi_k \quad k=1,2,\dots \quad (1.2.19)$$

onde $P(k)$ designa a matriz de correlação parcial⁽¹⁾.

Vejam agora alguns exemplos:

Exemplo 1.

Seja o processo estocástico bivariado estacionário:

$$\mathbf{X}_t = \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} \quad (1.2.20)$$

em que $\{X_{1t}\}$ e $\{X_{2t}\}$ são não correlacionados, ou seja:

$$\gamma_{12}(k) = \text{cov}(X_{1t}, X_{2,t-k}) = 0 \quad \text{para todo o } k=0, \pm 1, \pm 2, \dots \quad (1.2.21)$$

Nestas condições, $\Gamma(k)$ é:

$$\Gamma(k) = \begin{bmatrix} \gamma_{11}(k) & 0 \\ 0 & \gamma_{22}(k) \end{bmatrix} \quad (1.2.22)$$

Por sua vez, $\rho(k)$ assume a expressão:

$$\rho(k) = \begin{bmatrix} \rho_{11}(k) & 0 \\ 0 & \rho_{22}(k) \end{bmatrix} \quad (1.2.23)$$

Exemplo 2.

Suponha-se que $\{X_{1t}\}$ e $\{X_{2t}\}$ são ambos formados a partir do mesmo processo puramente aleatório $\{\varepsilon_{1t}\}$ da seguinte forma:

$$\begin{aligned} X_{1t} &= \varepsilon_t \\ X_{2t} &= \varepsilon_t + 0.4\varepsilon_{t-5} \end{aligned} \quad (1.2.24)$$

com $E(\varepsilon_t) = 0$ e $V(\varepsilon_t) = \sigma_\varepsilon^2$. Uma vez que $E(X_{1t}) = E(X_{2t}) = 0$, tem-se:

$$E(X_t) = \mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (1.2.25)$$

Deste modo, conclui-se que $\Gamma(k)$ assume a forma:

$$\Gamma(k) = \begin{cases} \begin{bmatrix} 0 & 0.4\sigma_\varepsilon^2 \\ 0 & 0.4\sigma_\varepsilon^2 \end{bmatrix} & k = -5 \\ \begin{bmatrix} \sigma_\varepsilon^2 & \sigma_\varepsilon^2 \\ \sigma_\varepsilon^2 & 1.16\sigma_\varepsilon^2 \end{bmatrix} & k = 0 \\ \begin{bmatrix} 0 & 0 \\ 0.4\sigma_\varepsilon^2 & 0.4\sigma_\varepsilon^2 \end{bmatrix} & k = 5 \\ 0 & \text{outros valores de } k \end{cases} \quad (1.2.26)$$

A matriz de correlação é então facilmente determinada, assumindo a expressão:

$$\rho(k) = \begin{cases} \begin{bmatrix} 0 & 0.37 \\ 0 & 0.34 \end{bmatrix} & k = -5 \\ \begin{bmatrix} 1 & 0.93 \\ 0.93 & 1 \end{bmatrix} & k = 0 \\ \begin{bmatrix} 0 & 0 \\ 0.37 & 0.34 \end{bmatrix} & k = 5 \\ 0 & \text{outros valores de } k \end{cases} \quad (1.2.27)$$

2.2. Transformações de uma série cronológica

Quando uma série cronológica multivariada é não estacionária, é necessário aplicar-lhe determinadas transformações que visam a sua estacionarização, uma vez que se irá recorrer a processos estocásticos estacionários de segunda ordem para a sua modelização, à semelhança do que sucede no caso univariado.

Assim, designando uma série cronológica por X_t e as suas componentes por X_{it} ($i=1, \dots, m$), as transformações que irão ser consideradas são de dois tipos:

(i) Visando linearizar a tendência e estabilizar a variância de cada uma das componentes de uma série cronológica multivariada, aplica-se-lhes as transformações de Box e Cox (1964) propostas para o caso univariado:

$$X_{it}^{\lambda_i} = \begin{cases} X_{it}^{\lambda_i} & , \lambda_i \neq 0 \\ \log X_{it} & , \lambda_i = 0 \end{cases} \quad i=1, \dots, m \quad (1.2.28)$$

Assim, quando se referir que o parâmetro de transformação de uma dada série cronológica multivariada é λ , significa que este é o valor máximo dos diferentes parâmetros λ_i ($i=1, \dots, m$).

(ii) Quando as componentes da série cronológica multivariada apresentam tendências não neutras, torna-se necessário proceder à estacionarização

destas, para o que se utiliza o operador diferença, aplicando-o a cada uma dessas componentes:

$$\nabla X_{it} = X_{it} - X_{i,t-1} \quad i=1, \dots, m \quad (1.2.29)$$

Utilizando o operador atraso B, que é tal que $B^j X_{it} = X_{i,t-j}$, é possível expressar esta igualdade da forma:

$$\nabla X_{it} = X_{it} - BX_{it} = (1-B)X_{it} \quad i=1, \dots, m \quad (1.2.30)$$

Considerando o vector X_t , a aplicação do operador diferença produz o resultado:

$$\nabla X_t = X_t - X_{t-1} \quad i=1, \dots, m \quad (1.2.31)$$

Por sua vez, a aplicação do operador atraso a X_t conduz a:

$$B^j X_t = \begin{bmatrix} B^j X_{1t} \\ \vdots \\ B^j X_{mt} \end{bmatrix} = \begin{bmatrix} X_{1,t-j} \\ \vdots \\ X_{m,t-j} \end{bmatrix} = X_{t-j} \quad (1.2.32)$$

o que leva à igualdade:

$$\nabla X_t = X_t - BX_t = (1-B)X_t \quad (1.2.33)$$

Por outro lado, pode ser necessário aplicar o operador diferença várias vezes para estacionarizar a tendência da série cronológica, o que conduz às expressões:

$$\begin{aligned} \nabla^{d_i} X_{it} &= (1-B)^{d_i} X_{it} \quad i=1, \dots, m \\ \nabla^d X_t &= B^d X_t \end{aligned} \quad (1.2.34)$$

onde ∇^d e B^d são matrizes diagonais de dimensão $(m \times m)$ cujos elementos genéricos são ∇^{d_i} e $(1-B)^{d_i}$ respectivamente e d_i é o número de vezes que o

operador diferença é aplicado a X_{it} , sendo d_i o máximo dos diferentes d_i ($i=1, \dots, m$). Convém realçar o facto de as expressões (1.2.34) permitirem que o número de vezes que o operador diferença é aplicado não seja o mesmo para todas as componentes do vector X_t , o que proporciona maior flexibilidade e pretende evitar certos problemas como o da sobrediferenciação, que será analisado mais à frente.

No entanto, a tendência pode ainda apresentar uma componente periódica, tornando necessária para a sua estacionarização a diferenciação sazonal, ou seja, a aplicação do operador diferença de acordo com o período sazonal s , designado de operador diferença sazonal, o que conduz a:

$$\begin{aligned} \nabla_s X_{it} &= (1-B^s)X_{it} - X_{i,t-s} & i=1, \dots, m \\ \nabla_s X_t &= (1-B^s)X_t - X_{t-s} \end{aligned} \quad (1.2.35)$$

onde ∇_s é o operador diferença sazonal e B^s o operador atraso sazonal, que é tal que $B^s X_{it} = X_{i,t-s}$ e $B^s X_t = X_{t-s}$.

Pode também ser necessário aplicar este operador por diversas vezes, o que produz as expressões:

$$\begin{aligned} \nabla_s^{d_i} X_{it} &= (1-B^s)^{d_i} X_{it} & i=1, \dots, m \\ \nabla_s^{d_i} X_t &= B X_t \end{aligned} \quad (1.2.36)$$

onde $\nabla_s^{d_i}$ e B^{d_i} são matrizes diagonais de dimensão $(m \times m)$ cujos elementos genéricos são respectivamente $\nabla_s^{d_i}$ e $(1-B)^{d_i}$ e d_{is} é o número de vezes que se aplica a diferenciação sazonal a X_{it} , sendo o respectivo número máximo designado por d_i . Do mesmo modo, também neste caso se admite que o grau de diferenciação sazonal seja diferente conforme a série univariada X_{it} .

No caso de a componente periódica incluir mais do que um período sazonal, torna-se necessário proceder à diferenciação sazonal de acordo com cada período. Assim, se, por exemplo, se revelar a existência de dois períodos sazonais, é necessário aplicar as transformações:

$$\begin{aligned} \nabla_{s_1}^{d_{i_1}} \nabla_{s_2}^{d_{i_2}} X_{it} &= (1 - B^{s_1})^{d_{i_1}} (1 - B^{s_2})^{d_{i_2}} X_{it} \quad i = 1, \dots, m \\ \nabla_{s_1}^{d_{i_1}} \nabla_{s_2}^{d_{i_2}} X_t &= B_{s_1}^{d_{i_1}} B_{s_2}^{d_{i_2}} X_t \end{aligned} \quad (1.2.37)$$

onde:

s_1 e s_2 são os períodos sazonais;

d_{i_1} e d_{i_2} representam o número de diferenciações sazonais aplicadas a X_{it} com respeito a s_1 e a s_2 respectivamente sendo d_{s_1} e d_{s_2} os correspondentes valores máximos;

$\nabla_{s_1}^{d_{i_1}}$ e $\nabla_{s_2}^{d_{i_2}}$ são matrizes diagonais de dimensão $(m \times m)$ cujos elementos genéricos são $\nabla_{s_1}^{d_{i_1}}$ e $\nabla_{s_2}^{d_{i_2}}$ respectivamente;

$B_{s_1}^{d_{i_1}}$ e $B_{s_2}^{d_{i_2}}$ são matrizes diagonais de dimensão $(m \times m)$ cujos elementos genéricos são $(1 - B^{s_1})^{d_{i_1}}$ e $(1 - B^{s_2})^{d_{i_2}}$ respectivamente.

Deste modo, as transformações acabadas de analisar podem ser resumidas na expressão:

$$\begin{aligned} \nabla_{s_1}^{d_{i_1}} \nabla_{s_2}^{d_{i_2}} (X_{it}^{\lambda_i}) &= (1 - B^{s_1})^{d_{i_1}} (1 - B^{s_2})^{d_{i_2}} X_{it}^{\lambda_i} \quad i=1, \dots, m \\ \nabla_{s_1}^{d_{i_1}} \nabla_{s_2}^{d_{i_2}} (X_t^{\lambda_i}) &= B_{s_1}^{d_{i_1}} B_{s_2}^{d_{i_2}} X_t^{\lambda_i} \end{aligned} \quad (1.2.38)$$

onde se considerou a existência de apenas um período sazonal .

No entanto, a diferenciação de cada série univariada X_{it} , ($i=1, \dots, m$) tendo em vista a estacionarização das respectivas tendências, pode ser inadequada para a análise de uma série multivariada (Hillmer e Tiao, 1979; Lütkepohl, 1982). Esta questão será novamente analisada mais à frente.

II. MODELOS PROBABILÍSTICOS

Tendo em vista a modelização de uma série cronológica multivariada, é conveniente considerar certas classes de processos estocásticos multivariados estacionários, designados por modelos probabilísticos, que permitem proceder a essa modelização.

Assim, irá ser considerado em primeiro lugar o ruído branco ou processo puramente aleatório, em seguida o processo auto-regressivo, posteriormente o processo de médias móveis e, por fim, o processo misto auto-regressivo e de médias móveis, em que irá ser incorporada a possibilidade de existência de não estacionaridade, originando o processo misto auto-regressivo e de médias móveis integrado.

Por outro lado, a análise destes modelos probabilísticos não irá ser alargada aos processos sazonais destinados a modelizar séries cronológicas em que se revela exclusivamente a presença de uma componente sazonal, nem aos processos multiplicativos sazonais não-sazonais. Com efeito, tal extensão da análise pareceu desnecessária, uma vez que constitui uma generalização directa do que se encontra estabelecido para o caso univariado e resulta facilmente dos diferentes modelos probabilísticos que irão agora ser apresentados. Esta metodologia irá também ser adoptada nos capítulos subsequentes, conforme será oportunamente referido.

Por fim, refira-se que se admite que o valor esperado do processo estocástico considerado é o vector nulo.

1. Processos estocásticos multivariados

Esta classe de processos destina-se a modelizar séries cronológicas onde não exista qualquer componente periódica, sendo facilmente generalizável ao caso em que exista sazonalidade.

1.1. Ruído branco ou processo puramente aleatório multivariado

$\{X_t, t=0, \pm 1, \pm 2, \dots\}$ diz-se um ruído branco ou processo puramente aleatório multivariado se for estacionário e a sua matriz de covariância assumir a forma:

$$\Gamma(k) = \begin{cases} \Sigma & \text{se } k = 0 \\ \mathbf{0} & \text{se } k \neq 0 \end{cases} \quad (2.1.1)$$

onde Σ é a matriz de variâncias e covariâncias das componentes do vector X_t , sendo uma matriz simétrica, definida positiva e de dimensão $(m \times m)$. Habitualmente, admitir-se-à a hipótese de o vector dos valores esperados deste tipo de processos ser o vector nulo, o que significa que $\Sigma = E(X_t X_t^T)$.

As componentes do vector X_t são processos puramente aleatórios univariados que não estão correlacionados entre si para momentos diferentes, mas que podem estar contemporaneamente correlacionados, o que significa que Σ é um matriz não necessariamente diagonal.

Um ruído branco ou processo puramente aleatório multivariado, com vector dos valores esperados nulo, passará a ser designado por $\{\varepsilon_t, t=0, \pm 1, \pm 2, \dots\}$.

1.2. Processo auto-regressivo

Seja $\{\varepsilon_t, t=0, \pm 1, \pm 2, \dots\}$ um processo puramente aleatório multivariado.

$\{X_t, t=0, \pm 1, \pm 2, \dots\}$ diz-se um processo auto-regressivo multivariado de ordem p , designando-se por $AR(p)$, se puder ser representado da forma:

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t \quad (2.1.2)$$

onde ϕ_1, \dots, ϕ_p são matrizes reais de dimensão $(m \times m)$. Utilizando o operador atraso, este processo pode também ser representado da forma:

$$\phi(B)X_t = \varepsilon_t \quad (2.1.3)$$

onde $\phi(B)X_t = I_m - \phi_1 B - \dots - \phi_p B^p$ é um polinómio matricial, sendo cada elemento da matriz $\phi(B)$ é um polinómio de coeficientes reais e grau não superior a p , designando I_r genericamente uma matriz identidade de dimensão $(r \times r)$. Tal como o processo $AR(p)$ univariado, este processo é invertível, sendo estacionário se, representando $|M|$ o determinante de uma qualquer matriz quadrada M , as raízes Z_r ($r=1, \dots, p$) da equação matricial:

$$|Z^p I_m - Z^{p-1} \phi_1 - \dots - \phi_p| = 0 \quad (2.1.4)$$

forem todas inferiores a 1 em valor absoluto, isto é, a r -ésima raiz Z_r deve ser inferior a 1 em valor absoluto, para todo o $r=1, \dots, p$. De forma equivalente, este processo diz-se estacionário se as raízes B_r ($r=1, \dots, p$) da equação matricial:

$$|\phi(B)| = 0 \quad (2.1.5)$$

forem todas superiores a 1 em valor absoluto, ou seja, se estiverem fora do círculo unitário, isto é, a r -ésima raiz B_r deve obedecer à condição $|B_r| > 1$, para todo o $r=1, \dots, p$. Este resultado prova-se por generalização do que se passa no caso univariado. No entanto, se houver raízes da equação (2.1.4) que estejam sobre o círculo unitário, admite-se que o

processo tem o seu início num qualquer momento finito do tempo, com valores iniciais fixos. É importante referir ainda que esta condição assegura que os elementos da matriz de covariância deste processo sejam finitos (Jenkins e Alavi, 1981).

As equações de Yule-Walker podem também ser generalizadas ao caso multivariado a partir da representação de um processo AR(p) estacionário:

$$\mathbf{X}_t = \phi_1 \mathbf{X}_{t-1} + \dots + \phi_p \mathbf{X}_{t-p} + \boldsymbol{\varepsilon}_t \quad (2.1.6)$$

sendo $\{\mathbf{X}_t\}$ um processo estacionário, prova-se facilmente que:

$$E(\mathbf{X}_t) = \mathbf{0} \quad (2.1.7)$$

Assim, multiplicando ambos os membros da equação (2.1.6) por \mathbf{X}_{t-k}^T e calculando valores esperados, obtém-se a relação:

$$\Gamma(k) = \phi_1 \Gamma(k-1) + \dots + \phi_p \Gamma(k-p) \quad k=1,2,\dots \quad (2.1.8)$$

As equações resumidas em (2.1.8) são equações de Yule-Walker no caso multivariado e podem ser resolvidas recorrendo às relações:

$$\Gamma(-k) = \Gamma^T(k)$$

$$\Gamma(0) = \phi_1 \Gamma(-1) + \dots + \phi_p \Gamma(-p) + \Sigma \quad (2.1.9)$$

Por outro lado, as equações de Yule-Walker podem ser utilizadas para determinar os parâmetros ϕ_1, \dots, ϕ_p do seguinte modo:

$$\begin{bmatrix} \phi_1^T \\ \vdots \\ \phi_p^T \end{bmatrix} = \begin{bmatrix} \Gamma(0) & \dots & \Gamma(p-1) \\ \vdots & & \vdots \\ \Gamma(1-p) & \dots & \Gamma(0) \end{bmatrix}^{-1} \begin{bmatrix} \Gamma^T(1) \\ \vdots \\ \Gamma^T(p) \end{bmatrix} \quad (2.1.10)$$

No que respeita a este tipo de processo, é importante referir ainda a propriedade de decréscimo gradual quer dos elementos das matrizes de covariância, quer dos elementos das matrizes de correlação quando um $|k|$ aumenta. Assim, quando se está em presença de um processo auto-regressivo, os valores das funções de autocovariância e de autocorrelação, bem como os das funções de covariância e de correlação cruzadas, apresentam um decréscimo gradual para zero à medida que $|k|$ aumenta.

Por outro lado, designando a matriz de correlação parcial por $P(k)$, a sua definição é:

$$P(k) = \phi_k \quad (2.1.11)$$

onde ϕ_k é a última matriz de parâmetros da equação que define uma regressão de X_t sobre X_{t-l} ($l=1, \dots, k$). Verifica-se então que, para um modelo AR(p), esta função possui a propriedade:

$$P(k) = \begin{cases} \phi_k & \text{se } k = p \\ 0 & \text{se } k > p \end{cases} \quad (2.1.12)$$

Assim, à semelhança do que sucede no caso univariado, esta função anula-se a partir da ordem do processo, o que a torna num instrumento privilegiado para a identificação de processos auto-regressivos, uma vez que, se aquela sofrer um corte abrupto para certo "lag", significa que se está em presença de um processo deste tipo, sendo a respectiva ordem igual ao valor desse "lag".

Em suma, a função de correlação parcial permite detectar a presença de um processo auto-regressivo e determinar a sua ordem.⁽²⁾

A título de exemplo, examinemos o processo AR(1) bivariado, ou seja, quando $m=2$, cuja representação assume a forma:

$$\begin{aligned}
X_{1t} &= \phi_{11,1} X_{1,t-1} + \phi_{12,1} X_{2,t-1} + \varepsilon_{1t} \\
X_{2t} &= \phi_{21,1} X_{1,t-1} + \phi_{22,1} X_{2,t-1} + \varepsilon_{2t}
\end{aligned}
\tag{2.1.13}$$

onde $\phi_{ij,l}$ ($i,j=1,2 ; l=1$) designa o parâmetro auto-regressivo relativo à j -ésima variável da i -ésima equação e para o momento $t-l$. Utilizando notação matricial, obtém-se a representação:

$$\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} \phi_{11,1} & \phi_{12,1} \\ \phi_{21,1} & \phi_{22,1} \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}
\tag{2.1.14}$$

o que é equivalente a:

$$X_t = \phi_1 X_{t-1} + \varepsilon_t
\tag{2.1.15}$$

Admitindo que $\{X_t\}$ é um processo estacionário, a sua matriz de covariância pode ser obtida através das equações de Yule-Walker, que assumem a forma:

$$\Gamma(k) = \phi_1 \Gamma(k-1) \quad k=1,2,\dots
\tag{2.1.16}$$

Trata-se de uma equação às diferenças matricial, cuja solução é:

$$\Gamma(k) = \phi_1^k \Gamma(0) \quad k=1,2,\dots
\tag{2.1.17}$$

sendo o $\Gamma(0)$ determinado em (2.1.9), o que, neste caso, assume a expressão:

$$\Gamma(0) = \phi_1 \Gamma(0) \phi_1^T + \Sigma
\tag{2.1.18}$$

chegando-se a este resultado com o recurso a (2.1.17).

Harvey (1981, pág.49) sugere duas formas possíveis de determinar a solução desta equação, de modo a obter $\Gamma(0)$. Assim, uma hipótese consiste na adopção de um método iterativo que calcule aproximações

sucessivas a partir de um valor inicial para $\Gamma(0)$. Alternativamente, é possível obter uma solução a partir da identidade:

$$\text{vec}[\Gamma(0)] = [\mathbf{I}_{m^2} - \phi_1 \otimes \phi_1]^{-1} \text{vec} \Sigma \quad (2.1.19)$$

onde \otimes designa o produto de Kronecker e $\text{vec}(\cdot)$ significa "vectorização", isto é, que as colunas de uma matriz de dimensão $(n \times l)$ são colocadas na vertical, formando um vector de dimensão $[(nl) \times 1]$.

1.3. Processo de médias móveis multivariado

Seja $\{\varepsilon_t, t=0, \pm 1, \pm 2, \dots\}$ um processo puramente aleatório multivariado.

$\{X_t, t=0, \pm 1, \pm 2, \dots\}$ diz-se um processo de médias móveis multivariado de ordem q , designando-se por MA(q) (de "Moving Averages"), se puder ser representado da forma:

$$X_t = \theta_0 \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (2.1.20)$$

onde $\theta_0, \theta_1, \dots, \theta_q$ são matrizes reais de dimensão $(m \times m)$. Habitualmente, considerar-se-à $\theta_0 = \mathbf{I}_m$. Recorrendo ao operador atraso, este processo pode também ser representado da forma:

$$X_t = \theta(B) \varepsilon_t \quad (2.1.21)$$

onde $\theta(B) = \mathbf{I}_m + \theta_1 B + \dots + \theta_q B^q$ é um polinómio matricial, sendo cada elemento da matriz $\theta(B)$ um polinómio de coeficientes reais e grau não superior a q . Tal como o processo MA(q) univariado, trata-se de um processo estacionário. Com efeito, o vector $E(X_t)$ é o vector nulo e facilmente se prova que a matriz de covariância de $\{X_t\}$ tem a expressão genérica:

$$\Gamma(k) = \begin{cases} \sum_{i=k}^q \theta_i \Sigma \theta_{i-k}^T & k = 0, 1, \dots, q \\ \mathbf{0} & k > q \\ \Gamma^T(-k) & k < 0 \end{cases} \quad (2.1.22)$$

com $\theta_0 = \mathbf{I}_m$.

Tal como sucede no caso univariado, verifica-se um corte abrupto da função a partir da ordem do processo, o que a converte, bem como à função de correlação, num instrumento privilegiado para a identificação de processos de médias móveis. De facto, esta propriedade permite, por um lado, distinguir um processo de médias móveis e um processo auto-regressivo, uma vez que, conforme já foi referido, tanto a função de covariância como a de correlação deste último decrescem para zero e, por outro lado, permite determinar a ordem de um processo de médias móveis.

Em resumo, a função de correlação, bem como a de covariância, permite também distinguir entre processos auto-regressivos e processos de médias móveis, possibilitando ainda a determinação da ordem destes últimos.

Por outro lado, este tipo de processos é invertível se as raízes Z_r ($r=1, \dots, q$) da equação matricial:

$$|Z^q \mathbf{I}_m + Z^{q-1} \theta_1 + \dots + \theta_q| = 0 \quad (2.1.23)$$

forem todas inferiores a 1 em valor absoluto, isto é, a r -ésima raiz Z_r deve verificar a condição $|Z_r| < 1$ para todo o $r=1, \dots, q$. De forma equivalente, conclui-se que este tipo de processos é invertível se as raízes B_r ($r=1, \dots, q$) da equação matricial:

$$|\theta(B)| = 0 \quad (2.1.24)$$

forem todas superiores a 1 em valor absoluto, ou seja, se estiverem fora do círculo unitário, isto é, a r -ésima raiz B_r deve obedecer à condição $|B_r| > 1$, para todo o $r=1, \dots, q$. Este resultado prova-se também por generalização do que se passa no caso univariado.

A título de exemplo, analisemos o processo MA(1) multivariado, que assume a forma:

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} \quad (2.1.25)$$

Para este processo, obtém-se os resultados:

$$\Gamma(k) = \begin{cases} \Sigma + \theta_1 \Sigma \theta_1^T & k = 0 \\ \theta_1 \Sigma & k = 1 \\ \Sigma \theta_1^T & k = -1 \\ \mathbf{0} & |k| \geq 2 \end{cases} \quad (2.1.26)$$

É importante referir que $\Gamma(-1) = \Gamma^T(1)$ e que a função se anula a partir da ordem do processo, ou seja, para $|k| \geq 2$, tal como sucede com o processo MA(1) univariado.

Admitindo que se trata de um processo bivariado, a sua representação matricial assume a expressão:

$$\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} + \begin{bmatrix} \theta_{11,1} & \theta_{12,1} \\ \theta_{21,1} & \theta_{22,1} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{bmatrix} \quad (2.1.27)$$

onde $\theta_{ij,l}$ ($i, j=1, 2, ; l=1$) designa o parâmetro de médias móveis relativo à j -ésima variável da i -ésima equação para o momento $t-l$. Esta representação pode ser expressa da forma equivalente:

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} = \theta(B) \varepsilon_t \quad (2.1.28)$$

onde $\theta(B) = I_2 + \theta_1 B$. É ainda possível representar este processo da forma alternativa:

$$\begin{aligned} X_{1t} &= \varepsilon_{1t} + \theta_{11,1} \varepsilon_{1,t-1} + \theta_{12,1} \varepsilon_{2,t-1} \\ X_{2t} &= \varepsilon_{2t} + \theta_{21,1} \varepsilon_{1,t-1} + \theta_{22,1} \varepsilon_{2,t-1} \end{aligned} \quad (2.1.29)$$

O processo MA(q) envolve um número finito de parâmetros, o mesmo sucedendo com o processo AR(p), uma vez que se admite que tanto q como p são finitos.

No entanto, é possível representar um processo AR(p) multivariado estacionário como um processo de médias móveis invertível de ordem infinita (Brockwell e Davis, 1987, pág.408). Assim, considere-se o processo AR(p) estacionário:

$$\phi(B)X_t = \varepsilon_t \quad (2.1.30)$$

A condição de estacionaridade deste tipo de processos assegura que $\phi(B)$ é uma matriz invertível para os valores de B contidos no círculo unitário, uma vez que implica que os zeros de $|\phi(B)|$ estejam todos fora desse círculo. Deste modo, invertendo $\phi(B)$, obtém-se:

$$X_t = \phi(B)^{-1} \varepsilon_t \quad (2.1.31)$$

sendo possível provar que o desenvolvimento em série de $\phi(B)^{-1}$ é:

$$\phi(B)^{-1} = \sum_{l=0}^{\infty} \psi_l B^l \quad |B| \leq 1 \quad (2.1.32)$$

com $\psi_0 = I_m$ (Jenkins e Alavi, 1981; Wilson, 1973) e sendo o desenvolvimento em série de $\phi(B)^{-1}$ calculado a partir da relação:

$$\phi(B)^{-1} = \phi^A(B) / |\phi(B)^{-1}| \quad (2.1.33)$$

designando ϕ^A a matriz adjunta de $\phi(B)$.

Assim, a representação de um processo AR(p) estacionário como um processo de médias móveis invertível de ordem infinita resulta de (2.1.31) e (2.1.32) e assume a forma:

$$X_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} = \psi(B) \varepsilon_t \quad (2.1.34)$$

onde $\psi_0 = I_m$ e $\psi(B) = \sum_{i=0}^{\infty} \psi_i B^i$.

Utilizando o exemplo do processo AR(1), obtém-se a representação:

$$X_t = \sum_{i=0}^{\infty} \phi_i^l \varepsilon_{t-i} \quad (2.1.35)$$

Por sua vez, um processo MA(q) invertível multivariado pode ser representado como um processo auto-regressivo estacionário de ordem infinita (Brockwell e Davis, 1987, pág.408). Seja então o processo MA(q) invertível multivariado:

$$X_t = \theta(B) \varepsilon_t \quad (2.1.36)$$

Tratando-se de processo invertível, os zeros de $|\theta(B)|$ situam-se todos fora do círculo unitário, o que significa que a matriz de $\theta(B)$ é invertível para os valores de B contidos no interior de círculo. Deste modo, invertendo $\theta(B)$ obtém-se:

$$\theta(B)^{-1} X_t = \varepsilon \quad (2.1.37)$$

provando-se neste caso que o desenvolvimento em série de $\theta(B)^{-1}$ é (Jenkins e Alavi, 1981; Wilson, 1973):

$$\theta(B)^{-1} = I_m - \sum_{i=1}^{\infty} \pi_i B^i \quad |B| \leq 1 \quad (2.1.38)$$

sendo o desenvolvimento em série de $\theta(B)^{-1}$ determinado a partir da relação:

$$\theta(B)^{-1} = \theta^A(B) / |\theta(B)^{-1}| \quad (2.1.39)$$

onde θ^A é a matriz adjunta de $\theta(B)$.

Deste modo, a representação de um processo MA(q) invertível como um processo auto-regressivo estacionário de ordem infinita é obtida a partir de (2.1.37) e (2.1.38) e assume a forma:

$$X_t - \sum_{i=1}^{\infty} \pi_i X_{t-i} = \pi(B) X_t = \varepsilon_t \quad (2.1.40)$$

onde $\pi(B) = I_m - \sum_{i=1}^{\infty} \pi_i B^i$. Exemplificando com o processo MA(1), obtém-se:

$$X_t + \sum_{i=1}^{\infty} (-1)^i \theta_i' X_{t-i} = \varepsilon_t \quad (2.1.41)$$

É evidente que estes dois resultados, expressos nas relações (2.1.34) e (2.1.40), constituem generalização dos correspondentes resultados para o caso univariado.

1.4. Processo misto auto-regressivo e de médias móveis multivariado

Seja $\{\varepsilon_t, t=0, \pm 1, \pm 2, \dots\}$ um processo puramente aleatório multivariado.

$\{X_t, t=0, \pm 1, \pm 2, \dots\}$ diz-se processo misto auto-regressivo e de médias móveis multivariado de ordens (p, q) designando-se por ARMA (p, q) , se puder ser representado da forma:

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \theta_0 \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (2.1.42)$$

onde ϕ_1, \dots, ϕ_p e $\theta_0, \theta_1, \dots, \theta_q$ são matrizes reais de dimensão $(m \times m)$. Habitualmente, considerar-se-à $\theta_0 = I_m$. Utilizando o operador atraso, obtém-se a representação equivalente:

$$\phi(B)X_t = \theta(B)\varepsilon_t \quad (2.1.43)$$

onde:

$$\begin{aligned} \phi(B) &= I_m - \phi_1 B - \dots - \phi_p B^p \\ \theta(B) &= I_m + \theta_1 B + \dots + \theta_q B^q \end{aligned} \quad (2.1.44)$$

O operador de médias móveis $\theta(B)$ é um polinómio matricial cujos elementos são polinómios de coeficientes reais e grau não superior a q , tal como para um processo MA (q) . Por sua vez, o operador auto-regressivo $\phi(B)$ é um polinómio matricial cujos elementos são polinómios de coeficientes reais e grau não superior a p , à semelhança do que foi definido para um processo AR (p) .

Note-se que o processo ARMA (p, q) se converte num processo MA (q) se $p=0$ e num processo AR (p) se $q=0$.

Para que um processo ARMA (p, q) multivariado seja estacionário, é necessário que as raízes da equação matricial:

$$|Z^p I_m - Z^{p-1} \phi_1 - \dots - \phi_p| = 0 \quad (2.1.45)$$

sejam todas inferiores a 1 em valor absoluto ou, o que é equivalente, que as raízes da equação matricial:

$$|\phi(B)| = 0 \quad (2.1.46)$$

estejam todas fora do círculo unitário, assegurando-se desta forma a estacionaridade da componente auto-regressiva. No entanto, se houver raízes da equação (2.1.46) sobre o círculo unitário, admite-se que o processo se inicia num dado momento do tempo, com valores iniciais fixos. Esta condição assegura também para este tipo de processo que os elementos da matriz de covariância deste sejam finitos. Trata-se, como é óbvio, de condições idênticas às que foram colocadas para o processo auto-regressivo.

Por outro lado, para que este processo seja invertível, é necessário que as raízes da equação matricial:

$$|Z^q I_m + Z^{q-1} \theta_1 + \dots + \theta_q| = 0 \quad (2.1.47)$$

sejam todas inferiores a 1 em valor absoluto ou, de forma equivalente, que as raízes da equação matricial:

$$|\theta(B)| = 0 \quad (2.1.48)$$

estejam todas fora do círculo unitário, de modo a assegurar a invertibilidade da componente de médias móveis do processo ARMA(p,q). É também uma condição idêntica à que foi colocada sobre a invertibilidade do processo de médias móveis.

Qualquer processo ARMA(p,q) multivariado invertível e estacionário, assumindo a forma (2.1.43), pode ser representado como um processo auto-regressivo de ordem infinita, tal como um processo

MA(q), sendo necessária a verificação das mesmas condições (Brockwell e Davis, 1987, pág.408). Para o efeito, considere-se a seguinte expressão, obtida a partir de (2.1.43):

$$\theta(B)^{-1}\phi(B)X_t = \varepsilon_t \quad (2.1.49)$$

O desenvolvimento em série de $\theta(B)^{-1}$ permite obter a referida representação que assume a forma:

$$X_t - \sum_{l=1}^{\infty} \pi_l X_{t-l} = \pi(B)X_t = \varepsilon_t \quad (2.1.50)$$

onde $\pi(B) = I_m - \sum_{l=1}^{\infty} \pi_l B^l$ (Jenkins e Alavi, 1981; Wilson, 1973) e as matrizes π_l ($l=1,2,\dots$) são determinadas igualando os coeficientes das potências de B na equação (Tiao e Tsay, 1983):

$$\phi(B) = \theta(B) \pi(B) \quad (2.1.51)$$

Por outro lado, este processo pode também ser representado como um processo de médias móveis de ordem infinita, tal como um processo AR(p), sendo ainda necessária a verificação das mesmas condições (Brockwell e Davis, 1987, pág.408). Para o efeito, considere-se a expressão seguinte, deduzida a partir de (2.1.43):

$$X_t = \phi(B)^{-1}\theta(B)\varepsilon_t \quad (2.1.52)$$

O desenvolvimento em série de $\phi(B)^{-1}$ conduz à referida representação, que assume a forma:

$$X_t = \sum_{l=0}^{\infty} \psi_l \varepsilon_{t-l} = \psi(B)\varepsilon_t \quad (2.1.53)$$

onde $\psi_0 = I_m$ e $\psi(B) = \sum_{l=0}^{\infty} \psi_l B^l$ (Jenkins e Alavi, 1981; Wilson, 1973) e as matrizes ψ_l ($l=1,2,\dots$) são obtidas igualando os coeficientes das potências de B na relação (Tiao e Tsay, 1983):

$$\phi(B)\psi(B) = \theta(B) \quad (2.1.54)$$

A expressão (2.1.53) define a única solução estacionária para a equação (2.1.43), que, por sua vez, define o processo ARMA(p,q).

Utilizando a representação definida em (2.1.53), é possível expressar X_{t-k} da forma:

$$X_{t-k} = \psi(B)\epsilon_{t-k} = \sum_{l=0}^{\infty} \psi_l \epsilon_{t-k-l} \quad (2.1.55)$$

Assim, verifica-se a relação:

$$E(\epsilon_{t-h} X_{t-k}^T) = \begin{cases} \Sigma \psi_l^T & k+l=h \\ 0 & k+l \neq h \end{cases} \quad (2.1.56)$$

Note-se que $\Sigma \psi_l^T = \Sigma$ para $k=h$, ou seja, para $l=0$. Esta expressão é de extrema utilidade para a caracterização da matriz de covariância de um processo ARMA(p,q), sendo útil para o efeito considerar previamente a igualdade:

$$(X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p}) X_{t-k}^T = (\epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}) X_{t-k}^T \quad (2.1.57)$$

Calculando valores esperados em ambos os membros, obtém-se a seguinte expressão para a matriz de covariância de um processo ARMA(p,q) invertível e estacionário:

$$\Gamma(k) = \begin{cases} \sum_{u=1}^r \phi_u \Gamma(k-u) + \sum_{u=k}^r \theta_u \Sigma \psi_{u-k}^T & k = 0, 1, \dots, r \\ \sum_{u=1}^r \phi_u \Gamma(k-u) & k = r+1, r+2, \dots \end{cases} \quad (2.1.58)$$

onde $\theta_0 = \mathbf{I}_m$, $r = \max(p, q)$ e:

(i) $\phi_{p+1} = \dots = \phi_r = 0$, se $p < q$;

(ii) $\theta_{q+1} = \dots = \theta_r = 0$, se $q < p$;

Note-se que, em particular, se se tiver $p=0$, obtém-se a expressão da matriz de covariância de um processo MA(q) e, se $q=0$, resultam de (2.1.58) as equações de Yule-Walker, tal como foram caracterizadas em (2.1.8) para o processo AR(p).

Observando a expressão de $\Gamma(k)$ em (2.1.58), conclui-se imediatamente que a função de correlação não sofre qualquer corte abrupto, antes se verificando um decréscimo gradual para zero dos seus elementos à medida que $|k|$ aumenta, sucedendo o mesmo com a função de correlação parcial (Tiao e Box, 1979, 1981). Um tal padrão de comportamento das funções de correlação e de correlação parcial de um dado processo, permite detectar a presença de um processo ARMA(p,q).

A título de exemplo, considere-se o processo ARMA(1,1) bivariado, cuja representação matricial é:

$$\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} \phi_{11,1} & \phi_{12,1} \\ \phi_{21,1} & \phi_{22,1} \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} + \begin{bmatrix} \theta_{11,1} & \theta_{12,1} \\ \theta_{21,1} & \theta_{22,1} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{bmatrix} \quad (2.1.59)$$

onde $\phi_{ij,l}$ designa os parâmetros auto-regressivos, tal como no caso de um processo auto-regressivo, e $\theta_{ij,l}$ designa os parâmetros de médias móveis, tal como para um processo de médias móveis ($i, j=1, 2; l=1$). Esta representação pode também ser expressa da forma:

$$\mathbf{X}_t = \phi_1 \mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t + \theta_1 \boldsymbol{\varepsilon}_{t-1} \quad (2.1.60)$$

o que, utilizando o operador atraso é equivalente a:

$$(\mathbf{I}_2 - \phi_1 B) \mathbf{X}_t = (\mathbf{I}_2 + \theta_1 B) \boldsymbol{\varepsilon}_t \quad (2.1.61)$$

ou seja:

$$\phi(B)X_t = \theta(B)\varepsilon_t \quad (2.1.62)$$

Este processo pode ainda ser expresso da forma:

$$\begin{aligned} X_{1t} &= \phi_{11,1}X_{1,t-1} + \phi_{12,1}X_{2,t-1} + \varepsilon_{1t} + \theta_{11,1}\varepsilon_{1,t-1} + \theta_{12,1}\varepsilon_{2,t-1} \\ X_{2t} &= \phi_{21,1}X_{1,t-1} + \phi_{22,1}X_{2,t-1} + \varepsilon_{2t} + \theta_{21,1}\varepsilon_{1,t-1} + \theta_{22,1}\varepsilon_{2,t-1} \end{aligned} \quad (2.1.63)$$

Por outro lado, a função de covariância de um processo deste tipo assume a expressão:

$$\Gamma(k) = \begin{cases} \phi_1 \Gamma^T(1) + \Sigma + \theta_1 \Sigma \psi_1^T & k = 0 \\ \phi_1 \Gamma(0) + \theta_1 \Sigma & k = 1 \\ \phi_1 \Gamma(k-1) & k = 2, 3, \dots \end{cases} \quad (2.1.64)$$

1.5. Processo misto auto-regressivo e de médias móveis integrado multivariado

Generalizando o que sucede no caso univariado, este tipo de processo pretende modelizar séries cronológicas multivariadas não estacionárias, sendo por isso um processo não estacionário.

Assim, admita-se um processo estocástico multivariado que assume a representação:

$$\varphi(B)X_t = \theta(B)\varepsilon_t \quad (2.1.65)$$

onde a equação matricial:

$$|\varphi(B)| = 0 \quad (2.1.66)$$

admite d raízes unitárias. Conclui-se então que, neste caso, o processo estocástico $\{X_t\}$ é não estacionário, designando-se $\varphi(B)$ por operador auto-regressivo não estacionário. Com efeito, revela-se extemamente útil

representar processos não estacionários permitindo a existência de raízes unitárias na equação (2.1.66).

Deste modo, seja $\{\epsilon_t, t=0, \pm 1, \pm 2, \dots\}$ um processo puramente aleatório multivariado.

$\{X_t, t=0, \pm 1, \pm 2, \dots\}$ diz-se um processo misto auto-regressivo e de médias móveis integrado multivariado, designando-se por ARIMA(p,d,q) (de "Autoregressive Integrated Moving Averages") se puder ser representado da forma:

$$\varphi(B)X_t = \theta(B)\epsilon_t \quad (2.1.67)$$

onde $\varphi(B)$ representa o operador auto-regressivo não estacionário, cujo determinante admite d zeros iguais a 1, e que assume a expressão:

$$\varphi(B) = (I_m - \phi_1 B - \dots - \phi_p B^p) B^d = \phi(B) B^d \quad (2.1.68)$$

sendo $\phi(B)$ o operador auto-regressivo, definido do mesmo modo que para o processo ARMA(p,q), e B^d uma matriz diagonal de dimensão (m x m) cujo elemento genérico é $(1-B)^{d_i}$, $i=1, \dots, m$, com $d_i \leq d$, tal como definido atrás;

$\theta(B)$ representa o operador de médias móveis, sendo definido tal como para o processo ARMA(p,q).

Uma vez que $B^d = \nabla^d$, o processo ARIMA(p,d,q) acabado de apresentar pode ser representado da forma:

$$\phi(B)\nabla^d X_t = \theta(B)\epsilon_t \quad (2.1.69)$$

onde ∇^d é uma matriz diagonal de dimensão (m x m) cujo elemento genérico é ∇^{d_i} ($i=1, \dots, m$), conforme já foi definido atrás.

Note-se que, se $d=0$, obtém-se um processo ARMA(p,q); se $q=d=0$, obtém-se um processo AR(p) e, se $p=d=0$, obtém-se um processo MA(q).

Por outro lado, é possível representar este tipo de processo através da expressão:

$$\phi(B)Y_t = \theta(B)\varepsilon_t \quad (2.1.70)$$

onde $Y_t = \nabla^d X_t$, cujo elemento genérico é $Y_{it} = \nabla^d X_{it}$, $i=1, \dots, m$. A expressão (2.1.70) revela que o processo $\{Y_t\}$ é um processo ARMA(p,q), verificando todas as características deste e sendo estacionário.

Deste modo, conclui-se que, através da aplicação adequada do operador diferença, é possível transformar um processo não estacionário de modo a convertê-lo num processo estacionário. Por outro lado, Hillmer e Tiao (1979) indicam que é muitas vezes possível que combinações lineares das componentes do processo multivariado sejam estacionárias.

No entanto, tal como refere Lütkepohl (1982a) e Tiao e Tsay (1983), a aplicação do operador diferença a cada um dos processos univariados componentes do processo multivariado, apesar de permitir a estacionarização dos primeiros, não implica necessariamente que o segundo seja também convertido num processo estacionário. Assim, Hillmer e Tiao (1979) e Lütkepohl (1982a) indicam que este procedimento pode conduzir à sobre-diferenciação, ou seja, a uma diferenciação excessiva, o que pode ter como consequências a não invertibilidade do operador de médias móveis e a introdução de complicações desnecessárias no modelo. A adopção de um operador diferença matricial para a estacionarização do processo tem também em vista evitar o problema da sobre-diferenciação, uma vez que a utilização do mesmo grau de diferenciação para todos os processos univariados, o que corresponderia à adopção de um operador diferença escalar, acarretaria frequentemente o referido problema.

Em resumo, é importante tomar em consideração que a aplicação do operador diferença, apesar de poder ser adequada para a estacionarização de processos univariados não estacionários, pode não ser necessária ou aconselhável quando se considera diversos processos conjuntamente, ou seja, pode não ser apropriada para a estacionarização de um processo multivariado não estacionário.

Tendo em conta estas dificuldades, Tjøstheim e Paulsen (1982) distinguem o caso em que $|\varphi(B)|$ possui zeros que se situam exactamente sobre o círculo unitário, daquele em que esses zeros estão apenas próximos deste círculo. No primeiro caso, $\{X_t\}$ é designado de não estacionário e, no segundo, de quase não estacionário. Quando $\{X_t\}$ é quase não estacionário, a sua diferenciação não é adequada para o converter num processo estacionário, devendo então ser considerada uma transformação do tipo (Tjøstheim e Paulsen, 1982):

$$W_t = X_t - \alpha X_{t-1} \quad (2.1.71)$$

onde a matriz α , de dimensão $(m \times m)$, deverá ser escolhida de modo a que o determinante do operador auto-regressivo de $\{W_t\}$ não tenha zeros próximos do círculo unitário.

Assim, considere-se um processo $\{X_t\}$ para o qual $|\varphi(B)|$ possui r zeros próximos do círculo unitário ($r \leq m$), o que é designado por quase não estacionaridade de grau r , sendo $\{W_t\}$ em (2.1.71) estacionário. Isto implica $\varphi(B)$ possa ser decomposto da forma:

$$\varphi(B) = \phi(B) (I_m - \alpha B) \quad (2.1.72)$$

onde nenhum dos zeros de $|\phi(B)|$ está próximo do círculo unitário. Decomposições do tipo de (2.1.72) são estudadas por Stensholt e Tjøstheim (1981), que mostram que, se os zeros de $|\varphi(B)|$ forem todos distintos, será sempre possível decompor $\varphi(B)$ numa expressão que

inclua o factor $(I_m - \alpha B)$. No entanto, mesmo se $r \leq m$, pode não ser possível determinar uma matriz α tal que nenhum dos zeros de $|\phi(B)|$ esteja próximo do círculo unitário. Conclui-se então que, mesmo se $r \leq m$, pode ser necessário repetir a transformação (2.1.71) por diversas vezes, de modo a obter um processo isento de quase não estacionaridade (Tjøstheim e Paulsen, 1982). Estes autores referem também que a diferenciação de cada processo univariado pode não ser adequada para a estacionarização de um processo multivariado. Por outro lado, a forma de determinação da matriz α adequada à transformação (2.1.71) será analisada mais à frente.

III. ESTIMAÇÃO

Após a identificação de um modelo para o processo estocástico subjacente à série cronológica em análise, torna-se necessário propor valores adequados para os parâmetros desse modelo, ou seja, estimá-los. Com efeito, uma vez que esses parâmetros são desconhecidos, é conveniente inferir sobre eles, procedendo-se assim ao ajustamento do modelo previamente identificado às observações que constituem a série cronológica.

No entanto, a estimação dos parâmetros de um modelo não é suficiente, sendo necessário também estimar o valor esperado de $\{X_t\}$ e as respectivas matrizes de covariância e de correlação, uma vez que são instrumentos imprescindíveis na análise de qualquer série cronológica, nomeadamente na fase de identificação e de confirmação do diagnóstico.

Por outro lado, paralelamente à estimação dos parâmetros das matrizes que compõem o operador auto-regressivo e o de médias móveis, é útil estimar também as ordens desses operadores, uma vez que estas são grandezas desconhecidas. De qualquer modo, a estimação dos parâmetros é tratada em primeiro lugar, admitindo-se para o efeito que as ordens dos referidos operadores são conhecidas, sendo a estimação destas analisada em seguida.

Assim, é útil considerar também a distribuição dos estimadores propostos quer para os parâmetros, quer para as funções de covariância e de correlação, dada a importância de que essa distribuição se reveste, nomeadamente para fins inferenciais e, em particular, na fase de confirmação do diagnóstico.

Finalmente, e uma vez que do ajustamento de um modelo a uma série cronológica resultam sempre os correspondentes resíduos de estimação, é ainda necessário analisar a distribuição dos estimadores da

função de covariância e de correlação residuais, uma vez que ambos desempenham um papel importante na fase de crítica do modelo.

Um aspecto que importa focar é o de que os métodos de estimação vão ser analisados apenas para modelos não sazonais, uma vez que a sua generalização ao caso em que se verifica a existência de sazonalidade não oferece dificuldade. Esta metodologia será adoptada ao longo da análise de todas as fases de modelização de uma série cronológica multivariada.

Refira-se ainda que se irá tratar apenas a estimação no domínio tempo.

1. Estimação do valor esperado, da matriz de covariância, da matriz de correlação e estimação preliminar dos parâmetros

A análise de qualquer série cronológica necessita da prévia estimação das matrizes de covariância e de correlação do processo estocástico que a gerou, o que obriga também a estimar o respectivo valor esperado. Assim, paralelamente à análise da estimação dessas grandezas, refere-se também as propriedades mais importantes dos estimadores resultantes dos procedimentos propostos, bem como alguns dos problemas com eles relacionados.

Finalmente, passa-se ao exame da estimação preliminar dos parâmetros dos modelos auto-regressivos e dos de médias móveis, propondo-se assim uma forma mais simples de proceder a essa estimação que faz recurso aos outros estimadores agora referidos e que pretende fornecer valores iniciais para a posterior estimação pelo método da máxima verosimilhança.

1.1. Estimação do valor esperado

Admitindo que $E(X_i) \neq 0$ e dispondo de N observações X_1, X_2, \dots, X_N que constituem a série cronológica, um estimador centrado no valor esperado do processo $\{X_i\}$ é o vector da média da série multivariada, ou seja, é um vector composto pelas médias das diferentes séries univariadas, assumindo a forma:

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i \quad (3.1.1)$$

Refira-se a propósito que o valor esperado de cada processo μ_i ($i=1, \dots, m$) é estimado pela média de cada série univariada:

$$\bar{X}_i = \frac{1}{N} \sum_{i=1}^N X_{it} \quad (3.1.2)$$

A partir da imposição de algumas hipóteses sobre $\gamma_{ii}(k)$, a consistência de \bar{X} resulta facilmente da consistência de cada uma das suas componentes, o que constitui um resultado estabelecido para o caso univariado, sendo assim possível estabelecer a seguinte proposição (Brockwell e Davis, 1987, pág.396):

Proposição 3.1.1. seja $\{X_t\}$ um processo estocástico multivariado estacionário com valor esperado μ e matriz de covariância $\Gamma(k)$. Então, quando $N \rightarrow \infty$, tem-se:

$$E(\bar{X} - \mu)^T (\bar{X} - \mu) \rightarrow 0 \quad \text{se} \quad \gamma_{ii}(N) \rightarrow 0 \quad i = 1, \dots, m$$

$$NE(\bar{X} - \mu)^T (\bar{X} - \mu) \rightarrow \sum_{i=1}^m \sum_{k=-\infty}^{\infty} \gamma_{ii}(k) \quad \text{se} \quad \sum_{k=-\infty}^{\infty} |\gamma_{ii}(k)| < \infty \quad i = 1, \dots, m$$

No caso particular de $\{X_t\}$ ser um processo de médias móveis multivariado e sob hipóteses mais restritivas, é possível ainda obter um outro resultado (Brockwell e Davis, 1987, pág.396):

Proposição 3.1.2. Seja $\{X_t\}$ um processo estocástico estacionário tal que:

$$X_t = \mu + \sum_{i=-\infty}^{\infty} H_i \varepsilon_{t-i}$$

onde $\{\varepsilon_t\}$ é um processo puramente aleatório multivariado com matriz de variâncias e covariâncias Σ e $\{H_i = [H_i(i,j)], i,j=1,\dots,m\}$ é uma sequência de matrizes de dimensão $(m \times m)$ tais que:

$$\sum_{i=-\infty}^{\infty} |H_i(i,j)| < \infty \quad (i,j=1,\dots,m).$$

Então, \bar{X} segue distribuição assintótica Normal:

$$\bar{X} \sim N(\mu; N^{-1} (\sum_{i=-\infty}^{\infty} H_i) \Sigma (\sum_{i=-\infty}^{\infty} H_i^T))$$

1.2. Estimação da matriz de covariância, da matriz de correlação e da matriz de correlação parcial

Uma vez que a matriz de covariância assume a expressão:

$$\Gamma(k) = E[(X_t - \mu)(X_{t-k} - \mu)^T] \quad (3.1.3)$$

propõe-se como seu estimador:

$$C(k) = \begin{cases} \frac{1}{N} \sum_{t=k+1}^N (X_t - \bar{X})(X_{t-k} - \bar{X})^T & k = 0, 1, \dots, N-1 \\ \frac{1}{N} \sum_{t=1}^{N+k} (X_t - \bar{X})(X_{t-k} - \bar{X})^T & k = -(N-1), \dots, -2, -1 \end{cases} \quad (3.1.4)$$

O (i,j)-ésimo elemento desta matriz assume a forma:

$$c_{ij}(k) = \begin{cases} \frac{1}{N} \sum_{t=k+1}^N (X_{it} - \bar{X}_i)(X_{j,t-k} - \bar{X}_j)^T & k = 0, 1, \dots, N-1 \\ \frac{1}{N} \sum_{t=1}^{N+k} (X_{it} - \bar{X}_i)(X_{j,t-k} - \bar{X}_j)^T & k = -(N-1), \dots, -2, -1 \end{cases} \quad (3.1.5)$$

ou seja $c_{ii}(k)$ é o estimador da função de autocovariância do processo $\{X_{it}\}$ e, para $i \neq j$, $c_{ij}(k)$ é o estimador da função de covariância cruzada entre $\{X_{it}\}$ e $\{X_{jt}\}$, com $i, j = 1, \dots, m$.

Por sua vez, a matriz de correlação assume a expressão:

$$\rho(k) = \Gamma_0^{-1/2} \Gamma(k) \Gamma_0^{-1/2} \quad (3.1.6)$$

o que leva a que o estimador proposto para $\rho(k)$ seja:

$$R(k) = C_0^{-1/2} C(k) C_0^{-1/2} \quad (3.1.7)$$

onde $C_0 = \text{diag} \{c_{11}(0), \dots, c_{mm}(0)\}$, sendo o elemento genérico da matriz $R(k)$:

$$r_{ij}(k) = \frac{c_{ij}(k)}{[c_{ii}(0)c_{jj}(0)]^{1/2}} \quad (3.1.8)$$

ou seja, $r_{ii}(k)$ é o estimador da função de autocorrelação do processo $\{X_{it}\}$ e, para $i \neq j$, $r_{ij}(k)$ é o estimador da função de correlação cruzada entre $\{X_{it}\}$ e $\{X_{jt}\}$, com $i, j = 1, \dots, m$.

Os estimadores acabados de apresentar possuem uma importante propriedade, aplicável ao caso de processos de médias móveis de ordem infinita e referida no seguinte teorema (Brockwell e Davis, 1987, pág.398):

Teorema 3.1.1. seja $\{X_t\}$ um processo estocástico multivariado que assume a forma:

$$X_t = \sum_{i=-\infty}^{\infty} H_i \varepsilon_{t-i}$$

onde os vectores ε_t são independentes e identicamente distribuídos com valor esperado igual ao vector nulo e matriz de variâncias e covariâncias finita Σ , sendo $\{H_t = [H_t(i, j)]\}$ uma sequência de matrizes com :

$$\sum_{i=-\infty}^{\infty} |H_t(i, j)| < \infty \quad (i, j = 1, \dots, m).$$

Então, quando $N \rightarrow \infty$, verificam-se as seguintes convergências em probabilidade:

$$\text{plim } c_{ij}(k) = \gamma_{ij}(k)$$

$$\text{plim } r_{ij}(k) = \rho_{ij}(k) \quad i, j = 1, \dots, m$$

para $k \geq 0$. Assim, sendo o significado da convergência em probabilidade de uma matriz estocástica o da convergência em probabilidade de todos os

seus elementos, é possível exprimir este resultado de uma forma alternativa:

$$\text{plim } C(k) = \Gamma(k)$$

$$\text{plim } R(k) = \rho(k)$$

para $k \geq 0$.

A propósito das propriedades destes estimadores, é útil referir ainda que ambos são assintoticamente centrados.

Se, além das condições referidas no teorema 3.1.1., os cumulantes de quarta ordem das componentes do vector ε_t forem finitos, Hannan (1970, pág.210) mostra que se verifica a convergência em média quadrática de $c_{ij}(k)$.

Por outro lado, é possível estimar a matriz de correlação parcial através da utilização do método dos mínimos quadrados multivariado a partir do ajustamento de modelos auto-regressivos de ordem $K=1,2,\dots$, sendo também possível estimar os desvios-padrão dos estimadores assim obtidos.

Para o efeito, considere-se o modelo AR(k):

$$X_t = \phi_1 X_{t-1} + \dots + \phi_k X_{t-k} + \varepsilon_t \quad (3.1.9)$$

Transpondo esta expressão, obtém-se o resultado:

$$X_t^\tau = X_{t-1}^\tau \phi_1^\tau + \dots + X_{t-k}^\tau \phi_k^\tau + \varepsilon_t^\tau \quad (3.1.10)$$

No entanto, para estimar um modelo auto-regressivo de ordem k a partir das N observações X_1, \dots, X_N que compõem a série cronológica, é necessário desprezar as k primeiras observações, uma vez que X_t depende de X_{t-1}, \dots, X_{t-k} , conforme é visível em (3.1.9), sendo esta questão tratada detalhadamente mais à frente.

Assim, considere-se as matrizes:



$$\begin{aligned}
\mathbf{Y} &= \begin{bmatrix} \mathbf{X}_{k+1}^\top \\ \vdots \\ \mathbf{X}_N^\top \end{bmatrix} & ; & \quad \mathbf{Z} = \begin{bmatrix} \mathbf{X}_k^\top & \dots & \mathbf{X}_1^\top \\ \vdots & & \vdots \\ \mathbf{X}_{N-1}^\top & \dots & \mathbf{X}_{N-k}^\top \end{bmatrix} \\
\boldsymbol{\lambda} &= \begin{bmatrix} \phi_1^\top \\ \vdots \\ \phi_k^\top \end{bmatrix} & ; & \quad \boldsymbol{\eta} = \begin{bmatrix} \boldsymbol{\varepsilon}_{k+1}^\top \\ \vdots \\ \boldsymbol{\varepsilon}_N^\top \end{bmatrix}
\end{aligned} \tag{3.1.11}$$

Deste modo, considerando a expressão (3.1.10) para $t=k+1, \dots, N$, obtém-se o modelo linear multivariado:

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\lambda} + \boldsymbol{\eta} \tag{3.1.12}$$

Nestas condições, os estimadores de mínimos quadrados de $\boldsymbol{\lambda}$ podem ser facilmente determinados (Tiao e Box; Tiao e Tsay, 1983). Adoptando este procedimento para $k=1, 2, \dots$, obtém-se o estimador da matriz de correlação parcial para o "lag" k , designado por $\hat{\mathbf{P}}(k)$ e definido da forma:

$$\hat{\mathbf{P}}(k) = \hat{\phi}_k \quad k = 1, 2, \dots \tag{3.1.13}$$

tendo sido ajustado um modelo AR(k). É também possível determinar desta forma os estimadores dos desvios-padrão dos elementos de $\hat{\mathbf{P}}(k)$ ⁽³⁾.

1.3. Estimação preliminar dos parâmetros

Uma vez que os parâmetros de qualquer dos modelos analisados são desconhecidos, é útil propor valores adequados para eles. Assim, utilizando fundamentalmente os estimadores acabados de analisar, é possível a sua imediata estimação. No entanto, tal não constitui mais do que uma estimação preliminar, destinada a fornecer valores iniciais que podem vir a ser posteriormente utilizados num processo de estimação mais adequado, que será analisado em seguida. Contudo, aqui apenas será

tratada a estimação preliminar dos parâmetros de modelos auto-regressivos e de modelos de médias móveis.

1.3.1. Estimação preliminar dos parâmetros de modelos auto-regressivos

O modelo auto-regressivo de ordem p assume a expressão:

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \epsilon_t \quad (3.1.14)$$

Utilizando as equações de Yule-Walker, definidas atrás em (2.1.8) e substituindo $\Gamma(k)$ pelo seu estimador $C(k)$, é possível obter estimadores iniciais para através da expressão:

$$C(k) = \phi_1 C(k-1) + \dots + \phi_p C(k-p) \quad k=1, \dots, p \quad (3.1.15)$$

Consideradas no seu conjunto, estas expressões permitem obter os referidos estimadores através da resolução do sistema:

$$\begin{bmatrix} \tilde{\phi}_1^T \\ \vdots \\ \tilde{\phi}_p^T \end{bmatrix} = \begin{bmatrix} C(0) & \dots & C(p-1) \\ \vdots & & \vdots \\ C(1-p) & \dots & C(0) \end{bmatrix} \begin{bmatrix} C^T(1) \\ \vdots \\ C^T(p) \end{bmatrix} \quad (3.1.16)$$

onde $\tilde{\phi}_1, \dots, \tilde{\phi}_p$ representam os estimadores preliminares. Jenkins e Alavi (1981) referem que os estimadores determinados desta forma constituem aproximações muito boas aos estimadores que serão determinados mais à frente pelo método da máxima verosimilhança.

1.3.2. Estimação preliminar dos parâmetros de modelos de médias móveis

O modelo de médias móveis de ordem q pode ser representado da forma:

$$X_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} \quad (3.1.17)$$

É possível propor estimadores iniciais para $\theta_1, \dots, \theta_q$ através da utilização da expressão da função de covariância deste tipo de modelo, substituindo $\Gamma(k)$ pelo seu estimador $C(k)$, o que constitui um método linearmente convergente (Jenkins e Alavi, 1981). Este procedimento conduz às expressões:

$$\begin{aligned}\tilde{\Sigma} &= C(0) - \tilde{\theta}_1 \tilde{\Sigma} \tilde{\theta}_1^T - \dots - \tilde{\theta}_q \tilde{\Sigma} \tilde{\theta}_q^T \\ \tilde{\theta}_k &= C(k) \tilde{\Sigma}^{-1} - \tilde{\theta}_{k+1} \tilde{\Sigma} \tilde{\theta}_1^T \tilde{\Sigma}^{-1} - \dots - \tilde{\theta}_q \tilde{\Sigma} \tilde{\theta}_{q-k}^T \tilde{\Sigma}^{-1} \quad k=1, \dots, q\end{aligned}\quad (3.1.18)$$

Estas equações podem ser resolvidas iterativamente de modo a obter $\tilde{\Sigma}, \tilde{\theta}_q, \tilde{\theta}_{q-1}, \dots, \tilde{\theta}_1$ por esta ordem, utilizando as estimativas mais recentes em cada iteração e tomando $\tilde{\theta}_i = 0$ inicialmente.

A título de exemplo, considere-se o caso $q=2$, em que as equações (3.1.18) assumem a forma:

$$\begin{aligned}\tilde{\Sigma} &= C(0) - \tilde{\theta}_1 \tilde{\Sigma} \tilde{\theta}_1^T - \tilde{\theta}_2 \tilde{\Sigma} \tilde{\theta}_2^T \\ \tilde{\theta}_2 &= C(2) \tilde{\Sigma}^{-1} \\ \tilde{\theta}_1 &= C(1) \tilde{\Sigma}^{-1} - \tilde{\theta}_2 \tilde{\Sigma} \tilde{\theta}_1^T \tilde{\Sigma}^{-1}\end{aligned}\quad (3.1.19)$$

No entanto, Jenkins e Alavi (1981) referem que as propriedades de convergência deste método não são satisfatórias, sobretudo se $q > 1$.

2. Estimação dos parâmetros

Iremos agora tratar o problema da estimação dos parâmetros de um dado modelo, apresentando os métodos utilizados e as principais dificuldades que se colocam para cada um dos tipos de modelos, concluindo-se que, em todos os casos, os procedimentos propostos constituem generalizações do que se passa no caso univariado. De qualquer modo, surgem complicações adicionais decorrentes do facto de se estar a considerar vectores aleatórios m -dimensionais.

Assim, admitir-se-à que o valor esperado de \mathbf{X}_t é o vector nulo e que os vectores $\boldsymbol{\varepsilon}_t$ são independentes e identicamente distribuídos seguindo distribuição $N(\mathbf{0}; \boldsymbol{\Sigma})$, o que possibilita o recurso ao método da máxima verosimilhança.

2.1. Estimação dos parâmetros de modelos auto-regressivos

Considere-se o modelo auto-regressivo estacionário de ordem p :

$$(\mathbf{I}_m - \phi_1 B - \dots - \phi_p B^p) \mathbf{X}_t = \boldsymbol{\varepsilon}_t \quad (3.2.1)$$

Tendo em vista obter a função de verosimilhança, iremos introduzir as seguintes notações e definições:

$$\mathbf{X}^T = (\mathbf{X}_1^T, \dots, \mathbf{X}_N^T) ;$$

$\mathbf{X}_{(0)}$ representa o vector dos valores desconhecidos ($\mathbf{X}_{1-p}, \dots, \mathbf{X}_0$):

$$\mathbf{X}_{(0)}^T = (\mathbf{X}_{1-p}^T, \dots, \mathbf{X}_0^T) ;$$

$\boldsymbol{\varepsilon}$ representa o vector das variáveis residuais:

$$\boldsymbol{\varepsilon}^T = (\boldsymbol{\varepsilon}_1^T, \dots, \boldsymbol{\varepsilon}_N^T) ;$$

D é a matriz de dimensão ($Nm \times Nm$):

$$D = \begin{bmatrix} \mathbf{I}_m & & & & & \mathbf{0} \\ -\phi_1 & & & & & \\ \vdots & & & & & \\ -\phi_p & & & & & \\ & \ddots & & & & \\ \mathbf{0} & & -\phi_p & \dots & -\phi_1 & \mathbf{I}_m \end{bmatrix};$$

E é a matriz de dimensão (Nm x pm):

$$E = \begin{bmatrix} \mathbf{I}_{pm} \\ \mathbf{0} \end{bmatrix};$$

M é a matriz de dimensão (pm x pm):

$$M = \begin{bmatrix} \phi_p & \phi_{p-1} & \dots & \phi_1 \\ & \ddots & & \vdots \\ & & \ddots & \phi_{p-1} \\ \mathbf{0} & & & \phi_p \end{bmatrix};$$

Neste contexto, vai-se considerar o modelo (3.2.1) para o conjunto das observações \mathbf{X} na forma (Hillmer e Tiao, 1979):

$$D\mathbf{X} = \boldsymbol{\varepsilon} + E\mathbf{M}\mathbf{X}_{(0)} \quad (3.2.2)$$

Por outro lado, é útil definir também $\mathbf{X}^T = (\mathbf{X}_{(1)}^T, \mathbf{X}_{(2)}^T)$, em que $\mathbf{X}_{(1)}$ é composto pelos p primeiros elementos de \mathbf{X} , os quais dependem dos elementos desconhecidos que constituem $\mathbf{X}_{(0)}$.

Além disso, sendo $\mathbf{W} = D\mathbf{X}$, define-se ainda $\mathbf{W}^T = (\mathbf{W}_{(1)}^T, \mathbf{W}_{(2)}^T)$, em que $\mathbf{W}_{(1)}$ é composto pelos p primeiros vectores de \mathbf{W} .

Deste modo, a função densidade de probabilidade de \mathbf{X} pode ser expressa da forma:

$$p(\mathbf{X}) = p(\mathbf{X}_{(1)})p(\mathbf{X}_{(2)} | \mathbf{X}_{(1)}) \quad (3.2.3)$$

Uma vez que os vectores \mathbf{E}_t seguem distribuição $N(\mathbf{0}; \Sigma)$ e que o Jacobiano da transformação de \mathbf{E} para \mathbf{X} é igual à unidade, a função de verosimilhança para os parâmetros (ϕ, Σ) , sendo $\phi = (\phi_1, \dots, \phi_p)$, é (Hillmer e Tiao, 1979):

$$l(\phi, \Sigma | \mathbf{X}) = l_1(\phi, \Sigma | \mathbf{X}) l_2(\phi, \Sigma | \mathbf{X}) \quad (3.2.4)$$

em que l_1 é a função de verosimilhança correspondente à função densidade marginal de $\mathbf{X}_{(1)}$, assumindo a expressão:

$$l_1(\phi, \Sigma | \mathbf{X}) = (2\pi)^{-\frac{mp}{2}} |V + (\mathbf{I}_p \otimes \Sigma)|^{-1/2} \exp\left\{-\frac{1}{2} \mathbf{W}_{(1)}^T [V + (\mathbf{I}_p \otimes \Sigma)]^{-1} \mathbf{W}_{(1)}\right\} \quad (3.2.5)$$

onde V é a matriz de covariância de $M\mathbf{X}_{(0)}$, sendo a obtenção de expressões explícitas para esta matriz muito complicada (Hillmer e Tiao, 1979). Por sua vez, l_2 é a função de verosimilhança associada à função densidade condicionada de $\mathbf{X}_{(2)}$ dado $\mathbf{X}_{(1)}$, sendo a sua expressão:

$$l_2(\phi, \Sigma | \mathbf{X}) = (2\pi)^{-\frac{(N-p)m}{2}} |\Sigma|^{-\frac{N-p}{2}} \exp\left\{-\frac{1}{2} \mathbf{W}_{(2)}^T (\mathbf{I}_{N-p} \otimes \Sigma)^{-1} \mathbf{W}_{(2)}\right\} \quad (3.2.6)$$

O problema que se coloca para a maximização de (3.2.4) é o do desconhecimento dos valores de $\mathbf{X}_{1-p}, \dots, \mathbf{X}_0$.

Contudo, Hillmer e Tiao (1979) mostram que, à medida que N aumenta, l_2 tende a dominar l_1 , pelo que, para um valor elevado de N , a contribuição de l_1 pode ser ignorada e a função de verosimilhança convenientemente aproximada por l_2 . Deste modo, evita-se o problema do desconhecimento de $(\mathbf{X}_{1-p}, \dots, \mathbf{X}_0)$ utilizando apenas a função densidade condicionada $p(\mathbf{X}_{(2)} | \mathbf{X}_{(1)})$, o que equivale a desprezar as p primeiras observações $(\mathbf{X}_{1-p}, \dots, \mathbf{X}_0)$, ou seja, a considerar \mathbf{X}_{p+1} como o primeiro valor da série.

Assim, (Hillmer e Tiao, 1979), os parâmetros ϕ_1, \dots, ϕ_p podem ser estimados pelo método dos mínimos quadrados e, como é conhecido, estes estimadores possuem as propriedades assintóticas convenientes.

Por outro lado, Σ pode ser estimada utilizando os resíduos de estimação obtidos da forma:

$$\hat{\varepsilon}_t = X_t - \hat{\phi}_1 X_{t-1} - \dots - \hat{\phi}_p X_{t-p} \quad t = p+1, \dots, N \quad (3.2.7)$$

onde $\hat{\phi}_1, \dots, \hat{\phi}_p$ designam os estimadores dos parâmetros e $\hat{\varepsilon}_t$ são os resíduos de estimação, a partir dos quais $\hat{\Sigma}$ é facilmente determinada:

$$\hat{\Sigma} = \frac{1}{N} \sum_{t=p+1}^N \hat{\varepsilon}_t \hat{\varepsilon}_t^T \quad (3.2.8)$$

2.2. Estimação dos parâmetros de modelos de médias móveis

Seja o modelo de médias móveis de ordem q :

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (3.2.9)$$

Neste tipo de modelos, a função de verosimilhança é obtida generalizando a abordagem de Box e Jenkins (1976) para o caso univariado (Osborn, 1977).

Assim, seja ε o vector das variáveis residuais, cuja dimensão é $[m(N+q) \times 1]$:

$$\varepsilon^T = (\varepsilon_{1-q}^T, \dots, \varepsilon_0^T, \varepsilon_1^T, \dots, \varepsilon_N^T) \quad (3.2.10)$$

Considere-se também o vector das q primeiras variáveis residuais de 3.2.10) e o vector das observações:

$$\begin{aligned} \varepsilon'^T &= (\varepsilon_{1-q}^T, \dots, \varepsilon_0^T) \\ X^T &= (X_1^T, \dots, X_N^T) \end{aligned} \quad (3.2.11)$$

O vector $\boldsymbol{\varepsilon}$ segue distribuição Normal de valor esperado $\mathbf{0}$ e matriz de variâncias e covariâncias $\boldsymbol{\xi} = (\mathbf{I}_{[(N+q) \times (N+q)]} \otimes \boldsymbol{\Sigma})$, pelo que a sua função densidade de probabilidade é:

$$p(\boldsymbol{\varepsilon}) = (2\pi)^{-\frac{m(N+q)}{2}} |\boldsymbol{\xi}|^{-1/2} \exp\left\{-\frac{1}{2} \boldsymbol{\varepsilon}^T \boldsymbol{\xi}^{-1} \boldsymbol{\varepsilon}\right\} \quad (3.2.12)$$

Por outro lado, as equações resumidas em (3.2.9) podem ser expressas da forma:

$$\begin{aligned} \boldsymbol{\varepsilon}_{1-q} &= \boldsymbol{\varepsilon}_{1-q} \\ \vdots & \quad \quad \quad \vdots \\ \boldsymbol{\varepsilon}_0 &= \boldsymbol{\varepsilon}_0 \\ \boldsymbol{\varepsilon}_1 &= \mathbf{X}_1 - \theta_1 \boldsymbol{\varepsilon}_0 - \dots - \theta_q \boldsymbol{\varepsilon}_{1-q} \\ \vdots & \quad \quad \quad \vdots \\ \boldsymbol{\varepsilon}_N &= \mathbf{X}_N - \theta_1 \boldsymbol{\varepsilon}_{N-1} - \dots - \theta_q \boldsymbol{\varepsilon}_{N-q} \end{aligned} \quad (3.2.13)$$

Por substituição sucessiva dos vectores $\boldsymbol{\varepsilon}_t$, $t \geq 1$, no segundo membro de (3.2.13), cada vector $\boldsymbol{\varepsilon}_t$ pode ser expresso como função das variáveis residuais $\boldsymbol{\varepsilon}_{1-q}, \dots, \boldsymbol{\varepsilon}_0$ e o das observações $\mathbf{X}_1, \dots, \mathbf{X}_N$.

Assim, utilizando (3.2.10), o sistema (3.2.13) pode ser expresso da forma:

$$\boldsymbol{\varepsilon} = \mathbf{LX} + \mathbf{T}\boldsymbol{\varepsilon}' \quad (3.2.14)$$

onde as matrizes \mathbf{L} e \mathbf{T} têm dimensão $[m(N+q) \times mN]$ e $[m(N+q) \times mq]$ respectivamente e são determinadas apenas por $\theta_1, \dots, \theta_q$.

A equação (3.2.14) define uma transformação de $\boldsymbol{\varepsilon}$ para \mathbf{X} e $\boldsymbol{\varepsilon}'$ e, uma vez que o respectivo Jacobiano é igual à unidade, a partir de (3.2.12) obtém-se a expressão:

$$p(\mathbf{X}, \boldsymbol{\varepsilon}') = (2\pi)^{-\frac{m(N+q)}{2}} |\boldsymbol{\xi}|^{-1/2} \exp\left\{-\frac{1}{2} (\mathbf{LX} + \mathbf{T}\boldsymbol{\varepsilon}')^T \boldsymbol{\xi}^{-1} (\mathbf{LX} + \mathbf{T}\boldsymbol{\varepsilon}')\right\} \quad (3.2.15)$$

O vector $\boldsymbol{\varepsilon}'$ é desconhecido, podendo ser estimado pelo método dos mínimos quadrados generalizados, do que resultam os estimadores:

$$\begin{aligned}\tilde{\boldsymbol{\varepsilon}}' &= -(\mathbf{T}^T \boldsymbol{\xi}^{-1} \mathbf{T})^{-1} \mathbf{T}^T \boldsymbol{\xi}^{-1} \mathbf{LX} \\ \tilde{\boldsymbol{\varepsilon}} &= \mathbf{LX} + \mathbf{T} \tilde{\boldsymbol{\varepsilon}}'\end{aligned}\quad (3.2.16)$$

Utilizando as propriedades deste estimador, a forma quadrática $\boldsymbol{\varepsilon}'^T \boldsymbol{\xi}^{-1} \boldsymbol{\varepsilon}$ pode ser decomposta da forma:

$$\boldsymbol{\varepsilon}'^T \boldsymbol{\xi}^{-1} \boldsymbol{\varepsilon} = (\mathbf{LX} + \mathbf{T} \tilde{\boldsymbol{\varepsilon}}')^T \boldsymbol{\xi}^{-1} (\mathbf{LX} + \mathbf{T} \tilde{\boldsymbol{\varepsilon}}') + (\boldsymbol{\varepsilon}' - \tilde{\boldsymbol{\varepsilon}}')^T \mathbf{T}^T \boldsymbol{\xi}^{-1} \mathbf{T} (\boldsymbol{\varepsilon}' - \tilde{\boldsymbol{\varepsilon}}') \quad (3.2.17)$$

onde a primeira forma quadrática do segundo membro é função apenas das observações \mathbf{X} e não das variáveis residuais "de partida" que compõem o vector $\boldsymbol{\varepsilon}'$. Deste modo, a partir da relação:

$$p(\mathbf{X}, \boldsymbol{\varepsilon}') = p(\mathbf{X}) p(\boldsymbol{\varepsilon}' | \mathbf{X}) \quad (3.2.18)$$

e utilizando (3.2.15) e (3.2.17), obtém-se expressões:

$$p(\boldsymbol{\varepsilon}' | \mathbf{X}) = (2\pi)^{-\frac{m}{2}} |\mathbf{T}^T \boldsymbol{\xi}^{-1} \mathbf{T}|^{1/2} \exp\left\{-\frac{1}{2} (\boldsymbol{\varepsilon}' - \tilde{\boldsymbol{\varepsilon}}')^T \mathbf{T}^T \boldsymbol{\xi}^{-1} \mathbf{T} (\boldsymbol{\varepsilon}' - \tilde{\boldsymbol{\varepsilon}}')\right\} \quad (3.2.19)$$

$$\begin{aligned}p(\mathbf{X}) &= (2\pi)^{-\frac{m}{2}} |\boldsymbol{\xi}|^{-1/2} |\mathbf{T}^T \boldsymbol{\xi}^{-1} \mathbf{T}|^{-1/2} \exp\left\{-\frac{1}{2} (\mathbf{LX} + \mathbf{T} \tilde{\boldsymbol{\varepsilon}}')^T \boldsymbol{\xi}^{-1} (\mathbf{LX} + \mathbf{T} \tilde{\boldsymbol{\varepsilon}}')\right\} = \\ &= (2\pi)^{-\frac{m}{2}} |\boldsymbol{\xi}|^{-1/2} |\mathbf{T}^T \boldsymbol{\xi}^{-1} \mathbf{T}|^{-1/2} \exp\left\{-\frac{1}{2} \tilde{\boldsymbol{\varepsilon}}'^T \boldsymbol{\xi}^{-1} \tilde{\boldsymbol{\varepsilon}}'\right\} = \\ &= (2\pi)^{-\frac{m}{2}} |\boldsymbol{\xi}|^{-1/2} |\mathbf{T}^T \boldsymbol{\xi}^{-1} \mathbf{T}|^{-1/2} \exp\left\{-\frac{1}{2} \sum_{i=1}^N \tilde{\boldsymbol{\varepsilon}}_i^T \boldsymbol{\Sigma}^{-1} \tilde{\boldsymbol{\varepsilon}}_i\right\}\end{aligned}\quad (3.2.20)$$

A equação (3.2.20) é a função de verosimilhança exacta (Osborn, 1977).

Assim, os estimadores de máxima verosimilhança exactos de $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)$ e de $\boldsymbol{\Sigma}$ podem ser obtidos através da minimização da função:

$$L_1(\theta, \Sigma | \mathbf{X}) = (N + q) \log |\Sigma| + \log |\mathbf{T}^T \xi^{-1} \mathbf{T}| + \sum_{i=1}^N \tilde{\mathbf{E}}_i^T \Sigma^{-1} \tilde{\mathbf{E}}_i \quad (3.2.21)$$

No entanto, uma vez que as equações resultantes de igualar as derivadas de L_1 a zero são não lineares e extremamente difíceis de resolver, é necessário utilizar um método de otimização não linear apropriado.

Mas, como a maximização da função de verosimilhança exacta, ou seja, a minimização da função L_1 , pode ser extremamente custosa, é possível, desde que o modelo (3.2.9) seja invertível, adoptar um procedimento mais simples que consiste em maximizar a função de verosimilhança aproximada, do que resultam estimadores de máxima verosimilhança aproximados. Esta função resolve o problema do desconhecimento das componentes de \mathbf{E}' fixando-os em zero.

Deste modo, a função de verosimilhança aproximada é (Wilson, 1973; Osborn, 1977):

$$l_2(\theta, \Sigma | \mathbf{X}) = (2\pi)^{-\frac{N}{2}} |\Sigma|^{-\frac{N}{2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^N \mathbf{E}_i^T \Sigma^{-1} \mathbf{E}_i \right\} \quad (3.2.22)$$

Então, os estimadores são obtidos a partir da minimização da função:

$$L_2(\theta, \Sigma | \mathbf{X}) = \log |\Sigma| + \frac{1}{N} \sum_{i=1}^N \mathbf{E}_i^T \Sigma^{-1} \mathbf{E}_i \quad (3.2.23)$$

Uma vez que $\mathbf{E}_{1-q}, \dots, \mathbf{E}_0$ são tomados como valores fixos, pois são igualados a zero, l_2 é uma função de verosimilhança condicionada, sendo a sua maximização, ou seja, a minimização de L_2 , conseguida também através da utilização de um método de otimização não linear conveniente.

Comparando (3.2.22) com (3.2.20), comprova-se que a primeira é uma aproximação à segunda. Os dois níveis de aproximação resultam da fixação em zero das variáveis residuais "de partida" $\mathbf{E}_{1-q}, \dots, \mathbf{E}_0$ e da utilização de $|\Sigma|^{-N/2}$ em vez da expressão exacta:

$$|\xi|^{-1/2} |\mathbf{T}^T \xi^{-1} \mathbf{T}|^{-1/2} = |\Sigma|^{-\frac{N+q}{2}} |\mathbf{T}^T (\mathbf{I}_{[(N+q) \times (N+q)]} \otimes \Sigma^{-1}) \mathbf{T}|^{-1/2}$$

O efeito sobre \mathcal{E} da fixação em zero das variáveis residuais "de partida" em (3.2.14) é determinado pela matriz \mathbf{T} , da qual, cada coluna de blocos⁽⁴⁾ segue a relação recursiva (Osborn, 1977):

$$\begin{aligned} \mathbf{K}_{t+q,s} &= -\theta_1 \mathbf{T}_{t+q-1,s} - \theta_2 \mathbf{T}_{t+q-2,s} - \dots - \theta_q \mathbf{T}_{t,s} & t &= 1, \dots, N \\ & & s &= 1, \dots, q \end{aligned} \quad (3.2.24)$$

cujos valores de partida, dados por (3.2.13) são:

$$\mathbf{T}_{rs} = \delta_{rs} \mathbf{I}_{[(N+q) \times (N+q)]} \quad (3.2.25)$$

onde δ_{rs} representa o delta de Kronecker.

Assim (Osborn, 1977), se a equação às diferenças (3.2.24) for estável, ou seja, se o modelo (3.2.9) for invertível, $\mathbf{T}_{t+q,s} \rightarrow \mathbf{0}$ quando $t \rightarrow \infty$, pelo que, em termos assintóticos, o efeito das variáveis residuais "de partida" é negligível.

No que se refere ao segundo nível de aproximação (Osborn, 1977), considere-se o termo $|\mathbf{T}^T (\mathbf{I}_{[(N+q) \times (N+q)]} \otimes \Sigma^{-1}) \mathbf{T}|$ de (3.2.20). O (r,s) -ésimo bloco de $\mathbf{T}^T (\mathbf{I}_{[(N+q) \times (N+q)]} \otimes \Sigma^{-1}) \mathbf{T}$ tem a forma:

$$\sum_{u=1}^{N+q} \mathbf{T}_{ur}^T \Sigma^{-1} \mathbf{T}_{us} \quad (3.2.26)$$

Uma vez que os q primeiros blocos linha de \mathbf{T} formam uma matriz identidade, tem-se:

$$\sum_{u=1}^{N+q} \mathbf{T}_{ur}^T \Sigma^{-1} \mathbf{T}_{us} = \delta_{rs} \Sigma^{-1} + \mathbf{T}_{q+1,r}^T \Sigma^{-1} \mathbf{T}_{q+1,s} + \mathbf{T}_{q+2,r}^T \Sigma^{-1} \mathbf{T}_{q+2,s} + \dots + \mathbf{T}_{q+N,r}^T \Sigma^{-1} \mathbf{T}_{q+N,s} \quad (3.2.27)$$

Deste modo, se o modelo (3.2.9) for invertível, os termos $\mathbf{T}_{t+q,r}^T \Sigma^{-1} \mathbf{T}_{t+q,s} \rightarrow \mathbf{0}$ quando $t \rightarrow \infty$, pelo que a soma (3.2.27) converge para uma matriz (limite) finita quando $N \rightarrow \infty$. Então, é de esperar que a contribuição dos termos de (3.2.27) para a função de verosimilhança seja

relativamente pequena para valores elevados de N , pelo que a utilização da função de verosimilhança aproximada l_2 significa incluir apenas o termo $\delta_{rs} \Sigma^{-1}$ da soma (3.2.27).

Portanto, a utilização da função de verosimilhança aproximada é legítima desde que o modelo seja invertível, contrariamente à função exacta, para cuja utilização essa propriedade não é necessária.

No que se refere às propriedades dos estimadores resultantes, Dunsmuir e Hannan (1976) mostram que os obtidos a partir da função de verosimilhança exacta são consistentes quer para modelos invertíveis quer para não invertíveis. Por outro lado, Reinsel (1976) indica que os estimadores obtidos a partir da função aproximada são também consistentes.

No entanto, Phadke e Kedem (1978) referem que, para dimensões finitas da série, quando o modelo é não invertível, o erro quadrático médio dos estimadores obtidos a partir da função de verosimilhança aproximada pode ser muito superior ao dos obtidos a partir da função exacta. Estudos de simulação apresentados por estes autores e por Hillmer e Tiao (1979) reforçam estas conclusões: quando o modelo é não invertível, os estimadores baseados na função exacta apresentam um comportamento bastante superior aos dos obtidos a partir da função aproximada; em particular, o erro quadrático médio destes últimos é superior e o seu enviesamento, embora decresça à medida que N aumenta, só assume valores baixos para dimensões muito elevadas da série, enquanto o enviesamento dos primeiros assume valores muito baixos mesmo para pequenos valores de N . O estudo dos primeiros autores sugere ainda que, quando as raízes de $|\theta(B)|$ estão bem afastadas do círculo unitário, o erro quadrático médio dos estimadores baseados na função aproximada é muito semelhante ao dos obtidos a partir da função exacta.

De qualquer modo, embora a função de verosimilhança exacta seja preferível em termos teóricos, a dificuldade da sua maximização pode impedir a sua utilização em algumas circunstâncias (Osborn, 1977). Assim, para ultrapassar essa dificuldade, Phadke e Kedem (1978) sugerem uma optimização em duas fases: numa primeira fase, proceder-se-ia à maximização da função de verosimilhança aproximada e, na segunda fase, utilizar-se-ia as estimativas obtidas na fase anterior como valores iniciais para a maximização da função de verosimilhança exacta.

2.3. Estimação dos parâmetros de modelos mistos auto-regressivos e de médias móveis

Considere-se o modelo misto auto-regressivo e de médias móveis de ordem (p,q) invertível e estacionário:

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (3.2.28)$$

Vai-se considerar apenas a função de verosimilhança aproximada, uma vez que, no caso de modelos mistos, as dificuldades computacionais são muito maiores, tornando o recurso a esta função muito mais vantajoso ou mesmo necessário. É claro que esse recurso só é legítimo se o modelo for invertível mas, tal como referem Jenkins e Alavi (1981), se isso não acontecer os estimadores resultantes poderão estar altamente correlacionados e os seus desvios-padrão perdem o significado, seja qual for o tipo de função de verosimilhança utilizada para a estimação, o que impossibilita a interpretação das estimativas.

Deste modo, um modelo misto auto-regressivo e de médias móveis é tratado da seguinte forma:

a) em primeiro lugar, o modelo é tratado como um modelo auto-regressivo, o que significa que se despreza as p primeiras observações;

b) em segundo lugar, as $(N-p)$ observações restantes são tratadas como sendo provenientes de um modelo de médias móveis, o que significa que se fixa as variáveis residuais "de partida" $\varepsilon_{p-q+1}, \dots, \varepsilon_p$ em zero.

Assim, a função de verosimilhança aproximada é (Wilson, 1973):

$$l(\phi, \theta, \Sigma | \mathbf{X}) = (2\pi)^{-\frac{mN}{2}} |\Sigma|^{-\frac{N}{2}} \exp\left\{-\frac{1}{2} \sum_{i=1}^N \varepsilon_i^T \Sigma^{-1} \varepsilon_i\right\} \quad (3.2.29)$$

Logaritmizando (3.2.29), obtém-se:

$$\log l(\phi, \theta, \Sigma | \mathbf{X}) = -\frac{1}{2} mN \log(2\pi) - \frac{N}{2} L(\phi, \theta, \Sigma | \mathbf{X}) \quad (3.2.30)$$

Então, os estimadores dos parâmetros são obtidos minimizando a função:

$$L(\phi, \theta, \Sigma | \mathbf{X}) = \log|\Sigma| + \frac{1}{N} \sum_{i=1}^N \varepsilon_i^T \Sigma^{-1} \varepsilon_i \quad (3.2.31)$$

Dada a dificuldade de minimização desta função, torna-se também necessário recorrer a um método de otimização não linear apropriado. Wilson (1973) apresenta um algoritmo adequado para o efeito, indicando ainda que os estimadores resultantes da minimização de (3.2.31) são consistentes, resultado que também é apontado por Reinsel (1976), e que $\hat{\phi}$ e $\hat{\theta}$ são assintoticamente não correlacionados com $\hat{\Sigma}$.

3. Estimação da ordem do modelo

O problema da estimação dos parâmetros de um modelo foi analisado admitindo-se que as ordens dos operadores auto-regressivo e de médias móveis eram conhecidas. Mas, na realidade, essas ordens são habitualmente desconhecidas, constituindo por isso parâmetros adicionais sobre os quais se torna necessário inferir a partir das observações, de modo a propor valores adequados.

Assim, vai-se apresentar os critérios que são habitualmente utilizados para este efeito.

3.1. Estimação da ordem de modelos auto-regressivos

Para a determinação da ordem deste tipo de modelos, têm sido propostos diversos critérios:

a) Critério da razão de verosimilhança (Hannan, 1970, pág. 340)

Trata-se de um procedimento que consiste em efectuar testes de hipóteses de uma forma sequencial (Lütkepohl, 1985). Admitindo-se que a verdadeira ordem do operador auto-regressivo, designada por p , é finita, fixa-se um valor máximo para essa ordem, designado por d , e testa-se sequencialmente as hipóteses $H_d: \phi_d=0$, $H_{d-1}: \phi_d=\phi_{d-1}=0$, etc. Se alguma destas hipóteses for rejeitada, o mesmo sucede a todas as hipóteses seguintes, terminando-se o procedimento quando isso acontece. A estatística utilizada para testar H_k , $k=0,1,\dots,d$ é:

$$LR(k) = N(\log|\hat{\Sigma}_{k-1}| - \log|\hat{\Sigma}_d|) \quad (3.3.1)$$

designando $\hat{\Sigma}_k$ o estimador de Σ obtido a partir do ajustamento de um modelo AR(k). Por outro lado, designando o estimador de p por \hat{p} , fixa-se $LR(0)$ de modo a obter $\hat{p} \geq 0$. A estatística $LR(k)$ segue uma distribuição assintótica $\chi^2_{m^2(d-k+1)}$, sendo conveniente notar que o nível de significância

num procedimento sequencial deste tipo difere do que é utilizado para cada teste individual (Anderson, 1971, secção 3.2.2.).

Jenkins e Alavi (1981) utilizam um procedimento ligeiramente diferente que consiste em testar sequencialmente $H_d^*: \phi_d=0$, $H_{d-1}^*: \phi_{d-1}=0$, etc. A estatística utilizada para testar H_k^* , $k=0,1,\dots,d$, designada por $LR^*(k)$, segue uma distribuição assintótica χ_m^2 , fixando-se também $LR^*(0)$ de modo a que $p \geq 0$.

Os critérios seguintes consistem na minimização de uma certa função. Definindo-se um valor máximo para p , que continuará a ser designado por d , determina-se o valor dessa função para cada ordem do operador auto-regressivo k , com $k=0,1,\dots,d$. O estimador de p , que continuará também a ser designado por \hat{p} , é aquele para o qual o modelo $AR(p)$ conduz ao valor mínimo da função.

Deste modo, os critérios habituais são:

b) FPE ("Final Prediction Error")

A generalização deste critério ao caso multivariado é proposta por Akaike (1971), a partir da sua proposta para modelos univariados (Akaike, 1969), assumindo a forma:

$$FPE(k) = \left(\frac{N + km}{N - km} \right)^m |\hat{\Sigma}_k| \quad (3.3.2)$$

Este critério não é consistente, sobreestimando p se d for superior a p (Lütkepohl, 1985).

c) AIC ("Akaike's Information Criterion")

Generalizando a proposta de Akaike (1969) para o caso univariado, o critério AIC assume a seguinte forma no caso multivariado (Parzen, 1977):

$AIC(k) = -2 \log(\text{máxima verosimilhança}) + 2$ (número de parâmetros do modelo ajustados independentemente) =

$$= \log |\hat{\Sigma}_k| + \frac{2m^2k}{N} \quad (3.3.3)$$

Paulsen e Tjøstheim (1985) provam que este critério também não é consistente, sobreestimando p se $d > p$.

d) CAT ("Criterion Autoregressive Transfer-Function")

Este critério é apresentado por Parzen (1977) generalizando a sua proposta para o caso univariado e assume a forma:

$$CAT(k) = \text{tr} \left\{ \frac{m}{N} \sum_{r=1}^k \left(\frac{N-rm}{N} \hat{\Sigma}_r^{-1} - \frac{N-km}{N} \hat{\Sigma}_k^{-1} \right) \right\} \quad (3.3.4)$$

Tal como o anterior, este critério não é consistente, ambos conduzindo habitualmente ao mesmo valor de \hat{p} (Parzen, 1977).

Por outro lado, verifica-se ainda a seguinte relação entre estes dois critérios (Parzen, 1977):

$$\frac{1}{m} CAT(k) \leq -\exp \left\{ -\frac{1}{m} AIC(k) \right\} \quad (3.3.5)$$

e) Critério de Shibata

A partir do critério proposto por Shibata (1980) para o caso univariado, Lütkepohl (1985) apresenta a sua generalização ao caso multivariado:

$$S(k) = \left(1 + \frac{2mk}{N} \right)^m |\hat{\Sigma}_k| \quad (3.3.6)$$

Este critério também não é consistente, sobreestimando o valor de p (Lütkepohl, 1985).

f) BIC ("Bayesian Information Criterion")

Trata-se de uma modificação Bayesiana do critério AIC, sendo a generalização ao caso multivariado da proposta de Schwarz (1978) para o caso univariado apresentada por Quinn (1980):

$$\text{BIC}(k) = \log|\hat{\Sigma}_k| + \frac{m^2 k \log N}{N} \quad (3.3.7)$$

Quinn (1980) mostra que este critério é consistente. Por outro lado, verifica-se a relação (Lütkepohl, 1985):

$$\hat{p}(\text{BIC}) \leq \hat{p}(\text{AIC}) \quad \text{se} \quad N \geq 8 \quad (3.3.8)$$

onde $\hat{p}(\cdot)$ designa o estimador de p utilizando o critério especificado dentro de parêntesis.

g) Critério ϕ

Este critério é proposto por Quinn (1980) a partir da proposta de Hannan e Quinn (1979) para o caso univariado:

$$\phi(k) = \log|\hat{\Sigma}_k| + \frac{2kb \log \log N}{N}, \quad b > 1 \quad (3.3.9)$$

Quinn (1980) mostra que se trata de um critério consistente desde que b seja superior a 1. Se $b=1$ e a dimensão da série for inferior a $\exp\{\exp m^2\}$ — o que até para $m=2$ é muito elevado — obtém-se a relação:

$$\hat{p}(\phi) = \hat{p}(\text{AIC}) \quad (3.3.10)$$

Por este motivo, Quinn (1980) sugere que se utilize $b=m^2$, do que resulta:

$$\hat{p}(\text{BIC}) \leq \hat{p}(\phi) \leq \hat{p}(\text{AIC}) \quad (3.3.11)$$

Por outro lado Lütkepohl (1985) indica ainda que se verifica a relação $\hat{p}(\phi) \leq \hat{p}(\text{AIC})$ se $N \geq 16$.

Lütkepohl (1985) apresenta um estudo de simulação no qual o critério BIC conduz mais frequentemente ao valor correcto de p , sendo imediatamente seguido pelo critério ϕ , aproximando-se os resultados de ambos à medida que a dimensão da série aumenta. Além disso, estes revelam-se como os critérios mais parcimoniosos. Por outro lado, os restantes critérios revelam-se bastante inferiores, sobreestimando o valor de p , não sendo no entanto possível hierarquizá-los, uma vez que a relação entre eles não se mantém constante. Por outro lado, é de referir ainda que os resultados do estudo de simulação indicam que a distribuição assintótica de $LR(k)$ e de $LR^*(k)$ não é adequada para pequenas amostras. No entanto, mesmo que o fosse, seria de esperar que o valor máximo para a ordem do operador auto-regressivo fosse escolhida em $\alpha 100\%$ dos casos, sendo α o nível de significância utilizado no primeiro teste. Assim, mesmo que N seja elevado e o nível de significância seja baixo, um procedimento sequencial deste tipo sobreestima a verdadeira ordem do modelo com probabilidade positiva, desde que essa ordem seja inferior ao valor máximo utilizado. Então, sobretudo quando o nível de significância é elevado, não é de surpreender que este procedimento conduza a más estimativas para a ordem do modelo auto-regressivo.

3.2. Estimação das ordens de modelos mistos auto-regressivos e de médias móveis

No caso de se tratar de um modelo ARMA(p,q), calcula-se o critério para uma grelha de valores das ordens dos operadores auto-regressivo e de médias móveis, retendo-se os que conduzirem ao valor mínimo desse critério.

Assim, Hannan (1981) propõe o seguinte critério para a estimação das ordens dos operadores de um modelo ARMA(p,q):

$$A(k+l) = \log |\hat{\Sigma}_{k+l}| + \frac{(k+l)H(N)}{N} \quad (3.3.12)$$

onde $H(N) \rightarrow \infty$ quando $N \rightarrow \infty$.

Hannan (1981) mostra que este critério é consistente, sendo necessário que $H(N) \rightarrow \infty$ quando $N \rightarrow \infty$ e que $H(N)/N \rightarrow 0$. É importante notar que, se se fizer $H(N)=2m^2$ em (3.3.12), se obtém o critério AIC, enquanto se se tiver $H(N)=m^2 \log N$, o critério resultante é o BIC e se for $H(N)=2m^2 \log \log N$, obtém-se o critério ϕ . Enquanto os dois últimos casos cumprem a condição de consistência para $A(k+l)$, o que significa que os critérios BIC e ϕ são consistentes, tal como já foi referido para o caso de modelos auto-regressivos, o mesmo já não se passa no primeiro caso, levando a concluir que o critério AIC não é consistente, tal como também já foi analisado.

Por outro lado, os critérios apresentados para a estimação da ordem de modelos AR(p) são generalizáveis ao caso ARMA(p,q).

Assim, Newbold e Hotopp (1986) propõem um procedimento em duas fases. Na primeira fase, o modelo ARMA é aproximado por um modelo AR de ordem suficientemente elevada, cujo valor máximo é habitualmente fixado em 10, visando-se estimar as variáveis residuais daquele. A ordem do modelo auto-regressivo é estimada através da utilização do critério AIC. Deste modo, as variáveis residuais do modelo ARMA, designadas por ϵ_t , são estimadas da forma:

$$\hat{\epsilon}_t = X_t - \hat{\phi}_1 X_{t-1} - \dots - \hat{\phi}_r X_{t-r} \quad (3.3.13)$$

onde $\hat{\phi}_s$ ($s=1, \dots, r$) são os estimadores dos parâmetros do modelo AR(r) utilizado para aproximar o modelo ARMA. Como já foi referido, o critério AIC não é consistente, sobreestimando a ordem do modelo, mas, uma vez que o objectivo nesta fase é apenas estimar os vectores ϵ_t e não

seleccionar um modelo, poderá ser vantajoso utilizar um critério que, neste sentido, é conservador.

Na segunda fase, efectua-se a regressão de X_t sobre X_{t-u} ($u=1, \dots, k$) e \hat{E}_{t-v} ($v=1, \dots, l$) para diversos valores de k e de l , sendo \hat{E}_t obtido a partir de (3.3.13). Assim, estima-se modelos da forma:

$$X_t = \hat{\phi}_1 X_{t-1} + \dots + \phi_k X_{t-k} + \theta_1 \hat{E}_{t-1} + \dots + \theta_l \hat{E}_{t-l} + \eta_{k,l,t} \quad (3.3.14)$$

onde $\eta_{k,l,t}$ é um vector de dimensão ($m \times 1$), constituindo o termo residual. Uma vez que \hat{E}_t são conhecidos, é possível estimar os parâmetros em (3.3.14) utilizando o método dos mínimos quadrados. Designando por $\hat{\Omega}_{k,l}$ o estimador da matriz de variâncias e covariâncias residual de (3.3.14), utiliza-se o critério BIC para seleccionar um modelo ARMA, ou seja, minimiza-se:

$$\text{BIC}(k, l) = \log |\hat{\Omega}_{k,l}| + \frac{(k+l)m^2 \log N}{N} \quad (3.3.15)$$

É importante sublinhar que estes critérios não são procedimentos destinados a fornecer definitivamente os valores para as ordens dos operadores de um modelo. Pelo contrário, trata-se apenas de instrumentos de apoio que auxiliam na selecção de um modelo adequado.

4. Distribuição assintótica dos estimadores

É importante estudar a distribuição dos estimadores propostos, nomeadamente para fins inferenciais. De facto, tal como será analisado mais à frente, a distribuição dos estimadores dos parâmetros desempenha um papel fundamental no desenvolvimento de métodos de crítica do modelo previamente ajustado à série cronológica.

Assim, tratar-se-à primeiro a distribuição dos estimadores da função de correlação cruzada e da função de correlação parcial, passando-se depois para a distribuição dos estimadores dos parâmetros e, finalmente, para a distribuição dos estimadores das funções de covariância e de correlação residuais.

4.1. Distribuição dos estimadores da função de correlação e da função de correlação parcial

Antes de proceder à apresentação da distribuição da função de correlação cruzada estimada, designada por $r_{ij}(k)$, é importante referir a expressão obtida por Bartlett (1978, pág.352) para a covariância assintótica entre $r_{ij}(k)$ e $r_{ij}(k+l)$ quando $\{X_{it}\}$ e $\{X_{jt}\}$ são processos estocásticos reais com distribuição Normal:

admitindo-se que $\sum_{k=-\infty}^{\infty} |\gamma_{ij}(k)| < \infty$

tem-se:

$$\begin{aligned} \text{Cov}(r_{ij}(k), r_{ij}(k+l)) = & (N-k)^{-1} \sum_{r=-\infty}^{\infty} \{ \rho_{ii}(r)\rho_{jj}(r+l) + \rho_{ij}(r+l+2k)\rho_{ji}(r) + \\ & + \rho_{ij}(k)\rho_{ij}(k+l) \left[\rho_{ij}^2(r) + \frac{1}{2}\rho_{ii}^2(r) + \frac{1}{2}\rho_{jj}^2(r) \right] - \\ & - \rho_{ij}(k) \left[\rho_{ii}(r)\rho_{ij}(r+k+l) + \rho_{ji}(r)\rho_{jj}(r+k+l) \right] - \\ & - \rho_{ij}(k+l) \left[\rho_{ii}(r)\rho_{ij}(r+k) + \rho_{ji}(r)\rho_{jj}(r+k) \right] \} \end{aligned}$$

$$\begin{aligned}
\text{Cov}(r_{ij}(k), r_{ij}(k+l)) = & (N-k)^{-1} \sum_{r=-k}^{\infty} \{ \rho_{ii}(r) \rho_{ij}(r+l) + \rho_{ij}(r+l+2k) \rho_{jj}(r) + \\
& + \rho_{ij}(k) \rho_{ij}(k+l) \left[\rho_{ij}^2(r) + \frac{1}{2} \rho_{ii}^2(r) + \frac{1}{2} \rho_{jj}^2(r) \right] - \\
& - \rho_{ij}(k) \left[\rho_{ii}(r) \rho_{ij}(r+k+l) + \rho_{jj}(r) \rho_{ij}(r+k+l) \right] - \\
& - \rho_{ij}(k+l) \left[\rho_{ii}(r) \rho_{ij}(r+k) + \rho_{jj}(r) \rho_{ij}(r+k) \right] \}
\end{aligned}
\tag{3.4.1}$$

Em particular, fazendo $l=0$, obtém-se:

$$\begin{aligned}
V(r_{ij}(k)) = & (N-k)^{-1} \sum_{r=-k}^{\infty} \{ \rho_{ii}(r) \rho_{ij}(r) + \rho_{ij}(r+2k) \rho_{jj}(r) + \\
& + \rho_{ij}^2(k) \left[\rho_{ij}^2(r) + \frac{1}{2} \rho_{ii}^2(r) + \frac{1}{2} \rho_{jj}^2(r) \right] - \\
& - 2\rho_{ij}(k) \left[\rho_{ii}(r) \rho_{ij}(r+k) + \rho_{jj}(r) \rho_{ij}(r+k) \right] \}
\end{aligned}
\tag{3.4.2}$$

Na realidade, é preferível ter $(N-|k|)^{-1}$ em vez de $(N-k)^{-1}$ na equação (3.4.2). Por outro lado, se os dois processos forem independentes, a hipótese da sua normalidade deixa de ser exigida, sendo imediato verificar que estas expressões vêm consideravelmente simplificadas:

$$\text{Cov}(r_{ij}(k), r_{ij}(k+l)) = (N-|k|)^{-1} \sum_{r=-k}^{\infty} \rho_{ii}(r) \rho_{ij}(r+l)
\tag{3.4.3}$$

$$V(r_{ij}(k)) = (N-|k|)^{-1} \sum_{r=-k}^{\infty} \rho_{ii}(r) \rho_{ij}(r)
\tag{3.4.4}$$

Além disso, verifica-se ainda que, se ambos os processos forem ruídos brancos independentes, $\text{Cov}(r_{ij}(k), r_{ij}(k+l)) \approx O(N^{-1})$, onde O designa "infinitésimo", e que a variância assintótica de $r_{ij}(k)$ se reduz a $(N-|k|)^{-1}$, o que é habitualmente aproximado a N^{-1} , admitindo-se que N é grande relativamente a k .

Estas expressões são de extrema utilidade para a caracterização da distribuição assintótica de $r_{ij}(k)$, que é muito complicada de obter em termos gerais (Brockwell e Davis, 1987, pág.400). No entanto, é possível defini-la quando $\{X_{it}\}$ e $\{X_{jt}\}$ são processos de médias móveis independentes, tendo-se então o seguinte teorema (Hannan, 1970, pág.230):

Teorema 3.4.1. Suponha-se que:

$$X_{it} = \sum_{r=-\infty}^{\infty} \alpha_r \varepsilon_{i,t-r}$$

$$X_{jt} = \sum_{r=-\infty}^{\infty} \lambda_r \varepsilon_{j,t-r}$$

onde $\{\varepsilon_{it}\}$ e $\{\varepsilon_{jt}\}$ são independentes e identicamente distribuídos e independentes entre si, com:

$$V(\varepsilon_{it}) = \sigma_i^2;$$

$$V(\varepsilon_{jt}) = \sigma_j^2;$$

$$\sum_{r=-\infty}^{\infty} |\alpha_r| < \infty;$$

$$\sum_{r=-\infty}^{\infty} |\lambda_r| < \infty.$$

Então:

$$r_{ij}(k) \sim N(0; N^{-1} \sum_{r=-\infty}^{\infty} \rho_{ii}(r) \rho_{jj}(r))$$

Por outro lado, no que respeita à função de correlação parcial estimada, Tiao e Box (1979,1981) indicam que, para um modelo AR(p), a distribuição assintótica conjunta de $\hat{\phi}_1^T, \dots, \hat{\phi}_p^T$ é Normal. Assim, considere-se a seguinte matriz das somas de quadrados e de produtos cruzados residuais resultantes da estimação de um modelo AR(k):

$$D(k) = \sum_{t=k+1}^N (X_t - \hat{\phi}_1 X_{t-1} - \dots - \hat{\phi}_k X_{t-k})(X_t - \hat{\phi}_1 X_{t-1} - \dots - \hat{\phi}_k X_{t-k})^T \quad (3.4.5)$$

Deste modo, é possível testar as hipóteses nulas $\phi_k=0$ contra as alternativas $\phi_k \neq 0$, $k=1, \dots, p$, a partir da estimação de um modelo AR(k), através do critério da razão de verosimilhança, sendo a necessária estatística definida pelo seguinte quociente de determinantes:

$$\lambda = |D(k)| / |D(k-1)| \quad (3.4.6)$$

Utilizando uma aproximação devida a Bartlett, Tiao e Box (1979, 1981) propõem então a seguinte estatística para a realização do teste:

$$\zeta(k) = -\left(n - \frac{1}{2} - km\right) \log \lambda \quad (3.4.7)$$

onde $n=N-p$ (se o modelo incluir um termo constante, tem-se $n=N-p-1$). Sob a hipótese nula atrás definida, $\zeta(k)$ segue uma distribuição assintótica de χ_m^2 , resultando daqui um outro procedimento para estimar a ordem de um modelo auto-regressivo.⁽⁵⁾

4.2. Distribuição dos estimadores dos parâmetros

Designando $\beta = (\beta_1, \beta_2, \dots, \beta_w)^T$ o vector dos parâmetros de um modelo ARMA(p,q), exceptuando-se os elementos de Σ , vai-se analisar a distribuição do estimador de β , que será designado por $\hat{\beta}$, tendo $\hat{\phi}$ e $\hat{\theta}$ sido obtidos a partir de (3.2.31).

Assim, Wilson (1973) indica que a distribuição assintótica de $\hat{\beta}$ é:

$$\hat{\beta} \sim N(\beta; N^{-1} \Delta^{-1}) \quad (3.4.8)$$

onde Δ é uma matriz com elementos $\Delta_{uv} = \frac{1}{2} E\{\partial^2 L / \partial \beta_u \partial \beta_v\}$, $u, v=1, 2, \dots, w$, e L é a função definida em (3.2.31). Por outro lado (Wilson, 1973), um estimador consistente de Δ , designado por $\hat{\Delta}$ pode ser obtido da forma:

$$\hat{\Delta}_{uv} = \frac{1}{N} \sum_{t=1}^N \left(\frac{\partial \hat{\boldsymbol{\varepsilon}}_t}{\partial \hat{\boldsymbol{\beta}}_u} \right)^T \hat{\boldsymbol{\Sigma}}^{-1} \left(\frac{\partial \hat{\boldsymbol{\varepsilon}}_t}{\partial \hat{\boldsymbol{\beta}}_v} \right) \quad (3.4.9)$$

onde $\hat{\Delta}_{uv}$ designa o elemento de ordem (u,v) da matriz $\hat{\Delta}$.

4.3. Distribuição dos estimadores da função de covariância e da função de correlação residuais

As funções de covariância e de correlação residuais estimadas possuem distribuições assintóticas que se revestem da maior importância, em particular na análise da adequabilidade do modelo ajustado.

Assim, no que respeita à primeira função, é necessário distinguir o estimador da matriz de covariância das variáveis residuais $\boldsymbol{\varepsilon}_t$ e o estimador da matriz de covariância dos resíduos de estimação $\hat{\boldsymbol{\varepsilon}}_t$. O primeiro será designado por $C_\varepsilon(k)$ e apresenta a forma:

$$C_\varepsilon(k) = \frac{1}{N} \sum_t \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_{t-k}^T \quad (3.4.10)$$

O segundo, designado por $\hat{C}_\varepsilon(k)$, tem a expressão:

$$\hat{C}_\varepsilon(k) = \frac{1}{N} \sum_t \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_{t-k}^T \quad (3.4.11)$$

Por exemplo, $\hat{C}_\varepsilon(0) = \frac{1}{N} \sum_t \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_t^T = \hat{\boldsymbol{\Sigma}}$

Deste modo, surge o teorema (Hosking, 1980a; Chitturi, 1976):
Teorema 3.4.4. Uma vez que $\{\boldsymbol{\varepsilon}_t\}$ é um processo puramente aleatório multivariado, verifica-se os dois resultados seguintes:

a) $C_\varepsilon(k)$ segue distribuição assintótica normal com valor esperado $\boldsymbol{\Sigma}_{\delta_{0k}}$ e covariâncias:

$$\text{Cov}(c_{ij}(k), c_{ij}(l)) = N^{-2} (N - k) \text{Cov}(\boldsymbol{\varepsilon}_u, \boldsymbol{\varepsilon}_v) V(\boldsymbol{\varepsilon}_j) \delta_{kl} \quad (3.4.12)$$

ou, assintoticamente:

$$\text{Cov}(c_{ij}(k), c_{ij}(l)) = N^{-1} \text{Cov}(\varepsilon_u, \varepsilon_\pi) V(\varepsilon_x) \delta_{kl} \quad (3.4.13)$$

b) Por outro lado, definido ainda $C_\varepsilon = [C_\varepsilon(1) C_\varepsilon(2) \dots C_\varepsilon(R)]$ sendo R um inteiro qualquer, tem-se o seguinte resultado assintótico:

$$\text{vec} C_\varepsilon \sim N(0; N^{-1}W)$$

sendo $W = I_R \otimes \Sigma \otimes \Sigma$.

No que respeita à distribuição de $\hat{C}_\varepsilon(k)$, é necessário definir primeiro as matrizes:

$$G_k = \sum_{u=0}^{\infty} \{ \Sigma \psi_u^T \otimes \pi_{k-u} \} \quad (3.4.15)$$

onde $\psi(B) = \phi^{-1}(B)\theta(B)$ e $\pi(B) = \theta^{-1}(B)$;

$$H_k = \Sigma \otimes \pi_k$$

É de sublinhar que $G_0 = H_0 = \Sigma \otimes I_m$ para $k < 0$.

É necessário definir também:

$$Y = \begin{bmatrix} G_0 & 0 & 0 & \dots & 0 \\ G_1 & G_0 & 0 & \dots & 0 \\ G_2 & G_1 & G_0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ G_{R-1} & G_{R-2} & G_{R-3} & \dots & G_{R-p} \end{bmatrix};$$

$$Z = \begin{bmatrix} H_0 & 0 & 0 & \dots & 0 \\ H_1 & H_0 & 0 & \dots & 0 \\ H_2 & H_1 & H_0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ H_{R-1} & H_{R-2} & H_{R-3} & \dots & H_{R-p} \end{bmatrix};$$

$U=(Y-Z)$ uma matriz de dimensão $m^2R \times m^2 (p+q)$;

$$Q=(U^T W^{-1} U)^{-1} U^T W^{-1} \quad (3.4.17)$$

Definindo ainda $\hat{C}_\epsilon = [\hat{C}_\epsilon(1) \hat{C}_\epsilon(2) \dots \hat{C}_\epsilon(k)]$, tem-se o teorema (Hosking,1980a):

Teorema 3.4.2. A distribuição assintótica de $\text{vec}\hat{C}_\epsilon$ é:

$$\text{vec}\hat{C}_\epsilon \sim N(0; N^{-1} (I_{m^2R} - Q)W)$$

Relativamente à função de correlação residual estimada, é também necessário definir o estimador da matriz de correlação das variáveis residuais \mathcal{E}_t e o estimador da matriz de correlação dos resíduos de estimação $\hat{\mathcal{E}}_t$. O primeiro será designado por $R_\epsilon(k)$ e o segundo por $\hat{R}_\epsilon(k)$, sendo as suas expressões facilmente deduzidas a partir de (3.4.10) e de (3.4.11) respectivamente:

$$R_\epsilon(k) = C_{\epsilon_0}^{-1/2} C_\epsilon(k) C_{\epsilon_0}^{-1/2} \quad (3.4.18)$$

$$\hat{R}_\epsilon(k) = \hat{C}_{\epsilon_0}^{-1/2} \hat{C}_\epsilon(k) \hat{C}_{\epsilon_0}^{-1/2} \quad (3.4.19)$$

onde $C_{\epsilon_0} = \text{diag}\{c_{\epsilon_{11}}(0) \dots c_{\epsilon_{mm}}(0)\}$ e $\hat{C}_{\epsilon_0} = \text{diag}\{\hat{c}_{\epsilon_{11}}(0) \dots \hat{c}_{\epsilon_{mm}}(0)\}$.

sendo $R_\epsilon = [R_\epsilon(1) R_\epsilon(2) \dots R_\epsilon(R)]$ e $\hat{R}_\epsilon = [\hat{R}_\epsilon(1) \hat{R}_\epsilon(2) \dots \hat{R}_\epsilon(R)]$

tem-se o teorema (Li e McLeod, 1981):

Teorema 3.4.3. A distribuição assintótica de $\text{vec}R_\epsilon$ é:

$$\text{vec}R_\epsilon \sim N(0; N^{-1}W) \quad (3.4.20)$$

tendo W sido definida no teorema 3.4.1..

Por outro lado, para a distribuição de \hat{R}_ε , é necessário definir primeiro a matriz de dimensão $[(p+q)m^2 \times Rm^2]$:

$$S^T = \begin{bmatrix} A_{11} & \dots & A_{1R} \\ \vdots & & \\ A_{p+q} & \dots & A_{p+qR} \end{bmatrix} \quad (3.4.21)$$

onde:

$$A_{lr} = \begin{cases} \frac{1}{N} \sum_t E \left[(T_{l,r})_{ij} \varepsilon_{l,r}^T \right]_{gh} & (l \geq r, r \leq p; l \geq r-p, r > p) \\ 0 & \text{outros } (l,r) \end{cases} \quad (3.4.22)$$

Com $i,j,g,h=1,\dots,m$ e onde $(T_{l,r})_{ij}$ corresponde a $-\partial \varepsilon_l / \partial \phi_{ij,r}$ ou a $\partial \varepsilon_l / \partial \theta_{ij,r}$, sendo $\phi_{ij,r}$ e $\theta_{ij,r}$ respectivamente o (i,j) -ésimo elemento da matriz de parâmetros ϕ_r e da matriz θ_r .

Assim, tem-se o teorema (Li e Mcleod, 1981):

Teorema 3.4.4. A distribuição assintótica de $\text{vec} \hat{R}_\varepsilon$ é:

$$\text{vec} \hat{R}_\varepsilon \sim N(0; N^{-1} \Lambda)$$

onde $\Lambda = (W - SI^{-1}S^T)$, sendo I a matriz de informação:

$$I = \frac{1}{2N} \lim_{N \rightarrow \infty} E \{ \partial^2 V / \partial \beta \partial \beta^T \}$$

com $V = \sum_{t=1}^N \hat{\varepsilon}_t^T \hat{\Sigma}^{-1} \hat{\varepsilon}_t$.

IV. CONFIRMAÇÃO DO DIAGNÓSTICO

Após a estimação do modelo identificado, torna-se necessário proceder à sua crítica, ou seja, à confirmação do diagnóstico, isto é, vai-se averiguar se o modelo é adequado à série cronológica. Assim, se não se encontrar discrepâncias relevantes, conclui-se que o modelo representa convenientemente a série cronológica, podendo-se estar mais confiante na sua utilização.

Deste modo, o objectivo de todos os métodos que vão ser apresentados é não só o de averiguar se o modelo é adequado, mas também o de evidenciar quaisquer características do verdadeiro processo multivariado subjacente à série cronológica que não tenham sido tomadas em conta, ou seja, modelizadas. Coloca-se assim o modelo identificado em questão e, se não houver evidência suficiente para o rejeitar, é possível utilizá-lo com maior segurança.

1. Teste de independência de dois processos estocásticos estacionários univariados

Para avaliar a qualidade do modelo identificado, é conveniente averiguar se um dado processo univariado componente do vector $\{X_t\}$ deve ser incluído na especificação de uma qualquer equação. Para o efeito, é possível estudar se existe qualquer relação entre esse processo univariado e o processo a que respeita a equação em causa, surgindo como instrumento mais indicado para estudar a existência de tal relação a respectiva função de correlação cruzada que, por ser desconhecida, conduz à necessidade de utilizar o seu estimador, que já foi definido.

Logo, o teorema (3.4.1) tem um papel importante quando se pretende testar a hipótese de independência entre dois processos univariados. Se ambos os processos forem ruídos brancos e independentes, tem-se o resultado:

$$r_{ij}(k) \sim N(0; N^{-1}) \quad (4.1.1)$$

tornando-se muito fácil testar a hipótese $\rho_{ij}(k)=0$, para o que basta comparar $N^{1/2}r_{ij}(k)$ com o valor adequado da distribuição $N(0;1)$.

Mas, se os processos estiverem correlacionados, um valor elevado de $N^{1/2}r_{ij}(k)$ não indica necessariamente que $\rho_{ij}(k)$ é significativamente diferente de zero, sendo necessário ter em conta a natureza de cada um dos processos. De facto, as expressões (3.4.1) e (3.4.2) mostram à evidência que os sucessivos valores da função de correlação cruzada estimada estão correlacionados e que a variância desta depende da estrutura da autocorrelação de cada um dos processos, impedindo a realização do teste com base nos valores da referida função. Mesmo detectando a presença de tal dificuldade, seria sempre extremamente difícil inferir correctamente a relação entre dois processos através da utilização da respectiva função de correlação cruzada estimada.

Por outro lado, observando as expressões (3.4.3) e (3.4.4), verifica-se ainda que podem ocorrer valores elevados, todos eles espúrios, da função de correlação cruzada estimada entre dois processos não correlacionados, apenas como resultado de valores elevados da função de autocorrelação de cada um desses processos.

Deste modo, é possível chegar a conclusões erradas através da atribuição de significado a padrões aparentes da função de correlação cruzada que, na realidade, são resultado das propriedades amostrais do estimador utilizado, o que torna este difícil de interpretar. Isto pode acontecer mesmo quando dois processos são independentes, o que será analisado mais à frente.

Assim, a melhor forma de ultrapassar esta dificuldade é converter ambos os processos em ruídos brancos através da aplicação de filtros apropriados e só depois estimar a respectiva função de correlação cruzada. Essa conversão, designada de "prewhitening" pelos autores

anglo-saxónicos, é realizada tratando cada processo univariado separadamente e ajustando--lhe um modelo adequado. As variáveis residuais resultantes deste procedimento constituem o novo processo que, se se tratar do modelo correcto, é um ruído branco.

Desta forma, para testar a hipótese de independência entre dois processos $\{X_{it}\}$ e $\{X_{jt}\}$, é conveniente notar primeiro que, se essa hipótese for verdadeira, os processos resultantes das respectivas operações de filtragem $\{\varepsilon_{it}\}$ e $\{\varepsilon_{jt}\}$ são também independentes. Admite-se então que $\{X_{it}\}$ e $\{X_{jt}\}$ podem ser separadamente modelizados pelos modelos ARMA univariados:

$$\phi_i(B)X_{it} = \theta_i(B)U_{it} \quad (4.1.2)$$

$$\phi_j(B)X_{jt} = \theta_j(B)U_{jt} \quad (4.1.3)$$

onde $\phi_i(B)$ e $\phi_j(B)$ são os operadores auto-regressivos e $\theta_i(B)$ e $\theta_j(B)$ os operadores de médias móveis do modelo ajustado a $\{X_{it}\}$ e a $\{X_{jt}\}$ respectivamente, e $\{U_{it}\}$ e $\{U_{jt}\}$ são processos puramente aleatórios univariados.

Mas, quando se este método a uma série cronológica univariada, uma vez que se desconhece o verdadeiro processo a que lhe está subjacente, as variáveis residuais de U_{it} e U_{jt} são também desconhecidas, tornando impossível a obtenção da sua função de correlação cruzada estimada, que será designada por $r_{U_{jt}}(k)$. Logo, é necessário recorrer aos resíduos de estimação dos modelos univariados \hat{U}_{it} e \hat{U}_{jt} , que constituem as séries após as operações de filtragem, não sendo mais do que aproximações a processos puramente aleatórios. No entanto, Haugh (1976) mostra que, se $\{X_{it}\}$ e $\{X_{jt}\}$ forem independentes, a função de correlação cruzada estimada entre $\{\hat{U}_{it}\}$ e $\{\hat{U}_{jt}\}$, que será designada por $\hat{r}_{U_{jt}}(k)$, tem a mesma distribuição assintótica que $r_{U_{jt}}(k)$, ou seja:

$$\hat{r}_{u_y}(k) \sim N(0; N^{-1}) \quad (4.1.4)$$

Assim, um teste de independência entre dois processos univariados $\{X_{it}\}$ e $\{X_{jt}\}$ a um nível de significância aproximado de 5%, por exemplo, resulta da comparação dos valores de $|\hat{r}_{u_y}(k)|$ com $2N^{-1/2}$.

No entanto, a partir de estudos de simulação, Haugh (1976) conclui que a aproximação $V(\hat{r}_{u_y}(k)) \sim N^{-1}$ é adequada apenas para valores k/N até 1/10 e que, à medida que k se torna elevado relativamente a N , se verifica um decréscimo de $V(\hat{r}_{u_y}(k))$. Esta última característica deve-se ao facto de:

$$V(\hat{r}_{u_y}(k)) = N^{-2}(N - |k|) \quad (4.1.5)$$

Portanto, para valores de $|k|$ elevados relativamente a N , a aproximação:

$$V(\hat{r}_{u_y}(k)) \sim N^{-1}(1 - |k|/N) \quad (4.1.6)$$

parece ser mais correcta do que a anteriormente proposta. Por outro lado, os mesmos estudos indicam que não é necessário ser muito exigente quanto à dimensão das séries cronológicas para que estas aproximações sejam válidas.

É importante referir ainda que a conversão em ruído branco de apenas um dos processos, $\{X_{it}\}$, por exemplo, leva a que, sob a hipótese de independência entre os dois processos, a função de correlação cruzada estimada entre $\{U_{it}\}$ e $\{X_{jt}\}$ siga também distribuição assintótica Normal de valor esperado nulo e variância N^{-1} , o que resulta do teorema 3.4.1.. Assim, sob a hipótese de independência entre $\{X_{it}\}$ e $\{X_{jt}\}$, os valores da função de correlação cruzada estimada entre $\{U_{it}\}$ e $\{U_{jt}\}$, designada por $r_{u_i, x_j}(k)$, continuam a estar contidos nas bandas $\pm 2N^{-1/2}$ com uma probabilidade aproximada de 95% (admitindo-se que o nível de significância do teste continua a ser de 5%). Mas, considerando a

estimação da função para dois "lags" diferentes, $r_{u_i, x_j}(k)$ e $r_{u_i, x_j}(k + l)$, a covariância assintótica entre estes é:

$$\text{Cov}(r_{u_i, x_j}(k), r_{u_i, x_j}(k + l)) = N^{-1} \rho_{ij}(l) \quad (4.1.7)$$

o que significa que a respectiva correlação é a seguinte:

$$\rho(r_{u_i, x_j}(k), r_{u_i, x_j}(k + l)) = \rho_{ij}(l) \quad (4.1.8)$$

isto é, no caso em que dois processos são não correlacionados e apenas um deles é puramente aleatório, a respectiva função de correlação cruzada estimada tem a mesma função de autocorrelação que o processo que não foi convertido em ruído branco. Portanto, conclui-se que a função de correlação cruzada estimada entre dois processos, sendo um deles um ruído branco, e mesmo que eles sejam não correlacionados, variará em torno de um valor médio de zero com um desvio-padrão de $N^{-1/2}$ segundo um padrão sistemático igual ao comportamento da função de autocorrelação do processo que não é puramente aleatório.

Logo, só no caso de ambos os processos serem puramente aleatórios e não estarem correlacionados entre si, é que a covariância e, portanto, a correlação entre os valores da respectiva função de correlação cruzada estimada assume o valor zero.

O teste acabado de apresentar utiliza individualmente a função de correlação cruzada estimada para cada "lag". No entanto, é útil utilizar também um conjunto de valores dessa função para diferentes "lags", o que permitirá revelar com maior rigor a natureza da relação entre os processos em análise. Assim, tirando partido da distribuição assintótica de $\hat{r}_{u_{ij}}(k)$, chega-se a duas estatísticas com distribuição assintótica do qui-quadrado e que cumprem esse objectivo.

1.2 Testes globais

Antes de proceder à apresentação destes testes, é necessário definir a distribuição de um conjunto de pontos da função de correlação cruzada estimada, o que é feito no teorema (Hannan, 1970, pág. 230):

Teorema 4.1.1. Considere-se dois processos estocásticos univariados $\{X_{it}\}$ e $\{X_{jt}\}$ obedecendo às mesmas condições do teorema 3.4.1..

Então, qualquer conjunto finito de valores $\sqrt{N}r_{ij}(k)$ tem distribuição assintótica normal de valor esperado 0 e com matriz de variâncias e covariâncias cujo elemento genérico para os "lags" k e $(k+l)$ é:

$$\sum_{r=-\infty}^{\infty} \rho_{ii}(r)\rho_{jj}(r+l)$$

Mas, como já foi referido, é necessário recorrer a $\hat{r}_{U_{ij}}(k)$, pelo que se tem o teorema (Haugh, 1976):

Teorema 4.1.2. Considere-se dois processos estocásticos univariados $\{X_{it}\}$ e $\{X_{jt}\}$ conjuntamente estacionários e nas condições do teorema anterior.

Suponha-se que se dispõe de N observações de $\{X_{it}\}$ e de $\{X_{jt}\}$ constituindo duas séries cronológicas que podem ser modelizadas por (4.1.2) e (4.1.3).

Se as duas séries são independentes, então $\sqrt{N}\hat{r}_U$ e $\sqrt{N}r_U$ têm a mesma distribuição assintótica $N(0;I_S)$, onde $r_U=(r_{U_{ij}}(k_1), \dots, r_{U_{ij}}(k_s))^T$ e k_1, \dots, k_s são números inteiros diferentes.

Nestas condições, já é possível apresentar os testes referidos, sendo o primeiro o teste de Haugh e o segundo um teste que tem em conta a possibilidade de existência de um padrão de comportamento da função de

correlação cruzada, constituindo este último um aperfeiçoamento do primeiro.

1.2.1. Teste de Haugh

Haugh (1976) propõe uma estatística para testar globalmente a hipótese de independência entre dois processos, utilizando a informação fornecida pela função de correlação cruzada para diferentes "lags" e definindo a distribuição dessa estatística a partir dos teoremas 4.1.1. e 4.1.2..

A estatística para a realização do teste é a seguinte:

$$S_M = N \sum_{k=-M}^M \hat{r}_{U_{ij}}^2(k) \quad (4.1.9)$$

onde M é um número inteiro positivo considerado adequado. Sob a hipótese de independência entre $\{X_{it}\}$ e $\{X_{jt}\}$, S_M segue uma distribuição assintótica de $\chi_{(2M+1)}^2$.

No caso de $|k|$ ser elevado relativamente a N , e uma vez que, nessa situação é a aproximação (4.1.6) que deve ser utilizada, a estatística para a realização do teste vem ligeiramente diferente:

$$S_M^* = N^2 \sum_{k=-M}^M (N-|k|)^{-1} \hat{r}_{U_{ij}}(k) \quad (4.1.10)$$

O valor de M a utilizar deve ser suficientemente elevado de modo a incluir qualquer possível relação entre os dois processos, devendo ter-se em atenção que, se for elevado relativamente a N , o que pode ser considerado para $M > N/10$, é mais adequado utilizar S_M^* .

É importante referir ainda que a distribuição assintótica utilizada pressupõe a independência dos dois processos, não sendo apropriada se existirem valores elevados da função de correlação cruzada.

1.2.2 Teste de Koch e Yang

O teste de Haugh ignora a possibilidade de existência de um padrão de comportamento dos sucessivos valores da função de correlação cruzada, podendo não rejeitar a hipótese de independência se a relação entre os dois processos se verificar ainda para "lags" elevados, sendo pouco intensa para cada "lag". Assim, utilizando o teste de Haugh, Pierce (1977) conclui que é fraca a evidência estatística que permite rejeitar a hipótese de independência entre pares de diversas variáveis monetárias e macroeconómicas que estariam fortemente relacionadas segundo a teoria económica.

Koch e Yang (1986) propõem então uma estatística que visa ultrapassar esse problema incorporando a possibilidade de existência de um padrão de comportamento da função de correlação cruzada. A sua abordagem baseia-se no facto de, para dois processos independentes, a função de correlação cruzada estimada entre os resíduos de estimação dos modelos univariados $\hat{r}_v(k)$, dever assumir valores baixos e distribuídos aleatoriamente em torno de zero. A estatística de Haugh, atribuindo o mesmo peso a todos os valores, revela-se incapaz de distinguir uma função de correlação cruzada estimada cujos valores são baixos e distribuídos aleatoriamente em torno de zero, de uma outra com valores baixos que seguem um padrão de comportamento próprio. O primeiro caso ocorre quando se trata de dois processos independentes, verificando-se o segundo para dois processos cuja relação, sendo pouco intensa em cada "lag", permanece até "lags" elevados. Portanto, a possibilidade de existência de correlação entre sucessivos valores da função de correlação cruzada estimada é ignorada no teste de Haugh, embora seja incompatível com hipótese de independência entre dois processos estocásticos univariados.

Assim, a estatística proposta por Koch e Yang (1986) e que visa ter esta questão em linha de conta, é:

$$r_h^* = N \sum_{k=-M}^{M-h} \left[\sum_{l=0}^h \hat{r}_{U_y}(k+l) \right]^2 \quad h = 0, 1, 2, \dots, M-1 \quad (4.1.11)$$

Se $h=0$, obtém-se a estatística de Haugh:

$$r_0^* = S_M = N \sum_{k=-M}^M \hat{r}_{U_y}^2(k) \quad (4.1.12)$$

Fazendo $h=1$, incorpora-se a informação sobre a covariação entre os valores da função de correlação cruzada estimada diferindo de um "lag":

$$r_1^* = N \sum_{k=-M}^{M-1} \left[\hat{r}_{U_y}(k) + \hat{r}_{U_y}(k+1) \right]^2 \quad (4.1.13)$$

É fácil verificar que o padrão de comportamento da função de correlação cruzada estimada vem reflectido na estatística r_h^* .

Para definir a distribuição de r_h^* , é conveniente notar primeiro que esta estatística é uma forma quadrática em $\sqrt{N}\hat{r}_U$, isto é:

$$r_h^* = N\hat{r}_U^T A_h \hat{r}_U \quad (4.1.14)$$

onde $A_h = D_h^T D_h$ é uma matriz simétrica de dimensão $[(2M+1) \times (2M+1)]$ e D_h tem dimensão $[(2M+1-h) \times (2M+1)]$ com cada linha composta por $(h+1)$ elementos iguais a 1 e $(2M-h)$ elementos nulos.

A estatística de Haugh pode ser expressa da forma:

$$r_0^* = N\hat{r}_U^T A_0 \hat{r}_U \quad (4.1.15)$$

onde $A_0 = I_{2M+1}$.

Koch e Yang (1986) referem que a distribuição assintótica de r_h^* depende dos valores próprios da matriz A_h e, portanto, de M e de h , o que constitui uma dificuldade para a obtenção do valor crítico que permite extrair a conclusão do teste. No entanto, este é fácil de efectuar a partir do

resultado estabelecido por estes autores que mostram que a distribuição aproximada de r_h^* é $\alpha_h \chi^2_{(v_h)}$, sendo:

$$v_h = \frac{[\text{tr}(A_h)]^2}{\text{tr}(A_h A_h)} \quad ; \quad \alpha_h = \frac{\text{tr}(A_h A_h)}{\text{tr}(A_h)} \quad (4.1.16)$$

Assim, é necessário obter v_h e α_h , para o que é necessário dispor de $\text{tr}(A_h)$ e de $\text{tr}(A_h A_h)$ cuja forma de cálculo é indicada por Koch e Yang (1986):

$$\text{tr}(A_h) = 2(M+1)(h+1) - (h+1)^2 \quad (4.1.17)$$

$$\text{tr}(A_h A_h) = 2 \sum_{n=1}^h [2M+h-3(n-1)]n^2 + (2M+1-2h)(h+1)^2 \quad h=0,1,2,\dots,M-1 \quad (4.1.18)$$

O teste conclui-se comparando o valor de $(1/\alpha_h)r_h^*$ com o ponto crítico adequado da distribuição $\chi^2_{(v_h)}$, para $h=0,1,2,\dots,M-1$.

Para a escolha do valor de h , no caso de não haver qualquer informação sobre a natureza da relação entre os dois processos e, nomeadamente, sobre o número de "lags" em que esta se verifica, é proposto o cálculo de r_h^* para $h=0,1,2,\dots,M-1$. No caso de se dispor de informação que indique L "lags" consecutivos, positivos ou negativos, na relação referida, o valor proposto é $h=L$. Por sua vez, o valor de M deve ser escolhido tal como no teste de Haugh.

Finalmente, é importante ainda referir que estudos de simulação efectuados por Koch e Yang (1986) revelam que, quando os dois processos estão fracamente relacionados ao longo de um número elevado de "lags", este teste mostra-se mais potente que o teste de Haugh.

2. Critérios gerais

Há várias características que o modelo estimado deve possuir para ser considerado um bom modelo e que são as seguintes:

(i) Critério de parcimónia: o modelo não deve incluir parâmetros desnecessários, o que é tanto mais importante quanto mais elaborado ele for. Este critério, que já é de fundamental importância no caso univariado, torna-se ainda mais relevante para modelos multivariados, uma vez que estes incluem geralmente um grande número de parâmetros. Os modelos parcimoniosos produzem habitualmente melhores previsões.

(ii) Invertibilidade e estacionaridade do modelo: os parâmetros estimados deverão obedecer a estas condições. Uma vez que os verdadeiros parâmetros do modelo são desconhecidos, esta condição tem que ser verificada utilizando as estimativas obtidas.

(iii) Qualidade dos parâmetros estimados: este critério engloba três condições.

A primeira é a da significância estatística dos parâmetros estimados, segundo a qual o modelo estimado não deve incluir parâmetros que não sejam significativos. Dado que a distribuição dos estimadores dos parâmetros é assintoticamente Normal, é possível testar a sua significância estatística da forma habitual.

A segunda condição é a inexistência de elevados valores da correlação entre parâmetros estimados, uma vez que, se se verificarem tais valores, as estimativas são de má qualidade. Esta condição é muito importante em modelos multivariados devido à existência de um grande número de parâmetros.

Finalmente, a terceira condição é a da inexistência de parâmetros redundantes, uma vez que, quando existe redundância entre os parâmetros, as estimativas são de má qualidade.

3. Análise dos resíduos

Se o modelo diagnosticado for adequado para a série cronológica analisada, os resíduos de estimação devem comportar-se como um ruído branco, sendo habitualmente propostas diversas formas para testar se isso se verifica:

(i) Significância estatística dos resíduos

O primeiro passo desta análise consiste em verificar se existem resíduos de elevado valor, para o que se compara os valores de $\hat{\epsilon}_i$ com $\pm 2\hat{C}_{\epsilon_i}(0)$, com $i=1, \dots, m$, sendo $\hat{C}_{\epsilon_i}(0)$ o i -ésimo elemento da diagonal principal da matriz $\hat{C}_{\epsilon}(0)$ definida em (3.4.11). Esta questão é de extrema importância, pois a existência de resíduos de elevado valor pode provocar sérias distorções na estrutura do modelo, nas estimativas obtidas para os parâmetros, nas funções de autocorrelação e de correlação cruzada residuais e nas previsões geradas a partir do modelo (Jenkins e Alavi, 1981).

Mas, as variáveis residuais ϵ_{it} ($i=1, \dots, m$) podem estar contemporaneamente correlacionadas, o que dificulta a realização de testes independentes para cada ϵ_{it} . Assim, designando por λ_i ($i=1, \dots, m$) o i -ésimo valor próprio da matriz $\hat{\Sigma}$ e por V a matriz dos vectores próprios associados, Jenkins e Alavi (1981) propõem a transformação $\epsilon'_i = V\epsilon_t$ para a realização de testes independentes, o que se torna possível porque, conforme indicam estes autores, ϵ'_i são não correlacionados. Sendo $V(\epsilon'_i) = \lambda_i$, utiliza-se $\pm 2\sqrt{\lambda_i}$ para testar à significância estatística dos resíduos transformados.

(ii) Análise das matrizes de correlação e de correlação parcial residuais estimadas

Uma vez que os resíduos de estimação se devem comportar como um ruído branco, os elementos das respectivas matrizes de correlação

estimadas não devem ser estatisticamente significativas a não ser para o "lag" zero, sendo habitualmente suficiente realizar este teste para os primeiros "lags". Se se concluir pela má qualidade do modelo ajustado, o exame das matrizes de correlação de correlação parcial residuais pode fornecer indicações sobre possíveis melhoramentos, a partir da identificação de um modelo para os resíduos e da sua combinação com o original, produzindo um novo modelo que será estimado e submetido a confirmação.

4. Testes de qualidade do ajustamento

Além das análises anteriores, é útil averiguar se o modelo se adequa convenientemente à série cronológica utilizando outros métodos, o que permite estudar outras facetas daquele que possam não ter sido ainda tomadas em conta na especificação adoptada.

O primeiro a ser apresentado será o método de sobreparametrização, ao que se seguirão diversos testes de natureza semelhante.

4.1 Sobreparametrização

É um método vulgarmente utilizado para testar a especificação de um modelo ARMA(p,q), sendo designado de "over-fitting" na terminologia anglo-saxónica e consistindo em ajustar um modelo mais elaborado e averiguar se este se adapta melhor à série em análise. Em particular, a significância estatística dos parâmetros adicionais é um critério que deve ser observado pelo modelo mais elaborado, uma vez que, se aqueles não forem significativos, podem ser retirados do modelo, permitindo depositar maior confiança no modelo inicial. Por outro lado, é importante não introduzir redundâncias no modelo devidas aos parâmetros adicionais.

Um princípio que pode ser contrariado por este método é o da parcimónia, o que obriga a um cuidado extremo na forma como se põe este procedimento em prática, não devendo adicionar-se simultaneamente parâmetros auto-regressivos e de médias móveis. Esta questão é de extrema relevância, pois um modelo multivariado inclui habitualmente um elevado número de parâmetros.

Por outro lado, os testes mais vulgarmente utilizados para averiguar a adequabilidade do(s) modelo(s) ajustado(s) a uma série

cronológica são geralmente classificados em duas categorias distintas que estão, no entanto, relacionadas. A primeira inclui testes globais que pretendem julgar se o modelo é ou não adequado sem comparar com qualquer modelo alternativo, constituindo generalizações do que sucede no caso univariado, e a segunda é composta por testes que visam averiguar se existe algum modelo alternativo que se adapte melhor à série do que o modelo identificado.

4.2 Testes globais

Hosking (1980a) generalizou ao caso multivariado a estatística de Box e Pierce (1970) que pretende testar a qualidade do ajustamento de um modelo a uma série univariada.

Tal como no caso univariado, o teste realiza-se com recurso à distribuição do qui-quadrado, sendo necessário utilizar a distribuição de $\text{vec}\hat{C}_\varepsilon$ definida no teorema 3.4.2.. Assim, surge o resultado (Hosking, 1980a):

Teorema 4.4.1. Uma vez que a matriz $(\mathbf{I}_R \otimes \Sigma \otimes \Sigma)$ é simétrica e definida positiva, onde R é um número inteiro positivo qualquer, então:

$$N(\text{vec}\hat{C}_\varepsilon)^T (\mathbf{I}_R \otimes \Sigma \otimes \Sigma)^{-1} \text{vec}\hat{C}_\varepsilon \sim \chi_{m^2(R-p-q)}^2$$

Mas, é impossível realizar o teste utilizando a estatística deste teorema, uma vez que $(\mathbf{I}_R \otimes \Sigma \otimes \Sigma)$ é desconhecida. No entanto, Hosking (1980a) apresenta o teorema seguinte, permitindo ultrapassar essa dificuldade:

Teorema 4.4.2. Se $(\mathbf{I}_R \otimes \Sigma \otimes \Sigma)^{-1}$ for substituído por um seu estimador consistente, ou seja, por $(\mathbf{I}_R \otimes \hat{\Sigma}^{-1} \otimes \hat{\Sigma}^{-1}) = (\mathbf{I}_R \otimes \hat{C}_\varepsilon^{-1}(0) \otimes \hat{C}_\varepsilon^{-1}(0))$, o resultado do teorema anterior não vem alterado:

$$N(\text{vec}\hat{C}_\varepsilon)^T (\mathbf{I}_R \otimes \hat{C}_\varepsilon^{-1}(0) \otimes \hat{C}_\varepsilon^{-1}(0)) \text{vec}\hat{C}_\varepsilon \sim \chi_{m^2(R-p-q)}^2$$

Deste modo:

$$Q_1 = N(\text{vec}\hat{C}_\varepsilon)^T (I_R \otimes \hat{C}_\varepsilon^{-1}(0) \otimes \hat{C}_\varepsilon^{-1}(0)) \text{vec}\hat{C}_\varepsilon \quad (4.4.1)$$

é a estatística a utilizar para a realização do teste, seguindo distribuição assintótica $\chi^2_{m^2(R-p-q)}$ sob a hipótese de adequabilidade do modelo, sendo esta rejeitada quando Q_1 assume valores elevados. É útil referir que no caso univariado ($m=1$), Q_1 se reduz à estatística de Box e Pierce (1970).

No entanto, para que a distribuição assintótica de Q_1 seja válida, é necessário que o valor de R obedeça às condições seguintes (Hosking 1980a):

1. $R=O(N^{1/2})$, o que significa que a distribuição assintótica de Q_1 é válida se $R/N \rightarrow 0$ quando $R, N \rightarrow \infty$.
2. $G_k, k > R-p$, e $H_k, k > R-q$, são $O(N^{-1/2})$, onde G_k e H_k são definidos em (3.4.15) e (3.4.16) respectivamente.

Em particular, se o valor de R não for pequeno relativamente ao de N , a condição 1. não se verifica. Deste modo, Poskitt e Tremayne (1986), propõem $R=[1+m(p+q)/2] \log N$ ou $R=[1+(p+q)/2] \log N$, tratando-se em ambos os casos de regras que verificam a condição 1., optando pela segunda por razões de dimensão e de potência do teste a partir dos resultados obtidos num estudo de simulação com base em modelos bivariados.

Por outro lado, é possível exprimir esta estatística sob uma forma alternativa que se revela muito útil para o cálculo do seu valor observado e que é (Hosking, 1980 a):

$$Q_1 = N \sum_{k=1}^R \text{tr}[\hat{C}_\varepsilon^T(k) \hat{C}_\varepsilon(0)^{-1} \hat{C}_\varepsilon^T(k) \hat{C}_\varepsilon(0)^{-1}] \quad (4.4.2)$$

o que é equivalente a (Hosking, 1981b):

$$Q_1 = N \sum_{k=1}^R [\text{vec}\hat{C}_\varepsilon(k)]^T [\hat{C}_\varepsilon(0)^{-1} \otimes \hat{C}_\varepsilon(0)^{-1}] \text{vec}\hat{C}_\varepsilon(k) \quad (4.4.3)$$

Mas, à semelhança do que se passa no caso univariado, a ocorrência de valores elevados Q_1 é menos frequente do que seria de esperar quando o número de observações disponíveis é pequeno ou moderado, tornando-se o teste pouco sensível a desvios em relação ao modelo. Isto deve-se ao facto de Q_1 só seguir distribuição $\chi_{m^2(R-p-q)}^2$ se a covariância entre os elementos de $C_\varepsilon(k)$ for dada por (3.4.13), uma vez que, para pequenos valores de N , é a expressão exacta (3.4.12) que deve ser utilizada, o que conduz à estatística modificada:

$$Q_1 = N \sum_{k=1}^R (N-k)^{-1} \text{tr}[\hat{C}_\varepsilon^T(k) \hat{C}_\varepsilon(0)^{-1} \hat{C}_\varepsilon(k) \hat{C}_\varepsilon(0)^{-1}] \quad (4.4.4)$$

sendo de esperar que, em pequenas amostras, Q_1 tenha uma distribuição mais próxima de $\chi_{m^2(R-p-q)}^2$ do que Q_1 . Estudos de simulação efectuados por Hosking (1980 a) apontam no sentido da correcção destas conclusões.

É também possível expressar a estatística do teste utilizando a matriz $\hat{R}_\varepsilon(k)$, tal como foi definida em (3.4.19), e recorrer à distribuição de $\text{vec}\hat{R}_\varepsilon(k)$ apresentada no teorema (3.4.4), obtendo-se (Hosking, 1981b):

$$Q_2 = N \sum_{k=1}^R [\text{vec}\hat{R}_\varepsilon(k)]^T [\hat{R}_\varepsilon(0)^{-1} \otimes \hat{R}_\varepsilon(0)^{-1}] \text{vec}\hat{R}_\varepsilon(k) \quad (4.4.5)$$

Mas Hosking (1981b) prova que Q_2 é igual a Q_1 , pelo que a distribuição da primeira coincide com a da segunda. Por outro lado, Li e McLeod (1981) propõem ainda a estatística:

$$Q_3 = N \sum_{k=1}^R [\text{vec}\hat{R}_\varepsilon^T(k)]^T [\hat{R}_\varepsilon(0)^{-1} \otimes \hat{R}_\varepsilon(0)^{-1}] \text{vec}\hat{R}_\varepsilon^T(k) \quad (4.4.6)$$

No entanto, Hosking (1981b) mostra que Q_2 e Q_3 são iguais, o que significa que esta última é também igual a Q_1 ⁽⁶⁾.

Tal como sucede com Q_1 , é conveniente proceder à modificação de Q_3 quando a dimensão da série for pequena, obtendo-se a estatística Q_3^1 , cuja distribuição assintótica é também $\chi_{m^2(R-p-q)}^2$:

$$Q_3^1 = Q_3 + \frac{m^2 R(R+1)}{2N} \quad (4.4.7)$$

Esta estatística é proposta por Li e McLeod (1981) que apresentam um estudo de simulação que sugere que Q_3^1 é de facto preferível para pequenos valores de N . Por outro lado, os resultados do estudo de simulação apresentados por Poskitt e Tremayne (1986) a que já se fez referência, indicam que a dimensão do teste realizado utilizando quer Q_1^1 quer Q_3^1 está de acordo com o que é definido em termos assintóticos, mesmo para valores de N razoavelmente pequenos, ou seja, Q_1^1 e Q_3^1 , ajustam-se satisfatoriamente às suas distribuições assintóticas, mesmo para pequenas dimensões da amostra.

4.3. Testes de multiplicadores de Lagrange

Para testar a especificação de um dado modelo ARMA(p,q) ajustado a uma série cronológica, um método que se afigura conveniente é o de sobreparametrização, tal como já foi referido.

Uma forma alternativa possível de realizar esse estudo consiste em estimar o modelo mais elaborado e utilizar o critério da razão de verosimilhança. No entanto, a estimação de modelos ARMA multivariados incluindo muitos parâmetros pode apresentar grandes dificuldades, o que é tanto mais importante quanto mais elaborado for o modelo alternativo.

Deste modo, a utilização de um teste de multiplicadores de Lagrange parece preferível, uma vez que a estimação do modelo alternativo pode assim ser evitada.

Estes testes são discutidos por Poskitt e Tremayne (1982) e por Hosking (1981a), constituindo generalizações dos testes propostos para o caso univariado pelos mesmos autores (Poskitt e Tremayne, 1980, e Hosking, 1981b, respectivamente). O objectivo é testar a hipótese de a especificação correcta ser a de um modelo ARMA(p,q) contra a alternativa de um modelo ARMA(p,q+S) ou de um modelo ARMA(p+S,q), sendo S um número inteiro positivo.

Os testes de multiplicadores de Lagrange baseiam-se nas derivadas do logaritmo da função de verosimilhança do modelo mais geral em ordem aos parâmetros, calculadas utilizando os estimadores de máxima verosimilhança sob a hipótese nula.

Tal como já foi referido, o logaritmo da função de verosimilhança de um modelo ARMA p,q) multivariado pode ser expresso na forma:

$$\log L = -\frac{mN}{2} \log(2\pi) - \frac{N}{2} \log |\Sigma| - \frac{1}{2} \sum_t \varepsilon_t^T \Sigma^{-1} \varepsilon_t \quad (4.4.8)$$

Assim (Granger e Newbold, 1986):

$$\frac{\partial \log L}{\partial \phi_{ij,l}} = \sum_t \varepsilon_t^T \Sigma^{-1} V_{ij,t-l} \quad (4.4.9)$$

onde $\phi_{ij,l}$ é o (i,j)-ésimo elemento de ϕ_l ($i,j=1,\dots,m; l=1,\dots,p$) e :

$$V_{ij,t} + \theta_1 V_{ij,t-1} + \dots + \theta_q V_{ij,t-q} = E_{ij} X_t \quad (4.4.10)$$

sendo E_{ij} uma matriz de dimensão (m x m) com o (ij)-ésimo elemento igual à unidade e todos os outros elementos nulos . Do mesmo modo:

$$\frac{\partial \log L}{\partial \theta_{ij,l}} = \sum_t \varepsilon_t^T \Sigma^{-1} U_{ij,t-l} \quad (4.4.11)$$

onde $\theta_{ij,l}$ é o (ij) -ésimo elemento de θ_l ($i,j=1,\dots,m; l=1,\dots,q$) e:

$$U_{ij,t} + \theta_1 U_{ij,t-1} + \dots + \theta_q U_{ij,t-q} = E_{ij} \varepsilon_t \quad (4.4.12)$$

Para realizar o teste, obtém-se os estimadores $\hat{V}_{ij,t}$ e $\hat{U}_{ij,t}$ substituindo em (4.4.10) e em (4.4.12) os parâmetros $\theta_1, \dots, \theta_q$ pelos seus estimadores de máxima verosimilhança sob a hipótese nula e as variáveis residuais ε_t pelos resíduos de estimação $\hat{\varepsilon}_t$.

Poskitt e Tremayne (1982) mostram que é possível derivar a estatística necessária à realização do teste através do ajustamento de uma regressão auxiliar. Se o objectivo for testar um modelo ARMA(p,q) contra um modelo ARMA(p,q+S), a estatística é obtida considerando a regressão do vector $\hat{\varepsilon}^T = (\hat{\varepsilon}_N^T, \hat{\varepsilon}_{N-1}^T, \dots)$ sobre:

$$\hat{Z} = \begin{bmatrix} \hat{V}_{N-1} & \dots & \hat{V}_{N-p} & \hat{U}_{N-1} & \dots & \hat{U}_{N-q-S} \\ \hat{V}_{N-2} & \dots & \hat{V}_{N-p-1} & \hat{U}_{N-2} & \dots & \hat{U}_{N-q-S-1} \\ \vdots & & \vdots & \vdots & & \vdots \end{bmatrix}$$

Sob a hipótese nula, a matriz de covariância residual desta regressão é consistentemente estimada por $(I_m \otimes \hat{\Sigma})$, sendo $\hat{\Sigma}$ o estimador de máxima verosimilhança da matriz de covariância das variáveis residuais do modelo ARMA(p,q).

Assim, a estatística para a realização do teste é:

$$LM_1 = \hat{\varepsilon}^T (I_m \otimes \hat{\Sigma}^{-1}) \hat{Z} \left[\hat{Z}^T (I_m \otimes \hat{\Sigma}^{-1}) \hat{Z} \right]^{-1} \hat{Z}^T (I_m \otimes \hat{\Sigma}^{-1}) \hat{\varepsilon} \quad (4.4.13)$$

Sob a hipótese nula, esta estatística tem a seguinte distribuição assintótica:

$$LM_1 \sim \chi_{(m^2 S)}^2 \quad (4.4.14)$$



sendo a hipótese nula rejeitada quando LM_1 assume valores elevados. Poskitt e Tremayne (1982) mostram ainda que, em termos assintóticos, o teste é idêntico quando a especificação proposta sob a hipótese alternativa é a de um modelo ARMA(p+S,q). Por outro lado, os resultados do estudo de simulação efectuado por Poskitt e Tremayne (1986) com base em modelos bivariados e já referido, indicam que a distribuição assintótica de LM_1 é válida mesmo para dimensões amostrais bastante pequenas.

Hosking (1981a) realiza o teste de multiplicadores de Lagrange considerando como alternativa não apenas modelos ARMA(p,q+S) ou ARMA(p+S,q), mas antes a especificação de modelos mais gerais que os incluem como casos particulares.

Sendo um modelo ARMA(p,q) definido da forma:

$$\phi(B)X_t = \theta(B)\varepsilon_t \quad (4.4.15)$$

toma-se como hipótese nula:

$$H_0 : \{\varepsilon_t\} \text{ é um ruído branco} \quad (4.4.16)$$

e como alternativa:

$$H_1 : \varepsilon_t + \sum_{r=1}^s \Xi(B)\Lambda_r Y(B)\varepsilon_{t-r} = \hat{\varepsilon}_t, \quad \{\hat{\varepsilon}_t\} \text{ é um ruído branco} \quad (4.4.17)$$

Esta hipótese alternativa consiste no ajustamento de um modelo auto-regressivo ao processo $\{\varepsilon_t\}$, sendo as matrizes de dimensão (m x m) $\Lambda_1, \dots, \Lambda_s$ os parâmetros adicionais do modelo alternativo. Os polinómios matriciais $\Xi(B)$ e $Y(B)$ devem incluir apenas potências não negativas de B, sendo as matrizes que os compõem funções de ϕ_r e de θ_r . Além disso, as raízes de $|\Xi(B)|=0$ e de $|Y(B)|=0$ devem situar-se fora do círculo unitário e $\Xi(B)$ e $Y(B)$ devem ser independentes de Λ_r , sendo ainda necessário que as suas derivadas em ordem a ϕ_r e a θ_r sejam finitas e limitadas quando

$N \rightarrow \infty$ e que Ξ_r e Y_r converjam exponencialmente para zero quando $r \rightarrow \infty$.

Considere-se ainda as matrizes:

$$F_r = \sum_{u=0}^r \Sigma Y_{r-u}^T \otimes \Xi_u \quad (4.4.18)$$

$$J_r = \sum_{u=0}^r \Sigma \Psi_{r-u}^T \otimes \pi_u \quad (4.4.19)$$

$$H_r = \Sigma \otimes \pi_r \quad (4.4.20)$$

sendo $\Psi(B) = \phi(B)^{-1} \theta(B) = \sum_{v=0}^{\infty} \psi_v B^v$ e $\pi(B) = \theta(B)^{-1} = \sum_{v=0}^{\infty} \pi_v B^v$. Seja ainda $\phi = (\phi_1, \dots, \phi_p)$ e $\theta = (\theta_1, \dots, \theta_q)$.

Nestas condições, Hosking (1981a) apresenta o seguinte teorema, que constitui a generalização do resultado correspondente para o caso univariado:

Teorema 4.4.3. Para o teste da hipótese especificada em (4.4.16) contra a alternativa especificada em (4.4.17), o estimador da matriz dos multiplicadores de Lagrange correspondentes a Λ_r é:

$$\hat{D}_r = N \sum_{u=0}^{\infty} \sum_{v=0}^{\infty} \hat{\Xi}_v^T \hat{\Sigma}^{-1} \hat{C}_\epsilon (r+u+v) Y_u^T$$

Sob H_0 , a distribuição assintótica de \hat{D}_r é:

$$N^{-1/2} \text{vec} \hat{D} \sim N(0; A - KB^{-1}K^T)$$

onde:

$$\hat{D} = (\hat{D}_1, \dots, \hat{D}_s);$$

A é uma matriz de dimensão $(m^2S \times m^2S)$ cujo bloco de ordem (r, u) é a matriz de dimensão $(m^2 \times m^2)$:

$$\sum_{v=0}^{\infty} F_{v-r}^T (\Sigma^{-1} \otimes \Sigma^{-1}) F_{v-u};$$

$\mathbf{K}=[\mathbf{K}_1 \ \mathbf{K}_2]$, sendo \mathbf{K}_1 de dimensão $(m^2S \times m^2p)$ e \mathbf{K}_2 de dimensão $(m^2S \times m^2q)$, cujos blocos de ordem (r,u) são respectivamente as matrizes de dimensão $(m^2 \times m^2)$:

$$\sum_{v=0}^{\infty} \mathbf{F}_{v-r}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \mathbf{J}_{v-u};$$

e

$$\sum_{v=0}^{\infty} \mathbf{F}_{v-r}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \mathbf{H}_{v-u};$$

\mathbf{B} é a matriz de informação para $\begin{bmatrix} \text{vec}\phi \\ \text{vec}\theta \end{bmatrix}$ sob H_0 , ou seja:

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{12}^T & \mathbf{B}_{22} \end{bmatrix}$$

sendo \mathbf{B}_{11} uma matriz de dimensão $(m^2p \times m^2p)$, \mathbf{B}_{12} de dimensão $(m^2p \times m^2q)$ e \mathbf{B}_{22} de dimensão $(m^2q \times m^2q)$, cujos blocos de ordem (r,u) são respectivamente as matrizes de dimensão $(m^2 \times m^2)$:

$$\sum_{v=0}^{\infty} \mathbf{J}_{v-r}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \mathbf{J}_{v-u};$$

$$\sum_{v=0}^{\infty} \mathbf{J}_{v-r}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \mathbf{H}_{v-u};$$

e

$$\sum_{v=0}^{\infty} \mathbf{H}_{v-r}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \mathbf{H}_{v-u};$$

É importante referir que existe um limite superior para S , sendo tal que $S=O(N^{1/2})$.

A hipótese alternativa H_1 especificada em (4.4.17) consiste no ajustamento de um modelo auto-regressivo ao processo $\{\mathcal{E}_t\}$, conforme foi referido, sendo também possível ajustar-lhe um modelo de médias móveis, o que conduz à especificação alternativa:

$$\varepsilon_t = \dot{\varepsilon}_t + \sum_{r=1}^s \Xi(B)\Lambda_r Y(B)\dot{\varepsilon}_{t-r}, \quad \{\dot{\varepsilon}_t\} \text{ é um ruído branco} \quad (4.4.21)$$

Generalizando o que se passa no caso univariado, Hosking (1981a) refere que o teste de H_0 contra H_1 ou de H_0 contra H_1^* são idênticos, embora os sinais das matrizes \hat{D} e \hat{K} sejam opostos.

Assim, sendo as matrizes $P=A-KB^{-1}K^T$, a estatística a utilizar em ambos os casos é:

$$LM_2 = N^{-1} (\text{vec}\hat{D})^T \hat{P}^{-1} \text{vec}\hat{D} \quad (4.4.22)$$

onde \hat{P} é um estimador consistente de P .

Tal como LM_1 , a estatística LM_2 tem a seguinte distribuição sob H_0 :

$$LM_2 \sim \chi_{(m^2s)}^2 \quad (4.4.23)$$

rejeitando-se H_0 quando LM_2 assume valores elevados.

Os modelos correspondentes quer a H_1 quer a H_1^* podem ser expressos de diversas formas. Assim, o modelo especificado em (4.4.17) é equivalente a:

$$\theta(B)\varepsilon_t + \sum_{r=1}^s \theta(B)\Xi(B)\Lambda_r Y(B)\varepsilon_{t-r} = \theta(B)\dot{\varepsilon}_t \quad (4.4.24)$$

ou seja:

$$\phi(B)X_t + \sum_{r=1}^s \theta(B)\Xi(B)\Lambda_r Y(B)\pi(B)\phi(B)X_{t-r} = \theta(B)\dot{\varepsilon}_t \quad (4.4.25)$$

que é o modelo especificado na hipótese nula acrescido de uma parcela no membro auto-regressivo.

Do mesmo modo, (4.4.21) é equivalente a:

$$\phi(B)X_t = \theta(B)\dot{\varepsilon}_t + \sum_{r=1}^s \theta(B)\Xi(B)\Lambda_r Y(B)\dot{\varepsilon}_{t-r} \quad (4.4.26)$$

Tal como já foi referido, (4.4.24) e (4.4.25) são expressões muito gerais, permitindo a especificação de uma vasta gama de modelos. Assim, se se definir:

$$\begin{aligned}\Xi(B) &= \pi(B) \\ Y(B) &= B^p \psi(B)\end{aligned}\tag{4.4.27}$$

a expressão (4.4.25) vem:

$$\phi(B)X_t + \sum_{r=1}^s \Lambda_r X_{t-p-r} = \theta(B)\dot{\epsilon}_t\tag{4.4.28}$$

Portanto, se $\Xi(B)$ e $Y(B)$ forem definidas por (4.4.27), o teste de H_0 contra H_1 significa testar a hipótese de a especificação correcta ser a de um modelo ARMA(p,q) contra a de ser um modelo ARMA(p+S,q).

Por outro lado, se se definir:

$$\begin{aligned}\Xi(B) &= B^q \pi(B) \\ Y(B) &= I_m\end{aligned}\tag{4.4.29}$$

a expressão (4.4.26) vem:

$$\phi(B)X_t = \theta(B)\dot{\epsilon}_t + \sum_{r=1}^s \Lambda_r \dot{\epsilon}_{t-q-r}\tag{4.4.30}$$

Deste modo, conclui-se que, quando $\Xi(B)$ e $Y(B)$ são definidos por (4.4.29), o modelo especificado em H_1^* é um modelo ARMA(p,q+S).

Em ambos os casos, e tal como já acontecia com o método proposto por Poskitt e Tremayne (1982), a estatística a utilizar para o teste é a mesma, ou seja, utiliza-se LM_2 quer a hipótese alternativa seja H_1 quer seja H_1^* .

É importante referir ainda que a estatística utilizada para testar a qualidade de ajustamento do modelo, designada por Q_1 e definida em (4.4.1), pode ser obtida a partir do teste de multiplicadores de Lagrange. De facto, se se tiver $\Xi(B)=Y(B)=I_m$ em (4.4.17), é possível estabelecer os resultados:

$$\hat{D}_r = -N\hat{\Sigma}^{-1}\hat{C}_\varepsilon(r); \quad (4.4.31)$$

$$\text{vec}\hat{D}_r = -N(I_m \otimes \hat{\Sigma}^{-1})\text{vec}\hat{C}_\varepsilon(r); \quad (4.4.32)$$

$$\text{vec}\hat{D} = -N(I_s \otimes I_m \otimes \hat{\Sigma}^{-1})\text{vec}\hat{C}_\varepsilon; \quad (4.4.33)$$

sendo neste caso $\hat{C}_\varepsilon = [\hat{C}_\varepsilon(1) \dots \hat{C}_\varepsilon(S)]$;

$$A = I_s \otimes \Sigma \otimes \Sigma^{-1}; \quad (4.4.34)$$

$$K_1 = (I_s \otimes I_m \otimes \hat{\Sigma}^{-1}) \begin{bmatrix} J_0 & 0 & 0 & \dots & 0 \\ J_1 & J_0 & 0 & \dots & 0 \\ J_2 & J_1 & J_0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ J_{s-1} & J_{s-2} & J_{s-3} & \dots & J_{s-p} \end{bmatrix} \quad (4.4.35)$$

$$K_2 = -(I_s \otimes I_m \otimes \hat{\Sigma}^{-1}) \begin{bmatrix} H_0 & 0 & 0 & \dots & 0 \\ H_1 & H_0 & 0 & \dots & 0 \\ H_2 & H_1 & H_0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ H_{s-1} & H_{s-2} & H_{s-3} & \dots & H_{s-p} \end{bmatrix} \quad (4.4.36)$$

Neste contexto, admitindo que:

$$\begin{aligned} J_r &= O(N^{-1/2}) \text{ para } r > S-p \\ H_r &= O(N^{-1/2}) \text{ para } r > S-q \end{aligned} \quad (4.4.37)$$

Hosking (1981a) mostra que a distribuição assintótica de $\text{vec}\hat{D}$ é:

$$N^{-1/2}\text{vec}\hat{D} \sim (0; \{I_{m^2s} - K(K^T A^{-1} K)^{-1} K^T A^{-1}\} A) \quad (4.4.38)$$

sendo $I_{m^2s} - K(K^T A^{-1} K)^{-1} K^T A^{-1}$ uma matriz idempotente de característica $m^2(S-p-q)$ e A uma matriz simétrica definida positiva, tem-se o seguinte resultado assintótico:

$$N^{-1}(\text{vec}\hat{D})^T \hat{A}^{-1} \text{vec}\hat{D} \sim \chi_{m^2(s-p-q)}^2 \quad (4.4.39)$$

Finalmente (Hosking, 1981a), daqui resulta:

$$\begin{aligned} N^{-1}(\text{vec}\hat{D})^T \hat{A}^{-1} \text{vec}\hat{D} &= \\ &= N(\text{vec}\hat{C}_\varepsilon)^T (\mathbf{I}_s \otimes \mathbf{I}_m \otimes \hat{\Sigma}^{-1}) (\mathbf{I}_s \otimes \hat{\Sigma}^{-1} \otimes \hat{\Sigma}) (\mathbf{I}_s \otimes \mathbf{I}_m \otimes \hat{\Sigma}^{-1}) \text{vec}\hat{C}_\varepsilon = \\ &= N(\text{vec}\hat{C}_\varepsilon)^T (\mathbf{I}_s \otimes \hat{\Sigma}^{-1} \otimes \hat{\Sigma}^{-1}) \text{vec}\hat{C}_\varepsilon \end{aligned} \quad (4.4.40)$$

Uma vez que $\hat{\Sigma} = \hat{C}_\varepsilon(0)$, (4.4.40) é igual à expressão de Q_1 definida em (4.4.1), ou seja, o teste de multiplicadores de Lagrange da hipótese nula definida em (4.4.16) contra a alternativa (4.4.17) é assintoticamente equivalente ao teste global realizado a partir de (4.4.1) e utilizando S covariâncias residuais.

A hipótese alternativa resultante deste teste, quando $\Xi(B) = Y(B) = \mathbf{I}_m$, pode ser encarada de diversas formas. A partir de (4.4.17), a hipótese alternativa é a de um modelo auto-regressivo de ordem S ajustado a $\{\varepsilon_t\}$ enquanto, a partir de (4.4.21), é a de um modelo de médias móveis de ordem S . Por outro lado, a hipótese alternativa definida em (4.4.26) especifica o modelo ARMA($p+q, S$):

$$\phi(B)X_t = \theta(B)\Lambda(B)\hat{\varepsilon}_t \quad (4.4.41)$$

sendo $\Lambda(B) = \mathbf{I}_m + \sum_{r=1}^S \Lambda_r B^r$.

No entanto, a hipótese alternativa não corresponde a nenhum modelo ARMA($p+S, q$) excepto se $q=0$, do que resulta para (4.4.25):

$$\Lambda(B)\phi(B)X_t = \hat{\varepsilon}_t \quad (4.4.42)$$

o que corresponde a um modelo AR($p+S$). Para que o modelo especificado na hipótese alternativa seja o modelo ARMA($p+S, q$):

$$\Lambda(B)\phi(B)X_t = \theta(B)\hat{\varepsilon}_t \quad (4.4.43)$$

seria necessário que, em (4.4.25) se definisse:

$$\Xi(B) = \pi(B)Y(B) = \theta(B) \quad (4.4.44)$$

No estudo de simulação com base em modelos bivariados a que já se fez referência, Poskitt e Tremayne (1986) procedem à comparação da dimensão e da potência dos testes globais e do teste de multiplicadores de Lagrange, sendo os resultados obtidos de molde a sugerir que este último revela características superiores, o que o tornaria preferível na detecção da má especificação de um dado modelo. Estes autores referem ainda que os resultados obtidos se devem ao facto de a distribuição assintótica das estatísticas utilizadas nos testes globais requerer que os respectivos graus de liberdade aumentem sem limite, à medida que N tende para o infinito, embora mais lentamente, o que não se verifica para a estatística utilizada no teste de multiplicadores de Lagrange. Se, pelo contrário, o número de parâmetros adicionais especificado na hipótese alternativa deste segundo tipo de testes aumentar da mesma forma que o número de correlações utilizadas nos testes globais, os dois tipos de testes tornam-se equivalentes. Assim, as diferenças na dimensão e na potência de ambos os tipos de testes só podem ser devidas a variações nos graus de liberdade das distribuições utilizadas, embora, tal como Poskitt e Tremayne (1986) afirmam, a dimensão do teste de multiplicadores de Lagrange não seja grandemente afectada pelo aumento do número de parâmetros adicionais do modelo especificado na hipótese alternativa, o que seria apoiado por alguma evidência empírica.

4.4. Teste informal

Além de todos os testes já propostos, é útil averiguar a qualidade do ajustamento de um modelo através de um procedimento informal

proposto por Granger e Newbold (1986). Esse método consiste em comparar os modelos univariados para os elementos de X_t , implícitos no modelo ARMA(p,q) multivariado, com os modelos resultantes da análise de cada série isoladamente. Na base desta comparação está a suposição de que é mais fácil diagnosticar correctamente um modelo para uma série univariada do que para uma multivariada. Apesar de este procedimento não envolver testes de hipóteses formais, pode detectar falhas ainda não detectadas na especificação do modelo.

Vejamus então como é possível obter os modelos univariados para os diversos elementos componentes de X_t a partir do modelo ARMA(p,q) multivariado:

$$\phi(B)X_t = \theta(B)\varepsilon_t \quad (4.4.45)$$

Daqui resulta:

$$|\phi(B)|X_t = \phi^A(B)\theta(B)\varepsilon_t \quad (4.4.46)$$

onde $\phi^A(B)$ é a matriz adjunta de $\phi(B)$, tendo-se $\phi(B)^{-1} = |\phi(B)|^{-1}\phi^A(B)$. O operador do primeiro membro de (4.4.6) é um polinómio escalar de grau mp em B, sendo o operador auto-regressivo de cada um dos modelos univariados; por sua vez, a matriz $\phi^A(B)\theta(B)$ é de grau (m-1)p+q em B. Assim, a ordem do operador de médias móveis de cada modelo univariado é (m-1)p+q. Deste modo, conclui-se que, se X_t for gerado por um modelo ARMA(p,q) multivariado, os seus elementos individuais X_{it} , $i=1, \dots, m$, são gerados por modelos ARMA(mp, (m-1)p+q) univariados assumindo a forma:

$$|\phi(B)|X_{it} = \theta_i(B)\varepsilon_{it} \quad i=1, \dots, m \quad (4.4.47)$$

onde $\{\varepsilon_{it}\}$ é um ruído branco e os (m-1)p+q parâmetros de médias móveis incluídos no polinómio $\theta_i(B)$ são funções das matrizes que compõem os

operadores auto-regressivo e de médias móveis do modelo multivariado (4.4.45).

A expressão (4.4.47) parece indicar que os operadores auto-regressivos de cada modelo univariado são iguais. No entanto, isso não acontece porque $|\phi(B)|$ possui habitualmente factores comuns, ou quase comuns, a $\theta_i(B)$.

A título de exemplo, considere-se o modelo AR(1) bivariado:

$$\phi(B)X_t = \varepsilon_t \quad (4.4.48)$$

ou seja:

$$(I_2 - \phi_2 B)X_t = \varepsilon_t \quad (4.4.49)$$

É possível expressar (4.4.49) da forma alternativa:

$$\begin{aligned} X_{1t} &= \phi_{11,1} X_{1,t-1} + \phi_{12,1} X_{2,t-1} + \varepsilon_{1t} \\ X_{2t} &= \phi_{21,1} X_{1,t-1} + \phi_{22,1} X_{2,t-1} + \varepsilon_{2t} \end{aligned} \quad (4.4.50)$$

Multiplicando ambos os membros de (4.4.49) pela matriz adjunta de $\phi(B)$, obtém-se o resultado:

$$\left[(1 - \phi_{11,1} B)(1 - \phi_{22,1} B) - \phi_{12,1} \phi_{21,1} B^2 \right] \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} 1 - \phi_{22,1} B & -\phi_{12,1} B \\ -\phi_{21,1} B & 1 - \phi_{11,1} B \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \quad (4.4.51)$$

Conclui-se então que cada série univariada é modelizada por um modelo ARMA(2,1) univariado. No entanto, note-se que essa é a ordem máxima para cada série individual e que os operadores auto-regressivos não são necessariamente iguais, o que poderia parecer da observação de (4.4.51). Por exemplo, se $\phi_{12,1} = \phi_{21,1} = 0$, cada série univariada obedeceria a um modelo AR(1).

Em conclusão, este procedimento consiste em construir os modelos univariados implícitos para cada série após o ajustamento do modelo

multivariado e compará-los com os que seriam diagnosticados através da análise de cada série univariada. Esta comparação é informal mas, admitindo maior capacidade na modelização de uma série univariada, pode conduzir à detecção de pontos fracos no modelo multivariado diagnosticado.

V. PREVISÃO

Um dos principais objectivos da modelização de uma série cronológica consiste em prever (ou predizer) os seus valores futuro a partir dos valores disponíveis em dado momento, sendo este um problema de grande importância. Assim, dispondo das observações X_t , X_{t-1} , X_{t-2}, \dots , pretende-se prever no momento t o valor da série l períodos à frente, ou seja, X_{t+l} . Deste modo, a questão consiste na escolha da função que permite obter o previsor a partir das observações disponíveis. Para o efeito, serão apresentados dois métodos: o primeiro consiste numa generalização ao caso multivariado do método de Box e Jenkins (1976) e o segundo consiste na abordagem de Granger e Newbold (1986). Enquanto o primeiro utiliza como critério para a obtenção do previsor a minimização do determinante da matriz de covariância do erro de previsão, o segundo baseia-se na definição no conceito de menor matriz de covariância desse erro. Ambos pretendem chegar a um previsor linear óptimo com base no critério adoptado, verificando-se no entanto que o previsor obtido é o mesmo, o que significa que existe equivalência entre os dois critérios para gerar previsões.

1. Extensão do método de Box e Jenkins

Esta abordagem constitui uma generalização ao caso multivariado do método proposto por Box e Jenkins (1976) para o caso univariado, pretendendo esta última minimizar o erro quadrático médio do previsor.

Assim, designando por $\mathbf{e}_t(l)$ o vector do erro de previsão, tem-se a relação:

$$\mathbf{e}_t(l) = \mathbf{X}_{t+l} - \hat{\mathbf{X}}_t(l) \quad (5.1.1)$$

onde $\hat{\mathbf{X}}_t(l)$ designa o previsor em que a origem da previsão é o momento t e l o número de passos da previsão.

Por outro lado, define-se ainda u_{t+l} como a combinação linear dos valores futuros das componentes da série multivariada para o momento $t+l$, assumindo a expressão:

$$u_{t+l} = w_1 X_{1,t+l} + w_2 X_{2,t+l} + \dots + w_m X_{m,t+l} \quad (5.1.2)$$

O erro de previsão de u_{t+l} é:

$$g_t(l) = u_{t+l} - \hat{u}_t(l) = \sum w_i \mathbf{e}_{it}(l) = \mathbf{W}^T \mathbf{e}_t(l) \quad (5.1.3)$$

onde $\mathbf{e}_{it}(l) = X_{i,t+l} - \hat{X}_{it}(l)$, $i=1, \dots, m$ e $\mathbf{W}^T = (w_1, \dots, w_m)$.

O erro quadrático médio de $g_t(l)$ assume a forma:

$$E[g_t^2(l)] = E[\mathbf{W}^T \mathbf{e}_t(l) \mathbf{e}_t^T(l) \mathbf{W}] = \mathbf{W}^T \mathbf{V}(l) \mathbf{W} \quad (5.1.4)$$

onde $\mathbf{V}(l) = E[\mathbf{e}_t(l) \mathbf{e}_t^T(l)]$ é designada por matriz de covariância do vector do erro de previsão $\mathbf{e}_t(l)$ para um número de passos de previsão de l , sendo o seu elemento genérico $\text{cov}[\mathbf{e}_{it}(l), \mathbf{e}_{jt}(l)]$, $i, j=1, \dots, m$. Uma vez que $\mathbf{V}(l)$ é uma matriz definida positiva, a minimização do erro quadrático médio de $g_t(l)$ é equivalente à minimização do determinante

da matriz de covariância do vector do erro de previsão $|V(l)|$. É este o critério adoptado por este método para a determinação do previsor (Jenkins e Alavi, 1981).

Assim, representando um modelo ARMA(p,q) invertível e estacionário como um modelo de médias móveis de ordem infinita:

$$X_{t+l} = \psi(B) \varepsilon_{t+l} = \sum_{u=0}^{\infty} \psi_u \varepsilon_{t+l-u} \quad (5.1.5)$$

prova-se, generalizando a abordagem de Box e Jenkins (1976) ao caso multivariado, que o previsor que minimiza $|V(l)|$ é (Jenkins e Alavi, 1981):

$$\hat{X}_t(l) = \psi_l \varepsilon_t + \psi_{l+1} \varepsilon_{t-1} + \dots \quad (5.1.6)$$

ou seja, é o valor esperado condicionado de X_{t+l} tomado na origem da previsão de t:

$$\hat{X}_t(l) = E(X_{t+l} | X_t, X_{t-1}, \dots) \quad (5.1.7)$$

Deste modo, o vector de erro de previsão é:

$$e_t(l) = \varepsilon_{t+l} + \psi_l \varepsilon_{t+l-1} + \dots + \psi_{l+1} \varepsilon_{t+1} \quad (5.1.8)$$

assumindo a sua matriz de covariância a expressão:

$$V(l) = \Sigma + \psi_l \Sigma \psi_l^T + \dots + \psi_{l+1} \Sigma \psi_{l+1}^T \quad (5.1.9)$$

Tal como no caso univariado, as previsões são facilmente geradas a partir da equação que define um modelo ARMA(p,q) multivariado invertível e estacionário, que pode ser expresso da forma:

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (5.1.10)$$

admitindo-se que os vectores ε_t são independentes.

No momento $t+1$, tem-se:

$$X_{t+1} - \phi_1 X_t - \dots - \phi_p X_{t+1-p} = \varepsilon_{t+1} + \theta_1 \varepsilon_t + \dots + \theta_q \varepsilon_{t+1-q} \quad (5.1.11)$$

Tomando valores esperados condicionados em ambos os membros, obtém-se o resultado:

$$\hat{X}_t(1) = \phi_1 X_t - \dots - \phi_p X_{t+1-p} = E^*(\varepsilon_{t+1}) + \theta_1 E^*(\varepsilon_t) + \dots + \theta_q E^*(\varepsilon_{t+1-q}) \quad (5.1.12)$$

onde E^* designa o valor esperado condicionado em X_t, X_{t-1}, \dots , sendo de notar que $E^*(X_{t-u}) = X_{t-u}$, para $u \geq 0$. Mas, uma vez que se pode exprimir X_t como uma função linear de $\varepsilon_t, \varepsilon_{t-1}, \dots$, e vice-versa, os valores esperados condicionados por X_t, X_{t-1}, \dots , são iguais aos valores esperados condicionados por $\varepsilon_t, \varepsilon_{t-1}, \dots$. Assim, tem-se o resultado:

$$E^*(\varepsilon_{t+1}) = E^*(\varepsilon_{t+1} | \varepsilon_t, \varepsilon_{t-1}, \dots) = 0 \quad (5.1.13)$$

uma vez que se admite que os vectores ε_t são independentes. Por outro lado, verifica-se ainda:

$$E^*(\varepsilon_{t-u}) = \varepsilon_{t-u} \quad u \geq 0 \quad (5.1.14)$$

Substituindo (5.1.13) e (5.1.14) em (5.1.12), obtém-se:

$$\hat{X}_t(1) - \phi_1 X_t - \dots - \phi_p X_{t+1-p} = \theta_1 \varepsilon_t + \dots + \theta_q \varepsilon_{t+1-q} \quad (5.1.15)$$

Comparando (5.1.11) e (5.1.13), chega-se à conclusão:

$$X_{t+1} - \hat{X}_t(1) = \varepsilon_{t+1} \quad (5.1.16)$$

De igual modo, tem-se as relações:

$$\begin{aligned}
\varepsilon_t &= X_t - \hat{X}_{t-1}(1) \\
\varepsilon_{t-1} &= X_{t-1} - \hat{X}_{t-2}(1) \\
\varepsilon_{t-2} &= X_{t-2} - \hat{X}_{t-3}(1) \\
&\vdots
\end{aligned}
\tag{5.1.17}$$

Substituindo estas expressões em (5.1.15), obtém-se o resultado:

$$\hat{X}_t(1) - \phi_1 X_t - \dots - \phi_p X_{t+1-p} = \theta_1 \{X_t - \hat{X}_{t-1}(1)\} + \dots + \theta_q \{X_{t+1-q} - \hat{X}_{t-q}(1)\}
\tag{5.1.18}$$

ou seja:

$$\begin{aligned}
&\hat{X}_t(1) + \theta_1 \hat{X}_{t-1}(1) + \dots + \theta_q \hat{X}_{t-q}(1) = \\
&= (\theta_1 + \phi_1) X_t + \dots + (\theta_q + \phi_q) X_{t+1-q} + \phi_{q+1} X_{t-q} + \dots + \phi_p X_{t+1-p}
\end{aligned}
\tag{5.1.19}$$

admitindo-se que $p \geq q$.

A equação (5.1.19) permite calcular $\hat{X}_t(1)$ recursivamente, dados os valores de $\hat{X}_{t-1}(1), \hat{X}_{t-2}(1), \dots, \hat{X}_{t-q}(1), X_t, X_{t-1}, \dots, X_{t+1-p}$. Esta equação permite ainda actualizar permanentemente a previsão a um passo, incorporando a observação que se torna disponível em cada período.

De um modo geral, obtém-se $\hat{X}_t(l)$ recursivamente a partir de $\hat{X}_t(1), \hat{X}_t(2), \dots, \hat{X}_t(l-1)$ e das observações X_t, X_{t-1}, \dots , substituindo-se na equação definidora do modelo ARMA(p,q) os valores futuros X_{t+l} pelos seus previsores $\hat{X}_t(l)$ e os valores futuros ε_{t+l} pelos seus valores esperados condicionados, ou seja, pelo vector nulo (Jenkins e Alavi, 1981). Isto significa que, na prática, o vector das previsões $\hat{X}_t(l)$ é obtido recursivamente a partir da expressão:

$$\hat{X}_t(l) = \phi_1 E^*(X_{t+l-1}) + \dots + \phi_p E^*(X_{t+l-p}) + E^*(\epsilon_{t+l}) + \theta_1 E^*(\epsilon_{t+l-1}) + \dots + \theta_q E^*(\epsilon_{t+l-q}) \quad (5.1.20)$$

onde:

$$E^*(X_{t+u}) = \begin{cases} \hat{X}_t(u) & \text{se } u \geq 1 \\ X_{t+u} & \text{se } u \leq 0 \end{cases} \quad \text{e} \quad E^*(\epsilon_{t+u}) = \begin{cases} 0 & \text{se } u \geq 1 \\ \epsilon_{t+u} & \text{se } u \leq 0 \end{cases} \quad (5.1.21)$$

Para a necessária definição dos valores iniciais, é possível generalizar a proposta de Box e Jenkins (1976), igualando a 0 , ou seja, ao seu valor esperado não condicionado, todos os vectores ϵ_t que não possam ser obtidos a partir das observações, isto é, aqueles para os quais os vectores X_t e $\hat{X}_{t-1}(l)$ correspondentes não estão disponíveis. Por outro lado, uma vez que as matrizes de parâmetros são desconhecidas, na prática é necessário substituí-las pelos seus estimadores em (5.1.20) de modo a tornar possível a geração das previsões. Esta observação também é aplicável ao método seguinte.

2. Método de Granger e Newbold

Granger e Newbold (1986) utilizam um critério diferente para a obtenção do previsor, partindo da hipótese de conhecimento exacto do modelo ARMA(p,q) que serve para a determinação desse previsor.

O objectivo consiste agora em determinar o previsor que minimize a matriz de covariância do erro de previsão. Para o efeito, sendo V_1 e V_2 duas matrizes reais definidas positivas de dimensão ($m \times m$), diz-se, segundo a definição de Granger e Newbold (1986), que V_1 é menor do que V_2 se, para qualquer vector d não nulo de dimensão ($m \times 1$), se tiver a relação:

$$d^T V_1 d < d^T V_2 d \quad (5.2.1)$$

Esta relação será designada por $V_1 < V_2$. Se $\{V(s)\}$ representar um conjunto de matrizes definidas positivas de dimensão ($m \times m$), uma dada matriz desse conjunto V_0 , é definida como a menor se $V_0 < V_r$ para qualquer $V_r \neq V_0$ contida em $\{V(s)\}$ (Granger e Newbold, 1986, pág. 228).

Assim, sendo o modelo ARMA(p,q) invertível e estacionário :

$$\phi(B)X_t = \theta(B)\varepsilon_t \quad (5.2.2)$$

a sua representação como um modelo MA de ordem infinita, conforme já foi referido, assume a forma:

$$X_t = \sum_{u=0}^{\infty} \psi_u \varepsilon_{t-u} \quad (\psi_0 = I_m) \quad (5.2.3)$$

sendo:

$$\theta(B) = \phi(B)\psi(B) \quad (5.2.4)$$

onde:

$$\phi(B) = I_m - \sum_{u=1}^p \phi_u B^u; \quad \theta(B) = I_m + \sum_{u=1}^q \theta_u B^u; \quad \psi(B) = I_m + \sum_{u=1}^{\infty} \psi_u B^u \quad (5.2.5)$$

Assim, igualando os coeficientes de B^u em (5.2.4), obtém-se a relação:

$$\psi_u - \sum_{r=1}^u \phi_r \psi_{u-r} = \theta_u \quad u=1,2,\dots \quad (5.2.6)$$

Deste modo, a previsão X_{t+l} feita no momento t é baseada em X_t, X_{t-1}, \dots , sendo então da forma:

$$\hat{X}_t(l) = \sum_{u=0}^{\infty} \lambda_{u,l} X_{t-u} \quad (5.2.7)$$

onde $\lambda_{u,l}$ é uma matriz de dimensão $(m \times m)$. Conclui-se então que (5.2.7) pode ser expresso da forma:

$$\hat{X}_t(l) = \lambda_l(B) X_t \quad (5.2.8)$$

o que é equivalente a:

$$\hat{X}_t(l) = \alpha_l(B) \varepsilon_t \quad (5.2.9)$$

sendo:

$$\alpha_l(B) = \lambda_l(B) \psi(B) \quad (5.2.10)$$

Por outro lado, o vector do erro de previsão assume a forma:

$$e_{t,l} = X_{t+l} - \hat{X}_t(l) \quad (5.2.11)$$

com matriz de covariância:

$$V(l) = E(e_{t,l} e_{t,l}^T) \quad (5.2.12)$$

uma vez que $E(e_{t,l}) = 0$.

Nestas condições, Granger e Newbold (1986) definem $\hat{X}_t(m)$ como o previsor (linear) óptimo se a matriz $V(l)$ correspondente for a menor de todas as matrizes de covariância do erro de previsão. É um critério diferente do utilizado para a previsão de Box e Jenkins, que consiste na minimização do determinante de $V(l)$.

Utilizando (5.2.3) e (5.2.9), a expressão (5.2.11) é equivalente a:

$$\mathbf{e}_{:,l} = [\Psi(B) - B^l \alpha_l(B)] \boldsymbol{\varepsilon}_{1:l} = \sum_{u=0}^{l-1} \Psi_u \boldsymbol{\varepsilon}_{1+l-u} + \sum_{u=0}^{\infty} (\Psi_{u+l} - \alpha_{l,u}) \boldsymbol{\varepsilon}_{1-u} \quad (5.2.13)$$

Assim, a partir de (5.2.12):

$$V(l) = \sum_{u=0}^{l-1} \Psi_u \Sigma \Psi_u^T + \sum_{u=0}^{\infty} (\Psi_{u+l} - \alpha_{l,u}) \Sigma (\Psi_{u+l} - \alpha_{l,u})^T \quad (5.2.14)$$

À medida que $\alpha_l(B)$ varia, o conjunto de matrizes definidas positivas assim obtidas tem o valor mínimo:

$$V(l) = \sum_{u=0}^{l-1} \Psi_u^T \Sigma \Psi_u \quad (5.2.15)$$

sendo este valor atingido quando:

$$\alpha_{l,u} = \Psi_{u+l} \quad (5.2.16)$$

Deste modo, o previsor óptimo é:

$$\hat{X}_t(l) = \sum_{u=0}^{\infty} \Psi_{u+l} \boldsymbol{\varepsilon}_{t-u} \quad (5.2.17)$$

o que origina o vector do erro de previsão:

$$\mathbf{e}_{:,l} = \sum_{u=0}^{l-1} \Psi_u \boldsymbol{\varepsilon}_{1+l-u} \quad (5.2.18)$$

É importante referir que se verifica a partir de (5.2.15) que $V(l_1) \geq V(l_2)$ para $l_1 > l_2$, o que significa que, tal como seria de esperar, a qualidade da previsão é tanto pior quanto maior o seu passo.

Por outro lado, (5.2.17) pode também ser expressa da forma:

$$\hat{X}_t(l) = \psi_l \varepsilon_t + \sum_{u=0}^{\infty} \psi_{u+l+1} \varepsilon_{t-1-u} \quad (5.2.19)$$

obtendo-se então o seguinte resultado:

$$\hat{X}_t(l) = \hat{X}_{t-1}(l+1) + \psi_l (X_t - \hat{X}_{t-1}(1)) \quad (5.2.20)$$

Esta é a equação que permite a actualização das previsões em cada período; sendo dada a sequência de previsões feita no momento (t-1) e o valor mais recente da série X_t , é possível gerar as novas previsões para o período (t+l) utilizando (5.2.20).

Por outro lado, utilizando (5.2.6), considere-se ainda:

$$\hat{X}_t(l) - \sum_{u=1}^p \phi_u \hat{X}_t(l-u) = \sum_{r=0}^{\infty} \left\{ \psi_{r+l} \varepsilon_{t-r} - \sum_{u=1}^p \phi_u \psi_{r+l-u} \varepsilon_{t-r} \right\} = \sum_{r=0}^{\infty} \theta_{r+l} \varepsilon_{t-r} \quad \theta_r=0, r>q \quad (5.2.21)$$

Esta equação permite gerar previsões quando l aumenta e para um dado momento t fixo, substituindo-se $\hat{X}_t(l)$ por X_{t+l} para $l \leq 0$ em (5.2.21).

A título de exemplo, o previsor a um passo para o modelo AR(p) é:

$$\hat{X}_t(1) = \sum_{u=1}^p \phi_u X_{t-u+1} \quad (5.2.22)$$

De modo semelhante, o previsor para o momento (t+l) do modelo AR(1) é dado por:

$$\hat{X}_t(l) = \phi_l' X_t \quad (5.2.23)$$

É importante referir que, para utilizar este procedimento, é necessário admitir que o modelo utilizado para realizar as previsões é o verdadeiro processo que gerou a série cronológica em análise, apesar de tal suposição estar longe de poder ser considerada realista.

Por outro lado, é ainda interessante verificar que esta abordagem conduz ao mesmo previsor do método anterior.

De facto, a representação do modelo ARMA(p,q) para o momento (t+l) é:

$$\mathbf{X}_{t+l} = \sum_{u=1}^p \phi_u \mathbf{X}_{t+l-u} + \sum_{u=0}^q \theta_u \varepsilon_{t+l-u} \quad (5.2.24)$$

Assim, a equação de recorrência para a geração das previsões a partir do momento t é obtida do seguinte modo:

- (i) Substitui-se os valores desconhecidos \mathbf{X}_{t+r} pelos seus previsores $\hat{\mathbf{X}}_t(r)$ para $r > 0$;
- (ii) Mantém-se os vectores \mathbf{X}_{t+r} para $r \leq 0$, uma vez que se trata de valores conhecidos;
- (iii) Substitui-se ε_{t+r} pelo vector nulo $\mathbf{0}$ para $r > 0$;
- (iv) Mantém-se os vectores ε_{t+r} para $r \leq 0$, uma vez que se trata de valores conhecidos (admitindo que se conhece o verdadeiro modelo).

Deste modo, conclui-se também que :

$$\mathbf{e}_{t,l} = \mathbf{X}_{t+l} - \hat{\mathbf{X}}_t(l) = \mathbf{e}_t(l) \quad (5.2.25)$$

Esta forma recursiva de obter as previsões é precisamente a que foi descrita para o método de Box e Jenkins.

Desta forma, conclui-se que ambos os métodos conduzem ao mesmo predictor, ou seja, são equivalentes, o que significa que, para gerar previsões, é indiferente adoptar o critério de minimização do determinante da matriz de covariância do erro de previsão ou o de chegar à menor matriz de covariância desse erro (segundo a definição de Granger e Newbold; 1986). Na realidade, conforme já foi referido, Jenkins e Alavi (1981) indicam que, uma vez que $\mathbf{V}(l)$ é definida positiva, a minimização de $\mathbf{d}^T \mathbf{V}(l) \mathbf{d}$, que constitui o critério de Granger e Newbold, é equivalente à minimização de $|\mathbf{V}(l)|$, que é o critério que serve de base à previsão de Box e Jenkins para o caso multivariado.

VI. IDENTIFIABILIDADE E REDUNDÂNCIA

Considere-se o modelo ARMA(p,q) multivariado:

$$\phi_0 X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \theta_0 \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

ou:

$$\phi(B)X_t = \theta(B)\varepsilon_t$$

Um dos principais problemas relacionados com este modelo é o da identifiabilidade da sua estrutura, o que consiste em averiguar se é possível determinar de forma única os valores de p e de q e as matrizes $\phi_0, \phi_1, \dots, \phi_p, \theta_0, \theta_1, \dots, \theta_q$, a partir da função de covariância de $X_t^{(n)}$. Na realidade, a não identifiabilidade de um modelo resulta numa função de verosimilhança que não tem um máximo único. Neste sentido, um modelo ARMA não é indentifiável na forma em que é apresentado acima. Para comprovar isso, é suficiente notar que, se se multiplicar ambos os membros das expressões acima pelo mesmo polinómio matricial em B, a estrutura da covariância de X_t não sofre qualquer alteração. Com efeito, existe uma "classe de equivalência" de modelos correspondente a uma dada estrutura de covariância, isto é, existem diferentes modelos, designados de equivalentes, que correspondem a uma mesma função de covariância.

O problema da identifiabilidade, que é por vezes designado por problema da multiplicidade de modelos (Granger e Newbold, 1977, pág. 219), consiste então em definir um conjunto de regras que permita seleccionar um único modelo representativo de cada classe de equivalência. Isto significa que é necessário impôr certas condições a $\phi(B)$, $\theta(B)$ e aos seus elementos matriciais de modo a que, em cada classe de equivalência, apenas um único modelo as satisfaça (Priestley, 1981, pág. 801). São essas restrições que se pretende agora analisar.

1. Regras de identiabilidade

Antes de considerar as referidas restrições, é necessário colocar algumas hipóteses (Hannan, 1969, 1970, 1971, 1975, 1976, 1979):

a) Admite-se que:

$$\phi_0 = \theta_0 = I_m \quad (6.1.1)$$

o que constitui uma hipótese que tem sido colocada habitualmente.

É importante notar que a hipótese $\theta_0 = I_m$ não implica perda de generalidade desde que se permita que a matriz de variâncias e covariâncias de \mathcal{E}_t , designada por Σ , assuma uma forma genérica (não necessariamente diagonal). Assim, como alternativa a esta restrição colocada sobre θ_0 , é possível impôr a condição $\Sigma = I_m$, o que é facilmente conseguido através da transformação $\eta_t = P^{-1}\mathcal{E}_t$, sendo P uma matriz não singular arbitrária que verifica a relação $PP^T = \Sigma$, e reescrevendo em termos de η_t o segundo membro da equação definidora do modelo ARMA(p,q). No entanto, a restrição imposta a P não é suficiente para a determinação de uma única matriz, sendo necessário colocar a hipótese adicional de, por exemplo, P ser triangular inferior, o que implica que a matriz $P\theta_0$ também o seja. Assim, em vez da condição $\theta_0 = I_m$, é possível utilizar como alternativa as hipóteses $\Sigma = I_m$ e de θ_0 ser triangular inferior. Qualquer destas alternativas permite estimar os parâmetros de médias móveis e a sua não verificação significa que o operador $\theta(B)$ inclui parâmetros redundantes. Com efeito, note-se que, para um modelo MA(q), existem $(q+1)m$ autocovariâncias não nulas e $(2q+1) \times \frac{1}{2}m(m-1)$ covariâncias cruzadas, constituindo no total $\left\{ (q+1)m^2 - \frac{1}{2}m(m-1) \right\}$ grandezas conhecidas. Por outro lado, as matrizes $\theta_0, \theta_1, \dots, \theta_q$ com $\Sigma = I_m$, envolvem $(q+1)m^2$ parâmetros

escalares. Logo, conclui-se que existem $\frac{1}{2}m(m-1)$ parâmetros redundantes, que podem ser eliminados impondo a condição $\theta_0 = I_m$ e permitindo que Σ , sendo simétrica, assuma uma forma genérica ou, alternativamente, impondo a restrição $\Sigma = I_m$ e que θ_0 seja triangular inferior (Priestley, 1981, pág. 803).

b) $\phi(B), \theta(B)$ não têm divisores à esquerda comuns a não ser divisores unimodulares, ou seja, se o polinômio matricial $D(B)$ for tal que $\phi(B) = D(B)\phi_1(B)$ e $\theta(B) = D(B)\theta_2(B)$, sendo $\phi_1(B)$ e $\theta_2(B)$ polinômios matriciais, então $|D(B)|$ é uma constante.

c) As raízes de $|\phi(B)|$ e $|\theta(B)|$ de situam-se fora do círculo unitário, ou seja, o modelo é invertível e estacionário. No entanto, mesmo que $|\theta(B)|$ possua raízes sobre o círculo unitário, é ainda possível conseguir-se a identifiabilidade do modelo (Hannan, 1970, pág. 370; Priestley, 1981, pág.802).

Nestas condições, um modelo ARMA(p,q) é identificável se verificar qualquer das seguintes regras (Hannan, 1969, 1970, 1971, 1975, 1976):

1. De cada classe de equivalência, escolhe-se a sub-classe de modelos com o menor valor possível de p e desta escolhe-se o modelo com o menor valor de q. Para que $\phi(B), \theta(B)$ sejam determinados de forma única, é necessário e suficiente que a matriz $[\theta_q \ \phi_p]$ tenha característica igual a m, sendo m, como habitualmente, o número de processos univariados componentes do vector $\{X_t\}$. É útil referir que este resultado também se verifica se as posições de p e de q forem trocadas.

Por outro lado, Hannan (1975) indica que esta condição é sobre-identificável, o que significa que existem classes de equivalência para as quais ela não é satisfeita.

2. É possível utilizar uma representação "escalar", isto é, uma forma do modelo ARMA(p,q) em que os parâmetros ϕ_r , $r=1, \dots, p$, assumem a forma $\phi_r = \phi_r \mathbf{I}_m$. Assim, o modelo é identificável se $\phi_p \neq 0$.
3. Uma outra hipótese consiste em utilizar uma representação "triangular", isto é, uma forma do modelo em que $\phi(B)$ é uma matriz triangular inferior. Deste modo, sendo $\phi_{ij}(B)$ o (i,j)-ésimo elemento de $\phi(B)$, $i, j=1, \dots, m$, o modelo é identificável se $\phi_{ij}(B)$ for de grau não superior a $\phi_{jj}(B)$.
4. O modelo é identificável se o único vector λ tal que $\lambda^T (\phi_p - \theta_q) = 0$ for $\lambda = 0$. Esta condição verifica-se necessariamente se ϕ_p ou θ_q forem matrizes não singulares.

O procedimento descrito na condição 1. é o mais simples e o mais utilizado em termos práticos.

É importante ainda sublinhar que, tal como apontam Granger e Newbold (1977, pág. 223), é necessário que o modelo esteja correctamente diagnosticado para que seja possível analisar se estas condições se verificam.

Por outro lado, é de notar que se admitiu a identificabilidade do modelo em todas as questões tratadas até agora, ou seja, na estimação e na confirmação do diagnóstico de um modelo e na sua utilização para previsão.

VII. IDENTIFICAÇÃO

Sendo a mais delicada de todo o processo de modelização de uma série cronológica, a fase da identificação pode também ser qualificada como a mais importante. Com efeito, a identificação é determinante para o desenrolar de todas as outras fases seguintes: dela depende a obtenção de bons resultados na estimação e o não aparecimento de sinais importantes de falta de adequação do modelo identificado à série cronológica. Só nessas condições será possível obter a geração de previsões de boa qualidade, o que constitui o objectivo último da modelização de uma série cronológica.

Assim, a perspectiva adoptada para a identificação depende daquilo que é conhecido e daquilo que se está disposto a admitir. Em algumas aplicações, poderá ser possível uma especificação inicial adequada a partir do eventual conhecimento da natureza do problema em análise, permitindo a identificação da estrutura subjacente de uma forma directa.

No entanto, a perspectiva que vai ser adoptada é a oposta : admitir-se-à a completa inexistência de conhecimento prévio sobre o sistema de relações dinâmicas que originou a série em análise, sendo esta a situação mais comum na prática, particularmente nos domínios da Economia e de outras ciências sociais.

Neste contexto, o objectivo nesta fase é o de, dado um conjunto de observações que constituem a série cronológica, conseguir uma especificação inicial da estrutura do processo estocástico que esteja subjacente à geração dessa série. Pretende-se assim chegar a um modelo probabilístico que envolva um número de parâmetros tão pequeno quanto é possível e, simultaneamente, represente adequadamente o sistema de relações dinâmicas em análise. De qualquer modo, não deixa de ser útil reforçar a ideia de que não será utilizado qualquer conhecimento "a priori" sobre esse sistema, deixando-se as observações "falar por si

próprias", ou seja, fazendo depender a determinação da estrutura subjacente aos dados daquilo que eles próprios revelarem, sem recurso a formas externas de informação, adoptando-se assim o que se designa habitualmente por perspectiva de séries cronológicas.

Tendo em conta estes princípios, as diferentes abordagens que irão ser analisadas foram ordenadas em três grupos: o primeiro inclui abordagens que, tendo sido pioneiras na identificação de uma série cronológica multivariada, não se revelam capazes de o fazer satisfatoriamente quando o número de componentes desta é elevado ou mesmo apenas moderado. No entanto, são métodos que, embora apresentando alguns problemas, se revelam aceitavelmente adequados para identificar séries cronológicas multivariadas com um pequeno número de componentes e criaram as condições necessárias ao aparecimento de métodos mais sofisticados e elaborados. Nessa linha de raciocínio, apresenta-se em seguida um conjunto de propostas que constituem generalizações ao caso multivariado do método proposto por Box e Jenkins (1976) para a identificação de séries cronológicas univariadas. Recorrendo à classe de modelos ARMA, este conjunto de métodos parece revelar-se bastante satisfatório, cumprindo o duplo objectivo de representar adequadamente a teia de relações que eventualmente caracterize uma série cronológica multivariada e o de chegar a um modelo parcimonioso, ou seja, que envolva um número de parâmetros tão pequeno quanto possível. Com efeito, a utilização da referida classe de processos é muito útil, uma vez que possibilita a representação de uma vasta gama de processos estocásticos e, assim, o resumo da estrutura de uma série num modelo simples, eliminando características não fundamentais que poderiam distorcer um modelo. De qualquer modo, todos estes métodos apresentam as suas dificuldades, não podendo ainda ser considerados como perfeitos. Tendo isso em linha de

conta, e uma vez que um dado modelo pertencente à classe ARMA pode, apesar dos aperfeiçoamentos introduzidos na metodologia por algumas das abordagens analisadas, apresentar uma estrutura um pouco complexa ou de difícil modelização, exigindo meios que podem não estar ao alcance de qualquer analista, o último grupo de métodos engloba abordagens simplificadoras. O objectivo destas abordagens consiste em simplificar a estrutura identificada, quer através da introdução de simplificações num dado modelo já identificado, quer mesmo através da consideração e novas classes de modelos probabilísticos, que pretendem uma mais fácil e acessível modelização de uma série cronológica multivariada, embora isso nem sempre seja conseguido.

Finalmente, é conveniente referir que se deixou a análise desta fase para o fim por diversos motivos, embora ela seja sempre a primeira na modelização de qualquer série cronológica. Com efeito, afigurou-se útil fornecer uma panorâmica prévia das outras fases, de modo permitir a apresentação e análise de diversos instrumentos necessários à identificação, o que possibilita um melhor enquadramento e compreensão desses instrumentos e, por outro lado, de forma a proporcionar a percepção da importância de que se reveste a correcta identificação de uma série cronológica. A razão para essa importância reside, conforme foi já referido, no facto de a fase de identificação ser determinante para as fases seguintes, uma vez que as precede e, conseqüentemente, para a geração de previsões de boa qualidade. Assim, trata-se da etapa que, sem sombra de dúvida, constitui o maior desafio para qualquer analista que pretenda a modelização de uma série cronológica. Esta constatação assume ainda maior importância quando se toma em consideração a circunstância de se pretender a identificação de séries cronológicas multivariadas, ou seja, de se estar a trabalhar num espaço de dimensão

possivelmente elevada, o que é sinónimo de dificuldades acrescidas, aumentando por isso o desafio que se coloca ao analista.

É útil referir ainda que os métodos de identificação irão ser analisados apenas para modelos não sazonais, uma vez que a sua generalização ao caso de modelos que incluem componentes sazonais é imediata e não oferece dificuldades adicionais dignas de nota, tal como já foi indicado atrás.

1. Modelos de pequena dimensão

Existem métodos de modelização de uma série cronológica multivariada vocacionados para modelos de pequena dimensão no sentido de o número de séries univariadas envolvido ser pequeno. Assim, procede-se à descrição dos métodos habitualmente utilizados nestas circunstâncias, analisando-se inicialmente dois métodos de âmbito genérico e dando-se posteriormente particular ênfase ao caso bivariado.

1.1. Método residual

Este método, proposto por Parzen (1969), parte da representação de cada componente de $\{X_t\}$, que se admite ser um processo estacionário, como um modelo de médias móveis de ordem infinita, assumindo a forma:

$$X_{it} = \varepsilon_{it} + \psi_1^i \varepsilon_{i,t-1} + \psi_2^i \varepsilon_{i,t-2} + \dots \quad (7.1.1)$$

onde ψ_l^i ($i=1,\dots,m; l=1,2,\dots$) representa o l -ésimo parâmetro de X_{it} em (7.1.1) e $\{\varepsilon_{it}\}$ é um processo puramente aleatório, tendo-se $\varepsilon_{it} = X_{it} - E(X_{it} | X_{i,t-1}, X_{i,t-2}, \dots)$.

Todo o método de modelização de Parzen (1969) assenta na representação (7.1.1) e confere um papel central às variáveis residuais ε_{it} , o que justifica a designação que lhe é atribuída.

Neste método, são propostos dois tipos de abordagem: uma abordagem individual e uma conjunta, que serão agora analisadas.

a) Abordagem individual

As variáveis residuais individuais, que passarão a ser designadas por ε_{it}^{ind} , são definidas a partir do valor esperado condicionado de X_{it} dados $X_{i,t-1}, X_{i,t-2}, \dots$:

$$\varepsilon_{it}^{ind} = X_{it} - E(X_{it} | X_{i,t-r}, r > 0) \quad (7.1.2)$$

A equação (7.1.1) representa X_{it} em termos das suas variáveis residuais individuais. Nestas condições, a abordagem individual procede à modelização de X_t em duas fases:

(i) Em primeiro lugar, procede-se à estimação de cada processo $\{\varepsilon_{it}^{ind}\}$, o que pode ser feito a partir do ajustamento de um modelo auto-regressivo univariado a X_{it} , cuja ordem pode ser estimada através da utilização dos critérios apropriados. Em seguida, passa-se à estimação dos parâmetros da equação:

$$X_{it} = \varepsilon_{it}^{ind} + \psi_{1ind}^i \varepsilon_{i,t-1}^{ind} + \psi_{2ind}^i \varepsilon_{i,t-2}^{ind} + \dots \quad (7.1.3)$$

(ii) A segunda fase consiste na estimação dos parâmetros da equação.

$$\varepsilon_t^{ind} = \eta_t + \psi_1^{ind} \eta_{t-1} + \psi_2^{ind} \eta_{t-2} + \dots \quad (7.1.4)$$

onde $\varepsilon_t^{ind} = (\varepsilon_{1t}^{ind}, \dots, \varepsilon_{mt}^{ind})^T$ e o vector $\{\eta_t\}$ é um processo puramente aleatório multivariado que representa as variáveis residuais da representação de ε_t^{ind} como um processo de médias móveis de ordem infinita especificada em (7.1.4). No entanto, uma vez que os elementos do vector ε_t^{ind} são desconhecidos, é necessário substituí-los pelos seus estimadores $\hat{\varepsilon}_t^{ind} = (\hat{\varepsilon}_{1t}^{ind}, \dots, \hat{\varepsilon}_{mt}^{ind})^T$ obtidos na fase anterior.

b) Abordagem conjunta

Esta abordagem parte da definição das variáveis residuais conjuntas, designadas por ε_{it}^{conj} , que é:

$$\varepsilon_{it}^{conj} = X_{it} - E(X_{it} | X_{j,t-r}, r > 0, j = 1, \dots, m) \quad (7.1.5)$$

ou seja, ε_{it}^{conj} é definida a partir do valor esperado condicionado de X_{it} dados os valores passados de todos os elementos do vector X_t . O vector $\varepsilon_t^{conj} = (\varepsilon_{1t}^{conj}, \dots, \varepsilon_{mt}^{conj})$ é um ruído branco com matriz de variâncias e covariâncias Σ^{conj} .

Assim, o processo de modelização nesta abordagem passa também por duas fases:

(i) Na primeira fase, procede-se à estimação da matriz Σ^{conj} a partir do ajustamento de um modelo auto-regressivo multivariado de ordem suficientemente elevada a X_t , sendo essa ordem estimada pelos critérios adequados.

(ii) Em seguida, passa-se à estimação das matrizes de parâmetros da equação:

$$X_t = \varepsilon_t^{\text{conj}} + \psi_1^{\text{conj}} \varepsilon_{t-1}^{\text{conj}} + \psi_2^{\text{conj}} \varepsilon_{t-2}^{\text{conj}} + \dots \quad (7.1.6)$$

Fica assim completo o processo de modelização de uma série cronológica multivariada utilizando esta abordagem.

No entanto, este método não parece ser capaz de proceder à modelização de uma série cronológica multivariada com um número de componentes elevado ou mesmo apenas moderado, devido principalmente aos seguintes motivos:

a) Está dependente da definição das variáveis residuais a partir dos modelos univariados ajustados às componentes de X_t , sendo susceptível de conduzir a um modelo sobreparametrizado, uma vez que assenta nas representações (7.1.3) e (7.1.4) na primeira abordagem, ou na representação (7.1.6) no caso da segunda abordagem.

b) Por outro lado, não indica qual a forma de determinação do número de parâmetros a incluir nas expressões (7.1.3), (7.1.4) e (7.1.6), podendo por isso conduzir à necessidade de utilização de um elevado número de parâmetros em qualquer dessas equações. Assim, a utilização que é feita dos modelos univariados, em particular na abordagem individual, e o tipo de representação que é proposto, parecem dificultar ou até inviabilizar a

modelização de uma série multivariada com um número de componentes elevado ou mesmo apenas moderado.

c) O tipo de representação utilizado pode ser inadequado. Com efeito, a determinação da forma do modelo é feita "a priori", não dependendo da série cronológica em análise, o que não é conveniente.

Apesar de tudo, é útil referir este método, apesar da sua complexidade e dos problemas que acarreta, pois trata-se de um método pioneiro, constituindo um primeiro esforço no sentido de um procedimento genérico e que permita modelizar uma série cronológica com um qualquer número de componentes, o que justifica o interesse da sua análise.

1.2. Método comparativo

Wallis (1977) e Chan e Wallis (1978) propõem um método que passa pela construção prévia de modelos univariados para as componentes da série cronológica multivariada, de modo a sugerir valores adequados para as ordens dos operadores do modelo multivariado a partir das ordens dos operadores daqueles modelos univariados. Assim, o método desenrola-se de acordo com as seguintes fases:

(i) Em primeiro lugar, procede-se à modelização isolada de cada série univariada, construindo-se os modelos ARMA invertíveis e estacionários:

$$\phi_i(B)X_{it} = \theta_i(B)\varepsilon_{it} \quad i = 1, \dots, m \quad (7.1.7)$$

onde $\{\varepsilon_{it}\}$ são ruídos brancos e $\phi_i(B)$ e $\theta_i(B)$ são respectivamente o operador auto-regressivo e o de médias móveis do i -ésimo modelo.

(ii) Admitindo que a série cronológica multivariada X_t pode ser modelizada pelo seguinte modelo ARMA (p,q) invertível e estacionário:

$$\phi(B)X_t = \theta(B)\varepsilon_t \quad (7.1.8)$$

é necessário representá-lo na forma alternativa:

$$|\phi(B)|X_i = H(B)\varepsilon_i \quad (7.1.9)$$

onde $H(B) = \phi^{\wedge}(B)\theta(B)$, representando $\phi^{\wedge}(B)$ a adjunta da matriz $\phi(B)$, como tem sido habitual.

Obtém-se deste modo uma representação alternativa que permite chegar a modelos univariados para as diferentes componentes do vector X_i , tal como já foi referido atrás.

(iii) O ponto fulcral deste procedimento consiste precisamente em comparar as ordens dos operadores dos diferentes modelos univariados com as dos operadores do modelo multivariado expresso sob a forma (7.1.9), de modo a determinar os valores das segundas a partir dos valores das primeiras, o que justifica a designação que foi atribuída a este método. Assim, a determinação das ordens do modelo multivariado processa-se do seguinte modo:

a) A ordem de $H(B)$ é determinada a partir das ordens dos operadores $\theta_i(B)$ dos diferentes modelos univariados (7.1.7), sendo geralmente escolhido o valor máximo destas.

b) Em seguida, considere-se o modelo (7.19) expresso sob a forma:

$$D(B)X_i = H(B)\varepsilon_i \quad (7.1.10)$$

onde $D(B) = |\phi(B)|xI_m$ é uma matriz diagonal cujos elementos são polinómios em B . Designando a ordem de $D(B)$ por l e mantendo a ordem de $H(B)$ fixa, procede-se à estimação de modelos da forma (7.1.10) para $l=r, r-1, \dots$, sendo r um valor máximo para l previamente fixado e escolhido habitualmente como o máximo das ordens dos operadores auto-regressivos dos modelos univariados (7.1.7). Com base em testes da razão de verosimilhança e em testes de significância estatística dos

parâmetros estimados individuais, escolhe-se o valor apropriado para l , mantendo sempre fixa a ordem do operador de médias móveis.

c) O passo seguinte consiste em testar se os elementos da matriz $D(B)$ são iguais, ou seja, da forma (7.1.9), através de um teste da razão de verossimilhança.

(iv) Finalmente, os valores de p e de q são propostos a partir das matrizes $H(B)$ e $D(B)$ ou de $|\phi(B)|$ estimados, determinando-se para o efeito qual o modelo da forma (7.1.8) que conduz à representação da forma (7.1.10) ou (7.1.9) obtida.

Este método apresenta diversos problemas, revelando grandes dificuldades para modelizar uma série cronológica multivariada com um número de componentes elevado ou apenas moderado. Com efeito, a sua eficácia parece muito questionável, levando a concluir que só é adequado para uma série com um pequeno número de componentes, o que é reconhecido pelos seus autores. Assim, as principais dificuldades que se lhe colocam são resumidamente as seguintes:

a) A ordem da matriz $H(B)$ em (7.1.9) pode ser superior à ordens de qualquer dos operadores $\theta_i(B)$ dos modelos univariados (7.1.7). A título de exemplo, suponha-se $m=2$ e que a matriz $H(B)$ assume a forma:

$$H(B) = \begin{bmatrix} 1 & h_1 B \\ h_2 B & 1 \end{bmatrix} \quad (7.1.11)$$

Admita-se também que \mathcal{E}_{1t} e \mathcal{E}_{2t} são independentes. Então, a ordem dos operadores de médias móveis é nula, mas seria um erro concluir que a ordem de $H(B)$ é também nula (Tiao e Box, 1981). Assim, a componente de médias móveis do modelo (7.1.9) poderá ser frequentemente mal identificada.

b) A representação (7.1.9) não é parcimoniosa para modelos ARMA multivariados. Ignorando a matriz de covariância Σ , o número máximo

de parâmetros na forma original (7.1.8) é de $m^2(p+q)$, enquanto na forma (7.1.9) é de $mp+[(m-1)p+q]m^2$, representando um excesso de $pm(m-1)^2$ parâmetros, podendo este ser ainda maior se se utilizar a forma (7.1.10). Assim, mesmo admitindo que a ordem de $H(B)$ é especificada correctamente, será necessário estimar um grande número de parâmetros adicionais para identificar correctamente um modelo auto-regressivo, mesmo de baixa ordem como $p=1$ ou $p=2$, o que tanto mais grave quanto maior o número de componentes da série cronológica. Isto significa que, geralmente, a estrutura resultante deste procedimento será sobreparametrizada.

Este é o principal obstáculo que se coloca à capacidade deste método para modelizar uma série cronológica de dimensão moderada ou elevada, revelando-se já um sério problema quando $m=3$ ou $m=4$.

c) A imposição de uma matriz diagonal como operador auto-regressivo pode conduzir a uma incorrecta especificação desta componente do modelo identificado.

d) Uma vez que a correspondência entre os graus do polinómio $|\phi(B)|$ e de $H(B)$ e os valores de (p,q) não é necessariamente unívoca, não é claro como se determina p e q em (7.1.8) a partir da forma (7.1.9).

Em resumo, este método não se afigura adequado para modelizar uma série cronológica com um número de componentes elevado ou mesmo apenas moderado, recorrendo a um procedimento bastante complicado e que necessita de um grande número de fases para identificar um modelo, por mais simples que ele seja.

1.3. Modelos bivariados

Um caso particular dos modelos de pequena dimensão é o caso bivariado, em que se relaciona apenas duas séries cronológicas univariadas. Trata-se de um caso particularmente importante, não só pela

sua grande relevância em termos práticos, como também pela existência de um considerável número de métodos que procuram solucionar o problema da modelização de uma série bivariada. Por outro lado, estes métodos revelam-se pouco adequados ou mesmo incapazes no caso de uma série cronológica de dimensão mais elevada, em que as dificuldades metodológicas seriam de tal ordem que inviabilizariam a construção do modelo, pois são especialmente concebidos para o caso bivariado. De qualquer modo, é possível analisar todas as questões relevantes no processo de modelização através destes métodos quando se dispõe apenas de um par de séries univariadas.

Assim, será primeiramente analisado o caso em que existe uma relação bidireccional entre as séries univariadas, ou seja, em que a relação se verifica nos dois sentidos, procedendo-se depois ao tratamento do caso em que a relação é unidireccional, ou seja, verifica-se apenas num sentido.

Deste modo, considerando-se o modelo ARMA (p,q) multivariado invertível e estacionário:

$$\phi(B)X_t = \theta(B)\varepsilon_t \quad (7.1.12)$$

é possível expressá-lo da forma:

$$\theta^{\wedge}(B)\phi(B)X_t = |\theta(B)|\varepsilon_t \quad (7.1.13)$$

onde $\theta^{\wedge}(B)$ designa a adjunta da matriz $\theta(B)$: $\theta(B)^{-1} = \theta^{\wedge}(B)/|\theta(B)|$.

Desta expressão resultam equações da forma:

$$X_{it} = \sum_{j=1}^m \frac{w'_{ij}(B)}{\delta'_{ij}(B)} X_{jt} + \frac{\psi'_i(B)}{\phi'_i(B)} \varepsilon_{it} \quad i, j = 1, \dots, m \quad (7.1.14)$$

onde $w'_{ij}(B)$, $\delta'_{ij}(B)$, $\psi'_i(B)$, e $\phi'_i(B)$ são polinómios escalares em B de ordem finita. Modelos da forma de (7.1.14) são habitualmente

designados por modelos de função transferência-ruído (Box e Jenkins, 1976, pág. 363; Granger e Newbold, 1977a, pág.217, 1977b).

Considere-se então o caso bivariado, ou seja, o caso em que $m=2$. Tal como já foi referido, tratar-se-à primeiro o caso em que existe uma relação bidireccional e, em seguida, analisar-se-á aquele em que a relação é unidireccional.

1.3.1. Relação bidireccional

Neste caso o modelo (7.1.14) pode ser expresso da forma:

$$\begin{aligned} X_{1t} &= V_1'(B)X_{2t} + U_1'(B)\varepsilon_{1t} \\ X_{2t} &= V_2'(B)X_{1t} + U_2'(B)\varepsilon_{2t} \end{aligned} \quad (7.1.15)$$

onde $V_i'(B)$ e $U_i'(B)$, $i=1,2$, são polinómios em B de ordem infinita, admitindo-se que X_{1t} e X_{2t} são conjuntamente estacionários e que $E(X_{1t})=E(X_{2t})=0$. Por outro lado, é possível estabelecer as relações:

$$V_i'(B) = \frac{w_i'(B)}{\delta_i'(B)} \quad ; \quad U_i'(B) = \frac{\psi_i'(B)}{\phi_i'(B)} \quad i=1,2 \quad (7.1.16)$$

onde $w_i'(B)$, $\delta_i'(B)$, $\psi_i'(B)$ e $\phi_i'(B)$, $i=1,2$, são polinómios em B de ordem finita, com $w_2'(0) = 0$, admitindo-se ainda a hipóteses de as raízes das equações seguintes se situarem todas fora do círculo unitário:

$$\delta_i'(B) \neq 0 \quad ; \quad \psi_i'(B) \neq 0 \quad ; \quad \phi_i'(B) \neq 0 \quad i=1,2 \quad (7.1.17)$$

O modelo (7.1.15) está expresso sob a forma recursiva de um modelo de função transferência-ruído, sendo $V_i'(B)$ a função transferência e $U_i'(B)\varepsilon_{it}$ a componente ruído. Quanto às variáveis residuais ε_{1t} e ε_{2t} , admite-se:

$$\begin{aligned} E(\varepsilon_{it}\varepsilon_{i,t-l}) &= 0 && i=1,2 \text{ para todo o } l \neq 0; \\ E(\varepsilon_{1t}\varepsilon_{2s}) &= 0 && \text{para todo o } t,s \end{aligned} \quad (7.1.18)$$

Neste contexto, torna-se agora possível desenvolver procedimentos conducentes à modelização de uma série cronológica bivariada, seguindo-se o método de Granger e Newbold (1977a,b), que se desenvolve percorrendo as seguintes fases:

(i) Em primeiro lugar, é necessário ajustar modelos ARMA invertíveis e estacionários univariados a cada série, conduzindo às expressões:

$$\phi_i(B)X_{it} = \theta_i(B)\eta_{it} \quad i=1,2 \quad (7.1.19)$$

onde $\{\eta_{1t}\}$ e $\{\eta_{2t}\}$ são processos puramente aleatórios.

O objectivo desta fase é o de converter X_{1t} e X_{2t} em ruídos brancos de modo a eliminar a estrutura de autocorrelação de ambos. Desta forma, já é possível extrair conclusões a partir da função de correlação cruzada, como já foi referido, sendo da maior importância neste caso poder inferir se a relação existente entre X_{1t} e X_{2t} é unidireccional ou bidireccional. Assim, com vista a analisar essa relação, procede-se à estimação da função de correlação cruzada entre os resíduos de estimação dos modelos univariados, designados por $\hat{\eta}_{1t}$ e $\hat{\eta}_{2t}$. Com efeito, a relação entre η_{1t} e η_{2t} é do mesmo tipo da existente entre X_{1t} e X_{2t} , isto é, uma relação unidireccional entre estas implica a existência de uma relação unidireccional entre aquelas, sucedendo o mesmo no caso de uma relação bidireccional. É possível então determinar o tipo de relação existente entre X_{1t} e X_{2t} através da função de correlação cruzada entre η_{1t} e η_{2t} , sendo necessário recorrer ao respectivo estimador. No caso de se tratar de uma relação unidireccional, a função de correlação cruzada deverá apresentar valores significativos apenas para "lags" negativos ou para "lags" positivos, o que revelaria respectivamente variações de sentido inverso ou no mesmo sentido, mas nunca para "lags" positivos e negativos simultaneamente. Pelo contrário, se a relação for bidireccional, a função

referida deverá apresentar valores significativos tanto para "lags" positivos como negativos.

Deste modo, Pierce (1977) propõe uma medida global da intensidade da relação entre X_{1t} e X_{2t} , assumindo a forma:

$$R^2 = \sum_{k \in A} \hat{r}_{\eta_{1t} \eta_{2t}}^2(k) \quad (7.1.20)$$

onde $\hat{r}_{\eta_{1t} \eta_{2t}}(k)$ designa a função de correlação cruzada estimada entre os resíduos de estimação dos modelos univariados (7.1.19) e o conjunto A é composto pelos "lags" para os quais os valores dessa função são significativos. A designação desta medida por R^2 resulta da consideração da regressão:

$$\eta_{1t} = \sum_{l \in A} \beta_l \eta_{2,t-l} + \zeta_t \quad (7.1.21)$$

sendo R^2 um estimador consistente do coeficiente de determinação desta regressão (Pierce, 1977). Analogamente à sua interpretação em modelos da regressão, R^2 pode ser encarado aqui como a percentagem da variância η_{1t} explicada pelos valores passados, presentes e futuros de η_{2t} . A grande diferença relativamente aos modelos de regressão habituais consiste, devido à consideração de η_{1t} em (7.1.21), em eliminar da possível explicação por X_{2t} aquela parte de X_{1t} que é explicada apenas pelo seu próprio passado.

Assim, após o ajustamento dos modelos univariados (7.1.19), calcula-se e normaliza-se os resíduos de estimação de cada modelo, de modo a que as suas variâncias sejam unitárias.

(ii) Constrói-se em seguida um modelo bivariado relacionando as variáveis residuais dos modelos (7.1.19):

$$\begin{aligned}\eta_{1t} &= V_1(B)\eta_{2t} + U_1(B)\xi_{1t} = \frac{w_1(B)}{\delta_1(B)}\eta_{2t} + \frac{\psi_1(B)}{\phi_1(B)}\xi_{1t} \\ \eta_{2t} &= V_2(B)\eta_{1t} + U_2(B)\xi_{2t} = \frac{w_2(B)}{\delta_2(B)}\eta_{1t} + \frac{\psi_2(B)}{\phi_2(B)}\xi_{2t}\end{aligned}\quad (7.1.22)$$

onde η_{1t} e η_{2t} são normalizadas de modo a que as suas variâncias sejam unitárias, sendo $w_2(0)=0$ e ξ_{1t} e ξ_{2t} são ruídos brancos que verificam as hipóteses:

$$\begin{aligned}E(\xi_{it}, \xi_{i,t-k}) &= 0 \quad i=1,2 \quad \text{para todo o } k; \\ E(\xi_{1t}, \xi_{2t}) &= 0 \quad \text{para todo o } t, l\end{aligned}\quad (7.1.23)$$

Deste modo, sendo $\gamma_{\eta_{12}}(k)=E(\eta_{1t}\eta_{2,t-k})$, a função de covariância cruzada entre η_{1t} e η_{2t} , e vez que ξ_{1t} e ξ_{2t} são não correlacionados, resulta de (7.1.22) a relação:

$$E[(\eta_{1t}-V_1(B)\eta_{2t})(\eta_{2,t-k}-V_2(B)\eta_{1,t-k})]=0 \quad (7.1.24)$$

retirando-se desta equação as relações:

$$\begin{aligned}\gamma_{\eta_{12}}(k) &= V_{1k} - \sum_{r=0}^{\infty} \sum_{s=1}^{\infty} V_{1r} V_{2s} \gamma_{\eta_{12}}(r-s-k) \quad k=0,1,2,\dots \\ \gamma_{\eta_{12}}(-k) &= V_{2k} - \sum_{r=0}^{\infty} \sum_{s=1}^{\infty} V_{1r} V_{2s} \gamma_{\eta_{12}}(k+r-s) \quad k=1,2,\dots\end{aligned}\quad (7.1.25)$$

onde V_{ik} ($i=1,2$) representa o k -ésimo elemento do polinómio $V_i(B)$.

A partir do estimador de $\gamma_{\eta_{12}}(k)$, é possível estimar V_{1k} ($k=0,1,\dots,M$) e V_{2k} ($k=1,2,\dots,M$) através de um método iterativo que permita resolver as equações (7.1.25), nas quais os somatórios são truncados em M , sendo este um valor inteiro para o qual é proposto $M=10$ ou $M=15$. Uma via possível consiste em tomar $\hat{V}_{2k}^{(1)} = \hat{C}_{\eta_{12}}(-k)$, $k=1,2,\dots,M$, substituir V_{2k} pelo estimador assim obtido, e resolver o sistema linear resultante de modo a determinar V_{1k} , obtendo-se desta

forma o estimador $\hat{V}_{ik}^{(1)}$, $k=0,1,2,\dots,M$, designando $\hat{C}_{\eta_{12}}(k)$ o estimador da função de covariância cruzada entre os resíduos de estimação dos modelos univariados (7.1.19) $\hat{\eta}_{1k}$ e $\hat{\eta}_{2k}$. Em seguida, V_{1k} é substituído pelo seu estimador nas equações (7.1.25), resolvendo-se o sistema linear resultante de modo a determinar V_{2k} , produzindo um novo estimador $\hat{V}_{2k}^{(2)}$ ($k=1,2,\dots,M$). O procedimento é então repetido iterativamente até ser obtida a sua convergência.

No entanto, tal como referem os seus autores, este método apresenta duas dificuldades. Por um lado, não existem garantias de convergência do algoritmo e, por outro, os erros de estimação de $\gamma_{\eta_{12}}(k)$ têm tendência por vezes a provocar grandes distorções nas estimativas obtidas para V_{ik} ($i=1,2$). Assim, introduz-se algumas alterações no método, de modo a diminuir as consequências destes problemas. As equações (7.1.25) são então resolvidas da forma referida, mas tomando $M=2,3,4,\dots$. Os estimadores iniciais $\hat{V}_{2k}^{(1)}$ ($k=1,2,\dots,M-1$) são tomados como os estimadores finais obtidos na fase anterior, com $\hat{V}_{2M}^{(1)}=0$. Por outro lado, tendo em vista aumentar as hipóteses de convergência, realiza-se cada iteração utilizando diferentes valores para um ponderador, designado por b , de modo a determinar os estimadores:

$$\hat{V}_{ik}^{(h)} = b\hat{V}_{ik}^{(h-1)} + (1-b) \text{ (Valor obtido através da resolução das equações)}$$

Granger e Newbold (1977a,b) referem que, geralmente, tomam $b=0$ como valor inicial do ponderador e vão aumentando-o em passos de 0.2 até chegar a $b=0.8$. Deste modo, dispõe-se de várias estimativas para os elementos dos polinómios $V_i(B)$, $i=1,2$, para $M=2,3,4,\dots$, sendo possível julgar se alterações abruptas nos valores dessas estimativas, quando M aumenta, são provocadas por erros na estimação ou se, na realidade, reflectem alguma relação subjacente.

A partir da estimação dos polinômios $V_i(B) = V_{i0} + V_{i1}B + V_{i2}B^2 + \dots$, já é possível identificar as formas das funções transferência $w_i(B)/\delta_i(B)$, $i=1,2$. Consegue-se assim aproximar cada um dos polinômios de ordem infinita $V_i(B)$ por um quociente de polinômios de ordem finita, sendo $w_i(B) = w_{i0} + w_{i1}B + \dots + w_{in}B^n$ e $\delta_i(B) = 1 - \delta_{i1}B - \dots - \delta_{in}B^n$. Assim, se, por exemplo, se verificar:

$$V_{1k} = \delta_{11} V_{1,k-1} \quad k=1,2,3,\dots \quad (7.1.26)$$

com $V_{10} = w_{10}$, o modelo $V_1(B) = w_{10}/(1 - \delta_{11}B)$ seria adequado, ou seja, ter-se-ia $w_1(B) = w_{10}$ e $\delta_1(B) = 1 - \delta_{11}B$, resultando a igualdade:

$$(1 - \delta_{11}B)(V_{10} + V_{11}B + V_{12}B^2 + V_{13}B^3 + \dots) = w_{10} + w_{11}B \quad (7.1.27)$$

o que permite obter as relações :

$$V_{10} = w_{10} ; V_{11} = \delta_{11}w_{10} + w_{11} ; V_r = \delta_{11}V_{r-1} \quad (r=2,3,\dots) \quad (7.1.28)$$

No entanto, se (7.1.26) se verificar apenas para $k=2,3,\dots$, a forma apropriada para a função transferência será:

$$V_1(B) = \frac{w_{10} + w_{11}B}{1 - \delta_{11}B} \quad (7.1.29)$$

tendo-se neste caso $w_1(B) = w_{10} + w_{11}B$.

Por outro lado, se se verificar um decréscimo de segunda ordem, que tanto pode ser exponencial como sinusoidal, nos elementos de $V_1(B)$, ou seja, se estes verificarem a relação:

$$V_{1k} = \delta_{11}V_{1,k-1} + \delta_{12}V_{1,k-2} \quad (7.1.30)$$

então, ter-se-ia $\delta_1(B) = 1 - \delta_{11}(B) - \delta_{12}B^2$.

Genericamente, a identificação da forma dos polinômios $w_i(B)$ e $\delta_i(B)$ a partir de $V_i(B)$ processa-se da forma que se passa a descrever

(Box e Jenkins, 1976, pág. 346). A partir de $V_i(B) = w_i(B) / \delta_i(B)$, obtém-se a igualdade:

$$\begin{aligned} (1 - \delta_{i1}B - \delta_{i2}B^2 - \dots - \delta_{in}B^n)(V_{i0} + V_{i1}B + V_{i2}B^2 + \dots) = \\ = (w_{i0} + w_{i1}B + \dots + w_{il}B) \end{aligned} \quad (7.1.31)$$

Igualando os coeficientes de B, obtém-se as equações:

$$\begin{aligned} V_{ik} &= 0 & k < 0 \\ V_{ik} &= \delta_{i1}V_{i,k-1} + \delta_{i2}V_{i,k-2} + \dots + \delta_{in}V_{i,k-n} + w_{ik} & k = 0, 1, \dots, l \\ V_{ik} &= \delta_{i1}V_{i,k-1} + \delta_{i2}V_{i,k-2} + \dots + \delta_{in}V_{i,k-n} & k > l \end{aligned} \quad (7.1.32)$$

Os elementos $V_{i0}, V_{i,1}, \dots, V_{i,l-n+1}$ constituem n valores iniciais da equação às diferenças:

$$\delta(B)V_{ik} = 0 \quad k > l \quad (7.1.33)$$

A solução desta equação aplica-se a todos os valores V_{ik} para $k \geq l-n+1$. Assim, conclui-se que os elementos V_{ik} assumem a forma:

- $V_{ik} = 0$ quando $k < 0$;
- Os $l-n+1$ elementos de $V_{i0}, V_{i1}, \dots, V_{i,l-n}$ não obedecem a nenhum padrão fixo (tais valores não ocorrem se $l > n$);
- Os elementos V_{ik} , para $k \geq l-n+1$, seguem um padrão determinado pela equação às diferenças de ordem n cujos n valores iniciais são $V_{i0}, V_{i,1}, \dots, V_{i,l-n+1}$. Os valores iniciais V_{ik} para $k < 0$ são nulos, como é evidente.

(iii) A terceira fase consiste em identificar a estrutura dos polinómios das variáveis residuais do modelo (7.1.22), ou seja, $\psi_i(B) / \phi_i(B)$, $i=1,2$. Para o efeito, considere-se a primeira equação de (7.1.22). Sendo

$e_{it} = \frac{\psi_i(B)}{\phi_i(B)} \xi_{it}$, a função de autocovariância de e_{1t} , que será designada por $\gamma_{e_{11}}(k)$, é:

$$\gamma_{e_{1t}}(k) = E[(\eta_{1t} - V_1(B)\eta_{2t})(\eta_{1,t-k} - V_1(B)\eta_{2,t-k})] \quad k=0,1,2,\dots \quad (7.1.34)$$

Deste modo, uma vez que η_{1t} e η_{2t} têm variâncias unitárias, obtém-se as equações:

$$\begin{aligned} \gamma_{e_{1t}}(0) &= 1 + \sum_{r=0}^{\infty} V_{1r}^2 - 2 \sum_{r=0}^{\infty} V_{1r} \gamma_{\eta_{12}}(r) \\ \gamma_{e_{1t}}(k) &= \sum_{r=k}^{\infty} V_{1r} V_{1,r-k} - \sum_{r=0}^{\infty} V_{1r} \gamma_{\eta_{12}}(r-k) - \sum_{r=0}^{\infty} V_{1r} \gamma_{\eta_{12}}(r+k) \quad k=1,2,\dots \end{aligned} \quad (7.1.35)$$

Em termos práticos, os somatórios de ambas as equações são truncados num número inteiro M.

As equações (7.1.35) servem para estimar $\gamma_{e_{1t}}(k)$, $k=0,1,2,\dots$, substituindo-se para o efeito $\gamma_{\eta_{12}}(k)$ e V_{1r} pelos respectivos estimadores. No entanto, é preferível utilizar os valores de V_{1r} determinados a partir da função transferência identificada, em vez dos valores calculados a partir de (7.1.25). Isto significa que, tendo-se a relação:

$$\delta_1(B) = w_1(B) V_1(B) \quad (7.1.36)$$

esta equação é resolvida igualando os coeficientes de B, fornecendo as primeiras equações estimadores para os parâmetros de $\delta_1(B)$ e de $w_1(B)$ a partir dos estimadores dos primeiros coeficientes de $V_1(B)$, obtidos em (7.1.25). Deste modo, os estimadores dos coeficientes de $V_1(B)$ seguintes são então obtidos a partir dos estimadores dos parâmetros de $\delta_1(B)$ e de $w_1(B)$, igualando os coeficientes de B em (7.1.36) para potências mais elevadas de B.

Assim, a partir da estimação da função de autocovariância de e_{it} , é possível estimar a sua função de autocorrelação, permitindo identificar, através dos métodos de modelização de séries univariadas, a estrutura dos polinómios $\psi_i(B)$ e $\phi_i(B)$, $i=1,2$, uma vez que se tem a igualdade:

$$\phi_i(B) e_{it} = \psi_i(B) \xi_{it} \quad i=1,2 \quad (7.1.37)$$

e dado que $\{\xi_{it}\}$ são ruídos brancos. Identificado o modelo (7.1.22), passa-se à fase seguinte.

(iv) Após a sua identificação, o modelo (7.1.22) é estimado, para o que Granger e Newbold (1977a,b) propõem um método simplificador. Uma vez que ξ_{1t} e ξ_{2t} são não correlacionados entre si e que $w_2(0)=0$, é possível estimar cada equação separadamente, exprimindo-se para o efeito as equações (7.1.22) da forma:

$$\begin{aligned}\delta_1(B)\varphi_1(B)\eta_{1t} &= \varphi_1(B)w_1(B)\eta_{2t} + \delta_1(B)\psi_1(B)\xi_{1t} \\ \delta_2(B)\varphi_2(B)\eta_{2t} &= \varphi_2(B)w_2(B)\eta_{1t} + \delta_2(B)\psi_2(B)\xi_{2t}\end{aligned}\quad (7.1.38)$$

Estas equações podem ser expressas do modo seguinte:

$$\begin{aligned}\xi_{1t} &= \eta_{1t} - a_{11}\eta_{1,t-1} - \dots - a_{1H_1}\eta_{1,t-H_1} - b_{10}\eta_{2t} - \dots - b_{1I_1}\eta_{2,t-I_1} + d_{11}\xi_{1,t-1} + \dots + d_{1J_1}\xi_{1,t-J_1} \\ \xi_{2t} &= \eta_{2t} - a_{21}\eta_{2,t-1} - \dots - a_{2H_2}\eta_{2,t-H_2} - b_{20}\eta_{1t} - \dots - b_{2I_2}\eta_{1,t-I_2} + d_{21}\xi_{2,t-1} + \dots + d_{2J_2}\xi_{2,t-J_2}\end{aligned}\quad (7.1.39)$$

onde H_i , I_i e J_i ($i=1,2$) são números inteiros conhecidos e os parâmetros $a_{11}, \dots, a_{1H_1}, b_{10}, \dots, b_{1I_1}, d_{11}, \dots, d_{1J_1}$ são funções conhecidas dos parâmetros de $\delta_1(B)$, $w_1(B)$, $\varphi_1(B)$, e $\psi_1(B)$; por sua vez, $a_{21}, \dots, a_{2H_2}, b_{20}, \dots, b_{2I_2}, d_{21}, \dots, d_{2J_2}$ são funções conhecidas dos parâmetros de $\delta_2(B)$, $w_2(B)$, $\varphi_2(B)$, e $\psi_2(B)$.

Os estimadores dos parâmetros de (7.1.22) são obtidos estimando através das equações (7.1.39) os valores de ξ_{it} como função de $\hat{\eta}_{it}$ ($i=1,2$) e dos parâmetros estimados, sendo $t=\max(H_1+1, J_1+1), \dots, N$ para a primeira equação, ou seja, para estimar ξ_{1t} e $t=\max(H_2+1, J_2+1), \dots, N$ para estimar ξ_{2t} . Os "valores de partida" da primeira equação ξ_{1t} para $t=\max(H_1, J_1), \dots, \max(H_1, J_1)-I_1+1$, são fixados em zero, o mesmo acontecendo aos da segunda equação ξ_{2t} para

$t = \max(H_2, J_2), \dots, \max(H_2, J_2) - I_2 + 1$. Assim, os parâmetros (7.1.22) são estimados através do método dos mínimos quadrados não linear.

Após a estimação, é necessário proceder à confirmação do diagnóstico, onde deve ser conferida especial atenção à estratégia de sobreparametrização. De qualquer modo, deve ser sempre tomada grande precaução quanto à possibilidade de existência de parâmetros redundantes ou não significativos durante todo o processo de modelização. Em particular, dado o grande número de parâmetros envolvidos, mesmo quando a ordem dos diferentes polinómios é baixa, deve sempre ser considerada a possibilidade de existência de factores comuns. As simplificações que podem assim ocorrer no modelo são um dos principais motivos para a realização deste passo intermédio que consiste na estimação e confirmação do modelo que relaciona as variáveis residuais dos modelos univariados (7.1.19). Tal procedimento permitirá simplificar significativamente a forma do modelo que será construído na fase que se segue e, por outro lado, a existência de quaisquer deficiências da sua estrutura é mais facilmente rectificadas em termos do modelo que relaciona as variáveis residuais, uma vez que se consegue obter assim uma ideia muito mais clara sobre onde se encontram essas falhas.

(v) O passo seguinte consiste em combinar o modelo (7.1.22), que relaciona as variáveis residuais, com os modelos (7.1.19), de modo a sugerir uma estrutura adequada forma (7.1.15) que permita descrever a relação entre as séries originais.

Essa combinação tem em vista a identificação, baseada nas fases anteriores, do modelo que relaciona as séries originais e processa-se expressando os modelos (7.1.19) da forma:

$$\eta_u = \frac{\phi_i(B)}{\theta_i(B)} X_u \quad i = 1, 2 \quad (7.1.40)$$

Em seguida, substitui-se as expressões obtidas para η_{it} no modelo (7.1.22), produzindo o seguinte resultado:

$$\begin{aligned} X_{1t} &= h_1 \frac{\theta_1(B)w_1(B)\phi_2(B)}{\phi_1(B)\delta_1(B)\theta_2(B)} X_{2t} + \frac{\theta_1(B)\psi_1(B)}{\phi_1(B)\phi_1(B)} \xi_{1t} \\ X_{2t} &= h_2 \frac{\theta_2(B)w_2(B)\phi_1(B)}{\phi_2(B)\delta_2(B)\theta_1(B)} X_{1t} + \frac{\theta_2(B)\psi_2(B)}{\phi_2(B)\phi_2(B)} \xi_{2t} \end{aligned} \quad (7.1.41)$$

sendo $h_1=[V(\eta_{1t})/V(\eta_{2t})]^{1/2}$ e $h_2=h_1^{-1}$ uma vez que o modelo (7.1.22) foi ajustado após a normalização de η_{1t} e de η_{2t} .

Este modelo constitui o resultado final de todo o processo de modelização, sendo agora necessário estimá-lo, o que pode ser feito tal como foi descrito para o modelo (7.1.22), bem como proceder à confirmação do diagnóstico de modo a obter-se um modelo que descreva adequadamente a relação entre X_{1t} e X_{2t} . Uma vez que, como é bem visível, se corre um forte risco de (7.1.41) conduzir a uma estrutura sobreparametrizada, é necessário ter um cuidado especial durante a confirmação do diagnóstico com a possibilidade de eliminação de factores comuns, de modo a evitar redundâncias, e com a possível existência de parâmetros não significativos.

Em resumo, conclui-se que, à semelhança do caso univariado, o objectivo da modelização de uma série cronológica bivariada consiste em transformar um par de séries cronológicas univariadas num par de ruídos brancos não correlacionados entre si. Na prática, ajusta-se à série bivariada um modelo que dê origem a resíduos de estimação cujas propriedades estejam tão próximas quanto possível das de dois processos puramente aleatórios não correlacionados entre si. No entanto, a construção de um modelo não pode ser baseada apenas nas indicações fornecidas, e cegamente seguidas, por um conjunto de momentos de segunda ordem estimados, sendo fundamental conseguir distinguir quais

as relações que são plausíveis e que estão subjacentes a determinados fenómenos observados, das que não o são. Assim, o objectivo de produzir modelos que dêem origem a resíduos de estimação que se aproximem muito de ruídos brancos, pode conflitar com o de produzir modelos razoáveis, conduzindo à necessidade de uma escolha extremamente delicada. Neste contexto, Granger e Newbold (1977a, pág. 262) propõem que, além do critério da parcimónia, que se destina a conduzir a uma "parametrização parcimoniosa", se adopte também o critério da plausibilidade, destinado a conduzir a uma "parametrização plausível". É evidente que qualquer princípio cegamente obedecido não tem qualquer valor, pelo que um modelo "plausível" é um fraco substituto de um "implausível" que sistematicamente lhe seja superior no que respeita à qualidade das previsões. De qualquer modo, o critério da plausibilidade será um princípio importante a observar na modelização de qualquer série cronológica, revestindo-se de grande importância no caso multivariado geral e, em particular, no caso bivariado.

Estas considerações são extremamente relevantes em todo o processo de modelização acabado de analisar. Com efeito, quer o critério de parametrização parcimoniosa quer o de parametrização plausível deverão ser cuidadosamente observados ao por em prática a proposta de Granger e Newbold, de modo a evitar estruturas incluindo um número excessivo de parâmetros, com todas as consequências nefastas que isso acarreta, pois é um risco que se coloca ao modelizar qualquer série cronológica bivariada através da utilização deste método. Deste modo, é importante dedicar grande atenção à possibilidade de simplificação de um dado modelo, de forma a evitar a identificação de um modelo implausível ou não parcimonioso, o que é tanto mais importante quanto mais elaborado ele for, uma vez que este método conduz habitualmente a modelos de estrutura pesada, envolvendo um elevado número de

parâmetros. No entanto, o resultado a que a observação destes critérios possa conduzir tem que ser validado, o que pode ser feito através da análise da qualidade das previsões geradas pelo modelo obtido.

Por outro lado, tal como já foi referido, este método é muito limitado, não sendo adequado para modelizar uma série de elevada dimensão, o que é reconhecido pelos seus autores. Na realidade, não parece poder ir muito além do caso bivariado, adaptando-se bem apenas neste caso. Assim, descrever-se-à as principais dificuldades que se colocam a esta metodologia, tornando-a incapaz de tratar o caso multivariado geral, e que são :

a) Em primeiro lugar, basta observar às equações (7.1.25) para compreender que a sua utilização se torna extremamente pesada ou mesmo inviável quando m aumenta, uma vez que, quando isso acontece, o número das suas parcelas aumenta muito mais rapidamente. Assim, o esquema recursivo delineado para a resolução destas equações não parece permitir elevados valores de m , até porque isso provocaria grandes dificuldades de convergência desse esquema. Além disso, a existência de um maior número de correlações cruzadas provocaria maior distorção nos valores estimados dos elementos dos polinómios $V_i(B)$, $i=1, \dots, m$, por um lado e, por outro, a utilização de $M=2,3,4, \dots$ acarretaria ainda maiores problemas de obtenção da solução das equações (7.1.25).

De modo semelhante, torna-se muito mais difícil obter a solução das equações (7.1.35) quando m aumenta, uma vez que o número das suas parcelas cresce a um ritmo muito superior. Por outro lado, quer a existência de maior número de correlações cruzadas estimadas, quer os maiores erros na estimação de V_{ik} ($i=1, \dots, m$), que já foram referidos, são factores que provocam maiores distorções na estimação da função de autocovariância das variáveis residuais e_{it} , o que torna a qualidade da identificação da estrutura dos respectivos polinómios bastante inferior.

b) Por outro lado, uma questão de extrema acuidade é a possibilidade da obtenção de estruturas sobreparametrizadas, com todas as consequências daí resultantes, o que é tanto mais grave quanto maior m , tratando-se já de um problema importante quando $m=2$. Assim, o número de parâmetros do modelo (7.1.15) aumenta rapidamente, o mesmo sucedendo a (7.1.22), relativamente ao qual é possível mostrar que, quando o modelo adequado ao vector X_t é um modelo ARMA(p,q), mesmo que seja de baixa ordem, o modelo correspondente para o vector composto pelas variáveis residuais dos diferentes modelos univariados pode ser muito complexo e difícil de identificar na prática, estando os seus parâmetros sujeitos a várias e complicadas restrições não lineares (Tiao e Box, 1979, 1981).

No entanto, este problema assume a sua maior gravidade para o modelo (7.1.41), ou seja, para o modelo resultante da combinação dos modelos univariados com o modelo que relaciona as variáveis residuais η_{it} . Com efeito, este último modelo apresenta, em termos genéricos, uma estrutura de tal forma pesada, sobreparametrizada, que se torna difícil de tratar, mesmo que haja factores comuns que seja possível eliminar. Logo, se já no caso bivariado o problema da obtenção de uma estrutura com um número de parâmetros demasiado elevado assumia grande gravidade, tal como já foi aflorado, isso tem muito maior relevância para valores superiores de m .

Em resumo, este método conduz a modelos que geralmente não respeitam nem o princípio da parcimónia nem o da plausibilidade.

c) Finalmente, coloca-se o problema do enfraquecimento das relações entre as variáveis residuais dos modelos univariados. Intuitivamente, parece admissível que as relações entre as variáveis residuais η_{it} ($i=1, \dots, m$) sejam em geral mais fracas do que as relações entre as séries X_{it} . Isto acontece essencialmente porque, quando X_{it} está individualmente relacionada com os seus próprios valores passados $X_{i,t-1}, X_{i,t-2}, \dots$, estes

servem como aproximação aos valores passados das outras séries X_{jt} ($i \neq j$), o que permite que as relações dinâmicas entre as diferentes séries sejam parcialmente levadas em linha de conta pelos valores passados das séries individuais. Deste modo, tal como é apontado por Tiao e Box (1979), conclui-se que a estrutura que relaciona as séries η_{it} é bastante mais fraca, pelo que parece poder afirmar-se que é mais eficaz modelizar X_t directamente do que através da modelização do vector das variáveis residuais dos modelos univariados. Este problema volta a colocar-se no caso da relação unidireccional, vindo também a ser analisado nesse contexto.

Deste modo, conclui-se que este método se revela inapropriado para a modelização de séries cronológicas de dimensão elevada ou mesmo apenas moderada.

Após esta análise, tratar-se-à agora o caso em que a relação existente entre as duas séries univariadas se verifica apenas num sentido, ou seja, é unidireccional.

1.3.2. Relação unidireccional

A situação mais simples possível na análise de séries cronológicas multivariadas é o caso de uma série bivariada em que existe uma relação num único sentido, ou seja, em que ela é unidireccional. Trata-se de uma análise que entronca com a teoria dos sistemas lineares, adoptando-se no entanto uma perspectiva de modelização de uma série cronológica, o que significa que não se vai abordar os métodos que respeitem àquela teoria.

Assim, sendo X_{1t} e X_{2t} conjuntamente estacionários, com $E(X_{1t})=E(X_{2t})=0$, o tipo de modelos que irá ser analisado assume a forma genérica:

$$X_{1t} = V'(B)X_{2t} + U'(B)\varepsilon_t \quad (7.1.42)$$

onde $\{\varepsilon_t\}$ é um processo puramente aleatório, tendo-se as relações :

$$V'(B) = \frac{w'(B)}{\delta'(B)} \quad ; \quad U'(B) = \frac{\psi'(B)}{\phi'(B)} \quad (7.1.43)$$

onde $w'(B)$, $\delta'(B)$, $\psi'(B)$ e $\phi'(B)$ são polinômios em B de ordem finita, cujos zeros se situam fora do círculo unitário e assumindo as expressões:

$$\begin{aligned} V'(B) &= V_0' + V_1'B + V_2'B^2 + \dots \\ U'(B) &= U_0' + U_1'B + U_2'B^2 + \dots \\ w'(B) &= w_0' + w_1'B + \dots + w_l'B^l \\ \delta'(B) &= 1 - \delta_1'B - \dots - \delta_n'B^n \end{aligned} \quad (7.1.44)$$

Habitualmente, os modelos da forma de (7.1.42) são designados por modelos de regressão dinâmica. Por outro lado, segundo a terminologia adoptada na teoria dos sistemas lineares, X_{2t} é designado de "input", X_{1t} de "output", e_t' de ruído, sendo $e_t' = U'(B)\epsilon_t$, e os elementos V_k' ($k=0,1,2,\dots$) constituem o que é designado de função impulso-resposta. É evidente que é possível alargar o modelo (7.1.44) de modo a incluir vários "inputs" com um só "output", ou vários "inputs" causando vários "outputs", conduzindo este segundo caso à definição de um sistema multi-equacional. No entanto, em qualquer dos casos, isso constituiria um aprofundamento de uma matéria que pertence à teoria dos sistemas lineares, não sendo por isso analisado aqui, resumindo-se o estudo ao caso mais simples possível de existência de um único "input" e de um único "output".

Deste modo, serão analisados quatro métodos visando a construção de modelos da forma (7.1.42): o de Sims (1972), o de Box e Jenkins (1976), o de Priestley (1971) e o de Granger e Newbold (1977a).

1.3.2.1. Método de Sims

Sims (1972) apresenta um método de tratamento de modelos do tipo de (7.1.42), segundo o qual as variáveis são transformadas através da

aplicação de um mesmo filtro, definido "a priori", cujo objectivo é eliminar a autocorrelação existente em e'_t na seguinte equação:

$$X_{1t} = V'(B)X_{2t} + e'_t \quad (7.1.45)$$

onde $e'_t = U'(B)\varepsilon_t$ e se admite que X_{2t} não está correlacionado com e'_t . Assim, este método possui duas características principais: por um lado, o filtro acabado de referir é escolhido antecipadamente, em vez de a sua escolha ser feita a partir de um estudo empírico dos dados em análise e, por outro, o operador $V'(B)$ é truncado num dado valor M considerado suficientemente elevado para incluir todos os parâmetros que se espera sejam significativos.

Desta forma, designando o filtro a aplicar a (7.1.45) por $P(B)$, este procedimento conduz a um modelo da forma:

$$Y_{1t} = Q(B)Y_{2t} + e_t \quad (7.1.46)$$

onde $Y_{it} = P(B)X_{it}$ ($i=1,2$), $e_t = P(B)e'_t$ e $Q(B)$ representa o polinómio $V'(B)$ truncado no ponto M .

Sims (1972) utiliza o filtro $P(B) = (1 - 0.75B)^2$ na maior parte dos casos, admitindo que a sua aplicação consiga eliminar ou, pelo menos, atenuar fortemente a autocorrelação existente em e'_t . No entanto, se permanecer autocorrelação em e_t , os estimadores de mínimos quadrados dos parâmetros de $Q(B)$ continuarão a ser consistentes, mas os estimadores das suas variâncias serão não centrados, o que acarreta diversos problemas. De qualquer modo, uma vez que o filtro $P(B)$ se destina a eliminar a autocorrelação existente em e'_t , é perfeitamente possível que permaneça uma forte autocorrelação em Y_{2t} e Y_{1t} , pelo que a que permanece nesta última aparecerá necessariamente em e_t .

Em resumo, este método apresenta dois grandes problemas:

a) Em primeiro lugar, a escolha do ponto de truncagem M torna necessário decidir previamente o número de parâmetros a incluir em (7.1.46), em vez de tal decisão ser tomada a partir do estudo das observações disponíveis.

b) Em segundo lugar, a decisão tomada "a priori" sobre o filtro $P(B)$ pode conduzir à escolha de um filtro que não seja o mais adequado, em vez de essa decisão ser tomada com base nas observações disponíveis, de modo a que a forma de $P(B)$ seja a mais apropriada. A escolha assim realizada conduz à necessidade de realização de testes sobre a existência de autocorrelação em e_t , o que tem ainda o inconveniente de esses testes poderem não ser os mais adequados para o padrão da autocorrelação que, num dado caso concreto, eventualmente permaneça em e_t .

Conclui-se então que é necessário determinar o filtro $P(B)$ a partir da série cronológica em análise, o mesmo sucedendo com a forma do polinómio $V'(B)$, em vez de essas escolhas serem realizadas "a priori". Os métodos que irão ser analisados a seguir pretendem tomar essas questões em linha de conta.

1.3.2.2. Método de Box e Jenkins

Box e Jenkins (1976) propõem um método bastante simples para a identificação de modelos (7.1.42), tendo sido pioneiros na introdução deste tipo de modelos ⁽⁹⁾.

O método proposto por estes autores processa-se de acordo com as seguintes fases:

(i) Em primeiro lugar, é necessário estimar os elementos de $V'(B)$. Uma via possível para o efeito consiste em utilizar o modelo expresso da forma:

$$X_{1t} = V_0'X_{2t} + V_1'X_{2,t-1} + V_2'X_{2,t-2} + \dots + e_t' \quad (7.1.47)$$

onde $e_t' = U'(B)\varepsilon_t$.

branco. Logo, procede-se ao ajustamento de um modelo ARMA, que se admite ser invertível e estacionário, a X_{2t} , de que resulta:

$$\phi_2(B)X_{2t} = \theta_2(B)\eta_{2t} \quad (7.1.51)$$

onde $\{\eta_{2t}\}$ é um processo puramente aleatório que se admite ser não correlacionado com \mathcal{E}_t .

Em seguida, define-se a variável:

$$Z_t = \theta_2(B)^{-1} \phi_2(B)X_{2t} \quad (7.1.52)$$

Multiplicando (7.1.42) por $\theta_2(B)^{-1}\phi_2(B)$ obtém-se o resultado:

$$Z_t = V'(B)\eta_{2t} + \alpha_t \quad (7.1.53)$$

onde $\alpha_t = \theta_2(B)^{-1} \phi_2(B)e_t$. Multiplicando agora (7.1.53) por $\eta_{2,t-k}$ e calculando valores esperados, uma vez que η_{2t} e \mathcal{E}_t são não correlacionados, obtém-se a igualdade:

$$\gamma_{z\eta_2}(k) = V'_k V(\eta_{2t}) \quad k = 0, 1, 2, \dots \quad (7.1.54)$$

onde $\gamma_{z\eta_2}$ designa a função de covariância cruzada entre Z_t e η_{2t} e $V(\eta_{2t})$ a variância de η_{2t} . A expressão (7.1.54) permite obter os elementos de $V'(B)$ da forma:

$$V'_k = \frac{\gamma_{z\eta_2}(k)}{V(\eta_{2t})} \quad k = 0, 1, 2, \dots \quad (7.1.55)$$

Utilizando funções de correlação, esta expressão vem:

$$V'_k = \rho_{z\eta_2}(k) \left[\frac{V(Z_t)}{V(\eta_{2t})} \right]^{1/2} \quad k = 0, 1, 2, \dots \quad (7.1.56)$$

onde $V(Z_t)$ designa a variância de Z_t .

Portanto, conclui-se que a função de correlação cruzada entre Z_t e η_{2t} é directamente proporcional aos elementos do polinómio $V'(B)$. No entanto, como $\rho_{z\eta_2}(k)$, $V(Z_t)$ e $V(\eta_{2t})$ são desconhecidas, é necessário estimá-las, o que conduz ao estimador de V'_k seguinte:

$$\hat{V}'_k = \hat{r}_{z\eta_2}(k) \left[\frac{\hat{V}(\hat{Z}_t)}{\hat{V}(\hat{\eta}_{2t})} \right]^{1/2} \quad (7.1.57)$$

onde $\hat{r}_{z\eta_2}(k)$ designa o estimador da função de correlação cruzada entre \hat{Z}_t e $\hat{\eta}_{2t}$, sendo $\hat{Z}_t = \hat{\theta}_2(B)^{-1} \hat{\phi}_2(B) X_{1t}$ e designando $\hat{\eta}_{2t}$ os resíduos de estimação do modelo univariado (7.1.51).

Os estimadores preliminares \hat{V}'_k assim obtidos não são geralmente eficientes, mas constituem uma base para a identificação da estrutura dos polinómios $w'(B)$ e $\delta'(B)$, o que constitui a segunda fase do método.

(ii) A partir da estimação do polinómio $V'(B)$, é possível propor valores para l e n , ou seja, identificar a estrutura de $w'(B)$ e de $\delta'(B)$. Tal é realizado utilizando o mesmo processo da fase (ii) descrita no caso da relação bidireccional, sendo em particular necessário recorrer às equações (7.1.32) e à análise subsequente.

(iii) Finalmente, é necessário identificar a componente de ruído, ou seja, a componente residual. Assim, a partir da estimação do polinómio $V'(B)$, obtém-se o modelo:

$$X_{1t} = \hat{V}'_k(B) X_{2t} + e'_t \quad (7.1.58)$$

A partir de (7.1.58), é possível estimar a variável residual e'_t , ou seja, determinar os resíduos de estimação do modelo (7.1.47), o que é feito do seguinte modo:

$$\hat{e}'_t = X_{1t} - \hat{V}'_k(B) X_{2t} \quad (7.1.59)$$

Uma forma alternativa de estimar e'_t consiste na estimação prévia de $w'(B)$ e de $\delta'(B)$, a partir de $V'(B)$ e no cálculo posterior de \hat{e}'_t da forma:

$$\hat{e}'_t = X_{1t} - \frac{\hat{w}'(B)}{\hat{\delta}'(B)} X_{2t} \quad (7.1.60)$$

Assim, obtém-se \hat{e}'_t a partir da igualdade:

$$\begin{aligned} \hat{e}'_t = X_{1t} + \hat{\delta}'_1 (\hat{e}'_{t-1} - X_{1,t-1}) + \dots + \hat{\delta}'_n (\hat{e}'_{t-n} - X_{1,t-n}) - \\ - \hat{w}'_0 X_{2t} - \hat{w}'_1 X_{2,t-1} - \dots - \hat{w}'_l X_{2,t-l} \end{aligned} \quad (7.1.61)$$

Os parâmetros de $w'(B)$ e de $\delta'(B)$ podem ser preliminarmente estimados utilizando as equações (7.1.32), onde os elementos V'_k são substituídos pelos seus estimadores \hat{V}'_k determinados na fase (i). Os estimadores de $w'(B)$ e de $\delta'(B)$ assim obtidos são utilizados para o cálculo de \hat{e}'_t em (7.1.61)

Após a estimação de \hat{e}'_t , quer seja a partir de (7.1.59), quer de (7.1.61), é possível estimar a sua função de autocorrelação, o que permite identificar um modelo ARMA univariado para \hat{e}'_t que assume a forma :

$$\phi'(B)e'_t = \psi'(B)\varepsilon_t \quad (7.1.62)$$

uma vez que $\{\varepsilon_t\}$ é um processo puramente aleatório. Fica assim identificada a componente residual do modelo (7.1.42), sendo no entanto possível conseguir este objectivo de uma forma alternativa, que consiste em utilizar as funções de autocorrelação do modelo (7.1.53), que pode expresso do seguinte modo:

$$Z_t = Y_t + \alpha_t \quad (7.1.63)$$

sendo $Y_t = V'(B)\eta_{2t}$. Uma vez que se verifica a relação:

$$\alpha_t = \theta_2(B)^{-1} \phi_2(B)e'_t \quad (7.1.64)$$

é possível, utilizando (7.1.64), identificar a estrutura da componente residual do modelo (7.1.42) a partir de um modelo identificado para α_t , o que implica a necessidade de conhecimento da função de autocorrelação desta última. Assim, uma vez que se admite que Y_t e α_t não estão correlacionadas, o que resulta da hipótese de inexistência de correlação entre X_{2t} e e_t e, logo, entre X_{2t} e ε_t , é possível obter a partir de (7.1.63) a igualdade:

$$\gamma_{zz}(k) = \gamma_{yy}(k) + \gamma_{\alpha\alpha}(k) \quad (7.1.65)$$

onde $\gamma_{zz}(k)$, $\gamma_{yy}(k)$ e $\gamma_{\alpha\alpha}(k)$ designam respectivamente as funções de autocovariância de Z_t , de Y_t e de α_t . Uma vez que $\{\eta_{2t}\}$ é um ruído branco, $\gamma_{yy}(k)$ pode ser expressa da forma (Box e Jenkins, 1976, pág. 81):

$$\gamma_{yy}(k) = V(\eta_{2t}) \sum_{r=0}^{\infty} V_r' V_{r+k}' = \frac{1}{V(\eta_{2t})} \sum_{r=0}^{\infty} \gamma_{z\eta_2}(r) \gamma_{z\eta_2}(r+k) \quad (7.1.66)$$

onde se utilizou (7.1.54). Deste modo, utilizando (7.1.65), obtém-se os resultados:

$$\begin{aligned} \gamma_{\alpha\alpha}(k) &= \gamma_{zz}(k) - \frac{1}{V(\eta_{2t})} \sum_{r=0}^{\infty} \gamma_{z\eta_2}(r) \gamma_{z\eta_2}(r+k) \\ \gamma_{\alpha\alpha}(0) &= \gamma_{zz}(0) - \frac{1}{V(\eta_{2t})} \sum_{r=0}^{\infty} \gamma_{z\eta_2}^2(r) \\ \rho_{\alpha\alpha}(k) &= \frac{\rho_{zz}(k) - \sum_{r=0}^{\infty} \rho_{z\eta_2}(r) \rho_{z\eta_2}(r+k)}{1 - \sum_{r=0}^{\infty} \rho_{z\eta_2}^2(r)} \end{aligned} \quad (7.1.67)$$

O conhecimento desta função de autocorrelação permite a identificação de um modelo para α_t , possibilitando a dedução de uma estrutura adequada para a componente residual do modelo (7.1.42). Mas, na prática, é

necessário substituir as funções de (7.1.67) pelos seus estimadores, o que significa que \hat{Z}_t e $\hat{\eta}_{2t}$ terão que ser utilizados em vez de Z_t e de η_{2t} respectivamente, estimando-se desta forma a função de autocorrelação de α_t .

A identificação do modelo (7.1.42) fica assim completa, permitindo a sua estimação através da utilização das técnicas habituais.

Este método apresenta um inconveniente importante, que consiste na possibilidade de a identificação do modelo (7.1.42) falhar ou, pelo menos, conduzir a maus resultados, sugerindo que X_{1t} e X_{2t} não estão relacionados, mesmo em casos em que exista forte evidência sobre essa relação. Downing e Pack (1982) afirmam que isso pode acontecer quando a variabilidade de X_{1t} é mais determinada pela variabilidade da componente residual do que pela de X_{2t} . Assim, lembre-se que os parâmetros de $V(B)$ dependem das variâncias das séries resultantes de X_{1t} e de X_{2t} após as operações de filtragem, tendo essas séries sido designadas respectivamente por Z_t e η_{2t} , e que o mesmo se verifica quando essas grandezas são estimadas, o que é bem evidente em (7.1.56) e (7.1.57). Conclui-se então que é necessário que a contribuição da componente residual para a variabilidade de X_{1t} seja bastante inferior à de X_{2t} , para que seja possível obter uma boa imagem da realidade através deste método, ou seja, para conseguir uma correcta identificação.

1.3.2.3. Método de Priestley

Priestley (1971) propõe um método que designa de "contração de covariância", partindo da especificação de um modelo com um número finito de parâmetros e assumindo a forma:

$$X_{1t} - \delta'_1 X_{1,t-1} - \dots - \delta'_n X_{1,t-n} = w'_0 X_{2t} + w'_1 X_{2,t-1} + \dots + w'_l X_{2,t-l} + \alpha_t \quad (7.1.68)$$

A preferência por um modelo deste tipo, em vez de um da forma (7.1.42), deve-se ao facto de modelos deste segundo tipo envolverem um número de parâmetros arbitrariamente elevado, assumindo uma forma que poderá não ser adequada para ajustar uma relação linear entre X_{1t} e X_{2t} a partir de um número finito de observações. Em particular, um modelo como (7.1.42) poderá revelar-se inadequado se se pretender prever valores futuros de X_{1t} a partir de observações de X_{2t} .

Assim, (7.1.68) pode ser expresso do seguinte modo:

$$\delta'(B)X_{1t} = w'(B)X_{2t} + \alpha_t \quad (7.1.69)$$

onde $\delta'(B) = 1 - \delta_1 B - \dots - \delta_n B^n$; $w'(B) = w_0 + w_1 B + \dots + w_l B^l$.

É útil notar que (7.1.68) pode ser reescrito na forma alternativa :

$$X_{1t} = \delta^{-1}(B)w'(B)X_{2t} + \delta^{-1}(B)\alpha_t \quad (7.1.70)$$

Quando os valores de n e de l são conhecidos, é possível proceder à estimação dos parâmetros de $\delta'(B)$ e de $w'(B)$ para o que Priestley (1971) propõe o método dos mínimos quadrados, uma vez que este é em muitos casos razoavelmente eficiente, apesar de geralmente se verificar a existência de autocorrelação em α_t .

Deste modo, uma vez que n e l são desconhecidos, é proposto o método de "contração de covariância" para a identificação da estrutura de $\delta'(B)$ e de $w'(B)$, desenvolvendo-se de acordo com as fases:

(i) Em primeiro lugar, é necessário ajustar os seguintes modelos ARMA univariados invertíveis e estacionários:

$$\phi_i(B)X_{it} = \theta_i(B)\eta_{it} \quad i=1,2 \quad (7.1.71)$$

onde $\{\eta_{it}\}$, $i=1,2$, são processos puramente aleatórios.

(ii) O passo seguinte consiste no ajustamento de um modelo relacionando as variáveis residuais dos modelos (7.1.71), assumindo a forma:

$$\eta_{1t} - \delta_1 \eta_{1,t-1} - \dots - \delta_g \eta_{1,t-g} = w_0 \eta_{2t} + w_1 \eta_{2,t-1} + \dots + w_r \eta_{2,t-r} + W_t \quad (7.1.72)$$

ou seja:

$$\delta(B)\eta_{1t} = w(B)\eta_{2t} + W_t \quad (7.1.73)$$

admitindo-se que W_t não está correlacionado com η_{2t} .

É necessário agora determinar a estrutura de $\delta(B)$ e de $w(B)$, o que é possível de conseguir através da função de covariância cruzada entre $\{\eta_{1t}\}$ e $\{\eta_{2t}\}$ uma vez que ambos são ruídos brancos. Assim, considere-se a equação:

$$\chi_t = \eta_{1t} - \delta_1 \eta_{1,t-1} - \dots - \delta_g \eta_{1,t-g} = \delta(B)\eta_{1t} \quad (7.1.74)$$

Esta igualdade permite expressar o modelo (7.1.73) da forma :

$$\chi_t = w_0 \eta_{2t} + w_1 \eta_{2,t-1} + \dots + w_r \eta_{2,t-r} + W_t \quad (7.1.75)$$

A função de covariância cruzada entre χ_t e η_{2t} , designada por $\gamma_{\chi\eta_2}(k)$, assume a forma:

$$\gamma_{\chi\eta_2}(k) = E(\chi_t \eta_{2,t-k}) = \begin{cases} w_k & k = 0, 1, \dots, r \\ 0 & \text{outros valores de } k \end{cases} \quad (7.1.76)$$

A expressão (7.1.76) resulta do facto de $\{\eta_{2t}\}$ ser um ruído branco e não estar correlacionado com W_t . Portanto, a covariância cruzada entre χ_t e η_{2t} para o "lag" k é igual ao parâmetro w_k , assumindo valores não nulos apenas para os "lags" $k=0,1,\dots,r$ desde que o polinómio $w(B)$ seja finito.

Por outro lado, a função de covariância cruzada entre η_{1t} e η_{2t} , designada por $\gamma_{\eta_{12}}(k)$, não assumirá geralmente o valor zero após um número finito de termos, pois o desenvolvimento em série de

$\{\delta^{-1}(B)w(B)\}$ origina uma série infinita. No entanto, existe a seguinte relação entre $\gamma_{\eta_{12}}(k)$ e $\gamma_{\chi\eta_2}(k)$:

$$\gamma_{\chi\eta_2}(k) = \delta(B)\gamma_{\eta_{12}}(k) \quad (7.1.77)$$

onde o operador B actua sobre k, isto é, $B\gamma_{\eta_{12}}(k) = \gamma_{\eta_{12}}(k-1)$.

A demonstração de (7.1.77) é muito simples:

$$\begin{aligned} \gamma_{\chi\eta_2}(k) &= E(\chi_t \eta_{2,t-k}) = \\ &= E[\delta(B)\eta_{1t} \eta_{2,t-k}] = \\ &= E(\eta_{1t} \eta_{2,t-k} - \sum_{n=1}^k \delta_n \eta_{1,t-n} \eta_{2,t-k}) = \\ &= \gamma_{\eta_{12}}(k) - \sum_{n=1}^k \delta_n \eta_{1,t-n} \eta_{2,t-k}(k-n) = \\ &= \delta(B)\gamma_{\eta_{12}}(k) \end{aligned} \quad (7.1.78)$$

e modo, o polinómio $\delta(B)$ pode ser encarado como o operador que "contraí" $\gamma_{\eta_{12}}(k)$ até a tornar numa função da forma (7.1.76), sendo necessário na prática recorrer ao estimador de $\gamma_{\eta_{12}}(k)$. Assim, se apresentar um decréscimo lento e suave, $\hat{r}_{\eta_{12}}(k)$ pode ser contraída, ou seja, transformada na forma desejada, através de aplicações sucessivas do operador diferença $(1-B)$, o que significa que, neste caso, $\delta(B)$ assumiria a forma:

$$\delta(B) = (1-B)^d \quad (7.1.79)$$

onde d é o número de vezes que o operador diferença é aplicado. Se, por outro lado, $\hat{r}_{\eta_{12}}(k)$ apresentar uma forma "oscilatória", pode ser contraída utilizando por exemplo um filtro de médias móveis, o que conduziria à estrutura de $\delta(B)$:

$$\delta(B) = \frac{1}{2d+1} (B^{-d} + B^{-d-1} + \dots + 1 + B + \dots + B^d) \quad (7.1.80)$$

Em ambos os casos, é possível seleccionar um valor adequado para d , iniciando o procedimento com um valor baixo (habitualmente $d=1$) e aumentando o valor de d até que a função de covariância cruzada transformada evidencie o grau de "contração" requerido.

Uma vez determinada a forma de $\delta(B)$, é possível inferir imediatamente a de $w(B)$ a partir da função de covariância cruzada contraída e utilizando o resultado (7.1.76). Na prática, os elementos de $w(B)$ são obtidos através da determinação dos "lags" para os quais $\hat{r}_{\chi\eta_{12}}(k)$ diferem significativamente de zero.

Após a identificação da estrutura de $\delta(B)$ e de $w(B)$, os parâmetros respectivos podem ser estimados através do método dos mínimos quadrados, isto é, através da regressão de η_{1t} sobre $\eta_{1,t-1}, \dots, \eta_{1,t-g}, \eta_{2t}, \dots, \eta_{2,t-r}$ embora $\{W_t\}$ não seja geralmente um ruído branco.

(iii) Finalmente, após a identificação dos operadores $\delta(B)$ e $w(B)$ e da estimação dos seus parâmetros, resta identificar o modelo que relaciona X_{1t} e X_{2t} .

Assim, visando a identificação da estrutura de $\delta(B)$ e de $w(B)$, exprime-se η_{1t} e η_{2t} em função de X_{1t} e X_{2t} respectivamente, utilizando os modelos (7.1.71), e substitui-se a expressão assim obtida em (7.1.73):

$$\begin{aligned} \eta_{1t} &= \theta_1(B)^{-1} \phi_1(B) X_{1t} \\ \eta_{2t} &= \theta_2(B)^{-1} \phi_2(B) X_{2t} \end{aligned} \quad (7.1.81)$$

A substituição destas expressões em (7.1.73) conduz ao resultado:

$$\delta(B) \theta_1(B)^{-1} \phi_1(B) X_{1t} = w(B) \theta_2(B)^{-1} \phi_2(B) X_{2t} + W_t \quad (7.1.82)$$

ou seja:

$$\delta(B)\theta_2(B)\phi_1(B)X_{1t} = w(B)\theta_1(B)\phi_2(B)X_{2t} + \alpha_t \quad (7.1.83)$$

onde $\alpha_t = \theta_1(B)\theta_2(B)W_t$. Uma vez que W_t e η_{2t} não estão correlacionados, o mesmo acontece com α_t e X_{2t} . O modelo (7.1.83) pode ainda ser expresso da forma:

$$\delta'(B)X_{1t} = w'(B)X_{2t} + \alpha_t \quad (7.1.84)$$

onde:

$$\begin{aligned} \delta'(B) &= \delta(B)\theta_2(B)\phi_1(B) = 1 - \delta'_1 B - \dots - \delta'_n B^n \\ w'(B) &= w(B)\theta_1(B)\phi_2(B) = w'_0 + w'_1 B - \dots - w'_l B^l \end{aligned} \quad (7.1.85)$$

A identificação da estrutura de $\delta'(B)$ e de $w'(B)$ é assim conseguida, pelo que os seus parâmetros podem agora ser estimados mais eficientemente através da regressão de X_{1t} sobre $X_{1,t-1}, \dots, X_{1,t-n}, X_{2t}, X_{2,t-1}, \dots, X_{2,t-l}$.

À primeira vista, pode parecer que o ajustamento descrito em (ii) de um modelo relacionando η_{1t} e η_{2t} envolve exactamente as mesmas dificuldades que surgem no ajustamento de um modelo relacionando as séries originais X_{1t} e X_{2t} . No entanto, a estratégia acabada de analisar baseia-se na consideração de que a forma dos operadores $\delta'(B)$ e $w'(B)$ dependem da estrutura de autocorrelação de X_{1t} e de X_{2t} bem como da estrutura da respectiva correlação cruzada. Ao ajustar modelos univariados a X_{1t} e a X_{2t} , converte-se ambas as séries aproximadamente em ruídos brancos ou seja, elimina-se as respectivas estruturas de autocorrelação, pelo que é de esperar que, em geral, as formas dos operadores $\delta(B)$ e $w(B)$ sejam consideravelmente mais simples que as de $\delta'(B)$ e de $w'(B)$. Por outro lado, conforme já foi referido, as estruturas dos operadores $\delta(B)$ e $w(B)$ podem ser indicadas a partir da função de correlação cruzada entre η_{1t} e η_{2t} , uma vez que ambos são ruídos brancos.

Assim, o passo intermédio de ajustamento de um modelo relacionando η_{1t} e η_{2t} é uma fase de fundamental importância para a identificação do modelo que relaciona X_{1t} e X_{2t} .

No entanto, este método apresenta algumas desvantagens, tal como referem Pierce e Haugh (1977), que são resumidamente:

a) Em primeiro lugar, a estrutura da autocorrelação da variável residual é ignorada em várias fases do método, tal como em (7.1.73), o que pode originar diversos problemas no caso de se verificar a existência de autocorrelação em W_t , conforme já foi referido a propósito do método de Sims.

b) Em segundo lugar, se o operador $\delta(B)$ necessário para contrair $\hat{\eta}_{12}(k)$ incluir $(1-B)$, conclui-se que o processo $\{\eta_{1t}\}$ não é geralmente estacionário, o que contraria os resultados da primeira fase do método, em $\theta_1(B)^{-1}\phi_1(B)$ é identificado e ajustado de modo a que $\{\eta_{1t}\}$ seja um processo puramente aleatório estacionário.

Por outro lado, este método tem o mérito de utilizar filtros diferentes para X_{1t} e X_{2t} , adequados para a conversão em ruídos brancos e determinados a partir da série cronológica em análise, o que é também uma característica do método seguinte.

1.3.2.4. Método de Granger e Newbold

O método proposto por Granger e Newbold (1977a), sendo do mesmo género que o método de Haugh e Box (1977), é o mesmo que foi aplicado no caso da relação bidireccional, apresentando ainda algumas simplificações, como seria de esperar.

Assim, a especificação do modelo é a seguinte:

$$\begin{aligned} X_{1t} &= V'(B)X_{2t} + U'(B)\varepsilon_t \\ \phi_2(B)X_{2t} &= \theta_2(B)\eta_{2t} \end{aligned} \quad (7.1.86)$$

o que significa que se ajusta um modelo ARMA univariado invertível e estacionário a X_{2t} , admitindo-se que $\{\varepsilon_t\}$ e $\{\eta_{2t}\}$ são ruídos brancos não correlacionados. Deste modo, o método processa-se percorrendo as seguintes fases:

(i) O primeiro passo consiste em ajustar também um modelo ARMA univariado invertível e estacionário a X_{1t} assumindo a expressão:

$$\phi_1(B)X_{1t} = \theta_1(B)\eta_{1t} \quad (7.1.87)$$

(ii) Em seguida, procede-se ao ajustamento de um modelo relacionando η_{1t} e η_{2t} , o que constitui uma fase intermédia. Uma vez que se admite a existência de uma relação unidireccional entre X_{1t} e X_{2t} , verifica-se o mesmo para η_{1t} e η_{2t} , o que conduz ao modelo:

$$\begin{aligned} \eta_{1t} &= V(B)\eta_{2t} + U(B)\xi_t = \\ &= \frac{w(B)}{\delta(B)}\eta_{2t} + \frac{\psi(B)}{\varphi(B)}\xi_t \end{aligned} \quad (7.1.88)$$

onde η_{1t} e η_{2t} são normalizados de modo a que as suas variâncias sejam unitárias e $\{\xi_t\}$ é um processo puramente aleatório não correlacionado com η_{2t} .

Multiplicando agora o modelo (7.1.88) por $\eta_{2,t-k}$ e calculando valores esperados, obtém-se a seguinte relação, uma vez que ξ_t não está correlacionada com η_{2t} e esta tem variância unitária:

$$\gamma_{\eta_{12}}(k) = E(\eta_{1t}\eta_{2,t-k}) = V_k \quad k=0,1,2,\dots \quad (7.1.89)$$

Conclui-se então que V_k é estimado através do cálculo do estimador de $\gamma_{\eta_{12}}(k)$, o que conduz à equação:

$$\hat{V}_k = \hat{r}_{\eta_{12}}(k) \quad k=0,1,2,\dots \quad (7.1.90)$$

onde $\hat{r}_{\eta_{12}}(k)$ designa como habitualmente o estimador da função de correlação cruzada entre $\hat{\eta}_{1t}$ e $\hat{\eta}_{2t}$.

A partir da estimação dos parâmetros de $V(B)$, é possível identificar a estrutura de $w(B)$ e de $\delta(B)$ utilizando as equações (7.1.32) e toda a análise que lhes é relativa, conforme foi descrito para o método de Box e Jenkins.

(iii) Por outro lado, a identificação da estrutura de $\psi(B)$ e de $\varphi(B)$ pode ser realizada de acordo com o método proposto para a relação bidireccional, sendo no entanto possível recorrer a um procedimento mais simples, que faz apelo ao teorema seguinte (Granger e Newbold, 1977a, pág.28):

Teorema 7.1.1. Se $\{X_{1t}\}$ e $\{X_{2t}\}$ são respectivamente processos ARMA(p,q) e ARMA(u,v) independentes, então $\{Z_t\}$ é um processo ARMA(g,h), onde:

$$\begin{aligned} Z_t &= X_{1t} + X_{2t} \\ g &\leq p + u \\ h &\leq \max\{p + v, q + u\} \end{aligned}$$

O referido procedimento conducente à identificação de $\psi(B)$ e de $\varphi(B)$ parte da constatação de que $\delta(B)$ e $\varphi(B)$ são necessariamente iguais, o que pode ser provado começando por multiplicar (7.1.88) por $\delta(B)$, do que resulta:

$$\delta(B)\eta_{1t} = w(B)\eta_{2t} + \frac{\delta(B)\psi(B)}{\varphi(B)}\xi_t \quad (7.1.91)$$

O primeiro membro desta equação, $\delta(B)\eta_{1t}$, é composto por um operador de médias móveis de ordem finita, sendo o segundo membro a soma de duas componentes não correlacionadas, das quais a primeira é também uma média móvel de ordem infinita. Então, resulta do teorema 7.1.1. que

isso só é possível se a segunda componente do segundo membro de (7.1.91) for também uma média móvel de ordem finita, o que permite concluir que $\varphi(B)$ é um factor de $\delta(B)\psi(B)$. No entanto, ao definir o modelo (7.1.88), admite-se que $\psi(B)$ e $\varphi(B)$ não têm factores comuns, o que implica que $\varphi(B)$ é um factor de $\delta(B)$.

Multiplicando agora (7.1.88) por $\varphi(B)$, obtém-se a equação:

$$\varphi(B)\eta_{1t} = \frac{\varphi(B)w(B)}{\delta(B)}\eta_{2t} + \psi(B)\xi_t \quad (7.1.92)$$

Um raciocínio exactamente análogo ao acabado de realizar permitiria concluir que $\delta(B)$ é um factor de $\varphi(B)$. Conjugando as duas conclusões, infere-se que $\delta(B)$ e $\psi(B)$ são idênticos, pelo que (7.1.88) pode ser expresso da forma:

$$\delta(B)\eta_{1t} = w(B)\eta_{2t} + \psi(B)\xi_t \quad (7.1.93)$$

Por outro lado, resulta ainda do teorema 7.1.1. que a ordem de $\varphi(B)$ em (7.1.93) deve ser igual à maior das ordens de $\delta(B)$ e de $w(B)$, se estas forem diferentes, e inferior ou igual à ordem de ambos, se estas forem da mesma ordem.

Estas conclusões permitem a fácil identificação da estrutura de $\varphi(B)$ e de $\psi(B)$ a partir da prévia identificação de $\delta(B)$ e de $w(B)$, só sendo válidas no caso de se tratar de uma relação unidireccional.

Terminada a identificação do modelo (7.1.88), procede-se à sua estimação e à confirmação do diagnóstico do mesmo modo que foi proposto no caso da relação bidireccional.

(iv) A fase final consiste na combinação dos modelos univariados ajustados a X_{2t} e a X_{1t} em (7.1.86) e (7.1.87) respectivamente, com o modelo (7.1.88), do que resulta:

$$X_{1t} = h_1 \frac{\theta_1(B)w(B)\phi_2(B)}{\phi_1(B)\delta(B)\theta_2(B)} X_{2t} + \frac{\theta_1(B)\psi(B)}{\phi_1(B)\varphi(B)} \xi_t \quad (7.1.94)$$

onde $h_1 = [V(\eta_{1t})/V(\eta_{2t})]^{1/2}$.

Após a estimação e a confirmação do diagnóstico de (7.1.94), o processo de modelização no caso de uma relação unidireccional está completo. Tal como para o modelo (7.1.88), e analogamente ao que sucede para uma relação bidireccional, é necessário agora dedicar especial atenção à possibilidade de eliminação de factores comuns ou de parâmetros não significativos pois, a menos que os modelos univariados incluam um número muito reduzido de parâmetros, existe um forte risco de (7.1.94) resultar numa estrutura sobreparametrizada, com grande abundância de parâmetros e extremamente pesada.

Por outro lado, este método, tal como o proposto por Box e Jenkins, está também sujeito a conduzir a maus resultados na identificação do modelo (7.1.86) se a variabilidade de X_{1t} for predominantemente determinada pela da componente residual, o que é tanto mais susceptível de ocorrer quanto mais semelhantes forem os modelos univariados ajustados a X_{1t} e a X_{2t} . Assim, é possível que seja esta a justificação para o facto de Pierce (1977) ter concluído pela fraqueza das relações entre diversos pares de variáveis macroeconómicas e financeiras que estariam fortemente relacionadas segundo a teoria económica, conforme já foi referido. Com efeito, foi este o método seguido por Pierce (1977) para averiguar a existência dessas relações e, tal como indicam Downing e Pack (1982), este fenómeno de fraqueza da evidência sobre a existência de relações entre séries cronológicas ocorre frequentemente no caso de dados relativos a variáveis económicas. Trata-se de um fenómeno idêntico ao que foi analisado no caso da relação bidireccional.

Deste modo, surge um dilema: por um lado, a estimação da função de correlação cruzada entre duas séries cronológicas na sua forma original pode conduzir a conclusões erradas devido à existência de autocorrelação do estimador dessa função e devido ao facto de a variância deste depender da estrutura da autocorrelação de ambas as séries; por outro lado, a prévia conversão em ruído branco de uma ou de ambas as séries, tendo como objectivo eliminar a autocorrelação nelas existente, pode levar a que permaneça apenas uma correlação cruzada de pequena intensidade entre as séries transformadas por essa operação de filtragem, o que também não permite a identificação correcta da estrutura do modelo que as relaciona.

2. Extensões do método de Box e Jenkins

A estratégia de modelização de uma série cronológica univariada proposta por Box e Jenkins (1976) foi generalizada ao caso multivariado por Tiao e Box (1979,1981), por Jenkins e Alavi (1981) e por Tjøstheim e Paulsen (1982) paralelamente e visa conseguir essa modelização através da consideração de modelos ARMA(p,q), de expressão genérica:

$$\phi(B)X_t = \theta(B)\varepsilon_t \quad (7.2.1)$$

onde X_t representa a série cronológica, que se admite ser estacionária e de valor esperado igual ao vector nulo, $\{\varepsilon_t\}$ representa um ruído branco e $\phi(B)$ e $\theta(B)$ representam respectivamente o operador auto-regressivo e o de médias móveis, conforme já foi definido e amplamente analisado. A equação (7.2.1) significa que X_t é gerada por um processo ARMA(p,q).

No entanto, a perspectiva da abordagem dos autores referidos não é a mesma: enquanto Tiao e Box (1979, 1981) e Jenkins e Alavi (1981) utilizam funções matriciais para a identificação de uma série cronológica multivariada, Tjøstheim e Paulsen (1982) recorrem a funções escalares. De qualquer modo, a identificação da estrutura de um modelo baseia-se nos mesmos padrões de comportamento das referidas funções, constituindo generalizações do que foi estabelecido por Box e Jenkins (1976) para o caso univariado.

Assim, no que respeita à identificação de uma série cronológica, também designada por Tiao e Box (1979, 1981) de tentativa de especificação, é útil ter em atenção os seguintes aspectos:

a) Frequentemente, modelos com baixos valores das ordens dos seus operadores constituem aproximações adequadas à série cronológica em análise, conseguindo a apropriada modelização desta última.

- b) O conhecimento do sistema, no caso de existir, pode permitir simplificações "a priori" no modelo, embora seja necessária a cuidadosa verificação da adequabilidade dessas simplificações.
- c) Após o ajustamento de um modelo inicial, é quase sempre possível conseguir simplificações consideráveis.

Neste contexto, vai-se analisar em primeiro lugar o método de Tiao e Box (1979, 1981) e de Jenkins e Alavi (1981), uma vez que constituem generalizações mais directas do método de Box e Jenkins (1976), descrevendo-se em seguida um método proposto por Jenkins e Alavi (1981) que recorre a modelos univariados e, por fim, passa-se à abordagem de Tjøstheim e Paulsen (1982), que utilizam as funções matriciais dos outros autores, embora com a perspectiva de extrair delas funções escalares que permitam a identificação de uma série cronológica multivariada.

2.1. Utilização de funções matriciais

Na perspectiva de Tiao e Box (1979, 1981) e de Jenkins e Alavi (1981), a identificação de uma série cronológica passa pela utilização de estatísticas que possam ser facilmente obtidas a partir das observações e que permitam a selecção de uma sub-classe de modelos que se revele adequada para posterior tratamento. Deste modo, os instrumentos utilizados são:

- (i) Matrizes de correlação estimadas;
- (ii) Matrizes de correlação parcial estimadas a partir do ajustamento de modelos auto-regressivos de ordem sucessivamente maior;
- (iii) Matrizes de correlação residuais estimadas, resultantes do ajustamento de cada modelo auto-regressivo.

Uma vez que a determinação das ordens de um modelo ARMA(p,q) apresenta diversos problemas quando se utiliza estes instrumentos, Jenkins e Alavi (1981) propõem ainda a utilização de matrizes de correlação parcial q - condicionadas estimadas e Tiao e Tsay (1983) a utilização de matrizes de correlação alargada estimadas, sendo ambas definidas mais à frente, tendo em vista a resolução desses problemas. Desta forma, dispõem-se de dois instrumentos adicionais que permitem refinar a análise e identificar mais correctamente uma série cronológica:

- (iv) Matrizes de correlação parcial q - condicionadas estimadas;
- (v) Matrizes de correlação alargada estimadas.

Estas funções matriciais não foram ainda apresentadas devido ao facto de constituírem instrumentos especificamente destinados à resolução deste problema, não tendo sido anteriormente propostos e fazendo por isso parte integrante dos métodos que irão presentemente ser analisados, o que explica o diferimento até agora da sua apresentação.

A forma como estes instrumentos são utilizados visando a identificação de uma série cronológica vai então ser agora analisada:

- (i) Matrizes de correlação estimadas

Uma vez que os elementos das matrizes de correlação de um processo MA(q) se anulam a partir do "lag" q, contrariamente ao que sucede para um processo AR(p), em que esses elementos decrescem lentamente para zero, o padrão de comportamento apresentado por estas matrizes permite distinguir entre os dois tipos de processos e, para o primeiro, determinar a sua ordem. No entanto, uma vez que o verdadeiro processo subjacente à série cronológica é desconhecido, o mesmo acontece com as suas matrizes de correlação, sendo por isso necessário estimá-las. Assim, é de esperar que o padrão de comportamento das grandezas desconhecidas venha reflectido no dos seus estimadores. Com efeito,

prova-se (Tiao e Tsay, 1983) que, para um modelo MA(q), se tem $\text{plim}R(k)=0$ para $k>q$, designando plim o limite em probabilidade e $R(k)$ a matriz de correlação estimada para o "lag" k, verificando-se deste modo que esta matriz apresenta um padrão de comportamento muito semelhante ao da verdadeira matriz de correlação de um processo MA(q), ou seja $\rho(k)=0$ para $k>q$.

Deste modo, o exame dos valores obtidos para as funções $r_{ij}(k)$, $i,j=1,\dots,m$, e a análise da sua representação gráfica apenas podem ter utilidade na sua modelização de uma ou de duas séries univariadas, tornando-se progressivamente mais complicados e incómodos à medida que m aumenta. Com efeito, será necessário analisar simultaneamente 10 gráficos se $m=4$ e 15 gráficos no caso de $m=5$, o que permite concluir imediatamente que essa análise só é exequível para valores muito pequenos de m.

Assim, Tiao e Box (1979, 1981) propõem listar a sequência das matrizes de correlação estimadas $R(k)$ em função de k, para $k=1,2,\dots$. A ausência de elementos destas matrizes estatisticamente significativos a partir de um certo valor de k permite detectar a presença de um processo de médias móveis e indicar directamente a sua ordem. Se, pelo contrário, os elementos destas matrizes se mantiverem persistentemente significativos, decrescendo lentamente em valor absoluto à medida que k aumenta, diagnostica-se um modelo auto-regressivo para a série cronológica. Mas, uma vez que o exame dos valores dos elementos de $R(k)$ não é também muito propício a extrair facilmente conclusões, estes autores propõem ainda que, ao listar as matrizes $R(k)$ em função de k, se substitua:

- 1) Os elementos positivos cujo valor é estatisticamente significativo por um sinal + ;

- 2) Os elementos negativos cujo valor é estatisticamente significativo por um sinal - ;
- 3) Os elementos cujo valor é estatisticamente não significativo por um ponto.

Neste contexto, tal como já foi referido, $r_{ij}(k)$ será considerado estatisticamente significativo se o seu valor absoluto for superior a $2N^{-1/2}$, o que significa que os autores referidos propõem a realização deste teste a nível de significância de aproximadamente 5%, tal como é vulgar. Deste modo, $r_{ij}(k)$ será considerado positivo se o seu valor ultrapassar $2N^{-1/2}$, e negativo se for inferior a $-2N^{-1/2}$. Tiao e Box (1979, 1981) sugerem também uma forma alternativa dos mesmos resultados, que consiste na colocação em linha e em coluna das séries X_{1t}, \dots, X_{mt} que compõem o vector X_t , de modo a formar um quadro de dupla entrada. Neste quadro, o cruzamento de duas séries X_{it} e X_{jt} ($i, j=1, \dots, m$) constitui um sector onde se lista os pontos, os sinais + e os sinais - relativos a cada $r_{ij}(k)$ para todos os valores de k considerados, consoante os valores de $r_{ij}(k)$ sejam respectivamente não significativos, significativos e positivos ou significativos e negativos. Conseguem-se desta forma resumir de um modo alternativo a informação referente a cada elemento da matriz $R(k)$, apresentando a sequência dos símbolos respeitantes aos valores de cada $r_{ij}(k)$ para todos os valores de k e para $i, j=1, \dots, m$. Pretende-se assim complementar a informação anterior em que se apresenta a matriz $R(k)$ para cada k , com $k=1, 2, \dots$.

Em suma, o padrão dos símbolos revelado pela sequência das matrizes $R(k)$, representados sob qualquer das duas formas alternativas, permite distinguir facilmente entre um modelo auto-regressivo e um modelo de médias móveis e, para este último, seleccionar a sua ordem.

(ii) Matrizes de correlação parcial estimadas

Tal como já foi referido, os elementos da matriz de correlação parcial de um processo AR(p) anulam-se a partir da ordem do processo, ou seja, a partir do "lag" p, não se verificando tal padrão de comportamento para um processo de médias móveis. Logo, conclui-se imediatamente que o padrão apresentado por estas matrizes permite distinguir entre processos de médias móveis e processos auto-regressivos e determinar a ordem destes últimos.

No entanto, uma vez que os elementos destas matrizes são também desconhecidos, torna-se necessário estimá-los, o que é feito através da estimação de modelos auto-regressivos de ordem k, com $k=1,2,\dots$, conforme já foi analisado. Deste modo, uma vez que, para um modelo AR(p), a matriz de correlação parcial se anula para $k > p$, é de esperar que os elementos de $\hat{P}(k)$ assumam pequenos valores (em valor absoluto) a partir do "lag" p, ou seja, é de esperar que a matriz de correlação parcial estimada apresente um comportamento semelhante ao da sua congénere teórica.

Deste modo, para extrair conclusões é útil normalizar os elementos de $\hat{P}(k)$, dividindo-se estes pelos respectivos desvios-padrão estimados. O padrão de comportamento das matrizes de correlação parcial estimadas pode então ser resumido atribuindo o sinal + aos coeficientes normalizados cujo valor ultrapasse 2, o sinal - àqueles cujo valor é inferior a -2 e um ponto aos que se situarem entre -2 e 2. A informação assim resultante deve também ser representada sob as duas formas alternativas especificadas para o caso das matrizes de correlação estimadas.

Por outro lado, é também possível utilizar a estatística $\zeta(k)$, definida em (3.4.7), para a determinação da ordem de um modelo auto-regressivo, conforme já foi definido atrás. Esta estatística serve para

testar as hipóteses nulas $\phi_k=0$ contra alternativas $\phi_k \neq 0$, a partir da estimação de um modelo AR(k), $k=1,2,\dots$. A ordem do modelo será o valor máximo de k para o qual a hipótese nula é rejeitada.

Finalmente, Tiao e Box (1979, 1981) propõem ainda recorrer à análise dos elementos da diagonal principal da matriz de covariância residual estimada Σ_k resultante da estimação dos sucessivos modelos AR(k), pois aqueles podem indicar como varia a qualidade do ajustamento à medida que se vai aumentando a ordem do modelo, uma vez que quanto menor o valor desses elementos, melhor é o ajustamento.

Em conclusão, o padrão revelado pela matriz de correlação parcial estimada, a estatística $\zeta(k)$ e a análise dos elementos da diagonal principal da matriz de covariância residual estimada são extremamente úteis para distinguir entre um modelo de médias móveis e um auto-regressivo e para seleccionar a ordem adequada para este último.

(iii) Matrizes de correlação residuais estimadas resultantes do ajustamento de modelos auto-regressivos.

Após o ajustamento de cada modelo AR(l), $l=1,\dots,p$, as matrizes de correlação residuais podem ser imediatamente estimadas, obtendo-se $\hat{R}_\varepsilon(k)$. Assim, uma vez que, se o modelo estiver correctamente identificado, os resíduos de estimação $\hat{\varepsilon}_t$ se devem comportar aproximadamente como um ruído branco, os elementos de $\hat{R}_\varepsilon(k)$, designados genericamente por $\hat{r}_{\varepsilon_{ij}}(k)$, $i,j=1,\dots,m$, devem ser estatisticamente não significativos para valores de k diferentes de zero. Deste modo, atribuindo o sinal - aos valores de $\hat{r}_{\varepsilon_{ij}}(k)$ inferiores a $-2N^{-1/2}$, o sinal + àqueles que forem superiores a $2N^{-1/2}$ e um ponto aos que estiverem entre $-2N^{-1/2}$ e $2N^{-1/2}$, ou seja, aos que forem não significativos, é possível representar a informação assim resultante sob as duas formas alternativas referidas para os elementos das matrizes de correlação e de correlação parcial estimadas.

O padrão de comportamento assim revelado por estas matrizes permite verificar se um modelo auto-regressivo é adequado ou se, pelo contrário, será preferível considerar um modelo de médias móveis ou até um modelo ARMA. No caso de persistirem elementos de $\hat{R}_\varepsilon(k)$ significativos ($k=1,2,\dots$) após a estimação de um dado modelo auto-regressivo, é necessário proceder ao ajustamento de modelos AR de ordem mais elevada. Se isso não for ainda adequado, ou seja, se continuarem a existir elementos significativos será necessário considerar outros tipos de modelos.

Assim, se, por exemplo, após o ajustamento de um modelo auto-regressivo de ordem elevada, houver elementos de $\hat{R}_\varepsilon(k)$ significativos para $k=1,2$, será útil abandonar os modelos auto-regressivos e considerar antes um modelo MA(2) como ponto de partida para uma nova tentativa de identificação.

A utilidade da análise destas matrizes que, tal como foi oportunamente referido, também deve ser feita durante a confirmação do diagnóstico, advém do facto de elas habitualmente captarem as falhas de adequação do modelo e permitirem sugerir características que não tenham ainda sido levadas em conta.

No que se refere aos modelos mistos auto-regressivos e de médias móveis em geral, quer os elementos das matrizes de correlação, quer os das matrizes de correlação parcial decrescem lentamente para zero, tornando bastante difícil a determinação das ordens dos operadores desta sub-classe de modelos através da utilização destes instrumentos. De qualquer forma, o padrão de comportamento revelado pela matriz de correlação residual estimada após o ajustamento de um modelo auto-regressivo pode indicar os valores dessas ordens. Suponha-se então, a título de exemplo, que, após o ajustamento de um modelo AR(p),

existem alguns elementos significativos na matriz $\hat{R}_\epsilon(1)$, sendo não significativos os elementos destas matrizes para valores superiores de k . Tal padrão e comportamento apontaria no sentido de se considerar um modelo ARMA(p,1).

No entanto, tal como é apontado por Tiao e Box (1979,1981), este procedimento conduzirá geralmente a uma incorrecta identificação do modelo, devido ao enviesamento dos estimadores das matrizes que constituem o operador auto-regressivo quando $q > 0$.

Assim, para obviar a este problema, Jenkins e Alavi (1981) propõem uma solução alternativa, introduzindo o conceito de matriz de correlação parcial q -condicionada, cujo estimador se converte num instrumento adicional para a identificação de modelos ARMA:

(iv) Matrizes de correlação parcial q -condicionadas estimadas.

Esta proposta parte da constatação de que a relação válida para a matriz de covariância de um processo AR(p) é também válida para um processo ARMA(p,q) para $k > q$, tal como já foi referido, assumindo a forma:

$$\Gamma(k) = \phi_1 \Gamma(k-1) + \dots + \phi_p \Gamma(k-p) \quad k = q+1, q+2, \dots \quad (7.2.2)$$

Assim, Jenkins e Alavi (1981) propõem o seguinte procedimento:

a) Uma vez que o resultado (7.2.2) é válido para modelos ARMA(p,q), fixa-se $q=1$ condicionada e obtém-se ϕ_k no sistema de equações:

$$\Gamma(h) = \phi_1 \Gamma(h-1) + \dots + \phi_k \Gamma(h-k) \quad (7.2.3)$$

para $h=2, \dots, k+1$. A matriz ϕ_k assim obtida é a matriz de correlação parcial $q=1$ condicionada e designa-se por $P_1(k)$, com $k=1, 2, \dots$.

b) Se as matrizes de correlação parcial $q=1$ condicionadas $P_1(k)$ não sofrerem nenhum corte abrupto, admite-se que a ordem do operador de médias móveis é $q=2$ e resolve-se as equações (7.2.3) para $h=3, \dots, k+2$. A

matriz de ϕ_k obtida deste modo é a matriz de correlação parcial $q=2$ condicionada e designa-se por $P_2(k)$, com $k=1,2,\dots$.

Deste modo, propõe-se um modelo ARMA(p,q) se a matriz de correlação parcial q -condicionada $P_q(k)$ se anular $k > p$. Note-se que, segundo esta definição, $P(k)$ é igual ao que seria agora designado por $P_0(k)$.

De uma forma geral, as matrizes de correlação parcial q -condicionadas $P_q(k)$ podem ser estimadas através da resolução do sistema de equações:

$$\begin{bmatrix} \hat{\phi}_1^T \\ \hat{\phi}_2^T \\ \vdots \\ \hat{\phi}_k^T \end{bmatrix} = \begin{bmatrix} C(-q) & C(1-q) & \dots & C(k-q-1) \\ C(-q-1) & C(-q) & \dots & C(k-q-2) \\ \vdots & \vdots & \vdots & \vdots \\ C(-q+1-k) & C(-q+2-k) & \dots & C(-q) \end{bmatrix}^{-1} \begin{bmatrix} C(-q-1) \\ C(-q-2) \\ \vdots \\ C(-q-k) \end{bmatrix} \quad (7.2.4)$$

para $k=1,2,\dots$, representando $\hat{\phi}_k$, obtido a partir das equações (7.2.4), o estimador da matriz de correlação parcial q -condicionada que passará a ser designado por $\hat{P}_q(k)$.

Jenkins e Alavi (1981) propõem ainda um procedimento paralelo, destinado a fornecer indicações iniciais de carácter geral que podem ser muito úteis no processo de modelização acabado de analisar. Tal procedimento consiste, em vez de examinar as matrizes $R(k)$, $\hat{P}(k)$ e $\hat{P}_q(k)$, em analisar os seus determinantes, para $k=1,2,\dots$, que deverão observar padrões de comportamento muito semelhantes ou mesmo idênticos aos das matrizes respectivas. É possível assim conseguir linhas de orientação iniciais que podem auxiliar na análise dessas matrizes em si. Este procedimento é desenvolvido por Tjøstheim e Paulsen (1982) e será analisado mais à frente.

Por outro lado, conforme já foi referido, Tiao e Tsay (1983) propõem ainda um outro método com vista á resolução do problema da determinação das ordens de um modelo ARMA(p,q):

(v) Matrizes de correlação alargada estimadas

Se, para um dado modelo ARMA(p,q), for possível determinar estimadores consistentes $\hat{\phi}_1, \dots, \hat{\phi}_p$ das matrizes que compõem o operador auto-regressivo e sendo W_t definido da forma:

$$W_t = (I_m - \hat{\phi}_1 B - \dots - \hat{\phi}_p B^p) X_t \quad (7.2.5)$$

verifica-se que $\{W_t\}$ é aproximadamente o modelo MA(q) seguinte (Tiao e Tsay, 1983):

$$W_t \simeq (I_m + \theta_1 B + \dots + \theta_q B^q) \varepsilon_t \quad (7.2.6)$$

Daqui resulta que a matriz de correlação de $\{W_t\}$ é nula para $k > q$. O ponto de partida deste procedimento consiste assim na obtenção de estimadores consistentes para as matrizes $\hat{\phi}_1, \dots, \hat{\phi}_p$ através de uma sequência de regressões iteradas, o que conduz ao conceito de matriz de correlação alargada estimada, que é determinada através de um método recursivo:

a) Regressão iterada

A primeira fase deste procedimento consiste em estabelecer a regressão AR(p):

$$X_t = \sum_{i=1}^p \phi_{i(p)}^{(0)} X_{t-i} + e_{p,t}^{(0)} \quad t=p+1, \dots, N \quad (7.2.7)$$

onde (0) significa que se trata de uma regressão vulgar e o índice (p) representa a sua ordem, sendo $e_{p,t}^{(0)}$ o vector residual correspondente. A partir de (7.2.7), facilmente se determina os estimadores de mínimos

quadrados $\hat{\phi}_{l(p)}^{(0)}$, sendo estes estimadores consistentes para ϕ_l ($l=1, \dots, p$) se X_t for gerado por um processo AR(p), ou seja, se $q=0$ (Tiao e Tsay, 1983). Quando $q>0$, os estimadores $\hat{\phi}_{l(p)}^{(0)}$ geralmente não são consistentes e, mesmo para elevados valores de N , os resíduos de estimação não são um ruído branco, assumindo a forma:

$$e_{p,t}^{(0)} = X_t - \sum_{l=1}^p \hat{\phi}_{l(p)}^{(0)} X_{t-l} \quad (7.2.8)$$

Deste modo, os vectores $\hat{e}_{p,t-k}^{(0)}$ contêm alguma informação sobre X_t , o que leva a definir a primeira regressão AR(p) iterada da forma:

$$X_t = \sum_{l=1}^p \phi_{l,p}^{(1)} X_{t-l} + \beta_{l(p)}^{(1)} \hat{e}_{p,t-1}^{(0)} + e_{p,t}^{(1)} \quad (7.2.9)$$

onde (1) designa a primeira regressão iterada e $e_{p,t}^{(1)}$ o vector residual correspondente. Através da sua transposição, a equação (7.2.9) pode ser expressa sob a forma de um modelo linear multivariado:

$$X_t^T = \sum_{l=1}^p X_{t-l}^T \phi_{l(p)}^{(1)T} \hat{e}_{p,t-1}^{(0)T} \beta_{l(p)}^{(1)T} + e_{p,t}^{(1)} \quad (7.2.10)$$

Os estimadores de mínimos quadrados $\hat{\phi}_{l(p)}^{(1)}$ e $\hat{\beta}_{l(p)}^{(1)}$ podem então ser facilmente determinados, sendo possível provar que $\hat{\phi}_{l(p)}^{(1)}$ são estimadores consistentes para ϕ_l ($l=1, \dots, p$) quando $q \leq 1$ e X_t verifica algumas condições de regularidade (Tiao e Tsay, 1983). De modo semelhante ao que acontecia com a regressão anterior, quando os estimadores $\hat{\phi}_{l(p)}^{(1)}$ são inconsistentes, considera-se a segunda regressão iterada, que assume a expressão:

$$X_t = \sum_{l=1}^p \phi_{l(p)}^{(2)} X_{t-l} + \beta_{l(p)}^{(2)} \hat{e}_{p,t-1}^{(1)} + \beta_{2(p)}^{(2)} \hat{e}_{p,t-2}^{(0)} + e_{p,t}^{(2)} \quad t=p+3, \dots, N \quad (7.2.11)$$

onde $\hat{e}_{p,t-2}^{(0)}$ e $\hat{e}_{p,t-1}^{(1)}$ são respectivamente os resíduos de estimação de (7.2.7) e de (7.2.9). É também possível mostrar que os estimadores de mínimos quadrados $\hat{\phi}_{l(r)}^{(2)}$, determinados a partir de (7.2.11), são estimadores consistentes para ϕ_l ($l=1, \dots, p$) quando $q \leq 2$ e assim sucessivamente (Tiao e Tsay, 1983), partindo destes resultados, consideram um conjunto genérico de regressões iteradas. Especificamente, a k -ésima regressão AR(r) iterada de um processo ARMA $\{X_t\}$ é definida da forma:

$$X_t = \sum_{l=1}^r \phi_{l(r)}^{(k)} X_{t-l} + \sum_{u=1}^k \beta_{u(r)}^{(k)} \hat{e}_{r,t-u}^{(k-u)} + e_{r,t}^{(k)} \quad t=r+k+1, \dots, N \quad (7.2.12)$$

onde:

$$\hat{e}_{r,t}^{(v)} = X_t - \sum_{l=1}^r \hat{\phi}_{l(r)}^{(v)} X_{t-l} - \hat{\beta}_{u(r)}^{(v)} \hat{e}_{r,t-u}^{(k-u)} \quad (7.2.13)$$

é o vector dos resíduos de estimação da v -ésima regressão AR(r) iterada e $\hat{\phi}_{l(r)}^{(v)}$ e $\hat{\beta}_{u(r)}^{(v)}$ são os estimadores de mínimos quadrados correspondentes. É também possível que, se X_t for gerado por um processo ARMA(p, q) e nenhuma das combinações das componentes de $Y_{p,t} = (X_t^T, \dots, X_{t-p+1}^T)^T$ for gerada por um processo de médias móveis de ordem inferior a q , se verifica a convergência em probabilidade (Tiao e Tsay, 1983):

$$\text{plim} \hat{\phi}_{l(r)}^{(k)} = \phi_l \quad l=1, \dots, r \quad (7.2.14)$$

para 1º) $r \geq p$ e $k=q$, ou 2º) $r=p$ e $k>q$, sendo $\phi_l = 0$ para $l > p$.

b) Método Recursivo

Os estimadores $\hat{\phi}_{l(r)}^{(k)}$ obtidos a partir das regressões iteradas, podem ser determinados recursivamente através do ajustamento de modelos

auto--regressivos. Com efeito, considere-se a primeira regressão AR(p) iterada em (7.2.9), que pode ser expressa da forma alternativa:

$$\mathbf{X}_t = \sum_{l=1}^p \phi_{l,p}^{(1)} \mathbf{X}_{t-l} + \beta_{1(p)}^{(1)} [\mathbf{X}_{t-1} - \sum_{l=1}^p \hat{\phi}_{l(p)}^{(0)} \mathbf{X}_{t-1-l}] + \mathbf{e}_{p,t}^{(1)} \quad (7.2.15)$$

ou seja:

$$\mathbf{X}_t = \sum_{l=1}^{p+1} \Lambda_{l(p+1)}^{(0)} \mathbf{X}_{t-l} + \mathbf{e}_{p,t}^{(1)} \quad (7.2.16)$$

onde:

$$\Lambda_{l(p+1)}^{(0)} = \phi_{l,p}^{(1)} - \beta_{l(p)}^{(1)} \hat{\phi}_{l-1(p)}^{(0)} \quad l=1, \dots, p \quad (7.2.17)$$

e

$$\Lambda_{l(p+1)}^{(0)} = -\beta_{l(p)}^{(1)} \hat{\phi}_{l-1(p)}^{(0)} \quad (7.2.18)$$

com $\hat{\phi}_{0(p)}^{(0)} = -\mathbf{I}_m$. Note-se que, na realidade, (7.2.16) é uma regressão AR(p+1) vulgar, assumindo esta a forma:

$$\mathbf{X}_t = \sum_{l=1}^{p+1} \phi_{l(p+1)}^{(0)} \mathbf{X}_{t-l} + \mathbf{e}_{p+1,t}^{(0)} \quad (7.2.19)$$

Logo, conclui-se:

$$\hat{\phi}_{l(p+1)}^{(0)} = \hat{\Lambda}_{l(p+1)}^{(0)} \quad l=1, \dots, p+1 \quad (7.2.20)$$

Utilizando as expressões (7.2.17), (7.2.18) e (7.2.20), obtém-se o resultado:

$$\hat{\phi}_{l(p)}^{(1)} = \hat{\phi}_{l(p+1)}^{(0)} - \hat{\phi}_{p+1(p+1)}^{(0)} \left[\hat{\phi}_{p(p)}^{(0)} \right]^{-1} \hat{\phi}_{l-1(p)}^{(0)} \quad l=1, \dots, p \quad (7.2.21)$$

Esta equação significa que os estimadores $\hat{\phi}_{l(p)}^{(1)}$ podem ser obtidos a partir de uma regressão AR(p) e de uma regressão AR(p+1). Deste modo, é possível chegar à expressão geral (Tiao e Tsay, 1983):

$$\hat{\phi}_{l(r)}^{(k)} = \hat{\phi}_{l(r+1)}^{(k-1)} - \hat{\phi}_{r+1(r+1)}^{(k-1)} \left[\hat{\phi}_{z(r)}^{(k-1)} \right]^{-1} \hat{\phi}_{l-1(r)}^{(k-1)} \quad l=1, \dots, r \quad (7.2.22)$$

com $\hat{\phi}_{0(r)}^{(k-1)} = -\mathbf{I}_m$. Esta equação indica que os estimadores da k-ésima regressão iterada podem ser determinados recursivamente a partir dos estimadores de mínimos quadrados das regressões vulgares AR(r), ..., AR(r+k).

c) Propriedades assintóticas da matriz de correlação alargada estimada

Tendo sido estabelecida a consistência de $\hat{\phi}_{l(r)}^{(k)}$ em (7.2.14), é possível agora definir a matriz de correlação alargada estimada, que será designada por $R_{(r)}(k)$. Para o efeito suponha-se primeiro $r=1$, o que conduz a:

$$W_{1,t}^{(k)} = X_t - \hat{\phi}_{1(1)}^{(k)} X_{t-1} \quad (7.2.23)$$

A partir de (7.2.14) verifica-se que, se X_t for gerada por um processo ARMA(1,q), $\{W_{1,t}^{(k)}\}$ será aproximadamente um processo MA(q) quando $k \geq q$ e para N elevado. Assim, é possível definir a matriz $R_{(1)}(k)$ como a matriz de correlação estimada de $W_{1,t}$ para o "lag" k, com $k=1,2,\dots$ (Tiao e Box, 1983), o que conduz ao resultado:

$$\text{plim} R_{(1)}(k) = \begin{cases} Q & k = q \\ 0 & k > q \end{cases} \quad (7.2.24)$$

onde Q é a matriz de correlação para o "lag" q do processo MA(q):

$$Z_t = (\mathbf{I}_m + \theta_1 B + \dots + \theta_q B^q) \varepsilon_t \quad (7.2.25)$$

A expressão (7.2.24) permite concluir imediatamente que, para modelos ARMA(1,q) as matrizes $R_{(1)}(k)$ apresentam exactamente o mesmo padrão de comportamento que as matrizes de correlação estimadas habituais dos modelos MA(q), ou seja, anulam-se para $k > q$. Deste modo, $R_{(1)}(k)$ é a primeira matriz de correlação alargada estimada para o "lag" k. De um modo geral, define-se a r-ésima matriz de correlação alargada estimada $R_{(r)}(k)$ como a matriz de correlação estimada de $W_{r,t}^{(k)}$ para o "lag" k, sendo $W_{r,t}^{(k)}$ definido da forma (Tiao e Tsay, 1983):

$$W_{r,t}^{(k)} = X_t - \sum_{l=1}^r \hat{\phi}_{l(r)}^{(k)} X_{t-l} \quad (7.2.26)$$

A partir de (7.2.14) verifica-se que, para um processo ARMA(p,q), se tem o resultado:

$$\text{plim} R_{(r)}(k) = \begin{cases} Q & \text{para } r=p \text{ e } k=q \\ 0 & \text{para } r=p \text{ e } k > q \end{cases} \quad (7.2.27)$$

Portanto, conclui-se que, para processos ARMA(p,q), as p-ésimas matrizes de correlação alargada estimadas $R_{(p)}(k)$ apresentam exactamente o mesmo padrão de comportamento que as matrizes de correlação estimadas de processos MA(q), o que as converte num instrumento de grande utilidade para a determinação de q quando p é conhecido.

Sem perda de generalidade, define-se $R_{(0)}(k) = R(k)$. De qualquer modo, é conveniente sublinhar que, excepto para $r=0$, as matrizes de correlação alargada estimadas não são as matrizes de correlação estimadas de qualquer série obtida como transformação de X_t .

No entanto, surge aqui uma dificuldade: o padrão de comportamento das matrizes $R_{(r)}(k)$, segundo o qual estas se anulam a

partir de $k=q$, admite que $r=p$. Contudo, surgem dificuldades se a ordem r do modelo $AR(r)$ ajustado for superior à verdadeira ordem p do operador auto-regressivo, que é desconhecida, e o número de iterações k for superior à verdadeira ordem do operador de médias móveis q , que é também desconhecida (Tiao e Tsay, 1983). O caso em que $X_t = \varepsilon_t$, ou seja, em que $p=q=0$, sendo muito simples, é também de grande utilidade para a exemplificação deste problema. Nestas circunstâncias, tem-se a convergência em probabilidade (Tiao e Tsay, 1983):

$$\text{plim} \hat{\phi}_{l(r)}^{(0)} = 0 \quad l=1, \dots, r \quad ; \quad r=1, 2, \dots$$

(7.2.28)

Considere-se agora o estimador $\hat{\phi}_{1(1)}^{(1)}$ da primeira regressão iterada, determinado em (7.2.21). A partir de (7.2.28), conclui-se que tanto $\hat{\phi}_{1(1)}^{(0)}$ como $\hat{\phi}_{2(2)}^{(0)}$, no segundo membro de (7.2.21), convergem em probabilidade para uma matriz nula, o que significa que, ao determinar $\hat{\phi}_{1(1)}^{(1)}$, se está numa situação semelhante à de dividir zero por zero. É também possível mostrar que, para as segundas regressões iteradas, se chega à convergência em probabilidade:

$$\text{plim} \hat{\phi}_{l(r)}^{(1)} = \begin{cases} E_{(r)}(l) & l=1, 2 \\ 0 & l=2, \dots, r \end{cases} \quad ; \quad r=1, 2, \dots \quad (7.2.29)$$

onde $E_{(r)}(l)$ é uma matriz estocástica (Tiao e Tsay, 1983). A partir de (7.2.29) e da relação recursiva (7.2.22), é possível mostrar que, para as segundas regressões iteradas, se chega à convergência em probabilidade:

$$\text{plim} \hat{\phi}_{l(r)}^{(2)} = \begin{cases} \mathbf{E}_{(r)}(l) & l=1,2 \\ \mathbf{0} & l=3,\dots,r \end{cases} ; \quad r=1,2,\dots \quad (7.2.30)$$

onde $\mathbf{E}_{(r)}(l)$, $l=1,2$, são matrizes estocásticas, e assim sucessivamente (Tiao e Tsay, 1983).

A partir da expressão (7.2.29), verifica-se que $\{\mathbf{W}_{1,t}^{(1)}\}$, definido em (7.2.23), é assintoticamente um processo MA(1), pelo que $\mathbf{R}_{(1)}(1)$ converge para uma matriz não nula. Do mesmo modo, verifica-se a partir de (7.2.30) que $\{\mathbf{W}_{1,t}^{(2)}\}$ também é aproximadamente um processo MA(1), o que implica que a matriz $\mathbf{R}_{(1)}(2)$, sendo a matriz de correlação estimada de $\{\mathbf{W}_{1,t}^{(2)}\}$ para $k=2$, convirja para uma matriz nula. Assim, conclui-se que, se a ordem da regressão AR iterada, ou seja, o valor de r , for superior a $p=0$, a ordem do operador de médias móveis do modelo de $\{\mathbf{W}_{1,t}^{(k)}\}$ sofre também um aumento. De um modo geral, é possível mostrar que, para um modelo ARMA(p,q), as matrizes $\mathbf{R}_{(r)}(k)$ possuem a propriedade assintótica:

$$\text{plim} \mathbf{R}_{(r)}(k) = \begin{cases} \mathbf{Q}_{(r)}(k) & 0 \leq k - q \leq r - p \\ \mathbf{0} & 0 \leq r - p \leq k - q \\ & r=1,2,\dots; k=1,2,\dots \end{cases} \quad (7.2.31)$$

onde $\mathbf{Q}_{(r)}(k)$ representa uma matriz cujos elementos não são todos nulos (Tiao e Tsay, 1983).

Neste contexto, é possível definir a estratégia conducente à determinação das ordens de um modelo ARMA(p,q) a partir da propriedade assintótica das matrizes $\mathbf{R}_{(r)}(k)$ num quadro de dupla entrada como o seguinte:

MA \ AR	0	1	2	3	...
0	$R_{(0)}(1)$	$R_{(0)}(2)$	$R_{(0)}(3)$	$R_{(0)}(4)$...
1	$R_{(1)}(1)$	$R_{(1)}(2)$	$R_{(1)}(3)$	$R_{(1)}(4)$...
2	$R_{(2)}(1)$	$R_{(2)}(2)$	$R_{(2)}(3)$	$R_{(2)}(4)$...
3	\vdots	\vdots	\vdots	\vdots	
\vdots					

(7.2.32)

Na primeira linha desta tabela estão as matrizes de correlação estimadas, na segunda estão as primeiras matrizes de correlação alargada estimadas e assim sucessivamente. As linhas e as colunas têm a numeração 0,1,2,... significando respectivamente a ordem do operador auto-regressivo e a do operador de médias móveis do modelo ARMA(p,q).

A título de exemplo, suponha-se que o verdadeiro modelo é um ARMA(1,1). As matrizes de correlação estimadas $R_{(0)}(k)$ verificam a propriedade assintótica:

$$\text{plim } R_{(0)}(k) = \rho(k) \quad k \geq 1 \quad (7.2.33)$$

onde $\rho(k)$ é a matriz de correlação de $\{X_t\}$, sendo uma matriz não nula para um modelo ARMA(1,1). Por outro lado, a partir de (7.2.31), chega-se ao seguinte resultado para as primeiras matrizes de correlação alargada estimadas:

$$\text{plim } R_{(1)}(k) = 0 \quad k \geq 2 \quad (7.2.34)$$

No que respeita às segundas matrizes de correlação alargada estimadas, obtém-se:

$$\text{plim } R_{(2)}(k) = 0 \quad k \geq 3 \quad (7.2.35)$$

e assim sucessivamente (Tiao e Tsay, 1983). Estes resultados são completamente descritos no quadro seguinte:

MA		0	1	2	3	4	5	...
AR								
0		Q	Q	Q	Q	Q	Q	...
1		Q	0	0	0	0	0	...
2		Q	Q	0	0	0	0	...
3		Q	Q	Q	0	0	0	...
4		Q	Q	Q	Q	0	0	...
⋮		⋮	⋮	⋮	⋮	⋮	⋮	

(7.2.36)

onde Q é uma matriz não nula. Verifica-se imediatamente que as matrizes nulas formam um triângulo limitado pelas rectas $r=1$ e $k-r=1$. As coordenadas da linha e da coluna do vértice desse triângulo correspondem à ordem do operador auto-regressivo e à do operador de médias móveis respectivamente, ou seja, $p=1$ e $q=1$, conduzindo correctamente à identificação de um modelo ARMA(1,1).

De um modo geral, a construção de um quadro da forma de (7.2.36) permite constatar a existência de um triângulo composto por matrizes assintoticamente nulas e limitado pelas rectas $r=a_1 \geq 0$ e $k-r=a_2 \geq 0$, especificando-se então $p=a_1$ e $q=a_2$ como as ordens do modelo ARMA que se pretende identificar. No entanto, para uma dimensão finita da série cronológica, os elementos das matrizes $R_{(r)}(k)$ não são exactamente nulos, sendo possível determinar as suas variâncias assintóticas através da fórmula de Bartlett. De qualquer modo, pode-se tomar $(N-r-k)^{-1}$ como estimador aproximado dos valores dessas variâncias, sob a hipótese de $\{W_{r,t}^{(k)}\}$ ser um processo puramente aleatório, embora tal subestime esses valores, devendo por isso ser considerado apenas como indicação (Tiao e Tsay, 1983). Tendo em

vista condensar a informação fornecida por estas matrizes, atribui-se aos seus elementos os sinais (+, ., -) conforme sejam significativos positivos, não significativos ou significativos negativos respectivamente, tal como já foi sugerido para as matrizes de correlação e de correlação parcial estimadas.

Finalmente, quando as ordens de um modelo ARMA(p,q) estão identificadas, é possível utilizar os estimadores resultantes da regressão AR iterada correspondentes como valores iniciais para as matrizes que compõem o operador auto-regressivo, tendo em vista a posterior estimação pelo método da máxima verosimilhança. Tal utilização é correcta, uma vez que, conforme foi estabelecido em (7.2.14), $\hat{\phi}_{i(r)}^{(k)}$ são estimadores consistentes para as matrizes $\phi_l (l=1, \dots, p)$, quando p está correctamente identificado (Tiao e Tsay, 1983).

Por outro lado, Tiao e Tsay (1983) indicam ainda que, em geral, as matrizes de correlação alargada estimadas não necessitam da prévia estacionarização de uma série cronológica não estacionária para conduzirem à correcta identificação de um modelo ARMA(p,q), o que é mais uma vantagem decorrente da sua utilização.

Por sua vez, Jenkins e Alavi (1981) propõem ainda um outro método que recorre a uma análise do tipo das acabadas de descrever, mas que utiliza simultâneamente os modelos univariados individuais, à semelhança dos métodos aplicáveis ao caso bivariado.

Assim, admita-se que é possível ajustar modelos ARMA invertíveis e estacionários univariados a cada série individual:

$$X_{it} = U_i'(B)\eta_{it} \quad i=1 \dots, m \quad (7.2.37)$$

Estes modelos podem ser representados conjuntamente através da expressão matricial:

$$X_t = U' \eta_t \quad (7.2.38)$$

onde U' é uma matriz diagonal cujo i -ésimo elemento da diagonal principal é $U'_i(B)$ e $\eta_t = (\eta_{1t}, \dots, \eta_{mt})^T$. Em seguida, utiliza-se as matrizes de correlação e de correlação parcial estimadas do vector η_t para identificar um modelo que lhe seja adequado e que deverá assumir a forma:

$$\eta_t = U \xi_t \quad (7.2.39)$$

onde $\xi_t = (\xi_{1t}, \dots, \xi_{mt})^T$. Finalmente, os modelos (7.2.38) e (7.2.39) são combinados, conduzindo ao modelo final:

$$X_t = U' U \xi_t \quad (7.2.40)$$

Este método, conjugando o recurso aos modelos univariados e a utilização da metodologia proposta por Tiao e Box (1979, 1981) e por Jenkins e Alavi (1981), permitirá modelizar séries com maior número de componentes do que os métodos apresentados para o caso bivariado. No entanto, tem também os inconvenientes que aí foram referidos e que, de uma forma abreviada, são:

- a) Existência do perigo de obtenção de estruturas sobreparametrizadas. Com efeito, o modelo final (7.2.40) deverá conter habitualmente um elevado número de parâmetros devido à forma como é obtido.
- b) Enfraquecimento das relações entre as variáveis residuais dos modelos univariados, o que poderá conduzir a grandes dificuldades na identificação da estrutura do modelo (7.2.39), distorcendo-a, devido à

possibilidade de incorrecção das informações fornecidas pelas matrizes de correlação e de correlação parcial estimadas.

2.2. Utilização de funções escalares

Baseando-se em funções escalares de algumas das funções matriciais utilizadas pelos métodos que foram analisados, Tjøstheim e Paulsen (1982) pretendem ultrapassar o inconveniente do recurso a matrizes e resumir a informação nelas contida de uma forma sucinta e mais fácil de tratar. Para o efeito, estabelecem alguns novos resultados relativos às funções matriciais já amplamente examinadas e definem uma nova função matricial, centrando a atenção no traço e nos valores próprios dessas matrizes, bem como na definição de algumas estatísticas. Assim, descrever-se-à esses resultados para os modelos auto-regressivos e, em seguida, para os modelos de médias móveis, passando-se finalmente à análise do procedimento proposto tendo em vista a identificação de uma série cronológica. Coloca-se à partida a hipótese de X_t seguir uma distribuição Normal e ser gerada pelo que se designa frequentemente de processo linear, que assume a forma (Hannan, 1970 pág. 209):

$$X_t = \sum_{l=-\infty}^{\infty} \psi_l \varepsilon_{t-l}, \quad \sum_{l=-\infty}^{\infty} \|\psi_l\| < \infty \quad (7.2.41)$$

onde os vectores ε_t são independentes e identicamente distribuídos com matriz de variâncias e covariâncias finita Σ e $\|M\|$ designa uma qualquer norma de M no espaço das matrizes ($m \times m$).

a) Modelos auto-regressivos

Conforme já foi referido, a matriz de correlação parcial $P(k)$ de um processo AR(p) anula-se para $k > p$. Antes da explanação do método

proposto, é necessário definir o vector de dimensão ($m^2 \times 1$) $\text{hvecP}(k)$, que assume a forma:

$$\text{hvecP}(k) = (P_{11}(k) P_{12}(k) \dots P_{1m}(k) P_{21}(k) \dots P_{2m}(k) \dots P_{m1}(k) \dots P_{mm}(k))^T \quad (7.2.42)$$

ou seja, $\text{hvecP}(k)$ é formado a partir da colocação numa única coluna das linhas de $P(k)$, cujo elemento genérico é designado por $P_{ij}(k)$, $i, j=1, \dots, m$. Assim, Tjøstheim e Paulsen (1982) indicam que $\text{hvec}\hat{P}(k)$ segue uma distribuição assintótica Normal com matriz de covariância assintótica:

$$N^{-1} \Sigma \otimes \Delta^{-1} \quad (7.2.43)$$

para $k > p$, sendo $\text{hvec}\hat{P}(k)$ obtido a partir da matriz $\hat{P}(k)$ de acordo com a definição (7.2.42); Δ é a matriz de covariância do vector η_t , sendo este o vector das variáveis residuais da seguinte representação de X_t :

$$X_t - A_1 X_{t-1} - \dots - A_p X_{t-p} = \eta_t \quad (7.2.44)$$

onde A_1, \dots, A_p são matrizes de dimensão ($m \times m$) e $\{\eta_t\}$ é um ruído branco, tendo-se $E(\eta_t \eta_t^T) = \Delta$. A expressão (7.2.43) indica que, dependendo da estrutura de Σ e de Δ , alguns dos elementos de $\text{hvec}\hat{P}(k)$ podem ter desvios-padrão assintóticos elevados para $k > p$, o que deve ser tomado em linha de conta se esses elementos forem directamente utilizados para a identificação. Deste modo, a dedução da normalidade assintótica de $\text{hvec}\hat{P}(k)$ tem como objectivo determinar a distribuição da estatística (Tjøstheim e Paulsen, 1982):

$$\zeta_p^1(k) = N(\text{hvec}\hat{P}(k))^T (\Sigma \otimes \Delta^{-1})^{-1} \text{hvec}\hat{P}(k) \sim \chi_{m^2}^2 \quad (7.2.45)$$

ou seja, a estatística $\zeta_p^1(k)$ segue uma distribuição assintótica χ_m^2 , para $k > p$. Na prática, Σ e Δ têm que ser substituídas pelos seus estimadores $\hat{\Sigma}_k$ e $\hat{\Delta}_k$ respectivamente, sendo o primeiro determinado a partir do ajustamento de um modelo auto-regressivo à série cronológica e o segundo de um ajustamento de um modelo do tipo de (7.2.44). Tjøstheim e Paulsen (1982) indicam que ambos os estimadores são consistentes quando $N \rightarrow \infty$, sendo possível simplificar a expressão da estatística $\zeta_p^1(k)$, o que conduz à expressão alternativa:

$$\zeta_p(k) = N(\text{vec}\hat{P}(k))^T (\hat{\Sigma}_k^{-1} \otimes \hat{\Delta}_k) \text{vec}\hat{P}(k) = N \text{tr} \left\{ \hat{\Sigma}_k^{-1} \hat{P}(k) \hat{\Delta}_k \hat{P}^T(k) \right\} \quad (7.2.46)$$

onde se utilizou $\text{tr}(M_1 M_2 M_3 M_2^T) = (\text{vec} M_2)^T (M_1^T \otimes M_3) (\text{vec} M_2)$. Esta estatística pode assim ser utilizada para a identificação da ordem de um modelo auto-regressivo, através de um procedimento idêntico ao que foi descrito para a utilização de $\zeta(k)$, definida em (3.4.7), e constituindo uma alternativa a esta última.

Uma outra função escalar que pode ser útil é o traço da matriz $\hat{P}(k)$, para o qual Tjøstheim e Paulsen (1982) indicam a variância assintótica que assume a expressão:

$$V[\text{tr}\hat{P}(k)] = N^{-1} \text{tr}(\Sigma \Delta^{-1}) \quad (7.2.47)$$

para $k > p$. Deste modo, é possível determinar os valores estatisticamente significativos de $\text{tr}[\hat{P}(k)]$.

b) Modelos de médias móveis

Para esta classe de modelos, a matriz de correlação $\rho(k)$ anula-se para $k > q$, sendo q a ordem do modelo MA(q). Assim, definindo $\text{vec}\rho(k)$ e

$\text{vecR}(k)$ de acordo com (7.2.42), Tjøstheim e Paulsen (1982) definem a covariância assintótica:

$$\text{cov}\{\text{vecR}(k), (\text{vecR}(r))^T\} = N^{-1} \sum_{u=-q}^q \rho(u) \otimes \rho(u+r-k) \quad (7.2.48)$$

para $k, r > q$. Uma vez que, conforme indicam Tjøstheim e Paulsen (1982), $\text{vecR}(k)$ segue uma distribuição assintótica Normal, é útil definir a estatística:

$$\zeta_p^1(k) = N(\text{vecR}(k))^T \left[\sum_{u=-q}^q \rho(u) \otimes \rho(u) \right]^{-1} \text{vecR}(k) \quad (7.2.49)$$

Esta estatística segue distribuição assintótica χ_m^2 (Tjøstheim e Paulsen, 1982). Uma vez que $\rho(u)$ e q são desconhecidos, os autores de $\zeta_p^1(k)$ propõem antes a utilização da estatística:

$$\zeta_p(k) = N(\text{vecR}(k))^T \left[\sum_{u=-(k-1)}^{k-1} \hat{R}(u) \otimes \hat{R}(u) \right]^{-1} \text{vecR}(k) \quad (7.2.50)$$

Esta estatística segue uma distribuição aproximada χ_m^2 , para $k > q$ e N elevado, uma vez que $\hat{R}(u)$ é um estimador consistente para $\rho(u)$, tendo-se os seguintes resultados, conforme já foi referido:

$$\text{plim} \hat{R}(u) = \begin{cases} \rho(u) & u \leq q \\ 0 & u > q \end{cases} \quad (7.2.51)$$

Uma vez que $\rho(-r) = \rho^T(r)$, Tjøstheim e Paulsen (1982) definem também:

$$\text{cov}\{\text{trR}(k), \text{trR}(r)\} = N^{-1} \sum_{u=-q}^q \text{tr}\{\rho(u) \dot{\rho}(k-r-u)\} \quad (7.2.52)$$

para $k, r > q$, o que permitirá também julgar da significância estatística dos elementos de $\text{trR}(k)$.

Por outro lado, Tjøstheim e Paulsen (1982) propõem ainda uma outra função matricial, designada por matriz de covariância normalizada, que assume a forma:

$$G(k) = \Gamma(k) \Gamma(0)^{-1} \quad (7.2.53)$$

onde $G(k)$ designa a matriz de covariância normalizada para o "lag" k , admitindo-se que $\Gamma(0)$ é não singular. Em termos genéricos, os autores desta proposta definem também a covariância assintótica:

$$\begin{aligned} \text{cov} \{ \text{hvec} \hat{G}(k), (\text{hvec} \hat{G}(r))^T \} = \\ = (\mathbf{I}_m \otimes \Gamma(0)^{-1}) \text{cov} \{ \text{hvec} \hat{F}(k), (\text{hvec} \hat{F}(r))^T \} (\mathbf{I}_m \otimes \Gamma(0)^{-1})^T \end{aligned} \quad (7.2.54)$$

sendo:

$$\begin{aligned} \hat{G}(k) = C(k)C(0)^{-1} \\ \text{hvec} \hat{F}(k) = \text{hvec} C(k) - (C(k) \otimes \mathbf{I}_m) \text{hvec} C(0) \end{aligned} \quad (7.2.55)$$

obtendo-se uma expressão para $\text{cov} \{ \text{hvec} \hat{F}(k), \text{hvec} \hat{F}(r) \}$ a partir do resultado (Tjøstheim e Paulsen, 1982):

$$\begin{aligned} \text{cov} \{ c_{ij}(k), c_{ln}(r) \} = N^{-1} \sum_{u=N+1}^{N-1} (1-|u|N^{-1}) \times \\ \times [c_{iu}(u) c_{jn}(u+r-k) + c_{ln}(u+r) c_{ji}(u-k)] \end{aligned} \quad (7.2.56)$$

onde $i, j, l, n = 1, \dots, m$. Estes autores indicam ainda que é possível mostrar que $\hat{G}(k)$ segue uma distribuição assintótica Normal.

No entanto, para um modelo MA(q), a expressão (7.2.54) vem consideravelmente simplificada:



$$\begin{aligned} \text{cov} \{ \text{hvec} \hat{G}(k), (\text{hvec} \hat{G}(r))^T \} &= \\ &= (\mathbf{I}_m \otimes \Gamma(0)^{-1}) \text{cov} \{ \text{hvec} C(k), (\text{hvec} C(r))^T \} (\mathbf{I}_m \otimes \Gamma(0)^{-1})^T \end{aligned} \quad (7.2.57)$$

para $k, r > q$. É possível exprimir (7.2.57) da forma alternativa:

$$\begin{aligned} \text{cov} \{ \text{hvec} \hat{G}(k), (\text{hvec} \hat{G}(r))^T \} &= \\ &= N^{-1} (\mathbf{I}_m \otimes \Gamma(0)^{-1}) \left[\sum_{u=-q}^q \mathbf{G}(u) \otimes \mathbf{G}(u+r-k) \right] (\Gamma(0) \otimes \mathbf{I}_m) \end{aligned} \quad (7.2.58)$$

onde se utilizou a expressão (7.2.56) válida para este caso e a regra $(\mathbf{M}_1 \otimes \mathbf{M}_2) (\mathbf{M}_3 \otimes \mathbf{M}_4) = \mathbf{M}_1 \mathbf{M}_3 \otimes \mathbf{M}_2 \mathbf{M}_4$.

Nestas condições é agora possível definir a seguinte estatística (Tjøstheim e Paulsen, 1982):

$$\zeta_G(k) = (\text{hvec} \hat{G}(k))^T [\text{côv} \{ \text{hvec} \hat{G}(k), (\text{hvec} \hat{G}(k))^T \}]^{-1} \text{hvec} \hat{G}(k) \quad (7.2.59)$$

e prova-se que $\zeta_G(k) = \zeta_p(k)$ (Tjøstheim e Paulsen, 1982). Com efeito, sendo $\mathbf{R}(k) = \mathbf{C}_0^{-1/2} \mathbf{C}(k) \mathbf{C}_0^{-1/2}$, com $\mathbf{C}_0 = \text{diag} \{ c_{11}(0), \dots, c_{mm}(0) \}$, chega-se ao resultado:

$$\hat{G}(k) = \mathbf{C}_0^{1/2} \mathbf{R}(k) \mathbf{C}_0^{1/2} \mathbf{C}(0)^{-1} \quad (7.2.60)$$

utilizando a regra $\text{hvec}(\mathbf{M}_1 \mathbf{M}_2 \mathbf{M}_3) = (\mathbf{M}_1 \otimes \mathbf{M}_3^T) \text{hvec} \mathbf{M}_2$, pode-se exprimir $\zeta_G(k)$ sob a forma alternativa:

$$\zeta_G(k) = N (\text{hvec} \mathbf{R}(k))^T (\mathbf{C}_0^{1/2} \otimes \mathbf{C}_0^{1/2} \mathbf{C}(0)^{-1}) (\mathbf{C}(0) \otimes \mathbf{I}_m)^{-1} \mathbf{x}$$

$$x(\sum_{\underline{u}} \hat{G}(u) \otimes \hat{G}(u))^{-1} (I_m \otimes C(0)^{-1})^{-1} (C_0^{1/2} \otimes C(0)^{-1} C_0^{1/2}) \text{vec} R(k) \quad (7.2.61)$$

Através de nova utilização da regra $(M_1 \otimes M_2)(M_3 \otimes M_4) = M_1 M_3 \otimes M_2 M_4$, (7.2.60) assume a expressão:

$$\zeta_G(k) = N(\text{vec} R(k))^T (\sum_{\underline{u}} (C_0^{-1/2} \otimes C_0^{-1/2})(C_0^{1/2} R(u) C_0^{1/2} C(0)^{-1} \otimes C_0^{1/2} R(u) C_0^{1/2} C(0)^{-1})(C(0) C_0^{-1/2} \otimes C(0) C_0^{-1/2}))^{-1} \text{vec} R(k) = \zeta_p(k) \quad (7.2.61)$$

Finalmente, pode-se mostrar, através da utilização de (7.2.58), que a expressão correspondente a (7.2.52) válida para este caso é (Tjøstheim e Paulsen, 1982):

$$\text{cov} \{ \text{tr} \hat{G}(k), \text{tr} \hat{G}(r)^T \} = N^{-1} \sum_{u=q}^k \text{tr} \{ G(u) \otimes G(k-r-u) \} \quad (7.2.63)$$

para $k, r > q$, o que permitirá testar a significância estatística de $\text{tr} \hat{G}(k)$.

Conforme foi referido, os valores próprios de todas estas matrizes seriam também instrumentos de grande utilidade na identificação de um modelo. No entanto, Tjøstheim e Paulsen (1982) revelam a sua incapacidade para obter os respectivos desvios-padrão, o que retira a possibilidade de testar a significância estatística dos elementos destas funções escalares e, em consequência, significa que a sua análise será pouco exacta, limitando-se às características gerais dos seus padrões de comportamento.

Assim, o procedimento proposto por estes autores baseia-se nos resultados acabados de analisar: em primeiro lugar, examina-se os

valores dos traços das matrizes $R(k)$, $P(k)$ e $\hat{G}(k)$ para os diferentes valores de k . Admitindo as hipóteses atrás colocadas, qualquer destas grandezas segue distribuição assintótica Normal, sendo assim possível testar a sua significância estatística a partir dos resultados anteriores (Tjøstheim e Paulsen, 1982).

Deste modo, se $\text{tr}R(k)$ ou $\text{tr}\hat{G}(k)$ se anulam a partir de $k > q$, devendo q ser um valor pequeno, identifica-se um modelo $MA(q)$. Se, pelo contrário, isso não acontece com estas matrizes, mas se verifica para $\text{tr}\hat{P}(k)$, com $k > p$, identifica-se um modelo $AR(p)$. O conhecimento das expressões dos desvios-padrão de cada uma destas variáveis permite imediatamente testar se se verifica qualquer comportamento deste tipo. Contudo, esta análise deve ser feita com uma certa precaução, uma vez que é possível que $\text{tr}M=0$ mesmo que M seja uma matriz não nula, o que significa que o exame dos valores dos traços daquelas matrizes pode fornecer indicações erradas.

Tendo em vista ultrapassar esta dificuldade, é proposto o cálculo das estatísticas $\zeta_p(k)$ e $\zeta_q(k)$ por um lado e, por outro, a análise do valor absoluto dos valores próprios de $R(k)$, $\hat{G}(k)$ e de $\hat{P}(k)$, devendo estes ser previamente ordenados por ordem decrescente. No entanto, conforme já foi referido, não é possível testar a significância estatística dos elementos desses valores próprios devido ao desconhecimento dos desvios-padrão, o que obriga a que a sua análise seja meramente indicativa. De qualquer modo, apesar da sua falta de exactidão, é possível que este exame forneça indicações úteis, pois é plausível que os valores próprios apresentem os mesmos padrões de comportamento que as respectivas matrizes ou os seus traços.

Por outro lado, os autores desta proposta não indicam como identificar as ordens de um modelo $ARMA(p,q)$ uma vez que, nesse caso, nenhuma das funções escalares analisadas sofrerá um corte

abrupto. No entanto, parece ser admissível afirmar que o valor de p será o "lag" para o qual $\text{tr}\hat{P}(k)$ começa a apresentar um decréscimo lento para zero e o valor de q será o "lag" a partir do qual $\text{tr}R(k)$ ou $\text{tr}\hat{G}(k)$ apresentam um comportamento desse tipo, devendo as características dos padrões de comportamento apresentados pelos valores próprios das respectivas matrizes serem as mesmas.

Finalmente, é conveniente referir que, se $\{X_t\}$ for um processo quase não estacionário, ou seja, se $|\phi(B)|$ possuir zeros próximos do círculo unitário, Tjøstheim e Paulsen (1982) propõem a estacionarização daquele através da transformação:

$$W_t = X_t - \alpha X_{t-1} \quad (7.2.64)$$

onde α é uma matriz de dimensão $(m \times m)$, conforme já foi definido atrás em (2.1.71). É possível agora determinar a forma da matriz α .

Assim, se a função de autocorrelação de cada processo univariado componente de $\{X_t\}$ assumir valores persistentemente elevados em valor absoluto, será geralmente necessária a aplicação de uma transformação do tipo de (7.2.64), uma vez que a quase não estacionaridade de um processo multivariado impede que aquela função assumia valores não significativos para pequenos "lags". A título de exemplo, considere-se o processo AR(1):

$$X_t - \phi_1 X_{t-1} = \varepsilon_t \quad (7.2.65)$$

Para este processo $G(k) = \phi_1^k = [G(1)]^k$ e os valores próprios de $G(k)$ e de ϕ_1 , designados respectivamente por $\lambda_{G_i}(k)$ e por λ_{ϕ_i} ($i=1, \dots, m$) estão relacionados pela equação (Tjøstheim e Paulsen, 1982):

$$\lambda_{G_i}(k) = \lambda_{\phi_i}^k \quad (7.2.66)$$

Esta expressão sugere a utilização dos valores próprios de $\hat{G}(k)$ para a detecção da existência de quase não estacionaridade, uma vez que a quase não estacionaridade de grau r ($r \leq m$) corresponde à existência de r valores próprios de ϕ_1 próximos do círculo unitário (Tjøstheim e Paulsen, 1982).

Deste modo, os valores próprios de $\hat{G}(k)$, designados por $\hat{\lambda}_{G_i}(k)$, são em geral complexos, mas Tjøstheim e Paulsen (1982) mostram que:

$$|\lambda_{G_i}(k)| \leq 1$$

$$|\hat{\lambda}_{G_i}(k)| \leq 1 \quad i = 1, \dots, m \quad (7.2.67)$$

Uma vez que, conforme indicam os referidos autores, os valores próprios de $\rho(k)$ e de $P(k)$ não verificam a propriedade (7.2.67), conclui-se imediatamente que, devido ao facto de $\lambda_{G_i}(k)$ serem desconhecidos, $|\hat{\lambda}_{G_i}(k)|$ podem ser de extrema utilidade para a detecção de quase não estacionaridade, pois estes últimos possuem a propriedade (7.2.67). Com efeito, estes autores propõem o exame dos valores $\hat{\lambda}_G^{(i)}(k)$ como função de k , sendo $\hat{\lambda}_G^{(i)}(k)$ o i -ésimo maior valor próprio, em valor absoluto, da matriz $\hat{G}(k)$. Conforme já foi referido, a análise destes valores próprios pode também ser útil para a identificação de uma série cronológica.

Nestas condições, a existência de quase não estacionaridade de grau r leva a que $|\hat{\lambda}_G^{(i)}(k)|$, $i \leq r$ assumam valores comparativamente mais elevados do que $|\hat{\lambda}_G^{(i)}(k)|$ para $i > r$. Esta análise pode também revelar se é necessário utilizar uma transformação do tipo de (7.2.64) e qual deverá ser. Assim, se $|\hat{\lambda}_G^{(i)}(k)|$, $i = 1, \dots, m$, apresentam todos um decréscimo lento para zero, $\alpha = I_m$ deverá ser a transformação escolhida, correspondendo à diferenciação separada de cada componente do vector

X_t . Se não houver qualquer forte indicação sobre esta escolha, é possível utilizar $\hat{\alpha} = \hat{G}(1)$, embora em tal situação possa ser vantajoso analisar a série sem qualquer transformação destinada a eliminar a quase não estacionaridade (Tjøstheim e Paulsen, 1982).

Por outro lado, uma vez que os métodos acabados de analisar conduzem geralmente à identificação da forma completa de um modelo ARMA(p,q) representada em (7.2.1), envolvendo todos os parâmetros, será geralmente necessário proceder a uma ou mais iterações após a estimação do modelo completo (7.2.1), de modo a eliminar da sua estrutura todos os parâmetros que se revelem não significativos. Assim, enquanto houver parâmetros estimados não significativos, é necessário eliminá-los do modelo, o que conduz a um novo modelo, mais simplificado, que deverá ser novamente estimado e testado e assim sucessivamente. No fim deste procedimento, obtém-se um modelo mais simples, em que todos os parâmetros estimados deverão ser significativos, evitando-se deste modo o problema da sobreparametrização, que é particularmente sensível num modelo multivariado. Portanto, após a identificação da estrutura inicial de um modelo, torna-se geralmente necessário proceder a diversas simplificações nessa estrutura, observando-se deste modo o princípio da parcimónia, o que significa que deve ser conferida particular relevância ao critério de significância estatística dos parâmetros estimados, permitindo assim a identificação de um modelo mais adequado.

Na realidade, os símbolos atribuídos aos elementos das matrizes utilizadas para a identificação não devem ser considerados exactos, apenas constituindo instrumentos bastante imperfeitos que pretendem auxiliar na selecção de um modelo inicial. Com efeito, a título de exemplo, Tiao e Box (1981) indicam que as variâncias dos elementos de

$R(k)$ podem ser bastante superiores a $N^{-1/2}$, no caso de as séries univariadas estarem altamente autocorrelacionadas, o que pode conduzir à sobreparametrização do modelo. Tomados no seu conjunto, todos estes indicadores podem ser extremamente úteis na percepção do padrão geral da correlação apresentado pela série, não devendo ser interpretados como testes de hipóteses formais, mas apenas como formas de orientação para a identificação preliminar de um modelo.

De qualquer modo, e em suma, estes métodos e os seus refinamentos (com a excepção da proposta de Jenkins e Alavi, 1981, que recorre aos modelos univariados) pretendem proceder à identificação de um modelo para uma série cronológica multivariada de uma forma directa, isto é, considerando à partida um modelo que relacione directamente os elementos que compõem o vector X_t , ou seja, um modelo do tipo de (7.2.1). Apesar de constituírem generezizações da metodologia de Box e Jenkins (1976) para o caso univariado, estes métodos parecem assim responder bastante bem às exigências que se colocam ao processo de modelização de uma série multivariada, nomeadamente no que respeita à capacidade de suportar valores elevados de m , ou seja, do número de elementos de X_t . Com efeito, tomando em linha de conta a totalidade das relações dinâmicas entre as m séries univariadas, são utilizados diversos instrumentos, tendo em vista fazer luz sobre a estrutura dessas relações.

3. Identificação de uma estrutura simplificada

Conforme já foi constatado, os modelos ARMA revelam-se extremamente vantajosos na modelização de séries cronológicas multivariadas com um qualquer número de componentes, uma vez que são capazes de representar parcimoniosamente uma vasta gama de processos estocásticos e, portanto, resumir a estrutura destes num modelo mais simples. No entanto, apresentam o inconveniente de a sua identificação, estimação e crítica poderem não ser muito acessíveis, implicando a disponibilidade de meios computacionais nem sempre existente.

Assim, têm sido propostos diversos métodos que pretendem simplificar a identificação de uma série cronológica multivariada, quer através da introdução de simplificações num modelo ARMA previamente identificado, quer através da consideração de novas classes de processos que originam novas estruturas, quer ainda através de análises paralelas que pretendem fazer luz sobre a estrutura subjacente à geração de uma série cronológica.

Neste contexto, proceder-se-à primeiro à análise de um método que, após a identificação de um modelo ARMA, pretende determinar quais os seus parâmetros significativos, facilitando grandemente a posterior estimação. Em seguida, descreve-se um método que recorre à análise factorial, visando a identificação de factores ocultos subjacentes à série cronológica, permitindo a determinação de uma transformação simplificadora para a sua representação e pretendendo evitar o aparecimento de uma estrutura de relações complexa e mal definidas que surgiria no caso de se tentar representar a série através de modelos ARMA. Em terceiro lugar, considera-se modelos índice auto-regressivos, cujo objectivo consiste em representar uma série cronológica através de um modelo auto-regressivo expresso em termos

de uma outra série, designada de índice, com um menor número de componentes e que contém toda a informação relevante sobre a série original, possibilitando assim uma representação mais parcimoniosa sem perda de informação. No seguimento deste método, analisa-se modelos auto-regressivos de polinómio descontado, cujo objectivo é serem de identificação e tratamento mais simples que os modelos ARMA, recorrendo à representação de uma série cronológica como um modelo auto-regressivo de ordem infinita e determinando uma forma de limitar o seu número de parâmetros. Por fim, passa-se à análise canónica que, sendo uma análise paralela à identificação de um dado modelo, pretende obter maior informação sobre este, detectando a existência de diversos tipos de relações entre as componentes da série multivariada, quais são as responsáveis pela eventual não estacionaridade da série e propondo uma transformação que pretende ilustrar estes aspectos.

A grande questão que aqui se coloca é a de saber se as propostas que agora irão ser analisadas têm êxito, verificando-se que, afinal, algumas são de resultados duvidosos, acabando por conduzir à conclusão de que, no fim de contas, o recurso a modelos ARMA é ainda preferível, embora se possa recorrer marginalmente a um ou a outro método simplificador de um modelo pertencente a essa classe.

Note-se ainda que se admitirá em todas as abordagens que o valor esperado do processo estocástico considerado é o vector nulo.

3.1. Identificação dos elementos significativos nos operadores de um modelo ARMA

A complexidade associada à identificação da estrutura de um modelo para uma série cronológica multivariada reside não só na selecção

das ordens dos operadores, mas também na determinação dos elementos significativos destes últimos. Na realidade, conforme foi referido atrás, os métodos de identificação até agora analisados obrigam a efectuar a estimação do modelo o número de vezes que for necessário para se obter uma estrutura com todos os parâmetros estimados significativos. Assim, se, por exemplo, se tiver $p=q=1$ e $m=5$, é necessário na primeira fase estimar 65 parâmetros (25 valores para cada operador e 15 valores para a matriz das variâncias e covariâncias). No entanto, na maior parte das situações práticas, o número de parâmetros estimados significativos é muito baixo, comparativamente ao número total de parâmetros de um modelo. Consequentemente, a estimação de um modelo que inclua a totalidade dos parâmetros é uma forma muito pouco eficiente de utilizar os meios computacionais.

Por esta razão, Koreisha e Pukkila (1987) propõem um método que visa identificar os elementos significativos na estrutura de um modelo ARMA, o que permitirá também gerar estimativas iniciais para os parâmetros, tendo em vista a posterior estimação através do método da máxima verodimilhança. Esta identificação / estimação preliminar dos elementos significativos permite reduzir substancialmente o tempo de computação necessário à subsequente estimação dos parâmetros, possibilitando a experimentação de várias estruturas sem grande aumento de custos. Com efeito, os autores desta proposta indicam que o seu procedimento pode reduzir pelo menos 10 vezes o tempo necessário à estimação de um modelo que inclua todos os parâmetros, o que engloba não só os elementos dos operadores mas também os seus desvios-padrão. É a referida proposta que se irá analisar.

Deste modo, ir-se-à explicar o método para o caso de um modelo ARMA(1,1), o que permitirá generalizar em seguida para o caso de modelos ARMA(p,q).

É útil sublinhar que este método pressupõe a identificação prévia das ordens de um modelo ARMA(p,q), que assume a forma genérica:

$$\phi(B) X_t = \theta(B) \varepsilon_t \quad (7.3.1)$$

onde $\phi(B) = I_m - \phi_1 B - \dots - \phi_p B^p$ e $\theta(B) = I_m + \theta_1 B + \dots + \theta_q B^q$, conforme já foi previamente definido. Admitindo que o modelo (7.3.1) é invertível e estacionário, é possível representá-lo sob qualquer das formas:

$$X_t = \psi(B) \varepsilon_t \quad ; \quad \psi(B) = \sum_{l=0}^{\infty} \psi_l B^l \quad \psi_0 = I_m \quad (7.3.2)$$

$$\pi(B) X_t = \varepsilon_t \quad ; \quad \pi(B) = I_m - \sum_{l=1}^{\infty} \pi_l B^l \quad (7.3.3)$$

(i) Utilização do método para modelos ARMA(1,1)

O método desenvolve-se em duas fases: em primeiro lugar, procede-se à estimação preliminar dos parâmetros e, em seguida, testa-se a sua significância estatística, o que permitirá inferir qual a estrutura final do modelo para a subsquente estimação de máxima verosimilhança.

a) Estimação preliminar dos parâmetros

Para um modelo ARMA(1,1), é fácil demonstrar a existência das relações:

$$\pi_1 = \psi_1 = \phi_1 + \theta_1$$

$$\pi_l = \theta_1 \pi_{l-1} \quad \geq 2$$

$$\psi_l = \phi_1 \psi_{l-1} \quad \geq 2 \quad (7.3.4)$$

Se for possível determinar os estimadores de π_1, \dots, π_M e de ψ_1, \dots, ψ_M , sendo M um inteiro positivo, consegue-se obter estimadores lineares para ϕ_1 e θ_1 a partir de (7.3.4) através da minimização em ordem a ϕ_1 e θ_1 da soma de quadrados:

$$S(\phi_1, \theta_1) = \text{tr}(\hat{\pi}_1 - \phi_1 - \theta_1)(\hat{\pi}_1 - \phi_1 - \theta_1)^T +$$

$$+ \sum_{i=2}^M \text{tr}(\hat{\pi}_i - \theta_1 \hat{\pi}_{i-1})(\hat{\pi}_i - \theta_1 \hat{\pi}_{i-1})^T + \sum_{i=2}^M \text{tr}(\hat{\psi}_i - \phi_1 \hat{\psi}_{i-1})(\hat{\psi}_i - \phi_1 \hat{\psi}_{i-1})^T \quad (7.3.5)$$

onde $\hat{\pi}_i$ e $\hat{\psi}_i$ designam os estimadores de π_i e de ψ_i respectivamente ($i=1, \dots, M$).

Igualando a zero o resultado da diferenciação de (7.3.5) em ordem a ϕ_1 e a θ_1 , obtém-se o sistema de equações que permite obter os estimadores preliminares destas matrizes:

$$\begin{bmatrix} \hat{\phi}_1 & \hat{\theta}_1 \end{bmatrix} \begin{bmatrix} \mathbf{V}_{11} & \mathbf{S}_{11} \\ \mathbf{T}_{11} & \mathbf{U}_{11} \end{bmatrix} = -[\mathbf{E}_1 \quad \mathbf{F}_1] \quad (7.3.6)$$

onde:

$$\mathbf{V}_{11} = \mathbf{I}_m + \sum_{i=2}^M \hat{\psi}_{i-1} \hat{\psi}_{i-1}^T ; \quad \mathbf{S}_{11} = -\mathbf{I}_m$$

$$\mathbf{T}_{11} = -\mathbf{I}_m ; \quad \mathbf{U}_{11} = \mathbf{I}_m + \sum_{i=2}^M \hat{\pi}_i \hat{\pi}_i^T$$

$$\mathbf{E}_1 = \hat{\pi}_1 + \sum_{i=2}^M \hat{\psi}_i \hat{\psi}_{i-1}^T ; \quad \mathbf{F}_1 = \hat{\pi}_1 + \sum_{i=2}^M \hat{\pi}_i \hat{\pi}_{i-1}^T$$

Deste modo, os autores referidos designam este método de estimação por método SQ. É necessário agora determinar estimadores para as matrizes π_i e, conseqüentemente, para ψ_i . Para o efeito, aproxima-se o modelo (7.3.1) através de um modelo auto-regressivo de ordem elevada:

$$\mathbf{X}_t = \sum_{i=1}^M \pi_i \mathbf{X}_{t-i} + \boldsymbol{\varepsilon}_t \quad (7.3.7)$$

onde se admite que as matrizes π_i se anulam a partir do "lag" M. Para a estimação de π_1, \dots, π_M , Koreisha e Pukkila (1987) propõem dois métodos: Auto-Regressão Longa (ARL, de forma abreviada) e Auto-Regressões Sucessivas (ARS). O primeiro método consiste em recorrer às equações de Yule-Walker para um modelo AR(M) e, no segundo, os estimadores são determinados através do ajustamento de vários modelos

auto-regressivos. Neste segundo método, determina-se o estimador $\hat{\pi}_{11}$ através das equações de Yule-Walker para um modelo AR(1) e, em seguida, procede-se a um ajustamento de um modelo AR(2) aos resíduos resultantes do ajustamento anterior, o que permite obter os estimadores $\hat{\pi}_{21}$ e $\hat{\pi}_{22}$ através das equações de Yule-Walker para um modelo AR(2). A aplicação sucessiva deste procedimento fornece os estimadores $\hat{\pi}_{p1}, \dots, \hat{\pi}_{pp}$, sendo p a ordem máxima estabelecida para a auto-regressão. Os estimadores $\hat{\pi}_1, \dots, \hat{\pi}_M$ podem então ser obtidos a partir de:

$$\hat{\pi}(B) = (I_m - \hat{\pi}_{p1} B - \dots - \hat{\pi}_{pp} B^p) \dots (I_m - \hat{\pi}_{21} B - \hat{\pi}_{22} B^2) (I_m - \hat{\pi}_{11} B) \quad (7.3.8)$$

Os autores deste método referem que fixam habitualmente $p = \lceil \sqrt{N} / 2 \rceil$, designando $[z]$ o maior inteiro não superior a z e que, com base em resultados de simulação, tomam $M=10$ em (7.3.5), uma vez que os parâmetros auto-regressivos se anulavam para "lags" superiores a 10. Paralelamente, os mesmos resultados indicavam que as estimativas obtidas para os parâmetros do modelo ARMA não sofriam alterações dignas de nota quando o número de matrizes auto-regressivas incluídas na função objectivo (7.3.5) era superior a 10. Com efeito, este procedimento pode ser encarado como um método de conversão de $\{X_t\}$ num ruído branco, o que é conseguido após um pequeno número de aplicações dos filtros auto-regressivos, o que significa que a ordem do modelo (7.3.7) não necessita de ser muito elevada. Por outro lado, os autores referidos preferem o método ARS, uma vez que este produz estimadores para os parâmetros auto-regressivos com menor enviesamento do que o método ARL.

Finalmente, estes autores afirmam ainda que, para que os estimadores resultantes de (7.3.5) sejam consistentes, é necessário que o valor de M em (7.3.7) varie a um ritmo dependente da dimensão da série cronológica. Quando M tende para infinito a um ritmo dependente de N ,

os estimadores π_l e ψ_l tendem em probabilidade para os verdadeiros valores dos parâmetros, ou seja, verificam-se as convergências em probabilidade:

$$\begin{aligned} \text{plim } \hat{\pi}_l &= \pi_l \\ \text{plim } \hat{\psi}_l &= \psi_l \quad l = 1, 2, \dots \end{aligned} \quad (7.3.9)$$

Isto implica ainda que todos os elementos em (7.3.6) convirjam em probabilidade para os seus limites teóricos, isto é, o valor de M pode ser substituído por infinito e os estimadores dos parâmetros pelos verdadeiros valores correspondentes. No limite, a solução da equação (7.3.6) permite obter ϕ_1 e θ_1 porque, por um lado, o valor mínimo da soma de quadrados (7.3.5), ou seja, 0, é atingido quando se substitui os estimadores obtidos na função objectivo e, por outro, os verdadeiros valores dos parâmetros minimizam aquela soma.

A partir da estimação dos parâmetros, torna-se agora possível determinar quais os que são significativos.

b) Identificação dos elementos não nulos

O objectivo principal desta fase consiste na geração de estimadores iniciais para os parâmetros cujo valor é não nulo, de modo a reduzir o custo computacional da subsequente estimação pelo método da máxima verosimilhança.

Assim, a identificação preliminar dos elementos não nulos de ϕ_1 e de θ_1 pode ser feita através da obtenção dos desvios-padrão dos estimadores resultantes de (7.3.6). No entanto, uma vez que, conforme indicam os autores referidos, ainda não foi possível determinar as expressões das variâncias e covariâncias assintóticas destes estimadores, torna-se necessário recorrer a uma forma de cálculo aproximado dos desvios-padrão desejados do tipo da que foi apresentada para estimação de

máxima verosimilhança dos parâmetros de um modelo ARMA(p,q), conforme apresentado em (3.4.8) e (3.4.9).

Por outro lado, da estimação da matriz de variâncias e covariâncias dos estimadores dos parâmetros resultam também novos estimadores para os próprios parâmetros, o que permite actualizar essa matriz. Este procedimento pode ser repetido até ser obtida a sua convergência com base num qualquer critério como, por exemplo, a variação do valor do determinante da matriz de variâncias e covariâncias residual. De qualquer modo, os autores deste método indicam que os resultados do estudo de simulação já referido apontam no sentido de ser necessário um reduzido número de iterações, uma vez que a qualidade das estimativas obtidas nesta fase preliminar é muito elevada. Paralelamente, e corroborando estas conclusões, as estimativas obtidas no referido estudo não sofrem grandes alterações após algumas iterações realizadas tomando como valores iniciais os resultantes de (7.3.6). Além disso, a identificação dos elementos não nulos também se mantém relativamente estável após a realização de uma ou duas iterações adicionais.

Neste contexto, a identificação preliminar dos elementos não nulos é feita em primeiro lugar através da verificação de quais as estatísticas t cujo valor absoluto é superior a 2. Em seguida, a comparação dos elementos significativos de $\hat{\pi}_1$, sendo $\hat{\pi}_1 = \hat{\phi}_1 + \hat{\theta}_1$, com os resultados da identificação preliminar, permite eliminar ou incorporar os parâmetros na estrutura do modelo que será posteriormente estimado. Essa comparação é feita do seguinte modo: um parâmetro incluído na fase preliminar é eliminado se o correspondente (i,j)-ésimo elemento da matriz $\hat{\pi}_1$ for não significativo, mesmo que se verifique que o (i,j)-ésimo elemento do operador auto-regressivo ou o (i,j)-ésimo elemento do operador de médias móveis é significativo e desde que tal não se verifique simultaneamente para os dois; um parâmetro eliminado na fase

preliminar é incorporado no modelo se o correspondente (i,j) -ésimo elemento da matriz $\hat{\pi}_1$ for significativo, não o sendo nenhum dos (i,j) -ésimos elementos do operador auto-regressivo e do operador de médias móveis. O motivo para a inclusão desta verificação adicional é o facto de as estimativas obtidas pelos referidos autores para as matrizes π_1 estarem muito próximas dos valores simulados. Daqui se conclui que não utilizar esta informação seria um desperdício, uma vez que ela se revela de grande exactidão. Com efeito, estes autores referem que acontece frequentemente que, quando apenas um parâmetro estimado na (i,j) -ésima posição é significativo (quer seja auto-regressivo, quer seja de médias móveis) mas, por qualquer razão, ambos os (i,j) -ésimos parâmetros são incorporados no modelo, as estimativas resultantes indicam geralmente que nenhum dos parâmetros deve ser incluído no modelo. Isto conduz o analista a ignorar erradamente o problema ou a ter que proceder a mais iterações, o que é desnecessário.

O estudo de simulação realizado por estes autores indica que o seu método reduz substancialmente o número de parâmetros que é necessário estimar através do método da máxima verosimilhança, raramente identificando de uma forma errada um parâmetro não nulo. No entanto, nota-se ainda uma certa tendência para a sobre-estimação do número de parâmetros não nulos, bem como para falhar a identificação de parâmetros de baixo valor, embora tal tendência seja na realidade muito ligeira. Por outro lado, o tempo de computação total necessário à estimação dos parâmetros e da matriz de variâncias e covariâncias dos estimadores é muito inferior quando se utiliza as estimativas obtidas na fase preliminar como valores iniciais e a identificação dos elementos não nulos aí efectuada, relativamente à estimação de máxima verosimilhança em que essa informação não é utilizada.

Após esta análise, é agora possível descrever o método proposto para o caso genérico de modelos ARMA(p,q).

(ii) Utilização do método para modelos ARMA(p,q) genéricos

A extensão do método para modelos ARMA(p,q) em geral, e que inclui os modelos MA(q) e AR(p), é simples e não parece acarretar problemas especiais, revelando-se tanto mais vantajoso quanto maior o número de componentes da série cronológica.

Assim, os estimadores preliminares das matrizes que compõem os operadores $\phi(B)$ e $\theta(B)$ de um qualquer modelo ARMA(p,q) podem ser obtidos a partir do sistema de equações:

$$\begin{bmatrix} \hat{\phi} & \hat{\theta} \end{bmatrix} \begin{bmatrix} \mathbf{V} & \mathbf{S} \\ \mathbf{T} & \mathbf{U} \end{bmatrix} = -[\mathbf{E} \quad \mathbf{F}] \quad (7.3.10)$$

onde:

$$\begin{bmatrix} \hat{\phi} & \hat{\theta} \end{bmatrix} = \begin{bmatrix} \hat{\phi}_1 \dots \hat{\phi}_p & \hat{\theta}_1 \dots \hat{\theta}_q \end{bmatrix}$$

$$[\mathbf{E} \quad \mathbf{F}] = \begin{bmatrix} \mathbf{E}_1 \dots \mathbf{E}_p & \mathbf{F}_1 \dots \mathbf{F}_q \end{bmatrix}$$

As matrizes \mathbf{V} , \mathbf{S} , \mathbf{T} e \mathbf{U} têm como elementos genéricos os blocos de dimensão $(m \times m)$ \mathbf{V}_{uv} , \mathbf{S}_{uv} , \mathbf{T}_{uv} , e \mathbf{U}_{uv} respectivamente, assumindo estes a forma:

$$\mathbf{V}_{11} = \mathbf{I}_m + \sum_{l=2}^M \hat{\psi}_{l-1} \hat{\psi}_{l-1}^T \quad ; \quad \mathbf{V}_{uu} = 2\mathbf{I}_m + \sum_{l=u+1}^M \hat{\psi}_{l-u} \hat{\psi}_{l-u}^T \quad u = 2, \dots, p$$

$$\mathbf{V}_{uv} = \hat{\psi}_{v-u}^T + \sum_{l=v+1}^M \hat{\psi}_{l-u} \hat{\psi}_{l-v}^T \quad u < v; \quad \mathbf{V}_{vu} = \mathbf{V}_{uv} \quad u < v$$

$$\mathbf{U}_{11} = \mathbf{I}_m + \sum_{l=2}^M \hat{\pi}_{l-1} \hat{\pi}_{l-1}^T \quad ; \quad \mathbf{U}_{uu} = 2\mathbf{I}_m + \sum_{l=u+1}^M \hat{\pi}_{l-u} \hat{\pi}_{l-u}^T \quad u = 2, \dots, q$$

$$\mathbf{U}_{uv} = -\hat{\pi}_{v-u}^T + \sum_{l=v+1}^M \hat{\pi}_{l-u} \hat{\pi}_{l-v}^T \quad u < v; \quad \mathbf{U}_{vu} = \mathbf{U}_{uv}^T \quad u = 2, \dots, q$$

$$\begin{aligned}
S_{11} &= -I_m \quad ; \quad S_{uu} = -2I_m \quad u > 1 \\
S_{uv} &= \hat{\psi}_{v-u} \quad u < v \quad ; \quad S_{vu} = \hat{\pi}_{v-u}^T \quad u < v \\
T_{uv} &= S_{vu} \quad u = 1, \dots, q \quad v = 1, \dots, p \\
E_1 &= -\hat{\pi}_1 - \sum_{l=2}^M \hat{\psi}_l \hat{\psi}_{l-1}^T \quad ; \quad E_u = -\hat{\pi}_u - \hat{\psi}_u - \sum_{l=u+1}^M \hat{\psi}_l \hat{\psi}_{l-u}^T \quad u = 2, \dots, p \\
F_1 &= -\hat{\pi}_1 - \sum_{l=2}^M \hat{\pi}_l \hat{\pi}_{l-1}^T \quad ; \quad F_u = -\hat{\pi}_u + \hat{\psi}_u - \sum_{l=u+1}^M \hat{\pi}_l \hat{\pi}_{l-u}^T \quad u = 2, \dots, q
\end{aligned} \tag{7.3.11}$$

É visível que as dificuldades adicionais são relativamente pequenas. Os autores desta proposta indicam que as vantagens da utilização deste método são mais notórias quando $m > 3$.

O potencial deste método parece ser realmente grande, podendo este ser um bom auxiliar de qualquer método conducente à identificação de uma série cronológica e permitindo ter maior confiança na selecção dos parâmetros incluídos na estrutura do modelo que será posteriormente estimado.

3.2. Modelo factorial

Quando uma série cronológica multivariada é função de um conjunto de factores ocultos, é necessário um grande número de parâmetros para obter uma representação adequada do sistema e os parâmetros estimados tornam-se altamente correlacionados. Assim, pode surgir uma estrutura de relações complexa e mal definida quando, na realidade, o que se verifica é a existência de um modelo mais simples e parcimonioso dependente apenas de alguns factores comuns (Peña e Box, 1987).

Deste modo, Peña e Box (1987) propõem um método de identificação de factores ocultos, indicando o seu número, e de

determinação de uma transformação simplificadora para representar a série. É esse método que vai ser agora analisado.

Em termos genéricos, o modelo que servirá para identificar os factores comuns de uma série cronológica multivariada assume a forma:

$$X_t = WY_t + \xi_t \quad (7.3.12)$$

onde X_t designa a série cronológica com m componentes, W é uma matriz de dimensão $(m \times r)$ de parâmetros desconhecidos, sendo r o número de factores, $\{Y_t\}$ é um processo estocástico não observável de dimensão $(r \times 1)$ e $\{\xi_t\}$ é um processo puramente aleatório de dimensão $(m \times 1)$ com matriz de covariância Σ_ξ , que se admite ter característica m . A matriz W é designada por matriz dos pesos dos factores, cujo elemento genérico W_{il} representa o peso do l -ésimo factor em X_{it} ($i=1, \dots, m$; $l=1, \dots, r$). Este modelo só é relevante se $r < m$ ou se $r=m$ e $\Sigma_\xi = 0$. No primeiro caso, ($r < m$), é possível obter uma redução da dimensão sem perda de informação e, no segundo caso ($r=m$), um problema importante consiste em determinar uma transformação linear da série que permita uma representação mais simples do sistema.

Neste contexto, irá ser analisado primeiro o modelo com factores independentes e, em seguida, o modelo com factores dependentes. Em ambos os casos, será apresentada uma transformação que visa conseguir uma simplificação da estrutura do modelo.

(i) Modelo com factores independentes

Em primeiro lugar, proceder-se-á à caracterização do modelo, o que permitirá determinar em seguida a respectiva transformação simplificadora.

a) Caracterização do modelo

Considere-se o modelo (7.3.12), onde se admite que a série cronológica X_t é estacionária e é gerada por r ($r \leq m$) factores que formam o vector

$\{Y_t\}$ que, por sua vez, é um processo ARMA(p_y, q_y) invertível e estacionário que assume a forma:

$$\phi_y(B)Y_t = \theta_y(B)\eta_t \quad (7.3.13)$$

onde:

$$\begin{aligned} \phi_y(B) &= I_m - \phi_{1y}B - \dots - \phi_{p_y}B^{p_y} \\ \theta_y(B) &= I_m - \theta_{1y}B - \dots - \theta_{q_y}B^{q_y} \end{aligned} \quad (7.3.14)$$

são polinômios matriciais em B e as matrizes ϕ_{ly} ($l=1, \dots, p_y$) e θ_{uy} ($u=1, \dots, q_y$) têm dimensão ($r \times r$). Admite-se ainda que $\{\eta_t\}$ é um processo puramente aleatório seguindo distribuição Normal, com valor esperado igual ao vector nulo e matriz de variâncias e covariâncias Σ_η , sobre a qual a única restrição imposta é a de ser definida positiva.

Sendo A uma matriz não singular e de dimensão ($r \times r$) qualquer, a equação geradora (7.3.12) pode ser expressa da forma:

$$X_t = W'Y_t' + \xi_t \quad (7.3.15)$$

onde $W' = WA^{-1}$ é a nova matriz de parâmetros e $Y_t' = AY_t$ é uma transformação linear dos factores. Multiplicando (7.3.13) por A, obtém-se:

$$A\phi_y(B)A^{-1}AY_t = A\theta_y(B)A^{-1}A\eta_t \quad (7.3.16)$$

ou seja:

$$p \lim \phi'_y(B)Y_t' = \theta'(B)\eta_t' \quad (7.3.17)$$

o que significa que o modelo para o novo conjunto de factores é também um modelo ARMA(p_y, q_y) com r componentes e os seguintes parâmetros:

$$\phi'_y(B) = A\phi_y(B)A^{-1}$$

$$\theta'_y(B) = A\theta_y(B)A^{-1}$$

$$\Sigma'_\eta = A\Sigma_\eta A^T \quad (7.3.18)$$

Para eliminar esta indeterminação, ou seja, para tornar o modelo identificável, admite-se que $W^T W = I_m$.

Designando as matrizes de covariância de $\{X_t\}$ e de $\{Y_t\}$ por $\Gamma_x(k)$ e $\Gamma_y(k)$ respectivamente, tem-se:

$$\Gamma_x(0) = W\Gamma_y(0)W^T + \Sigma_\xi$$

$$\Gamma_x(k) = W\Gamma_y(k)W^T \quad k \geq 1 \quad (7.3.19)$$

sendo a característica de $\Gamma_x(k)$ para $k \geq 1$ igual ao número de factores, que é de r . Se os factores forem independentes para todos os valores de k e a matriz Σ_η for diagonal, as matrizes $\Gamma_y(k)$ serão todas diagonais, pelo que as matrizes $\Gamma_x(k)$ serão simétricas para $k \geq 1$ e as colunas de W serão os vectores próprios de $\Gamma_x(k)$ com valores próprios $\gamma_{\mu_y}(k)$, $\mu = 1, \dots, r$, sendo estes elementos da diagonal principal de $\Gamma_y(k)$.

A matriz de correlação parcial assume a forma:

$$\rho_x^T(1) = \Gamma_x(0)^{-1} \Gamma_x(1)$$

$$\rho_x^T(k) = [\Gamma_x(0) - \mathbf{b}^T(k)E(k)^{-1} \mathbf{b}(k)]^{-1} =$$

$$= [\Gamma_x(k) - \mathbf{b}^T(k)E(k)^{-1} \mathbf{d}(k)] \quad k > 1 \quad (7.3.20)$$

onde:

$$\mathbf{b}^T(k) = \left[\Gamma_x^T(k-1) \quad \Gamma_x^T(k-2) \dots \Gamma_x^T(k-1) \right]$$

$$\mathbf{d}^T(k) = \left[\Gamma_x^T(1) \quad \Gamma_x^T(2) \dots \Gamma_x^T(k-1) \right]$$

$$\begin{bmatrix} \Gamma_x(0) & \Gamma_x^T(1) & \dots & \Gamma_x^T(k-2) \\ \Gamma_x(1) & \Gamma_x(0) & \dots & \Gamma_x^T(k-3) \\ \vdots & \vdots & & \vdots \\ \Gamma_x(k-2) & \Gamma_x(k-3) & \dots & \Gamma_x(0) \end{bmatrix} \quad (7.3.21)$$

o que significa que a matriz de correlação parcial pode ser expressa da forma:

$$P^T(k) = F(k) \left[\Gamma_x(k) - \sum_{u=1}^{k-1} \sum_{v=1}^{k-1} \Gamma_x^T(k-u) E(u,v,k)^{-1} \Gamma_x(v) \right] \quad (7.3.22)$$

onde $F(k) = [\Gamma_x(0) - b^T(k)E(k)^{-1}b(k)]^{-1}$ e $E(u,v,k)^{-1}$ é o (u,v) -ésimo elemento da matriz $E(k)^{-1}$. Então:

$$P^T(k) = F(k) W \left[\Gamma_y(k) - \sum_{u=1}^{k-1} \sum_{v=1}^{k-1} \Gamma_y^T(k-1) W^T E(u,v,k)^{-1} W \Gamma_y(v) \right] W^T \quad (7.3.23)$$

Designando agora por ν o espaço nulo de W definido por todos os vectores que verificam a relação $W^T \delta = 0$ e sendo V uma matriz de dimensão $[m \times (m-r)]$ cujas colunas são uma base do espaço ν , tem-se:

$$P^T(k) V = H W^T V = 0. \quad (7.3.24)$$

Esta relação mostra que a característica de $P^T(k)$ é no máximo de r e que os vectores associados aos seus valores próprios nulos pertencem ao espaço ν .

Estas propriedades das matrizes de covariância e de correlação impõem fortes restrições sobre o tipo de processo que pode gerar a série observada X_t e sobre as características dos seus parâmetros. Assim, os autores referidos provam os dois teoremas seguintes:

Teorema 7.3.1. Suponha-se que $X_t = W Y_t + \xi_t$, onde $\{Y_t\}$ é um processo ARMA(p_y, q_y) com r componentes, W é uma matriz de dimensão $(m \times r)$ e

característica $r(r \leq m)$ e $\{\xi_t\}$ é um processo puramente aleatório de dimensão $(m \times 1)$ com matriz de variâncias e covariâncias Σ_ξ .

Então, $\{X_t\}$ é um processo ARMA(p_x, q_x) com m componentes, sendo $p_x = p_y$ e $q_x = \max(p_y, q_y)$.

No entanto, a representação obtida para X_t não é única, o que é demonstrado no segundo teorema:

Teorema 7.3.2. Se X_t é uma série cronológica multivariada gerada por r factores comuns conforme especificado em (7.3.12) e estes verificam o modelo (7.3.13), então X_t pode também ser representado da forma:

$$\phi'_x(B)X_t = \theta'_x(B)U_t$$

onde $\phi'_{i_x} = \phi_{i_x} + J$ e $\theta'_{v_x} = \theta_{v_x} + J$ são os novos parâmetros, sendo J uma matriz não nula de característica $m-r$.

Uma importante conclusão que resulta do teorema 7.3.1. é o facto de as matrizes que compõem o operador auto-regressivo do modelo de X_t verificarem a relação:

$$\phi_{i_x} W = W \phi_{i_y} \quad (7.3.25)$$

cuja solução geral é:

$$\phi_{i_x} = W \phi_{i_y} W^{-1} + J(I_m - WW^{-1}) \quad (7.3.26)$$

onde W^{-1} é uma qualquer matriz inversa generalizada de W que satisfaça a condição $WW^{-1}W = W$ e J é uma qualquer matriz arbitrária, com a única restrição de os zeros de $|\phi_x(B)|$ estarem todos fora do círculo uniário.

Uma vez que a matriz ϕ_{i_y} é diagonal, a equação (7.3.25) mostra que as colunas de W são vectores próprios de ϕ_{i_x} , sendo os respectivos valores próprios os elementos da diagonal principal ϕ_{i_y} . Contudo, a característica

da matriz ϕ_{l_x} pode assumir qualquer valor devido à presença da matriz arbitrária J . Por outro lado, o teorema 7.3.2. indica que, para qualquer solução possível de ϕ_{l_x} , existe um conjunto de matrizes θ_{l_x} que devem verificar certas restrições. Com efeito, os autores deste método referem que, sendo V uma matriz qualquer que verifica a relação:

$$V^T W = 0 \quad (7.3.27)$$

então:

$$V^T \xi_t = V^T \varepsilon_t; \quad V^T \xi_t = V^T \psi_x(B) \varepsilon_t; \quad V^T \psi_v = 0 \quad v \geq 1 \quad (7.3.28)$$

onde $\{\varepsilon_t\}$ é um processo puramente aleatório com matriz de variâncias e covariâncias Σ_ε , constituindo o vector de variáveis residuais do processo $\{X_t\}$:

$$\phi_x(B) X_t = \theta_x(B) \varepsilon_t \quad (7.3.29)$$

Por outro lado, sendo $\{X_t\}$ um processo invertível e estacionário, é possível representá-lo como um processo de médias móveis de ordem infinita, cujas matrizes de parâmetros são as matrizes ψ_v :

$$X_t = \psi(B) \varepsilon_t \quad (7.3.30)$$

Os autores referidos indicam ainda que existe sempre uma matriz V^T que verifica (7.3.27), sugerindo que as primeiras $m-r$ linhas sejam vectores próprios de $W W^T$ linearmente independentes associados a valores próprios nulos. É possível assim colocar nas outras linhas quaisquer r vectores do subespaço gerado pelas primeiras $m-r$ linhas.

Deste modo se, por exemplo, $\{Y_t\}$ for um processo AR(1), ou seja, se $p_y=1$ e $q_y=0$, o sistema de equações que conduz à determinação dos parâmetros é:

$$\Sigma_\xi + W \Sigma_\eta W^T = \Sigma_\varepsilon - \phi_{l_x} \Sigma_\varepsilon \theta_{l_x}^T + \theta_{l_x} \Sigma_\varepsilon \theta_{l_x}^T$$

$$\Sigma_{\xi} \phi_{1_x}^T = \Sigma_{\varepsilon} \theta_{1_x}^T$$

$$\phi_{1_x} W = W \phi_{1_y} \quad (7.3.31)$$

onde ϕ_{1_x} e θ_{1_x} são as matrizes de parâmetros de $\{X_t\}$ que, de acordo com o teorema 7.3.1., é um processo ARMA(1,1). A matriz ϕ_{1_x} é obtida a partir de (7.3.26), o que significa que existem $m(m+1)/2$ e m^2 equações na primeira e na segunda expressões de (7.3.31) respectivamente, o que iguala o número de parâmetros desconhecidos. Por outro lado, a última equação de (7.3.28) é neste caso:

$$V^T \psi_{1_x} = V^T (\phi_{1_x} + \theta_{1_x}) = 0 \quad (7.3.32)$$

uma vez que, para um processo ARMA(1,1), se verifica a relação $\psi_1 = \phi_1 + \theta_1$. A equação (7.3.32) implica que, apesar de as características de ϕ_{1_x} e de θ_{1_x} , poderem assumir quaisquer valores, a sua soma deve ter uma característica não superior ao número de factores. Além disso, supondo que h é um vector próprio da matriz $\phi_{1_x} + \theta_{1_x}$, que o valor próprio associado a h é λ , sendo λ não nulo, e se se verificar a relação:

$$V_u^T (\phi_{1_x} + \theta_{1_x}) h = \lambda V_u^T h = 0 \quad (7.3.33)$$

onde V_u^T é uma qualquer linha de V , então h deve ser ortogonal ao subespaço gerado pela matriz V , pelo que pertence ao subespaço gerado por W .

Em resumo, as matrizes de parâmetros devem verificar as seguintes restrições:

- 1ª) Todas as matrizes auto-regressivas têm r vectores próprios comuns que definem o subespaço dos factores W ;
- 2ª) Todas as matrizes ψ_v devem ter característica igual a r e as suas colunas devem pertencer ao subespaço gerado por W .

Torna-se agora possível determinar uma transformação que permitirá simplificar a estrutura do modelo gerador da série cronológica em análise.

b) Uma transformação simplificadora

É importante determinar uma transformação da série cronológica X_t que permita a simplificação da estrutura subjacente à sua geração. Para essa determinação, é importante notar que, sendo W^{-1} uma matriz inversa generalizada de W , se verifica:

$$Y_t = W^{-1}X_t - W^{-1}\xi_t \quad (7.3.34)$$

Admita-se em seguida que se aplica uma transformação linear a X_t , designada por M , sendo:

$$M = \begin{bmatrix} W^{-1} \\ Q \end{bmatrix} \quad (7.3.35)$$

Então obtém-se um novo vector Z_t , que assume a forma:

$$Z_t = MX_t = \begin{bmatrix} W^{-1}X_t \\ QX_t \end{bmatrix} = \begin{bmatrix} Y_t + W^{-1}\xi_t \\ QWY_t + Q\xi_t \end{bmatrix} = \begin{bmatrix} Z_{1t} \\ Z_{2t} \end{bmatrix} \quad (7.3.36)$$

Para isolar o efeito dos factores, Q deve ser escolhido de modo a que $QW=0$. Assim, as primeiras r componentes de Z_t serão iguais ao vector Y_t adicionado de um vector residual e as outras $m-r$ componentes serão apenas um ruído branco. Uma matriz Q que satisfaça esta condição pode ser obtida colocando como suas linhas os $m-r$ vectores próprios de WW^T associados a valores próprios nulos. Designando esses vectores próprios por V_u , tem-se:

$$WW^T[V_1 \dots V_{m-r}] = 0 = WW^TV \quad (7.3.37)$$

o que implica que $V^T W = 0$, pelo que $Q = V^T$ verifica a propriedade requerida.

Apesar de W^- poder ser arbitrária, parece preferível adoptar uma matriz inversa generalizada que conduza a variáveis transformadas tão simples quanto possível em (7.3.36). Uma propriedade conveniente, tal como indicam os autores referidos, é que Z_{1t} e Z_{2t} sejam independentes, embora tal não seja geralmente possível a não ser que $\Sigma_\xi = I_m$. Admitindo que isso se verifica, tem-se:

$$E(Z_{1t} Z_{2t}^T) = W^- \Sigma_\xi Q^T = W^- Q^T \quad (7.3.38)$$

Se se escolher W^- como a inversa generalizada de W de Moore-Penroe, dada por $(W^T W)^{-1} W^T$, a matriz de variâncias e covariâncias será nula e as duas componentes do vector Z_t serão independentes.

Esta transformação pretende chegar a uma representação mais simples do sistema, tirando partido do facto de o sistema subjacente à geração de X_t ser da forma de (7.3.12). Paralelamente, a determinação do número de factores presentes faz-se pela característica das matrizes de covariância ou de correlação da série observada X_t ou pela das matrizes ψ_v . Por sua vez, a matriz W é determinada tendo em conta que as suas colunas são os vectores próprios das matrizes de covariância de X_t .

(ii) Modelo com factores dependentes

As principais diferenças relativamente ao modelo com factores independentes são:

- 1ª) A matriz $\Gamma_x(k)$ não é simétrica;
- 2ª) Os vectores próprios de $\Gamma_x(k)$ não são colunas de W ;
- 3ª) Os vectores próprios de ϕ_x não são vectores próprios de ϕ_y .

A segunda equação de (7.3.19) e a equação (7.3.25) mantêm-se, conforme indicam os autores desta proposta, embora as conclusões que

daí podem ser extraídas sejam diferentes. Por exemplo, a primeira das equações referidas:

$$\Gamma_x(k) = W\Gamma_y(k)W^T \quad k \geq 1 \quad (7.3.39)$$

Admitindo que $\Gamma_y(k) = D_k \Lambda_k D_k^{-1}$, onde Λ_k é uma matriz diagonal, verifica-se que $\Gamma_x(k)$ possui os mesmos valores próprios que $\Gamma_y(k)$, embora com vectores próprios WD_k e V , pertencendo as colunas desta última matriz ao espaço dos vectores que verificam a relação:

$$WW^T \delta = 0 \quad (7.3.40)$$

ou seja, ao espaço nulo de WW^T . Com efeito, considere-se a equação:

$$\Gamma_x(k) = [WD_k \quad V] \begin{bmatrix} \Lambda_k & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} D_k^{-1}W^T \\ V^T \end{bmatrix} \quad (7.3.41)$$

Se $\Gamma_y(k)$ for diagonal, $D_k = I_m$ e cai-se no caso do modelo com factores independentes.

Da mesma forma, a equação (7.3.25) mantém-se:

$$\phi_{l_x} W = W \phi_{l_y} \quad (7.3.42)$$

Admitindo que ϕ_{l_y} possui vectores próprios independentes e tendo:

$$\phi_{l_y} = L_l S_l L_l^{-1} \quad (7.3.43)$$

onde S_l é uma matriz diagonal e contém os valores próprios de ϕ_{l_y} , então tem-se a equação:

$$\phi_{l_x} W L_l^{-1} = W L_l S_l \quad (7.3.44)$$

o que mostra que os valores próprios de ϕ_{l_y} são também valores próprios de ϕ_{l_x} , estando associados aos vectores próprios $W L_l^{-1}$.

Tendo em vista a obtenção da transformação simplificadora, é necessário determinar o espaço nulo de WW^T . Assim, designando por K a matriz composta pelos vectores próprios de ϕ_x , estes autores indicam que r colunas de K são WL_1 e as outras $m-r$ colunas são arbitrárias. Uma vez que se verifica a relação:

$$(WL_1)(L_1^{-1}W^T) = WW^T \quad (7.3.45)$$

o espaço nulo de $(WL_1)(L_1^{-1}W^T)$ é o mesmo de WW^T . Assim, para seleccionar quais as colunas de K que constituem WL_1 , é necessário ter em consideração que o seu número r é conhecido, uma vez que é a característica de $\Gamma_x(k)$, $P(1)$ e de todas as matrizes ψ_v ; por outro lado, essas colunas são aproximadamente os vectores próprios de $P(1)$ associados a valores próprios não nulos. Então, dispondo de (WL_1) , é possível obter $L_1^{-1}W^T$ a partir de K^{-1} . Paralelamente, designando por V a matriz dos vectores próprios associados aos valores próprios não nulos de $(WL_1)(L_1^{-1}W^T)$, é possível decompor o vector X_t em duas componentes através da transformação:

$$M = \begin{bmatrix} L_1^{-1}W^T \\ V^T \end{bmatrix} \quad (7.3.46)$$

A primeira componente contém os factores adicionados de uma componente residual e a segunda é um ruído branco multivariado, tal caso de factores independentes:

$$Z_t = MX_t = \begin{bmatrix} L_1^{-1}Y_t + L_1^{-1}\xi_t \\ V^T\xi_t \end{bmatrix} = \begin{bmatrix} Z_{1t} \\ Z_{2t} \end{bmatrix} \quad (7.3.47)$$

Uma propriedade importante desta transformação é a de as componentes de Z_{1t} estarem apenas fracamente relacionadas. Por exemplo, tal como

indicam os autores referidos, se $\{Y_t\}$ for um processo AR(1), $\{L^{-1}Y_t\}$ será também um processo AR(1), sendo a sua matriz dos parâmetros auto-regressivos uma matriz diagonal.

Em conclusão, este procedimento propõe uma transformação simplificadora do modelo aplicável a uma série cronológica multivariada, de modo a evitar o inconveniente da obtenção de estruturas extremamente complexas quando se procede à identificação de um modelo ARMA. Com efeito, é possível obter uma representação simples através da utilização de um pequeno conjunto de factores, quando isso for adequado. Paralelamente, ignorar tal representação pode acarretar problemas, pois a interpretação da dinâmica interna que é possível inferir do modelo é completamente diferente nos dois casos. No entanto, o método agora proposto afigura-se muito indirecto e com um elevado grau de complexidade. Além da impossibilidade de determinar com exactidão num caso correcto se se está perante uma situação em que os factores são independentes ou dependentes, o facto de o recurso aos modelos ARMA resultar numa estrutura complexa não significa necessariamente que se está perante um modelo com factores ocultos. Na realidade, a explicação para esse facto deverá muito provavelmente ser encontrada na identificação realizada, que poderá não ter sido correcta, devendo repetir-se essa fase com o auxílio dos instrumentos utilizados na fase de confirmação do diagnóstico.

Assim, o recurso ao método agora analisado deve ser cauteloso, devendo previamente esgotar-se todas as possibilidades de uma correcta modelização através dos modelos ARMA.

3.3. Modelo índice auto-regressivo

Este tipo de modelos visa a identificação de séries cronológicas multivariadas através da consideração de um número mais reduzido de "séries índice", sendo estas seleccionadas de modo a fornecer um resumo tão completo quanto possível da informação contida na série original e necessária para a geração de previsões. Assim, pretende-se conseguir a referida identificação com um reduzido número de parâmetros, considerando para o efeito "modelos índice", que podem também ser úteis para examinar a existência de relações estruturais entre as séries originais. Este tipo de modelos e, em seguida, o procedimento conducente à identificação de séries cronológicas são caracterizados por Reinsel (1983), passando-se à sua análise.

Deste modo, descrever-se-à primeiro o tipo de modelos e, em seguida, o procedimento proposto por este autor para a identificação de séries cronológicas multivariadas:

(i) Caracterização do modelo

Considere-se um modelo auto-regressivo multivariado estacionário:

$$\phi(B)X_t = \varepsilon_t \quad (7.3.48)$$

com $\phi(B) = I_m - \phi_1 B - \dots - \phi_p B^p$, admitindo-se que os vectores ε_t são independentes com distribuição $N(0; \Sigma)$.

Tendo em vista reduzir o número de parâmetros incluídos em $\phi(B)$, considera-se uma classe de modelos em que $I_m - \phi(B) = \sum_{i=1}^p \phi_i B^i$ pode ser decomposto da forma:

$$I_m - \phi(B) = F(B)H(B) \quad (7.3.49)$$

sendo $F(B) = F_1 B + \dots + F_{p_1} B^{p_1}$ e $H(B) = H_0 + H_1 B + \dots + H_{p_2} B^{p_2}$ matrizes de dimensão $(m \times r)$ e $(r \times m)$ respectivamente, com $r \leq m$ e $p_1 + p_2 = p$. Assim, o modelo (7.3.48) pode ser expresso da forma:

$$X_t = F(B)H(B)X_t + \varepsilon_t = \sum_{u=1}^{p_1} \sum_{v=0}^{p_2} F_u H_v X_{t-u-v} + \varepsilon_t \quad (7.3.50)$$

Neste contexto as "variáveis índice" são:

$$Y_t = H(B)X_t = H_0 X_t + H_1 X_{t-1} + \dots + H_{p_2} X_{t-p_2} \quad (7.3.51)$$

sendo Y_t um vector de dimensão ($r \times 1$). Estas variáveis têm a interpretação de que os p_1 valores passados $Y_{t-1}, \dots, Y_{t-p_1}$ fornecem um resumo completo de toda a informação passada necessária à previsão do vector X_t , isto é, (7.3.50) pode ser representado como:

$$X_t = F(B)Y_t + \varepsilon_t = \sum_{u=1}^{p_1} F_u Y_{t-u} + \varepsilon_t \quad (7.3.52)$$

onde as matrizes F_u ($u=1, \dots, p_1$) têm dimensão ($m \times r$). Deste modo, conclui-se que um dos principais interesses dos modelos índice consiste em analisar até que ponto a série X_t pode ser "explicada" ou prevista pelos valores passados de uma outra série com menor número de componentes Y_t . Uma vez que se está a considerar situações em que r pode ser muito inferior a m , as variáveis índice Y_t representam uma apreciável redução relativamente ao número de componentes de X_t e, conseqüentemente, o modelo índice (7.3.50) pode significar uma redução considerável do número de parâmetros em relação ao modelo não restringido (7.3.48).

A estrutura do modelo auto-regressivo (7.3.50) pode ser especificada em termos das ordens p_1, p_2 e de r , designando-se simbolicamente tal modelo por $AR_I(p_1, p_2, r)$.

Por outro lado, tendo em vista assegurar a identificabilidade do modelo (7.3.50), é útil considerar $H_v = [H_{v_1} \ H_{v_2}]$, com $v=0, 1, \dots, p_2$, onde o bloco H_{v_1} tem dimensão ($r \times r$) e se admite que a característica de H_0 e de H_{p_2} é r . Assim, impõe-se a condição $H_{0_1} = I_r$, o que permite assegurar a identificabilidade do modelo.

Convém ainda referir que o modelo (7.3.50) pode ser estimado pelo método da máxima verosimilhança através das técnicas habituais. Para o efeito, é útil exprimir (7.3.50) sob a forma:

$$\mathbf{X}_t = \mathbf{F}\mathbf{Z}_{t-1} + \boldsymbol{\varepsilon}_t \quad (7.3.53)$$

onde $\mathbf{F} = [\mathbf{F}_1 \dots \mathbf{F}_{p_1}]$, sendo as matrizes \mathbf{F}_u de dimensão $(m \times r)$, com $r \leq m$ ($u=1, \dots, p_1$) e:

$$\mathbf{Z}_{t-1}^T = (\mathbf{Y}_{t-1}^T, \dots, \mathbf{Y}_{t-p_1}^T) = (\mathbf{X}_{t-1}^T, \dots, \mathbf{X}_{t-p}^T) \sum_{v=0}^{p_2} (\mathbf{E}_v \otimes \mathbf{H}_v^T) \quad (7.3.54)$$

sendo \mathbf{H}_v matrizes de dimensão $(r \times m)$, conforme já foi definido, e \mathbf{E}_v matrizes diagonais de dimensão $(p \times p_1)$ com os elementos da diagonal principal iguais à unidade ($v=0, 1, \dots, p_2$). O autor referido prova que os estimadores de máxima verosimilhança assim obtidos possuem as propriedades assintóticas desejáveis, nomeadamente a consistência e a normalidade assintótica.

Nestas condições, é possível particularizar o modelo (7.3.50), assumindo particular interesse o caso em que $p_2=0$, $p_1=p$, o que conduz a:

$$\mathbf{X}_t = \sum_{l=1}^{p_1} \mathbf{F}_l \mathbf{H}_0 \mathbf{X}_{t-l} + \boldsymbol{\varepsilon}_t = \mathbf{F}\mathbf{Z}_{t-1} + \boldsymbol{\varepsilon}_t \quad (7.3.55)$$

onde $\mathbf{F} = [\mathbf{F}_1 \dots \mathbf{F}_p]$, sendo as matrizes \mathbf{F}_l ($l=1, \dots, p$) e \mathbf{H}_0 de dimensão $(m \times r)$ e $(r \times m)$ respectivamente, admitindo-se que a característica de \mathbf{H}_0 é r ($r \leq m$); por outro lado, tem-se:

$$\mathbf{Z}_{t-1}^T = (\mathbf{X}_{t-1}^T \mathbf{H}_0^T, \dots, \mathbf{X}_{t-p}^T \mathbf{H}_0^T) = (\mathbf{X}_{t-1}^T, \dots, \mathbf{X}_{t-p}^T) (\mathbf{I}_p \otimes \mathbf{H}_0^T) \quad (7.3.56)$$

Considere-se ainda $\mathbf{H}_0 = [\mathbf{H}_1 \mathbf{H}_2]$, sendo \mathbf{H}_1 de dimensão $(r \times r)$. Com o objectivo de assegurar a identificabilidade do modelo (7.3.55), impõe-se a condição $\mathbf{H}_1 = \mathbf{I}_r$, do que resulta $\mathbf{H}_0 = [\mathbf{I}_r \mathbf{H}_2]$.

O número de parâmetros incluídos no operador auto-regressivo de (7.3.55) é de $m^2p + r(m-r) = m^2p - (m-r)(mp-r)$, o que é bastante inferior ao número de parâmetros do operador do modelo não restringido (7.3.48), que é de m^2p . O vector $Y_t = H_0 X_t$, de dimensão $(r \times 1)$, pode ser encarado como um vector de índices que comandam a série cronológica X_t , podendo assim possuir alguma interpretação física.

O modelo (7.3.55) pode também ser estimado com recurso ao método da máxima verosimilhança, tal como sucedia com (7.3.50), possuindo os estimadores resultantes as mesmas propriedades assintóticas.

Finalmente, o autor deste método indica ainda um outro caso particular de (7.3.50) que pode ter algum interesse e que resulta de ter $p_1=1, p_2=p-p_1$, o que conduz a:

$$X_t = \sum_{i=0}^{p-1} F_1 H_i X_{t-1-i} + \varepsilon_t = F_1 \sum_{i=0}^{p-1} H_i X_{t-1-i} + \varepsilon_t \quad (7.3.57)$$

No entanto, este autor não lhe confere importância digna de nota.

Neste contexto, é possível agora delinear o método de identificação de uma série cronológica proposto.

(ii) Identificação do modelo

Com o objectivo de seleccionar o modelo adequado a uma série cronológica, o autor referido indica previamente a forma do previsor a um passo para o modelo (7.3.55):

$$\hat{X}_t(1) = \hat{F}(I_p \otimes \hat{H}_0) W_t \quad (7.3.58)$$

onde $W_t^T = (X_t^T, \dots, X_{t+1-p}^T)$. Conferindo particular atenção ao modelo (7.3.55), ou seja, ao caso particular de (7.3.50) em que $p_2=0, p_1=p$, este autor propõe um método de identificação apenas para este caso. Assim, um estimador consistente para a matriz do erro quadrático médio de

$\hat{X}_t(1)$, ou seja, para $E[(X_{t+1} - \hat{X}_t(1))(X_{t+1} - \hat{X}_t(1))^T]$ baseado nas observações disponíveis é:

$$\hat{\Sigma} + \frac{1}{N} \sum_{t=1}^N \hat{V}_t [\hat{V}_t^T (I_N \otimes \hat{\Sigma}^{-1}) \hat{V}_t]^{-1} V_t^T \quad (7.3.59)$$

onde:

$$V_t = \left[\left\{ I_m \otimes W_{t-1}^T (I_p \otimes H_0) \right\} \sum_{i=1}^p (F_i \otimes X_{2,t-i}^T) \right]$$

$$V^T = (V_1^T, \dots, V_N^T) \quad (7.3.60)$$

sendo \hat{V}_t resultante da substituição em (7.3.60) das matrizes H_0 e F_i ($i=1, \dots, p$) pelos seus estimadores de máxima verosimilhança; por outro lado, X_{2t} resulta de particionar X_t da mesma forma que H_0 , ou seja, de ter $X_t^T = (X_{1t}^T, X_{2t}^T)$, sendo X_{1t} de dimensão $(r \times r)$.

Deste modo, a utilização de alguma medida que resuma (7.3.59) permite a selecção da ordem e da característica do modelo índice (7.3.55), de modo semelhante ao procedimento aplicado a modelos auto-regressivos não restringidos ($r=m$), como é o caso dos critérios de estimação da ordem do modelo atrás analisados.

Assim, o autor deste método propõe que se considere, para cada ordem do operador auto-regressivo, um valor mínimo da característica do modelo índice, ou seja, da característica desse operador, sendo este valor aumentado até atingir m . Para cada um dos modelos resultantes, o valor do critério utilizado é então calculado. A ordem do operador auto-regressivo é em seguida aumentada e o procedimento repetido até se atingir um valor dessa ordem considerado suficientemente elevado. Visando reduzir o peso computacional, o autor referido indica ainda que é possível seleccionar um modelo não restringido através dos critérios atrás

analisados e considerar apenas para essa ordem o ajustamento de modelos índice de característica inferior a m . Em cada uma das etapas, é útil comparar os resultados obtidos na estimação de um modelo índice com os da estimação de um modelo não restringido. Paralelamente, é ainda vantajoso, sobretudo para séries com um elevado número de componentes em que o modelo auto-regressivo não restringido inclui um elevado número de parâmetros, comparar os resultados de um modelo índice de baixa característica, com os de um modelo índice de característica mais elevada.

Estas comparações e a escolha final do modelo, ou seja, a identificação da série, podem ser feitas directamente com base numa qualquer função, como é o caso do determinante ou do traço da matriz (7.3.59), ou em qualquer critério de estimação da ordem de modelos não restringidos. O modelo seleccionado deve ser aquele para o qual o critério utilizado atinge o seu valor mínimo, tratando-se portanto de um método idêntico ao estabelecido para modelos não restringidos.

Finalmente, este autor refere ainda que este procedimento também é útil para o caso do modelo mais geral (7.3.50). Por outro lado, afirma que o procedimento acabado de analisar pode também ser aplicado ao vector dos resíduos obtidos a partir da prévia conversão em ruído branco das séries originais através do ajustamento de modelos auto-regressivos univariados de baixa ordem.

No entanto, apesar de pretender conseguir a identificação de uma estrutura com um reduzido número de parâmetros, o método agora analisado incorre no risco de fracassar no seu objectivo, pois a consideração apenas de modelos auto-regressivos pode ser inadequada para uma dada série cronológica. Isso verifica-se, por exemplo, no caso em que o modelo mais adequado seja um ARMA. Esta falta de adequação

pode resultar na obtenção de uma estrutura sobreparametrizada, com um número excessivo de parâmetros, o que contraria o objectivo do método. Por outro lado, coloca-se também o problema da truncagem do procedimento, isto é, a definição do valor máximo para a ordem do modelo índice que permite terminar o procedimento proposto. Nada garante que o valor máximo seleccionado seja o mais adequado. Finalmente, a utilização de alguns critérios de estimação da ordem do modelo pode não ser a mais adequada, como é caso do critério AIC, uma vez que, conforme já foi analisado, este critério não é consistente.

3.4. Modelo auto-regressivo de polinómio amortecido

Este tipo de modelo pretende evitar, por um lado, o recurso a modelos ARMA, uma vez que tal recurso pode não ser exequível e, por outro lado, a obtenção de estruturas sobreparametrizadas a que o recurso a modelos auto-regressivos, embora mais acessível, poderá conduzir.

Assim, o método agora proposto pode ser encarado como um compromisso entre o recurso a modelos AR e a modelos ARMA de que resulta a utilização de uma sub-classe de modelos ARMA de carácter mais genérico do que os modelos AR, sem no entanto, ser de aplicação complexa.

Neste contexto, e com o duplo objectivo agora referido, Lütkepohl (1982b) propõe o recurso a modelos auto-regressivos de polinómio amortecido para a identificação de séries cronológicas multivariadas. Deste modo, após a caracterização genérica do modelo, passar-se-á à análise da sua estimação, finalizando-se com o método proposto para a especificação do modelo:

(i) Caracterização do modelo

Seja $\{X_t\}$ um processo estocástico que pode ser representado como um processo auto-regressivo de ordem infinita:

$$\pi(B) X_t = \varepsilon_t \quad (7.3.61)$$

onde $\{\varepsilon_t\}$ é um ruído branco com matriz de variâncias e covariâncias não singular Σ e $\pi(B) = \sum_{l=0}^{\infty} \pi_l B^l$ é uma matriz de dimensão $(m \times m)$ cujo elemento genérico é:

$$\pi_{ij}(B) = \sum_{l=0}^{\infty} \pi_{ij,l} B^l \quad (i,j=1, \dots, m) \quad (7.3.62)$$

admitindo-se também que:

$$\sum_{l=0}^{\infty} |\pi_{ij,l}| < \infty \quad (i,j=1, \dots, m) \quad (7.3.63)$$

É ainda necessário que as raízes da equação matricial:

$$|\pi(B)| = 0 \quad (7.3.64)$$

estejam todas fora do círculo unitário. Note-se que a representação (7.3.61) é ligeiramente diferente da que foi definida atrás em (2.1.40), sendo mais conveniente para a caracterização do tipo de modelos que vai ser apresentado.

Com o objectivo de assegurar a identifiabilidade do modelo (7.3.61) o autor deste método admite que π_0 é triangular superior e que Σ é uma matriz diagonal.

Uma vez que o número infinito de parâmetros incluídos em (7.3.61), não pode ser estimado a partir de um número finito de observações, é necessário impor alguma restrição que limite a dimensão do espaço dos parâmetros. Para o efeito, é possível admitir que $\pi_{i,j,l} = 0$ $(i,j=1, \dots, m)$ para $l > l_0$, sendo l_0 um valor considerado adequado ou, alternativamente, aproximar o polinómio matricial de ordem infinita $\pi(B)$ através de $\theta(B)^{-1} \phi(B)$, sendo $\theta(B)$ e $\phi(B)$ polinómios matriciais de

ordem finita. Como é evidente, esta segunda opção equivale à consideração de modelos ARMA.

Pretendendo evitar os inconvenientes acarretados por ambas as alternativas, este autor sugere a consideração de processos auto-regressivos com matrizes de parâmetros da forma:

$$\pi_t = \tau' H(l) \quad (7.3.65)$$

onde τ é um número arbitrário e possivelmente fixo, assumindo um valor absoluto entre 0 e 1 e $H(l)$ é uma matriz de dimensão $(m \times m)$ com elemento genérico:

$$h_{ij}(l) = \sum_{n=0}^{\lambda_{ij}} h_{ij,n} l^n \quad (i,j=1,\dots,m) \quad (7.3.66)$$

É vantajoso que os polinómios $h_{ij}(l)$ tenham graus diferentes, de modo a possibilitar a redução do número de parâmetros.

Nestas condições, um processo auto-regressivo com matrizes de parâmetros da forma (7.3.65) é designado por processo auto-regressivo de polinómio amortecido e representa-se por $AP(\lambda)$, sendo λ o valor máximo de λ_{ij} ($i,j=1,\dots,m$).

Sendo $0 < |\tau| < 1$ e $H(l)$ uma matriz de polinómios cujo máximo grau é λ , é possível mostrar que existe uma outra matriz de polinómios com o mesmo grau máximo que verifica:

$$\sum_{l=0}^{\infty} \tau^l H(l) B^l = Q(B) / (1 - \tau B)^{\lambda+1} \quad (7.3.67)$$

Deste modo, conclui-se que os modelos AP são uma sub-classe dos modelos ARMA e que, para um valor de τ próximo de zero, são aproximadamente modelos auto-regressivos de ordem finita, podendo então constituir uma aproximação a um qualquer modelo ARMA. Na realidade, isso verifica-se não só para τ próximo de zero, mas para qualquer valor de τ entre zero e um, tendo-se o teorema:

Teorema 7.3.3. Para uma qualquer sequência de matrizes π_l , de dimensão $(m \times m)$ e com elemento genérico $\pi_{ij,l}$ que verifica a condição (7.3.63), para qualquer $0 < \tau < 1$ e qualquer $\delta > 0$, existe uma matriz de polinómios $H(l)$ com elementos $h_{ij}(l)$ que verifica:

$$\sum_{i,j=1}^m \sum_{l=1}^{\infty} |\pi_{ij,l} \tau^l h_{ij}(l)| < \delta \quad H(0) = \pi_0$$

Este resultado significa que, se um dado processo admitir uma representação do tipo de (7.3.61), então existem processos AP com parâmetros auto-regressivos arbitrariamente próximos dos parâmetros auto-regressivos, possivelmente em número infinito, do processo dado, não sendo para isso necessária a verificação da condição (7.3.64), nem da condição que impõe que π_0 seja triangular superior e Σ diagonal.

Apesar de ter sido utilizado apenas um factor de amortecimento τ na formulação do modelo AP e no teorema 7.3.3., a expressão (7.3.67) permite concluir que os graus λ_{ij} dos polinómios $h_{ij}(l)$ necessários para uma boa aproximação da estrutura subjacente dependem de τ . Consequentemente, para determinar uma aproximação parcimoniosa pode ser conveniente utilizar um factor de amortecimento diferente para cada polinómio $h_{ij}(l)$, $i, j = 1, \dots, m$. Isso equivaleria a considerar processos auto-regressivos cujas matrizes π_l têm elementos:

$$\tau_{ij}^l h_{ij}(l) \quad i, j = 1, \dots, m \quad (7.3.68)$$

Visando aumentar a flexibilidade do modelo, seria ainda possível tratar τ_{ij} como parâmetros desconhecidos, com $0 < |\tau_{ij}| < 1$, em vez de serem considerados constantes fixas. No entanto, este autor prefere considerar τ fixo para não introduzir complicações desnecessárias no modelo.

Caracterizado o modelo, vai-se analisar agora a sua estimação.

(ii) Estimação

Admitindo que os valores λ_{ij} ($i, j = 1, \dots, m$) são conhecidos e que τ é fixo, sendo $0 < |\tau| < 1$, é possível estimar os parâmetros do modelo AP.

Assim, a hipótese de π_0 ser triangular e de Σ ser diagonal, permite estimar cada equação separadamente, produzindo estimadores assintoticamente eficientes se $\{\epsilon_t\}$ for um processo Gaussiano. Para o efeito, exprime-se a i -ésima equação da forma:

$$\begin{aligned} X_{it} &= - \sum_{n=i+1}^m h_{in}(0) X_{nt} - \sum_{n=1}^m \sum_{l=1}^m \tau^l h_{in}(l) X_{n,t-l} + \eta_{it} = \\ &= - \sum_{n=i+1}^m h_{in,0} X_{nt} + \sum_{n=1}^m \sum_{j=0}^{\lambda_n} h_{in,j} g_{nj,t} - \tau^t \sum_{j=0}^{\lambda_i} u_{ij} t^j + \eta_{it} \end{aligned} \quad (7.3.69)$$

onde:

$$g_{nj,t} = - \sum_{l=1}^{t-1} \tau^l h_{nj,l} X_{n,t-l} \quad \lambda_i = \max(\lambda_{in}, n=1, \dots, m) \quad (7.3.70)$$

sendo as variáveis u_{ij} independentes de t . A partir de (7.3.69), obtém-se a equação de regressão:

$$X_{it} - g_{i0,t} = \sum_{n=i+1}^m h_{in,0} (g_{n0,t} - X_{nt}) + \sum_{n=1}^m \sum_{j=1}^{\lambda_n} h_{in,j} g_{nj,t} - v_{it} + \eta_{it} \quad (7.3.71)$$

onde;

$$v_{it} = \sum_{j=0}^{\lambda_i} u_{ij} \tau^t t^j \quad (7.3.72)$$

Deste modo, a estimação pode ser efectuada como se se tratasse de um modelo de regressão linear.

Não é possível estimar consistentemente os valores iniciais de u_{ij} ($j=1, \dots, \lambda_i$), uma vez que as variáveis v_{it} tendem rapidamente para zero quando t tende para infinito, pelo que a substituição desses elementos por zero em (7.3.71) não afecta as propriedades dos estimadores dos parâmetros

$h_{in,j}$, cuja estimação se pretende. No entanto, em pequenas amostras, a inclusão ou não dos valores iniciais produz efeitos não negligíveis. Uma vez que o impacto de v_{it} é relativamente importante para pequenos valores de t , é conveniente não utilizar as primeiras observações, no caso de os valores iniciais terem sido desprezados, sendo então essas observações utilizadas apenas para o cálculo de $g_{nj,t}$. Por sua vez, a matriz de variâncias e covariâncias dos estimadores dos parâmetros pode ser estimada da forma habitual a partir da regressão ajustada. Por outro lado, os parâmetros auto-regressivos podem também ser facilmente estimados, uma vez que estão linearmente relacionados com $h_{in,j}$ quando τ é fixo. Se $E(X_t) \neq 0$, é ainda possível incluir um termo independente em (7.3.31)

Conforme foi referido atrás, a estimação necessita da prévia especificação do modelo, ou seja, tem de ser precedida pela identificação que irá ser agora analisada.

(ii) Identificação

A identificação do modelo engloba não só a selecção de valores adequados para λ_{ij} ($i,j=1,\dots,m$), mas também a selecção do factor de amortecimento τ . Procede-se inicialmente à análise da selecção dos primeiros e, em seguida, à da determinação do segundo.

No entanto, uma vez que a inclusão dos valores iniciais na equação de regressão (7.3.71) aumenta o peso computacional, o autor desta proposta sugere a substituição de v_{it} por zero durante a fase de identificação, se o número de observações for moderado ou elevado. A posterior inclusão dos valores iniciais para a estimação final pode revelar-se útil para um número moderado de observações.

a) Selecção dos graus dos polinómios

Principiando com o caso univariado, é útil começar por referir que só λ_{11} tem que ser especificado, pelo que o grau do único polinómio existente será designado por λ apenas por razões de simplificação. Uma

vez que o modelo foi transformado num modelo de regressão linear, a alteração de λ resulta na alteração do número de regressores. Logo, a selecção do valor de λ pode ser realizada através da utilização de um critério de selecção de variáveis no âmbito de modelos de regressão linear múltipla.

Nesta fase, é conveniente utilizar um critério simples para a selecção de λ , uma vez que o modelo identificado será posteriormente sujeito a confirmação, o que leva o autor referido a propor a escolha prévia de um valor máximo M para o grau do polinómio e a estimação posterior de todos os modelos, com λ variando de 1 a M . Tendo em vista a selecção do valor óptimo para λ , procede-se à determinação de intervalos de confiança para os parâmetros, sendo aquele valor escolhido como o maior grau para o qual o intervalo de confiança de $h_{11,\lambda}$ não inclui zero. Um outro critério possível seria seleccionar o valor λ que conduza ao modelo com variância residual mínima. Finalmente, este autor sugere ainda o recurso a um critério de estimação da ordem de modelos univariados, que constituem particularizações dos critérios atrás analisados para o caso multivariado.

No que respeita ao caso multivariado, é necessário seleccionar os valores de λ_{ij} ($i,j=1,\dots,m$). O procedimento proposto para o efeito pelo autor deste método será exemplificado para o caso trivariado ($m=3$) e para a primeira equação apenas, permitindo a imediata percepção do tratamento a adoptar para as outras equações, uma vez que é idêntico, bem como a generalização para valores mais elevados de m , o que é imediato.

Assim, admita-se que se escolheu um valor máximo M para λ_{11} , λ_{12} e λ_{13} . O procedimento desenvolve-se de acordo com os seguintes passos:

1ª) Selecciona-se um valor óptimo para λ_{11} entre 0 e M , designado por $\lambda_{11}^{(1)}$, condicionado por $h_{12}(l) \equiv h_{13}(l) \equiv 0$, onde (1) indica que se trata da

primeira iteração. Em cada passo, pode ser utilizado um dos critérios de selecção propostos para o caso univariado.

2º) Selecciona-se um valor óptimo para λ_{12} , designado por $\lambda_{11}^{(2)}$, condicionado por $\lambda_{11}=\lambda_{11}^{(1)}$ e $h_{13}(l)\equiv 0$. Em seguida, procede-se à escolha de um valor óptimo para λ_{11} entre 0 e M, designado por $\lambda_{11}^{(2)}$, condicionado por $\lambda_{12}=\lambda_{12}^{(2)}$ e $h_{13}(l)\equiv 0$. Alternativamente é possível utilizar $h_{12}(l)\equiv 0$. A título de exemplo, se a significância estatística dos parâmetros estimados for o critério de selecção utilizado, toma-se $h_{12}(l)\equiv 0$ se os parâmetros estimados deste polinómio forem não significativos para todos os valores de λ_{12} entre 0 e M. Nesse caso, a segunda variável seria eliminada no passo seguinte.

3º) Determina-se um valor óptimo $\lambda_{13}=\lambda_{13}^{(3)}$ entre 0 e M, condicionado por $\lambda_{11}=\lambda_{11}^{(2)}$ e $\lambda_{12}=\lambda_{12}^{(2)}$. Em seguida, selecciona-se um valor óptimo para λ_{11} , designado por $\lambda_{11}^{(3)}$, condicionado por $\lambda_{12}=\lambda_{12}^{(2)}$, $\lambda_{13}=\lambda_{13}^{(3)}$ e, finalmente, selecciona-se um valor óptimo $\lambda_{12}=\lambda_{12}^{(3)}$, condicionado por $\lambda_{11}=\lambda_{11}^{(3)}$, $\lambda_{13}=\lambda_{13}^{(3)}$.

A ordem pela qual as variáveis são adicionadas pode ser trocada, isto é, é possível escolher no 2º passo $\lambda_{13}^{(3)}$ em vez de $\lambda_{12}^{(2)}$ e assim por diante.

Uma vez que se admite que a matriz π_0 é triangular superior, os termos de ordem nula nos polinómios $h_{ij}(l)$, com $i,j=1,\dots,m$ e $i<j$, correspondem às correlações contemporâneas entre as componentes do processo multivariado. No entanto, estes elementos merecem um tratamento especial, pelo que o autor deste método propõe a sua eliminação se as estimativas correspondentes não forem significativas, o que pode ser averiguado através de um teste t.

Assim, a identificação dos graus dos diferentes polinómios $h_{ij}(l)$ fica concluída, faltando apenas a selecção do factor de amortecimento τ , o que constitui a fase seguinte do método.

b) Selecção do factor de amortecimento

Conforme já foi anteriormente mencionado, a escolha do factor de amortecimento τ pode ser útil para a redução do número de parâmetros do modelo, uma vez que os graus adequados λ_{ij} ($i, j=1, \dots, m$) dependem de τ . Com efeito, é possível utilizar um factor diferente para cada polinómio, ou seja:

$$\pi_{ij}(B) = \sum_{l=0}^{\lambda_{ij}} \tau_{ij}^l h_{ij}(l) B^l \quad i, j=1, \dots, m \quad (7.3.73)$$

Neste caso, se v_{it} forem utilizados em (7.3.71), devem ser substituídas por:

$$\bar{v}_i = \sum_{n=1}^m \sum_{j=0}^{\lambda_n} u_{in,j} \tau_{in}^j t^j \quad (7.3.74)$$

onde $u_{in,j}$ são independentes de t . Se o factor de amortecimento diferir apenas para equações diferentes, já não é necessário substituir v_{it} por \bar{v}_i .

É evidente que o peso computacional aumenta quando se experimenta diversos valores para o factor de amortecimento, uma vez que, para cada diferente combinação, é necessário percorrer todo o procedimento de identificação. Deste modo, o autor referido sugere que uma boa estratégia seria utilizar um pequeno valor de τ , comum para todos os elementos, propondo $\tau=0.1$ numa primeira tentativa de identificação, mantendo-se esse valor se os graus dos diferentes polinómios $h_{ij}(l)$ forem baixos. Se tal não acontecer, é aconselhável fazer variar τ em pequenos passos nas equações com elevados graus dos polinómios.

Na fase de confirmação do diagnóstico, este autor propõe, além dos critérios habituais, o recurso a testes da razão de verosimilhança para averiguar a eventual significância de grupos adicionais de parâmetros, o que pode ser útil para a detecção de falhas do modelo.

A principal crítica a que este método está sujeito é a de correr o risco de fracassar no seu duplo objectivo. Com efeito, é muito provável que a estrutura obtida seja sobreparametrizada, quer por partir da representação de um processo estocástico multivariado como um processo auto-regressivo de ordem infinita, quer por considerar os polinómios $h_{ij}(l)$, cujos graus podem ser bastante elevados. O primeiro problema agrava-se quando se utiliza diferentes τ_{ij} para cada elemento ($i,j=1,\dots,m$). Por outro lado, o objectivo de adopção de um método simples de identificação também não parece ser totalmente atingido, uma vez que o método proposto se revela um pouco complexo e pouco directo, não primando pela acessibilidade. Por isso, parece ser preferível o recurso à classe mais geral de modelos ARMA, em vez de utilizar a sub-classe de modelos auto-regressivos de polinómio amortecido.

3.5. Análise canónica

A abordagem que irá agora ser analisada propõe uma transformação canónica de um processo auto-regressivo estacionário, ordenando as componentes do processo transformado conforme a sua previsibilidade. As componentes menos previsíveis estão frequentemente próximas de serem ruídos brancos e podem reflectir relações contemporâneas entre as variáveis originais. Por sua vez, as componentes mais previsíveis podem ser quase não estacionárias, representando o crescimento dinâmico de uma série cronológica multivariada.

Trata-se de uma análise que pretende fornecer novas indicações sobre uma série cronológica multivariada, procurando fazer luz sobre a sua dinâmica interna. Tal como indicam Tiao e Box (1979, 1981), este tipo de análise tem um duplo objectivo: por um lado, permitir a detecção de relações lineares entre as componentes da série multivariada e, por outro, auxiliar na interpretação e compreensão de um modelo ajustado.

Assim, passar-se-à agora a tratar das diversas facetas que compõem esta análise, proposta por Box e Tiao (1977).

Neste contexto, considere-se o modelo auto-regressivo de ordem p multivariado estacionário:

$$\mathbf{X}_t = \hat{\mathbf{X}}_{t-1}(1) + \boldsymbol{\varepsilon}_t \quad (7.3.75)$$

onde $\hat{\mathbf{X}}_{t-1}(1)$ é o valor esperado de \mathbf{X}_t condicionado pelos seus valores passados até ao momento $t-1$, ou seja:

$$\hat{\mathbf{X}}_{t-1}(1) = E(\mathbf{X}_t | \mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \dots) = \sum_{l=1}^p \pi_l \mathbf{X}_{t-l} \quad (7.3.76)$$

sendo $\pi_l (l=1, \dots, p)$ matrizes de dimensão $(m \times m)$ e $\{\boldsymbol{\varepsilon}_t\}$ um ruído branco multivariado com matriz de variâncias e covariâncias Σ e independente de $\hat{\mathbf{X}}_{t-1}(1)$.

Suponha-se em primeiro lugar que $m=1$, ou seja, que se está no caso univariado. Se o processo for estacionário, verifica-se:

$$E(\mathbf{X}_t^2) = E[\hat{\mathbf{X}}_{t-1}(1)]^2 + E(\boldsymbol{\varepsilon}_t^2) \quad (7.3.77)$$

ou seja:

$$\sigma_x^2 = \sigma_{\hat{x}}^2 + \sigma_{\varepsilon}^2 \quad (7.3.78)$$

onde $\sigma_x^2 = V(\mathbf{X}_t)$, $\sigma_{\hat{x}}^2 = V[\hat{\mathbf{X}}_{t-1}(1)]$ e $\sigma_{\varepsilon}^2 = V(\boldsymbol{\varepsilon}_t)$

É então possível definir uma medida da previsibilidade de um processo a partir do seu próprio passado, designado por λ e assumindo a forma:

$$\lambda = \sigma_{\dot{x}}^2 \sigma_x^{-2} = 1 - \sigma_e^2 \sigma_x^{-2} \quad (7.3.79)$$

Suponha-se que se pretende analisar diferentes séries cronológicas univariadas que estão relacionadas e que revelam crescimentos dinâmicos. É possível admitir que essas séries possam ser representadas como o agregado de alguns "inputs" comuns, dos quais alguns podem estar próximos de ser não estacionários e outros podem ser estacionários ou ruídos brancos. Isto conduz a considerar relações lineares da forma $U_t = V^T X_t$, sendo V um vector de dimensão $(m \times 1)$. Os elementos U_t que dependem mais fortemente do seu passado, ou seja, que têm um valor de λ elevado, podem ser úteis como indicadores compostos do crescimento global das diferentes séries. Opostamente, os elementos possuindo um valor de λ próximo de zero podem reflectir relações contemporâneas estáveis entre as séries originais.

A análise agora proposta determina m componentes canónicas ordenadas por ordem crescente de previsibilidade. As componentes mais previsíveis estarão frequentemente próximas de serem não estacionárias e as menos previsíveis serão estacionárias ou independentes. É possível decompor o espaço de dimensão m da observação X_t em subespaços independentes, estacionários e não estacionários. As variáveis que pertencem ao espaço não estacionário representam crescimento dinâmico, enquanto as incluídas no espaço estacionário e no independente podem reflectir relações que permanecem estáveis no tempo.

Feito este preâmbulo, é possível agora caracterizar a análise canónica:

(i) Selecção das variáveis canónicas

Admita-se que $\{X_t\}$ é um processo estacionário. Designando por $\Gamma_x(k)$ a matriz de covariância de $\{X_t\}$, tem-se:

$$\Gamma_x(0) = \sum_{l=1}^p \pi_l \Gamma_x(l) + \Sigma = \Gamma_x(0) + \Sigma \quad (7.3.80)$$

onde $\Gamma_x(0)$ representa a matriz de variâncias e covariâncias de $\hat{X}_{t-1}(1)$. Admite-se ainda que Σ e, portanto, $\Gamma_x(0)$ são definidas positivas.

Considere-se agora a combinação linear $U_t = V^T X_t$, o que leva à equação:

$$U_t = \hat{U}_{t-1}(1) + \eta_t \quad (7.3.81)$$

onde $\hat{U}_{t-1}(1) = V^T \hat{X}_{t-1}(1)$ e $\eta_t = V^T \varepsilon_t$. A previsibilidade de U_t a partir do seu passado é então medida por:

$$\lambda = \sigma_u^2 \sigma_u^{-2} = \{V^T \Gamma_x(0) V\} / \{V^T \Gamma_x(0) V\} \quad (7.3.82)$$

Daqui se conclui que, para a máxima previsibilidade, λ deve ser o maior valor próprio de $\Gamma_x(0)^{-1} \Gamma_x(0)$ e V o respectivo vector próprio. De modo semelhante, o vector próprio associado ao menor valor próprio conduz à combinação de X_t menos previsível.

É possível agora determinar a transformação canónica. Assim, sejam $\lambda_1, \dots, \lambda_m$ os m valores próprios reais da matriz $\Gamma_x(0)^{-1} \Gamma_x(0)$, ordenados por ordem crescente, ou seja, $\lambda_1 < \lambda_2 < \dots < \lambda_m$. Suponha-se também que os vectores próprios linearmente independentes que lhes estão associados, designados por h_1^T, \dots, h_m^T , formam as m linhas de uma matriz H . É possível então construir um processo transformado $\{Y_t\}$ que verifica:

$$Y_t = \hat{Y}_{t-1}(1) + \xi_t \quad (7.3.83)$$

onde $Y_t = H X_t$, $\xi_t = H \varepsilon_t$, $\hat{Y}_{t-1}(1) = \sum_{i=1}^p \tilde{\pi}_i Y_{t-i}$ e $\tilde{\pi}_i = H \pi_i H^{-1}$. A relação correspondente a (7.3.80) é:

$$\Gamma_y(0) = \Gamma_x(0) + \tilde{\Sigma} \quad (7.3.84)$$

onde $\Gamma_y(0) = H \Gamma_x(0) H^T$, $\Gamma_x(0) = H \Gamma_x(0) H^T$ e $\tilde{\Sigma} = H \Sigma H^T$.

É importante constatar que:

$$H^{T-1} \Gamma_x(0)^{-1} \Gamma_x(0) H^T = \Lambda, \quad H^{T-1} \Gamma_x(0)^{-1} \Sigma H^T = I_m - \Lambda \quad (7.3.85)$$

onde Λ é uma matriz diagonal de dimensão $(m \times m)$ cujos elementos da diagonal principal são $\lambda_1, \dots, \lambda_m$. Por outro lado, verifica-se que $0 \leq \lambda_i < 1 (i=1, \dots, m)$ e que, para $i \neq j$, $h_i^T \Sigma h_j = h_i^T \Gamma_x(0) h_j = 0$. Por outras palavras, $H \Sigma H^T$, $H \Gamma_x(0) H^T$ e, portanto, $\Gamma_x(0) H^T$ são todas matrizes diagonais.

Portanto, conclui-se que a transformação (7.3.83) produz m componentes $\{Y_{1t}, \dots, Y_{mt}\}$ que estão ordenadas por ordem crescente de previsibilidade, são contemporaneamente independentes, têm componentes previsíveis $\{\hat{Y}_{1,t-1}(1), \dots, \hat{Y}_{m,t-1}(1)\}$ que são também contemporaneamente independentes e, finalmente, têm componentes imprevisíveis $\{\xi_{1t}, \dots, \xi_{mt}\}$, que são independentes contemporaneamente e para momentos diferentes.

Nestas condições, é possível deduzir a existência de diversos tipos de relações entre as componentes do processo $\{X_t\}$:

a) Relações lineares não exactas contemporâneas

Este tipo de relação surge quando existem valores próprios de λ_i nulos. Quando existem m_1 valores próprios nulos, $\lambda_1, \dots, \lambda_{m_1}$, a matriz $\Gamma_x(0)$ em (7.3.84) pode ser expressa na forma:



$$\Gamma_y(0) = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix} \quad (7.3.86)$$

onde \mathbf{D} é uma matriz diagonal de dimensão $(m_2 \times m_2)$, com $m_2 = m - m_1$.

Sendo:

$$\mathbf{Y}_t^T = (\mathbf{Y}_{1t}^T, \mathbf{Y}_{2t}^T), \quad \xi_t^T = (\xi_{1t}^T, \xi_{2t}^T) \quad (7.3.87)$$

onde \mathbf{Y}_{1t} e ξ_{1t} são vectores de dimensão $(m_1 \times 1)$, verifica-se que $\mathbf{Y}_{1t} = \xi_{1t}$ com probabilidade um. Assim, particionando a matriz $\tilde{\pi}_l$ do seguinte modo:

$$\tilde{\pi}_l = \begin{bmatrix} \tilde{\pi}_{11,l} & \tilde{\pi}_{12,l} \\ \tilde{\pi}_{21,l} & \tilde{\pi}_{22,l} \end{bmatrix} \quad l=1, \dots, p \quad (7.3.88)$$

onde $\tilde{\pi}_{11,l}$ é uma matriz de dimensão $(m_1 \times m_1)$, o processo transformado $\{\mathbf{Y}_t\}$ pode ser expresso da forma:

$$\begin{bmatrix} \mathbf{Y}_{1t} \\ \mathbf{Y}_{2t} \end{bmatrix} = \sum_{l=1}^p \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \tilde{\pi}_{21,l} & \tilde{\pi}_{22,l} \end{bmatrix} \mathbf{B}' \begin{bmatrix} \mathbf{Y}_{1t} \\ \mathbf{Y}_{2t} \end{bmatrix} + \begin{bmatrix} \xi_{1t} \\ \xi_{2t} \end{bmatrix} \quad (7.3.89)$$

Conclui-se então que a transformação canónica decompõe o processo original $\{\mathbf{X}_t\}$ em duas partes: uma parte $\{\mathbf{Y}_{1t}\}$ que é um ruído branco de dimensão $(m_1 \times 1)$ e uma outra $\{\mathbf{Y}_{2t}\}$ que é estacionária, cuja parte previsível depende de $\mathbf{Y}_{1,t-l}$ e de $\mathbf{Y}_{2,t-l}$ para $l=1, \dots, p$.

A importância fundamental da equação (7.3.89) deve-se ao facto de ela implicar a existência de m_1 relações entre as componentes do vector original $\{\mathbf{X}_t\}$ da forma "estática":

$$h_{i1}X_{1t} + \dots + h_{im}X_{mt} = \xi_{it} \quad (i=1, \dots, m_1) \quad (7.3.90)$$

onde ξ_{it} são independentes contemporaneamente e para momentos diferentes.

b) Relações lineares exactas contemporâneas

Foi admitido até agora que $\Gamma_x(0)$ é positiva definida, o que pode não se verificar para um caso concreto. Assim, estas relações ocorrem quando existem m_1 valores próprios de $\Gamma_x(0)$ nulos, ou seja, quando $\Gamma_x(0)$ é uma matriz singular. Nesse caso, existem m_1 relações independentes da forma (Tiao e Box, 1979):

$$w_{i1}X_{1t} + \dots + w_{im}X_{mt} = \xi_{it} \quad (i=1, \dots, m_1) \quad (7.3.91)$$

onde $w_i^T = (w_{i1}, \dots, w_{im})$, com $i=1, \dots, m_1$, são os vectores próprios de $\Gamma_x(0)$ associados aos valores próprios nulos. Relações do tipo de (7.3.91) ocorrem quando uma ou mais componentes de $\{X_t\}$ são construídas a partir de valores contemporâneos das outras. Seria ideal que, perante um caso concreto, o analista conhecesse a proveniência das suas observações suficientemente bem, de modo a estar consciente da existência de tais relações. No entanto, como isso raramente acontece, Tiao e Box (1979) sugerem que se proceda sempre a esta verificação na fase inicial de especificação de um modelo. Assim, é possível detectar a existência de relações desta forma e, através da limitação da análise canónica agora discutida ao subespaço das componentes de $\{X_t\}$ linearmente independentes, em número de $m_2 = m - m_1$, os procedimentos visando a estimação não saem frustrados pela existência de singularidade. A existência de valores próprios próximos de zero pode também constituir um aviso sobre a existência de relações lineares contemporâneas aproximadas.

c) Relações lineares exactas desfazadas

Foi também admitido até agora que a matriz Σ é não singular, podendo no entanto surgir situações em que $\Gamma_x(0)$ é positiva definida e Σ é singular. Admita-se então que a característica de Σ é $m_1 < m$, o que significa que o processo $\{X_t\}$ é na realidade determinado por um

processo aleatório não singular de dimensão $(m_1 \times 1)$. Deste modo, os m_2 valores próprios de $\Gamma_x(0)^{-1} \Gamma_x(0)$, $\lambda_{m_1+1}, \dots, \lambda_m$, são iguais à unidade, assumindo a matriz $\tilde{\Sigma}$, definida em (7.3.84), a forma:

$$\tilde{\Sigma} = \begin{bmatrix} D_1 & 0 \\ 0 & 0 \end{bmatrix} \quad (7.3.92)$$

onde D_1 é uma matriz diagonal de dimensão $(m_1 \times m_1)$ e com elementos positivos. Particionando Y_t , ξ_t e $\tilde{\pi}_t$ como em (7.3.87) e (7.3.88), verifica-se que $\xi_t = 0$ com probabilidade um. Logo, o modelo transformado é:

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \sum_{l=1}^p \begin{bmatrix} \tilde{\pi}_{11l} & \tilde{\pi}_{12l} \\ \tilde{\pi}_{21l} & \tilde{\pi}_{22l} \end{bmatrix} B' \begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} + \begin{bmatrix} \xi_{1t} \\ 0 \end{bmatrix} \quad (7.3.93)$$

Esta expressão significa que o vector Y_{2t} , de dimensão $(m_2 \times 1)$, é completamente previsível a partir dos valores passados $Y_{1,t-l}$ e $Y_{2,t-l}$ ($l=1, \dots, p$). A título de exemplo, Tiao e Box (1979) indicam que é possível que duas séries que aparentam ser muito diferentes, sejam na realidade geradas por variáveis residuais idênticas submetidas à aplicação de diferentes filtros. Na prática, Σ pode ser estimada inicialmente a partir do ajustamento de um modelo auto-regressivo de ordem adequada.

Por outro lado, embora toda esta análise tenha sido feita para um processo AR(p), Tiao e Box (1979) indicam que ela pode ser directamente aplicada a um processo ARMA(p, q).

(ii) O modelo AR(1)

Quando $\{X_t\}$ é um processo auto-regressivo de ordem 1, a transformação canónica apresenta algumas particularidades de interesse. Assim, sendo $p=1$ e fazendo $\pi_1 = \phi_1$, (7.3.75) vem:

$$X_t = \hat{X}_{t-1}(1) + \varepsilon_t = \phi_1 X_{t-1} + \varepsilon_t \quad (7.3.94)$$

Uma vez que $\Gamma_x^T(1) = \phi_1 \Gamma_x(0)$, conclui-se a partir de (7.3.80) que $\Gamma_x(1) = \phi_1 \Gamma_x(0) \phi_1^T + \Sigma$, pelo que λ_i e h_i são respectivamente os m valores próprios e vectores próprios da matriz:

$$\Gamma_x(0)^{-1} \phi_1 \Gamma_x(0) \phi_1^T \quad (7.3.95)$$

Sendo $\tilde{\phi}_1 = H \phi_1 H^{-1}$, o processo transformado pode ser expresso da forma:

$$Y_t = \tilde{\phi}_1 Y_{t-1} + \xi_t \quad (7.3.96)$$

É útil agora analisar os diferentes casos:

a) Séries não estacionárias e raízes unitárias

Foi admitido até agora que $\{X_t\}$ é um processo estacionário, o que nem sempre se verifica. Assim, para o modelo (7.3.94), designando os valores próprios de ϕ_1 por $\alpha_1, \dots, \alpha_m$, tem-se a igualdade:

$$|I_m - \phi_1 B| = \prod_{i=1}^m (1 - \alpha_i B) \quad (7.3.97)$$

o que significa que os zeros de $|I_m - \phi_1 B|$ são $\alpha_1^{-1}, \dots, \alpha_m^{-1}$. Se um ou mais valores próprios $\alpha_i (i=1, \dots, m)$ estiverem sobre o círculo unitário, $\Gamma_x(0)$ não existe, tornando a transformação canónica impossível. No entanto, tem interesse analisar a situação em que m_2 dos valores próprios α_i estão próximos do círculo unitário. Para o efeito define-se:

$$Y_t^T = (Y_{1t}^T, Y_{2t}^T), \quad \xi_t^T = (\xi_{1t}^T, \xi_{2t}^T), \quad \tilde{\phi}_1 = \begin{bmatrix} \tilde{\phi}_{11,1} & \tilde{\phi}_{12,1} \\ \tilde{\phi}_{21,1} & \tilde{\phi}_{22,1} \end{bmatrix} \quad (7.3.98)$$

onde Y_{1t} e ξ_{1t} são vectores de dimensão $(m_1 \times 1)$ e $\tilde{\phi}_{11,1}$ é uma matriz de dimensão $(m_1 \times m_1)$, sendo $m_1 = m - m_2$. Nestas condições, os autores deste método provam o teorema:

Teorema 7.3.4. Suponha-se que $\{X_t\}$ é um processo estocástico estacionário da forma de (7.3.94), onde a matriz de variâncias e covariâncias de $\{\varepsilon_t\}$, designada por Σ , é definida positiva. Uma condição necessária e suficiente para que m_2 valores próprios de:

$$\Gamma_x(0)^{-1} \phi_1 \Gamma_x(0) \phi_1^T$$

tendam para a unidade, é que m_2 valores próprios de ϕ_1 estejam próximos do círculo unitário.

Por outro lado, estes autores provam ainda que, no limite, o modelo transformado para Y_t é:

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} \tilde{\phi}_{11,1} & 0 \\ \tilde{\phi}_{21,1} & \tilde{\phi}_{22,1} \end{bmatrix} B \begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} + \begin{bmatrix} \xi_{1t} \\ \xi_{2t} \end{bmatrix} \quad (7.3.99)$$

Conclui-se portanto que a transformação canónica decompõe $\{Y_t\}$ em duas partes: uma parte $\{Y_{1t}\}$ que é um processo auto-regressivo de ordem um estacionário e uma parte $\{Y_{2t}\}$ que é quase não estacionária e também depende de $Y_{1,t-1}$. O significado deste resultado é o de que as componentes de $\{Y_{2t}\}$ podem ser úteis como indicadores compostos do crescimento dinâmico global do processo original.

b) Raízes nulas e unitárias

Suponha-se que, no contexto do modelo (7.3.94), m_1 dos valores próprios α_i são nulos, m_3 estão próximos da unidade e os restantes $m_2 = m - m_1 - m_3$ assumem valores intermédios. Assim, a partir dos resultados (7.3.89) e (7.3.99) e particionando Y_t , ξ_t e $\tilde{\phi}_1$ da forma:

$$Y_t^T = (Y_{1t}^T, Y_{2t}^T, Y_{3t}^T), \quad \xi_t^T = (\xi_{1t}^T, \xi_{2t}^T, \xi_{3t}^T), \quad \tilde{\phi}_1 = \begin{bmatrix} \tilde{\phi}_{11,1} & \tilde{\phi}_{12,1} & \tilde{\phi}_{13,1} \\ \tilde{\phi}_{21,1} & \tilde{\phi}_{22,1} & \tilde{\phi}_{23,1} \\ \tilde{\phi}_{31,1} & \tilde{\phi}_{32,1} & \tilde{\phi}_{33,1} \end{bmatrix} \quad (7.3.100)$$

onde Y_{1t} e ξ_{1t} são vectores de dimensão $(m_1 \times 1)$, Y_{2t} e ξ_{2t} são vectores de dimensão $(m_2 \times 1)$ e $\tilde{\phi}_{11,t}$ e $\tilde{\phi}_{22,t}$ são matrizes de dimensão $(m_1 \times m_1)$ e $(m_2 \times m_2)$ respectivamente, o processo transformado $\{Y_t\}$ assume a forma:

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \\ Y_{3t} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ \tilde{\phi}_{21,t} & \tilde{\phi}_{22,t} & 0 \\ \tilde{\phi}_{31,t} & \tilde{\phi}_{32,t} & \tilde{\phi}_{33,t} \end{bmatrix} B \begin{bmatrix} Y_{1t} \\ Y_{2t} \\ Y_{3t} \end{bmatrix} + \begin{bmatrix} \xi_{1t} \\ \xi_{2t} \\ \xi_{3t} \end{bmatrix} \quad (7.3.101)$$

Logo, tem-se um ruído branco $\{Y_t\}$ de dimensão $(m_1 \times 1)$, um processo estacionário $\{Y_{2t}\}$ de dimensão $(m_2 \times 1)$ e tal que a parte previsível de Y_{2t} depende apenas de $Y_{1,t-1}$ e de $Y_{2,t-1}$, e ainda um processo próximo da não estacionaridade $\{Y_{3t}\}$ de dimensão $(m_3 \times 1)$ e tal que a parte previsível de Y_{3t} depende de $Y_{1,t-1}$, de $Y_{2,t-1}$ e de $Y_{3,t-1}$.

c) Transformação canónica individualizadora

Sendo $\{X_t\}$ um processo AR(1) e se o número de vectores próprios de ϕ_1 linearmente independentes for de m , Tiao e Box (1979) indicam que existe uma matriz real F de dimensão $(m \times m)$ que verifica a relação:

$$F\phi_1 F^{-1} = \Lambda \quad (7.3.102)$$

onde Λ é uma matriz diagonal composta por blocos, tendo estes dimensão unitária correspondendo a valores próprios reais ou dimensão igual a 2 para pares de valores próprios complexos. Deste modo, obtém-se:

$$Z_t = \Lambda Z_{t-1} + \xi_t \quad (7.3.103)$$

onde $Z_t = FX_t$ e $\xi_t = F\xi_t$.

Quando um valor próprio de ϕ_1 é real, o processo univariado transformado correspondente é "não emparelhado", no sentido de que pode ser previsto de forma óptima a partir do seu próprio passado. A

existência de um par de valores próprios complexos implica que só é possível o "não emparelhamento" de pares de componentes do processo multivariado, ou seja, que um par de componentes do processo transformado pode ser previsto pelo passado apenas desse par. É neste sentido que esta transformação foi designada de "individualizadora" (Tiao e Box, 1979).

Estes autores referem ainda que esta análise pode também ser generalizada ao caso de um processo MA(1).

Pretendendo elucidar sobre a dinâmica interna de uma série cronológica multivariada, esta análise apresenta ainda uma outra faceta importante: a identificação das componentes não estacionárias pode evitar a diferenciação de todas as componentes da série, o que, conforme foi oportunamente referido, acarreta diversos problemas. Assim, a detecção de quais as componentes de uma série multivariada que são responsáveis pelo seu crescimento dinâmico pode ser muito útil para a sua modelização. Com efeito, o crescimento dinâmico de uma série multivariada pode ser devido a um pequeno conjunto de componentes não estacionárias, podendo existir relações contemporâneas estáveis entre as outras. A identificação das características dessas diferentes componentes pode permitir evitar a sobre-diferenciação e todos os problemas que lhe são inerentes.

Em resumo, a análise canónica afigura-se extremamente promissora para a compreensão do padrão de comportamento global apresentado por uma série multivariada, constituindo um instrumento de grande utilidade.

3.6. Conclusão

Todos os métodos simplificadores acabados de analisar têm um objectivo comum: facilitar a identificação de uma série cronológica multivariada. No entanto, de uma forma geral, os seus resultados parecem ser de eficácia duvidosa, introduzindo um tal grau de complexidade na análise, que se torna legítimo questionar se a sua utilização é aconselhável.

Assim, embora por vezes possa vir a ser útil, o método de identificação dos elementos significativos nos operadores de um modelo ARMA, proposto por Koreisha e Pukkila (1987), é susceptível de acarretar dificuldades computacionais acrescidas, que poderão não ser compensadas pelas suas vantagens. Com efeito, os métodos utilizados na fase de confirmação do diagnóstico após a identificação e estimação sem recurso a este procedimento, deverão conduzir a resultados idênticos. De qualquer modo, o potencial desta proposta parece ser grande, podendo a sua utilização constituir um bom auxílio na fase de identificação.

O emprego de um modelo factorial, conforme proposto por Peña e Box (1987), levanta imediatamente a questão de ser necessário admitir a existência de factores ocultos subjacentes à geração de uma série cronológica, o que é muito questionável. Além disso, revela-se de grande complexidade e de difícil utilização, sendo muito indirecto. A forma como a abordagem é proposta conduz a uma estrutura interna dos modelos muito complexa e de difícil compreensão, levando a uma duvidosa aproximação prática. Na realidade, tal como já foi referido, a causa da complexidade da estrutura de um dado modelo ARMA identificado deverá ser encontrada na forma como essa identificação foi realizada e não na existência de factores ocultos.

A adopção da proposta de Reinsel (1983), que pretende modelizar uma série cronológica através da consideração de modelos índice auto-regressivos, acarreta também diversos problemas. Em primeiro lugar, a

consideração de um único tipo de modelos não parece ser correcta, devendo ser a série cronológica a ditar a estrutura do modelo mais apropriada. Isto pode resultar em falta de adequação do modelo à série cronológica, conduzindo à obtenção de uma estrutura sobreparametrizada, o que significaria o fracasso do método.

Por sua vez, os efeitos da adopção de modelos auto-regressivos de polinómio amortecido, propostos por Lütkepohl (1982b), podem não ser os esperados. Na realidade, pretendendo ser um compromisso entre o recurso a modelos ARMA e a utilização de modelos auto-regressivos, de modo a evitar os problemas acarretados pelo emprego destes últimos e a obtenção de estruturas pesadas resultantes dos primeiros, este método acaba por conduzir a um procedimento bastante complexo. Por outro lado, é muito susceptível de conduzir também a estruturas sobreparametrizada, o que significa o fracasso do seu duplo objectivo.

Finalmente, a análise canónica pode ser de grande utilidade, embora a sua aplicação não seja fácil. Tratando-se de uma análise paralela, as indicações que pode fornecer são bastante válidas, sendo no entanto de difícil interpretação. De qualquer modo, parece ser um tipo de análise de elevado potencial.

Em resumo, todas estas abordagens não parecem atingir o seu objectivo, no seu estágio actual de desenvolvimento, não constituindo alternativas credíveis às extensões ao caso multivariado do método de Box e Jenkins (1976). Simultaneamente, revelam-se de difícil aplicação prática e de resultado duvidoso ou, pelo menos, incerto. No entanto, poderão eventualmente ser de alguma utilidade e constituir instrumentos auxiliares válidos, embora as dificuldades computacionais que acarretam possam tornar a sua utilização pouco interessante. De qualquer modo, estes métodos não deixam de assumir uma certa relevância, tendo por isso sido analisados, e podendo constituir pontos de partida para o desenvolvimento

futuro de novos procedimentos que se venham a revelar eficazes e válidos na simplificação de métodos conducentes à modelização de uma série cronológica multivariada.

NOTAS

(1) No entanto, esta não é a única definição possível de matriz de correlação parcial.

Assim, considere-se agora a regressão X_t sobre X_{t-l} ($l = 1, \dots, k$):

$$X_t = \phi_{k1} X_{t-1} + \phi_{k2} X_{t-2} + \dots + \phi_{kk} X_{t-k} + U_{kt}$$

onde ϕ_{kl} ($l = 1, \dots, k$) são matrizes de parâmetros, de dimensão $(m \times m)$ e U_{kt} tem matriz de covariância D_k , sendo o termo residual da regressão.

Considere-se também a regressão de X_t sobre X_{t+l} ($l = 1, \dots, k$):

$$X_t = \phi'_{k1} X_{t+1} + \phi'_{k2} X_{t+2} + \dots + \phi'_{kk} X_{t+k} + U'_{kt}$$

sendo D'_k a matriz de covariância de U'_{kt} .

Designando por A_k e por A'_k as raízes quadradas simétricas de U_{kt} e de U'_{kt} respectivamente, a matriz de correlação parcial de ordem $k + 1$ pode também ser definida de uma forma alternativa e que é assintoticamente equivalente (Ansley e Newbold, 1979):

$$P_{k+1} = A_k^{-1} E(U_{kt} U'_{k,t-k-1}) A_k^{-1}$$

Ansley e Newbold (1979) mostram que P_{k+1} pode também ser expresso da forma:

$$P_{k+1} = A_k^{-1} \phi_{k+1,k+1} A'_k = A_k \phi_{k+1,k+1}^{\prime T} A_k^{-1}$$

De qualquer modo, a definição de Tiao e Box (1979, 1981) e de Jenkins e Alavi (1981) parece ser preferível, quer por ser mais fácil de utilizar, quer por constituir uma extensão da função de correlação parcial válida no caso univariado.

(2) É possível extrair as mesmas conclusões utilizando a definição de matriz de correlação parcial de Ansley e Newbold (1979) e que assume a expressão:

$$P_{k+1} = A_k^{-1} E(U_k U_{k+1-k}^T) A_k^{-1}$$

De facto, é imediato verificar que a função de correlação parcial, quando definida desta forma, também se anula a partir da ordem do processo, uma vez que U_k e U_{k+1-k} são não correlacionados.

(3) Se se utilizar a definição alternativa de matriz de correlação parcial, é necessário recorrer à função $C(k)$ para a sua estimação, o que pode ser feito através da utilização de um método recursivo proposto por Ansley e Newbold (1979) que permite obter o estimador de P_{k+1} , designado por \hat{P}_{k+1} , a partir dos estimadores de A_k , de A_k' e de $\phi_{k+1,k+1}$, designados respectivamente por \hat{A}_k , \hat{A}_k' e $\hat{\phi}_{k+1,k+1}$.

(4) Um bloco de dimensão $(m \times m)$ de uma qualquer matriz A será designado por A_{rs} , incluindo as linhas $m(r-1)+1$ a mr e as colunas $m(s-1)+1$ a ms da matriz A . Além disso, referir-se-à ainda a l -ésima coluna de blocos de A , que inclui todas as linhas das colunas $m(l-1)+1$ a ml , e a r -ésima linha de blocos, que inclui todas as colunas das linhas $m(r-1)+1$ a mr .

(5) No que se refere à matriz de correlação parcial estimada, segundo a definição de Ansley e Newbold (1979), Hannan (1970, pág. 398) mostra que, sob a hipótese de $\{X_t\}$ ser um processo AR(k), os elementos de \hat{P}_{k+l} , $l=1, 2, \dots$, são independentes, seguindo distribuição assintótica Normal de valor esperado nulo e variância N^{-1} . Logo, (Ansley e Newbold, 1979; Newbold e Hottop, 1986), as estatísticas:

$$D_{k+l} = \text{Ntr} \hat{\mathbf{P}}_{k+l} \hat{\mathbf{P}}_{k+l}^T$$

são também independentes, seguindo distribuição assintótica χ_m^2 . Uma vez que se verifica a relação:

$$\hat{\mathbf{P}}_{k+l}^T = \hat{\mathbf{A}}_{k+l-1}^{-1} \hat{\boldsymbol{\phi}}_{k+l,k+l} \hat{\mathbf{A}}_{k+l-1}'$$

é possível expressar D_{k+l} sob a forma alternativa (Ansley e Newbold, 1979):

$$\begin{aligned} D_{k+l} &= \text{Ntr} \hat{\mathbf{P}}_{k+l} \hat{\mathbf{P}}_{k+l}^T = \text{Ntr} \hat{\mathbf{A}}_{k+l-1}^{-1} \hat{\boldsymbol{\phi}}_{k+l,k+l} \hat{\mathbf{A}}_{k+l-1}' \hat{\mathbf{A}}_{k+l-1}^{-1} \hat{\boldsymbol{\phi}}_{k+l,k+l}' \hat{\mathbf{A}}_{k+l-1} = \\ &= \text{Ntr} \hat{\boldsymbol{\phi}}_{k+l,k+l} \hat{\boldsymbol{\phi}}_{k+l,k+l}'^{-1} \end{aligned}$$

Logo, conclui-se que as estatísticas D_l podem também ser utilizadas na identificação de processos auto-regressivos.

(6) Neste contexto, têm sido propostas diversas definições para $\hat{\mathbf{R}}_\varepsilon(k)$, inventariadas por Hosking (1981b), que mostra que as estatísticas resultantes das diversas definições são iguais a Q_2 , à exceção da proposta em c):

a) $\hat{\mathbf{R}}_\varepsilon(k) = \hat{\mathbf{C}}_{\varepsilon_0}^{-1/2} \hat{\mathbf{C}}_\varepsilon(k) \hat{\mathbf{C}}_{\varepsilon_0}^{-1/2}$ que é a definição habitual e que é também adoptada por Li e McLeod (1981).

b) $\hat{\mathbf{R}}_\varepsilon(k) = \mathbf{L}^T \hat{\mathbf{C}}_\varepsilon(k) \mathbf{L}$ apresentada por Hosking (1980a), onde \mathbf{L} é uma matriz triangular inferior tal que $\mathbf{L}\mathbf{L}^T = \hat{\mathbf{C}}_\varepsilon(0)^{-1}$. Adoptando esta definição, resulta ainda:

$$Q_2 = N \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^R \hat{r}_{\varepsilon_{ij}}^2(k)$$

c) $\hat{R}_\varepsilon(k) = \hat{C}_\varepsilon(k)\hat{C}_\varepsilon(0)^{-1}$ adoptada por Chitturi (1974), que apresenta uma estatística para testar a qualidade do ajustamento de modelos auto-regressivos.

Utilizando esta definição, obtém-se o resultado seguinte (Chitturi, 1974 ; Hosking 1980a):

$$Q_2^* = N \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^R \hat{r}_{\varepsilon_{ij}}(k) \hat{r}_{\varepsilon_{ij}}(-k)$$

Esta estatística pode também ser expressa da forma (Hosking, 1989):

$$Q_2^* = N \sum_{k=1}^R \left[\text{vec} \hat{R}_\varepsilon(k) \right]^T \text{vec} \hat{R}_\varepsilon^T(-k)$$

Assim, conclui-se que, quando se adopta esta definição para $\hat{R}_\varepsilon(k)$, a estatística do teste que se obtém já não é igual a Q_2 .

(7) Trata-se de um problema distinto do problema de identificação de um modelo, que consiste no diagnóstico do modelo que melhor se ajuste a uma série cronológica e que será tratado mais à frente. Assim, afigurou-se preferível utilizar o termo "identifiabilidade" para este problema, embora haja autores, como é o caso de Hannan (1975, 1976), que empregam a palavra "identificação" para designar ambas as questões. No entanto, para evitar confusões que daí poderiam surgir, utilizou-se designações diferentes para as duas questões, tal como Priestley (1981). Granger e Newbold (1977) preferem utilizar a designação "identificação de séries cronológicas" ("time series identification" ou, abreviadamente, "TS-identification") para o que se apelidou de identificação, e o termo "identificação econométrica" ("econometric identification" ou, de forma abreviada, "E-identification") para o que se designou por identifiabilidade.

(8) Granger e Newbold (1977a,b) designam a relação unidireccional por "casualidade unidireccional" e a relação bidireccional por "feedback". Tais designações resultam da definição de causalidade adoptada por estes autores, sendo uma questão que não irá ser analisada. Pierce e Haugh (1977), Granger e Newbold (1977a; 1986) ou Newbold e Hottop (1986) fornecem adequada discussão deste tema.

(9) Na realidade, Box e Jenkins (1976) definem o modelo (7.1.42) da forma:

$$X_{1t} = V'(B)X_{2,t-b} + U'(B)\varepsilon_t$$

onde b é um parâmetro adicional que é designado de "atraso" do sistema e sendo um número inteiro não negativo.

BIBLIOGRAFIA

Ahn, S. K. (1988)

"Distribution for Residual Autocovariances in Multivariate Autoregressive Models with Structured Parameterization"
Biometrika 75, 590-593.

Akaike, H. (1969)

"Fitting Autoregressive Models for Prediction"
Annals of the Institute of Statistical Mathematics 21, 243-247.

Akaike, H. (1971)

"Autoregressive Model for Fitting and Control"
Annals of the Institute of Statistical Mathematics 23, 163-180.

Akaike, H. (1973)

"Maximum Likelihood Identification of Gaussian Autoregressive Moving Average Models"
Biometrika 60, 255-265.

Ali, M. M. (1989)

"Tests for Autocorrelation and Randomness in Multiple Time Series"
Journal of the American Statistical Association 84, 533-540.

Anderson, T. W. (1971)

"The Statistical Analysis of Time Series"
John Wiley and Sons, Nova Iorque.

Anderson, T. W. (1980)

"Maximum Likelihood Estimation for Vector Autoregressive Moving Average Models"
"Directions of Time Series" Eds. D. R. Brillinger e G. C. Tiao, Institute of Mathematical Statistics, 49-59.

Anderson, T. W. (1984)

"An Introduction to Multivariate Statistical Analysis"
John Wiley and Sons.

- Ansley, C. F. e Newbold, P. (1979)
"Multivariate Partial Autocorrelations"
Proceedings of the Business and Economics Statistics, American
Statistical Association, 349-353.
- Ash, R. B. (1972)
"Real Analysis and Probability"
Academic Press, Nova Iorque.
- Baillie, R. T. (1979)
"Asymptotic Prediction Mean Squared Error for Vector Autoregressive
Models"
Biometrika 66, 675-678.
- Barone, P. (1987)
"A Method for Generating Independent Realizations of a Multivariate
Normal Stationary and Invertible ARMA(p,q) Process"
Journal of Time Series Analysis 8, 125-130.
- Bartlett, M. S. (1978)
"An Introduction to Stochastic Processes with Special Reference to
Methods and Applications"
Cambridge University Press, Cambridge.
- Bartlett, M. S. e Rajalakshman, D. V. (1953)
"Goodness of Fit Tests for Simultaneous Autoregressive Series"
Journal of the Royal Statistical Society B, 15, 107-124.
- Billingsley, P. (1986)
"Probability and Measure"
John Wiley and Sons, Nova Iorque.
- Box, G. E. P. e Cox, D. R. (1964)
"An Analysis of Transformations"
Journal of the Royal Statistical Society B, 26, 211-252.
- Box, G. E. P. e Jenkins, G. M. (1976)
"Time Series Analysis : Forecasting and Control"
Holden-Day, San Francisco, California.

- Box, G. E. P. e Newbold, P. (1971)
"Some Comments on a Paper of Coen, Gomme and Kendall"
Journal of the Royal Statistical Society A, 134, 229-240.
- Box, G. E. P. e Pierce, D. (1970)
"Distribution of Residual Autocorrelations in Autoregressive-Integrated
Moving Average Time Series Models"
Journal of the American Statistical Association 65, 1509-1526.
- Box, G. E. P. e Tiao, G. C. (1977)
"A Canonical Analysis of Multiple Time Series"
Biometrika 64, 355-365.
- Brockwell, P. J. e Davis, R. A. (1987)
"Time Series : Theory and Methods"
Springer-Verlag, Nova Iorque.
- Caines, P. E. e Chan, C. W. (1975)
"Feedback Between Stationary Stochastic Processes"
I.E.E.E. Transactions on Automatic Control AC-20, 498-508.
- Chan, W.Y.T. e Wallis, K.F. (1978)
"Multiple Time Series Modelling : Another Look at the Mink-Muskrat
Interaction"
Applied Statistics 27, 168-175.
- Chatfield, C. (1989)
"The Analysis of Time Series : An Introduction"
Chapman and Hall, Londres.
- Chitturi, R. V. (1974)
"Distribution of Residual Autocorrelations in Multiple Autoregressive
Schemes"
Journal of the American Statistical Association 69, 928-934.
- Chitturi, R. V. (1976)
"Distribution of Multivariate White Noise Autocorrelations"
Journal of the American Statistical Association 74, 223-226.

- Cipra, T. e Tlustý, P. (1987)
"Estimation in Multiple Autoregressive-Moving Average Models Using Periodicity"
Journal of Time Series Analysis 8, 293-300.
- Cox, D. R. e Miller, H. D. (1965)
"The Theory of Stochastic Processes"
Chapman and Hall, Londres.
- Cramer, R. H. e Miller, R. B. (1976)
"Dynamic Modeling of Multivariate Time Series for Use in Bank Analysis"
Journal of Money, Credit and Banking 8, 85-96.
- Deistler, M., Dunsmuir, W. e Hannan, E. J. (1978)
"Vector Linear Time Series Models : Connections and Extensions"
Advances in Applied Probability 10, 360-372.
- Deistler, M. e Hannan, E. J. (1981)
"Some Properties of the Parameterization of ARMA Systems with Unknown Order"
Journal of Multivariate Analysis 11, 474-484.
- Deistler, M., Ploberger, W. e Pötscher, B. M. (1982)
"Identifiability and Indeference in ARMA Systems"
"Time Series Analysis : Theory and Practice 2", Ed. O. D. Anderson,
North-Holland, Amsterdão, 43-60.
- Downing, D. J. e Pack, D. J. (1982)
"The Vanishing Transfer Function"
"Time Series Analysis : Theory and Practice 2", Ed. O. D. Anderson,
North-Holland, Amsterdão, 227-246.
- Dunsmuir, W. (1979)
"A Central Limit Theorem for Parameter Estimation in Stationary Vector Time Series and its Application to Models for a Signal Observed with Noise"
Annals of Statistics 7, 490-506.

- Dunsmuir, W. e Hannan, E. J. (1976)
"Vector Linear Time Series Models"
Advances in Applied Probability 8, 339-364.
- Geweke, J. (1981)
"A Comparison of Tests of the Independence of Two Covariance-Stationary Time Series"
Journal of the American Statistical Association 76, 363-373.
- Geweke, J. (1982)
"Measurement of Linear Dependence and Feedback Between Multiple Time Series"
Journal of the American Statistical Association 77, 304-313.
- Geweke, J. (1984)
"Measures of Conditional Linear Dependence and Feedback Between Time Series"
Journal of the American Statistical Association 79, 907-915.
- Gottman, J. (1981)
"Time-Series Analysis
A Comprehensive Introduction for Social Scientists"
Cambridge University Press, Nova Iorque.
- Granger, C. W. J. e Hatanaka, M. (1964)
"Spectral Analysis of Economic Time Series"
Princeton University Press, Princeton, New Jersey.
- Granger, C. W. J. e Newbold, P. (1977a)
"Forecasting Economic Time Series"
Academic Press, Orlando, Florida.
- Granger, C. W. J. e Newbold, P. (1977b)
"Identification of Two-Way Causal Systems"
"Frontiers of Quantitative Economics", Ed. M. D. Intriligator, Vol. 3,
North-Holland, 337-359.
- Granger, C. W. J. e Newbold, P. (1986)
"Forecasting Economic Time Series"
Academic Press, San Diego.

- Hallin, M. (1989)
 "Modèles non Stationnaires, Séries Univariées et Multivariées"
 "Séries Chronologiques - Théorie et Pratique des Modeles ARIMA",
 Ed. J. -J. Dreesbeke, B. Fichet e P. Tassi, Economica, Paris, 157-196.
- Hannan, E. J. (1969a)
 "The Identification of Vector Mixed Autoregressive-Moving Average
 Systems"
 Biometrika 56, 223-225.
- Hannan, E. J. (1969b)
 "The Estimation of Mixed Moving Average Autoregressive Systems"
 Biometrika 56, 579-593.
- Hannan, E. J. (1970)
 "Multiple Time Series"
 John Wiley and Sons, Nova Iorque.
- Hannan, E. J. (1971)
 "The Identification Problem for Multiple Equation Systems with Moving
 Average Errors"
 Econometrica 39, 751-766.
- Hannan, E. J. (1975)
 "The Estimation of ARMA Models"
 Annals of Statistics 3, 975-981.
- Hannan, E. J. (1976)
 "The Identification and Parameterization of ARMAX and State Space
 Forms"
 Econometrica 44, 713-723.
- Hannan, E. J. (1978)
 "Multivariate ARMA Theory"
 "Stability and Inflation - A Volume of Essays to Honour the Memory
 of A. W. H. Phillips", Eds. A. R. Bergstrom, A. J. L. Catt, M. H. Peston
 e B. D. J. Silverstone, John Wiley and Sons, Nova Iorque, 201-212.

- Hannan, E. J. (1979)
"The Statistical Theory of Linear Systems"
"Developments in Statistics" vol. 2, ed. P. Krishnaiah, Academic Press,
Nova Iorque, 83-121.
- Hannan, E. J. (1981)
"Estimating the Dimension of a Linear System"
Journal of Multivariate Analysis 11, 450-473.
- Hannan, E. J., Dunsmuir, W. e Deistler, M. (1980)
"Estimation of Vector ARMAX Models"
Journal of Multivariate Analysis 10, 275-295.
- Hannan, E. J. e Kavalieris, L. (1984)
"Multivariate Linear Time Series Models"
Advances in Applied Probability 16, 492-561.
- Hannan, E. J. e Quinn, B. G. (1979)
"The Determination of the Order of an Autoregression"
Journal of the Royal Statistical Society B, 41, 190-195.
- Harvey, A. C. (1981)
"Time Series Models"
Phillip Allan Publishers Limited, Oxford.
- Haugh, L. D. (1976)
"Checking the Independence of Two Covariance-Stationary Time
Series: A Univariate Residual Cross-Correlation Approach"
Journal of the American Statistical Association 71, 378-385.
- Haugh, L. D. e Box, G. E. P. (1977)
"Identification of Dynamic Regression (Distributed Lag)
Models Connecting Two Time Series"
Journal of the American Statistical Association 72, 121-130.
- Hillmer, S. C. e Tiao, G. C. (1979)
"Likelihood Function of Stationary Multiple Autoregressive Moving
Average Models"
Journal of the American Statistical Association 74, 652-661.

- Hosking, J. R. M. (1980a)
"The Multivariate Portmanteau Statistic"
Journal of the American Statistical Association 75, 602-608.
- Hosking, J. R. M. (1980b)
"Lagrange-Multiplier Tests of Time Series Models"
Journal of the Royal Statistical Society B, 42, 170-181.
- Hosking, J. R. M. (1981a)
"Lagrange-Multiplier Tests of Multivariate Time Series Models"
Journal of the Royal Statistical Society B, 43, 219-230.
- Hosking, J. R. M. (1981b)
"Equivalent Forms of the Multivariate Portmanteau Statistic"
Journal of the Royal Statistical Society B, 43, 261-262.
- Hosking, J. R. M. (1989)
"Corrigendum: Equivalent Forms of the Multivariate Portmanteau Statistic"
Journal of the Royal Statistical Society B, 51, 303.
- Jenkins, G. M. e Alavi, A. S. (1981)
"Some Aspects of Modelling and Forecasting Multivariate Time Series"
Journal of Time Series Analysis 2, 1-47.
- Jenkins, G. M. e Watts, D. G. (1968)
"Spectral Analysis and its Applications"
Holden-Day, San Francisco.
- Judge, G. , Griffiths, W. E., Hill, R. C. , Lütkepohl, H. e Lee, T.-C. (1985)
"The Theory and Practice of Econometrics"
John Wiley and Sons, Nova Iorque.
- Karlin, S. (1968)
"A First Course in Stochastic Processes"
Academic Press, Londres.
- Karlin, S. e Taylor, H. M. (1975)
"A First Course in Stochastic Processes"
Academic Press, Nova Iorque.

Kendall, M. G., Stuart, A. e Ord, J. K.

"The Advanced Theory of Statistics" vol. 3, Charles Griffin & Co. Ltd.,
Londres.

Khabie-Zeitoune, E. (1982)

"Identification, Estimation and Prediction of Linear and Nonlinear
Multivariate Models in Time Series"

"Time Series Analysis : Theory and Practice 1", Ed. O. D. Anderson,
North-Holland, Amsterdão, 607-620.

Koch, P. D. e Yang, S. S. (1986)

"A Method for Testing the Independence of Two Time Series that
Accounts for a Potential Pattern in the Cross-Correlation Function"
Journal of the American Statistical Association 81, 533-544.

Kohn, R. (1977)

"Note Concerning the Akaike and Hannan Estimation Procedures for an
Autoregressive Moving Average Process"
Biometrika 64, 622-625.

Kohn, R. (1978)

"Asymptotic Properties of Time Domain Gaussian Estimators"
Advances in Applied Probability 10, 339-359.

Kohn, R. (1979)

"Asymptotic Estimation and Hypothesis Testing Results for Vector
Linear Time Series Models"
Econometrica 47, 1005-1030.

Koreisha, S. G. e Pukkila, T. M. (1987)

"Identification of Nonzero Elements in the Polynomial Matrices of Mixed
VARMA Processes"
Journal of the Royal Statistical Society B, 112-116.

Koster, F. H. (1982)

"An Application of Transfer Function-Noise Models in the Financial
Sector"

"Time Series Analysis : Theory and Practice 2", Ed. O. D. Anderson,
North-Holland, Amsterdão, 75-84.

Li, W. K. (1985)

"Distribution of Residual Autocorrelations in Multivariate Autoregressive Index Models"

Biometrika, 72, 686-688.

Li, W. K. e McLeod, A. I. (1981)

"Distribution of the Residual Autocorrelations in Multivariate ARMA Time Series Models"

Journal of the Royal Statistical Society B, 43, 231-239.

Lii, K. S. (1985)

"Transfer Function Model Order and Parameter Estimation"

Journal of Time Series Analysis 6, 153-169.

Lütkepohl, H. (1982a)

"Differencing Multiple Time Series : Another Look at the Canadian Money and Income Data"

Journal of Time Series Analysis 3, 235-243.

Lütkepohl, H. (1982b)

"Discounted Polynomials for Multiple Time Series Model Building"

Biometrika 69, 107-115.

Lütkepohl, H. (1985)

"Comparison of Criteria for Estimating the Order of a Vector Autoregressive Process"

Journal of Time Series Analysis 6, 35-52.

McLeod, A. I. (1979)

"Distribution of the Residual Cross-Correlation in Univariate ARMA Time Series Models"

Journal of the American Statistical Society 74, 849-855.

Morf, M., Viera, A. e Kailath, T. (1978)

"Covariance Characterization by Partial Autocorrelation Matrices"

The Annals of Statistics 6, 643-648.

- Nelson, C. R. e Schwert, G. W. (1982)
"Tests for Predictive Relationships Between Time Series Variables : A Monte Carlo Investigation"
Journal of the American Statistical Association 77, 11-18.
- Newbold, P. e Hottop, S. M. (1986)
"Testing Causality Using Efficiently Parametrized Vector ARMA Models"
Applied Mathematics and Computation, 20, 329-348.
- Nicholls, D. F. (1976)
"The Efficient Estimation of Vector Linear Time Series Models"
Biometrika 63, 381-390.
- Nicholls, D. F. (1977)
"A Comparison of Estimation Methods for Vector Linear Time Series Models"
Biometrika 64, 85-90.
- Nicholls, D. F. e Hall, A. D. (1979)
"The Exact Likelihood Function of Multivariate Autoregressive Moving Average Models"
Biometrika 66, 259-264.
- Osborn, D. R. (1977)
"Exact and Approximate Maximum Likelihood Estimators for Vector Moving Average Processes"
Journal of the Royal Statistical Society B, 39, 114-118.
- Pagano, M. (1978)
"On Periodic and Multiple Autoregressions"
The Annals of Mathematical Statistics 6, 1310-1317.
- Pankratz, A. (1983)
"Forecasting with Univariate Box - Jenkins Models"
John Wiley and Sons, Nova Iorque.
- Parzen, E. (1962)
"Stochastic Processes"
Holden-Day, San Francisco.

Parzen, E. (1969)

"Multiple Time Series Modeling"

"Multivariate Analysis II", Ed. P. Krishnaiah, Academic Press, Nova Iorque, 389-409.

Parzen, E. (1977)

"Multiple Time Series : Determining the Order of Approximating Autoregressive Schemes"

"Multivariate Analysis IV", Ed. P. Krishnaiah, North-Holland, Amsterdão, 283-295.

Parzen, E. e Newton, H. J. (1979)

"Multiple Time Series Modelling II"

"Multivariate Analysis V", Ed. P. Krishnaiah, North-Holland, Amsterdão, 181-197.

Paulsen, J. (1984)

"Order Determination of Multivariate Autoregressive Time Series with Unit Roots"

Journal of Time Series Analysis 5, 115-127.

Paulsen, J. e Tjøstheim, D. (1985)

"On the Estimation of Residual Variance and Order in Autoregressive Time Series"

Journal of the Royal Statistical Society B, 216-228.

Peña, D. e Box, G. E. P. (1987)

"Identifying a Simplifying Structure in Time Series"

Journal of the American Statistical Association 82, 836-843.

Phadke, M. S. e Kedem, G. (1978)

"Computation of the Exact Likelihood Function of Multivariate Moving Average Models"

Biometrika 65, 511-519.

Pham-Dihn, T. (1978)

"On the Fitting of Multivariate Processes of the Autoregressive-Moving Average Type"

Biometrika 65, 99-107.

- Pierce, D. (1968)
"Distribution of Residual Correlations in Dynamic/Stochastic Time Series Models"
Technical Report N° 173, Department of Statistics, University of Wisconsin.
- Pierce, D. (1975)
"Forecasting in Dynamic Models with Stochastic Regressors"
Journal of Econometrics 3, 349-374.
- Pierce, D. (1977)
"Relationships - and the Lack Thereof - Between Economic Time Series, with Special Reference to Money and Interest Rates"
Journal of the American Statistical Association 72, 11-22
- Pierce, D. e Haugh, L. D. (1977)
"Causality in Temporal Systems : Characterizations and a Survey"
Journal of Econometrics 5, 265-293.
- Poskitt, D. S. (1989)
"A Method for the Estimation and Identification of Transfer Function Models"
Journal of the Royal Statistical Society B, 29-46.
- Poskitt, D. S. e Tremayne, A. R. (1980)
"Testing the Specification of a Fitted ARMA model"
Biometrika 67, 359-363.
- Poskitt, D. S. e Tremayne, A. R. (1982)
"Diagnostic Tests for Multiple Time Series Models"
Annals of Statistics 10, 114-120.
- Poskitt, D. S. e Tremayne, A. R. (1985)
"Some Aspects of the Performance of Diagnostic Checks in Bivariate Time Series Models"
Journal of Time Series Analysis 7, 217-233.
- Priestley, M. B. (1971)
"Fitting Relationships Between Time Series"
Bulletin of the International Statistical Institute 34, 295-321.

- Priestley, M. B. (1981)
"Spectral Analysis and Time Series"
Academic Press, Londres.
- Quinn, B. G. (1980)
"Order Determination for a Multivariate Autoregression"
Journal of the Royal Statistical Society B, 42, 182-185.
- Reinsel, G. (1976)
"A Note on Approximations for the Inverse of the Covariance Matrix of a Stationary Vector Process"
Technical Report N° 470, Department of Statistics, University of Wisconsin.
- Reinsel, G. (1980)
"Asymptotic Properties of Prediction Errors for the Multivariate Autoregressive Model Using Estimated Parameters"
Journal of the Royal Statistical Society B, 328-333.
- Reinsel, G. (1983)
"Some Results on Multivariate Autoregressive Index Models"
Biometrika 70, 145-156.
- Rissanen, J. e Caines, P. E. (1979)
"The Strong Consistency of Maximum Likelihood Estimators for ARMA Processes"
Annals of Statistics 7, 297-315.
- Ross, S. M. (1983)
"Stochastic Processes"
John Wiley and Sons, Nova Iorque.
- Samaranayake, V. A. e Hasza, D. P. (1987)
"The Asymptotic Properties of the Sample Autocorrelations for a Multiple Autoregressive Process with One Unit Root"
Journal of Time Series Analysis 8, 79-93.

- Samaranayake, V. A. e Hasza, D. P. (1988)
"Properties of Predictors for Multivariate Autoregressive Models with
Estimated Parameters"
Journal of Time Series Analysis 9, 361-383.
- Schwarz, G. (1978)
"Estimating the Dimension of a Model"
The Annals of Statistics 6, 461-464.
- Shea, B. L. (1987)
"Estimation of Multivariate Time Series"
Journal of Time Series Analysis 8, 95-109.
- Shea, B. L. (1988)
"A Note on the Generation of Independent Realizations of a Vector
Autoregressive-Moving Average Process"
Journal of Time Series Analysis 9, 403-410.
- Shea, B. L. (1989)
"The Exact Likelihood of a Vector Autoregressive Moving Average
Process"
Journal of the Royal Statistical Society C, 38, 161-184.
- Shibata, R. (1980)
"Asymptotically Efficient Selection of the Order of the Model for
Estimating Parameters of a Linear Process"
The Annals of Statistics 8, 147-164.
- Sims, C. A. (1972)
"Money, Income and Causality"
American Economic Review 62, 540-552.
- Solo, V. (1984)
"The Exact Likelihood for a Multivariate ARMA Model"
Journal of Time Series Analysis 14, 164-173.
- Spliid, H. (1983)
"A Fast Estimation Method for the Vector Autoregressive Moving
Average Model with Exogenous Variables"
Journal of the American Statistical Association 78, 843-849.

- Stam, A. e Hillmer, S. C. (1988)
 "Marginals of Multivariate First-Order Autoregressive Time Series Models"
 Journal of Time Series Analysis 9, 89-97.
- Stensholt, E. e Tjøstheim, D. (1981)
 "Factorizing Multivariate Time Series Operators"
 Journal of Multivariate Analysis 11, 244-249.
- Teräsvirta, T. (1985)
 "Mink and Muskrat Interaction : A Structural Analysis"
 Journal of Time Series Analysis 6, 171-180.
- Tiao, G. C. e Box, G. E. P. (1979)
 "An Introduction to Applied Multiple Time Series Analysis"
 Technical Report N° 582, Department of Statistics, University of Wisconsin.
- Tiao, G. C. e Box, G. E. P. (1981)
 "Modeling Multiple Time Series with Applications"
 Journal of the American Statistical Association 76, 802-816.
- Tiao, G. C. e Tsay, R. S. (1983)
 "Multiple Time Series Modeling and Extended Sample Cross-Correlations"
 Journal of Business and Economic Statistics 1, 43-56.
- Tjøstheim, D. (1981)
 "Granger Causality in Multiple Time Series"
 Journal of Econometrics 17, 157-176.
- Tjøstheim, D. e Paulsen, J. (1982)
 "Empirical Identification of Multiple Time Series"
 Journal of Time Series Analysis 3, 265-282.
- Troutman, B. M. (1979)
 "Some Results in Periodic Autoregressions"
 Biometrika 66, 219-228.

- Wahba, G. (1971)
"Some Tests of Independence for Stationary Multivariate Time Series"
Journal of the Royal Statistical Society B, 33, 153-167.
- Wallis, K. F. (1977)
"Multiple Time Series Analysis and the Final Form of Econometric Models"
Econometrica 45, 1481-1497.
- Whittle, P. (1953)
"The Analysis of Multiple Stationary Time Series"
Journal of the Royal Statistical Society B, 15, 125-139.
- Whittle, P. (1963)
"On the Fitting of Multivariate Autoregressions, and the Approximate Canonical Factorization of a Spectral Density Matrix"
Biometrika 50, 129-134.
- Wilson, G. T. (1973)
"The Estimation of Parameters in Multivariate Time Series Models"
Journal of the Royal Statistical Society B, 35, 76-85.
- Yamamoto, T. (1981)
"Predictions of Multivariate Autoregressive-Moving Average Models"
Biometrika 68, 485-492.
- Zellner, A. e Palm, F. (1974)
"Time Series Analysis and Simultaneous Equation Econometric Models"
Journal of Econometrics 2, 17-54.