Peter Hackl
(Editor)

Statistical Analysis and Forecasting of Economic Structural Change

Springer-Verlag

International Institute for
Applied Systems Analysis

Peter Hackl (Ed.)

Statistical Analysis and Forecasting of Economic Structural Change

With 98 Figures

Springer-Verlag Berlin Heidelberg New York
London Paris Tokyo Hong Kong

Professor Dr. Peter Hackl
Institut für Statistik
Wirtschaftsuniversität
Augasse 2-6
A-1090 Wien

ISBN 3-540-51454-6 Springer-Verlag Berlin Heidelberg New York Tokyo
ISBN 0-387-51454-6 Springer-Verlag New York Berlin Heidelberg Tokyo

Library of Congress Cataloging-in-Publication Data
Statistical analysis and forecasting of economic structural change / Peter Hackl (ed.).
Includes bibliographical references.
ISBN 0-387-51454-6
1. Economic forecasting --Statistical methods. I. Hackl, Peter.
HB3730.S735 1989
330'.01'51--dc20

Printed in Germany

Printing: Weihert-Druck GmbH, Darmstadt
Bookbinding: J. Schäffer GmbH u. Co. KG., Grünstadt
2142/7130-543210

Preface

In 1984, the University of Bonn (FRG) and the International Institute for Applied System Analysis (IIASA) in Laxenburg (Austria), created a joint research group to analyze the relationship between economic growth and structural change. The research team was to examine the commodity composition as well as the size and direction of commodity and credit flows among countries and regions. Krelle (1988) reports on the results of this "Bonn-IIASA" research project.

At the same time, an informal IIASA Working Group was initiated to deal with problems of the statistical analysis of economic data in the context of structural change: What tools do we have to identify nonconstancy of model parameters? What type of models are particularly applicable to nonconstant structure? How is forecasting affected by the presence of nonconstant structure? What problems should be anticipated in applying these tools and models? Some 50 experts, mainly statisticians or econometricians from about 15 countries, came together in Lodz, Poland (May 1985); Berlin, GDR (June 1986); and Sulejov, Poland (September 1986) to present and discuss their findings. This volume contains a selected set of those conference contributions as well as several specially invited chapters.

After a euphoric period in the 1960s, model builders in economics became aware of a need of model diagnostics and, in particular, for methods to detect — and cope with — structural changes. Statisticians' interest grew slowly after a few early contributions, especially R.E. Quandt's (1958) paper on switching regression. Related problems were discussed and adaptable methods were developed in areas such as probability theory (the change point problem), continuous sampling inspection (the CUSUM technique), or engineering (recursive estimation, filtering). Hackl and Westlund's (1985) bibliography, *Statistical Analysis of Structural Change*, contains some 300 titles; in a revised version to be published in 1989, the number of entries has increased by about 100 papers.

Nevertheless, in practical model building exercises, the methods recommended by statisticians seem not to be extensively used. Reasons for this might be the scarcity of corresponding computer programs and also the lack of a systematic survey. Broemeling and Tsurumi's (1986) book is the most comprehensive volume on the subject, in spite of the fact that it is based on a Bayesian paradigm. The same basis underlies a 1982 supplement of the *Journal of Econometrics*, edited by Broemeling, which also covers a broad area of the subject. Specialized books have been published on multiphase regression

(Schulze, 1986), spline function-based models (Poirier, 1976), and the analysis of residuals (Hackl, 1980). Several volumes contain special sections on model building, such as those by Broemeling (1985) and by Krämer and Sonnberger (1986). Accessible bibliographies have been compiled by Shaban (1980) and by Johnson (1977, 1980).

The present volume contains an introduction and three sections:

I. Introduction

The introductory chapter, "What can statistics contribute to the analysis of economic structural change?" by G.J. Anderson and G.E. Mizon, discusses not only the role of statistics in the detection and assimilation of structural changes, but also the relevance of respective methods in the evaluation of econometric models. Trends in the development of these methods are indicated, and the contributions to the present volume are put into a broader context of empirical economics to help to bridge the gap between economists and statisticians.

II. Identification of Structural Change

The chapters combined under this heading are concerned with the detection of parameter nonconstancy. The procedures discussed range from classical methods, such as the CUSUM test, to new concepts, particularly those based on nonparametric statistics. Several chapters assess the conditions under which these methods can be applied and their robustness under such conditions.

In "Testing for structural change in simultaneous equation models", A.C. Harvey and G.D.A. Phillips develop exact tests to detect changes in the coefficients of structural equations based upon k-class estimation. B. Schips and Y. Abrahamsen demonstrate in a Monte Corlo case study ("Specification and stability tests versus jackknifing: Some illustrative examples") the superiority of jackknifing-type criteria over classical inference statistics, given nonconstancy of linear model parameters. W. Krämer shows in Chapter 4, "The robustness of the Chow test to autocorrelation among disturbances", that independence of the error terms is crucial for deviations between true and nominal significance level.

P. Hackl and W. Katzenbeisser compare, in "Tests against nonconstancy in linear models based on counting statistics", various tests and conclude that some simple ones are strong competitors to the well-known CUSUM procedure. M. Hušková and P.K. Sen in "Nonparametric tests for shift and change in regression at an unknown time point", cover various nonparametric and robust tests available and suggest an adaptive procedure that fulfills an (asymptotic) optimality criterion. H. Tsurumi's "Detection of join point in regression models" reviews methods for estimating the join point and proposes a new procedure based on the mean squared errors of post-sample forecasts.

Tools to characterize parameter changes in linear regression models in various patterns of nonconstancy are studied by A.H. Westlund and B. Törnkvist in Chapter 8: "On the identification of time of structural changes by MOSUM-SQ and CUSUM-SQ procedures". Conditions for the CUSUM-SQ test to have nontrivial local power are given by W. Ploberger in "The local power of the CUSUM-SQ test against heteroscedasticity". J. Praagman's "Bahadur efficiency of tests for a shift in location of normal populations"

treats two generalized forms of the most relevant test statistics.

Finally, in Chapter 11, Z. Wasilewski demonstrates "The use of graphical displays in the analysis of structural change" for the investigation of regression residuals.

III. Model Building in the Presence of Structural Change

This section addresses models that are in some sense generalizations of constant-parameter models, so that they can assimilate structural changes.

In Chapter 12, "Adaptive estimation and structural change in regression and time series models", J. Ledolter reviews heuristic and model-based approaches to adaptive estimation of regression parameters and discusses in detail the case where the parameters follow ARMA processes. The use of exponential weights for adaptive estimation is treated in "An adaptive method of regression analysis" by Y.P. Lukashin. J. Dziechciarz, in "Changing and random coefficient models. A survey", reviews comprehensively the related literature (about 200 references).

P.M. Robinson discusses, in "Nonparametric estimation of time-varying parameters", the construction and properties of a kernel-based estimator of the regression coefficient. V.V. Fedorov, in "Latent variables in regression analysis", treats two types of regression models with unobservable variables, together with reformulations that can be handled by traditional regression analysis techniques.

L.D. Broemeling returns in Chapter 17, "Structural change and time series analysis", to demonstrate a Bayesian approach in analyzing a time series model for data so that the trend or the autocovariance function changes at an unknown time point. H. Tong, in "Thresholds, stability, nonlinear forecasting, and irregularly sampled data", examines threshold models, i.e., global models composed of submodels for areas delineated by thresholds. "Forecasting in situations of structural change: A general approach", by F.X. Diebold and P. Pauly, presents a method of combining forecasts to compensate for poor primary forecasts on the basis of time-varying weighting. J. Kleffe, in "Updating parameters of linear change point models", discusses an algorithm for efficiently updating the residual sum of squares applied in two-phase regression with a shifting change point.

IV. Data Analysis and Modeling

This section deals with real-life structural change situations.

P.K. Sen, in "Change point problem relating to the poverty structure", constructs and analyzes poverty indices, based on income distribution. T. Ozaki and V.H. Ozaki, in "Statistical identification of nonlinear dynamics in macroeconomics using nonlinear time series models", describe a model representing both Keynesian and monetarist viewpoints *vis à vis* the dynamics of the Hicksian IS–LM concept by a difference in model parameters. A. Keller's Chapter 23, "Econometrics of technical change: Techniques and problems", surveys studies concerning "technical progress" — an essential notion of economic growth with implications and pitfalls for data observations, model specifications, and estimation procedures.

On the basis of interest rates for Austria, W. Polasek demonstrates, in "Local autoregression models for detection of changes in causality", an approach to analyze nonstation-

arity by applying local stationary autoregressive processes.

Finally, in Chapter 25, J.-M. Dufour presents an empirical study, "Investment, taxation, and econometric policy evaluation: Some evidence of the Lucas critique", which discusses Lucas's arguments that parameters in econometric relationships reflect economic agents' decision rules.

It is hoped that this volume will be useful and stimulating to both statisticians and economists.

I wish to thank the former deputy project leader, Dr. Anatoly Smyshlyeav, now back in Moscow, the instigator of this project, for his encouragement and interest; Professor Wilhelm Krelle, for his help and patience; the scientists who contributed by participating in the workshops, by delivering chapters for this volume, or by acting as referees, for their efforts; and both the International Institute for Applied Systems Analysis and the Austrian Ministry of Science and Research (BMfWF) for their financial support.

Peter Hackl
Institut für Statistik,
Wirtschaftsuniversität Wien,
and IIASA, Laxenburg

References

Broemeling, L.D. (1985), *Bayesian Analysis of Linear Models.* New York: M. Dekker.

Broemeling, L.D. and Tsurumi, H. (1986), *Econometrics and Structural Change.* New York: M. Dekker.

Hackl, P. (1980), *Testing the Constancy of Regression Relationships over Time.* Göttingen: Vandenhoeck & Ruprecht.

Hackl, P. and Westlund, A.H. (1985), *Statistical Analysis of "Structural Change": An Annotated Bibliography.* Collaborative Paper CP-85-31. Laxenburg: International Institute for Applied Systems Analysis.

Johnson, L.W. (1977), Stochastic parameter regression: An annotated bibliography. *International Statistical Review*, **45**, 257-272.

Johnson, L.W. (1980), Stochastic parameter regression: An additional annotated bibliography. *International Statistical Review*, **48**, 95-102.

Krämer, W. and Sonnberger, H. (1986), *The Linear Regression Model under Test.* Heidelberg: Physica.

Krelle, W. (Ed.) (1989), *The Future of the World Economy: Economic Growth and Structural Change* (Results of the Bonn–IIASA Research Project). Berlin: Springer.

Poirier, D.J. (1976), *The Econometrics of Structural Change.* Amsterdam: North-Holland.

Quandt, R.E. (1958), The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association*, **53**, 873-880.

Schulze, U. (1986), *Mehrphasenregression: Stabilitätsprüfung, Schätzung, Hypothesenprüfung.* Berlin: Akademie-Verlag.

Shaban, S.A. (1980), Change point problem and two-phase regression: An annotated bibliography. *International Statistical Review*, **48**, 83-93.

Foreword

As Professor Hackl has already pointed out in his preface, this work on statistical identification of economic structural change was originally conceived as part of a common research project in which the other part was concerned with economic analysis and forecasting of economic growth and structural change. For reasons beyond our control, these projects had to be separated. This was unfortunate for several reasons. The most important one is that the results of the statistical project were to have been applied and tested in the economic project, and the practical problems encountered in the economic project were to have been analyzed and solved in the statistical project. Although this ideal arrangement was not possible, we strove to compensate for this loss. Within the Sonderforschungsbereich 303 (Special Research Unit 303) at Bonn University, three research projects have been carried out. The corresponding research reports are given by C. Weihs (1987) ("Auswirkungen von Fehlern in den Daten auf Parameterschätzungen und Prognosen", in K.-A. Scheffer *et al.*, Arbeiten zur Angewandten Statistik, Vol. 30. Heidelberg: Physica-Verlag); and by A. Kirchen (1988) ("Schätzung zeitveränderlicher Strukturparameter in ökonometrischen Prognosemodellen", Frankfurt/Main: Athenäum); the third project, performed by Mr. Körösi, on latent variable approaches, is not yet completed. We were able to use these results for our research.

The economic research team consisted of a central group at Bonn University, composed of five scholars in 1985 and 1986 and two scholars in the first half of 1987, and collaborating groups in almost all important world market countries and international institutions. The results of their work are contained in *The Future of the World Economy: Economic Growth and Structural Change*, edited by Wilhelm Krelle and published by Springer-Verlag.

In Part I of that book, the forecasts to the year 2000 resulting from the solution of an econometric world model are given. This work was carried out by the central group: W. Krelle, H. Ross, and H. Welsch (all from the FRG); R. Dobrinsky (Bulgaria); I. Székély (Hungary); and J. Gajda and J. Sztaudynger (Poland). In Part II of the book, these forecasts are collated with those made by the collaborating country groups: B. Hickman and P. Pauly (USA); S. Dubovsky and O. Eismont (USSR); C. Moriguchi (Japan); J. Donges, H. Klodt, K.-D. Schmidt, W. Krelle, and H. Sarrazin (FRG); J.-L. Brillet (France); M. Landesmann and A. Snell (UK); O. Panov and J. Djarova (Bulgaria); A. Czyzewski and W. Welfe (Poland); K. Zeman and I. Sujan (Czechoslovakia); A. Simon (Hungary); and S. Schleicher (Austria). In Part III, world trade and the debt situation are analyzed in more detail by G. Erber (FRG); P. Pauly (USA); and J. Richtering (UNCTAD). In Part

IV, some general and methodological problems related to economic growth and structural change are dealt with by S. Menshikov and K. Klimenko (USSR); G. d'Alcantara (Belgium); I. Tchijov (USSR); J. Ceska (Czechoslovakia); and R. Dobrinsky (Bulgaria).

It is hoped that the results of the statistical group, headed by Professor Hackl and published in this book, can be used to improve the results obtained by the economic group within the framework of a new approach to the problem of economic growth and structural change, and vice versa.

Wilhelm Krelle
Institut für Gesellschafts- u.
Wirtschaftswissenschaften,
Universität Bonn

Contents

Part I

Introduction

CHAPTER 1

What Can Statistics Contribute to the Analysis of Economic Structural Change?

Gordon J. Anderson and Grayham E. Mizon

Summary

The role of statistics in the detection and assimilation of structural change in econometric models is analyzed. Detection of structural change has been made much easier and more sophisticated by recent developments in graphical analysis and recursive estimation and testing techniques, particularly for use on microcomputers. A typology of models incorporating structural change is presented, and methods for discriminating between these models are considered. It is also argued that statistical tests for the hypothesis of structural constancy play an important role in the evaluation of econometric models. In addition, it is noted that major changes in the sample correlations between variables, rather than being a nuisance for econometric model builders, is in fact an important stimulus to model evaluation and improvement.

1.1 Introduction

Structural change is endemic in model building, and as such it is the role of statistical analysis to detect its presence, to find ways to assimilate it in models, and to find methods of statistical inference that are robust to its presence. In this chapter the contribution of statistics in these three areas is analyzed. However, it is noted that structural change has been both a concern and a fascination for economists.

Economists are concerned with structural change because, in their search for simple and readily interpretable models to capture economic fundamentals, they have been haunted by the spectrum of pernicious "shocks" to the economic system. Indeed, at times the spectre has been transformed into the chimera for which there is no Bellerophon: q.v., the Lucas critique (Lucas, 1976). The monster, when carefully and skillfully handled though, enables economists to evaluate the usefulness and durability of their models, thus helping them to discriminate between alternative models for the same phenomenon, and occasionally indicating ways in which models can be modified to extend their range of applicability.

For example, the oil price "shock" of 1972/1973 provided a major test that many economic models failed to pass, and yet the widespread model failure associated with this shock provided an important stimulus to model evaluation and improvement. Moreover, the fact that traditional models of the demand for money in the USA had succumbed to the structural changes associated with the period of "missing money" (Goldfeld, 1976) and that of the "great velocity decline", revealed the inadequacies of these models and encouraged researchers, such as Baba *et al.* (1987), to develop models that encompass previous ones and provide an adequate characterization of the observed data over a long historical period.

Hence, sudden and unexpected changes in the sample correlation structure between variables can be of great value in proving and improving models. Like Bellerophon, the econometrician can ride the Pegasus of skillful and careful modeling to overcome the challenge provided by the chimera of structural change, and in the process greatly enhance his reputation and rewards.

There is also a long history of economists being fascinated by structural change as a vehicle for analyzing how growth and prosperity can be introduced and maintained in the economy. By examining technical progress and other stimulants to growth, the economist hopes to initiate and control the evolution of structural change, rather than responding to the "shocks" and innovations of unanticipated structural change. Keller (Chapter 23 of this volume) provides an example of such research in economics. It is in the analysis of long-run historical data that the economist can attempt to model the major changes in the economy, which, despite the fact that they are often labeled revolutions, have been gradual adjustments to, and evolutions of, new technologies.

Similarly, the traditional treatment of taste changes and technical progress as simple functions of time are examples of structural change being incorporated in models. This type of structural change usually receives less attention in the econometrics literature, and will not be a major concern in this chapter. However, it should be noted that, in order to assimilate such changes in economic models, it is necessary to be able to identify the epochs associated with each structure or technology, and to have a deep enough understanding of the workings of the economic system to be able to model the transition between these epochs.

In attempting to achieve a deeper understanding of the nature of the economic system, economists employ models to represent the relationship between relevant variables. While

such models, of necessity, can only be approximate characterizations of the process that actually generated the data, the hope is that they capture the important features of the relationship, and that they do this with a constant structure. Clearly, models that have continuously, or frequently, changing structure, which is unpredictable, are of limited value. Hence, economists who seek to characterize in simple models aspects of the economic system, require there to be a large degree of constancy or stationarity, and yet we know that there are important changes in basic behavior from time to time. In the process of learning, and as a result of research and development activities, economic agents do change their production, consumption, and other patterns of behavior. There are also "shocks" induced by institutional changes, such as moves from fixed to floating exchange rates, and changes in government economic policy from faith in monetarist policies to reliance on more Keynesian policies. Whether these changes in economic behavior are evolutionary or "shocks", they are in principle ones that economists aim to represent in their models: though, by definition, "shocks" will be represented in models *post hoc*.

The complexity of the economic system, our limited knowledge of it, and the scarcity of data can result in the attendant inadequacies of econometric models manifesting themselves in apparent structural change. Indeed, one feature of macroeconometric models is the common occurrence of periods of predictive failure. For example, the Phillips curve, which had been a cornerstone of many of these models in the 1960s, exhibited extensive predictive failure in the more turbulent times of the 1970s and 1980s. Whether these model failures revealed important structural changes in the economy, or inadequacies of the models independent of any structural change, is a moot point. However, there are occasions when regime changes, in particular sectors of the economy that can be clearly identified *ex post* (though not easily predicted *ex ante*), result in predictive failure in models concerned with behavior sufficiently removed, or isolated, from those sectors that there was no reason to expect the model failure, either *ex ante* or *ex post*. Predictive failure in such cases is more likely to be the result of model misspecification (omitted variables, ignored nonlinearities, dynamic misspecification, invalid exogeneity assumptions etc.), than being directly attributable to a regime change. Hence, tests for structural change in econometric models are likely to be potent checks for model misspecification, even though tests of structural constancy *per se* will rarely identify the precise misspecification.

In the next section we present a view of econometric modeling, indicating the impact and role of structural change in it. Section 1.3 provides a typology of structural change and of the models and approaches used to represent it. Section 1.4 is concerned with alternative ways of testing for structural change, and discusses the role of such test statistics in the evaluation of econometric models. An important constituent in the statistical analysis of structural change is the determination of the number and location of regime break points: this is discussed in Section 1.5. Section 1.6 is concerned with alternative ways of assimilating structural change in econometric models. Section 1.7 summarizes the chapter and offers some conclusions.

1.2 A View of Econometric Modeling

Whatever process generates the data that are used by economists, it is likely to be characterized by nonlinearities, dynamic and simultaneous interactions, measurement errors, and other complications, which make the data generating process (DGP) unknown and unknowable, given the limited information available to economists. Hence, in econometric modeling we are not searching for the "truth", but rather trying to find statistically and economically sound models that are of value for the problem in hand, and that are congruent with the available information, which comes from four principal sources: (i) *a priori* theory; (ii) sample data; (iii) the measurement system; and (iv) rival models.

In discussing the way in which the information from these sources can be used in modeling, emphasis will be placed on parametric models, particularly those using time series data. Since structural change is essentially a time series phenomenon (though nonhomogeneities in models using cross-section data, particularly when they are indexed by an appropriate scale variable, such as income or household size, constitute an analogous phenomenon), and economic problems are usually representable in the framework of parametric (or at least semi-parametric) models, this is unlikely to be a serious limitation.

1.2.1 A class of models

Let the variables thought to be relevant for a particular problem, and for which T time series observations are available, be $\boldsymbol{x}(t)$. Typically, not all these variables will be precisely the variables from economic theory on which we would like observations. However, since this latent variable problem does not impinge critically on our analysis of structural change, we will ignore it. Further, we assume that the sequential stochastic process used in modeling the generation of the observations $X(1,T) = [\boldsymbol{x}(1)', \ldots, \boldsymbol{x}(T)']$ takes the form:

$$D[X(1,T)/X(0,0);\Psi] = \prod_{t=1}^{T} D[\boldsymbol{x}(t)/X(0,t-1);\Psi] \tag{1.1}$$

where $D(.)$ is a generic density function, and $X(0,0)$ is a matrix of known initial conditions.

The statistical analysis of structural change in this class of parametric models is concerned with the possibility that Ψ varies with t. In order for the statistical analysis to be feasible, we assume that there is sufficient parameter constancy (e.g., only one or two separate regimes, or that evolution can be represented by a low-dimensional parameterization), plus adequate *a priori* information, for Ψ to be finite-dimensional, identifiable, and to constitute a sufficient parameterization.

Letting $\boldsymbol{x}(t)'$ be partitioned into $[y(t)', z(t)']$, we note that $D(.)$ in (1.1) can be factored:

$$D[\boldsymbol{x}(t)|X(0,t-1);\Psi] = D[y(t)|z(t), X(0,t-1);\Theta]D[z(t)|X(0,t-1);\lambda] \tag{1.2}$$

and that economists are — typically — concerned with the conditional model for $y(t)$, i.e., $D[y(t)|z(t), X(0,t-1); \Theta]$, in which the variables $y(t)$ and $z(t)$ are treated as endogenous and exogenous variables, respectively. In order for it to be appropriate to ignore the information in the marginal model $D[z(t)|X(0,t-1); \lambda]$, $z(t)$ must be a vector of weakly exogenous variables for the parameters of interest Φ. This requires that Φ be functions of Θ only, and that Θ and λ be variation-free. Note that if invalid exogeneity assumptions are made, and there are structural changes in the marginal process $D[z(t)|X(0,t-1); \lambda]$, it is most likely that the parameters Θ of the conditional model will exhibit nonconstancy. Hence, the rejection of the hypothesis of parameter constancy in a conditional model can be the result of invalid exogeneity assumptions.

For the analysis of models in which structural change is allowed, the concept of weak exogeneity (which is appropriate for valid conditioning *within regimes*) can be extended to that of super exogeneity (which is appropriate for valid conditioning *across regimes*). We return to this point below, but note now that one of the objectives of econometric modeling is to find relatively *invariant* parameters Θ characterizing the conditional model, even though the parameters λ of the marginal model may be *transient* – see Engle *et al.* (1983).

1.2.2 Model evaluation

No matter from what origins, or by what modeling strategies, econometric models are developed, their relevance and potential value for the analysis of economic problems can only be determined by evaluating their congruence with the available information. The origin of a model *per se* is a determinant of research efficiency, not model validity – see Hendry and Mizon (1985). The relative merits of alternative econometric modeling strategies have been the subject of much recent discussion (see, e.g., Pagan, 1987), but the route by which a model is arrived at does not determine, though it can clearly influence, the usefulness of a model: the latter is determined by rigorous model evaluation. It is argued in Hendry and Richard (1982, 1983) and in Hendry and Mizon (1985) that rigorous evaluation of a model requires the testing of its congruence with all the available information.

Congruence with a priori theory

Congruence with *a priori* information requires a model to be consistent with *a priori* theory, usually economic theory. A major advantage of this requirement is that, in addition to being able to present the statistical credentials of a model, it will be possible to use economic theory to interpret the model and discuss its merits with other economists. Note, though, that although this could be achieved by entertaining only models that are theory-consistent, this will often result in models that are noncongruent with information from some of the other sources, and it may prevent the discovery of other models that are congruent with all other information and can be rendered theory-consistent. Alternatively, recognizing that models are human artifacts and, hence, susceptible to being "designed", we can use the requirement of *a priori* theory consistency as a model selection criterion.

Manifestly, theory used in this way is not tested. If it is desired to test a theory, then obviously that theory cannot be used as a model selection criterion. Furthermore, since a valid basis for conditional inference is required as a framework within which to test any theory, it is inappropriate not to question the model's congruence with information from sources other than *a priori* theory. For further arguments against confining attention solely to economic theory-based models, see Hendry and Mizon (1985).

Congruence with sample information

For a given sample of data, the choices of model specification and estimation technique determine the properties of the model's residuals. Since the residuals are the difference between the observed values of the endogenous variables and their within-sample model predictions, it is not surprising that detailed analysis of the properties of the residuals can be used to check a model's congruence with the sample data – see, for example, Pagan and Hall (1983). If it is found that the residuals are serially correlated, heteroscedastic, and correlated with variables in the modeler's information set but not currently included in the model, then there is evidence that the sample data contain potentially valuable information that the current model does not exploit.

Modifying a model until its residuals no longer exhibit systematic behavior, at least as judged by the particular diagnostic test statistics being employed, is one way to achieve congruence with the sample information. This strategy of designing a model to be congruent with the sample information is discussed in detail by Hendry and Richard (1982, 1983), who also classify diagnostic statistics according to whether they are powerful in testing a model's noncongruence with past, present, or future data. Tests for parameter constancy and predictive failure, when employed as model design criteria, are examples of tests for structural change being used to achieve congruence with future sample information.

Congruence with the measurement system

Another valuable source of information lies in the known properties of the data. This includes both the basic time series properties of the data, and the properties of the measurement system. In econometric modeling using time series data, an important benchmark against which to assess the performance of an econometric model is the best-fitting time series model for each endogenous variable: unless an econometric model performs at least as well as a pure time series model, there is clearly scope for improving it. Similarly, an econometric model should be capable of generating fitted values and forecasts for the endogenous variables that share the major charateristics of the data revealed by descriptive statistics, e.g., trend, cyclical, and seasonal components. In addition, if it is known that a variable cannot be negative (e.g., aggregate consumption), or that it has finite upper and lower bounds (e.g., % unemployment rate), a model of that variable that does not yield fitted values and forecasts that also have these properties is not congruent with the measurement system. In such situations, the choice of alternative functional forms, which will give fitted values and forecasts consistent with the properties of the measurement system, is a strategy that has potential for model improvement.

1.2.3 Robustness: new data

When econometric modeling proceeds along the lines outlined above, there is the possibility that the preferred model(s) are too finely tuned to the information contained in the economic theory, the sample data, and the measurement system. In particular, such models may lack robustness to minor changes in the information contained in any of these sources. The major way in which such potential model fragility can be tested is to check the performance of the model against new information, i.e., information not previously used in the development of the model. One source of new information is additional, or future, observations on the variables of the model. Such information can then be used to test the model for predictive failure or parameter nonconstancy. Hence, tests for structural constancy across the sample and prediction period can be genuine tests, rather than being an integral part of the model selection process, and as such they provide information on the robustness of the model to changes (in this case, extensions) in the data set. Of course, if the hypothesis of no-structural-change is rejected, and as a result the econometrician modifies the model until such time as there is no indication of structural change using the "new" (which, clearly, is no longer new) data, then this test also ceases to be a genuine test statistic and becomes subsumed in the collection of model selection criteria being used in designing the model.

1.2.4 Robustness: rival models

A second source of new information that is less susceptible, though not immune, to this difficulty is that contained in rival models. It is rarely the case that there is only one class of model available in the economics literature relevant for the analysis of a particular problem. Usually, the alternative or rival models involve different variables, different functional forms, different statistical distributions, or all three, and so it is important for the economics profession as a whole, as well essential to the evaluation of the robustness of any particular model, that rigorous comparison of the performance of rival models be undertaken, rather than allowing a collection of alternative models (which may have very different policy implications) for the same phenomenon to coexist.

One of the most persuasive ways to establish the credentials of a model is to demonstrate that it is robust to minor changes in specification, and that it performs at least as well as any other available model. This latter property, that of requiring a model to encompass its rivals, when added to the list of criteria that an adequate model should satisfy, imposes a high performance standard for any model. By the same token, any encompassing model is an impressive one. In particular, the adoption of the encompassing principle as a part of a modeling strategy helps to ensure that:

1. Econometric modeling is progressive in that useful models are discarded only in favor of those that inferentially dominate them.

2. Alternative models for the same phenomenon do not coexist without having their relative merits compared.

3. New evidence is used to evaluate models "outside" of their design.

4. "Later" models account for the properties and findings of "earlier" models.

For further discussion of the nature and role of the encompassing principle in econometric modeling, see Mizon (1984) and Mizon and Richard (1986).

1.2.5 Model congruence

Having argued that necessary conditions for a model to be adequate for statistical and econometric purposes are that the model be congruent with information from four main sources, it is relevant to consider how model congruence can be achieved. We must immediately emphasize that there is no unique way to find congruent models; to quote Gilbert (1986), "Scientific discovery is necessarily an innovative and imaginative process, and cannot be automated". The value of models is determined solely by model evaluation, not by the route followed in their discovery. Discussion of the merits of alternative modeling strategies, as opposed to the list of necessary conditions for model congruence, is concerned with research efficiency, not model validity.

We note that all models are human artifacts, some of which are well designed, and others not. However, even well designed models are no more than models, and in particular are not the DGP. Furthermore, although the DGP, were it known, would provide a valid framework within which to evaluate and compare models, this is not the case for all models. Indeed, it is well known that invalid and misleading inferences can result from testing hypotheses in the context of a misspecified model, e.g., a linear regression model with serially correlated or heteroscedastic errors, or both. An illustration of this point, clearly discussed by Spanos (1987), is the estimation of a simple linear regression model such as:

$$y(t) = a + bx(t) + u(t) \tag{1.3}$$

It is well known that use of a standard OLS computer program to estimate a and b in this model will yield incorrect standard errors and, hence, invalid t-ratios, if the errors are serially correlated. What is less well appreciated is that, if the serial correlation in the errors is not generated by an autoregressive process [i.e., the common factor restrictions are not valid – see Hendry and Mizon (1978)], then the OLS estimators of a and b also will be inconsistent! Hence, as a first step in econometric modeling, it seems appropriate to determine a valid basis for conditional inference, subject to the limitations imposed by the quantity and quality of the available information.

Noting that relatively general models are more likely to provide such a valid basis for inference than very specific and simple models, especially since the consequence of overparameterization is inefficiency rather than inconsistency, a general-to-specific modeling strategy seems appropriate. Such a strategy would first isolate a statistically adequate general model using tests of misspecification, and then employ specification tests to determine parsimonious models that remain congruent with the available information. For

more discussion of the distincton between misspecification and specification tests, and their use in a general-to-specific modeling strategy, see Mizon (1977) and Hendry (1983).

Of particular relevance in the present context is the fact that tests for parameter constancy and predictive failure have an extremely important role as tests of model misspecification. Since the power properties of the test statistics employed for this purpose will depend on the particular forms of structural change that are being considered, we now discuss the major types of parameter nonconstancy.

1.3 Models of Parametric Structural Change

The introduction of structural change into the conditional model $D[y(t)|z(t), X(0,t-1);\Theta]$ requires the specification of the way in which Θ varies with t. One obvious possibility is that elements of Θ vary seasonally or as a function of trend. The inclusion of $(0,1)$ seasonal dummy variables in linear models represents structural change via seasonal changes in the intercept. These changes in the intercept are deterministic and season-specific. Deterministic changes in the intercept that are evolutionary, rather than season-specific, can be captured by the introduction of polynomials in t. The introduction of fixed effects into models using panel data is a similar representation of deterministic differences in parametric structure. A more extreme form of seasonal difference in structure is accommodated by the inclusion of interactive seasonal dummy variables, which then allows each season to have a different structure. However, structural changes that can be characterized in this way are regular and constant, and as such are less interesting than irregular economic structural changes that involve an element of surprise, at least *ex ante*. However, these comments on seasonally varying structure do suggest that other forms of parametric structural change might be representable via the introduction of specific effect variables, such as dummy variables and random effect variables.

A whole class of models in this category is that associated with the combination of cross-section and time series data, including panel data. Linear regression models of this form are given by:

$$\begin{aligned} y_{it} &= x_{it}' b_{it} + u_{it} \\ b_{it} &= b^* + v_i + w_t \end{aligned} \tag{1.4}$$

for $i = 1,\ldots,N$ and $t = 1,\ldots,T$, where x_{it}, b_{it}, v_i and w_t are $k \times 1$ vectors, which can be random or deterministic (fixed); b^* is a constant $k \times 1$ vector. When the changes in the regression coefficients are deterministic, analysis in these models can be undertaken by using individual specific and time-specific dummy variables for v_i and w_t, respectively, and treating the regression coefficients b_{it} as constants that vary deterministically over i and t. When v_i and w_t are random variables, restrictions on the covariance structure of v_i, w_{it} and u_{it} are required. For a summary of the literature concerning these models, including a table (Table 8.1 on page 326) listing the many types of model in this category,

see Judge *et al.* (1980). Chapter 9 in that same source also contains a discussion of regression models for either cross-section or time series data with coefficients that vary randomly around a constant mean.

In classifying parametric representations of economic structural change, a distinction can be made between, on the one hand, deterministic and stochastic changes, and, on the other, event-specific, epoch-specific, regime-specific, and evolutionary changes. Event-specific changes in structure, which are typically deterministic changes, arise when there is a shift in parameters (possibly only the intercept in a linear regression model) of short duration, after which the parameters revert to their original values. For example, a correctly anticipated (or preannounced) tax change might induce economic agents to bring forward (or delay) by one or two periods purchases of particular goods. Such a change can be represented by the inclusion of a dummy variable of the form $(0,\ldots,0,+1,-1,0,\ldots,0)$, which has the effect of bringing forward one period the dependent variable's partial response. Another example would be provided by economic agents changing their behavior during a period of war or during a strike, but then reverting back to their normal behavior afterward. The inclusion of a dummy variable, which takes the value 1 during the war or the strike and 0 elsewhere, in isolation or interacting with other variables in the model, will capture this change.

Epoch-specific changes are also typically deterministic changes, but of longer duration. Models of epoch-specific change rely on the identification of at least two regimes, for which there is constant structure within a regime, but structural change across regimes. Recall that, in the context of the factorization in (1.2), if there are regime changes in the process generating the conditioning variables $z(t)$ – i.e., λ changes between regimes – the parameters of the conditional model $D[y(t)/z(t), X(0,t-1);\Theta]$ are invariant to the regime changes if $z(t)$ are super exogenous variables, but they will usually change if $z(t)$ are weakly exogenous.

Richard (1980) analyzes models in which there are regime changes and changes in the exogeneity status of the $z(t)$ variables. Deterministic regime changes in the conditional model, whether induced by structural changes in the marginal process generating the weakly exogenous variables $z(t)$, or a direct consequence of structural change in the conditional model, can be accommodated by the inclusion of regime shift dummy variables (see Salkever, 1976) and can be tested for using analysis of variance test statistics or Chow's (1960) prediction test statistics. Note, though, since it is possible that regime changes in the conditional model could result from invalidly conditioning on $z(t)$, especially when there are structural changes in the process generating $z(t)$, it will not always be appropriate to accommodate the structural change, rather than regarding it as an indication of model misspecification.

In the context of linear regression models, a further class of deterministic epoch-specific structural changes is generated by restricting the conditional mean of the regressand evaluated at each join point between adjacent regimes to be equal for both structures. Models employing spline functions (see Poirier, 1976) are a particular example of this class of piecewise regression models with join points. It is also possible to have models with separate regimes, without each regime having observations that are sequential in time. Such

models are better described as having regime-specific, rather than epoch-specific structural change. For models of regime-specific structural change it is important to specify the process by which observations are allocated to the different regimes. In contrast to epoch-specific structural change for which the allocation is typically deterministic, regime-specific structural change can have deterministic or stochastic allocation mechanisms – see Goldfeld and Quandt (1973) for a survey of switching regime models.

The final class of parametric models used to represent structural change captures evolutionary structural change. One of the main types of model in this class is the so called "time-varying parameters" model. This is an unfortunate term because parameters are constant by definition. However, the name does evoke the idea that some of the characteristics of a model, which in the absence of evolutionary structural change would be parameters, can be described by time-varying processes.

Though the models described above in (1.5) have coefficients that vary randomly through time when w_t is random, the process specified for these variations is almost invariably stationary. Indeed, (1.5) with v_i deterministic (often zero) and with w_t being a zero mean, constant covariance, serially independent, normal variate yields the random coefficient model. Similarly, if $w_t = inv\{A(L)\}B(L)e_t$ when $A(L)$ and $B(L)$ are matrix polynomials of orders p and q, respectively, in the lag operator L, $inv\{.\}$ is the inverse operator, and e_t is a normally, independently, and identically distributed random vector, then (1.5) gives the stationary-coefficient regression model. For further discussion of models in this class, see Chow (1984).

However, models in which coefficients are generated by stationary stochastic processes, while they are capable of representing the effects of evolutionary structural change, offer little scope for the incorporation of *a priori* information from economic analysis of the nature of structural change. Time-varying coefficient models, on the other hand, typically employ nonstationary stochastic processes, e.g., coefficients following a random walk. The impact of assuming that the coefficients follow a nonstationary process, especially when $A(L)$ has unit roots, is to introduce "persistence" in the process of "shocks" to the system. Insofar as models are being used to represent economic structural change, which presumably is of more interest if the change in structure persists, this class of model is of potential value. Many of the above models capable of describing evolutionary structural change can be cast in the framework of the Kalman filter, which then offers efficient computational procedures for statistical inference; for more details, see Chow (1984) and Judge *et al.* (1980).

Most of the above discussion was concerned with the linear regression model, reflecting the emphasis on this model in the econometrics literature on the treatment of structural change. However, tests for epoch-specific structural change in linear and nonlinear simultaneous equations models have been proposed, e.g., Anderson and Mizon (1983), and employed by, among others, Carr-Hill and Stern (1974). Parameter constancy tests and the estimation of single equations from linear simultaneous equations systems have been analyzed by Barten and Bronsard (1970). Prediction tests have been proposed, for equations estimated by instrumental variables, by Lo and Newey (1985), and for limited dependent variable models, by Anderson (1987). Furthermore, Andrews and Fair (1987)

have presented a framework for analyzing epoch-specific structural change, with known break points, in nonlinear dynamic simultaneous equations models with errors possibly generated by dependent and heteroscedastic processes.

The major concern in the models discussed above has been with parametric representations of changes in the conditional mean. However, structural change can affect the second moments of variables as well as the first moments. Linear models in which epoch-specific structural change induces changes in the unconditional error variances have long been considered. A more recently developed class of models involves changes in the conditional error variances, and these also offer scope for econometric analysis of the effects of economic structural change. The ARCH model of Engle (1982), and the many closely related extensions of it, provide a rich class of models capable of capturing the effects of volatility changes.

In this section we have mentioned a range of parametric models that are capable of being used to characterize economic structural change; and though this range is wide, it is by no means exhaustive. In addition, the class of models of potential relevance in modeling economic structural change can be further widened by considering nonparameteric and semi-parametric models. Though we do not discuss nonparametric approaches, we note that Robinson (Chapter 15 in this volume) illustrates their value in the context of structural change. No matter what class of models is considered, statistical inference can assist the econometrician in detecting the presence of structural change, locating the breakpoints between regimes; and assimilating economic structural change into econometric models and discriminating between alternative representation of it. In the next three sections, we briefly consider each of these contributions of statistics to the analysis of economic structural change.

1.4 Detecting the Presence of Structural Change: Misspecification Tests

One of the most important requirements for statistical and econometric modeling is the existence of relatively constant relationships between variables. In econometric modeling the aim is usually to represent these statistical regularities parametrically in marginal propensities and elasticities. An important check on model adequacy is therefore a test of the hypothesis of parameter constancy. Indeed, some of the most commonly used methods for detecting structural change are those associated with testing hypotheses of parameter constancy and absence of predictive failure. The second group of commonly used methods for detecting structural change are those employing recursive estimation and associated graphical display techniques. Finally, nonparametric methods might also be used to detect structural change in models.

Probably the best-known test statistic for the hypothesis of parameter constancy is the analysis of variance F test statistic. This test statistic is an optimal one, in the context of the normal linear regression model, for testing constancy of the regression coefficients across two or more regimes, each of which has enough data points to allow reliable es-

timation of the regression coefficients, conditional on the error variances being constant across regimes – see Chow (1960). It is therefore sensible to test the hypothesis that the error variances are constant before testing the hypothesis of regression coefficient constancy. Indeed, the fact that the variance ratio and the analysis of variance test statistics are statistically independent (see Phillips and McCabe, 1983) means that it is possible to control the probability of Type I error for this testing procedure. In situations where there are insufficient observations to allow separate estimation of the model parameters for each regime, Chow (1960) proposed an alternative test statistic that is optimal for the hypothesis that there is no change in the conditional mean of the dependent variable across the regimes. This test statistic, therefore, is effectively a test for the absence of predictive failure, and has different power properties from the analysis of variance test statistic; on this point, see Rea (1978), Anderson and Mizon (1983), and Breusch (1986).

In the 1970s considerable attention was given to the analysis of the robustness (or lack of it) of these test statistics to heteroscedasticity and serial correlation in the error terms. Further research in this vein is contained in Krämer (Chapter 4 of this volume). Generalizations of these three test statistics applicable in dynamic linear or nonlinear simultaneous equations models are given in Anderson and Mizon (1983); and for an even more general context, in Andrews and Fair (1987). It is to be expected, though, that the more widely applicable test statistics of Andrews and Fair (1987) will lack power against simpler and more specific alternatives where it is *known* that the disturbances are white noise. Hence, a case can be made for the use of test statistics of differing degrees of robustness, particularly when these test statistics are being used to design models to be congruent with the available sample information. This point is especially relevant when the tests for parameter constancy are being used as tests of model misspecification.

As mentioned above, one important situation in which model misspecification can induce apparent structural change, is when variables are invalidly treated as exogenous in a model. It is therefore extremely important to test the validity of exogeneity assumptions. Unfortunately, the direct testing of exogeneity assumptions requires the specification of the *joint* process generating the endogenous and exogenous variables, and this removes the advantage of being able to analyze the "partial" likelihood rather than the likelihood for the joint model. Hence, not only is it important to test the validity of exogeneity assumptions since invalid ones can induce parameter nonconstancy, but also statistics designed to test hypotheses of parameter constancy are themselves important indirect tests of valid conditioning. Lubrano *et al.* (1986) provide an example in which there is evidence of changes in the exogeneity status of the interest rate, and of structural change associated with competition and credit, in a UK money demand function.

Invalid conditioning is not the only model misspecification that can induce parameter nonconstancy. The choice of an inappropriate functional form for a model, and the closely related misspecification of omitting relevant variables from the model, can result in apparent structural change. An example of the choice of an inappropriate functional form for a model, which has received much attention recently, is the estimation of a relationship between nonstationary variables that are not cointegrated – see Engle and Granger (1987).

Model fragility, arising from a model being too finely tuned to fit a particular sample of data, can also be detected by tests of parameter constancy across the sample data and "new" data not used in designing the model.

The statistics typically used for testing parameter constancy and the absence of predictive failure, particularly when used as a part of the diagnostic checking of model adequacy, require the potential break points to be nominated by the researcher. In the case of *ex post* analysis of the impact of events, such as the oil price "shock" in 1972–1973, the choice of break point should provide no difficulty. However, when there are no obvious discontinuities, the selection of arbitrary break points could produce misleading information about a model's properties. Hence, there is an important role for statistics in detecting potential break points, and in identifying separate regimes when there is structural change.

1.5 Detecting Breakpoints

Inspection of the graphs of variables plotted against time is a well-known, and obvious, method for spotting structural breaks. Indeed, it was argued above in Section 1.2 that an important part of model evaluation is checking whether the model is congruent with the known time series properties of the data. The recent development of econometric software for use on personal computers (which, in turn, have become more powerful and relatively less expensive) has, by incorporating powerful and easy-to-use graphing options, added significantly to the tools available to the econometrician. The routine inspection of the graphs of the major variables involved in a modeling exercise should provide valuable information about trend, seasonal, and cyclical behavior, as well as identifying potential structural breaks.

Wasilewski (Chapter 11 in this volume) discusses the use of graphical methods in the detection of structural breaks. Armed with this information, it should be easier for economists to design models that are congruent with the sample data. However, in using graphical information, it must be remembered that it is the relationship between variables, rather than their univariate properties alone, that is relevant for econometric models. While it is possible to plot variables against each other, this is limited to two dimensions, and ultimately it is analysis of the joint distribution, or the distribution of the endogenous variables conditional on the predetermined variables, that is relevant.

A very powerful way to identify structural breaks in the context of the relationships between variables, rather than relying solely on univariate graphical analysis, is to adopt recursive estimation techniques and to analyze recursive residuals. Ever since the seminal paper by Brown *et al.* (1975), the value of computing recursive residuals and analyzing them using CUSUM and related test statistics has been increasingly appreciated, so much so that recursive estimation is now a main estimation technique in a number of econometrics programs. Indeed a number of chapters in this volume are concerned with the analysis of recursive residuals – see Dufour (Chapter 25), Hackl and Katzenbeisser (Chapter 5), Ploberger (Chapter 9), and Westlund and Törnkvist (Chapter 8). An alternative, but closely related, technique for identifying structural breaks is based on the behavior of the

mean squared error of post-sample forecasts, and Tsurumi (Chapter 7) illustrates its use. Recursive as well as nonrecursive methods have been used in the analysis of the properties of nonparametric tests to determine break points by Hušková and Sen (Chapter 6). These techniques provide a striking example of what statistics can contribute to the analysis of structural change.

In the absence of *a priori* information to identify points of structural change, statistical analysis to determine the location of break points *ex post* can be invaluable. However, once potential break points have been isolated, it has to be decided whether these are genuine structural turning points or the result of model misspecification. In the former case, the model has to be modified to assimilate the structural change; and as the wide range of models mentioned in Section 1.3 makes clear, there are many alternative ways of modeling structural change. Hence, we now turn to the problem of model choice and discrimination.

1.6 Assimilating Structural Change

In attempting to model structural change, it is essential to decide what type of change is involved: event-specific, epoch-specific, regime-specific, or evolutionary. It is also important to decide whether the structural change affects all the parameters of the model, just the parameters involved in the conditional means of the endogenous variables, or just the parameters of the unconditional second moments. There will also be cases where the structural change is in the conditional second moments, as in ARCH processes.

These decisions are of the utmost importance because the range of possible models for representing structural change is enormous, as the discussion in Section 1.3 indicates. Hence, the use of *a priori* information concerning known institutional or market changes, plus any information about the nature of the change gained in using statistical techniques to detect structural change, will greatly facilitate the choice of an appropriate type of model to assimilate the structural change. However, even with the limitations on the range of model to be considered induced by using knowledge of the general type and location in time of the structural change, there will be considerable scope for choosing alternative models. In line with the view of modeling outlined in Section 1.2, it is therefore recommended that a general model of the chosen type (i.e., event-, epoch-, or regime-specific, or evolutionary) be adopted, and that this model be rigorously evaluated using tests of misspecification to check its congruence with the sample data.

Once a general model, in the appropriate class, has been found that has residuals that do not exhibit serial correlation, heteroscedasticity, ARCH effects, excess skewness, or excess kurtosis, and which appears to have a constant structure (though, of course, part of this structure is describing changes in basic features of interest to the economist, such as marginal propensities and elasticities), it is then appropriate to use it as basis for conditional inference. In particular, by having found a statistically adequate general model, it is then possible to use significance test statistics to identify statistically acceptable simplifications of the general model.

At this stage, the econometrician should have a model characterizing structural change of the appropriate type, which is also congruent with the information from three sources: *a priori* theory, the properties of the data and the measurement system, and the sample data. Tempting though it may be, this is not the time to relax. It has to be recognized that as a part of the modeling exercise many decisions have been taken, both on the class of model to consider in the first place, and on how to modify models when they are revealed to be inadequate. At each of these decision points, different decisions could have been taken and, despite this, a different model congruent with information from these same three sources could have been found. It is therefore of paramount importance that any model found to be congruent with these three sources of information be rigorously evaluated in terms of its ability to encompass rival models. This can be done using Cox-type one degree of freedom nonnested test statistics, the complete parametric-encompassing test statistic (if degrees of freedom permit), and the forecast-encompassing test statistics, for example. More details on each of these test statistics, including discussion of their relative merits, is contained in Mizon (1984), Mizon and Richard (1986), and Chong and Hendry (1986).

In connection with forecast-encompassing tests, we note that if it is found necessary to combine the forecasts of several models in order to get improved forecasts, and in particular to obtain forecasts that are less susceptible to structural change, this is *prima facie* evidence that each model is inadequate, and that none of them is capable of forecast-encompassing the others. If the purpose in modeling is forecasting, as is the case in Diebold and Pauly (Chapter 19 in this volume), then forecast combination may be the sensible way to proceed. In general, though, when it is found necessary to combine forecasts, it will be preferable to look for a "new" model, designed to capture the good features of the set of models being considered, so that it will then forecast-encompass the previous models. When possible, it is better to "combine" models, rather than combining forecasts. A "combined" or encompassing model should then dominate in forecasting performance, as well as being an improved model available for more general inferences.

1.7 Summary and Conclusions

Statistics plays an important role in the analysis of economic structural change. Ignored structural change and other forms of model misspecification induce predictive failure and parameter nonconstancy in models. Thus, statistical tests for the absence of these phenomena are important constituents in the evaluation of model performance. The potency of prediction tests in revealing model inadequacies is well illustrated by the large number of simple empirical implementations of economic laws that have performed badly outside narrow samples, or when there is a change in the sample correlations between variables that are not expected to have a major impact. The widespread predictive failures of money demand equations and Phillips curve models serve as examples. Statistical tests for the null hypothesis of no-structural-change, which have power against a wide range of alternative hypotheses involving structural change, are essential in econometric modeling.

Statistical methods for detecting the presence of different regimes, the number of regimes, and their location in time, are also major aids in the analysis of economic structural change. While *a priori* knowledge of institutional changes, and of economic theory associated with change, will always have a crucial role in the analysis of structural change, very often it is statistical analysis using recursive estimation and the Kalman filter that identifies structural change. Univariate and bivariate graphical methods have also proved valuable, though care must be exercised in interpreting information produced by such procedures, especially when it is used to infer multivariate relationships.

Statistics has also provided an extremely rich collection of models capable of representing structural change. These models range, on the one hand, from those incorporating deterministic shift dummy variables, to the random coefficient, stationary coefficient, and time-varying coefficient models. This diversity of models capable of characterizing structural change means that there is unlikely to be a single, uniquely best approach for modeling a particular economic phenomenon within structural change, and so statistics again must provide methods for comparing the properties of alternative models.

Finally, it might be thought that "shocks" to the economic system, changes in institutional arrangements, and changes in government economic policies are all irritants to the economic modeler. After all, such changes cause econometric models to fail! On the contrary, these are precisely the changes that are required in order to "prove" models; and by helping to identify model inadequacies, they are also instrumental in model improvement. For, as argued in Hendry and Mizon (1985), econometric evaluation is essentially a destructive activity — hazarding models to failure — but it is destruction with a constructive purpose. New and improved models are typically sought and discovered when existing models are plagued with changing structures.

References

Anderson, G.J. (1987), Prediction tests in limited dependent variable models: A note. *Journal of Econometrics*, **34**, 253-261.

Anderson, G.J. and Mizon, G.E. (1983), Parameter constancy tests: Old and new. Discussion Paper in Economics and Econometrics No.8325. Economics Department, University of Southampton, UK.

Andrews, D.W.K. and Fair, R.C. (1987), Inference in models with structural change. Cowles Foundation, Yale University, New Haven, CT.

Baba, Y., Hendry, D.F., and Starr, R.M. (1987), US money demand 1960-1984. Discussion Paper, University of California, San Diego, CA.

Barten, A.P. and Bronsard, L.S. (1970), Two-stage least squares estimation with shifts in the structural form. *Econometrica*, **38**, 938-941.

Breusch, T.S. (1986), Hypothesis testing in unidentified models. *Review of Economic Studies*, **53**, 635-651.

Brown, R.L., Durbin, J., and Evans, J.M. (1975), Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society*, **B-37**, 149-192.

Carr-Hill, R.A. and Stern, N.H. (1974), An econometric model of the supply and control of offences in England and Wales. *Journal of Public Economics*, **2**, 289-318.

Chong, Y.Y. and Hendry, D.F. (1986), Econometric evaluation of linear macroeconomic models. *Review of Economic Studies*, **53**, 671-690.

Chow, G.C. (1960), Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, **28**, 591-605.

Chow, G.C. (1984), Random and changing coefficient models, Ch. 21 in: Z. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics*, Volume 2. Amsterdam: North-Holland.

Engle, R.F. (1982), Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica*, **38**, 410-421.

Engle, R.F. and Granger, C.W.J. (1987), Cointegration and error correction: representation, estimation and testing. *Econometrica*, **55**, 251-276.

Engle, R.F., Hendry, D.F., and Richard, J.-F. (1983), Exogeneity. *Econometrica*, **51**, 277-304.

Gilbert, C.L. (1986), Professor Hendry's econometric methodology. *Oxford Bulletin of Economics and Statistics*, **48** (3), 283-306.

Goldfeld, S.M. (1976), The case of the missing money. *Brookings Papers in Economic Activity*, **3**, 683-730.

Goldfeld, S.M. and Quandt, R.E. (1973), The estimation of structural shifts by switching regressions. *Annals of Economic and Social Measurement*, **2**, 475-485.

Hendry, D.F. (1983), Econometric modelling: The consumption function in retrospect. *Scottish Journal of Political Economy*, **30**, 193-220.

Hendry, D.F. and Mizon, G.E. (1978), Serial correlation as a convenient simplification not a nuisance: A comment on a study of the demand for money by the Bank of England. *Economic Journal*, **88**, 549-563.

Hendry, D.F. and Mizon, G.E. (1985), Procrustean econometrics: Or stretching and squeezing data. Discussion Paper No. 68, *Centre for Economic Policy Research*, London, UK. (ISSN 0265-8003).

Hendry, D.F. and Richard, J.-F. (1982), On the formulation of empirical models in dynamic econometrics. *Journal of Econometrics*, **20**, 3-33.

Hendry, D.F. and Richard, J.-F. (1983), The econometric analysis of economic time series. *International Statistical Review*, **51**, 111-163.

Judge, G.G., Griffiths, W.E., Hill, R.C., and Lee, T.C. (1980), *The Theory and Practice of Econometrics*, New York: John Wiley.

Lo, A.W. and Newey, W.K. (1985), A large sample Chow test for the linear simultaneous equation. *Economics Letters*, **18**, 351-353.

Lubrano, M., Pierse, R.G., and Richard, J.-F. (1986), Stability of a UK money demand equation: A Bayesian approach to testing exogeneity. *Review of Economic Studies*, **53**, 603-634.

Lucas, R.E. (1976), Econometric policy evaluation: A critique, pp. 19-46 in: K. Brunner and A.H. Meltzer (eds.), *The Phillips Curve and Labor Markets*, New York: Carnegie-Rochester Conference on Public Policy 1.

Mizon, G.E. (1977), Inferential procedures in nonlinear models: An application in a UK cross section study of factor substitution and returns to scale. *Econometrica*, **45**, 1221-1242.

Mizon, G.E. (1984), The encompassing approach in econometrics, pp. 135-172 in: D.F. Hendry and K.F. Wallis (eds.), *Econometrics and Quantitative Economics*. Oxford: Basil Blackwell.

Mizon, G.E. and Richard, J.-F. (1986), The encompassing principle and its application to nonnested hypotheses. *Econometrica*, **54**, 657-678.

Pagan, A.R. (1987), Three econometric methodologies: a critical appraisal. *Journal of Economic Surveys*, **1** (1), 3-24.

Pagan, A.R. and Hall, A.D. (1983), Diagnostic tests as residual analysis. *Econometric Reviews*, **2** (2), 159-218.

Phillips, G.D.A. and McCabe, B.P. (1983), The independence of structural change tests in regression models. *Economics Letters*, **12**, 283-287.

Poirier, D.J. (1976), *The Economics of Structural Change.* Amsterdam: North-Holland.
Rea, J.D. (1978), Indeterminacy of the Chow test when the number of observations is insufficient. *Econometrica*, **46**, 229.
Richard, J-F. (1980), Models with several regimes and changes in exogeneity. *The Review of Economic Studies*, **47**, 1-20.
Salkever, D.S. (1976), The use of dummy variables to compute predictions, prediction errors, and confidence intervals. *Journal of Econometrics*, **4**, 393-397.
Spanos, A. (1987), On the statistical implications of modelling the error term: common factors, Granger causality and unit roots. Discussion Paper No.87/9, Birkbeck College, University of London.

Part II

Identification of Structural Change

CHAPTER 2

Testing for Structural Change in Simultaneous Equation Models

Andrew C. Harvey and Garry D.A. Phillips

Summary

Tests for changes in the coefficients of linear regression models, particularly the analysis of covariance and the Chow tests, are well known to econometricians and they are widely used. This paper demonstrates that analogous tests can also be constructed in static simultaneous equation models when equations are estimated by common k-class estimators, e.g., OLS, 2SLS, and LIML. The tests are based on the residuals obtained when the estimated endogenous part of a simultaneous equation is regressed on all the exogenous variables in the system. The tests have many of the characteristics of the regression based tests although the nature of the residuals used makes it more difficult to analyse their power properties.

2.1 Introduction

Regression analysis based upon time series data usually proceeds under the assumption that the regression relationship remains constant over time. Sometimes, particularly in economic applications, this assumption is open to question, and it may be necessary to examine for changes in the regression coefficients.

Tests for changes in the regression coefficients have been developed in single-equation models and are frequently used by econometricians. In the case where sufficient observations are available both before and after the suspected change to enable estimates of the regresssion coefficients to be made, the apropriate test is the analysis of covariance

test. In the case where there are insufficient observations to compute a second regression the Chow test can be used. These tests are discussed in Chow (1960) and Fisher (1970).

There has not been a corresponding development of tests for changes in the coefficients of structural equations in simultaneous-equations models, although tests have been developed for other misspecifications. Durbin (1957) showed that the bounds test for serial correlation is available, while Bouman (1971) and Harvey and Phillips (1980) showed how to obtain exact tests. In each case the tests are based on residuals from a regression of the estimated endogenous part of the structural equation on all the exogenous variables of the model. Harvey and Phillips (1981) extended their earlier work to develop exact tests for heteroscedasticity as well.

A feature of the Harvey and Phillips work is the demonstration that, when the endogenous part of the structural equation is estimated by two stage-least squares, the theoretical development of exact test procedures is particularly straightforward. Following their work, Giles (1981) proposed CUSUM and CUSUM-SQ tests, and Erlat (1983) discussed a Chow test. In both sources the proposed tests were based upon two-stage least squares estimation.

In this chapter we develop exact tests for structural change in the context of simultaneous-equation models based upon k-class estimation, e.g., two-stage least squares, ordinary least squares, and limited information maximum likelihood, of the endogenous part of the structural equation. These new tests have many of the characteristics of the analysis of covariance and Chow tests of the classical regression model; and they may be regarded as, essentially, simultaneous-equation analogues of those tests.

2.2 Testing for Structural Change in Simultaneous Equation Models: Some Preliminary Results

We shall consider the equation

$$\mathbf{y}_1 = Y_2\beta + Z_1\gamma + \mathbf{u}_1 \tag{2.1}$$

where $\mathbf{y}_1$ and Y_2 are, respectively, a $T \times 1$ vector and $T \times g$ matrix of observations on $g+1$ endogenous variables, and Z_1 is a $T \times r$ matrix of observations on r exogenous variables.

The reduced form of the system includes

$$Y_1 = Z\Pi_1 + V_1 \tag{2.2}$$

where $Y_1 = (\mathbf{y}_1 \vdots Y_2)$, $Z = (Z_1 \vdots Z_2)$ is a $T \times R$ matrix of exogenous variables, $\Pi_1 = (\pi_1 \vdots \Pi_2)$ is a $R \times (g+1)$ matrix of reduced form parameters, and $V_1 = (\mathbf{v}_1 \vdots V_2)$ is a $T \times (g+1)$ matrix of reduced form disturbances.

It is assumed that:

1. The rows of V_1 are independently and normally distributed with mean $\mathbf{0}'$ and non-singular covariance matrix $\Omega = \{\omega_{ij}\}$matrix
2. The $T \times R$ matrix Z is nonstochastic and of rank R $(< T)$.

The k-class estimators of the unknown parameters of (2.1) are given by

$$\begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} Y_2'Y_2 - k\hat{V}_2'\hat{V}_2 & Y_2'Z_1 \\ Z_1'Y_2 & Z_1'Z_1 \end{pmatrix}^{-1} \begin{pmatrix} Y_2 - k\hat{V}_2 \\ Z_1 \end{pmatrix}' \mathbf{y}_1 \tag{2.3}$$

where $\hat{V}_2 = Y_2 - Z\hat{\Pi}_2$ is the $T \times g$ matrix of reduced form residuals from regressing Y_2 on Z. Equation (2.3) may be rewritten in the form

$$\begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} \hat{\Pi}_2'Z'Z\hat{\Pi}_2 + (1-k)\hat{V}_2'\hat{V}_2 & \hat{\Pi}_2'Z'Z_1 \\ Z_1'Z\hat{\Pi}_2 & Z_1'Z_1 \end{pmatrix}^{-1} \times \left[\begin{pmatrix} \hat{\Pi}_2'Z'Z\hat{\pi}_1 \\ Z_1'Z\hat{\pi}_1 \end{pmatrix} + \begin{pmatrix} (1-k)\hat{V}_2'\hat{\mathbf{v}}_1 \\ 0 \end{pmatrix} \right] \tag{2.4}$$

where $\hat{\Pi}_2$ is the matrix of reduced-form coefficient estimates from regressing Y_2 on Z, $\hat{\pi}_1$ is the vector of reduced-form coefficient estimates from regressing $\mathbf{y}_1$ on Z, and $\hat{\mathbf{v}}_1$ is the residual vector from this latter regression.

From (2.4) it is seen that when $k = 1$, the estimator depends only on $\hat{\Pi}_1 = (\hat{\pi}_1 : \hat{\Pi}_2)$ but when $k \neq 1$ and is nonstochastic, the estimator depends on $\hat{\Pi}_1$, $\hat{V}_2'\hat{V}_2$, and $\hat{V}_2'\hat{\mathbf{v}}_1$. Note that both $\hat{V}_2'\hat{V}_2$ and $\hat{V}_2'\hat{\mathbf{v}}_1$ are contained in $\hat{V}_1'\hat{V}_1$ where $\hat{V}_1 = (\hat{\mathbf{v}}_1 : \hat{V}_2)$ so that for nonstochastic $k \neq 1$, the estimator depends on $\hat{\Pi}_1$ and $\hat{V}_1'\hat{V}_1$.

When k is stochastic, the estimator of primary interest is the limited information maximum likelihood (LIML) estimator when k is the smallest root, μ, of the determinantal equation

$$|Y_1'(I - P_{z_1})Y_1 - \mu Y_1'(I - P_z)Y_1| = 0$$

Here, $P_{z_1} = Z_1(Z_1'Z_1)^{-1}Z_1'$ and $P_z = Z(Z'Z)^{-1}Z'$. Notice that k depends on the components of $Y_1'(I - P_{z_1})Y_1$ and $Y_1'(I - P_z)Y_1 = \hat{V}_1'\hat{V}_1$. Since

$$Y_1'(I - P_{z_1})Y_1 = Y_1'(I - P_z)Y_1 + Y_1'(P_z - P_{z_1})Y_1 = \hat{V}_1'\hat{V}_1 + Y_1'(P_z - P_{z_1})Y_1$$

where $Y_1'(P_z - P_{z_1})Y_1$ itself depends on the stochastic matrix $(P_z - P_{z_1})V_1$, it follows that the LIML estimator depends on $\hat{\Pi}_1$, $\hat{V}_1'\hat{V}_1$, and $(P_z - P_{z_1})V_1$.

Lemma 2.1 *(See, for example, Feller, 1971, p. 134) If the m-tuple $(X_1, \ldots, X_m)$ is independent of the n-tuple $(W_1, \ldots, W_n)$, then $h(X_1, \ldots, X_m)$ and $\ell(W_1, \ldots, W_n)$ are independent for any pair of functions h and ℓ.*

In the model considered here, the matrix of reduced-form coefficient estimates is distributed independently of the residual matrix $\hat{V}_1$. This last point, together with the foregoing observations and lemma, enable the following comments to be made:

1. The two-stage least squares (2SLS) estimators (i.e., $k = 1$) of the unknown parameters of (2.1) depend only on $\hat{\Pi}_1$, and they are distributed independently of $\hat{V}_1$.

2. The ordinary least squares (OLS) estimators (i.e., $k = 0$) of the unknown parameters of (2.1) are not independent of $\hat{V}_1$. However, when $\hat{V}_1'\hat{V}_1$ is conditionally fixed, the OLS estimators depend only on $\hat{\Pi}_1$ and so are conditionally distributed independently of $\hat{V}_1$. By the term "conditionally fixed", we mean that we consider a partitioning of the sample space of the matrices $\hat{V}_1$ according to $\hat{V}_1'\hat{V}_1 = M$ (say). The conditional distributions defined for $\hat{V}_1$ are the distributions over each of these subsets of the sample space taken separately.

3. The limited information maximum likelihood (LIML) estimators (i.e., $k = \mu$) of the unknown parameters of (2.1) are not independent of $\hat{V}_1$. However, when $\hat{V}_1'\hat{V}_1$ is conditionally fixed, the LIML estimators depend only on $\hat{\Pi}_1$ and $(P_z - P_{z_1})V_1$, both of which are independent of $\hat{V}_1$. Thus, the LIML estimators are conditionally distributed independently of $\hat{V}_1$.

Next, comparing the first reduced form equation of (2.2) with (2.1) and equating stochastic parts, we have

$$\mathbf{v}_1 = V_2\beta + \mathbf{u}_1$$

and it will be convenient to write this in the form

$$\mathbf{u}_1 = V_1\beta_0 \tag{2.5}$$

where $\beta_0 = (1, -\beta')'$.

Consider the vector

$$\mathbf{u}_1^\star = \hat{V}_1\beta_0^\star \tag{2.6}$$

where $\beta_0^\star = (1, -\beta'^\star)'$ and $\beta^\star$ is the 2SLS estimator of β. Since $\beta^\star$ is distributed independently of $\hat{V}_1$, we have

$$E(\hat{V}_1\beta_0^\star \mid \beta^\star) = 0 \tag{2.7}$$

and

$$E(\hat{V}_1\beta_0^\star\beta_0'^\star\hat{V}_1' \mid \beta_0^\star) = w^2(I - P_z) \tag{2.8}$$

where $w^2 = \beta_0^{\star\prime}\Omega\beta_0^{\star}$.

It follows that $\hat{V}_1\beta_0^{\star} \mid \beta^{\star} \sim N[0, w^2(I - P_z)]$, so that, conditional on $\beta^{\star}$, the residual vector in (2.6) behaves like an OLS residual vector in the general linear model, a result which enables the development of tests for changes in the parameters of (2.1) in a relatively straightforward fashion. An analogous result does not hold, however, if the estimator of β employed in (2.6) is the OLS estimator because it would not be distributed independently of $\hat{V}_1$. Consequently, a different approach will be used to develop tests based on OLS or LIML estimation.

2.3 Tests Based upon Two-stage Least Squares (2SLS) Estimation

Suppose that two samples of observations of size T_1 and T_2 are available, and it is required to test for changes in the coefficients of (2.1) during the period to which the T_2 observations refer. Suppose first that $T_1, T_2 > R$ so that a regression of the estimated endogenous part of (2.1) on all the exogenous variables of the model can be carried out. We shall write the $T_i \times 1$ vector $\mathbf{y}_1^{(i)}$ to indicate the T_i observations of the vector $\mathbf{y}_1$, $i = 1, 2$, and a similar notation will be adopted for the matrices V_1 and Z. Next suppose that the T_i observations are used to obtain the 2SLS estimator of β. We shall write this estimator as $\beta^{\star(i)}$ so that the estimated endogenous part of the equation based on the T_i observations is $\mathbf{y}_1^{(i)} - Y_2^{(i)}\beta^{\star(i)}$, $i = 1, 2$.

Let $\hat{V}_1^{(i)}$ be the $T_i \times g$ matrix of reduced-form residuals obtained from the regression of Y_1 on Z using the T_i observations. Then, the sum of squared residuals from the regression of $\mathbf{y}_1^{(i)} - Y_2^{(i)}\beta^{\star(i)}$ on T_i observations of all the exogenous variables is given by

$$SS_i^{\star} = \beta_0^{\star(i)\prime}\hat{V}_1^{(i)\prime}\hat{V}_1^{(i)}\beta_0^{\star(i)}, \quad i = 1, 2, \tag{2.9}$$

where $\beta_0^{\star(i)} = [1, -\beta^{\star(i)}]'$.

We note that $\hat{V}_1^{(1)}$ is distributed independently of $\hat{V}_1^{(2)}$, and $\beta^{\star(1)}$ is distributed independently of $\beta^{\star(2)}$ since distinct sets of observations are used to obtain them. Also, the independence of $\hat{V}_1^{(i)}$ and $\beta^{\star(i)}$ means that $\hat{V}_1^{(i)}\beta_0^{\star(i)} \mid \beta^{\star(i)} \sim N[0, w_i^2(I_{T_i} - P_z^{(i)})]$, $i = 1, 2$, where $P_z^{(i)}$ is a version of $Z(Z'Z)^{-1}Z'$ based on T_i observations and $w_i^2 = \beta_0^{\star(i)\prime}\Omega\beta_0^{\star(i)}$.

It follows that, conditional on $\beta^{\star(i)}$, the sums of squares $SS_i^{\star}$ in (2.9) are distributed as $w_i^2\chi^2_{T_i-R}$, $i = 1, 2$, independently of each other. Finally, when all the observations are used, we have

$$SS^{\star} = \beta_0^{\star\prime}\hat{V}_1'\hat{V}_1\beta_0^{\star}$$

and $SS^{\star}$ is distributed, conditional on $\beta^{\star}$, as $w^2\chi^2_{T-R}$ where $w^2 = \beta_0^{\star\prime}\Omega\beta_0^{\star}$.

The analysis of covariance test considered here is basically the same as that used in the general linear model, and so the rationalization will be the same. All the exogenous variables are included in the regression that produces the residuals from which the test statistic is derived, however, and this is necessary to obtain a test statistic with the appropriate null hypothesis distribution. Under the alternative hypothesis, the residuals in $\hat{V}_1$ have a nonzero mean.

A simultaneous equation analogue of the analysis of covariance test statistic, as used in the general linear model, is given by

$$\begin{aligned} F^{\star} &= \frac{(\beta_0^{\prime\star}\hat{V}_1'\hat{V}_1\beta_0^{\star} - [\beta_0^{\prime\star(1)}\hat{V}_1^{\prime(1)}\hat{V}_1^{(1)}\beta_0^{\star(1)} + \beta_0^{\prime\star(2)}\hat{V}_1^{\prime(2)}\hat{V}_1^{(2)}\beta_0^{\star(2)}])/R}{((\beta_0^{\prime\star(1)}\hat{V}_1^{\prime(1)}\hat{V}_1^{(1)}\beta_0^{\star(1)} + \beta_0^{\prime\star(2)}\hat{V}_1^{\prime(2)}\hat{V}_1^{(2)}\beta_0^{\star(2)})/(T_1+T_2-2R)} \\ &= \frac{(SS^{\star} - [SS_1^{\star} + SS_2^{\star}])/R}{(SS_1^{\star} + SS_2^{\star})/(T_1+T_2-2R)} \end{aligned} \tag{2.10}$$

Conditional on $\beta^{\star(i)}$, $i = 1,2$, and $\beta^{\star}$, the components forming (2.10) are distributed as

$$\frac{(w^2\chi^2_{T-R} - [w_1^2\chi^2_{T_1-R} + w_2^2\chi^2_{T_2-R}])/R}{(w_1^2\chi^2_{T_1-R} + w_2^2\chi^2_{T_2-R})/(T_1+T_2-2R)}$$

where, in general, w^2, w_1^2, and w_2^2 would differ because $\beta^{\star(i)}$, $i = 1,2$, and $\beta^{\star}$ are not usually the same, and so the distribution of (2.10) is unknown.

If w^2, w_1^2, and w_2^2 were always equal, the conditional distribution of (2.11), below, would approximate $F_{(R,T_1+T_2-2R)}$. [The theoretical analysis is complicated by the fact that $\hat{V}_1$ is not independent of $\beta^{\star(i)}$, $i = 1,2$, so that the conditional distribution may not be exactly $F_{(R,T_1+T_2-2R)}$. We are grateful to a referee for drawing our attention to this point.] Since this would hold for all $\beta^{\star(i)}$, $i = 1,2$, and $\beta^{\star}$, the result would hold unconditionally and so a test for structural change could be based on (2.10). If w^2, w_1^2, and w_2^2 are not all equal, but merely approximately equal, tests can still be carried out, but "exact" inferences are not possible.

In practice, if neither T_1 nor T_2 is small, then it is probably safe to assume that w^2, w_1^2, and w_2^2 are close enough to base inferences on (2.10), assuming that the $F_{(R,T_1+T_2-2R)}$ distribution is approximately correct. The effects of heteroscedasticity on the test statistics used in testing for structural change in the general linear model have been studied by Toyoda (1974) and Schmidt and Sickles (1977). Their results, which are also relevant in this context, suggest that, if the heteroscedasticity is not serious, the size of the test is little affected.

Of course, it is possible to derive a test statistic that has the $F_{(R,T_1+T_2-2R)}$ distribution under the null hypothesis, simply by choosing appropriately the same estimate of β in obtaining $SS^{\star}$, $SS_1^{\star}$, and $SS_2^{\star}$. Which estimate of β to use is dictated by the need to make (2.10) as large as possible under the alternative hypothesis so as to maximize the power

of the test. There are no clear rules for choosing which estimate to use, but heuristic reasoning would suggest choosing $\beta^{\star(1)}$ or $\beta^{\star}$ when, as is usually the case, T_1 is greater than T_2. If $\beta^{\star(1)}$ is chosen, we run into the same difficulties noted above for the case where $\hat{V}_1$ is not independent of $\beta^{\star(i)}$, $i = 1, 2$. If $\beta^{\star}$ is used, then the analysis of covariance test statistc becomes

$$F^{\star} = \frac{\beta_0^{\prime\star}(\hat{V}_1'\hat{V}_1 - [\hat{V}_1^{\prime(1)}\hat{V}_1^{(1)} + \hat{V}_1^{\prime(2)}\hat{V}_1^{(2)}])\beta_0^{\star}/R}{\beta_0^{\prime\star}(\hat{V}_1^{\prime(1)}\hat{V}_1^{(1)} + \hat{V}_1^{\prime(2)}\hat{V}_1^{(2)})\beta_0^{\star}/(T_1 + T_2 - 2R)} \tag{2.11}$$

Since $\beta^{\star}$ is distributed independently of $\hat{V}_1^{(1)}$, $\hat{V}_1^{(2)}$, and $\hat{V}_1$, it is apparent that the conditional distribution of (2.11) is $F_{(R,T_1+T_2-2R)}$ under the null hypothesis. Since this distribution holds for all $\beta^{\star}$, it holds unconditionally and an *exact* test for structural change can be based on (2.11). The requirement of using the same estimate of β in obtaining $SS^{\star}$, $SS_1^{\star}$, and $SS_2^{\star}$ may reduce, to an unknown degree, the power of the test against certain types of alternative hypothesis, particularly those that involve a change in β. However, this has to be accepted if an exact test is to be obtained.

In the case where insufficient observations are available to compute the second regression, i.e., $T_2 < R$, the simultaneous equation analogue of the Chow test statistic is

$$\begin{aligned} F^{\star\star} &= \frac{(\beta_0^{\prime\star}\hat{V}_1'\hat{V}_1\beta_0^{\star} - \beta_0^{\prime\star(1)}\hat{V}_1^{\prime(1)}\hat{V}_1^{(1)}\beta_0^{\star(1)})/T_2}{\beta_0^{\prime\star(1)}\hat{V}_1^{\prime(1)}\hat{V}_1^{(1)}\beta_0^{\star(1)}/(T_1 - R)} \\ &= \frac{(SS^{\star} - SS_1^{\star})/T_2}{SS_1^{\star}/(T_1 - R)} \end{aligned} \tag{2.12}$$

[The Chow test considered here is basically the same as that used in the general linear model. Again, all the exogenous variables are included in the regression.] Conditional on $\beta^{\star(1)}$ and $\beta^{\star}$, the components of the statistic (2.12) are distributed as

$$\frac{(w^2\chi^2_{T-R} - w_1^2\chi^2_{T_1-R})/T_2}{w_1^2\chi^2_{T_1-R}/(T_1 - R)}$$

which will not be distributed as $F_{(T_2,T_1-R)}$ unless $w^2 = w_1^2$. Since by assumption T_2 is small, it will often be the case that $w^2 \simeq w_1^2$, whereupon it will usually be safe to assume that (2.12) is approximately distributed as $F_{(T_2,T_1-R)}$.

An exact test can be obtain by choosing the estimate $\beta^{\star}$ when calculating $SS^{\star}$ and $SS_1^{\star}$. If this is done, the statistic becomes

$$F^{\star\star} = \frac{\beta_0^{\prime\star}(\hat{V}_1'\hat{V}_1 - \hat{V}_1^{\prime(1)}\hat{V}_1^{(1)})\beta_0^{\star}/T_2}{\beta_0^{\prime\star}\hat{V}_1^{\prime(1)}\hat{V}_1^{(1)}\beta_0^{\star}/(T_1 - R)} \tag{2.13}$$

Since $\beta_0^\star$ is distributed independently of $\hat{V}_1^{(1)}$ and $\hat{V}_1$, it follows that, conditional on $\beta^\star$, the statistic (2.13) is distributed as $F_{(T_2,T_1-R)}$. Since the result holds for all $\beta^\star$, it holds unconditionally, and so an exact test for structural change can be based on (2.13). [This test is discussed by Erlat (1983).]

While the test statistics used in (2.11) and (2.13) are analogous to those used for the analysis of covariance test and the Chow test in the general linear model, the performances of the tests are more difficult to analyze theoretically. In the general linear model, it is usually assumed that the disturbance variance is unchanged after the structural change has occurred and that, under the alternative, the OLS residuals based on all the observations have a nonzero mean, and this fact crucially affects the power of the test. In the simultaneous-equation case, the situation is more complicated if the alternative hypothesis involves a change in β since then the distribution of the reduced-form disturbances must change if the structural disturbances are to remain unchanged. So apart from the residuals involved in $SS^\star$ having nonzero means, if a structural change has occurred, the structural change also affects the variance of the disturbances involved. Such changes may serve to increase or decrease the powers of the tests considered.

2.4 Tests Based upon Ordinary Least Squares (OLS) and Limited Information Maximum Likelihood (LIML) Estimation

In this section we consider tests for structural change in which the endogenous part of the equation is estimated using OLS. It is still commonplace for simultaneous-equation models to be estimated by OLS, even though it is well-known that OLS estimators are biased and inconsistent. Often, models are estimated by more than one technique and, when this is the case, OLS estimators may be compared with theoretically more attractive estimators before a choice is made. Since OLS is still used frequently, it is of some interest that tests for structural change can be based on OLS estimation. The LIML estimator has not been used as often as 2SLS or OLS, but there has been renewed interest in LIML in recent years.

In developing our tests, we shall consider

$$C\tilde{\mathbf{u}}_1 = C\hat{V}_1\tilde{\beta}_0 \tag{2.14}$$

where $\tilde{\mathbf{u}}_1 = \hat{V}_1\tilde{\beta}_0$, $\tilde{\beta}_0 = (1, -\tilde{\beta}')'$, $\tilde{\beta}$ is the OLS or LIML estimator of β, and C is a $(T-R)\times T$ matrix as defined in the appendix. Consider the matrix

$$C = \begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix}$$

where $C_1C_2' = 0$, $C_1C_3' = 0$, $C_2C_3' = 0$, and C has the properties: $CZ = 0$, $C'C = I_T - P_z$ and $CC' = I_{T-R}$. The quantities C_1V_1, C_2V_1, and C_3V_1 are distributed independently of each other and, for samples in which $\hat{V}_1'\hat{V}_1 = \hat{V}_1'C'CV_1$ is constant, $\tilde{\beta}$ is distributed independently of CV_1, i.e., distributed independently of C_1V_1, C_2V_1, and C_3V_1.

Durbin (1957) examined the LIML case and showed that, when $\tilde{\beta}_0$ and $\hat{V}_1'\hat{V}_1$ are fixed, $CV_1\tilde{\beta}_0$ is uniformly distributed on the sphere $\tilde{\beta}_0'V_1'C'CV_1\tilde{\beta}_0 =$ constant; and if we put $\varepsilon = CV_1\tilde{\beta}_0$ and let $\varepsilon_i^* = \varepsilon_i/\sqrt{\varepsilon'\varepsilon}$, $i = 1,\ldots,T-R$, then ε_i^* is, conditionally, uniformly distributed on the unit sphere. However, ε_i^* has this distribution for all $\hat{V}_1'\hat{V}_1$, and so will have this distribution unconditionally. This analysis carries through directly for the OLS case as well. Durbin noted that if q is a χ^2 variate with $T-R$ degrees of freedom, distributed independently of ε_i^*, then the variates

$$\eta_i = \sqrt{q}\,\varepsilon_i^*, \quad i = 1,\ldots,T-R$$

are $iidN(0,1)$, and Bouman (1971) provides an explicit proof of this result.

Putting

$$\varepsilon^* = \frac{1}{\tilde{\beta}_0'V_1'C'CV_1\tilde{\beta}_0}\begin{pmatrix} C_1V_1\tilde{\beta}_0 \\ C_2V_1\tilde{\beta}_0 \\ C_3V_1\tilde{\beta}_0 \end{pmatrix} \tag{2.15}$$

it is seen that

$$q\,\frac{\tilde{\beta}_0'V_1'C_1'C_1V_1\tilde{\beta}_0}{\tilde{\beta}_0'V_1'C'CV_1\tilde{\beta}_0} = q\sum_{i=1}^{T_1-R}\varepsilon_i^{*2} = \sum_{i=1}^{T_1-R}\eta_i^2$$

is distributed as a χ^2 variate with $T_1 - R$ degrees of freedom. Similarly,

$$q\,\frac{\tilde{\beta}_0'V_1'C_2'C_2V_1\tilde{\beta}_0}{\tilde{\beta}_0'V_1'C'CV_1\tilde{\beta}_0} \quad \text{and} \quad q\,\frac{\tilde{\beta}_0'V_1'C_3'C_3V_1\tilde{\beta}_0}{\tilde{\beta}_0'V_1'C'CV_1\tilde{\beta}_0}$$

are distributed as $\chi^2_{(T_2-R)}$ and $\chi^2_{(R)}$, respectively, and all three χ^2 variates are mutually independent.

The appropriate analysis of covariance test statistic can now be formed as

$$\tilde{F} = \frac{\tilde{\beta}_0'(\hat{V}_1'\hat{V}_1 - [\hat{V}_1'^{(1)}\hat{V}_1^{(1)} + \hat{V}_1'^{(2)}\hat{V}_1^{(2)}])\tilde{\beta}_0/R}{\tilde{\beta}_0'(\hat{V}_1'^{(1)}\hat{V}_1^{(1)} + \hat{V}_1'^{(2)}\hat{V}_1^{(2)})\tilde{\beta}_0/(T_1+T_2-2R)} \tag{2.16}$$

which is distributed as $F_{(R,T_1+T_2-2R)}$, under the null hypothesis.

Note that (2.16) is formed from the χ^2 variates given above, using the fact that the numerator quadratic form is equal to $\tilde{\beta}_0'V_1'C_3'C_3V_1\tilde{\beta}_0$, and then cancelling q between numerator and denominator.

When insufficient observations are available to compute the second regression, an exact Chow test is available, as in the 2SLS case, and the required statistic is obtained by substituting $\tilde{\beta}_0$ for β_0^* in (2.12).

2.5 Conclusions

In the foregoing analysis we have shown that, when the endogenous part of a simultaneous equation is estimated using two-stage least squares, ordinary least squares, or limited information maximum likelihood, the residuals obtained from regressing the estimated endogenous part on all the exogenous variables of the model can be used to construct exact tests for structural change. These tests can be thought of as simultaneous-equation counterparts of the analysis of covariance and Chow tests.

Appendix

Let $Z = (Z_1 \vdots Z_2)$ be a $T \times R$ matrix containing the observations on the R nonstochastic exogenous variables; then it is possible to find a $(T - R) \times T$ matrix C, say, such that

$$\begin{aligned} CZ &= 0 \\ CC' &= I_{T-R} \\ C'C &= I - Z(Z'Z)^{-1}Z' \end{aligned} \tag{2.17}$$

Such "C" matrices are much discussed in the theory of exact tests for serial correlation in the general linear model; see Theil (1965). The explicit form of an appropriate matrix is given in Phillips and Harvey (1974).

Let P_i, $i = 1, 2$, be $T \times T$ pattern matrices such that $P_1 Z = \begin{pmatrix} Z^{(1)} \\ 0 \end{pmatrix}$ and $P_2 Z = \begin{pmatrix} 0 \\ Z^{(2)} \end{pmatrix}$, where $Z^{(1)}$ contains the first T_1 rows and $Z^{(2)}$ contains the last T_1 rows, of Z. Noting that $P_1 Z(Z'P_1'P_1Z)^{-1}Z'P_1'$ and $P_2 Z(Z'P_2'P_2Z)^{-1}Z'P_2'$ are both symmetric and idempotent, consider the matrix

$$A = P_1 Z(Z'P_1'P_1Z)^{-1}Z'P_1' + P_2 Z(Z'P_2'P_2Z)^{-1}Z'P_2' - Z(Z'Z)^{-1}Z' \tag{2.18}$$

Noting that $P_1'P_2 = 0$ and $(P_1 + P_2)Z = Z$, it is easily seen that A is symmetric and idempotent, and $AZ = 0$.

Now if we put $I_T^{(1)} = P_1 I_T = \begin{pmatrix} I_{T_1} & 0 \\ 0 & 0 \end{pmatrix}$ and $I_T^{(2)} = P_2 I_T = \begin{pmatrix} 0 & 0 \\ 0 & I_{T_2} \end{pmatrix}$, we may write $I_T - Z(Z'Z)^{-1}Z' = [I_T^{(1)} - P_1 Z(Z'P_1'P_1Z)^{-1}Z'P_1'] + [I_T^{(2)} - P_2 Z(Z'P_2'P_2Z)^{-1}Z'P_2'] + A$, where the three matrices on the r.h.s. are symmetric and idempotent and the first two are orthogonal to A.

Let C_i, $i = 1, 2$, be a $(T_i - R) \times T$ matrix such that

$$\begin{aligned} C_i P_i Z &= 0 \\ C_i C_i' &= I_{T_i - R} \\ C_i' C_i &= I_T^{(i)} - P_i Z (Z' P_i' P_i Z)^{-1} Z' P_i' \end{aligned} \tag{2.19}$$

and let C_3 be an $R \times T$ matrix such that $C_3 Z = 0$, $C_3 C_3' = I_R$, and $C_3' C_3 = A$, then we can define $C = (C_1' \vdots C_2' \vdots C_3')'$ which is a $(T - R) \times T$ matrix with the properties in (2.17). Note that $C_1 C_2' = 0$, $C_1 C_3' = 0$, and $C_2 C_3' = 0$.

With this particular definition of C, we may write

$$V_1' C_i' C_i V_1 = V_1^{\prime(i)} [I_{T_i} - Z^{(i)} (Z^{\prime(i)} Z^{(i)})^{-1} Z^{\prime(i)}] V_1^{(i)} = \hat{V}_1^{\prime(i)} \hat{V}_1^{(i)}, \quad i = 1, 2 \tag{2.20}$$

and

$$\begin{aligned} V_1' C_3' C_3 V_1 &= V_1' [I - Z(Z'Z)^{-1} Z'] V_1 - \\ &\sum_{i=1}^{2} V_1^{\prime(i)} [I_{T_i} - Z^{(i)} (Z^{\prime(i)} Z^{(i)})^{-1} Z^{\prime(i)}] V_1^{(i)} = \hat{V}_1' \hat{V}_1 - \hat{V}_1^{\prime(1)} \hat{V}_1^{(1)} - \hat{V}_1^{\prime(2)} \hat{V}_1^{(2)} \end{aligned} \tag{2.21}$$

The terms in (2.20) and (2.21) are used in forming (2.16).

Acknowledgment

We are grateful to the E.S.R.C. for financial support in connection with our overall project, "Testing for Specification Error in Econometric Models", at the University of Kent, 1975-1977. The first version of this chapter (Harvey and Phillips, 1978) was written as part of this project.

References

Bouman, J. (1971), Testing for autocorrelation in simultaneous equation models. Report 7104, Econometric Institute, Erasmus University, Rotterdam.

Chow, G.C. (1960), Tests of equality between subsets of coefficients in two linear regressions. *Econometrica*, **28**, 591-605.

Durbin, J. (1957), Testing for serial correlation in systems of simultaneous regression equations. *Biometrika*, **44**, 370-377.

Erlat, H. (1983), A note on testing for structural change in a single equation belonging to a simultaneous equation system. *Economics Letters*, **13**, 185-189.

Feller, W. (1971), *An Introduction to Probability Theory and Its Applications*, Vol. 2, 2nd edition, New York: John Wiley.

Fisher, F.M. (1970), Tests of equality between sets of coefficients in two linear regressions: An expository note. *Econometrica*, **38**, 361-366.

Giles, D.E.A. (1981), Testing for parameter stability in structural econometric relationships. *Economics Letters*, **7**, 323-326.

Harvey, A.C. and Phillips, G.D.A. (1974), A simple test for serial correlation in regression models. *Journal of the American Statistical Association*, **69**, 935-939.

Harvey, A.C. and Phillips, G.D.A. (1978), Testing for structural change in simultaneous equation models. University of Kent (unpublished paper)

Harvey, A.C. and Phillips, G.D.A. (1980), Testing for serial correlation in simultaneous equation models. *Econometrica*, **48**, 747-759.

Harvey, A.C. and Phillips, G.D.A. (1981), Testing for heteroscedasticity in simultaneous equation models. *Journal of Econometrics*, **15**, 311-340.

Schmidt, P. and Sickles, R. (1977), Some further evidence on the use of the Chow test under heteroscedasticity. *Econometrica*, **45**, 1293-1298.

Theil, H. (1965), The analysis of disturbances in regression analysis. *Journal of the American Statistical Association*, **60**, 1067-1078.

Toyoda, T. (1974), Use of the Chow test under heteroscedasticity. *Econometrica*, **42**, 601-608.

CHAPTER 3

Specification and Stability Tests versus Jackknifing: Some Illustrative Examples

Bernd Schips and Yngve Abrahamsen

Summary

We present results of Monte Carlo simulations for model misspecification. Several variations of Stone's (1974) Q^2 statistic are computed using the jackknife procedure. These statistics measure the relative error in predicting the omitted observations using simple extrapolations. We compare the ability of these "predictive" statistics to detect model misspecification with the usual test statistics. The misspecifications we examined include omitted variables, nonlinearities, and structural change (shifts in coefficients during the sample period). Although none of the predictive statistics perform uniformly better over the entire range of misspecifications, in some cases, they detect departures from the "true model" with substantially greater frequency.

In the process of estimating unknown coefficients of the structural form of interdependent econometric models, classical methods are only asymptotically justified. It is unknown how large the sample must be in order to reach the asymptotic features of the chosen estimators and tests (Zellner, 1979). In addition, we must consider the fact that proving asymptotic qualities presupposes the complete and correct specification of the model itself. Yet models are in fact no more than a fair approximation of reality.

In theory econometric model building can be divided into three sequential steps: (a) the specification of the model, (b) the numeric determination of the model's structure, and (c) the evaluation of the model. In practice, the strict sequence of these three steps does not exist. A model emerges in an iterative process that is dominated by the principle of trial and error.

Textbook presentations cannot account for this fact. They are therefore limited to

two aspects: First, they demonstrate methods for estimating unknown parameters of a given model structure that is supposed to be correct. Second, they demonstrate methods to verify assumptions used in the specification process. Such textbook presentations therefore assume that the structure of the model, whose parameters we want to estimate, are specified correctly. In other words, they assume that the set of alternative hypotheses contains a model structure that is specified correctly. Only if these assumptions are true can the desired qualities of the normally used estimators and tests be proved.

From the viewpoint of classical inference statistics, the traditional procedure and presentation is quite obvious. As soon as the distribution of the random variables is determined, the possibilities for deducing estimators with the desired qualities are given. The same is true for the deduction of tests for the verification of hypotheses with respect to the distribution.

Disregarding the technical problems in the process of deducing estimators and tests (which may be considerable), it seems as though quite efficient methods would result from the application of classical inference statistics. Yet one of the main problems of these methods lies in the fact that statements about the model's structure are dependent upon the distributive assumptions that are presupposed. These distributive assumptions cannot be verified themselves.

Tests concerning the stochastic specification of an econometric model are always applied in connection with the implied economic hypotheses and are limited to the verification of specific distributive assumptions within a certain set of distributions. The fundamental problem of specification test (i.e., tests with respect to specification errors) lies in the fact that only certain aspects of the hypotheses — or only parts of competing hypotheses — can be examined. Other aspects of the hypotheses are assumed to be correct and remain untested. The process of specifying an econometric model, therefore, involves a decision with respect to which assumptions will remain untested and which will be put to verification. From this point of view, we should evaluate an econometric model as a whole.

The process of evaluating a model is normally limited to

1. The economic (but not econometric) verification of estimated parameters.
2. The multipliers based on these parameters.
3. The analysis of the test statistics concerning the *ex post* (and in some cases the *ex ante*) predictive qualities of the estimated structure.

These evaluation procedures do not eliminate all doubt (Chong and Hendry, 1986). These doubts become obvious if we take into account that the estimated models are normally put to further tests. For example, it is common to do dynamic *ex post* simulations. In this process it is often ignored that the results of these dynamic simulations are also sensitive to exogeneous variables that are not part of the model.

On the other hand, the *ex ante* predictive relevance can always be tested. Therefore, we recommend using a combination of jackknifing and other measures that shed light upon

Table 3.1: Types of misspecification: omitted variable and incorrect functional form.

Error code	*Data-generating model* [a]	*Estimation model* [b]
F1	$C_t = u_t$	$C_t = b_1 + b_2 Y_t + v_t$
F2	$C_t = 10 + 0.5Y_t + u_t$	$C_t = b_1 + b_2 Y_t + v_t$
F3	$C_t = 10 + 0.2Y_t + 0.375C_{t-1} + u_t$	$C_t = b_1 + b_2 Y_t + b_3 C_{t-1} + v_t$
F4	$C_t = 10 + 0.3Y_t + 0.15Y_{t-1} + u_t$	$C_t = b_1 + b_2 Y_t + v_t$
F5	$C_t = 10 + 0.2Y_t + 0.375C_{t-1} + u_t$	$C_t = b_1 + b_2 Y_t + v_t$
F6	$C_t = 10 + 0.2Y_t + 0.375C_{t-1} + u_t$	$C_t = b_1 + b_2 C_{t-1} + v_t$
F7	$C_t = 10 + 0.3Y_t + 0.015Y_t^2 + u_t$	$C_t = b_1 + b_2 Y_t + v_t$
F8	$C_t = 10 + 0.3Y_t^{0.9} + u_t$	$C_t = b_1 + b_2 Y_t + v_t$

[a] $u_t \sim N(0,1)$; the variances of the error terms correspond to the estimated standard deviation of residuals in econometric modeling.

[b] Here, v_t is assumed to be distributed as $N(0, \sigma_v^2)$.

the predictive quality of a model. Such a combination has been suggested by Stone (1974) and Geisser (1974), similar to the Q^2 that was introduced by Ball (1963). The efficiency of such a combination will be demonstrated by using single-equation models that are specified incorrectly in different ways. For these types of model, there exists a large number of specification tests or tests regarding specification errors. The types of incorrect specifications examined are listed in *Table 3.1.*

For the analysis of Monte Carlo simulations, we used — besides the well-known ordinary criteria — modifications of the Q^2 measure suggested by Stone (1974) and Geisser (1974). Instead of comparing the predictive quality of the values for Y_t^* (these are the predictions for the observations, which have been omitted in the process of jackknifing) with the predictive quality of the mean value of the observations:

$$Q_{SG}^2 = 1 - \frac{\Sigma_t (Y_t - Y_t^*)^2}{\Sigma_t (Y_t - \overline{Y}_t)^2}$$

we made comparisons with others' "naive" predictors. First, we compared the values for Y_t^* with a random-walk model, which means that the prediction is equal to the last value observed:

$$Q_{RW}^2 = 1 - \frac{\Sigma_t (Y_t - Y_t^*)^2}{\Sigma_t (Y_t - Y_{t-1})^2}$$

Second, we compared Y_t^* with a type of prediction that extrapolates the last change of the observation:

$$Q_{LC}^2 = 1 - \frac{\Sigma_t (Y_t - Y_t^*)^2}{\Sigma_t (Y_t - 2Y_{t-1} + Y_{t-2})^2}$$

Table 3.2: Criteria for detecting misspecification.

Criterion code	*Procedure*
K1	t-test; H_0: $b_1 = 0$
K2	JK-t-test; H_0: $b_1 = 0$
K3	Durbin-Watson test or Durbin's h-test
K4	$R^2 < 0.8$
K5	Bartlett's M specification error test (BAMSET)
K6	Regression specification error test (RESET)
K7	$Q^2_{SG} < 0.8$
K8	$Q^2_{RW} < 0.8$
K9	$Q^2_{LC} < 0.8$

Table 3.3: Evaluation of the Monte Carlo simulations: Number of indicated misspecifications out of 100 simulations (significance level of the tests: 0.05).

Criterion	*Error code*							
code	F1	F2	F3	F4	F5	F6	F7	F8
K1	98	0	89	0	0	28	91	0
K2	94	0	87	0	0	31	94	0
K3	5	5	32	7	39	2	11	5
K4	100	0	0	0	0	0	0	0
K5	20	20	20	12	12	11	23	19
K6	6	6	19	5	15	3	23	6
K7	100	0	0	0	0	0	0	0
K8	100	59	100	94	100	100	50	87
K9	36	7	81	37	87	100	8	13

In Monte Carlo simulations the realizations of independent normal distributed random variables were used as error terms. It is well-known that, in cases of autocorrelated dependent or nonnormal distributed error variables, the jackknifed estimators have some advantages compared with classical methods. In a first step the Hausman test and the Lagrange multiplier test have not been taken into consideration. These tests require knowledge about the special type of misspecification. Instead, we used the RESET and the BAMSET tests. The practicability of their combined application was shown by Ramsey and Gilbert (1972).

In *Table 3.2* and *3.3* we present the adopted evaluation criteria in detail and the results of the analysis of the Monte Carlo simulations.

Further, we analyzed the problem of a possible instability of the model's structure (i.e., a rupture in the structure caused by a numeric change of the coefficients) in the same

Table 3.4: Types of misspecification: structural change.

Error code	*Time period* [a]	*Data-generating model* [b]	*Estimation model* [c]
F9	A	$C_t = 10 + 0.45Y_t + u_t$	
	B	$C_t = 10 + 0.43Y_t + u_t$	$C_t = b_1 + b_2Y_t + v_t$
F10	A	$C_t = 10 + 0.45Y_t + u_t$	
	B	$C_t = 10 + 0.41Y_t + u_t$	$C_t = b_1 + b_2Y_t + v_t$
F11	A	$C_t = 10 + 0.2Y_t + 0.375C_{t-1} + u_t$	
	B	$C_t = 10 + 0.19Y_t + 0.36C_{t-1} + u_t$	$C_t = b_1 + b_2Y_t + b_3C_{t-1} + v_t$
F12	A	$C_t = 10 + 0.2Y_t + 0.375C_{t-1} + u_t$	
	B	$C_t = 10 + 0.18Y_t + 0.345C_{t-1} + u_t$	$C_t = b_1 + b_2Y_t + b_3C_{t-1} + v_t$

[a] Time period A: $t = 1, \ldots, 15$; B: $t = 16, \ldots, 25$.

[b] $u_t \sim N(0,1)$; the variances of the error terms correspond to the estimated standard deviation of residuals in econometric modeling.

[c] Here, v_t is assumed to be distributed as $N(0, \sigma_v^2)$.

Table 3.5: Criteria for detecting structural changes.

Criterion code	*Procedure*
K10	Chow test; H_0: $b_A = b_B = b$
K11	Sequence of Chow tests
K12	CUSUM test
K13	$Q^2_{SG} \leq \max[Q^2_{SG}(A), Q^2_{SG}(B)]$
K14	$Q^2_{RW} \leq \max[Q^2_{RW}(A), Q^2_{RW}(B)]$
K15	$Q^2_{LC} \leq \max[Q^2_{LC}(A), Q^2_{LC}(B)]$

way (*Table 3.4*). Besides the well-known criteria, we investigated the number of cases in which the Q^2 measures for the whole estimation period did not exceed the Q^2 measures for both subperiods (*Table 3.5*). If a Q^2 measure for a subperiod exceeds the corresponding measure for the complete period, we get a clear indication to suspect an instability in the model's structure. The Q^2 measures normally tend to increase according to the number of the observations in a correctly specified model.

These Monte Carlo simulations demonstrate the low degree of sensitivity of ordinary testing procedures (*Table 3.6*). The Q^2 measures used are at least an equivalent instrument for revealing an incorrect specification. From this point of view, we can say, at least, that nothing speaks against the use of prediction-oriented estimating methods in connection with prediction-oriented measures for model evaluation. On the contrary, these Q^2 measures seem to be more sensitive than the normally used tests originating from classical inference statistics.

Table 3.6: Evaluation of the Monte Carlo simulations: Number of indicated misspecifications out of 100 simulations (significance level of the tests: 0.05).

Criterion	*Error code*			
code	F9	F10	F11	F12
K1	0	0	54	43
K2	0	0	57	49
K3	5	7	22	18
K4	0	0	0	0
K5	16	16	10	11
K6	7	7	4	8
K7	0	0	0	0
K8	79	81	97	99
K9	9	12	71	77
K10	5	6	5	8
K11	21	21	21	27
K12	5	3	3	4
K13	21	22	5	4
K14	71	70	59	75
K15	65	31	39	46

We have yet to answer the questions whether and to what extent is a jackknifing procedure preferable to, e.g., bootstrapping. An advantage of bootstrapping lies in the fact that, normally, the true variance-covariance of the error terms (and, consequently, the variance-covariances of the unknown coefficients of the structural form) are estimated correctly; whereas jackknifing normally overestimates the variance-covariance matrices, as theoretical reasoning and Monte Carlo simulations show (Bluemel, 1986).

Using jackknifing we find ourselves, so to speak, always on the safe side, which is certainly an advantage if we take into account the theoretical deficits in econometric model building. In addition, the Q^2 measures are a sort of by-product of jackknifing estimations. The evaluation of the model and the estimation of the parameters take place *uno actu*.

References

Ball, R.J. (1963), The significance of simultaneous methods of parameter estimates in econometric models. *Applied Statistics*, **12**, 14-25.

Bluemel, B. (1986), Das Bootstrapverfahren angewandt auf die Fixpunkt-Methode: Theorie und Anwendungen der Systemschätzung, Ruhr-Universität Bochum (mimeo).

Chong, Y.Y. and Hendry, D.F. (1986), Econometric evaluation of linear macro-economic models. Applied Economics Discussion Paper No. 10, Institute of Economics and Statistics, Oxford University.

Geisser, S. (1974), A predictive approach to the random effect model. *Biometrika*, **61**, 101-107.

Ramsey, J.B. and Gilbert, R. (1972), A Monte Carlo study of some small sample properties

of tests for specification errors. *Journal of the American Statistical Association*, **67**, 180-186.

Stone, M. (1974), Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society*, **B-36**, 111-133.

Zellner, A. (1979), Statistical analysis of econometric models. *Journal of the American Statistical Association*, **74**, 628-643.

CHAPTER 4

The Robustness of the Chow Test to Autocorrelation among Disturbances

Walter Krämer

Summary

The chapter considers the rejection probability of the Chow test when there is unaccounted for autocorrelation among the disturbances in a linear regression model. The Chow test proved to be extremely nonrobust to autocorrelation. Its true size can even be one for the special case of an AR(1) disturbance process.

4.1 Introduction

Consider the standard linear regression model

$$y_t = x_t'\beta + u_t \ , t = 1, \ldots, T \tag{4.1}$$

where notation is obvious and where, for the sake of finite sample results, the regressors x_t $(K \times 1)$ are fixed and the disturbances u_t are $nid(0, \sigma^2)$. The familiar matrix notation for the model (4.1) is

$$Y = X\beta + u \tag{4.2}$$

where the $(T \times K)$-matrix X $(T > K)$ has full column rank.

The assumptions embodied in (4.2) constitute at the same time the null hypothesis under test. The alternative is that the regression coefficients β change somewhere in the sample, at $t = \tau$, say. The model (4.2) can then be rewritten as

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ X_2 & X_2 \end{pmatrix} \begin{pmatrix} \beta \\ \Delta\beta \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = Z\delta + u \tag{4.3}$$

where $Y_1 = (y_1, \ldots, y_\tau)'$, $Y_2 = (y_{\tau+1}, \ldots, y_T)'$, $X_1 = (x_1, \ldots, x_\tau)'$, $X_2 = (x_{\tau+1}, \ldots, x_T)'$, and where the null hypothesis is $H_0 : \Delta\beta = 0$.

The well-known Chow test is based on the statistic

$$F = \frac{(SS - SS_1 - SS_2)/K}{(SS_1 + SS_2)/(T - 2K)} = \frac{(SS - \widetilde{SS})/K}{\widetilde{SS}/(T - 2K)} \tag{4.4}$$

where SS is the residual sum of squares from (4.2), $\widetilde{SS}$ is the residual sum of squares from (4.3), and where SS_i is the residual sum of squares from regressing Y_i on X_i $(i = 1, 2)$.

For $\tau < K$ or $T - \tau < K$, F cannot be computed, and the statistic

$$F_1 = \frac{(SS - SS_1)/(T - \tau)}{SS_1/(\tau - K)} \tag{4.5}$$

applies.

This chapter is concerned with the robustness of the null distribution of the test statistics (4.4) and (4.5) to autocorrelation among disturbances. The rationale is: if the null distribution were robust to autocorrelation, there would be no need to circumvent this problem, either by first eliminating the disturbance autocorrelation via some Cochrane-Orcutt-type transformation or by searching for more robust alternatives, as White (1980) has done so successfully for heteroscedasticity.

Since (4.4) is the standard F-test for $H_0 : \Delta\beta = 0$ in the model (4.3), general results on the robustness of the F-test to autocorrelated disturbances, dating back to Box (1954), can be brought to bear on the present problem. Unfortunately, the type of autocorrelation considered by Box is rather untypical in economic contexts. Below I follow Corsi *et al.* (1982) in assuming a stationary AR(1) disturbance process:

$$u_t = \rho u_{t-1} + \varepsilon_t \tag{4.6}$$

The null hypothesis is thus enlarged to allow for a nonzero ρ in the range $(-1, 1)$, and the problem can be rephrased as to whether or not the size of the test remains intact when the nuisance parameter ρ takes values different from zero. Corsi *et al.* considered the mean of the test statistic (4.4) and found that it is affected by ρ. They also showed via Monte Carlo that the true size of the test exceeds the nominal one by considerable margins, confirming Box (1954).

Below, I make this more precise and show for many design matrices X that the size of the test equals exactly 1. This means that the rejection probability under H_0 can be made arbitrarily large by an appropriate choice of the nuisance parameter ρ. This in turn serves as a warning that the Chow test might wrongly indicate a structural change when the real culprit is "only" autocorrelation. Some empirical evidence for this is available in,

e.g., Corsi *et al.* (Section 10.4), or Krämer and Sonnberger (1986, Chapter 6.a), where the significance of the Chow test declined drastically after applying a Cochrane-Orcutt transformation to the data.

The result below therefore imply that the Chow test is extremely nonrobust to autocorrelation, calling for amendments similar to those suggested by Jayatissa (1977) in the case of heteroscedasticity. The recent results by Newey and West (1987) may also be used to construct autocorrelation-consistent versions of any F-test, and thus the Chow test in particular. Such issues are, however, beyond the scope of the present chapter.

4.2 Rejection Probabilities under Autocorrelation

Let $\hat{\beta} = (X'X)^{-1}X'Y$, $\hat{\delta} = (Z'Z)^{-1}Z'Y$, and $\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'Y_1$ be the OLS coefficient estimates in (4.2), (4.3), and in the model $Y_1 = X_1\beta_1 + u_1$, respectively. The resulting residuals are denoted by $\hat{u} = Y - X\hat{\beta} = M_xY = M_xu$, $\tilde{u} = Y - Z\hat{\delta} = M_zY = M_zu$, and $\hat{u}_1 = Y_1 - X_1\hat{\beta}_1 = M_1Y_1 = M_1u_1$, where $M_x = I - X(X'X)^{-1}X'$, $M_z = I - Z(Z'Z)^{-1}Z'$, and $M_1 = I - X_1(X_1'X_1)^{-1}X_1$.

The test statistics F and F_1 can then be rewritten as

$$F = \frac{Y'(M_x - M_z)Y/K}{Y'M_zY/(T-2K)} \tag{4.7}$$

and

$$F_1 = \frac{Y'(M_x - M_1)Y/(T-\tau)}{Y'M_1Y/(\tau - K)} \tag{4.8}$$

Let $\langle X\rangle$, $\langle Z\rangle$ and $\langle X_1\rangle$ denote the column spaces of X, Z and X_1 respectively. Let $i = (1,\ldots,1)'$ denote a T-vector (or τ-vector, depending on the context) of 1's, and let $e = (1,-1,\ldots)'$ denote a corresponding vector with the t-th component equal to $(-1)^{t+1}$. My first result concerns the limiting behavior of F and F_1 as $\rho \to 1$ or $\rho \to -1$.

Theorem 4.1 *Under H_0 [i.e., $\Delta\beta = 0$ in (4.3)], the following relationships hold:*

if $i \notin \langle Z\rangle$:

$$\operatorname*{plim}_{\rho\to 1} F = \frac{i'(M_x - M_z)i/K}{i'M_zi/(T-2K)} = \overline{F}^{+} \tag{4.9}$$

if $e \notin \langle Z\rangle$:

$$\operatorname*{plim}_{\rho\to -1} F = \frac{e'(M_x - M_z)e/K}{e'M_ze/(T-2K)} = \overline{F}^{-} \tag{4.10}$$

if $i \notin \langle X_1\rangle$:

$$\operatorname*{plim}_{\rho \to 1} F_1 = \frac{i'(M_x - M_1)i/(T-\tau)}{i'M_1 i/(\tau - K)} = \overline{F}_1^+ \tag{4.11}$$

if $e \notin \langle X_1 \rangle$:

$$\operatorname*{plim}_{\rho \to -1} F_1 = \frac{e'(M_x - M_1)e/(T-\tau)}{e'M_1 e/(\tau - K)} = \overline{F}_1^- \tag{4.12}$$

Proof: Under H_0, we can put u in place of Y in (4.7) and (4.8). From (4.6), we have

$$\sigma_u^2 = (1-\rho)^{-1}\sigma_\varepsilon^2 \tag{4.13}$$

Since σ_u^2 cancels out in both test statistics, we can without loss of generality keep σ_ε^2 fixed as ρ varies, which implies that σ_u^2 tends to infinity as $\rho \to 1$ or $\rho \to -1$.

The point of departure and basic idea is now to note that

$$\operatorname*{plim}_{\rho \to 1} u/u_1 = i \tag{4.14}$$

and

$$\operatorname*{plim}_{\rho \to -1} u/u_1 = e \tag{4.15}$$

where u_1 is the first element of u (any other component would do as well). Equations (4.14) and (4.15) follow immediately from

$$u_t/u_1 = \rho^{t-1} + (\sum_{i=0}^{t-2} \rho^i \varepsilon_{t-i})/u_1 \tag{4.16}$$

and the fact that the second term in (4.16) tends to zero in probability (see Krämer, 1985; or Krämer and Sonnberger, 1986, p.20).

Equations (4.14) and (4.15) imply that the numerators and denominators of F and F_1 tend to the corresponding expression in (4.9)–(4.12). The conditions imposed on i and e ensure that the denominators are nonzero, which completes the proof of the theorem.

Theorem 4.1 implies that, given certain conditions, the test statistics of the Chow test tend to nonstochastic constants as correlation among disturbances increases in absolute value. This in turn has immediate implications for the size of the test as ρ varies.

Let $F_{\alpha,n,m}$ denote the $1-\alpha$ percentile of the central F distribution with n and m degrees of freedom. The next result sheds light on the true size of the Chow test with nominal significance level α.

Theorem 4.2 *The Chow test (4.7) has true size equal to* 1 *if either of the following holds:*

$$i \notin \langle Z \rangle \text{ and } \overline{F}^{+} > F_{\alpha,K,T-2K} \tag{4.17}$$

$$e \notin \langle Z \rangle \text{ and } \overline{F}^{-} > F_{\alpha,K,T-2K} \tag{4.18}$$

Similarly, the Chow-test (4.8) has true size equal to one if either of the following holds:

$$i \notin \langle X_1 \rangle \text{ and } \overline{F}_1^{+} > F_{\alpha,T-\tau,\tau-K} \tag{4.19}$$

$$e \notin \langle X_1 \rangle \text{ and } \overline{F}_1^{-} > F_{\alpha,T-\tau,\tau-K} \tag{4.20}$$

Proof: Immediate from *Theorem* 4.1, since the true rejection probability tends to 1 as $\rho \to 1$ (4.17, 4.19) or $\rho \to -1$ (4.18, 4.20).

The empirical significance of *Theorem* 4.2 depends on the prevalence of conditions (4.17)–(4.20) in practice. For instance, it is easily seen that $i \notin \langle Z \rangle$ (and similarly $i \notin \langle X_1 \rangle$) imply that there must not be a constant in the regression. On the other hand, $e \notin \langle Z \rangle$ or $e \notin \langle X_1 \rangle$ holds more often, and the probability limits of the test statistics exceed the respective critical F-value for many design matrices.

The above analytical results agree well with the Monte Carlo findings of Corsi *et al.* (1982, Tables 10.2 and 10.3), who considered version (4.7) of the test, where $K = 1$, $T = 50$, $\tau = 25$, and where the regressors are likewise generated as AR(1). This implies that both $i \notin M_z$ and $e \notin M_z$ hold with probability 1. Keeping X fixed in repeated runs, Corsi *et al.* found that the true rejection probability increases with increasing ρ when the x_t's are positively correlated, and decreases with increasing ρ when the x_t's are negatively correlated, whereas the opposite occurs when $\rho \to -1$: the true size increases when the x_t's are negatively correlated, and decreases when the correlation among the observations of the independent variable is positive.

This is exactly what one would expect in the light of the above Theorems. Unfortunately, Corsi *et al.* do not reproduce their x-series, so I could not check (4.17) and (4.18). I therefore generated two artifical x-series of length 50 myself, one with positive correlation of 0.8, and one with negative correlation of -0.8. The first one produced $\overline{F}^{+} = 6.83$ and $\overline{F}^{-} = 0.46$, whereas the corresponding figures for the second are $\overline{F}^{+} = 0.57$ and $\overline{F}^{-} = 5.92$.

4.3 A Numerical Example

Section 4.2 has given the limiting rejection probabilities as $|\rho| \to 1$. Next I show for a concrete example what happens for intermediate values of ρ. Exact rejection probabilities are computed with Koerts and Abrahamse's (1969) FQUAD subroutine. To keep things simple, and to facilitate replication of the results, I confine myself to the case $T = 20$,

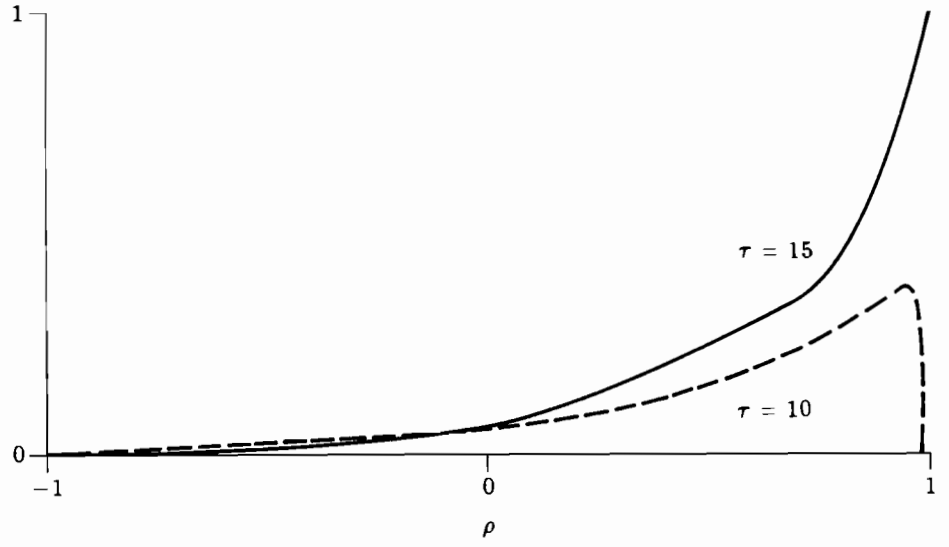

Figure 4.1: True rejection probabilities without a constant.

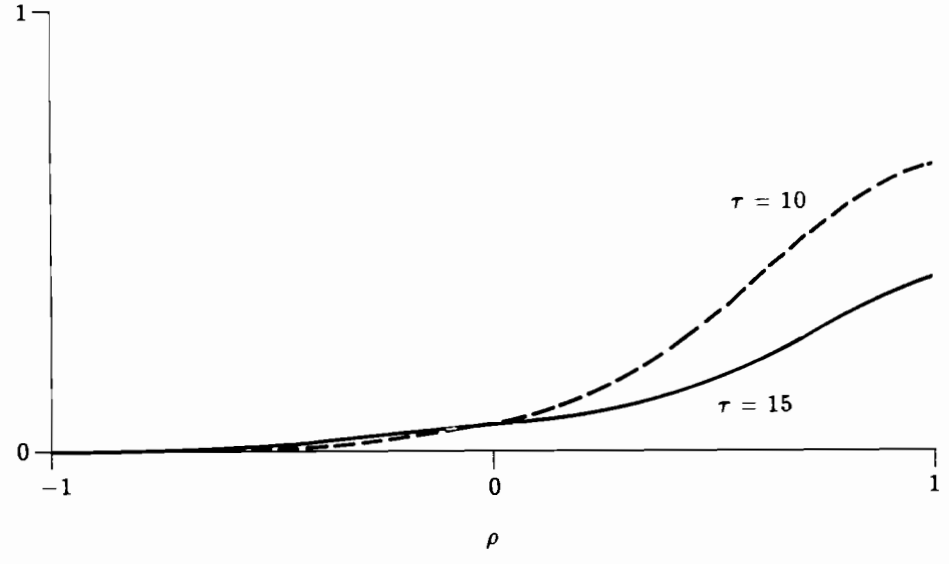

Figure 4.2: True rejection probabilities with a constant.

$x_t = t$ (i.e., $K = 1$ or 2, depending upon whether there is a constant in the regression) and $\tau = 10$ and 15. In addition, I consider only the version (4.7) of the test for ease of comparison with Corsi *et al.* (1982).

Figure 4.1 shows the true rejection probabilities as a function of ρ when there is no constant in the regression, for $\tau = 10$ and 15, given a nominal significance level of $\alpha = 0.05$. In both cases, neither i nor e is in $\langle Z \rangle$. For $\tau = 10$, we have $\overline{F}^+ = 3.94$ and $\overline{F}^- = 0.26$. In view of $F_{0.05;1;18} = 4.41$, this implies that the true rejection probability tends to zero as $\rho \to -1$ and $\rho \to 1$. For $\tau = 15$, we have $\overline{F}^+ = 8.11$ and $\overline{F}^- = 0.02$, so the true rejection probability tends to 1 as $\rho \to 1$.

Figure 4.2 shows the analogous results when there is a constant in the regression. Since i is now in $\langle Z \rangle$, (4.9) does not apply here. However, the true rejection probability tends to zero as $\rho \to -1$ in view of $e \notin \langle Z \rangle$, $\overline{F}^- = 0.19$ ($\tau = 10$), $\overline{F}^- = 0.53$ ($\tau = 15$), and $F_{0.05;2;16} = 3.63$. Although we cannot say much analytically for $\rho \to 1$, *Figure 4.2* indicates that the nominal significance level understates the true rejection probability by a substantial margin when correlation is large and positive.

4.4 Conclusion

This chapter corroborates previous evidence on the nonrobustness of F-tests to autocorrelation among disturbances. In particular, I give sufficient conditions for the true size of the Chow test to be equal to 1. Whether these conditions apply can easily be determined, given a particular design matrix X and a nominal significance level α, in which case a significant Chow test might well be due to autocorrelation. I conclude that the Chow test can be trusted only when there is reason to assume that the disturbances are indeed independent.

Acknowledgments

This chapter originated in a research project financed by the DFG, whose assistance is gratefully acknowledged. Jörg Breitung did most of the work on the numerical examples, and the remaining computations were done with the econometric software package IAS.

References

Box, G.E.P. (1954), Some theorems on quadratic forms applied in the study of analysis of variance problems, II: Effects of inequality of variance and the correlation between errors in the two-way classification. *Annals of Mathematical Statistics*, **30**, 484-498.

Corsi, P., Pollack, R.E., and Prakken, J.C. (1982), The Chow-test in the presence of serially correlated errors, in: G.C. Chow and P. Corsi (eds.), *Evaluating the reliability of macroeconomic models.* New-York: John Wiley.

Jayatissa, W.A. (1977), Tests of equality between two sets of coefficients in two linear regressions when disturbance variances are unequal. *Econometrica*, **45**, 1291-1292.

Krämer, W. (1985), The power of the Durbin-Watson-test for regressions without an intercept. *Journal of Econometrics*, **28**, 363-370.

Krämer, W. and Sonnberger, H. (1986), *The Linear Regression Model under Test.* Heidelberg: Physica-Verlag.

Koerts, J. and Abrahamse, A.P.J. (1969), *On the Theory and Application of the General Linear Model.* Rotterdam: University Press, Rotterdam.

Newey, W.K. and West, K.D. (1987), A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance-matrix. *Econometrica*, **55**, 703-708.

White, H. (1980), A heteroskedasticity-consistent covariance-matrix estimator and a direct test for heteroskedasticity. *Econometrica*, **48**, 817-838.

CHAPTER 5

Tests against Nonconstancy in Linear Models Based on Counting Statistics

Peter Hackl and Walter Katzenbeisser

Summary

Procedures based on various types of counting statistics are considered for testing a sequence of independent random variables against trend alternatives. Power comparisons with standard parametric tests, are also performed, partly by analytical means and partly by Monte Carlo estimation. Some of the tests turn out to be strong competitors to the CUSUM procedure. Finally, the use of the nonparametric tests for the detection of parameter nonconstancy in regression models is discussed.

5.1 Introduction

Let us consider the varying parameter model $y_t = x_t'\beta_t + u_t$, $t = 1, \ldots, T$, where, at time t, y_t is the observation of the dependent variable, x_t is a $(k \times 1)$ vector of nonstochastic regressors, and β_t is a $(k \times 1)$ vector of unknown regression coefficients. The disturbances u_t, $t = 1, \ldots, T$, follow the classical assumption $u_t \sim \text{iid}N(0, \sigma^2)$.

For this general model we want to test the null hypothesis

$$H_0 : \beta_1 = \ldots = \beta_T \tag{5.1}$$

against the general alternative

$$H_1 : \beta_i \neq \beta_j \text{ for at least one pair } (i, j), i \neq j \tag{5.2}$$

Under H_0, the regression (R-)model can be written in the classical form $y = X\beta + u$, where $y = (y_1, \ldots, y_T)'$ and X is a $(T \times k)$ matrix containing the observations of the exogeneous variables. For testing the null hypothesis, a number of procedures is available, which can be classified into:

1. Methods based on overfitting.
2. Methods based on the analysis of residuals.

In this chapter we consider procedures that belong to the second group. This group contains, e.g., the well-known CUSUM and CUSUM-SQ procedures of Brown *et al.* (1975), and the MOSUM procedures of Hackl (1980). These procedures are defined on the basis of recursive (R-) residuals, which are defined by

$$w_t = r_t(y_t - x_t'\hat{\beta}_{t-1}) \tag{5.3}$$

where $\hat{\beta}_{t-1}$ is the OLS estimator of β, based on observations prior to t; and $r_t = (1 + x_t'(X_{t-1}'X_{t-1})^{-1}x_t)^{-1/2}$, where X_{t-1} is the submatrix of X corresponding to the observations prior to t.

The use of R-residuals is motivated by their simple stochastic properties: Under H_0, the $(T-k)$ vector w of R-residuals is normally distributed with a diagonal, nonsingular covariance matrix: $w \sim N(0, \sigma^2 I_{T-k})$, i.e., the w_t are iid$N(0,\sigma^2)$ distributed. On the other hand, in the special case of a two-phase R-model,

$$y_t = \begin{cases} x_t'\beta_1 + u_t & , t = 1, \ldots, \tau \\ x_t'\beta_2 + u_t & , t = \tau + 1, \ldots, T \end{cases} \tag{5.4}$$

with $\beta_2 = \beta_1 + \delta$, the R-residuals can be written as (Hackl, 1980)

$$w_t = E\{w_t\} + \varepsilon_t, t = k+1, \ldots, T \tag{5.5}$$

with uncorrelated error terms ε_t, obeying

$$\left.\begin{array}{rcl} E\{\varepsilon_t\} & = & 0 \\ var\{\varepsilon_t\} & = & \sigma^2 \end{array}\right\} \; t = k+1, \ldots, T \tag{5.6}$$

and expectations

$$E\{w_t\} = \begin{cases} 0 & , t = k+1, \ldots, \tau \\ z_t'\delta & , t = \tau+1, \ldots, T \end{cases} \tag{5.7}$$

where $z_t = r_t x_t'(X_{t-1}'X_{t-1})^{-1}X_\tau'X_\tau$; the covariance matrix of the R-residuals does not show any effect of this type of nonconstancy. Thus, the R-residuals w_t follow for $t \leq \tau$ the null distribution even in the case of nonconstancy; only for $t > \tau$ the distribution of the R-residuals is affected by the nonconstancy.

Example 5.1: Consider the following simple test situation:

$$\begin{aligned} H_0 : y_t &= \beta + u_t \,, \quad t = 1, \ldots, T \\ H_1 : y_t &= \begin{cases} \beta + u_t \,, & t = 1, \ldots, \tau \\ \beta + \delta + u_t \,, & t = \tau + 1, \ldots, T \end{cases} \end{aligned} \tag{5.8}$$

where $u_t \sim \text{iid}N(0, \sigma^2)$. The $(T-1)$ vector w of R-residuals obeys under H_0: $w \sim N(0, \sigma^2 I)$, and under H_1: $w \sim N(\mu, \sigma^2 I)$; the elements of the vector μ being

$$\mu_t = \begin{cases} 0 \,, & t = 2, \ldots, \tau \\ \delta\tau / \sqrt{(t(t-1))} \,, & t = \tau + 1, \ldots, T \end{cases}$$

A further shift of the vector β in (5.4) after, say, $t > \tau$ causes a further summand in (5.7) of analogous structure as $z_t' \delta$.

Testing the constancy of regression coefficients over time by means of an analysis of R-residuals can therefore be embedded in the more general problem of testing a sequence of independent random variables X_i, $i = 1, \ldots, n$, with continuous distribution functions $F_i(x)$ for "randomness". The corresponding null hypothesis is

$$H_0 : F_i(x) = F(x), \quad i = 1, \ldots, n \tag{5.9}$$

Many authors discuss procedures for testing this null hypothesis against suitably chosen "trend" alternatives. Examples of such alternatives are:

1. A sudden shift in the location of the X's:

$$H_1^{(1)} : F_1(x) = \ldots = F_{m-1}(x) = F_m(x - \Delta) = \ldots = F_n(x - \Delta) \tag{5.10}$$

   with, for example, $\Delta > 0$. Typically, the change point is unknown.

2. A trend alternative, i.e., the Xs constitute a stochastically increasing (or decreasing) sequence:

$$H_1^{(2)} : F_1(x) \geq \ldots \geq F_n(x) \tag{5.11}$$

3. An important class of alternatives, called "trend in location" alternative, is described by

$$H_1^{(3)} : F_i(x) = F(x - \Delta_i), \quad i = 1, \ldots, n \tag{5.12}$$

   where, e.g., $0 < \Delta_1 \leq \ldots \leq \Delta_n$, with at least one strict inequality — here, the Δs are unknown location shift parameters.

In the context of the R-model, simple alternatives, such as (5.10) to (5.12), will in general not adequately describe the effect of nonconstant regression coefficients. As is seen in the case of *Example 5.1*, the shift δ of the intercept causes a sequence of location parameters $\mu_t \neq 0$ for $t > \tau$ with equal signs and a decreasing trend. In the general R-model, the pattern of these location parameters will be more complex.

There are numerous parametric as well as nonparametric procedures available for testing randomness in a sequence against "trend" alternatives; for a survey of nonparametric procedures see, e.g., Bhattacharyya (1984). In the following sections, nonparametric test statistics, based on counts, will be discussed; these are attractive for applications owing to their simplicity.

The chapter is organized as follows:

1. After a presentation of some nonparametric procedures for testing randomness of a sequence of random variables against shift alternatives (Section 5.2), we discuss distributional properties of the respective test statistics (Section 5.3) and report some power calculations (Section 5.4).

2. Some of these procedures are applied in testing the constancy of regression coefficients over time, and power comparisons with the familiar CUSUM procedure based on a simple R-model will be reported (Section 5.5).

5.2 The Test Procedures

In this section we report some nonparametric procedures for testing randomness in a sequence against the alternative of a "trend". The test procedures are based on simple counting statistics and can be classified as follows:

1. Tests based on N_n = [number of outstanding variables among the Xs]. The variable X_k is called outstanding if $X_k > \max\{X_1, \ldots, X_{k-1}\}$; X_1 is outstanding. A test based on N_n was suggested by Brunk (1969), with reference to a theorem derived by Andersen (1954).

2. The second group contains test procedures, based on the signs of differences of some or all pairs (X_i, X_j), which have the general form

$$T = \Sigma_{i,j} c_{ij} \psi_{ij}$$

with

$$\psi_{ij} = \begin{cases} 1 & \text{if } X_i < X_j \\ 0 & \text{elsewhere} \end{cases}$$

and weights $c_{ij} \geq 0$. Some prominent members of this group are the tests recommended by Mann (1945), Daniels (1950), Cox and Stuart (1955), and Wolfe and Schechtman (1984), which are based on the respective test statsistics

$$\begin{aligned} M &= \Sigma\Sigma_{i<j}\psi_{ij} \\ D &= \Sigma\Sigma_{i<j}(j-i)\psi_{ij} \\ CS_1 &= \sum_{i=1}^{n/2}(n-2i+1)\psi_{i,n-i+1} \\ CS_2 &= \sum_{i=1}^{n/3}\psi_{i,\frac{2}{3}n+i} \\ WS_1 &= \Sigma_i\Sigma_j(i-1)\psi_{ij} \end{aligned}$$

The WS_1 test can be seen as a special case of a nonparametric test proposed by Bhattacharyya and Johnson (1968) and is similar to a linear rank statistic with Wilcoxon scores.

3. Another test proposed by Wolfe and Schechtman (1984) is based on the test statistic

$$WS_2 = \sum_{i=1}^{n}(i-1)\psi\{X_i - \text{median}_{1\leq j\leq n}\{X_j\}\}$$

where $\psi\{t\} = 1$ and 0 if $t > 0$ and $t \leq 0$, respectively.

4. The fourth group of tests considered are based on Wilcoxon- and median-type statistics. In the notation of Wolfe and Schechtman (1984), let

$$U_{k,n-k} = \sum_{i=k+1}^{n}\sum_{j=1}^{k}\psi\{X_i - X_j\}$$

and

$$M_{k,n-k} = \sum_{i=k+1}^{n}\psi\{X_i - \text{median}_{i\leq j\leq n}\{X_j\}\}$$

Sen and Srivastava (1975) suggested tests based on

$$D_1 = \max_{1\leq k\leq n-1}\{(M_{k,n-k} - E_0\{M_{k,n-k}\})/\sqrt{\text{Var}_0\{M_{k,n-k}\}}\}$$

and

$$D_2 = \max_{1\leq k\leq n-1}\{(U_{k,n-k} - E_0\{U_{k,n-k}\})/\sqrt{\text{Var}_0\{U_{k,n-k}\}}\}$$

where $E_0\{M_{k,n-k}\}$ and $\mathrm{Var}_0\{M_{k,n-k}\}$ are the null expectation and variance, respectively, of $M_{k,n-k}$. Pettitt (1979) considered a test based on

$$\begin{aligned} K &= \max_{1\le k\le n-1}\{\sum_{i=1}^{k}\sum_{j=k+1}^{n} Q_{ij}\} = \\ &= 2\max_{1\le k\le n-1}\{U_{k,n-k} - k(n-k)/2\} \end{aligned}$$

where $Q_{ij} = sign\{X_j - X_i\}$.

5.3 Distributional Properties

In this section distributional properties of the respective test statistics will be discussed.

5.3.1 Random variable N_n

To obtain the distribution of the random variable N_n, the following *Lemma 5.1* will be useful:

Lemma 5.1 : *Let X_i, $i = 1,\ldots,n$, be a sequence of independent random variables with continuous distribution functions $F_i(x)$ and density functions $f_i(x)$. Then, the probability $P\{X_k$ is outstanding$\} := \varphi_k$ is given by*

$$\varphi_k = \int_{-\infty}^{\infty}\prod_{j=1}^{k-1} F_j(z)f_k(z)dz\ ,\quad k = 1,\ldots,n$$

Proof: Let $\varphi_k = P\{X_k$ is outstanding$\} = P\{X_k > \max\{X_1,\ldots,X_{k-1}\}\} = P\{Y > 0\}$, where $Y = X_k - \max\{X_1,\ldots,X_{k-1}\}$. Because of the independence of the Xs, the joint density of X_k and $max\{X_1,\ldots,X_{k-1}\} = Z$ is

$$f_{X_k,Z}(x,z) = f_k(x)\sum_{i=1}^{k-1} f_i(z)\prod_{j=1,j\neq i}^{k-1} F_j(z)$$

The transformation $Y = X_k - Z$ and $Z = Z$ with Jacobian unity yields

$$\begin{aligned} P\{Y>0\} &= \sum_{i=1}^{k-1}\int_{-\infty}^{\infty} f_i(z)[1-F_k(z)]\prod_{j=1,j\neq i}^{k-1} F_j(z)dz = \\ &= \int_{-\infty}^{\infty}\prod_{j=1}^{k-1} F_j(z)f_k(z)dz \end{aligned}$$

With the aid of *Lemma 5.1*, the null distribution of N_n can be derived as follows: If $F_1(x) = \ldots = F_n(x) = F(x)$, we get

$$\varphi_k = \int_{-\infty}^{\infty} F(z)^{n-1} f(z) dz$$

The substitution $F(z) = u$ yields

$$\varphi_n = 1/n$$

This result has been derived by Renyi (cf. Gupta and Panchapakesan, 1973, p. 237f.).

Theorem 5.1 : *Let X_i, $i = 1, \ldots, n$, be a sequence of independent random variables with identical continuous distribution function $F(x)$. The null distribution of N_n is given by $P\{N_n = k\} := p_{n,k}$ and the $p_{n,k}$s obey the recursion*

$$p_{n,k} = (1 - \varphi_n) p_{n-1,k} + \varphi_n p_{n-1,k-1}$$

with initial conditions $p_{n,0} = 0$, $n \geq 1$; $p_{1,1} := 1$; $p_{n,k} = 0$, $k > n$.

Proof: Consider the decomposition $\{N_n = k\} = \{N_{n-1} = k$ and X_n is not outstanding$\} \bigcup \{N_{n-1} = k - 1$ and X_n is outstanding$\}$. Because of the iid property of the Xs, we immediately get the result.

The use of $\varphi = 1/n$ gives

$$p_{n,k} = \frac{n-1}{n} p_{n-1,k} + \frac{1}{n} p_{n-1,k-1}$$

Introduction of the probability generating function (pgf)

$$P_n(z) = \Sigma_{k \geq 0} p_{n,k} z^k$$

leads to the recursion

$$P_n(z) = \frac{z + n - 1}{n} P_{n-1}(z)$$

with $P_0(z) = 1$. By iteration, the pgf for N_n is obtained:

$$P_n(z) = \binom{z + n - 1}{n} = \frac{1}{n!} \Sigma_{k \geq 0} S(n,k) z^k$$

where $S(n,k)$ denotes the Stirling number of the first kind. Therefore, under the null hypothesis

$$P_0\{N_n = k\} = \frac{1}{n!} S(n,k)$$

5.3.2 Test statistics M and D

The null distribution of the test statistics M and D can be derived by considering the pairs (i, X_i), $i = 1, \ldots, n$, as a sequence of bivariate random variables. The statistic M is related to Kendall's τ, whereas the statistic D is related to Spearman's rank correlation coefficient. Consider first $M = \Sigma_i \Sigma_{j>i} \varphi_{ij}$ which is the number of pairs (X_i, X_j) with $i < j$ and $X_i < X_j$. By means of Q = [number of inversions of a random permutation of n], M can be rewritten as

$$M = \binom{n}{2} - Q$$

it follows

$$P_0\{M = k\} = P_0\{Q = \binom{n}{2} - k\}$$

Denote by $b(n, k)$ the number of permutations of n elements with k inversions. It is well known that under the null hypothesis

$$P_0\{Q = k\} = \frac{1}{n!} b(n, k)$$

where the $b(n, k)$s satisfy the relations $b(n, k) = b(n-1, k) + b(n, k-1)$, $k < n$, and $b(n, k) = b(n, \binom{n}{2} - k)$ (cf. Comtet, 1974, p.239f.), which are useful for computing the critical values. On the other hand, Kendall's τ is defined as $\tau = 1 - 2Q/\binom{n}{2}$, and therefore

$$P_0\{M = k\} = P_0\{\tau = 2k/\binom{n}{2} - 1\}$$

Tables of the null distribution of Kendall's τ can be used to obtain critical values for M.

Spearman's rank correlation coefficient is defined as

$$r_s = \frac{12}{n^3 - n} \sum_{i=1}^{n} \left(i - \frac{n+1}{2}\right) R_i$$

where $R_1, \ldots, R_n$ denote the ranks of $X_1, \ldots, X_n$. It can be shown (cf. Bhattacharyya, 1984, p. 92) that

$$D = \frac{n^3 - n}{12}(1 + r_s)$$

Therefore,

$$P_0\{D = k\} = P_0\{r_s = \frac{12k}{n^3 - n} - 1\}$$

and tables of the null distribution of Spearman's correlation coefficient can be used to obtain critical values for D.

The null distribution of the statistics CS_1 and CS_2 can be derived on the basis of their respective pgf's: CS_2 is based on

$$\psi_{i,\frac{2}{3}n+i} = \begin{cases} 1 & \text{if } X_i < X_{\frac{2}{3}n+i} \\ 0 & \text{elsewhere} \end{cases}$$

for $i = 1, \ldots, n/3$, with $\psi_{i,\frac{2}{3}n+i} \sim B(1, p_{i,\frac{2}{3}n+i})$ where $p_{i,\frac{2}{3}n+i} = P\{X_i < X_{\frac{2}{3}n+i}\}$ and $B(n,p)$ denotes the binomial distribution with parameters n and p. The corresponding pgf is

$$P_i(z) = 1 - p_{i,\frac{2}{3}n+i} + p_{i,\frac{2}{3}n+i} z$$

Because of the independence of the Xs (and therefore of the ψs), the pgf of CS_2

$$P_{CS_2}(z) = \prod_{i=1}^{n/3} P_i(z) \tag{5.13}$$

can be used to derive the distribution of CS_2:

$$P\{CS_2 = k\} = [z^k] P_{CS_2}(z) := P_{CS_2}(k)$$

Here $[z^k]P(z)$ denotes the coefficient of $[z^k]$ in $P(z)$. Under the null hypothesis

$$p_{i,\frac{2}{3}n+i} = 1/2$$

and CS_2 follows

$$CS_2 \sim B(n/3, 1/2).$$

Consider next the statistic CS_1. Let us define $Z_{i,n-i+1} = (n - 2i + 1)\psi_{i,n-i+1}$, so that

$$Z_{i,n-i+1} = \begin{cases} n - 2i + 1 & \text{if } X_i < X_{n-i+1} \\ 0 & \text{elsewhere} \end{cases}$$

The Zs are independent random variables with distribution

$$\begin{aligned} P\{Z_{i,n-i+1} = n - 2i + 1\} &= p_{i,n-i+1} \\ P\{Z_{i,n-i+1} = 0\} &= 1 - p_{i,n-i+1} \end{aligned}$$

where $p_{i,n-i+1} = P\{X_i < X_{n-i+1}\}$. The corresponding pgf is

$$Q_i(z) = 1 - p_{i,n-i+1} + p_{i,n-i+1} z^{n-2i+1}$$

because of the independence of the Xs (and therefore of the Zs). Again the pgf for CS_1

$$Q_{CS_1}(z) = \prod_{i=1}^{n/2} Q_i(z) \tag{5.14}$$

can be used to derive the distribution of CS_1:

$$P\{CS_1 = k\} = [z^k] Q_{CS_1}(z) := P_{CS_1}(k) \tag{5.15}$$

Under the null hypothesis, $p_{i,n-i+1} = 1/2$, and the pgf for CS_1 is given by

$$Q_{CS_1}(z) = 2^{-n/2} \prod_{i=1}^{n/2} (1 + z^{n-2i+1})$$

For small values of n, the coefficients $[z^k]P(z)$ can be obtained by multiplication of the relevant factors in (5.13) and (5.14).

5.3.3 Other test statistics

The null distributions of the other test statistics, viz., WS_1, WS_2, D_1, D_2, and K, can hardly be derived analytically. To obtain critical values, the permutational distribution of these statistics can be used; for an example, cf. Schechtman (1982). This procedure, however, is prohibitively cumbersome for any reasonably large sample size.

Pettitt's test statistic has asymptotically the distribution of the Kolmogorov-Smirnov one sample test statistic, so that approximate critical limits can be obtained. Results for the asymptotic case are discussed by Sen (1978). In the sequel, Monte Carlo estimates for the critical values were used.

5.4 The Power of the Tests against Trend Alternatives

In this section some power comparisons are reported. The first intention was to study the power of the mentioned tests against various "trend" alternatives. The null hypothesis (5.9) is to be tested against the alternatives expressed by means of suitably chosen shift parameters:

1. A shift in location ("shift" alternative)

$$H_1^{(1)}: \begin{cases} F_i(x) = F(x) & i = 1, \ldots, m \\ F_i(x) = F(x - \Delta) & i = m+1, \ldots, n \end{cases}$$

with $m = n/2$.

2. A trend in location ("trend" alternative)

$$H_1^{(2)}: \quad F_i(x) = F(x - c_i\Delta), \quad i = 1, \ldots, n$$

The actual choice for c_i is $i - 1$, $i = 1, \ldots, n$; other specifications for the c_is can be used, e.g., for the case with more than one change point.

Under H_0, the distributions of the respective test statistics are independent of the distributional assumptions for the Xs; however, this is not true under H_1. In view of our principal goal, viz., testing the constancy of regression coefficients over time, we base our considerations on sequences of normally distributed random variables: under H_0, the Xs are assumed to be iid$N(0,1)$ distributed.

As the alternatives are one-sided, H_0 has to be rejected in favor of H_1 if $T \geq z_\alpha$ for the respective test statistics T, where z_α is a level α critical value. As the test statistics are discrete random variables, randomization was used to assure that all tests achieve the desired significance level; the power function will be denoted by $P(\Delta)$.

5.4.1 Exact power functions

To obtain the exact power functions for the tests based on CS_1 and CS_2 we rely upon the useful

Lemma 5.2 *Let X and Y be independent and continuously distributed random variables with distribution functions F and G and density functions f and g, respectively. Then*

$$P\{X < Y\} = \int_{-\infty}^{\infty} g(y)F(y)dy \tag{5.16}$$

In the case of the "shift" alternative, *Lemma 5.2* leads to

$$p_{i,n-i+1} = p_{i,\frac{2}{3}n+i} = \Phi\left(\frac{\Delta}{\sqrt{2\sigma^2}}\right) \tag{5.17}$$

For the "shift" alternative, the power function for CS_1 can be obtained from $P(\Delta) = \Sigma_{k \geq z_\alpha} p_{CS_1}(k)$, where $p_{CS_1}(k)$ is given in (5.15).

The power function of CS_2 can be derived by observing that, under the "shift" alternative, the $p_{CS_2}(k)$s are independent of i, and CS_2 follows the distribution $B(n/3, p_{i,\frac{2}{3}n+i})$. Therefore, under the "shift" alternative,

$$P(\Delta) = \Sigma_{k \geq z_\alpha} \binom{n/3}{k} p_{i,\frac{2}{3}n+i}^k (1 - p_{i,\frac{2}{3}n+i})^{n/3-k} \tag{5.18}$$

For the "trend" alternative, *Lemma 5.2* gives

Table 5.1: Estimated power functions for the tests N_n, M, D, WS_2, and K and exact power functions for the tests CS_1 and CS_2 against shift alternatives ($n = 20$, $\alpha = 0.05$):

$$H_0: \quad X_i \sim \text{ iidN}(0,1)\ , i = 1,\ldots,20$$
$$H_1: \quad X_i \sim \begin{cases} \text{iidN}(0,1)\ , & i = 1,\ldots,10 \\ \text{iidN}(\Delta,1)\ , & i = 11,\ldots,20 \end{cases}$$

Test	Δ					
	0.5	1.0	1.5	2.0	2.5	3.0
N_n	0.093	0.172	0.238	0.285	0.342	0.399
CS_1	0.188	0.440	0.721	0.891	0.970	0.994
CS_2	0.149	0.326	0.534	0.724	0.861	0.936
M	0.209	0.514	0.820	0.946	0.988	0.992
D	0.213	0.539	0.846	0.963	0.996	1.000
WS_2	0.179	0.463	0.734	0.910	0.970	0.998
K	0.217	0.583	0.887	0.978	1.000	1.000

$$p_{i,n-i+1} = \Phi\left(\frac{(n-2i+1)\Delta}{\sqrt{2\sigma^2}}\right)$$

and

$$p_{i,\frac{2}{3}n+i} = \Phi\left(\frac{\sqrt{2n}\Delta}{3\sigma}\right)$$

Here again, for the test based on CS_1, $P(\Delta) = \Sigma P_{CS_1}(k)$, where the $P_{CS_1}(k)$s are given in (5.15). For the test based on CS_2, the ps are independent of i and, therefore, $P(\Delta)$ can be calculated using (5.18).

Some values of the exact power functions for the tests based on CS_1 and CS_2 are given in *Tables 5.1* and *5.2*

5.4.2 Monte Carlo experiments

Because we could not derive the exact distributions of the other test statistics under either H_0 or H_1, we performed Monte Carlo studies in order to estimate their respective power functions.

For the tests based on N_n, CS_1, CS_2, M, and D exact critical values were used; for the tests based on WS_2 and K we used critical values estimated from 10,000 replications of the sampling experiment. The pseudo-random numbers were generated using the SAS routine RANNOR. For the test based on K, the estimated critical values show a very good agreement with those obtained via the asymptotic distribution given by Pettitt (1979). Values of the power functions were estimated for $n = 20$ on the basis of 1,000 replications

Table 5.2: Estimated power functions for the tests N_n, M, D, WS_2, and K and exact power functions for the tests CS_1 and CS_2 against trend alternatives ($n = 20$, $\alpha = 0.05$):

H_0: $X_i \sim \text{iidN}(0,1)$, $i = 1,\ldots,20$
H_1: X_i are independent $\text{N}[(i-1)\Delta, 1]$, $i = 1,\ldots,20$

Test	Δ				
	0.05	0.1	0.2	0.3	0.4
N_n	0.128	0.281	0.664	0.895	0.970
CS_1	0.249	0.585	0.959	0.998	1.000
CS_2	0.210	0.494	0.911	0.994	0.996
M	0.321	0.768	1.000	1.000	1.000
D	0.319	0.782	1.000	1.000	1.000
WS_2	0.264	0.634	0.979	1.000	1.000
K	0.287	0.697	0.994	1.000	1.000

(cf. *Tables 5.1* and *5.2*). Analogous but not reported results, as mentioned above, were obtained for $n = 40$.

5.4.3 Results

Estimates of the power function are given in *Tables 5.1* and *5.2*; graphs of the power functions are shown in *Figures 5.1* and *5.2*. They represent exact values for the tests based on CS_1 and CS_2 and Monte Carlo estimates for the other procedures. The following conclusions can be drawn:

1. In all experiments, the tests based on K, M, and D are of about the same power and more powerful than the other tests. A second group of tests of about equal power contains the tests based on WS_2 and CS_1. The tests based on CS_2 and N_n are in all experiments of less power than the others.

2. The performance of the tests is not uniform for the two alternatives; all tests considered are more powerful for the "trend" alternative.

3. The test based on N_n is uniformly less powerful as compared with the other tests.

5.5 The Power of the Tests against Nonconstancy in R-Models

To assess the performance of the tests for testing the constancy of regression coefficients over time, a further Monte Carlo experiment was conducted.

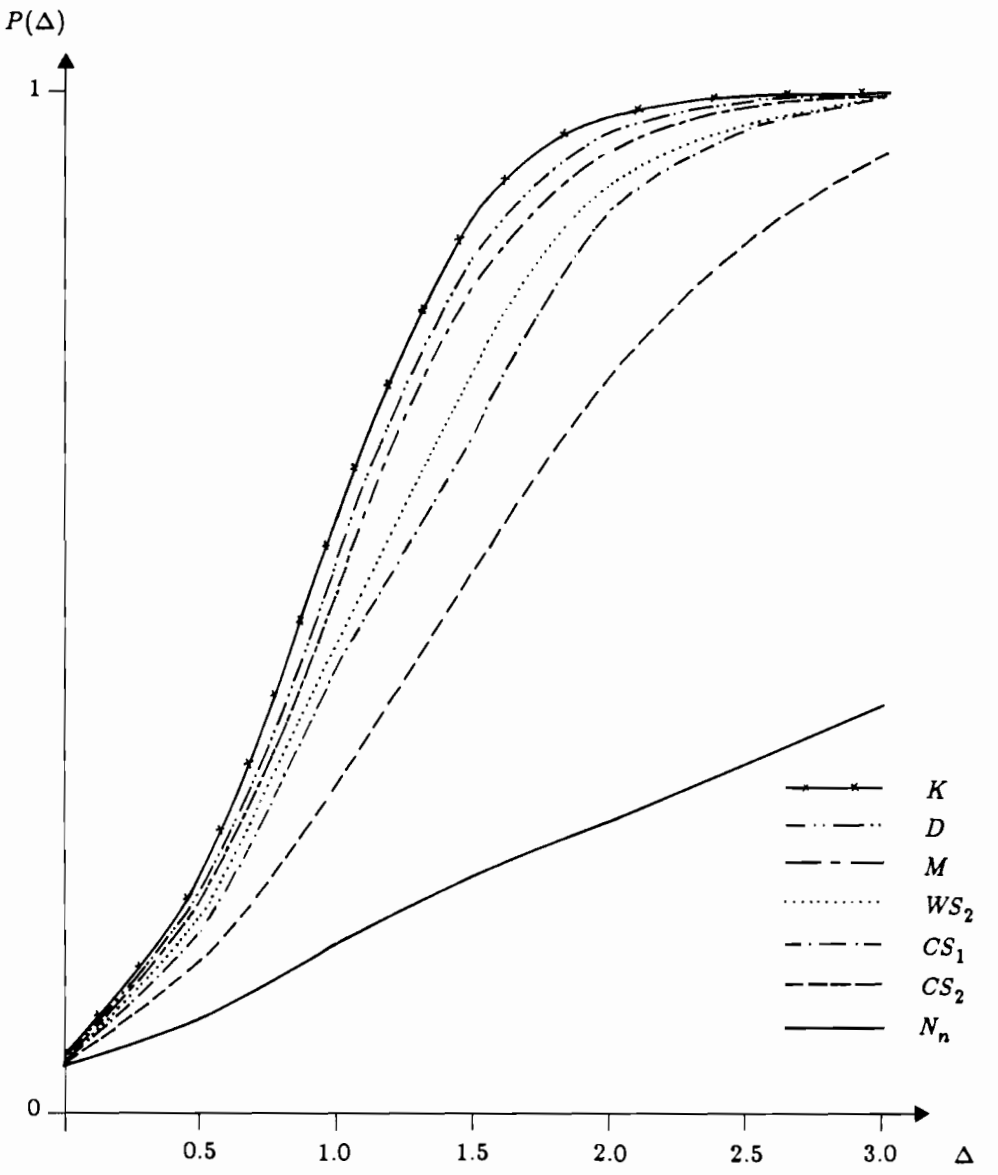

Figure 5.1: Power function for the tests N_n, M, D, WS_2, and K (estimated) and for the tests CS_1 and CS_2 (exact): Testing against shift alternatives.

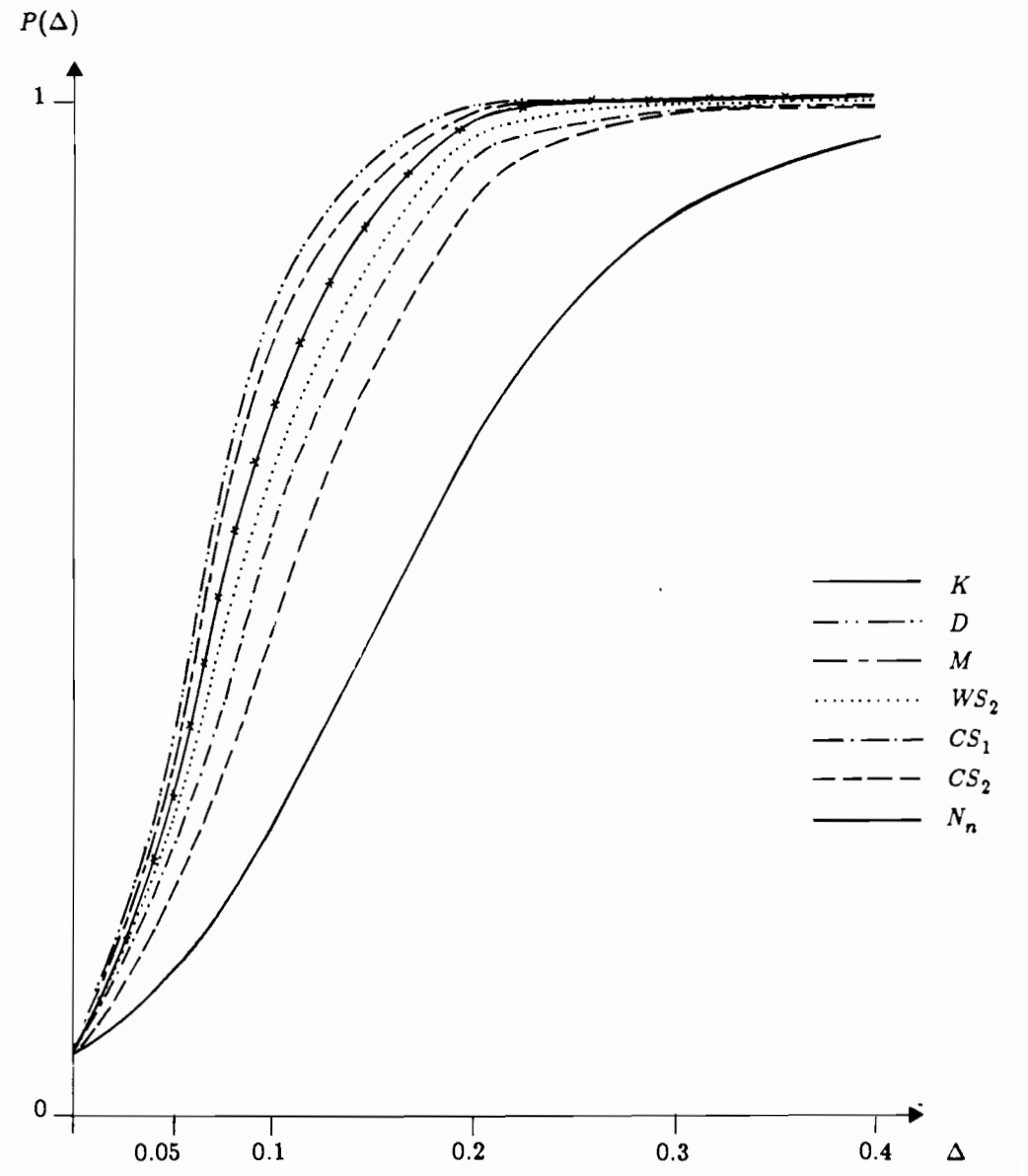

Figure 5.2: Power function for the tests N_n, M, D, WS_2, and K (estimated) and for the tests CS_1 and CS_2 (exact): Testing against trend alternatives.

Table 5.3: Estimated power functions for the tests N_n, CS_1, CS_2, M, D, WS_2, and K for testing the constancy of regression relationships over time ($\alpha = 0.05$):

$$H_0: \quad y_t = \beta + u_t \,, t = 1, \ldots, 20$$
$$H_1: \quad y_t = \begin{cases} \beta + u_t \,, & t = 1, \ldots, \tau \\ \beta + \delta + u_t \,, & t = \tau + 1, \ldots, 20 \end{cases}$$

with $u_t \sim \text{iid}N(0, \sigma^2)$ and $\beta = 0$.

σ^2	δ	τ	*Tests*								
			C	C_m	N_n	CS_1	CS_2	M	D	WS_2	K
1	1	10	0.141	0.163	0.000	0.150	0.112	0.185	0.216	0.150	0.395
1	2	10	0.446	0.481	0.000	0.426	0.325	0.486	0.589	0.417	0.703
1	2	5	0.572	0.600	0.000	0.280	0.263	0.308	0.400	0.282	0.536
1	2	15	0.075	0.095	0.001	0.365	0.319	0.465	0.549	0.315	0.593
2	2	10	0.148	0.174	0.027	0.186	0.136	0.260	0.301	0.230	0.322

5.5.1 Procedure

We base our considerations on the model (5.8) of *Example 5.1*, where $u_t \sim \text{iid}N(0, \sigma^2)$, $\sigma^2 = 1$, and, without loss of generality, $\beta = 0$, was used. In varying the parameters δ, σ^2, and the change point τ, various types of nonconstancy were considered (cf. *Table 5.3*). In addition to the nonparametric tests, the familiar CUSUM procedure, denoted by C, was, as a prominent parametric competitor, included in this study. Moreover, a modified version of the CUSUM test, C_m, was used; in this modification (Hackl, 1986) the individual cumulative sums are tested simultaneously, the critical limits being adjusted by means of Hunter's (1976) inequality. The pseudo-random numbers again were generated by the SAS routine RANNOR, and the power estimates are based on 1,000 replications of the sampling experiment.

The alternatives considered in Section 5.4 are one-sided. When testing the constancy of regression coefficients over time, two-sided alternative hypotheses must be used. In the two-sided versions of the nonparametric tests based on N_n, CS_1, CS_2, M, D, and WS_2, the null hypothesis was rejected if both small or large values were observed. In the two-sided version of Pettitt's test, the null hypothesis was rejected if $\max \Sigma\Sigma Q_{ij}$ was large or $\min \Sigma\Sigma Q_{ij}$ was small.

5.5.2 Results

The results of the Monte Carlo experiments, shown in *Table 5.3*, suggest the following conclusions:

1. For the pure shift alternatives ($\delta > 0$, $\sigma^2 = 1$ for $t > \tau$), the tests based on the simple counting statistics K, M, and D show very good power, compared with the other tests including the CUSUM procedures.

2. It is well known that the position of the change point influences the power of the CUSUM test. Shift points near the beginning of the sampling period lead to more powerful tests; the power is reduced drastically, if the shift point is moved towards the end of the sampling period. This pattern is again shown in *Table 5.3*. Only for $\tau = T/4$, the CUSUM procedures are more powerful than the best nonparametric procedure, the Pettitt test. The nonparametric tests show most power if $\tau = T/2$. The decrease of power if τ changes from $T/2$ to $T/4$ or $3T/4$ turns out to be not so drastic as for CUSUM procedures.

3. An additional change in the variance of the disturbances ($\sigma^2 = 2$ for $t > \tau$) leads to a reduction of the power as compared to the change of the intercept only. This behavior is well known in connection with the CUSUM test; the nonparametric tests show similar behavior. However, the decrease of power is again not so drastic for the nonparametric tests.

5.6 Summary

The purpose of this chapter was to investigate the performance of some nonparametric tests, based on simple counting statistics, for the problem of testing the constancy of regression coefficients over time. The power comparisons suggest that, at least, the tests based on the statistics M, D, and K are very strong competitors to the well-known CUSUM procedure. However, some possible drawbacks should be noted:

1. The results are based on a very simple R-model. An obvious question is how strongly these results depend on this model.

2. The exact distribution of the statistic K and WS_2 is not known, and estimated critical values were used. The effect of the estimation of the critical values cannot be assessed.

Acknowledgment

We are grateful to Professor Pranab K. Sen for his interest and his valuable comments.

References

Andersen, E.S. (1954), On the fluctuation of sums of random variables. *Mathematica Scandinavica*, **2**, 171-181.

Bhattacharyya, G.K. (1984), Tests for randomness against trend or serial correlation, in: P.R. Krishnaiah and P.K. Sen (eds.), *Nonparametric Methods, Handbook of Statistics*, Vol. 4. Amsterdam: North-Holland.

Bhattacharyya, G.K., and Johnson, R.A. (1968), Nonparametric tests for shift at an unknown time point. *Annals of Mathematical Statistics*, **39**, 1731-1743.

Brown, R.L., Durbin, J., and Evans, J.M. (1975), Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society*, **B-37**, 149-192.

Brunk, H.D. (1960), On a theorem of E. Sparre Andersen and its applications to tests against trend. *Mathematica Scandinavica*, **8**, 305-326.

Comtet, L. (1974), *Advanced Combinatorics*. Dordrecht and Boston: D. Reidel.

Cox, D.R. and Stuart, A. (1955), Some quick sign tests for trend in location and dispersion. *Biometrika*, **42**, 80-95.

Daniels, H.E. (1950), Rank correlation and population models. *Journal of the Royal Statistical Society*, **B-12**, 171-181.

Gupta, S.S. and Panchapakesan, S. (1973), On order statistics and some applications of combinatorial methods in statistics, in: J.N. Srivastava *et al.* (eds.), *A Survey of Combinatorial Theory*. Amsterdam: North-Holland.

Hackl, P. (1980), *Testing the constancy of regression models over time*. Göttingen: Vandenhoeck & Ruprecht.

Hackl, P. (1986), *A Note on the critical limits of the CUSUM test*. Working Paper WP-86-26. Laxenburg: International Institute for Applied Systems Analysis.

Hunter, D. (1976), An upper bound for the probability of a union. *Journal of Applied Probability*, **13**, 597-603.

Mann, H.B. (1945), Nonparametric tests against trend. *Econometrica*, **13**, 245-259.

Pettitt, A.N. (1979), A nonparametric approach to the change point problem. *Applied Statistics*, **28**, 126-135.

Sen, P.K. (1978), Invariance principles for linear rank statistics revisited. *Sankhya*, **A-40**, 215-236.

Sen, A., and Srivastava, M.S. (1975), On tests for detecting changes in the mean. *Annals of Statistics*, **3**, 98-108.

Schechtman, E. (1982), A nonparametric test for detecting changes in location. *Communications in Statistics — Theory and Methods*, **11**, 1475-1482.

Wolfe, D.A., and Schechtman, E. (1984), Nonparametric statistical procedures for the changepoint problem. *Journal of Statistical Planning and Inference*, **9**, 389-396.

CHAPTER 6

Nonparametric Tests for Shift and Change in Regression at an Unknown Time Point

Marie Hušková and Pranab K. Sen

Summary

A variety of nonparametric and robust tests for the "change point" model relating to the location as well as regression problems is available in the literature. The main objective of the current study is to provide a broad coverage of the main stream of these developments, encompassing both rank (R-) and maximum likelihood type (M-) procedures in a recursive as well as nonrecursive setup. Along with some interpretations of (asymptotic) optimality properties of nonparametric and robust tests for the change-point problems, suitable adaptive procedures are suggested, which achieve this optimality in a meaningful sense.

6.1 Introduction

Let $X_1, \ldots, X_n$ be n independent random variables observed at ordered time points $t_1, \ldots, t_n$, respectively. We denote the distribution function (d.f.) of X_i by F_i, $i = 1, \ldots, n$, and assume that all the F_is belong to a common family $\mathcal{F}$.

In a conventional statistical inference model, one generally assumes that all the F_is are the same (or they are structurally related to a common F with known algebraic constants and unknown parameters) and within this framework, one seeks to draw statistical inference on F or its parameters). There are, however, problems [typically arising in continuous sampling inspection plans (viz., Page, 1957) or in some other economic problems

(viz., Chapter 21 in this volume)] where a change of the d.f. (or its associate parameters) may occur at some unknown time point τ (the *change-point*). Hence, one may desire to test for the equality of the F_i (or the associated parameters) against a composite alternative that a change has occurred. The change point τ may assume the values $t_1, \ldots, t_n$; we may exclude the case of $\tau = t_1$, as then the homogeneity is not violated. It is, of course, not necessary to assume that τ equals one of the n time points $t_1, \ldots, t_n$. It suffices to take $\tau \in (t_k, t_{k+1}]$, for some $k = 1, \ldots, n-1$.

The *change-point problem*, as has been posed above, is essentially a *nonsequential* one (where $t_1, \ldots, t_n$ and n are all given). However, a variant of this model relates to the so called *sequential detection problem* (viz., Shirjajev, 1963, 1978), where one encounters an infinite sequence $\{X_i; i \geq 1\}$ of random variables (r.v.), gathered over another sequence $\{t_i; i \geq 1\}$ of ordered time points, such that for some integer K (possibly equal to $+\infty$), the $F_i, i \leq K$ are homogeneous and the $F_i, i \geq K+1$ are homogeneous (in the sense of the classical change point model), but F_K and F_{K+1} are not the same. In this context, one would like to choose a *stopping number* N, such that if K is actually finite (i.e., $K < \infty$), then the excess of N over K should be small (in a well-defined sense), while if $K = +\infty$, then the probability of a *false alarm* (i.e., $N < +\infty$) should be small. Thus the sequential detection problem may be regarded as a variant of the change point model where n is not specified in advance.

In the context of statistical and economic identification of structural change, one generally has a finite (and prespecified) time domain, and, hence, the change point is more relevant than the sequential detection problem. Whereas in the sequential detection problem N needs to be defined in a precise manner and the testing procedures are genuinely sequential in nature, in the classical change point model the tests need not be (quasi-) sequential in nature. Nevertheless, *recursive estimates* (arising typically in the sequential detection problem) may be used with advantage for the change point model, too. Thus, we shall confine ourselves to nonparametric and robust tests for the change point model based on recursive as well as terminal estimators.

The change point model for the *"shift" alternative with a specified location* is introduced in Section 6.2. This model is more likely to arise in quality control problems where the targets are specified in advance. In the current context of statistical and economic identification of structural change, even for the shift model, we generally have an *unknown location*, and this case, deserving more detailed analysis is treated in Section 6.3. For both the models, suitable rank tests and M-tests are considered, and in that context the choice of optimal score functions is also discussed. In the parametric case, as a natural extension of the "shift" model, one may consider a more general linear model (with the regression vectors identified from extraneous factors governing the economic-statistical structures), and this is generally referred to as the "*constancy of regression relationships over time*" model [viz., Brown *et al.*, (1975) , Hackl (1980), and others]. Nonparametric and robust counterparts are discussed in Section 6.4.

In each section along with a basic review of the existing tests, emphasis has been placed on the optimality properties in some well-defined manner. Some general remarks are made in the concluding section.

6.2 Nonparametric Change Point Test for Shift: Location Known

We start with the model that F_i, the d.f. of X_i, is given by

$$F_i(x) = F(x-\theta)\ ,\quad i = 1,\ldots,n \tag{6.1}$$

with F symmetric.

For some specified location θ_0, the null hypothesis of no change point is framed as

$$H_0\colon\ \theta_1 = \ldots = \theta_n = \theta_0 \tag{6.2}$$

Also, we denote by

$$H_{1k}\colon\ \theta_i = \theta_0 \text{ for } i \le k,\quad \theta_i = \theta^\star \text{ for } i > k\ ,\quad \theta^\star \neq \theta_0 \tag{6.3}$$

for $k = 1,\ldots,n$. Then, the alternative hypothesis $\mathcal{H}_1$ is a composite one:

$$\mathcal{H}_1 = \bigcup_{k=1}^{n-1} H_{1k} \tag{6.4}$$

i.e., a shift in the location θ_i occurs at some point in $T_n = (t_1, t_n]$. Note that $\mathcal{H}_1$ may also be expressed as

$$\mathcal{H}_1\colon\ \theta_0 = \ldots = \theta_k \neq \theta_{k+1} = \ldots \theta_n \quad \text{for some } k\colon\ 1 \le k \le n-1 \tag{6.5}$$

Thus, $\tau \in (t_k, t_{k+1}]$ represents the change point and is not known. The null hypothesis is that $\tau \notin T_n = (t_1, t_n]$, while $\mathcal{H}_1$ relates to $\tau \in T_n$.

Page (1955) considered a simple sign test for this problem. Let $S_k = \sum_{i=1}^{k} \text{sign}(X_i - \theta_0)$, for $k = 1,\ldots,n$. Conventionally, we let $S_0 = 0$. Then the Page test is based on the *cumulative sums* S_k, $k \le n$. Since the $\text{sign}(X_i - \theta_0)$ are independent (and identically distributed with mean zero and unit variance under the null hypothesis), the CUSUMs S_k, $k \ge 0$ have the simple random walk property. Hence, the test can be carried out relatively simply. However, this test totally ignores the magnitudes of the $|X_i - \theta_0|$, and thus is generally less efficient than other rank or M-tests.

Bhattacharyya and Johnson (1968) considered a general class of locally optimal rank tests for this model. They had in mind the Bayesian setup of Chernoff and Zacks (1964), although the underlying normality of F is not required in this setup. Suppose that there exist nonnegative numbers $d_1,\ldots,d_n$ with $\sum_{i=1}^{n} d_i = 1$ such that

$$P(\tau = t_k) = d_k\ ,\quad k = 1,\ldots n \tag{6.6}$$

Let

$$D_k = \sum_{i=k}^{n} d_i \ , \quad k = 1, \ldots n \tag{6.7}$$

Also, let R_{ni}^{+} be the rank of $|X_i - \theta_0|$ among the $|X_k - \theta_0|$, $k = 1, \ldots, n$. If the F_i are continuous, then ties among the X_i may be neglected with probability 1 so that the R_{ni}^{+} are the natural integers $1, \ldots, n$, permuted in some order. Also consider a score function $\varphi^{+} = \{\varphi^{+}(u),\ 0 < u < 1\}$, defined by $\varphi^{+}(u) = \varphi[(u+1)/2]$, $0 < u < 1$, where $\varphi(u)$ is skew-symmetric about $1/2$ [i.e., $\varphi(u) + \varphi(1-u) = 0, \forall u \in (0,1)$]. Let then $U_{n1} < \ldots < U_{nn}$ stand for the ordered r.v.'s of a sample of size n from the uniform distribution over $(0,1)$.

Define the scores

$$a_n^{+}(k) = a_n^{+}(k, \varphi) = E\varphi^{+}(U_{nk}) \ , \ k = 1, \ldots, n \tag{6.8}$$

Then, their proposed test statistic is

$$B_n = \sum_{i=1}^{n} D_i \operatorname{sign}(X_i - \theta_0) a_n^{+}(R_{ni}^{+}) \tag{6.9}$$

In particular, if τ is assumed to have a (discrete) uniform distribution on T_n, we have $D_i = n - i + 1$, $1 \leq i \leq n$, so that B_n in (6.9) reduces to $\sum_{i=1}^{n}(n-i+1)\operatorname{sign}(X_i - \theta_0) a_n^{+}(R_{ni}^{+})$. Also, if we choose the scores as $a_n^{+}(k) = k/(n+1)$, $1 \leq k \leq n$ (i.e., the *Wilcoxon* scores), then (6.9) further reduces to

$$B_n^{(w)} = \sum_{i=1}^{n}(n - i + 1)\ \operatorname{sign}(X_i - \theta_0)\, R_{ni}^{+}(n+1)^{-1} \tag{6.10}$$

Similarly, choosing $a_n^{+}(k) = 1$, $\forall k \leq n$, (6.9) reduces to

$$B_n^{(p)} = \sum_{i=1}^{n}(n - i + 1) \operatorname{sign}(X_i - \theta_0) \tag{6.11}$$

which is related to the Page (1955) statistic — i.e., $B_n^{(p)} = S_1 + \ldots + S_n$. If we choose $a_n^{+}(k)$ to be the expected value of the kth order statistic of a sample of size n from the $\chi^2(1)$ distribution, for $k = 1, \ldots, n$, then $B_n = B_n^{(N)}$ relates to the so called *normal* scores statistic.

If we now assume that the d.f. F has an absolutely continuous and symmetric probability density function (p.d.f.) f with a finite Fisher information $I(f) = \int [f'(x)]^2 / f(x) dx$, where $f'(x) = (d/dx) f(x)$ exists a.e., and if we let

$$\varphi(u) = \varphi_f(u) = -f'[F^{-1}(u)] / f[F^{-1}(u)] \ , \quad 0 < u < 1 \tag{6.12}$$

then in the Bayesian setup of Chernoff and Zacks (1964), B_n in (6.9) enjoys the local optimality property, interpreted in the light of best average power (locally). Note that $\varphi_f(u)$ is scale-equivariant, so that if f has an unknown scale parameter, that does not affect B_n. Thus, B_n is scale-invariant and locally optimal if the correct p.d.f. f is used in the definition of φ_f and if the prior probabilities in (6.6) describes the change-point in a proper Bayesian setup. The statistic B_n in (6.9) is, under H_0, genuinely distribution-free [for any given set of the D_i and the scores $a_n^+(k)$], so that even if the true density f is not specified, B_n may be used with suitably chosen $a_n^+(k)$, $1 \le k \le n$. Thus, it has a broad scope for practical adaptation where f need not be of some specified form (viz., normal, logistic, double exponential, etc.). Usually the d_i are all taken to be equal (so that τ is equally likely to be equal to any one of the t_i) and this simplifies (6.9), too.

Hušková and Sen (1986) have considered an adaptive procedure for signed-rank statistics leading to an asymptotically most efficient rank test and estimate for the location model. Toward this end, they introduced an (aligned) signed-rank statistic

$$S_n^+(a;\varphi) = \sum_{i=1}^{n} \operatorname{sign}(X_i - \theta_0 - a)\, a_n^+[R_{ni}^+(a);\varphi] \tag{6.13}$$

where $a \in R_1$, $\varphi \in L_2(0,1)$, $R_{ni}^+(a)$ is the rank of $|X_i - \theta_0 - a|$ among $|X_1 - \theta_0 - a|, \ldots, |X_n - \theta_0 - a|$ for $i = 1, \ldots, n$. If the Fisher information is finite, then φ_f can be estimated by

$$\hat{\varphi}_n^+(u) = \sum_{s=0}^{K_n^+ + r_n} \hat{\gamma}_{n,s}^+ P_{2s+1}(u)\ ,\ u \in (0,1) \tag{6.14}$$

$$\hat{\gamma}_{n,s}^+ = (2an)^{-1/2}[S_n^+(-a\sqrt{n}, P_{2s+1}) - S_n^+(a\sqrt{n}, P_{2s+1})]\ ,\ s = 0,1,2,\ldots \tag{6.15}$$

$$K_n^+ = \min\{k \ge k_0 : \sum_{j=k+1}^{k+r_n} (\hat{\gamma}_{n,j}^+)^2 \le \epsilon_n\} \tag{6.16}$$

where $\{P_k;\, k = 1, \ldots, \infty\}$ is the Legendre polynomial system, $a \neq 0$, k_0 is a predetermined positive integer, $r_n > 0$ is either fixed or increasing to infinity, and ϵ_n is small. In other words, φ_f is estimated by first k_n members of the Fourier expansion whose coefficients are estimated through asymptotic linearity of signed-rank statistics (proved by van Eeden, 1972). The Legendre polynomial system can be replaced by any complete orthogonal system on $(0,1)$ fulfilling some smoothness conditions (e.g., the trigonometric system). The estimator $\hat{\varphi}_n^+$ was originally developed for the classical testing hypothesis in location model. However, it appears that the procedure based on $S_n(\hat{\varphi}_n^+)$ for our problem is also asymptotically optimal among procedures based on $S_n^+(\varphi)$, $\varphi \in L_2(0,1)$. Though asymptotic in character, this linearity of signed-rank statistics holds for moderately large values of n, and thus the described procedure works out well even for moderate values of n.

For this change point problem, M-tests have also been considered in the literature. In this context, it is assumed that the d.f. F is symmetric about 0, and one considers

a score function ψ: $R \to R$, such that (i) $\psi(x)$ is skew-symmetric about 0; (ii) $\psi'(x)$ is bounded on $[-Q, Q]$ (for some $0 < Q < +\infty$) and $\psi'(x) \equiv 0$, $\forall x \notin [-Q, Q]$; and (iii) $\psi(x)$ is nondecreasing. Note that parallel to the rank statistic B_n in (6.9), one may then consider an M-statistic

$$M_n = \sum_{i=1}^{n} D_i \psi(X_i - \theta_0) \tag{6.17}$$

and the test is then based on M_n. The choice of some particular ψ has been discussed in detail by Huber (1981), in the context of local minimaxity and/or other robustness criteria. It appears that the same choice of ψ pertains to the change point model under parallel local optimality criteria.

In passing, we may remark that whereas the rank tests are scale-invariant, the M-tests are not so, in general. Hence, the choice of an optimal ψ, presumably, may depend on the particular scale factor one has in mind. One way to eliminate this problem is to use the so-called "one-step Huber" version for ψ or to use a studentized version. In either case, the simplicity of the distribution of M_n may have to be compromised. In the particular case of $\psi(u) = \operatorname{sign} u$, $u \in R$, M_n will reduce to the Page-type statistic in (6.11) and is scale-equivariant. Though asymptotic properties of M_n remain in fact for a broad class of F, the optimality result may hold only for a local departure from the assumed form of F. In this sense, the rank tests may present a better picture.

6.3 Nonparametric Change Point Tests for Shift: Locations Unknown

We consider here the same model as in (6.1) through (6.5), with the notable exception that θ_0 is unknown. Bhattacharyya and Johnson (1968) treated this problem in the same Bayesian setup, and they considered test statistics of the form

$$L_n = L_n(\varphi) = \sum_{i=1}^{n} D_i a_n(R_{ni}) \tag{6.18}$$

where the D_i are defined as in (6.6)–(6.7), R_{ni} is the rank of X_i among $X_1, \ldots, X_n$, for $i = 1, \ldots, n$, and the scores $a_n(k)$ are defined by

$$a_n(k) = a_n(k, \varphi) = E\varphi(U_{nk}) , \quad 1 \le k \le n \tag{6.19}$$

where the U_{nk}s are defined as in the context of (6.7) and $\varphi = \{\varphi(u), 0 < u < 1\}$ is a suitable score function (not necessarily skew-symmetric about 1/2). Moreover, the d.f. F need not be symmetric about 0. The test based on L_n is genuinely distribution-free (under H_0) and is invariant under any translation: $X_i \to X_i + c$, $c \in R$, $i = 1, \ldots, n$. If we choose $\varphi(u) = \operatorname{sign}(u - 1/2)$, $\varphi(u) = u$, and $\varphi(u) = \Phi^{-1}(u)$, the inverse of the standard

normal distribution, then the corresponding L_n relates to variants of the so-called median test, Wilcoxon test, and normal scores test. If the form of the true p.d.f. f is assumed to be given and $\varphi(u)$ is defined by (6.12), then L_n is a locally optimal invariant test for the change point (shift) model (viz., Bhattacharyya and Johnson, 1968). These rank tests are all scale-equivariant.

Hušková and Sen (1985) have considered asymptotically efficient adaptive scores based on the Legendre polynomial system and the Jurečková-linearity of rank statistics, i.e., φ_f is estimated by

$$\hat{\varphi}_n(u) = \sum_{s=0}^{K_n+r_n} \hat{\gamma}_{n,s} P_s(u) , \quad u \in (0,1) \tag{6.20}$$

$$\hat{\gamma}_{n,s} = \frac{1}{2a} \sum_{i=1}^{n} D_{in}(a_n[R_{ni}(-a), P_s] - a_n[R_{ni}(a), P_s]) \tag{6.21}$$

where $D_{in} = (D_i - \overline{D}_n)[\sum_{j=1}^n (D_j - \overline{D}_n)^2]^{-1/2}$, $\overline{D}_n = \frac{1}{n}\sum_{j=1}^n D_j$, K_n is defined by (6.16) with $\hat{\gamma}^+_{n,j}$ replaced by $\hat{\gamma}_{n,j}$, $R_{ni}(a)$ is the rank of $X_i - aD_{in}$ among $X_1 - aD_{1n}, \ldots, X_n - aD_{nn}$, $a \neq 0$.

Though the estimator $\hat{\varphi}_n$ was considered for the two-sample (and regression) models for the classical estimation testing procedures, it appears that the asymptotic optimality [among the procedures based on $L_n(\varphi)$, $\varphi \in L_2(0,1)$] remains in fact for the change point model, too.

Let us have a pseudo two-sample look at this problem. Under H_{ik} in (6.2), $X_1, \ldots, X_k$ are iid r.v.s with the d.f. $F(x - \theta^*)$; there are $n-1$ such pseudo two-sample cases (for $k = 1, \ldots, n-1$). Based on the pair $(X_1, \ldots, X_k)$ and $(X_{k+1}, \ldots, X_n)$, let us consider a typical two-sample rank statistic $L_{k,n-k}(\varphi)$. For example, $L_{k,n-k}(\varphi)$ may be the median statistic [when $\varphi(u) = \text{sign}(u - 1/2)$], Wilcoxon rank sum statistic [when $\varphi(u) = u$] or the normal scores statistic [when $\varphi(u) = \Phi^{-1}(u)$]. If the d_i in (6.6) are all taken to be equal, then it is easy to verify that $L_n(\varphi)$ in (6.18) is the sum $\sum_{k=1}^{n-1} L_{k,n-k}$. Instead of this sum, a variant form may be

$$\max_{1 \leq k \leq n} |L_{k,n-k} - E_0(L_{k,n-k})| / \sqrt{Var_0(L_{k,n-k})} \tag{6.22}$$

or another form:

$$\max_{1 \leq k \leq n-1} |L_{k,n-k} - E_0(L_{k,n-k})| / \sqrt{Var_0(L_{n/2,n/2})} \tag{6.23}$$

where $n/2$ may be replaced by $(n+1)/2$ if n is odd and E_0 and Var_0 stand for the expectation and variance under H_0.

For the particular cases of median and rank sum statistics, (6.22) has been studied by A. Sen and M.S. Srivastava (1975); and (6.23), by Pettitt (1979). Whereas in (6.22),

the denominator varies from k to k, in (6.23), it is a function of n, independent of k. Schechtman and Wolfe (1981, 1984) have also considered these statistics, and through simulations have obtained numerical ideas about their critical levels. For (6.23), under H_0, a martingale characterization is due to Sen (1978), and this enables one to use the usual weak invariance principle for martingales to provide good approximations for the critical levels in terms of the classical Brownian bridge processes. Actually, following Sen (1978), we may consider a test statistic of the form

$$\max_{1\le k\le n-1}\{|\sum_{i=1}^{n}(c_{ik}-\bar{c}_{nk})a_n(R_{ni})q(k/n)|\} \tag{6.24}$$

where

$$c_{ik}=\begin{cases}1 & 1\le i\le k\\ 0 & i>k\end{cases}\quad,\quad \bar{c}_{nk}=k/n\ ,\quad k=1,\ldots,n-1 \tag{6.25}$$

and $q=\{q(t),0<t<1\}$ is a square-integrable U-shaped function. For example, if we let $q(t)\equiv 1$, we get (6.23); while for (6.22), we have $q(t)=[t(1-t)]^{-1/2}$, which does not satisfy the square integrability condition. To remove this difficulty, usually a small neighborhood of 0 and 1 is excluded, i.e., $q(t)=0$ for $t\le\varepsilon_n$ or $t\ge 1-\varepsilon_n$ for some $\varepsilon_n>0$. But then in (6.24) we have an effective range of k: $\varepsilon_n\le Var_0(L_{k,n-k})/Var_0(L_{n/2,n/2})\le 1-\varepsilon_n$. From a practical point of view, this amounts to basing the test not on all the $n-1$ pseudo-samples but on $n-1-2k_n$ such pseudo-samples, where k_n is small compared to n. For such truncated cases, Sinha and Sen (1979) have done some simulation studies for the critical levels, while for square integrable q, under H_0, (6.24) converges in law to

$$\sup_{0<t<1}\{|Z^0(t)q(t)|\}=D_q^0 \tag{6.26}$$

say, where $Z^0=\{Z^0(t),0\le t\le 1\}$ is a Brownian bridge on $[0,1]$. For D_q^0, too, simulation studies can be made, as in Sinha and Sen (1979), or numerical solutions can be obtained as in DeLong (1981).

Whenever F is assumed to be symmetric about 0, aligned signed rank statistics may also be used, as in Sen (1977). Let $R_{ni}^{+}(t)$ be the rank of $|X_i-t|$ among $|X_1-t|,\ldots,|X_n-t|$, for $i=1,\ldots,n$ and $t\in R_1$. Also, let θ_n be a $\sqrt{n}$-consistent estimator of θ [i.e., $n^{1/2}|\theta_n-\theta|=O_p(1)$]. Consider then the statistic

$$\max_{1\le k\le n}\{|S_k(\tilde{\theta}_n,\varphi)-[\sum_{i=1}^{n}a_n^2(i)]^{-1}[\sum_{i=1}^{k}a_k^2(i)]S_n(\tilde{\theta}_n,\varphi)|\} \tag{6.27}$$

where $S_k(a,\varphi)$ is defined by (6.13) with $n=k$, $\theta_0=0$. Then defining Z^0 as before, it follows from Sen (1977) that, under H_0, (6.27), normalized by $[\sum_{i=1}^{n}a_n^2(i)]^{-1/2}$, converges in law to $\sup\{Z^0(t):0\le t\le 1\}$, and this has therefore a simple distribution. For $\tilde{\theta}_n$

either an L-estimator (e.g., median) or, under the second moment condition, the usual least squares estimator suffices. The rank estimator of θ based on the same score function may also be used, and in that case, the second factor in (6.27) drops out. However, computation of an R-estimator may involves cumbrous iteration steps that in view of (6.27) may not be really needed.

Sen (1983a) used another alignment procedure based on *recursive rank residuals*. If $\{\tilde{\theta}_k, k \geq 1\}$ be any sequence of estimators, such that for every $\varepsilon > 0$, there exists an $m_0 = m_0(\varepsilon)$ for which

$$P\{\max_{m_0 \leq k \leq n} k^{1/2}(\log k)^{-1}|\tilde{\theta}_k - \theta| \geq 1\} < \varepsilon \tag{6.28}$$

then one may consider the test statistic

$$\max_{1 \leq k \leq n} \{|\sum_{i=1}^{k} \text{sign}(X_i - \tilde{\theta}_{i-1}) a_i^+[R_{ii}^+(\tilde{\theta}_{i-1})]|\} \tag{6.29}$$

Under H_0, when normalized by $(\sum_{i=1}^n (a_i^+[i])^2)^{-1/2}$ or by $[n \int_0^1 \varphi^{+2}(u)du]^{-1/2}$, (6.29) converges in law to $\sup\{|W(t)| : 0 \leq t \leq 1\}$, where W is a standard Wiener process on $[0,1]$. Note that (6.28) is satisfied by the usual least squares estimators (under the second moment condition) and R- and M-estimators, even under weaker moment conditions. Nonnull distribution theory (under local alternatives) have also been studied by Sen (1983a).

It appears that the adaptive procedure of Hušková and Sen (1986) may also be used for the signed-rank statistics in (6.27) and (6.29), and they may be justified on the grounds of asymptotic optimality (Pitman efficiency). Namely, in (6.27) we use the estimator $\hat{\varphi}_n^+$ from (6.14) with θ_0 replaced by $\tilde{\theta}_n$. In case (6.29), the same estimator can be applied; however, it should be independent on $X_1, \ldots, X_n$, which demands an increase in the total number of observations for the asymptotic theory to work out in practice.

Finally, in (6.24), (6.27), or elsewhere, one may replace the "max" norm by a "Cramer-von Mises"-type norm, and based on the same invariance principles, consider the functional $\int_0^1 [Z^0(t)]^2 q^2(t)dt$ or $\int_0^1 W^2(t)q^2(t)dt$ for seeking the asymptotic solutions to the desired critical levels.

Let us now consider M-procedures. The score function $\psi : R \to R$ is defined as in (6.17). Then, the following nonrecursive procedure was considered by Sen (1984). Let $\tilde{\theta}_n(\psi)$ be an M-estimator of θ based on the score function ψ, i.e., $\tilde{\theta}_n(\psi)$ is a solution to

$$\sum_{i=1}^{n} \psi(X_i - t) = 0 \tag{6.30}$$

Then the M-test statistc is

$$\max_{1 \leq k \leq n} \{|\sum_{i=1}^{k} \psi[X_i - \tilde{\theta}_n(\psi)](\sum_{i=1}^{n} \psi^2[X_i - \tilde{\theta}_n(\psi)])^{-1/2}|\} \tag{6.31}$$

It is also possible to replace the M-estimator $\tilde{\theta}_n(\psi)$ by an arbitrary $\sqrt{n}$-consistent estimator $\bar{\theta}_n$ and take the test statistic as

$$\max_{1\le k\le n}\{|\sum_{i=1}^{n}\psi^2(X_i-\bar{\theta}_n)^{-1/2}[\sum_{i=1}^{k}\psi(X_i-\bar{\theta}_n)-\frac{k}{n}\sum_{j=1}^{n}\psi(X_j-\bar{\theta}_n)]|\} \tag{6.32}$$

As in the case of rank statistics, dealt with in (6.27), for either (6.31) or (6.32), under H_0, the critical level can be obtained from the well-known distribution theory of $\sup\{|Z^0(t)| : 0 \le t \le 1\}$, where Z^0 is a standard Brownian bridge on $[0,1]$. Side by side, recursive M-tests may also be based on

$$\max_{1\le k\le n}\{|(\sum_{i=1}^{k}\psi^2[X_i-\tilde{\theta}_{i-1}(\psi)])^{-1/2}\sum_{i=1}^{k}\psi[X_i-\tilde{\theta}_{i-1}(\psi)]\sqrt{k/n}|\} \tag{6.33}$$

where $\tilde{\theta}_{i-1}(\psi)$ is an M-estimator of θ based on the score function ψ and on $X_1,\ldots,X_{i-1}$, for $i \ge 2$; $\tilde{\theta}_0(\psi) = 0$ conventionally. As in the case with the recursive rank statistics in (6.29), under H_0, the critical level of (6.33) can well be approximated by the corresponding level of $\sup\{|W(t)| : 0 \le t \le 1\}$ where W is a standard Wiener process on $[0,1]$. Here also, for either (6.31) or (6.33), the "max norm" may be replaced by a "Cramer-von Mises"-type norm, and one may, with advantage, use the functionals $\int_0^1[Z^0(t)]^2q^2(t)dt$ or $\int_0^1 W^2(t)q^2(t)dt$ for approximating the critical levels.

For local alternatives, the asymptotic distribution theory of such M-procedures has also been studied by Sen (1984). To formulate such local alternatives, we assume that

1. The number of observations, n, is large.
2. The change point $\tau \in (t_{q_n}, t_{q_n+1}]$, where q_n increases with n, in such a way that

$$n^{-1}q_n \to \pi : 0 < \pi < 1 \tag{6.34}$$

3. $\theta_1 = \ldots = \theta_{q_n}$, $\theta_{q_n+1} = \ldots = \theta_n$, where

$$n^{1/2}(\theta_{q_1+1} - \theta_{q_n}) \to \lambda \tag{6.35}$$

Thus, the local character of the alternative hypothesis is preserved by (6.35), while (6.34) asserts that the number of observations before the change point occurs is large, as is the number of observations following the change point. From the mathematical point of view, these conditions ensure the contiguity of probability measures under the alternative to those under the null hypothesis, so that the Hájek-LeCam theory can be imported to study the desired nonnull distribution theory in a relatively simpler manner. For the nonrecursive procedure, we have a segmental linear drift function; while for the recursive procedure, we have a null drift up to the point π and a logarithmic drift on $[\pi,1]$. As such,

we are not in a position to compare the recursive and nonrecursive procedures in the light of the usual Pitman efficiency. Nevertheless, within each class, the asymptotic relative efficiency may be studied by simply comparing the drift functions (which are proportional to each other), and this has been exploited by Sen (1984) in a systematic manner. These results have also been shown to be true for rank tests.

Further, tests for change points based on U-statistics have also been considered by Sen (1982). These tests may pertain to models more general then the "shift alternatives", and they contain the simple tests in (6.22)–(6.23) for the Wilcoxon scores as special cases. The theory runs parallel to that of rank or M-procedures, and both nonrecursive and recursive U-statistics have been considered in the same generality. In economic models dealing with multivariate observations, such U-statistics based change point tests may be very fruitful when one is interested in a change point model relating to association or other functionals. Locations on shift functionals are, of course, included as special cases.

6.4 Nonparametric Change Point Tests for Constancy of Regression

As a natural extension of (6.1), we consider here the model:

$$F_i(x) = F(x - \beta_i' \mathbf{c}_i), \quad i = 1, \ldots, n \tag{6.36}$$

where the $\mathbf{c}_i$ are known vectors of regression constants and the β_is are vectors of unknown (regression) parameters. The null hypothesis relates to the constancy of the regression relationship over time, i.e.,

$$H_0\colon \beta_1 = \ldots = \beta_n = \beta \tag{6.37}$$

β unknown, against the composite alternative that

$$\mathcal{H} = \bigcup_{m=1}^{n-1} \mathcal{H}_m; \quad \mathcal{H}_m\colon \beta_1 = \ldots = \beta_m \neq \beta_{m+1} = \ldots = \beta_n, \quad 1 \leq m \leq n-1 \tag{6.38}$$

Thus, here also the change point $\tau \in (t_m, t_{m+1}]$ for some $m\,(= 1, \ldots, n-1)$.

We denote by $\mathbf{C}_m = \sum_{i=1}^{m} \mathbf{c}_i \mathbf{c}_i'$, $m = 1, \ldots, n$, and assume that $m^{-1}\mathbf{C}_m \to \mathbf{C}_0$ (p.d.) as m increases. We may assume without any loss of generality that $\mathbf{C}_0$ is known. First, we consider some rank tests for this problem. Let $R_{ki}(\mathbf{t}, a)$ be the rank of $X_i - \mathbf{t}'\mathbf{c}_i - ad_i$ among $X_\alpha - \mathbf{t}'\mathbf{c}_\alpha - ad_\alpha$, $\alpha = 1, \ldots, k$, for $1 \leq i \leq k$; $k \geq 1$, $\mathbf{t} \in R^q$, $a \in R^1$. Also, let β_n be a $\sqrt{n}$-consistent estimator of β, based on the entire data set and define the scores $a_k(i, \varphi)$, $1 \leq i \leq k$, for $k \geq 1$ as in (6.19). Let then

$$\mathcal{L}_{nk} = \mathcal{L}_{nk}(\varphi) = \tag{6.39}$$
$$[\mathbf{S}_k(\hat{\beta}_n, 0; \varphi) - \frac{k}{n}\mathbf{S}_n(\hat{\beta}_n, 0; \varphi)]' \, (\mathbf{C}_n - n\bar{\mathbf{c}}_n\bar{\mathbf{c}}_n')^{-1} \, [\mathbf{S}_k(\hat{\beta}_n, 0; \varphi) - \frac{k}{n}\mathbf{S}_n(\hat{\beta}_n, 0; \varphi)]$$

$$\mathbf{S}_k(\mathbf{t},0;\varphi) = \sum_{j=1}^{n} \mathbf{c}_j a_k[R_{kj}(\mathbf{t},0);\varphi] \tag{6.40}$$

for $k = 1,\ldots,n$, where $\bar{\mathbf{c}}_n = n^{-1}\sum_{i=1}^{n}\mathbf{c}_i$. Further, let

$$\mathcal{L}_n^* = \mathcal{L}_n^*(\varphi) = \max_{1\le k\le n}\{\mathcal{L}_{nk}(\varphi)\,(\frac{1}{n}\sum_{k=1}^{n}[a_n(k,\varphi)-\bar{a}_n(\varphi)]^2)^{-1}\} \tag{6.41}$$

Note that if we use an R-estimator of β based on the same score function, then $\mathbf{S}_n(\hat{\beta}_n,0;\varphi) \approx 0$ so that (6.39) simplifies further. However, other estimators may be used as well. If Z_k^0, $k = 1,\ldots,q$ are independent copies of a standard Brownian bridge over $[0,1]$ and we define $Z^* = \{Z^*(t) = \sum_{k=1}^{q}[Z_k^0(t)]^2,\ 0 \le t \le 1\}$ then, under H_0,

$$\mathcal{L}_n^* \xrightarrow{D} \sup_{0\le t\le 1} Z^*(t) = Z^{**} \tag{6.42}$$

say. This weak convergence to the Bessel process (tied down at 0 and 1) provides the access to the close approximation for the critical values of $\mathcal{L}_n^*$.

Recursive signed-rank test statistics may also be used as in Sen (1983a). Define the scores as in (6.8) and let

$$\mathbf{S}_k^* = \sum_{i=1}^{k}\text{sign}(X_i - \hat{\beta}_{i-1}'\mathbf{c}_i)\,a_i^+[R_{ii}^+(\hat{\beta}_{i-1})]\mathbf{c}_i \tag{6.43}$$

for $k = 1,\ldots,n$, where $\hat{\beta}_k = \beta(X_1,\ldots,X_k)$ is a $\sqrt{k}$-consistent estimator of β. Then we may use the test statistic

$$\max_{1\le k\le n}\{(\mathbf{S}_k^*)'\mathbf{C}_n^{-1}\mathbf{S}_k^*\,[\int_0^1 \varphi^{+2}(u)du]^{-1}\} = S_n^{\max} \tag{6.44}$$

Under H_0,

$$S_n^{\max} \xrightarrow{D} \sup_{0\le t\le 1}[\sum_{j=1}^{q} W_j^2(t)] \tag{6.45}$$

where W_j are independent copies of a standard Wiener process on $[0,1]$. In the Bessel process, approximation provides access to the simplified formula for the critical levels of $S_n^{\max}$.

Parallel results for M-procedures have been worked out in Sen (1984). In (6.39), replace the sum $\sum_{i=1}^{k}\mathbf{c}_i a_k[R_{ki}(\hat{\beta}_n,0)]$ by $\sum_{i=1}^{k}\mathbf{c}_i\psi(X_i - \tilde{\beta}_n'\mathbf{c}_i)$, $1 \le k \le n$; and in (6.41) $n^{-1}\sum_{i=1}^{n}[a_n(i)-\bar{a}_n]^2$ by $n^{-1}\sum_{i=1}^{n}\psi^2(X_i - \tilde{\beta}_n'\mathbf{c}_i)$, and in (6.44), $\int_0^1\varphi^{+2}(u)du$ by $k^{-1}\sum_{i=1}^{k}\psi^2(X_i - \tilde{\beta}_{i-1}'\mathbf{c}_i)$. The rest of the theory holds.

Note that for the test in (6.41) [or in (6.44)], within the class of rank tests, the optimal score function is the same as discussed in Section 6.2 and 6.3. Thus, the Hušková and Sen (1985, 1986) procedure can be used to provide an adaptive procedure to achieve the asymptotic normality (within the respective class). Namely, in (6.41), φ_f can be estimated by (see Hušková(1988))

$$\hat{\varphi}_n(u) = \sum_{s=1}^{K_n+r_n} \hat{\gamma}_{n,s} P_s(u), \quad u \in (0,1) \tag{6.46}$$

$$\hat{\gamma}_{n,s} = \frac{2}{t\sqrt{n}} \sum_{i=1}^{[n/2]} \{a_n[R_{ni}(\hat{\beta}_n, -\mathbf{t}n^{-1/2}, P_s] - a_n[R_{ni}(\hat{\beta}_n, \mathbf{t}n^{-1/2}, P_s]\} \tag{6.47}$$

where $d_i = 1/2$ for $1 \le i \le [n/2]$, $d_i = -1/2$ for $i > [n/2]$; K_n is defined by (6.16) with $\hat{\gamma}^+_{n,s}$ replaced by $\hat{\gamma}_{n,s}$. As for (6.44), one can repeat the words for the situation (6.29).

6.5 General Remarks

We have observed that in the nonparametric case, R- and M-procedures for the change point problem have been developed on parallel lines. However, there is a qualitative difference in the basic formulation of the two procedures. If the error variables are expected to follow a specified distribution (F_0) subjected to small amount of contaminations (or, in other words, if we have "local" departures from a specified d.f. F_0), then the M-procedures may be used, and with the advantage that they are "nearly" optimal for F_0 and remain robust for local departures from F_0 as well. The Huber scores for normal contamination models are a classical example of this type.

On the other hand, if this error distribution is not sufficiently known, but can be assumed to be a member of a given class of distributions, then the rank (R-) procedures can be formulated in such a way that they remain "globally" robust for this broader class and enjoy high efficiency within the same framework. The rank procedures are scale-equivariant, while the M-procedures are generally not. However, from the computational aspect, the M-estimators are generally simpler then the R-estimators (for the linear models at least), although, in either case, good iterative procedures are available (when an initial $\sqrt{n}$-consistent estimator is used to initiate an iterative process). In this context, ordinary least squares estimators or their trimmed versions may be used conveniently toward this initial estimator. In the context of linear models, one may not need to assume, for R-procedures, that the errors have a symmetric distribution, while for M-procedures this is a prerequisite. For the location model (θ_0 specified), however, both procedures depend on the assumed symmetry of F. Thus, in any live application, this assumption needs to be critically examined and, in the light of that assessment, the choice between an R- and M-procedure should be made.

Finally, inlike the classical location or simple regression model, there may not be, in general, any uniformly (or even locally) most powerful test for the change point model.

The optimality of such tests may usually be studied in the light of "Bayesian optimality" as by Chernoff and Zacks (1964); and, in this setup, linear (or signed-linear) rank statistics and parallel M-statistics may be formulated to achieve this local optimality. Within this framework, the choice of asymptotically efficient score functions, (viz., Hušková and Sen 1985, 1986), work well. On the other hand, the "pseudo two-sample" approach may also be justified on other grounds, and R- and M-procedures akin to this approach enjoy the same interpretations and share the common property. Within this class also, local optimality can be studied in terms of maximization of local (asymptotic) power. Again, in that context, the choice of an asymptotically efficient score function (Hušková and Sen, 1985, 1986) leads to a locally optimal rank procedure. The adoption of the pseudo Bayesian or the pseudo two-sample approach in the change point model in a particular case has to be judged on the basis of other considerations; once this decision is made, the choice of an adaptive nonparametric procedure can be made on the general grounds explored in the earlier sections of this chapter.

Acknowledgments

The second author's work was supported by the Office of Naval Research, Contract No. N00014-83-K-0387.

References

Bhattacharyya, G.K. and Johnson, R.A. (1968), Nonparametric tests for shift at an unknown time point. *Annals of Mathematical Statistics*, **39**, 1731-1743.

Brown, R.L., Durbin, J., and Evans, J.M. (1975), Techniques for testing the constancy of regression relationships over time (with discussion). *Journal of the Royal Statistical Society*, **B-37**, 149-192.

Chernoff, H. and Zacks, S. (1964), Estimating the current mean of a normal distribution which is subjected to changes in time. *Annals of Mathematical Statistics*, **35**, 999-1018.

Eeden, C. van (1972), An analogue for signed rank statistics of Jurečková's asymptotic linearity theorems for rank statistics. *Annals of Mathematical Statistics*, **43**, 751-802.

Hackl, P. (1980), *Testing the Constancy of Regression Relationships over Time.* Göttingen: Vandenhoeck & Ruprecht.

Huber, P.J. (1981), *Robust Statistics.* New York: John Wiley.

Hušková, M. (1988), Adaptive procedures for detection of change. *Statistics and Decisions*, **6**, 137-148.

Hušková, M. and Sen, P.K. (1985), On sequentially adaptive asymptotically efficient rank statistics. *Sequential Analysis*, **4**, 125-151.

Hušková, M. and Sen, P.K. (1986), On sequentially adaptive signed-rank statistics. *Sequential Analysis*, **5**, 237-252.

Jurečková, J. (1969), Asymptotic linearity of a rank statistic in regression parameter. *Annals of Mathematical Statistics*, **40**, 1889-1900.

DeLong, D. (1981), Crossing probabilities for a square root boundary by a Bessel process. *Communications in Statistics, Theory and Methods*, **A10**, 2197-2213.

Page, E.S. (1955), A test for a change in a parameter occurring at an unknown point. *Biometrika*, **42**, 523-527.

Page, E.S. (1957), On problems in which a change of parameters occurs at an unknown time point. *Biometrika*, **44**, 248-252.

Pettitt, A.N. (1979), A nonparametric approach to the change-point problem. *Applied Statistics*, **28**, 126-135.

Schechtman, E. and Wolfe, D.A. (1981), Distribution-free tests for the change-point problem. Technical Report, Ohio State University, Columbus, OH.

Sen, A. and Srivastava, M.S. (1975), On tests for detecting changes in mean. *Annals of Statistics*, **3**, 98-108.

Sen, P.K. (1977), Tied down Wiener process approximations for aligned rank order processes and some applications. *Annals of Statistics*, **5**, 1107-1123.

Sen, P.K. (1978), Invariance principles for linear rank statistics revisited. *Sankhya*, **A-40**, 215-236.

Sen, P.K. (1980), Asymptotic theory of some tests for a possible change in the regression slope occurring at an unknown time point. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **52**, 203-218.

Sen, P.K. (1982), Asymptotic theory of some tests for constancy of regression relationships over time. *Statistics*, **13**, 21-31.

Sen, P.K. (1983a), On some recursive residual rank tests for change-points, pp. 371-391 in M.H. Rizvi (ed.), *Recent Advances in Statistics: Papers in Honor of Herman Chernoff's Sixtieth Birthday*. New York: Academic Press.

Sen, P.K. (1983b), Tests for change-points based on recursive U-statistics. *Sequential Analysis*, **1**, 263-284.

Sen, P.K. (1984), Recursive M-tests for the constancy of multivariate regression relationships over time. *Sequential Analysis*, **3**, 191-211.

Sen, P.K. (1985), *Theory and Applications of Sequential Nonparametrics*. Philadelphia, PA: SIAM.

Shirjajev, A.N. (1963), On optimum methods in quickest detection problems. *Theory of Probability and Their Applications*, **10**, 348-354.

Shirjajev, A.N. (1978), *Optimal Stopping Rules*. New York: Springer-Verlag.

Sinha, A.N. and Sen, P.K. (1979), Progressively censored tests for multiple regression based on weighted empirical distributions. *Calcutta Statistical Association Bulletin*, **28**, 57-82.

Wolfe, D.A. and Schechtman, E. (1984), Nonparametric statistical procedures for the changepoint problem. *Journal of Statistical Planning and Inference*, **9**, 389-396.

CHAPTER 7

Detection of Join Point in Regression Models

Hiroki Tsurumi

Summary

A Bayesian predictive density for the mean squared errors of post-sample forecasts is derived within the linear regression framework. The kernel of the predictive density is an F distribution. In the process of deriving the predictive density, we use a degenerate hyperbolic function to express the distribution of quadratic forms in normal variables. The Bayesian predictive density is then used to detect a join point by the highest posterior density interval criterion. Numerical examples are given to compare the Bayesian predictive density procedure with the maximum likelihood and Bayesian posterior density procedures for detecting the join point. When the join point is at either the beginning or ending edge of the sample period, the Bayesian predictive density procedure detects the join point whereas the maximum likelihood and Bayesian posterior density procedures cannot.

7.1 Introduction

With linear models that abruptly switch regimes, essentially two problems arise: detecting the join point (i.e., the point at which regression shifts), and making inference about regression parameters.

In this chapter we deal with the first problem, and propose a procedure to detect the join point.

The proposed procedure is a prediction criterion based on the mean squared errors of post-sample forecasts (MSEF). If the actual MSEF falls outside of the highest posterior

density interval (HPDI) consistently for a number of forecasting periods, we judge that the regression model switched regimes. This predictive criterion differs from the criteria so far proposed in the literature. Quandt (1958) estimated the join point by maximum likelihood. Ferreira (1975), and Choy and Broemeling (1980) used posterior probability density functions to estimate and make inference on the join point. The prediction criterion we propose in this chapter has an advantage over these existing procedures when the join point occurs at either end of the sample period.

The organization of the chapter is as follows. In Section 7.2, we derive the predictive density for the MSEF and suggest an HPDI procedure to detect the join point. A numerical example is given in Section 7.3 to compare the HPDI criterion with the maximum likelihood and Bayesian posterior density procedures for detecting the join point. Concluding remarks are given in Section 7.4.

7.2 Bayesian Predictive Density of the MSEF

Let the linear model be given by

$$y = X\beta + u \tag{7.1}$$

where y is an $(n \times 1)$ vector of observations on the dependent variable, X is an $(n \times k)$ matrix of observations on the explanatory variables with rank k, u is an $(n \times 1)$ vector of error terms, and β is a $(k \times 1)$ vector of unknown regression coefficients. Assume that $u \sim N(0, \sigma^2 I_n)$ and that β is estimated by $\hat{\beta} = (X'X)^{-1}X'y$.

The mean squared error for the post-sample period, $n+1, \ldots, n+m$, is computed using the post-sample actual observations on y and X. Let $y_\star$ and $X_\star$ be, respectively, an $(m \times 1)$ vector and an $(m \times k)$ matrix of post-sample observations and assume that the rank of $X_\star$ is $\min(n, k)$. Then the MSEF is

$$\text{MSEF} = \frac{1}{m}(\hat{y}_\star - y_\star)'(\hat{y}_\star - y_\star) \tag{7.2}$$

where $\hat{y}_\star = X_\star\hat{\beta}$.

We shall derive the Bayesian predictive density of the MSEF in (7.2). This can be attained in either of the following two ways. We first reduce the MSEF as a quadratic form in normal variables by substituting (7.1) and $\hat{\beta} = \beta + (X'X)^{-1}X'u$ and, realizing that this quantity contains σ^2, we integrate it out by using the posterior pdf (probability density function) of σ^2. This approach is given in Tsurumi and Wago (1988). Alternatively, we may start with the joint density for $y_\star$, β, and σ^2, and integrate β first, and transform $y_\star$ into the MSEF given σ^2. Finally, we integrate σ^2 out. This second approach is used in the theorem below.

Theorem 7.1 *The Bayesian predictive density of the MSEF, z, is given by*

$$p(z|s^2,\nu,m) \propto \frac{z^{(m/2)-1}}{(\nu s^2+mz/\mu_m)^{(m+\nu)/2}} \times \sum_{p=0}^{\infty}\Gamma(\frac{m+\nu}{2}+p)\,2^p c(m,p)[\frac{mz}{\nu s^2+mz/\mu_m}]^p \tag{7.3}$$

where $c(m,p)$ *is the recursive coefficient given by*

$$c(m,p)=\frac{\Gamma(p+\frac{m-1}{2})}{\Gamma(p+\frac{m}{2})}\sum_{j=0}^{p}\frac{c(m-1,j)a_m^{p-j}}{(p-j)!} \qquad \text{for } m\geq 2 \tag{7.4}$$

and $c(1,0)=1$, $c(1,j)=0$ *for* $j\geq 1$; $a_m=(\mu_m^{-1}-\mu_{m-1}^{-1})/2$, $\mu_1\geq\ldots\geq\mu_m$, *for* $m\geq 2$, $a_1=(2\mu_1)^{-1}$, *and* $a_i^0=1$ *for all* $i=1,\ldots,m$. *The* μ_i*s are the nonzero characteristic roots of* $B'B$, *where* $B=(A,-I_m)$, $A=X_\star(X'X)^{-1}X_\star'$, *and* s^2 *is given by* $\nu s^2=y'(I-X(X'X)^{-1}X')y$, $\nu=n-k$.

Proof: The predictive pdf of $y_\star$ is given by

$$p(y_\star|y,X,X_\star)=\int\int p(y_\star|\beta,\sigma^2,X_\star)\,p(\beta,\sigma|y,X)\,d\sigma d\beta$$

where

$$p(y_\star|\beta,\sigma^2,X_\star)\propto\sigma^{-m}\exp\{-\frac{1}{2\sigma^2}(y_\star-X_\star\beta)'(y_\star-X_\star\beta)\} \tag{7.5}$$

and we shall use the posterior pdf of β and σ that is given by

$$p(\beta,\sigma|y,X)\propto\sigma^{-(n+1)}\exp\{-\frac{1}{2\sigma^2}[\nu s^2+(\beta-\hat\beta)'X'X(\beta-\hat\beta)]\}$$

Thus $p(y_\star|\beta,\sigma^2,X_\star)p(\beta,\sigma|y,X)$ becomes

$$\begin{aligned}p(y_\star|\beta,\sigma^2,X_\star)&p(\beta,\sigma|y,X)\\ &\propto\ \sigma^{-(n+m+1)}\exp\{-\frac{1}{2\sigma^2}[y_\star'(I-X_\star X_\star^+)y_\star+(\beta-\hat\beta_\star)'X_\star'X_\star(\beta-\hat\beta_\star)]\}\\ &\quad\times\exp\{-\frac{1}{2\sigma^2}[\nu s^2+(\beta-\hat\beta)'X'X(\beta-\hat\beta)]\}\end{aligned} \tag{7.6}$$

where $X_\star^+$ is the Moore-Penrose generalized inverse of $X_\star$, and $\hat\beta_\star=X_\star^+y_\star+(I-X_\star X_\star^+)r$, and r is an arbitrary $(m\times 1)$ vector. Arranging the sum of two quadratic forms in β into a quadratic form in β, and integrating β out, we derive from equation (7.5)

$$p(y_\star|y, X, X_\star) \propto \int \sigma^{-(\nu+m+1)} \exp\{-\frac{\nu s^2}{2\sigma^2}\} \exp\{-\frac{1}{2\sigma^2}(y_\star - X_\star\hat{\beta})' H (y_\star - X_\star\hat{\beta})\} d\sigma \quad (7.7)$$

where $H = [I + X_\star(X'X)^{-1}X_\star']^{-1} = (I + A)^{-1}$. The right-hand side of equation (7.7) shows that, given σ, $y_\star - X_\star\beta$ is distributed as $N(0, \sigma^2 H^{-1})$. Let $w = y_\star - X_\star\beta$, and $H = R'R$, where R is a nonsingular matrix. Then $m.z = w'w - \eta'(RR')^{-1}\eta$, with $\eta = Rw \sim N(0, \sigma^2 I_m)$. Since the nonzero characteristic roots of $H^{-1} = (R'R)^{-1}$ are the same as those of $(RR')^{-1}$, we see that $m.z = \sum_{i=1}^{m} \mu_i \varepsilon_i$, with $\varepsilon_i \sim N(0, \sigma^2)$, and μ_i is the i-th characteristic root of H^{-1}. Hence, given σ, $m.z$ has the distribution of a quadratic form in normal variables. Using the degenerate hyperbolic function, Tsurumi and Wago (1988) derive the density function of a quadratic form in normal variables, q, as

$$f(q|m, \sigma^2) = c_1 q^{\frac{1}{2}m-1} \sigma^{-m} \exp\{-\frac{m}{2\mu_m} q\sigma^{-2}\} \sum_{p=0}^{\infty} c(m, p)\, m^p\, q^p \sigma^{-2p} \quad (7.8)$$

where

$$c_1 = \frac{m^{m/2}}{2^{\frac{1}{2}m} \pi^{\frac{1}{2}} \prod_{i=1}^{k} \mu_i^{\frac{1}{2}}}$$

Using equation (7.8) the predictive density for $z = \frac{1}{m}\sum_{i=1}^{m} \mu_i \varepsilon_i$ becomes

$$p(z|s^2, m) \propto z^{\frac{1}{2}m-1} \int_0^{\infty} \sigma^{-m} \exp\{-\frac{m}{2\mu_m} z\sigma^{-2}\} \sum_{p=0}^{\infty} c(m, p)\, m^p\, z^p \sigma^{-2p} d\sigma \quad (7.9)$$

and interchanging the integration and summation signs and integrating σ out from each term of the summation, we obtain the desired result.

Remark 1: The distribution of quadratic forms or ratios of quadratic forms has been investigated by many; some of the earlier works are by McCarthy (1939), von Neumann (1941), and Bhattacharyya (1943). Bhattacharyya (1945) and Hotelling (1948) employed Laguerre expansions, and Gurland (1953) and Johnson and Kotz (1970) further refined the convergent Laguerre expansions. The degenerate hyperbolic function, which we used above, is convenient for computational purposes.

Remark 2: *Theorem 7.1* is the Bayesian predictive density, of z, given s^2. The sampling distribution of $u = z/(\mu_m s^2)$ can be derived, and this becomes the same as equation (7.3), except that the expression $m.z/(\nu s^2 + mz/\mu_m)$ in (7.3) is now replaced by $u/(1 + mu/\nu)$. In the sampling distribution both z and s^2 are random, whereas in the Bayesian predictive density z is random and s^2 is a fixed number (a realized random variable.)

Remark 3: The characteristic roots μ_i of $B'B$ can be given by $\mu_i = 1 + \lambda_i$, $i = 1, \ldots, m$, for $m \leq k$ and $\mu_i = 1 + \lambda_i$, $i = 1, \ldots, k$; $\mu_i = 1$, $i = k+1, \ldots, m$ for $m > k$, where λ_i is the ith nonzero characteristic root of $X_*(X'X)^{-1}X_*'$.

The MSEF may be used for identifying the join point, t^*, of the switching regression model

$$\begin{aligned} y_t &= \beta_1 + \beta_2 x_{t2} + \ldots + \beta_k x_{tk} + \varepsilon_t \ , \quad t = 1, \ldots, t^* \\ y_t &= \beta_1^* + \beta_2^* x_{t2} + \ldots + \beta_k^* x_{tk} + \varepsilon_t \ , \quad t = t^* + 1, \ldots, n \end{aligned}$$

or

$$\begin{aligned} y_1 &= X_1\beta_1 + \varepsilon_1 \\ y_2 &= X_2\beta_2 + \varepsilon_2 \end{aligned} \tag{7.10}$$

where $\varepsilon_t \sim N(0, \sigma^2)$; $y_1 = (y_1, \ldots, y_{t^*})'$, $y_2 = (y_{t^*+1}, \ldots, y_n)'$; X_1 is a $(t^* \times k)$ matrix of observations on the k explanatory variables in regime 1; X_2 is $(n - t^*) \times k$ matrix of observations on the k explanatory variables in regime 2; β_i is a $(k \times 1)$ vector of regression coefficients in regime i $(i = 1, 2)$, and ε_i is the error vector in regime i.

First, we estimate β_i and σ in regime 1 by using the first n_1 observations. The sample size, n_1, may be chosen with some external information that the regression equation was stable in the first n_1 periods; n_1 must be less than or equal to t^*. Then we compute the MSEF each for $m = n_1 + 1, \ldots, n$: $\text{MSEF}_1 = (\hat{y}_{n_1+1} - y_{n_1+1})^2$, $\text{MSEF}_2 = [(\hat{y}_{n_1+1} - y_{n_1+1})^2 + (\hat{y}_{n_1+2} - y_{n_1+2})^2]/2$, ..., $\text{MSEF}_{n-n_1} = \frac{1}{n-n_1}\sum_{j=n_1+1}^{n}(\hat{y}_j - y_j)^2$. If the actual MSEF_i is within the $(1-\alpha)$ % highest posterior density interval (HPDI), we judge that, up to period i, the regression did not switch. If MSEF_i's are consistently out of $(1-\alpha)$ % HPDI's for $i = r, \ldots, n$, then we say that the join point is at r.

7.3 Numerical Example

Using a numerical example, let us illustrate the HPDI procedure in the previous section and compare it with Quandt's procedure and Bayesian predictive density procedure. Quandt (1958) proposed to detect the join point, t^*, by evaluating the likelihood function

$$\log L(t^*|\text{data}) = -\frac{n}{2}[\log(2\pi) + 1] - \frac{1}{2}t^* \log \hat{\sigma}_1^2 - \frac{1}{2}(n - t^*)\log \hat{\sigma}_2^2 \tag{7.11}$$

where $t^*\hat{\sigma}_1^2 = y_1'[I - X_1(X_1'X_1)^{-1}X_1']y_1$, $(n - t^*)\hat{\sigma}_2^2 = y_2'[I - X_2(X_2'X_2)^{-1}X_2']y_2$. For each t^* $(k < t^* < n - k)$, he suggested choosing the value of t^* that maximized $\log L(t^*|\text{data})$. Equation (7.11) is derived assuming that σ_1 and σ_2 in (7.10) have unequal variances: $\varepsilon_1 \sim N(0, \sigma_1^2 I)$ and $\varepsilon_2 \sim N(0, \sigma_2^2 I)$. If we assume that $\sigma_1^2 = \sigma_2^2$, the log likelihood function becomes

$$\log L(t^*|\text{data}) = -\frac{n}{2}[\log(2\pi) + 1] - \frac{n}{2}\log s^2 \tag{7.12}$$

where $\nu s^2 = y'[I - W(W'W)^{-1}W']y$, $W = (X, V)$, $X = (X_1', X_2')'$, $V = (0, X_2')'$, and $y = (y_1', y_2')'$, and $\nu = n - 2k$. Assuming $\sigma_1^2 = \sigma_2^2$, Ferreira (1975) derives the posterior pdf for t^* as

$$p(t^*|\text{data}) \propto p(t_1^*)|W'W|^{-\frac{1}{2}}(\nu s^2)^{-(n-2k)/2} \tag{7.13}$$

Ferreira suggested three expressions for $p(t^*)$: $p_1(t^*) \propto$ constant for $k \leq t^* \leq n - k$; $p_2(t^*) \propto [t^*(n - t^*)]^{1/2}$ for $k \leq t^* \leq n - k$; and $p_3(t^*) \propto \mid W'W \mid^{1/2}$ for $k \leq t^* \leq n - k$. The use of $p_3(t^*)$ leads to the posterior mode that corresponds to the maximum of $\log L(t^*|\text{data})$ in (7.12). Choy and Broemeling (1980) employed a gamma-normal prior pdf for β_1, β_2, and σ^2 so that they could extend the interval for t^* to $1 \leq t^* \leq n - 1$.

For a numerical example, we shall use the investment equation of Klein's Model I (see Theil, 1970):

$$I_t = \beta_1 P_t + \beta_2 P_{t-1} + \beta_3 K_{t-1} + \beta_4 + \epsilon_t \ , \quad t = 1, \ldots, 21 \tag{7.14}$$

where I_t is net investment in year t; P_t is profits in year t, and K_t is the stock of capital goods at the end of year t. Klein's model is based on pre-war data from 1921 to 1941. Since post-war data for the model for 1947 to 1967 are available (Schink, 1971), let us use the post-war data for the explanatory variables, P_t, P_{t-1}, K_{t-1}, and treat equation (7.13) as the classical linear model. We use the parameter estimated using the pre-war data by the mixed three-stage least squares method. The parameter estimates are given in Theil (1970, p. 517):

$$I_t = 0.1451P_t + 0.6134P_{t-1} - 0.1593K_{t-1} + 20.72 + \epsilon_t$$

At the join point, t^*, we change the coefficient of profits from $\beta_1 = 0.1451$ in regime 1 to $\beta_1^* = 0.2902$ in regime 2. We generated 21 observations for I_t by drawing ϵ_t from $N(0, 1.3832)$. Although only one of the four parameters in (7.14) changes, we shall assume that all of them change so that as far as the detection of the join point is concerned, we use equation (7.10). We do this since, in practice, we hardly have information on which parameters vary, and thus we may have to assume that all of them changed.

Table 7.1 presents the actual MSEF's and 95% HPDI's. The join point, t^*, is set at $t^* = 11$, and equation (7.14) is estimated using the first 11 observations ($n_1 = 11$). In the column under the MSEF, two actual MSEF's are presented for each m, the forecasting period. The "Shift" column gives the actual MSEF's with the dependent variable, I_t, being computed using the shifted coefficient $\beta_1^* = 0.2902$ for the second regime ($t = 12, \ldots, 21$), whereas the "No Shift" column presents the actual MSEF's with I_t being computed using the coefficient of $\beta_1 = 0.1451$ for the second regime. The 95% HPDI's for the MSEF's

Table 7.1: The actual MSEF's and 95% HPDI with $n_1 = 11$ and $t^* = 11$.

		Actual MSEF		
m	*n*	*Shift*	*No shift*	*95% HPDI*
1	12	16.01	1.35	(0, 12.71)
2	13	28.84	1.25	(0, 10.63)
3	14	33.58	1.31	(0, 10.14)
4	15	30.22	1.26	(0, 9.52)
5	16	33.07	1.14	(0, 8.14)
6	17	34.54	0.98	(0, 12.21)
7	18	37.70	1.02	(0, 11.39)
8	19	40.02	0.93	(0, 9.77)
9	20	41.62	0.83	(0, 8.95)
10	21	46.36	1.26	(0, 8.14)

Table 7.2: Posterior probability density function for t^* ($t^* = 11$).

t^*	*pdf*
6	0.00001
7	0.00028
8	0.00016
9	0.00012
10	0.00019
11	0.74930
12	0.00011
13	0.00003
14	0.00002

Posterior mean	11.0
Posterior standard deviation	0.090
Maximum of equation (7.12)	11

Table 7.3: The actual MSEF's and 95% HPDI's with $n_1 = 15$ and $t^* = 17$.

		MSEF		
m	*n*	*Shift*	*No shift*	*95% HPDI*
1	16	—	0.17	(0, 17.57)
2	17	—	0.10	(0, 13.92)
3	18	15.17	0.11	(0, 11.61)
4	19	21.42	0.17	(0, 9.81)
5	20	23.92	0.57	(0, 8.17)
6	21	30.17	0.55	(0, 7.54)

contain the actual MSEF's without shift in β_1, while the actual MSEF's with shift are all outside the 95% HPDI's for every period in regime 2.

Table 7.2 gives the posterior pdf for the join point t^* when t^* is set at 11. As a prior for t^*, we used $p(t^*) \propto$ constant. The maximum for equation (7.12) is also at $t^* = 11$, indicating that both the Bayesian procedure on t^* and Quandt's procedure detect the join point.

Table 7.3 presents the actual MSEF's and 95% HPDI's when the join point is $t^* = 17$ and the first 15 observations ($n_1 = 15$) are used to estimate equation (7.14) for the forecasting exercises. Again we see that the MSEF's with shift are all outside the 95% HPDI's, while the MSEF's without shift are well inside the 95% HPDI's. With the join point of $t^* = 17$, neither Quandt's procedure [equation (7.12)] nor the Bayesian posterior pdf for t^* [equation (7.13)] can be used since they require $t^* \leq n-5$. The MSEF procedure, on the other hand, can be used as long as $t^* \leq n-1$.

7.4 Concluding Remarks

In this chapter we derived a Bayesian predictive density for the MSEF and suggested that it can be used for detecting the join point of the linear model with abruptly switching regimes. Using numerical examples, we compared the MSEF criterion with the Bayesian posterior pdf and Quandt's procedure for detecting the join point. Our numerical examples show that when the join point is well in the middle point of the sample size ($k < t^* < n-k$), then Quandt's and Bayesian posterior pdf procedures detect the join point as well as the MSEF criterion. When the join point is at either edge of the sample ($t^* \leq k$ or $t \leq n-k$), Quandt's and Bayesian posterior pdf procedures cannot be used, while the MSEF criterion can. For the case of $t^* \leq k$, we can use a backward forecasting procedure to detect the join point by the MSEF criterion. We estimate the linear equation using observations from, say, r to n, and compute the actual MSEF's and HPDI's for $m = r-1, \ldots, 1$.

References

Bhattacharyya, A. (1945), A note on the distribution of the sum of chi-squares. *Sankhya*, **7**, 27-28.

Bhattacharyya, B.C. (1943), On an aspect of the Pearson system of curves, and a few analogies. *Sankhya*, **6**, 415-448.

Choy, C.J.H. and Broemeling, L.D. (1980), Some Bayesian inferences for a changing linear model. *Technometrics*, **22**, 71-78.

Ferreira, P.E. (1975), A Bayesian analysis of a switching regression model: Known number of regimes. *Journal of the American Statistical Association*, **70**, 370-374.

Gurland, J. (1953), Distribution of quadratic forms and ratios of quadratic forms. *Annals of Mathematical Statistics*, **24**, 416-427.

Hotelling, H. (1948), Some new methods for distributions of quadratic forms (abstract). *Annals of Mathematical Statistics*, **19**, 119.

Johnson, N.L. and Kotz, S. (1970), *Continuous Univariate Distributions*, Vol. 2. New York: John Wiley.

McCarthy, M.D. (1939), On the application of the z-test to randomized blocks. *Annals of Mathematical Statistics*, **10**, 337-359.

Neumann, J. von (1941), Distribution of the ratio of the mean square successive difference to variance. *Annals of Mathematical Statistics*, **12**, 367-395.

Quandt, R.E. (1958), The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association*, **53**, 873-880.

Schink, G.R. (1971), Small sample estimates of the variance-covariance matrix of forecast errors for large econometric models: The stochastic simulation technique. Unpublished Ph.D. dissertation, University of Pennsylvania.

Theil, H. (1970), *Principles of Econometrics*. New York: John Wiley.

Tsurumi, H. and Wago, H. (1988), A predictive density criterion for selecting non-nested linear models and comparison with other criteria (forthcoming in the *Journal of Econometrics*).

CHAPTER 8

On the Identification of Time for Structural Changes by MOSUM-SQ and CUSUM-SQ Procedures

Anders H. Westlund and Birgitta Törnkvist

Summary

One way to represent a structural change when modeling an economic system is to allow for parameter changes. Besides verifying the existence of such parameter changes, the ultimate purpose of the structural analysis will be to further characterize them. This chapter includes a partial analysis of such a characterization process. Approximate expected values and variances of CUSUM-SQ and MOSUM-SQ statistics are given for various cases of parameter changes, instantaneous as well as gradual. A numerical simulation study of the statistics is also given. The theoretical study and the simulations together demonstrate that identification of time for structural changes is more intricate in the cases of gradual changes, and when the changes occur early during the observed time period.

8.1 Introduction

Structural variability is in general a very important aspect of the economic planning process, but also a complicating one. The assumption of stability or invariance in structural analysis and forecasting of an economic system often entails improper oversimplifications. Allowing for structural variability, however, requires a quantitative methodology adapted to the presuppositions, in the structural analysis as well as for forecasting.

An analysis of structural variability also requires that this concept be considered to be related to the specific economic system "at work". Irrespective of the definition of

structural variability, it must be a measurable concept, in principle with a unique interpretation. Modeling a specific economic system characterized by structural variability involves allowance for assumptions of parameter variability. In line with definitions given by Poirier (1976), the structural change concept will, thus, be associated here with "considerable and low frequent" structural variability (see also Westlund and Zackrisson, 1986).

The first step in a structural analysis of parameter variability implies verification of its existence. This is followed by characterization, often with the ultimate purpose of modeling the parameter variability. Important aspects of such a characterization process involve time dating of the potential variability and identifying of the type of parameter variability. Often, a rough classification of the parameter changes as gradual or instantaneous is urgent.

A number of test techniques and test strategies have been suggested (see bibliography by Hackl and Westlund, 1985). The choice of test strategy is in general based on certain criteria, such as power maximization. Various studies indicate the relative power of competitive test statistics. Characterization, such as time dating, may also be based on these test statistics, e.g., by studying the observed path of the test statistic. Such a study might also give certain indications regarding type of parameter variability. In order to do that, knowledge about the test statistics and how their distributions will react to various alternative parameter variability models is necessary. Such studies are rare so far (see, however, Westlund and Törnkvist, 1985).

In this chapter, the distributions of some test statistics are further studied for different parameter variability models. In particular, our interest focuses on CUSUM-SQ and MOSUM-SQ tests. So far, knowledge of their distributions subject to certain parameter variability hypotheses is very restricted (and is not easy to acquire because of mathematical intractability). Here, the distributions are only partially studied, by focusing on expectation and variance. The exact formulae of these moments are implicit and very complicated, while only approximations are verified and studied. Other characteristics of their distributions, such as percentiles and skewness, might also be of interest here, but are not easily theoretically verified.

The approximate theoretical analysis is therefore supplemented here by a numerical study. Thus, the degree of approximation above will be indicated to some extent. Section 8.2 introduces the models and test statistics applied in the present study. A theoretical study, based on approximate expectation and variance of CUSUM-SQ and MOSUM-SQ tests for various parameter variability hypotheses, is given in Section 8.3. A related numerical study is then performed, estimating not only expectations and standard deviations of the test statistics, but also their skewness and certain percentiles. Design and some basic results are summarized in Section 8.4. Concluding Section 8.5 summarizes the analysis and formulates some operational conclusions to support empirical studies of economic structural variabiltiy. Some suggestions for further research are also given.

8.2 Models and Test Statistics

The analysis in this chapter is based on the unirelational model

$$\left.\begin{array}{rcl} y_t & = & x_t'\beta_t + u_t \\ u_t & \sim & N(0,\sigma^2) \end{array}\right\} \quad t = 1,\ldots,T \tag{8.1}$$

where at time t, y_t is the observation on the regressand, and x_t is the column vector of observations on k regressors, here assumed to be nonstochastic. The column vector of regression coefficients, $\beta_t = (\beta_{1t},\ldots,\beta_{kt})'$, is written with a subscript t to indicate that it may vary with time. The error terms, u_t, are assumed independent and normally distributed with means zero and variances σ^2.

In particular, the following special cases of (8.1) are considered:

M1 The intercept model

$$\left.\begin{array}{rcl} y_t & = & \beta_{1t} + u_t \\ u_t & \sim & iidN(0,\sigma^2) \end{array}\right\} \quad t = 1,\ldots,T$$

M2 The simple regression model with an intercept and one regressor, where one or both of the regression coefficients may vary with time.

The hypothesis of stability over time is

$$H_0 : \beta_t = \beta \ , \quad t = 1,\ldots,T$$

and the alternative hypothesis is

$$H_1 : \beta_{it} = \begin{cases} \beta_i & t = 1,\ldots,t_1 \\ \beta_i + \delta_i\sigma\frac{t-t_1}{t_2-t_1} & t = t_1+1,\ldots,t_2 \\ \beta_i + \delta_i\sigma & t = t_2+1,\ldots,T \end{cases}$$

for at least one $i = 1,\ldots,k$, where β_i and δ_i are constants, and $t_1 + 1$ and t_2 denote the time of the start and the end, respectively, of the parameter change. The quantity $\delta_i\sigma$ is the total change of β_{it} during the observed time period and is related to the standard deviation of the regressand y_t in (8.1).

In analyzing the structural variability of an economic system, it is often necessary to separate instantaneous and gradual parameter changes. As an example, when political and economic steps are taken, one interesting question is whether the response to these steps is instantaneous or gradual.

In H_1, β_{it} changes in an instantaneous way if $t_2 = t_1 + 1$. This special case of H_1 is here denoted PM1 (parameter model 1). For t_2 larger than $t_1 + 1$ the parameter model H_1 is denoted PM2, and represents a gradual linear change of β_{it}.

If the purpose of the analysis is to identify and characterize the parameter variability, studies of recursive residuals and some functions of these have proved useful.

The recursive residuals, w_t, are independent $N(0, \sigma^2)$ variables for stable parameters, and are defined as

$$w_t = \frac{y_t - x_t' b_{t-1}}{\sqrt{1 + x_t'(X_{t-1}' X_{t-1})^{-1} x_t}}, \quad t = k+1, \ldots, T \tag{8.2}$$

where $X_{t-1}' = (x_1, \ldots, x_{t-1})$, $b_{t-1} = (X_{t-1}' X_{t-1})^{-1} X_{t-1}' Y_{t-1}$, and $Y_{t-1}' = (y_1, ..., y_{t-1})$ (see Brown *et al.*, 1975). A change in the parameter vector β_t at $t = t_1 + 1$ only changes the expectation of w_t from $t = t_1 + 1$ onward. The variance of w_t and the covariances, $Cov(w_t, w_{t+h})$, $h \neq 0$, are not affected by a nonstochastic parameter change. These properties of the recursive residuals suggest that they may be a suitable tool for detecting parameter variabilities.

In this chapter the CUSUM-SQ (CUmulated SUms of SQuared recursive residuals) and the MOSUM-SQ (MOving SUms of SQuared recursive residuals) test statistics are analyzed. The CUSUM-SQ test statistics are defined as

$$CS_t = \sum_{r=k+1}^{t} w_r^2 \Big/ \sum_{r=k+1}^{T} w_r^2, \quad t = k+1, \ldots, T \tag{8.3}$$

and the MOSUM-SQ test statistics are defined as

$$MQ_t = \frac{\sum_{r=t-G+1}^{t} w_r^2}{\sum_{r=k+1}^{t-G} w_r^2 + \sum_{r=t+1}^{T} w_r^2} \frac{T-k-G}{G}, \quad t = G+k, \ldots, T \tag{8.4}$$

Under H_0, CS_t follows a Beta distribution with parameters $(t-k)/2$ and $(T-t)/2$ (see Brown *et al.*, 1975) and MQ_t follows an F-distribution with G and $T-k-G$ degrees of freedom (see Hackl, 1980). In Section 8.3 we show theoretically how parameter variability, represented by the parameter model H_1, influences the expectations and the variances of CS_t and MQ_t. This theoretical study and a related numerical study, in Section 8.4, may throw some light on how these test statistics can be used to detect the timing and the nature of the parameter variability.

8.3 Properties of the Test Statistics

The main purpose of the analysis of the test statistics CS_t and MQ_t is to demonstrate how the paths of the test statistics are affected by the time of the start, $t_1 + 1$, and the end, t_2, of a parameter change. In particular the instantaneous case, PM1, is compared to the noninstantaneous one, PM2.

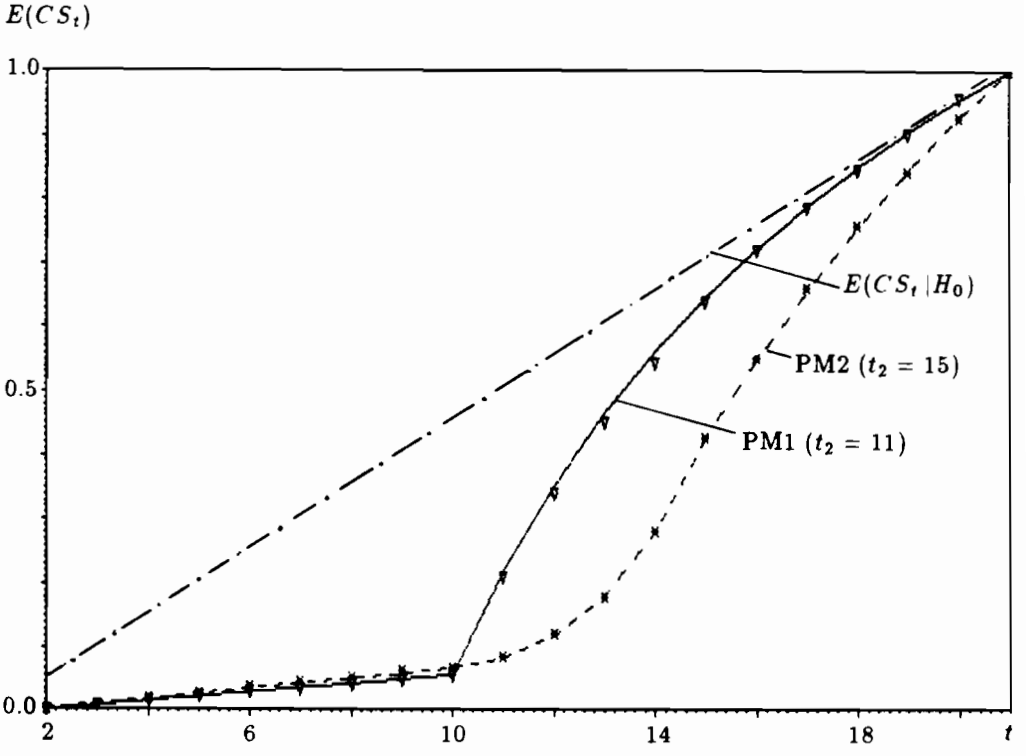

Figure 8.1: $E(CS_t)$, when $T = 21$, $t_1 = 10$; model M1. (Lines and symbols correspond to theoretical and estimated values, respectively.)

The distributions of the test statistics are known only for stable parameters. For this reason, the analysis is based on studies of their expectations and variances for different parameter models. As the exact moments are not known, approximations of these are analyzed. The size of these approximations is discussed in Section 8.4.

8.3.1 Properties of CS_t

Stable parameters

Under H_0, CS_t follows a Beta-distribution with parameters $(t-k)/2$ and $(T-t)/2$. The expectation is

$$E(CS_t) = (t-k)/(T-k)\,, \quad t = k+1, \ldots, T$$

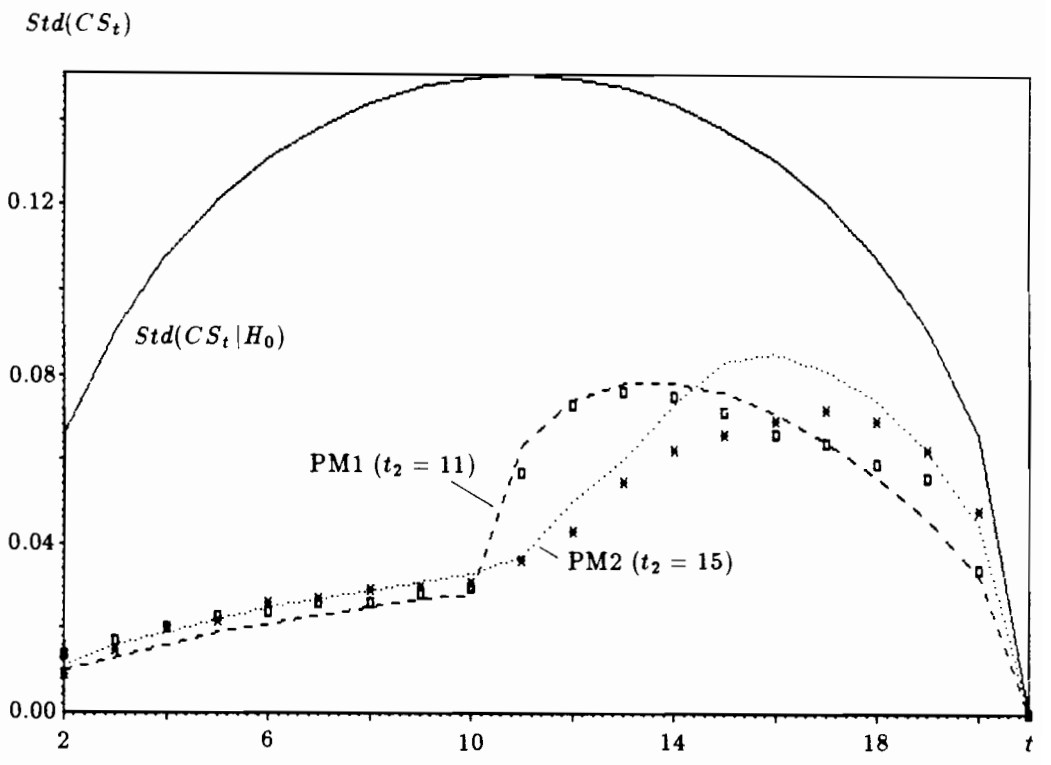

Figure 8.2: $Std(CS_t)$, when $T = 21$, $t_1 = 10$; model M1. (Lines and symbols correspond to theoretical and estimated values, respectively.)

i.e., an increasing function of t. The variance is

$$Var(CS_t) = 2(t-k)(T-t)/[(T-k+2)(T-k)^2] \ , \quad t = k+1, \ldots, T$$

i.e., an increasing function of t for $t = k+1, \ldots, (T+1)/2$, and a decreasing function of t for $t = (T+1)/2+1, \ldots, T$.

A change in the parameter vector β_t affects the distribution of the recursive residual w_t and, therefore, the expectation of CS_t. For that reason Brown *et al.* (1975) suggested the critical region:

$$|CS_t - \frac{t-k}{T-k}| > CS_\alpha \text{ for any } t = k+1, \ldots, T$$

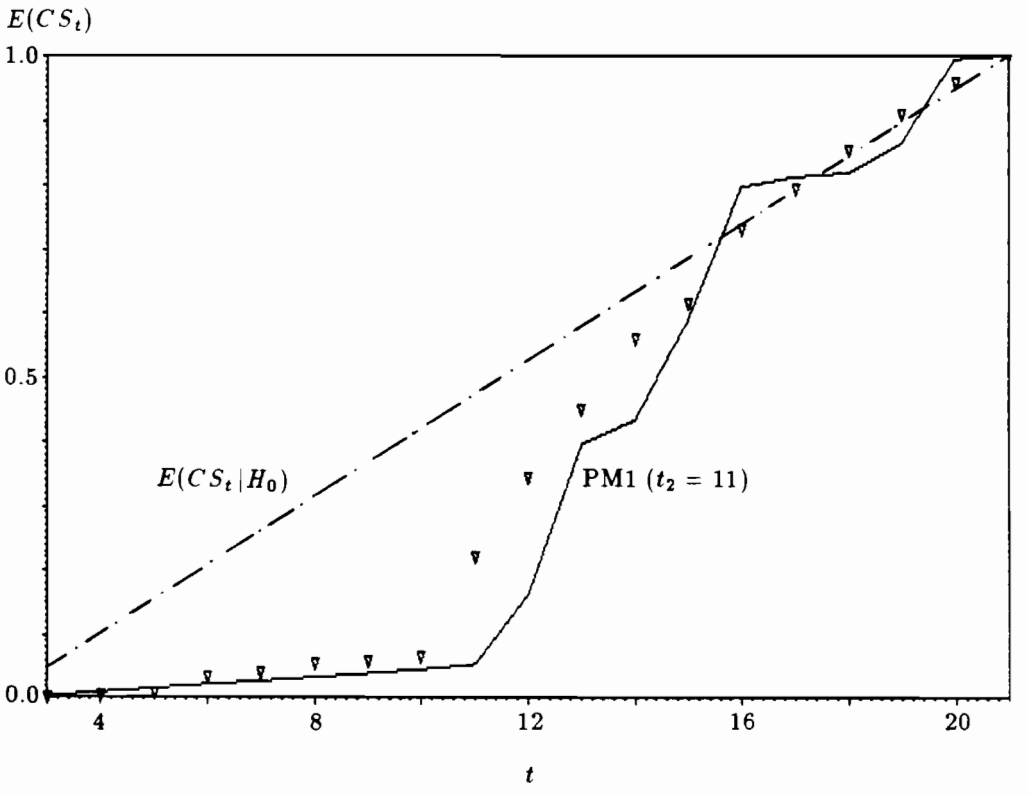

Figure 8.3: $E(CS_t)$, when $T = 21$, $t_1 = 10$; model M2 and β_{2t} varies. (Lines and symbols correspond to theoretical and estimated values, respectively.)

where CS_α depends on the significance level α.

Varying parameters

Gauss approximation formulae are used to derive the expectation and variance of CS_t for the parameter model H_1 (see Törnkvist, 1986a). For the intercept model (M1) we get

$$E(CS_t) \approx \frac{(t-1)D_t}{(T-1)D_T}\ , \quad t = 2, \ldots, T \tag{8.5}$$

and

$$Var(CS_t) \approx \frac{2(t-1)}{[(T-1)D_T]^2}\left[2D_t - 1 - \frac{(t-1)D_t}{(T-1)D_T}\left(2D_T - 2 + \frac{D_t}{D_T}\right)\right] \quad t = 2, \ldots, T \tag{8.6}$$

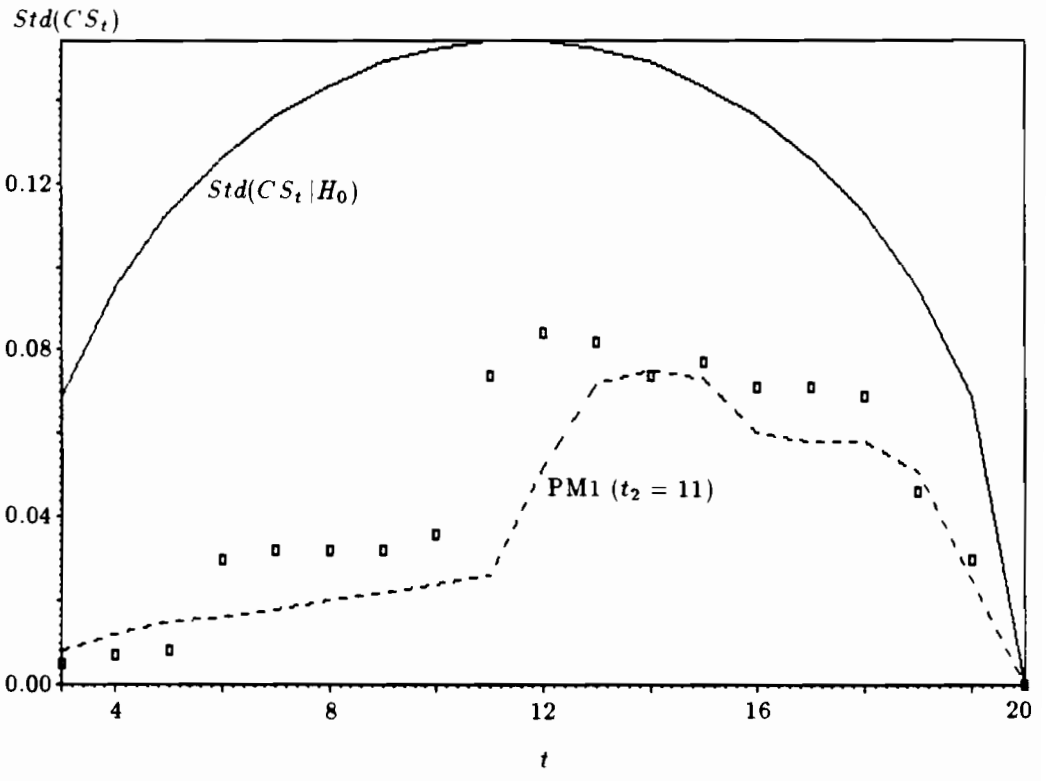

Figure 8.4: $Std(CS_t)$, when $T = 21, t_1 = 10$; model M2 and β_{2t} varies. (Lines and symbols correspond to theoretical and estimated values, respectively.)

where

$$D_t = \begin{cases} 1\ , \quad t = 2, \ldots, t_1 \\ 1 + \frac{\delta^2(t-t_1)}{4(t_2-t_1)^2(t-1)} \left[\frac{t(t+t_1)-(t_1-1)(5t_1-1)}{3} + \frac{t_1(t_1-1)^2}{t} \right] \\ \qquad\qquad t = t_1+1, \ldots, t_2 \\ 1 + \frac{\delta^2}{4(t-1)} \left[\frac{2t_2+4t_1-3+1/(t_2-t_1)}{1.5} - \frac{(t_2+t_1-1)^2}{t} \right] \\ \qquad\qquad t = t_2+1, \ldots, T \end{cases} \tag{8.7}$$

Approximate values of $E(CS_t)$ and the standard deviation $Std(CS_t)$ are calculated for $T = 21$, $\delta = 5$, $\sigma = 1$ for the intercept model (M1) and are illustrated in *Figures 8.1* and *8.2*.

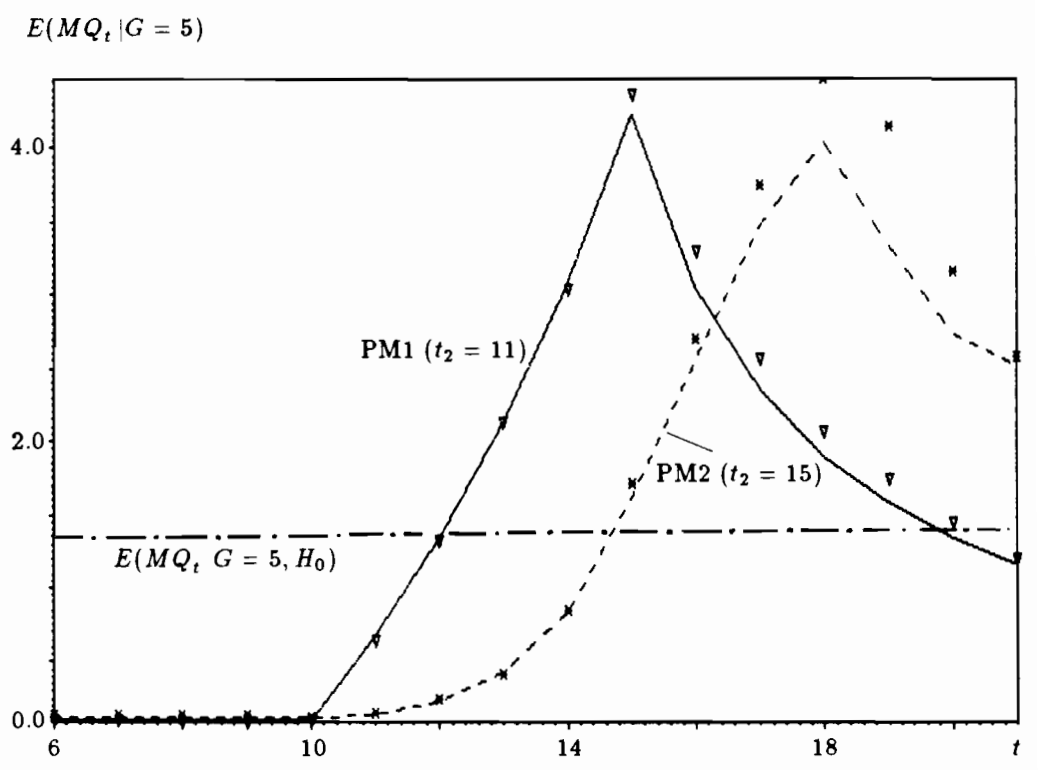

Figure 8.5: $E(MQ_t \mid G = 5)$, when $T = 21$, $t_1 = 10$; model M1. (Lines and symbols correspond to theoretical and estimated values, respectively.)

In *Figure 8.1* the paths of $E(CS_t)$ are illustrated for $t_1 = 10$, $t_2 = 11$ and $t_1 = 10$, $t_2 = 15$. $E(CS_t)$ is rather small for t up to $t = 10$, but increases rapidly after $t = 11$. When $t_2 = 11$, the $E(CS_t)$ increases faster than if $t_2 = 15$. This same pattern was indicated in a study by Törnkvist (1986a) with $T = 41$, $\delta = 2$, $\sigma = 1$, for different values of t_1 and t_2. The shorter the distance $t_2 - t_1$, the faster $E(CS_t)$ increases after $t = t_1$. *Figure 8.2* illustrates $Std(CS_t)$ for $t_1 = 10$, $t_2 = 11$ and $t_1 = 10$, $t_2 = 15$. Compared to $Std(CS_t)$ for stable parameters, these standard deviations are relatively small.

These illustrations seem to show that the time $t_1 + 1$ of the parameter change is easier to determine in the case of an instantaneous parameter change than in a noninstantaneous one. This is a result of the fact that the increase in $E(CS_t)$ from $t = t_1 + 1$ depends on the value of the parameter change per unit of time, $\delta\sigma/(t_2 - t_1)$.

For the regression model (M2), with one varying regression coefficient (the intercept is stable), which changes according to PM1, we get

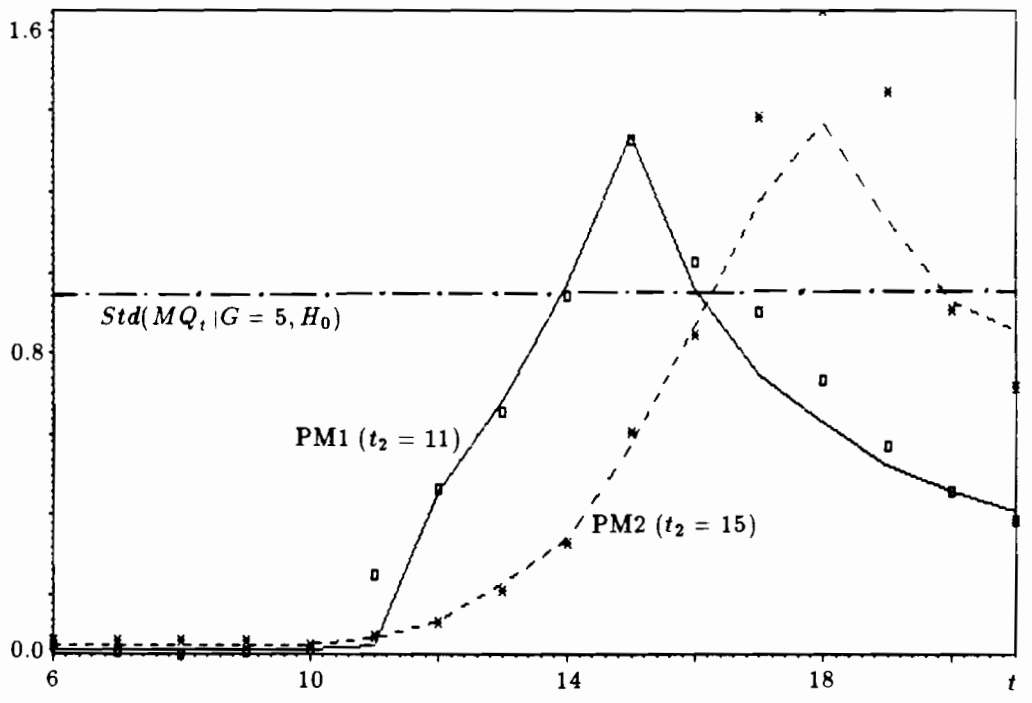

Figure 8.6: $Std(MQ_t \mid G = 5)$, when $T = 21$, $t_1 = 10$; model M1. (Lines and symbols correspond to theoretical and estimated values, respectively.)

$$E(CS_t) \approx \frac{(t-2)d_t}{(T-2)d_T}\ , \quad t = 3, \ldots, T \tag{8.8}$$

and

$$Var(CS_t) \approx \frac{2(t-2)}{[(T-2)d_T]^2}\left[2d_t - 1 - \frac{(t-2)d_t}{(T-2)d_T}\left(2d_t - 2 + \frac{d_t}{d_T}\right)\right] \quad t = 3, \ldots, T \tag{8.9}$$

where

$$d_t = \begin{cases} 1 & t = 3, \ldots, t_1 \\ 1 + \frac{\delta^2}{t-2}\sum_{r=t_1+1}^{t} S_r^2 & t = t_1+1, \ldots, T \end{cases} \tag{8.10}$$

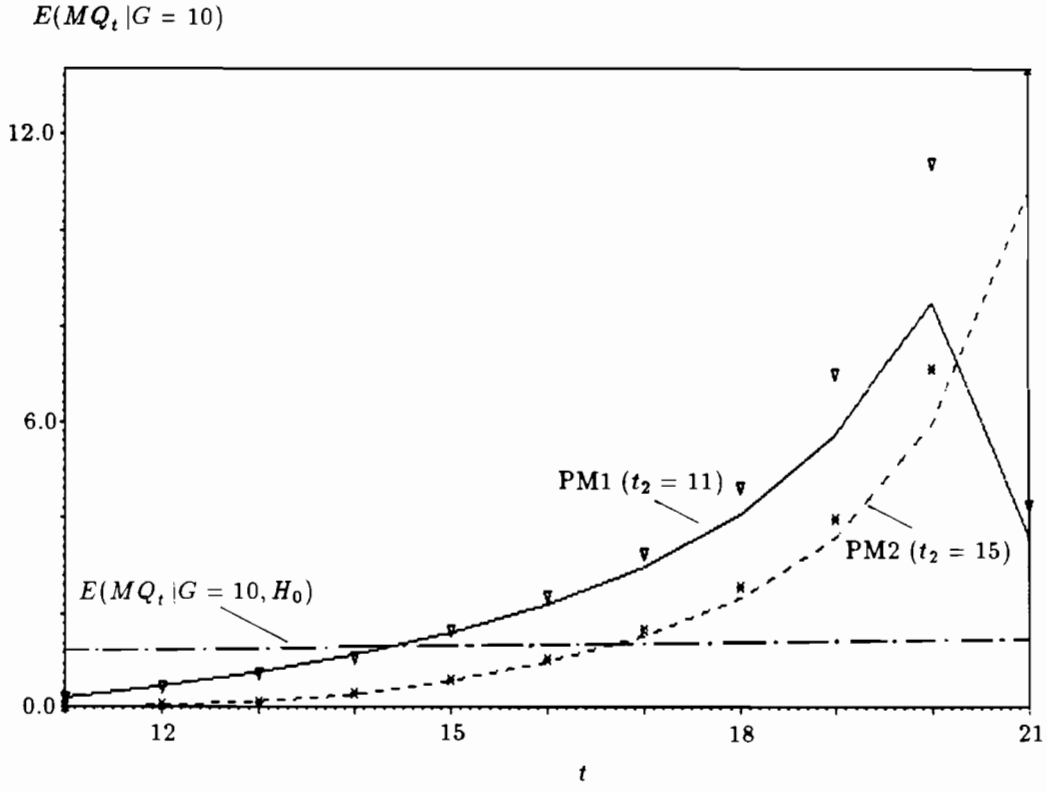

Figure 8.7: $E(MQ_t \mid G = 10)$, when $T = 21$, $t_1 = 10$; model M1. (Lines and symbols correspond to theoretical and estimated values, respectively.)

and

$$S_r = \left[x_r \left((r-1) \sum_{j=1}^{t_1} x_j^2 - \sum_{j=1}^{r-1} x_j \sum_{j=1}^{t_1} x_j \right) - \sum_{j=1}^{r-1} x_j \sum_{j=1}^{t_1} x_j^2 + \sum_{j=1}^{r-1} x_j^2 \sum_{j=1}^{t_1} x_j \right] / \left\{ \left[(r-1) \sum_{j=1}^{r-1} x_j^2 - \left(\sum_{j=1}^{r-1} x_j \right)^2 \right] \left[r \sum_{j=1}^{r} x_j^2 - \left(\sum_{j=1}^{r} x_j \right)^2 \right] \right\}^{1/2} \quad (8.11)$$

Figures 8.3 and *8.4* illustrate the approximate values of $E(CS_t)$ and $Std(CS_t)$, respectively, for model (M2) when $T = 21$, $\delta = 5$, $\sigma = 1$. The regression coefficient changes in an instantaneous way at $t = 11$. The values of x_t are generated from a normal distribution

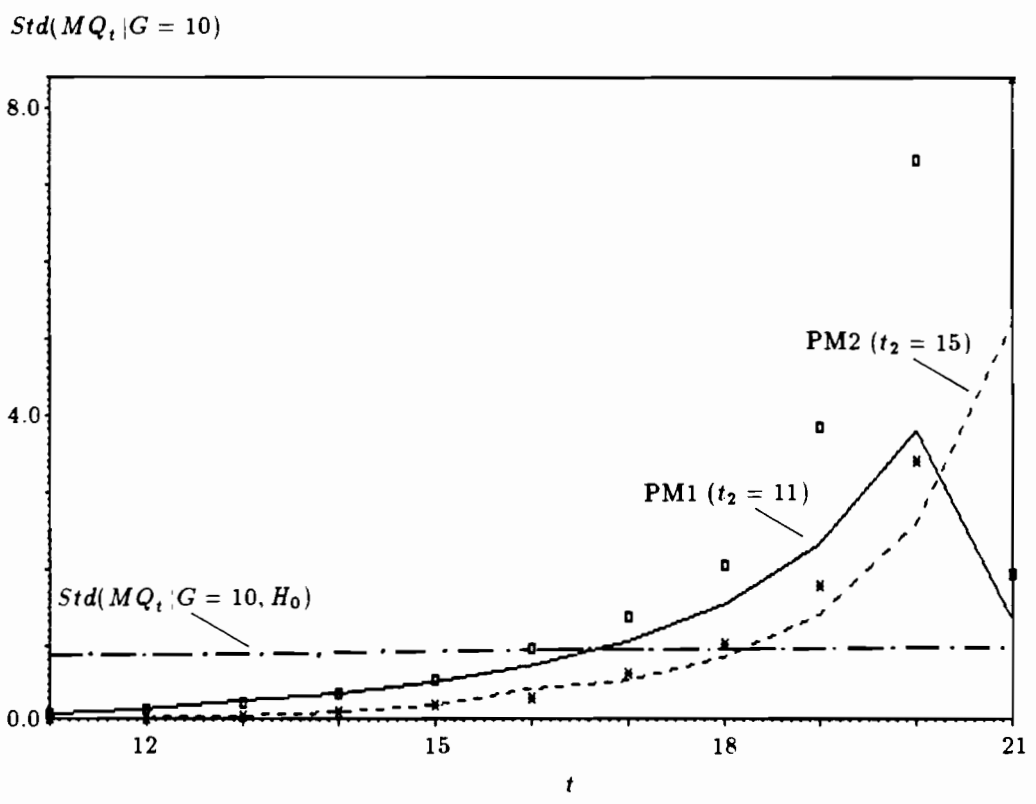

Figure 8.8: $Std(MQ_t \mid G = 10)$, when $T = 21$, $t_1 = 10$; model M1. (Lines and symbols correspond to theoretical and estimated values, respectively.)

with expectation 0 and variance 1. The increase of $E(CS_t)$ per unit of time is constant up to $t = t_1$, in the same way as for the intercept model, and is very small compared to $E(CS_t)$ for stable parameters. From $t = t_1 + 1$ the increase depends, except on $\delta\sigma$, also on the value of the regressor x_t. Large absolute values of x_t give large increases in $E(CS_t)$. $Std(CS_t)$ is affected by x_t in the same way, and is much smaller for every t (except $t = T$) than $Std(CS_t)$ for stable parameters.

8.3.2 Properties of MQ_t

Stable parameters

For stable parameters, the test statistic MQ_t follows an F-distribution with G and $T-k-G$ degrees of freedom. The expectation is

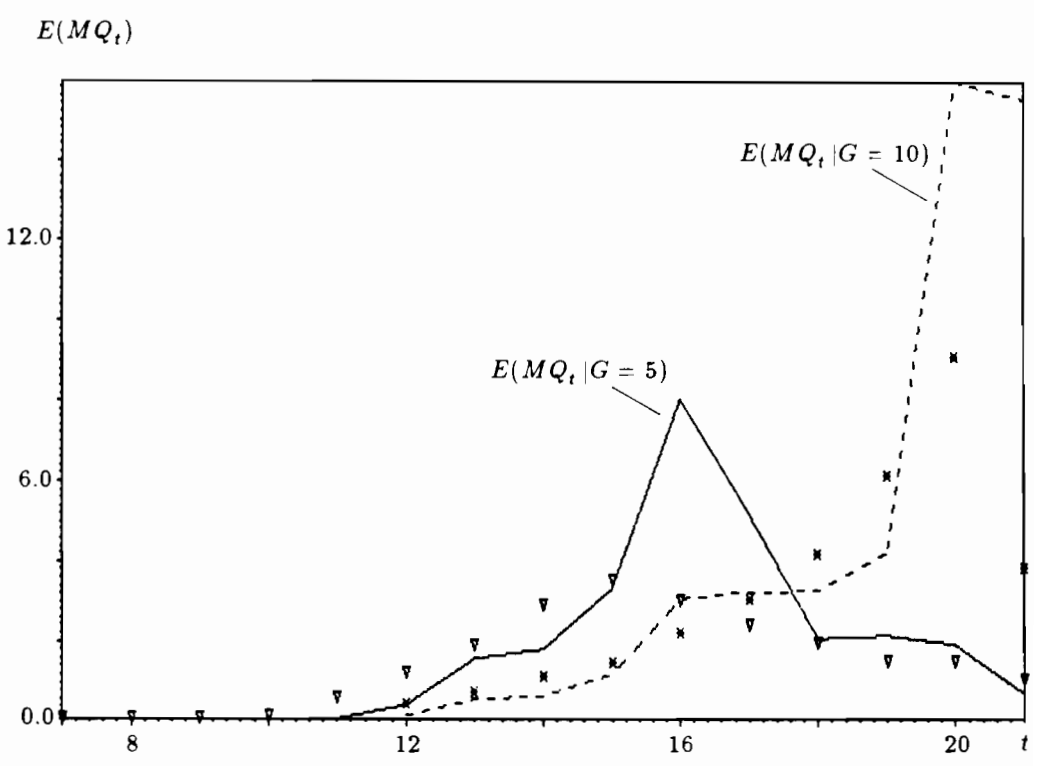

Figure 8.9: $E(MQ_t)$, when $T = 21$, $t_1 = 10$, and PM1 ($t_2 = 11$). (Lines and symbols correspond to theoretical and estimated values, respectively.)

$$E(MQ_t) = \frac{T-k-G}{T-2-G-k}\,, \quad t = G+k,\ldots,T$$

and the variance is

$$Var(MQ_t) = \left(\frac{T-k-G}{T-2-G-k}\right)^2 \frac{2}{G}\,\frac{T-k-2}{T-k-G-4}\,, \quad t = G+k,\ldots,T$$

i.e., expectation and variance are both constant for stable parameters.

Hackl (1980) suggests the following critical region:

$$MQ_t < mq_{l\alpha} \text{ or } MQ_t > mq_{u\alpha} \quad \text{for any } t = G+k,\ldots,T$$

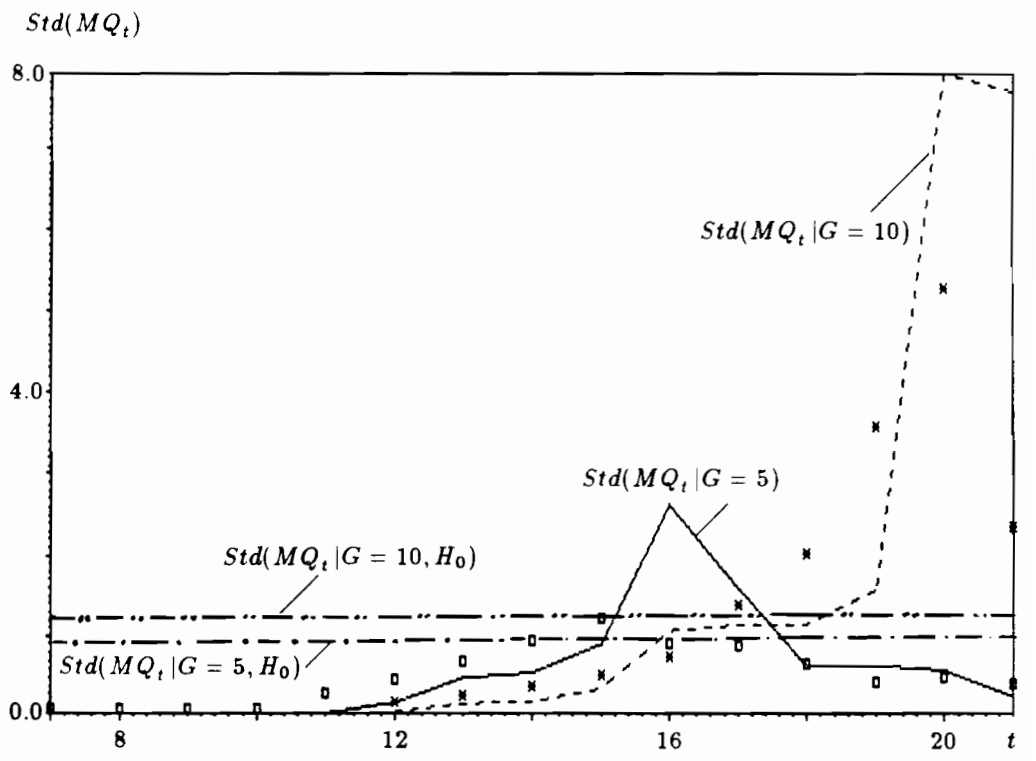

Figure 8.10: $Std(MQ_t)$, when $T = 21$, $t_1 = 10$, and PM1 ($t_2 = 11$). (Lines and symbols correspond to theoretical and estimated values, respectively.)

where

$$F(mq_{l\alpha}) = \frac{\alpha}{2(T-k-G+1)}$$
$$F(mq_{u\alpha}) = 1 - \frac{\alpha}{2(T-k-G+1)}$$

and F is the distribution function of Snedecor's F-distibution with G and $T-k-G$ degrees of freedom. This procedure gives a conservative test (see Hackl, 1980).

Varying parameters

As for CS_t, the Gauss approximation formulae are used to derive the expectation and variance of MQ_t for varying parameters. For parameter model H_1 and the intercept model (M1) we get

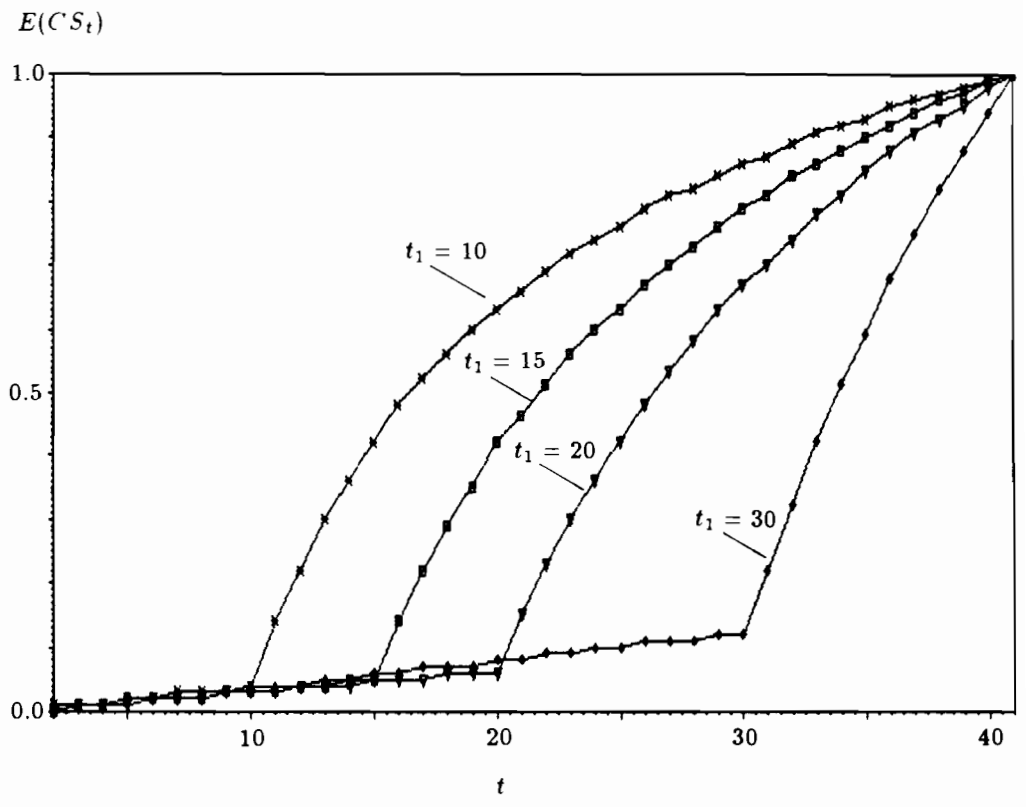

Figure 8.11: $E(CS_t)$, when M1, PM1, and $T = 41$. (Lines and symbols correspond to theoretical and estimated values, respectively.)

$$E(MQ_t) \approx \frac{(T-1-G)D_{t,G}}{(T-1)D_T - GD_{t,G}}\ , \quad t = G+1,\ldots,T \tag{8.12}$$

and

$$Var(MQ_t) \approx \frac{2(T-1-G)^2}{[(T-1)D_T - GD_{t,G}]^2}\,\frac{1}{G} \left[2D_{t,G} - 1 + \frac{GD_{t,G}^2}{(T-1)D_T - GD_{t,G}}\left(2 - \frac{T-1-G}{(T-1)D_T - GD_{t,G}}\right)\right] \quad t = G+1,\ldots,T \tag{8.13}$$

where

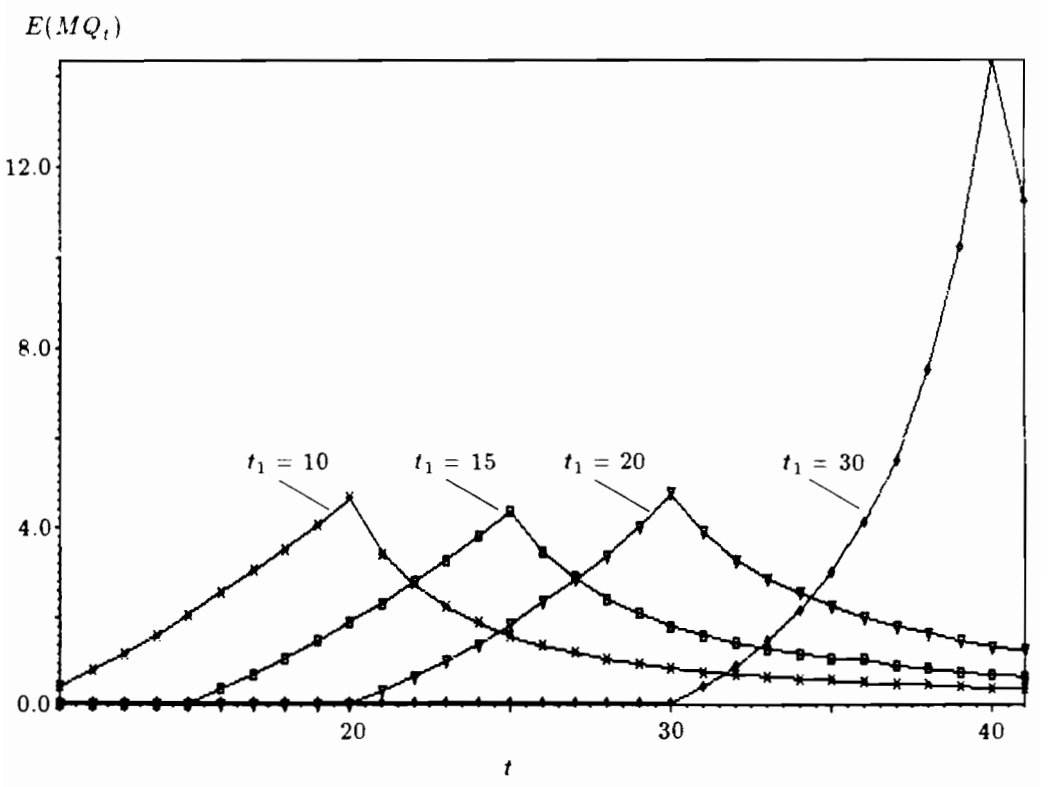

Figure 8.12: $E(MQ_t \mid G = 10)$, when M1, PM1, and $T = 41$. (Lines and symbols correspond to theoretical and estimated values, respectively.)

$$D_{t,G} = \begin{cases} 1, \quad t = G+1, \ldots, t_1 \\ 1 + \frac{\delta^2(t-t_1)}{4G(t_2-t_1)^2}\left[\frac{1}{3}[t(t+t_1) - (t_1-1)(5t_1-1)] + \frac{t_1(t_1-1)^2}{t}\right] \\ \qquad\qquad t = t_1+1, \ldots, t_m^- \\ 1 + \frac{\delta^2}{4G}\left[\frac{2}{3}[2t_2 + 4t_1 - 3 + \frac{1}{t_2-t_1}] - \frac{(t_2+t_1-1)^2}{t}\right] \\ \qquad\qquad t_2 < t_1 + G\ , \quad t = t_2+1, \ldots, t_1+G \\ 1 + \frac{\delta^2}{4(t_2-t_1)^2}\left[t(t-G) + \frac{G^2-1}{3} - 2t_1(t_1-1) + \frac{t_1^2(t_1-1)^2}{t(t-G)}\right] \\ \qquad\qquad t_2 > t_1 + G\ , \quad t = t_1+G+1, \ldots, t_2 \\ 1 + \frac{\delta^2}{4G}\left\{\frac{t_2-t+G}{(t_2-t_1)^2}\left[\frac{1}{3}\left[t(t+t_2-2G) + t_2(t_2-G) + G^2 - 1\right] - \right.\right. \\ \qquad \left.\left. 2t_1(t_1-1) + \frac{t_1^2(t_1-1)^2}{t_2(t-G)}\right] + \frac{(t_2+t_1-1)^2(t-t_2)}{tt_2}\right\} \\ \qquad\qquad t = t_m^+ + 1, \ldots, t_2 + G - 1 \\ 1 + \frac{\delta^2(t_2+t_1-1)^2}{4t(t-G)}\ , \quad t = t_2+G, \ldots, T \end{cases} \tag{8.14}$$

where $t_m^- = \min(t_1+G, t_2)$, $t_m^+ = \max(t_1+G, t_2)$, and D_T is defined as (8.7).

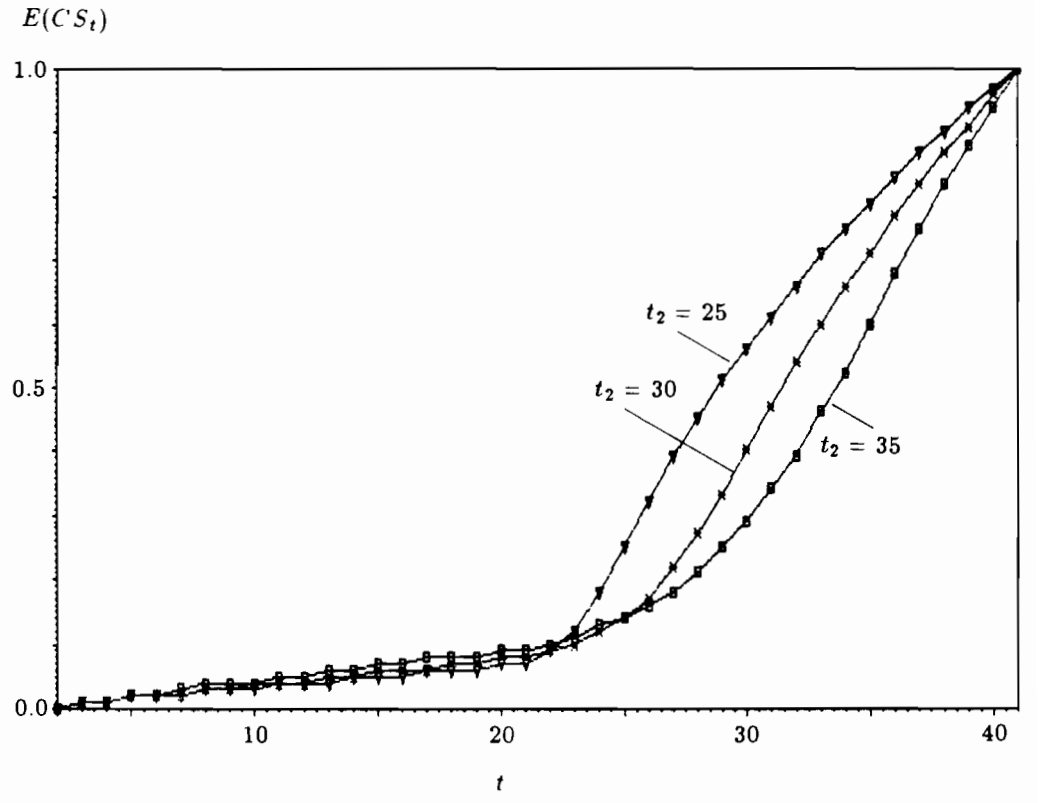

Figure 8.13: $E(CS_t)$, when M1, PM2, $t_1 = 20$ and $T = 41$. (Lines and symbols correspond to theoretical and estimated values, respectively.)

Approximate values of $E(MQ_t)$ and $Std(MQ_t)$ are, as for CS_t, calculated for $T = 21$, $\delta = 5$, $\sigma = 1$ and illustrated in *Figures 8.5–8.8*. To study the importance of the length of the moving window G, $E(MQ_t)$ and $Std(MQ_t)$ are calculated for two values of G: 5 and 10.

Figures 8.5 and *8.6* illustrate the paths of $E(MQ_t)$ and $Std(MQ_t)$, respectively, for $G = 5$ and the parameter models with $t_1 = 10$, $t_2 = 11$ and $t_1 = 10$, $t_2 = 15$. $E(MQ_t)$ is constant up to $t = t_1$, increases to its maximum at $t \geq t_1 + G$, and then decreases. The maximum is attained at $t = t_1 + G$ for $t_2 = t_1 + 1$. The larger the value of t_2, the later the maximum is attained.

$Std(MQ_t)$ increases and decreases with $E(MQ_t)$. It is about 65% of $E(MQ_t)$ for $t = 6, \ldots, 10$ and about 32% at $t = t_1 + G$, for $t_2 = 11$. For $t_2 = 15$ it is about 66% for $t = 6, \ldots, 10$, 36% at $t = t_1 + G$, and 34% at $t = t_1 + G + 3$, where $E(MQ_t)$ attains its maximum.

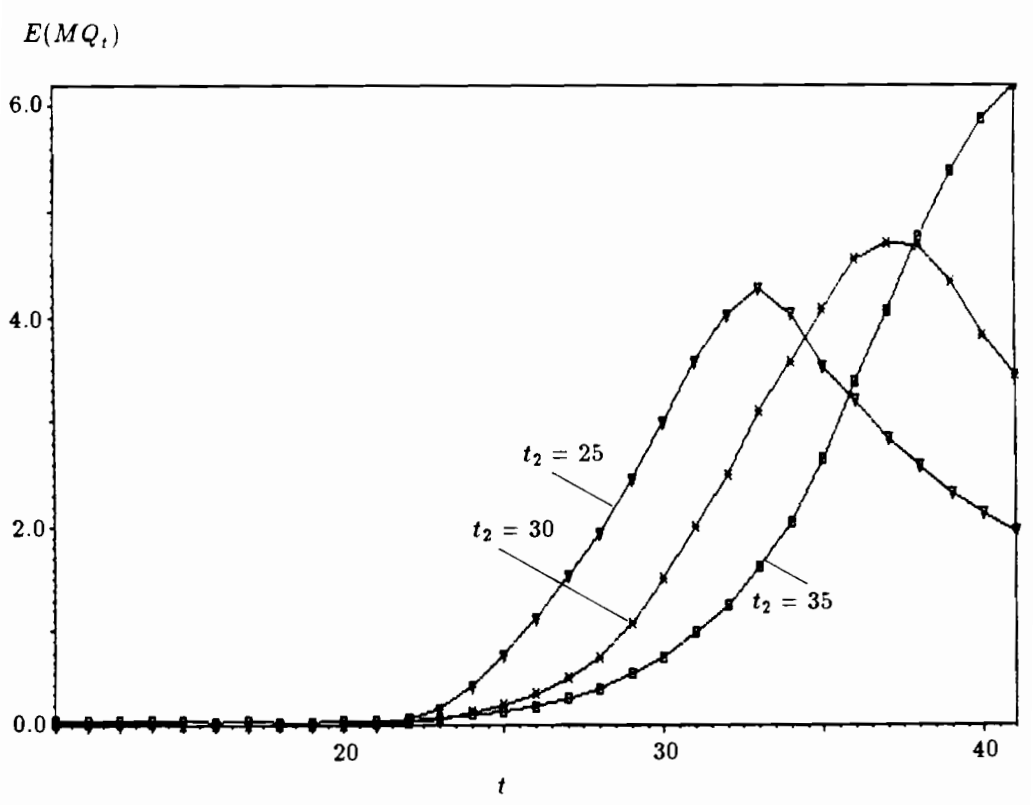

Figure 8.14: $E(MQ_t \mid G = 10)$, when M1, PM2, $t_1 = 20$ and $T = 41$. (Lines and symbols correspond to theoretical and estimated values, respectively.)

For $G = 10$ (see *Figures 8.7* and *8.8*), $E(MQ_t)$ increases from $t = G+1$ to its maximum at $t = t_1 + G$, for $t_2 = 11$, and at $t = T$, for $t_2 = 15$. In both these cases, t_1 less than $G + 1$ and the maximum attained at $t = T$, it is rather difficult to determine the values of t_1 and t_2. The "best" choice of value on G thus depends, except on T, on t_1 and t_2. As these values are not known in practice, the MQ_ts should be calculated for at least two values on G.

For the regression model (M2) with a stable intercept and one regressor, with a varying regression coefficient that changes according to PM1, we get

$$E(MQ_t) \approx \frac{(T-2-G)d_{t,G}}{(T-2)d_T - Gd_{t,G}}\ , \quad t = G+2, \ldots, T \tag{8.15}$$

and

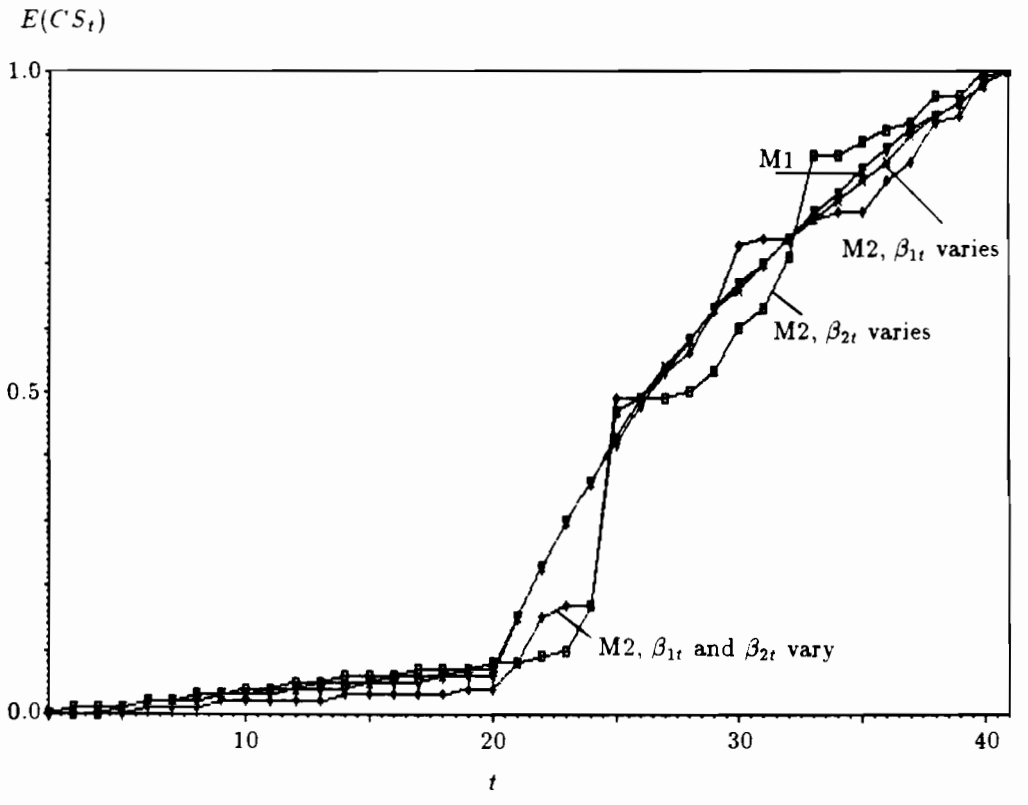

Figure 8.15: $E(CS_t)$, when PM1, $t_1 = 20$ and $T = 41$. (Lines and symbols correspond to theoretical and estimated values, respectively.)

$$Var(MQ_t) \approx \frac{2(T-2-G)^2}{[(T-2)d_T - Gd_{t,G}]^2}\frac{1}{G}$$
$$\left[2d_{t,G} - 1 + \frac{Gd_{t,G}^2}{(T-2)d_T - Gd_{t,G}}\left(2 - \frac{T-2-G}{(T-2)d_T - Gd_{t,G}}\right)\right]$$
$$t = G+2,\ldots,T \qquad (8.16)$$

where

$$d_{t,G} = \begin{cases} 1, & t = G+2,\ldots,t_1 \\ 1 + \frac{\delta^2}{G}\sum_{\tau=t_1+1}^{t} S_\tau^2, & t = t_1+1,\ldots,t_1+G \\ 1 + \frac{\delta^2}{G}\sum_{\tau=t-G+1}^{t} S_\tau^2, & t = t_1+G+1,\ldots,T \end{cases} \qquad (8.17)$$

and d_T and S_τ are defined as in (8.10) and (8.11), respectively.

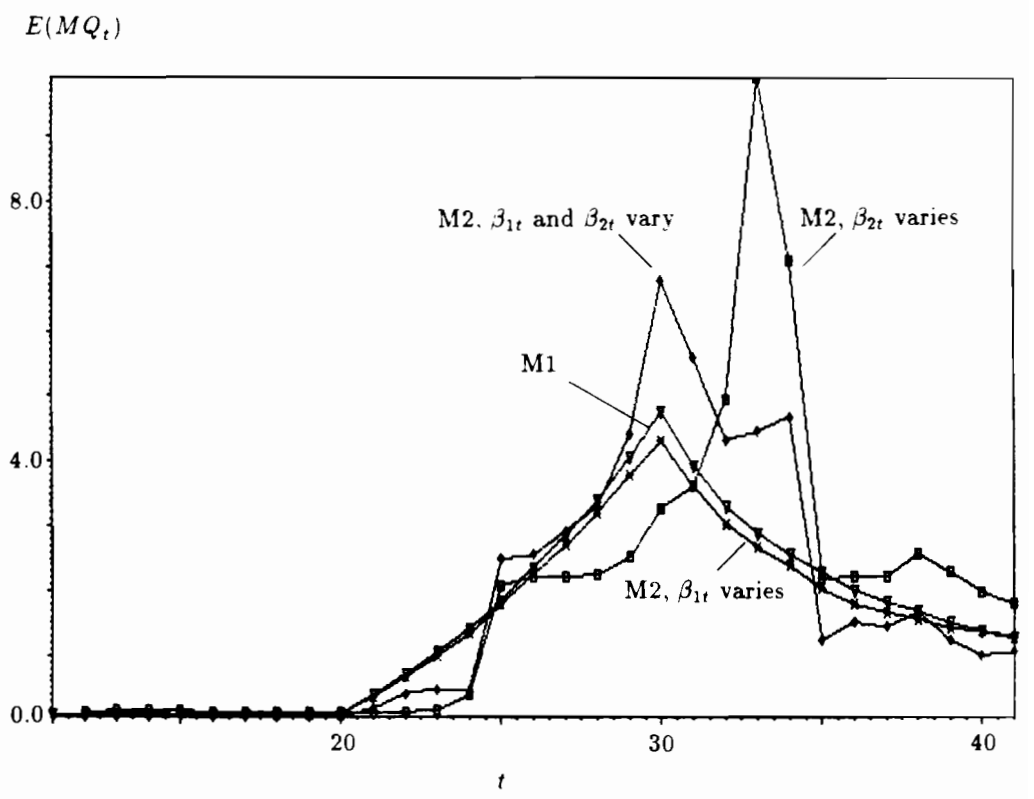

Figure 8.16: $E(MQ_t \mid G = 10)$, when PM1, $t_1 = 20$ and $T = 41$. (Lines and symbols correspond to theoretical and estimated values, respectively.)

Figures 8.9 and *8.10* illustrate $E(MQ_t)$ and $Std(MQ_t)$, respectively, for this regression model. The regression coefficient changes at $t_1 = 11$. The values of x_t are the same as in Section 8.3.1 (varying parameters). In *Figure 8.9*, $E(MQ_t \mid G = 5)$ is compared to $E(MQ_t \mid G = 10)$. In the first case the maximum is attained at $t = t_1 + G + 1$ and in the second case at $t = t_1 + G$. When studying these paths, we also have to study the values of the regressor. As the absolute value of x_t is much larger for $t = 16$ ($x_{16} = -1.46859$) than for $t = 11$ ($x_{11} = 0.12558$), the maximum is attained at $t = t_1 + G + 1$ and not at $t = t_1 + G$, for $G = 5$. For $G = 10$, the maximum is attained at $t = t_1 + G$ because $x_{21} = 0.03651 < x_{11}$.

As for the intercept model, $E(MQ_t \mid G = 5)$ is constant up to $t = t_1$. The increase of $E(MQ_t)$ depends on the absolute values of x_t, in the same way as $E(CS_t)$. It may be more difficult to determine the time of the parameter change for model (M2) than for model (M1), as the values of x_t must be kept in mind.

8.3.3 Summary

The quantities $E(CS_t)$ and $Std(CS_t)$ are small, compared with corresponding values for stable parameters, for t up to $t = t_1$. From $t = t_1 + 1$, $E(CS_t)$ increases rather rapidly. The shorter the distance $t_2 - t_1$, the faster $E(CS_t)$ increases after $t = t_1$.

The quantites $E(MQ_t)$ and $Std(MQ_t)$ are constant up to $t = t_1$, increase from $t = t_1+1$ to $t = t_1 + G$ and then decrease for PM1. For PM2 the maximum is attained after $t_1 + G$. The increase in $E(MQ_t)$ depends on t_2, in the same way as the increase in $E(CS_t)$. Furthermore, the larger the value of t_2, the later in the time period the maximum is attained.

8.4 Structural Change Characterization: A Numerical Study

The following study of structural variability test statistics has two basic objectives. First, the study is designed to indicate the degree of approximation typical for the different theoretical analyses given in the previous section. Expectations and standard deviations of the test statistics are thus estimated and compared with those determined analytically. Second, the numerical study will supplement the theoretical analysis by estimating other characteristics of these statistics. Characteristics that are mathematically intractable to handle analytically, e.g., skewness, median, and different percentiles, are estimated.

The design of the numerical study is first briefly outlined. Then, some of the basic results of the study are presented and discussed.

8.4.1 The design

1. The data are generated according to two models which are special cases of (8.1):

   M1 — an intercept model, with $k = 1$, and β_{1t} varying over time.

   M2 — a simple linear regression model, with $k = 2$, where β_{1t} and/or β_{2t} may vary over time.

2. The parameters β_{1t} in (M1) and β_{1t} and β_{2t} in (M2) vary according to two parameter variability models:

   PM1 — instantaneous changes at $t = t_1 + 1$.

   PM2 — gradual linear changes, starting at $t = t_1 + 1$ and continuing up to $t = t_2$.

   For (M2), three different cases are studied: (i) only β_{1t} changes, (ii) only β_{2t} changes, and (iii) β_{1t} and β_{2t} change according to the same parameter variability model. The total parameter change, $\delta\sigma$, is in all cases equal to 5, where $\delta = 5$ and $\sigma = 1$.

3. Three different time series length cases are studied, viz. $T = 21$, 41 and 81.

4. The distributions of the following two test statistics are studied:

   - CUSUM-SQ [see (8.3)]
   - MOSUM-SQ [see (8.4)]

   The "window length" G in MOSUM-SQ varies for different T values as follows:

   - $T = 21$: $G = 5, 10$
   - $T = 41$: $G = 10, 15$
   - $T = 81$: $G = 10, 20$

   Both test statistics are here based on recursive least-squares residuals.

5. The time for parameter changes is varied as follows:

   - $T = 21$: $t_1 = 10$ and $t_2 = 11, 15$
   - $T = 41$: $t_1 = 10$ and $t_2 = 11, 15, 20, 25$
     $t_1 = 15$ and $t_2 = 16, 20, 25, 30$
     $t_1 = 20$ and $t_2 = 21, 25, 30, 35$
     $t_1 = 30$ and $t_2 = 31, 35, 40$
   - $T = 81$: $t_1 = 40$ and $t_2 = 41, 45, 55$
     $t_1 = 60$ and $t_2 = 61, 65, 75$

6. The characteristics studied numerically are:

   - expectation
   - standard deviation
   - skewness
   - median
   - percentiles (P_{01}, P_{05}, P_{25}, P_{75}, P_{95}, P_{99}, where P_ν denotes the νth percentile).

7. All estimates of these characteristics are based on 100 replicates.

8.4.2 Basic results

It is evident that numerical simulation studies cannot provide a strict evaluation of the degree of approximation typical of the analytical study, as the simulation results are characterized by a numerical uncertainty. However, the simulation results will indicate to some extent whether analytical approximation errors are negligible or not, by comparing analytically derived and numerically determined estimates.

As for the estimation of expected values of the test statistics, the chief impression is that numerical and analytical approaches provide very similar estimates. This is particularly

true for the CUSUM-SQ test (see *Figures 8.1* and *8.3*). For model M2, however, the numerical estimates of $E(CS_t)$ increases faster at $t = t_1 + 1$ and for a brief period after the parameter shift. This means that to a certain extent the theoretical evaluation might overestimate the possibility to discover the parameter shift.

There are some discrepancies when estimating $E(MQ_t/G)$, but only close to the maximum of $E(MQ_t/G)$. Such discrepancies seem to increase with G (see *Figures 8.5*, *8.7*, and *8.9*). These statements will do for PM1 as well as for PM2. As for CS_t, *Figure 8.9* indicates certain risks for theoretically overestimating the possibility to detect discrete parameter shifts by MQ_t in the case of model M2 with varying β_{2t}. Here the differences between theoretical and numerical estimates are actually quite large around the maximum of $E(MQ_t)$.

In the intercept model case (M1), we observe when using MQ_t some tendencies to theoretically underestimate the possibility to detect discrete intercept shifts, although discrepancies are rather small between numerically estimated and theoretically determined expectations.

Even when estimating variances of test statistics, we cannot find any conspicuous differences between analytical and numerical estimates, at least not for the CUSUM-SQ test (see *Figures 8.2* and *8.4*). For model M2 (when β_{2t} varies) the theoretical standard deviation is slightly smaller than the numerically estimated standard deviation. For the MOSUM-SQ tests, corresponding approximation errors seem to be negative, and are in a few cases [close to the maximum of $Std(MQ_t)$] quite considerable. This is particularly true for large values of G, and in the case of PM2 (see *Figures 8.6*, *8.8* and *8.10*). However, in the M2 case (when β_{2t} varies), the theoretically determined maximum of $Std(MQ_t)$ seems clearly overestimated.

Besides indicating the degree of approximation typical for the theoretical analysis in Section 8.3, this numerical study also generates results with respect to the general ability of CUSUM-SQ and MOSUM-SQ tests to characterize certain structural variability.

The first situation considered is when the structural variability is *a priori* known to have caused a discrete parameter shift. The remaining problem is then to determine the time of that parameter change. All numerical results seem to show that this time dating problem is reasonably solvable. This is so, even if the estimated expectations and standard deviations are considered simultaneously, and t_1 seems to be easily identified by CUSUM-SQ as well as MOSUM-SQ procedures (see, e.g., *Figure 8.11* and *8.12*). The time dating ability increases with the length of the time series and also, of course, with the size of the parameter shift.

The empirically more common situation is that it cannot *a priori* be decided whether the structural variability implies instantaneous or gradual parameter changes. Judging by the numerical results, it is obviously rather difficult to identify and distinguish different types of parameter shift models. This is so in particular when applying the CUSUM-SQ test procedure, and if expectations as well as standard deviations are considered (see *Figures 8.13* and *8.14*). It seems somewhat easier to identify various cases of PM2 by the MOSUM-SQ test, basically because the maximum of MQ_t varies with t_2.

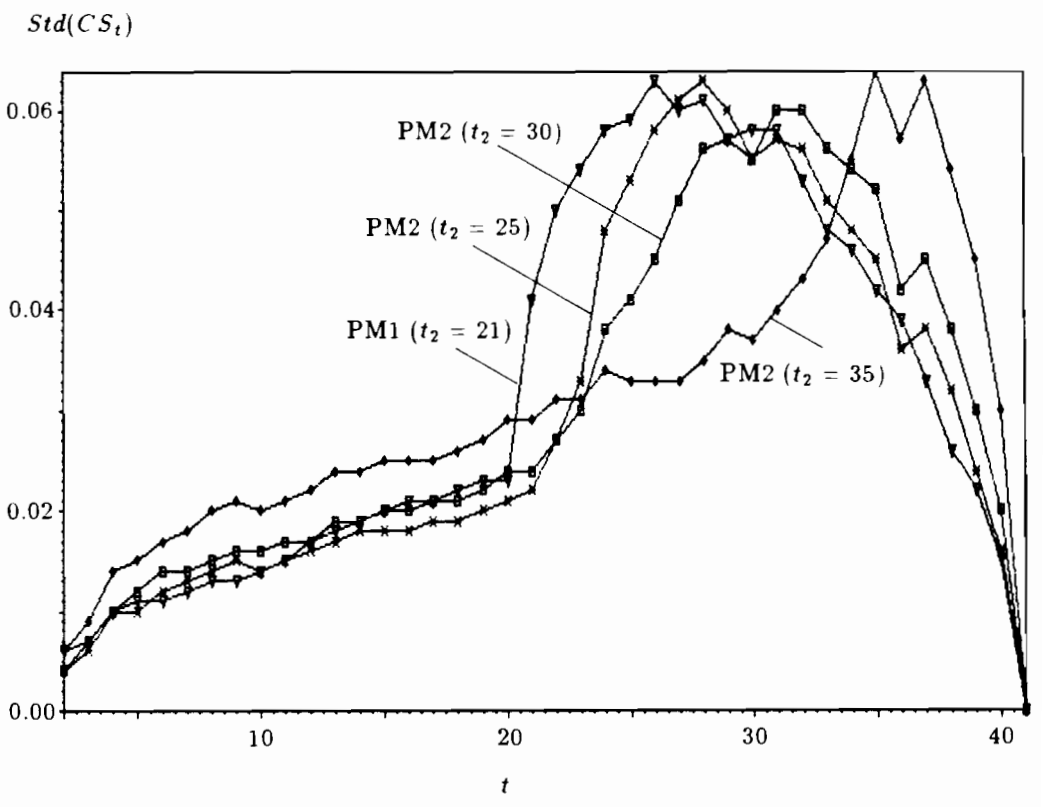

Figure 8.17: $Std(CS_t)$, when M1, $t_1 = 20$ and $T = 41$. (Lines and symbols correspond to theoretical and estimated values, respectively.)

In general, models involve several structural parameters. Structural variability will not necessarily be represented by parameter variability with respect to all parameters, but rather be concentrated to a subset of the parameters. It is then essential to be able to identify those parameters that actually will vary. One aspect of that problem is investigated to some extent within the present numerical study. A simple regression model with two parameters is examined. Three cases, characterized by different parameter variability situations, are separated. It is found to be nearly impossible to separate the three cases of M2 (see *Figures 8.15* and *8.16*) by the means of the CUSUM-SQ or the MOSUM-SQ test. When the intercept is constant (but β_{2t} varies), however, the case seems to differ from the other two alternatives in the sense that it will imply different test paths.

It is stated above that it seems difficult to identify PM1 and different PM2. That is further illustrated by the estimated standard deviations in *Figure 8.17*. *Figure 8.18*

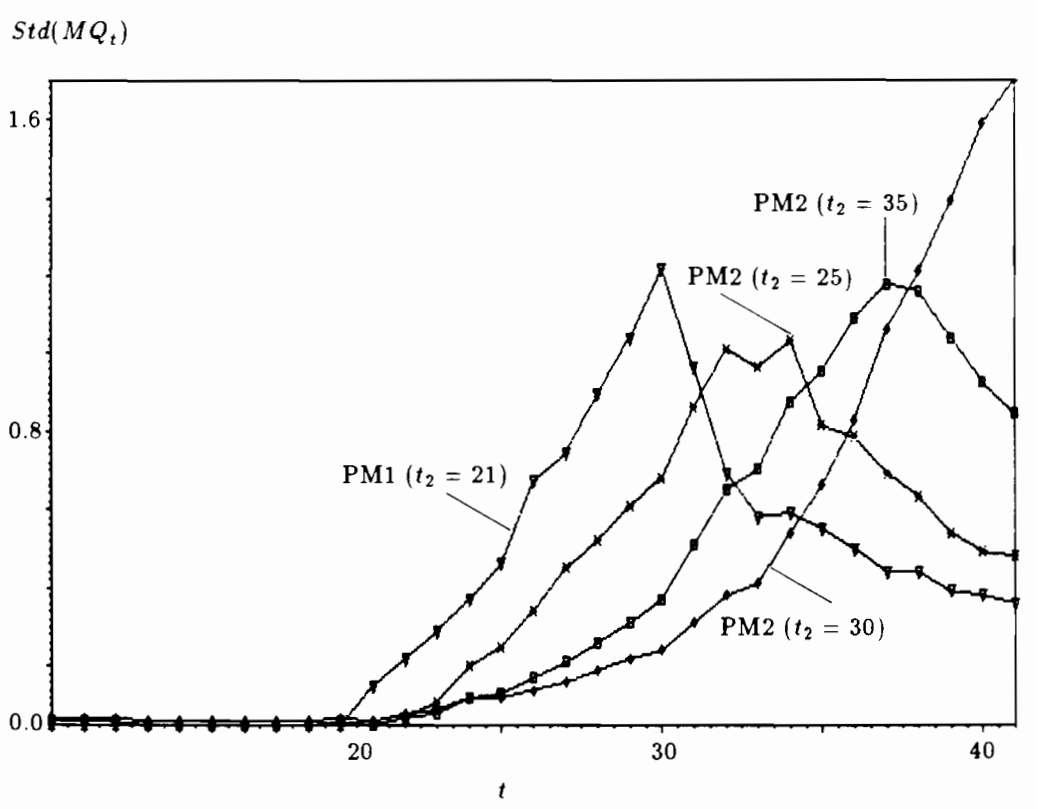

Figure 8.18: $Std(MQ_t \mid G = 10)$, when M1, $t_1 = 20$, and $T = 41$. (Lines and symbols correspond to theoretical and estimated values, respectively.)

supports the idea that MOSUM-SQ statistics might better identify different parameter variability models. Finally, *Figure 8.19* corroborates what was said earlier about difficulties in identifying different cases of M2.

The observed estimated expectations and standard deviations will indicate the possibilities of characterizing parameter variability to some extent. It is known, however, that the distributions of CUSUM-SQ as well as MOSUM-SQ test statistics are skew, a fact that further complicates the characterization process. The degree of skewness subject to the alternative hypothesis H_1 is not theoretically known, but numerically estimated within the framework of the present study.

In general, the degrees of skewness are considerable, above all for the MOSUM-SQ statistic, which is always positively skew. The CUSUM-SQ test statistic is positively skew up to the parameter shift period, after which the skewness will be negative, and the size partly decreases. *Figures 8.20* and *8.21* illustrate the general picture with respect to observed skewness for a few cases.

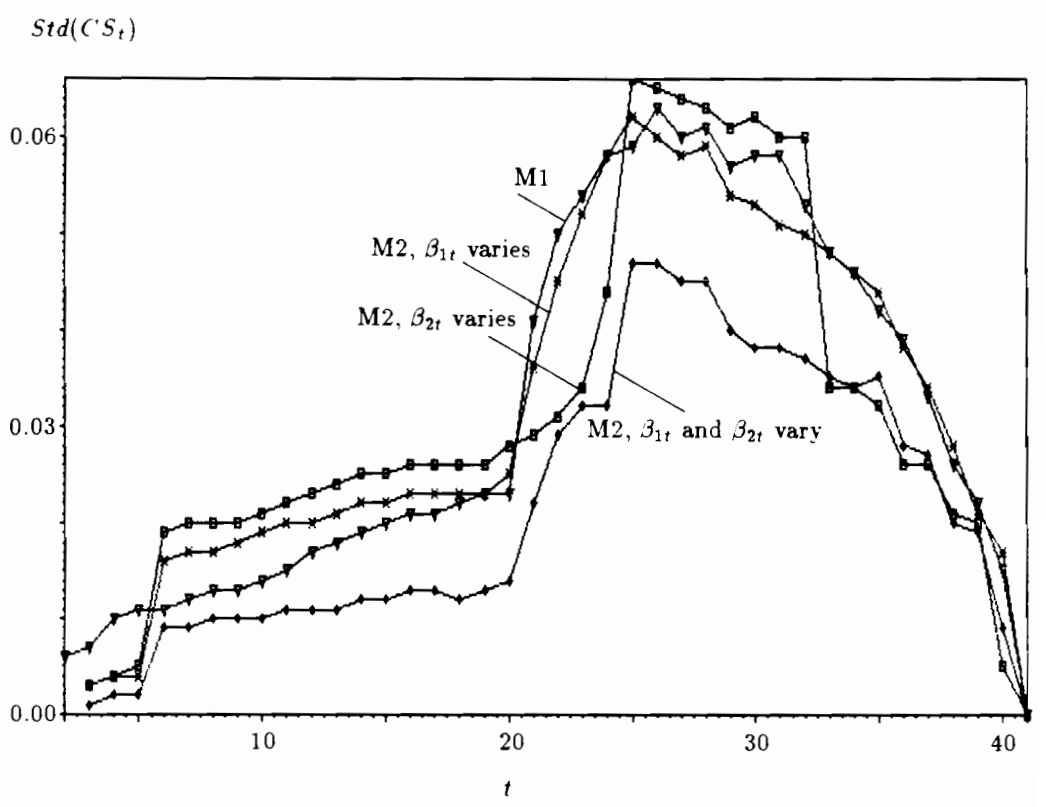

Figure 8.19: $Std(CS_t)$, when PM1, $t_1 = 20$ and $T = 41$. (Lines and symbols correspond to theoretical and estimated values, respectively.)

8.5 Conclusions

Hypotheses about structural variability in economic and social systems are sometimes formulated on a purely theoretical basis. Often, however, empirical support is needed. Is it possible to use empirical studies in order to generate such hypotheses? If so, how may this be done?

The questions are partially answered if, from empirical studies, it is possible to identify when the variability is introduced or changed. In order to contribute to this kind of knowlegde, we have focused here on examining the potential of test statistics for testing the hypothesis of structural stability.

We have analyzed the CUSUM-SQ and the MOSUM-SQ test statistics. Several approaches can be used for this kind of analysis, e.g., theoretical studies of the distributions of the test statistics, Monte Carlo experiments, and empirical studies. The distributions

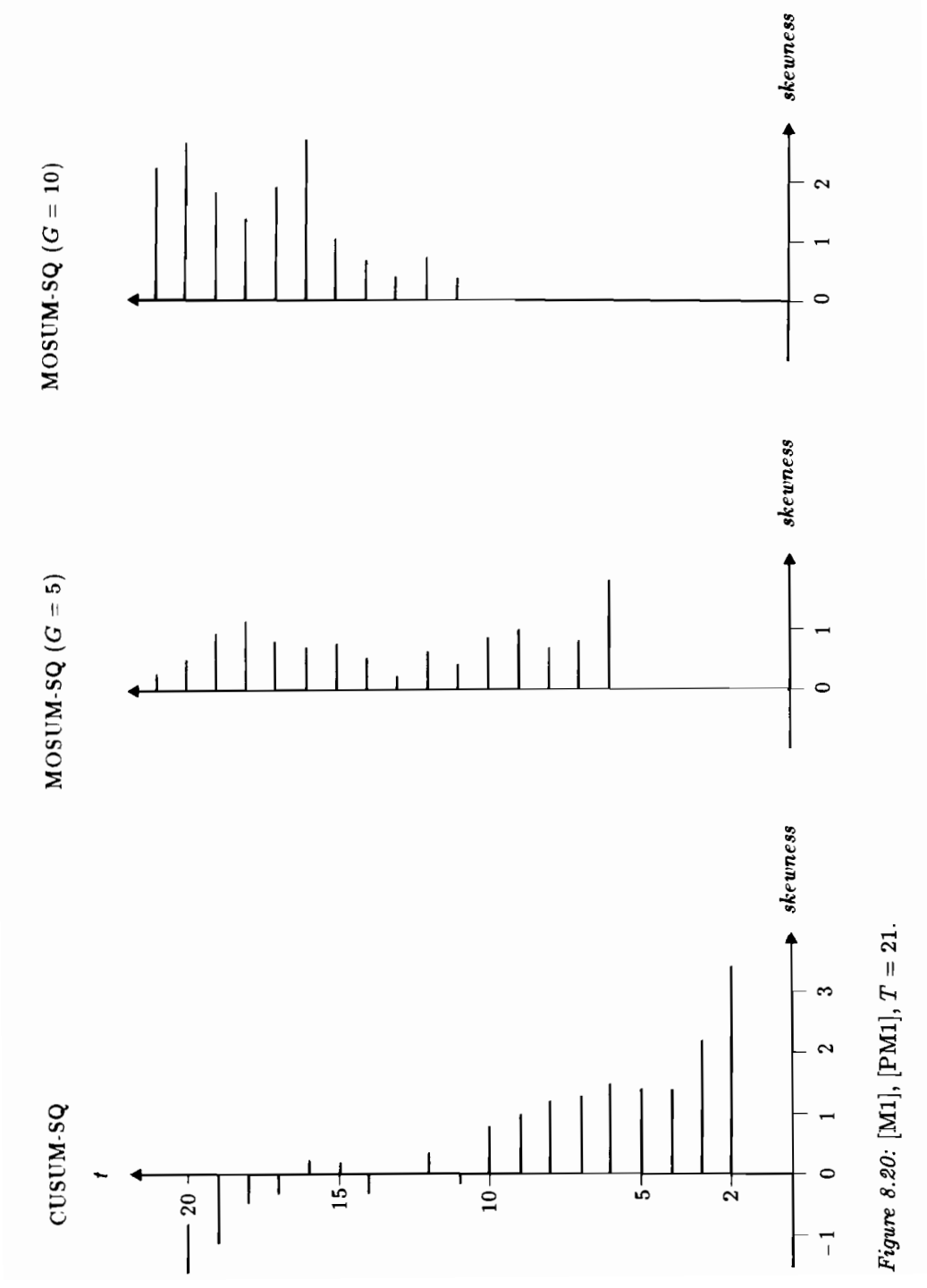

Figure 8.20: [M1], [PM1], $T = 21$.

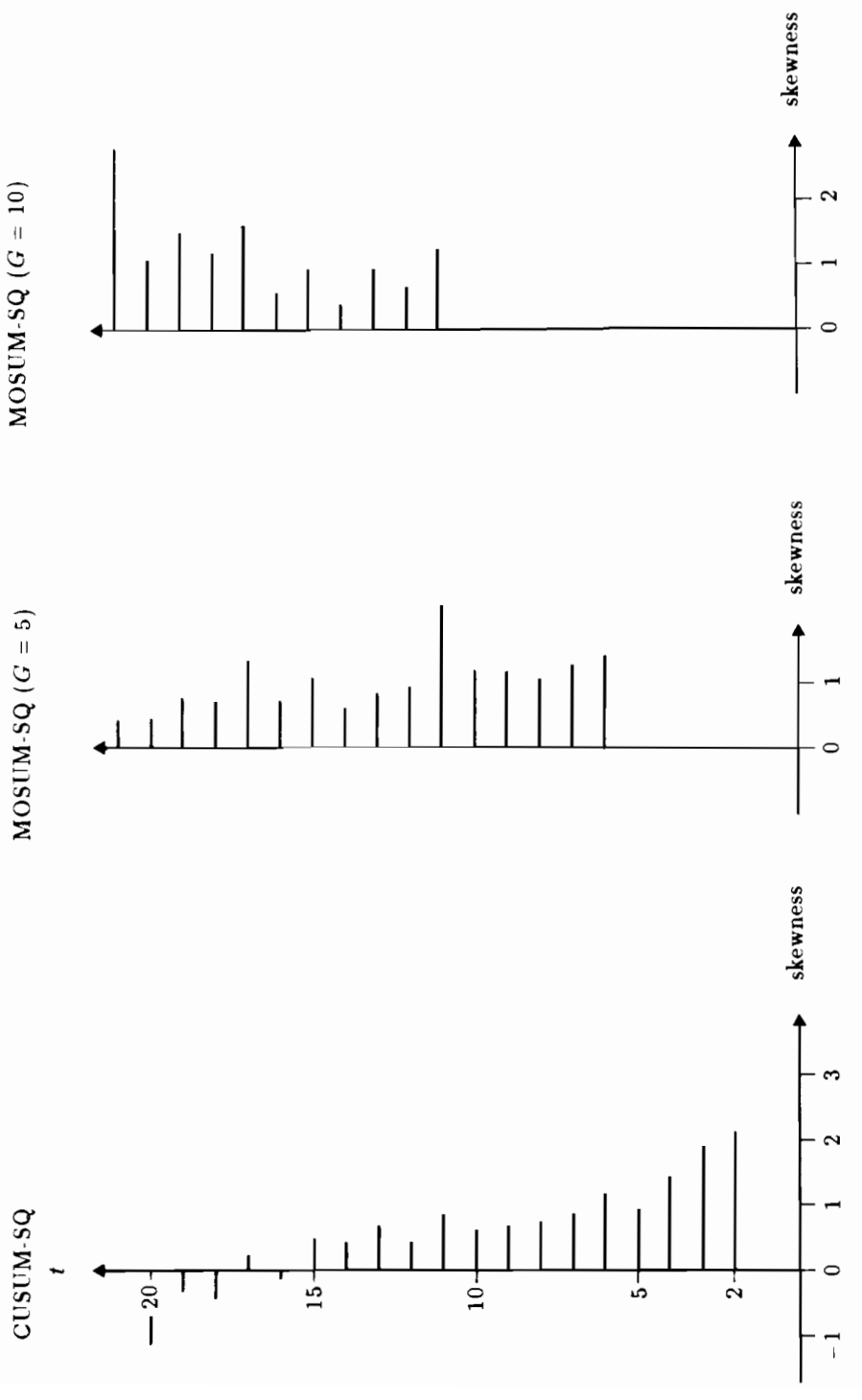

Figure 8.21: [M1], [PM2], $T = 21$.

of the test statistics for varying parameters are in general rather complicated. The main disadvantage of the Monte Carlo technique is the limited possibilities of generalizations beyond the framework of the study. To overcome part of that disadvantage, we have chosen to coordinate theoretical and numerical studies.

The distributions of the CUSUM-SQ and MOSUM-SQ test statistics are not known for varying parameters. Thus, we have focused on the expectations and variances of the test statistics. The exact formulae of these moments are implicit and rather complicated, while just approximations were studied here. The degree of approximation is indicated by the numerical study. The chief impression is that numerical and theoretical approaches provide very similar estimates of the expectations as well as the variances, particularly in the case of the CUSUM-SQ test statistic.

Two linear regression models were studied — the intercept model, M1, and the simple regression model, M2, with an intercept and one regressor. The regression coefficients in these models either change instantaneously, PM1, or gradually, PM2. In the case of M2, either one or both of the regression coefficients are varying.

The basic conclusions to be drawn from these theoretical and numerical analyses are as follows. When the structural variability is *a priori* known to have caused an instantaneous parameter change, all results seem to show that the time dating problem is reasonably solvable by CUSUM-SQ as well as MOSUM-SQ procedures. The time dating ability increases with the length of the time series and also, of course, with size of parameter change.

In the case where it cannot be *a priori* decided whether structural variability implies instantaneous or gradual parameter changes, it is rather difficult to identify the type of parameter model, in particular when applying the CUSUM-SQ test statistic. The problem of time dating seems to increase with the distance between the start, $t_1 + 1$, and the end, t_2, of the parameter change. It seems somewhat easier to identify various PM2 by the MOSUM-SQ test statistic, basically because the time when the MOSUM-SQ test statistic attains its maximum depends on t_2. As this time point also depends on the "window length" G, the choice of G affects the ability of the MOSUM-SQ test statistic to state the time when the parameter change ends.

In the case of model M2, it was found, as expected, to be almost impossible to distinguish between different parameter variability situations by the test statistics studied.

The characterization process indicated above is complicated by the fact that the distributions of the CUSUM-SQ as well the MOSUM-SQ test statistics are skew. The degree of skewness was only numerically estimated here. In general, the degrees of skewness are considerable, above all for the MOSUM-SQ statistic, which is always positively skew. Thus, other properties of the distributions may be of interest, such as the percentiles. (The skewness and the percentiles are the subject of a forthcoming paper.) Further studies should be carried out on the CUSUM-SQ and the MOSUM-SQ test statistics in cases of more complicated regression models and other parameter models.

References

Brown, R.L., Durbin, J., and Evans, J.M. (1975), Techniques for testing the constancy of regression relationships over time (with discussion). *Journal of the Royal Statistical Society*, **B-37**, 149-192.

Hackl, P. (1980), *Testing the Constancy of Regression Models over Time.* Göttingen: Vandenhoeck & Ruprecht.

Hackl, P. and Westlund, A.H. (1985), *Statistical Analysis of "Structural Change" — An Annotated Bibliography.* Collaborative Paper CP-85-31. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Poirier, D.J. (1976), *The Econometrics of Structural Change.* Amsterdam: North-Holland.

Törnkvist, B. (1986a), On the identification and characterization of parameter variability using the CUSUMSQ(R)-test statistic. Unpublished manuscript. Umeå University.

Törnkvist, B. (1986b), Karaktärisering av parametervariabilitet med hjälp av MOSUMSQ (R,G)-test. Unpublished manuscript. Umeå University.

Westlund, A.H. and Törnkvist, B. (1985), On the identification of time for parameter variabilities. *Metron*, **63**, 21-40.

Westlund, A.H. and Zackrisson, U. (1986), On the prediction of structurally varying systems. *Technological Forecasting and Social Change*, **30**, 63-72.

CHAPTER 9

The Local Power of the CUSUM-SQ Test against Heteroscedasticity

Werner Ploberger

Summary

This chapter considers the limiting behavior of the well-known CUSUM-SQ test for suitably defined sequences of local alternatives describing heteroscedasticity. It is shown that under certain circumstances the asymptotic behavior of the CUSUM-SQ test can be computed even for alternatives describing changes of the conditional variance of the error term (e.g., ARCH-processes).

9.1 Introduction

The CUSUM-SQ test for structural change in the parameters, in the form suggested by Brown *et al.* (1975), has been widely employed in empirical work. Although this procedure was developed as a descriptive diagnostic, it seems interesting to compute its power.

We examine this question from an asymptotic point of view, analogous to that of Ploberger and Krämer (1985). In that paper the asymptotic behavior of the CUSUM-SQ test was investigated for alternatives describing changes of the regression parameters, and it was shown that the CUSUM-SQ test has only trivial local power for these alternatives (in contrast to the nontrivial local power of the CUSUM test).

Here we investigate changes of the variance, and the main result is that the CUSUM-SQ test [or a "Studentized" procedure as described in Ploberger and Krämer (1986)] has nontrivial local power for alternatives describing heteroscedasticity.

9.2 The Model and Notation

Let us consider a triangular sequence of standard linear regression models

$$y_{t,T} = x_t'\beta_0 + u_{t,T} \ , \quad 1 \leq t \leq T \ , \quad T = 1, 2, \ldots \tag{9.1}$$

where at time t, $y_{t,T}$ is the observation on the dependent variable, x_t is a $K \times 1$ vector of observations on the independent variables, $u_{t,T}$ is the disturbance term, and β_0 is the K-dimensional vector of regression coefficients. The null hypothesis H_0 to be tested is that $Eu_{t,T}^2$ is constant $(= \sigma^2)$.

We will make the following assumptions:

A1 All random variables in (9.1) are defined on a common probability space $(\Omega, \mathcal{A}, P)$.

A2 The disturbances are martingale differences with

$$E(u_{t,T} \mid \mathcal{A}_{t,T}) = 0 \tag{9.2}$$

where $\mathcal{A}_{t,T}$ is the σ-field generated by

$$\{x_{t-s}, y_{t-s-1}, u_{t-s-1} \mid s > 0\}. \tag{9.3}$$

A3 The regressors x_t are (possibly) stochastic and satisfy

$$\lim_{T\to\infty} (1/T)\Sigma_{1\leq t\leq T} x_t x_t' = \lim_{T\to\infty} (1/T)\Sigma_{1\leq t\leq T} E(x_t x_t') = R \quad P - a.s. \tag{9.4}$$

for some nonsingular and constant matrix R. For simplicity, we also assume that $R(T) \equiv \Sigma_{1\leq t\leq T} x_t x_t'$ is nonsingular for $T = K$ and therefore for all $T > K$. Failing this, our proofs could easily be adapted as (9.4) guarantees that $R(T)$ is nonsingular for all but finitely many T. Note that this condition also allows for lagged dependent variables among the regressors, but excludes trended data.

A4 The regressors in addition satisfy $\limsup_{T\to\infty}(1/T)\Sigma_{1\leq t\leq T}\|x_t\|^4 < \infty$, where $\|.\|$ denotes the usual Euclidian norm.

Concerning the disturbances, we will assume that the following set of assumptions is fulfilled: (9.5) is a Lindeberg-type condition and (9.6) a norming condition.

A5 The $u_{t,T}$s are integrable up to the fourth order and

$$\lim_{M\to\infty} \lim_{T\to\infty} (1/T)\Sigma_{1\leq t\leq T} E(u_{t,T}^4 I\{|u_{t,T}| > M \mid \mathcal{A}_{t,T}\}) = 0 \tag{9.5}$$

Furthermore there exists a μ_4 so that for all z with $0 \leq z \leq 1$

$$\operatorname{plim}_{T\to\infty}(1/T)\Sigma_{1\le t\le Tz}u_{t,T}^4 = \mu_4 z \tag{9.6}$$

A sufficient and more familiar criterion for (9.5) would be that for some $\tau > 0$ the empirical moments of order $(4+\tau)$ of the $u_{t,T}$s remain bounded, i.e., $\limsup_{T\to\infty}(1/T)$ $\Sigma_{1\le t\le T}|u_{t,T}|^{(4+\tau)} < \infty$.

H_0 is fulfilled if $E(u_{t,T}^2 \mid \mathcal{A}_{t,T}) = \sigma^2$. Therefore, it seems reasonable to consider the following local alternatives. Let us define the processes $g_{t,T}$ by

$$E(u_{t,T}^2 \mid \mathcal{A}_{t,T}) = \sigma^2 + g_{t,T} \tag{9.7}$$

Then we will assume that the following conditions are satisfied:

A6 There exists a continuous function $G(z)$ so that for $0 \le z \le 1$

$$\operatorname{plim}_{T\to\infty}(1/\sqrt{T})\Sigma_{1\le t\le Tz}g_{t,T} = G(z) \tag{9.8}$$

uniformly in z.

A7 There exists a random variable M so that $\Sigma_{1\le s\le t}g_{s,T}^2 \le M$.

At a first glance, these conditions might look a bit abstract. In Section 9.4, however, some examples will be discussed.

In order to clarify our notation, we will suppress in the sequel the index T when it is clear from the context; e.g., we will write u_t instead of $u_{t,T}$.

Let us now briefly describe the CUSUM-SQ test procedure suggested by Brown *et al.* (1975). It is for a fixed T based on the test statistic

$$\hat{S}_T \equiv \max_{K\le t\le T}|\hat{S}_T(t) - (t-K)/(T-K)|$$

where $\hat{S}_T(t) = (\Sigma_{s\le t}\hat{u}_s^2)/(\Sigma_{s\le T}\hat{u}_s^2)$; here, $\hat{u}_s = (y_s - x_s'\beta^{(s-1)})/f_s$ is the sth recursive residual and $\hat{\beta}^{(s-1)} = R(s-1)^{-1}\Sigma_{t\le s}x_t y_t$ is the OLS estimate for β_0 from the first $s-1$ observations. The normalization factor f_s in the above equation, defined as $f_s = \sqrt{1 + x_s'R(s-1)^{-1}x_s}$ is introduced so that the recursive residuals have (provided H_0 is true and the regressors x_t are fixed) identical variance σ^2. It is easily seen that $\lim_{t\to\infty} f_t = 1$ $P-a.s.$, a fact that we will use in the sequel freely.

The CUSUM-SQ test rejects H_0 in the case where $\hat{S}_T$ becomes "too large". In Brown *et al.* (1975), the limiting distribution of $\sqrt{T}\hat{S}_T$ has been derived for normal u_t. For the case of nonnormal distributions, see Ploberger and Krämer (1986).

9.3 The Local Power of the CUSUM-SQ Test

Now let us consider a triangular sequence of models (9.1) that satisfy assumptions (A1)–(A7).

Theorem 9.1 *The statistic* $\sqrt{T}\,\hat{S}_T$ *has the limiting distribution*

$$\sup_{0\le z\le 1} |W(z) - zW(1) + G(z) - zG(1)|$$

where $W(z)$ *is a one-dimensional Wiener process with variance* $(\mu_4 - \sigma^4)/\sigma^4$ *[i.e., a Gaussian process with zero mean and correlation* $r(z_1, z_2) = \min(z_1, z_2)$ *].*

Proof: Let us first show that

$$\operatorname{plim}_{T\to\infty}(1/\sqrt{T})\sup_{t\le T}\sum_{s\le t}(\hat{u}_s^2 - u_s^2) = 0 \tag{9.9}$$

To this purpose write $\hat{u}_s$ as

$$\hat{u}_s = (1/f_s)[y_s - x_s'(\beta_0 - \beta_0 + \hat{\beta}^{(s-1)})] = (1/f_s)[u_s - x_s'R(s-1)^{-1}\Sigma_{1\le i\le s-1}x_iu_i]$$

Hence, we have to show that for $T \to \infty$

$$\operatorname{plim}_{T\to\infty}\sup_{t\le T}(1/\sqrt{T})\Sigma_{s\le t}u_s^2(1 - f_s^2) = 0 \tag{9.10}$$

$$\operatorname{plim}_{T\to\infty}\sup_{t\le T}(1/\sqrt{T})\Sigma_{s\le t}u_sd_s/f_s = 0 \tag{9.11}$$

$$\operatorname{plim}_{T\to\infty}\sup_{t\le T}(1/\sqrt{T})\Sigma_{s\le t}d_s^2/f_s = 0 \tag{9.12}$$

where $d_s = x_s'(\beta_0 - \hat{\beta}^{(s-1)})$.

Let us first show (9.10). As $(1 - f_s^2) \le x_s'R(s-1)^{-1}x_s$ and $\|R(s-1)^{-1}\| \le C/s$ for a random variable C, it is sufficient to show that $\operatorname{plim}_{T\to\infty}(1/\sqrt{T})\Sigma_{1\le t\le T}u_t^2\|x_t\|^2/t = 0$. Now observe that $(1/\sqrt{T})E(\Sigma_{1\le t\le T}u_t^2\|x_t\|^2/t) = (1/\sqrt{T})\Sigma_{1\le t\le T}\sigma^2E\|x_t\|^2/t + (1/\sqrt{T})$ $\Sigma_{1\le t\le T}g_t\|x_t\|^2/t$. Perfectly analogous to Ploberger and Krämer (1985), one can show that the first summand on the rhs of the above equation is $O(\ln T/\sqrt{T})$. For the second term, one can use the Cauchy-Schwartz inequality and the assumptions of our theorem.

Now let us show (9.12): With C as defined above, it is easily seen that

$$d_t^2 \le C^2\|x_t\|^2\|\Sigma_{1\le s\le t}u_sx_s\|^2/t^2 \tag{9.13}$$

Hence, (9.12) follows from *Theorem 9.2* in the Appendix A.

Finally, we have to prove (9.11): By *Lemma 9.1* (Appendix B), it is sufficient to show that

$$\operatorname*{plim}_{T\to\infty}(1/T)\Sigma_{1\le t\le T}(\sigma^2+g_t)d_t^2=0$$

Again, we can use the upper bound (9.13) for d_t^2 and then apply *Theorem 9.2.*

We have now shown (9.9). Let us now introduce $P_T(t)=\Sigma_{s\le t}u_s^2$ and $\hat{P}_T(t)=\Sigma_{s\le t}\hat{u}_s^2$. As $\sigma^2=\operatorname{plim}_{T\to\infty}(1/T)P(T)$ and, therefore, also $\sigma^2=\operatorname{plim}_{T\to\infty}(1/T)\hat{P}(T)$ it is easily seen that $\operatorname{plim}_{T\to\infty}\sqrt{T}\sup_{t\le T}|P(t)/P(T)-\hat{P}(t)/\hat{P}(T)|=0$. Therefore, the limiting distribution of $\sqrt{T}\sup_{t\le T}|P(t)/P(T)-(t-k)/(T-k)|$ has to be calculated.

Let us now define $e_{t,T}=u_{t,T}^2-g_{t,T}-\sigma^2$. Then it can easily be verified that $e_{t,T}$ is a triangle array, martingale difference scheme that satisfies the assumptions of the standard invariance principle for martingale differences on $D[0,1]$ (cf., e.g., Billingsley, 1968). We may therefore conclude that the random elements $E_T(z)=(1/\sqrt{T})\Sigma_{t\le Tz}e_{t,T}$ (defined for $0\le z\le 1$) converge in distribution (in $D[0,1]$) to a Wiener process $W(z)$ with variance $\mu_4-\sigma^4$. As $P_T(t)=E_T(t/T)+\Sigma_{s\le t}g_{s,T}+t\sigma^2$, we may conclude that the random elements Q_T, defined by $Q_T(z)=[P_T(Tz)-t\sigma^2]/\sqrt{T}$, converge in distribution in $D[0,1]$ to $W(z)+G(z)$ [$G(.)$ has been assumed to be continuous!].

It can easily be seen that $\sqrt{T}\sup_{t\le T}|(t-K)/(T-K)-t\sigma^2/(T\sigma^2)|\to 0$ and $P(t)/P(T)=\sqrt{T}\{Q(t/T)/(T\sigma^2[1+Q(1)/(\sqrt{T}\sigma^2)])\}+t\sigma^2/(T\sigma^2[1+Q(1)/(\sqrt{T}\sigma^2)])$. Furthermore, $\operatorname{plim}_{T\to\infty}\sqrt{T}\{1/[1+Q(1)/\sqrt{T}][1-Q(1)]\}=1$. Therefore, we may conclude that $\sup_{t\le T}\sqrt{T}|P(t)/P(T)-t/T-[Q(t/T)-Q(1)(t/T)]/\sigma^2|\to 0$ in probability, which finishes the proof of our theorem.

9.4 Some Examples

In this section we will discuss some examples of *Theorem 9.1.* Let us assume throughout this section that $u_{t,T}=e_t\sqrt{1+g_{t,T}}$, where e_t is independent of $\mathcal{A}_{t,T}$, stationary, and $E(e_t^4)<\infty$. We discuss three examples.

Example 1: Let us assume that $g_{t,T}=(1/\sqrt{T})h(t/T)$, where h is a (deterministic) function from $[0,1]$ to the real numbers which is the uniform limit of step functions (i.e., functions constant on intervals). It should be noted that all continuous functions satisfy this condition. Then, after some tedious calculations, one can see that the assumptions of *Theorem 9.1* depending on $u_{t,T}$ are fulfilled and $G(z)=\int_0^z h(w)dw$. Therefore, for nonconstant h, the test has local power against these alternatives.

Example 2: Let us assume that $g_{t,T}=(1/\sqrt{T})h(t/T)u_{t,T}^2$, where h satisfies the assumptions of *Example 1.* [I.e., we investigate alternatives describing ARCH-type processes (cf. Engle, 1982)] . In these cases also (in the case of Gaussian e_t when $\sigma^2=1$) the conditions of *Theorem 9.1* can easily be checked and $G(z)$ equals the expression on *Example 1.*

Therefore, we may conclude that the CUSUM-SQ test has local power even for alternatives describing changes in conditional heteroscedasticity (nonconstant h), but not against alternatives describing conditional heteroscedasticity (constant h).

Example 3: Now we will discuss a "negative" example. Let us assume that $g_{t,T} = (1/\sqrt{T})v_t$, where the v_t form an *iid* sequence of square integrable random variables completely independent of $\mathcal{A}_{t,T}$. Then, one can easily see that $G(z) = Ev_t$ is constant and that, therefore, the CUSUM-SQ test has zero power against these alternatives — which we might describe as "simple misspecification" — in contrast to *Example 1* or *2* where heteroscedasticity did arise from a time-dependent change of parameters of the error processes.

Appendix A

Theorem 9.2 *Let $A_{t,T}$, $0 \leq t \leq T$, be a triangular sequence of nonnegative $\mathcal{A}_{t,T}$-measurable random variables so that $A_{t,T} \geq A_{t-1,T}$, $A_{0,T} = 0$, and there exists a random variable M so that*

$$(1/t)A_{t,T} \leq M \tag{9.14}$$

Let $a_{t,T} = A_{t,T} - A_{t-1,T}$ $(A_{0,T} = 0)$. Then $\operatorname{plim}_{T\to\infty}(1/\sqrt{T})\Sigma_{1\leq t\leq T}a_t\|\Sigma_{s\leq t-1}x_s u_s\|^2/t^2 = 0$.

Proof: Let $B_{t,T} = \|\Sigma_{s\leq t-1}x_s u_s\|^2/t$, $B_{0,T} = 0$, $b_{t,T} = B_{t,T} - B_{t-1,T}$, and $C(t) = \Sigma_{1\leq s\leq t-1}x_s u_s$. We shall split the proof of the theorem into two propositions.

Proposition 9.1 $(1/\sqrt{T})\Sigma_{1\leq t\leq T}a_t B_t \leq O_p(1)|\Sigma_{1\leq t\leq T-1}u_t x_t' C(t)/[(t+1)\sqrt{T})]\ | + o_p(1)$.

Proposition 9.2 $\Sigma_{1\leq t\leq T-1}u_t x_t' C(t)/[(t+1)\sqrt{T}] = o_p(1)$.

Proof of Proposition 9.1: Observe the following relation:

$$\Sigma_{1\leq t\leq T}a_t B_t = A_T B_T + (-\Sigma_{1\leq t\leq T}A_{t-1}b_t)$$

Therefore, we have to show that both terms on the rhs of the above equation are $o_p(\sqrt{T})$. As a_t, $B_t \geq 0$, it is sufficient to give upper bounds for these terms.

For the first one, this can easily be seen. Rewrite $A_T B_T$ as

$$(A_T/T)(\|\Sigma_{1\leq s\leq T}x_s u_s\|^2)/T$$

Then the first factor remains bounded by (9.14). Immediately from our assumptions, it can be seen that the expectations of the second factor are uniformly (in T) bounded.

Now let us investigate the second summand. For $t \geq 2$:

$$\begin{aligned} b_t &= \|C(t)\|^2/t^2 - \|C(t-1)\|^2/(t-1)^2 \\ &= [\|C(t)\|^2 - \|C(t-1)\|^2]/(t-1)^2 + \|C(t)\|^2[1/t^2 - 1/(t-1)^2] \\ &\geq 2(x_{t-1}u_{t-1})'C(t-1)/(t-1)^2 + \|C_t\|^2[1/t^2 - 1/(t-1)^2] \end{aligned} \tag{9.15}$$

By (9.14) and (9.15)

$$\begin{aligned} A_{t-1}b_t \leq u_{t-1}(-2M)(1/t-1)x'_{t-1}C(t-1) \\ + M(t-1)[-1/t^2 + 1/(t-1)^2]\|C_t\|^2 \end{aligned}$$

As $b_1 = 0$ and for some constant K, $(t-1)[-1/t^2 + 1/(t-1)^2] \leq K(1/t^2)$, it remains to show that $\Sigma_{2\leq t\leq T}\|C_t\|^2/t^2$ is $o_p(\sqrt{T})$.

This can be seen by computing the expectations. One can easily conclude that $E\|C_t\|^2/t$ is uniformly bounded. Therefore, $E\{\Sigma_{2\leq t\leq T}[\|C_t\|^2/t](1/t)\}$ remains $O(\ lnT)$ and, consequently, these terms are $o_p(\sqrt{T})$.

Proof of proposition 9.2: For showing that $\Sigma_{1\leq t\leq T-1}u_t x'_t C(t)/[(t+1)\sqrt{T}\,)]$ converges to zero in probability it is, by Lemma 9.1, sufficient to show that $\text{plim}_{T\to\infty}\,\Sigma_{t\leq T}(\sigma^2 + g_{t,T})(x'_t C(t))^2/[(t+1)^2 T] = 0$. This can be shown after some tedious calculations, partly analogous to those by Ploberger and Krämer (1986).

Appendix B

Lemma 9.1 *Let $A_{t,T}$ be a sequence of random variables adapted to $\mathcal{A}_{t,T}$ so that* $\text{plim}_{T\to\infty}$ $\Sigma_{t\leq T}(\sigma^2 + g_{t,T})A^2_{t,T} = 0$. *Then* $\text{plim}_{T\to\infty}\sup_{t\leq T}\Sigma_{s\leq t}u_{s,T}A_{s,T} = 0$.

Proof: The proof is perfectly analogous to the proof of Lemma 1 in Ploberger and Krämer (1985).

References

Billingsley, P. (1968), *Convergence of Probability Measures.* New York: John Wiley.

Brown, R.L., Durbin, J., and Evans, J.M. (1975), Techniques for testing the constancy of regression relationships over time (with discussion). *Journal of the Royal Statistical Society*, **B-37**, 149-163.

Engle, R.F. (1982), Autoregressive conditional heteroscedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica*, **50**, 987-1007.

Ploberger, W. and Krämer, W. (1985), The local power of the CUSUM and CUSUM of squares test. Unpublished manuscript.

Ploberger, W. and Krämer, W. (1986), On studentizing a test for structural change. *Economics Letters*, **20** (4), 341-344.

CHAPTER 10

Bahadur Efficiency of Tests for a Shift in Location of Normal Populations

Jaap Praagman

Summary

A sequence of independent random variables $X_1, \ldots, X_N$ is said to have a change point n, if $X_1, \ldots, X_n$ have a common distribution F and $X_{n+1}, \ldots, X_N$ have a common distribution G, $G \neq F$. Consider the problem of testing the null hypothesis of no change, against the alternative of a one-sided change at an unknown change point n, when both F and G are normal with equal variance σ^2. Most of the test statistics for this problem can be interpreted as generalizations of two-sample statistics (n known). In this chapter we derive the Bahadur efficiencies for two classes of statistics that are generalizations of the two-sample likelihood ratio statistics. The asymptotic results are compared with some small-sample power estimates based on Monte Carlo experiments.

10.1 Introduction

Let $X_1, \ldots, X_N$ be a sequence of independent random variables. Then this sequence is said to have a change point at n, $1 \leq n < N$, if $X_1, \ldots, X_n$ have a common distribution F and $X_{n+1}, \ldots, X_N$ have a common distribution G; $G \neq F$. We consider the problem of testing the null hypothesis of no change, against the alternative of a one-sided location change at an unknown change point n, when F is normal.

Hence,

$$\mathcal{H}_0: \quad X_i \sim \mathcal{N}(\mu; \sigma^2) \quad i = 1, \ldots, N$$

$$\mathcal{H}_a: \qquad X_i \sim \begin{cases} \mathcal{N}(\mu;\sigma^2) & i = 1,\ldots,n \\ \mathcal{N}(\mu+\delta\sigma;\sigma^2) & i = n+1,\ldots,N \end{cases} \qquad (10.1)$$

$$\text{for some } n\,, \quad 1 \le n < N \text{ and } \delta > 0$$

Both μ and δ are unknown, while the case with σ^2 known as well as with σ^2 unknown will be studied.

Various test statistics have been proposed for (10.1), and thus mutual comparisons are of interest. For most of these statistics, however, no managable expressions exist for the distribution under $\mathcal{H}_0$ or under the alternative hypothesis. As a result, simple power comparisons cannot be made, unless we resort to Monte Carlo experiments. (In fact, the power comparisons reported in the literature are all partly based on such experiments.) Therefore, we will rely on asymptotics and derive the Bahadur efficiencies for two classes of test statistics.

These classes of statistics are based on the close relationship between the change point problem (c.p.p.) and the well-known two-sample problem. Both problems have the same null hypothesis; and the alternative for the c.p.p. can be conceived as the union of the alternatives for the $N-1$ two-sample problems with $n = 1,2,\ldots,N-1$, respectively. Hence, when $T_{N,k}$ is a two-sample statistic (samples $X_1,\ldots,X_k$ and $X_{k+1},\ldots,X_N$), there are two obvious ways to define a statistic for the c.p.p. : as a weighted sum or a weighted maximum of the $T_{N,k}, k = 1,\ldots,N-1$.

Thus, we consider a sum-type statistic

$$S_N = \sum_{k=1}^{N-1} c_{N,k}\, T_{N,k} \qquad (10.2)$$

and a max-type statistic

$$M_N = \max_{1\le k<N} c_{N,k}\, T_{N,k} \qquad (10.3)$$

where the $c_{N,k}$ are nonnegative weight coefficients. Note that, from a Bayesian point of view, $c_{N,k}/\sum_{k=1}^{N-1} c_{N,k}$ can be interpreted as the prior probability that X_{k+1} is the initial shifted variable.

Almost every statistic proposed for (10.1) can be written in one of the forms (10.2) and (10.3) or as a minor modification. Chernoff and Zacks (1964) used a Bayesian approach to derive a sum-type statistic in case of known variance σ^2: $\sum_{i=1}^{N}(i-1)(X_i - \overline{X}_N)$, where $\overline{X}_N = N^{-1}\sum_{i=1}^{N} X_i$. Since this statistic is a linear function of normal random variables, it is easy to obtain the critical value for a size α test and the power function. Chernoff and Zacks give these in their paper with some numerical illustrations. Furthermore, they derive similar results for the case when μ is known.

Gardner (1969) used the same approach to tackle the two-sided problem.

When σ^2 is unknown, a corresponding statistic is obtained if the Chernoff-Zacks statistic is divided by a suitable variance estimator. Sen and Srivastava (1975a) used both $\sum_{i=1}^{N}(X_i - \overline{X}_N)^2/(N-1)$ and $\sum_{i=1}^{N-1}(X_{i+1} - X_i)^2/(2N-2)$ to estimate the variance and compared the powers of the two resulting statistics by Monte Carlo methods, showing only small differences for the sample sizes involved ($N \leq 50$). For the first statistic, the $\mathcal{H}_0$-distribution was found.

Max-type statistics — more particulary, the likelihood ratio statistic for testing $\mathcal{H}_0$ against $\mathcal{H}_a$ — also were investigated by Sen and Srivastava (1975a). In this paper the power of the likelihood ratio test when σ^2 is unknown was estimated, too; while in a second paper (1975b) power comparisons, again based on Monte Carlo experiments, of the likelihood ratio test when σ^2 is known, with the Chernoff-Zacks statistic have been reported. These comparisons indicate that the Chernoff-Zacks statistic is generally more powerful when n is close to $N/2$. On the other hand, when n is close to 1 or to N, the likelihood ratio test is more powerful.

Hawkins (1977) considered the likelihood ratio test statistic for the two-sided case, $\delta \neq 0$, and found its distribution under the null hypothesis for the case of known σ^2. The null distribution for the unknown σ^2 has been given by Worsley (1979), who pointed out that the earlier result of Hawkins for this case is incorrect.

As far as we know, no results on asymptotic efficiency have been reported, apart from the Pitman efficiency of one of the studentized forms of the Chernoff-Zacks statistic (Bhattacharyya and Johnson, 1986), and an investigation of the likelihood ratio test using Chernoff's approach by Deshayes and Picard (1982).

In this chapter, we will derive the Bahadur slopes for the classes of sum- and max-type statistics, taking for $T_{N,k}$ the two-sample likelihood ratio statistics. After some preliminaries in Section 10.2, we consider in Section 10.3, the case when σ^2 is known and, in Section 10.4, when σ^2 is unknown. Some examples are presented in the fifth section, and comparisons with power estimates for small and moderate sample sizes in the last section. For related results concerning c.p.p. statistics based on two sample rank tests, see Praagman (1986) .

10.2 Preliminaries

The derivation of the Bahadur slopes of the test statistics in this chapter follow the strategy suggested by the following theorem by Bahadur (1971, th. 7.2). For a detailed description of the concepts of Bahadur slope and efficiency, see Bahadur (1971) or Groeneboom and Oosterhoff (1977).

Theorem 10.1 *(Bahadur) Let $\{T_N\}$ be a sequence of statistics for testing $\theta \in \Theta_0$ against $\theta \in \Theta_a$.*

Suppose that

$$\lim_{N\to\infty} N^{-1}T_N = b(\theta) \quad a.s.\ P_\theta, \quad \theta \in \Theta_a$$

where $-\infty < b(\theta) < \infty$*; and that for each* t *in an open* I*,*

$$\lim_{N\to\infty} -N^{-1}\log[\sup\{P(T_N \geq Nt); \theta \in \Theta_0\}] = a(t)$$

where a *is a nonnegative function on* I*, continuous at* $t = b(\theta)$*; then, the Bahadur slope of* $\{T_N\}$ *at* θ *equals* $2a(b(\theta))$*.*

Thus, to find the Bahadur slope for a particular alternative, the almost sure limit under that alternative and a large deviation result under $\mathcal{H}_0$ are needed. The *Theorems 10.2–10.4* all bear on these aspects.

Theorem 10.2 *Let* $X_1, X_2, \ldots$ *be iid random variables such that* $E[X_1] = 0$*. Then, for* $1 \leq \alpha \leq 2$*,* $E[|X_1|^\alpha] < \infty$ *if and only if,*

$$\lim_{N\to\infty} N^{-1/\alpha} \sum_{k=1}^{N} a_{N,k} X_k = 0, \ \ a.s.$$

for each array $a_{N,k}$ *of real numbers such that* $\limsup_{N\to\infty} \sum_{k=1}^{N} a_{N,k}^2 < \infty$*.*

Proof: See Chow and Lai (1973, th.9). □

Our first large deviation result is a theorem by Killeen *et al.* (1972). Let $\{T_N\}$ be a sequence of random variables and $\{\varepsilon_N\}$, $\{t_N\}$ and $\{u_N\}$ nonnegative sequences of real numbers with $N^{-1}\log \varepsilon_N = o(1)$ and $N^{-1}\log u_N = o(1)$ as $N \to \infty$.

Theorem 10.3 *Suppose, for each* $N = 1, 2, \ldots, T_N$ *has an absolutely continuous distribution with density* $f_N(t)$*. If there exists a natural number* N_1 *such that for* $N \geq N_1$*,* $f_N(t)$ *is nonincreasing for* $t \in [t_N, \infty)$ *and if*

$$N^{-1}\log[f_N(t_N + \varepsilon_N)/f_N(t_N)] = o(1) \quad (N \to \infty)$$

and

$$\limsup_{N\to\infty} N^{-1}\log[P(T_N \geq u_N)/f_N(t_N)] \leq 0$$

Then,

$$N^{-1}\log f_N(t_N) - N^{-1}\log P(T_N \geq t_N) = o(1) \quad (N \to \infty)$$

Proof: See Killeen *et al.* (1972, th. 2.1.1). □

For large deviations of max-type statistics, the following generalization of theorem 3.1 in Killeen and Hettmansperger (1972) will turn out to be very useful.

Theorem 10.4 *For each N, let $T_{N,1},\dots,T_{N,m_N}$ be identically distributed random variables; $c_{N,1},\dots,c_{N,m_N}$ nonnegative real numbers; and $M_N = \max_{1\le k\le m_N} c_{N,k}T_{N,k}$.*

Suppose an interval $(t^-,t^+)\subset R$ exists, such that $\lim_{N\to\infty} N^{-1}\log P(T_{N,1}\ge t_N) = h(t)$, for each sequence $\{t_N\}$ with $t_N\to t\in(t^-,t^+)$ and there exists an integer ℓ such that, for sufficiently large N, $1\le m_N\le N^\ell$. Then,

$$\lim_{N\to\infty} N^{-1}\log P(M_N\ge t_N) = h(tc)$$

for all $\{t_N\}$ such that $\lim_{N\to\infty}\min_{1\le k\le m_N} t_N c_{N,k}^{-1} = t_c$ and $t_c\in(t^-,t^+)$.

Proof: Since $\{c_{N,k}T_{N,k}\ge t_N\}\subset\{M_N\ge t_N\} = \bigcup_{k=1}^{m_N}\{c_{N,k}T_{N,k}\ge t_N\}$ it follows that

$$\max_{1\le k\le m_N} P(c_{N,k}T_{N,k}\ge t_N)\le P(M_N\ge t_N)\le\sum_{k=1}^{m_N} P(c_{N,k}T_{N,k}\ge t_N)$$

Hence,

$$P(T_{N,1}\ge\min_k t_N c_{N,k}^{-1})\le P(M_N\ge t_N)\le N^\ell P(T_{N,1}\ge\min_k t_N c_{N,k}^{-1})$$

and thus, since $\min_k t_N c_{N,k}^{-1}\to t_c\in(t^-,t^+)$, therefore

$$\lim_{N\to\infty} N^{-1}\log P(M_N\ge t_N) = \lim_{N\to\infty} N^{-1}\log P(T_{N,1}\ge\min_k t_N c_{N,k}^{-1}) = h(t_c)\quad\square$$

Theorem 10.5 *Let f and g be measurable and square integrable functions on $[0,1]$, with $\int f(x)\,dx = \int g(x)\,dx$ and $\int f^2(x)\,dx = \int g^2(x)\,dx$. Suppose f is nonincreasing and $\int_0^y f(x)\,dx\le\int_0^y g(x)\,dx$, for all $y\in[0,1]$. Then, $\int[f(x)-g(x)]^2\,dx = 0$.*

Proof: (Unless otherwise stated, integration is over $[0,1]$).

$$\int(f-g)^2\,dx = \int[g^2-f^2-2(g-f)f]\,dx = -2\int(g-f)f\,dx \tag{10.4}$$

Define $h:[0,1]\to R$ by $h(y) = \int_0^y(g-f)\,dx$. Then, h is nonnegative and $h(0) = h(1) = 0$.

Thus, by substitution of h and partial integration

$$\int(g-f)f\,dx = \int f\,dh = -\int h\,df\ge 0$$

because f is nonincreasing and h is nonnegative. Together with (10.4), this completes the proof. $\square$

10.3 Variance Known

When σ^2 is known, the two-sample likelihood ratio statistic for samples $X_1, \ldots, X_k$ and $X_{k+1}, \ldots, X_N$ is

$$T_{N,k} = \sqrt{\frac{k(N-k)}{N}} \frac{\bar{X}'_{N-k} - \bar{X}_k}{\sigma} \tag{10.5}$$

with $\bar{X}_k = \frac{1}{k}\sum_{i=1}^{k} X_i$ and $\bar{X}'_{N-k} = \frac{1}{N-k}\sum_{i=k+1}^{N} X_i$. Without losing generality, put σ^2 equal to one.

10.3.1 Sum-type statistics

With $T_{N,k}$ as in (10.5), the sum-type statistics defined by $\sum_{k=1}^{N-1} c_{N,k} T_{N,k}$ can be written as

$$S_N^{(1)} = \sum_{i=1}^{N} d_{N,i} X_i \tag{10.6}$$

where $d_{N,i} = \sum_{k=1}^{i-1} [k(N-k)^{-1} N^{-1}]^{\frac{1}{2}} c_{N,k} - \sum_{k=i}^{N-1} [k^{-1}(N-k)N^{-1}]^{\frac{1}{2}} c_{N,k}$, $i = 1, \ldots, N$. Note that

$$\sum_{i=1}^{N} d_{N,i} = 0 \tag{10.7}$$

and that, from $c_{N,k} \geq 0$, $k = 1, \ldots, N-1$, it follows that

$$d_{N,1} \leq d_{N,2} \leq \ldots \leq d_{N,N} \tag{10.8}$$

Furthermore, we assume, without loss of generality, that

$$N^{-1} \sum_{i=1}^{N} d_{N,i}^2 = 1 \tag{10.9}$$

Just like the $c_{N,k}$, the $d_{N,i}$ will be called weight coefficients.

Now, we come to the formulation of theorems on large deviations (*Theorem 10.6*) and the almost sure limit under an alternative (δ, λ) (*Theorem 10.7*).

Theorem 10.6 *Let $X_i \sim \mathcal{N}(\mu, 1)$, $i = 1, \ldots, N$, all X_i independent, $\mu \in R$; and let $S_N^{(1)}$ be defined by (10.6) with $d_{N,i}$ satisfying (10.7) and (10.9). Then,*

$$\lim_{N\to\infty} -N^{-1}\log[\sup\{P(S_N^{(1)} \geq Nt); \mu \in R\}] = \frac{1}{2}t^2 \tag{10.10}$$

Proof: For all $\mu \in R$, $S_N^{(1)} \sim \mathcal{N}(0, N)$. Thus $P[S_N^{(1)} \geq Nt] = P(Z \geq tN^{\frac{1}{2}})$, with $Z \sim \mathcal{N}(0,1)$. Thus, the use of $P(Z > z) = (z\sqrt{2\pi})^{-1}\exp(-\frac{1}{2}z^2)[1 + o(1)]$, $(z \to \infty)$ (see, e.g., Feller, 1957, p.166), yields, for all $\mu \in R$,

$$P[S_N^{(1)} \geq Nt] = t^{-1}(2\pi N)^{-\frac{1}{2}}\exp(-\frac{1}{2}t^2N)[1 + o(1)] \quad (N \to \infty)$$

and (10.10) follows. □

Note that only the standardization of the $d_{N,i}$ according to (10.7) and (10.9) is assumed here and that no conditions whatsoever are imposed on the relation between the $d_{N,i}$ for different values of N. Such a condition is needed for the next theorem on the almost sure limit under a particular alternative. For the testing problem (10.1), for each N a simple alternative is characterized by a pair $(\delta, n/N)$ and for $N \to \infty$ we will characterize a fixed alternative by (δ, λ) with $\lambda \in (0,1)$ such that $n/N \to \lambda$.

Theorem 10.7 *Let $X_i \sim \mathcal{N}(\mu, 1)$, $i = 1, \ldots, n$; $X_i \sim \mathcal{N}(\mu + \delta, 1)$, $i = n+1, \ldots, N$; all X_i independent, $\delta > 0$, $\mu \in R$ and $n/N \to \lambda \in (0,1)$.*

Take $S_N^{(1)}$ as in Theorem 10.6 *and suppose*

$$\lim_{N\to\infty} N^{-1}\sum_{i=n+1}^{N} d_{N,i} = \zeta_\lambda < \infty \tag{10.11}$$

Then

$$\lim_{N\to\infty} N^{-1}S_N^{(1)} = \delta\zeta_\lambda \ a.s.$$

Proof: Introduce $Y_i = X_i - E[X_i]$ and $a_{N,i} = N^{-\frac{1}{2}}d_{N,i}$, $i = 1, \ldots, N$. Then $Y_1, ..., Y_N$ i.i.d., $E[Y_i] = 0$, $E[Y_i^2] < \infty$, and, according to (10.9), $\sum_{i=1}^{N} a_{N,i}^2 = 1$. Thus, by *Theorem 10.2*,

$$\lim_{N\to\infty} N^{-\frac{1}{2}}\sum_{i=1}^{N} a_{N,i}Y_i = 0$$

i.e., $\lim_{N\to\infty}(N^{-1}\sum_{i=1}^{N} d_{N,i}X_i - N^{-1}\sum_{i=n+1}^{N} d_{N,i}\delta) = 0$, a.s. And, since $N^{-1}\sum_{i=n+1}^{N} d_{N,i} \to \zeta_\lambda$, the theorem follows. □

Remark 10.1 *In the proof, no use is made of the normality of the X_i. Indeed, the theorem holds for all X_i, with a finite second moment.* □

Together with *Theorem 10.1*, the foregoing theorems immediately lead to the Bahadur slope of $\{S_N^{(1)}\}$:

Theorem 10.8 *Let $X_i \sim \mathcal{N}(\mu, \sigma^2), i = 1, \ldots, n; X_i \sim \mathcal{N}(\mu + \delta\sigma, \sigma^2), i = n+1, \ldots, N$; all X_i independent, $\delta > 0$, $\mu \in R$ and $n/N \to \lambda \in (0,1)$.*

Then, with $S_N^{(1)}$ and ζ_λ as in Theorem 10.7, *the Bahadur slope of $\{S_N^{(1)}\}$ at the alternative (δ, λ) is given by $\delta^2\zeta_\lambda^2$.*

Remark 10.2 *The mutual Bahadur efficiency at (δ, λ) of two sequences of $S_N^{(1)}$-type test depends, therefore, on the location of the change, λ, and the weights, through ζ_λ, but not on the magnitude of the change, δ.*

Example 10.1 *The Chernoff-Zacks statistic $S_N^{(cz)} = \sum_{i=1}^N (i-1)(X_i - \bar{X}_N) = \frac{1}{2}\sum_{i=1}^N (2i - N - 1)X_i$. Scaling the weights to meet (10.9) — (10.7) is met already — we get $\zeta_\lambda^2 = 3\lambda^2(1-\lambda)^2$. Thus, the Bahadur slope of $S_N^{(cz)}$ at the alternative (δ, λ) equals $3\delta^2\lambda^2(1-\lambda)^2$.*

Example 10.2 *The two-sample likelihood ratio statistic $T_{N,k}$, for alternatives (δ, λ_1) with a particular λ_1, can be seen as a special degenerate case of $S_N^{(1)}$,*

$$T_{N,k} = \sqrt{\frac{k(N-k)}{N}}(\bar{X}'_{N-k} - \bar{X}_k) = -\sqrt{\frac{N-k}{kN}}\sum_{i=1}^{k} X_i + \sqrt{\frac{k}{(N-k)N}}\sum_{i=k+1}^{N} X_i$$

with $k/N \to \lambda_1$.

It turns out that, writing $a \wedge b$ for $\min(a, b)$,

$$\zeta_\lambda = [\lambda_1(1-\lambda_1)]^{\frac{1}{2}}\left(\frac{1-\lambda}{1-\lambda_1} \wedge \frac{\lambda}{\lambda_1}\right)$$

and, thus, the Bahadur slope at the alternative (δ, λ_1) is $\delta^2\lambda_1(1-\lambda_1)$. By a straightforward calculation, we see that Raghavachari's (1970) upper bound for the Bahadur slope equals $\delta^2\lambda(1-\lambda)$, for the problem under consideration in this section. So the slope of the $T_{N,k}$ test does attain this upper bound for $\lambda = \lambda_1$, i.e., $\{T_{N,k}\}$ is Bahadur-optimal at $\lambda = \lambda_1$.

Theorems 10.6 and *10.7* (and thus *10.8*) do not use the monotonicity of the $d_{N,i}$ as stated in equation (10.8). The statistics considered in this section, however, do have this monotonicity as an immediate implication of the obvious restriction to nonnegative $c_{N,k}$ coefficients. The following theorem shows that when the $d_{N,i}$ also meet (10.8), the class of statistics $S_N^{(1)}$ contains no statistic that is Bahadur-optimal for all alternatives (δ, λ).

For the convergence of the weights, we introduce a condition that differs a little bit from (10.11). Define the weight functions $\psi_N : (0,1] \to R$ by

$$\psi_N(v) := d_{N,i} \quad \text{for} \quad v \in \left(\frac{i-1}{N}, \frac{i}{N}\right]$$

and suppose a measurable and square integrable function $\psi : [0,1] \to R$ exists, such that $\int \psi^2(v)\,dv = 1$ and,

$$\int |\, \psi_N(v) - \psi(v) \,|^2 \; dv \to 0 \tag{10.12}$$

Note that $\zeta_\lambda = \int_\lambda^1 \psi(v)\,dv$.

Theorem 10.9 *Within the class of statistics $S_N^{(1)}$ as defined by (10.6) with weights $d_{N,i}$ satisfying (10.7), (10.8), (10.9) and (10.12), no pair of statistics exists in which one is uniformly better, in the sense of Bahadur, than the other.*

Proof: Consider two sequences of $S_N^{(1)}$ statistics with limit weight functions ψ_1 and ψ_2, respectively. First, note that due to the conditions on the $d_{N,i}$, $\int \psi_1^2 = \int \psi_2^2 = 1, \int \psi_1 = \int \psi_2 = 0$, and both ψ_1 and ψ_2 are nondecreasing on $[0,1]$. And since $\int_\lambda^1 \psi_1(v)\,dv \geq \int_\lambda^1 \psi_2(v)\,dv$ for all $\lambda \in (0,1)$, when the first sequence of statistics is uniformly better than the second, the result directly follows from *Theorem 10.4*. □

Now, let us focus on one particular λ_1 and look for those statistics that have maximal slope for the alterntives (δ, λ) with $\lambda = \lambda_1$. Thus, we want to find the statistics $S_N^{(1)}$ such that $N^{-1} \sum_{i=k+1}^N d_{N,i}$ is maximal under the restrictions that $\sum_{i=1}^N d_{N,i} = 0$ and $\sum_{i=1}^N d_{N,i}^2 = N$, where $k/N \to \lambda_1$. Using the generalized Lagrange multiplier theorem, we find that this maximum is attained for $d_{N,i} = -(N-k)^{\frac{1}{2}} k^{-\frac{1}{2}}$ if $i \leq k$, and $d_{N,i} = (N-k)^{-\frac{1}{2}} k^{\frac{1}{2}}$ if $i > k$. Indeed, this is the optimal two-sample likelihood ratio test (see Example 10.2), for samples $X_1, \ldots, X_k$ and $X_{k+1}, \ldots, X_N$.

Therefore, we have the next corollary.

Corollary: Apart from the two-sample statistics, the class of statistics $S_N^{(1)}$ defined by (10.6), with weights $d_{N,i}$ satisfying (10.7) and (10.9), does not contain a statistic that is optimal in the sense of Bahadur for any alternative (δ, λ). □

10.3.2 Max-type statistics

In this section, we will study the statistics

$$M_N^{(1)} = \max_{1 \leq k < N} c_{N,k} \sqrt{\frac{k(N-k)}{N}} \, (\bar{X}'_{N-k} - \bar{X}_k) \tag{10.13}$$

with $c_{N,k} \geq 0$. If $c_{N,k} = 1, k = 1, \ldots, N$, this becomes the likelihood ratio statistic for the c.p.p.

Theorem 10.10 *Let $X_i \sim \mathcal{N}(\mu,1)$, $i = 1,\ldots,N$, all X_i independent, $\mu \in R$; and let $M_N^{(1)}$ be defined by (10.13).*

Suppose

$$\lim_{N\to\infty} \max_{1\le k<N} c_{N,k} = \gamma_{\max} < \infty \tag{10.14}$$

Then,

$$\lim_{N\to\infty} -N^{-1}\log(\sup\{P[M_N^{(1)} \ge N^{\frac{1}{2}}t];\mu \in R\}) = \frac{t^2}{2\gamma_{\max}^2}$$

Proof: For each N, and $1 \le k < N, T_{N,k} \sim \mathcal{N}(0,1)$. Hence, using the same relationship as in *Theorem 10.6*, we get, for each $\mu \in R$,

$$\lim_{N\to\infty} -N^{-1}\log P(T_{N,k} \ge N^{\frac{1}{2}}t) = \frac{1}{2}t^2$$

Thus, application of *Theorem 10.4*, together with (10.14), results in

$$\lim_{N\to\infty} -N^{-1}\log P[M_N^{(1)} \ge N^{\frac{1}{2}}t] = \frac{t^2}{2\gamma_{\max}^2}$$

and, since this holds for each $\mu \in R$, the proof is complete. □

We proceed to the almost sure limit under an alternative (δ,λ). Let $D = D[0,1]$ be the space of functions on $[0,1]$ that are right-continuous and have left-hand limits. Endow D with the Skorohod-topology (see, e.g., Billingsley, 1968, Ch. 3). Introduce the function $\gamma_N \in D$ by (take $c_{N,0} = c_{N,N} = 0$)

$$\gamma_N(u) := c_{N,[Nu]} \quad 0 \le u \le 1 \tag{10.15}$$

And define, for each $\lambda \in (0,1)$,

$$b_1(u;\lambda) := [u(1-u)]^{\frac{1}{2}}\left(\frac{1-\lambda}{1-u} \wedge \frac{\lambda}{u}\right) \tag{10.16}$$

Theorem 10.11 *Let $X_i \sim \mathcal{N}(\mu,1), i = 1,\ldots,n$; $X_i \sim \mathcal{N}(\mu+\delta,1), i = n+1,\ldots,N$; all X_i independent, $\delta > 0$, $\mu \in R$, $n/N \to \lambda \in (0,1)$; and let $M_N^{(1)}$ be defined by (10.13).*

Suppose there exists a function $\gamma : [0,1] \to R^+$, such that $\gamma \in D$ and $\gamma_N \to \gamma$ in the Skorohod topology. Then,

$$\lim_{N\to\infty} N^{-\frac{1}{2}} M_N^{(1)} = \delta \max_{0\le u\le 1} \gamma(u)\, b_1(u;\lambda) \quad a.s.$$

Proof: Introduce for $u \in [0,1]$, (take $T_{N,0} = T_{N,N} = 0$), $T_N(u) := N^{-\frac{1}{2}} T_{N,[Nu]}$, then $N^{-\frac{1}{2}} c_{N,k} T_{N,k} = \gamma_N(k/N) T_N(k/N)$. Using the strong law of large numbers for independent random variables, it is easy to show that for each $u \in [0,1]$, $T_N(u) \to \delta b_1(u;\lambda)$, a.s. However, to justify the interchange of max and lim, in addition to the given Skorohod convergence of the γ_N functions, the uniform convergence of $T_N(u)$ for $u \in [0,1]$ is needed. That is to say, we need to show that for almost every realization $x_1, x_2, \ldots$ of $X_1, X_2, \ldots$, the corresponding functions $t_N(u)$ converge uniformly to $\delta b_1(u;\lambda)$.

To that end, first, we will show that for each $u \in [0,1]$ and for each $\varepsilon > 0$, a neighborhood B_u of u exists, such that for almost every realization

$$\exists_{N_u \in N} \forall_{N > N_u} \forall_{v \in B_u} [|t_N(u) - t_N(v)| < \varepsilon] \tag{10.17}$$

For notational convenience, introduce $z_N(u) = (N - [Nu])^{\frac{1}{2}} [Nu]^{-\frac{1}{2}}$. Fix $\varepsilon > 0$ and $u \in (0,1)$; suppose $w > u$ (for $w < u$ the proof is similar), then for all $v \in [u,w]$

$$\begin{aligned} T_N(u) - T_N(v) = {} & [z_N(u) - z_N(v)] N^{-1} \sum_{i=1}^{[Nu]} (-X_i) \\ & + [z_N^{-1}(u) + z_N(v)] N^{-1} \sum_{i=[Nu]+1}^{[Nv]} X_i + [z_N^{-1}(u) + z_N^{-1}(v)] N^{-1} \sum_{i=[Nv]+1}^{N} X_i \end{aligned}$$

Successively, we will consider X_1 to $X_{[Nu]}$, $X_{[Nw]+1}$ to X_N, and $X_{[Nu]+1}$ to $X_{[Nw]}$.

By the strong law of large numbers,

$$\begin{aligned} & [z_N(u) - z_N(w)] N^{-1} \sum_{i=1}^{[Nu]} (-X_i) \\ & \qquad \to \left(-\sqrt{\frac{1-u}{u}} + \sqrt{\frac{1-w}{w}} \right) [u\mu + (u - \lambda) 1_{[\lambda,1]}(u) \delta] \text{ a.s.} \end{aligned}$$

So, for $w - u$ small enough — say, $w - u < \triangle_1$ — this limit will be smaller than $\varepsilon/6$ and, hence, for almost every realization $x_1, x_2, \ldots$

$$| [z_N(u) - z_N(w)] N^{-1} \sum_{i=1}^{[Nu]} (-x_i) | < \varepsilon/3$$

for N sufficiently large. But then, the same inequality, with w replaced by v, also holds true for every $v \in (u,w)$.

In a similar way, it follows that for $w - u$ smaller than some $\triangle_2$, for almost every realization and N sufficiently large,

$$| [z_N^{-1}(u) - z_N^{-1}(v)]N^{-1} \sum_{i=[Nw]+1}^{N} x_i | < \varepsilon/3$$

for all $v \in (u, w)$.

For the third and final part, the Cauchy-Schwarz inequality will be used. Define $Y_i := X_i - E[X_i]$, then $Y_i \sim \mathcal{N}(0,1)$, and, hence, $N^{-1} \sum_{[Nu]+1}^{[Nw]} Y_i^2 \to (w-u)$ a.s. Again, this limit is small for $w-u$ small, and thus for almost every realization $x_1, x_2, \ldots$,

$$N^{-1} \sum_{[Nu]+1}^{[Nw]} Y_i^2 < \varepsilon/6 \tag{10.18}$$

for N sufficiently large.

On the other hand, for all $v \in [u, w]$,

$$\begin{aligned} &\sum_{[Nu]+1}^{[Nv]} N^{-1}[z_N^{-1}(u) + z_N(v)]^2 + \sum_{[Nv]+1}^{[Nw]} N^{-1}[z_N^{-1}(u) - z_N^{-1}(v)]^2 \\ &\quad \leq \ ([Nv] - [Nu])N^{-1}[z_N^{-1}(u) + z_N(u)]^2 + ([Nw] - [Nv])N^{-1}[z_N^{-1}(u) - z_N^{-1}(w)]^2 \end{aligned}$$

Letting $N \to \infty$, the right-hand side tends to

$$(v-u)\left[\left(\frac{u}{1-u}\right)^{\frac{1}{2}} + \left(\frac{1-u}{u}\right)^{\frac{1}{2}}\right]^2 + (w-v)\left[\left(\frac{u}{1-u}\right)^{\frac{1}{2}} - \left(\frac{w}{1-w}\right)^{\frac{1}{2}}\right]^2$$

which can be made arbitrarily small by taking $w-u$ small enough. Hence, for N large enough, the right-hand side of the inequality will be smaller than $\varepsilon/6$, for every $v \in (u, w)$. And thus, together with (10.18) and Cauchy-Schwarz, we get that for $w-u$ smaller than $\triangle_3$, for almost every realization and for every $v \in [u, w]$

$$| [z_N^{-1}(u) + z_N(v)]N^{-1} \sum_{[Nu]+1}^{[Nv]} y_i + [z_N^{-1}(u) - z_N^{-1}(v)]N^{-1} \sum_{[Nv]+1}^{[Nw]} y_i | \leq \varepsilon/6$$

if N is sufficiently large.

It easily follows that the same inequality holds with the y_i replaced by x_i and the right-hand side by $\varepsilon/3$; which completes the proof of (10.17) for $u \in (0,1)$. Using the same kind of argument, (10.17) can be proved for $u = 0$ or 1.

Furthermore, the continuity of $b_1(u;\lambda)$ implies that for each $u \in [0,1]$, a neighborhood exists such that $|\delta b_1(u;\lambda) - \delta b_1(v;\lambda)| < \varepsilon$, for each v in that neighborhood. Let B'_u be the intersection of B_u with that neighborhood.

In addition, $T_N(u)$ converges a.s. to $\delta b_1(u;\lambda)$; thus, for almost every realization, there is an N'_u such that for $N > N'_u$, $|t_N(u) - \delta b_1(u;\lambda)| < \varepsilon$.

Consequently, since

$$|t_N(v) - \delta b_1(v;\lambda)| \leq |t_N(v) - t_N(u)| + |t_N(u) - \delta b_1(u;\lambda)| + |\delta b_1(u;\lambda) - \delta b_1(v;\lambda)|$$

we get that, for almost every realization $x_1, x_2, \ldots$

$$\forall_{u\in[0,1]}\exists_{N^*_u\in I\!N}\exists_{B'_u}\forall_{N\geq N^*_u}\forall_{v\in B'_u}[|t_N(v) - \delta b_1(v;\lambda)| < 3\varepsilon]$$

Due to the compactness of $[0,1]$ the open cover, $\bigcup_{u\in[0,1]} B'_u$, has a finite subcover $\{B_{u(i)}\}$, $i = 1,\ldots,\ell$. Thus, for $N > \max_{1\leq i\leq \ell} N^*_{u(i)}$, we have that for all $v \in [0,1]$ $|t_N(v) - \delta b_1(\delta;\lambda)| < 3\varepsilon$, i.e., the a.s. convergence of $T_N(u)$ is uniform. □

Remark 10.3 *For the max-type statistics in this and in the next section, Skorohod convergence is needed for the weight functions γ_N; L_2- (or L_1-) convergence, as used for the sum-type statistics, is not sufficient here, since for the max-type statistics, asymptotically one particular $T_{N,k}$ dominates. (See* Theorems 10.10 *and* 10.11.*)* □

Remark 10.4 *The weight functions γ_N and γ are bounded (as are all the elements of D). Hence, without loss of generality, it may be assumed that $\max_{0\leq u\leq 1} \gamma_N(u) = 1$, for all N, and $\max_{0\leq u\leq 1} \gamma(u) = 1$.* □

Theorem 10.12 *Let $X_i \sim \mathcal{N}(\mu,1), i = 1,\ldots,n; X_i \sim \mathcal{N}(\mu+\delta,1), i = n+1,\ldots,N$; all X_i independent, $\delta > 0$, $\mu \in R$, $n/N \to \lambda \in (0,1)$. Then, with $M_N^{(1)}$ and γ as in* Theorem 10.11, *and $\max_{0\leq u\leq 1} \gamma(u) = 1$, the Bahadur slope of $\{M_N^{(1)}\}$ at the alternative (δ,λ) equals*

$$\delta^2 \left[\max_{0\leq u\leq 1} \gamma(u) b_1(u;\lambda)\right]^2$$

Proof: Directly from *Theorems 10.1, 10.10* and *10.11.* □

Remark 10.5 *As with the $S_N^{(1)}$ statistics — see Remark 10.2 — the mutual efficiency of two sequences of $M_N^{(1)}$-type tests does not depend on the magnitude of the change δ. And the same holds true for the efficiency of an $S_N^{(1)}$ with respect to an $M_N^{(1)}$ test.*

Example 10.3 *The likelihood ratio statistic for the c.p.p. with σ^2 known is the special case of $M_N^{(1)}$, when $c_{N,k} = 1$, for all k and all N. Thus, $\gamma(u) = 1$, for $u \in [0,1]$ and from* Theorem 10.12, *the Bahadur slope of the likelihood ratio test at the alternative (δ, λ) is $\delta^2\lambda(1-\lambda)$. This slope equals Raghavachari's upper bound for all alternatives, which complies with the Bahadur optimality of the likelihood ratio test.*

Remark 10.6 *This example shows, that the class of statistics $M_N^{(1)}$ contains at least one statistic that is optimal for all alternatives (δ, λ). Moreover, from* Theorem 10.12, *it is clear that a sequence of statistics $\{M_N^{(1)}\}$ with weight functions γ_N is optimal for an alternative (δ, λ), if and only if γ_N converges to γ and $\gamma(\lambda) = \max_{0 \le u \le 1} \gamma(u)$.*

Thus, we may conclude that:

1. For every sequence of statistics $\{M_N^{(1)}\}$, at least one $\lambda' \in (0,1)$ exists such that $\{M_N^{(1)}\}$ is Bahadur-optimal at (δ, λ') for all $\delta > 0$.
2. Ignoring statistics with other γ_N, but converging to the same limit function γ, the likelihood ratio statistic treated in Example 10.3 is the only one that is optimal for all alternatives with $\lambda \in (0,1)$.

10.4 Variance Unknown

When σ^2 is unknown, the two-sample likelihood ratio statistic for samples $X_1, \ldots, X_k$ and $X_{k+1}, \ldots, X_N$ is

$$T_{N,k} = \left[\frac{k(N-k)}{N}\right]^{\frac{1}{2}} \frac{\bar{X}'_{N-k} - \bar{X}_k}{\left[\frac{(k-1)s_k^2 + (N-k-1)s'^2_{N-k}}{N-2}\right]^{\frac{1}{2}}} \tag{10.19}$$

where $s_k^2 = \frac{1}{k-1}\sum_{i=1}^{k}(X_i - \bar{X}_k)^2$ and $s'^2_{N-k} = \frac{1}{N-k-1}\sum_{i=k+1}^{N}(X_i - \bar{X}'_{N-k})^2$. In addition, we will consider a modification of (10.19),

$$T'_{N,k} = \left[\frac{k(N-k)}{N}\right]^{\frac{1}{2}} \frac{(\bar{X}' - \bar{X}_k)}{s_N} \tag{10.20}$$

where $s_N^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X}_N)^2$. Sum-type statistics, using this modified likelihood ratio, were suggested by Sen and Srivastava (1975a).

Remark 10.7 *Since $(N-1)s_N^2 = (k-1)s_k^2 + (N-k-1)s'^2_{N-k} + \frac{k(N-k)}{N}(\bar{X}_k - \bar{X}'_{N-k})^2$, or $SS_{total} = SS_{within} + SS_{between}$, the statistics (10.19) and (10.20) have a functional relationship*

$$T_{N,k} = \left(\frac{N-2}{N-1}\right)^{\frac{1}{2}} \left(1 - \frac{(T'_{N,k})^2}{N-1}\right)^{-\frac{1}{2}} T'_{N,k}$$

And thus, the monotonicity of $x(1-x^2)^{-\frac{1}{2}}$ *on* $[0,1]$ *along with*

$$(T'_{N,k})^2 = (N-1)\frac{SS_{between}}{SS_{within}} \leq N-1$$

implies that tests for the two-sample problem, based on $T_{N,k}$ *and* $T'_{N,k}$ *are, in fact, the same.* □

10.4.1 Sum-type statistics

The sum-type statistics based on (10.19) are rather intractable. This is the main reason for the introduction of the statistics $T'_{N,k}$. Only sum-type statistics based on the latter will be treated here.

Consider $S_N^{(2)} = \sum_{k=1}^{N-1} c_{N,k} T'_{N,k}$, which, like (10.6) can be written as

$$S_N^{(2)} = \frac{\sum_{i=1}^{N} d_{N,i} X_i}{[\sum_{i=1}^{N}(X_i - \bar{X})^2]^{\frac{1}{2}}} \tag{10.21}$$

with

$$\sum_{i=1}^{N} d_{N,i} = 0 \tag{10.22}$$

and

$$d_{N,1} \leq d_{N,2} \leq \ldots \leq d_{N,N} \tag{10.23}$$

Again, assume without loss of generality that

$$N^{-1} \sum_{i=1}^{N} d_{N,i}^2 = 1 \tag{10.24}$$

The standardization of $S_N^{(2)}$ differs from that of $S_N^{(1)}$, Section 10.3.1, because the denominator of (10.21) equals the sum of squares and is not a variance estimator. A large deviation result for $S_N^{(2)}$ is formulated in *Theorem 10.14*. The proof of this theorem refers to *Theorem 10.3*. Therefore, the distribution of $S_N^{(2)}$ has to be known; this is stated in *Theorem 10.13*.

Theorem 10.13 *Let $X_i \sim \mathcal{N}(\mu, \sigma^2)$, $i = 1, \ldots, N$, all X_i independent, $\mu \in R$, $\sigma^2 > 0$, and let $S_N^{(2)}$ be defined by (10.21). Suppose the $d_{N,i}$ satisfy (10.22) and (10.24). Then, the density function f_{SN} of $N^{-\frac{1}{2}} S_N^{(2)}$ is*

$$f_{SN}(s) = \frac{1}{B(\frac{1}{2}, \frac{1}{2}N - 1)}(1 - s^2)^{(N-4)/2}, \quad |s| \leq 1 \tag{10.25}$$

Proof: Sen and Srivastava (1975a) proved that $[N^{-\frac{1}{2}} S_N^{(2)}]^2$ has a Beta-distribution with parameters $\frac{1}{2}$ and $\frac{1}{2}N - 1$. Together with the symmetry of the distribution of $S_N^{(2)}$, this leads directly to (10.25). □

Remark 10.8 *Note that the distribution of $N^{-\frac{1}{2}} S_N^{(2)}$ is independent of the weights $d_{N,i}$, as long as the conditions (10.22) and (10.24) are met.* □

We are ready now to prove the necessary large deviation theorem.

Theorem 10.14 *Let $X_i \sim \mathcal{N}(\mu, \sigma^2), i = 1, \ldots, N$; all X_i independent, $\mu \in R$, $\sigma^2 > 0$; and let $S_N^{(2)}$ be defined by (10.21). Suppose the $d_{N,i}$ satisfy (10.22) and (10.24). Then, for $0 < t < 1$,*

$$\lim_{N \to \infty} -N^{-1} \log(\sup\{P[S_N^{(2)} \geq N^{\frac{1}{2}} t]; \mu \in R\}) = -\frac{1}{2}\log(1 - t^2) \tag{10.26}$$

Proof: Introduce for all N, $t_N := u_N := t$ and $\varepsilon_N := N^{-1}$. Then, $N^{-1} \log \varepsilon_N = o(1)$ and $N^{-1} \log u_N = o(1)$. From (10.25), it is evident that for $N \geq 5$, f_{SN} is decreasing on $[0, 1]$.

Furthermore,

$$N^{-1} \log \left[\frac{f_{SN}(t + N^{-1})}{f_{SN}(t)}\right] = \frac{N - 4}{2N} \log \left[1 - \frac{2Nt + 1}{N^2(1 - t^2)}\right] = o(1) \quad (N \to \infty)$$

Finally,

$$P(N^{-\frac{1}{2}} S_N^{(2)} \geq t) = \int_t^1 f_{SN}(s)\, ds < f_{SN}(t)$$

Hence,

$$\limsup_{N \to \infty} N^{-1} \log \left[\frac{P(N^{-\frac{1}{2}} S_N^{(2)} \geq t)}{f_{SN}(t)}\right] \leq 0$$

So, all the conditions for *Theorem 10.3* are fulfilled. Thus,

$$\lim_{N\to\infty} -N^{-1}\log P[S_N^{(2)} > N^{\frac{1}{2}}t] = \lim_{N\to\infty} -N^{-1}\log f_{SN}(t)$$
$$= \lim_{N\to\infty}[N^{-1}\log B(\frac{1}{2},\frac{1}{2}N-1) - \frac{N-4}{2N}\log(1-t^2)]$$

Now, by definition, $B(\frac{1}{2},\frac{1}{2}N-1) = \Gamma(\frac{1}{2})\Gamma(\frac{1}{2}N-1)/\Gamma(\frac{1}{2}N-\frac{1}{2})$, and $\Gamma(\frac{1}{2}N-1) < \Gamma(\frac{1}{2}N-\frac{1}{2}) < \Gamma(\frac{1}{2}N) = (\frac{1}{2}N-1)\Gamma(\frac{1}{2}N-1)$. So for all $N > 2$,

$$(\frac{1}{2}N-1)^{-1} = \frac{\Gamma(\frac{1}{2}N-1)}{\Gamma(\frac{1}{2}N)} < \frac{\Gamma(\frac{1}{2}N-1)}{\Gamma(\frac{1}{2}N-\frac{1}{2})} < 1$$

and thus,

$$\lim_{N\to\infty} N^{-1}\log\frac{\Gamma(\frac{1}{2}N-1)}{\Gamma(\frac{1}{2}N-\frac{1}{2})} = 0$$

Consequently,

$$\lim_{N\to\infty} -N^{-1}\log P[S_N^{(2)} \geq N^{\frac{1}{2}}t] = -\frac{1}{2}\log(1-t^2)$$

which holds for all $\mu \in R$ and, thus, the proof is complete. □

Our next concern is the almost sure limit of $S_N^{(2)}$ under an alternative (δ,λ).

Theorem 10.15 *Let* $X_i \sim \mathcal{N}(\mu,\sigma^2), i = 1,\ldots,n; X_i \sim \mathcal{N}(\mu+\delta\sigma,\sigma^2), i = n+1,\ldots,N;$ *all* X_i *independent,* $\mu \in R, \sigma > 0, \delta > 0,$ *and* $n/N \to \lambda \in (0,1)$. *Take* $S_N^{(2)}$ *as in* Theorem 10.14 *and suppose*

$$\lim_{N\to\infty} N^{-1}\sum_{i=n+1}^{N} d_{N,i} = \zeta_\lambda < \infty$$

Then,

$$\lim_{N\to\infty} N^{-\frac{1}{2}}S_N^{(2)} = \frac{\delta\zeta_\lambda}{[1+\lambda(1-\lambda)\delta^2]^{\frac{1}{2}}} \qquad a.s.$$

Proof: It is straightforward to show that

$$N^{-1}\sum_{i=1}^{N}(X_i-\bar{X}_N)^2 \to \sigma^2[1+\lambda(1-\lambda)\delta^2] \qquad \text{a.s}$$

and then *Theorem 10.15* is an immediate consequence of *Theorem 10.4*. □

The next theorem combines the results of *Theorems 10.14* and *10.15* according to *Theorem 10.1*.

Theorem 10.16 *Let* $X_i, S_N^{(2)}$ *and* ζ_λ *be as in* Theorem 10.14. *Then, the Bahadur slope of* $S_N^{(2)}$ *at the alternative* (δ, λ) *equals*

$$-\log\left[1 - \frac{\delta^2\zeta_\lambda^2}{1+\lambda(1-\lambda)\delta^2}\right] \qquad \square$$

Just as in the known variance case, the monotonicity of the $d_{N,i}$ has not been used in the foregoing theorems. Results similar to those stated in *Theorem 10.9* and its corollary can easily be shown to hold here, by application of the next corollary.

Corollary: Let $S_N^{(1)}$ be defined by (10.6), $S_N^{(2)}$ by (10.21) with the same $d_{N,i}$ coefficients, and suppose $N^{-1}\sum_{i=n+1}^{N} d_{N,i} \to \zeta_\lambda$. Then

$$\log[1+\delta^2\lambda(1-\lambda)] - BS[S_N^{(2)};\delta,\lambda] = \log[1+\delta^2\lambda(1-\lambda) - BS[S_N^{(1)};\delta,\lambda)]$$

where $BS[S_N^{(i)};\delta,\lambda]$ denotes the Bahadur slope of $\{S_N^{(i)}\}$, $i = 1,2$ at (δ,λ).

Proof: Directly from *Theorems 10.8* and *10.15*. □

10.4.2 Max-type statistics

In this section, we deal with max-type statistics based on (10.19) and (10.20). Contrary to the sum-type statistics, max-type statistics based on (10.19) can be handled. For the sake of completeness, those based on (10.20) will be taken into consideration, too.

Thus, with $T_{N,k}$ and $T'_{N,k}$ according to (10.19) and (10.20),

$$M_N^{(2)} = \max_{1\le k<N} c_{N,k}T_{N,k} \tag{10.27}$$

and

$$\tilde{M}_N^{(2)} = \max_{1\le k<N} c_{N,k}T'_{N,k} \tag{10.28}$$

Again, with $c_{N,k} = 1$, $M_N^{(2)}$ equals the likelihood ratio statistic for our problem. And, due to the equivalence of $T_{N,k}$ and $T'_{N,k}$ (see Remark 10.7), if $c_{N,k} = 1$, $M_N^{(2)}$ and $\tilde{M}_N^{(2)}$ are equivalent too (Worsley, 1979).

Once again, we start with a large deviation theorem.

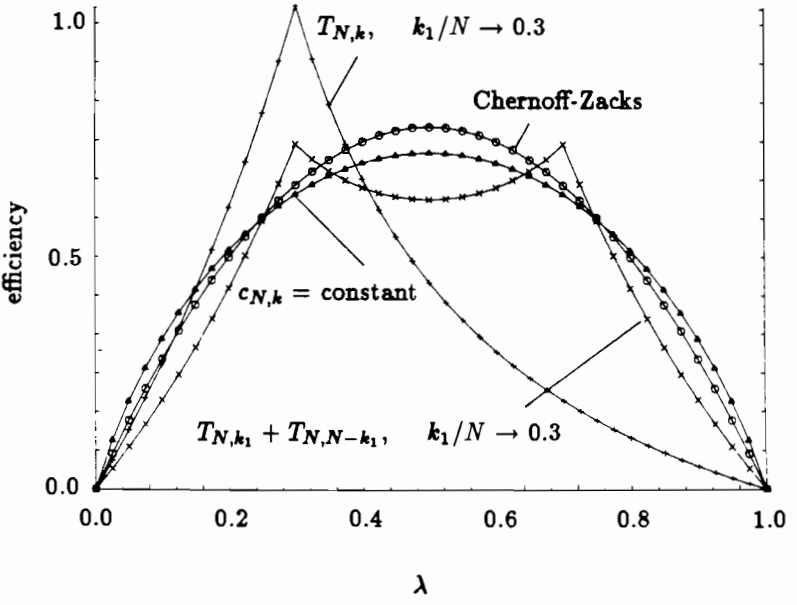

Figure 10.1: Bahadur efficiency of $S_N^{(1)}$ statistics.

Theorem 10.17 *Let* $X_i \sim \mathcal{N}(\mu, \sigma^2), i = 1, \ldots, N$, *all* X_i *independent,* $\mu \in R$, $\sigma^2 > 0$. *Let* $M_N^{(2)}$ *and* $\tilde{M}_N^{(2)}$ *be defined by (10.27) and (10.28). Suppose* $\lim_{N\to\infty} \max_{1\le k<N} c_{N,k} = \gamma_{\max} < \infty$. *Then,*

$$\lim_{N\to\infty} -N^{-1} \log(\sup\{P[M_N^{(2)} \ge N^{\frac{1}{2}}t]; \mu \in R\}) = \frac{1}{2}\log\left(1 + \frac{t^2}{\gamma_{\max}^2}\right) \tag{10.29}$$

and

$$\lim_{N\to\infty} -N^{-1} \log(\sup\{P[\tilde{M}_N^{(2)} \ge N^{\frac{1}{2}}t]; \mu \in R\}) = -\frac{1}{2}\log\left(1 - \frac{t^2}{\gamma_{\max}^2}\right) \tag{10.30}$$

Proof: For all k, $1 \le k < N$, $T_{N,k}$ has a t-distribution with $N-2$ degrees of freedom; thus,

$$\lim_{N\to\infty} -N^{-1} \log P(T_{N,k} \ge N^{\frac{1}{2}}t) = \frac{1}{2}\log(1 + t^2)$$

See Bahadur (1971, example 5.1), which, along with *Theorem 10.4*, leads to (10.29).

Table 10.1: Bahadur slopes of $S_N^{(1)}$ statistics.

Statistic (weights)	*Slope at* (δ, λ)
Chernoff-Zacks [a]	$3\delta^2\lambda^2(1-\lambda)^2$
$c_{N,k} = c,\ k = 1, \ldots, N-1$	$4\delta^2(\pi^2-\delta)^{-1}[-\sqrt{\lambda(1-\lambda)} + (1-\lambda)\arccos\sqrt{1-\lambda} + \lambda \arccos\sqrt{\lambda}]$
$c_{N,k} = \begin{cases} 1, & k = k_1, k_1/N \to \lambda_1 \\ 0, & k \neq k_1 \end{cases}$	$\delta^2(1-\lambda_1)\lambda_1 \left(\frac{\lambda}{\lambda_1} \wedge \frac{1-\lambda}{1-\lambda_1}\right)^2$
$c_{N,k} = \begin{cases} 1, & k = k_1 \text{ or } N-k_1; \\ 0, & \text{elsewhere; } k_1/N \to \lambda_1 < \frac{1}{2} \end{cases}$	$\frac{1}{2}\delta^2\lambda_1 \left(\frac{\lambda}{\lambda_1} \wedge 1 \wedge \frac{1-\lambda}{\lambda_1}\right)^2$

[a](1964, Example 2.3.1).

To prove (10.30), first note that $(N-1)^{-\frac{1}{2}}T'_{N,k}$ is a special case of $N^{-\frac{1}{2}}S_N^{(2)}$ (compare with Example 10.2). One consequence (Remark 10.8) being that the distribution of $T'_{N,k}$ is independent of k. Thus, *Theorem 10.4* applies again.

Secondly, by *Theorem 10.14*,

$$\lim_{N\to\infty} -N^{-1}\log P[(N-1)^{-\frac{1}{2}}T'_{N,k} \geq t] = -\frac{1}{2}\log(1-t^2)$$

and applying *Theorem 10.4* completes the proof. □

Recall the definition of the γ_N functions, $\gamma_N(u) := c_{N,[Nu]}$, with $c_{N,0} = c_{N,N} = 0$, and introduce, analogous to $b_1(u;\lambda)$,

$$b_2(u;\delta,\lambda) = \frac{\delta[u(1-u)]^{\frac{1}{2}}\left(\frac{1-\lambda}{1-u} \wedge \frac{\lambda}{u}\right)}{\left[1+\delta^2 \mid u-\lambda \mid \left(\frac{1-\lambda}{1-u} \wedge \frac{\lambda}{u}\right)\right]^{\frac{1}{2}}} \tag{10.31}$$

Theorem 10.18 *Let* $X_i \sim \mathcal{N}(\mu,\sigma^2), i = 1,\ldots,n;\ X_i \sim \mathcal{N}(\mu+\delta\sigma,\sigma^2), i = n+1,\ldots,N;$ *all* X_i *independent,* $\delta > 0, \sigma^2 > 0$, $\mu \in R$, *and* $n/N \to \lambda \in [0,1]$. *Let* $M_N^{(2)}$ *and* $\tilde{M}_N^{(2)}$ *be defined by (10.27) and (10.28). Suppose a function* γ: $[0,1] \to R^+$ *exists such that* $\gamma \in D$ *and* $\gamma_N \to \gamma$ *in the Skorohod topology.*

Then,

$$\lim_{N\to\infty} N^{-\frac{1}{2}}M_N^{(2)} = \max_{0\leq u\leq 1} \gamma(u)b_2(u;\delta,\lambda) \qquad a.s.$$

and

$$\lim_{N\to\infty} N^{-\frac{1}{2}}\tilde{M}_N^{(2)} = \max_{0\leq u\leq 1} \frac{\delta\gamma(u)b_1(u;\lambda)}{[1+\lambda(1-\lambda)\delta^2]^{\frac{1}{2}}} \qquad a.s.$$

Proof: This proof follows the same line of argument as used to prove *Theorem 10.11* and, hence, will be omitted. □

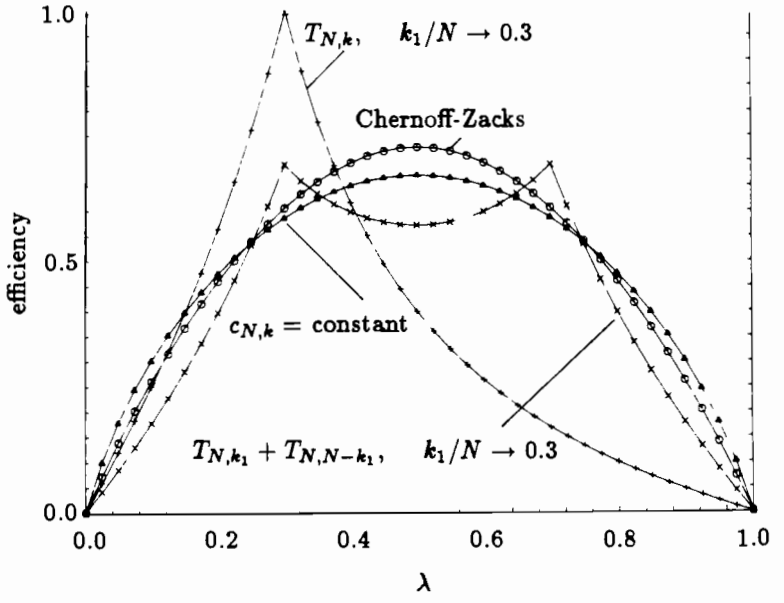

Figure 10.2: Bahadur efficiency of $S_N^{(2)}$ statistics, $\delta^2 = 1$; same weights as in *Figure 10.1.*

Again (see Remark 10.4), the weight functions are bounded; therefore, assume that $\max_{0\le u\le 1}\gamma(u) = 1$.

Theorem 10.19 *Let $X_i \sim \mathcal{N}(\mu,\sigma^2)$, $i = 1,\ldots,n$; $X_i \sim \mathcal{N}(\mu+\delta\sigma,\sigma^2)$, $i = n+1,\ldots,N$; all X_i independent, $\delta > 0$, $\sigma^2 > 0$, $\mu \in R$, and $n/N \to \lambda \in (0,1)$. Then, the Bahadur slope of $\{M_N^{(2)}\}$ at the alternative (δ,λ) equals*

$$\log\left\{1 + \left[\max_{0\le u\le 1}\gamma(u)b_2(u;\delta,\lambda)\right]^2\right\}$$

and, that of $\{\tilde{M}_N^{(2)}\}$ equals

$$-\log\left\{1 - \frac{\delta^2[\max_{0\le u\le 1}\gamma(u)b_1(u;\lambda)]^2}{1+\lambda(1-\lambda)\delta^2}\right\}$$

Proof: Directly from *Theorems 10.1, 10.17* and *10.18.* □

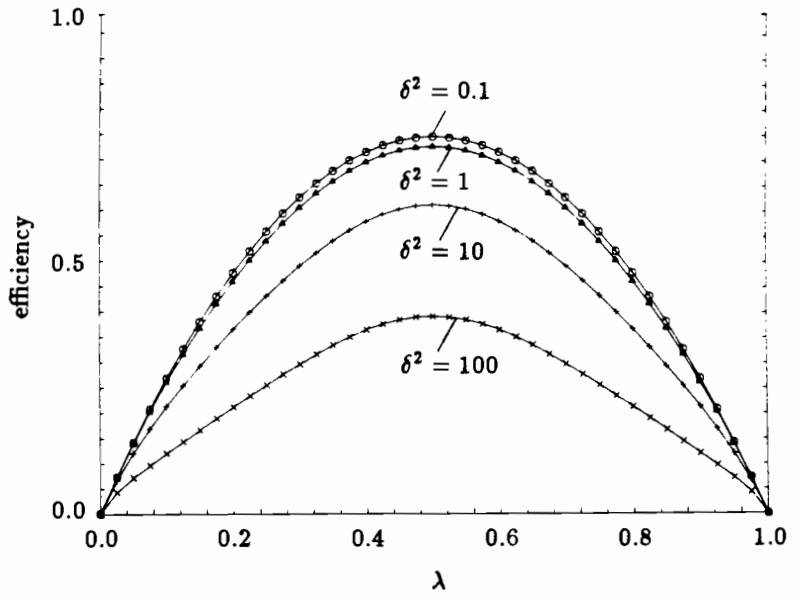

Figure 10.3: Bahadur efficiency of $S_N^{(2)}$ statistic with Chernoff-Zacks weights.

Example 10.4 *As said previously, $M_N^{(2)}$ with $c_{N,k} = 1$ for all k is the likelihood ratio statistic for the c.p.p. According to Theorem 10.19, we can find the Bahadur slope at (δ, λ):*

$$\log\left\{1 + \left[\max_{0\le u\le 1} b_2(u;\delta,\lambda\right]^2\right\} = \log\left[1 + b_2^2(\lambda;\delta,\lambda)\right] = \log[1 + \delta^2\lambda(1-\lambda)]$$

Again, as in Example 10.3, this slope equals Raghavachari's upperbound, showing the optimality of the likelihood ratio test. Furthermore, note that since $M_N^{(2)}$ and $\tilde{M}_N^{(2)}$ are equivalent when $c_{N,k} = 1$ for all k, the latter is Bahadur-optimal, too. □

Remark 10.9 *Again, these two statistics are the only optimal ones for all (δ, λ) within the classes $M_N^{(2)}$ and $\tilde{M}_N^{(2)}$; and, just as in the known variance case, $\{M_N^{(2)}\}$ (or $\{\tilde{M}_N^{(2)}\}$) is optimal for some alternative (δ, λ), if and only if $\gamma(\lambda) = \max_{0\le u\le 1}\gamma(u)$.* □

From *Theorems 10.12* and *10.19*, it is clear that the relation between the Bahadur slopes of $S_N^{(1)}$ and $S_N^{(2)}$, as stated in the corollary to *Theorem 10.16*, holds for the statistics $M_N^{(1)}$ and $\tilde{M}_N^{(2)}$, too.

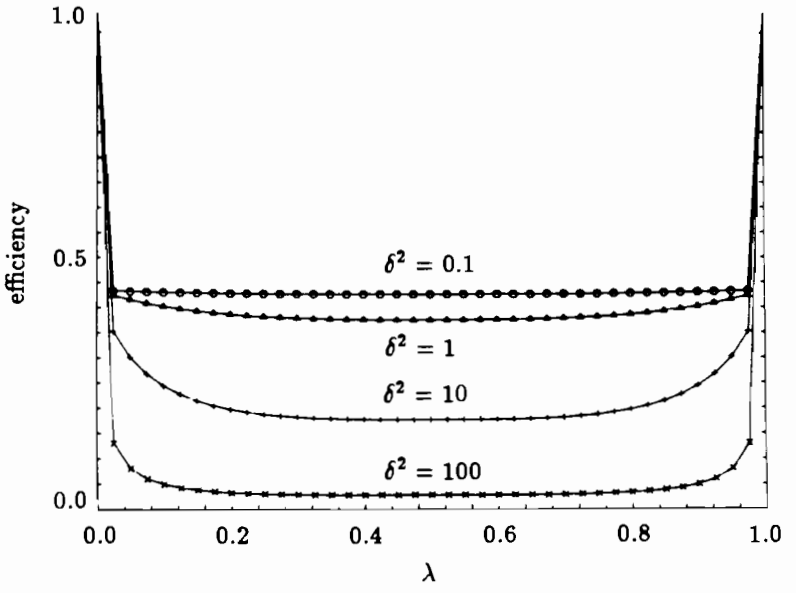

Figure 10.4: Bahadur efficiency of $S_N^{(2)}$ with respect to $S_N^{(1)}$ statistics with Chernoff-Zacks weights; same δ^2 values as in *Figure 10.3.*

Corollary: Let $M_N^{(1)}$ be defined by (10.13) and $\tilde{M}_N^{(2)}$ by (10.28). Suppose both have the same weight coefficients, $c_{N,k}$, and the necessary convergence conditions are met. Then,

$$\log[1 + \delta^2\lambda(1-\lambda)] - BS[\tilde{M}_N^{(2)}; \delta, \lambda] = \log\{1 + \delta^2\lambda(1-\lambda) - BS[M_N^{(1)}; \delta, \lambda]\} \quad \square$$

Of course, it is interesting to consider the efficiency of an $\tilde{M}_N^{(2)}$ statistic with respect to the $M_N^{(2)}$ statistic with the same weights $c_{N,k}$.

Theorem 10.20 *Let $M_N^{(2)}$ be defined by (10.27) and $\tilde{M}_N^{(2)}$ by (10.28). Suppose that both have the same weights, $c_{N,k}$, and the necessary convergence conditions are met; then, for all (δ, λ),*

$$BS[M_N^{(2)}; \delta, \lambda] \geq BS[\tilde{M}_N^{(2)}; \delta, \lambda]$$

with equality, if and only if $\gamma(u)b_1(u;\lambda)$ and $\gamma(u)b_2(u;\delta,\lambda)$ attain their maxima for the same $u = u^\star$ and $\gamma(u^\star) = 1$.

Table 10.2: Bahadur slopes of $M_N^{(1)}$ statistics.

Statistic (weights)	*Slope at* (δ, λ)
$\gamma_1(u) = \left(\frac{u}{\lambda_1} \wedge \frac{1-u}{1-\lambda_1}\right)$, $\lambda_1 < \frac{1}{2}$	$\begin{cases} \delta^2\lambda^2(1-\lambda_1)\lambda_1^{-1} & 0 < \lambda \leq \lambda_1 \\ \delta^2\lambda(1-\lambda)^3(1-\lambda_1)^{-2} & \lambda_1 < \lambda \leq \frac{1}{2} \\ \delta^2(1-\lambda)^2(1-\lambda_1)^{-2} & \frac{1}{2} < \lambda < 1 \end{cases}$
$\gamma_2(u) = \left[\frac{u(1-\lambda_1)}{(1-u)\lambda_1} \wedge \frac{(1-u)\lambda_1}{u(1-\lambda_1)}\right]^{\frac{1}{2}}$	$\delta^2(1-\lambda_1)\lambda_1\left(\frac{\lambda}{\lambda_1} \wedge \frac{1-\lambda}{1-\lambda_1}\right)^2$
$\gamma_3(u) = 1_{[\lambda_1,\lambda_2]}(u)$	$\delta^2\lambda(1-\lambda)\left[\frac{\lambda(1-\lambda_1)}{(1-\lambda)\lambda_1} \wedge 1 \wedge \frac{(1-\lambda)\lambda_2}{\lambda(1-\lambda_2)}\right]$

Proof: First note that,

$$b_2(u;\delta,\lambda) = \frac{\delta b_1(u;\lambda)}{[1+\delta^2\lambda(1-\lambda)-\delta^2 b_1^2(u;\lambda)]^{\frac{1}{2}}}$$

Thus,

$$\begin{aligned} \max_u \gamma(u) b_2(u;\delta,\lambda) &\geq \max_u \frac{\delta\gamma(u)b_1(u;\lambda)}{[1+\delta^2\lambda(1-\lambda)-\delta^2\gamma^2(u)b_1^2(u;\lambda)]^{\frac{1}{2}}} \\ &\geq \frac{\delta \max_u \gamma(u)b_1(u;\lambda)}{[1+\delta^2\lambda(1-\lambda)-\delta^2 \max_u \gamma^2(u)b_1^2(u;\lambda)]^{\frac{1}{2}}} \end{aligned}$$

and since

$$BS[\tilde{M}_N^{(2)};\delta,\lambda] = \log\left\{1 + \frac{\delta^2[\max_u \gamma(u)b_1(u;\lambda)]^2}{1+\delta^2\lambda(1-\lambda)-\delta^2[\max_u \gamma(u)b_1(u;\lambda)]^2}\right\}$$

This proves the theorem. □

10.5 Some Comparisons and Examples

In this section, the results of the foregoing sections are applied to derive the slopes and efficiencies of some specific statistics from the general classes under consideration.

Figure 10.1 presents the efficiencies of four $S_N^{(1)}$ statistics — see also *Table 10.1* —

with respect to Raghavachari's upperbound, as a function of λ. It should be recalled that these efficiencies do not depend on δ, the size of change. This figure clearly complies with *Theorem 10.9* and its corollary.

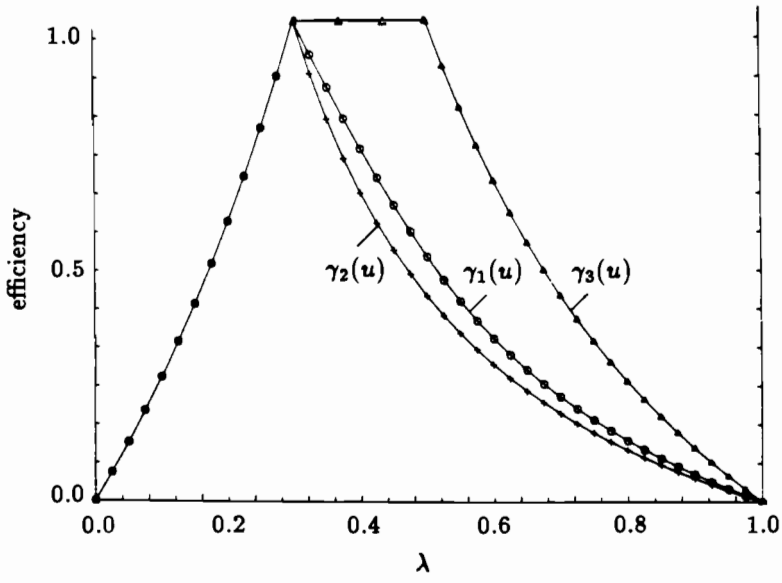

Figure 10.5: Bahadur efficiency of $M_N^{(1)}$ statistics, weight functions as in *Table 10.2*; $\lambda_1 = 0.3$, $\lambda_2 = 0.5$.

According to the corollary of *Theorem 10.16*, the mutual relations between the corresponding $S_N^{(2)}$ statistics — i.e., those with the same weight functions — are similar; see *Figure 10.2*. However, they now depend on δ^2. For $\delta^2 \to 0$, the efficiency of an $S_N^{(2)}$ test (w.r.t. Raghavachari's upper bound) tends to the efficiency of the $S_N^{(1)}$ test with the corresponding weights. And, if $\delta^2 \to \infty$, this efficiency goes to zero for all λ, for which $S_N^{(2)}$ is not optimal, i.e., for all λ where $\zeta_\lambda^2 < \lambda(1-\lambda)$. This effect is brought about by the bad behavior for large δ-values of the variance estimator s_N^2.

In *Figure 10.3*, the efficiency of the $S_N^{(2)}$ statistic with the Chernoff-Zacks weights is presented for several δ^2-values. It appears that the difference from the known variance case remains small, for small and moderate δ^2.

Figure 10.4 shows the loss in efficiency because σ^2 is unknown. Again, with the Chernoff-Zacks weights, the efficiency of the $S_N^{(2)}$ statistic with respect to the $S_N^{(1)}$ statistic is given for some values of δ^2. Although this efficiency tends to 1 when $\delta^2 \to 0$ (or $\lambda \downarrow 0$, or $\lambda \uparrow 1$), even for small values of δ^2, (or λ-values close to 0 or 1), the loss in efficiency is considerable.

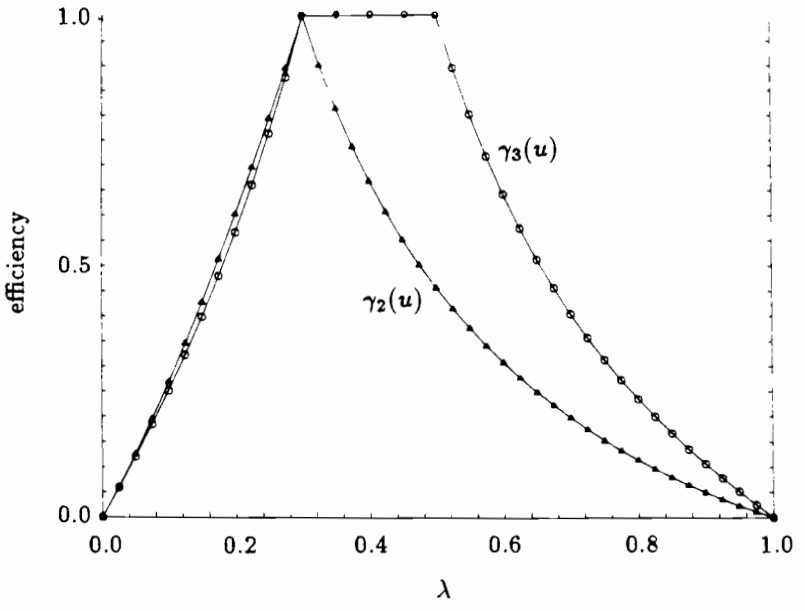

Figure 10.6: Bahadur efficiency of $M_N^{(2)}$ statistics; $\delta^2 = 1$; $\lambda_1 = 0.3$, $\lambda_2 = 0.5$.

For the $M_N^{(1)}$ statistics, we easily get from *Theorem 10.12* that if $\gamma(\lambda_1) = 1$ [$= \max_u \gamma(u)$], then:

1. For all alternatives (δ, λ) with $\lambda < \lambda_1$, the Bahadur slope is independent of $\gamma(u)$ for $u > \lambda_1$.

2. And, provided that $\gamma(u) \leq \left(\frac{1-\lambda_1}{\lambda_1}\frac{u}{1-u}\right)^{\frac{1}{2}}$ for $u \in [0, \lambda_1]$, it does not depend on $\gamma(u)$ for $u < \lambda_1$, either.

Of course, a similar statement holds for alternatives with $\lambda > \lambda_1$. This property is illustrated in *Figure 10.5* (see also *Table 10.2*). For $\lambda < \lambda_1 = 0.3$, the included statistics all have the same slope. (Moreover, the two-sample statistic $T_{N,k}$, with $k/N \to \lambda_1$, also has this slope for $\lambda < \lambda_1$; see *Table 10.1*).

In view of the corollary to *Theorem 10.19*, effects similar to those illustrated for the $S_N^{(2)}$ statistics in *Figures 10.2–10.4* also hold for the $\tilde{M}_N^{(2)}$ statistics and their relations with the $M_N^{(1)}$ statistics.

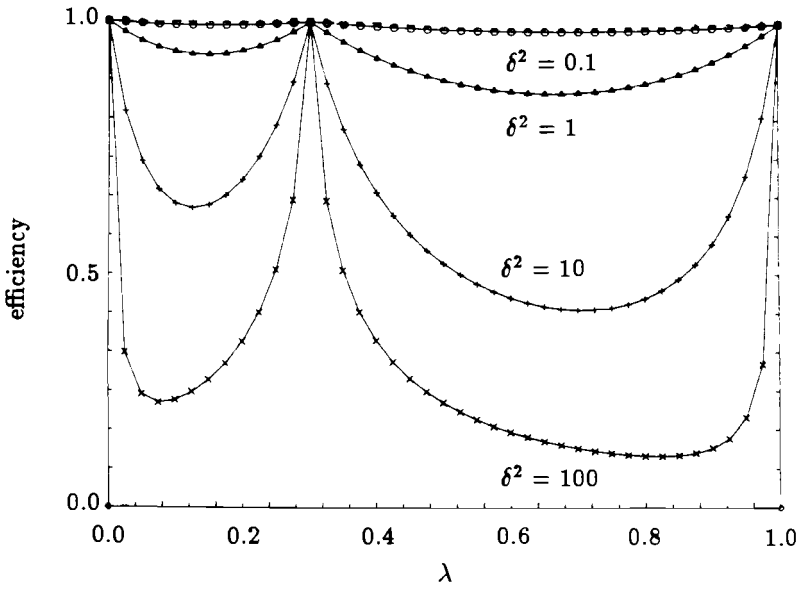

Figure 10.7: Bahadur efficiency of $\tilde{M}_N^{(2)}$ with respect to $M_N^{(2)}$ statistics; weights $\gamma_2(u)$; $\lambda_1 = 0.3$; δ^2 values as in *Figure 10.3.*

For both the $M_N^{(2)}$ and the $\tilde{M}_N^{(2)}$ statistics, the statement above concerning the influence of γ on the slope of the $M_N^{(1)}$ statistics is relevant; only for the $M_N^{(2)}$ statistics the second part of it needs a slightly stronger condition on the weight function $\gamma(u)$. In *Figure 10.6*, this can be seen: although the weight functions of both the included statistics comply with the previous restriction, their efficiencies differ for $\lambda < \lambda_1$. This difference depends on δ^2.

The efficiency of $\tilde{M}_N^{(2)}$ with respect to $M_N^{(2)}$ statistics is considered in *Figure 10.7*, for the case where both have the limit weight function $\gamma(u) = [(1-\lambda_1)\lambda_1^{-1}u(1-u)^{-1}]^{\frac{1}{2}}$ for $u \leq \lambda_1$, and $\gamma(u) = [(1-\lambda_1)^{-1}\lambda_1 u^{-1}(1-u)]^{\frac{1}{2}}$ for $u > \lambda_1$. According to *Theorem 10.20*, this efficiency tends to 0 if $\delta^2 \to \infty$, for all λ with $\gamma(\lambda) \neq 1$.

10.6 Power Estimates for Small Sample Sizes

In this final section, the performance of the c.p.p. tests is considered for samples of practical size. Therefore, the power of the tests will be investigated for the alternatives of a shift of size $\delta = 0.5\sigma$, occurring between X_n and X_{n+1}, with $n = [\lambda N]$, $\lambda = 0.1(0.1)0.5$, and for $N = 20, 50, 100$, and $\alpha = 0.05$.

Table 10.3: Power of $S_N^{(1)}$ and $S_N^{(2)}$ statistics at (δ, λ); all X_i normal, $\text{Var}(X_i) = 1$; $\alpha = 0.05$; $\delta = 0.5$.

N	λ	*Statistics*				
		$S_N^{(1)}$ [a]	$S_N^{(2)}$ [b]	$S_N^{(2)}$ [c]	$S_N^{(2)}$ [d]	$S_N^{(2)}$ [e]
20	0.1	0.098	0.096	0.102	0.081	0.082
	0.2	0.153	0.142	0.139	0.165	0.131
	0.3	0.203	0.204	0.199	0.203	0.193
	0.4	0.238	0.232	0.224	0.228	0.218
	0.5	0.250	0.232	0.236	0.149	0.215
50	0.1	0.137	0.129	0.141	0.144	0.109
	0.2	0.253	0.235	0.262	0.263	0.234
	0.3	0.360	0.352	0.346	0.483	0.378
	0.4	0.430	0.418	0.398	0.425	0.355
	0.5	0.455	0.450	0.412	0.280	0.375
100	0.1	0.193	0.193	0.207	0.184	0.148
	0.2	0.398	0.397	0.402	0.449	0.350
	0.3	0.569	0.560	0.533	0.717	0.615
	0.4	0.668	0.664	0.621	0.611	0.582
	0.5	0.698	0.684	0.650	0.497	0.594

[a] with $d_{N,i+1} - d_{N,i} = \text{const.}$ (Chernoff-Zacks).

[b] with $d_{N,i+1} - d_{N,i} = \text{const.}$

[c] with $\forall_k c_{N,k} = c_N$.

[d] with $c_{N,k} = 1$ if $k = 0.3N$, $c_{N,k} = 0$ elsewhere.

[e] with $c_{N,k} = c_{N,N-k} = 1$ if $k = 0.3N$, $c_{N,k} = 0$ elsewhere.

Apart from the $S_N^{(1)}$ statistics, which are normally distributed under $\mathcal{H}_0$ as well as under $\mathcal{H}_a$, and the $\mathcal{H}_0$-distribution of $S_N^{(2)}$ (Section 10.4), the distribution under $\mathcal{H}_0$ and/or $\mathcal{H}_a$ of our statistics are unknown. In these cases, the power is estimated using Monte Carlo experiments.

Critical values ($\alpha = 0.05$) were estimated based upon 5,000 samples of size N. Next, power estimates were obtained using, for each alternative, 1,000 samples of size N.

All the simulations were carried out on a Burroughs B7900 computer. Uniform $[0,1)$ variables were generated by the linear congruential random number generator, which is an intrinsic feature of the B7900. Transformation to normally distributed r.v.'s then followed by the Box-Muller method (see Law and Kelton, 1982, Ch.7).

In *Tables 10.3* and *10.4*, results are given of various tests, based upon sum- and max-type statistics, for the $S_N^{(2)}$ statistics with the same weights as in Section 10.5.

Table 10.4: Power of $M_N^{(1)}$, $M_N^{(2)}$ and $\tilde{M}_N^{(2)}$ statistics at (δ, λ); all X_i normal, $\mathrm{Var}(X_i) = 1$; $\alpha = 0.05$; $\delta = 0.5$.

N	λ	*Statistics*				
		$M_N^{(1)}$ [a]	$M_N^{(2)}$ [b]	$M_N^{(2)}$ [c]	$M_N^{(2)}$ [d]	$\tilde{M}_N^{(2)}$ [e]
20	0.1	0.108	0.099	0.073	0.088	0.106
	0.2	0.125	0.138	0.130	0.135	0.184
	0.3	0.181	0.149	0.206	0.170	0.247
	0.4	0.188	0.150	0.201	0.218	0.278
	0.5	0.190	0.170	0.192	0.230	0.221
50	0.1	0.144	0.140	0.131	0.116	0.133
	0.2	0.239	0.220	0.287	0.216	0.282
	0.3	0.308	0.299	0.447	0.386	0.463
	0.4	0.328	0.332	0.443	0.446	0.484
	0.5	0.365	0.340	0.395	0.469	0.397
100	0.1	0.258	0.199	0.170	0.158	0.199
	0.2	0.409	0.379	0.450	0.345	0.484
	0.3	0.497	0.461	0.678	0.582	0.722
	0.4	0.604	0.508	0.712	0.700	0.712
	0.5	0.586	0.552	0.653	0.718	0.628

[a] $\forall_k c_{N,k} = 1$

[b] $\forall_k c_{N,k} = 1$

[c] $c_{N,k} = \frac{k}{0.3N} \wedge \frac{N-k}{0.7N}$

[d] $c_{N,k} = 1,\ 0.3N \leq k \leq 0.7N$

[e] $c_{N,k}$ as (c)

Comparing the $S_N^{(1)}$ and $S_N^{(2)}$ statistics, both with the Chernoff-Zacks weights, i.e., (a) and (b) in *Table 10.3*, shows that the difference in power of these tests is almost negligible, whereas from *Figure 10.4*, we know that the Bahadur efficiency of $S_N^{(2)}$ with respect to $S_N^{(1)}$ is less than 0.5 for all the alternatives involved.

On the other hand, the differences in power between the four $S_N^{(2)}$ statistics comply reasonably well with their Bahadur slopes (see *Figures 10.2* and *10.3*), when $N = 50$ or 100. For $N = 20$, however, the power of the statistics (d) and (e), when compared with those of the other $S_N^{(2)}$ statistics, is smaller than would be expected from the Bahadur slopes.

For the max-type statistics, $M_N^{(1)}$, $M_N^{(2)}$, and $\tilde{M}_N^{(2)}$, the power was estimated for the same alternatives and sample sizes (see *Table 10.4*).

Again, the difference between the powers of the $M_N^{(1)}$ and $M_N^{(2)}$ tests are smaller than would be expected from the Bahadur efficiency. Moreover, the Bahadur optimality of an $M_N^{(2)}$ statistic at those alternatives (δ, λ) for which $\gamma(\lambda) = 1$ (see Remark 10.9) is not reflected in the estimated powers. For example, at alternatives with $\lambda \in (0.3, 0.7)$, the power of the test with $c_{N,k} = 1_{[0.3,0.7]}(k/N)$ is much better than the power of the likelihood ratio test, (b).

A comparison of the powers in *Table 10.3* with those in *Table 10.4* also shows that the likelihood ratio test is not the most powerful for small sample sizes: the sum-type statistic $S_N^{(1)}$, with Chernoff-Zacks weights, has greater power than $M_N^{(1)}$ with $c_{N,k} = 1$ when $0.3 \leq \lambda \leq 0.7$. This confirms the results of Sen and Srivastava (1975b).

All in all, the likelihood ratio test is not as good as could be expected from the Bahadur efficiency, or as was supposed by Haccou *et al.* (1985), who considered only the likelihood ratio test and a minor modification of it for the case of exponentially distributed X_i.

Finally, consider the statistics $M_N^{(2)}$ and $\tilde{M}_N^{(2)}$, both with weights $c_{N,k} = \frac{k}{0.3N} \wedge \frac{N-k}{0.7N}$, i.e., columns (c) and (e) in *Table 10.4*. [Recall that with the weights chosen in (b) or (d), the test based upon $M_N^{(2)}$ is equivalent to that based upon $\tilde{M}_N^{(2)}$.] From Section 10.4, we know that the Bahadur efficiency of $M_N^{(2)}$ with respect to $\tilde{M}_N^{(2)}$ is greater than 1, when $\lambda \leq 0.3$ (*Theorem 10.19*); although the differences remain small, the power of $M_N^{(2)}$ is rather smaller than the power of $\tilde{M}_N^{(2)}$ for $N = 20, 50,$ or 100.

References

Bahadur, R.R. (1971), *Some Limit Theorems in Statistics.* Philadelphia, PA: SIAM.

Bhattacharyya, G.K. and Johnson, R.A. (1968), Nonparametric tests for shift at an unknown time point. *Annals of Mathematical Statistics*, **39**, 1731-1743.

Billingsley, P. (1968), *Convergence of Probability Measures.* New York: John Wiley.

Chernoff, H. and Zacks, S. (1964), Estimating the current mean of a normal distribution which is subjected to changes in time. *Annals of Mathematical Statistics*, **35**, 999-1018.

Chow, Y.S. and Lai, T.L. (1973), Limiting behaviour of weighted sums of random variables. *Annals of Probability*, **1**, 810-824.

Deshayes, J. and Picard, D. (1982), Tests of disorder of regression: Asymptotic comparison. *Theory of Probability and its Applications*, **27**, 100-115.

Feller, W. (1957), *An Introduction to Probability Theory and Its Applications*, vol. 1, second edn. New York: John Wiley.

Gardner, L.A. (1969), On detecting changes in the mean of normal variates. *Annals of Mathematical Statistics*, **40**, 116-126.

Groeneboom, P. and Oosterhoff, J. (1977), Bahadur efficiency and probabilities of large deviations. *Statistica Neerlandica.*, **31**, 1 -24.

Haccou, P., Meelis, E., and van de Geer, S. (1985), On the likelihood ratio test for a change point in a sequence of independent exponentially distributed random variables. Report MS-R8507. Amsterdam: Centre for Mathematics and Computer Science.

Hawkins, D.M. (1977), Testing a sequence of observations for a shift in location. *Journal of the American Statistical Association*, **72**, 180-186.

Killeen, T.J. and Hettmansperger, T.P. (1972), Bivariate tests for location and their Bahadur efficiencies. *Annals of Mathematical Statistics*, **43**, 1507-1516.

Killeen, T.J., Hettmansperger, T.P., and Sievers, G.L. (1972), An elementary theorem on the probability of large deviations. *Annals of Mathematical Statistics*, **43**, 181-192.

Law, A.M. and Kelton, W.D. (1982), *Simulation Modeling and Analysis*. New York: McGraw-Hill.

Praagman, J. (1986), *Efficiency of Change-Point Tests*, Unpublished Ph.D. thesis, Eindhoven University of Technology.

Raghavachari, M. (1970), On a theorem of Bahadur on the rate of convergence of test statistics. *Annals of Mathematical Statistics*, **41**, 1695-1699.

Sen, A. and Srivastava, M.S. (1975a), Some one-sided tests for change in level. *Technometrics*, **17**, 61-64.

Sen, A. and Srivastava, M.S. (1975b), On tests for detecting change in mean. *Annals of Statistics*, **3**, 98-108.

Worsley, K.J. (1979), On the likelihood ratio test for a shift in location of normal populations. *Journal of the American Statistical Association*, **74**, 365-367.

CHAPTER 11

The Use of Graphical Displays in the Analysis of Structural Change

Zbigniew Wasilewski

Summary

Recently developed techniques for graphic analysis of residuals have made them more legible. Things to look for in a plot are: (a) the largest residuals; (b) progressive change in the variability of the residuals; (c) a curved regression of residuals on fitted values or the number of cases; and (d) the subsets of successive residuals with significantly different configuration. In this chapter we concentrate on (d) which indicates possible structural changes. Using Monte Carlo experiments, we examine smoothing scatterplots of the transformed least squares residuals. To interpret the smoothed scatterplots of standardized residuals, we propose plotting a kind of confidence envelope around the smoothed curve, based on Atkinson (1981).

11.1 Introduction

Validation of a regression model is mainly based on investigating to what extent the assumptions underlying model specification are fulfilled. Relatively little attention has been paid to the structural analysis of regression data and residuals. The drawback of test procedures is that they are usually based on many assumptions that are often difficult to verify; and even when such procedures indicate model failure, they do not provide its real reason, owing to the mutual relation between certain failures. The typical example here is the indication of outliers, structural change, and nonlinearity. Such failures can usually be detected by applying suitable tests, but very often only all these tests together indicate the failure.

The use of graphical displays can be very useful in such a situation. The plots show a variety of features and make it possible to see, how certain features relate to one another. In comparison with tests, they provide much more qualitative information (see, e.g., Anscombe, 1973; Daniel and Wood, 1971).

The aim of this chapter is to show how graphical displays can be used to investigate nonconstancy of regression parameters. To make the graphical analysis more legible, a number of transformations of OLS residuals and different smoothing procedures will be considered. The use of scatterplots, normal probability plots, and partial leverage regression plots is illustrated with both simulated and real data.

11.2 Transformed Residuals Used in the Plots

Owing to rather poor correspondence between the regression disturbances and the OLS residuals, their use in judgment about model failures is strongly limited. For diagnostic purposes, several transformations of the OLS residuals can be used to overcome some of their shortcomings (see, e.g., Cook and Weisberg, 1982; and Belsley *et al.*, 1980). These transformations are described briefly below.

11.2.1 Internally studentized residuals

In least squares regression, the internally studentized residuals r_i, $i = 1, \ldots, n$, are derived from the OLS residuals e_i as

$$r_i = \frac{e_i}{s\sqrt{1 - h_{ii}}} \tag{11.1}$$

where $s^2 = \sum_{i=1}^{n} e_i^2/(n-k)$, $h_{ij} = x_i'(X'X)^{-1}x_j$, and x_i' denotes the ith row of the $n \times k$ matrix X. The quantity $r_i^2/(n-k)$ follows a Beta-distribution with parameters $1/2$ and $(n-k-1)/2$, and $E(r_i) = 0$, $Var(r_i) = 1$, $Corr(r_i, r_j) = -h_{ij}/\sqrt{(1-h_{ii})(1-h_{jj})}$ for $i \neq j$ (see Cook and Weisberg, 1982, Atkinson 1982).

The residuals $r_i = e_i/(s\sqrt{1-h_{ii}})$ reflect a contamination of the ith observation better than the OLS residuals, when $h_{ii} \to 1$, but problems can still arise when $h_{ii} \to 0$.

The internally studentized residuals are often used as a replacement for the OLS residuals in graphical procedures, such as scatterplots or probability plots (see, e.g., Anscombe and Tukey, 1963; Cook and Weisberg, 1982; Atkinson, 1982).

11.2.2 Externally studentized residuals

These residuals, denoted by t_i in the following, are defined similarly to the internally studentized residuals, but the estimator of σ^2 is modified so that it does not make use of the ith observation:

$$s^2(i) = \frac{1}{n-k-1}\sum_{j\neq i}[y_j - x_j b(i)]^2 \tag{11.2}$$

where $b(i)$ is the estimate of the regression coefficients β when x_ith and y_i have been deleted from the set of data to be fitted. A simple formula for $s^2(i)$ results from

$$(n-k-1)s^2(i) = (n-k)s^2 - e_i^2/(1-h_{ii}) \tag{11.3}$$

Substituting (11.2) into (11.1) and using (11.3), we obtain

$$t_i = \frac{e_i}{s(i)\sqrt{1-h_{ii}}} = r_i\sqrt{\frac{n-k-1}{n-k-r_i^2}} \tag{11.4}$$

which shows that t_i^2 is a monotone transformation of r_i^2.

In comparison with the internally studentized residuals, the t_is reflect large deviations much more dramatically, as $t_i^2 \to \infty$ for $r_i^2 \to n-k$. The residual t_i can also be interpreted as a test statistic for the significance of a dummy variable in point "i"; under normality of the errors, t_i follows Student's t-distribution with $n-k-1$ degrees of freedom.

11.2.3 Predicted residuals

Both OLS and studentized residuals are based on a fit of the regression model to all the data. The predicted residuals are, in contrast, given by

$$e_{(i)} = y_i - x_i' b(i) = y_i - x_i'\left[b - \frac{e_i(X'X)^{-1}x_i}{1-h_{ii}}\right] = \frac{e_i}{1-h_{ii}} \tag{11.5}$$

and can be interpreted as prediction errors, because the ith predicted residual is based on a fit of the model to the data with the ith case deleted.

Under normality of the errors, the predicted residuals e_i are normally distributed with mean zero and variance equal to $\sigma^2/(1-h_{ii})$ and have the same correlation structure as OLS residuals. They tend to emphasize the cases with large h_{ii} value. When divided by the OLS estimate of the standard error of prediction, based on the reduced set of data, predicted residuals lead to the externally studentized residuals.

11.2.4 Recursive residuals

Assuming that the index of the observations stands for any ordering, the ith recursive residual, $i = k+1, \ldots, n$, is defined (Brown *et al.*, 1975) as,

$$q_i = \frac{y_i - x_i' b_{i-1}}{\sqrt{1 + x_i'(X_{i-1}'X_{i-1})^{-1}x_i}} \tag{11.6}$$

where b_{i-1} is computed using the first $i-1$ observations only, and X_{i-1} contains the first $i-1$ rows of X. The term "recursive" is used because b_i can be computed from b_{i-1} by means of updating formulas (Plackett, 1950; Bartlett, 1951):

$$\begin{aligned} b_i &= b_{i-1} + (X_i'X_i)^{-1}x_i(y_i - x_i'b_{i-1}) \\ (X_i'X_i)^{-1} &= (X_{i-1}'X_{i-1})^{-1} - d_i d_i'/(1 + x_i'd_i) \end{aligned} \tag{11.7}$$

where $d_i = (X_{i-1}'X_{i-1})^{-1}x_i$. It is obvious that only $n-k$ recursive residuals can be calculated, as k points form the basis for the calculations. Brown *et al.* (1975) suggest starting the calculation on the basis of the first k data points, but in fact any other arbitrarily selected basis can be used.

Under standard assumptions, the recursive residuals are normally distributed with mean zero and variances σ^2. They are easily interpreted as showing the effect of successively deleting points from the data set. If there is a single outlier among the data at point $r > k$, q_r will be large and positive, the recursive residuals q_i for $i > r$ will tend to be biased toward negative values, and the recursive residuals q_i for $k < i < r$ will be random.

Because of their attractive properties, we use scatterplots of recursive residuals for judging break points and nonlinearity. The computation of the residuals is started on the basis of the first and last k observations, the respective residuals being denoted as foreward $[q_i^{(f)}]$ and backward $[q_i^{(b)}]$ recursive residuals. For a comprehensive study of the use of recursive residuals in linear regression model validation, see Galpin and Hawkins (1984).

11.3 Smoothing Residuals

To enhance the visual information on a scatterplot, three techniques for discovering and summarizing smooth residual patterns were used.

The first one is suggested by Cleveland (1979), Cleveland and McGill (1984), and Cleveland and Kleiner (1975). They proposed to plot smoothed points in conjunction with a scatterplot of the data (x_i, y_i), $i = 1, \ldots, n$. These smoothed points are designed to enhance the perception of the pattern of dependence of y on x.

For a given set of data points (x_i, y_i), $i = 1, \ldots, n$, corresponding smoothed values $\hat{y}_i$ can be obtained using locally weighted polynomial regression of degree d of y on x. This is done in the following steps (see Cleveland, 1979):

1. For a given total number n of observations choose a number r, $1 \leq r < n$, of the observations that form the neighborhood of (x_i, y_i) to be used to compute $\hat{y}_i$.

2. For each point x_i form the weights $w_k(x_i) = W[h_i^{-1}(x_k - x_i)]$ where $W(x) > 0$ for $|x| < 1$; $W(x) = 0$ for $|x| \geq 1$; $W(x) = W(-x)$; and h_i is the rth smallest number among $|x_k - x_i|$.

3. For each i, compute the corresponding fitted value

$$\hat{y}_i = \sum_{j=0}^{d} c_{ij} x_i^j$$

where the coefficients c_{ij} are obtained by minimizing the weighted sum of squares

$$\sum_{k=1}^{n} w_k(x_i)(y_k - c_{i0} - c_{i1}x_k - \ldots - c_{id}x_k^d)^2$$

4. Take $(x_i, \hat{y}_i,)$ as the smoothed point at x_i.

Cleveland (1979) proposed a robustified version of this procedure, iteratively adjusting the weights used in step (3) by means of a robust weight function based on the residuals. Such of robustification implies that outliers play a smaller role in the determination of the smoothed points. In this study we use a nonrobust version of the smoothing procedure, taking $d = 1$, $r = \lfloor 0.7n \rfloor$ and a "tricube" weight function W defined by,

$$W(x) = \begin{cases} (1 - |x|^3)^3, & \text{for } |x| < 1 \\ 0, & \text{for } |x| \geq 1 \end{cases}$$

In order to aid the interpretation of the smoothed scatterplots of the studentized residuals, we plot a kind of confidence envelope around the smoothed curve. This is achieved in the way proposed by Atkinson (1981) for probability plots, using small Monte Carlo experiments. Given the matrix X, 30 pseudo random samples of standardized residuals were generated, using standardized normal deviates as disturbances. For each sample the corresponding smoothed values of the residuals were computed, and the smallest and the largest from these smoothed values, among the 30 samples, were chosen. As a result, we obtained two sets of smoothed points that form a kind of envelope around the smoothed residuals.

Two other smoothing techniques are based on combinations of running medians and running weighted averages. These were described by Tukey (1971) and Velleman and Hoaglin (1981) (see also Mosteller and Tukey 1977). Using compact notation for elementary smoothing operations, these two techniques will be refered as ⟨4253H, twice⟩ and ⟨3RSSH, twice⟩.

The combination ⟨4253H, twice⟩ of smoothers starts with a running median of four, recentered with a running median of two. The result is resmoothed by a running median of five, and once more by the running median of three. Finally, the running weighted average is used (denoted here by H for hanning). The result of this smoothing is then polished by computing residuals, applying the same smoother to them, and adding the result to the first pass. Algebraically, the above sequence of operations replaces the data values r_t by x_t:

1. $z_t = (med\{r_{t-2}, r_{t-1}, r_t, r_{t+1}\} + med\{r_{t-1}, r_t, r_{t+1}, r_{t+2}\})/2$
2. $v_t = med\{z_{t-2}, z_{t-1}, z_t, z_{t+1}, z_{t+2}\}$
3. $u_t = med\{v_{t-1}, v_t, v_{t+1}\}$
4. $x_t = 0.25u_{t-1} + 0.5u_t + 0.25u_{t+1}$

then the differences $w_t = r_t - x_t$ are computed, steps (1)–(4) are repeated for the w_ts, and the result added to the smoothed values x_t.

The technique ⟨3RSSH, twice⟩ is the combination of the repeated running median of three elements (3R) followed by a splitting operation (S). After splitting, the entire sequence is resmoothed (R). The last two operations are repeated, and the result is resmoothed by running weighted average method (H). All the above operations are then repeated for the residuals. The ⟨3RSSH, twice⟩ technique can be described by the following sequence of operations on the data values r_t:

1. $z_t = med\{r_{t-1}, r_t, r_{t+1}\}$,
2. $v_t = med\{z_{t-1}, z_t, z_{t+1}\}$
3. $u_t = med\{3v_{t+1} - 2v_{t+2}, v_t, v_{t+1}\}$, $u_{t-1} = med\{v_{t-2}, v_{t-1}, 3v_{t-2} - 2v_{t-3}\}$
4. $w_t = med\{u_{t-1}, u_t, u_{t+1}\}$,
5. $x_t = med\{w_{t-1}, w_t, w_{t+1}\}$
6. $u_t = med\{3x_{t+1} - 2x_{t+2}, x_t, x_{t+1}\}$, $u_{t-1} = med\{x_{t-2}, x_{t-1}, 3x_{t-2} - 2x_{t-3}\}$
7. $w_t = med\{u_{t-1}, u_t, u_{t+1}\}$,
8. $x_t = med\{w_{t-1}, w_t, w_{t+1}\}$
9. $z_t = 0.25x_{t-1} + 0.5x_t + 0.25x_{t+1}$

then, the operations (1)–(9) are applied to the differences $w_t = r_t - z_t$; the final result is obtained as the sum of the values z_t and smoothed differences.

11.4 Using Plots in the Analysis of Structural Changes

To illustrate how helpful graphical displays can be in investigating the nonconstancy of the regression parameters, four examples are considered. Two of them are based on the artificially generated data with break point. Two other examples are based on real data.

Figures 11.1 – 11.3 reflect the case of a change in the parameter vector used to generate the data. The standard linear regression model was used, with the matrix X containing one constant term and three exogenous variables (without high leverage points). The error

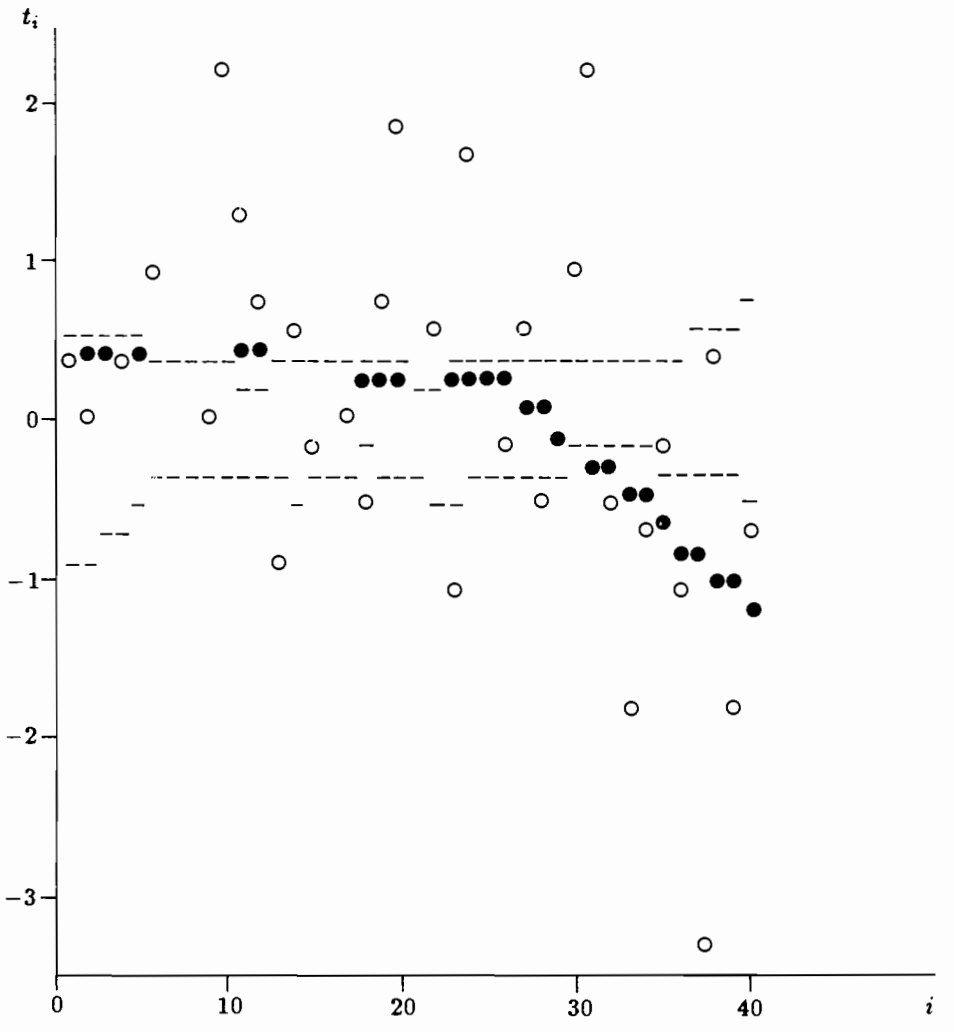

Figure 11.1: Scatterplot of the externally studentized residuals t_i and their confidence envelope against the index i for example 1.

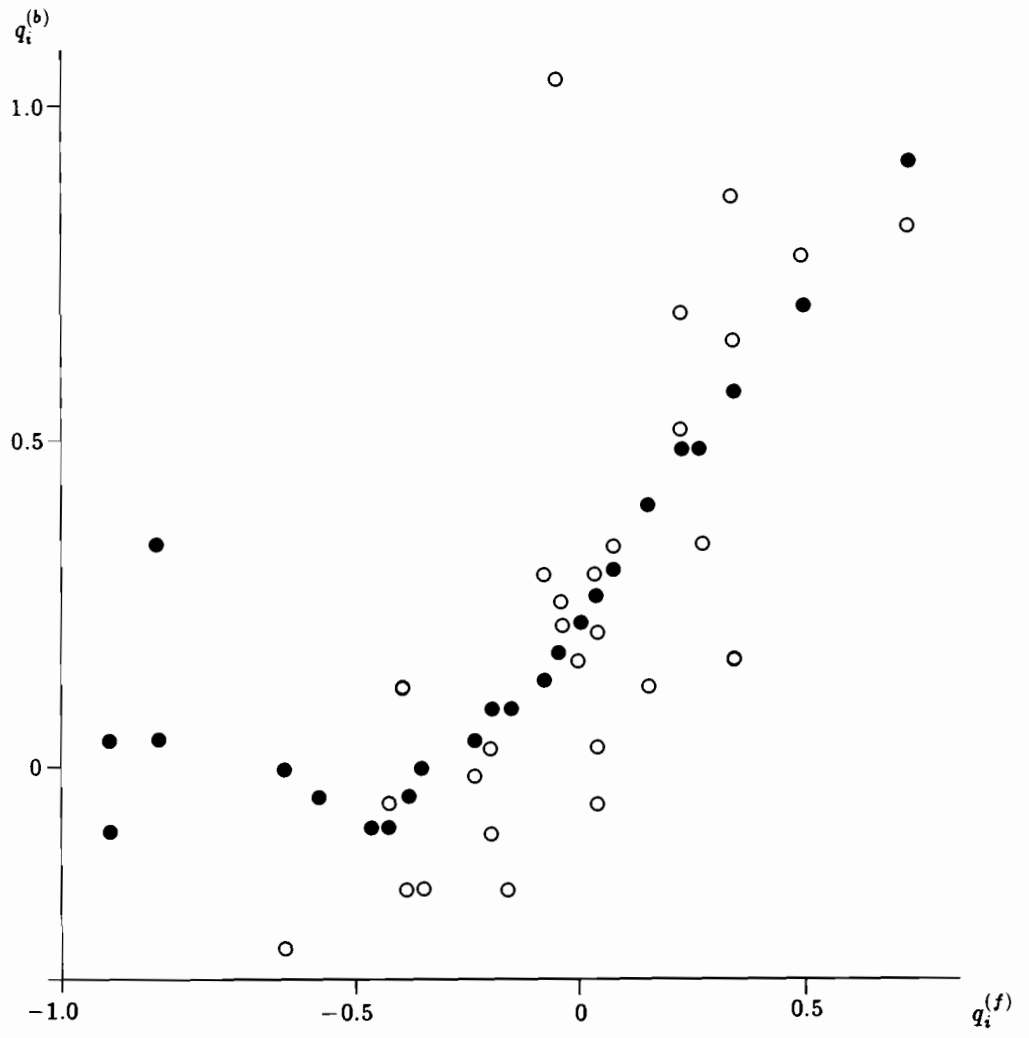

Figure 11.2: Scatterplot of the foreward $[q_i^{(f)}]$ and backward $[q_i^{(b)}]$ recursive residuals for example 1.

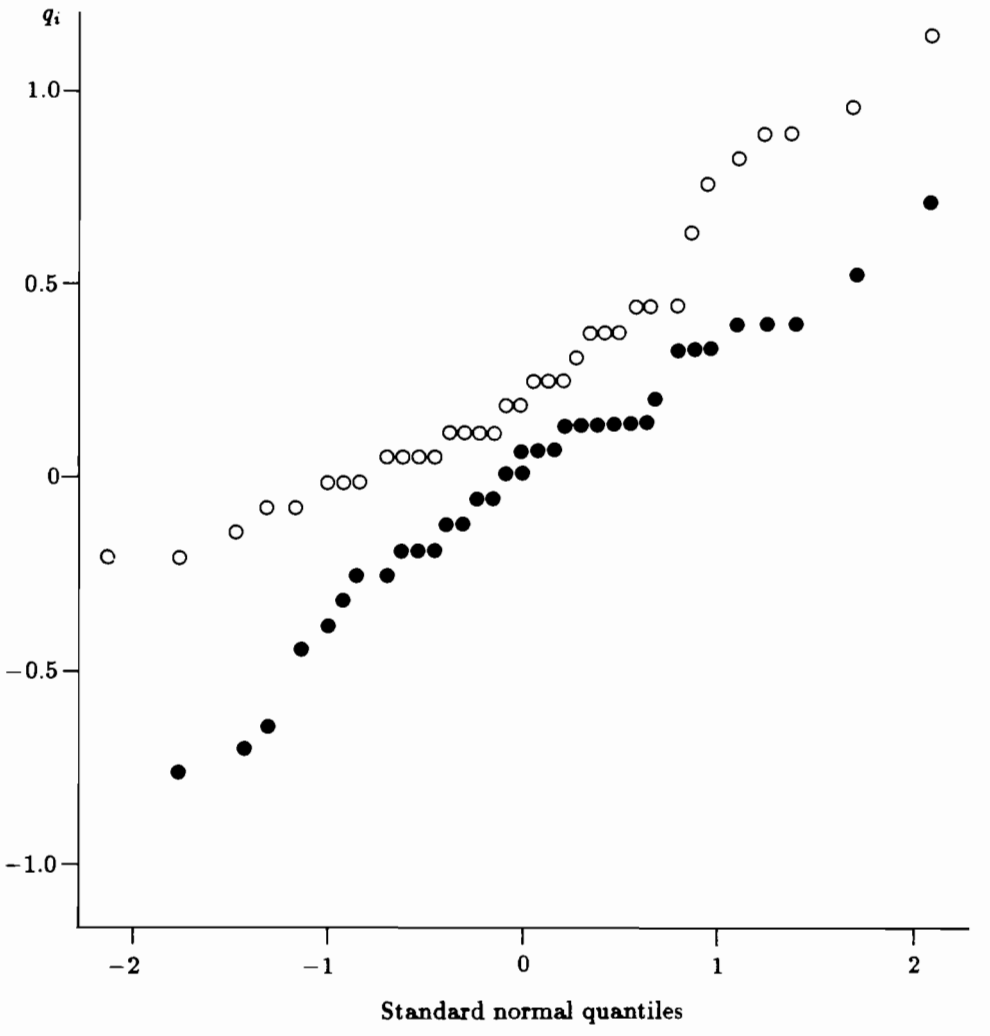

Figure 11.3: Normal probability plot of foreward (∗) and backward (+) recursive residuals q_i for example 1.

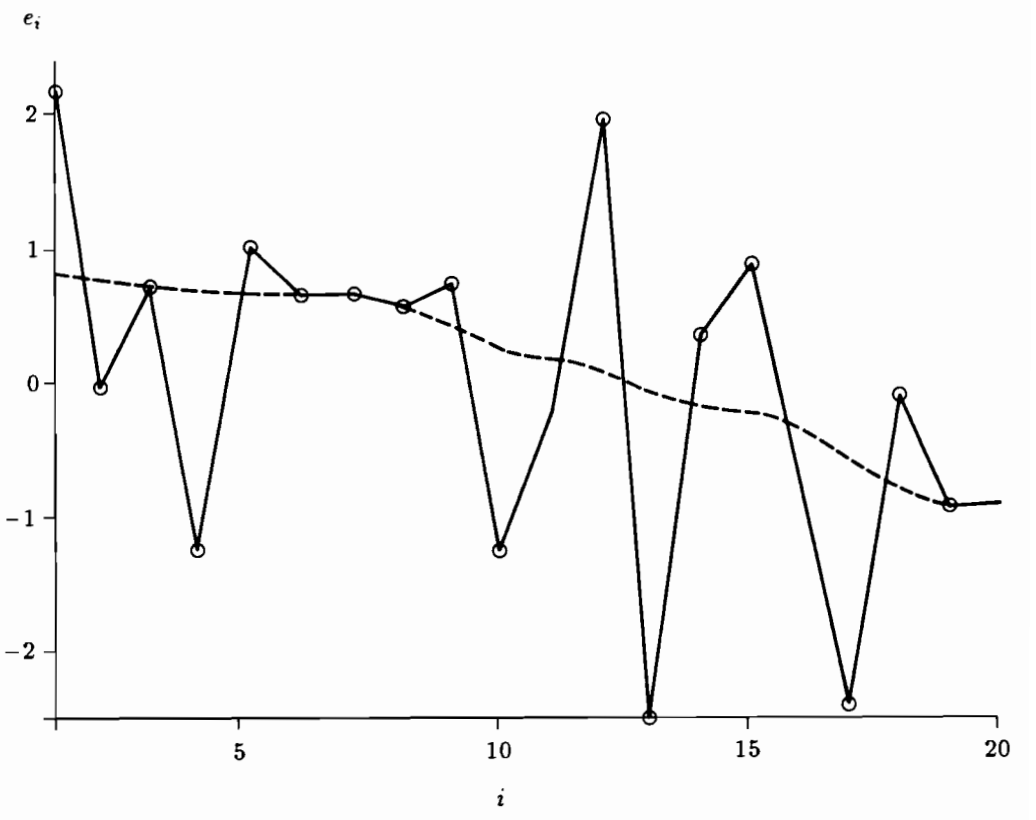

Figure 11.4: OLS residuals e_i and smoothed OLS residuals against the index i for Quandt's data.

term followed a normal distribution with mean zero and variance equal to 0.2 times the variance of $X\beta$. Beginning from the 30th observation, the last 10 data were generated using changed values of the parameters.

This is quite visible in *Figure 11.1*, where externally studentized residuals are plotted against the case number. Thus, by the 30th residual, the smoothed curve is shifted down beyond the lower envelope bound. [The simulated envelope bounds determine a kind of confidence intervals for the smoothed residuals. Fluctuations of the smoothed curve appearing inside these bounds are still acceptable for the given matrix X].

The plot of backward and forward recursive residuals, plotted against each other in *Figure 11.2*, also indicates the two different regimes used to generate the sample data. Note that it is possible to see this mainly because the smoothed points were plotted out. The normal probability plot of the recursive residuals (*Figure 11.3*) shows a great distance between the curves which correspond to the forward and backward recursive residuals.

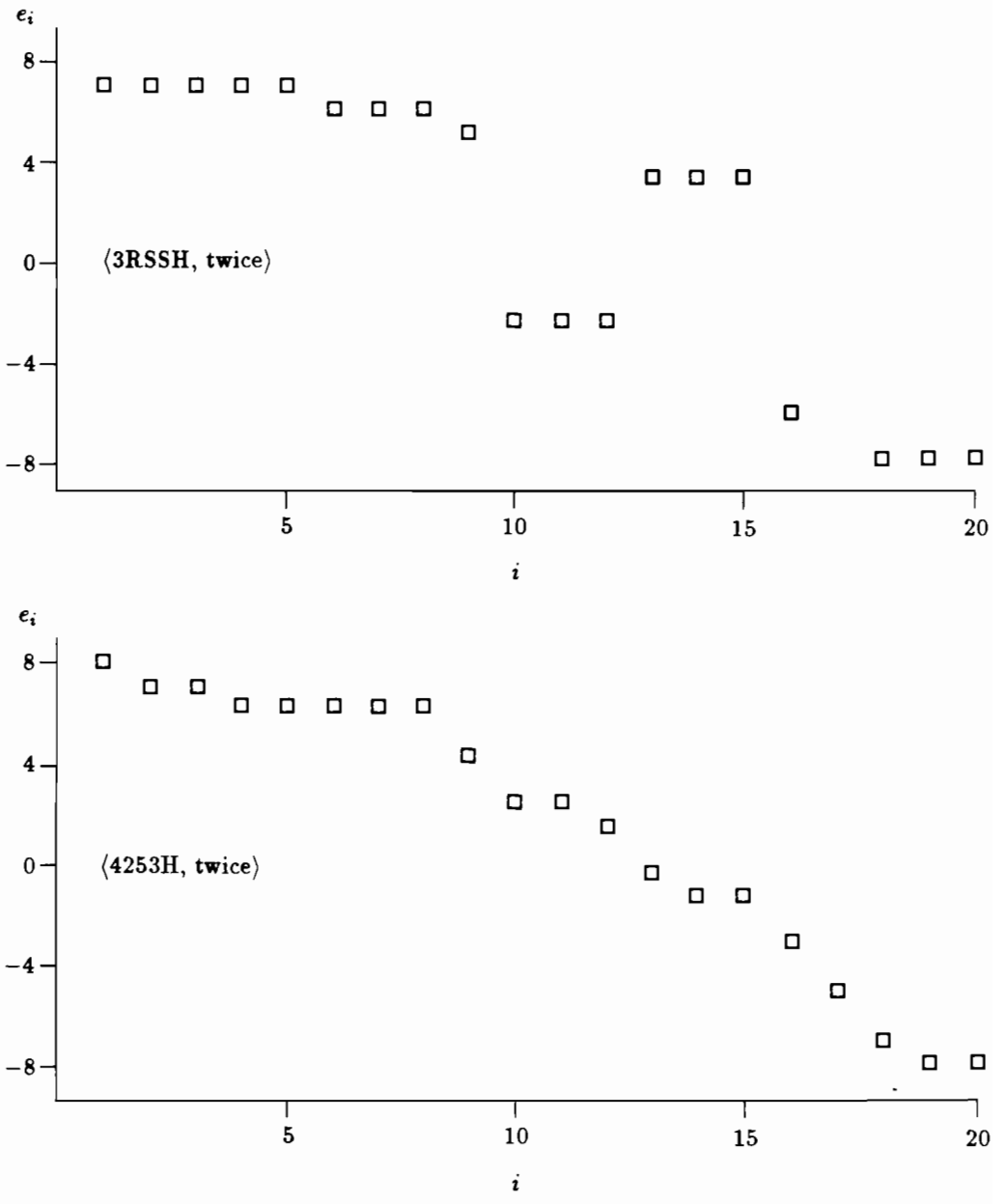

Figure 11.5: Scatterplot of the OLS residuals e_i smoothed by the ⟨3RSSH, twice⟩ and ⟨4253H, twice⟩ techniques, against the index i for Quandt's data.

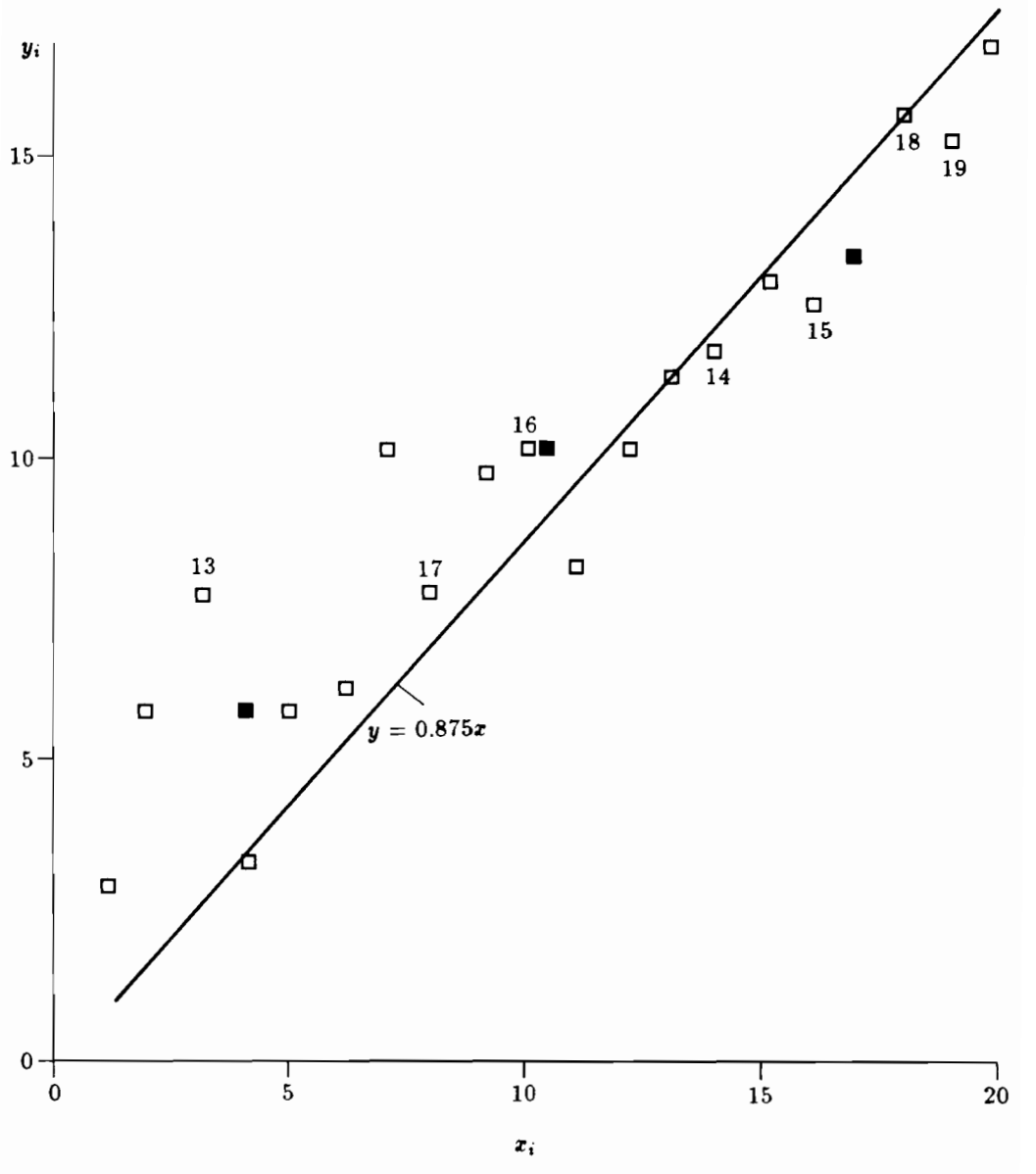

Figure 11.6: Scatterplot of Quandt's data; the solid points are median summary points for the left, middle, and right thirds of the data according to the order of the x-values.

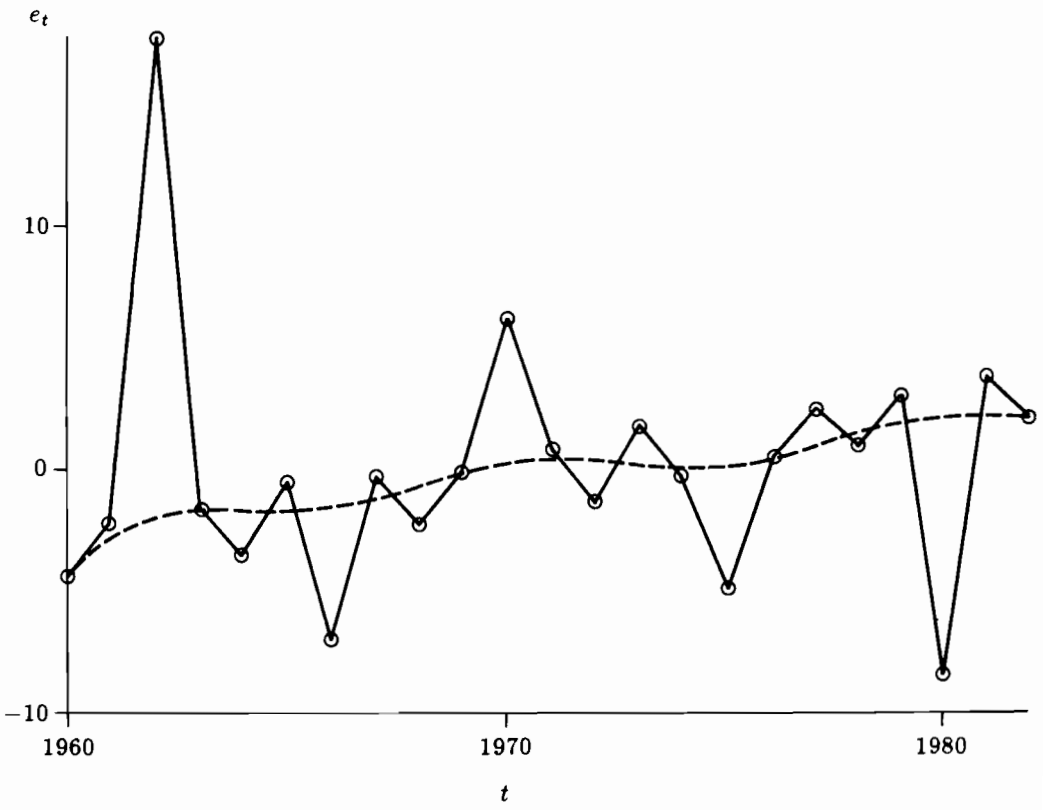

Figure 11.7: OLS residuals e_t and smoothed OLS residuals against the year t of the observation for example 3.

Figures 11.4–11.6 are based on the data set given by Quandt (1958) which contains a structural change by the 13th observation. The smoothed OLS residuals (*Figure 11.4*) indicate a systematic decreasing tendency, starting at the 8th–10th observations. *Figure 11.5* shows the results of applying the smoothing techniques ⟨4253H, twice⟩ and ⟨3RSSH, twice⟩ on these data. The smoothed residuals obtained by the first of these techniques (the upper plot in *Figure 11.5*) form two different configurations on the plot, separated by the points 10, 11, and 12. The ocurrence of sequences of three points with the same level is due to the repeated resmoothing used in this technique. Additional valuable information concerning the structure of the data yields a simple scatterplot of y values versus x values, which also contains the corresponding linear regression line (see *Figure 11.6*). Numbering the points on this plot, one can see that the observations 13–20 probably would yield a slightly different fitted line when used alone.

The last two examples are based on real data from the Austrian economy. *Figures 11.7–*

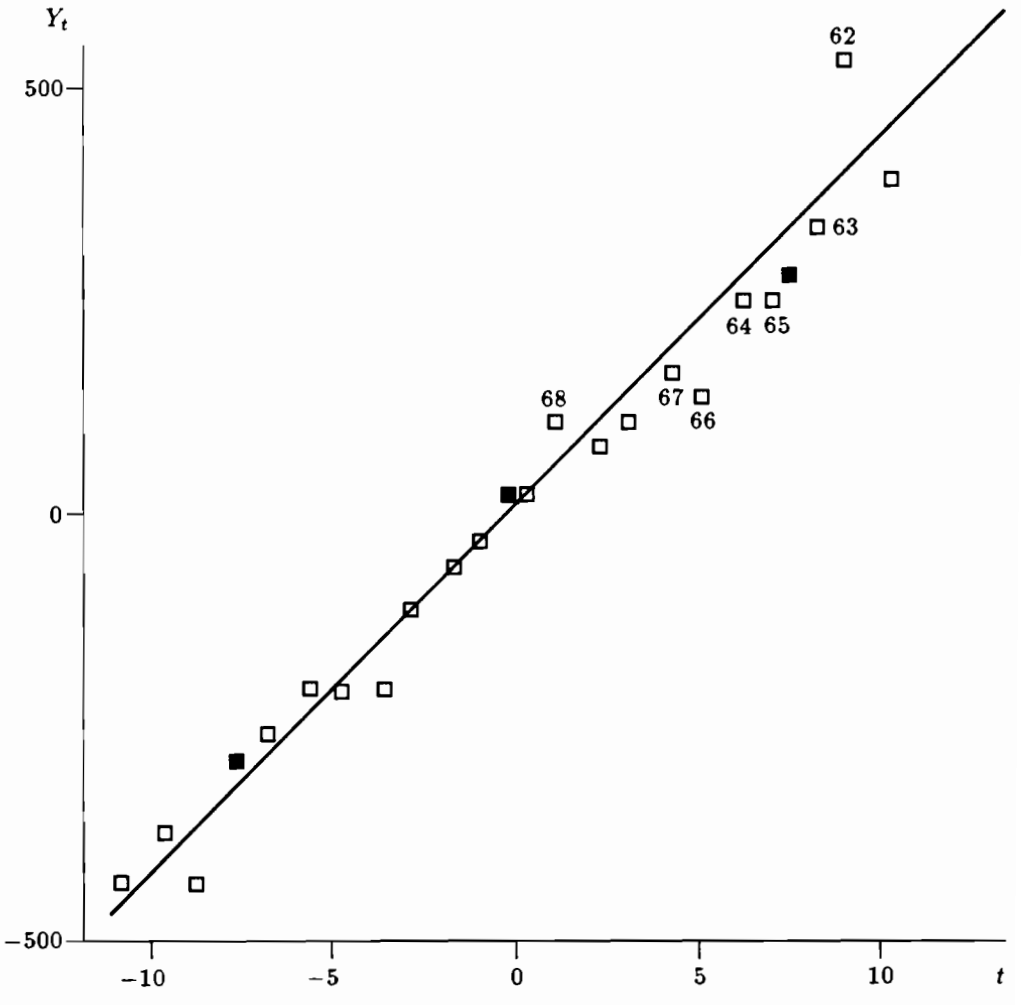

Figure 11.8: Leverage plot for the trend variable t and fitted line for example 3.

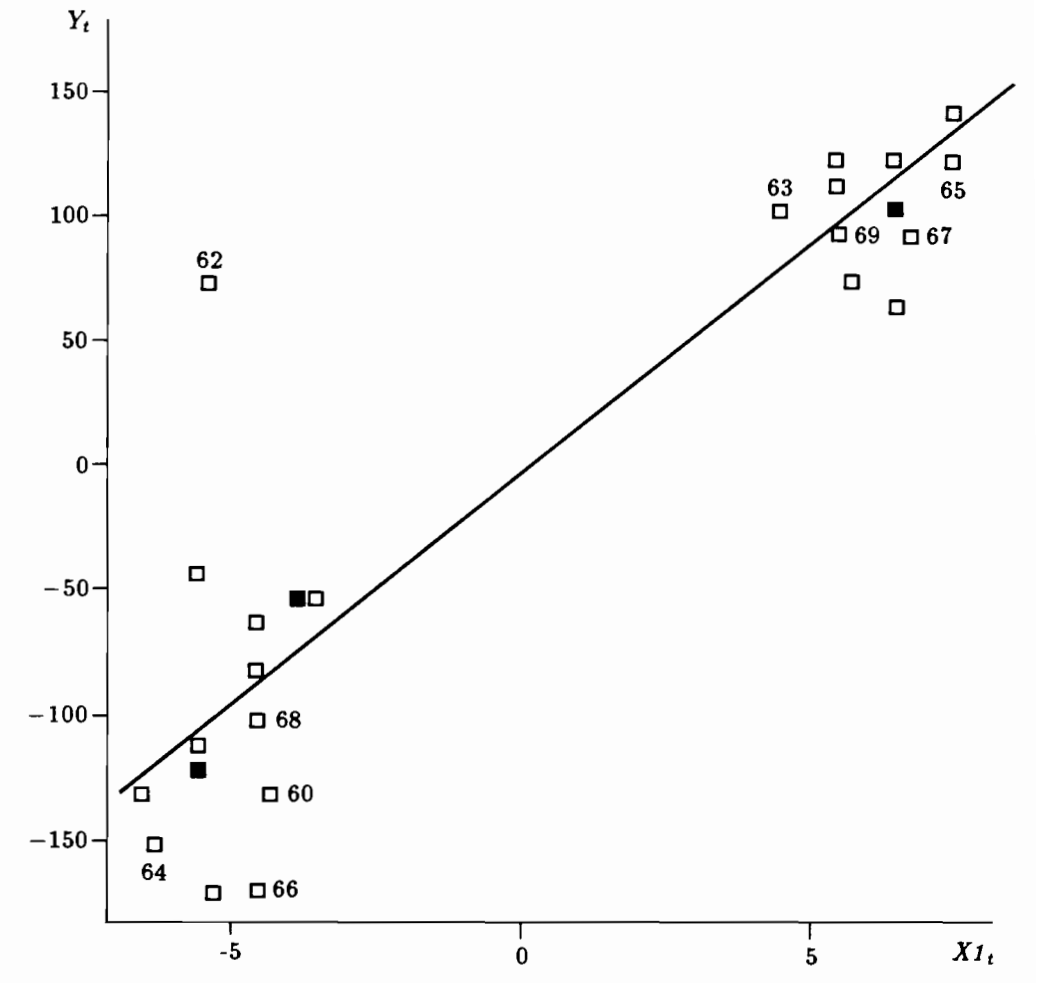

Figure 11.9: Leverage plot for the variable $X1_t$ and fitted line for example 3.

11.9 are based on the data used in examples given in the test processor of the IAS-SYSTEM (see Sonnberger *et al.*, 1986). The 23 observations from 1960 to 1982 were used to estimate the parameters of the equation

$$Y_t = a_1 + a_2 t + a_3 X1_t + u_t$$

The Quandt test indicates a change of the parameters in 1964. The CUSUM test does not reject the null hypothesis that the parameters remain constant over time. The CUSUM-SQ test indicates two regimes before and after 1964. The same phenomenon is indicated by Ploberger's (1983) fluctuation test.

Examination of the scatterplot for OLS residuals and their smoothed values (the ⟨4253H, twice⟩ technique was used) does not show any significant change in the configuration of the residuals (see *Figure 11.7*). The smoothed curve seems to rise a bit, starting in 1965 but this may be due to the fact that there is an outlier in 1962 and probably also in 1980. Analysis of the partial regression residuals (or "leverage") plots for the two exogenous variables (*Figures 11.8* and *11.9*) allows us to look for any special subconfiguration of the given subset of successive data points. The leverage plot for the variable $X1$ only indicates that the observations from 1964 to 1968 would lead to different parameter estimates than the global estimation process.

The graphical analysis for the last example is based on the following equation for demand for labor in the Austrian building industry given in Sonnberger *et al.* (1983):

$$Y_t = b_0 + b_1 t + b_2 X2_t + u_t$$

where Y is the logarithmic transform of employment in the construction sector and $X2$ is the logarithmic transform of the real output in the construction sector. A scatterplot of the externally studentized residuals (*Figure 11.10*) shows not only an extremely negative residual for 1975, but also indicates two other relatively high positive residuals for 1967 and 1973. The same plot indicates a possible break point in 1972 or 1973. [In fact, it is possible to distinguish two subsequences of increasing residuals within the loop between 1973 and 1974. The smoothed values of residuals show the change in tendency of the residuals' distribution.]

The scatterplot of the backward recursive residuals plotted against the forward recursive residuals (*Figure 11.11*) confirms this supposition. There is no linear dependency between these two types of recursive residuals. This means that predictions made from two different edges of the sample do not correspond to each other. *Figure 11.12* and *11.13* show that there are no single influential points playing a crucial role in the determination of the parameter estimates as well as extremely outlying observations. The effect of the possible outlier in 1975 seems negligible, and influential points in 1967 and 1973 offset each other's leverage on the estimated coefficients. The possible occurrence of the break point in 1972–1973 is no longer quite clear in the light of *Figure 11.12* and *11.13*. The observations for 1972–1978 might lead to a different slope of the regression line in these

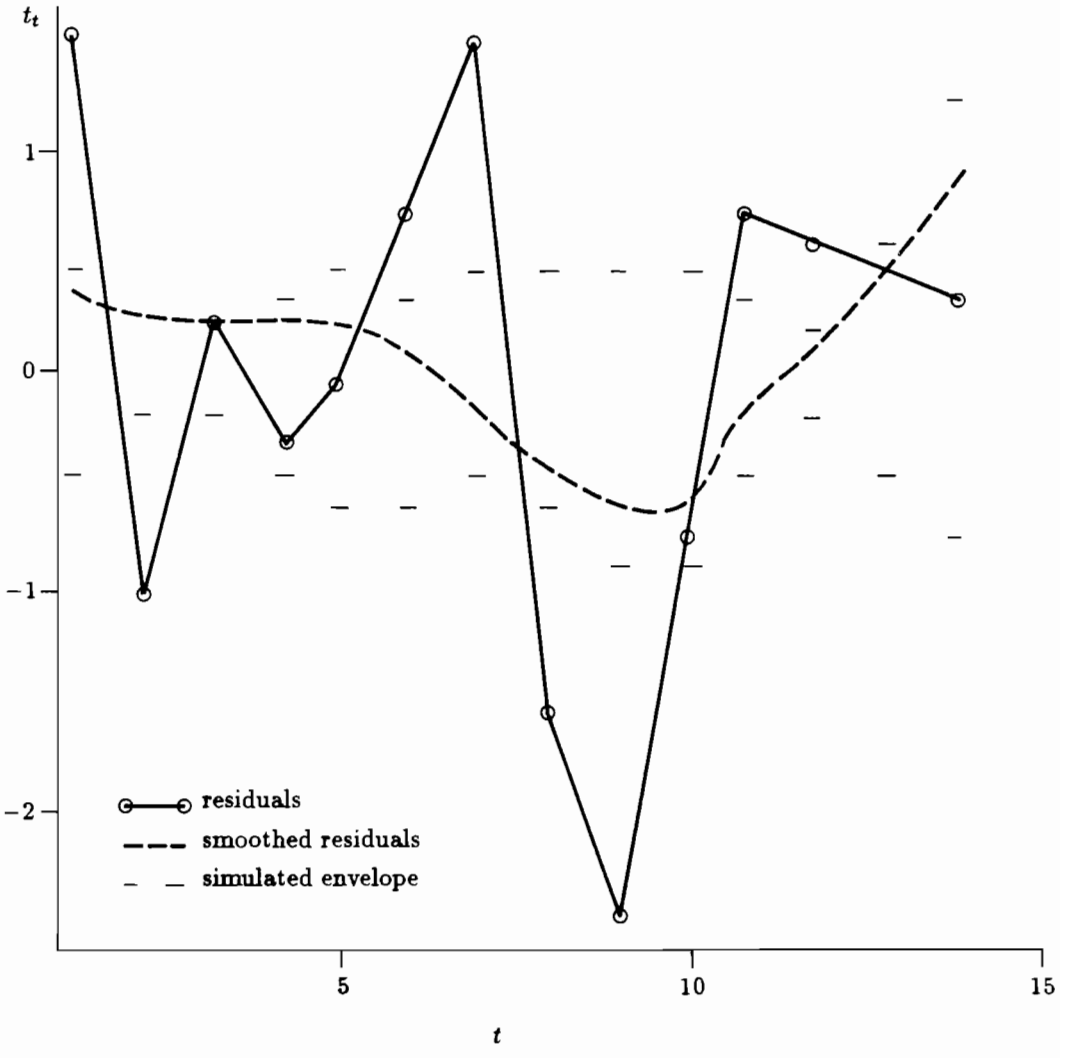

Figure 11.10: Scatterplot of externally studentized residuals t_t and smoothed values against the index t for example 4.

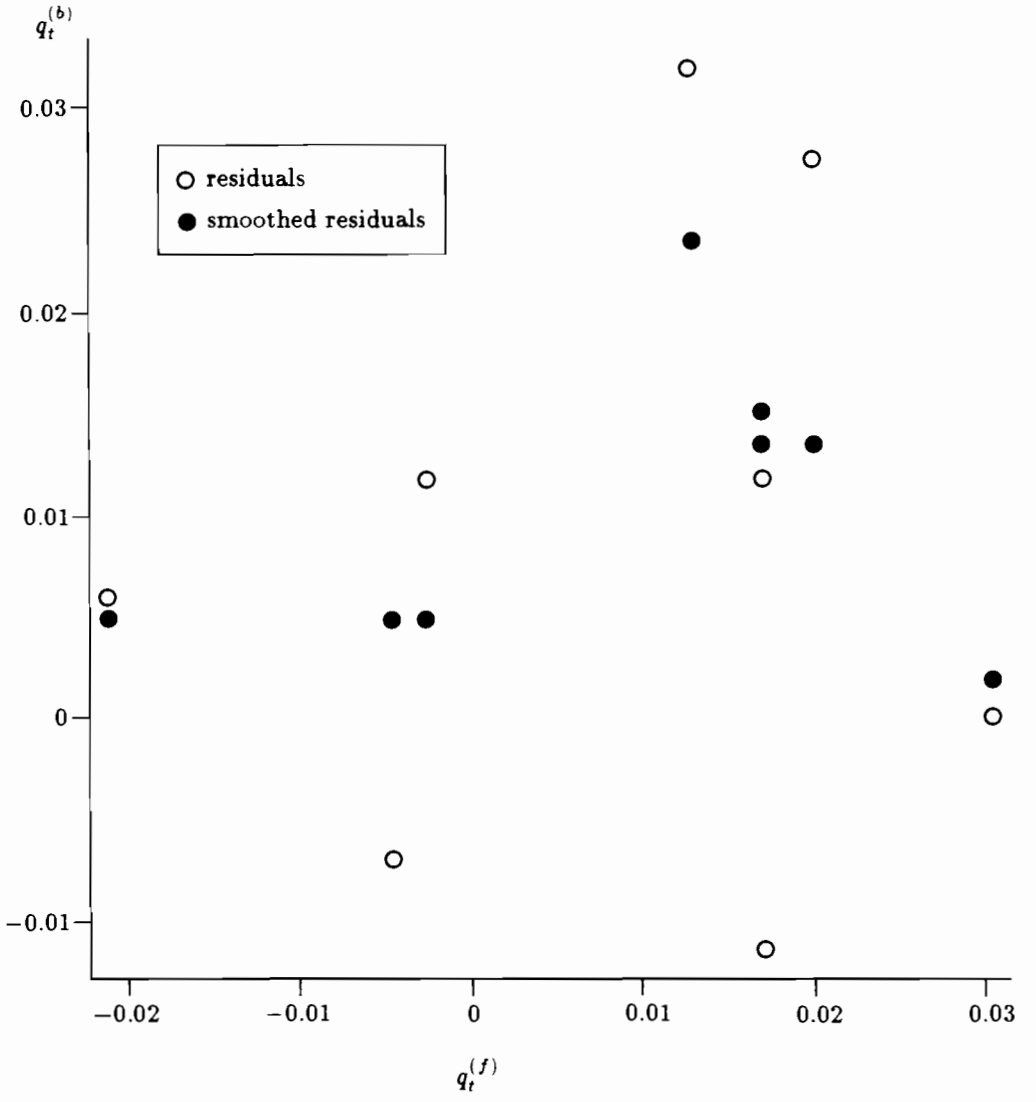

Figure 11.11: Scatterplot of backward $[q_t^{(b)}]$ against forward $[q_t^{(b)}]$ recursive residuals for example 4.

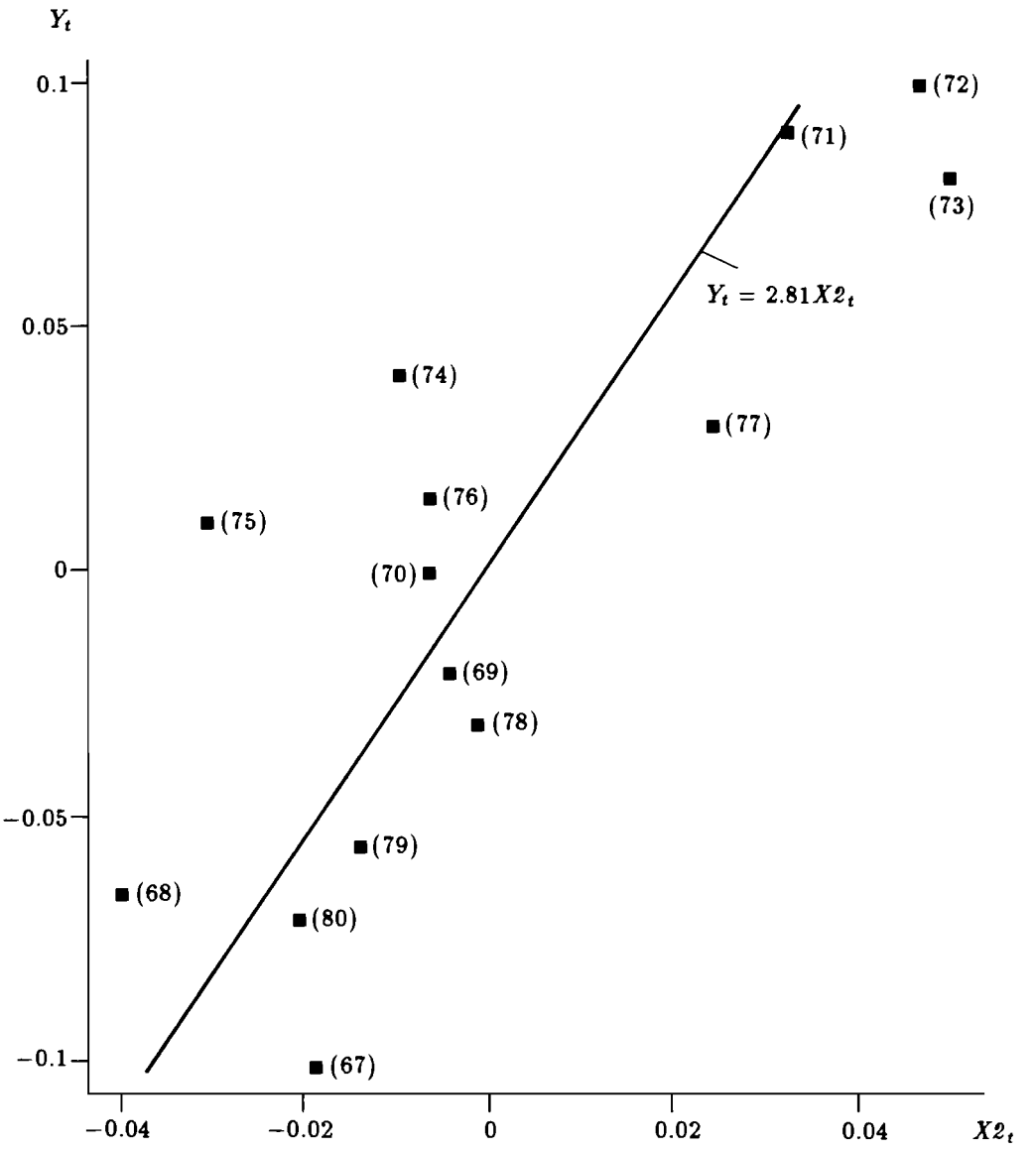

Figure 11.12: Leverage plot for variable $X2$ and fitted line for example 4.

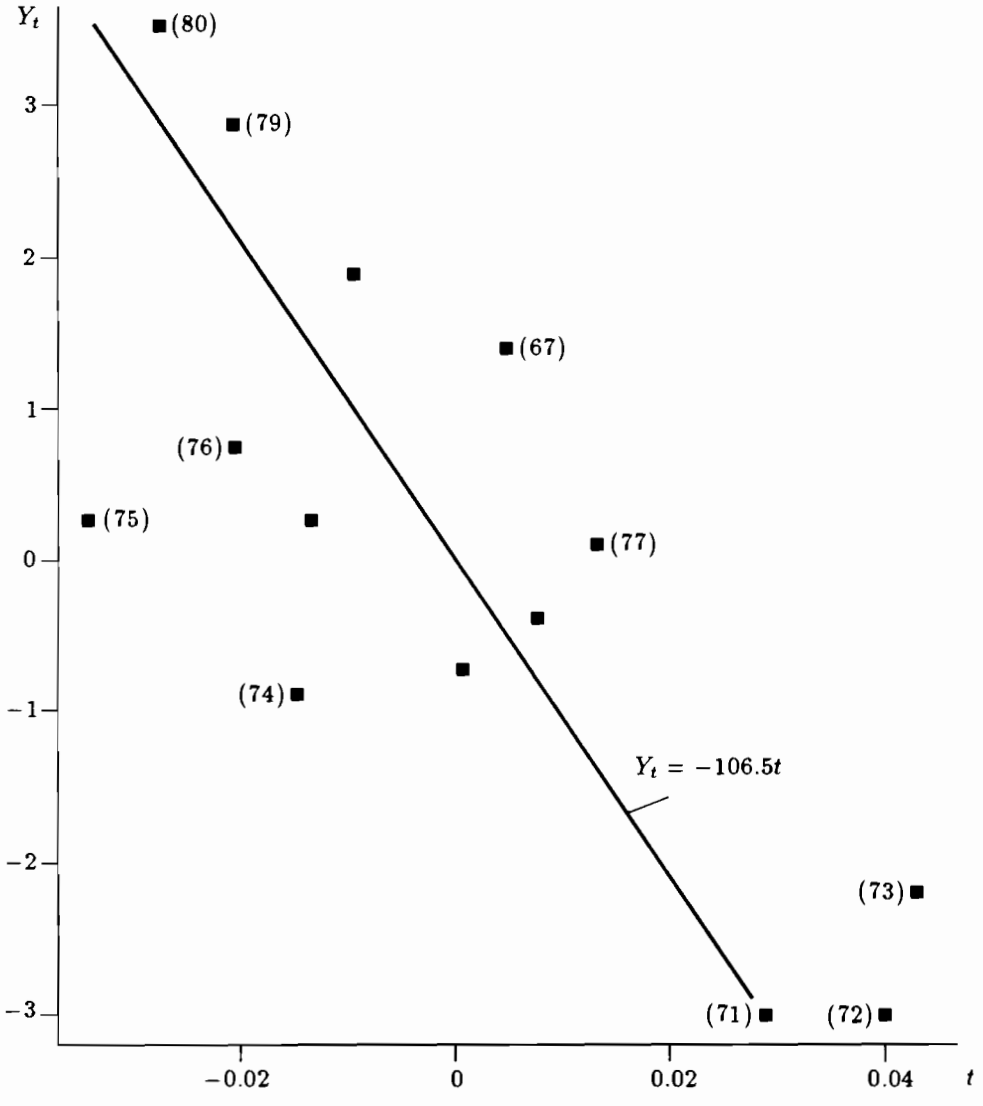

Figure 11.13: Leverage plot for the trend variable t and fitted line for example 4.

figures when considered separately. [This slope is equal to the value of the coefficient's estimate by variable t and $X2$, respectively.] The observations in 1979 and 1980 still remain in agreement with the old line. This can be interpreted as the return to the old state after a temporary change of the regression regime which occurred in 1972.

11.5 Final Remarks

The four examples considered in Section 11.4 give an idea of the usefulness of graphical displays in residual analysis. They provide important qualitative information about the data structure and its effects in the estimation process. The plots of the smoothed residuals show the general tendency in the residual configuration; and with partial leverage regression plots, one can recognize influential subsets of data. These two features seem to be especially important in the investigation of structural change.

References

Anscombe, F.J. (1973), Graphs in Statistical Analysis. *The American Statistician*, **27** (1), 17-21.

Anscombe, F.J. and Tukey, J.W. (1963), The examination and analysis of residuals. *Technometrics*, **5**, 141-160.

Atkinson, A.C. (1981), Robustness, transformations and two graphical displays for outlying and influential observations in regression. *Biometrika*, **68**, 13-20.

Atkinson, A.C. (1982), Regression diagnostics, transformations and constructed variables. *Journal of the Royal Statistical Society*, **B-44**, 1-36.

Bartlett, M.S. (1951), An inverse matrix adjustment arising in discriminant analysis. *Annals of Mathematical Statistics*, **22**, 107-111.

Belsley, D.A., Kuh, E., and Welsch, R.E. (1980), *Regression Diagnostics.* New York: John Wiley.

Brown, R.L., Durbin, J., and Evans, J.M. (1975), Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society*, **B-37**, 149-192.

Cleveland, W.S. (1979), Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, **74**, 829-836.

Cleveland, W.S. and Kleiner, J. (1975), A graphical technique for enhancing scatterplots with moving statistics. *Technometrics*, **17**, 447-454.

Cleveland, W.S. and McGill, R. (1984), The many faces of the scatterplot. *Journal of the American Statistical Association*, **79**, 531-554.

Cook, R.D. and Weisberg, S. (1982), *Residuals and Influence in Regression.* New York: Chapman and Hill.

Daniel, C. and Wood, F.S. (1971), *Fitting Equations to Data.* New York: John Wiley.

Galpin, J.S. and Hawkins, D.M. (1984), The use of recursive residuals in checking model fit in linear regression. *The American Statistician*, **38** (2), 94-105.

Gnanadesikan, R. and Wilk, M.B. (1968), Probability plotting methods for the analysis of data. *Biometrika*, **55**, 1-17.

Mosteller, F. and Tukey, J.W. (1977), *Data Analysis and Regression.* Reading, MA: Addison-Wesley.

Quandt, R.E. (1958), The estimation of the parameters of a linear regression system obeing two separate regimes. *Journal of the American Statistical Association*, **53**, 873-880.

Plackett, R.L. (1950), Some theorems in least squares. *Biometrika*, **37**, 149-157.

Ploberger, W. (1983), Testing the constancy of parameters in linear models. Technical Report, Vienna: Technical University.

Sonnberger, H., Havlik, P., Krämer, W., and Maurer, J. (1983), *Testing for Autocorrelation and Model Specification with the IAS-SYSTEM.* Research Memorandum 186. Vienna: Institute for Advanced Studies.

Sonnberger, H., Krämer, W., Maurer, J., and Reschenhofer, E. (1986), *IAS-SYSTEM Reference Manual. Part 2: Estimation and Testing.* Vienna: Institute for Advanced Studies.

Tukey, J.W. (1971), *Exploratory Data Analysis.* Reading, MA: Addison-Wesley.

Velleman, P.F. and Hoaglin, D.C. (1981), *Applications, Basics, and Computing of Exploratory Data Analysis.* Boston: Duxbury Press.

Part III

Model Building in the Presence of Structural Change

CHAPTER 12

Adaptive Estimation and Structural Change in Regression and Time Series Models

Johannes Ledolter

Summary

Heuristic and model-based approaches to adaptive estimation in regression models are reviewed in this chapter. We describe a model-based approach that introduces time-varying coefficients explicitly and assumes that the coefficients follow certain autoregressive integrated moving average time series processes. We show how these time-varying coefficient models can be written in state space form, we illustrate how the Kalman filter approach can be used to update the coefficient estimates and forecasts, and we discuss why the resulting estimates are more responsive to structural change than the standard least squares estimates. The parameters in the underlying stochastic processes that generate the time-varying coefficients are needed to update the coefficient estimates. It is shown how these parameters can be estimated from historic observations. These parameters determine how adaptive the resulting coefficient estimates are to changes in the coefficients.

12.1 Introduction

The coefficients in statistical models (such as regression, time series, transfer-function, ARMAX, and econometric models) are usually assumed to be *constant*. Under the assumption of parametric stationarity, statisticians have developed methods for the efficient estimation of the underlying parameters; for example, least squares, maximum likelihood, or Bayesian methods can be used to estimate the underlying coefficients.

In applied work it is often recognized and anticipated that relationships change over time. For example, the relationships before and after a certain event or intervention may differ; the behavioral characteristics may change and drift with time; and linear approximations to complex, nonlinear, and poorly defined phenomena may exhibit time-varying structures.

If the parameters are expected to shift at a given point in time from one value, say β, to some other value, say $\beta^* = \beta + \delta$, then it is usually quite easy to formulate a more general model, estimate the change in the coefficients, and test whether the change is significant. For example, one can use F- and t-statistics to test the equality between two sets of regression coefficients; see Chow (1960). But, if the time period at which the coefficients change is unknown, the problem becomes more difficult as one has to make additional inferences to detect the unknown change point. There are many approaches to this problem, and many papers have been written on this topic. The reader may refer to the extensive bibliographies compiled by Shaban (1980) and Hackl and Westlund (1985). Papers on *two-phase* or *switching regression* [see, for example, Quandt (1958, 1960, 1972), Goldfeld and Quandt (1973, 1976), Poirier (1976), Hinkley (1969, 1971)]; *Bayesian methods* for change detection [see Ferreira (1975), Broemeling (1985, Chapter 7)], and various *cumulative sum procedures* [see Brown *et al.* (1975), Hackl (1980)] fall into this category.

Parameters may also change *continuously over time*, and this is the topic that is addressed in this chapter. Instead of assuming that the coefficients are constant or shift from one value to another (step changes in the coefficients), we assume that they drift and vary continuously over time. *Recursive estimation* plays an important role as recursive parameter estimates provide information on the existence of nonstationarity. Furthermore, it is quite easy to modify recursive estimation procedures such that more weight is given to the most recent observations and less to the ones in the distant past. Modifications may be heuristic, such as various weighted least squares approaches where the weights decrease with the age of the observations. Or, one can assume that the parameters follow certain stochastic models. For example, one can assume that the coefficients change smoothly according to a random walk, $\beta_t = \beta_{t-1} + \mathbf{v}_t$, where the $\mathbf{v}_t$ are independent disturbances, and can derive at each time period t the optimal estimates of the coefficients β_t in this more general model. Also, one may want to test whether the coefficients are in fact constant or whether they exhibit time-variability.

12.2 Heuristic Approaches to Recursive Estimation in Regression Models

Let us consider the multiple linear regression model

$$y_t = x_{t1}\beta_1 + \ldots + x_{tp}\beta_p + a_t = \mathbf{x}_t'\beta + a_t \tag{12.1}$$

Here y_t, for $t = 1, \ldots, n$, is the response in an experiment where the levels of the p explanatory (or regressor) variables are set at $x_{t1}, \ldots, x_{tp}$. If an intercept term is desired in

the model, one sets $x_{t1} = 1$. The disturbance terms a_t, for $t = 1, \ldots, n$, are independent random variables with mean zero and constant variance σ^2. The vectors $\mathbf{x}_t$ and β in model (12.1) are $p \times 1$ column vectors, defined as $\mathbf{x}_t = (x_{t1}, \ldots, x_{tp})'$ and $\beta = (\beta_1, \ldots, \beta_p)'$. The levels of the explanatory variables are assumed to be either nonrandom constants or random variables which are independent of the error terms a_t. In this chapter we assume that t stands for time or run order. This means that the responses become available sequentially in time. For example, they may refer to monthly, quarterly, or annual observations on certain economic indicators or other variables of interest.

The coefficients in the regression model (12.1) can be estimated by least squares such that the sum of the squared deviations $S(\beta) = \sum_{t=1}^{n}(y_t - \mathbf{x}_t'\beta)^2$ is minimized. These least squares estimates (which are maximum likelihood estimates if the errors are assumed normal) can be written in matrix form as

$$\hat{\beta}_n = (X_n'X_n)^{-1}X_n'\mathbf{y}_n \tag{12.2}$$

where X_n is an $(n \times p)$ design matrix that consists of the levels of the explanatory variables, and $\mathbf{y}_n$ is an $(n \times 1)$ column vector of the responses. More specifically,

$$X_n = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \quad , \qquad \mathbf{y}_n = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

12.2.1 Recursive least squares

Let us assume that the observations become available sequentially in time, and let us express the least squares estimate $\hat{\beta}_n$ as a function of the previous estimate at time $n-1$, that is $\hat{\beta}_{n-1}$, and the new information at time n. We can write the least squares estimate as

$$\hat{\beta}_n = P_n(X_{n-1}'\mathbf{y}_{n-1} + \mathbf{x}_n y_n)$$

where

$$\begin{aligned} P_n &= (X_n'X_n)^{-1} = (X_{n-1}'X_{n-1} + \mathbf{x}_n\mathbf{x}_n')^{-1} = (P_{n-1}^{-1} + \mathbf{x}_n\mathbf{x}_n')^{-1} \\ &= P_{n-1} - P_{n-1}\mathbf{x}_n(1 + \mathbf{x}_n'P_{n-1}\mathbf{x}_n)^{-1}\mathbf{x}_n'P_{n-1} \end{aligned} \tag{12.3}$$

Here we have used a well-known result about the inverses of symmetric matrices; for example, see page 29 of Rao (1965). Thus,

$$\begin{aligned}
\hat{\beta}_n &= [P_{n-1} - P_{n-1}\mathbf{x}_n(1 + \mathbf{x}_n' P_{n-1}\mathbf{x}_n)^{-1}\mathbf{x}_n' P_{n-1}](X_{n-1}'\mathbf{y}_{n-1} + \mathbf{x}_n y_n) \\
&= \hat{\beta}_{n-1} - P_{n-1}\mathbf{x}_n(1 + \mathbf{x}_n' P_{n-1}\mathbf{x}_n)^{-1}\mathbf{x}_n'\hat{\beta}_{n-1} + P_{n-1}\mathbf{x}_n y_n \\
&\qquad - P_{n-1}\mathbf{x}_n(1 + \mathbf{x}_n' P_{n-1}\mathbf{x}_n)^{-1}\mathbf{x}_n' P_{n-1}\mathbf{x}_n y_n \\
&= \hat{\beta}_{n-1} + P_{n-1}\mathbf{x}_n(1 + \mathbf{x}_n' P_{n-1}\mathbf{x}_n)^{-1}[-\mathbf{x}_n'\hat{\beta}_{n-1} \\
&\qquad + (1 + \mathbf{x}_n' P_{n-1}\mathbf{x}_n)y_n - \mathbf{x}_n' P_{n-1}\mathbf{x}_n y_n] \\
&= \hat{\beta}_{n-1} + P_{n-1}\mathbf{x}_n(1 + \mathbf{x}_n' P_{n-1}\mathbf{x}_n)^{-1}[y_n - \mathbf{x}_n'\hat{\beta}_{n-1}] \\
&= \hat{\beta}_{n-1} + P_n\mathbf{x}_n[y_n - \mathbf{x}_n'\hat{\beta}_{n-1}] \qquad (12.4)
\end{aligned}$$

since

$$\begin{aligned}
P_n\mathbf{x}_n &= [P_{n-1} - P_{n-1}\mathbf{x}_n(1 + \mathbf{x}_n' P_{n-1}\mathbf{x}_n)^{-1}\mathbf{x}_n' P_{n-1}]\mathbf{x}_n \\
&= P_{n-1}\mathbf{x}_n\{1 - [\mathbf{x}_n' P_{n-1}\mathbf{x}_n/(1 + \mathbf{x}_n' P_{n-1}\mathbf{x}_n)]\} \\
&= P_{n-1}\mathbf{x}_n(1 + \mathbf{x}_n' P_{n-1}\mathbf{x}_n)^{-1}
\end{aligned}$$

Equations (12.3) and (12.4) give us a recursive method for calculating and updating the least squares estimates. We can start with the first p equations of model (12.1) for $t = 1, \ldots, p$, and calculate $\hat{\beta}_p = P_p X_p'\mathbf{y}_p$. This estimate implies a certain fitted, or predicted, value for the next response y_{p+1}, that is $\hat{y}_{p+1} = \mathbf{x}_{p+1}'\hat{\beta}_p$. The one-step-ahead prediction error, $y_{p+1} - \mathbf{x}_{p+1}'\hat{\beta}_p$, is used to revise the estimate, and $\hat{\beta}_{p+1}$ is obtained. The vector of weights that is attached to the prediction error is given in equation (12.4). The matrix P_n is revised according to the relationship in equation (12.3). The updating relationships are very convenient as they involve no further matrix inversion. These equations are very intuitive as they use the latest prediction error to revise the estimate that is obtained at the previous stage. Also, the elements of the covariance matrix of the least squares estimates,

$$Var(\hat{\beta}_n) = \sigma^2(X_n' X_n)^{-1} = \sigma^2 P_n \qquad (12.5)$$

are easily updated from equation (12.3).

For nondegenerate explanatory variables (that is, when the elements of $\mathbf{x}_t$ are not zero all the time) the elements of the matrix $P_n = (X_n' X_n)^{-1}$ and of the adjustment vector $P_n\mathbf{x}_n$ in (12.4) approach zero. This illustrates that, eventually, when the sample size has become very large, the coefficients will be virtually unchanged and unaffected by a new observation.

These recursive updating equations have been known for a long time; see Plackett (1950) for a discussion of recursive least squares. As a matter of fact, the development of least squares theory in its recursive form goes back to Gauss and to the early nineteenth century [see Appendix 2 of Young (1984)].

12.2.2 Heuristic approaches to the estimation of time-varying coefficients

The least squares criterion gives each pair of observations $(\mathbf{x}_t, y_t)$ the same weight. This implies that after a while each new observation will change the least squares estimate only little. If observations are collected sequentially and if one believes that the recent observations are more relevant than the ones in the past, then one should weight the observations differently.

We can use one of several heuristic approaches. Under one approach, we estimate the coefficients at time n only from the $s \geq p$ most recent observations; that is, those observed at times $n, n-1, \ldots, n-s+1$. This amounts to using a moving rectangular window function. The other approach uses an exponential data weighting function and discounts the observations according to their age. This amounts to minimizing the weighted sum of squares $S_\star(\beta) = \sum_{t=1}^{n} \omega^{n-t}[y_t - \mathbf{x}_t'\beta]^2$, where $\omega = 1 - \alpha$ is a discount coefficient that is between zero and one; α is called the smoothing constant. For example, one could pick $\omega = 0.95$.

Moving rectangular window function

If we estimate the parameters from only the last s time periods, the parameter estimate at time n satisfies the normal equations

$$[\sum_{t=n-s+1}^{n} \mathbf{x}_t\mathbf{x}_t']\hat{\beta}_n = [\sum_{t=n-s+1}^{n} \mathbf{x}_t y_t] \tag{12.6}$$

We obtain this estimate from the one at time $n-1$, $\hat{\beta}_{n-1}$, by adding the new observation $(\mathbf{x}_n, y_n)$ and omitting the observation $(\mathbf{x}_{n-s}, y_{n-s})$. In equations (12.3) and (12.4) we learned how to revise estimates after adding a new observation. Adding the new observation $(\mathbf{x}_n, y_n)$ leads to the estimates

$$\hat{\beta}_n^\star = \hat{\beta}_{n-1} + P_{n-1}\mathbf{x}_n(1 + \mathbf{x}_n'P_{n-1}\mathbf{x}_n)^{-1}(y_n - \mathbf{x}_n'\hat{\beta}_{n-1}) \tag{12.7}$$

$$P_n^\star = P_{n-1} - P_{n-1}\mathbf{x}_n(1 + \mathbf{x}_n'P_{n-1}\mathbf{x}_n)^{-1}\mathbf{x}_n'P_{n-1} \tag{12.8}$$

Removing the observation $(\mathbf{x}_{n-s}, y_{n-s})$ leads to the revised new estimate

$$\begin{aligned}
\hat{\beta}_n &= (\sum_{t=n-s}^{n} \mathbf{x}_t\mathbf{x}_t' - \mathbf{x}_{n-s}\mathbf{x}_{n-s}')^{-1}(\sum_{t=n-s}^{n} \mathbf{x}_t y_t - \mathbf{x}_{n-s}y_{n-s}) \\
&= [(P_n^\star)^{-1} - \mathbf{x}_{n-s}\mathbf{x}_{n-s}']^{-1}(\sum_{t=n-s}^{n} \mathbf{x}_t y_t - \mathbf{x}_{n-s}y_{n-s}) \\
&= [P_n^\star + P_n^\star\mathbf{x}_{n-s}(1 - \mathbf{x}_{n-s}'P_n^\star\mathbf{x}_{n-s})^{-1}\mathbf{x}_{n-s}'P_n^\star](\sum_{t=n-s}^{n} \mathbf{x}_t y_t - \mathbf{x}_{n-s}y_{n-s}) \\
&= \hat{\beta}_n^\star + P_n^\star\mathbf{x}_{n-s}(\mathbf{x}_{n-s}'P_n^\star\mathbf{x}_{n-s} - 1)^{-1}(y_{n-s} - \mathbf{x}_{n-s}'\hat{\beta}_n^\star)
\end{aligned} \tag{12.9}$$

and

$$P_n = P_n^\star - P_n^\star \mathbf{x}_{n-s}(\mathbf{x}'_{n-s}P_n^\star \mathbf{x}_{n-s} - 1)^{-1}\mathbf{x}'_{n-s}P_n^\star \tag{12.10}$$

The four equations (12.7)–(12.10) are used recursively to update the estimates $\hat{\beta}_n$.

Exponential window function

The weighted least squares estimator that minimizes $\sum_{t=1}^{n} \omega^{n-t}(y_t - \mathbf{x}'_t\beta)^2$ satisfies the normal equation

$$[\sum_{t=1}^{n} \omega^{n-t}\mathbf{x}_t\mathbf{x}'_t]\,\hat{\beta}_n = \sum_{t=1}^{n} \omega^{n-t}\mathbf{x}_t y_t \tag{12.11}$$

It is given by

$$\hat{\beta}_n = P_n[\sum_{t=1}^{n} \omega^{n-t}\mathbf{x}_t y_t] = P_n[\omega \sum_{t=1}^{n-1} \omega^{n-1-t}\mathbf{x}_t y_t + \mathbf{x}_n y_n] \tag{12.12}$$

where

$$\begin{aligned} P_n &= [\sum_{t=1}^{n} \omega^{n-t}\mathbf{x}_t\mathbf{x}'_t]^{-1} = [\omega \sum_{t=1}^{n-1} \omega^{n-1-t}\mathbf{x}_t\mathbf{x}'_t + \mathbf{x}_n\mathbf{x}'_n]^{-1} = [\omega P_{n-1}^{-1} + \mathbf{x}_n\mathbf{x}'_n]^{-1} \\ &= \frac{1}{\omega}[P_{n-1} - P_{n-1}\mathbf{x}_n(\omega + \mathbf{x}'_n P_{n-1}\mathbf{x}_n)^{-1}\mathbf{x}'_n P_{n-1}] \end{aligned} \tag{12.13}$$

after applying the matrix inversion result in equation (12.3). Substituting (12.13) into the expression for $\hat{\beta}_n$, we find that

$$\begin{aligned} \hat{\beta}_n &= \hat{\beta}_{n-1} + P_{n-1}\mathbf{x}_n(\omega + \mathbf{x}'_n P_{n-1}\mathbf{x}_n)^{-1}[-\mathbf{x}'_n P_{n-1} \sum_{t=1}^{n-1} \omega^{n-1-t}\mathbf{x}'_t y_t \\ &\quad + (1/\omega)(\omega + \mathbf{x}'_n P_{n-1}\mathbf{x}_n)y_n - (1/\omega)\mathbf{x}'_n P_{n-1}\mathbf{x}_n y_n] \\ &= \hat{\beta}_{n-1} + P_{n-1}\mathbf{x}_n(\omega + \mathbf{x}'_n P_{n-1}\mathbf{x}_n)^{-1}(y_n - \mathbf{x}'_n\hat{\beta}_{n-1}) \\ &= \hat{\beta}_{n-1} + P_n\mathbf{x}_n(y_n - \mathbf{x}'_n\hat{\beta}_{n-1}) \end{aligned} \tag{12.14}$$

since

$$\begin{aligned} P_n\mathbf{x}_n &= \frac{1}{\omega}[P_{n-1} - P_{n-1}\mathbf{x}_n(\omega + \mathbf{x}'_n P_{n-1}\mathbf{x}_n)^{-1}\mathbf{x}'_n P_{n-1}]\,\mathbf{x}_n \\ &= P_{n-1}\mathbf{x}_n(\omega + \mathbf{x}'_n P_{n-1}\mathbf{x}_n)^{-1} \end{aligned}$$

We can use the equations in (12.13) and (12.14) recursively to update the coefficient estimates. We can start with the estimates $\hat{\beta}_p$ and P_p from the first p equations and update the subsequent estimates in a recursive fashion. The updating equations are similar to the ones for least squares estimates in equations (12.3) and (12.4), except that the recursion for P_n includes the discount coefficient ω.

We found that for least squares estimates the elements of the matrix P_n in (12.3) and the elements in the adjustment vector $P_n\mathbf{x}_n$ in (12.4) approach zero as the sample size n goes toward infinity. For an exponential data weighting function, this is no longer the case. There P_n approaches a steady state limit that is different from the zero matrix. This implies that some adjustment will be done, even though a large number of observations may already have been processed. Thus, these estimates are more responsive to changes in the underlying coefficients. They allow for coefficient variability.

One disadvantage with this scheme is that all p coefficients are treated the same. It is not possible to have one coefficient vary more than the other. A second disadvantage is that one has to pick a discount coefficient. This coefficient can be chosen heuristically (for example, one can set ω equal to 0.95 or 0.90), or one can use historic data and select it such that the sum of the squared one-step-ahead forecast errors is as small as possible.

Special case: Regression on functions of time

It is instructive to look at these updating equations for the special case when the explanatory variables in the regression model are functions of time. Let us consider models of the form

$$y_{n+j} = \beta_0 + \beta_1 j + \ldots + \beta_{p-1} j^{p-1} + a_{n+j} = \mathbf{f}'(j)\beta + a_{n+j} \tag{12.15}$$

where $\mathbf{x}_{n+j} = \mathbf{f}(j) = (1, j, \ldots, j^{p-1})'$. Note that the coefficients in this trend model are parameterized with respect to time origin n. These trend models relate the values of a time series to functions of time. They are traditionally used for the prediction of time series observations. The prediction of a future value y_{n+j} from time origin n is given by

$$\hat{y}_n(j) = \mathbf{f}'(j)\hat{\beta}_n \tag{12.16}$$

where $\hat{\beta}_n$ is the estimate of β in the model

$$y_t = \mathbf{f}'(-t)\beta + a_t \text{ for } \quad t = 1, \ldots, n$$

One could use the least squares estimate, but this would give each observation the same weight, irrespective of its age. If more recent observations should get more weight than old observations, one could determine the estimate by minimizing the discounted sum of squares

$$S(\beta) = \sum_{t=1}^{n} \omega^{n-t}[y_t - \mathbf{f}'(t-n)\beta]^2$$

The resulting estimate is given by

$$\hat{\beta}_n = [\sum_{t=1}^{n} \omega^{n-t} \mathbf{f}(t-n)\mathbf{f}'(t-n)]^{-1} [\sum_{t=1}^{n} \omega^{n-t} \mathbf{f}(t-n)\, y_t] \tag{12.17}$$

If we use polynomial fitting (forecast) functions $\mathbf{f}(t)$, the matrix

$$P_n = [\sum_{t=1}^{n} \omega^{n-t} \mathbf{f}(t-n)\mathbf{f}'(t-n)]^{-1} \tag{12.18}$$

approaches a steady state limit as n goes to infinity; let us call this matrix $P_\star$. This follows because, with increasing n, the terms in $\omega^n \mathbf{f}(-n)\mathbf{f}'(-n)$ go to zero, as the decrease in an exponential is faster than the increase in a polynomial. Therefore,

$$\hat{\beta}_n = P_\star [\mathbf{f}(0) y_n + \omega \sum_{t=1}^{n-1} \omega^{n-1-t} \mathbf{f}(t-n)\, y_t] \tag{12.19}$$

Polynomial trend functions $\mathbf{f}(j)$ are such that they satisfy the difference equation

$$\mathbf{f}(j) = L\mathbf{f}(j-1)$$

where L is a certain transition matrix. For example, consider the mean model with a single constant forecast function $f(j) = 1$. In this case $L = 1$ is a scalar and $f(0) = 1$. Or, take the linear trend model with $\mathbf{f}(j) = (1, j)'$. There the transition matrix and the vector of initial fitting functions are given by

$$L = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{f}(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Or, consider the quadratic trend model with $\mathbf{f}(j) = (1, j, j^2)'$. There

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{f}(0) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

Using the special nature of these polynomial fitting functions, we can write the second term in equation (12.19) as

$$\omega P_\star \sum_{t=1}^{n-1} \omega^{n-1-t} \mathbf{f}(t-n)\, y_t = [L' - P_\star \mathbf{f}(0)\mathbf{f}'(0)L']\hat{\beta}_{n-1}$$

For a proof, see page 103 in Abraham and Ledolter (1983). Therefore,

$$\begin{aligned}\hat{\beta}_n &= P_\star \mathbf{f}(0) y_n + L'\hat{\beta}_{n-1} - P_\star \mathbf{f}(0)\mathbf{f}'(1)\hat{\beta}_{n-1} \\ &= L'\hat{\beta}_{n-1} + P_\star \mathbf{f}(0)\,[y_n - \mathbf{f}'(1)\,\hat{\beta}_{n-1}] \end{aligned} \tag{12.20}$$

The new parameter estimate is a linear combination of the previous parameter estimate, $\hat{\beta}_{n-1}$, and the latest one-step-ahead forecast error $y_n - \hat{y}_{n-1}(1) = y_n - \mathbf{f}'(1)\hat{\beta}_{n-1}$. The only purpose of the factor L' in the first component is to update the coefficients for the new time origin. Even if we predict y_n perfectly, we have to update the β vector as it is always parameterized with respect to the time origin of the latest available observation.

It is instructive to look at special cases and consider the mean model with $f(j) = 1$; that is, $y_{n+j} = \mu + a_{n+j}$. Then with $L = 1$ and $f(0) = 1$,

$$\hat{\mu}_n = \hat{\mu}_{n-1} + (1-\omega)\,[y_n - \hat{\mu}_{n-1}] = (1-\omega)\,y_n + \omega\hat{\mu}_{n-1} \tag{12.21}$$

since

$$P_\star = (\sum_{j\geq 0} \omega^j)^{-1} = 1 - \omega$$

The forecasts of all new observations made from forecast origin n are given by $\hat{y}_n(j) = \hat{\mu}_n$. Equation (12.21) shows how they are updated with each new observation. The new mean level is a weighted average of the latest observation and the previous mean level. Or, equivalently, it is obtained by adjusting the previous mean level by a fraction of the latest one-step-ahead forecast error. Note that each new observation carries some weight, even if we already have observed a very long record. In the forecasting literature, this procedure is known as *simple exponential smoothing* with smoothing constant $\alpha = 1 - \omega$.

For the linear trend model, $y_{n+j} = \beta_0 + \beta_1 j + a_{n+j}$, one can show by substitution into equation (12.20) that the updating equations for the discounted least squares coefficients in the forecasts $\hat{y}_n(j) = \hat{\beta}_0(n) + \hat{\beta}_1(n)j$ are given by

$$\begin{aligned}\hat{\beta}_0(n) &= \hat{\beta}_0(n-1) + \hat{\beta}_1(n-1) + (1-\omega^2)[y_n - \hat{y}_{n-1}(1)] \\ \hat{\beta}_1(n) &= \hat{\beta}_1(n-1) + (1-\omega)^2[y_n - \hat{y}_{n-1}(1)] \end{aligned} \tag{12.22}$$

or, equivalently, by

$$\begin{aligned}\hat{\beta}_0(n) &= (1-\omega^2)y_n + \omega^2[\hat{\beta}_0(n-1) + \hat{\beta}_1(n-1)] \\ \hat{\beta}_1(n) &= [(1-\omega)/(1+\omega)][\hat{\beta}_0(n) - \hat{\beta}_0(n-1)] + [2\omega/(1+\omega)]\hat{\beta}_1(n-1) \end{aligned} \tag{12.23}$$

The estimate of the mean level at time n, $\hat{\beta}_0(n)$, is found as the weighted average of the most recent observation, y_n, and $[\hat{\beta}_0(n-1) + \hat{\beta}_1(n-1)]$, which is our best estimate of that mean level based on observations up to time $n-1$. Similarly, to get the new estimate of the slope, $\hat{\beta}_1(n)$, we take a weighted average of the slope estimate at time $n-1$, $\hat{\beta}_1(n-1)$, and the most recent estimate of slope that is obtained as the difference of the mean levels, $[\hat{\beta}_0(n) - \hat{\beta}_0(n-1)]$. This procedure is known as *double exponential smoothing* with smoothing constant $\alpha = 1 - \omega$. The discount coefficient determines how these components are weighted.

The updating equations in (12.23) illustrate how each new observation is used in the revision of the previous estimates and show how the forecasts are updated. The discount coefficient makes the estimates more responsive to changes in the trend of the time series. The smaller the discount coefficient, the more adaptive the coefficient estimates will be. The nature of this procedure is still mostly heuristic as no model for coefficient changes is ever specified. Also, there is only one discount coefficient and, in this simple situation, one cannot distinguish between coefficients that are stable over time and coefficients that are time-varying. [Extensions of this method that allow for different smoothing constants do exist; see Abraham and Ledolter (1983, Chapter 3) for a discussion.]

There are very simple relationships between the forecasts from this discounted least squares approach (or exponential smoothing approach, as it is called in the forecasting literature) and the forecasts that are implied by the autoregressive integrated moving average time series models discussed by Box and Jenkins (1976). For example, it can be shown that the forecasts from simple exponential smoothing with smoothing constant $\alpha = 1-\omega$ are equivalent to the forecasts from the ARIMA(0, 1, 1) model $y_t = y_{t-1} + a_t - \omega a_{t-1}$; the a_ts in this model are independent mean zero random variables. Similarly, it can be shown that the forecasts from double exponential smoothing in (12.23) are equivalent to the forecasts from the ARIMA(0, 2, 2) model $y_t = 2y_{t-1} - y_{t-2} + a_t - 2\omega a_{t-1} + \omega^2 a_{t-2}$. In general, it can be shown that the forecasts from the polynomial trend model in equation (12.15), in which the coefficients are estimated with the moving exponential window function, are the same as the forecasts from a certain restricted ARIMA(0, p, p) model. [For a detailed discussion of these relationships we refer the reader to the paper by Abraham and Ledolter (1986).] This shows that ARIMA models can be thought of as polynomial trend-type models in which the trend coefficients are allowed to change with time. This explains why these models have been so successful in describing and forecasting time series data.

12.3 Model-Based Adaptive Recursive Estimation

The disadvantages with heuristic approaches (either the moving rectangular or the moving exponential window function) are (1) that each coefficient is treated the same way, and (2) that they require an *ad hoc* choice of additional constants — either the length s of the moving window, or the exponential discount coefficient ω. Instead of such heuristic approaches, one can adopt a model-based approach to recursive estimation. There we

assume a probabilistic model for the time-varying coefficients and suppose that the regression coefficients β_t follow certain ARIMA time series processes. For example, in the simplest case we assume that they follow a random walk,

$$\beta_t = \beta_{t-1} + \mathbf{v}_t \tag{12.24}$$

where $\mathbf{v}_t$ are independent random variables with mean vector zero and a certain covariance matrix. Random walks generate smoothly time-varying coefficients. The most useful case in practice is the one where the covariance matrix is diagonal; this implies that the coefficients vary independently from one another. The diagonal elements control the variability of the coefficients; if a diagonal element is zero, then the corresponding coefficient is constant. If it is different from zero, the coefficients vary smoothly over time.

The random walk is a special but very useful model. In econometric applications, where we deal with mostly short series, it is usually difficult to identify more complicated models for the underlying unknown regression coefficients.

In theory one can always work with more elaborate model specifications and assume that the coefficients in the regression model

$$y_t = \mathbf{x}_t'\beta_t + a_t \tag{12.25}$$

follow the difference equation model

$$\beta_t = T\beta_{t-1} + \mathbf{v}_t \tag{12.26}$$

This is a multiple first-order autoregressive model and includes the random walk model in (12.24) as a special case (namely, when $T = I$, where I is the identity matrix). In fact, it is a very general model as it can be shown that any multiple ARMA model of the form

$$\beta_t = \phi_1\beta_{t-1} + \ldots + \phi_p\beta_{t-p} + \mathbf{v}_t - \theta_1\mathbf{v}_{t-1} - \ldots - \theta_q\mathbf{v}_{t-q} \tag{12.27}$$

can be written as such a difference equation. Here the matrices $\phi_1, \ldots, \phi_p$, $\theta_1, \ldots, \theta_q$ contain the autoregressive and moving average parameters. [For a detailed discussion of multiple ARMA models, we refer the reader to the book by Hannan (1970) and the paper by Tiao and Box (1981).] It is well known that we can rewrite an ARMA model as an AR(1) model, but of larger dimension; see, for example, Abraham and Ledolter (1983) or Harvey (1981). As simple substitution shows, we can write the ARMA(p,q) model in (12.27) as

$$\begin{pmatrix} \beta_t \\ \beta^*_{t,2} \\ \vdots \\ \beta^*_{t,k-1} \\ \beta^*_{t,k} \end{pmatrix} = \begin{pmatrix} \phi_1 & I & \mathrm{O} & \ldots & \mathrm{O} \\ \phi_2 & \mathrm{O} & I & \ldots & \mathrm{O} \\ \vdots & \vdots & \vdots & & \vdots \\ \phi_{k-1} & \mathrm{O} & \mathrm{O} & \ldots & I \\ \phi_k & \mathrm{O} & \mathrm{O} & \ldots & \mathrm{O} \end{pmatrix} \begin{pmatrix} \beta_{t-1} \\ \beta^*_{t-1,2} \\ \vdots \\ \beta^*_{t-1,k-1} \\ \beta^*_{t-1,k} \end{pmatrix} + \begin{pmatrix} I \\ -\theta_1 \\ \vdots \\ -\theta_{k-2} \\ -\theta_{k-1} \end{pmatrix} \mathbf{v}_t$$

or

$$\beta_t^\star = T\beta_{t-1}^\star + \mathbf{v}_t^\star \tag{12.28}$$

where $k = \max(p, q+1)$ and $\phi_j = O$ (matrix of zeros) for $j > p$ and $\theta_j = O$ for $j > q$.

After extending the coefficient vector β_t to $\beta_t^\star$, we can write the regression model with the ARMA time-varying coefficients given in (12.27) as

$$\begin{aligned} y_t &= (\mathbf{x}_t', \mathbf{0}', \ldots, \mathbf{0}')\beta_t^\star + a_t \\ \beta_t^\star &= T\beta_{t-1}^\star + \mathbf{v}_t^\star \end{aligned} \tag{12.29}$$

where T, $\beta_t^\star$, and $\mathbf{v}_t^\star$ are defined above. To simplify the notation in our presentation, we assume that the model is given by equations (12.25) and (12.26). However, as we have shown here, we can always generalize this model to more elaborate situations.

Two questions now arise:

1. How do we estimate and update the regression coefficient estimates at time n, knowing that these coefficients are not constant, but follow this more general time-varying coefficients model?

2. How do we make inferences about the underlying coefficient variability, and how can we test whether the coefficients are in fact constant or time-varying?

Kalman filter equations can be used to estimate and update the coefficient estimates; the reader is referred to the work by Kalman (1960) and Kalman and Bucy (1961). The books by Jazwinski (1970) and Young (1984), and the paper by Meinhold and Singpurwalla (1983) include excellent reviews of this topic. There are two equations to the system in (12.25) and (12.26). The first, equation (12.25), is called the measurement equation; it describes the generation of the observations from a given state vector, which in our case is the vector of unknown coefficients. The second equation, (12.26), is called the system equation; it describes the evolution of the state (coefficient) vector. There are two error (noise) components: the measurement and the process noise. It is assumed that a_t and $\mathbf{v}_t$ are two independent white noise sequences with zero means and variances σ^2 and $\sigma^2\Omega$, respectively. The matrix Ω is a matrix of variance ratios; it relates the variability in the coefficients to the variability of the measurement noise.

Let us assume that the noise sequences are from normal distributions and let us suppose that also the distribution of the initial state (coefficient) vector at time zero, β_0, follows a normal distribution with mean vector, say $\hat{\beta}_{0|0}$, and covariance matrix, say $\sigma^2 P_{0|0}$. We can think of this as the prior distribution. Then one can show that also the conditional distribution of β_n, given the data up to time $n-1$, i.e., $\mathbf{y}^{n-1} = \{y_{n-1}, y_{n-2}, \ldots, y_1\}$, and the conditional distribution of β_n given the data up to time n, i.e., $\mathbf{y}^n = \{y_n, y_{n-1}, \ldots, y_1\}$,

are normal. (Of course, this requires that the parameters T, σ^2, Ω, the initial values $\hat{\beta}_{0|0}$, $P_{0|0}$, and the values of the explanatory variables are known.) Let us denote the mean vector and the covariance matrix of the conditional distribution of β_n given $\mathbf{y}^n$ by $\hat{\beta}_{n|n}$ and $\sigma^2 P_{n|n}$, and the ones for the conditional distribution of β_n given $\mathbf{y}^{n-1}$ by $\hat{\beta}_{n|n-1}$ and $\sigma^2 P_{n|n-1}$. Then there exist convenient updating equations that revise the means and covariance matrices; they are known as the Kalman filter equations:

$$\begin{aligned}
\hat{\beta}_{n|n-1} &= T\hat{\beta}_{n-1|n-1} \\
P_{n|n-1} &= TP_{n-1|n-1}T' + \Omega \\
\hat{\beta}_{n|n} &= \hat{\beta}_{n|n-1} + \mathbf{k}_n(y_n - \mathbf{x}_n'\hat{\beta}_{n|n-1}) \\
P_{n|n} &= P_{n|n-1} - \mathbf{k}_n\mathbf{x}_n'P_{n|n-1} \\
\mathbf{k}_n &= P_{n|n-1}\mathbf{x}_n(1 + \mathbf{x}_n'P_{n|n-1}\mathbf{x}_n)^{-1}
\end{aligned} \tag{12.30}$$

To start the recursions, one has to choose the starting values $\hat{\beta}_{0|0}$ and $P_{0|0}$. To reflect ignorance about the parameters at time zero $P_{0|0}$ is usually taken as a diagonal matrix with large diagonal elements, and $\hat{\beta}_{0|0}$ is taken as the zero vector. The above equations show how to revise the estimates. The first equation provides the prediction of the next parameter estimate. From $\hat{\beta}_{n-1|n-1}$ (which is the estimate of β_{n-1} given data up to time $n-1$) we compute $\hat{\beta}_{n|n-1}$, which is the "projected" estimate of β_n given the data up to time $n-1$. The third equation shows how to update this estimate and illustrates the calculation of $\hat{\beta}_{n|n}$ after the most recent observation y_n has become available. The Kalman gain vector $\mathbf{k}_n$ in this recursive updating equation depends on T, Ω, and the past data. It determines how much weight is given to the most recent one-step-ahead forecast error $y_n - \mathbf{x}_n'\hat{\beta}_{n|n-1}$.

Here we have explained the recursive equations in (12.30) from a Bayesian viewpoint and have followed the approach by Ho and Lee (1964). Kalman has derived these equations from an orthogonal projection argument, while yet others (see Duncan and Horn, 1972) use a generalized least squares argument to derive these recursions.

The case when $T = I$ (that is, the regression coefficients follow a random walk) has been treated extensively by Athens (1974) and Harrison and Stevens (1976). Let us set $\hat{\beta}_n = \hat{\beta}_{n|n}$ and $P_n = P_{n|n}$ to simplify the notation. Then the Kalman filter equations become

$$\begin{aligned}
\hat{\beta}_n &= \hat{\beta}_{n-1} + [1 + \mathbf{x}_n'(P_{n-1} + \Omega)\mathbf{x}_n]^{-1}(P_{n-1} + \Omega)\mathbf{x}_n(y_n - \mathbf{x}_n'\hat{\beta}_{n-1}) \\
P_n &= (P_{n-1} + \Omega) - [1 + \mathbf{x}_n'(P_{n-1} + \Omega)\mathbf{x}_n]^{-1}(P_{n-1} + \Omega)\mathbf{x}_n\mathbf{x}_n'(P_{n-1} + \Omega)
\end{aligned} \tag{12.31}$$

In the regression model with constant coefficients (that is, $T = I$ and $\Omega = O$), they simplify to

$$\begin{aligned}\hat{\beta}_n &= \hat{\beta}_{n-1} + (1 + \mathbf{x}_n' P_{n-1} \mathbf{x}_n)^{-1} P_{n-1} \mathbf{x}_n (y_n - \mathbf{x}_n' \hat{\beta}_{n-1}) \\ P_n &= P_{n-1} - (1 + \mathbf{x}_n' P_{n-1} \mathbf{x}_n)^{-1} P_{n-1} \mathbf{x}_n \mathbf{x}_n' P_{n-1}\end{aligned} \tag{12.32}$$

They are the same as the updating equations for recursive least squares in (12.3) and (12.4).

Additional insight into these recursions can be gained by considering the model with just one independent variable, $y_t = \beta x_t + a_t$. For random walk coefficients, that is $\beta_t = \beta_{t-1} + v_t$ with $Var(v_t) = \sigma^2 \omega$ where ω is the variance ratio $Var(v_t)/Var(a_t)$, the updating equations are

$$\begin{aligned}\hat{\beta}_n &= \hat{\beta}_{n-1} + P_n x_n (y_n - x_n \hat{\beta}_{n-1}) \\ P_n &= (P_{n-1} + \omega)/[1 + x_n^2 (P_{n-1} + \omega)]\end{aligned} \tag{12.33}$$

For constant coefficients ($\omega = 0$) the second equation simplifies after repeated substitutions to

$$P_n = P_1 (1 + P_1 \sum_{t=2}^{n} x_t^2)^{-1} = (\sum_{t=1}^{n} x_t^2)^{-1} \tag{12.34}$$

Here we have set $P_1 = 1/x_1^2$. With this, we find that for constant coefficients the updating equation is given by

$$\hat{\beta}_n = \hat{\beta}_{n-1} + (\sum_{t=1}^{n} x_t^2)^{-1} x_n (y_n - x_n \hat{\beta}_{n-1}) \tag{12.35}$$

In the linear regression model with constant coefficients, the adjustment weights for $x_n(y_n - x_n \hat{\beta}_{n-1})$ depend only on the past x-values. In the regression model with time-varying coefficients, the weights P_n depend also on the parameter $\omega > 0$; they are always larger than the ones for constant coefficients. This indicates that, as expected, the time-varying coefficient model gives more weight to the most recent one-step-ahead forecast error. Thus, estimates in time-varying coefficient regression models adapt to changes faster than the constant-coefficient least squares estimates.

This discussion has shown that models with time-varying coefficients lead to a faster adaptation in the parameter estimates. However, in order to calculate these estimates one has to select the elements in the variance ratio matrix Ω. One can pick these coefficients heuristically, which is very similar to the *ad hoc* choice of the discount coefficient in the moving exponential window estimates in Section 12.2. This is usually difficult. Alternatively, one can use past data to estimate these unknown coefficients and, in addition, test whether the coefficients are in fact time-varying or constant.

Several estimation procedures are discussed in the literature. Here we describe one that is based on the maximum likelihood principle. Under the usual normal assumptions on the disturbances in equations (12.25) and (12.26) we know that also the conditional (predictive) distribution of $y_n = \mathbf{x}_n'\beta_n + a_n$, given the information on the past history up to time $n-1$ (that is, $y_{n-1}, y_{n-2}, \ldots, y_1$), is again normal with mean $\mathbf{x}_n'\hat{\beta}_{n|n-1}$ and variance $\sigma^2 + \sigma^2\mathbf{x}_n'P_{n|n-1}\mathbf{x}_n = \sigma^2 f_n$ where $f_n = 1 + \mathbf{x}_n'P_{n|n-1}\mathbf{x}_n$. Since a joint density of n random variables can always be written as the product of conditional densities, we find that

$$\begin{aligned} p(y_1, \ldots, y_n|\sigma^2, T, \Omega) &= p(y_1)p(y_2|y_1)\ldots p(y_n|y_{n-1}, \ldots, y_1) \\ &= \prod_{t=1}^{n}(2\pi\sigma^2 f_t)^{-1/2}\exp\{-\frac{1}{2\sigma^2 f_t}(y_t - \mathbf{x}_t'\hat{\beta}_{t|t-1})^2\} \end{aligned} \tag{12.36}$$

For given data on the response and the explanatory variables one can calculate the log-likelihood function

$$\begin{aligned} &\ell(T, \Omega, \sigma^2|\text{data}) \\ &\quad = \text{constant} - n\log\sigma - \frac{1}{2}\sum_{t=1}^{n}\log f_t - \frac{1}{2\sigma^2}\sum_{t=1}^{n}(y_t - \mathbf{x}_t'\hat{\beta}_{t|t-1})^2/f_t \end{aligned} \tag{12.37}$$

Setting the derivative of the log-likelihood function with respect to σ^2 equal to zero, we find that the maximum likelihood estimate of σ^2 is given by

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}(y_t - \mathbf{x}_t'\hat{\beta}_{t|t-1})^2/f_t \tag{12.38}$$

which is the average of the squared standardized one-step-ahead forecast errors. Substituting this expression into the log-likelihood function in (12.37) we obtain the concentrated log-likelihood function

$$\ell_c(T, \Omega|\text{data}) = \text{constant} - n\log\hat{\sigma} - \frac{1}{2}\sum_{t=1}^{n}\log f_t \tag{12.39}$$

Estimates of T and Ω, subject to the constraint that the symmetric matrix Ω is nonnegative definite, can be found by numerical maximization of the concentrated log-likelihood function. For given values of T and Ω, and for given starting values $\hat{\beta}_{0|0}$ and $P_{0|0}$, one can use the recursive Kalman filter equations and update $\hat{\beta}_{t|t-1}$, $P_{t|t-1}$, and $f_t = 1 + \mathbf{x}_t'P_{t|t-1}\mathbf{x}_t$. Usually, one takes the zero vector and a diagonal matrix with large diagonal elements as the starting values for $\hat{\beta}_{0|0}$ and $P_{0|0}$.

If the series are relatively short (as most economic series are), it will not be possible to specify very complicated models for the time-varying coefficients. Parsimonious models have to be considered to make inferences in these models feasible. The simplest model that allows for time-varying coefficients and includes constant coefficients as a special case is the random walk model with $T = I$,

$$\beta_t = \beta_{t-1} + \mathbf{v}_t$$

If we further assume that the coefficients vary independently (that is, Ω is a diagonal matrix), we have to maximize the concentrated log-likelihood function with respect to the p nonnegative diagonal elements $\omega = (\omega_1, \ldots, \omega_p)'$.

One may also want to test the null hypothesis that the coefficients β_t are constant against the alternative that they follow a certain stochastic model. The alternative, for example, may specify that the individual coefficients vary according to independent random walks. An approximate test of whether the coefficients are constant compares the likelihood ratio test statistic

$$G = 2[\ell_c(\hat{\omega}) - \ell_c(\omega = \mathbf{0})] \tag{12.40}$$

to the percentiles of the chi-square distribution with p degrees of freedom. Here $\ell_c(\hat{\omega})$ and $\ell_c(\omega = \mathbf{0})$ are the values of the concentrated log-likelihood function when evaluated at the maximum likelihood estimate $\hat{\omega}$ and at $\omega = \mathbf{0}$, respectively. However, simulations by Garbade (1977) and by Pagan and Tanaka (1979) have shown that, under the null hypothesis of constant coefficients, the distribution of this statistic is more concentrated toward zero than the chi-square distribution that we ordinarily expect in likelihood ratio tests. As a consequence, the likelihood ratio tests are conservative. Theoretical reasons why this can be expected, as well as other estimation and testing procedures, are discussed in a recent review of time-varying coefficient regression models by Nicholls and Pagan (1985).

12.4 Extensions and Concluding Remarks

In this chapter we discussed the recursive estimation of coefficients in regression and certain time series models and reviewed modifications of these recursions that make the estimates more adaptive to changes. Extensions of these procedures to more general models, such as ARMA time series models, transfer-function models, and ARMAX models, are possible. Ljung and Söderström (1983), Söderström and Stoica (1983), Young (1984, 1985), and Ljung (1985) discuss these models in detail and describe instrumental variable methods, estimation methods based on the extended Kalman filter, and various self-adaptive (self-tuning) estimation procedures. The interested reader is referred to these sources.

Acknowledgment

This research was supported by ONR Contract N00014-85-K-0386.

References

Abraham, B. and Ledolter, J. (1983), *Statistical Methods for Forecasting.* New York: John Wiley.

Abraham, B. and Ledolter, J. (1986), Forecast functions implied by ARIMA models and other related forecast procedures. *International Statistical Review*, **54**, 51-66.

Athens, M. (1974), The importance of Kalman filtering methods for economic systems. *Annals of Economic and Social Measurement*, **3**, 49-64.

Box, G.E.P. and Jenkins, G.M. (1976), *Time Series Analysis: Forecasting and Control*, Second edition. San Francisco: Holden-Day.

Broemeling, L.D. (1985), *Bayesian Analysis of Linear Models.* New York: Marcel Dekker.

Brown, R.L., Durbin, J., and Evans, J.M. (1975), Techniques for testing the constancy of regression relationships over time (with discussion). *Journal of the Royal Statistical Society*, **B-37**, 149-192.

Chow, G.C. (1960), Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, **28**, 591-605.

Duncan, D.B. and Horn, S.D. (1972), Linear dynamic regression estimation from the viewpoint of regression analysis. *Journal of the American Statistical Association*, **67**, 815-821.

Ferreira, P.E. (1975), A Bayesian analysis of a switching regression model. *Journal of the American Statistical Association*, **70**, 370-374.

Garbade, K. (1977), Two methods for examining the stability of regression coefficients. *Journal of the American Statistical Association*, **72**, 54-63.

Goldfeld, S.M. and Quandt, R.E. (1973), The estimation of structural shifts by switching regressions. *Annals of Economic and Social Measurement*, **2**, 475-485.

Goldfeld, S.M. and Quandt, R.E. (1976), Techniques for estimating switching regressions, in: *Studies in Nonlinear Estimation.* Cambridge, MA: Ballinger.

Hackl, P. (1980), *Testing the Constancy of Regression Relationships over Time.* Göttingen: Vandenhoeck and Ruprecht.

Hackl, P. and Westlund, A.H. (1985), *Statistical Analysis of Structural Change: An Annotated Bibliography.* Collaborative Paper CP-85-31. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Hannan, E.J. (1970), *Multiple Time Series.* New York: John Wiley.

Harrison, P.J. and Stevens, C.F. (1976), Bayesian Forecasting (with discussion). *Journal of the Royal Statistical Society*, **B-38**, 205-247.

Harvey, A.C. (1981), *The Econometric Analysis of Time Series.* Oxford, UK: Philip Allan.

Hinkley, D.V. (1969), Inference about the intersection in two-phase regression. *Biometrika*, **56**, 495-504.

Hinkley, D.V. (1971), Inference in two-phase regression. *Journal of the American Statistical Association*, **66**, 736-743.

Ho, Y.C. and Lee, R.C.K. (1964), A Bayesian approach to problems in stochastic estimation and control. *IEEE Transactions on Automatic Control*, **9**, 333-339.

Jazwinski, A.H. (1970), *Stochastic Processes and Filtering.* New York: Academic Press.

Kalman, R.E. (1960), A new approach to linear filtering and prediction problems.*ASME Transactions, Journal of Basic Engineering*, **D-82**, 35-45.

Kalman, R.E. and Bucy, R.S. (1961), New results in linear filtering and prediction theory. *ASME Transactions, Journal of Basic Engineering*, **D-83**, 95-107.

Ljung, L. (1985), Estimation of parameters in dynamical systems, in: E.J. Hannan, P.R. Krishnaiah, and M.M. Rao, (eds.), *Time Series in the Time Domain.* Amsterdam: North-Holland.

Ljung, L. and Söderström, T. (1983), *Theory and Practice of Recursive Identification.* Cambridge, MA: MIT Press.

Meinhold, R.J. and Singpurwalla, N.D. (1983), Understanding the Kalman filter. *Ameri-*

can Statistician, **37**, 123-127.

Nicholls, D.F. and Pagan, A.R. (1985), Varying Coefficient Regression, in E.J. Hannan, P.R. Krishnaiah, and M.M. Rao, (eds.), *Time Series in the Time Domain.* Amsterdam: North-Holland.

Pagan, A.R. and Tanaka, K. (1979), A further test for assessing the stability of regression coefficients, unpublished research paper, Australian National University.

Plackett, R.L. (1950), Some theorems on least squares. *Biometrika*, **37**, 149-157.

Poirier, D.J. (1976), *The Econometrics of Structural Change.* Amsterdam: North-Holland.

Quandt, R.E. (1958), The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association*, **53**, 873-880.

Quandt, R.E. (1960), Tests of hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistical Association*, **55**, 324-330.

Quandt, R.E. (1972), A new approach to estimating switching regressions. *Journal of the American Statistical Association*, **67**, 306-310.

Rao, C.R. (1965), *Linear Statistical Inference and its Applications.* New York: John Wiley.

Shaban, S.A. (1980), Change point problem and two-phase regression: An annotated bibliography. *International Statistical Review*, **48**, 83-94.

Söderström, T. and Stoica, P.G. (1983), *Instrumental Variable Methods for System Identification.* Berlin: Springer.

Tiao, G.C. and Box, G.E.P. (1981), Modeling multiple time series with applications. *Journal of the American Statistical Association*, **76**, 802-816.

Young, P.C. (1984), *Recursive Estimation and Time Series Analysis.* Berlin: Springer.

Young, P.C. (1985), Recursive identification, estimation and control, in: E.J. Hannan, P.R. Krishnaiah, and M.M. Rao, (eds.), *Time Series in the Time Domain.* Amsterdam: North-Holland.

CHAPTER 13

An Adaptive Method of Regression Analysis

Yuri P.Lukashin

Summary

A new adaptive method of analyzing a linear regression with time-varying coefficients is presented. The coefficients are adapted by means of exponentially weighted moving averages (EWMA). The coefficients' trajectories imply possible improvements of the model specification. Aspects of suitable preparation of the time series, such as the elimination of time trends and parameter estimation, are also considered. The method is illustrated on the basis of both artificially generated and real economic data.

13.1 Introduction

A basic assumption in ordinary discrete-time regression analysis is that the relation between the endogeneous variable y and the exogenous variables x_i, $i = 1,\ldots,p$, can be approximated by a linear equation with constant coefficients. These coefficients reflect the force of interrelation between the variables. The regression equation is

$$y_t = \sum_{i=1}^{p} \beta_i x_{ti} + e_t \tag{13.1}$$

where β_i are the unknown regression coefficients, e_t are the error terms of the model and the subscript $t = 1,\ldots,T$ denotes time points of observation.

In matrix notation

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{e} \tag{13.2}$$

where $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)'$ i is a $(T \times p)$ matrix, $\mathbf{y}$ and $\mathbf{e}$ are T-vectors, and β is a p-vector. Numerical values for the unknown coefficients are usually derived by means of the least squares estimator

$$b = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{13.3}$$

However, the relation between the variables may be nonconstant over time. Thus, the parameter estimates obtained under the hypothesis of constancy are in some sense averages over the observation period, and it is doubtful whether they are a good basis for analysis or forecasting. In other words, the linear regression with constant coefficients is too rigid in some cases and can lead to serious failures.

One way to take possible nonconstancies into account is to adapt the estimates of the coefficients to new observations. This would allow us to analyze the trajectories of the coefficients over time and, moreover, to forecast their future values. The most popular tool for this aim is the Kalman filter, originally used in engineering (Kalman, 1960; Kalman and Bucy, 1961; Mehra, 1972). For applications in economics, Cooley and Prescott (1973, 1976) proposed an adaptive estimation method for regression coefficients that obey a Markov process. A survey on time-varying coefficients regression analysis is given, for instance, by Raj and Ullah (1981).

The mentioned methods need model parameters to be known such as the covariance matrix of the error terms or the transition matrix of the Markov process. In applications to economic data, however, this must be considered as a serious shortcoming. To overcome the related difficulties, Lukashin (1979) suggested using an anti-gradient direction for the adaptation of the coefficients, an approach already suggested be Wheelwright and Makridakis (1973). However, in some situations the adaptation process converges only slowly. In addition, the corrections for the coefficients, being proportional to the forecast errors, may be highly correlated.

In this chapter, a modification of Brown's (1963) classical adaptive methods is suggested. The main concept of these methods is the exponentially weighted moving average (EWMA). The application of exponentially weighted moving averages for adapting regression coefficients considerably enlarges the family of adaptive models.

In Section 13.2 the suggested method of adapting the regression coefficients is exposed. In Section 13.3 the method is first applied to artificially generated data; then its practical application is demonstrated by analyzing the effect of the inflation rate on the interest rate. The chapter draws conclusions in Section 13.4.

13.2 Adaptive Estimation of Regression Coefficients

Let us consider the following regression equation with time-varying coefficients

$$y_t = \mathbf{x}_t'\beta_t + e_t \tag{13.4}$$

The estimate $\mathbf{b}$ of the parameter vector β may be expressed in terms of the sample averages of the mixed products of observations of the endogeneous and exogeneous variables, i.e.,

$$\mathbf{b} = [\frac{1}{T}\mathbf{X}'\mathbf{X}]^{-1}[\frac{1}{T}\mathbf{X}'\mathbf{y}] = \begin{pmatrix} m_{11} & \dots & m_{1p} \\ \vdots & & \vdots \\ m_{p1} & \dots & m_{pp} \end{pmatrix}^{-1} \begin{pmatrix} m_{1,p+1} \\ \vdots \\ m_{p,p+1} \end{pmatrix} \tag{13.5}$$

here $m_{ij} = \sum_t x_{ti}x_{tj}/T$, $i,j = 1,\dots,p+1$, and $x_{t,p+1} = y_t$.

The main feature of an adaptive regression analysis procedure is a suitable updating method for the averages m_{ij}. It is proposed to substitute in (13.5) the whole period averages m_{ij} by local (or current) averages $s_{ij,t}$. Examples of suitable concepts for $s_{ij,t}$ are: In the context of moving regression analysis, the quantities $s_{ij,t}$ are averages based on a moving window of the observations. A straightforward modification is to use exponentially weighted moving averages. A different approach of updating the $s_{ij,t}$ is to approximate the time trend shown by the $s_{ij,t}$ with known functions of time or related models and to extrapolate them.

As already noted, most methods of adaptive regression analysis make use of an *a priori* stated model for the dynamic coefficients; sometimes even the parameter values of this model are required to be known. Such approaches must be considered to be unrealistic in many situations, particularly in economics. The method suggested here is based only on the dynamic of the $s_{ij,t}$. These products form time series, which can be presented graphically and analyzed visually or with the help of analytical means. The set of these graphs represents the dynamic and structure of the process under study. For example, the graphs allow us to locate points of suspected changes in the regression coefficients: as can be seen from (13.4), a jump in parameter β causes a change in the level of y and, consequently, of all corresponding $s_{ij,t}$.

Thus, a multidimensional analysis is decomposed into $p(p+3)/2$ unidimensional ones. However, these unidimensional analyses should not be performed in isolation from the corresponding simultaneous studies. All unidimensional problems must be submitted to one global criterion.

The adaptive regression analysis will first be considered for the case where the variables do not show marked trends. For this case it is proposed to substitute for m_{ij} the exponentially weighted moving average $s_{ij,t}$. For all i,j, and t, the updating recursion for $s_{ij,t}$ is

$$s_{ij,t} = (1-\alpha)s_{ij,t-1} + \alpha(x_{ti}x_{tj}) \tag{13.6}$$

where α $(0 < \alpha < 1)$ is the smoothing constant. Initial values $s_{ij,0}$ may be determined as the arithmetic means of the first n_1 observations, i.e.,

$$s_{ij,0} = \frac{1}{n_1} \sum_{t=1}^{n_1} x_{ti} x_{tj} \tag{13.7}$$

The estimates of the coefficients β are calculated for each t on the basis of (13.5) with $s_{ij,t}$ substituted for m_{ij}.

As has been mentioned above, the number of quantities to be updated is $p(p+3)/2$. A simple version of the procedure is to use the same numbers n_1 and α for all pairs (i,j). To find best values for n_1 and α, a suitable optimality criterion, such as the mean-squared error of the one-step ahead forecasts, might be used:

$$Q(n_1, \alpha) = \hat{\mathbf{e}}'\hat{\mathbf{e}}/T$$

where $\hat{\mathbf{e}}$ is the vector of residuals. The minimum of this global criterion Q may be found iteratively by searching the minimal value of Q with respect to α for a suitable set of integers n_1.

Obviously, this method is based on the assumption that the mean levels of the quantities m_{ij} evolve only slightly. If this does not hold, the exponentially weighted moving averages may not attain the mean levels of the products. This will be the case if, e.g., one or several variables grow linearly in time. In such situations, polynomial models for the m_{ij} can be adapted using multiple exponential smoothing. Another way is to reformulate the model by taking suitable differences of the time series so that the trend is eliminated and a correct use of the method is possible. It should be noted that the results of the analysis depend on the specification of the evolution of the m_{ij}.

13.3 Illustrating Examples

Three examples are discussed in this section which demonstrate the applicability of the suggested method.

Example 1:

The data to be analyzed are a series of 120 values x_t, generated according to $iidN(1,1)$, and corresponding values $y_t = x_t\beta_t + e_t$, where

$$\beta_t = 3 + \sin(\pi t/10)$$

and the random components e_t follow $iidN(0, 0.01)$.

Figure 13.1 shows the trajectories of the true coefficient β and of its estimates b obtained on the basis of the proposed adaptive method. The plots show a good agreement although there are some particular deviations.

Example 2:

The difference here, in comparison with Example 1, is that the coefficients are generated according to

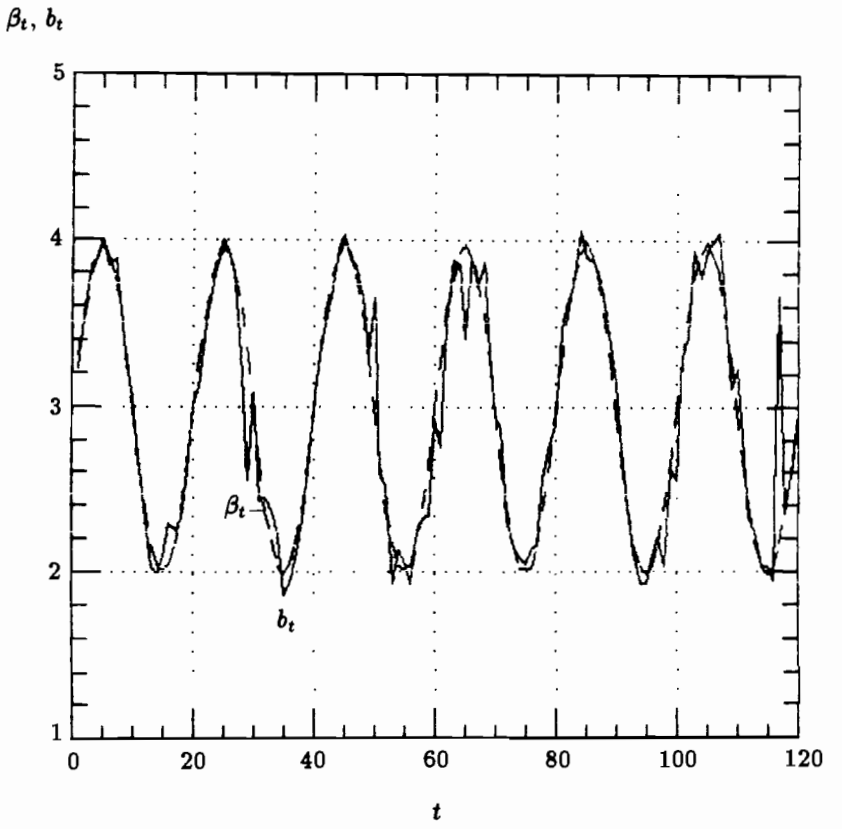

Figure 13.1: Example 1: Trajectories of the true parameter β_t and its estimate b_t of $y = x\beta + e$, with $x \sim N(1,1)$ and $e \sim N(0,\ 0.01)$.

$$\beta_t = \begin{cases} 3 & \text{if} \quad 1 \le t \le 39 \\ 6 & \text{if} \quad 40 \le t \le 79 \\ 3 & \text{if} \quad 80 \le t \le 120 \end{cases}$$

i.e., there are two shifts of the coefficient β. The trajectories of the true coefficient β and of its estimates are given in *Figure 13.2.* The agreement of the true values and the estimates again is quite good. The rather small inertness corresponds to a small value of α.

It should be noted that, in both examples 1 and 2, only little sensitivity of the results to changes in n_1 in a wide range has been observed.

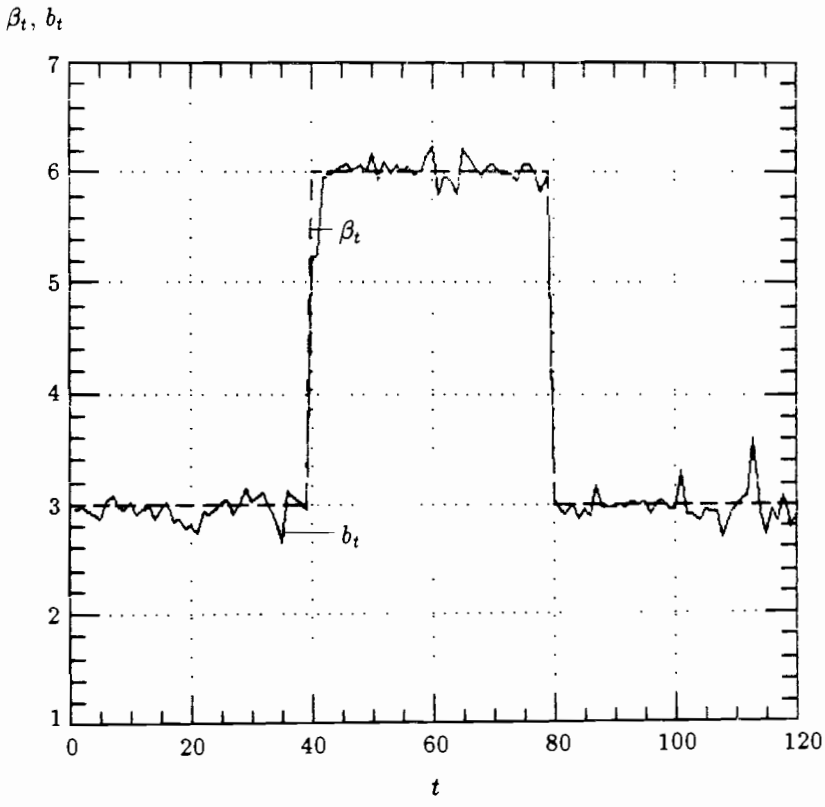

Figure 13.2: Example 2: Trajectories of the true parameter β_t and its estimate b_t of $y = x\beta + e$, with $x \sim N(1,1)$ and $e \sim N(0, 0.01)$.

Example 3:

As an application to real data, the effect of the inflation rate x, derived from the consumer price index, on the interest (commercial paper) rate y will be analyzed. The data to be analyzed are quarterly for the period 1960 to 1984 from the USA. The coefficients of interest are β_1 and β_2 from

$$y_t = \beta_1 + x_t\beta_2 + e_t$$

The coefficient β_1 may be interpreted as the rate of interest after eliminating the effect of inflation.

In *Figure 13.3* the graphs of the trajectories for the estimates of β_1 and β_2 are shown. The results seem to be meaningful in the sense that at least the turning points and the local trends do not contradict the experts' opinion.

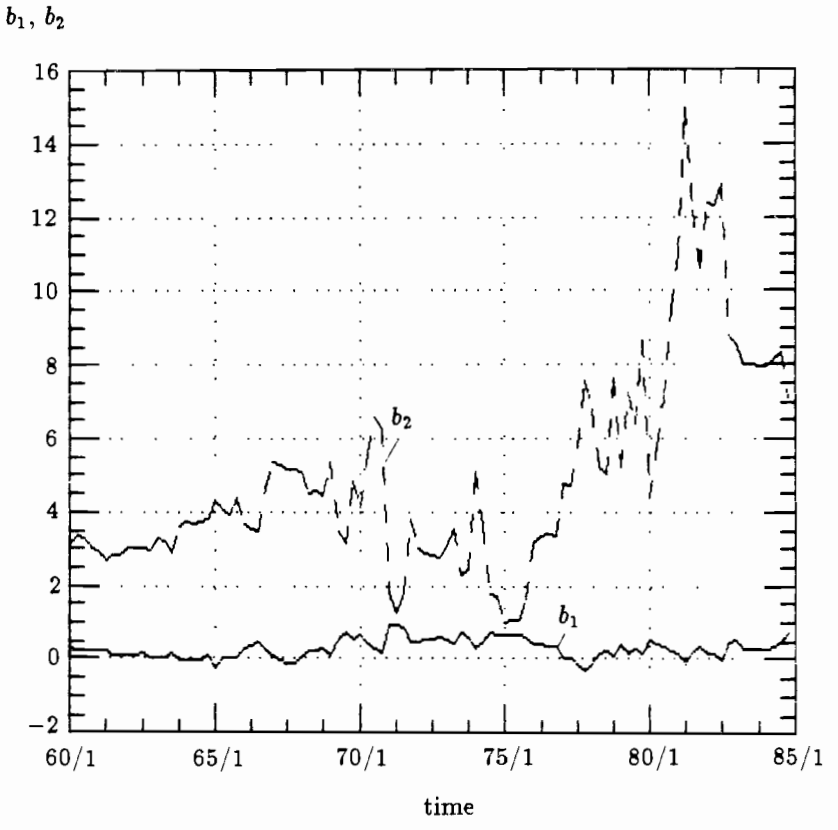

Figure 13.3: Example 3: Trajectories of the estimates b_1 and b_2 of β_1 and β_2 in $y = \beta_1 + x\beta_2 + e$ (y and x are the interest rate and the inflation rate, respectively).

13.4 Conclusions and Further Aspects

Often, changes in economic structures are not abrupt but appear, owing to the inertia of economic mechanisms, as evolutionary processes. Therefore the proposed adaptive regression analysis method may in many situations be an adequate tool for analyzing structural changes. The trajectories of the coefficients give insights on how to respecify a model so that the nonconstant behavior of the coefficients is taken into account.

The proposed method is rather intuitive; its main aim is to allow a simple, easily applicable and useful adaptive analysis of linear regression models. It is obvious that the method will not give good results in every case. The flexibility of the model gained from allowing the coefficients to be time-varying causes an increased sensitivity to errors and random fluctuations of the data.

Multicollinearity causes problems not only in the usual regression analysis. In the adaptive version, multicollinearity may result in high correlations of the estimated coefficients in some periods. In periods for which the precision of the adaptive estimates is doubtful (e.g., due to suspected multicollinearity, indicated by diverging trajectories of two coefficients), it is apparently better to rely on the "constant" estimates and to substitute them for the corresponding adaptive estimates. Therefore, the adaptive regression analysis method might be useful as a supplement to the usual regression analysis. Problems caused by multicollinearity are also known to occur with other adaptive methods, such as the Cooley-Prescott method. It should be noted that multicollinearity does not affect forecasts seriously. Errors in estimates of one of the coefficients tend to be compensated by errors in estimates of others, owing to the correlation between the coefficients.

The method outlined in Section 13.2 can be modified in several ways. An important modification concerns the automatic control of the smoothing parameter α (cf. Lukashin, 1979). For instance, this is possible on the basis of a tracking control signal, or by using an updating method that is also applied to α.

References

Brown, R.G. (1963), *Smoothing, Forecasting and Prediction of Discrete Time Series.* New York: Prentice-Hall.

Cooley, T.F. and Prescott, E.C. (1973), The adaptive regression model. *International Economic Review*, **14**, 364-371.

Cooley, T.F. and Prescott, E.C. (1976), Estimation in the presence of stochastic parameter variation. *Econometrica*, **44**, 167-184.

Kalman, R.E. (1960), A new approach to linear filtering and prediction problems. *Transactions ASME, Journal of Basic Engineering*, **D-82**, 35-45.

Kalman, R.E. and Bucy, R.S. (1961), New results in linear filtering and prediction theory. *Transactions ASME, Journal of Basic Engineering*, **D-83**, 95-108.

Lukashin, Y.P. (1979), *Adaptivnye Metody Kratkosrochnogo Prognozirovaniya.* Moskva: Statistika (in Russian).

Mehra, R.K. (1972), Approaches to adaptive filtering. *IEEE Transactions on Automatic Control*, **AC-17**, 693-698.

Raj, B. and Ullah, A. (1981), *Econometrics, A Varying Coefficients' Approach.* London: Croom Helm.

Wheelwright, S.C. and Makridakis, S. (1973), Forecasting with adaptive filtering. *Revue Française d'Automatique, Informatique et de Recherche Operationelle*, **V-7** (1).

CHAPTER 14

Changing and Random Coefficient Models. A Survey

Józef Dziechciarz

Summary

This chapter contains a survey of various econometric model formulations in which it is assumed that coefficients vary across time. Depending on the accepted parameter variation structure one may classify such models into two main groups: models with variable but nonstochastic parameters and models with randomly varying coefficients. The latter group consists of two types — models where coefficients are generated from stationary and models in which coefficients are generated from nonstationary stochastic processes. All three groups are surveyed. Several representative models from each group are shown with special emphasis on estimation, testing the specification and possible fields of implementation. Justification for the various model formulations is given. A detailed list of references ends the survey.

14.1 Introduction

In classical econometric modeling, it is assumed that an economic structure generating a statistical sample remains constant. This means that, explicitly or implicitly, the existence of several factors is assumed:

1. A unique parameter vector connecting the endogenous variable with the set of independent variables.
2. One set of parameters of stochastic processes generating the model's disturbances.

3. A unique functional form of the econometric model.

The assumptions listed above define a very important property of the econometric model -- namely "model stability". [A detailed discussion of econometric model stability is available in Dziechciarz (1980).] In other words, in an unstable model, the structural parameters, disturbance distribution, or the model's analytical form may not be the same for all sample observations. This chapter deals mainly with the stability of structural parameters.

Economics belongs to a group of nonexperimental sciences, and econometricians have to work with statistical samples that are generated in uncontrollable economic processes and frequently under unobservable conditions. Many attempts to model economic relationships have been unsuccessful, and some of those failures were caused by parameter stability problems. Researchers noticed that different samples yielded different sets of the model's parameter estimates. In this context, the constant-parameter assumption is not obvious.

An early way to deal with this problem was the attempt to isolate separate groups of homogeneous observations. The different treatment of pre- and postwar data can serve as a typical example. Unfortunately, although economists are aware of this factor, they do not consequently limit their attention to at least approximately homogeneous data sets, i.e., samples generated under stable economic circumstances. It has to be noted that some aspects of structural instability are commonly recognized. One of the early attempts to deal with this problem was the introduction of dummy variables in order to represent seasonal, institutional, or other structural differences. Although dummy variables make a convenient tool, their use often results in pure models and inaccurate forecasts.

Other early attempts to represent the instability of structural relations were adaptive models, where an incorporated mechanism adjusted models' shapes to changing relations. Widely known models in this group are the crowing trend, introduced by Hellwig (1967), and exponential smoothing, by Brown (1963). Since these models do not take into account the complex character of changing economic processes, such approaches are often unsatisfactory.

In recent years, much effort has been devoted to the creation of a more general approach to this problem of parameter variation, proper estimation, and testing techniques for new types of models. Quandt's papers (1958, 1960) initiated research to find new methods of uncovering and handling the instability of models.

14.2 Defining an Adequate Model

An adequate model means in general

1. That a model's equation represents reality, i.e., the equation does not contain any systematic error, either in time or cross-sectional space covered by the statistical sample used for the model's estimation.

2. That the necessary conditions to apply statistical tools in order to test the model's properties are fulfilled.

Checking the model's efficiency and accuracy means testing relative statistical hypotheses. For a very interesting discussion of econometric model evaluation criteria, see Dhrymes *et al.* (1972).

Many types of model shortcomings are considered in the literature. Hackl (1980) lists the following:

1. One or more important regressors are omitted.
2. The functional form of one or more regressors is incorrectly specified.
3. The model is unstable, i.e., structural parameters, disturbance distribution, or both vary for different observations.
4. Disturbances for different observations are correlated.
5. Disturbances have other than the normal distribution.

One cannot consider those problems separately; they are often closely connected. For example, an omitted variable can result in nonzero expected value of disturbance, instability of slopes, or variance of disturbances. Surveys of different types of model failure are provided in Ramsey (1969) and Hackl (1980). Several alternative techniques for ascertaining a model's imperfections based on analyses of errors are surveyed in Hackl and Katzenbeisser (1978).

Many scientists argue that recent theoretical works dealing with parameter variation represent a new fashion without significant importance for the practice of econometric modeling. It has to be clearly stated that the conventional econometric model with constant parameters is, and will remain, very important and a wholly useful tool. Its basic merit is simplicity in looking into economic interrelations not obscured by additional details describing changes in the outside world. Whenever it is possible to uncover stable economic relations, which satisfy at least approximately the assumptions of the conventional model, it can be used with confidence and convenience. Sometimes, however, the researcher is not lucky or skilled enough to specify stable regression equations, lack of which changes the relationships being studied.

Although classical models will remain the basic tool in econometrics, the need quite frequently arises to represent a more detailed picture of the modeled processes. It is extremely important, for that reason and from a strategic point of view, to choose properly between models with constant and with varying parameters. Such a decision is a compromise between complexity connected with accuracy and simplicity connected with inaccuracy.

The above discussion is true once there is agreement that varying parameter formulation is more general. It is assumed that, generally, model parameters are generated in

a nonstationary random process, which means that parameters do not have a constant mean and/or variance and can vary systematically. Such models are relevant mainly for modeling the systematic structural variation in time.

The special case of this general formulation is the model whose parameters are generated in a stationary random process. In this case parameters do have constant mean and variance, and therefore they cannot show systematic change for different observations (time, units). Such a model is relevant mainly in modeling cross-sectional data and the time series of cross-sections; see Rosenberg (1973a). In particular, it is assumed that cross-sectional units have the same regression regimes, which are constant in time. Individual reaction of particular units in different time points is treated as a random selection from parameter population with constant mean. In the simplest situation, parameter variance is zero, which yields models with constant parameters.

14.3 Arguments for Varying Parameter Specification

Such arguments are extensive, some of which are specified below:

1. Response of the dependent variable to a signal from the independent variables may not be the same for all observations, although true parameters are constant. Formulation with varying parameters can be relevant because of the nature of econometric models: they are abstractions with some simplifications caused by available data, time, financial resources, and a tendency to formulate a soluble model with acceptable results. Under such circumstances specification errors, which can result from poor quality descriptors, and forecasts have to be taken into account. Introduction of the proper varying-parameter structure can neutralize specification errors. The most frequent specification errors are omission of an important regressor, introduction of the proxy variable and aggregated data, and neglecting nonlinearity. Omission of the important explanation variable can be brought about by inadequate theory, inaccessible data, or by the search for simplicity.

2. Such omitted variables are often connected with structural changes caused by fashion evolution, technological progress, and institutional and organizational changes. In the conventional model, it is assumed that the effect of those omitted factors can be treated as random and that their distribution is constant in time. Additionally, it is assumed that omitted variables are orthogonal to those used in the model. This assumption implies that omitted regressors do not affect parameter estimates explicitly taken into consideration in the model. In reality, the time series of such variables show nonstationary behavior and are nonorthogonal to regressors used in the model. In these circumstances, it is very likely that estimated response parameters will be unstable in time. Minimal expectation is that omitted variables with nonzero effect will result in the intercept variation.

3. Since statistical data for many economic variables are inaccessible, proxy variables are quite frequently applied in the econometric model construction. Proxy variables are also introduced into models — especially dynamic models — in order to represent factors that are difficult or impossible to measure: patterns of future projections or consumption trends, for example. Unfortunately, proxy variables reflect only partially the variation

of the represented processes. In addition, the relationships between proxy variables and their counterpart variables can vary. Under such circumstances, changes of real variables controlling development of the process being modeled lead to instability of proxy variables' parameters.

4. The possible instability of parameters in models using aggregate data is widely discussed in literature; see, for example, Zellner (1962). Since aggregated data are measured by weighting the relative importance of heterogeneous subsets of economic units, parameters in estimated aggregated equations are constant only as long as those weights are constant. In time series, the constant weights assumption, i.e., the constant relative importance of individual units in the aggregate, is not easy to satisfy. Because changes of aggregating weights are a rule, rather than an exception, parameters connected with aggregated variables are usually not stable.

5. The inexact specification of the functional form of the relationship being studied is another possible source of the parameter variation. For example, if on the pretext of Taylor's series, a linear model is estimated as an approximation of a real nonlinear relation, the assumption about constant parameters can be proved only if regressors vary in some narrow range. Once regressor variation exceeds that range, where the linear approximation is acceptable, parameters become unstable; see Rausser *et al.* (1982). Observation of the long-range evolution of many economic time series shows that any construction based on an assumption of a narrow range of variation must be rejected. Approximation of nonlinear true relationships with simple linear constructions, along with observations of variations outside the assumed narrow range, gives one of the strongest arguments for the varying-parameter structure formulation.

6. Economic theory substantiates varying-parameter models. In many situations, economic theory allows for the expectation that relationships will vary over time. For example, changes in economic policy result in changes in the economic environment of the economic units. On the assumption that those units act according to the rules of rational behavior, changes in economic policy result in changes of equations describing their behavior. Indeed, dynamic economic theory and the theory of rational behavior provide no arguments for model formulations with constant parameters.

7. Frequently, the relationship is properly specified, but it is different for some subsets of the sample. It is clear that common parameters represent none of the existing subsets in the available sample. Division of the sample and introduction of more than one regression regime can improve the accuracy of the model and resulting forecasts.

14.4 Types of Varying-Parameter Models

As stated earlier, varying-parameter models may be classified into three main groups:

1. Random parameters from a nonstationary process:

    - the Cooley-Prescott model

- Rosenberg's convergent-parameters model
- Kalman's filter models

2. Random parameters from a stationary process:

   - the Hildreth-Houck model
   - Harvey-Phillips' return-to-normality model
   - Swamy's random coefficient model
   - Hsiao's random coefficient model

3. Varying but nonstochastic parameters:

   - models with constant slope coefficients and intercept that varies over individuals
   - models with constant slope coefficients and intercept that varies over time and individuals
   - seemingly unrelated equations
   - systematically varying parameters
   - seasonality models
   - switching regression models

To the first two types of models in group (3) belong dummy variable models and a type that is often referred to as variance or error component models. The latter also may be considered in a stochastic models framework. The switching regression model represents a number of different formulations, which assume in common that there are groups of observations where parameters are constant. The parameters may vary across subsets of observations within the sample and are nonstochastic. Our survey will begin with the nonstochastically varying coefficient models.

Assuming that parameters can vary across observations or subsets of observations, but not in stochastic way, the structure of the parameter variation must be specified. One of the possible solutions is to relate the parameter variation to one or more explanatory variables.

14.4.1 Fixed coefficient models

One of the most popular models that allows for differences in behavior over cross-sectional units or any differences in behavior over time for a given cross-sectional unit is the model where all coefficients are constant and the disturbance is assumed to capture differences over time and individuals.

These models may be classified further, depending upon whether the variable coefficient is assumed to be random or fixed. The fixed assumption leads to *dummy variable models* and the *seemingly unrelated regression model*, while the random assumption leads

to the model referred to as the *error components model.* This type of model is also sometimes called the *variance component model.* Since models with constant slopes and variable intercept are not the topic here, only some general remarks and literature references will be made about them. Introductory discussions concerning the model with constant parameters and a variable intercept may be found in the works of Maddala (1971), Nerlove (1971a), Swamy (1971), Mundlak (1978c), Hausman and Taylor (1981), and Judge *et al.* (1980, 1982, and 1985).

The model with constant coefficients and a variable intercept (the *error components model* version) may be regarded as the one with random parameters (but some of which — the slopes — are constant), or as one where all coefficients are constant and the disturbance covariance matrix is identical for all individuals. Disturbances in different time periods for the same individual are correlated, but this correlation is constant over time and it is identical for all individuals. Alternatively, the assumption that the intercept may vary over individuals and time may be accepted.

Maddala (1971), Nerlove (1971a), Swamy (1971), and Arora (1973) recommend the estimation technique, which may be regarded as some generalization of the dummy variable estimator. Balestra and Nerlove (1966), Nerlove (1971b), Swamy (1971), Arora (1973), and Fuller and Battese (1973) discuss some convenient transformations for the estimation of this model. Lee and Griffiths (1979) and Taub (1979) suggest a best linear unbiased predictor for the random components. Battese and Fuller (1982) consider a "best constrained predictor". A number of variance component estimators are suggested; these include estimators based on the ordinary least squares residuals, as seen in Wallace and Hussain (1969), Maddala (1971), Swamy (1971), and Arora (1973). Other estimators are proposed by Henderson (1953, 1975), Fuller and Battese (1973, 1974), Rao (1970, 1972), and Kelejian and Stephan (1983). The maximum likelihood version comes from works by Amemiya (1971), Nerlove (1971a), and Maddala (1971). Swamy (1971) and Fuller and Battese (1973) consider the distribution of the various estimators. An important work is that of Searle (1979). Finite sample properties are investigated by Swamy and Mehta (1979) and Taylor (1980). Arora (1973), Maddala and Mount (1973), and Baltagi (1981) study some estimators in Monte Carlo experiments.

Breusch and Pagan (1980) suggest a test based on the Lagrange multiplier statistic for testing hypotheses that state that the intercept is constant for all observations. This is an alternative approach to the classical procedure where a dummy variable estimator is employed jointly with the F-test based on restricted and unrestricted residual sums of squares.

The choice between the assumption that variable component is random or fixed is crucial for the choice of the estimation procedure. Mundlak (1978b), Chamberlain (1978, 1979, and 1983), and Hausman and Taylor (1981) consider this problem. Wallace and Hussain (1969), Swamy (1971), Nerlove (1971a), Swamy and Arora (1972), and Mundlak (1978c) examine a statistical test that helps to choose between a dummy variable and an error components model. Other sources include Lee (1978b), Chamberlain and Griliches (1975), Hausman (1978), and Pudney (1978).

Mundlak and Yahav (1981) investigate a combined model integrating fixed and random effects. Quite a number of different extensions of the model may be found in the literature. Balestra and Nerlove (1966), Nerlove (1967 and 1971b), Maddala (1971), Trognon (1978), Berzeg (1979), Chamberlain (1979, 1983), Anderson and Hsiao (1981, 1982), Nickell (1981), Sevestre and Trognon (1982), and Bhargava and Sargan (1983) explore problems that occur when a lagged dependent variable is included. An alternative disturbance covariance structure is considered in studies by Hause (1977 and 1980), Glejser (1978), Lillard and Willis (1978), Lee (1978a), Pudney (1978), Lillard and Weiss (1979), Revankar (1979), Kiefer (1980), Bhargava *et al.* (1982), MaCurdy (1982), and Schmidt (1983).

Models with discrete and truncated dependent variables are evaluated by Chamberlain (1978, 1979, and 1983), Heckman (1978), Flinn and Heckman (1982), and Singer (1982), as well as by Griliches *et al.*(1978), Hausman and Wise (1979), Kiefer and Neumann (1981), and Maddala (1978). Error components models with heteroscedasticity are investigated by Mazodier and Trognon (1978); and nonlinear error components models with heteroscedasticity, by Griffiths and Anderson (1982). Other works include Avery (1977), Jöreskog (1978), Baltagi (1980), Magnus (1982), Prucha (1984), Reinsel (1982), and Biorn (1981). Mundlak (1978a) proposes the use of biased estimators with a lower mean square error.

14.4.2 Systematically Varying Parameter Models

The most general systematically varying parameter model may be formulated in the following way.

$$y_{it} = \mathbf{x}_{it}'\beta_{it} + e_{it} \tag{14.1}$$

where y_{it} is the observation of the ith unit in the tth time period, $\mathbf{x}_{it}$ is the $(K \times 1)$ vector of nonstochastic explanatory variables, β_{it} is the $(K \times 1)$ coefficient vector, and e_{it} is the disturbance term, the set of which are normally and independently distributed random variables with zero means and variances $\sigma^2 > 0$, $i = 1, \ldots, N$; $t = 1, \ldots, T$.

Since $(KNT + 1)$ parameters are to be estimated with only NT observations available, some additional information is required to make the problem tractable. In general this additional information places some structure on how coefficients vary across observations. Without this, the problem is not tractable. Some typical nonsample information can be introduced. Following Belsley (1973a, 1973b, and 1973c), let nonsample information be described by K linear relations:

$$\beta_{it} = \mathbf{Z}_{it}\tau \tag{14.2}$$

where $\mathbf{Z}_{it}$ is the $(K \times M)$ matrix containing observations on variables explaining the way parameters, β_{it} vary across observations, and τ is the $(M \times 1)$ vector of coefficients associated with variables in the matrix $\mathbf{Z}_{it}$.

In the nonstochastic formulation of the Belsley model under consideration, the $\mathbf{Z}_{it}$ is a known, nonstochastic matrix. This means that equation (14.1) is an exact, rather than a stochastic, relation. Combining (14.1) with (14.2) results in

$$y_{it} = \mathbf{x}'_{it}\beta_{it} + e_{it} = \mathbf{w}'_{it}\tau + e_{it} \tag{14.3}$$

where $\mathbf{w}'_{it} = \mathbf{x}'_{it}\mathbf{Z}_{it}$ is the $(1 \times M)$ vector of observations of the interaction variables. With the assumptions made about e_{it}, the least square estimator of the τ and β_{it} is best, linear, unbiased. This is true whenever $\mathbf{Z}_{it}$ is known and nonstochastic.

Frequently, the variation structure in this model is assumed to be stochastic, the resulting formula being given by the following stochastic equation system:

$$\beta_{it} = \mathbf{Z}_{it}\tau + \mathbf{v}_{it} \tag{14.4}$$

where $\mathbf{v}_{it}$ is the normally distributed disturbance vector with means zero and covariance matrix $\mathbf{V}_v$. In the nonstochastic formulation of Belsley's model, $\mathbf{Z}_{it}$ is a known, nonstochastic matrix and $\mathbf{v}_{it}$ is a zero vector. The resulting model is

$$y_{it} = \mathbf{x}'_{it}\beta_{it} + u_{it} = \mathbf{w}'_{it}\tau + u_{it} \tag{14.5}$$

where $\mathbf{w}'_{it} = \mathbf{x}'_{it}\mathbf{Z}_{it}$ is the $(1 \times M)$ vector of observations of interaction variables, and $u_{it} = x'_{it}\mathbf{v}_{it} + e_{it}$; the disturbances u_{it} have zero mean and variance $E[u_{it}^2] = \mathbf{x}'_{it}\mathbf{V}_v\mathbf{x}_{it} + \sigma^2$.

Disturbances u_{it} are heteroscedastic; the least squares estimator for τ is consistent, but ineffective in comparison with the Aitken's estimator. If the matrix $\mathbf{x}_{it}$ contains a unit vector, its coefficient (intercept) and the estimate for equation disturbance will be indistinguishable. This formulation is a special case of the model presented by Hsiao (1975). In the case $T = 1$, techniques of the Hildreth-Houck model could be applied to estimate parameters with the generalized least squares method. The case of $N = 1$ is discussed by Singh *et al.* (1976). They consider a model in which $\mathbf{Z}_{it}$ contains functions of calendar time and $\mathbf{v_t}$ is normal with mean zero and a diagonal covariance matrix. Singh *et al.* develop both modified Hildreth-Houck and maximum likelihood estimators.

Some additional remarks may be made about relationships in equation (14.2). Whenever parameters β_{it} are thought to be dependent on the set of the same variables, related rows of the matrix $\mathbf{Z}_{it}$ will be identical. On the contrary, if all parameters are functions of different variables, in the matrix $\mathbf{Z}_{it}$ there will be zeros in proper places. It is unlikely that all elements of β_{it} will be the same linear function of the same set of explanatory variables. Given that $\mathbf{Z}_{it}$ contains zeros in appropriate places, equation (14.2) is general enough to cover different possible forms of those functions. The following situations may be distinguished:

1. $\mathbf{Z}_{it}$ may contain functions of variables included in the matrix $\mathbf{x}_{it}$. That means that the true form of relationships in (14.1) are nonlinear.

2. $\mathbf{Z}_{it}$ may contain functions of variables other than those appearing in $\mathbf{x}_{it}$ — for example, calendar time. The justification for using calendar time for an explanation of the structural parameters β_{it} variation is the same as that used when such variables are included in regression models. Calendar time acts as a surrogate for all the unknown

time-related dynamic forces within the economy. Caution is always recommended in using trend-related variables because very often they tempt one to engage in gross, curve-fitting exercises.

3. $\mathbf{Z}_{it}$ may contain qualitative variables. Those variables may be both stochastic or nonstochastic. In such a case, one can expect the existence of different regression regimes.

Alternative formulations of matrix $\mathbf{Z}_{it}$, listed above, includes several specifications of models considered in the literature. An interesting model can be formulated by assuming that the first column in matrix $\mathbf{Z}_{it}$ is a unit vector, and $\mathbf{v}_t' = (v_{t1}, 0, \ldots, 0)$, that means that only the intercept is random. In this case v_{t1} serves as the model's disturbance and all other parameters are deterministic functions of variables from matrix $\mathbf{Z}_{it}$. Each parameter can be written as

$$\beta_{tk} = \tau_1 + \tau_2 z_{2t_k} + \ldots + \tau_M z_{Mt_k} \tag{14.6}$$

where $k = 2, \ldots, N$; $t = 1, \ldots, T$.

14.4.3 Switching regression models

These contain a wide spectrum of different constructions based on the general assumption that there are several regression regimes controlling the process being modeled for relative groups (subsets, partitions) of the sample, which allows constancy of the regression parameters within parts of the sample. Parameters are different between subsets. This is a modification of the previously recognized assumption, where structural parameters were allowed to be different for each observation. Models of this type can be considered in the framework of previously introduced terminology as containing in matrix $\mathbf{Z}_{it}$ qualitative variables sorting observations into different subsets.

One of such a group of models allows the systematic parameter variation for different seasonal periods. The other examples are dummy variables models, and a family of segmented (piecewise) regression models discussed by Quandt (1958), McGee and Carlton (1970), Hinkley (1971a and 1971b), Gallant and Fuller (1973), Goldfeld and Quandt (1973a, 1973b), Poirier (1973, 1976), and many others. In particular, one can distinguish two basic situations. In the first of them, the switch points are known; in the second, they have to be estimated. In the sample, the piecewise regression model can be continuous or not. Because this type of model is used mostly in a time series context, for the sake of convenience and interpretation, this context is mainly considered.

Many economic variables show a seasonal variation that is connected with time. Variation is observed especially with variables such as output, consumption, employment, and others reported systematically — every week, month, quarter, or year. The models with seasonally varying economic variables are referred to as *seasonal models*. The situation where a sample can be divided into two or more subsamples in connection with some seasonal variable will be considered:

$$y_t = x_{t1}\beta_1 + x_{t2}\beta_2 + \ldots + x_{tK}\beta_K + e_t \tag{14.7}$$

where $t = 1, \ldots, T$.

It is assumed that, for some subsamples, values of model parameters can be different. For the sake of simplicity, it will be assumed that the regression structure is constant and different in two parts of the sample, i.e., for $t = 1, \ldots, t_0$ (first subsample), and for $t = t_0 + 1, \ldots, T$ (second subsample). Additionally, it is assumed that not all structural parameters vary, only β_i does, where $i = p+1, \ldots, K$; $p \in \{0, \ldots, (K-1)\}$, $K = p + q$. A standard model with dummy variable D defined as follows:

$$D = \begin{cases} 0 & \text{if } t = 1, \ldots, t_0 \\ 1 & \text{if } t = t_0 + 1, \ldots, T \end{cases}$$

takes the form

$$y_t = \sum_{k=1}^{K} x_{tk}\beta_k + (x_{t,p+1}D)\delta_{p+1} + \ldots + (x_{t,K}D)\delta_K + e_t \tag{14.8}$$

where $t = 1, \ldots, T$.

Model (14.8) may be rewritten in two parts: one for the first subsample:

$$E[y_t] = \sum_{k=1}^{K} x_{tk}\beta_K$$

and one for the second part of the sample:

$$E[y_t] = \sum_{k=1}^{p} x_{tk}\beta_k + \sum_{k=p+1}^{K} x_{tk}(\beta_k + \delta_k)$$

Parameter δ_k measures the incremental change of the structural parameter connected with variable x_{tk} in the second part of the sample. Judge *et al.* (1980, in chapter 14) introduce an interesting alternative — a dummy variable-related approach that is sometimes easier and more convenient than a classical one and gives equivalent estimates of the parameters. (In chapter 16 of the same work, the authors provide a detailed discussion of a general dummy variable model. The parameter and variance estimation, as well as some alternative parameterization and testing techniques, are considered. An illustrative example is shown.) Of course, it is possible to extend the above model for more than two subsamples. Application of such a model is based on the assumption that there is more than one regression regime with constant parameters in each of them. It is also assumed that structural changes are rapid and abrupt.

If data with seasonal patterns are exogenously determined, and the data generating process is stable, then Zellner's (1962) seemingly unrelated regression model framework

gives a good and convenient tool for statistical modeling of a seasonal variation. Judge *et al.* (1982) discuss the example in which the seemingly unrelated model framework is employed to model statistical data with quarterly seasonality; several estimation methods and tests are compared. Kmenta (1986) and Johnston (1984 pp. 234-239) argue that the problem of seasonally varying parameters can be solved by means of dummy variables. Seemingly unrelated regressions seems to be easier and more straightforward in estimation and inference. Dummy variable models require additional calculations in order to obtain estimates of original parameters and their variances. The problem of seasonal variation of structural parameters can be easily solved whenever it is possible to identify economic or noneconomic factors that determine the seasonal variation. In such cases, model (14.1) can be directly employed with those factors used as explanatory variables.

Some variables are published by government agencies after deseasonalization by using a moving average process. Because deseasonalized variables contain very little or no information about the seasonal parameter variation, such a variable should not be used in dynamic statistical models. See, for example, Wallis (1974), Sims (1974), and Judge *at al.* (1985, chapter 10) for detailed discussions of consequences of using deseasonalized variables in econometric modeling. Havenner and Swamy (1981) consider consequences of using deseasonalized variables in the random coefficient model. In Zellner (1984) the procedures of deseasonalizing are reviewed.

For expository purposes, the time series context with two partitions will be considered without loss of generality, i.e., $N = 1$ will be set in model (14.1). Such a *switching regression model* can be easily extended to situations where there are both more than one statistical unit (time series of cross-sections) and/or a greater number of partitions. In the simplest case, observations for which different regression regimes hold are known. In other words, the sample may be split into two groups of T_1 and T_2 observations with $T = T_1 + T_2$. Those partitions may be both sequential or not in time.

A *piecewise regression model with known join point* may be formulated as

$$\begin{pmatrix} \mathbf{y_1} \\ \mathbf{y_2} \end{pmatrix} = \begin{pmatrix} \mathbf{X_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{X_2} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \mathbf{e_1} \\ \mathbf{e_2} \end{pmatrix} \tag{14.9}$$

Segments of model (14.9) are not necessarily joined. Some restrictions may be imposed to guarantee appropriate properties. For example, if some coefficients are expected not to change, condition that guarantee equality of the corresponding elements of β_1 and β_2 across the sample partitions may be imposed. A very important problem is that of joining the segments of model (14.9). Whenever two regression regimes are assumed to join at the point $t_0 \in [1, T]$, model (14.9) would be estimated subject to the condition

$$\mathbf{x}'_{t_0}(\beta_1 - \beta_2) = 0 \tag{14.10}$$

Smoothness restrictions can be formulated in some other way. A very important alternative formulation of (14.10) assumes that in the joining point the first and second derivates are equal. Models with segments being cubic polynomials in a single variable

(time) are widely known as cubic splines. Poirier (1973 and 1976) gives a detailed theoretical review of cubic splines methodology; Buse and Lim (1977) discuss some alternative techniques. Cubic splines have been extensively used by physical scientists, and in economics there were also several very interesting applications. Although cubic splines are very flexible functions and may approximate available data very well, they give very little information about the nature of both the process being modeled and the nature of structural change.

In *piecewise regression models with unknown join point*, the point of structural change is unknown and therefore is treated as an unknown parameter to be estimated. Goldfeld and Quandt (1973b) assume that in model (14.9) e_{1t} and e_{2t} are normally distributed with mean zero and variance σ_1^2 and σ_2^2, respectively. They assumed also that $(\beta_1, \sigma_1^2) \neq (\beta_2, \sigma_2^2)$. There are several possible ways to choose between the two regression regimes:

1. The deterministic choice is based on some variable compared with some unknown threshold. The basis of the choice may be the trend variable or other economic variable.

2. Choice is based on some unknown probabilities.

In the deterministic case where the switch occurs on the basis of a time index, it is assumed that the first regime holds for $t \leq t_0$ and the second elsewhere. The estimate of t_0 may be obtained by maximizing the likelihood function

$$L(\beta_1, \beta_2, \sigma_1^2, \sigma_2^2 | t_0) = (2\pi)^{-T/2} \sigma_1^{-t_0} \sigma_2^{-(T-t_0)} \exp\{-\frac{1}{2}\sigma_1^2 \sum_{t=1}^{t_0} (y_t - \mathbf{x}_t'\beta_1)^2 - \frac{1}{2}\sigma_2^2 \sum_{t=t_0+1}^{T} (y_t - \mathbf{x}_t'\beta_2)^2\} \tag{14.11}$$

The likelihood ratio test is used to determine whether two regression regimes are equal, i.e., whether there is one single regression over the entire sample. Brown *et al.* (1975) investigate the constancy of the regression relations over time by considering some functions of recursive residuals generated in moving regressions. They provide a test that answers the questions whether the regression is stable and where the instability occurs. Farley and Hinich (1970) and Farley *et al.* (1975) suggest some alternative techniques for investigating regression regimes equivalency. Assuming that the model

$$y_t = \mathbf{x}_t'\beta_t + e_t$$

where $\beta_t = \overline{\beta} + t\delta$, i.e., the model may be rewritten as

$$y_t = \mathbf{x}_t'\overline{\beta} + t\mathbf{x}_t'\delta + e_t$$

The likelihood ratio test of the hypothesis $\delta = 0$ is used to test the constancy of the structural parameters. Farley *et al.* (1975) look at properties of some stability tests by means of Monte Carlo experiments. Their test exhibits robustness with respect to gradual parameter shifts in one or more parameters. The power of this test is not great, and the test is reliable only with a large sample or a great shift.

The procedure for the case where shifts are determined by time can be applied to cases where a single economic variable (other than time) determines switching regression. Sample data have to be reordered according to increasing values of that variable. The problem becomes more complex in the presence of autocorrelation and/or lagged dependent variables. In recent years, much effort was devoted to evaluate Bayesian techniques for solving switching regression problems. See, for example, Ferreira (1975), Choy and Broemeling (1980), Smith and Cook (1980), Booth and Smith (1982), Holbert (1982), Ohtani (1982), Tsurumi (1982), and others. In particular, there is the question how *a priori* information can be incorporated in order to improve the quality of estimate of both switching points and structural parameters.

Whenever the deterministic choice of regression regimes is based on a variable other than time, frequently it is assumed that more than one variable exists with observations $z_{1t}, \ldots, z_{mt}$, $t = 1, \ldots, T$. Regimes are selected according to whether $\mathbf{z}_t'\tau \leq 0$, or $\mathbf{z}_t'\tau > 0$, where τ is an unknown coefficient vector. Goldfeld and Quandt (1973b) suggest the introduction of a dummy variable with values $D_t = 0$ if $\mathbf{z}_t'\tau \leq 0$, and $D_t = 1$ if $z_t'\tau > 0$. The two-regimes switching regression model became

$$y_t = \mathbf{x}_t'[(1 - D_t)\beta_1 + D_t\beta_2] + (1 - D_t)e_{1t} + D_t e_{2t} \tag{14.12}$$

where β_1, β_2, σ_1^2, σ_2^2, and D_t have to be estimated. Those unknown parameters can be found by maximizing the log likelihood function

$$\begin{aligned} \ell \;=\; & -\frac{1}{2}T\ln 2\pi - \frac{1}{2}\sum_{t=1}^{T}\ln[\sigma_1^2(1 - D_t)^2 + \sigma_2^2 D_t^2] \\ & - \frac{1}{2}\sum_{t=1}^{T}\frac{\{y_t - \mathbf{x}_t'[\beta_1(1 - D_t) + \beta_2 D_t]\}^2}{\sigma_1^2(1 - D_t)^2 + \sigma_2^2 D_t^2} \end{aligned} \tag{14.13}$$

where D_t is approximated by a continuous function — for example, by a probit function

$$D_t = \int_{-\infty}^{\mathbf{z}_t'\tau} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2\sigma^2}u^2\}\, du$$

The log likelihood function (14.13) is maximized with respect to β_1, β_2, τ, σ_1^2, and σ_2^2 upon replacing D_t by its approximation function. Frequently estimated values of $D_t = f(\mathbf{z}_t'\tau)$ are not exactly zero and one; the simplest solution is to partition the sample according to whether $D_t \leq 1/2$. Goldfeld and Quandt (1973b) suggest that, in the case when D_t is

not exactly zero or one (the discrimination is not perfect), one of the possible solutions is then to create two subsamples on the basis of whether $\mathbf{z}_t'\tau \leq 0$, or $\mathbf{z}_t'\tau > 0$. The likelihood ratio test may provide the answer whether there are separate regimes in each subsample.

Given that observations belong to different regression regimes with some unknown probabilities α and $(1-\alpha)$, the stochastic choice of regimes may be introduced. The log likelihood function is

$$\ell = \sum_{t=1}^{T} \ln g(y_t|\mathbf{x}_t)$$

where $g(y_t|\mathbf{x}_i)$ is the density function of y_t:

$$\begin{aligned} g(y_t|\mathbf{x}_t) &= \alpha f_1(y_t|\mathbf{x}_t) + (1-\alpha) f_2(y_t|\mathbf{x}_t) \\ &= \frac{\alpha}{\sqrt{2\pi\sigma_1^2}} \exp\{-\frac{1}{2\sigma_1^2}(y_t - \mathbf{x}_t'\beta_1)^2\} + \frac{1-\alpha}{\sqrt{2\pi\sigma_2^2}} \exp\{-\frac{1}{2\sigma_2^2}(y_t - \mathbf{x}_t'\beta_2)^2\} \end{aligned}$$

and may be maximized with respect to β_1, β_2, σ_1^2, σ_2^2, and α. In a more complex case, α can be the function of some exogenous variables.

Goldfeld and Quandt (1973a) propose an alternative method in which a Markov chain with explicitly specified transition probabilities is employed as a mechanism of choice of regression regimes. They consider several possible solutions where transition probabilities were fixed or nonstationary and were functions of exogenous variables. The likelihood function has to be maximized with respect to relevant variables. Tishler and Zang (1979) develop computationally simple approximation functions to the likelihood function. Swamy and Mehta (1975b) offer the Bayesian approach and some other generalizations. Lee and Porter (1984) suggest a model for the case in which sample separation information is imperfect.

14.5 Random-Parameter Models

In Section 14.4 the structural parameters were allowed to vary in a nonstochastic way. A model in which parameters are assumed to be random draws from some stochastic process will be considered in this section. First, it will be assumed that the process generating structural parameters is stationary, in the sense that it has a constant mean and variance. Such models are referred to as *random coefficient models.* They represent an improved alternative over dummy variable models because the number of parameters to be estimated is reduced.

14.5.1 Hildreth-Houck random coefficient models

Such models have the form [see Hildreth and Houck (1968)]

$$y_i = \mathbf{x}_i'\beta_i \tag{14.14}$$

where $i = 1, \ldots, N$; $\beta_i = \mathbf{Z}_i\tau + \mathbf{v}_i$, where $\mathbf{Z}_i = \mathbf{I}_k$ and $\tau = \overline{\beta}$; and

$$\beta_i = \overline{\beta} + \mathbf{v}_i \tag{14.15}$$

The parameter vector β_i contains population average parameter $\overline{\beta}$ and random disturbances $\mathbf{v}_i$. Disturbances $\mathbf{v}_i$ are independently distributed with zero means and covariance matrix $\mathbf{V}$. Note that the equation disturbance is indistinguishable from intercepts disturbance v_{1i} and therefore does not appear in equation (14.14). Model (14.14) may be considered as a variation of the previous model with $T = 1$ and with parameters assumed to be stochastic in the sense of (14.15). Combining (14.14) with (14.15) gives

$$y_i = \mathbf{x}_i'\overline{\beta} + e_i \tag{14.16}$$

where $e_i = \mathbf{x}_i'\mathbf{v}_i$, with distribution $e_i \sim (0, \sigma_i^2)$, $\sigma_i^2 = \mathbf{x}_i'V\mathbf{x}_i$. It is assumed that $E[\mathbf{v}_i] = \mathbf{0}$, $E[\mathbf{v}_i\mathbf{v}_i'] = \mathbf{V}$, and $E[\mathbf{v}_i\mathbf{v}_j'] = \mathbf{0}$ if $i \neq j$. Whenever $\mathbf{V}$ is the matrix with known elements, generalized least squares may be applied for the estimation of $\overline{\beta}$ in this model. The best, linear, unbiased, generalized least squares estimator for $\overline{\beta}$ may be obtained from

$$\hat{\overline{\beta}} = (\sum_{i=1}^{N} \sigma_i^{-2}\mathbf{x}_i\mathbf{x}_i')^{-1}(\sum_{i=1}^{N} \sigma_i^{-2}\mathbf{x}_i y_i) \tag{14.17}$$

It has the covariance matrix $(\sum_{i=1}^{N} \sigma_i^{-2}\mathbf{x}_i\mathbf{x}_i')^{-1}$. Griffiths (1972), Swamy and Mehta (1977), and Lee and Griffiths (1979) show that the best, linear, unbiased predictor for the vector of individual coefficients may be obtained from

$$\hat{\beta}_i = \overline{\beta} + \mathbf{V}\mathbf{x}_i(\mathbf{x}_i'\mathbf{V}\mathbf{x}_i)^{-1}(y_i - \mathbf{x}_i'\overline{\beta}) \tag{14.18}$$

However, since elements of matrix $\mathbf{V}$ are unknown rather than known, the way of finding their values is to be developed. Let $\mathbf{W}$ contain distinct, unique elements of $\mathbf{V}$, and let $\sigma_i^2 = \mathbf{x}_i'\mathbf{V}\mathbf{x}_i = \mathbf{Z}_i'\mathbf{W}$. Matrix $\mathbf{Z}_i$ may be found by calculating Kronecker product $\mathbf{x}_i' \otimes \mathbf{x}_i'$ and combining identical elements; matrix $\mathbf{Z}_i$ also contains explanatory variables, their second powers, and their combinations. [Note that the variance of y_i is a linear function of a set of exogenous variables, i.e., the Hildreth-Houck model has heteroscedastic errors. Such models require special treatment. Judge *et al.* (1985, in chapter 11), provide an excellent survey of those problems.]

Another problem connected with variance estimation arises. Since elements of $\mathbf{V}$ are variances and covariance, the estimation has to be restricted to positive values of the estimates. This problem is very difficult to handle because of its nonlinear nature. Hildreth-Houck (1968) consider the case of diagonal matrix $\mathbf{V}$; they advise replacing the negative estimates with zeros or using a quadratic programming estimator. The other

possible solution is an *ad hoc* adjustment in the estimated matrix $\mathbf{V}$, such that it is nonnegative definite. Schwaille (1982) finds a reparameterization to be useful in this case. Several different estimators are proposed by Swamy and Mehta (1975a, 1975b) and Srivastava *et al.* (1981).

A generally good estimator of the covariance matrix for random coefficients in the Hildreth-Houck model does not exist. Even provided this matrix is diagonal, usual estimators may not be nonnegative. Important for the application of the random coefficients models is testing for the randomness in the coefficients; the Breusch-Pagan (1979 and 1980) test seems to be the best for this purpose.

Several other tests in use with the heteroscedastic models are also relevant. Chow (1984, pp. 1239-1242) discusses the applicability of his test, the likelihood ratio test [see Chernoff (1954), Moran (1970), and Gourieroux *et al.* (1982)]; the Lagrangian multiplier test of Silvey (1959) along with the score test of Rao (1972, p. 417); the test of Pagan and Tanaka (1979); and the test of LaMotte and McWhorter (1978). Chow concludes that none of them is good enough and further work is required to obtain computationally simple, uniformly most powerful test statistic with known distribution in a small sample. Raj*et al.* (1980) consider distribution moments in a finite sample. Griffiths *et al.* (1979), Liu (1981), and Liu and Hanssens (1981) evaluate the Bayesian approach to the Hildreth-Houck model.

14.5.2 Return-to-normality models

Harvey and Phillips (1982) suggest a model referred to as the *return to normality model*, which is a generalization of the Hildreth-Houck random coefficient model. This term was first used in the work of Schaefer *et al.* (1975), who were investigating a model with the parameters generated by a first-order autoregressive process. It is specifically suited for use with time series data. It enables releasing the assumption that the coefficients in the regression model are constant over time — an assumption that is frequently unreasonable.

The dynamic parameters of this model follow a stationary process with a fixed but unknown mean. Harvey and Phillips consider the model

$$y_t = \mathbf{x}_t'\beta_t \tag{14.19}$$

where

$$\beta_t - \overline{\beta} = \Phi(\beta_{t-1} - \overline{\beta}) + e_t \tag{14.20}$$

with $t = 1, \ldots, T$; y_t is the observation on the dependent variable; $\mathbf{x}_t$ is a $(K \times 1)$ vector of nonstochastic observations; β_t is a $(K \times 1)$ vector of stochastic parameters, including the fixed component $\overline{\beta}$; Φ is the $(K \times K)$ matrix of parameters with characteristic roots less than 1 in absolute value; $\mathbf{e}_t$ is the $(K \times 1)$ vector of disturbances, which is assumed to follow a multivariate normal distribution with mean vector zero and a covariance matrix $E[\mathbf{e}_t\mathbf{e}_t'] = \sigma^2 Q$, $E[\mathbf{e}_t\mathbf{e}_s'] = \mathbf{0}$ for $t \neq s$.

Note that model (14.19) is written without an error term. It is assumed that the first element of $\mathbf{x}_t$ is unity so that the variance of the first parameter is indistinguishable from the equation error variance, and thus it is ignored. Not all of the elements in coefficient vector β_t need to be time-varying. Some of them may be fixed, others may be random. In the case of the existence of fixed parameters, they drop out of expression (14.20). Note that if $\mathbf{\Phi} = \mathbf{0}$, (i.e., the parameters are random rather than dynamic), the model reduces to the Hildreth-Houck random coefficient model, and thus represents a dynamic generalization.

Harvey and Phillips (1982) suggest a full maximum likelihood method and two-step estimation procedures, both of them based on Kalman filtering [Kalman (1960) and Kalman and Bucy (1961)], linked with the recursive residuals technique suggested by Phillips and Harvey (1974), and by Brown*et al.* (1975). In a Monte Carlo experiment, Harvey and Phillips compare small-sample properties of the maximum likelihood estimator gained over ordinary least squares and two-step estimated generalized least squares. The latter estimator provided substantial improvement over the OLS.

A very important generalization of (14.19) is obtained by using

$$A(L)(\beta_t - \overline{\beta}) = \mathbf{e}_t \tag{14.21}$$

instead of (14.20) as the parameter generating process; here, A(L) is a rational function of finite polynomials, which implies that $(\beta_t - \overline{\beta})$ follows a multivariate ARMA process. This model covers a number of important special cases. Burnett and Guthrie (1970), Rosenberg (1972, 1973a, 1973b, and 1973c), Cooley and Prescott (1973a, 1973b, 1973c, and 1976), Harvey and Phillips (1979), and Swamy and Tinsley (1980) consider models of this class. Pagan (1980) discusses sufficient conditions for asymptotic identification of such models, assuming $(\beta_t - \overline{\beta})$ is stationary. He also establishes sufficient conditions for the consistency and asymptotic normality of maximum likelihood estimators without assuming stationarity, but by assuming asymptotic identifiability. Liu and Hanssens (1981) consider estimation of (14.19) from the Bayesian perspective using noninformative priors.

14.5.3 Swamy random coefficient models

In the Swamy random coefficient model, a $(K \times 1)$ response parameter vector for each individual, β_i, is regarded as a random vector drawn from the probability distribution with mean β and the covariance matrix $\mathbf{V}$; see Swamy (1970, 1971, 1973, and 1974). Note the similarity of this model with the Hildreth-Houck model. The latter was designed to model cross-sectional data; the Swamy random coefficient model is relevant for time series of cross-sectional data. In both it is assumed that the process generating values of the dependent variable vary and that this variation can be confined in the structural parameters of the linear model. Because of the estimation requirements, in both models a certain structure of the parameter variation is to be specified. The nature of parameter variation is continuous rather than an abrupt, unique switch.

The Swamy random coefficient model for ith unit may be written as

$$\mathbf{y}_i = \mathbf{X}_i(\overline{\beta} + \mu_i) + \mathbf{e}_i \tag{14.22}$$

$$\beta_i = \overline{\beta} + \mu_i \tag{14.23}$$

where $i = 1, \ldots, N$, $E[\mu_i] = \mathbf{0}$, $E[\mu_i \mu_i'] = \mathbf{V}$, and $E[\mu_i \mu_j'] = 0$ for $j \neq i$.

Every unit in the sample has a unique parameter vector β_i. The parameters for each individual are constant in time and have a common mean parameter vector $\overline{\beta}$, but a different disturbances vector μ_i. Several alternative sets of assumptions about $\mathbf{e}_i$ yield different model variations. Judge *et al.* (1985, chapter 12) list a number of such assumptions of varying degrees of complexity in the context of seemingly unrelated regression equations, which is a nonstochastic counterpart of Swamy's random coefficient model. All sets of assumptions listed there may be also used for Swamy's model. Here a set of assumptions defining a relatively simple case will be considered.

Let $E[\mathbf{e}_i \mathbf{e}_i'] = \sigma_i^2 \mathbf{I}$ and $E[\mathbf{e}_i \mathbf{e}_j'] = \mathbf{0}$ for $j \neq i$. This means that the disturbances are heteroscedastic across individuals, but that the disturbances corresponding to different units are uncorrelated; there is no serial correlation. In such a case if β_i were a fixed parameter vector, the least squares estimator $\mathbf{b}_i = (\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{y}_i$ would be best, linear, unbiased. In order to estimate the mean parameter vector $\overline{\beta}$ and predict the individual parameter vectors β_i, first one must estimate the variances upon which the generalized least squares (GLS) estimator for $\overline{\beta}$ and the best, linear, unbiased predictor for β_i depend. Testing the hypothesis that $\mathbf{V} = \mathbf{0}$ may answer whether structural parameters vary.

After including all NT observations, model (14.22) may be rewritten as

$$\mathbf{y} = \mathbf{X}\overline{\beta} + \mathbf{Z}\mu + \mathbf{e} \tag{14.24}$$

where $\mathbf{Z}$ is a $(NT \times NK)$ block diagonal matrix with blocks $\mathbf{X}_i$; $i = 1, \ldots, N$; $\mathbf{y}$ is an $(1 \times NT)$ observation vector on dependent variable; $\mu' = \{\mu_i'\}$, $i = 1, \ldots, N$; $\overline{\beta}$ is a $(1 \times NK)$ vector of unknown, fixed parameters to be found; and $\mathbf{e} = \{\mathbf{e}_i'\}$, $i = 1, \ldots, N$. Composite disturbance $(\mathbf{Z}\mu + \mathbf{e})$ has the block diagonal covariance matrix $\Phi = E[(\mathbf{Z}\mu + \mathbf{e})(\mathbf{Z}\mu + \mathbf{e})']$, with the diagonal block given by $\Phi_{ii} = \mathbf{Z}_i \mathbf{V} \mathbf{Z}_i' t + \sigma_{ii}^2 \mathbf{I}$. It is convenient to write the GLS estimator for $\overline{\beta}$ as

$$\hat{\overline{\beta}} = (\mathbf{X}'\Phi^{-1}\mathbf{X})\mathbf{X}'\Phi^{-1}\mathbf{y} = [\sum_{j=1}^{N}(\mathbf{X}_j'\Phi_{jj}^{-1}\mathbf{X}_j)]^{-1}\sum_{i=1}^{N}(\mathbf{X}_i'\Phi_{ii}^{-1}\mathbf{y}_i) = \sum_{i=1}^{N}\mathbf{W}_i\mathbf{b}_i \tag{14.25}$$

where

$$\begin{aligned} \mathbf{W}_i &= \{\sum_{j=1}^{N}[\mathbf{V} + \sigma_{jj}(\mathbf{X}_j'\mathbf{X}_j)^{-1}]^{-1}\}^{-1}[\mathbf{V} + \sigma_{ii}(\mathbf{X}_i'\mathbf{X}_i)^{-1}]^{-1} \\ \mathbf{b}_i &= (\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{y}_i \end{aligned} \tag{14.26}$$

Estimator (14.25) has the usual GLS properties. Judge *et al.* (1985) argue that the predictor of β_i given in equality (14.26), based on the matrix results of Rao (1965a, p. 29), is convenient for computational purposes. It requires a matrix inversion of the order K, which is especially important with large T. The GLS estimator may be interpreted as the matrix-weighted average of estimators (14.26) with weights inversely proportional to their covariance matrices. Mundlak (1978a) provides a different interpretation of this estimator.

An important task is to predict individual components of β_i. Having done this, it is possible to predict future values of dependent variable for each individual and to describe its behavior. Several predictors have been proposed in the literature.

Swamy (1970, 1971) and Lee and Griffiths (1979) suggest some alternative approaches. Most widely known are their best, linear, unbiased predictors — BLUP estimators; others may be obtained by minimizing some quadratic function with respect to $\overline{\beta}$ and β_i. Smith (1973) and Leamer (1978) consider Bayesian solutions. Both parameter vectors $\overline{\beta}$ and β_i are dependent on the unknown variances $\mathbf{V}$ and σ_{ii}. Therefore their estimates are required.

Using results of Rao (1965b), Swamy (1970) suggests consistent estimators for both variances. The GLS estimator for $\overline{\beta}$, based on those variance estimates, is consistent and asymptotically effective. The problem is that the estimator for $\mathbf{V}$ may not be nonnegative definite; see Dielman *et al.* (1980). Swamy (1971) discusses this problem and gives some suggestions about how to handle this situation. He argues that negative variance estimates may result from incorrectly specified assumptions about the form of the disturbance covariance matrix (for example, about homoscedasticity, serial correlation, or contemporaneous correlation). Swamy (1974) suggests appropriate corrections for the violations of accepted assumptions.

The other possible source of negative variance estimates is that certain coefficients are not random. The model containing both fixed and random coefficients is referred to as a mixed random coefficient model. It was first proposed by Swamy (1971, pp. 143-155) and is analytically examined by Rosenberg (1973b), Mundlak (1978a, 1978b, 1978c), and Dielman (1980). The solutions given by Swamy are not obvious and can destroy properties of estimators. Dielman (1980) reviews available statistical procedures to test for the possibility that parameters are not random. Swamy (1970, 1971) gives some alternative suggestions about how to test if individual coefficient vectors are not random and are all identical to the mean. Rao (1972) gives conditions under which the estimator for $\overline{\beta}$ has a finite mean and is unbiased. Swamy (1971, 1973) and Rosenberg (1973b) discuss the possibility of applying maximum likelihood techniques for the parameter estimation.

The assumptions that have been made about $\mathbf{e}_i$ could be regarded as fairly restrictive. Parks (1967) adopts an alternative set of assumptions that relax some of these restrictions. He assumes that the disturbance for each unit follows a first-order autoregressive process, and that there exists a contemporaneous correlation. (Parks adopts this set of assumptions for the fixed coefficient model.) Swamy (1973, 1974) considers Parks's assumptions when introduced into the random coefficient model. Another extension is given by Swamy (1974)

for the case when X contains a lagged dependent variable. Rosenberg (1973b) considers estimation when the covariance matrix could be singular. Swamy (1973, 1974) attempts to evaluate estimators that may be biased, but with a lower mean square error.

Mundlak (1978a, 1978b, 1978c) suggests that the regression coefficients can be always regarded as random; but in the case when parameters β_i are regarded as being fixed and different (in a seemingly unrelated regression framework), the inference is conditional on the coefficients in the sample. The random coefficient model uses additional information provided by an assumption about the randomness of the coefficients. It should be expected that if the assumption is true, the estimates will be more efficient. In some cases, variable coefficients perform a correlation with the explanatory variables; then Swamy's assumptions are unreasonable and the GLS estimator for mean coefficient vector $\overline{\beta}$ will be biased. Mundlak (1978a) advises one to incorporate into the model any dependence of the coefficients on the explanatory variables.

Pudney (1978) provides the procedure to test whether variable coefficients and explanatory variables are uncorrelated. Chamberlain (1982) considers further properties of the estimators under these circumstances. Zellner (1969a, 1969b) shows that if the coefficient vectors of the individual units satisfy Swamy's assumptions, a macro coefficient estimator will not possess an aggregation bias. Applications of Swamy's model can be found in Swamy (1971), Boot and Frankfurter (1972), Feige and Swamy (1974), Boness and Frankfurter (1977), Mehta*et al.* (1978), and Hendricks*et al.* (1979). Johnson and Lyon (1973) describe a simulation study with stochastic explanatory variables. Swamy (1971, pp. 1-23) and Spjotvoll (1977) survey other random coefficient models.

14.5.4 The Hsiao random coefficient model

This is an extension of Swamy's model where all coefficients may vary both over time and over individuals; see Hsiao (1974, 1975).

$$y_{it} = \sum_{k=1}^{K} (\beta_k + \mu_{ki} + \pi_{kt}) x_{kit} + e_{it} \tag{14.27}$$

with $i = 1, \ldots, N$; $t = 1, \ldots, T$. More compactly for the ith unit, equation (14.27) may be rewritten as

$$\mathbf{y}_i = \mathbf{X}_i \overline{\beta} + \mathbf{X}_i \mu_i + \mathbf{Z}_i^o \pi + \mathbf{e}_i \tag{14.28}$$

where $\mathbf{y}_i$ and $\mathbf{X}_i$ have dimensions $(T \times 1)$ and $(T \times K)$, respectively; $\mu_i = (\mu_{1i}, \ldots, \mu_{Ki})'$; $\pi' = (\pi_1', \ldots, \pi_T')$; $\pi_t' = (\pi_{1t}, \ldots, \pi_{Kt})'$; $\mathbf{e}_i = (e_{i1}, \ldots, e_{iT})'$; and $x_{it}' = (x_{1it}, \ldots, x_{Kit})'$ is an element of block diagonal matrix $\mathbf{Z}_i^o$ which is of order $(T \times TK)$. Hsiao assumes that $E[\mathbf{e}_i] = \mathbf{0}$, $E[\mu_i] = \mathbf{0}$, $E[\pi_t] = \mathbf{0}$, $E[\mathbf{e}_i, \mathbf{e}_i'] = \sigma_e^2 \mathbf{I}$, $E[\mathbf{e}_i, \mathbf{e}_j'] = \mathbf{0}$ for $i \neq j$, $E[\mu_i, \mu_i'] = \mathbf{V}$, $E[\mu_i, \mu_j'] = \mathbf{0}$ for $i \neq j$, $E[\pi_t, \pi_t'] = \mathbf{A}$, $E[\pi_t, \pi_s'] = \mathbf{0}$ for $t \neq s$. It is assumed also that μ_i, π_t, and $\mathbf{e}_i$ are all uncorrelated, and that covariance matrices $\mathbf{V}$ and $\mathbf{A}$ are diagonal with elements v_k and α_k, respectively.

Rewriting (14.28) more compactly to include all NT observations yields

$$\mathbf{y} = \mathbf{X}\overline{\beta} + \mathbf{Z}\mu + \mathbf{Z}^o\pi + e \tag{14.29}$$

where $\mathbf{y}' = (\mathbf{y}_1', \ldots, \mathbf{y}_N')$; $\mathbf{X}_t' = (X_1', \ldots, X_N')$; $\mathbf{Z}$ is block diagonal with $\mathbf{X}_i$ as ith diagonal block; $\mathbf{Z}^{o\prime} = (\mathbf{Z}^{o\prime}{}_1, \ldots, \mathbf{Z}^{o\prime}{}_N)$; $\mu' = (\mu_1', \ldots, \mu_N')$; and $\mathbf{e}' = (\mathbf{e}_1', \ldots, \mathbf{e}_N')$. Provided that equation (14.26) is reparameterized to eliminate redundant parameters, and that NT is sufficiently large, if μ and π are regarded as fixed parameters, μ_i, π, and $\overline{\beta}$ may be estimated by applying the ordinary least squares to (14.26). Since the matrix $(\mathbf{X}, \mathbf{Z}, \mathbf{Z}^o)$ is of dimension $[NT \times (T + N + 1)K]$, and of rank $[(T + N - 1)K]$, $2K$ parameters are redundant. It is convenient to drop $(\mu_{1N}, \ldots, \mu_{KN}, \pi_{1T}, \ldots, \pi_{KT})$, provided that corresponding columns of $\mathbf{Z}$ and $\mathbf{Z}^o$ are also eliminated. The minimal number of observations required is $NT > [(T + N - 1)K]$.

Such estimation when μ and π are random, with the above-listed assumptions, requires an estimate of the covariance matrix of the composite disturbance. Covariance matrices $\mathbf{V}$ and $\mathbf{A}$ and variance σ_e^2 are assumed to be known. With these assumptions, an estimate of the covariance matrix of the composite disturbance may be obtained from

$$\begin{aligned} \Phi &= E[(\mathbf{Z}\mu + \mathbf{Z}^o\pi + \mathbf{e})(\mathbf{Z}\mu + \mathbf{Z}^o\pi + \mathbf{e})'] \\ &= \mathbf{Z}(I_N \otimes \mathbf{V})\mathbf{Z}' + \mathbf{Z}^o(I_T \otimes \mathbf{A})\mathbf{Z}'^o + \sigma_e^2\mathbf{I}_{NT} \end{aligned} \tag{14.30}$$

Then the GLS estimator for $\hat{\overline{\beta}} = (\mathbf{X}'\Phi^{-1}\mathbf{X})^{-1}\mathbf{X}'\Phi^{-1}\mathbf{y}$ is the best, linear, unbiased one for $\overline{\beta}$, and has the covariance matrix $(\mathbf{X}'\Phi^{-1}\mathbf{X})^{-1}$. When NT is large, inversion of Φ could be computationally difficult. Hsiao (1974) provides a computational procedure where the largest order of inversion is reduced to $\max\{NK, NT\}$, which still may be quite large. Wansbeek and Kapteyn (1982) suggest a convenient inversion procedure for their model.

Lee and Griffiths (1979) show the best, linear, unbiased predictor for the random component associated with each cross-sectional unit:

$$\hat{\mu} = (\mathbf{I}_N \otimes \mathbf{V})\mathbf{Z}'\Phi^{-1}(\mathbf{y} - \mathbf{X}\overline{\beta}) \tag{14.31}$$

Covariance matrices $\mathbf{V}$ and $\mathbf{A}$ and variance σ_e^2 are assumed to be known, although typically they are unknown. Hsiao has found a minimum norm, quadratic, unbiased estimator (MINQUE) and provided maximum likelihood procedure for variances estimation. Those estimators are used to construct a feasible Aitken estimator of $\overline{\beta}$.

Alternatively, Hildreth-Houck techniques may be used. Hildreth and Houck (1968) list conditions under which their variance estimators are consistent. Hsiao (1974) gives sufficient conditions for the consistency of the estimated GLS estimator for $\overline{\beta}$, based on Hildreth-Houck's variance estimates.

Kelejian and Stephan (1983) extend some of Hsiao's asymptotic results. Problems with those estimators are the same as those discussed earlier: they may have negative values.

All comments made above on this problem also apply here. As before, μ_{ki} and π_{kt} may be assumed to be either fixed parameters to be estimated or random variables. Generally, it is assumed that they are random; sometimes it is convenient to estimate them as fixed parameters. Pudney (1978) points out that it may be reasonable to assume that one of the components is fixed and the other is random. To estimate variances, N and T have to be sufficiently large. Otherwise, any estimate of the variance will be unreliable. In such a situation it would be preferable to include appropriate dummy variables, and an inference conditionally on the sample variance. Note that if the time effects are replaced by dummy variables, the model becomes identical to the Swamy random coefficient model.

There are several alternative models considered in the literature. Singh and Ullah (1974), Swamy and Mehta (1975a, 1975b, 1977), and Pudney (1978) discuss the model

$$y_{it} = \sum_k (\beta_k + \mu_{ki} + e_{kit}) x_{kit} \tag{14.32}$$

The disturbance e_{1it} replaces the model's disturbance. Time effect π_{kt} has been replaced by random component e_{kit}; this element is not restricted to be the same for units in a given time period. Swamy and Mehta (1975a, 1975b, 1977) modify Hsiao's assumption that covariance matrix $\mathbf{V}$ is diagonal; i.e., they allow contemporaneous correlation among the coefficients. Swamy and Mehta construct an approximate minimum average risk linear estimator for $\overline{\beta}$ in their model. (The estimator is approximate because it uses the estimates of the variance and covariance components, not the true values.)

Both models mentioned above assume that random coefficients vary around a constant mean. Rosenberg (1973c), Johnson and Rausser (1975), Harvey (1978, 1981, and 1982), and Liu and Hanssens (1981) analyze other models where parameters vary systematically over time. A special case of Hsiao's model, where random coefficients are associated only with time invariant and individual invariant variables, is studied by Wansbeek and Kapteyn (1978, 1981, 1982).

Other references on random coefficient models with parameters generated in a stationary process include Burnett and Guthrie (1970), Belsley (1973c), Cooper (1972, 1973), Sarris (1973), Sant (1977), Pagan (1980), Rausser *et al.* (1982), and Chow (1984, chapter 21), as well as Chamberlain (1984, chapter 22). Important collections of papers on the topic are a special issue of the *Annals of Economic and Social Measurement* (1973, no. 2); a special issue of the *Annales de l'INSEE* (1978), entitled "The Econometrics of Panel Data", edited by Mazodier; and a special issue of the *Journal of Econometrics* (1982), entitled "Econometrics of Longitudinal Data", edited by Heckman and Singer.

14.6 Nonstationary Random-Parameter Models

Up to now it was assumed that parameters in the econometric model have a constant mean. This assumption can now be replaced by the assumption that parameters are generated in a nonstationary random process. Contrary to the models with parameters

generated in a stationary random process, here coefficients do not have a constant mean and variance. Thus, they may vary systematically over observations. This means that a less restrictive structure is placed on the parameter variation. Such models are suitable to describe systematic variation over time.

14.6.1 Cooley-Prescott models

Cooley and Prescott (1973a, 1973b, 1973c, 1976) suggest a model where parameters vary from one time period to another on the basis of a nonstationary process. They consider the model

$$y_t = \mathbf{x}_t' \beta_t \tag{14.33}$$

where $t = 1, \ldots, T$, $\mathbf{x}_t$ is a $(K \times 1)$ vector of nonstochastic observations, and β_t is a $(K \times 1)$ parameter vector subject to stochastic variation. Parameter variation is modeled as

$$\beta_t = \beta_t^P + \mathbf{u}_t \tag{14.34}$$

where

$$\beta_t^P = \beta_{t-1}^P + \mathbf{v}_t \tag{14.35}$$

It is assumed that the parameter variation is of two types: permanent and transitory. Permanent component β_t^P, of the vector β_t, allows some tendency in the parameter variation. The terms $\mathbf{u}_t$ and $\mathbf{v}_t$ are independent, normal, random vectors with mean vectors zero and covariance matrices $\mathbf{U}_u$ and $\mathbf{V}_v$, where $E[\mathbf{u}_t \mathbf{u}_t'] = (1-\tau)\sigma^2 \mathbf{U}_u$ and $E[\mathbf{v}_t \mathbf{v}_t'] = \tau \sigma^2 \mathbf{V}_v$. Covariance matrices $\mathbf{U}_u$ and $\mathbf{V}_v$ are assumed to be known up to the scale factor and normalized, i.e., the element corresponding to the intercept is unity — the first regressor is the constant term. The transitory component of the corresponding parameter's variation plays the role of the additive disturbance in the regression equation. Element τ measures the relative importance of the permanent and transitory changes. The close-to-one value of τ means relatively large permanent changes and relatively small transitory changes.

Cooley and Prescott evaluate the maximum likelihood estimation procedure, which provides consistent estimates of τ and asymptotically efficient estimates of $\beta_{t+1}^P(\tau)$. The nature of the model precludes any notion of the consistent estimation of β_{t+1}^P. Cooley and Prescott (1976, pp. 172-173) discuss the possibility of testing hypotheses about τ, and they evaluate asymptotic distribution of τ. Although it is relatively simple to estimate, interpret, and infer in the Cooley-Prescott model, its application is not straightforward. In particular the need to specify matrices $\mathbf{U}_u$ and $\mathbf{V}_v$ may be very complicated. They have to be assumed on the basis of theoretical considerations, which in turn presumes the ability to specify the relative variability of the parameters.

Similar models have been considered by Belsley (1973a, 1973b), Cooper (1973), Sarris (1973), Sant (1977), Rausser and Mundlak (1978), and Rausser *et al.* (1982).

14.6.2 Convergent parameter models

Rosenberg (1973c) evaluates a convergent parameter model, which is devoted to investigating the time series of cross-sections. It is assumed that parameters of each unit may vary randomly in time, but they tend to converge to some value (population parameter). This feature differentiates Rosenberg's model from that of Cooley-Prescott, where parameters also vary over time but in some systematic way and were not convergent to any value. In the interest of simplicity, the one-unit variant will be shown.

The basic model has the form

$$y_t = \mathbf{x}_t'\beta_t + e_t \tag{14.36}$$

where $t = 1, \ldots, T$; e_t are independent with the same normal distribution with mean zero and $E[e_t^2] = \sigma^2$. Rosenberg models parameter variation in the following way:

$$\beta_t = \overline{\beta} + \mathbf{A}(\beta_{t-1} - \overline{\beta}) + \mathbf{v}_t = \overline{\beta}(\mathbf{I} - \mathbf{A}) + \mathbf{A}\beta_{t-1} + \mathbf{v}_t \tag{14.37}$$

where $\overline{\beta}$ is a $(K \times 1)$ vector of the mean parameters in the population; $\mathbf{A}$ is a $(K \times 1)$ convergence matrix with elements $0 \leq \delta_i \leq 1$, $i = 1, \ldots, K$. Convergence rates δ_i show the relative difference between $\overline{\beta}$ and β_{t-1}, which still exist at time point t. The $(K \times 1)$ disturbance vector $\mathbf{v}_t$ has mean vector $E[\mathbf{v}_t] = \mathbf{0}$ and the contemporaneous covariance matrix $E[\mathbf{v}_t\mathbf{v}_t'] = \mathbf{V}_v$.

Rosenberg evaluates maximum likelihood and Bayesian estimation techniques for the general model. For the sake of simplicity, it will be assumed that $\delta_i = \delta$ for all $i = 1, \ldots, K$; i.e., formula (14.37) could be rewritten:

$$\beta_t = (1 - \delta)\overline{\beta} + \delta\beta_{t-1} + \mathbf{v}_t \tag{14.38}$$

or

$$(1 - \delta L)\beta_t = (1 - \delta)\overline{\beta} + \mathbf{v}_t \tag{14.39}$$

which eventually leads to a model

$$y_t = \mathbf{x}_t'[(1 - \delta)\overline{\beta}] + \delta y_{t-1} + w_t \tag{14.40}$$

where

$$w_t = \mathbf{x}_t'\mathbf{v}_t + e_t - \delta e_{t-1} \tag{14.41}$$

Estimation of this kind of model is similar to the models with infinite geometric lag but with a much more complicated error structure. The estimation of infinite geometric lag models is described, for example, in Judge *et al.* (1980, chapter 16).

14.6.3 Kalman filter models

One of the most general models with random parameters generated in a nonstationary stochastic process is the Kalman filter model. Many models which have been presented here may be regarded as special cases of the Kalman filter models; see, for example, Belsley (1973c), Cooper (1973), Sarris (1973), Sant (1977), and Rausser and Mundlak (1978).

The basic model may be written as

$$y_t = \mathbf{x}_t'\beta_t + e_t \tag{14.42}$$

where $t = 1, \ldots, T-1$; with the following general parameter variation structure:

$$\beta_{t+1} = \Phi\beta_t + \mathbf{v}_{t+1} \tag{14.43}$$

where Φ is the $(K \times K)$ matrix of transition probabilities; $E[\mathbf{v}_t] = \mathbf{0}$; $E[\mathbf{v}_t\mathbf{v}_t'] = \mathbf{V}_v$; e_s and $\mathbf{v}_t$ are uncorrelated for all t and s. Assuming that T, $\mathbf{V}_v$, and β_0 are known and performing some calculations, (14.43) may be rewritten in compact matrix form as

$$\beta = \Phi_1\beta_0 + \Phi_2\mathbf{v} \tag{14.44}$$

For estimation purposes, the values of Φ, $\mathbf{V}_v$, and β_0 have to be known. Not much could be said about the first two matrices, although Sarris (1973) provides some guidelines on how they could be specified. Provided they are somehow known, being specified in a theoretical or in another way, without knowledge about β_0 one is in a situation where the nature of the development is known, but the starting point of the path is not known. Although vector β_0 cannot be specified in any theoretical way, on the basis of *a priori* knowledge, there are some possibilities for finding starting points on the basis of the available statistical sample. Cooper (1973) suggests a comfortable reparameterization of the model that enables one to make a maximum likelihood estimation of unknown parameters, provided that matrices Φ and $\mathbf{V}_v$ are known. Generally, it may be stated that estimation of the Kalman filter model is not satisfactorily solved.

14.7 Summary and Conclusions

In modeling economic processes by means of the varying or random coefficient model, several problems have to be taken into consideration. For modeling changes over time, one may

1. Assume a constant correlation structure of the disturbances and use the error component model.
2. Assume that disturbances are generated in some autoregressive or moving average process.

3. Choose a dummy variable model and regard inference as conditional on changes in the sample.

A decision has to be made about the nature of the model: are all coefficients likely to vary, or is it reasonable to assume that some of them are constant, or is only the intercept able to capture behavioral differences in the sample? If the coefficients are likely to vary, does this variation depend on the explanatory variables? In such a situation, an applied researcher may choose a dummy variable model or the seemingly unrelated equations. Once there exists a correlation between individual effects and the explanatory variables, an error component model or Swamy's random coefficient model may be reasonable.

A very important factor is the size of the sample. Having a small sample means that the random assumptions requiring some variance estimates are unlikely to be reliable. It may be better to treat the coefficients as fixed even when the random assumption is reasonable. In the model where the parameters are assumed to be random the relative size of the sample has an important bearing on the finite sample reliability of variance estimates and, consequently, on the estimated generalized least squares estimators for the slope coefficients.

A number of model specification tests may be used to help choose between model specifications. Once a model has been specified, there are additional problems concerning the most efficient estimation procedure and the testing of hypotheses about parameters. The problem of testing the constancy of the coefficients, with the varying coefficient model serving as the alternative, is not completely resolved, although many tests have been suggested.

It is necessary to remember that there is a great danger of misspecification. The model chosen is only as good as the structural information introduced on the parameters' variation. Theoretically, by introducing more information about the nature of the process being modeled, the model should be more informative; but because the information imposed may not be true, the danger of misspecification is great. The general recommendation is that in applied work a judicious use of tests with *a priori* knowledge about the nature of the problem should be combined.

The problem of estimating models with changing parameters deserves further study. In particular, finite-sample properties as well as pretest estimator properties have to be further investigated. The same may be applied to finite-sample properties of a number of test statistics used in the regression model with varying parameters. Since the variances necessary to estimate parameters by means of generalized least squares are unknown rather than known — and maximum likelihood estimates are used — their finite-sample properties and the sampling distributions of the slope estimates based on them are not sufficiently known.

Periodical Abbreviations

AESM	Annals of Economic and Social Measurement
AI	Annales de l'INSEE
AMS	Annals of Mathematical Statistics
ANS	Annals of Statistics
APS	Applied Statistics
B	Biometrika
BS	Biometrics
E	Econometrica
EL	Economic Letters
ESQ	Economis Studies Quarterly
IER	International Economic Review
JASA	Journal of the American Statistical Association
JBE	Journal of Basic Engineering
JE	Journal of Econometrics
JET	Journal of Economic Theory
JF	Journal of Finance
JFQA	Journal of Financial and Quantitative Analysis
JISA	Journal of the Indian Statistical Association
JMCB	Journal of Money, Credit and Banking
JSCS	Journal of Statistical Computation and Simulation
MOR	Methods of Operations Research
MOSS	Mathematische Operationsforschung, ser. Statistik
PS	Przeglad Statystyczny
RES	Reviev of Economic Studies
RSSB	Journal of the Royal Statistical Association, Series B
SH	Statistische Hefte
T	Technometrics
TRES	The Review of Economics and Statistics

References

Amemiya, T. (1971), The estimation of variances in a variance-components model, IER, **12**, 1-13.

Anderson, T.W. and Hsiao, C. (1981), Estimation of dynamic models with error components. JASA, **76**, 598-606.

Anderson, T.W. and Hsiao, C. (1982), Formulation and estimation of dynamic models using panel data. JE, **18**, 47-82.

Arora, S.S. (1973), Error components regression models and their applications. AESM, **2**, 451-461.

Avery, R.B. (1977), Error components and seemingly unrelated regressions. E, **45**, 199-209.

Balestra, P. and Nerlove, M. (1966), Pooling cross-section and time-series data in the estimation of a dynamic model: The demand for natural gas. E, **34**, 585-612.

Baltagi, B. (1980), On seemingly unrelated regressions with error components. E, **48**, 1547-1551.

Baltagi, B. (1981), Pooling: An experimental study of alternative testing and estimation procedures in a two-way error component model. JE, **17**, 21-49.

Battese, G.E. and Fuller, W.A. (1982), *An Error Components Model for Prediction of County Crop Areas Using Survey and Satelitte Data.* University of New England Working Paper No. 15, Armidale, Australia.

Belsley, D.A. (1973a), A test for the systematic variation in regression coefficients. AESM, **2**, 494-499.

Belsley, D.A. (1973b), On the determination of the systematic parameter variation in the linear regression model. AESM, **2**, 487-494.

Belsley, D.A. (1973c), The applicability of the Kalman filter in the determination of systematic parameter variation. AESM, **2**, 531-533.

Berzeg, K. (1979), The error components model: Conditions for the existence of the maximum likelihood estimates. JE, **10**, 99-102.
Bhargava, A. and Sargan, J.D. (1983), Estimating dynamic random effects models from panel data covering short time period. E, **51**, 1635-1659.
Bhargava, A., Franzini, L., and Narendranathan, W. (1982), Serial correlation and the fixed effects model. RES, **49**, 533-550.
Biorn, E. (1981), Estimating economic relations from incomplete cross-section/time-series data. JE, **16**, 221-236.
Boness, J.A. and Frankfurter, G.M. (1977), Evidence of non-homogeneity of capital costs within risk classes. JF, **32**, 775-787.
Booth, N. and Smith, A.F.M. (1982), A Bayesian approach to retrospective identification of change points. JE, **19**, 7-22.
Boot, J. and Frankfurter, G.M. (1972), The dynamics of corporate debt management, decision rules, and some empirical evidence. JFQA, **7**, 1957-1965.
Breusch, T.S. and Pagan, A.R. (1979), A simple test for heteroscedasticity and random coefficient variation. E, **47**, 1287-1294.
Breusch, T.S. and Pagan, A.R. (1980), The Lagrange multiplier test and its application to model specification in econometrics. RES, **47**, 239-253.
Broemeling L.D. and Tsurumi, H. (1987), *Econometrics and Structural Change.* New York: Marcel Dekker.
Brown, R.G. (1963), *Smoothing, Forecasting and Prediction of Discrete Time Series.* New York: John Wiley.
Brown, R.L., Durbin, J., and Evans, J.M. (1975), Techniques for testing the constancy of regression relationships over time. RSSB, **37**, 149-163.
Burnett, T.D. and Guthrie, D. (1970), Estimation of stationary stochastic regression parameters. JASA, **65**, 1547-1553.
Buse, A. and Lim, L. (1977), Cubic splines as a special case of restricted least squares. JASA, **72**, 64-68.
Chamberlain, G. (1978), Omitted variable bias in panel data: Estimating returns to schooling. AI, **30-31**, 49-82.
Chamberlain, G. (1979), *Heterogeneity, Omitted Variable Bias, and Duration Dependence.* Harvard University Discussion Paper No. 691, Cambridge, MA.
Chamberlain, G. (1982), Multivariate regression models for panel data. JE, **18**, 5-46.
Chamberlain, G. (1983), A characterization of the distributions that imply mean-variance utility functions. JET, **29**, 185-201.
Chamberlain, G. (1984), Panel Data, in: Z. Griliches and M.D. Intriligator, (eds.), *Handbook of Econometrics,* Amsterdam: North-Holland.
Chamberlain, G. and Griliches, Z. (1975), Unobservables with a variance components structure: Ability, schooling and the economic success of brothers. IER, **16**, 422-450.
Chernoff, H. (1954), On the distribution of the likelihood ratio. AMS, **25**, 573-578.
Chow, G.C. (1984), Random and changing coefficient models, in: Z. Griliches, and M.D. Intriligator (eds.), *Handbook of Econometrics.* New York: Elsevier.
Choy, C.J.H. and Broemeling, L.D. (1980), Some Bayesian inferences for a changing linear model. T, **22**, 71-78.
Cooley, T.F. and Prescott, E.C. (1973a), An adaptive regression model. IER, **14**, 364-371.
Cooley, T.F. and Prescott, E.C. (1973b), Tests of an adaptive regression model. TRES, **55**, 248-256.
Cooley, T.F. and Prescott, E.C. (1973c), Varying parameter regression: A theory and some applications. AESM, **2**, 463-474.
Cooley, T.F. and Prescott, E.C. (1976), Estimation in the presence of stochastic parameter variation. E, **44**, 167-184.
Cooper, J.P. (1972), Two approaches to polynomial distributed lags estimation: An ex-

pository note and comment. ANS, , 32-35.
Cooper, J.P. (1973), Time-varying regression coefficients: A mixed estimation approach and operational limitations of the general Markov structure. AESM, **2**, 525-530.
Dhrymes, P.J., Howrey, E.P., Hymans, S.H., Kmenta, J., Leamer, E.E., Quandt, R.E., Ramsey, J.B., Shapiro, H.T., and Zarnovitch, V. (1972), Criteria for evaluation of econometric models. AESM, **1**, 291-324.
Dielman, T.E. (1980), *Pooled Data for Financial Markets.* Ann Arbor, MI: UMI Research Press.
Dielman, T.E., Nantell, T., and Wright, R. (1980), Price effects of stock repurchasing: A random coefficient regression approach. JFQA, **15**, 175-189.
Dziechciarz, J. (1980), Metody analizy i przewidywania wielkosci zuzycia energii elktrycznej w dlugich okresach, Ph.D. Dissertation, Academy of Economics, Wroclaw.
Farley, J. and Hinich, M. (1970), Testing for a shifting slope coefficient in a linear model. JASA, **65**, 1320-1329.
Farley, J., Hinich, M. and McGuire, T. (1975), Some comparisons of tests for a shift in the slopes of a multivariate linear time series model. JE, **3**, 297-318.
Feige, E.L. and Swamy, P.A.V.B. (1974), A random coefficient model of the demand for liquid assets. JMCB, **6**, 241-252.
Ferreira, P.E. (1975), A Bayesian analysis of a switching regression model: Known number of regimes. JASA, **70**, 370-374.
Flinn, C. and Heckman, J.J. (1982), New methods for analyzing structural models of labor force dynamics. JE, **18**, 115-168.
Fuller, W.A. and Battese, G.E. (1973), Transformations for estimation of linear models with nested error structure. JASA, **68**, 626-632.
Fuller, W.A. and Battese, G.E. (1974), Estimation of linear models with crossed-error structure. JE, **2**, 67-78.
Gallant, A.R. and Fuller, W.A. (1973), Fitting segmented polynomial regression models whose joint points have to be estimated. JASA, **68**, 144-147.
Glejser, H. (1978), Truncated distributed lags in small sample panel models. AI, **30-31**, 131-135.
Goldfeld, S.M. and Quandt, R.E. (1973a), A Markov model for switching regressions. JE, **1**, 3-16.
Goldfeld, S.M. and Quandt, R.E. (1973b), The estimation of structural shifts by switching regressions. AESM, **2**, 475-485.
Gourieroux, C., Holly, A., and Montfort, A. (1982), Likelihood ratio test, Wald test and Kuhn-Tucker test in linear models with inequality constrains on the regression parameters. E, **50**, 63-80.
Griffiths, W.E. (1972), Estimation of actual response coefficients in the Hildreth-Houck random coefficients model. JASA, **67**, 633-635.
Griffiths, W.E. and Anderson, J.R. (1982), Using time-series and cross-section data to estimate a production function with positive and negative marginal risks. JASA, **77**, 529-536.
Griffiths, W.E., Drynan, R., and Prakash, S. (1979), Bayesian estimation of a random coefficient model. JE, **10**, 201-220.
Griliches, Z., Hall, B.H., and Hausman, J.A. (1978), Missing data and self-selection in large panels. AI, **30-31**, 137-176.
Hackl, P. (1980), *Testing the Constancy of Regression Models over Time*, Göttingen: Vandenhoeck and Ruprecht.
Hackl, P. and Katzenbeisser, W. (1978), Residuals in tests for adequacy of regression relationships. SH, **19**, 83-98.
Harvey, A.C. (1978), The estimation of time-varying parameters from panel data. AI, **30-31**, 203-226.
Harvey, A.C. (1981), *Time Series Models*, Oxford: Phillip Allan.

Harvey, A.C. (1982), The Kalman filter and its applications in econometrics and time series analysis. MOR, **44**, 3-18.
Harvey, A.C. and Phillips, G.D.A. (1979), Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. B, **66**, 49-58.
Harvey, A.C. and Phillips, G.D.A. (1982), The estimation of regression models with time-varying parameters, in:M. Deistler *et al.* (eds.), *Games, Economic Dynamics, and Time Series Analysis*, Vienna: Physica.
Hause, J. (1977), The covariance structure of earnings and the on-the-job training hypothesis. AESM, **6**, 335-365.
Hause, J. (1980), The fine structure of earnings and the on-the-job training hypothesis. E, **48**, 1013-1029.
Hausman, J.A. (1978), Specification tests in econometrics. E, **46**, 1251-1272.
Hausman, J.A. and Taylor, W.E. (1981), Panel data and unobservable individual effects. E, **49**, 1377-1398.
Hausman, J.A. and Wise, D.A. (1979), Attrition bias in experimental and panel data: The Gary income maintenance experiment. E, **47**, 455-473.
Havenner, A. and Swamy, P.A.V.B. (1981), A random coefficient approach to seasonal adjustment of economic time series. JE, **15**, 177-210.
Heckman, J.J. (1978), Simple statistical models for discrete panel data developed and applied to test the hypothesis of true state dependence against the hypothesis of spurious state dependence. AI, **30**, .
Hellwig, Z. (1967), Schemat budowy prognozy statystycznej za pomoca wag harmonicznych. PS, **2**, .
Henderson, C.R. (1953), Estimation of variance and covariance components. BS, **9**, 226-252.
Henderson, C.R. (1975), Best linear unbiased estimation and prediction under a selection model. BS, **31**, 423-447.
Hendricks, W., Koenker, R. and Poirier, D.J. (1979), Stochastic parameter models for panel data: An application to the Connecticut peak load pricing experiment. IER, **20**, 707-724.
Hildreth, C. and Houck, J.P. (1968), Some estimators for a linear model with random coefficients. JASA, **63**, 584-595.
Hinkley, D.V. (1971a), Inference about the change-point from cumulative sum tests. B, **58**, 509-523.
Hinkley, D.V. (1971b), Inference in two-phase regression. JASA, **66**, 736-743.
Holbert, D. (1982), A Bayesian analysis of a switching linear model. JE, **19**, 77-87.
Hsiao, C. (1974), Statistical inference for a model with both random cross sectional and time effects. IER, **15**, 12-30.
Hsiao, C. (1975), Some estimation methods for a random coefficient model. E, **43**, 305-325.
Jöreskog, K.G. (1978), An econometric model for multivariate panel data. AI, **30-31**, 355-366.
Johnson, K.H. and Lyon, H.L. (1973), Experimental evidence on combining cross-section and time series information. TRES, **55**, 465-474.
Johnson, S.R. and Rausser, G.C. (1975), An estimating method for models with stochastic, time varying parameters. American Statistical Association, *Proceedings of the Business and Economic Statistics Section*, 356-361.
Johnston, J. (1984), *Econometric Methods.* New York: McGraw-Hill.
Judge, G.G., Griffiths, W.E., Hill, R.C., and Lee, T.C. (1980), *The Theory and Practice of Econometrics.* New York: John Wiley.
Judge, G.G., Griffiths, W.E., Hill, R.C., Lee, T.C., and Lütkepohl, H. (1982), *Introduction to the Theory and Practice of Econometrics.* New York: John Wiley.
Judge, G.G, Griffiths, W.E., Hill, R.C., Lee, T.C., and Lütkepohl, H. (1985), *The Theory and Practice of Econometrics.* New York: John Wiley.

Kalman, R.E. (1960), A new approach to linear filtering and prediction problems. JBE, **82**, 35-45.

Kalman, R.E. and Bucy, R.S. (1961), New results in linear filtering and prediction theory. JBE, **82**, 95-108.

Kelejian, H.H. and Stephan, S.W. (1983), Inference in random coefficient and error component models: Some corrections and extensions of the literature. IER, **24**, 249-257.

Kiefer, N.M. (1980), Estimation of fixed effects models for time series of cross-sections with arbitrary intertemporal covariance. JE, **14**, 195-202.

Kiefer, N.M. and Neumann, G.R. (1981), Individual effects in a nonlinear model: Explicit treatment of heterogeneity in the empirical job-search model. E, **49**, 965-979.

Kmenta, J. (1986), *Elements of Econometrics.* New York: Macmillan.

LaMotte, L.R. and McWhorter, A., Jr. (1978), An exact test for the presence of random walk coefficients in a linear regression model. JASA, **73**, 816-820.

Leamer, E.E. (1978), *Specification Searches: Ad Hoc Inference with Nonexperimental Data.* New York: John Wiley.

Lee, L.F. (1978a), *Estimation of Error Components Models with ARMA (p,q) Time Component - An Exact GLS Approach*, University of Minnesota Discussion Paper No. 78-104, Minneapolis.

Lee, L.F. (1978b), *On the Issues of Fixed Effects vs. Random Effects Econometric Models with Panel Data*, University of Minnesota Discussion Paper No. 78-101, Minneapolis.

Lee, L.F. and Griffiths, W.E. (1979), *The Prior Likelihood and Best Linear Unbiased Prediction in Stochastic Coefficient Linear Models*, University of New England Working Paper No. 1, Armidale, Australia.

Lee, L.F. and Porter, R.D. (1984), Switching regression models with imperfect sample separation information — With an application on cartel stability. E, **52**, 427-448.

Lillard, L.A. and Weiss, Y. (1979), Components of variation in panel earnings data: American scientists 1960-1970. E, **47**, 437-454.

Lillard, L.A. and Willis, R.J. (1978), Dynamic aspects of earnings mobility. E, **46**, 985-1012.

Liu, L.M. (1981), Estimation of random coefficient regression models. JSCS, **13**, 27-39.

Liu, L.M. and Hanssens, D.M. (1981), A Bayesian approach to time-varying cross-sectional regression models. JE, **15**, 341-356.

MaCurdy, T.E. (1982), The use of time series processes to podel the error structure of earnings in a longitudinal data analysis. JE, **18**, 83-114.

Maddala, G.S. (1971), The use of variance components in pooling cross section and time series data. E, **39**, 341-358.

Maddala, G.S. (1978), Selectivity problems in longitudinal data. AI, **30-31**, 423-451.

Maddala, G.S. and Mount, T.D. (1973), A comparative study of alternative estimators for variance components models used in econometric applications. JASA, **68**, 324-328.

Magnus, J.R. (1982), Multivariate error components analysis of linear and nonlinear regression models by maximum likelihood. JE, **19**, 239-285.

Mazodier, P. and Trognon, A. (1978), Heteroscedasticity and stratification in error components models. AI, **30-31**, 451-482.

McGee, V.E. and Carlton, T.W. (1970), Piecewise regression. JASA, **65**, 1109-1124.

Mehta, J.S., Narasimham, G.V.L., and Swamy, P.A.V.B. (1978), Estimation of a dynamic demand function for gasoline with different schemes of parameter variation. JE, **7**, 263-279.

Moran, P.A.P. (1970), On asymptotically optimal tests of composite hypotheses. B, **57**, 47-55.

Mundlak, Y. (1978a), Models with variable coefficients: Integration and extension. AI, **30-31**, 483-509.

Mundlak, Y. (1978b), On variable coefficients models. AI, **30-31**, .

Mundlak, Y. (1978c), On the pooling of time series and cross section data. E, **46**, 69-85.

Mundlak, Y. and Yahav, J.A. (1981), Random effects, fixed effects, convolution, and separation. E, **49**, 1399-1416.
Nerlove, M. (1967), Experimental evidence on the estimation of dynamic economic economic relations from a time series of cross sections. ESQ, **18**, 42-74.
Nerlove, M. (1971a), A note on error components models. E, **39**, 383-396.
Nerlove, M. (1971b), Further evidence on the estimation of dynamic economic relations from a time series of cross sections. E, **39**, 359-382.
Nickell, S. (1981), Biases in dynamic models with fixed effects. E, **49**, 1417-1426.
Ohtani, K. (1982), Bayesian estimation of the switching regression model with autocorrelated errors. JE, **18**, 251-261.
Pagan, A.R. (1980), Some identification and estimation results for regression models with stochastically varying coefficients. JE, **13**, 341-363.
Pagan, A.R. and Tanaka, K. (1979), A further test for assessing the stability of regression coefficient. Australian National University, Canberra, (mimeo).
Parks, R.W. (1967), Efficient estimation of a system of regression equations when disturbances are both serially and contemporaneously correlated. JASA, **62**, 500-509.
Phillips, G.D.A. and Harvey, A.C. (1974), A simple test for serial correlation in regression analysis. JASA, **69**, 935-939.
Poirier, D.J. (1973), Piecewise regression using cubic splines. JASA, **68**, 515-524.
Poirier, D.J. (1976), *The Econometrics of Structural Change.* Amsterdam: North Holland and New York: American Elsevier.
Prucha, I.R. (1984), On the asymptotic efficiency of feasible Aitken estimators for seemingly unrelated regression models with error components. E, **52**, 203-208.
Pudney, S.E. (1978), The estimation and testing of some error components models, London School of Economics (mimeo).
Quandt, R.E. (1958), The estimation of the parameters of a linear regression system obeying two separate regimes. JASA, **53**, 873-880.
Quandt, R.E. (1960), Tests of the hypothesis that a linear regression system obeys two separate regimes. JASA, **55**, 324-330.
Raj, B. and Ullah, A. (1981), *Econometrics. A Varying Coefficients Approach.* London: Croom Helm.
Raj, B., Srivastava, V.K., and Upadhyaya, S. (1980), The efficiency of estimating a random coefficient model. JE, **12**, 285-299.
Ramsey, J.B. (1969), Tests for specification errors in classical linear least-squares regression analysis. RSSB, **31**, 350-371.
Rao, C.R. (1965a), *Linear Statistical Inference and its Applications.* New York: John Wiley.
Rao, C.R. (1965b), The theory of least squares when parameters are stochastic and its application to the analysis of growth curves. B, **52**, 447-458.
Rao, C.R. (1970), Estimation of heteroscedastic variances in linear model. JASA, **65**, 161-172.
Rao, C.R. (1972), Estimation of variance and covariance components in linear models. JASA, **67**, 112-115.
Rausser, G.C. and Mundlak, Y. (1978), Structural change, parameter variation, and agricultural forecasting. Harvard University, Cambridge, MA (mimeo).
Rausser, G.C., Mundlak, Y., and Johnson, S.R. (1982), Structural change, updating, and forecasting, in: G.C. Rausser (ed.), *New Directions in Econometric Modelling and Forecasting US Agriculture.* Amsterdam: North-Holland.
Reinsel, G. (1982), Multivariate repeated-measurement of growth curve models with multivariate random-effects covariance structure. JASA, **77**, 190-195.
Revankar, N.S. (1979), Error component models with serially correlated time effects. JISA, **17**, 137-160.

Rosenberg, B. (1972), The estimation of stationary stochastic regression parameters reexamined. JASA, **67**, 650-654.
Rosenberg, B. (1973a), A survey of stochastic parameter regression. AESM, **2**, 381-397.
Rosenberg, B. (1973b), Linear regression with randomly dispersed parameters. B, **60**, 65-72.
Rosenberg, B. (1973c), The analysis of a cross section of time series by stochastically convergent parameter regression. AESM, **2**, 399-428.
Sant, D. (1977), Generalized least squares applied to time-varying parameter models. AESM, **6**, 301-314.
Sarris, A. (1973), A Bayesian approach to estimation of time-varying regression coefficients. AESM, **2**, 501-523.
Schaefer, S., Brealey, R., Hodges, S., and Thomas, H. (1975), Alternative models of systematic risk, in E. Elton and M. Gruber (eds.), *International capital markets: An inter and intra country analysis.* Amsterdam: North-Holland.
Schmidt, P. (1983), A note on a fixed effect model with arbitrary interpersonal covariance. JE, **22**, 391-394.
Schwaille, D. (1982), Unconstrained maximum likelihood estimation of contemporaneous covariances. EL, **9**, 359-364.
Searle, S.R. (1979), Variance component estimation and animal breeding, in: Van Vleck, (ed.), *In Honour of Henderson.* Ithaca, NY: Cornell University Press.
Sevestre, K. and Trognon, A. (1982), *A Note on Autoregressive Error Components Models*, L'INSEE Working Paper No. 82-04, Paris.
Silvey, S.D. (1959), The Lagrangian multiplier test. AMS, **30**, 389-407.
Sims, C.A. (1974), Seasonality in regression. JASA, **69**, 618-626.
Singer, B. (1982), Aspects of non-stationarity. JE, **18**, 169-190.
Singh, B. and Ullah, A. (1974), Estimation of seemingly unrelated regressions with random coefficients. JASA, **69**, 191-195.
Singh, B., Nagar, A.L., Choudhry, N.K., and Raj, B. (1976), On the estimation of structural change: A generalization of the random coefficients regression model. IER, **17**, 340-361.
Smith, A.F.M. (1973), A general Bayesian linear model. RSSB, **35**, 67-75.
Smith, A.F.M. and Cook, R.D. (1980), Straight lines with a change point: A Bayesian analysis of some renal transplant data. APS, **29**, 180-189.
Spjotvoll, E. (1977), Random coefficient regression models: A review. MOSS, **8**, 69-93.
Srivastava, V.K., Mishva, G., and Chaturvedi, A. (1981), Estimation of linear regression model with random coefficients ensuring almost non-negativity of variance estimators. BJ, **23**, 3-8.
Swamy, P.A.V.B. (1970), Efficient inference in a random coefficient regression model. E, **38**, 311-323.
Swamy, P.A.V.B. (1971), *Statistical Inference in Random Coefficient Regression Models.* New York: Springer.
Swamy, P.A.V.B. (1973), Criteria, constrains and multicollinearity in random coefficient regression models. AESM, **2**, 429-450.
Swamy, P.A.V.B. (1974), Linear models with random coefficients, in: P. Zarembka (ed.), *Frontiers in Econometrics.* New York: Academic Press.
Swamy, P.A.V.B. and Arora, S.S. (1972), The exact finite sample properties of the estimators of coefficients in the error components regression models. E, **40**, 253-260.
Swamy, P.A.V.B. and Mehta, J.S. (1975a), Bayesian and non-Bayesian analysis of switching regressions and of random coefficient regression models. JASA, **70**, 593-602.
Swamy, P.A.V.B. and Mehta, J.S. (1975b), On Bayesian estimation of seemingly unrelated regressions when some observations are missing. JE, **3**, 157-169.
Swamy, P.A.V.B. and Mehta, J.S. (1977), Estimation of linear models with time and

cross-sectionally varying coefficients. JASA, **72**, 890-898.
Swamy, P.A.V.B. and Mehta, J.S. (1979), Estimation of common coefficient in two regression equations. JE, **10**, 1-14.
Swamy, P.A.V.B. and Tinsley, P.A. (1980), Linear prediction and estimation methods for regression models with stationary stochastic coefficients. JE, **12**, 103-142.
Taub, A.J. (1979), Prediction in the context of the variance components model. JE, **10**, 103-108.
Taylor, W.E. (1980), Small sample considerations in estimation from panel data. JE, **13**, 203-223.
Tishler, A. and Zang, I. (1979), A switching regression method using inequality conditions. JE, **11**, 259-274.
Trognon, A. (1978), Miscellaneous asymptotic properties of ordinary least squares and maximum likelihood estimators in dynamic error components models. AI, **30-31**, 631-657.
Tsurumi, H. (1982), A Bayesian and maximum likelihood analysis of a gradual switching regression in a simultaneous equation framework. JE, **19**, 165-182.
Wallace, T.D. and Hussain, A. (1969), The use of error components models in combining cross-sections with time series data. E, **37**, 55-72.
Wallis, K.F. (1974), Seasonal adjustment and relations between variables. JASA, **69**, 18-31.
Wansbeek, T.J. and Kapteyn, A. (1978), The separation of individual variation and systematic change in the analysis of panel data. AI, **30-31**, 659-680.
Wansbeek, T.J. and Kapteyn, A. (1981), Estimators of the covariance structure for longitudinal data, in: E.G.Charatsis (ed.), *Proceedings of the Econometric Society European Meeting, 1979.* Amsterdam: North-Holland.
Wansbeek, T.J. and Kapteyn, A. (1982), A class of decompositions of the variance-covariance matrix of a generalized error components model. E, **50**, 713-724.
Zellner, A. (1962), An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. JASA, **57**, 348-368.
Zellner, A. (1969a), Estimators for seemingly unrelated regression equations: Some exact finite sample results. JASA, **64**, 977-992.
Zellner, A. (1969b), On the aggregation problem. A new approach to a troublesome problem, in: J.K. Fox *et al.*, *Economic Models, Estimations and Risk Programming.* New York: Springer.
Zellner, A. (1984), *Basic Issues in Econometrics.* Chicago: University of Chicago Press.

CHAPTER 15

Nonparametric Estimation of Time-Varying Parameters

Peter M. Robinson

Summary

A sequence of observations $y_t, t = 1, 2, \ldots, N$, is generated by the time-varying multiple regression model

$$y_t = \beta_t' x_t + \sigma_t u_t \ , \quad t = 1, 2, \ldots, N,$$

where, for $t = 1, 2, \ldots, N$, u_t is an unobservable random variable with zero mean and unit variance, x_t is an observable p-vector-valued variable, and σ_t and β_t are, respectively, unobservable scalar and p-vector-valued parameters. No model (stochastic or nonstochastic) is assumed for the σ_t or β_t; instead they are assumed to be smoothly varying over t, in a certain sense. A class of estimators of the β_t, σ_t is proposed, for each value of t; the estimators optimize a criterion prompted by Gaussian maximum likelihood considerations, and may be viewed as analogous to certain nonparametric function fitting estimators, employing a kernel function and band-width parameter, both selected by the practitioner. Consistency and asymptotic normality are established in case of independent u_t, and a consistent estimator of the asymptotic covariance matrix of the β_t estimators is given. Such results are also possible for serially correlated u_t. We discuss questions of implementation, in particular the choice of kernel function and band-width. Generalization of the class of estimators to include certain robust estimators is possible, as is generalization of the methods to more general models involving time-varying parameters.

15.1 Introduction

To model the dependence of a scalar time series y_t on a $p \times 1$ column vector time series $\boldsymbol{x}_t$, we consider the time-varying multiple regression

$$y_t = \beta_t' \boldsymbol{x}_t + \sigma_t u_t \ , \quad t = 1, 2, \ldots, N \tag{15.1}$$

The $p \times 1$ vector β_t is unknown for each t, and is not assumed constant over t, and the prime denotes transposition. The sequence of random variables u_t is likewise unobservable. Throughout, we assume u_t is independent of $\boldsymbol{x}_s$, for all s and t, and

$$E(u_t) = 0 \ , \quad V(u_t) = 1$$

The sequence of residual variances σ_t^2 is unknown, and not assumed constant over t. Observations on $\boldsymbol{x}_t$ and y_t are available for $t = 1, \ldots, N$. It is of interest to estimate the β_t and σ_t.

If β_t and σ_t are modeled as known functions of t and finitely many unknown parameters, the problem reduces to a standard one of parametric estimation. For example, polynomials in t might be employed. So far as β_t is concerned, the parameters, θ, on which it depends might be estimated by linear or nonlinear least squares, depending on whether β_t is linear or nonlinear in θ. The parameters describing both β_t and σ_t might be simultaneously estimated by optimizing a Gaussian likelihood.

An alternative approach to the modeling of β_t and σ_t takes them to be generated by a finite parameter stochastic model. This approach has been employed so far as β_t is concerned (with $\sigma_t \equiv \sigma$ assumed) by many authors, taking the stochastic model for β_t to depend on finitely many unkown parameters — typically, a stationary autoregressive or autoregressive moving average model or a random walk. Again, after some manipulation, the estimation problem is of parametric type.

While various such parametric models for the β_t and σ_t — be they stochastic or nonstochastic — may prove computationally convenient and afford precise estimation when they are reasonable, the goal of explaining time-varying parameters so parsimoniously may be overly ambitious. While they do afford more generality than time-invariant regressions, they need not necessarily provide a good approximation to the actual data generating mechanism. In this chapter, we avoid finite parameterization of β_t and σ_t, relying only on smoothness assumptions. Our approach is nonparametric, in the sense that we can regard the number of free parameters as increasing slowly with N.

In order to provide asymptotic justification for our estimators, it is convenient to regard the β_t, σ_t as being generated from functions $\beta(t)$ and $\sigma(t)$ on $(0, 1)$:

$$\beta_t = \beta(\frac{t}{N}), \quad \sigma_t = \sigma(\frac{t}{N}) \ , \quad t = 1, 2, \ldots, N \tag{15.2}$$

Thus, the β_t, σ_t depend on N in this framework, though our notation does not emphasize this. The specifications (15.2) are analogous to ones standardly employed, for the same reason, in the nonparametric curve fitting literature; see, e.g., Nadaraya (1964), Benedetti (1977), Clarke (1977), Gasser *et al.* (1985) and Wong (1983). The model considered by these authors is more general in the sense that some of their results apply to irregularly spaced observations. However, their model is more special in that it applies only to the case where $\boldsymbol{x}_t$ consists only of an intercept, so $p = 1$, $\boldsymbol{x}_t \equiv 1$, and in that they assume σ_t is constant over t. In addition, whereas we, like these authors, assume the residuals u_t are independent, we also indicate how our results may be extended to allow for serially dependent u_t.

Condition (15.2) might seem strange because it makes the β_t, σ_t depend on sample size, N. The reason for this requirement is that an estimator of β_t or σ_t will not be consistent unless the amount of data on which it depends increases; and merely increasing the length of the series N will not necessarily improve estimation of β_t or σ_t at some fixed point t, even if some smoothness condition is imposed on the β_t, σ_t sequences. The amount of *local* information must increase suitably if variance and bias are to decrease suitably. A convenient way to achieve this is to regard the β_t, σ_t as ordinates of smooth functions $\beta(\cdot)$, $\sigma(\cdot)$ on an equally spaced grid over $(0,1)$, which becomes finer as $N \to \infty$, and consider estimation of $\beta(\tau)$, $\sigma(\tau)$ at fixed points τ, while defining u_t, $\boldsymbol{x}_t$, and y_t on the integers, as is conventional in time series analysis.

Notice that this problem does not arise in two related nonparametric estimation problems: spectral analysis and regression on a stochastic explanatory variable. In spectral estimation, a periodic function is being estimated, so its support can be taken as finite such that increasing N permits calculation of approximately independent discrete Fourier transforms over an ever-finer grid on the Nyquist frequencies, and thus estimation of the spectrum with both increasing resolution and precision. In stochastic nonparametric regression, the density of observations on the regressors increases with N, with similar effect. In our case, however, the natural support of the β_t, σ_t is the integers, which do not become more dense as N increases, so that as in the nonparametric curve fitting literature referred to above, the problem has to be rephrased.

Our representation does not, however, regard the sampling of the y_t, $\boldsymbol{x}_t$ as taking place on a grid on $(0,1)$, which would make the preservation of independence or weak dependence properties as N increases implausible. We note that the device of taking the y_t, $\boldsymbol{x}_t$ to be observations at intervals $1/N$ on a continuous process on $(0,1)$ that itself is independent of N would not work because it does not achieve the accumulation of new information as N increases that is necessary for consistency. Making the parameters dependent on sample size is not unknown elsewhere in the statistical literature (for example, in Pitman sequences), and any asymptotic scenario is open to the criticism that in many applications there is no possibility of increasing N, the motivation being merely to provide approximate justification for an inferential procedure based on a finite sample.

In the following section, estimators of the β_t, σ_t are proposed. In Section 15.3, we show them to be consistent. In Section 15.4, asymptotic normality of the β_t estimator is established. In Section 15.5, we discuss the implications of serially dependent u_t; describe

alternative estimators, including robust ones; and indicate that analogous estimators can be obtained for other models involving time-varying parameters.

15.2 Estimators of the β_t, σ_t

We introduce a kernel function k, that is a real-valued function heavily concentrated around the origin. Let h be a positive constant, dependent on N, and for $\tau \in (0,1)$ define

$$k_{\tau t} = k(\frac{N\tau - t}{Nh})$$

To estimate $\beta(\tau)$, $\sigma(\tau)$ for any $\tau \in (0,1)$ consider the Gaussian pseudo log likelihood function

$$\ell[\beta(\tau), \sigma(\tau)] = \sum_t k_{\tau t} \log f[y_t | x_t; \beta(\tau), \sigma(\tau)] \tag{15.3}$$

where $\sum_t$ is a sum over t from 1 through N and

$$f(y|x;\beta,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\{-\frac{1}{2}(\frac{y - \beta' x}{\sigma})^2\}$$

The first-order conditions for a maximum of $\ell[\beta(\tau), \sigma(\tau)]$ are

$$\sum_t k_{\tau t} x_t (y_t - \hat{\beta}(\tau)' x_t) = 0$$

$$\sum_t k_{\tau t} \left\{ \frac{1}{\hat{\sigma}(\tau)^2} - \frac{(y_t - \hat{\beta}(\tau)' x_t)^2}{\hat{\sigma}(\tau)^4} \right\} = 0$$

When solved, these yield the Gaussian pseudo maximum likelihood estimators of $\beta(\tau)$, $\sigma(\tau)^2$,

$$\hat{\beta}(\tau) = (\sum_t k_{\tau t} x_t x_t')^{-1} \sum_t k_{\tau t} x_t y_t$$

$$\hat{\sigma}(\tau)^2 = (\sum_t k_{\tau t})^{-1} \sum_t k_{\tau t} [y_t - \hat{\beta}(\tau)' x_t]^2$$

The implementation of this procedure depends on the choice of k and h. Consider

$$k(t) = \begin{cases} \frac{1}{2}, & |t| \leq 1 \\ 0, & |t| > 1 \end{cases} \tag{15.4}$$

In case (15.4), for sufficiently large h, $\hat{\beta}(\tau)$ and $\hat{\sigma}(\tau)^2$ reduce, for all τ, to the Gaussian estimators of β and σ^2 in the familiar time-invariant regression model

$$y_t = \beta' x_t + \sigma u_t$$

namely,

$$\hat{\beta} = (\sum_t x_t x_t')^{-1} \sum_t x_t y_t$$

$$\hat{\sigma}^2 = \frac{1}{N} \sum_t (y_t - \hat{\beta}' x_t)^2$$

For suitably small h, $\hat{\beta}(\tau)$ and $\hat{\sigma}(\tau)^2$ exhibit variation over τ. With k given by (15.4), the $\hat{\beta}_s = \hat{\beta}(s/N), s = 1, \ldots, N$, are the moving regression estimators of Brown *et al.* (1975)

$$\hat{\beta}_s = \left(\sum_{|t-s| \leq Nh} x_t x_t' \right)^{-1} \sum_{|t-s| \leq Nh} x_t y_t$$

We have for $\hat{\sigma}_s^2 = \hat{\sigma}(s/N)^2$

$$\hat{\sigma}_s^2 = \frac{1}{Nh} \sum_{|t-s| \leq Nh} (y_t - \hat{\beta}_s' x_t)^2$$

Notice that the formulae for $\hat{\beta}(\tau)$, $\hat{\sigma}(\tau)^2$ are computable for any $\tau \in (0,1)$, allowing for "interpolation". Viewed as a function of $\tau \in (0,1)$, for k given by (15.4), $\hat{\beta}(\tau)$, $\hat{\sigma}(\tau)^2$ are discontinuous at $\tau = s/N$, for integer s. A continuous $\hat{\beta}(\tau)$, $\hat{\sigma}(\tau)^2$ trajectory is obtained via use of $k(t)$ that is everywhere continuous, such as

$$k(t) = \begin{cases} \frac{3}{4}(1 - t^2), & |t| \leq 1 \\ 0, & |t| > 1 \end{cases}$$

see Epanechnikov (1969). Many other choices of k are possible, including ones with infinite support, such as the Gaussian kernel

$$k(t) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2} t^2), \quad -\infty < t < \infty \tag{15.5}$$

and ones discussed by Gasser *et al.* (1985).

The choice of band-width h is generally believed to be more crucial than that of k in kernel estimation. Statistically, small h tends to correspond to small bias in $\hat{\beta}(\tau)$, $\hat{\sigma}(\tau)^2$; large h to small variance. (Of course, the value of p will induce a lower bound on h.) In practice h is often chosen in an *ad hoc* fashion.

Alternatively, an automatic method, such as cross-validation, might be employed. To describe this, introduce

$$\hat{\beta}_{-s}(\tau) = (\sum_{t \neq s} k_{\tau t} x_t x_t')^{-1} \sum_{t \neq s} k_{\tau t} x_t y_t$$

A possible choice of h is $\hat{h}$ given by

$$CV(\hat{h}) = \min_h CV(h)$$

where

$$CV(h) = \sum_s [y_s - \hat{\beta}'_{-s}(s/N)\, x_s]^2$$

Wong (1983) considers this criterion in the "nonparametric curve fitting" special case of (15.1), where $p = 1, x_t \equiv 1, \sigma_t \equiv \sigma$, providing some asymptotic justification.

The computation of $\hat{\beta}(\tau)$, $\hat{\sigma}(\tau)^2$ for given h and many τ — for example, for all $\tau = s/N$ when N is large — is likely to be expensive. In the latter case, however, an approximate computational method, based on the fast Fourier transform, is available. Suppose we know K such that

$$k(t) = \int_{-\infty}^{\infty} K(u)\, e^{-itu}\, du$$

for example, $K(u) = \exp(-\frac{1}{2}u^2)$ in case (15.5). Then for any sequence z_t, $t = 1, \ldots, N$

$$\begin{aligned} \sum_t k_{s/N,t} z_t &= \sum_t k(\frac{s-t}{Nh}) z_t \qquad (15.6) \\ &= \int_{-\infty}^{\infty} K(u) \exp(\frac{-isu}{Nh}) \sum_t z_t \exp(\frac{itu}{Nh})\, du \\ &= \frac{h}{2\pi} \int_{-\infty}^{\infty} K(\frac{vh}{2\pi}) \exp(\frac{-isv}{2\pi N}) \sum_t z_t \exp(\frac{itv}{2\pi N})\, dv \end{aligned}$$

We may thus approximate (15.6) by passing the z_t through the fast Fourier transform, multiplying by the $K(\frac{vh}{2\pi})$, for integer v, and using the inverse fast Fourier transform. We can construct $\hat{\beta}_s$ and $\hat{\sigma}_s^2$ by using for z_t the functions $x_t x_t', x_t y_t$, and $(y_t - \hat{\beta}_s' x_t)^2$. The same type of approach has been used by Silverman (1982) in connection with kernel probability density estimation.

15.3 Consistency

We shall establish consistency of $\hat{\beta}(\tau)$, $\hat{\sigma}(\tau)^2$ for $\beta(\tau)$, $\sigma(\tau)^2$ under the assumption that the u_t are independent across t, which is standard in the nonparametric curve fitting

literature, though unlike in that literature that we do not assume $\sigma_t^2 \equiv \sigma^2$. We assume $E|u_t|^{2+\eta} \leq C < \infty$, for some $\eta > 0$, some C, independent of t. We allow for serial dependence in x_t (which is not present in the usual curve fitting problem).

Specifically, we assume x_t is stationary such that $M = E(x_t x_t')$ is positive definite, $E\|x_t\|^{4+\delta} < \infty$, some $\delta > 0$, and x_t satisfies the strong mixing condition [see, e.g., Deo (1973)] with mixing coefficient α_j satisfying

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \alpha_j^{\delta/(4+\delta)} = 0$$

where

$$\alpha_j = \sup_{A\in\mathcal{A}, B\in\mathcal{B}} |P(A \cap B) - P(A)P(B)|$$

such that $\mathcal{A}$ and $\mathcal{B}$ are the σ-fields of events generated by $x_t, t \leq s$, and $x_t, t \geq s + j$, respectively, for $j > 0$. We assume k is bounded and continuous except possibly at finitely many points, that it integrates to 1, and

$$\int_{-\infty}^{\infty} |k(u)|\, du < \infty$$

We assume $\beta(\tau)$ and $\sigma(\tau)$ are continuous and bounded on $(0,1)$.

We may write

$$\begin{aligned} \hat{\beta}(\tau) - \beta(\tau) &= \Big(\sum_t k_{\tau t} x_t x_t'\Big)^{-1} \sum_t k_{\tau t} x_t x_t' \{\beta_t - \beta(\tau)\} \\ &\quad + \Big(\sum_t k_{\tau t} x_t x_t'\Big)^{-1} \sum_t k_{\tau t} x_t \sigma_t u_t \end{aligned} \tag{15.7}$$

We shall show first that

$$\frac{1}{Nh} \sum_t k_{\tau t} x_t x_t' \xrightarrow{p} M \tag{15.8}$$

when $h \to 0$, $Nh \to \infty$, as $N \to \infty$. Let x_{it} be the ith element of x_t and m_{ij} be the (i,j)th element of M. For $B > 0$ denote

$$\sum_t{}' = \sum_{|\tau N - t| \leq BNh}, \qquad \sum_t{}'' = \sum_{|\tau N - t| > BNh}$$

For any i, j

$$E[|\frac{1}{Nh}\sum_t k_{\tau t}(x_{it}x_{jt} - m_{ij})|] \le \frac{1}{Nh}\left\{E[\sum_t{}' k_{\tau t}(x_{it}x_{jt} - m_{ij})]^2\right\}^{\frac{1}{2}} \tag{15.9}$$

$$+ \frac{1}{Nh}\sum_t{}'' |k_{\tau t}|\, E[|x_{it}x_{jt} - m_{ij}|] \tag{15.10}$$

Using inequality (2.1) of Deo (1973), the r.h.s. of (15.9) is bounded by

$$\frac{1}{Nh}\left\{\sum_t{}' k_{\tau t}^2 E[x_{it}^2 x_{jt}^2] + \sum_t{}'\sum_s{}' |k_{\tau t}k_{\tau s}|\,(E[|x_{it}x_{jt}|^{2+\delta/2}])^{4/(4+\delta)}\,\alpha_{|t-s|}^{\delta/(4+\delta)}\right\}^{\frac{1}{2}}$$

$$\le C\,[(Nh)^{-\frac{1}{2}} + B\,\{(BNh)^{-1}\sum_{j\le BNh}\alpha_j^{\delta/(4+\delta)}\}^{\frac{1}{2}}]$$

where C denotes a generic constant. We can bound (15.10) by

$$\frac{C}{Nh}\sum_t{}'' |k_{\tau t}| \to C\int_{|u|\ge B}|k(u)|\,du, \qquad \text{as } N\to\infty$$

by a slight extension of Lemma 1 of Benedetti (1977). Then (15.9) and (15.10) $\to 0$ by letting $N\to\infty$ then $B\to\infty$. Because

$$\frac{1}{Nh}\sum_t k_{\tau t} \to \int_{-\infty}^{\infty} k(u)\,du = 1, \qquad \text{as } N\to\infty \tag{15.11}$$

by a slight extension of Lemma 1 of Benedetti (1977), we have established (15.8). Next, with $\|\cdot\|$ denoting Euclidean norm,

$$E[\|\frac{1}{Nh}\sum_t k_{\tau t}x_t x_t'\{\beta_t - \beta(\tau)\}\|]$$

$$\le \sup_{|t-\tau|\le Bh}\|\beta(t)-\beta(\tau)\|\,\frac{1}{Nh}\sum_t{}' |k_{\tau t}|\,\mathrm{tr}(M) + \frac{C}{Nh}\sum_t{}'' |k_{\tau t}|\,\mathrm{tr}(M)$$

$$\le C[\sup_{|t-\tau|\le Bh}|\|\beta(t)-\beta(\tau)\| + \int_{|u|\ge B}|k(u)|\,du] + o(1)$$

which $\to 0$ as $h\to 0$ then $B\to\infty$. Finally,

$$E[\|\frac{1}{Nh}\sum_t k_{\tau t}x_t\sigma_t u_t\|^2] = \frac{1}{(Nh)^2}\sum_t k_{\tau t}^2\,\mathrm{tr}(M)\,\sigma_t^2 = 0(\frac{1}{Nh}) \to 0$$

We have thus established

$$\hat{\beta}(\tau) \xrightarrow{p} \beta(\tau)$$

To handle $\hat{\sigma}(\tau)^2$ we may write

$$y_t - \hat{\beta}(\tau)' x_t = \sigma_t u_t - v_t$$

where

$$v_t = \{\hat{\beta}(\tau) - \beta(\tau)\}' x_t + \{\beta(\tau) - \beta_t\}' x_t$$

Thus,

$$\left(\frac{\sum k_{\tau t}}{Nh}\right)\{\hat{\sigma}(\tau)^2 - \sigma(\tau)^2\} = \frac{1}{Nh}\sum_t k_{\tau t}\{(\sigma_t u_t - v_t)^2 - \sigma(\tau)^2\}$$

which is majorized by

$$\left|\frac{1}{Nh}\sum_t k_{\tau t}\sigma_t^2(u_t^2 - 1)\right| + \left|\frac{1}{Nh}\sum_t k_{\tau t}[\sigma_t^2 - \sigma(\tau)^2]\right|$$
$$+ 2\left(\frac{1}{Nh}\sum_t |k_{\tau t}|\sigma_t^2 u_t^2\right)^{\frac{1}{2}}\left(\frac{1}{Nh}\sum_t |k_{\tau t}| v_t^2\right)^{\frac{1}{2}} + \left|\frac{1}{Nh}\sum_t k_{\tau t} v_t^2\right|$$

Using von Bahr and Esseen (1965) , and taking $0 < \eta \leq 2$,

$$E\left[\left|\frac{1}{Nh}\sum_t k_{\tau t}\sigma_t^2(u_t^2 - 1)\right|\right] \leq \left(\frac{1}{(Nh)^{1+\eta/2}}\sum_t{}' |k_{\tau t}\sigma_t^2|^{1+\eta/2} E[|u_t|^{2+\eta}]\right)^{2/(2+\eta)}$$
$$+ \frac{2}{Nh}\sum_t{}'' |k_{\tau t}| E[u_t^2] \to 0$$

arguing as above. We have next

$$\left|\frac{1}{Nh}\sum_t k_{\tau t}(\sigma_t^2 - \sigma(\tau)^2)\right|$$
$$\leq \sup_{|t-\tau|\leq Bh} |\sigma(t)^2 - \sigma(\tau)^2| \frac{1}{(Nh)}\sum_t{}' |k_{\tau t}| + \frac{C}{Nh}\sum_t{}'' |k_{\tau t}| \to 0$$

Finally,

$$\frac{1}{Nh}\sum_t k_{\tau t}v_t^2 \leq 2\{\hat{\beta}(\tau)-\beta(\tau)\}'\frac{1}{Nh}\sum_t |k_{\tau t}|x_t x_t'\{\hat{\beta}(\tau)-\beta(\tau)\} + \frac{2}{Nh}\sum_t |k_{\tau t}|\{\beta(\tau)-\beta_t\}'x_t x_t'\{\beta(\tau)-\beta_t\} \xrightarrow{p} 0$$

in view of what has already been established. Thus, because of (15.11),

$$\hat{\sigma}(\tau)^2 \xrightarrow{p} \sigma(\tau)^2$$

15.4 Asymptotic Normality

We focus here on the asymptotic normality of $\hat{\beta}(\tau)$, centered at $\beta(\tau)$, and under the assumption $\sigma_t^2 \equiv \sigma^2$. We assume, in addition to the previous conditions, that $\beta(\tau)$ satisfies a Lipschitz condition of order α, $0 < \alpha \leq 1$, and assume in addition to $h \to 0, Nh \to \infty$, as $N \to \infty$,

$$Nh^{1+2\alpha} \to 0, \quad \text{as } N \to \infty$$

We also assume now, for simplicity, that k has compact support, which we take for convenience to be $[-1,1]$.

We wish to show that

$$(Nh)^{\frac{1}{2}}\{\hat{\beta}(\tau)-\beta(\tau)\} \xrightarrow{d} N(0,\, \sigma^2\int_{-\infty}^{\infty} k(u)^2\,du M^{-1}) \tag{15.12}$$

assuming the integral is finite. Using steps following (15.10), note that

$$E[\|\frac{1}{(Nh)^{\frac{1}{2}}}\sum_t k_{\tau t}x_t x_t'\{\beta_t-\beta(\tau)\}\|] \leq \sup_{|t-\tau|\leq h}\|\beta(t)-\beta(\tau)\|(Nh)^{\frac{1}{2}}\,\mathrm{tr}(M) = O[(Nh^{1+2\alpha})^{\frac{1}{2}}] \to 0$$

So far as the term

$$\frac{\sigma}{(Nh)^{\frac{1}{2}}}\sum_t k_{\tau t}x_t u_t \tag{15.13}$$

is concerned, note that $x_t u_t$ are martingale differences with finite $(2+\eta)$th moment, some $\eta > 0$, and the remaining relevant conditions of Scott (1973) may readily be checked to show that (15.13) is asymptotically $N(0, \sigma^2 M)$, using

$$\frac{1}{Nh}\sum_t k_{\tau t}^2 \to \int_{-\infty}^{\infty} k(u)^2\,du$$

by a slight extension of Lemma 1 of Benedetti (1977). The proof of (15.12) follows by invoking (15.8).

A consistent estimator of the limiting covariance matrix in (15.12) is given simply by

$$\int_{-\infty}^{\infty} k(u)^2\,du \left(\frac{1}{N}\sum_t x_t x_t'\right)^{-1}$$

Notice that (15.12) implies that the limiting distribution is independent of τ. Moreover, it may readily be shown that not only are $\hat{\beta}(\tau)$ and $\hat{\beta}(\zeta)$ asymptotically identically distributed, for fixed, distinct τ, ζ, but they are also asymptotically independent, because

$$|E[(\frac{1}{Nh}\sum_t k_{\tau t}x_t'u_t \sum_t k_{\zeta t}x_t u_t)]|$$
$$= \frac{\mathrm{tr}(M)}{Nh}\sum_t |k_{\tau t}k_{\zeta t}| \le \frac{C}{Nh}\sum_{|\tau N - t|\le Nh} |k_{\zeta t}| \le C \max_{|s|\le 1} k(s + \frac{\zeta - \tau}{h}) \to 0$$

as $h \to 0$.

It is possible to establish asymptotic normality of $\hat{\beta}(\tau)$ without the assumption that σ_t is constant. The limiting covariance matrix is no longer $\sigma^2 M^{-1}$, however, but it may be consistently estimated as in Eicker (1963).

15.5 Extensions

15.5.1 Serially correlated u_t

Consistency of $\hat{\beta}(\tau)$, $\hat{\sigma}(\tau)^2$ may also be established in the presence of a variety of serial dependence assumptions on u_t. Asymptotic normality also may be established under such assumptions, though the form of the asymptotic covariance matrix will reflect this. More efficient estimators than $\hat{\beta}(\tau)$, $\hat{\sigma}(\tau)^2$ can be constructed in this case, in a way analogous to that in constant-parameter models.

15.5.2 Alternative estimators

Concerning (15.3), the function f can be replaced by an alternative, non-normal, density function, if this seems appropriate. Alternatively, $\log f$ might be replaced by a form of robustified loss function, leading to estimators that are less sensitive to outlying observations [cf., Härdle and Gasser (1984)]. Such estimators will typically have to be computed by numerical iteration.

15.5.3 More general models

The estimators $\hat{\beta}(\tau)$, $\hat{\sigma}(\tau)^2$ can clearly be computed in the case where x_t includes lagged dependent y_t, though the asymptotic statistical theory will more difficult. Extensions to other models in the regression family are also apparent. Indeed, the same type of approach may also be applied to more general models, that are not necessarily of regression type.

Consider a vector random variable z_t, having density function $g(z_t|\theta_t)$, where θ_t is a vector parameter generated from a function $\theta(\tau)$ via $\theta_t = \theta(t/N)$. Then $\theta(\tau)$ can be estimated by maximizing

$$\sum_t k_{\tau t} \log g[z_t|\theta(\tau)]$$

Acknowledgments

This article is based on research funded by the Economic and Social Research Council (ESRC) reference number: B00232156.

References

Benedetti, J.K. (1977), On the nonparametric estimation of regression functions. *Journal of the Royal Statistical Society*, **B-39**, 248–253.

Brown, R.L., Durbin, J., and Evans, J.M. (1975), Techniques for testing the constancy of regression relationships over time (with discussion). *Journal of the Royal Statistical Society*, **B-39**, 149–192.

Clarke, R.M. (1977), Nonparametric estimation of a smooth regression function. *Journal of the Royal Statistical Society*, **B-39**, 107–113.

Deo, C.M. (1973), A note on empirical processes of strong mixing sequences. *Annals of Probability*, **5**, 870–875.

Eicker, F. (1963), Asymptotic normality and consistency of the least squares estimators for families of linear regressions. *Annals of Mathematical Statistics*, **34**, 447–456.

Epanechnikov, V.A. (1969), Nonparametric estimators of a multivariate probability density. *Theory of Probability and its Applications*, **14**, 153–158.

Gasser, T., Müller, H.-G., and Mammitzsch, V. (1985), Kernels for nonparametric curve estimation. *Journal of the Royal Statistical Society*, **B-47**, 238–252.

Härdle, W. and Gasser, T. (1984), Robust non-parametric function fitting. *Journal of the Royal Statistical Society*, **B-46**, 42–51.

Nadaraya, E.A. (1964), On estimating regression. *Theory of Probability and its Applications*, **9**, 141–142.

Scott, D.W. (1973), Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach. *Advances in Applied Probability*, **5**, 119–137.

Silverman, B.W. (1982), Kernel density estimation using the fast Fourier transform. *Applied Statistics*, **31**, 93–97.

Von Bahr, B. and Esseen, C.-G. (1965), Inequalities for the rth absolute moment of a sum of random variables, $1 \leq r \leq 2$. *Annals of Mathematical Statistics*, **36**, 299–303.

Wong, W.H. (1983), On the consistency of cross-validation in kernel nonparametric regression. *Annals of Statistics*, **11**, 1136–1141.

CHAPTER 16

Latent Variables in Regression Analysis

Valeri V. Fedorov

Summary

Attention is drawn to the fact that a number of results from econometric analysis of regression models with unobservable variables can be readdressed using traditional regression analysis techniques. This observation is of importance in the choice of comparatively simple methods for handling corresponding problems, particulary in cases when unobservable explain structural changes in the final regression models.

16.1 Introduction

Models with unobservable variables are widespread in econometric investigations. This chapter concerns two of the most popular models of this kind.

The first model [compare with Robinson and Ferrara (1977)] can be described by the following system:

$$\begin{aligned} y_i &= \vartheta_0' z_i + \varepsilon_i \\ z_i &= B_0' x_i + \nu_i \ , \quad i = 1, \ldots, n \end{aligned} \tag{16.1}$$

In (16.1) y_i is an observable variable (response), the vector $x_i \in R^l$ describes the conditions of observation and is supposed to be known; the vector $z_i \in R^m$ corresponds to unobservable variables; ε_i are *iid* random variables with mean zero and variances σ^2; ν_i are random vectors with zero means and variance matrix d; and vector ϑ_0 and matrix B_0

contain unknown components; "′" stands for transposition; the index "0" points out the true value of the parameters.

The second model under consideration is described by the following set of equations [compare with Zellner (1977)]:

$$\begin{aligned} y_i &= \vartheta_0' z_i + \varepsilon_i \\ z_i &= B_0' x_i \\ u_i &= z_i + \nu_i \end{aligned} \tag{16.2}$$

In (16.2) all notations and assumptions which take place for (16.1) are fullfiled. However, the relation between z_i and x_i does not contain random values, and the vector z_i is observed only through u_i with some additive random errors.

Both models can be generalized for multiresponse cases (y_i can be a vector). The generalization of the estimators is straightforward, and this is the reason why we consider the scalar case in this chapter. Our main purpose is to show that the models (16.1) and (16.2) can be transformed into some well-known regression models for which the properties and the numerical procedures are well studied, and which can efficiently be used after the appropriate adjustment.

16.2 The Regression Model with a Variance Containing Unknown Parameters

It is obvious that for model (16.1) the variables z_i can be eliminated:

$$y_i = \vartheta_0' B_0' x_i + \xi_i \tag{16.3}$$

where $E[\xi_i] = 0$ and $E[\xi_i \xi_j] = \delta_{ij}(\sigma^2 + \vartheta_0' d \vartheta_0)$, i.e., (16.1) is equivalent to a regression problem (nonlinear, if both elements B and ϑ are unknown) with a variance depending on parameters ϑ_0 [see, for instance, Carroll (1982); Fedorov (1974); and Malyutov (1982)].

When ϑ_0 and d_0 are given and some elements of the matrix B are unknown, model (16.3) can be transformed into

$$y_i = \gamma_0' x_i + \xi_i \tag{16.4}$$

where $\gamma_0 = B\vartheta_0$. This model is a traditional linear regression model with unknown variance $s^2 = \sigma^2 + \vartheta_0' d_0 \vartheta_0$. Therefore, *not more than l linear combinations* of the elements of B can be estimated and the least squares method provides the best linear unbiased estimators of these linear combinations.

When all elements of the matrices B_0 and d_0 are known, one still must deal with a rather trivial situation, which, nevertheless, is of some interest for applications. For the

sake of simplicity, let us assume that the rank of B_0 is m. Then model (16.3) can be transformed into the linear regression model with an unknown variance depending on the parameters ϑ [see Malyutov (1982)]:

$$y_i = \vartheta_0' w_i + \xi_i$$

where $w_i = B_0 x_i$, and ξ_i is defined in the comments to (16.3).

The estimation problem for model (16.3) becomes more difficult when both ϑ and B depend upon the same unknown parameters. Consider the most simple case with $l = m = 1$ and $\vartheta = \nu\gamma$, $B = b\gamma$, where ν and b are known. The model (16.1) can be reduced to

$$y_i = a\gamma^2 x_i = \xi_i$$

where

$$a = \nu b \ , \quad E[\xi_i] = 0 \ , \quad E[\xi_i, \xi_j] = \delta_{ij}(\sigma^2 \vartheta \gamma^2 d)$$

It is obvious that the parameter γ for (16.1) is consistently estimated if at least a sign of γ_0 is known *a priori*.

In the general case, when $\vartheta = \vartheta(\gamma)$, $B = B(\gamma)$, $l, m \geq 1$, and $\gamma \in R^k$, one has

$$y_i = \vartheta'(\gamma_0) B'(\gamma_0) x_i + \xi_i = \Psi'(\gamma_0) x_i + \xi_i$$

where

$$E[\xi_i] = 0 \ , \quad E[\xi_i, \xi_j] = \delta_{ij}(\sigma^2 \vartheta'(\gamma_0) d\vartheta(\gamma_0))$$

It is known [Wu (1981): Theorem 1] that if there exists a consistent estimator for all $\gamma_0 \in \Gamma \subset R^k$, where Γ is compact, then (under very mild conditions on the distribution of ξ_i)

$$[\Psi(\gamma) - \Psi(\gamma_0)]' M_n [\Psi(\gamma) - \Psi(\gamma_0)] \to \infty \tag{16.5}$$

as $n \to \infty$ for all $\gamma \neq \gamma_0$ in Γ. Here $M_n = \sum_{i=1}^n x_i x_i'$. Moreover, condition (16.5) provides the consistency of the least squares estimator of γ. In other words, an experimenter should appropriately choose a design $\{x_i\}_1^n$ as well as Γ. Note that condition (16.5) is not fulfilled if the set Γ includes $\gamma = -\gamma_0$ in (16.5). It is clear from (16.5) that no more than l parameters can be estimated if either σ^2 or d is unknown.

In cases when σ^2 is a function of x: $\sigma^2(x, \gamma)$, which can also depend upon γ, one has to apply iterative least square estimation [Fedorov (1974); Jobson and Fuller (1980); and Malyutov (1982)]:

$$\begin{aligned} \hat{\gamma} &= \lim_{s\to\infty} \gamma_s \\ \gamma_s &= \arg\min_{\gamma} \sum_{i=1}^{n} \frac{[y_i - \Psi'(\gamma)x_i]^2}{\sigma^2(x_1\gamma_{s-1}) + \vartheta'(\gamma_{s-1})d\vartheta(\gamma_{s-1})} \end{aligned} \tag{16.6}$$

The estimator $\hat{\gamma}$ will be consistent if (16.5) is fulfilled. It is worthwhile to note that the straightforward least square estimator

$$\gamma = \arg\min_{\gamma} \sum_{i=1}^{n} \frac{[y_i - \Psi'(\gamma)x_i]^2}{\sigma^2(x_1\gamma) + \vartheta'(\gamma)d\vartheta(\gamma)}$$

is not consistent.

A method similar to (16.6) can be used for *approximate* estimation of parameters of the generalized version of (16.1):

$$\begin{aligned} y_i &= \eta(\gamma_0, z_i) + \varepsilon_i \\ z_i &= \rho(\gamma_0, x_i) + \delta\nu_i = \rho_i + \delta\nu_i \ , \quad i = l, \ldots, n \end{aligned} \tag{16.7}$$

where γ stands for both parameters ϑ and B, $\eta \in R^1$, $\rho \in R^l$, and $\gamma \in R^k$; ε_i are *iid* random values with variance σ^2; ν_i are *iid* random vectors with unit variance matrix d; and δ is some constant.

It will be assumed that the function $\eta(\gamma, z)$ has derivates with respect to z up to the third ones for all $z_i = \rho(\gamma, x_i)$, $\gamma \in \Gamma \subset R^k$, where Γ is compact, and

$$E[|\nu_{ip}\nu_{iq}\nu_{ir}|] \le c < \infty \ , \quad i = 1, \ldots, n \ , \quad p, q, r = 1, \ldots, m$$

Similar to Fedorov (1974), one obtains

$$E[y_i] = E[\eta(\gamma_0, \rho_i + \delta\nu_i) + \varepsilon_i] = \varphi(\gamma_0, x_i) + O(\delta^3)$$

$$E[(y_i - E[y_i])^2] = \lambda^{-1}(\gamma_0, x_i) + O(\delta^3) \tag{16.8}$$

where

$$\varphi(\gamma, x) = \eta(\gamma, \rho) + \frac{\delta^2}{2} \operatorname{tr} d \left. \frac{\partial^2 \eta(\gamma, \rho)}{\partial \rho \partial \rho'} \right|_{\rho=\rho(\gamma,x)}$$

$$\lambda^{-1}(\gamma, x) = \sigma^2 + \delta^2 \left. \frac{\partial \eta(\gamma, \rho)}{\partial \rho'} d \frac{\partial \eta(\gamma, \rho)}{\partial \rho} \right|_{\rho=\rho(\gamma,x)}$$

Therefore the modified model (16.7) can be approximated by the regression model

$$y_i = \varphi(\gamma_0, x_i) + \mu_i$$

where $\lambda^{1/2}(\gamma_0, x_i)\mu_i$ are *iid* random values. The estimator of γ is defined, similar to (16.6), as a limit point of the following iterative procedure:

$$\hat{\gamma} = \lim_{s \to \infty} \gamma_s$$

$$\gamma_s = \arg\min_{\gamma \in \Gamma} \sum_{i=1}^{n} \lambda(\gamma_{s-1}, x_i)[y_i - \varphi(\gamma, x_i)]^2 \tag{16.9}$$

The estimator will be consistent within the frame of approximation (16.8) under very mild assumptions, the main one of which is

$$\sum_{i=1}^{n} \lambda(\gamma_0, x_i)[\varphi(\gamma, x_i) - \varphi(\gamma_0, x_i)]^2 \to \infty$$

as $n \to \infty$ for all $\gamma \neq \gamma_0$ in Γ. For "sufficiently" smooth functions $\varphi(\gamma, x_i)$, the estimator (16.9) is normally distributed asymptotically:

$$\tau(\hat{\gamma}_n - \gamma) \to N(0, M^{-1})$$

where $\tau_i \to \infty$ as $n \to \infty$, and

$$M = \lim_{n \to \infty} \tau_n^{-1} \sum_{i=1}^{n} \lambda(\gamma_0, x_i) \left. \frac{\partial \varphi(\gamma, x_i)}{\partial \gamma} \frac{\partial \varphi(\gamma, x_i)}{\partial \gamma'} \right|_{\gamma=\gamma_0} \tag{16.10}$$

Naturally, the existence of the limit in (16.10) is suggested. The last result is more general than a similar one from Fedorov (1974) and it is the obvious corollary of Theorem 5 from Wu (1981).

16.3 Multiresponse Regression Model

The model (16.2) can be treated as a specific case of the multiregression model

$$w_i = \eta(\gamma, x_i) + \mu_i \tag{16.11}$$

where $\eta(\gamma, x)$ is a given vector function, $\gamma \in R^k$; the random vectors μ_i are *iid* with zero means and the covariance matrix

$$E[\mu_i \mu_i'] = \Sigma = \begin{pmatrix} \sigma^2 & 0 \\ 0 & d_s \end{pmatrix}$$

Consider the case when all elements of B and ϑ are unknown. Model (16.2) can be transformed to model (16.11) if one assumes that

$$\gamma' = (\vartheta_1, \ldots, \vartheta_m, B_{11}, \ldots, B_{l1}, \ldots, B_{1m}, \ldots, B_{lm})$$

$$\eta_1(\gamma, x) = \sum_{\alpha=1}^{m} \vartheta_\alpha \sum_{\beta=1}^{l} B_{\beta\alpha} x_\beta \ , \quad \eta_2(\gamma, x) = \sum_{\beta=1}^{l} B_{\beta 1} x_\beta \ , \quad \eta_{m+1}(\gamma, x) = \sum_{\beta=1}^{l} B_{\beta m} x_\beta$$

Model (16.11) was studied, for instance, by Phillips (1976) and Fedorov (1977). Several slightly different estimators were suggested for γ and Σ (the case when Σ is given is too well-known to require discussion here). Similar to the estimator defined by (16.9), both γ and Σ can be estimated with the help of the following iterative procedure.

$$\begin{aligned} \hat{\gamma} &= \lim_{s \to \infty} \gamma_s \ , \quad \hat{\Sigma} = \lim_{s \to \infty} \Sigma_s \\ \gamma_s &= \arg\min_{\gamma \in \Gamma} \sum_{i=1}^{n} [w_i - \eta(\gamma, x_i)]' \Sigma_s^{-1} [w_i - \eta(\gamma, x_i)] \\ \Sigma_s &= \sum_{i=1}^{n} [w_i - \eta(\gamma_{s-1}, x_i)][w_i - \eta(\gamma_{s-1}, x_i)]' \end{aligned} \tag{16.12}$$

Unlike model (16.1), in the case under consideration, all elements of B and ϑ can be consistently estimated if the sequence $x_1, \ldots, x_n$ is appropriately chosen. The estimator (16.12) asymptotically coincides with the maximum likelihood estimator when the μ_i are normally distributed.

The estimator (16.12) can be improved if the structure of the covariance matrix Σ is taken into account, and in the iterative procedure instead of the matrices Σ_s, one uses

$$\Sigma_s^1 = \begin{pmatrix} \sigma^2 & 0 \\ 0 & d_s \end{pmatrix} = \sum_{i=1}^{n} \begin{pmatrix} (y_i - \vartheta_s' z_{si})^2 & 0 \\ 0 & (u_i - z_{si})(u_i - z_{si})' \end{pmatrix} \tag{16.13}$$

where $z_{si} = B_s' x_i$; $\tilde{\gamma}$ (or $\tilde{\vartheta}$ and $\tilde{B}$) and $\tilde{\Sigma}$ (or $\tilde{\sigma}^2$ and $\tilde{d}$) denotes the improved estimators of γ (or ϑ and B) and Σ (or σ^2 and d).

Let us introduce matrices

$$M_{11} = \sum_{i=1}^{n} \tilde{z}_i \tilde{z}_i' , \quad M_{12} = \sum_{i=1}^{n} x_i \tilde{z}_i' , \quad M_{22} = \sum_{i=1}^{n} x_i x_i'$$

where $\tilde{z}_i = \tilde{B}' x_i$. Using the standard techniques of regression analysis, it is possible to check that the consistent estimator $\tilde{D}$ of the variance matrix $D = E[\tilde{\gamma}\tilde{\gamma}']$ can be calculated in the following way:

$$\tilde{D} = (J_1 + J_2)^{-1}$$

where

$$J_1 = \tilde{\sigma}^{-1} \begin{pmatrix} M_{11} & \tilde{\vartheta}' \otimes M_{12}' \\ \tilde{\vartheta} \otimes M_{12} & \tilde{\vartheta}\tilde{\vartheta}' \otimes M_{22} \end{pmatrix} \quad \text{and} \quad J_2 = \begin{pmatrix} 0 & 0 \\ 0 & \tilde{d}^{-1} \otimes M_{22} \end{pmatrix}$$

The matrix J_1 can be interpreted as the information matrix corresponding to observation of y_i, and the matrix J_2 can be interpreted as the information matrix corresponding to observations of z_i.

When $\operatorname{rank}(B_0) = \operatorname{rank}(d_0) = m$, then the paramter ϑ and B will be consistenly estimated by (16.12) if

$$\lim_{n \to \infty} \frac{(\text{max eigenvalue of } M_{22})^{(1+c)/2}}{\text{max eigenvalue of } M_{22}} \leq c < \infty$$

for some $c > 0$. This fact is derived from the results of Wu (1981) and from the structure of the matrix $\tilde{D}$, which is mainly defined by the matrix M_{22}:

$$M_{11} = \tilde{B}' M_{22} \tilde{B} \quad \text{and} \quad M_{12} = M_{22} \tilde{B}$$

In conclusion, it should be emphasized that (16.9), (16.12), and (16.13) not only describe the estimators with some admissible statistical properties; they, moreover, deliver the effective numerical procedures that are based on the well-studied standard least square techniques.

References

Carroll, R.J. (1982), Adapting for heteroscedasticity in linear models. *Annals of Statistics*, **10**, 1224-1233.

Fedorov, V.V. (1974), Regression problems with controllable variables subject to errors. *Biometrika*, **61**, 49-56.

Fedorov, V.V. (1977), Estimation of regression parameters for the vector-response case, in: V.Natimov (ed.), *Regression Experiments.* Moscow: Moscow State University Press.

Jobson, J.D. and Fuller, W.A. (1980), Least squares estimation when the covariance matrix and parameter vector are functionally related. *Journal of the American Statistical Association*, **75**, 176-181.

Malyutov, M.B. (1982), Asymptotical properties and applications parameter estimators for generalization regression models. *Uspekhi Mathematicheskikh Nauk*, **37** (2), 209-210.

Phillips, P.C.B. (1976), The iterated minimum distance estimator and quasi-maximum likelihood estimator. *Econometrica*, **44**, 449-460.

Robinson, P.M. and Ferrara, M.C. (1977), The estimation of model for an unobservable variable with endogenous causes, in: D.J. Aigner and A.S. Goldberger (eds.), *Latent Variables in Socio-Economic Models.* Amsterdam: North-Holland.

Wu, C.F. (1981), Asymptotic theory of nonlinear least squares estimation. *Annals of Statistics*, **9**, 501-513.

Zellner, A. (1977), Estimation of regression relationships containing unobservable independent variables, in: D.J. Aigner and A.S. Goldberger (eds.), *Latent Variables in Socio-Economic Models.* Amsterdam: North-Holland.

CHAPTER 17

Structural Change and Time Series Analysis

Lyle D. Broemeling

Summary

This investigation introduces changing-parameter ARMA processes as a way to model a time series. Many time series exhibit a changing trend or a changing autocorrelation structure; that is to say, they have certain nonstationary characteristics that cannot be modeled by the usual ARMA representation. The analysis of a changing parameter process is accomplished by a Bayesian approach, where the posterior distributions of the parameters are derived, and the analysis is illustrated with a moving average model that has a changing autocorrelation function.

17.1 Introduction

Modern approaches to time series analysis assume that the realization or a tranformation of the data was generated by an ARMA process that is stationary and invertible. If the data exhibit a changing trend or an unstable covariance, differencing the data will often induce stationarity, and one can assume an ARMA process generated the data. This is the technique used in a Box-Jenkins (1970) time series. However, sometimes no transformation can induce stationarity.

In this chapter, the approach is to model the realization with ARMA processes that have changing parameters. If the data exhibit a changing trend or a changing covariance structure, these characteristics can be captured with such processes. In addition, such processes can be employed with a Box-Jenkins as diagnostic tests; however, such uses will be explored at a later date.

The following three sections will introduce: (a) ARMA processes with a changing trend, (b) ARMA processes with a changing autocorrelation function, and (c) a Bayesian analysis of these changing-parameter models, along with an example of a moving average process that has a changing covariance structure.

17.2 Trends in the Data

Consider the following model, where the observation $Y(t)$ is given by the stochastic difference equation

$$Y(t) = C_1 + (C_2 - C_1)S(t-m) + \theta Y(t-1) + e(t) - \phi e(t-1) \tag{17.1}$$

where t is any integer and the process $[e(t)\colon\ t \in I]$ is a sequence of independent $N(0, \tau^{-1})$ random variables. The parameters C_1, C_2, θ, and ϕ are real, and $|\theta| < 1$. In addition, m is an integer called the shift point and S is the function

$$S(t-m) = \begin{cases} 0, & t < m \\ 1, & t \geq m \end{cases} \tag{17.2}$$

The model (17.1) is a changing parameter Gaussian process with mean value function

$$\begin{aligned} E(t) &= C_1(1-\theta)^{-1}\ ,\ \ t \leq m \\ E(m+s) &= \theta^s C_1(1-\theta)^{-1} + C_2(1+\theta+\ldots+\theta^{s-1})\ ,\ \ s = 1,2,\ldots \end{aligned} \tag{17.3}$$

Thus, the mean is constant until time $m+1$, at which point it changes and either increases or decreases in value depending on C_1 and C_2. One can say the initial mean is $C_1(1-\theta)^{-1}$ while the "final" value of the mean is

$$\lim_{s \to \infty} E(m+s) = C_2(1-\theta)^{-1}$$

Now assuming that the variance of the process is a constant, then it must be

$$V(t) = \tau^{-1}(1 - \phi^2 - 2\theta\phi)(1-\theta^2)^{-1}\ ,\ \ t \in I$$

where $V(t) = Var[Y(t)]$. The covariance process is that of an ARMA$(1,1)$ model, and the lag-one value is

$$Cov[Y(t), Y(t+1)] = (1-\theta\phi)(\theta-\phi)(1-\theta^2)^{-1}$$

for all t.

This model has a stable autocorrelation function, but its mean is changing. In a later section, the Bayesian analysis will be demonstrated.

17.3 Changing Covariance Structure

How do we construct an ARMA-type model that has a constant mean but a changing covariance structure? One way to do this is by the stochastic difference equation

$$Y(t) = C + \theta Y(t-1) + e(t) - [\phi_1 + (\phi_2 - \phi_1)S(t-m)]\, e(t-1) \tag{17.4}$$

where everything is as before, but where ϕ_1 and ϕ_2 are real with $|\phi_i| < 1$, $i = 1, 2$.

It can be easily shown that the mean of the process is

$$E(t) = C(1-\theta)^{-1}\ ,\quad t \in I \tag{17.5}$$

and is thus constant. On the other hand, the variance of the process is

$$V(t) = \tau^{-1}(1 + \phi_1^2 - 2\theta\phi_1)(1-\theta^2)^{-1}\ ,\quad t \leq m \tag{17.6}$$

and

$$V(m+s) = \theta^{2s}V(m) + \tau^{-1}(1 + \phi_2^2 - 2\theta\phi_2)(1 + \ldots + \theta^{2s-2})$$

where $s = 1, 2, \ldots$.

The process begins with a variance that remains constant, then begins to change at time $m+1$, and the limiting value of the variance is

$$\lim_{s\to\infty} V(m+s) = \tau^{-1}(1 + \phi_2^2 - 2\theta\phi_2)(1-\theta^2)^{-1}$$

which is the same as that of an ARMA$(1,1)$ process with parameters θ and ϕ_2.

The autocovariance function is quite involved but the lag-one value is

$$\begin{aligned} Cov[Y(t), Y(t+1)] &= \tau^{-1}(1 - \theta\phi 1)(\theta - \phi_1)(1-\theta^2)^{-1}\ ,\quad t \leq m-1 \\ Cov[Y(m+s), Y(m+s+1)] &= \theta V(m+s) - \phi_2\tau^{-1}\ ,\quad s = 0, 1, \ldots \end{aligned} \tag{17.7}$$

where $V(m)$ is given by (17.6).

17.4 The Bayesian Analysis

The following analysis of a changing ARMA process is quite similar to that done by Broemeling and Shaarawy (1986) and Broemeling and Tsurumi (1986).

Consider the previous ARMA process (17.4) with a changing covariance structure; there the tth residual is

$$e(t) = Y(t) - C - \theta Y(t-1) + [\phi_1 + (\phi_2 - \phi_1)S(t-m)]\, e(t-1) \tag{17.8}$$

where $t \in I$. Let

$$D_n = [y(1), \ldots, y(n)]$$

be a realization of n observations and assume $1 \leq m \leq n-1$. Thus, the change in variance occurs somewhere during the period of observation, but we do not know at exactly which point. The likelihood function is

$$L(C, \theta, \phi_1, \phi_2, \tau, m | D_n) \propto \tau^{n/2} \exp\left\{-\frac{\tau}{2}\sum_{t=1}^{n} e^2(t)\right\} \tag{17.9}$$

but unfortunately the sum of squared residuals is nonlinear in the parameters and depends on the unobservable residuals. By letting $\epsilon(0) = 0$ and then estimating the parameters by minimizing

$$\sum_{t=1}^{n} e^2(t) = SS(m, \theta, \phi_1, \phi_2, C) \tag{17.10}$$

over the region where $m = 1, 2, \ldots, n-1$; $|\theta| < 1$, $|\phi_i| < 1$, $i = 1, 2$, and then letting

$$\hat{e}(t) = y(t) - C - \hat{\theta} y(t-1) + [\hat{\phi}_1 + (\hat{\phi}_2 - \hat{\phi}_1)S(t-m)]\, \hat{e}(t-1) \tag{17.11}$$

where $\hat{e}(0) = 0$ and $t = 1, 2, \ldots, n$, the likelihood function can be approximated by

$$\hat{L}(C, \theta, \phi_1, \phi_2, \tau, m | D_n) \propto \tau^{n/2} \exp\left\{-\frac{\tau}{2}\sum_{t=1}^{n} \check{e}^2(t)\right\} \tag{17.12}$$

where

$$\check{e}(t) = y(t) - C - \theta y(t-1) + [\phi_1 + (\phi_2 - \phi_1)S(t-m)]\, \hat{e}(t-1) \tag{17.13}$$

and $\hat{e}(t)$ is given by (17.11).

If this is done, the approximate likelihood function is of a normal-gamma form in the unknown parameters. Combining $\hat{L}$ with the prior density

$$\pi(m, \theta, \phi_1, \phi_2, \tau) \propto \tau^{-1} \tag{17.14}$$

where $m = 1, 2, \ldots, n-1$, $\theta \in R$, $\phi_i \in R$, and $\tau > 0$ will yield posterior inferences for the parameters. For example, by eliminating θ, ϕ_1, and ϕ_2, using the properties of the normal-gamma distribution will give the marginal posterior density of the shift point m, and one can estimate where the change in the variance occurred. Inferences about θ, ϕ_1, and ϕ_2 will be based on a mixture of trivariate t-distributions, where the mixing distribution is the marginal posterior mass function of the shift point m.

As an illustration, consider the Broemeling-Tsurumi (1986, pp. 178-180) example of a moving average process

$$Y(t) = e(t) + [0.5 - (0.5 + 0.5)S(t-m)]\, e(t-1)$$

where the $e(t) \sim N(0,1)$ and the moving average coefficient changed from $-0,5$ to 0.5 at $m = 40$. The realization contained 80 observations $y(t)$, $t = 1, 2, \ldots, 80$, and the initial least squares estimates were $\hat{\phi}_1 = -0.54$, $\hat{\phi}_2 = 0.66$, and $m = 42$. The Bayesian analysis gave the following posterior information

$E[M]$	$Var[M]$	mod$[M]$	$E[\phi_1]$	$E[\phi_2]$	$Var[\phi_1]$	$Var[\phi_2]$
41.96	10.12	42	-0.56	0.59	0.028	0.032

The above information was calculated from the appropriate marginal posterior distribution and more of the details can be found in Broemeling and Tsurumi (1986, Table 6.7). The results show that the Bayesian analysis does not give unreasonable inferences. The actual change was at $m = 40$, while the posterior mean was 41.96 and the posterior mode was 42.

Since this is a moving average process, the variance function is

$$V(t) = \begin{cases} \tau^{-1}(1+\phi_1^2)\,, & t = \ldots, m \\ \tau^{-1}(1+\phi_2^2)\,, & t = m+1, \ldots \end{cases}$$

and the correlation function is

$$\rho[Y(t), Y(t+s)] = \begin{cases} -\phi_1(1+\phi_1^2)^{-1}\,, & t = \ldots, m-1 \\ -\phi_2(1+\phi_1^2)^{-1/2}(1+\phi_2^2)^{-1/2}\,, & t = m \\ -\phi_2(1+\phi_2^2)^{-1}\,, & t = m+1, \ldots \end{cases}$$

where $s = 1$. The lag-two correlations are zero. A challenging problem here is to estimate the variance function directly by its posterior distribution. Note in the above example the correlation (lag-one) changed from a positive to a negative value.

17.5 Summary and Conclusions

The presentation here opens the door to many interesting problems. The most basic question is: does a changing-parameter ARMA process offer a viable alternative to the usual way of doing a time series analysis?

Another thought is the use of changing parameter representations as a diagnostic tool for the Box-Jenkins analysis. For example, suppose one tentatively identifies the series as being generated from the ARMA(1,1) process. The assumption in such an analysis is that the generating model is stationary; thus, it would have a constant mean and variance structure. A test for a constant mean could be based on the model (17.1), where one would extend the support of the shift point to $m = n$. A test of no-change in the mean could be made by computing the posterior probability that $m = n$.

These and additional questions will be studied in later investigations.

Acknowledgments

The author wishes to thank Joaquin Diaz-Siaz for his programming assistance and the Office of Naval Research for research support under contract N00082K0292.

References

Box, G.E.P. and Jenkins, G.M. (1970), *Time Series Analysis, Forecasting and Control.* San Francisco, CA: Holden-Day.

Broemeling, L.D. and Shaarawy, S. (1986), A Bayesian analysis of time series, Chapter 21 in: P. Goel and A. Zellner (eds.), *Bayesian Inference and Decision Techniques*, Amsterdam and New York: Elsevier.

Broemeling, L.D. and Tsurumi, H. (1986), *Econometrics and Structural Change.* New York: Marcel Dekker.

CHAPTER 18

Thresholds, Stability, Nonlinear Forecasting and Irregularly Sampled Data

Howell Tong

Summary

The central theme in this Chapter is unconventional analysis of time series data, the conventional one being that based on linear models (e.g., autoregressive/moving average models) and second-order moments (e.g., spectral analysis). After the natural emergence of thresholds, attention is focused on the stability of the global system in connection with that of each constituent subsystem delineated by the thresholds. Exotic results are obtained by relying on simple linear algebraic analysis of the system, which may be considered an application of symbolic dynamics. Some unexpected results are described in nonlinear forecasting, which expose a myth generated by linear mentality. Finally, comments are made about nonlinear modeling of irregularly sampled data.

18.1 Introduction

In conventional time series modeling, *linearity* is a common assumption, the adequacy and appropriateness of which in practice has been seriously challenged only quite recently. Under this assumption, the whole dynamic system, if globally and asymptotically stable, is destined to approach a steady state. This type of situation is actually the dullest, and often the most unrealistic, because one and the same mode of dynamics is assumed no matter where the current position of the system lies. Linearity will simply ignore saturation, starvation, etc., which exist in most (e.g., economic, ecological, etc.) systems and

which will almost certainly lead to a structural change, i.e., a change in the fundamental dynamics, somewhere in the state space.

As soon as we recognize the necessity of structural changes under certain conditions, we are effectively postulating the existence of *thresholds*, which may be either real and sharp (e.g., the effect of melting a glacier on riverflow at a *critical* temperature) or conceptual and blurred (e.g., an over-heated economy or a saturated animal population). A threshold is an expression of nonlinearity. It is, in fact, a form of "strong" nonlinearity in that a distinct mode of dynamics pertains to each side of the threshold. In other words, the whole state space is now divided into a number of *regimes*, each with its own dynamic. Once again, we can clearly visualize the great constraint imposed by the assumption of linearity: there the universe (i.e., the whole state space) is under the dictatorship of one single regime!

18.2 Threshold Inference

Can we infer something about the thresholds from real data? This is a crucial question, which often permits extremely interesting interpretations in different disciplines, such as the notion of a critical temperature in a meteorologic-hydrological system, a critical population size in demographic dynamics, the "take-off" point (state) of an economic system, etc.

The existence of a threshold in the dynamic implies a discontinuity of the dynamic, and this usually poses technical difficulties. One approach to statistical inference on the threshold from observations is to start with a net of smooth models that include the discontinuous case in the limit.

Let $F(.)$ denote a sufficiently smooth distribution function with a $F'(.)$ rapidly decaying at the tails. One possible choice is the standard normal distribution, which we now use for convenience of discussion. A net of smooth models may be constructed formally as:

$$\begin{aligned} X_t &= a_0 + a_1 X_{t-1} + \ldots + a_p X_{t-p} \\ &\quad + (b_0 + b_1 X_{t-1} + \ldots + b_p X_{t-p}) F\left(\frac{X_{t-d} - r}{z}\right) + e_t \end{aligned} \tag{18.1}$$

where d (an integer) ≥ 1, p (an integer) ≥ 0, $\{e_t\}$ is a sequence of *iid* random variables with e_t independent of X_s, $s < t$, and with zero mean and finite variance. The parameter z is real-valued and controls the amount of smoothing. As z varies over the real numbers we have a *net* of models. Clearly, the larger is z, the smoother is the transition, over states, from the linear dynamic specified by $(a_0, a_1, \ldots, a_p)$ to that specified by $(a_0 + b_0, a_1 + b_1, \ldots, a_p + b_p)$. The model becomes discontinuous at $z = 0$. The critical change point, i.e., the threshold, is the real-valued parameter r. If z is negligibly small, then r delineates two regimes, within each of which the simplest dynamic is postulated. (It is obviously possible to incorporate a more complex dynamic within the regimes, if so desired, but our

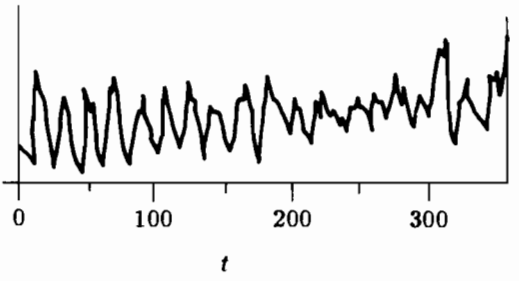

Figure 18.1: Time plot of square root of blowfly data.

focus here is on the threshold parameter.) Chan and Tong (1986) have given a rigorous discussion of conditions for ergodicity/stationarity of the models and other probabilistic aspects. As far as statistical inference is concerned, the key result is as follows.

Let θ denote the vector of parameters $(a_0, a_1, \ldots, a_p, b_0, b_1, \ldots, b_p, r, z)'$. Let θ_0 denote the true vector of parameters. Given a set of observations X_t, $t = 1, \ldots, n$, let $\hat{\theta}_n$ denote the conditional least squares estimate of θ_0, i.e., it is the minimizer of the conditional sum of squares

$$Q_n(\theta) = \sum_{t=m}^{n-1} [X_{t+1} - g(\theta, F_t)]^2 \tag{18.2}$$

where $m = \max(d, p)$, $g(\theta, F_t) = E_\theta[X_{t+1}|F_t]$, F_t being the sigma field generated by $X_1, X_2, \ldots X_t$. Then

$$\sqrt{n}\,(\hat{\theta}_n - \theta_0) \rightsquigarrow N(0, \sigma^2 V^{-1})$$

where

$$V = E_{\theta_0}\left[\frac{\partial g(\theta_0, F_m)}{\partial \theta_i}\,\frac{\partial g(\theta_0, F_m)}{\partial \theta_j}\right]$$

We now illustrate the approach with a real data set. Our analysis is quite preliminary, focusing on the threshold only.

In 1950, A.J. Nicholson started an experiment: Some blowflies were kept in cages, and a fixed amount of liver was provided daily. The population size of the blowflies was enumerated bi-daily for approximately two years. For more details about the experiment and other relevant data, see, e.g., Brillinger *et al.* (1980).

We have carried out analyses of the raw, the square root transformed, and the $\log_{10}$ transformed data. The general conclusions are similar, and we give details only for the

$\log_{10}$ transformed data. The results are summarized below. (Approximate standard errors are bracketed.)

First half of data set:

$\hat{a}_0$	$\hat{a}_1$	$\hat{a}_2$	$\hat{b}_0$	$\hat{b}_1$	$\hat{b}_2$	$\hat{r}$	$\hat{z}$	$\hat{\sigma}^2$
2.51	0.34	−0.03	−2.32	1.10	−0.49	3.103	0.084	0.021
(0.31)	(0.12)	(0.07)	(0.35)	(0.17)	(0.15)	(0.025)	(0.045)	(0.002)

Second half of data set:

$\hat{a}_0$	$\hat{a}_1$	$\hat{b}_0$	$\hat{b}_1$	$\hat{r}$	$\hat{z}$	$\hat{\sigma}^2$
0.85	0.77	−0.37	0.03	3.95	0.16	0.011
(0.19)	(0.05)	(0.70)	(0.22)	(0.18)	(0.13)	(0.001)

It is clear that for the first year the coefficients, the $\hat{b}$'s, are significantly different from zero, implying that different dynamics pertain to the different regimes delineated by the threshold parameter estimate $\hat{r}$. It would therefore seem reasonable to accept the population dynamicists' hypothesis of a threshold (i.e., a critical population size imposed by protein limitation), which is one of the main factors contributing to the observed population cycles in the first year (see *Figure 18.1*).

In contrast, the situation is quite different in the second year, where the $\hat{b}$'s are acceptably negligible, implying the universality of the linear dynamics specified by $\hat{a}$'s. Note that the "nuisance parameters" r and z are, strictly speaking, absent under the null hypothesis $b_i = 0$, all i. This gives rise to a rather interesting *nonstandard* problem, the formal solution of which is yet to be found. In conclusion, the threshold effect supported by observations in the first year seems to have disappeared after a prolonged period of captivity of the flies. For more details, see Tong (1987).

18.3 Stability

The idea of *innovations* is now well developed and may be formalized by the expression $X_t - E[X_t|F_{t-1}]$ where, as before, F_s denotes the sigma field generated by $X_s, X_{s-1}, \ldots$. Let the innovation sequence be denoted by $\{\epsilon_t\}$. Clearly, under general conditions (typically, we require the existence of the first moment), we may decompose a time series $\{X_t\}$ by

$$X_t = \hat{X}_t + \epsilon_t \tag{18.3}$$

where

$$\hat{X}_t = E[X_t|F_{t-1}] \,, \qquad \epsilon_t = X_t - \hat{X}_t$$

This may be put on a parallel with the classic decomposition

$$\text{Time Series} = \text{Trend} + \text{Random Series} \tag{18.4}$$

or the more recent and exotic decomposition

$$\text{Dynamical System} = \text{Slow Manifold} + \text{Fast Manifold} \tag{18.5}$$

We prefer to call

$$X_t = \hat{X}_t \tag{18.6}$$

the *skeleton* of the time series specified by (18.3); see, e.g., Tong (1986). Note that the skeleton is *deterministic*, but equation (18.3) specifies a *stochastic* process.

An interesting and fundamental question is: can we say something about the time series (18.3) by looking at its skeleton (18.6) only? Rather surprisingly, we can say quite a lot in the case of nonlinear autoregressive models, i.e., models of the form

$$X_t = g(X_{t-1}) + e_t \tag{18.7}$$

and its higher-order generalization. In this case

$$\hat{X}_t = g(X_{t-1})$$

and

$$\epsilon_t = e_t$$

Now, under very mild conditions on the distribution of e_t (milder than normality!) and on g (milder than continuity), if the skeleton is asymptotically stable at the origin in the sense that the recursion

$$x_t = g(x_{t-1}) \tag{18.8}$$

always tends to 0 as $t \to \infty$ regardless of initial x_0, then the time series defined by (18.7) is ergodic and, associated with equation (18.7), is a properly defined, strictly stationary time series. This very useful result may be greatly generalized and converse results may also be obtained. For references, see Chan and Tong (1985) and Chan (1986).

Now, except for the linear case, there is no systematic way of checking stability. For this reason, applied mathematicians and engineers have accumulated vast experiences, and it is up to us to "hitch hike" on them. Quite often a first-principle approach yields fascinating rewards. We now describe some of our experiences hitherto not reported in the West. (The discussion that follows is based on joint work with Dr. K.S. Chan, now at the University of Chicago, USA.) We are concerned here with *stability in the sense of Lagrange*, i.e., bounded trajectories of (18.8) and its higher-order generalization.

First, we consider the following very simple model:

$$X_n = \begin{cases} \phi_1 X_{n-1} & \text{if} \quad X_{n-1} > 0 \\ \phi_1' X_{n-1} & \text{otherwise} \end{cases} \tag{18.9}$$

It is known that the necessary and sufficient conditions for (18.9) to be Lagrange-stable is $\phi_1 \leq 1$, $\phi_1' \leq 1$ and $\phi_1 \phi_1' \leq 1$. [See, e.g., Petruccelli and Woolford (1984)]. For this system, the Lagrange stability of the whole system implies stability of at least one of the subsystems. Moreover, the Lagrange stability of the two sub-systems guarantees that of the whole system. The Petruccelli-Woolford condition is extremely interesting because either ϕ_1 or ϕ_1' (but not both) is allowed to be very large negatively without causing instability. Next, we give a class of examples showing that the whole system may be Lagrange-unstable even though all its subsystems are Lagrange-stable.

Consider

$$x_n = \begin{cases} \phi_1 x_{n-1} + \phi_2 x_{n-2} & \text{if} \quad x_{n-2} > 0 \\ \phi_1' x_{n-1} & \text{otherwise} \end{cases} \tag{18.10}$$

Suppose that $\phi_1 \geq 0$, $\phi_2 < 0$, $\phi_1' < 0$, and $\phi_1^2 + 4\phi_2 < 0$.

To understand the stability of (18.10), it is more convenient to use the following representation, the so-called state space representation. Let

$$\begin{aligned} \mathbf{Z}_n &= \begin{pmatrix} z_n^1 \\ z_n^2 \end{pmatrix} = \begin{pmatrix} x_n \\ x_{n-1} \end{pmatrix} \\ \mathbf{A} &= \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix} \\ \mathbf{B} &= \begin{pmatrix} \phi_1' & 0 \\ 1 & 0 \end{pmatrix} \end{aligned}$$

Note that $\mathbf{A}$ and $\mathbf{B}$ are both companion matrices, $\mathbf{B}$ being degenerate. Then, (18.10) is equivalent to

$$\begin{aligned} \mathbf{Z}_n &= \begin{cases} \mathbf{A}\mathbf{Z_{n-1}} & \text{if} \quad z_{n-1}^2 > 0 \\ \mathbf{B}\mathbf{Z_{n-1}} & \text{otherwise} \end{cases} \\ x_n &= (1,0)\mathbf{Z}_n \end{aligned} \tag{18.11}$$

Clearly, x_n is Lagrange-stable iff $\mathbf{Z}_n$ is so. Since $\phi_1^2 + 4\phi_2 < 0$, $\mathbf{A}$ is similar to a rotation followed by a contraction (or expansion).

Define

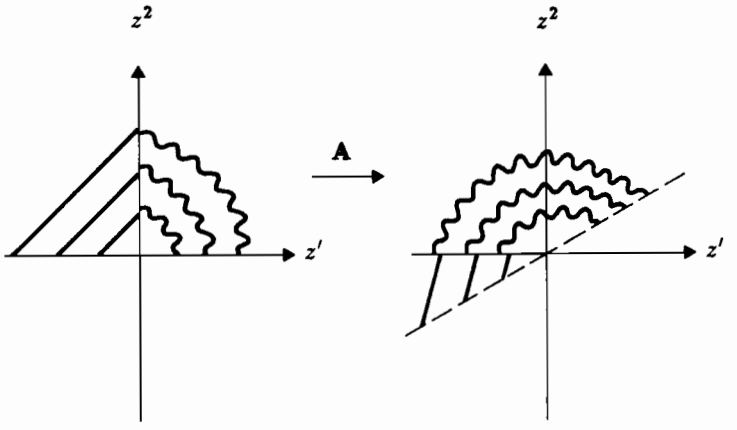

Figure 18.2: Action of **A** on U.

$$
\begin{aligned}
U &= \{(x,y) : y > 0\} \\
L &= \{(x,y) : y \leq 0\} \\
Q_1 &= \{(x,y) \in U : x > 0\} \\
Q_2 &= \{(x,y) \in U : x \leq 0\} \\
Q_3 &= \{(x,y) \in L : x < 0\} \\
Q_4 &= \{(x,y) \in L : x = 0\} \\
Q_5 &= \{(x,y) \in L : x > 0\}
\end{aligned}
$$

For every $\mathbf{Z}_n \in U$, $\mathbf{Z}_{n+1}$ is obtained by the action of **A** on $\mathbf{Z}_n$. We say that **A** controls U. Similarly, **B** controls L.

As

$$\mathbf{A}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \phi_1 \\ 1 \end{pmatrix} \quad \text{and} \quad \mathbf{A}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \phi_2 \\ 0 \end{pmatrix}$$

we have the pictorial representation of *Figure 18.2*, which displays the action of **A** on U.

Now, **A** is a rotation followed by a contraction or expansion after a change of basis. Hence, there exists a positive integer k, such that for every vector $\mathbf{x}_0 \in U$, there exists a positive integer $k(\mathbf{x}_0) < k$, such that

$$\mathbf{A}^{k(\mathbf{x}_0)}\mathbf{x}_0 \in L \quad \text{and} \quad \mathbf{A}^j\mathbf{x}_0 \notin L\ , \quad 0 \leq j < k(\mathbf{x}_0)$$

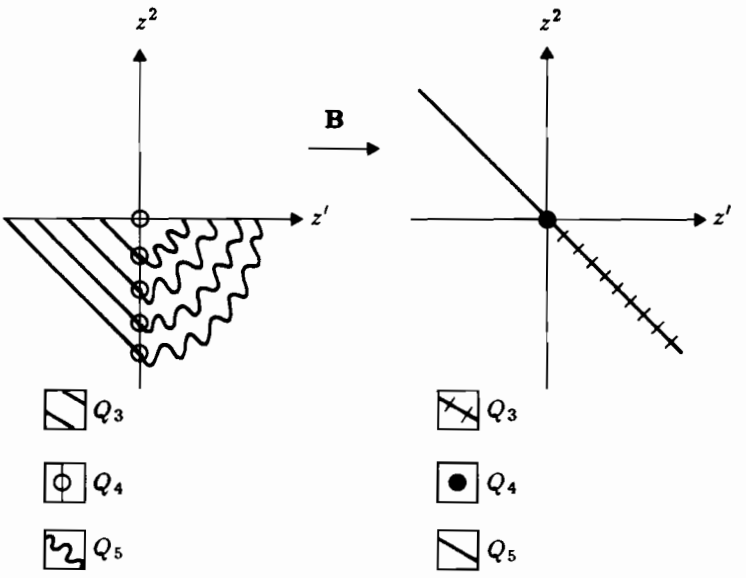

Figure 18.3: Action of **B** on L.

In particular, $\forall \mathbf{x}_0 \in Q_2$, $k(\mathbf{x}_0) = 1$. The action of **B** on L is displayed in *Figure 18.3*.

Let

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in R^2$$

and T: $R^2 \to R^2$ be defined by

$$T\mathbf{x} = \begin{cases} \mathbf{Ax} & \text{if} \quad x_2 > 0 \\ \mathbf{Bx} & \text{otherwise} \end{cases}$$

Then, (18.11) is equivalent to

$$\begin{aligned} \mathbf{Z}_n &= T(\mathbf{z}_{n-1}) \\ x_n &= (1,0)\mathbf{Z}_n \end{aligned} \tag{18.12}$$

Given $\mathbf{Z}_0$, $\{\mathbf{Z}_i,\ i = 0,1,2,\ldots\}$ is the trajectory under T with initial state $\mathbf{Z}_0$. We now classify these trajectories. Let $\mathbf{x}_0 \in U$; then let $k = k(\mathbf{x}_0)$. We say that the trajectory is of type $(S_1, S_2, S_3, \ldots, S_k; S'_0, S'_1, \ldots, S'_{\ell-1})$ iff for $j \geq 1$,

Table 18.1: Trajectories (semi-colon is omitted inside the parentheses where $k = 0$)

Region	*Type*
Q_1 and Q_2	$(\underbrace{\mathbf{A}, \mathbf{A}, \ldots, \mathbf{A}}_{k(\mathbf{x}_0) \text{ of them}}; \mathbf{B}, \mathbf{B}, \mathbf{A})$
Q_3	$(\mathbf{B}, \mathbf{B}, \mathbf{A})$
Q_4	$(\mathbf{B}; \mathbf{0})$
Q_5	$(\mathbf{B}, \mathbf{A}, \mathbf{B})$

$$T^j \boldsymbol{x} = \begin{cases} S_j T^{j-1} \boldsymbol{x} & \text{if } 1 \le j \le k \\ S'_{j-k-1 (\mathrm{mod}\, \ell)} T^{j-1} \boldsymbol{x} & \text{if } \quad j > k \end{cases}$$

where S_i and S'_i are $\mathbf{A}$ or $\mathbf{B}$ or $\mathbf{0}$ (the zero matrix), $T^0 \boldsymbol{x} = \boldsymbol{x}$.

Table 18.1 summarizes all possible types of trajectories under study. Types are not uniquely represented because

$$(\mathbf{B}, \mathbf{A}, \mathbf{B}) = (\mathbf{B}, \mathbf{A}; \mathbf{B}, \mathbf{B}, \mathbf{A})$$

Thus, it is readily seen that the "tail" of the trajectory starting from any initial (point), excepting those in Q_4, can be understood by examining the trajectory starting from

$$\mathbf{A} \begin{pmatrix} \phi'_1 \\ 1 \end{pmatrix}$$

Now

$$\mathbf{A}\mathbf{B}^2 = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \phi'_1 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \phi'_1 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \phi_1 \phi'^2_1 + \phi_2 \phi'_1 & 0 \\ \phi'^2_1 & 0 \end{pmatrix}$$

and

$$\mathbf{A} \begin{pmatrix} \phi'_1 \\ 1 \end{pmatrix} = \begin{pmatrix} \phi_1 \phi'_1 + \phi_2 \\ \phi'_1 \end{pmatrix}$$

giving

$$\mathbf{A}\mathbf{B}^2 \begin{pmatrix} \phi_1 \phi'_1 + \phi_2 \\ \phi'_1 \end{pmatrix} = \phi'_1 (\phi_1 \phi'_1 + \phi_2) \begin{pmatrix} \phi_1 \phi'_1 + \phi_2 \\ \phi'_1 \end{pmatrix} \tag{18.13}$$

Theorem 18.1 *If* $\{\mathbf{Z}_n\}$ *satisfying (18.12) has* $\phi_1 \ge 0$, $\phi_2 < 0$, $\phi'_1 < 0$ *and* $\phi_1^2 + 4\phi_2 < 0$, *then it is Lagrange-stable iff* $\phi'_1(\phi_1 \phi'_1 + \phi_2) \le 1$.

Proof: First note that the tail behavior of points excepting Q_4 is the same as that of

$$\begin{pmatrix} \phi_1\phi_1' + \phi_2 \\ \phi_1' \end{pmatrix}$$

and the latter is bounded iff $\phi_1'(\phi_1\phi_1' + \phi_2) \leq 1$. [Recall that, by assumption, $\phi_1'(\phi_1\phi_1' + \phi_2) \geq 0$.] For points in Q_4 they remain at the origin after the action of $\mathbf{B}$.

Corollary 1: If $\{X_n\}$ satisfying (18.10) has $\phi_1 \geq 0$, $\phi_2 < 0$, $\phi_1' < 0$, and $\phi_1^2 + 4\phi_2 < 0$, then it is Lagrange-stable iff $\phi_1'(\phi_1\phi_1' + \phi_2) \leq 1$.

Example 1: This is modified from an example given in Tong and Pemberton (1980).

$$X_n = \begin{cases} 1.8X_{n-1} - 0.9X_{n-2} & \text{if } X_{n-2} > 0 \\ -0.9X_{n-1} & \text{otherwise} \end{cases}$$

The two subsystems are stable. However $\phi_1'(\phi_1\phi_1' + \phi_2) = 2.268$. Therefore $\{X_n\}$ is not Lagrange-stable. This helps to explain the observation made by Tong and Pemberton (1980) to the effect that subsystem stability does not imply system stability.

Example 2:

$$X_n = \begin{cases} 0.6X_{n-1} - 0.1X_{n-2} & \text{if } X_{n-2} > 0 \\ -1.1X_{n-1} & \text{otherwise} \end{cases}$$

Here, the "upper" subsystem is stable, but the "lower" one is unstable. However, the whole system is stable since $\phi_1'(\phi_1\phi_1' + \phi_2) = 0.836 < 1$.

The conditions in *Theorem 18.1* are complicated. The characteristic equation of $\mathbf{A}$ has complex eigenvalues iff $\phi_1^2 + 4\phi_2 < 0$. $\mathbf{A}$ is then similar to a rotation followed by an expansion (or contraction). As this last property guarantees that, after a finite number of iterations, points in U will go to L, it is assumed to hold in subsequent discussion. In order that the discriminant of the characteristic equation be negative, ϕ_2 must be negative. If ϕ_1' is non-negative, then points once in $Q_3 \cup Q_4$ will always lie in $Q_3 \cup Q_4$. Thus, when $\phi_1' \geq 0$, the whole system is Lagrange-stable iff $\phi_1' \leq 1$. Now we consider the case when $\phi_1 < 0$ and $\phi_1' < 0$. The action of $\mathbf{A}$ on U is displayed in *Figure 18.4*.

Now, similar to the discussion preceding *Theorem 18.1*, it is clear that the trajectories starting with

$$\begin{pmatrix} \phi_1' \\ 1 \end{pmatrix}$$

gives us an idea about the Lagrange stability of the whole system. If $\mathbf{A}(\phi_1' \ 1)'$ lies in Q_3, then it is of type $(\mathbf{A}, \mathbf{B}, \mathbf{B})$. If $\mathbf{A}(\phi_1' \ 1)'$ lies in Q_4, then it is of type (0). If $\mathbf{A}(\phi_1' \ 1)'$ lies

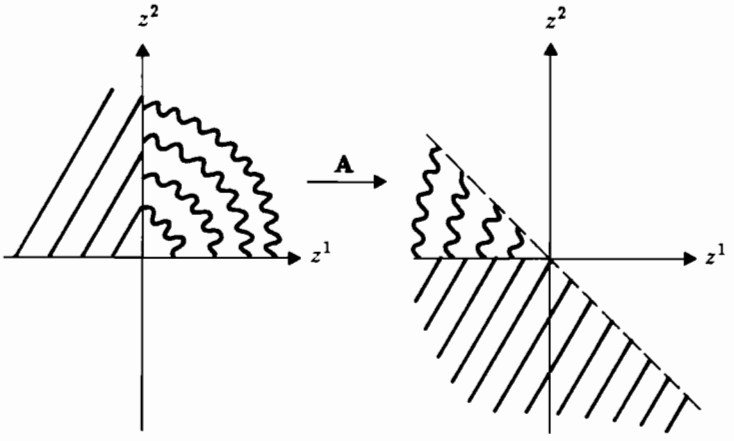

Figure 18.4: Action of **A** on U.

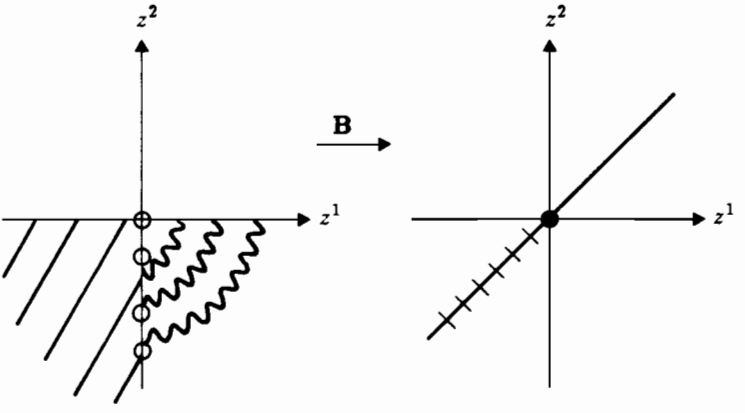

Figure 18.5: Action of **A** on L.

in Q_5, then it is of type $(\mathbf{A}, \mathbf{B})$. Since $\mathbf{A}(\phi_1' \ 1)' = (\phi_1\phi_1' + \phi_2 \ \ \phi_1')'$, it will lie in Q_3 or Q_4 or Q_5, according as $\phi_1\phi_1' + \phi_2 < 0$ or $= 0$ or > 0.

The action of $\mathbf{B}$ on L is displayed in *Figure 18.5*.

For the case $\phi_1 < 0$ and $\phi_1' \geq 0$, we note that, as discussed above, ϕ_1' must be less than or equal to 1 if the model is to be Lagrange-stable. Moreover, the trajectory of $(\phi_1' \ 1)'$ gives us information about the Lagrange stability of the model.

First, let $\phi_1' > 0$. Then it is clear that $\mathbf{A}(\phi_1' \ 1)' \in U$ and $\mathbf{A}^2(\phi_1' \ 1)' \in L$. Since

$$\mathbf{A}^2 = \begin{pmatrix} \phi_1^2 + \phi_2 & \phi_1\phi_2 \\ \phi_1 & \phi_2 \end{pmatrix}$$

$$\mathbf{A}^2 \begin{pmatrix} \phi_1' \\ 1 \end{pmatrix} = \begin{pmatrix} (\phi_1^2 + \phi_2)\phi_1' + \phi_1\phi_2 \\ \phi_1'\phi_1 + \phi_2 \end{pmatrix}$$

$\mathbf{A}^2(\phi_1' \ 1)' \in Q_3$ or Q_4 or Q_5, depending on whether $(\phi_1^2 + \phi_2)\phi_1' + \phi_1\phi_2 < 0$ or $= 0$ or > 0. Thus, the type of $(\phi_1' \ 1)'$ may be $(\mathbf{A}, \mathbf{A}; \mathbf{B})$ or $(\mathbf{A}, \mathbf{A}; \mathbf{0})$ or $(\mathbf{A}, \mathbf{A}, \mathbf{B})$.

Now,

$$\mathbf{BA}^2 = \begin{pmatrix} \phi_1' & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \phi_1^2 + \phi_2 & \phi_1\phi_2 \\ \phi_1 & \phi_2 \end{pmatrix} = \begin{pmatrix} \phi_1'(\phi_1^2 + \phi_2) & \phi_1'\phi_1\phi_2 \\ \phi_1^2 + \phi_2 & \phi_1\phi_2 \end{pmatrix}$$

$$\mathbf{BA}^2 \begin{pmatrix} \phi_1' \\ 1 \end{pmatrix} = \begin{pmatrix} \phi_1'\phi_1'(\phi_1^2 + \phi_2) + \phi_1'\phi_1\phi_2 \\ \phi_1'(\phi_1^2 + \phi_2) + \phi_1\phi_2 \end{pmatrix} = [\phi_1'(\phi_1^2 + \phi_2) + \phi_1\phi_2] \begin{pmatrix} \phi_1' \\ 1 \end{pmatrix}$$

If $\phi_1' = 0$, then $\mathbf{A}(\phi_1' \ 1)' \in Q_3$. Hence, the type of $(\phi_1' \ 1)'$ will be $(\mathbf{A}, \mathbf{B}; \mathbf{0})$. *Theorem 18.1* may therefore be extended to the following theorem.

Theorem 18.2 *Let $\{Z_n\}$ satisfy (18.12). Suppose $\phi_1^2 + 4\phi_2 < 0$. Then $\{Z_n\}$ is Lagrange-stable iff one of the following conditions holds*

1. $\phi_1' < 0$, $\phi_1 \geq 0$ and $\phi_1'(\phi_1\phi_1' + \phi_2) \leq 1$
2. $\phi_1' < 0$, $\phi_1 < 0$, $\phi_1\phi_1' + \phi_2 < 0$ and $\phi_1'(\phi_1\phi_1' + \phi_2) \leq 1$
3. $\phi_1' < 0$, $\phi_1 < 0$ and $\phi_1\phi_1' + \phi_2 = 0$
4. $\phi_1' < 0$, $\phi_1 < 0$, $\phi_1\phi_1' + \phi_2 > 0$ and $\phi_1\phi_1' + \phi_2 \leq 1$
5. $\phi_1' = 0$
6. $1 \geq \phi_1' > 0$, $\phi_1 \geq 0$
7. $1 \geq \phi_1' > 0$, $\phi_1 < 0$, $(\phi_1^2 + \phi_2)\phi_1' + \phi_1\phi_2 \leq 0$

8. $1 \geq \phi_1' > 0$, $\phi_1 < 0$, $(\phi_1^2 + \phi_2)\phi_1' + \phi_1\phi_2 > 0$, and $\phi_1'(\phi_1^2 + \phi_2) + \phi_1\phi_2 \leq 1$

Corollary 1 may be extended likewise.

Finally, we have since learned from Professor Zhu Zhao Xuan of the Department of Mechanics, Peking University, China, that our above analysis is an application of what is known as *symbolic dynamics*, the "symbols" being the matrix operators **A** and **B** in our case. We conjecture that similar but more complex analysis would lead to a complete solution of the general case involving an arbitrary (but finite) number of arbitrary companion matrices.

18.4 Nonlinear Forecasting

Once again, we consider a first-order nonlinear autoregressive model

$$X_t = h(X_{t-1}) + e_t \ , \quad t = 0, \pm 1, \pm 2, \ldots \tag{18.14}$$

where $\{e_t\}$ is a sequence of iid random variables with $E[e_t] = 0$ and $Var[e_t] = \sigma^2$, $(0 < \sigma^2 < \infty)$ for all t. Since we have here a Markov chain over R, we may recall the Chapman-Kolmogorov relation

$$f(x_{t+m}|x_t) = \int_{-\infty}^{\infty} f(x_{t+m}|x_{t+1}) f(x_{t+1}|x_t) dx_{t+1} \tag{18.15}$$

where $f(x_s|x_t)$ denotes the conditional probability density function of X_s given $X_t = x_t$ (assumed to exist). Suppose that model (18.14) is strictly stationary. Let g denote the probability density function of e_t. Let k denote a well-behaved (i.e., Baire) function of X_t and suppose $E[|k(X_t)|] < \infty$. Let $K_m(X_t)$ denote the conditional expectation $E[k(X_{t+m}|X_t)]$. Equation (18.15) gives immediately

$$K_m(x_t) = \int_{-\infty}^{\infty} K_{m-1}(x_{t+1}) f(x_{t+1}|x_t)\, dx_{t+1} \tag{18.16}$$

i.e.,

$$K_m(x) = \int_{-\infty}^{\infty} K_{m-1}(y)\, g(y - h(x))\, dy \tag{18.17}$$

Equation (18.17) gives, in particular, recursive formulae for conditional expectations and conditional variances, which are of the type recently discussed by, e.g., Al-Qassem and Lane (1987) and Pemberton (1987), as alternatives to the methods developed by Jones (1976). Except for special cases of h (e.g., the linear case), the integral in equation (18.17) does not readily admit analytic solution, and numerical integration is commonly the only solution. Experiences show that, for piecewise linear h, numerical techniques

work quite efficiently. Needless to say, if numerical integration is employed, care must be taken to avoid accumulation of rounding errors. An alternative techniques that is especially useful for higher-order autoregressive models is the Monte Carlo method. Here, a sufficiently long record of data is simulated in accordance with the model [i.e., equation (18.14) or its higher-order generalisation] and the sample estimate of $K_m(x)$ is taken as an approximation of $K_m(x)$. Detailed comparison will be given elsewhere. Suffice it to say that both methods give reasonable approximations.

On using a numerical integration technique with accuracy to more than two decimal places, *Table 18.2* gives the results of two experiments. In Experiment I, $x_0 = 4.0435$, and

$$h(x) = \begin{cases} 1.5 - 0.9x & \text{if} \quad x \leq 0 \\ -0.4 - 0.6x & \text{if} \quad x > 0 \end{cases} \tag{18.18}$$

and in Experiment II, $x_0 = 5.0$ and

$$h(x) = \begin{cases} 1.5 - 0.7x & \text{if} \quad x \leq 0 \\ 1.0 + 0.8x & \text{if} \quad x > 0 \end{cases} \tag{18.19}$$

In each experiment, $e_t \sim N(0, \sigma^2)$.

Several comments are in order:

1. Let $h^m(x)$ denote $h(h(\ldots(h(x))\ldots))$, the m-fold application of h. With decreasing signal-to-noise ratio, the difference between $\hat{X}_0(m)$ and $h^m(x_0)$ increases.

2. For model (18.19), which admits only a periodic attractor of period 1 [i.e., $h^m(x) \to 5.0$ as $m \to \infty$, for all x], $\hat{\sigma}(m)$ is a monotonic increasing function of m for all the three choices of σ.

3. In contrast, for model (18.18), which admits a periodic attractor of period 2 at $C = \{-2.8261, 4.9435\}$, [i.e., $h^m(x) \to C$ as $m \to \infty$, for all x], we observe that for $\sigma = 0.4$ and 1.0, $\hat{\sigma}(1) < \hat{\sigma}(3) < \hat{\sigma}(5)\ldots$ and $\hat{\sigma}(2) < \hat{\sigma}(4) < \hat{\sigma}(6)\ldots$. However, $\hat{\sigma}(2m) \not< \hat{\sigma}(2m+1)$, $m = 1, 2, \ldots$.

4. It is clear that the limiting behavior of $h^m(x)$ as $m \to \infty$ exerts important influence on the multistep forecasts and their precision (i.e., the conditional variances). The influence is progressively more transparent with reducing noise variance. In general, if h has a limiting r-cycle $c_1, c_2, \ldots c_r$, i.e.,

$$h : c_i \to c_{i+1} \quad \text{for} \quad i = 1, 2, \ldots, r-1 \quad \text{and} \quad h : c_r \to c_1$$

then we would expect the stationary probability density function to attach a weight $1/r$ to each of the regions around $c_1, c_2, \ldots, c_r$ and the density could be expressed as $[f_1(x - c_1) + \ldots + f_r(x - c_r)]/r$, where $f_1, \ldots, f_r$ are density functions with variances

Table 18.2: Conditional means and conditional variance $[\hat{\sigma}(m) = \sqrt{Var[X_m|X_0]}]$ of two threshold autoregressive models.

	$\sigma = 0.4$		$\sigma = 1.0$		$\sigma = 2.0$	
m	$\hat{X}_0(m)$	$\hat{\sigma}(m)$	$\hat{X}_0(m)$	$\hat{\sigma}(m)$	$\hat{X}_0(m)$	$\hat{\sigma}(m)$
Experiment I						
1	−2.8261	0.4000	−2.8261	1.0000	−2.8261	2.0000
2	4.0435	0.5381	4.0392	1.3571	3.9151	2.8762
3	−2.8261	0.5141	−2.8177	1.3076	−2.5339	2.9890
4	4.0435	0.6116	4.0092	1.6107	3.5602	3.6110
5	−2.8261	0.5428	−2.7736	1.4859	−2.1061	3.5869
6	4.0435	0.6314	3.9524	1.7772	3.1537	4.0278
7	−2.8261	0.5509	−2.7073	1.6497	−1.6943	3.9603
8	4.0435	0.6371	3.8805	1.9228	2.7753	4.2814
9	−2.8261	0.5533	−2.6309	1.8038	−1.3321	4.2007
10	4.0435	0.6387	3.8022	2.0557	2.4444	4.4415
11	−2.8261	0.5540	−2.5508	1.9454	−1.0225	4.3601
12	4.0435	0.6392	3.7217	2.1769	2.1615	4.5453
13	−2.8261	0.5542	−2.4701	2.0741	−0.7604	4.4685
14	4.0435	0.6393	3.6414	2.2871	1.9217	4.6137
15	−2.8261	0.5542	−2.3902	2.1905	−0.5394	4.5438
Experiment II						
1	5.0000	0.4000	5.0000	1.0000	5.0000	2.0000
2	5.0000	0.5123	5.0000	1.2806	5.0091	2.5477
3	5.0000	0.5727	5.0000	1.4315	5.0524	2.7939
4	5.0000	0.6082	5.0003	1.5198	5.1159	2.9150
5	5.0000	0.6299	5.0007	1.5731	5.1551	2.9802
6	5.0000	0.6433	5.0012	1.6058	5.1966	3.0177
7	5.0000	0.6518	5.0018	1.6260	5.2300	3.0401
8	5.0000	0.6572	5.0024	1.6387	5.2565	3.0542
9	5.0000	0.6606	5.0029	1.6466	5.2772	3.0633
10	5.0000	0.6628	5.0033	1.6515	5.2934	3.0693
11	5.0000	0.6642	5.0037	1.6546	5.3058	3.0734
12	5.0000	0.6651	5.0040	1.6566	5.3154	3.0763
13	5.0000	0.6657	5.0042	1.6578	5.3228	3.0784
14	5.0000	0.6660	5.0044	1.6586	5.3285	3.0799
15	5.0000	0.6663	5.0046	1.6591	5.3329	3.0810

$\sigma_1^2, \ldots, \sigma_r^2$, respectively. The f_is tend to the δ-functions as the noise variance tends to zero. Therefore, unless the σ_is are all equal (this would be trivially so if $r = 1$), monotonicity of $\hat{\sigma}(m)$ in m will not be obtained. It may be shown that σ_is are all equal only in exceptional cases.

The above comments fully expose the myth perpetuated by the linear mentality in believing that $\hat{\sigma}(m)$ is a *monotonic* increasing function of m, i.e., the myth that the further ahead we forecast, the less "reliable" is the forecast. [This section is based on joint work reported in Tong and Moenaddin (1987).]

18.5 Irregularly Sampled Data

In practice, we frequently come across irregularly sampled time series data. For example, some pollution data are known to be collected irregularly with higher frequencies in the high-risk periods; some medical data are also collected irregularly so as not to cause the patients undue inconvenience; and daily business data typically contain gaps corresponding to weekends, public holidays, or other extraneous factors. The analysis of such data has been attracting substantial attention recently. See, for instance, the proceedings edited by Parzen (1984). However, the analyses are overwhelmingly linear-model-based and second-order moment-orientated.

We now sketch a possible nonlinear-model-based approach. Let $\{X(t_k)\colon k = 1, \ldots, n\}$ denote the irregularly sampled data, i.e., the time differences $t_k - t_{k-1}$ are not all equal. Now, R.H. Jones [see, e.g., Parzen (1984)] introduced the ingenious idea of "embedding" these in a continuous-time *linear* autoregressive model of the form (written heuristically and to first order only for simplicity of discussion)

$$\frac{dX(t)}{dt} = FX(t) + GW(t) \tag{18.20}$$

where F and G are constants and $W(t)$ is a continuous-time white Gaussian process with zero mean and unit variance. Since the observations lie on the integral path of (18.20), they must satisfy the equation

$$X(t_{k+1}) = \phi(t_{k+1}, t_k)\, X(t_k) + W_k \tag{18.21}$$

where

$$\phi(t_{k+1}, t_k) = \exp\{F(t_{k+1} - t_k)\}$$

and $\{W_k\}$ is a white noise sequence with

$$W_k \sim N(0, Q_k)\ , \quad Q_k = \int_{t_k}^{t_{k+1}} [\phi(t_{k+1}, \tau)\, G]^2 d\tau$$

Jones [in Parzen (1984)] has developed an algorithm based on Kalman's filter to obtain maximum likelihood estimates of F and G through those of ϕ and Q_k.

Now, the *general threshold principle* (which states that *we should simplify a complex nonlinear system by a process of introducing thresholds to partition the state space into regimes within each of which the simplest dynamic obtains*) suggests that we should consider piecewise linear differential equations as our approximation to a complex nonlinear system. Thus, equation (18.20) may be generalized to

$$\begin{cases} \frac{dX(t)}{dt} = F_1 X(t) + G_1 W(t) & \text{if} \quad X(t) \leq r \\ \frac{dX(t)}{dt} = F_2 X(t) + G_2 W(t) & \text{if} \quad X(t) > r \end{cases} \tag{18.22}$$

We keep the number of regimes to two for simplicity of discussion. In principle, there is no difficulty in increasing the number of regimes. The beauty of piecewise linearization is that we can still integrate equation (18.22), piece by piece, to obtain a complete integral path upon which $\{X(t_k)\}$ are assumed to lie. Jones' algorithm can be lifted with minor modifications to suit the present case provided we make the not-unreasonable assumption that, if $X(t_j)$ and $X(t_{j+1})$ lie in different regimes, then the integral path will cross r only once over the time interval $[t_j, t_{j+1}]$. Details of our approach will be reported elsewhere.

18.6 Concluding Remarks

Nonlinearity should be the rule, rather than the exception, in any realistic analysis of real systems, and the notion of a threshold is generic for nonlinear analysis. Some have complained that nonlinear systems defy simple analysis. However, we have shown in this chapter that, by postulating the simplest type of dynamic (namely, linear) within each of the regimes delineated by the thresholds, we do achieve a remarkable conceptual simplification of what would otherwise be a very complex system.

We have followed up by illustrating the possibility of simple analysis (linear algebra being the prerequisite of every educated time series analyst or econometrician!), which often leads to previously unexpected vistas. If this chapter persuades readers to take a long and hard look at the shackles of linearity, then I consider my efforts in writing it well expended.

Acknowledgments

I am very grateful to Professor Dr. Peter Hackl for retaining his interest in an absentee participant and for his invitation to write the present chapter. I am also indebted to Dr. Russell Gerrard for supplying me with a detailed analysis, upon which comment (4) in Section 18.4 rests.

References

Al-Qassem, M.S. and Lane, J.A. (1987), Forecasting exponential autoregressive models of order 1. To appear in *Journal of Time Series Analysis.*

Brillinger, D.R., Guckenheimer, J., Guttorp, P.E., and Oster, G. (1980), Empirical modeling of population time series data: The case of age and density dependent vital rates. *Lectures on Mathematics in the Life Sciences*, (American Mathematical Society), **13**, 65-90.

Chan, K.S. (1986), *Topics in non-linear time series analysis*, unpublished Ph.D. thesis, Department of Mathematics, Princeton University, U.S.A.

Chan, K.S. and Tong, H. (1985), On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations. *Advanced Applied Probability*, **17**, 666-678.

Chan, K.S. and Tong, H. (1986), On estimating thresholds in autoregressive models. *Journal of Time Series Analysis*, **7**, 179-190.

Jones, D.A. (1976), *Non-linear Autoregressive Processes*, unpublished Ph.D. thesis, University of London, U.K.

Parzen, E. (1984), *Time Series Analysis of Irregularly Observed Data.* Lecture Notes in Statistics No. 25. New York: Springer.

Pemberton, J. (1987), Exact least-squares multi-step prediction from non-linear autoregressive models. To appear in *Journal of Time Series Analysis.*

Petruccelli, J.D. and Woolford, S.W. (1984), A threshold $AR(1)$ model. *Journal of Applied Probability*, **21**, 270-286.

Tong, H. (1986), Non-linear time series models of regularly sampled data: A review. Invited paper presented at the First World Congress of the Bernoulli Society, September, Tashkent, USSR. [Proceedings to be published by VNU, Holland, 1987.]

Tong, H. (1987), Non-linear time series modelling in population biology: a preliminary case study, in: R. Mohler (ed.), *Non-linear Time Series and Variables Structure in Signal Processing.* New York: Springer.

Tong, H. and Moenaddin, R. (1987), A note on the conditional variance of the error of a multi-step non-linear least squares predictor. Technical Report, Mathematical Institute, University of Kent at Canterbury, U.K.. To appear in *The Statisticians* 1988.

Tong, H. and Pemberton, J. (1980), On the stability and limit cycles of non-linear autoregression in discrete time. *Cahiers du CERO*, **22** (2), 137-148.

CHAPTER 19

Forecasting in Situations of Structural Change: A General Approach

Francis X. Diebold and Peter Pauly

Summary

The problem of optimal forecast combination is considered in situations of structural change. We develop a rather general approach, which combines the time-varying-parameter models of Diebold and Pauly (1987a) with allowance for prediction-error serial correlation as in Diebold (1988). The methodology is based on the regression-based paradigm of Granger and Ramanathan (1984), so that many earlier results emerge as special (and often restrictive) cases. Both deterministic and stochastic parameter variations are considered, with and without allowance for serial correlation. The results are illustrated in a series of examples.

19.1 Introduction

Forecasters are constantly challenged by continuous structural changes in the relationships of interest. As Makridakis *et al.* (1984) write, "... forecasting must ... accept that structural changes in the data are and will be taking place .... The major question, then, becomes how the various methods perform under a continuously changing environment." Naturally, modelers seek to identify structural changes in the process of model specification, and they generally attempt to incorporate extraneous adjustments in the forecast to account for those shifts not yet modeled in a nonparametric way. Nevertheless, forecasts remain susceptible to changes in the environment. Furthermore, various candidate models, such as different structural econometric models, nonstructural time series models, or

expert consensus forecasts, may turn out to be vulnerable to structural change in different degrees.

In this chapter, it is suggested that the techniques of forecast combination can be used successfully to partially alleviate the effects of structural changes on forecasting performance. In their pioneering work, Bates and Granger (1969) showed that if a number of unbiased forecasts of the same variable are available, then it is rarely (if ever) optimal to seek out the best of the competing forecasts and use it alone. Rather, the forecasts can always be combined in such a way that the composite forecast has (asymptotic) variance less than or equal to any of the competing forecasts; in that sense, all sources of information may prove valuable. Similar reductions in mean squared error may be achieved for (possibly) biased forecasts via the regression-based technique of Granger and Ramanathan (1984).

The basic concept of combining has been extended in various directions. Of most immediate concern in the present context are those efforts that are directed toward allowing the combining weights to be flexible over time. For our present purposes, we view the explicit modeling of nonconstancies in the combining weights as an attempt to compensate for the poor performance of the primary forecasts in situations of structural change of unknown form. In many situations, such an approach yields powerful increases in forecasting performance because the available primary forecasts do not adequately account for structural change. Furthermore, even if it is desired to model structural change explicitly in the primary forecasts, it is often difficult (or impossible) to locate and compensate for the changing structure, particularly in an ongoing forecasting organization where timely forecasts must be produced.

In Section 19.2, we review the basic theory of combining forecasts. In Section 19.3, we present alternative ways to model nonconstancy of weights within the class of regression-based combining methods, which include weighted least squares and various forms of varying-coefficient models. Those models are more general than, and include as special cases, time-varying variance-covariance methods. In Section 19.4, we outline testing procedures for various aspects of these models. The improvements in forecasting performance delivered by these methods are illustrated by a numerical example in Section 19.5. In Section 19.6, we summarize the major results and directions for future research.

19.2 The Basic Theory of Combining

In this section, we shall give a brief overview of the recent literature on the combination of forecasts; for a detailed exposition see Diebold and Pauly (1986,1987a).

19.2.1 Variance-covariance combining

Consider a set of m competing forecasts $f^1_{t|t-1}, \ldots, f^m_{t|t-1}$ of a variable y_t, made at time $t-1$, and examine linearly combined forecasts of the form

$$C_t = \beta_1 f^1_{t|t-1} + \beta_2 f^2_{t|t-1} + \ldots + (1 - \sum_{i=1}^{m-1} \beta_i) f^m_{t|t-1} \tag{19.1}$$

The one-step-ahead combined prediction error $e^C_t = y_t - C_t$ can be shown to satisfy the same equality

$$e^C_t = \beta_1 e^1_t + \beta_2 e^2_t + \ldots + (1 - \sum_{i=1}^{m-1} \beta_i) e^m_t \tag{19.2}$$

Thus,

$$Var(e^C_t) = \sum_{i=1}^{m} \beta_i^2 \sigma_i^2 + 2 \sum_{i=1}^{m} \sum_{\substack{j=1 \\ j \neq i}}^{m} \beta_i \beta_j \sigma_{ij} \tag{19.3}$$

Minimization of this expression leads to the optimal combining weight vector (Reid, 1969; or Granger and Newbold, 1974):

$$\beta^\star = (\Sigma^{-1}\mathbf{i})/(\mathbf{i}'\Sigma^{-1}\mathbf{i}) \tag{19.4}$$

where $\beta^\star$ is a $(m \times 1)$ vector, Σ is the variance-covariance matrix of the one-step-ahead forecast errors, and $\mathbf{i}$ is a conformable column vector of ones. We will refer to the calculation of the optimal weights in this fashion as the "variance-covariance" method. Note the intuitive results that

$$\left.\begin{array}{l} \lim_{\sigma_i^2 \to \infty} \beta_i^\star = 0 \\ \lim_{\sigma_j^2 \to \infty} \beta_i^\star = 1, \quad j = 1, \ldots, m \quad (j \neq i) \end{array}\right\} \text{ for each } i \tag{19.5}$$

Thus, the more reliable f^i, the more weight placed on it, and vice versa. Note also that the covariances σ_{ij} play an important role in determining the weights, and that the restriction that the combining weights sum to unity ensures that the combined forecast will be unbiased if the primary forecasts are unbiased. Under this assumption the minimum-variance combined forecast is also the minimum MSE combined forecast, which can, *ex post*, be no worse than the best individual forecast (Granger and Newbold, 1977). Alternatively, if all variances and covariances are known, the combined forecast can be no worse than the best primary forecast.

In practice we estimate $\beta^\star$ by replacing Σ with an estimate $\hat{\Sigma}$, where $\hat{\Sigma}_{ij} = \sum_{t=1}^{T} e_{it} e_{jt}$. Thus, the elements of Σ are viewed as fixed, but unknown, quantities to be estimated from the T sample observations. Even in a real-time forecasting environment, in which $\hat{\beta}_T$ is updated recursively and therefore changes as $T \to \infty$, the change is viewed not as structural, but rather represents the convergence in probability of $\hat{\beta}_T$ to β.

A number of authors have recognized that the true but unknown matrix Σ, and henced the vector β, may not be fixed over time. In such situations, the use of

$$\hat{\beta}_T = (\hat{\Sigma}^{-1}\mathbf{i})/(\mathbf{i}'\hat{\Sigma}^{-1}\mathbf{i}) \tag{19.6}$$

where

$$\hat{\Sigma}_{ij} = \frac{1}{T}\sum_{t=1}^{T} e_{it}e_{jt} \tag{19.7}$$

may be severely suboptimal.

Suboptimality of fixed-weight combinations occurs for many reasons. For example, differential learning speeds of different forecasting groups and/or forecasting techniques may lead to a particular forecast becoming progressively better over time, relative to others. Similarly, the design of various forecasting models may make them relatively better forecasting tools in some situations than in others. In such a situation, a truly optimal combining procedure should weight one or more forecasts progressively more heavily over time. Also, nonlinearity in the underlying economic structure leads directly to nonconstant forecast error variances and, hence, to the desirabilty of nonconstant combining weights, as argued by Greene *et al.* (1985). The major focus here is, however, on change in the macroeconomic environment, and certain forecasting techniques may be relatively more vulnerable to such change. Moreover, if many different types of structural change are simultaneously occuring, we would expect them to have differential effects on forecasts produced by different methods.

Most of the procedures that have been proposed to deal with the drift problems for the variance-covariance combining method are adaptive "real-time" algorithms for calculating the combining weights. These methods make use of a moving data subset (e.g., the V most recent observations) to calculate the weights. Thus,

$$\hat{\beta}_T = (\hat{\Sigma}_T^{-1}\mathbf{i})/(\mathbf{i}'\hat{\Sigma}_T^{-1}\mathbf{i}) \tag{19.8}$$

where

$$\hat{\Sigma}_{ij,T} = V^{-1}\sum_{t=T-V+1}^{T} e_{it}e_{jt} \tag{19.9}$$

This has the desirable properties of giving the most weight to those forecasts that have performed best in the recent past and allowing for the possibility of a nonstationary relationship over time between the primary forecasts. On the other hand, the choice of V is arbitrary, and its value will have substantial effects on the estimated combining weights. Furthermore, as noted by Bessler and Brandt (1981), most of these methods not only lead to convex combining weights (as opposed to weights that simply sum to unity), but also force each weight to lie in the interval $[0, 1/(m-1)]$, where m is the number of primary forecasts. This limitation is particularly severe if one primary forecast is substantially better than the others.

Granger and Newbold (1977) suggest the following possibilities, in addition to (19.8), which we list here for comparison with later results. Assuming that m forecasts are to be combined, we have

$$\hat{\beta}_{iT} = \Big(\sum_{t=T-V+1}^{T} e_{it}^2\Big)^{-1} \Big/ \Big[\sum_{j=1}^{m}\Big(\sum_{t=T-V+1}^{T} e_{jt}^2\Big)^{-1}\Big], \quad i = 1, \ldots, m \tag{19.10}$$

$$\hat{\beta}_{iT} = a\hat{\beta}_{i,T-1} + (1-a)\Big(\sum_{t=T-V+1}^{T} e_{it}^2\Big) \Big/ \Big[\sum_{j=1}^{m}\Big(\sum_{t=T-V+1}^{T} e_{jt}^2\Big)^{-1}\Big] \quad (0 < a < 1), \quad i = 1, \ldots, m \tag{19.11}$$

$$\hat{\beta}_T = (\hat{\Sigma}_T^{-1}\mathbf{i}) / (\mathbf{i}'\hat{\Sigma}_T^{-1}\mathbf{i}) \tag{19.12}$$

where the elements of $\hat{\Sigma}_T$ are: $\hat{\Sigma}_{ij,T} = (\sum_{t=1}^{T} \lambda^t e_{it} e_{jt}) / (\sum_{t=1}^{T} \lambda^t)$, $\lambda \geq 1$;

$$\hat{\beta}_{iT} = \Big(\sum_{t=1}^{T} \lambda^t e_{it}^2\Big)^{-1} \Big/ \Big[\sum_{j=1}^{m}\Big(\sum_{t=1}^{T} \lambda^t e_{jt}^2\Big)^{-1}\Big] \quad (\lambda \geq 1), \quad i = 1, \ldots, m \tag{19.13}$$

These formulae represent different ways of discounting past information, and treating covariances, when constructing combining weights. Clearly, (19.9) represents a moving sample approach using all variance and covariance information; ({eqdi.10) uses the same moving sample but ignores covariance information; (19.11) is an "adaptive" scheme which ignores covariance information; (19.12) uses the full sample but weights recent observations more heavily; and (19.13) is like (19.12) but ignores covariance information.

In a recent development, Engle *et al.* (1984) used the model of autoregressive conditional heteroscedasticity (ARCH), due to Engle (1982a), actually to model the evolution of prediction error variances and covariances over time. This approach makes use of the full sample to produce a sequence of time-varying weights in a rigorous and systematic fashion, rather than simply (and artifically) basing the weight calculations on a recent subset of observations. While this approach represents a notable contribution, it has problems of its own. First, it produces an extremely noisy weight sequence, as opposed to the smoothly changing weights argued for by Granger and Newbold (1974). Second, although their ARCH-combined forecast does improve upon the individual forecasts, it does not compare favorably with a fixed-weight combination. They note that this may be due to misspecification of the diagonal bivariate ARCH-model that they use, and that further research in this area is needed.

It should be noted that the Engle-Granger-Kraft approach requires the modeling of an entire conditional covariance matrix over time, which is a formidable task. Combination by a regression approach with time-varying parameters, on the other hand, may be more tractable since the evolution of only one parameter must be modeled. The regression estimator, while depending on all available variances and covariances, models their evolution implicitly rather than explicitly. We now consider such an approach in detail.

19.2.2 Regression-based combining

Granger and Ramanathan (1984) show that the above variance-covariance forecast combination theory has a regression interpretation, by estimating a linear regression model

$$\mathbf{y} = \mathbf{F}\beta + \mathbf{u} \tag{19.14}$$

with $\mathbf{y} = (y_1, y_2, \ldots, y_t)'$, $\mathbf{u} = (u_1, u_2, \ldots, u_t)'$, $\beta = (\beta_0, \beta_1, \ldots, \beta_m)'$, and $\mathbf{f}_t = (1, f^1_{t|t-1}, f^2_{t|t-1}, \ldots, f^m_{t|t-1})'$, where the $u_t \sim \text{iid}N(0, \sigma^2_u)$, and $\mathbf{F}$ is a matrix with tth row $\mathbf{f}'_t$, subject to linear constraints $\beta_0 = 0$, $\sum_{i=1}^m \beta_i = 1$.

The information set consists of actual realizations for the variable of interest up to period t, and of primary forecasts of this variable for the same sample period. The restriction can be expressed as $\mathbf{R}\beta = \mathbf{r}$, where

$$\mathbf{R} = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 1 & \ldots & 1 \end{pmatrix}, \qquad \mathbf{r} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Least squares estimation of (19.14), subject to the set of restrictions will generate a vector of weights

$$\beta^*_{RLS} = \hat{\beta} - (\mathbf{F}'\mathbf{F})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\beta} - \mathbf{r}) \tag{19.15}$$

where $\hat{\beta}$ is the unrestricted least squares estimate of the set of weights. The weights obtained from (19.4) are the same as the ones obtained from (19.15), and they lead to an unbiased combined forecast if the component forecasts are unbiased and the constraint is satisfied. It should be noted, however, that the variable being forecast must be stationary. Otherwise, an appropriate stationary-rendering transformation should be performed prior to analysis.

Generally, we will refer to the calculation of the optimal weights in this fashion as the "regression method". It is interesting to note that there exists also a well-defined Bayesian interpretation of these weights (Bunn, 1975; Bordley, 1982, 1986). Furthermore, a seemingly unrelated regressions estimator (SURE) for multivariate combining problems can easily be derived analogous to (19.15) (see, in a somewhat different context, Diebold and Pauly, 1987b).

Failure to impose the $\Sigma\beta_i = 1$ constraint in the regression method leads to a combined forecast that is biased unless $E(f^i_t) = y_t$, $\forall i$, and $\Sigma\hat{\beta}_i = 1$. The major virtue of (19.15) is that it leads to combining weights identical to those based on the variance-covariance method. Alternatively, one can ignore the constraints and obtain the unconstrained least squares combining vector

$$\hat{\beta}_{OLS} = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{y}$$

Any bias that may be present in the component forecasts may be eliminated by including an intercept, which amounts to adding the unconditional mean of $\mathbf{y}$ as a constituent forecast.

The unrestricted OLS predictor of y_{t+1}, given $\mathbf{f}_{t+1} = (1, f^1_{t+1}, \ldots, f^m_{t+1})'$ has a mean squared error

$$MSE(\hat{y}_{t+1}) = \sigma^2[\mathbf{f}'_{t+1}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{f}_{t+1} + 1]$$

while for the constrained predictor we obtain

$$MSE(y^\star_{t+1}) = \sigma^2\{\mathbf{f}'_{t+1}[(\mathbf{F}'\mathbf{F})^{-1} - (\mathbf{F}'\mathbf{F})^{-1}\mathbf{R}'(\mathbf{M} - \sigma^{-2}\mathbf{M}\delta\delta'\mathbf{M})\mathbf{R}(\mathbf{F}'\mathbf{F})^{-1}]\mathbf{f}_{t+1} + 1\}$$

where $\mathbf{M} = [\mathbf{R}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{R}']^{-1}$ and $\delta = \mathbf{R}\beta - \mathbf{r}$.

If $\mathbf{M} - \sigma^{-2}\mathbf{M}\delta\delta'\mathbf{M}$ is positive definite, $MSE(y^\star_{t+1}) < MSE(\hat{y}_{t+1})$ and the constrained combined predictor will be more efficient. This will be the case if $\mathbf{R}\beta = \mathbf{r}$, i.e., if the constraints are correct. If $\mathbf{R}\beta \neq \mathbf{r}$ or if the primary forecasts are not unbiased, unrestricted least squares combining is preferable. Clemen (1986) and Trenkler and Liski (1986) argue, however, that for certain values of β and σ^2, even with $\mathbf{R}\beta \neq \mathbf{r}$, the gain in efficiency from imposing the constraints in the presence of biased primary forecasts may offset the incurred bias. For the remainder of this chapter we shall deal only with unrestricted combining regressions.

Granger and Ramanathan (1984) note that, even if the primary forecast errors are white noise, there is no guarantee that the combined forecast errors will be white as well. The presence of serial correlation in the combined prediction error would make the estimates of combining weights inefficient and provide inconsistent standard errors. Furthermore, the least squares combined forecast would not be the best linear unbiased predictor. Diebold (1988) examines these issues and shows that in general the unrestricted regression-based combined prediction errors will be serially correlated if the variable to be forecast (y_t) is serially correlated or if one or more of the constituent forecast errors e_i, $i = 1, \ldots, m$, are serially correlated. He suggests that the latter is likely to be of minor importance, while the former may be of great importance. In fact, if y_t is *not* serially correlated, it is linearly unpredictable, and any attempt to combine different primary forecasts is useless as well.

In addition, Diebold (1988) shows that unless y_t and e_i $(i = 1, \ldots, m)$, strictly follow (possibly degenerate) finite moving average processes in deviations from their means, the combined prediction error will generally follow an ARMA (p, q) process, where both p and q are nonzero. Consequently, the linear combining model (19.14) will have to be generalized to

$$\mathbf{y} = \mathbf{F}\beta + \mathbf{u} \qquad \text{with } \Theta(L)u_t = \alpha(L)\varepsilon_t$$

where $\Theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \ldots - \theta_p L^p$ and $\alpha(L) = 1 + \alpha_1 L + \alpha_2 L^2 + \ldots + \alpha_q L^q$ are finite-order polynomials in the lag operator L.

Efficient estimation of the combining weights will therefore have to be based on a feasible generalized least squares (FGLS) procedure, i.e., the optimal weights are determined as

$$\hat{\beta}_{GLS} = (\mathbf{F}'\hat{\Sigma}^{-1}\mathbf{F})^{-1}\mathbf{F}'\hat{\Sigma}^{-1}\mathbf{y} \tag{19.16}$$

where $\hat{\Sigma}$ denotes a consistent first-stage estimate of the variance-covariance matrix. In practice, the structure of $\Sigma = \sigma_\varepsilon^2\Omega$ will be rather complex for most specifications of $\Theta(L)$ and $\alpha(L)$. Generally, however, an ARMA (1,1) will be sufficiently flexible. We then have $\Omega_{tt} = (1 + \alpha^2 + 2\alpha\theta)/(1 - \theta^2)$ and $\Omega_{ts} = [(\theta + \alpha)(1 + \theta\alpha)\theta^{|t-s|-1}]/(1 - \theta^2)$ for $t \neq s$. Estimation procedures for this specification are given in Ansley (1979) and Harvey and Phillips (1979).

19.3 Non-Constant Weights

19.3.1 Weighted least squares

The general success of time-varying weights constructed by the variance-covariance method should extend to weights produced by the regression method. The relaxation of the restriction that the weights sum to unity and the ability to handle biased forecasts are strong advantages of the regression approach, so that it is particularly desirable to explore the possibilities for time-varying weights in that framework. In this section we shall therefore explore the potential for an application of weighted least squares techniques.

Instead of choosing $\hat{\beta}$ to minimize $\mathbf{u}'\mathbf{u} = \sum_{t=1}^{T}(y_t - \sum_{i=0}^{m}\beta_i\mathbf{f}^i_{t|t-1})^2$ we instead chooose it to minimize the matrix-weighted average $\mathbf{u}'\mathbf{W}\mathbf{u}$, or

$$\sum_{t=1}^{T}\sum_{t'=1}^{T} w_{t,t'}u_t u_{t'} \tag{19.17}$$

where the $(T \times T)$ matrix of the quadratic form is given by $\mathbf{W} = [w_{t,t'}]$. For most applications it will be adequate to assume that the weighting matrix is diagonal, i.e., $\mathbf{W} = \text{diag}(w_{11}, \ldots, w_{T,T})$, which means that we minimize the weighted sum of squares

$$\sum_{t=1}^{T} w_{tt}(y_t - \sum_{i=0}^{m}\beta_i f^i_{t|t-1})^2$$

The least squares estimator is, of course,

$$\hat{\beta}_{WLS} = (\mathbf{F}'\mathbf{W}\mathbf{F})^{-1}\mathbf{F}'\mathbf{W}\mathbf{y} \tag{19.18}$$

Note that a "moving sample" estimator, analogous to the moving variance-covariance estimator in (19.9) earlier, emerges as a special case when $\mathbf{W} = \text{diag}(w_a, w_b)$ with $(T-V)$- and V-vectors $w_a = (0, \ldots, 0)'$ and $w_b = (1, \ldots, 1)'$, respectively.

A simple method for ensuring that the influence of past observations declines with their distance from the present is to specify

$$\mathbf{W} = \text{diag}(w_1, \ldots, w_T) \tag{19.19}$$

where $w_t \geq w_{t-1}$ for $t = 2, \ldots, T$.

There are, of course, insufficient degrees of freedom to maintain such generality, so that explicit parameterizations such as linearly or geometrically declining elements of the weighting matrix may prove extremely useful. In Diebold and Pauly (1987a), we discuss several such schemes; we summarize those results now, and we integrate them with the serially correlated disturbance structures discussed earlier.

Extracting a factor k ($k > 0$), a general nonlinearly declining weight specification, closely related to the Box-Cox transformation, is given by:

$$\mathbf{W} = k \begin{pmatrix} 1^{\lambda} & & 0 \\ & \ddots & \\ 0 & & T^{\lambda} \end{pmatrix} \tag{19.20}$$

or $\mathbf{W} = \text{diag}[w_{tt}] = [kt^{\lambda}]$, where $k, \lambda > 0$. Note that $dw_{tt}/dt = k\lambda t^{(\lambda-1)} > 0$, which guarantees that the recent past is weighted more heavily than the distant past. Furthermore,

$$\frac{d^2 w_{tt}}{dt^2} = k\lambda(1-\lambda)t^{(\lambda-2)} \quad \begin{cases} > 0 & \text{if } \lambda > 1 \\ < 0 & \text{if } \lambda < 1 \end{cases}$$

Thus, the sign of $(\lambda - 1)$ determines whether the weights decline at an increasing rate or at a decreasing rate (as we go further back into the past). A full Box-Cox transformation may also be undertaken by letting $\mathbf{W} = k\mathbf{W}^*(\lambda)$, where

$$w_{tt}^*(\lambda) = \begin{cases} (t^{\lambda} - 1)/\lambda & \text{if } 0 < \lambda \leq 1 \\ \ln t & \text{if } \lambda = 0 \end{cases}$$

However, with the Box-Cox weight structure, we can only obtain weights in the region bounded by linear and log-linear schemes, but others are excluded. The t^{λ} specification, on the other hand, appears quite attractive. First, note that constant weights emerge for $\lambda = 0$ and linear weights emerge for $\lambda = 1$. Furthermore, unlike a geometric specification, this specification can produce weights that decrease either at an increasing or a decreasing rate, which increases its potential usefulness and applicability. In *Figure 19.1*, we see the t^{λ} weights as λ ranges from 0 to 7. Without loss of generality, the weights are normalized by T^{λ}, so that $w_{TT} = 1$. Like geometric weights, t^{λ} weights are capable of dying out very quickly, for large λ.

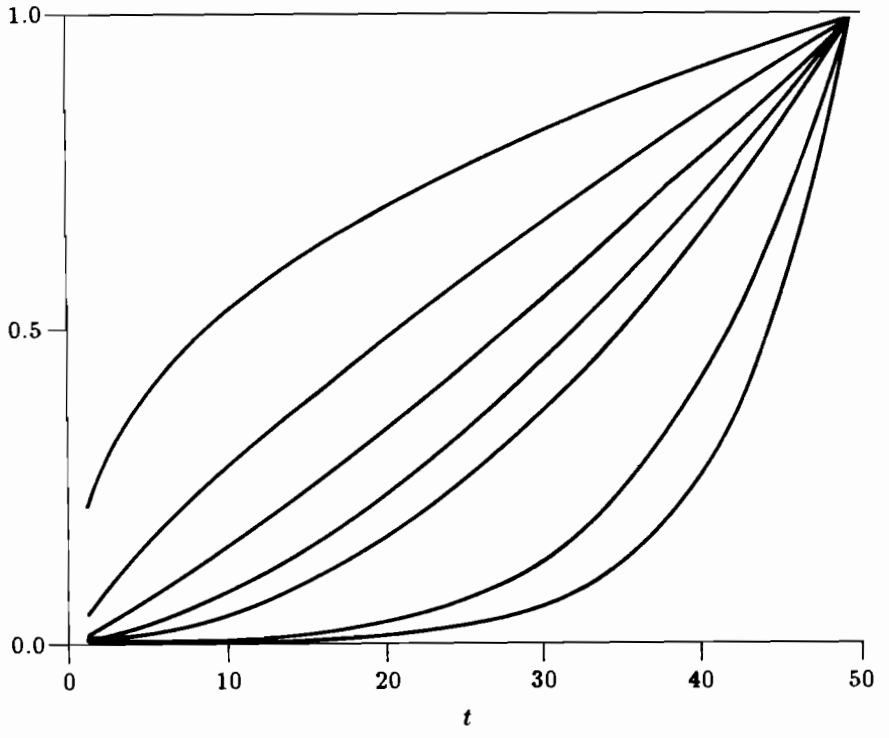

Figure 19.1: λ weight structures over time.

Finally, we need not pick $\mathbf{W}$ arbitrarily; rather, it too can be estimated. If $\mathbf{W} = k\mathbf{W}^*(\lambda)$, then we simply choose $\hat{\lambda}$ and $\hat{\beta}$ to solve

$$\min_{\lambda,\beta} \mathbf{u}'\mathbf{W}\mathbf{u}$$

or

$$\min_{\lambda,\beta} \sum_{t=1}^{T} w_{tt}^*(\lambda)(y_t - \sum_{i=0}^{m} \beta_i f_{t|t-1}^i)^2$$

Note that there is no need to choose k, since it cancels from the expression for the WLS estimator.

This analysis highlights the extreme restrictions imposed by the "moving sample" approach, since it restricts the weights on the V most recent observations to be constant

and equal to unity, while all other weights are restricted to be constant and equal to zero. The WLS approach, on the other hand, uses all the data and requires only that the weights be decreasing. Furthermore, these facts should lead the WLS method to produce a non-noisy sequence of combining weights.

We have already seen that the moving sample approach (19.9) to the variance-covariance method emerges as a special case of the WLS regression method for a particular $\mathbf{W}$ matrix. The WLS regression approach also sheds light on the "weighted" variance-covariance approach (19.12). In particular, it is equivalent to geometric WLS, with no intercept and subject to the restriction that $\sum \beta_i = 1$. It can be shown that the optimal weight resulting from (19.12) is identical to the expression for the restricted OLS estimator for data that have been transformed by $\sqrt{\lambda^t}$, i.e., it is our geometric WLS estimator.

The WLS regression approach also highlights the large amount of information that is lost in (19.10), (19.11), and (19.13) by ignoring covariance information, and the convenience of the WLS approach in terms of not having to explicitly compute all the elements of $\hat{\Sigma}$.

It is now straightforward to combine the WLS approach with a more general error specification that allows for the presence of serially correlated prediction erors. Analogous to (19.18), we can define the matrix $\Phi = \mathbf{W}\Omega^{-1}$. The FGLS estimator in the presence of parameterized weighting is then given as

$$\hat{\beta}^{W}_{GLS} = (\mathbf{F}'\hat{\Phi}\mathbf{F})^{-1}\mathbf{F}'\hat{\Phi}\mathbf{y} \tag{19.21}$$

Operationally, an iterative estimation procedure for (19.21) involves an initial WLS stage as in (19.18); based upon this initial estimate of $\mathbf{u}$, we can construct an estimate of Ω, which enables us to construct $\hat{\Phi}$. Alternatively, FGLS estimation of the combining equation on transformed variables can proceed as in (19.16) above, with $\mathbf{y}$ and $\mathbf{F}$ replaced by $\tilde{\mathbf{y}} = \mathbf{P}'\mathbf{y}$ and $\tilde{\mathbf{F}} = \mathbf{P}'\mathbf{F}$ to incorporate the weights ($\mathbf{P}'\mathbf{P} = \mathbf{W}$). A grid search over λ can be performed to obtain the global optimum.

19.3.2 Deterministic time-varying parameter models

While the WLS regression-based approach may offer substantial benefits relative to the moving sample variance-covariance approach, we may also want to consider a regression-based systematically time-varying parameter model, which makes use of the full sample. The simplest and straightforward member of this class has deterministically time-varying parameters. This gives the combining equation

$$\mathbf{y} = \mathbf{F}\beta + \mathbf{u} \tag{19.22}$$

where $\beta^i = P^i(t)$, $P^i(t) = p^i_0 + p^i_1 t + \ldots + p^i_r t^r$, $i = 0, \ldots, m$. Thus, the smoothly varying combining weights will be deterministic nonlinear (polynomial) functions of time.

If the evolution of the elements of β (due to the evolution of underlying forecast error variances and covariances) is well described by low-order deterministic time trends, then

exploitation of that fact may yield substantial increases in forecasting performance. The advantage of this approach relative to our earlier WLS regression-based approach is that it enables us to model explicitly any parameter evolution in the combining equation and to project that evolution when combining the forecasts. For example, consider the simple bivariate restricted combining equation

$$y_t - f_t^2 = \beta(f_t^1 - f_t^2)$$

Now, at time T, the following data will be available:

$$\{y_t\}_{t=1}^{T} \ , \ \{f_{t|t-1}^1\}_{t=1}^{T+1} \ , \ \{f_{t|t-1}^2\}_{t=1}^{T+1}$$

For constant parameter combination, $\hat{\beta}$ is obtained from the T-observation combining regression, and then the forecast of y_{T+1} is obtained as

$$\hat{y}_{T+1} = \hat{\beta} f_{T+1|T}^1 + (1 - \hat{\beta}) f_{T+1|T}^2 \tag{19.23}$$

With the (linearly) deterministically time-varying parameter model, on the other hand, the combining regression is

$$\begin{aligned} (y_t - f_{t|t-1}^2) &= (\beta_0 + \beta_1 t)(f_{t|t-1}^1 - f_{t|t-1}^2) \\ &= \beta_0(f_{t|t-1}^1 - f_{t|t-1}^2) + \beta_1 t(f_{t|t-1}^1 - f_{t|t-1}^2) \end{aligned} \tag{19.24}$$

The estimated parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ are then used to produce the forecast as

$$\hat{y}_{T+1|T} = [\hat{\beta}_0 + \hat{\beta}_1(T+1)] f_{T+1|T}^1 + [1 - \hat{\beta}_0 - \hat{\beta}_1(T+1)] f_{T+1|T}^2 \tag{19.25}$$

The extension to general polynomial trends and unrestricted regression-based combination is immediate.

For example, the unrestricted regression-based analog of the above example is

$$\begin{aligned} y_t &= (p_0^0 + p_1^0 t) + (p_0^1 + p_1^1 t) f_{t|t-1}^1 + (p_0^2 + p_1^2 t) f_{t|t-1}^2 \\ &= p_0^0 + p_1^0 t + p_0^1 f_{t|t-1}^1 + p_1^1 (t f_{t|t-1}^1) + p_0^2 f_{t|t-1}^2 + p_1^2 t f_{t|t-1}^2 \end{aligned} \tag{19.26}$$

After estimation of the parameters p_0^i and p_1^i $(i = 0, 1, 2)$, the predictor is obtained as

$$\hat{y}_{t+1|t} = [\hat{p}_0^0 + \hat{p}_1^0(T+1)] + [\hat{p}_0^1 + \hat{p}_1^1(T+1)] f_{T+1|T}^1 + [\hat{p}_0^2 + \hat{p}_1^2(T+1)] f_{T+1|T}^2 \tag{19.27}$$

In addition, while the use of time-varying parameters lessens the need to weight recent observations more heavily, it does not eliminate it. Thus, a WLS approach together with time-varying parameters may prove useful. Finally, if the combined prediction errors are serially correlated, the predictor (19.27) should be changed, as in Diebold (1988), to reflect that fact.

19.3.3 Stochastic time-varying parameter models

It may be more realistic to make the regression-based combining weights stochastic, rather than deterministic, functions of time. We now view the disturbance term in the standard combining regressions as arising from random coefficients, i.e.

$$\mathbf{y} = \mathbf{F}\beta = \sum_{i=0}^{m} \beta^i f_t^i \tag{19.28}$$

(where $f_t^0 = 1$ for all t) with $\beta_t^i = \overline{\beta}^i + \mu_t^i$, $E(\mu_t^i) = 0$, $Var(\mu_t^i) = \gamma^i$, for all $i = 0, \ldots, m$ and all t. This gives the (heteroscedastic) combining equation

$$y_t = \sum_{i=0}^{m} f_t^i(\overline{\beta}^i + \mu_t^i) = \sum_{i=0}^{m} \overline{\beta}^i f_t^i + v_t^i \tag{19.29}$$

where $v_t^i = \sum_{i=0}^{m} f_t^i \mu_t^i$. This model was studied by Hildreth and Houck (1968) and further refined by Crockett (1985). The model as stated represents purely random-coefficient variation, so it is inadequate for our purposes. However, making use of the results of Singh *et al.* (1976), we can produce a stochastic systematically varying-parameter model for the combining equation.

We retain $y_t = f_t\beta_t$ and write $\beta_t^i = g^i(t) + \mu_t^i$ where $g^i(t)$ is a function of time. Thus,

$$y_t = \sum_{i=0}^{m} [g^i(t) + \mu_t^i] f_t^i$$

Rewrite this as

$$y_t = \sum_{i=0}^{m} g^i(t) f_t^i + \omega_t$$

where $\omega_t = \sum_{i=0}^{m} f_t^i \mu_t^i$. Again, we assume that $E(\mu_t^i) = 0$ and $Var(\mu_t^i) = \gamma^i$. Thus, we have $E(\omega_t) = 0$ for all t and

$$\Omega = Cov(\omega) = \begin{pmatrix} \sum_{i=0}^{m} (f_1^i)^2 \gamma^i & & 0 \\ & \ddots & \\ 0 & & \sum_{i=0}^{m} (f_T^i)^2 \gamma^i \end{pmatrix} \tag{19.30}$$

We may estimate the stochastic systematically varying-parameter model quite simply by recalling that

$$y_t = \sum_{i=0}^{m} g^i(t) f_t^i + \omega_t \tag{19.31}$$

or $\mathbf{y} = \mathbf{X}\mathbf{g}+\omega$ where $\mathbf{x}_t = (1, t, \ldots, t^p, f_t^1, \ldots, t^p f_t^1, \ldots, f_t^m, \ldots, t^p f_t^m)'$ and $\mathbf{g} = (g_0^0, \ldots, g_p^0, \ldots, g_0^m, \ldots, g_p^m)'$. As usual, $\hat{\omega} = \mathbf{M}\mathbf{y}$, where $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Note that

$$\dot{\hat{\omega}} = \dot{\mathbf{M}}\dot{\mathbf{X}}\sqrt{\gamma} + \eta \equiv \mathbf{G}\sqrt{\gamma} + \eta \tag{19.32}$$

(where a "." indicates squaring all elements, and $\gamma = (\gamma^1, \ldots, \gamma^m)'$ is the vector of parameter disturbance variances). Thus, $\hat{\gamma}$ is immediately obtained as

$$\hat{\gamma} = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\dot{\hat{\omega}} \tag{19.33}$$

This enables us to obtain $\hat{\Omega}$ and then, finally,

$$\hat{\beta}_{GLS} = (\mathbf{X}'\hat{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\Omega}^{-1}\mathbf{y} \tag{19.34}$$

As discussed earlier, additional gains may be obtained by modeling serially correlated residuals and weighting the data.

19.4 Testing Procedures

In the preceding sections, we have shown that the standard regression-based model of combining can be generalized in various directions to increase its flexibility under structural instabilities. Naturally, systematic hypothesis-testing procedures can be employed to test these specifications against a given body of data. In general, test procedures based on likelihood ratio, Wald, and Lagrange multiplier (LM) principles are available. All three, of course, lead to asymptotically equivalent tests in the sense that they have the same asymptotic power characteristics. On practical grounds, the only difference between the test statistics is computational; for further discussion, the reader is referred to the surveys in Breusch and Pagan (1980), Pagan and Hall (1983), and Engle (1982b, 1985).

Based on the ease of estimation under the null hypothesis in the present context, we recommend the use of LM tests, which only require estimation of the model under the null. For all specification tests that arise here, LM tests can be computed. Moreover, in many instances the LM statistic can be computed as TR^2 from a regression of (a transformation of) the estimated residuals on their lagged values and possibly a vector of (transformations of) explanatory variables in the original combining regression. Specifically, in the present context one may wish to test for, among others,

1. Linear constraints on parameters, i.e., restricted versus unrestricted least squares (Engle, 1985, pp. 790-791).
2. Serial correlation (Godfrey, 1978).
3. Nonconstant variances (Breusch and Pagan, 1980, pp. 246-247).

All these tests are well developed and need not be discussed further. We shall, however, describe a test for the stochastic systematic parameter variation introduced above, which can be derived easily applying a result of Breusch and Pagan (1979) as modified by Koenker (1981).

Consider the heteroscedastic alternative

$$\Omega_{tt} = h(\mathbf{z}_t'\alpha)$$

where $\mathbf{z}_t = (1, [\mathbf{z}_t^*]')'$ and $\alpha = (\sigma^2, [\alpha^*]')'$ are $(m \times 1)$ vectors.

The m-dimensional null hypothesis is that $\alpha^* = 0$ or $\Omega_{tt} = \sigma^2$ for all t. Under the assumption of normal disturbances, the Lagrange multiplier test statistic does not depend on h and is given by

$$LM = \frac{\mathbf{q}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{q}}{2\hat{\sigma}^4} \tag{19.35}$$

where $\mathbf{q} = \hat{\omega} - \hat{\sigma}^2\mathbf{i}$, $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_T)'$ is a $T \times (m+1)$ matrix, and $\hat{\sigma}^2 = \hat{\omega}'\hat{\omega}/T$; $\mathbf{i}$ is a $(T \times 1)$ column vector of ones, $\hat{\omega}$ is the OLS residual vector, obtained under the null, and the "." operator squares all elements of a vector or matrix. Conveniently, the numerator of LM is equal to the explained sum of squares in a regression of $\hat{\omega}$ on $\mathbf{Z}$. Under the null, LM is asymptotically distributed as χ^2_m.

Koenker (1981) shows that the size and power of this test are extremely sensitive to the normality assumption, and he develops a robust LM test by replacing $2\hat{\sigma}^4$ with $\sum(\hat{\omega}_t^2 - \hat{\sigma}^2)^2)/T$. The reason for the robustness of the Koenker test to nonnormality is that, while both $2\hat{\sigma}^4$ and $\sum(\hat{\omega}_t^2 - \hat{\sigma}^2)^2)/T$ may be viewed as estimates of $Var(\omega_t^2)$, the former estimate is valid only under normality since only in that situation is $Var(\omega_t^2) = 2\sigma^4$. On the other hand, $\sum(\hat{\omega}_t^2 - \hat{\sigma}^2)^2)/T$ is consistent for $Var(\omega_t^2)$ under much more general conditions, which Koenker specifies. Furthermore, the modified LM test may be calculated as TR^2 in a regression of $\hat{\omega}$ on Z.

Thus, to implement the test, we proceed as follows. First, note that because in our case the functions $g^i(t)$ are time polynomials, we can write under the null

$$\begin{aligned} y_t &= \sum_{i=0}^{m} g^i(t) f_t^i + \omega_t \\ &= (g_0^0 + g_1^0 t + \ldots + g_p^0 t^p) + (g_0^1 + g_1^1 t + \ldots + g_p^1 t^p) f_t^1 + \ldots \\ &\qquad + (g_0^m + g_1^m t + \ldots + g_p^m t^p) f_t^m + \omega_t \end{aligned} \tag{19.36}$$

Thus y_t is regressed on an intercept, t, t^2, ..., t^p, f_t^1, tf_t^1, ..., $t^p f_t^1$, f_t^2, tf_t^2, ..., $t^p f_t^2$, ..., f_t^m, tf_t^m, ..., $t^p f_t^m$. The residuals from this regression are retained, and squared. Then, $\hat{\omega}$ is regressed on i.e., $(f_t^i)^2$, $i = 1, \ldots, m$. The LM statistic, given by the uncentered coefficient of determination from this regression multiplied by sample size, is then distributed as χ^2_m under the null.

19.5 Examples

In order to illustrate the results, four artificial data sets, each with different stochastic properties, were generated. Each of these data sets, which we will refer to as case 1 to case 4, respectively, consists of 80 observations on 3 variables, y_t, $f^{1i}_{t|t-1}$, and $f^{2i}_{t|t-1}$. The y_t variables are the same in each case (to aid in variance reduction) and are realizations of the stationary AR(1) process

$$(1 - 0.9L)(y_t - 20) = \varepsilon_t, \ , \ \varepsilon_t \sim iidN(0,1)$$

To obtain these, we set $y_0 = 0$ and generated 480 observations on the process; the last 80 observations were then used as our sample, to guarantee that the initial condition had absolutely no effect. In case i, $f^{1i}_{t|t-1}$ and $f^{2i}_{t|t-1}$ are two different forecasts of y_t, made at time $t-1$. For each case, f^{1i} and f^{2i} are generated as

$$f^{i}_{t|t-1} = 1 + e^{i}_{t|t-1} + y_t$$

Thus, each forecast (in each case) has a unit bias, and is equal to the true realized value y_t plus a one-step-ahead prediction error. It is the variance-covariance structures of e^{1i} and e^{2i} that are changed across the cases.

In case 1, the one-step-ahead prediction errors are uncorrelated and have constant, but different, variances throughout the sample ($\sigma_1 = 1$ and $\sigma_2 = 2$). Clearly, although we would expect f^{11} to receive more weight in the combination than f^{21}, we would not expect our time-varying coefficient methods to outperform the traditional constant coefficient methods here.

In case 2, we again have $\sigma_2 = 2$ throughout the sample, but now $\sigma_{1t} = 1, t = 1, \ldots, 50$, at which point it begins to grow linearly until it achieves a value of 5 by $t = 80$. The forecast errors are again independent throughout the entire sample. In such a situation, we would expect our time-varying methods to lead to substantial forecasting improvements.

Case 3 is identical to case 2 for $t = 1, \ldots, 50$, but σ_{1t} then grows linearly from $t = 50, \ldots, 65$, at which point it reaches its maximum of 5. It then retreats linearly back to 1 by $t = 80$. Again, we would expect our time-varying methods, particularly those with quickly decreasing weights and/or nonlinearly deterministically varying combining weights, to perform well.

In case 4, we explore the possibility of a changing covariance between the forecast errors. We hold $\sigma_1 = 3$ and $\sigma_2 = 4$ throughout the entire sample, while the covariance is held at 0 for $t = 1, \ldots, 50$, but grows linearly to 11.7 by $t = 80$.

The details of the generation of variances and covariances are described in Diebold and Pauly (1987a). For each case the following methods were used to produce time-varying weights:

M1 WLS, t^{λ} weights

Table 19.1: MSPE results with associated optimal λ

Method	*Case 1*	*Case 2*	*Case 3*	*Case 4*
(M1) WLS, t^λ				
OLS	0.638 (9.0)	1.595 (23.0)	1.461 (24.5)	1.654 (26.0)
AR(1)	0.542 (6.0)	0.801 (4.0)	0.862 (6.5)	1.230 (5.5)
(M2) t^λ, lin				
OLS	0.700 (3.5)	1.453 (12.0)	1.290 (14.0)	1.305 (17.0)
AR(1)	0.610 (3.0)	0.861 (2.0)	0.903 (4.5)	1.140 (3.0)
(M3) t^λ, qd				
OLS	0.828 (1.0)	1.497 (6.0)	1.347 (7.0)	1.244 (10.0)
AR(1)	0.654 (2.0)	0.941 (1.0)	0.919 (3.0)	1.060 (1.5)
(M4) OLS				
OLS	0.658	3.336	3.374	3.009
AR(1)	0.476	0.734	0.988	2.100
(M5) Var-Cov	2.130	8.010	7.240	10.690
f^1 alone	2.018	15.980	12.625	9.193
f^2 alone	7.569	7.569	7.569	24.475

M2 WLS, t^λ weights, linear deterministic time-varying parameters

M3 WLS, t^λ weights, quadratic deterministic time-varying parameters

M4 OLS (simple unrestricted regression-based combination)

M5 Restricted OLS (variance-covariance combination).

We begin the exercise in period 50, in which $\{y_t\}_{t=1}^{50}$, $\{f_{t|t-1}^{1i}\}_{t=1}^{51}$, and $\{f_{t|t-1}^{2i}\}_{t=1}^{51}$ are available, $i = 1, \ldots, 4$. These 50 observations of y are regressed on the first 50 observations of the two forecasts, and the combined forecast $\hat{y}_{51}$ is obtained as

$$\hat{y}_{51}^i = \hat{\beta}_0 + \hat{\beta}_1 f_{51|50}^{1i} + \hat{\beta}_2 f_{51|50}^{2i}, \quad i = 1, \ldots, 4$$

This process is then repeated recursively until the entire sample is exhausted. The end result, then, is four sets (corresponding to $i = 1, \ldots, 4$) of five forecasts (corresponding to $M1, \ldots, M5$). The mean squared one-step-ahead prediction errors of these forecasts (for optimal λ, calculated by a grid search) are given in *Table 19.1.* For methods (M1) to (M3), we report standard least squares results and estimates based on a model with AR(1) errors.

Some general characteristics of the results are at once apparent. First, the standard (i.e., unrestricted OLS) regression-based combined forecast absolutely dominates the primary forecasts f^{1i} and f^{2i} (as well as the restricted variance-covariance combination), cutting the MSPE by approximately 60%. In addition, our time-varying combination procedures lead to substantial further reductions in MSPE in cases 2–4.

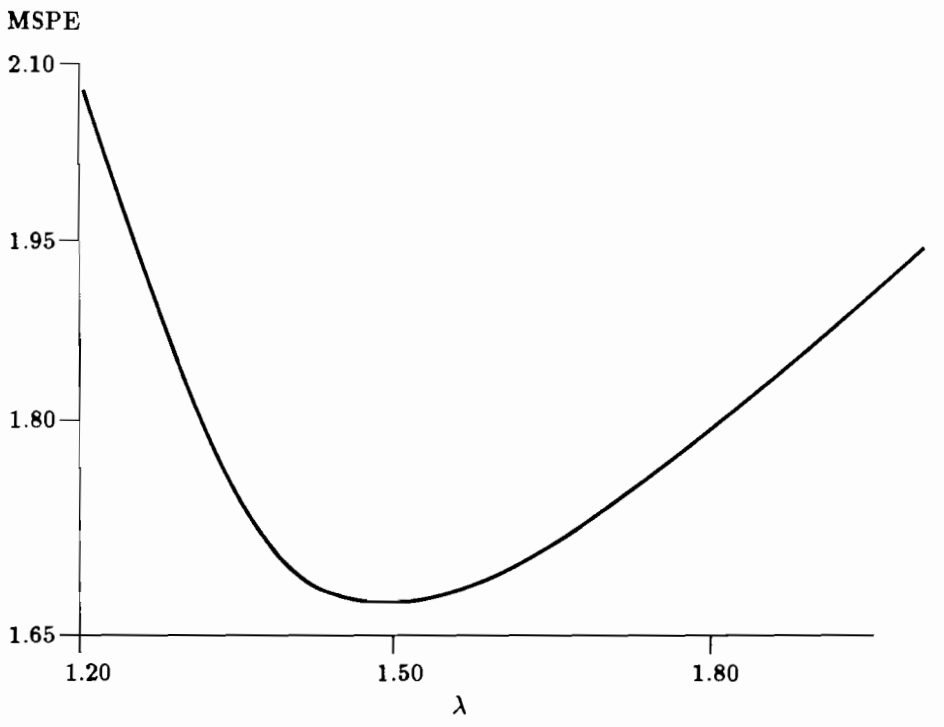

Figure 19.2: Case 2, Method M1: MSPE as a function of λ.

Recall that in case 1, in addition to the usual bias of 1.0 for both f^{11} and f^{21}, the forecasts are uncorrelated and $\sigma_1^2 = 1$, $\sigma_2^2 = 4$, for all t. The MSPE of f^{11} is 2.018, which is very close to the variance plus squared bias, while the MSPE of f^{21} is 7.569, which is somewhat above the expected MSPE of 5. The equally weighted combined forecast has MSPE of 0.658, which is less than 9% of the MSPE of f^{11} and less than 33% of the MSPE of f^{21}. Furthermore, the WLS method is not helpful, which is what we expected for case 1.

Moving to case 2, we see that the MSPE of the first forecast is 15.980 while that of the second forecast is 7.569, and the MSPE of the unrestricted regression-based combined forecast is a greatly reduced to 3.336. More importantly, the time-varying combining methods enable *further* reductions in MSPE of approximately 50%. Consider first WLS with t^{λ} weights. The optimal λ was found to be 23 and led to a combined MSPE of 1.595. This large value of λ implies quickly declining combining weights, which are needed to capture the quickly changing prediction error variance of f^{12}.

When we allow for linear deterministic time-varying combining weights in addition to the geometric WLS scheme (M2), the MSPE drops to 1.453, and λ^* drops to 12. The drop in MSPE is expected since the use of linearly time-varying combining weights enables us

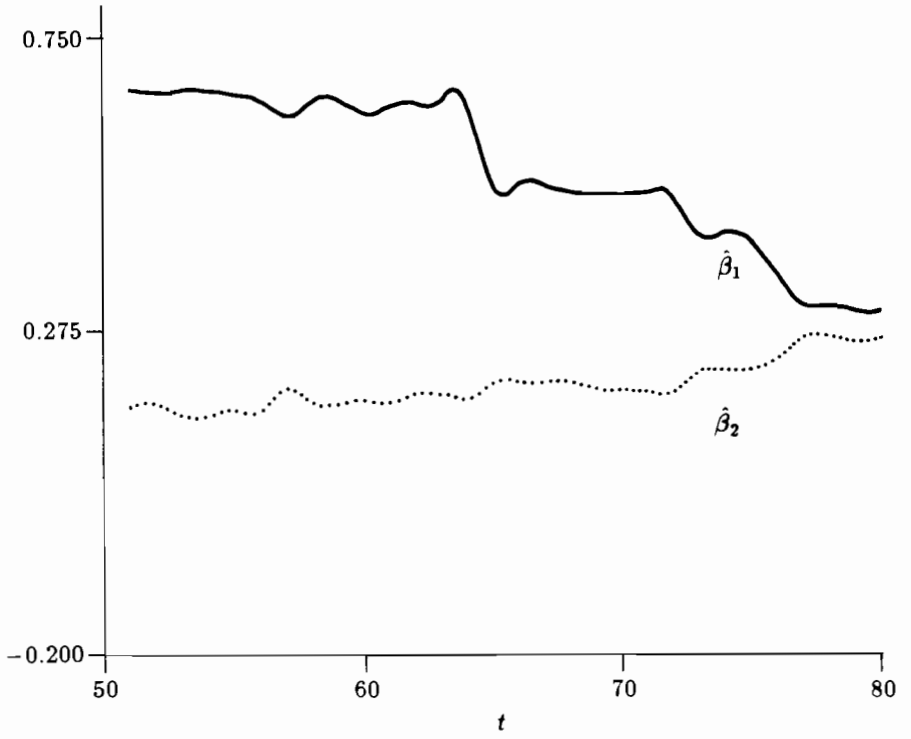

Figure 19.3: Case 2, Method M1: Weights $\hat{\beta}_1$ and $\hat{\beta}_2$ as functions of time.

to model and forecast the structural change which is occurring. The drop in λ^* is also to be expected, because once we *model* the evolution of the combining weights there is less need to heavily discount the past. Allowing for quadratic time-varying weights (M3) yields a slightly higher MSPE, which is explained by the estimation inefficiency incurred by including the unnecessary quadratic term. (Recall that σ_{1t}^2 is simply growing linearly over time.)

The results for cases 3 and 4 are very similar; a substantial decrease in MSPE is obtained through the use of time-varying coefficient methods. Whenever applied, a first-order autoregessive correction reduces the MSPE appreciably; the estimates of ρ are generally in the range of 0.6 to 0.8 and highly significant.

Figures 19.2, *19.3*, and *19.4* illustrate, respectively, for one of the examples (case 2, method 1), the dependence of MSPE on λ, the evolution of the weights, and the improvement in forecast performance in the combined forecast.

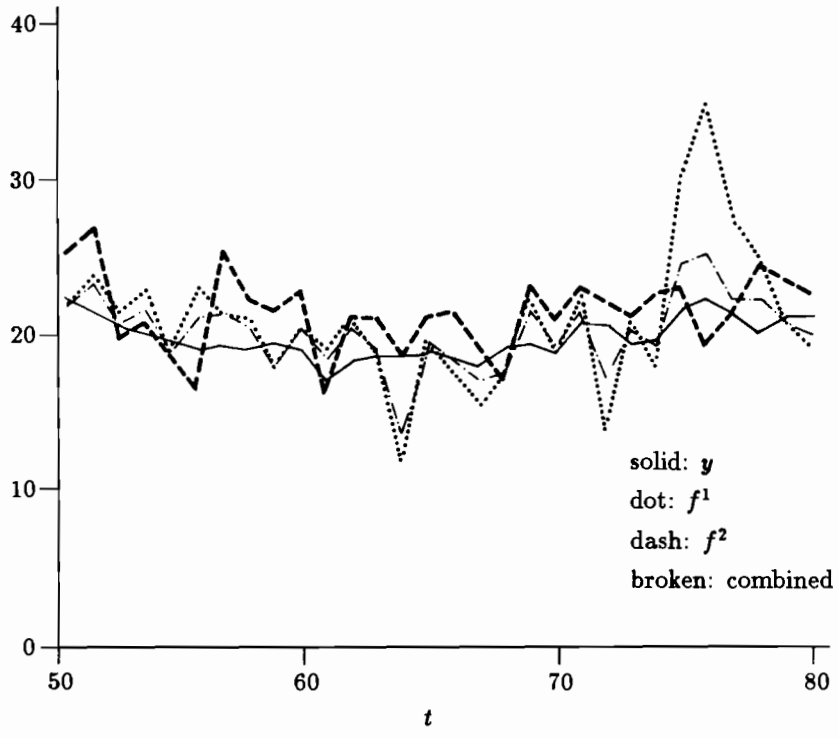

Figure 19.4: Case 2, Method M1: y, f^1, f^2, and combined forecasts as functions of time.

19.6 Conclusions and Directions of Future Research

We have developed and illustrated the potential usefulness of regression based WLS and time-varying parameter methods of forecast combination in the context of structural changes in the primary forecasts. It was shown that all of the earlier methods of forecast combination, based on the variance-covariance approach, emerge as special cases of the WLS regression approach, and that the suppression of explicit modeling of variances and covariances, facilitated by our approach, is particulary useful. In effect, our time-varying parameter models replace the explicit modeling of the evolution of variances and covariances, as in Engle *et al.* (1984), with the simpler modeling of time-varying regression parameters, for which a well-developed theory is available. In the example we presented, our combined forecasts had MSPE of as little as 10% of that of the worst primary forecast, and 40% of the unrestricted OLS regression-based combined forecast.

The research is currently proceeding in a number of directions. First, we are considering other systematically time-varying parameter models, such as the random walk parameter model, which can be conveniently estimated using the Kalman filter. The Kalman filter

approach also facilitates real-time parameter "updating" and can readily handle both stationary (e.g., ARMA) and nonstationary (e.g., integrated ARMA) parameter drift.

Second, nonlinear combining equations may lead to large decreases in MSPE, particulary if a number of linear forecasts are being combined, but the true (and unknown) process is nonlinear. An obvious possibility is to view the standard combining equation as a first-order Taylor expansion, and therefore to include higher-order terms. More generally, if we view forecast combination as a "production" process with the primary forecasts as "inputs", the use of flexible functional forms may prove worthwhile.

Finally, while combined weights obtained by our methods enable quick adaptation to structural change, they may be unduly influenced by outliers, so that robust estimation methods, such as least absolute deviations or m-estimation, may prove useful for the combining equation. Recent experience with the combination of real macroeconomic time series (see Kang, 1986; or Clemen and Winkler, 1986) indicates also that instability in the combining weights may be caused by severe multicollinearity of the primary forecasts due to overlapping information sets. In Diebold and Pauly (1987b), we explore these issues by using Bayesian shrinkage techniques to incorporate prior information into the estimation of combining weights.

References

Ansley, C. (1979), An algorithm for the exact likelihood of a mixed autoregressive-moving average process. *Biometrika*, **66**, 59-65.

Bates, J.M. and Granger, C.W.J. (1969), The Combination of Forecasts. *Operations Research Quarterly*, **20**, 451-468.

Bessler, D.A. and Brandt, J.A. (1981), Forecasting livestock prices with individual and composite methods. *Applied Economics*, **13**, 513-522.

Bordley, R.F. (1982), The combination of forecasts: A Bayesian approach. *Journal of the Operations Research Society*, **33**, 171-174.

Bordley, R.F. (1986), Linear combination of forecasts with an intercept: A Bayesian approach. *Journal of Forecasting*, **5**, 243-249.

Breusch, T.S. and Pagan, A.R. (1979), A simple test for heteroskedasticity and random coefficient variation. *Econometrica*, **47**, 1287-1294.

Breusch, T.S. and Pagan, A.R. (1980), The Lagrange multiplier test and its applications to model specification in econometrics. *Review of Economic Studies*, **47**, 239-253.

Bunn, D.W. (1975), A Bayesian approach to the linear combination of forecasts. *Operations Research Quarterly*, **26**, 325-329.

Clemen, R.T. (1986), Linear constraints and the efficiency of combined forecasts. *Journal of Forecasting*, **5**, 31-38.

Clemen, R.T. and Winkler, R.L. (1986), Combining economic forecasts. *Journal of Business and Economic Statistics*, **4**, 39-46.

Crockett, P.W. (1985), Asymptotic distribution of the Hildreth-Houck estimator. *Journal of the American Statistical Association*, **80**, 202-204.

Diebold, F.X. (1988), Serial correlation and the combination of forecasts. *Journal of Business and Economic Statistics*, **6**, 105-112.

Diebold, F.X. and Pauly, P. (1986), The combination of forecasts. *Prévision et Analyse Economique*, **7**, 7-31.

Diebold, F.X. and Pauly, P. (1987a), Structural change and the combination of forecasts. *Journal of Forecasting*, **6**, 21-40.

Diebold, F.X. and Pauly, P. (1987b), *The Use of Prior Information in Forecast Combination*, Special Studies Paper 218, Washington, DC: Board of Governors of the Federal Reserve System.

Engle, R.F. (1982a), Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica*, **50**, 987-1008.

Engle, R.F. (1982b), A general approach to Lagrange multiplier model diagnostics. *Journal of Econometrics*, **20**, 83-104.

Engle, R.F. (1985), Wald, likelihood ratio, and Lagrange multiplier tests in econometrics, in: Z. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics*, Vol. 2. Amsterdam: North-Holland.

Engle, R.F., Granger, C.W.J., and Kraft, D.F. (1984), Combining competing forecasts of inflation using a bivariate ARCH model. *Journal of Economic Dynamics and Control*, **8**, 151-165.

Godfrey, L.G. (1978), Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica*, **46**, 1293-1302.

Granger, C.W.J. and Morris, M.J. (1976), Time series modeling and interpretation. *Journal of the Royal Statistical Society*, **A-139**, 246-257.

Granger, C.W.J. and Newbold, P. (1974), Experience with forecasting univariate time series and the combination of forecasts. *Journal of the Royal Statistical Society Series*, **A-137**, 131-164.

Granger, C.W.J. and Newbold, P. (1977), *Forecasting Economic Time Series*. New York: Academic Press.

Granger, C.W.J. and Ramanathan, R. (1984), Improved methods of combining forecasts. *Journal of Forecasting*, **3**, 197-204.

Greene, M.N., Howrey, E.P., and Hymans, S.H. (1985), The use of outside information in econometric forecasting, in E. Kuh and D.A. Belsley (eds.), *Model Reliability*. Cambridge, MA: MIT Press.

Harvey, A.C. and Phillips, G.D.A. (1979), Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. *Biometrika*, **66**, 49-58.

Hildreth, C. and Houck, J.P. (1968), Some estimators for a linear model with random coefficients. *Journal of the American Statistical Association*, **63**, 584-595.

Kang, H. (1986), Unstable weights in the combination of forecasts. *Management Science*, **32**, 683-695.

Koenker, R. (1981), A note on studentizing a test for heteroskedasticity. *Journal of Econometrics*, **17**, 107-112.

Makridakis, S. *et al.*, (1984), *The Forecasting Accuracy of Major Time Series Methods*. New York: John Wiley.

Pagan, A.R. and Hall, A.D. (1983), Diagnostic tests as residual analysis. *Econometric Review*, **2** (2), 159-218.

Reid, D.J. (1969), A comparative study of time-series prediction techniques on economic data. Ph.D. thesis, Department of Mathematics, University of Nottingham, UK.

Singh, B., Nagar, A.L., Choudhry, N.K., and Raj, B. (1976), On the estimation of structural change: a generalization of the random coefficients regression model. *International Economic Review*, **17**, 340-361.

Trenkler, G. and Liski, E.P. (1986), Linear constraints and the efficiency of combined forecasts: A note. *Journal of Forecasting*, **5**, 197-202.

CHAPTER 20

Updating Parameters of Linear Change Point Models

Jürgen Kleffe

Summary

A method by Gill, Golub, Murray, and Saunders for updating matrix factorizations is used to improve the computations necessary for detecting a change point of the regression line in linear models. The method shows considerably higher speed and better numerical stability than using standard routines for linear regression. It is based on updating the residual sum of squares and the least squares estimators for the regression parameters if one regression equation gets added to or deleted from the model.

20.1 Introduction

We consider a linearchange point model

$$\begin{aligned} y_t &= x_t' b_1 + e_t \ , \quad t = 1, \ldots, m \\ y_t &= x_t' b_2 + e_t \ , \quad t = m+1, \ldots, n \end{aligned}$$

where y_t are independently distributed observations, x_t are given q-vectors of regressors, and e_t are *iid* error variables. The unknown parameters are b_1, b_2 and m, the change point index. This index is usually identified by minimizing the residual sum of squares

$$SSR(m) = SSR_1(m) + SSR_2(m) \tag{20.1}$$

subject to m, where

$$SSR_1(m) = \sum_{t=1}^{m} \|y_t - x_t' b_1(m)\|^2$$

$$SSR_2(m) = \sum_{t=m+1}^{n} \|y_t - x_t' b_2(m)\|^2$$

and $b_1(m)$ and $b_2(m)$ are least squares estimators based on the first m and on the remaining $n - m$ observations, respectively.

It is now common practice to compute all the above terms for each individual m separately, using standard packages for least squares estimation. We propose a method of computing both $SSR_1(m)$ and $SSR_2(m)$ in a more efficient manner, based on updating matrix factorizations. This method provides the entire series of values $SSR_1(1), \ldots, SSR_1(n)$ by the time the usual manner may take to get $b_1(n)$, only. Some linear change point models require the computation of $b_1(m)$ and $b_2(m)$ as least squares estimators that satisfy given linear constraints, often also depending on m. An extension to such problems is indicated.

20.2 Updating $SSR_1(m)$

Let $\mathbf{X}$ be the matrix formed by the rows x_t', $t = 1, \ldots, m$ and $\mathbf{Y} = (y_1, \ldots, y_m)'$. Then $b_1(m)$ satisfies the normal equation

$$\mathbf{X}'\mathbf{X} b_1(m) = \mathbf{X}'\mathbf{Y}$$

But for numerical calculations, this is known to be about the worst way of finding $b_1(m)$. Let T be an orthogonal transformation such that

$$T: \quad (\mathbf{X} \vdots \mathbf{Y}) \to \mathbf{R} = (\mathbf{R}_1 \vdots \mathbf{R}_2)$$

and $\mathbf{R}$ is a $m \times (q+1)$ upper triangular matrix. Some of the leading elements $r_{ii} \geq 0$ may vanish. But, by definition, let us agree

$$r_{ii} = 0 \quad \text{iff} \quad r_{ij} = 0, \quad j = 1, \ldots, q+1 \tag{20.2}$$

Such triangular factorization is unique, i.e., if $\mathbf{R}'\mathbf{R} = \mathbf{R}_\star'\mathbf{R}_\star$ for any two factorizations $\mathbf{R}$ and $\mathbf{R}_\star$, then $\mathbf{R} = \mathbf{R}_\star$. Further, orthogonality of T implies

$$\|\mathbf{X}b - \mathbf{Y}\|^2 = \|\mathbf{R}_1 b - \mathbf{R}_2\|^2$$

for any b, so that

$$SS\,R_1(m) = \min_b \|\mathbf{X}b - \mathbf{Y}\|^2 = r^2_{q+1,q+1}$$

Hence, computation of $SSR_1(m)$ does not require computation of $b_1(m)$. But, if wanted, $b_1(m)$ follows by solving the triangular system of equations

$$\mathbf{R}_1 b_1(m) = \mathbf{R}_2$$

where we ignore the $(q+1)$th equation, which is generally inconsistent with the first q equations.

It is now interesting to know how $\mathbf{R}$ changes as m increases by one. This problem is one of finding the triangular factorization $\mathbf{R}_\star$, say, belonging to

$$\begin{pmatrix} \mathbf{X} & \mathbf{Y} \\ x'_{m+1} & y_{m+1} \end{pmatrix}$$

But as $\mathbf{R}$ is already the result of an orthogonal transformation of $(\mathbf{X}\vdots\mathbf{Y})$, it remains to transform

$$\mathbf{D} = \begin{pmatrix} \mathbf{R}_1 & \mathbf{R}_2 \\ x'_{m+1} & y_{m+1} \end{pmatrix}$$

for which the following series of elementary orthogonal transformations has been proposed by Gill *et al.* (1974).

Denote the individual rows of $\mathbf{D}$ by $d'_1, \ldots, d'_{m+1}$. The first transformation annuls $x_{m+1,1}$ (the first component of x'_{m+1}), if necessary, supposing that $r_{11} \neq 0$. Otherwise, we make $d'_{m+1} = (x'_{m+1}, y_{m+1})$ the first row of $\mathbf{D}$. In general, we form

$$c = r_{11}/e\ ; \quad d = x_{m+1,1}/e\ ; \quad e = \sqrt{x^2_{m+1,1} + r^2_{11}}$$

and replace d'_1 and d'_{m+1} by new lines

$$d'_1 \leftarrow c d'_1 + d d'_{m+1}\ ; \quad d'_{m+1} \leftarrow c d'_{m+1} + d d'_1$$

The second transformation acts on the changed matrix $\mathbf{D}$ and annuls $x_{m+1,2}$, if necessary, using the second line d'_2. This process goes on until the complete triangular structure is obtained. The final matrix is denoted by $\mathbf{R}_\star$ and

$$SS\,R_1(m+1) = (r^\star_{q+1,q+1})^2$$

where $r^{\star}_{q+1,q+1}$ denotes the element in the $(q+1)$th row and $(q+1)$th column of $\mathbf{R}_{\star}$.

Note 1: Zero lines in $\mathbf{R}$ or $\mathbf{D}$ have no effect on the computations so that the computations need space for a $(q+1)\times(q+1)$ triangular matrix $\mathbf{R}$ and a $(q+1)$-vector, only.

Note 2: This method of updating $SSR_1(m)$ or $b_1(m)$ does not rely on any assumption made for $\mathbf{X}$, $\mathbf{Y}$, and the row (x'_{m+1}, y_{m+1}). We can actually start with just one regression equation incorporating each new line in turn. The initial value for $\mathbf{R}$ is (x'_1, y_1).

Note 3: It is obvious that the entire series of values $SSR_i(j)$, $j=1,\ldots,m$ is a by-product of factorizing $\mathbf{R}$.

Note 4: Several familiar updating techniques are based on the result

$$(\mathbf{H}+x_{m+1}x'_{m+1})^{-} = \mathbf{H}^{-} - a\mathbf{H}^{-}x_{m+1}x'_{m+1}\mathbf{H}^{-} \tag{20.3}$$

where $\mathbf{H}=\mathbf{X}'\mathbf{X}$ and $a = 1 + x'_{m+1}\mathbf{H}^{-}x_{m+1}$. Equation (20.3) holds if $x_{m+1}\in\mathcal{R}(\mathbf{X}')$, the column space of $\mathbf{X}'$. There are also many results on updating a least squares g-inverse of $\mathbf{X}$. But calculating along such formulae is generally more complex and involves a good deal of instability. These formulae are all in two parts, one for the case $x_{m+1}\in\mathcal{R}(\mathbf{X}')$ and one for $x_{m+1}\notin\mathcal{R}(\mathbf{X}')$. But all numerical criteria to distinguish these two cases are also based on matrix factorizations.

Note 5: $SSR_1(m)$ is not always a continuous function of $\mathbf{X}$. For instance the two regression equations

$$b_1 + b_2 = 1\ ;\quad b_1 + (1+w)b_2 = 2 \tag{20.4}$$

imply $SSR_1(2)=0$ for $w>0$, but $SSR_1(2)=1/2$ if $w=0$. Such situations occur in the proposed algorithm when we arrive at a small diagonal element $r_{ii}=w$. Replacing this by zero contradicts (20.2). It is then necessary to introduce a zero line in $\mathbf{R}$ and absorb the ith line into the lower part of $\mathbf{R}$. This results in a discontinuous increase of $SSR_1(m)$. For the regression equations (20.4) the triangular matrix $\mathbf{R}$ is

$$\mathbf{R} = \sqrt{1/2}\begin{pmatrix} 2 & 2 & 3 \\ 0 & w & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

if $w>0$, but

$$\mathbf{R} = \sqrt{1/2}\begin{pmatrix} 2 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

if $w=0$. We may avoid such effects by annulating each diagonal element of $\mathbf{R}$ that is less than a prespecified small value. In this way, we reduce the rank of $\mathbf{X}'\mathbf{X}$ when one of its eigenvalues is close to zero. In other words, we omit estimation of only purely estimable

parameters in favor of estimating others. This stabilizes the updating procedure. It is easily seen that each diagonal element r_{ii} of $\mathbf{R}$ grows monotonically by absorbing further regression equations — thus providing a measure of information about the ith regression parameter.

20.3 Updating $SSR_2(m)$

Updating $SSR_2(m)$ requires an operation inverse to what was described in Section 20.2. Let $\mathbf{X}$ be the matrix formed by the rows $x'_{m+1}, \ldots, x'_n$, $\mathbf{Y} = (y_{m+1}, \ldots, y_n)'$, and assume knowledge of an upper triangular matrix $\mathbf{R}$ such that

$$T: \quad \begin{pmatrix} x'_m & y_m \\ \mathbf{X} & \mathbf{Y} \end{pmatrix} \to \mathbf{R} = (\mathbf{R}_1 \vdots \mathbf{R}_2)$$

for some orthogonal transformtion T. We are interested in deriving an upper triangular matrix $\mathbf{R}_\star = (\mathbf{R}_1^\star \vdots \mathbf{R}_2^\star)$, which differs from $(\mathbf{X} \vdots \mathbf{Y})$ by an orthogonal transformation. Gill *et al.* (1974) suggested the following method.

Step 1: Find some solution p to

$$\mathbf{R}'p = (x'_m, y_m)' \; ; \quad p'p \leq 1$$

The above system of equations is consistent, and

$$(\mathbf{X} \vdots \mathbf{Y})'(\mathbf{X} \vdots \mathbf{Y}) = \mathbf{R}'(\mathbf{I} - pp')\mathbf{R} \geq 0$$

implies $p'p \leq 1$ for every solution $p \in \mathcal{R}(\mathbf{R})$. But every single component p_i of p is unique if $r_{ii} > 0$ and arbitrary if $r_{ii} = 0$. Therefore, a solution $p \in \mathcal{R}(\mathbf{R})$ is obtained by choosing $p_i = 0$ if $r_{ii} = 0$. The vector p has at most $q + 1$ nonzero elements.

Step 2: Let $p_0^2 = 1 - p'p$,

$$\mathbf{F} = \begin{pmatrix} p & \mathbf{R} \\ p_0 & 0 \end{pmatrix}$$

and denote by $f'_1, \ldots, f'_s$ the rows of $\mathbf{F}$. Now use elementary orthogonal transformations in order to annul the nonzero elements p_i in the first column of $\mathbf{F}$, but maintain the triangular form of the submatrix $\mathbf{R}$. These transformations introduce a 1 in the position of p_0 and certain nonzero elements in the last row of $\mathbf{F}$. We start with line f'_{q+1} by forming

$$c = p_{q+1}/e \; , \quad d = p_0/e \; , \quad e = \sqrt{p_{q+1}^2 + p_0^2}$$

and replace f'_{q+1} and f'_s by

$$f'_{q+1} \leftarrow df'_{q+1} - cf'_s \;, \quad f'_s \leftarrow cf'_{q+1} + df'_s$$

The new matrix $\mathbf{F}$ owns a zero in the position of p_{q+1}, and we continue annulling p_q in the same manner, based on lines f'_q and f'_s. If this process ends with a matrix

$$\mathbf{F}_\star = \begin{pmatrix} 0 & \mathbf{R}_\star \\ 1 & r'_\star \end{pmatrix}$$

say, the relation $\mathbf{F}'_\star \mathbf{F}_\star = \mathbf{F}'\mathbf{F}$ yields $r'_\star = (x'_m, y_m)$ and

$$\mathbf{R}'\mathbf{R} = \mathbf{R}'_\star \mathbf{R}_\star + \begin{pmatrix} x_m \\ y_m \end{pmatrix} (x'_m, y_m)$$

Consequently, $\mathbf{R}'_\star \mathbf{R}_\star = (\mathbf{X}\vdots\mathbf{Y})'(\mathbf{X}\vdots\mathbf{Y})$ and $\mathbf{R}_\star$ is the wanted triangular factorization since it is unique.

The effort required for deleting one row is about twice of that for adding one row. Also, the condition $p_0 \geq 0$ may cause this method to fail by rounding errors. But stability of computation improves if we avoid small diagonal elements of $\mathbf{R}$. They may lead to very large values for p_i, whereas p_i has to be zero if $r_{ii} = 0$. Therefore, when arriving at a small r_{ii}, we suggest replacing it by zero and modifying $\mathbf{R}$ accordingly. This procedure can be justified by the argument that the deleted line of observations carried all considerable information about the ith regression parameter. What is left is unsure information, possibly based on computational errors. It should not be used in the analysis.

Alternative methods based on g-inverses are computationally more complex and involve at least the same chance of breaking down. In difficult situations, we recommend computing the series $SSR_2(n), \ldots, SSR_2(1)$ backward, using the algorithm of Section 20.2 for incorporating the regression equations in the inverse order.

20.4 Example

A small basic programm was written to test the proposed algorithm. The data, taken from Table 2.3.1. in Schulze (1986), are reported in our *Table 20.1.*

We computed the series of 16 triangular matricies $\mathbf{R}$ incorporating all lines in turn, using the updating algorithm of Section 20.2, and continued computing backward, deleting lines in the inverse order. The same procedure was repeated, starting with line 16 and passing on to line 1. No differences due to computational errors appeared within 4 digits following the decimal point, i.e., we got the same answers for each triangular matrix $\mathbf{R}$, whatever the method of computation. The series of values for the $SSR(m)$, $SSR_1(m)$, and $SSR_2(m)$ are reported in *Table 20.2.*

Table 20.1: Some data from Schulze (1986).

t	x_{1t}	x_{2t}	y_t
1	1.0	0.00	0.0228
2	1.0	0.25	0.0277
3	1.0	0.50	0.0337
4	1.0	0.75	0.0408
5	1.0	1.00	0.0481
6	1.0	1.25	0.0600
7	1.0	1.50	0.0713
8	1.0	1.75	0.0865
9	1.0	2.00	0.1029
10	1.0	2.25	0.1228
11	1.0	2.50	0.1246
12	1.0	2.75	0.1256
13	1.0	3.00	0.1253
14	1.0	3.25	0.1239
15	1.0	3.50	0.1253
16	1.0	3.75	0.1228

Table 20.2: Residual sums of squares for Schulze's (1986) data.

m	$SSR_1(m)$	$SSR_2(m)$	$SSR(m)$
1	0.0000	0.0485	0.0485
2	0.0000	0.0485	0.0485
3	0.0004	0.0482	0.0482
4	0.0011	0.0473	0.0473
5	0.0016	0.0452	0.0453
6	0.0046	0.0411	0.0413
7	0.0069	0.0356	0.0363
8	0.0108	0.0271	0.0292
9	0.0164	0.0167	0.0227
10	0.0216	0.0029	0.0218
11	0.0217	0.0020	0.0218
12	0.0277	0.0016	0.0228
13	0.0267	0.0016	0.0268
14	0.0336	0.0016	0.0336
15	0.0403	0.0000	0.0403
16	0.0485	0.0000	0.0485

A more critical view indicated some trouble with the algorithm of Section 20.3 if there is a change in rank of $\mathbf{X}$. We observed that deleting all but the last two lines does not introduce errors within 7 digits after the decimal point. But then errors suddenly grow so that only 4 digits remain correct. These errors always start to appear after getting a small negative value for p_0, which we changed to zero in order to continue computation. This problem did not appear when we truncated each r_{ii} to zero, which was less than 10^{-6}.

An investigation of more complex data, obtained by adding the variable $x_{3t} = x_{2t}^2$, showed the same results.

20.5 Linear Constraints

In some problems $b_1(m)$ is a least squares estimator subject to linear constraints $\mathbf{A}b_1(m) = a$. $\mathbf{A}$ and a may depend on m, often through $b_2(m)$. Therefore, we suggest updating the factorization of the regression equations as described in Section 20.2 and correcting for the side constraints in each step, separately. This correction takes three steps:

1. Factorize the matrix $(\mathbf{A}\vdots a)$ using the same method as described in Section 20.2 and express rank($\mathbf{A}$) components of $b_1(m)$ in terms of the remaining ones.
2. Replace these rank($\mathbf{A}$) explicitly given components of $b_1(m)$ in the factorized system represented by $\mathbf{R}$ and reduce it to $q - \text{rank}(\mathbf{A})$ unknowns. The new system, represented by $\mathbf{R}_\star$, say, no longer has triangular structure.
3. Factorize $\mathbf{R}_\star$ anew and find $SSR_1(m)$ as the square of the $[q + 1 - rank(\mathbf{A})]$th diagonal element of $\mathbf{R}_\star$.

The constraint least squares estimator $b_1(m)$ follows partly from $\mathbf{R}_\star$ and the factorized constraints $(\mathbf{A}\vdots a)$.

References

Gill, P.E., Golub, G.H., Murray, W., and Saunders, M.A. (1974), Methods for modifying matrix factorizations. *Mathematics of Computation*, **28**, 505-535.

Schulze, U. (1986), *Mehrphasenregression*. Berlin: Akademie Verlag.

Part IV

Data Analysis and Modeling

CHAPTER 21

Change Point Problem Relating to the Poverty Structure

Pranab K. Sen

Summary

An index of poverty reflects the extent to which individuals in a society or community fall below a minimal acceptable standard of living. It is generally framed in terms of a set poverty line, the income distribution of the poor, and other social welfare functions relevant to the poverty structure; the Gini coefficient plays a vital role in this context. The income distribution and other measures based on this distribution rarely remain stationary over time, so that in studying the poverty structure over a period of time, one essentially encounters a time-dependent model that may be analyzed in a parametric or nonparametric manner. In this context, the change point problem is very relevant, and the related methodology is considered in a systematic manner.

21.1 Introduction

For a society or community, *poverty* is usually defined in terms of the extent to which individuals fall below a minimal acceptable standard of living. As such, for a given society, at a given point of time t, an *index of poverty* (π_t) may generally be framed in terms of a set *poverty line* (ω_t), the income distribution (of the poor) $\{F_t(y), y \geq 0\}$, and other *social welfare functions* $\{\mathbf{W}_t\}$ which may have relevance to poverty. Generally, ω_t, $F_t(.)$, and $\mathbf{W}_t$ may not remain stationary over time, and, as a result, π_t may vary over time (t) in a rather involved manner.

In a parametric setup, usually, the social welfare functions are taken into account in formulating an *adjusted income distribution* $F_t^\star(.)$ $[= \{F_t^\star(y), y \geq 0\}]$; and it is taken for

granted that $F_t^\star$ is of some specified functional form (viz., Pareto). In this setup, there are some *unknown parameters* θ_t, appearing as algebraic constants in $F_t^\star$, and, hence, given ω_t, a poverty index π_t $[= \pi_t(\theta_t, \omega_t)]$ can be solely expressed in terms of ω_t and θ_t. Thus, to study the poverty structure over a period of time T, in such a parametric setup, one needs to assume that, for the adjusted income distributions $\{F_t^\star, t \in T\}$, the same functional form remains intact, while the parameters θ_t, $t \in T$ may be time-dependent. Whenever such an assumption holds, the analysis can be done relatively simply (as we shall see in Section 21.3). However, such an assumption may not be very reasonable in all cases (particularly when the span of T is not small). Moreover, even relatively small departures from the assumed functional form of the $F_t^\star(.)$ may cause considerable distortion in the interpretation of the associated parameters θ_t, so that poverty indexes based solely on θ_t (and ω_t) are, in general, not robust against plausible departures from the assumed model. For this reason, other nonparametric indexes of poverty are often advocated in practice.

One of the simplest indexes is the following:

$$\alpha_t = F_t^\star(\omega_t) = \text{ proportion of people at time } t \text{ with adjusted income below the set poverty line } \omega_t \tag{21.1}$$

for $t \in T$. Note that α_t may be estimated in a parametric as well as nonparametric setup. In a parametric setup, α_t is a known function of θ_t and ω_t (and, thereby, it has the same limitations as discussed above). The parametric estimator may be considerably biased (and may even be inconsistent) for any departure of the assumed model from the correct one. In a nonparametric setup, α_t can simply be estimated from the empirical (adjusted) income distribution. Either way, α_t is generally a crude index of poverty as it does not take into account the income inequality of the poor people, and more meaningful indexes of poverty have been considered by various workers.

In this context, we may define the *income gap ratio* (β_t) of poor people (at time t) by

$$\beta_t = 1 - \omega_t^{-1}\{\alpha_t^{-1} \int_0^{\omega_t} y \, dF_t^\star(y)\} \ , \quad t \in T \quad . \tag{21.2}$$

This may be incorporated in the formulation of a second index of poverty — namely,

$$\pi_{tA} = \alpha_t \beta_t \ , \quad t \in T \quad . \tag{21.3}$$

A refined index of poverty, due to A.K. Sen (1976), is the following:

$$\pi_{tS} = \alpha_t\{\beta_t + (1 - \beta_t) G_{t(\alpha)}\} \ , \quad t \in T \quad , \tag{21.4}$$

where $G_{t(\alpha)}$ is the *Gini coefficient of the income distribution among the poor at time* t, and we shall formally define it in Section 21.2.

Takayama (1979) considered another index:

$$\pi_{tT} = G^c_{t(\omega)} \ , \quad t \in T \quad , \tag{21.5}$$

where $G^c_{t(\omega)}$ is the *Gini coefficient of the income distribution censored at* ω_t (to be defined in Section 21.2). There are certain undesirable properties of π_{tT}, while π_{tS} is generally higher than π_{tA} or π_{tT}. Based on certain robustness criteria, Sen (1986) has proposed the following poverty index:

$$\pi^\star_t = \alpha_t\{\beta_t^{1-G_{t(\alpha)}}\}, \quad t \in T \quad . \tag{21.6}$$

It has been shown there that for arbitrary income distributions and poverty lines:

$$0 \le \pi_{tT} \le \pi_{tA} \le \pi^\star_t \le \pi_{tS} \le \alpha_t\beta_t(2-\beta_t) \le \alpha_t \quad . \tag{21.7}$$

Further, it follows from Sen (1986) that

$$G^c_{t(\omega)} = \alpha_t G_{t(\alpha)} + \alpha_t(1-\alpha_t)(1-\alpha_t\beta_t)^{-1}\{\beta_t - G_{t(\alpha)}\} \ , \quad t \in T \quad , \tag{21.8}$$

where $G_{t(\alpha)} \le \beta_t$, $\forall t$. Given this picture, we can express each of these poverty indices in terms of α_t, β_t, and $G_{t(\alpha)}$. Thus, we consider a general poverty index of the form

$$\pi_t = \pi(\alpha_t, \beta_t, G_{t(\alpha)}) \ , \quad t \in T \quad . \tag{21.9}$$

Also, as in Sen (1986, Sec. 3), we assume that over the unit cube $E_3 = \{\mathbf{x} : x_i \in [0,1], i = 1,2,3\}$, $\pi(\mathbf{x})$ is monotone in each of its three arguments (when the other two are held fixed). Further, we introduce the following partial ordering of income distributions [considered in Sen (1986)]:

Two income distributions F_1 *and* F_2 *are said to be poverty index (PI-)ordered*, i.e., $F_1 >^{PI} F_2$, *if* $\{\alpha(F_1), \beta(F_1), G_{[\alpha(F_1)]}\} \ge \{\alpha(F_2), \beta(F_2), G_{[\alpha(F_2)]}\}$ *where* $\mathbf{x} \ge \mathbf{x}'$ *means* $x_i \ge x'_i$, *for* $i = 1,2,3$.

It follows from the above that if the (adjusted) income distributions $F^\star_t$ are PI-ordered and π_t, in (21.9), is a monotone function of each of its three arguments, then the π_t, $t \in T$, are ordered, too. In a nonparametric setup, this PI-ordering of the adjusted income distributions may be incorporated in a convenient manner to study the poverty structure in a coherent way. This will be mainly exploited in the current study.

Due to inflation and progressive changes in socioeconomic conditions, the poverty line ω_t, income gap ratio β_t, α_t as well as the Gini coefficient $G_{t(\alpha)}$ rarely remain stationary over T. On the other hand, from time to time, some extraneous factors may have profound impact on the socioeconomic condition, and, thereby, may lead to an instantaneous change in these measures. For example, the "oil embargo" (by the OPEC) in 1973–1974 produced a serious inflationary effect all over the globe (with more serious effects on the Third World countries); as a result, α_t, β_t, ω_t, $G_{t(\alpha)}$, etc., were all shaken up at that time. On the other hand, the recent "lowering of oil prices" by OPEC has not led to a similar effect (in

the opposite direction), as most of the governments have absorbed the difference in the form of direct or indirect taxes. Due to such considerations, it may be quite plausible to formulate the "change point model" relating to the poverty structure in such a way that a smooth change in the picture over the time period T constitutes the null hypothesis, while a change from this smooth pattern at some intermediate point (in T) is to be taken as the "change point". Thus, unlike the classical case, the constancy of the π_t (over $t \in T$) may not be the null hypothesis of interest.

Along with other preliminary notions, this basic formulation of the "change point" problem relating to poverty is considered in Section 21.2. Some parametric procedures are then presented in Section 21.3. The main emphasis is on suitable nonparametric procedures, which are presented in Section 21.4. The concluding section offers some general remarks.

21.2 Preliminary Notions

Keeping (21.4), (21.5), and (21.6) in mind, we proceed first to define the various versions of the Gini coefficient of income distribution, and present some of their basic properties. These results are mainly adapted from Sen (1986).

For the sake of notational simplicity, we denote an adjusted income distribution by $F = \{F(x), x \geq 0\}$ and let

$$\mu_F = \int_0^\infty x\, dF(x), \quad F^{-1}(u) = \inf\{x : F(x) \geq u\}, \quad 0 \leq u \leq 1 \quad ; \tag{21.10}$$

$$\lambda_F(u) = \mu_F^{-1}\{\int_0^{F^{-1}(u)} x\, dF(x)\}, \quad 0 \leq u \leq 1 \quad . \tag{21.11}$$

Also, for any given w (> 0), we let $\alpha = \alpha(w) = F(w)$. Then the *income distribution of the poor* is defined as the truncated distribution of the income on $[0, w]$, i.e.,

$$F_\alpha(x) = \begin{cases} \alpha^{-1}F(x)\,, & 0 \leq x \leq w \\ 1\,, & x > w \quad . \end{cases} \tag{21.12}$$

On the other hand, the *censored income distribution* (at w) is defined as

$$F_w^c(x) = \begin{cases} F(x)\,, & 0 \leq x < w \\ 1\,, & x \geq w \quad . \end{cases} \tag{21.13}$$

Plotting $\lambda_F(u)$ against u on $[0,1]^2$ gives the *Lorenz Curve* for the income distribution F, and the *Gini coefficient* is twice the area formed by the Lorenz curve and the diagonal line joining the lower and upper corner of $[0,1]^2$. Analytically, the Gini coefficient G may be expressed as

$$G = (2\mu_F)^{-1}E|Y_1 - Y_2| = 1 - 2\mu_F^{-1}\int_0^\infty y\,\overline{F}(y)\,dF(y) \tag{21.14}$$

where $\overline{F}(y) = 1 - F(y)$, $y \geq 0$, and Y_1, Y_2 are two independent r.v.'s, each having the d.f. F. If in (21.14), we replace the d.f. F by F_α (or F_w^c), then the corresponding measure is the Gini coefficient for F_α (or F_w^c); (21.12), (21.13), and (21.14) lead us to (21.8). Other properties of the versions of the Gini coefficient are studied in Sen (1986).

Note that if we have two income distributions, F_1 and F_2, such that for some arbitrary k (> 0),

$$F_1(x) = F_2(kx)\ ,\quad \forall x \in R^+ \tag{21.15}$$

then the Gini coeffficients for F_1 and F_2 are the same (for every $k > 0$). This scale-invariance of the Gini coefficient has an important role to play in our study. If the set $\{F_t^\star; t \in T\}$ of the adjusted income distributions satisfies the following:

$$F_t^\star(x) = F^\star(k_t x),\quad x \in R^+,\quad t \in T \tag{21.16}$$

where $\{k_t; t \in T\}$ is an arbitrary set of positive numbers, then the Gini coefficients for the $F_t^\star$ are all the same. In such a case, if poverty lines ω_t, $t \in T$ are adjusted accordingly, i.e.,

$$\omega_t = k_t^{-1}\omega_0\ ,\quad \forall t \in T \tag{21.17}$$

where the k_t satisfy (21.16), then we have by (21.1), (21.16), and (21.17),

$$\alpha_t = F_t^\star(\omega_t) = F^\star(\omega_0) = \alpha_0\ ,\quad \forall t \in T \tag{21.18}$$

and little analysis yields that by (21.2), (21.16), (21.17), and (21.18),

$$\beta_t = \beta_0 = 1 - \omega_0^{-1}\{\alpha_0^{-1}\int_0^{\omega_0} y\,dF^\star(y)\}\ ,\quad \forall t \in T\ . \tag{21.19}$$

Further, (21.16) and (21.18) may be used to verify easily that, for the income distributions of the poor, the Gini coefficients satisfy the following:

$$G_{t(\alpha)} = G_{(\alpha_0)}\ ,\quad \forall t \in T\ . \tag{21.20}$$

As a result, we conclude that under (21.16) and (21.17),

$$\pi_t = \pi_0\ ,\quad \forall t \in T\ . \tag{21.21}$$

Thus, if inflationary forces induce only scalar changes in the income distributions and if the poverty lines are adjusted by the same scale factors, than the poverty indices remain stationary over time. This may be the basic idea of *cost-of-living* adjustments followed in some countries, but it could be generally misleading if, besides the scale changes, there are other changes in the income distributions.

On the other hand, if the inflationary factors induce (in addition to possible scale changes) some other changes in the "shape" of the income distributions and/or the poverty lines are not adjusted for the inflationary factors, then (21.21) may not hold. In this framework, it is quite reasonable to frame a null hypothesis

$$H_0^{(1)}: \quad \pi_t = \pi_0 \text{ (unknown)}, \quad \forall t \in T \tag{21.22}$$

against an alternative

$$H^{(1)} = \bigcup_{\tau \in T} H_\tau^{(1)}; \quad H_\tau^{(1)} : \begin{array}{ll} \pi_t = \pi_0, & t < \tau \\ \pi_t = \pi_\tau \neq \pi_0, & t \geq \tau \end{array} . \tag{21.23}$$

Thus, τ represents the usual "change point" with respect to the poverty index. Both parametric and nonparametric tests for this change point problem will be considered here.

As we shall see in Section 21.3, even in the simplest parametric case, an income distribution is characterized by a scale parameter and a second parameter reflecting the income inequality. In such a case, (21.16) is not so general and a variation in the "income inequality" parameter may vitiate (21.21) even if the poverty lines are adjusted by the scale factors. Thus, a second "change point" model may be framed in terms of the homogeneity of the income inequality parameters against possible shifts at an unknown time point τ (belonging to T) where the scale parameters may be treated as nuisance parameters. Both parametric and nonparametric tests for this problem will be considered.

In a general framework, an income distribution may not be solely characterized by scale and "inequality of income" parameters. Nevertheless, whithin the subclass of PI-oredered family of income distributions, a change point problem may equivalently be posed in terms of the triplet $\{\alpha_t, \beta_t, G_{t(\alpha)}\}$. It may therefore be convenient to introduce a parameter (vector)

$$\xi_t = (\alpha_t, \beta_t, G_{t(\alpha)})' \ , \quad t \in T \tag{21.24}$$

and to pose the hypothesis in terms of the ξ_t. To be more general [than in (21.22)–(21.23)], we may conceive of a vector $\zeta_t = (\zeta_{1t}, \ldots, \zeta_{qt})'$ of known functions of the variable t (viz., the ζ_{jt} are the orthogonal polynomials of degree j in $t \in T$, $0 \leq j \leq q-1$, for some $q \geq 1$), and write

$$\xi_t = \Gamma_t \zeta_t \ , \quad t \in T \tag{21.25}$$

where $\mathbf{\Gamma}_t$ is a $3 \times q$ matrix of unknown constants. The relevant null hypothesis (of no-change-point) is framed as

$$H_0: \quad \mathbf{\Gamma}_t = \mathbf{\Gamma} \text{ (unknown)}, \quad \forall t \in T \tag{21.26}$$

against an alternative

$$H = \bigcup_{\tau \in T} H_\tau \; ; \; H_\tau : \begin{matrix} \mathbf{\Gamma}_t = \mathbf{\Gamma}_0, & t < \tau \\ \mathbf{\Gamma}_t = \mathbf{\Gamma}_\tau \neq \mathbf{\Gamma}_0, & t \geq \tau \end{matrix} \; . \tag{21.27}$$

Thus, τ represents the usual "change point" with respect to the structural relationship in (21.25). This problem is posed more in line with the "constancy of the regression relationship over time" problem treated (for linear models) in Brown *et al.* (1975) and others. The hypothesis in (21.18) through (21.20) can easily be framed by letting $q = 1$ and $\zeta_{1t} = 1$, $\forall t \in T$. Thus, introducing a more general ζ_t in (21.25) and drawing a parallel to the "constancy of regression relationship" problem, we are able to formulate a more general "change point" problem relating to the poverty structure, and the main purpose of the current study is to provide suitable statistical procedures relating to this problem.

21.3 Parametric Formulation

Corresponding to the income distribution $F = \{F(x);\ x \geq 0\}$, we define the survival distribution $\overline{F}$ by

$$\overline{F}(x) = 1 - F(x) \; , \quad x \geq 0 \quad . \tag{21.28}$$

In the parametric case, the most commonly used income distribution is the *Pareto distribution*, which corresponds to a survival function of the form

$$\overline{F}(x) = \begin{cases} (x/\sigma)^{-\gamma} \; , & x \geq \sigma \\ 1 \; , & x < \sigma \end{cases} \tag{21.29}$$

where $\sigma > 0$ is a *scale parameter* and $\gamma > 0$ is an *index of inequality*. Although (21.29) may quite adequately describe the "upper tail" of an income distribution, it may, generally, have some drawbacks for the lower tail; and, for the poverty indexes, we are confronted with this lower part of the income distribution. A *Burr distribution* having the survival function

$$\overline{F}(x) = \{1 + (x - \mu/\sigma)^{\delta}\}^{-\gamma}, \quad x \geq \mu (\geq 0) \tag{21.30}$$

is often used to provide a more flexible form of the income distribution of the poor. In this formulation, μ, δ, σ, and γ are all nonnegative parameters. The particular case of $\delta = 1$ (i.e., $\overline{F}(x) = k_1(x + c)^{-\gamma}$, $x \in R^+$,) is known as the "*Pareto distribution of the second kind*" or the *Lomax distribution* [viz., Johnson and Kotz (1970)]. In fact, there is also a "*Pareto distribution of the third kind*", for which the survival function is given by

$$\overline{F}(x) = k_2 e^{-bx}(x + c)^{-\gamma} \ , \quad k_2 > 0, b \geq 0, \gamma > 0 \ . \tag{21.31}$$

For a nice treatment of the statistical interpretations and estimation of the parameters of such (generalized) Pareto distributions, we may refer to Johnson and Kotz (1970, ch. 19), where other references are also cited.

For the original Pareto distribution, which is (21.29), the arguments made in Section 21.2 [viz., (21.15) through (21.23)] provide simple statistical procedures for the change point model. In fact, for the scale-adjusted poverty line, for the model (21.29), we have thus the reduced problem of testing the equality of the γ_t (i.e., the indexes of inequality) against a possible change at an unknown time point τ ($\in T$). Thus, if $\{\gamma_t, t \in T\}$ stands for the set of the index of inequality parameters of the F_t, $t \in T$ [each of which having the form (21.29) with possibly different σ], then based on independent samples from the F_t, maximum likelihood estimators (MLE) of the γ_t may easily be incorporated in the formulation of CUSUM or some other recursive tests for the change point problem. Based on the asymptotic normality and other related properties of these MLE (as well as the consistency properties of the estimated information matrices), the theory developed in Brown *et al.* (1975) and others can easily be adapted. Instead of the MLE, one may [as in Johnson and Kotz (1970)] also use quantile estimators of the σ_t; and these estimators may also be used in a similar manner for constructing the test statistic for the change point problem. If, however, the poverty lines ω_t, $t \in T$, do not satisfy (21.17), then the test based on the estimators of the γ_t, $t \in T$, gives only a partial picture.

This inadequacy becomes much more significant when the F_t follows the pattern in (21.30) or (21.31). In such a case, we have a multiparameter model, so that a change point needs to be posed in terms of all the parameters $(\mu_t, \sigma_t, \delta_t, \gamma_t)$ or $(k_{2t}, b_t, c_t, \gamma_t)$ as well as the ω_t, $t \in T$. This can, of course, be done in the same manner as in the nonparametric case (which we shall be considering in detail in Section 21.4). However, we have some reservations about advocating these parametric procedures over their nonparametric counterparts:

1. In the nonparametric case, we would be dealing with vectors ζ_t in (21.24) having three basically meaningful components α_t, β_t, $G_{t(\alpha)}$, while in the parametric case, for (21.30) or (21.31), we have four parameters, not all of which may be relevant to the poverty structure. This redundancy may generally entail some loss of power (or efficiency) of the tests for the parametric model (relative to their nonparametic counterparts).

2. The parametric model-based tests may not be that robust against plausible departures from the assumed model, while the nonparametric ones are not based on any such particular form of the F_t, and, hence, remain robust for a broad class of F_t's.

3. Finally, (21.29)–(21.31) are all known to be more appropriate for the higher income structure, while for the income distribution of the poor, under various social welfare functions, the income distributions may not conform to any one of them. In fact, it is not unlikely to have positive probability mass at the null income point (due to unemployment, etc.) or at some other levels, so that a generalized Pareto distribution may not be so appropriate.

For these reasons, we shall study the nonparametric case in detail, and append a small discussion on some parametric procedures that may be formulated along the same line.

21.4 Nonparametric Formulations

Based on the *PI-ordering* of the income distributions in Section 21.1, we may consider a general family F of income distributions, and within this family, the change point problem relating to the poverty structure may conveniently be studied in terms of the triplets α_t, β_t, $G_{t(\alpha)}$ in (21.24).

Note that $\alpha_t = F_t^\star(\omega_t)$ is the proportion of the poor people at time t, while, defining the truncated income distribution as in (21.12), $\omega_t(1-\beta_t)$ is the average income of poor people at time t and $G_{t(\alpha)}$ is the corresponding Gini coefficient. All of these are therefore estimable parameters of the income distribution $F_t^\star$ and its truncated version $F_{\alpha_t}^\star$. The natural estimator $(\hat{\alpha}_t)$ of α_t is the sample proportion of poor people at time t, and it is therefore a simple *U-statistic of degree* 1; we may refer to Hoeffding (1948) for the definition and other properties of the U-statistics. Similarly, $\hat{\beta}_t$, the estimator of β_t, is a function of two U-statistics, and $\hat{G}_{t(\alpha)}$, the estimator of $G_{t(\alpha)}$, is a function of two U-statistics (i.e., the symmetric, unbiased estimators of μ_t and $E|Y_{t1} - Y_{t2}|$, respectively, where the Y_{ti} are independent with the truncated d.f. $F_{\alpha_t}^\star$). This characterization enables us to incorporate the existing theory (in an asymptotic setup) in the formulation of the proposed tests for the change point model.

We denote the time domain T by

$$T = \{t_1, \ldots, t_K\}\ ,\ \ t_1 < \ldots < t_K\ \ , \tag{21.32}$$

where K is a positive integer. Suppose that at the time point t_k, we have a sample of n_k observations on the income variable; and based on this sample of size n_k, we have an estimator $\hat{\xi}_k = (\hat{\alpha}_t, \hat{\beta}_t, \hat{G}_k)$ of ξ_{tk} in (21.24). As has already been discussed earlier, each of $\hat{\alpha}_t$, $\hat{\beta}_t$, and $\hat{G}_k$ is a function of (at most two) U-statistics. This enables us to make use of the standard properties of U-statistics (including the asymptotic normality and moment convergence) for the study of the corresponding properties of the $\hat{\xi}_k$, $k \leq K$. Note that the coordinates of $\hat{\xi}_k$ are all nonnegative and bounded from above by 1. Thus, we have no problem in claiming the usual moment convergence results [on the $n_k^{1/2}(\hat{\xi}_k - \xi_{t_k})$] along with the usual asymptotic normality results.

Keeping this in mind [along with (21.25)], we assume that

$$\hat{\xi}_k = \Gamma_{t_k}\zeta_{t_k} + \mathbf{e}_k \ , \quad k = 1, \ldots, K \tag{21.33}$$

where the $\mathbf{e}_k$ have zero mean vector and dispersion matrix $n_k^{-1}\mathbf{V}_k$, $k = 1, \ldots, K$. Usually, in a nonparametric setup, these $\mathbf{V}_k$ are not known. Nevertheless, as in Sen (1977), suitable jackknifed estimators of them may easily be obtained. We denote the usual jackknifed estimator of $\mathbf{V}_k$ by $\mathbf{V}_k^*$, for $k = 1, \ldots, K$. Then, whenever the n_k are large, we may assume that for each k $(= 1, \ldots, K)$,

$$n_k^{1/2}\mathbf{e}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_k) \ , \tag{21.34}$$

$$\mathbf{V}_k^{-1}\mathbf{V}_k^* \xrightarrow{p} \mathbf{I}_3, \ \text{ as } n_k \to \infty \ . \tag{21.35}$$

The restriction that the n_k are all large does not pose a serious problem in this study (as large data sets are generally available). We define

$$N = n_1 + \ldots + n_K \tag{21.36}$$

and assume that there exist positive numbers $\lambda_1, \ldots, \lambda_K$, such that as N becomes large,

$$N^{-1}n_k \to \lambda_k \ , \quad k = 1, \ldots, K, \ \text{where} \ \sum_{k=1}^{K} \lambda_k = 1 \ . \tag{21.37}$$

21.4.1 The pseudo two-sample approach

As in the classical change point problem, the tests we shall formulate may be based on a "pseudo two-sample" approach or on the so-called "recursive residuals". We consider first the pseudo two-sample approach. For every k $(q \le k \le K - q)$, we consider the partition of $\hat{\xi}_k$ into two subsets

$$\{\hat{\xi}_i \ ; \ i \le k\} \quad \text{and} \quad \{\hat{\xi}_i \ ; \ k < i \le K\} \tag{21.38}$$

and assuming the homogeneity of the Γ_t, from each of these subsets, we consider the usual weighted least squares estimates. Thus, we consider the usual quadratic norm:

$$\sum_{i \le k} n_i(\hat{\xi}_i - \Gamma\zeta_{t_i})'(\mathbf{V}_i^*)^{-1}(\hat{\xi}_i - \Gamma\zeta_{t_i}) \tag{21.39}$$

$$\sum_{k < i \le K} n_i(\hat{\xi}_i - \Gamma\zeta_{t_i})'(\mathbf{V}_i^*)^{-1}(\hat{\xi}_i - \Gamma\zeta_{t_i}) \tag{21.40}$$

and minimize each with respect to the unknown Γ (having $3q$ elements). We denote these weighted least squares estimators by $\hat{\Gamma}_{N(k)}$ and $\hat{\Gamma}_{(k)N}$, respectively. Note that these estimators are linear (in the $\hat{\xi}_i$).

In the next step, we consider a suitable "norm" for the difference $\hat{\Gamma}_{N(k)} - \hat{\Gamma}_{(k)N}$, i.e., we take

$$\|\mathbf{A} - \mathbf{B}\| \tag{21.41}$$

where $\mathbf{A}$ and $\mathbf{B}$ are $3 \times q$ matrices. Actually, we may roll out $\mathbf{A}$ and $\mathbf{B}$ into $3q$-vectors, and then $\|\mathbf{A} - \mathbf{B}\|$ may be taken as the *"max norm"* (i.e., the maximum of the coordinatewise distances) or as a quadratic norm (often termed the *Mahalanobis distance*). Let then

$$D_{Nk} = N^{1/2}\|\hat{\mathbf{\Gamma}}_{N(k)} - \hat{\mathbf{\Gamma}}_{(k)N}\| \ , \quad k = q, \ldots, K - q \tag{21.42}$$

$$D_N^\star = \max_{q \le k \le K-q} D_{Nk} \ . \tag{21.43}$$

Note that for $k < q$ (or $k > K - q$), $\mathbf{\Gamma}$ is not estimable from the first (or second) subset in (21.38), and, hence, in (21.42)–(21.43), we confine ourselves to $q \le k \le K - q$.

Under the null hypothesis (of homogeneity of the $\mathbf{\Gamma}_t$), the D_{Nk} in (21.42) are all bounded in probability (and, hence, $D_N^\star$ is also so). If, however, this null hypothesis is not true, but (21.23) holds, then for at least one k ($q \le k \le K - q$), $\|\hat{\mathbf{\Gamma}}_{N(k)} - \hat{\mathbf{\Gamma}}_{(k)N}\|$ will be stochastically different from 0, and, hence, $D_N^\star$ will be large, in probability. This leads us to consider the following test procedure:

> Accept or reject the null hypothesis (of homogeneity of the $\mathbf{\Gamma}_t$) according as $D_N^\star$ is $\le$ or $> D_{N,\epsilon}^\star$, where $(0 < \epsilon < 1)$ is the desired *significance level* of the test and $D_{N,\epsilon}^\star$ is the corresponding *critical value.* (21.44)

Note that by virtue of (21.34)–(21.37) and the general theory of weighted least squares estimators, there exist stochastic matrices $\mathbf{Z}_k$, $q \le k \le K - q$, such that

$$\|N^{1/2}\{\hat{\mathbf{\Gamma}}_{N(k)} - \hat{\mathbf{\Gamma}}_{(k)N}\} - \mathbf{Z}_k\| \xrightarrow{p} 0 \ , \quad \forall k \ , \quad q \le k \le K - q \tag{21.45}$$

and under the null hypothesis,

$$\mathbf{Z} = (\mathbf{Z}_q, \ldots, \mathbf{Z}_{K-q}) \xrightarrow{D} \mathcal{N}_{3q(K-2q+1)}(\mathbf{0}, \mathbf{A}) \tag{21.46}$$

where $\mathbf{A}$ depends on the λ_k, $\mathbf{V_k}$ as the ζ_t, $t \in T$. Thus,

$$D_N^\star \xrightarrow{D} Z^\star = \max_{q \le k \le K-q} \|\mathbf{Z}_k\| \ . \tag{21.47}$$

Hence, to carry out the testing procedure, we need to find out the distribution of $Z^\star$ under the model (21.46). We shall make detailed comments on it later on. If the null hypothesis is not true but (21.27) holds, then $D_N^\star = O_p(N^{1/2})$, so that the test in (21.44) is consistent against $H^{(1)}$ in (21.27). The testing problem in (21.22)–(21.23) may be treated in a more simple manner by using the (scalar) estimates of the poverty indexes from the first k and last $K - k$ samples; and in this case, the theory in the classical change

point model applies directly. Since most of these results have been discussed in detail in Sen (1985, Ch. 3), we omit the details.

Let us also denote $\hat{\mathbf{\Gamma}}_{N(K)}$ by $\hat{\mathbf{\Gamma}}_N$ and let

$$\tilde{D}_{Nk} = N^{1/2}\|\hat{\mathbf{\Gamma}}_{N(k)} - \hat{\mathbf{\Gamma}}_N\| \ , \quad k = q, \ldots, K \tag{21.48}$$

$$\overline{D}_N^{\star} = \max_{q \le k \le K} \overline{D}_{Nk} \ . \tag{21.49}$$

Then, a variant form of the test in (21.44) may also be based on $\overline{D}_N^{\star}$ in (21.49). The asymptotic results in (21.45)–(21.47) may easily be extended to this case, and, hence, we may conclude that under the null hypothesis,

$$\tilde{D}_N^{\star} \xrightarrow{D} \tilde{Z}^{\star} = \max_{q \le k \le K} \|\tilde{\mathbf{Z}}_k\| \tag{21.50}$$

where the $\tilde{\mathbf{Z}}_k$ have jointly (asymptotically) a multinormal distribution.

Following the Bayesian formulation of the change point problem relating to the normal mean [viz., Chernoff and Zacks (1964)], we may also consider a different procedure based on the estimates $\hat{\mathbf{\Gamma}}_{N(k)}$, $k \le K$. Let ℓ_{Nk}, $k \le K$ be a monotone function of k (e.g., $\ell_{Nk} = N^{1/2}[k - (K+1)/2]$, $1 \le k \le K$) and let

$$\mathbf{D}_N^0 = \sum_{k \le K} \ell_{Nk}(\hat{\mathbf{\Gamma}}_{N(k)} - \hat{\mathbf{\Gamma}}_N) \ , \quad D_N^0 = \|\mathbf{D}_N^0\| \ . \tag{21.51}$$

Then, under the null hypothesis, the asymptotic multinormality of $\mathbf{D}_N^0$ may be obtained along the same lines as in (21.45)–(21.46), so that we have

$$D_N^0 \xrightarrow{D} Z^0 = \|\mathbf{Z}^0\| \ ; \quad \mathbf{Z}^0 \sim N_{3q}(\mathbf{0}, \mathbf{A}^0) \ . \tag{21.52}$$

The matrix $\mathbf{A}$ depends on the $\mathbf{V}_k$, λ_k, ζ_t, $t \in T$, as well as on the ℓ_{Nk}, $k \le K$. One possibility is to choose the ℓ_{Nk} in such a way that the test based on D_N has (locally) maximum average power, where τ is allowed to assume each of the points $t_1, \ldots, t_K$ with a given probability distribution $\pi^{(k)}$.

From the operational point of view, in each case, the basic problem is to derive the percentile points of the distribution of $\mathbf{Z}^{\star}$ (or $\tilde{\mathbf{Z}}$ or $\tilde{\mathbf{Z}}^0$), where these statistics do not generally have the normal, chi (square) or other simple forms of distribution. Though in some specific cases, the exact distribution of $Z^{\star}$ (or the other statistics) can be obtained in a closed form, in the general case, we may have to rely on numerical or simulation methods. The Bessel process results of DeLong (1981) may be used in some specific cases. However, we may consider the following simulation method in the general case.

From the given ζ_t, $t \in T$, $\hat{\xi}_k$, $k \le K$, and the jackknifed estimators $\mathbf{V}_k^{\star}$, $k \le K$, we derive a suitable estimator $\hat{\mathbf{A}}$ of the dispersion matrix $\mathbf{A}$ in (21.46). Note that there exists

a matrix $\mathbf{B}^*$ such that $\mathbf{A} = \mathbf{B}^*(\mathbf{B}^*)'$. Let $p = 3q(K - 2q + 1)$ and let $\mathbf{Y} = (Y_1, \ldots, Y_p)'$ be a vector having the normal distribution $N(\mathbf{0}, \mathbf{I}_p)$. We generate a large number (say, M) copies of $\mathbf{Y}$ (i.e., pM standard normal deviates), and denote these by $\mathbf{Y}_s$, $1 \leq s \leq M$. Let then $\mathbf{Y}_s^* = \mathbf{B}^*\mathbf{Y}_s$, $1 \leq s \leq M$ and let $Y_{(s)}^* = \|\mathbf{Y}_s^*\|$, the norm being defined as in (21.47). We arrange these $MY_{(s)}^*$ in ascending order of magnitudes, so that the $[M\varepsilon]$th observation from the top provides the desired simulated value for $D_{N,\varepsilon}^*$. Since the computation of $\mathbf{B}^*$ is needed once for all, this simulation procedure is relatively inexpensive and can be done for a broad range of situations. When K is large, Gaussian process approximations may also be used toward the same goal.

It is clear that this pseudo two-sample approach can also be used for the parametric model. We may have to estimate the parameters $(\mu_t, \sigma_t, \delta_t, \gamma_t)$ (or $k_{2t}, b_t, c_t, \gamma_t$) and their dispersion matrices, and with these, we may repeat the steps in (21.42) and (21.43) [or (21.48)–(21.49)]. The disadvantage here will be that instead of $3q$ parameters ($q \geq 1$), we would have $4q$ parameters, so that for $q > 1$, the nonparametric approach will be computationally simpler, too.

21.4.2 The recursive-residual approach

Define the estimates $\hat{\mathbf{\Gamma}}_{N(k)}$ as in (21.39)–(21.41). Consider then the recursive residuals:

$$\omega_{Nk} = \hat{\xi}_k - \hat{\mathbf{\Gamma}}_{N(k-1)}\zeta_{t_k} \quad \text{for } k = q+1, \ldots, K \tag{21.53}$$

Also, define the CUSUMs for these recursive residuals by

$$\mathbf{W}_{Nk} = \sum_{i \leq k} \omega_{Ni} \quad \text{for } k = q+1, \ldots, K \tag{21.54}$$

We may then define

$$W_{Nk} = \begin{cases} \|\mathbf{W}_{Nk}\|, & k = q+1, \ldots, K \\ 0, & k \leq q \end{cases} \tag{21.55}$$

$$W_N^* = \max_{k \leq K} W_{Nk} \ . \tag{21.56}$$

The test procedure is similar to that in (21.44). Thus, the basic problem is to obtain the critical values for W_N^*. For large K, invariance principles for recursive residuals considered by Sen (1982) and others may be incorporated to provide a Bessel process approximation to the CUSUMs in (21.54), so that boundary-crossing probabilities for such Bessel processes [studied by DeLong (1981) and others] can be used to provide the desired critical levels. For values of K not so large, it may be quite appropriate to consider a simulation method similar to the earlier case.

Note that under the null hypothesis, by (21.34), (21.35) and (21.53), (21.54),

$$N^{1/2}(\mathbf{W}_{Nq+1}, \ldots, \mathbf{W}_{NK}) \xrightarrow{D} N_{3(K-q)}(\mathbf{0}, \overline{\mathbf{A}}) \tag{21.57}$$

where the dispersion matrix $\overline{\mathbf{A}}$ can be consistently estimated by $\overline{\mathbf{A}}^{\star}$, using the jackknifed estimators $\mathbf{V}_k^{\star}$, $k \leq K$ and the ζ_t, $t \in T$. Thus, if we choose a $\overline{\mathbf{B}}^{\star}$ such that $\overline{\mathbf{A}}^{\star} = \overline{\mathbf{B}}^{\star}(\overline{\mathbf{B}}^{\star})'$, we may again generate normal (vectors) $\mathbf{Y}_r = (Y_{1r}, \ldots, Y_{pr})'$, $p = 3(K-q)$, $r \geq 1$, such that $Var(\mathbf{Y}_r) = \mathbf{I}_p$, and let $\mathbf{Y}_r^{\star} = \overline{\mathbf{B}}^{\star}\mathbf{Y}_r$, $r \geq 1$. We define then $Y_r^{\star} = \|\mathbf{Y}_r^{\star}\|$, $r \geq 1$. The M generated Y_r are arranged in ascending order of magnitudes and the $[M\varepsilon]$th value from the top provides a consistent (empirical) estimate of the critical level of $N^{1/2}W_N^{\star}$.

A similar recursive residual test may also be worked out for the parametric approach [using the estimates of the parameters k_{2t}, b_t, c_t, γ_t and forming the recursive residuals].

21.5 Some General Remarks

In Section 21.2, we characterized the change point model relating to the poverty structure in terms of all the parameters in the parametric case and α_t, β_t, and $G_{t(\alpha)}$ in the nonparmetric case. If, however, we choose a specific index of poverty, π_t, such as the one in (21.4), (21.5), or (21.6), then in the parametric case, π_t can be expressed as a real valued function of the parmeters [in (21.30) or (21.31)]. Then, in terms of the estimates $\hat{\pi}_t$ (derived through efficient estimators of these parameters), we would have a univariate model in (21.33), and the analysis would have been simpler. Similarly, in the nonparametric case, $\hat{\pi}_t$ would be a real valued function of three U-statistics [namely, the estimates of α_t, $\omega_t(1-\beta_t)$, and $\alpha_t G_{t(\alpha)}$], so that we could have written a model on these $\hat{\pi}_t$, $t \in T$, parallel to (21.33), where the $\mathbf{e_k}$ are scalar random variables having asymptotic normal distributions.

In either case, the deailed analysis made in Section 21.4 simplifies a great deal, and the normal theory model considered by Brown *et al.* (1975) can be readily amended to provide good approximation to the critical values of the test statistics. For robustness considerations, however, we prefer to use the multidimensional model in Section 21.3 or 21.4. This is because the real valued function of the ξ_t [such as the π_t in (21.9)] may not be sensitive to changes in all directions and, hence, may not be totally relevant to the broad spectrum of poverty structure. A better picture is conveyed by the vector ξ_i, $t \in T$, and, hence, the analysis made in Section 21.4 would reveal this picture to a greater extent than the one based on the $\hat{\pi}_t$ alone. Also, between the parametric and nonparametric models, as has been explained earlier, the nonparametric ones seem to be more meaningful and less dependent on the form of the underlying income distributions. Thus, on such robustness grounds, we prefer to use the nonparametric methods.

It is not uncommon to have indefinitely large sample sizes $n_1, \ldots, n_k$, so that in (21.33)–(21.34), the error terms $\mathbf{e}_k$ would have negligible covariance matrices. In this case, from (21.42) and (21.43), we may conclude that the test in (21.44) will be sensitive to any departure from the assumed homogeneity of the $\hat{\Gamma}_t$, $t \in T$. Thus, from a descriptive analysis point of view, the test in (21.44) provides a good interpretation of the basic

model and allows the normal theory procedures to have the leading role in the proposed analysis when the n_k are not small.

In Section 21.4 (or 21.3), we considered the case where k, the number of time points belonging to the interval T, is fixed. In economic analysis, it is not uncommon to have a large number of time points in T (as may be obtained by considering the monthly or quarterly pictures instead of the annual ones). In such a case, one may have to be a little extra careful in identifying the seasonal and/or other short-term fluctuations in the income pattern.

For example, for workers in the agricultural sector, the income pattern may vary drastically from the harvest season to the winter one, so that on a fixed poverty line, the same group of people may not always be under the poverty line. We may, of course, seasonally adjust the poverty line, but that is likely to impose a seaonal structure on the poverty indexes. Hence, in (21.25) or elsewhere, such seasonality effects should be taken into account. This would naturally make the analysis more complicated.

It is better to eliminate the seasonality and other short-term effects in the definition and interpretation of poverty structures and indexes, and then to carry out the proposed analysis on such adjusted estimators. For large values of k, the multinormal models worked out in Section 21.4 can easily be extended to multivariate Gaussian processes, and parallel results can be obtained by reference to the change point models pertaining to such Gaussian processes. Some of these developments are discussed in Sen (1985, Ch. 3), where other references are also cited.

Finally, for the parametric models in Section 21.3, we may use some convenient estimators of the parameters (other than the maximum likelihood estimators), such as the L-estimators discussed in Johnson and Kotz (1970, Ch. 19). In the nonparametric case, too, the $\hat{\xi}_k$ are based on (functions of) U-statistics. To reduce possible bias, we may as well use the corresponding jackknifed estimators. By virtue of the results in Sen (1977), these jackknifed estimators lead to the same asymptotic theory, and, hence, the results in Section 21.4 remain intact.

Instead of the weighted least squares method in Section 21.4, we may use some other robust methods, too. For example, keep in mind the linear model in (21.35); it is quite conceivable to use M-estimators of the $\boldsymbol{\Gamma}_t$. Recursive M-estimators in the change point model have already been studied by Sen (1984). For some further developments in this direction, we refer interested readers to Hušková and Sen (1986) and to Chapter 6 of this volume.

Acknowledgment

This work was supported by the Office of Naval Research, Contract No. N00014-83-K-0387.

References

Brown, R.L., Durbin, J., and Evans, J.M. (1975), Techniques for testing the constancy of regression relationships over time (with discussion). *Journal of the Royal Statistical Society*, **B-37**, 149-192.

Chernoff, H. and Zacks, S. (1964), Estimating the current mean of a normal distribution which is subjected to changes in time. *Annals of Mathematical Statistics*, **35**, 999-1018.

DeLong, D.M. (1981), Crossing probabilities for a square root boundary by a Bessel process. *Communications in Statistics, Theory and Methods*, **A10**, 2197-2213.

Hoeffding, W. (1948), A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics*, **19**, 293-325.

Hušková, M. and Sen, P.K. (1986), Nonparametric tests for shift and change in regression at an unknown time point. Paper presented at the IIASA meeting, Sulejow, Poland.

Johnson, N.L. and Kotz, S. (1970), Distributions in Statistics: Continuous Univariate Distributions – 1. New York: John Wiley.

Sen, A.K. (1976), The measurement of poverty: An axiomatic approach. *Econometrica*, **44**, 219-232.

Sen, P.K. (1977), Some invariance principles relating to jackknifing and their role in sequential analysis. *Annals of Statistics*, **5**, 315-329.

Sen, P.K. (1982), Invariance principles for recursive residuals. *Annals of Statistics*, **10**, 327-332.

Sen, P.K. (1984), Recursive M-tests for the constancy of regression relationship over time. *Sequential Analysis*, **3**, 151-211.

Sen, P.K. (1985), *Theory and Applications of Sequential Nonparametrics*. Philadelphia, PA: Siam.

Sen, P.K. (1986), The Gini coefficient and poverty indexes: Some reconciliations. *Journal of the American Statistical Association*, **81**, 1050-1057.

Takayama, N. (1979), Poverty, income inequality and their measures: Professor Sen's axiomatic approach reconsidered. *Econometrica*, **47**, 747-760.

CHAPTER 22

Statistical Identification of Nonlinear Dynamics in Macroeconomics Using Nonlinear Time Series Models

Tohru Ozaki and Valerie H. Ozaki

Summary

A mathematical model is introduced to explain the dynamics of the Hicksian IS-LM paradigm, in which the difference between the attitudes of Keynesians and monetarists is representable by a difference in the parameters of the model. Statistical identification procedures are introduced for both this model and for the nonstationary model of a time-varying Hicksian IS-LM structure. Application of the models to simulation data is also discussed, and numerical results are given.

22.1 Introduction

It is well known in macroeconomics that there is a sharp conflict between "monetarists" and "Keynesians". The difference between the two "schools" is generally held to center on whether money supply or fiscal variables are the major determinants of aggregate economic activity and, hence, on the most appropriate tool of stabilization policies (Morgan, 1978). The two groups also differ over the question of whether stabilization policies are needed (Modigliani, 1977).

Several explanations have been presented to justify Keynesians' or monetarists' views on macroeconomic policy. These explanations are, as seen in the succeeding section, mostly based on the Hicksian IS-LM paradigm. It is a pity that the monetarists and Keynesians

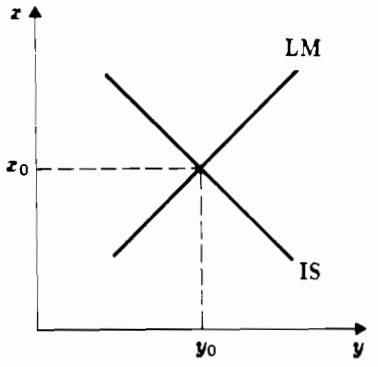

Figure 22.1: IS and LM curves.

use different assumptions about the IS-LM curves to come to different conclusions about effective economic policy. It would be useful if the rationality of these assumptions could be checked by some statistical method on some historical economic data.

Our purpose in this chapter is to introduce a dynamic model for the IS-LM paradigm, and an identification method for this model. With the identification method, we can estimate the IS-LM curves and the strength of the stabilizing power of the equilibrium point of the IS-LM paradigm. A Bayesian method of estimating IS-LM curve changes over time is also presented, together with some numerical results.

22.2 Dynamic Modeling of the Hicksian IS-LM Paradigm

The Hicksian IS-LM paradigm is a model that explains the dynamics of interest rate and output. The IS curve shows the relation between the equilibrium output in the goods market and interest rate. The LM curve shows the relation between the equilibrium of interest rate and output. The intersection of the two curves is the equilibrium point of both the output and interest rate (see *Figure 22.1*).

The difference between Keynesian's and monetarist's macroeconomic policies comes from different assumptions about the IS-LM curves. Keynesians assume thet the slope of the IS curve is large and almost vertical, while the slope of the LM curve is small and almost flat. Therefore, shifting the LM curve by increasing money supply is not effective in increasing output. On the other hand, shifting the IS curve by the fiscal policy of government spending is effective in increasing the equilibrium output (see *Figure 22.2*). Monetarists start from opposite assumptions: the slope of the LM curve is almost vertical, while the slope of the IS curve is almost flat. In this situation, fiscal policy is effective in increasing it (see *Figure 22.3*).

Unfortunately, there has been no objective method to check which of these contradictory assumptions are more realistic and preferable. This may be because there is no model that explicitly specifies the dynamics of IS-LM paradigm in mathematical form. If such a

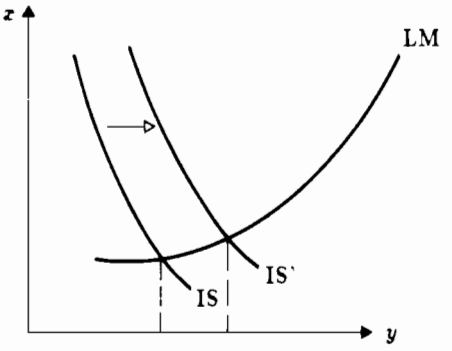

Figure 22.2: Keynesians' IS-LM curves.

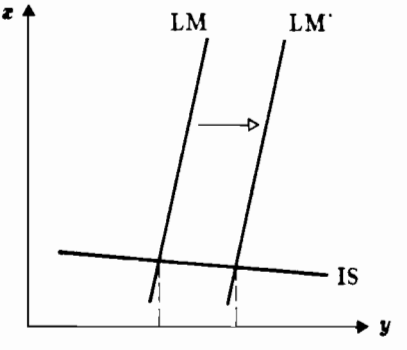

Figure 22.3: Monetarists' IS-LM curves.

mathematical model were given, we could use a statistical method to identify the model from the time series data of interest rate and output. To provide a method of checking the appropriateness of the competing assumptions about the IS-LM curve, we first introduce a mathematical model of the IS-LM paradigm.

The dynamice of the interest rate y and output x are expressed by the following dynamic system

$$\begin{aligned} \dot{x} &= f_1(x,y)\,x \\ \dot{y} &= f_2(x,y)\,y \end{aligned} \tag{22.1}$$

where $f_1(x,y)$ and $f_2(x,y)$ are some functions of x and y. At the equilibrium point of the interest rate, we have $\dot{y}=0$ for which $f_2(x,y)=0$. This relation between x and y gives us the LM curve. From the equilibrium condition of the output, $\dot{x}=0$, for which $f_1(x,y)=0$, we have the IS curve. Naturally, the intersection of $f_1(x,y)=0$ and $f_2(x,y)=0$ gives the equilibrium point of both the variables. The simplest functional form of $f_1(x,y)$ and $f_2(x,y)$ is the linear function. For example, let

$$\begin{aligned} f_1(x,y) &= a_1 + a_2 x + a_3 y \\ f_2(x,y) &= b_1 + b_2 x + b_3 y \end{aligned}$$

It can be proved that the equilibrium point (x,y) of the dynamical system (22.1) given by the intersection of the two lines, $a_1+a_2x+a_3y=0$ and $b_1+b_2x+b_3y=0$, is a stable equilibrium point (Hirsh and Smale, 1974). This means the trajectory of (x,y) starting from a point near (x_∞, y_∞) converges to (x_∞, y_∞) for $\Delta t \to \infty$.

In macroeconomics it may not be realistic to think that the system is deterministic — that the future behavior of the system is completely determined by its initial state. It may be more realistic to think of the system being disturbed by external, unpredictable, random shocks. This uncertainty is realized by making the system into a stochastic dynamical system by adding a random white noise. For example, let

$$\begin{aligned} \dot{x} &= f_1(x,y)\,x + n_1(t) \\ \dot{y} &= f_2(x,y)\,y + n_2(t) \end{aligned} \tag{22.2}$$

where $\mathbf{n}(t) = [n_1(t), n_2(t)]'$ is a bivariate Gaussian white noise whose variance-covariance matrix is $\Sigma = (\sigma_{i,j})$. The stochastic process $(x,y)'$ defined by (22.2) is a Markov diffusion process whose transition probability $p(\mathbf{z}|\mathbf{z}_0,t)$, from the state $\mathbf{z}_0$ at $t=0$ to the state $\mathbf{z}=(x,y)'$ at t, satisfies the following Fokker-Plank equation

$$\frac{\partial p(\mathbf{z}|\mathbf{z}_0,t)}{\partial t} = -\sum_{i=1}^{2} \frac{\partial}{\partial z_i}[f_i(\mathbf{z})\,p] + \frac{1}{2}\sum_{i,j=1}^{2} \frac{\partial^2}{\partial z_i \partial z_j}[\sigma_{i,j}\,p]$$

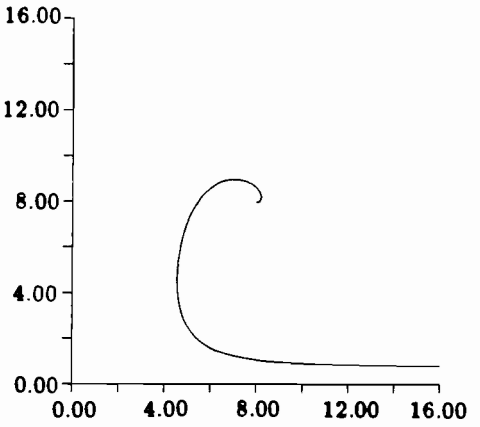

Figure 22.4: Trajectory $[x(t), y(t)]$ of example 1.

We can see how the dynamical systems (22.1) and (22.2) work as the model of the IS-LM paradigm in the following examples.

Example 1: $\dot{x} = (y - x)x$
$\dot{y} = (20 - x - y)y$

The equilibrium point of the model is $(10, 10)$. The trajectory of the model, starting from the initial state $\mathbf{z_0} = (20, 1)$, converges to the equilibrium point, as is seen in *Figure 22.4*.

Example 2: $\dot{x} = (20 - y - x)x$
$\dot{y} = (x - y)y$

The equilibrium point of the model is also $(10, 10)$. The trajectory, starting from the initial state $\mathbf{z}_0 = (1, 20)$, converges to the equilibrium point $(10, 10)$, but in a different way from the previous example (see *Figure 22.5*). When these systems have white noise input, as in model (22.2), we have trajectories that fluctuate around the stable equilibrium point $(10, 10)$ (see *Figure 22.6* and *22.7*).

22.3 Time Discretization

The difference between the assumptions of Keynesians and monetarists on the slopes of the IS and LM curves and the strength of the stabilization power of the equilibrium point can be realized by the difference in the parameters, a_1, a_2, a_3, b_1, b_2, and b_3 of the following model,

$$\begin{aligned} \dot{x} &= (a_1 + a_2 x + a_3 y)\, x + n_1(t) \\ \dot{y} &= (b_1 + b_2 x + b_3 y)\, x + n_2(t) \end{aligned} \tag{22.3}$$

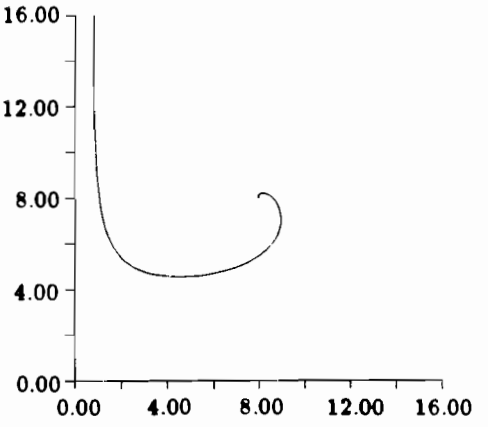

Figure 22.5: Trajectory $[x(t), y(t)]$ of example 2.

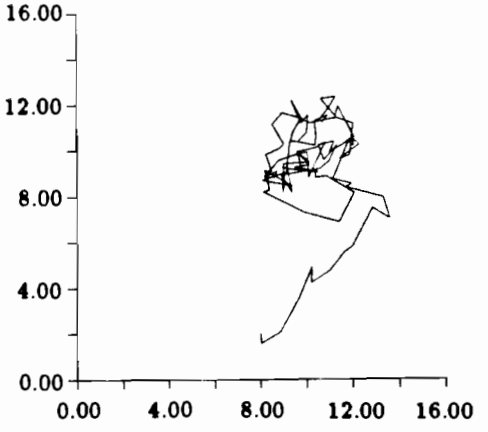

Figure 22.6: Trajectory $[x(t), y(t)]$ of the model of example 1 with white noise input.

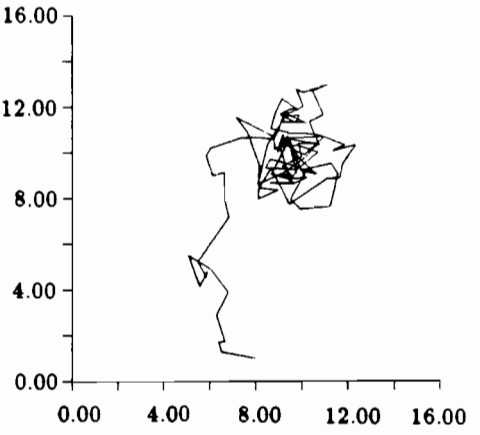

Figure 22.7: Trajectory $[x(t), y(t)]$ of the model of example 2 with white noise input.

It is well known that the estimation of parameters of a continuous-time stochastic dynamical system is not easy, except for linear cases. On the other hand, the estimation of the parameters of a discrete-time stochastic dynamical system, which we call a nonlinear time series model, is easily obtained by applying the maximum likelihood method. Our idea for estimating the parameters of a continuous-time dynamical system is to derive a discrete-time version of the continuous-time model and use the maximum likelihood method for the estimation of the discrete-time model. For this purpose we rewrite model (22.3) in vector form as follows:

$$\dot{\mathbf{z}} = \mathbf{f}(\mathbf{z}|\mathbf{a}, \mathbf{b}) + \mathbf{n}(t) \tag{22.4}$$

where $\mathbf{z} = (x, y)'$, $\mathbf{f}(\mathbf{z}|\mathbf{a}, \mathbf{b}) = [-f_1(\mathbf{z}|\mathbf{a}), -f_2(\mathbf{z}|\mathbf{b})]'$, and $\mathbf{n}(t) = [n_1(t), n_2(t)]'$. We assume that the white noise processes $n_1(t)$ and $n_2(t)$ are independent and their variances are equal to σ^2.

It is shown in Ozaki (1986) that when we assume that the dynamical system is locally linear, i.e., the Jacobian matrix

$$\left[\frac{\partial \mathbf{f}(\mathbf{z}|\mathbf{a}, \mathbf{b})}{\partial \mathbf{z}}\right]_{\mathbf{z}=\mathbf{z}_t} = \mathbf{J}_t$$

is constant on each short interval $[t,\ t + \Delta t)$, we have, from (22.4), the following discrete-time model:

$$\mathbf{z}_{t+\Delta t} = \exp\{\mathbf{K}_t \Delta t\}\mathbf{z}_t + \mathbf{w}_{t+\Delta t} \tag{22.5}$$

where $\mathbf{w}_{t+\Delta t} = \int_t^{t+\Delta t} \exp\{\mathbf{K}_t(t + \Delta t - u)\}\mathbf{n}(u)du$, $\mathbf{K}_t = \frac{1}{\Delta t}\log \mathbf{A}_t$, and $\mathbf{A}_t = \mathbf{I} + \mathbf{J}_t^{-1}$ $\times(\exp\{\mathbf{J}_t \Delta t\} - \mathbf{I})\mathbf{F}_t$; $\mathbf{I}$ is a unit matrix and $\mathbf{F}_t$ is a matrix satisfying $\mathbf{F}_t\mathbf{z} = \mathbf{f}(\mathbf{z}_t|\mathbf{a}, \mathbf{b})$.

The matrices $\exp\{\mathbf{K}_t(\Delta t)\}$ and $\log \mathbf{A}_t$ are defined respectively by $\exp\{\mathbf{K}_t\Delta t\} = \mathbf{I} + \sum_{i=1}^{\infty} \frac{(\mathbf{K}_t\Delta t)^i}{i!}$, $\log \mathbf{A}_t = \sum_{i=1}^{\infty} \frac{(-1)^i}{i!}(\mathbf{A}_t - \mathbf{I})^i$. The variance-covariance matrix of $\mathbf{w}_{t+\Delta t}$ is

$$
\begin{aligned}
& E[\int_t^{t+\Delta t} \exp\{\mathbf{K}_t(t+\Delta t-u)\}\,\mathbf{n}(u)du \int_t^{t+\Delta t} \mathbf{n}'(u)\exp\{\mathbf{K}_t'(t+\Delta t-u)\}du] \\
&= \int_t^{t+\Delta t} \exp\{\mathbf{K}_t(t+\Delta t-u)\} \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \exp\{\mathbf{K}_t'(t+\Delta t-u)\}du \\
&= \sigma^2 \mathbf{V}_t \begin{pmatrix} e_{11}s_{11} & e_{12}s_{12} \\ e_{21}s_{21} & e_{22}s_{22} \end{pmatrix} \mathbf{V}_t'
\end{aligned}
$$

where $\mathbf{V}_t$ is a matrix which satisfies $\mathbf{V}_t^{-1}\mathbf{K}_t\mathbf{V}_t = \begin{pmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{pmatrix}$ and is given by

$$
\mathbf{V}_t = \begin{pmatrix} \mu_1 & \mu_2 \\ 1 & 1 \end{pmatrix} \tag{22.6}
$$

The elements $e_{11}, \ldots, e_{22}$, $s_{11}, \ldots$, and s_{22} are defined by

$$
\begin{aligned}
e_{11} &= \frac{\exp\{2\mu_1\Delta t\} - 1}{2\mu_1} \\
e_{12} &= \frac{\exp\{(\mu_1+\mu_2)\Delta t\} - 1}{\mu_1+\mu_2} = e_{21} \\
e_{22} &= \frac{\exp\{2\mu_2\Delta t\} - 1}{2\mu_2} \\
s_{11} &= \frac{1+\mu_2^2}{(\mu_1-\mu_2)^2} \\
s_{12} &= \frac{1+\mu_1\mu_2}{(\mu_1+\mu_2)^2} = s_{21} \\
s_{11} &= \frac{1+\mu_1^2}{(\mu_1-\mu_2)^2}
\end{aligned}
$$

where λ_1, λ_2 are roots of the polynomial $\Lambda^2 - (e_{11}s_{11} + e_{22}s_{22})\Lambda + e_{11}s_{11}e_{22}s_{22} - e_{12}^2 s_{12}^2 = 0$. With $\mathbf{V}_t$ given in (22.6) and

$$
\mathbf{U}_t = \frac{1}{\sqrt{e_{12}^2 s_{12}^2 + (\lambda_1 - e_{11}s_{11})^2}} \begin{pmatrix} e_{12}s_{12} & \lambda_2 - e_{22}s_{22} \\ \lambda_i - e_{11}s_{11} & e_{12}s_{12} \end{pmatrix} \tag{22.7}
$$

the variance-covariance matrix of $w_{t+\Delta t}$ is

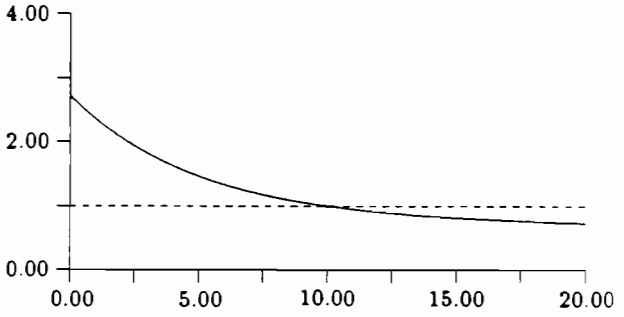

Figure 22.8: Function $A_{11}(x, \star)$.

$$\sigma^2 \mathbf{V}_t \mathbf{U}_t \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \mathbf{U}_t' \mathbf{V}_t'$$

From (22.5) we have the following nonlinear bivariate time series model:

$$\mathbf{z}_{t+\Delta t} = \mathbf{A}_t \mathbf{z}_t + \mathbf{B}_t \mathbf{n}_{t+\Delta t} \tag{22.8}$$

where

$$\mathbf{A}_t = \exp\{\mathbf{K}_t \Delta t\}\ , \quad \mathbf{B}_t = \mathbf{V}_t \mathbf{U}_t \begin{pmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{pmatrix}$$

and $\mathbf{n}_t$ is a discrete-time bivariate white noise with variance-covariance matrix $\sigma^2 \mathbf{I}$.

It is well known that most discretization schemas, although they are consistent, give us an unstable model for a fixed Δt. The present scheme gives us a stable discrete-time model from a stable continuous-time model (Ozaki, 1985b). When Δt is small, $\mathbf{B}_t$ is almost constant and we have $\mathbf{B}_t = \sqrt{\Delta t}\, \mathbf{I}$. $\mathbf{A}_t$ is very different from a constant matrix. We can see how the matrix $\mathbf{A}_t$ changes as a function of x_t and y_t using example 1 of the previous section. *Figure 22.8* shows the $(1,1)$-element $A_{11}(x_t, y_t)$ as a function of x_t for a fixed $y_t = 10$. *Figure 22.9* shows $A_{11}(x_t, y_t)$ as a function of y_t for a fixed $x_t = 10$. $A_{12}(x_t, y_t)$ and $A_{21}(x_t, y_t)$ are very small, compared with other elements, and can be ignored. The figures of $A_{22}(x_t, y_t)$ as a function of x_t and as a function of y_t are shown in *Figures 22.10* and *22.11*, respectively.

The information on the behavior of the matrix elements as a function of x_t and y_t are useful for the parameterization of the model in the discrete-time domain in the next section.

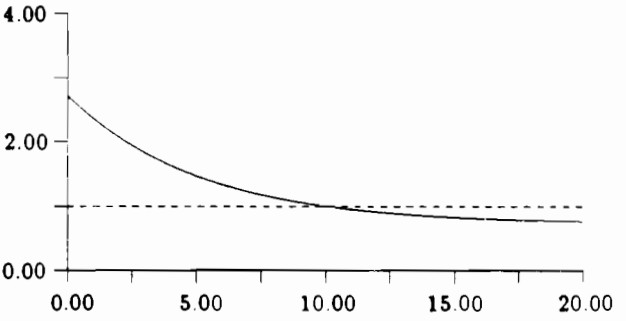

Figure 22.9: Function $A_{11}(\star, y)$.

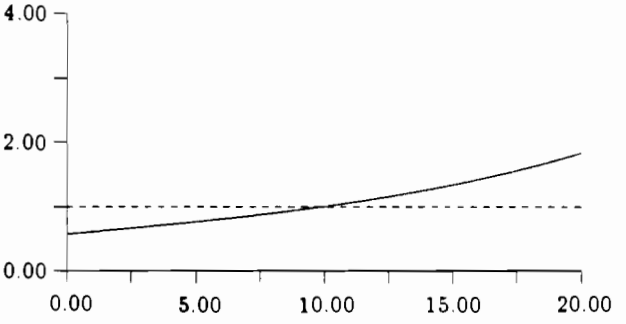

Figure 22.10: Function $A_{22}(x, \star)$.

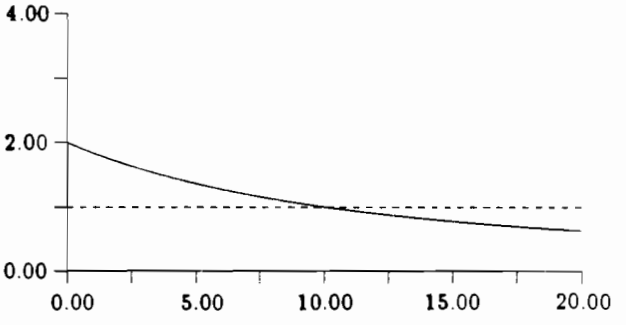

Figure 22.11: Function $A_{22}(\star, y)$.

22.4 Model Identification

To identify the stochastic dynamical system model (22.4), we first try to estimate the parameters of the discrete-time model (22.8) from the data, $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N$, where $\mathbf{z}_i = (x_i, y_i)'$. The log-likelihood of the model (22.8) is given as follows.

$$\begin{aligned}
&\log p(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N | \mathbf{a}, \mathbf{b}, \sigma^2) \qquad (22.9)\\
&= \log p(\mathbf{z}_2, \ldots, \mathbf{z}_N | \mathbf{z}_1, \mathbf{a}, \mathbf{b}, \sigma^2) + \log p(\mathbf{z}_1 | \mathbf{a}, \mathbf{b}, \sigma^2)\\
&= \log p(\mathbf{n}_2, \ldots, \mathbf{n}_N | \mathbf{z}_1, \mathbf{a}, \mathbf{b}, \sigma^2) + \log J(\mathbf{z}, \mathbf{n}) + \log p(\mathbf{z}_1 | \mathbf{a}, \mathbf{b}, \sigma^2)\\
&= -\sum_{t=2}^{N} \frac{\|\mathbf{B}_t^{-1}(\mathbf{z}_{t+1} - \mathbf{A}_t \mathbf{z}_t)\|^2}{2\sigma^2} - \frac{N-1}{2} \log |\sigma^2 \mathbf{I}| - \sum_{t=2}^{N} \log |\mathbf{B}_t| + \log p(\mathbf{z}_1 | \mathbf{a}, \mathbf{b}, \sigma^2)
\end{aligned}$$

where $J(\mathbf{z}, \mathbf{n})$ is the Jacobian of the transformation from $(\mathbf{z}_2, \ldots, \mathbf{z}_N)$ to $(\mathbf{n}_2, \ldots, \mathbf{n}_N)$, and $\|.\|$ means the Euclidian norm. For large N, the last term is negligibly small, compared with the rest, and can be ignored. Since the maximum likelihood estimate $\hat{\sigma}_N^2$ satisfies

$$\left[\frac{\partial \log p(\mathbf{z}_1, \ldots, \mathbf{z}_N | \mathbf{a}, \mathbf{b}, \sigma^2)}{\partial \sigma^2} \right]_{\sigma^2 = \hat{\sigma}_N^2} = 0$$

we have $\hat{\sigma}_N^2 = \min_{\mathbf{a},\mathbf{b}} \sigma^2(\mathbf{a}, \mathbf{b}|\mathbf{z})$, where $\sigma^2(\mathbf{a}, \mathbf{b}|\mathbf{z}) = \frac{1}{2(N-1)} \sum_{t=2}^{N} \|\mathbf{B}_t^{-1}(\mathbf{z}_{t+1} - \mathbf{A}_t \mathbf{z}_t)\|^2$. To minimize $\sigma^2(\mathbf{a}, \mathbf{b}|\mathbf{z})$ with respect to $\mathbf{a}$ and $\mathbf{b}$, some nonlinear numerical optimization procedure is needed (see, for example, Fletcher and Powell, 1963).

To see if the maximum likelihood method works numerically, we performed some simulation studies. We simulated the model (22.4) with the following parameters: $a_1 = 10$, $a_2 = -1$, $a_3 = -1$, $b_1 = 0$, $b_2 = 1$, and $b_3 = -1$. The variances of the white noise, $n_1(t)$ and $n_2(t)$, are equal to 1. We generated 400 data points $(x_1, y_1)', \ldots, (x_{400}, y_{400})'$, which are plotted in *Figure 22.12.*

The maximum likelihood estimates obtained by applying the above-mentioned method are $\hat{a}_1 = 9.8882$, $\hat{a}_2 = -0.9856$, $\hat{a}_3 = -0.9970$, $\hat{b}_1 = 0.1893$, $\hat{b}_2 = 0.9738$, $\hat{b}_3 = -0.9261$, and $\hat{\sigma}^2 = 1.0161$.

By the above maximum likelihood method, we obtained the maximum likelihood estimates $\hat{\mathbf{a}}$, $\hat{\mathbf{b}}$, and $\hat{\sigma}^2$ for the model $\mathbf{z}_{t+1} = \mathbf{A}_{\Delta t}(\mathbf{z}_t|\mathbf{a}, \mathbf{b})\mathbf{z}_t + \mathbf{B}_{\Delta t}(\mathbf{z}_t|\mathbf{a}, \mathbf{b})\mathbf{n}_{t+1}$ but *not* for $\dot{\mathbf{z}} = \mathbf{f}(\mathbf{z}|\mathbf{a}, \mathbf{b}) + \mathbf{n}(t)$. That means the nonlinear function that we have obtained by the maximum likelihood method is not $\mathbf{f}(\mathbf{z}|\hat{\mathbf{a}}, \hat{\mathbf{b}})$ but some function $\mathbf{f}_{\Delta t}(\mathbf{z}|\hat{\mathbf{a}}, \hat{\mathbf{b}})$, which satisfies the local linearization relationship

$$\mathbf{f}_{\Delta t}(\mathbf{z}|\hat{\mathbf{a}}, \hat{\mathbf{b}}) = \left[\exp\left\{ \frac{\partial \mathbf{f}_{\Delta t}(\mathbf{z}|\hat{\mathbf{a}}, \hat{\mathbf{b}})}{\partial \mathbf{z}} \Delta t \right\} - \mathbf{I} \right]^{-1} \frac{\partial \mathbf{f}_{\Delta t}(\mathbf{z}|\hat{\mathbf{a}}, \hat{\mathbf{b}})}{\partial \mathbf{z}} [\mathbf{A}_{\Delta t}(\mathbf{z}|\hat{\mathbf{a}}, \hat{\mathbf{b}}) - \mathbf{I}]\mathbf{z} \quad (22.10)$$

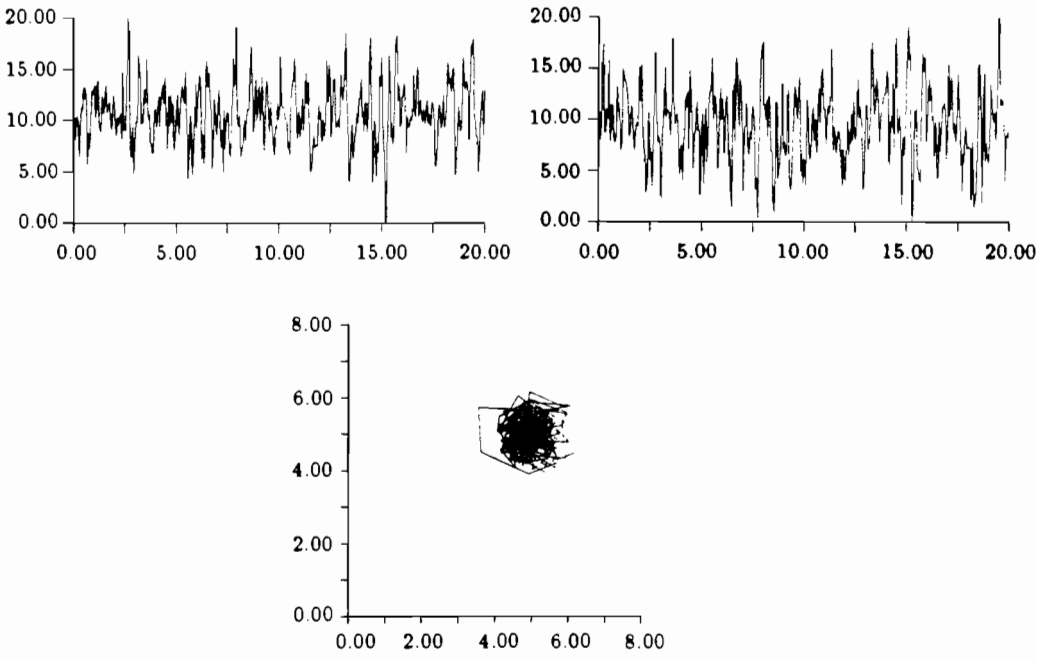

Figure 22.12: Simulated data $(x_1, y_1), \ldots, (x_{400}, y_{400})$.

with $\mathbf{A}_{\Delta t}(\mathbf{z}|\hat{\mathbf{a}}, \hat{\mathbf{b}})$. In general, we cannot give an explicit form for $\mathbf{f}_{\Delta t}(\mathbf{z}|\mathbf{a}, \mathbf{b})$. Instead, we can obtain $\mathbf{f}_{\Delta t}(\mathbf{z}|\mathbf{a}, \mathbf{b})$ by the following numerical iterative procedure:

$$\begin{aligned} &\mathbf{f}_{\Delta t}^{(h+1)}(\mathbf{z}_{i,j}|\mathbf{a}, \mathbf{b}) \qquad (22.11) \\ &\quad = (1 - \Delta s)\, \mathbf{f}_{\Delta t}^{(h)}(\mathbf{z}_{i,j}|\mathbf{a}, \mathbf{b}) + \Delta s[\exp\{J_{i,j}\Delta t\} - \mathbf{I}]^{-1}[\mathbf{A}_{\Delta t}(\mathbf{z}_{i,j}|\mathbf{a}, \mathbf{b}) - \mathbf{I}]\mathbf{z}_{i,j} \end{aligned}$$

for each point $\mathbf{z}_{i,j} = (x_i, y_j)'$, $(i = 1, 2, \ldots, j = 1, 2, \ldots)$ on any finite region. This is because $\mathbf{f}_{\Delta t}(\mathbf{z}_{i,j}|\mathbf{a}, \mathbf{b})$ can be considered as a limiting function $\mathbf{f}_{\Delta t}(\mathbf{z}, \infty|\mathbf{a}, \mathbf{b})$, characterized by

$$\frac{\partial \mathbf{f}_{\Delta t}(\mathbf{z}, s|\mathbf{a}, \mathbf{b})}{\partial s} = 0$$

of a deterministic spatial process $\mathbf{f}_{\Delta t}(\mathbf{z}, s|\mathbf{a}, \mathbf{b})$, defined by the partial differential equation

$$\begin{aligned} &\frac{\partial \mathbf{f}_{\Delta t}(\mathbf{z}, s|\mathbf{a}, \mathbf{b})}{\partial s} = -\mathbf{f}_{\Delta t}(\mathbf{z}, s|\mathbf{a}, \mathbf{b}) \qquad (22.12) \\ &\quad + \left[\exp\{\frac{\partial \mathbf{f}_{\Delta t}(\mathbf{z}, s|\mathbf{a}, \mathbf{b})}{\partial \mathbf{z}}\Delta t\} - \mathbf{I}\right]^{-1} \left[\frac{\partial \mathbf{f}_{\Delta t}(\mathbf{z}, s|\mathbf{a}, \mathbf{b})}{\partial \mathbf{z}}\right] [\mathbf{A}_{\Delta t}(\mathbf{z}|\mathbf{a}, \mathbf{b}) - \mathbf{I}]\mathbf{z} \end{aligned}$$

$J_{i,j}$ is a matrix obtained by approximating the Jacobian matrix

$$\left[\frac{\partial \mathbf{f}_{\Delta t}(\mathbf{z}|\mathbf{a},\mathbf{b})}{\partial \mathbf{z}}\right]_{\mathbf{z}=(x_i,y_j)'}$$

by some numerical differencing method with $\Delta\mathbf{z} = (\Delta x, \Delta y)'$. We have $\mathbf{f}_{\Delta t}^{(h+1)}(\mathbf{z}_{i,j}|\mathbf{a},\mathbf{b}) \to \mathbf{f}_{\Delta t}(\mathbf{z}_{i,j}|\mathbf{a},\mathbf{b})$ for $k \to \infty$ if Δs of (22.11) is sufficiently small, compared with Δx and Δy. For the initial function of the above iteration, the following function

$$\mathbf{f}_{\Delta t}^{(0)}(\mathbf{z}|\mathbf{a},\mathbf{b}) = \frac{1}{\Delta t}[\mathbf{A}_{\Delta t}(\mathbf{z}|\mathbf{a},\mathbf{b}) - \mathbf{I}]\mathbf{z}$$

which is obtained from the local linearization relationship (22.10) by employing the approximation

$$\exp\left\{\frac{\partial \mathbf{f}_{\Delta t}(\mathbf{z}|\mathbf{a},\mathbf{b})}{\partial \mathbf{z}}\Delta t\right\} \approx \mathbf{I} + \frac{\partial \mathbf{f}_{\Delta t}(\mathbf{z}|\mathbf{a},\mathbf{b})}{\partial \mathbf{z}}\Delta t$$

will be useful.

Of course, the function $\mathbf{f}_{\Delta t}(\mathbf{z}|\mathbf{a},\mathbf{b})$ obtained from $\mathbf{f}(\mathbf{z}|\mathbf{a},\mathbf{b})$ through the locally linearized $\mathbf{A}_{\Delta t}(\mathbf{z}|\mathbf{a},\mathbf{b})$ using

$$\mathbf{A}_{\Delta t}(\mathbf{z}|\mathbf{a},\mathbf{b}) = \mathbf{I} + \left[\frac{\partial \mathbf{f}(\mathbf{z}|\mathbf{a},\mathbf{b})}{\partial \mathbf{z}}\right]^{-1}\left[\exp\left\{\frac{\partial \mathbf{f}(\mathbf{z}|\mathbf{a},\mathbf{b})}{\partial \mathbf{z}}\Delta t\right\} - \mathbf{I}\right]F(\mathbf{z}|\mathbf{a},\mathbf{b})$$

is not equal to the original $\mathbf{f}(\mathbf{z}|\mathbf{a},\mathbf{b})$ since the local linearity approximation is used in deriving $\mathbf{A}_{\Delta t}(\mathbf{z}|\mathbf{a},\mathbf{b})$ from $\mathbf{f}(\mathbf{z}|\mathbf{a},\mathbf{b})$. We can easily see how fast $\mathbf{f}_{\Delta t}(\mathbf{z}|\mathbf{a},\mathbf{b})$ approaches the true $\mathbf{f}(\mathbf{z}|\mathbf{a},\mathbf{b})$ for $\Delta t \to 0$ by employing the above numerical procedure (see Ozaki, 1985a).

The above-mentioned numerical procedure suggests another possible method of identifying $\mathbf{f}(\mathbf{z}|\mathbf{a},\mathbf{b})$. Instead of giving a continuous-time parametric model, $\dot{\mathbf{z}} = \mathbf{f}(\mathbf{z}|\mathbf{a},\mathbf{b}) + \mathbf{n}(t)$ we can give a discrete-time model parametrization thus:

$$\mathbf{z}_{t+\Delta} = \mathbf{A}_{\Delta t}(\mathbf{z}_t|\varphi)\mathbf{z}_t + \mathbf{B}_{\Delta t}(\mathbf{z}_t|\varphi)\mathbf{n}_{t+\Delta t} \tag{22.13}$$

where $\mathbf{A}_{\Delta t}(\mathbf{z}_t|\varphi)$ is some matrix function, with parameter vector $\varphi = (\varphi_1, \varphi_2, \ldots, \varphi_k)'$. $\mathbf{B}_{\Delta t}(\mathbf{z}_t|\varphi)$ is determined from $\mathbf{A}_{\Delta t}(\mathbf{z}_t|\varphi)$ as follows:

$$\mathbf{B}_{\Delta t}(\mathbf{z}_t|\varphi) = \mathbf{V}_t(\mathbf{z}_t|\varphi)\mathbf{U}_t(\mathbf{z}_t|\varphi)\begin{pmatrix} \sqrt{\lambda_1(\mathbf{z}_t|\varphi)} & 0 \\ 0 & \sqrt{\lambda_2(\mathbf{z}_t|\varphi)} \end{pmatrix}$$

$\mathbf{V}_t(\mathbf{z}_t|\varphi)$, $\mathbf{U}_t(\mathbf{z}_t|\varphi)$, $\lambda_1(\mathbf{z}_t|\varphi)$, and $\lambda_2(\mathbf{z}_t|\varphi)$ are functions of $e_{i,j}$ and $s_{i,j}$ ($i = 1,2$ and $j = 1,2$), which in turn are functions of the eigenvalues $\mu_1(\mathbf{z}_t|\varphi)$ and $\mu_2(\mathbf{z}_t|\varphi)$ of the matrix $\log \mathbf{A}_{\Delta t}(\mathbf{z}_t|\varphi)$, as in the previous section. We note $\log \mathbf{A}_{\Delta t}(\mathbf{z}_t|\varphi)$ is well-defined when $\|\mathbf{A}_{\Delta t}(\mathbf{z}_t|\varphi) - \mathbf{I}\|$ is sufficiently small, which is always true when the data are sampled from the original continuous-time process with a sufficiently small sampling interval. For the parameterization of $\mathbf{A}_{\Delta t}(\mathbf{z}_t|\varphi)$, the behavior of the functions $A_{11}(x_t, y_t)$, $A_{12}(x_t, y_t)$, $A_{21}(x_t, y_t)$, and $A_{22}(x_t, y_t)$ of $\mathbf{A}_{\Delta t}(\mathbf{z}_t|\mathbf{a}, \mathbf{b})$, described at the end of the previous section, gives us useful information. $A_{12}(x_t, y_t)$ and $A_{21}(x_t, y_t)$ for the model (22.3) are almost zero compared with $A_{11}(x_t, y_t)$ and $A_{22}(x_t, y_t)$, which are smooth functions of x_t and y_t (see *Figure 22.8* to *22.11*. Therefore a reasonable parameterization would be

$$\begin{aligned} A_{11}(x_t, y_t) &= \pi_1 + \pi_2 x_t + \pi_3 y_t \\ A_{22}(x_t, y_t) &= \theta_1 + \theta_2 x_t + \theta_3 y_t \end{aligned} \tag{22.14}$$

or

$$\begin{aligned} A_{11}(x_t, y_t) &= \pi_1 + \pi_2 x_t^2 + \pi_3 y_t^2 \\ A_{22}(x_t, y_t) &= \theta_1 + \theta_2 x_t^2 + \theta_3 y_t^2 \end{aligned} \tag{22.15}$$

or

$$\begin{aligned} A_{11}(x_t, y_t) &= \pi_1 + \pi_2 e^{-x_t} + \pi_3 e^{-y_t} \\ A_{22}(x_t, y_t) &= \theta_1 + \theta_2 e^{-x_t} + \theta_3 e^{-y_t} \end{aligned} \tag{22.16}$$

or

$$\begin{aligned} A_{11}(x_t, y_t) &= \pi_1 + \pi_2 e^{-x_t^2} + \pi_3 e^{-y_t^2} \\ A_{22}(x_t, y_t) &= \theta_1 + \theta_2 e^{-x_t^2} + \theta_3 e^{-y_t^2} \end{aligned} \tag{22.17}$$

By applying the least squares estimation method, which is asymptotically equivalent to the maximum likelihood method, to the data of *Figure 22.1*, with the parameterization (22.14) we obtained the estimates $\hat{\pi}_1 = 1.7965$, $\hat{\pi}_2 = -0.0947$, $\hat{\pi}_3 = -0.0648$, $\hat{\theta}_1 = 0.9602$, $\hat{\theta}_2 = 0.0865$, $\hat{\theta}_3 = -0.0781$, and $\hat{\sigma}^2 = 0.0996$. For the matrix

$$\mathbf{A}_{\Delta t}(x_t, y_t|\hat{\pi}, \hat{\theta}) = \begin{pmatrix} \hat{\pi}_1 + \hat{\pi}_2 x_t + \hat{\pi}_3 y_t & 0 \\ 0 & \hat{\theta}_1 + \hat{\theta}_2 x_t + \hat{\theta}_3 y_t \end{pmatrix}$$

we have the approximate relation $\hat{\mathbf{f}}_{\Delta t}(x_t, y_t) \approx \frac{1}{\Delta t}[\mathbf{A}_{\Delta t}(x_t, y_t|\hat{\pi}, \hat{\theta}) - \mathbf{I}](x_t, y_t)'$.

From this relation, approximate IS-LM curves are obtained by

$$\begin{aligned} f_1(x_t, y_t) &= \frac{1}{\Delta t}(\hat{\pi}_1 - 1 + \hat{\pi}_2 x_t + \hat{\pi}_3 y_t)\, x_t = 0 \\ f_2(x_t, y_t) &= \frac{1}{\Delta t}(\hat{\theta}_1 - 1 + \hat{\theta}_2 x_t + \hat{\theta}_3 y_t)\, y_t = 0 \end{aligned}$$

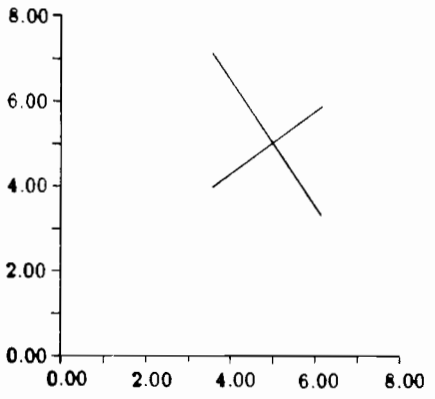

Figure 22.13: Estimated IS-LM curves by the model (14).

which are shown in *Figure 22.13*, where we see very close estimates to the original true lines. Using the parameterization (22.15), and applying the same least squares estimation method, we obtained the following estimates: $\hat{\pi}_1 = 1.3946$, $\hat{\pi}_2 = -0.0092$, $\hat{\pi}_3 = -0.0065$, $\hat{\theta}_1 = 0.9748$, $\hat{\theta}_2 = 0.0088$, $\hat{\theta}_3 = -0.0077$, and $\hat{\sigma}^2 = 0.0996$. From the estimates we obtained the approximate IS-LM curves, which are also close to the true curves (see *Figure 22.14*).

Since $\mathbf{B}_t(\mathbf{z}_t|\varphi)$ is almost constant and equal to $\sqrt{\Delta t}\mathbf{I}$, we have a linearly parameterized nonlinear discrete-time model $\mathbf{z}_{t+1} = \mathbf{A}_t(\mathbf{z}_t|\varphi) + \sqrt{\Delta t}\,\mathbf{n}_{t+1}$ where $\varphi = (\pi_1, \ldots, \theta_3)'$; $\mathbf{n}_{t+1}$ is a bivariate Gaussian white noise with variance-covariance matrix $\sigma^2\mathbf{I}$; $\mathbf{A}_t(\mathbf{z}_t|\varphi)$ is given by (22.14), (22.15), (22.16), or (22.17). The least squares estimate $\hat{\varphi}$, which is asymptotically equivalent to the maximum likelihood estimate, is obtained by solving a linear equation. For example, for the parameterization (22.14), we have only to solve the following equation:

$$X'Y = X'X\varphi$$

where $\varphi = (\pi_1, \pi_2, \pi_3, \theta_1, \theta_2, \theta_3)'$, $Y = (x_2, y_2, x_3, y_3, \ldots, x_N, y_N)'$, and

$$X = \begin{pmatrix} x_1 & x_1^2 & x_1y_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & y_1 & y_1x_1 & y_1^2 \\ x_2 & x_2^2 & x_2y_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & y_2 & y_2x_2 & y_2^2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{N-1} & x_{N-1}^2 & x_{N-1}y_{N-1} & 0 & 0 & 0 \\ 0 & 0 & 0 & y_{N-1} & y_{N-1}x_{N-1} & y_{N-1}^2 \end{pmatrix}$$

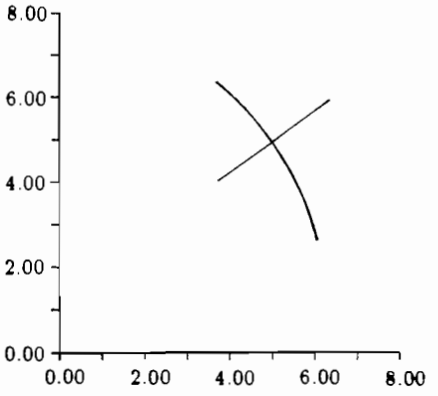

Figure 22.14: Estimated IS-LM curves by the model (15).

22.5 Bayesian Estimation of Time-varying IS-LM Curves

In the previous section, we introduced a stationary dynamic model for the IS-LM paradigm. However, a real economy is unlikely to be stationary. The slopes of the IS and LM curves may change over time. The strength of the stabilizing power may also change as time passes.

Let us allow the parameter $\varphi_t = (\pi_1, \pi_2, \pi_3, \theta_1, \theta_2, \theta_3)'_t$ of the following model

$$\begin{aligned} x_{t+1} &= \pi_1 x_t + \pi_2 x_t^2 + \pi_3 x_t y_t + \varepsilon_{t+1}^{(1)} \\ y_{t+1} &= \theta_1 y_t + \theta_2 x_t y_t + \theta_3 y_t^2 + \varepsilon_{t+1}^{(2)} \end{aligned} \tag{22.18}$$

to be stochastic and slowly change as $\varphi_{t+1} = \varphi_t + \mathbf{w_{t+1}}$, where $\mathbf{w_{t+1}}$ is a Gaussian white noise with variance-covariance matrix $\sigma_w^2\mathbf{I}$. Then we have the following state space representation

$$\begin{aligned} \varphi_{t+1} &= \varphi_t + \mathbf{w_{t+1}} \\ \mathbf{z}_{t+1} &= \mathbf{F}_{t+1}\varphi_{t+1} + \mathbf{v_{t+1}} \end{aligned} \tag{22.19}$$

here $\mathbf{z}_{t+1} = (x_{t+1}, y_{t+1})'$, $\mathbf{v}_{t+1}$ is a bivariate Gaussian white noise with variance-covariance matrix $\sigma_v^2\mathbf{I}$, and

$$\mathbf{F}_{t+1} = \begin{pmatrix} x_t & x_t^2 & x_t y_t & 0 & 0 & 0 \\ 0 & 0 & 0 & y_t & y_t x_t & y_t^2 \end{pmatrix}$$

The estimate of φ_t which minimizes $E(\|\tilde{\varphi}_t\|^2)$ is obtained recursively by applying the Kalman filtering algorithm. $\|\tilde{\varphi}_t\|^2$ means the sum of squares of the components of $\tilde{\varphi}_t$, the estimation error of φ_t given by $\tilde{\varphi}_t = \varphi_t - \hat{\varphi}_{t|t}$, where $\hat{\varphi}_{t|t}$ is an estimate of φ_t, given the observations $\boldsymbol{z}_t$, $\boldsymbol{z}_{t-1}$, $\ldots, \boldsymbol{z}_1$. The state estimates, $\hat{\varphi}_{1|1}$, $\hat{\varphi}_{2|2}$, $\ldots, \hat{\varphi}_{N|N}$, thus obtained by the Kalman filtering algorithm are equivalent to the estimates $\hat{\varphi}_1$, $\hat{\varphi}_2$, $\ldots, \hat{\varphi}_N$ obtained by maximizing

$$l(\varphi) = \exp\{-\frac{1}{2\sigma^2}[L(\varphi) + d^2\|\varphi_0 - \varphi\|_R^2]\}$$

and are given as the mean of the posterior distribution defined by the data distribution

$$f(\boldsymbol{z}|\sigma^2, \varphi) = \left(\frac{1}{2\pi}\right)^{\frac{N}{2}} \left(\frac{1}{\sigma}\right)^N \exp\{-\frac{1}{2\sigma^2}L(\varphi)\} \tag{22.20}$$

and the prior distribution

$$g(\varphi|d) = \left(\frac{1}{2\pi}\right)^{\frac{k}{2}} \left(\frac{1}{\sigma}\right)^k |d^2 R| \exp\{-\frac{d^2}{2\sigma^2}\|\varphi - \varphi_0\|_R^2\} \tag{22.21}$$

$L(\varphi)$ is a sum of squares of the prediction error of $\boldsymbol{x}_t$ and y_t for parameter φ; $\|.\|_R^2$ denotes the norm defined by a positive definite matrix R; $|R|$ denotes the determinant of R; k is the dimension of the parameter space, i.e., $k = 6$ in the present case. In the above model (22.19) we have $\|\varphi - \varphi_0\|_R^2 = \|\mathrm{c}_0 - \mathbf{D}\varphi\|^2$ where $\mathbf{D}$ is a properly chosen matrix, and $\mathrm{c}_0 = \mathbf{D}\varphi_0$. In the Kalman filtering algorithm σ_w^2 and σ_v^2 of (22.19) are assumed to be given. In real applications we need to decide their values. A practical Bayesian criterion for the choice of the variables, σ_v^2 and σ_w^2 which correspond to σ^2 and d^2 in (22.20) and (22.21), is given by Akaike (1980). (Details of the procedure and an algorithm for the estimation of the above model will be given in a future paper.)

Let us see how the procedure works in some simulations. We generate data from the model

$$\begin{aligned} \dot{x} &= [a_1(t) + a_2(t)x + a_3(t)y]\,x + n_1(t) \\ \dot{y} &= [b_1(t) + b_2(t)x + b_3(t)y]\,y + n_2(t) \end{aligned} \tag{22.22}$$

where the vectors $[a_1(t), \ldots, b_3(t)]'$ slowly change from $(21.25, -1, -0.25, 0, 1, -0.25)'$ for $t = 0$ to $(21.25, -0.25, -1, 0, 0.25, -1)'$ for $t = N\Delta t$; $N = 400$ and $\Delta t = 0.1$. The generated data $(x_1, y_1)', \ldots, (x_{400}, y_{400})'$ are plotted in *Figure 22.15*. The intersection of the two lines defined by $a_1(t) + a_2(t)x + a_3(t)y = 0$ and $b_1(t) + b_2(t)x + b_3(t)y = 0$ are plotted in *Figure 22.16*.

For this simulated data, by applying the above estimation procedure with $\sigma_v^2 (= \Delta t\, \sigma^2) = 0.1$ and $\sigma_w^2 = 10^{-5}$, we obtained a sequence of estimates $\hat{\varphi}_1, \ldots, \hat{\varphi}_N$. As we saw in

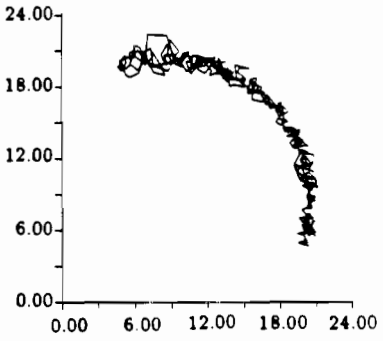

Figure 22.15: Simulated data ($N = 400$) of nonstationary model (22).

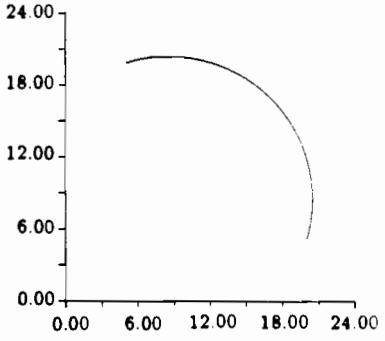

Figure 22.16: Trajectory of equilibrium points of nonstationary model (22) with $\sigma^2 = 1$.

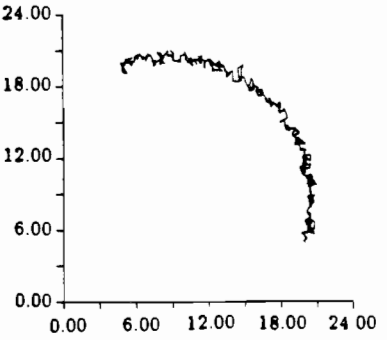

Figure 22.17: Estimated trajectory of the equilibrium points with $\sigma_w^2 = 10^{-5}$.

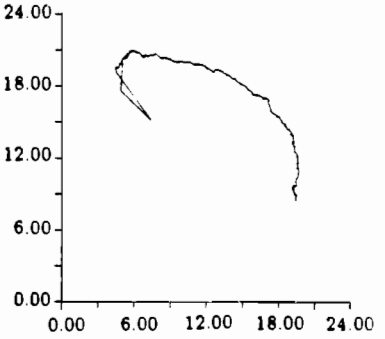

Figure 22.18: Estimated trajectory of the equilibrium points with $\sigma_w^2 = 10^{-7}$.

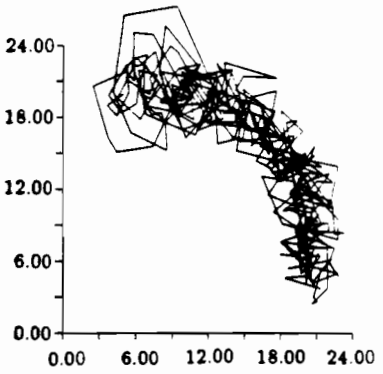

Figure 22.19: Simulated data ($N = 400$) of model (22) with $\sigma^2 = 12$.

the previous section, each estimate $\hat{\varphi}_i = [\hat{\pi}_1^{(i)}, \hat{\pi}_2^{(i)}, \hat{\pi}_3^{(i)}, \hat{\theta}_1^{(i)}, \hat{\theta}_2^{(i)}, \hat{\theta}_3^{(i)}]'$ defines two lines, $\hat{\pi}_1^{(i)} - 1 + \hat{\pi}_2^{(i)} x_t + \hat{\pi}_3^{(i)} y_t = 0$ and $\hat{\theta}_1^{(i)} - 1 + \hat{\theta}_2^{(i)} x_t + \hat{\theta}_3^{(i)} y_t = 0$, which are approximations of the IS and LM curves. The sequence of the intersections of the two lines is plotted in *Figure 22.17*. *Figure 22.18* shows the sequence of the intersections of the lines, where $\sigma_w^2 = 10^{-7}$. The procedure seems to work well even with quite a large noise variance σ^2. *Figure 22.19* shows the data ($N = 400$) generated from the same model as (22.22) but for $\sigma^2 = 12$. *Figure 22.20* shows the plotted sequence of the intersections of the lines defined by the estimated sequence of the parameters $\hat{\varphi}_1, \ldots, \hat{\varphi}_N$ obtained by applying the procedure with $\sigma_w^2 = 10^{-7}$ and $\sigma_v^2 (= \Delta t \sigma^2) = 1.2$.

22.6 Discussion

The dynamical system model (22.2) is valid not only as a model of the dynamics of interest rate and output, but also as a model of the dynamics of other variables in macroeconomics, such as inflation rate and unemployment rate, or labor demand and labor supply.

Dornbusch and Fischer (1978) introduced a model to explain the dynamics of inflation rate and output. In their model the equilibrium point of the inflation rate and output is realized by the intersection of the demand curve (AD-curve) and the supply curve (AS-curve). The dynamics explained by the Dornbusch-Fischer model are explicitly realized in a mathematical form in the model (22.2), where x is the inflation rate, y is the output, $f_1(x, y) = 0$ defines the AS-curve, and $f_2(x, y) = 0$ defines the AD-curve. Since the coefficients of $f_1(x, y)$ and $f_2(x, y)$ are dependent on the money supply and the government spending, the AD-AS curves and the equilibrium point could be shifted. This means that the Phillips curve, which is obtained from the AS-curve by changing the variable from output to unemployment rate using Okun's law, could also shift depending on monetary

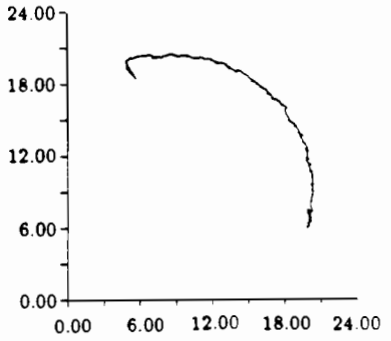

Figure 22.20: Estimated trajectory of the equilibrium points with $\sigma_w^2 = 10^{-7}$.

and fiscal policies.

Application of the present model and identification method to real economic data will be discussed in a future paper.

References

Akaike, H. (1980), Likelihood and the Bayes procedure, in: J.S. Bernardo, M.H. Degroot, D.V. Lindley, and A.F.M. Smith (eds.) *Bayesian Statistics.* Valencia: University Press.

Dornbusch, R. and Fischer, S. (1978), *Macroeconomics.* New York: McGraw-Hill.

Fletcher, R. and Powell, M.J.D. (1963), A rapidly convergent descent method for minimization. *Computer Journal*, **6**, 163-168.

Hirsh, M.W. and Smale, S. (1974), *Differential Equations, Dynamical System and Linear Algebra.* New York: Academic Press.

Modigliani, F. (1977), The monetarist controversy or: Should we forsake stabilization policies? *American Economic Review*, **67**, 1-19.

Morgan, B. (1978), *Monetarists and Keynesians.* London: Macmillan.

Ozaki, T. (1985a), Statistical identification of storage models with application to stochastic hydrology. *Water Resources Bulletin*, **21**, 663-675.

Ozaki, T. (1985b), Non-linear time series models and dynamical systems, pp. 25-83 in: E.J.Hannan *et al.* (eds.), *Handbook of Statistics*, Vol. 5. Amsterdam: North-Holland.

Ozaki, T. (1986), Local Gaussian modelling of stochastic dynamical systems in the analysis of nonlinear random vibrations, pp. 241-255 in: *Essays in Time Series and Allied Processes: Festschrift in Honour of Prof. E.J. Hannan.* Applied Probability Trust.

CHAPTER 23

Econometrics of Technical Change: Techniques and Problems

André Keller

Summary

Technical change greatly contributes to the explanation of economic structural changes. Numerous studies attempt to quantify and model this essential aspect of economic growth. This study surveys the models, their estimation techniques, and the problems and pitfalls in the applications. Particular emphasis is laid on aggregated production functions, factor productivity, and input-output approaches.

23.1 Introduction

Technical change and structural change are tightly interrelated in the economic growth process, as is shown in numerous theoretical and empirical studies. Denison (1980, 1983), in evaluating the 3.8% growth rate of national income from 1948 to 1973, attributes 15% to capital contributions and 37 % to advances in technologies, managerial skills, and organizational knowledge. Moreover, technical progress has a special importance for longrun simulations of economic models. For Kennedy and Thirlwall (1972), technical progress embraces two main aspects: the effects of changes in technology (macrostudies) quantified as the rate of technical progress, and the changes in technologies themselves (microstudies) to explain the process. Here, we are mainly concerned with the first interpretation.

Technical progress, which embraces the process of innovation and the longterm determinants of capital accumulation as well, has been designated as one of the driving forces of structural change along with population growth, capital accumulation, and the use

of resources [Krelle (1984)]. Pasinetti (1985) postulates a multisector economic system wherein technical progress occurs at different sectoral rates. He criticizes macroeconomic growth models only containing a single commodity (or a composite one with invariable composition) and where technical change is introduced in the form of an "overall rate of technical progress." But his approach differs from input-output models in the sense that he prefers to consider vertically integrated sectors.

Empirical studies aim at providing a longrun analysis of technical progress and structural change. For example, Førsund and Hjalmarsson (1983) estimate vintage models on the basis of microdata for individual firms in the cement industry. In response to rising fuel prices in the 1970s, the approach was extended to other branches of industry, such as the aluminium industry in Norway; see Førsund and Jansen (1983). To elucidate the process of technical change, these authors generally had to reformulate existing models.

Input-output techniques have been extensively explored in the survey of Sato and Ramachandran (1980). The authors describe the improvements that have been introduced. Vaccara (1970) and Carter (1970) compute measures of technological progress on the basis of both direct input-output matrix and Leontief inverse matrix, to study structural change in the US economy. Istvan (1974) introduces lags in a dynamic input-output model to evaluate the economic impact of projected technological changes. Carter (1974, 1976) combines these two propositions into a general equilibrium analysis. Craven (1983) discusses the properties of input-output systems in which the technical coefficients may change over time.

A shift in the aggregate production function over time is generally considered as an effect of technical progress because of greater efficiency in combining the inputs. These shifts are done in a variety of ways, including changes in the coefficients of labor and capital. These problems led econometricians, in particular, to question the stability and the consistency of estimated parameters in production functions. Difficulties arise when one is attempting to distinguish between movement along a given production function and shifts. The Divisia index of technical change helps to make this distinction as long as the production function is differentiable; see Haltmaier (1984).

Nelson (1980) tends to consider as a "fragile construct" the usual assumptions on production sets, their efficiency frontiers (constant return to scale of techniques, concave and differentiable isoquants), and technological knowledge (shift of the production function explained by R&D spending differences). Thus, the distinction between moving along and shifting the production function supposes that learning and doing are seperate activities.

Theoretical and empirical aspects of the technical progress have been extensively considered in numerous papers and books. Most of them attempt to model this essential aspect of economic growth. The surveys cover various aspects of technical change. Several important contributions discuss the economics of technical change — Mansfield (1968), Rosenberg (1971, 1974), Mensch (1975), and others — or industrial innovation, such as Freeman (1974). Production functions are extensively examined by Hildebrand and Liu (1965), Walters (1968), Ferguson (1969), and Sato (1975). Other surveys are based on the theory and empirical analysis of production and cost functions [see Walters (1963), Frisch (1965),

Solow (1967), Nerlove (1967)]; of factor productivity and R&D [see Nadiri (1970), Kamien and Schwartz (1982), Nelson (1981), and Griliches (1984)]. Technical progress problems are exposed by Kennedy and Thirlwall (1972). These surveys, which deal with technical change, show the specificity of econometric questions concerning availability and measurement of data, specification of equations and models with technical change, estimation techniques, and resolution procedures.

Following Rosenberg (1974), most empirical studies of technological change attempt to evaluate the contribution of technical progress to growth and to study the rate at which new inventions exert their impacts on productivity growth. The diffusion process and the pattern of inventive activity are considered. Gort and Wall (1986), for example, investigate the optimal path of investment into innovation.

Major econometric topics relating to technical change will be surveyed in this study. The restrictive approach still represents an ambitious task because of the various complexities and propositions that have been introduced into practice. Particular emphasis will be laid on aggregated production functions; but we will also consider input-output models, which are more appropriate to describe structural changes in economics. This focus on econometric techniques and problems in studying technical change will simply follow the different stages one usually passes through in applied econometrics: data availability and measurement problems are first considered; then we examine usual specification of equations, such as production functions and their duality forms; finally, we consider problems that arise in estimation and try to evaluate the sensitivity of results owing to alternative specifications, estimation methods, and errors in the variables.

23.2 Description and Effects of Technical Change

23.2.1 Characteristics and bias of technical change

The characteristics of technical change may be shown by the shifts of the unit isoquant toward the origin over time, but they are also known through their consequences. According to Nadiri (1970), better techniques allow for reducing the unit cost of all factors equally. A greater saving in one input than in others will result in a bias in technical change. Exchanging factors in the production process is measured by the elasticity of substitution. Then, a bias in technical change will be represented by a modification in the position of the isoquant and will lead, for example, to greater labor savings for all techniques. The isoquant may also change its curvature, which will increase the specificity of capital and labor when the elasticity of substitution σ is reduced.

The neutrality of technical change arises since such characteristics as capital/output ratio, output/labor ratio, factor proportions, and marginal productivities depend not only on technology, but also on factor proportions [Beckmann and Sato (1969)]. It is therefore necessary to neutralize the effects of any change in input factor proportions. This question leads us to consider unchanged relationships under technical change. Moreover, since increases in efficiency are reflected in increased productivity for all existing capital equip-

ment, technical progress is considered as being "disembodied" and corresponds to the first class of real growth models with technical progress considered by Ott *et al.* (1975). The second class, which is examined later, corresponds to the vintage models where technical improvements only concern certain kinds of capital equipment.

Disembodied technical progress embraces Hicks-neutral, Harrod-neutral, and Solow-neutral forms of technical change. Technical change is Hicks-neutral if the ratio of marginal products remains unchanged at a constant capital/labor ratio. Technical change is Harrod-neutral if it augments labor input, the rate of return of capital remaining unchanged at a constant output/capital ratio. Technical change is Solow-neutral if it augments capital input, leaving the wage rate unchanged, at a constant output/labor ratio over time. The properties of neutral forms of technical progress have been extensively discussed in the literature on growth models. See, for example, Hahn and Matthews (1964), and Ott *et al.* (1975). In particular, it has been shown that Hicksian and Harrodian definitions are equivalent when the elasticity of substitution equals unity. This is the case of a Cobb-Douglas production function. Beckmann and Sato (1969) derive other forms of technical change. They distinguish product augmenting, factor augmenting, and input decreasing. Following Nadiri (1970), bias in technical change, which is defined in different ways [Stiglitz and Uzawa (1969)], may be measured by the relative shares of the inputs. Thus, the Hicksian bias may be symbolically represented by the following partial timederivative

$$\left[\frac{\partial(F_K K)/(F_L L)}{\partial t}\right] \begin{matrix} > \\ = 0 \\ < \end{matrix} , \qquad \text{K/L constant} , \qquad \begin{matrix} \text{labor-saving} \\ \text{neutral} \\ \text{capital-saving} \end{matrix} \tag{23.1}$$

where F_L and F_K figure the partial derivatives of output with respect to capital L and labor K.

The underlying production function is

$$Q = AF(L, K) \tag{23.2}$$

where A represents the disembodied technical change. The function F states a homogeneous and differential production function. Neutrality is achieved when the derivatives equal zero; labor-saving technical change requires the derivative to be positive; and capital-saving technical change, negative. Unfortunately, because these characteristics of neutral technical change are highly interdependent, it becomes difficult to distinguish between them.

Brown and Popkin (1962) propose one specific statistical procedure to decompose neutral and nonneutral technical change and returns to scale. The method consists in isolating periods of no nonneutral technical change and then inferring measurement of output changes due to nonneutral technical change. Nonneutral technical change is measured by the ratio of marginal products of factors, say, $\alpha/(\alpha + \beta)$, where α and β design the elasticities with respect to labor and to capital, respectively. [Based on this method, Brown

and Popkin (1962) find, for the US economy, that the period 1890–1937 was labor-saving and the period 1919–1959, labor-using.] Salter and Reddaway (1969) attempt to seperate neutral from nonneutral technical change on the basis of aggregated data for industry. In their survey on technical progress, Kennedy and Thirlwall (1972) also express doubt as to the possibility of distinguishing bias on technical progress from factor substitution. Moreover, the test for bias in technical progress involves serious identification problems.

The embodiment hypothesis implies that new inputs are more apparent than older ones because of technological improvements. In this case, technical progress can only be embodied in new capital goods. New capital will be introduced by scrapping old capital equipment. Vintage models are generally classified into three groups, depending on assumptions about the substitutability of labor and capital. The "putty-clay" model allows for a substitution only when new equipment is introduced. After the installation of the equipment, the capital/labor ratio remains constant until new equipment is introduced. The "putty-putty" model is a vintage model with smooth substitution. In this case, new equipment continues to be used indefinitely with declining associated labor. In the "clay-clay" assumption, the capital/labor ratio remains constant, whatever new equipment may be introduced. The theory of vintage models includes technical progress as a function of the investment rate; see Kaldor (1957) and Solow (1960). Kaldor and Mirrlees' (1962) study is based on the rate of change of investment. Arrow's (1962) models of learning by doing introduce the notion of cumulative investment. The production function may be described by

$$Q = C[b(t)L]^{1-\alpha}K^{\alpha} \tag{23.3}$$

where $b(t)$ is the labor efficiency which depends on capital knowledge.

Nadiri (1970) mentions that the bias in technical change depends upon the elasticity of substitution and the differential rates of growth of labor and capital embodiment. The relation (23.2) may be written

$$Q = AF(\lambda_1 L, \lambda_2 K) \tag{23.4}$$

where λ_1 and λ_2 are the coefficients of factor augmentation. The direction of technical change then depends upon the ratio λ_1/λ_2. Technical change is Hicks-neutral when λ_1/λ_2 remains constant, Harrod-neutral with the constancy of λ_2 (labor-augmenting), and Solow-neutral with the constancy of λ_1 (capital-augmenting). Solow (1967) calculates the bias in the three following cases, so that we have

$$B = -\frac{1-\sigma}{\sigma}\left[\frac{d\lambda_1}{\lambda_1} - \frac{d\lambda_2}{\lambda_2}\right] \tag{23.5}$$

$$C = (1-\sigma)d\lambda_1/\lambda_1 \tag{23.6}$$

$$D = (1-\sigma)d\lambda_2\lambda_2 \tag{23.7}$$

where σ is the elasticity of substitution between K and L. Bias B is equally labor- and capital-augmenting; C is labor-augmenting; and D capital-augmenting.

Intriligator (1965) incorporates both embodied and disembodied technical change for the US manufacturing sector. Alternative measures are used for labor inputs (adjustment of the effectiveness of labor force), for embodied capital input. Mairesse (1977, 1978) objects that the following studies based on aggregated time series did not succeed in separating the influence of embodied and disembodied technical change: Intriligator (1965), Berglas (1965), Wickens (1970), and Barger (1976). Mairesse's studies, which are based on panel data for French manufacturing, determine that the embodied technical progress is an important factor explaining the differences in labor productivity between firms.

The *induced technical change* approach of Kennedy (1964) aims at incorporating factors that will determine the direction of the bias in new techniques. The innovative possibility curve (IPC), which he introduces, relates the proportional reduction in requirements of L and K due to new technologies. The curve is completely determined by the relation

$$\phi(\mu_L, \mu_K) = 0 \qquad \text{with } d\mu_L/d\mu_K < 0 \text{ and } d^2\mu_L/d\mu_K^2 > 0 \tag{23.8}$$

Here, the choice of techniques depends on the relative share of inputs, s_L and s_K, and technology represented by μ_L and μ_K. The bias depends upon whether μ_L is larger, equal or less than μ_K. If $s_L > s_K$, then $\mu_L > \mu_K$. This introduces a labor-saving bias in this case.

According to Nadiri (1970), this approach differs from the preceding Schumpeterian view of technical change, which was supposed to be autonomous, neutral, and growing at a constant rate because of the knowledge and inventions. Ahmad (1966), Fellner (1969), and Hicks (1964) considered the effect of relative prices on the direction of technical change, because some firms tend to anticipate the relative real factor prices and, consequently, to substitute factors. However, Nadiri (1970) observes that technical change may occur without any influence of relative prices by means of a learning process.

Endogenous technical change has been previously considered by Arrow (1962), Shell (1966), Schmookler (1966), Mansfield (1968), and Nelson (1968); see also Sato and Suzawa (1983) for an extensive presentation. The authors aimed at pinpointing the determinants and the direction of the accumulation of capital knowledge. For Lucas (1967) and Nordhaus (1969), the rule and timing of technical change could be determined by the discount rate at relative prices. Nordhaus (1969) stated a production function for technology, defined by

$$dA/A = N^{\beta_1} A^{-\beta_2} \tag{23.9}$$

where A is the stock of technology, and N is the number of inventions. The optimum value of N is given by

$$N = \frac{[\beta_1 Q A^{-\beta_2} \phi]^{1/(1-\beta_1)}}{D} \tag{23.10}$$

where ϕ is the discount factor, and D the cost per invention.

Griliches (1964, 1979) largely initiated empirical studies on the role and impact of R&D, as we shall see later.

23.2.2 Measurement of technical change and sources of errors

Productivity indexes have been widely used to evaluate technical change; see, for example, Kennedy and Thirlwall (1972), and Nadiri (1970) for a presentation of the approaches. Solow's geometric index is based on a Cobb-Douglas production function with constant returns to scale for an autonomous neutral technical change. In this approach, Solow (1967) determines technical change as a residual as follows

$$\frac{dA}{A} = \frac{dQ}{Q} - \alpha\frac{dL}{L} - \beta\frac{dK}{K} \tag{23.11}$$

Thus the rate of change of total productivity (dA/A) which is used as an indicator of technical progress, equals the difference between the rate of growth of the real product and real factor inputs variations. Kendrick's (1961) measure is consistent with the aggregate production function

$$Q = tKL(cL^{\rho} + dK^{\rho})^{-1/\rho} \tag{23.12}$$

where t stands for disembodied technical change, c and d for efficiency parameters. The equivalence of Kendrick's measure to Solow's is established by Levhari *et al.* (1966) for small variations in the inputs.

Several difficulties may arise from the previous method, since the production functions used in the calculations require a precise specification and, hence, an accurate estimation of the parameters α and β. Moreover, weighting inputs by their respective share in the total output supposes the existence of perfectly competitive markets, where prices equal marginal products. Massell (1962) mentions another statistical difficulty, resulting in an excess of technical change in the US manufacturing sector over a weighted average of individual firms, due to interindustry shifts. International and interregional comparisons are also difficult because of changing prices and lack of data. Econometric difficulties will be discussed later.

Errors in the variables certainly constitute the main source of difficulties: measurement of output and inputs, of quality changes, of nonmarketed items; influence of unobservables, such as the state of knowledge and technology, management, etc. [see Lindbeck (1983) and Morris (1983)]. Jorgenson and Griliches (1967), who discuss this major problem extensively, present two kinds of errors: first, errors of aggregation, coming from the combination of heterogeneous inputs and outputs; and second errors of measurement in the variables. Aggregation problems may result from modification in technical characteristics over time. Errors of aggregation are generally labelled "quality change" in the sense that differences in the rates of growth of quantities may occur within a given group. The

choice of an appropriate unit of measurement is difficult. A suitable approach consists in using engineering characteristics and transposing the aggregate measure into quantities. A further advantage is provided by the wide range of available observations, unlike time series or cross-section data. Chenery (1949, 1953) and Moore (1959) pioneered the use of engineering data to find out industrial production functions. Heady and Dillon (1962) adopted the approach for the agriculture sector; Ferguson (1951), for air transport [see Walters (1963a)]. To eliminate the error of aggregation, Solow (1957) suggests to replace the initial index of total inputs by a Divisia index of labor and capital input.

Moreover, difficulties may arise from the measurement techniques. Production is generally measured by a deflated index. Labor may be measured by the number of employees or manhours without difficulties. However, capital is more difficult to estimate since official data are generally not available. Capital stock may be derived from the sequence of investment expenditures and then be deflated. But capital in use seems to be more relevant as a variable. Mairesse (1978) precisely describes the sequence of corrections needed to elaborate significant measures of the variables. Thus, capital stock, which proceeds from the book value of gross fixed assets, is reevaluated by the price index of investment in equipment and structures. The consequences of measurement errors have been extensively discussed by authors who suggest that the residual change in the total factor productivity would be much smaller, since the evaluation of inputs are generally underestimated. Jorgenson and Griliches (1967) lay emphasis on other likely useful variables of technical change, such as R&D expenditures. Griliches (1979) considers the insufficiency of R&D data and especially the lack of high-skill labor-intensive deflators to be critical. Patent applications are only a rough indicator of technical advances, because of their heterogeneous composition. The economic value of patents has been discussed by Taylor and Silberston (1973), Stoneman (1983), and Schankerman and Pakes (1984). Nevertheless, Hall *et al.* (1986) attempt to capture the lag structure of the patents-R&D relationship.

23.2.3 Estimates of technical change for the US economy

Disembodied technical change has generally been obtained by means of a Cobb-Douglas production function with time trend term for different periods. Various methods of evaluation have been used to calculate the change in the total productivity, which is generally considered an indicator of technical progress. The results are shown in *Table 23.1*, which is only concerned with disembodied technical change. The rate of growth of disembodied technical change may be derived from a Cobb-Douglas production function with an exponential time trend

$$Q = Ce^{\lambda t}L^{\alpha}K^{\beta} \tag{23.13}$$

From (23.12) we deduce

$$\frac{\triangle Q}{Q} = \lambda + \alpha\frac{\triangle L}{L} + \beta\frac{\triangle K}{K} \tag{23.14}$$

The results of *Table 23.1* clearly show the sensitivity to the choice of the specifications such as first differences, according to Brown (1966), or levels, for others. When we consider the evolution of the results from one period to another, technical change tends to increase. More recent estimates of the residual by Kendrick (1981) give the following rates of growth of technical progress in the US economy: +0.5% in 1960–1973, +0.4% in 1973–1979.

Embodied technical change, which involves an increase in the effectiveness of factor inputs due to the improvement of quality or efficiency over time, may also be determined. Intriligator (1965) incorporates both embodied and disembodied technical change in a production function for the aggregate US manufacturing sector. The rate of embodied technical change, which corresponds to the "best" estimates, is 4% annually over the period 1929-1958. Mansfield (1968) and Solow (1962) also retain a rate of 4% as embodied technical change. Ignoring disembodied technical change in his equation, Solow (1960) obtains 2.5% per annum. [Mairesse (1978), for French industries, simultaneously estimates embodied and disembodied technical change by using a Solow production function. Both forms of technical progress then increase at a 2% annual rate over the period 1966 to 1975. Barger (1976) estimates that embodied and disembodied technical change would have been more rapid in the 1960s than in the 1950s in the nine European countries and in the US economy.]

The corresponding Cobb-Douglas production function that incorporates embodied technical changes may be rewritten

$$\frac{\triangle Q}{Q} = \lambda' + \alpha\frac{\triangle L}{L} + \beta\frac{\triangle K}{K} + (1-\alpha)\lambda_K - (1-\alpha)\lambda_K\triangle\bar{a} \tag{23.15}$$

where λ' is the disembodied technical progress, $\bar{a}$ is the average age of capital equipment, $\triangle\bar{a}$ is a rough measure of the gap existing between the "best practice technology" and the average level of technology, and λ_K is the growth rate of the average quality of capital.

This estimation form simply results from a Nelson-Solow version of the Cobb-Douglas with embodied technical change, as we shall see later [Nelson (1964)]. However, when gross investment both adds new capital goods to the existing capital stock and raises the quality of intermediary inputs, productivity calculations underestimate the importance of capital accumulation (a quality reduction is also due to obsolescence), according to Lindbeck (1983).

23.3 Econometric Specification of Technical Change

23.3.1 Specification of production functions with disembodied technical change

The specification of a production function and the form of technical change correspond to one another. Beckmann and Sato (1969) show that certain forms of production function preclude some forms of technical change. The model is based on the usual assumptions: one homogeneous production function of degree one with two factors, L and K; perfectly

Table 23.1: Estimates of technical change for the US manufacturing sector.

Period	*Concept* [a]	*Method*	*Estimates* (in annual rates [%])	*Reference*
1869–1928	TFP	Factor shares	1.1	Schmookler (1952)
1869–1928	TC	Cobb-Douglas	0.75	Valavanis-Vail (1955)
1870–1914	TFP	Cobb-Douglas with time trend	1.1	Tinbergen (1942)
1890–1906	TC	Cobb-Douglas in first differences	0.18	Brown (1966)
1890–1960	TC	Cobb-Douglas in first differences	0.61	Brown (1966)
1899–1953	TFP	Cobb-Douglas	1.7	Kendrick (1961)
1909–1949	TFP	SMAC production function	1.6	Kendrick and Sato (1963)
1909–1949	DTC	Embodied-disembodied model	1.6	Intriligator (1965)
1919–1955	TFP	Cobb-Douglas	1.5	Massell (1960)
1919–1957	TFP	Solow production function	1.5	Solow (1957)
1929–1958	DTC	Embodied-disembodied model	1.6	Intriligator (1965)
1950–1965	MFP	Productivity function	2.1	Baily (1984)
1950–1968	TC	Cobb-Douglas	1.62	Coen, Hickman (1980)
1960–1973	TFP	Cobb-Douglas	1.2	Åberg (see Lindbeck) (1982)
		(cross-country data)	1.8	Lindbeck (1983)
1965–1973	MFP	Productivity function	1.96	
1969–1978	TC	Cobb-Douglas	1.03	Coen, Hickman (1980)
1973–1978	TFP	Cobb-Douglas	0.6	Åberg (see Lindbeck) (1982)
		(cross-country data)	1.2	Lindbeck (1983)
1973–1981	MFP	Productivity function	0.76	Baily (1984)

[a] TFP: total factor productivity; TC: technical change; DTC: disembodied technical change; MFP: multifactor productivity

competitive markets in the long run; and exogenous technical change. Hence, the marginal productivities of labor and capital equal their respective prices — say, w and r

$$w = \partial F/\partial L \quad \text{and} \quad r = \partial F/\partial K \tag{23.16}$$

The role of the technical change may be introduced as usual. Hicks-neutrality (product-augmenting) is represented by

$$Q = A(t)F(L, K) \tag{23.17}$$

where $A(t)$ is the technical progress.

Harrod-neutrality (labor-augmenting) is given by

$$Q = F[A(t)L_0, K] \tag{23.18}$$

A formal proof of (23.18) is given by Uzawa (1960–1961).

Solow-neutrality (capital-augmenting) is given by

$$Q = F[L, A(t)K_0] \tag{23.19}$$

Alternative specifications and underlying production functions that may correspond to each major type of technical change are described in *Table 23.2*. According to this presentation, the implied production function has been obtained by integration of the differential equations that correspond to the retained econometric specification and previous assumptions. The production functions are close to a Cobb-Douglas form in the linear case and to a CES form in the nonlinear case. As to the empirical results obtained by Beckmann and Sato (1969), Solow-neutral technical change ranks first in the US economy when we consider the R^2 statistics (*Table 23.3*).

The accuracy of the specification of the production function will generally govern the quality of the productivity measurement. The differentiation of a Hicks-neutral production function, such as (23.17), with respect to time gives

$$\frac{dA}{A} = \frac{dQ}{Q} - \frac{LF_L}{Q}\frac{dL}{L} - \frac{KF_K}{Q}\frac{dK}{K} \tag{23.20}$$

In equation (23.20), the residual clearly depends upon the form of the production function, as determined by the partial derivatives F_L and F_K, by the measurement of the variables L and K and also the adjustment of their quality change, and by the omitted variables. [Morris (1983) also mentions interrelations and multiple causalities among capital, labor, and output; between real and financial variables; and between supply and demand factors.] The methods of estimation are the area for another source of error that we shall discuss later. A measure of the specification error is given by Nelson (1965), who compares a Cobb-Douglas production function with constant returns to scale, and Hicks-neutral technology with a CES production function. The measure of technical change by the CES function will differ by the following expression

Table 23.2: Correspondence between major types of technical change econometric specifications and implied production functions [according to Beckmann and Sato (1969)].

Type of techn. change	*Equation form*	*Econometric specification*	*Implied production function* Q/K [a]
Hicks	Linear	$r/w = a + b(L/K)$	$A(t)[a + (1+b)L/K]^{1/(1+b)}$
	Log-lin.	$\ln(r/w) = a' + b' \ln(L/K)$	$A(t)[e^{a'} + (L/K)^{1-b'}]^{1/(1-b')}$
Harrod	Linear	$r = a + b\,(Q/K)$	$\frac{[A(t)L/K]^{1-b}}{1-b} + \frac{a}{1-b}$
	Log-lin.	$\ln r = a' + b' \ln(Q/K)$	$\{e^{a'} + [A(t)(L/K)]^{1-b'}\}^{1/(1-b')}$
Solow	Linear	$w = a + b(Q/L)$	$(L/K)\{\frac{[A(t)K/L]^{1-b}}{1-b} + \frac{a}{1-b}\}$
	Log-lin.	$\ln w = a' + b' \ln(Q/L)$	$(L/K)\{e^{a'} + [A(t)(K/L)]^{1-b'}\}^{1/(1-b')}$

[a]The calculation may be illustrated for Harrod neutrality in the linear case. The production function is: $Q_t = F(L_t, K_t, t)$ or $\frac{Q}{K} = F(\frac{L}{K}, 1, t)$. With $q = Q/K$ and $x = L/K$ we get $q = f(x,t)$. Since the marginal productivity of capital (K) equals its price: $r = \partial F/\partial K = f - x f'_x$. The econometric specification gives $r = a + b_q$. Then we deduce the following differential equation: $(1-b)f - X f'_X = a$ where $X = A(t)x$. By integration, we get the corresponding production function which is given by: $Q/K = \frac{1}{1-b}\{[A(t)L/K]^{1-b} + a\}$.

Table 23.3: Importance of some technical change forms for the US, Japan, and German private nonfarm sectors: R^2 statistics from log-linear regressions [according to Beckmann and Sato (1969)].

Type of neutrality	*USA* (1909–1960) [a]	*Japan* (1930–1960)	*Germany* (1850–1959)
Hicks	0.831	0.785	0.708
Harrod	0.933	0.855*	0.422
Solow	0.944*	0.758	0.980*

[a]Period of the available data; the authors do not precise the period of fit. The asterisk indicates what type of technical progress ranks first.

$$E = +\frac{1}{2}\alpha(1-\alpha)[(\sigma-1)/\sigma][dK/K - dL/L]^2 \tag{23.21}$$

where σ is the elasticity of substitution between L and K. Hence, we have

$$\frac{dA}{A} = \frac{dQ}{Q} - \alpha\frac{dL}{L} - (1-\alpha)\frac{dK}{K} - E \tag{23.22}$$

The productivity changes dA/A will then be affected if σ differs from unity and also if the growth rates of inputs differ considerably.

Structural forms of production functions with neutral technical progress may differ. The easiest way is to allow the scale parameter A to vary without affecting the marginal rate of substitution. Then, for a Cobb-Douglas production function [Cobb and Douglas (1928)], we have

$$Q(t) = A(t)L^{\alpha}(t)K^{\beta}(t) \tag{23.23}$$

Tinbergen (1942) introduces an exponential time trend as a proxy variable for technical change, with

$$A(t) = A_0 e^{\lambda t} \tag{23.24}$$

In this case, technical change occurs smoothly over time. This approach has been extensively used in global macroeconometric models. However, Kopp and Smith (1982) support the use of explicit technological indicators in preference to a time trend variable. Nelson (1984) considers the time trend to be a "catchall variable" for various technological factors, such as learning by doing, organizational changes, etc. Bodkin and Klein (1967) introduce a time trend into a Cobb-Douglas production function. For example, the production function may be written as follows when the random error is supposed to be additive

$$Q = A10^{\lambda t}L^{\alpha}K^{\beta} + \varepsilon \tag{23.25}$$

where the variables Q, t, L, and K are vectors of data observations.

Bodkin and Klein (1967) also introduce a marginal productivity condition in the structural form to represent the costminimization behavior on the competitive factor markets. We deduce the system of equations

$$Q = A10^{\lambda t}L^{\alpha}K^{\beta} + \varepsilon \tag{23.26}$$

$$\frac{r}{w} = \frac{\beta}{\alpha}\frac{L}{K} + \varepsilon' \tag{23.27}$$

The parameters α and β are estimated simultaneously.

We also have to consider that one way to obtain nonneutral change is to allow the ratio α/β to vary over time. Hence, technical change will be capital-using if the parameter α rises relatively to β. The constant elasticity of substitution (CES) function, introduced by Arrow *et al.* (1961), presents interesting properties as they include the Cobb-Douglas and the Leontief production functions as special cases. This function has also been widely used in empirical works; see for example, Domar (1961), Dhrymes (1963), David and Van de Klundert (1965), Jorgenson and Griliches (1967). Kennedy and Thirlwall (1972) question whether the CES production function is useful, since a large diversity of results may be performed in time series and cross-section studies. Thus, the estimates of σ are frequently lower in time series studies than those issued from cross-section approaches. In their study, Bodkin and Klein (1967) introduce a likely time trend. Thus, they propose an estimation form defined

$$Q = A10^{\lambda t}[\delta K^{-\rho} + (1-\delta)L^{-\rho}]^{-\mu/\rho} + \varepsilon \tag{23.28}$$

where δ is a distribution parameter with $0 < \delta < 1$, ρ is a substitution parameter with $\rho \neq -1$, and μ is the degree of returns to scale. The marginal productivity relationships based on cost minimization introduces

$$\frac{r}{w} = \frac{\delta}{1-\delta}\left(\frac{K}{L}\right)^{-(\rho+1)} + \varepsilon' \tag{23.29}$$

The system (23.28)–(23.29) may then be used to find out the values of the parameters by means of techniques we shall present later. *Tables 23.4* and *23.5* give some of the results achieved by Bodkin and Klein (1967), using Cobb-Douglas and CES functions with different specifications. Here, we will consider the straight regression and the simultaneous estimation in the case of additive errors. Moreover, the parameters of the Cobb-Douglas production function are estimated without constrained returns to scale. The residuals of the production function are significantly autocorrelated as regards the low value of the von Neumann-Hart test (the critical value is about 1.52 at 5%). The parameter estimates of the straight equation differ from those calculated from the system. For either the Cobb-Douglas or CES production function, Bodkin and Klein found increasing returns to scale. In the Cobb-Douglas function, capital is not significant. This result is not consistant with CES results where δ significantly differs from unity. The rates of disembodied neutral technical change that we have indicated vary between 1.2 % and 1.6 % annually. Bodkin and Klein (1967) also indicate that increasing returns to scale generally allow for a slower pace of neutral technical change [see also Diwan (1963) and Walters (1968)].

A reduced form of production function may also be used to infer the value of the parameters of the structural form, provided that identification will then be possible. Thus, the underlying production function of the French Mogli model [see Courbis *et al.* (1980)] may be found in the following way

$$Q_m = Ce^{\lambda t}L_m^{\alpha}K^{\beta} \tag{23.30}$$

Table 23.4: Estimates of a Cobb-Douglas production function for the US economy over the period 1909–1949, according to Bodkin and Klein (1967). (The t-statistics are shown within parentheses.)

Method [a]	A	α	β	$\alpha+\beta$	λ	$\lambda \ln 10$ [b]	$\overline{R}^2$	$\frac{\delta^2}{s^2}$ [c]
SR	88.31	1.145	0.062	1.207	0.0069	1.6 %	0.99	1.07
	(15.6)	(16.5)	(0.8)	(18.9)	(16.0)			
SE	61.87	0.964	0.501	1.465	0.0053	1.2 %	0.98	0.64
	(79.3)	(14.8)	(14.7)	(14.9)	(8.3)			

[a]SR: straight regression; SE: simultaneous estimation; the errors are assumed to be additive.

[b]Evaluation of the growth rate of disembodied technical progress.

[c]Von Neumann-Hart statistic: a ratio of the mean square successive difference of the residuals to their total variance.

Table 23.5: Estimates of a CES production function for the US economy: 1909–1949, according to Bodkin and Klein (1967). (The t-statistics are shown within parentheses.)

Method [a]	A	ρ	$\sigma = \frac{1}{1+\rho}$	δ	μ	$\lambda \times 10^2$	$\lambda \ln 10$	$\overline{R}^2$	$\frac{\delta^2}{s^2}$
SR	40.46	10.18	0.0894	0.999	1.22	0.663	1.5 %	0.99	1.09
	(4.3)	(1.2)		(163.8)	(19.5)	(13.8)			
SE	58.39	0.475	0.6780	0.447	1.362	0.589	1.4 %	0.98	0.62
	(36.0)	(2.5)		(10.3)	(15.0)	(9.7)			

[a]SR: straight regression; SE: simultaneous estimation; the errors are assumed to be additive.

$$\frac{L}{L_m} = \left(\frac{Q}{Q_m}\right)^{\gamma} \quad \text{with} \quad 0 < \gamma < 1 \tag{23.31}$$

where Q_m, L_m are maximum values of Q and L. From equation (23.30), we deduce

$$L_m = C^{-1/\alpha} e^{-(\lambda/\alpha)t} K^{-\beta/\alpha} Q_m^{1/\alpha} \tag{23.32}$$

Equations (23.31)–(23.32) are determining the effective employment, L, so that

$$L = C^{-1/\alpha} e^{-(\lambda/\alpha)t} \left(\frac{K}{Q_m}\right)^{-\beta/\alpha} \left(\frac{Q}{Q_m}\right)^{\gamma} Q_m^{(1-\beta)/\alpha} \tag{23.33}$$

Taking $v = K/Q_m$ for the capital/output ratio and $C_p = Q/Q_m$ for the capacity utilization, we may calculate the average labor productivity so that we deduce from (23.32)

$$\frac{Q}{L} = C^{1/\alpha} e^{(\lambda/\alpha)t} v^{\beta/\alpha} C_p^{\frac{1-\beta}{\alpha}-\gamma} Q^{\frac{\alpha+\beta-1}{\alpha}} \tag{23.34}$$

where Q_m has been replaced by Q/C_p in (23.32). The estimation form may be written in differential terms

$$(\dot{Q/L}) = f(\dot{v}, \dot{C}_p, \dot{Q}) \tag{23.35}$$

where $\dot{X} = \frac{1}{X}\frac{dX}{dt}$. An application to French industry over the period 1961–1978 [Courbis (1983)] gives

$$\begin{array}{lccccc} (\dot{Q/L}) = & 0.032 & + 0.252\dot{v}_{-1} + & 0.209\dot{C}_p + & 0.406\,\dot{Q} - & 0.0188\,d_{63} \\ & (8.8) & (3.5) & (2.1) & (6.1) & (-3.1) \end{array} \tag{23.36}$$

where d_{63} is a dummy variable for the year 1963. Then, the structural parameters can be evaluated as $\alpha = 1.182$, $\beta = 0.298$, $\alpha + \beta = 1.48$, $\lambda = 0.038$ (or 3.8%) annually, and $\gamma = 0.385$.

Schmookler (1952) makes use of a factor shares method to calculate the estimates of the structural coefficient of the production function. Thus, under perfect competition, we can deduce the marginal productivity conditions of inputs — for example, labor

$$\beta\frac{Q}{L} = \frac{w}{p} \tag{23.37}$$

The equation (23.36) may be written

$$\ln\beta = \ln\left(\frac{wL}{pQ}\right) + \varepsilon \tag{23.38}$$

$e^{\widehat{\ln\beta}}$ is not an unbiased estimate, even if $\widehat{\ln\beta}$ is an unbiased estimator [Walters (1963a)]. Klein (1953) extensively used a variant of this method.

Duality forms of production functions result from the minimization problem where the total cost $wL + rK$ is minimized, given a production function $Q = F(L, K)$. Cost functions are usually easier to estimate than a generalized production function, when the latter is confined to two inputs. For the Cobb-Douglas $Q = AL^{\alpha}K^{\beta}$, the cost function according to Intriligator (1978) is

$$C = A'[w^{\alpha}r^{\beta}Q]^{1/(\alpha+\beta)} \tag{23.39}$$

The elasticities α and β can be estimated from the linear form

$$\ln C = a' + \frac{\alpha}{\alpha+\beta}\ln w + \frac{\beta}{\alpha+\beta}\ln r + \frac{1}{\alpha+\beta}\ln Q + \varepsilon \tag{23.40}$$

The rate of growth of technical change may be deduced from variations in the cost functions. Nelson (1984) includes a technology variable in the general cost function specification. Ohta (1974), and Berndt and Khaled (1979) have shown that the rate of technical change is related to the growth rate of total factor productivity.

23.3.2 Specification of production functions with embodied technical change

Capital vintage models have been introduced by Abramovitz (1952), Johansen (1959), Solow (1960), and Kaldor and Mirrlees (1962). Nelson (1964) found an appropriate approximation to the estimation problem raised by Solow (1960). In these models, the capital stock is no longer considered homogeneous. The previous definition for capital and labor was

$$K_t = \sum_{v=V}^{t} K_{vt} \quad \text{and} \quad L_t = \sum_{v=V}^{t} L_{vt} \tag{23.41}$$

where K_{vt} is the vintage stock of capital v, which is still in use at time t. This formulation for K_t is replaced by

$$J_t = \sum_{v=V}^{t} (1 + \lambda_K)^v K_{vt} \tag{23.42}$$

where λ_K is the rate of embodied technical change. We also have

$$K_{vt} = (1 - \delta)^{t-v} I_v \tag{23.43}$$

where δ is a depreciation rate, and I_v is the gross investment in equipment of time v.

Since we suppose a Cobb-Douglas production function with constant returns to scale,

$$Q_t = A e^{\gamma t} L_t^{\alpha} J_t^{1-\alpha} \tag{23.44}$$

then, we can replace L_t and J_t by their respective definitions in (23.41)–(23.43)

$$Q_t = A e^{\gamma t} L_t^{\alpha} [\sum_{v=V}^{t} (1 + \lambda_K)^v (1 - \delta)^{t-v} I_v]^{1-\alpha} \tag{23.45}$$

The expression (23.45) may be easily approximated by

$$Q_t = A e^{(\gamma - \delta\beta)t} L_t^{\alpha} [\sum_{v=V}^{t} [1 + (\delta + \lambda_K)]^v I_v]^{1-\alpha} \tag{23.46}$$

where γ is the rate of disembodied, λ_K is that of embodied technical change, and $\beta = 1 - \alpha$.

In this model, technical change is a combination of embodied and disembodied forms. All technical change is embodied when $\gamma = 0$. The neutrality depends on the elasticity of substitution σ between L and J. It would be neutral if $\sigma = 1$, labor-saving with $\sigma > 1$, and capital-saving with $\sigma < 1$. The assumption of constant return to scale can be modified [Westfield (1966), Solow (1960)].

A clay-clay vintage model has been retained for the French DMS model [INSEE (1978)]. The production function is described by the following four equations

$$Q_{vt} = \mu_{vt} L^{\star}_{vt} \tag{23.47}$$

$$Q_{vt} = \lambda_{vt} I_v \tag{23.48}$$

$$\mu_{vt} = \alpha'(1+\mu_0)^{t-v}(1+\mu_1)^v = \alpha'(1+a')^t(1+b')^v \tag{23.49}$$

$$\lambda_{vt} = \alpha(1+\lambda_0)^{t-v}(1+\lambda_1)^v(1-\xi)^{t-v} = \alpha(1+a)^t(1+b)^v \tag{23.50}$$

where $L^{\star}$ is the required labor for this production; μ and λ are average productivities of inputs.

In equations (23.49)–(23.50), the parameters λ_1 and μ_1 describe the embodied technical change. The evolution of the technical coefficients along with the age of equipment is measured by the parameters λ_0 and μ_0. The parameter ξ represents the depreciation rate of equipment. If we introduce a partial adjustment process on labor

$$L_t / L_{t-1} = (L^{\star}_t / L_{t-1})^{\lambda} \tag{23.51}$$

the system of equations is described by the equations (23.52)–(23.54):

$$Q(t) = \alpha(1+a)^t \sum_{t-m(t)}^{t-1} (1+b)^v I_v \tag{23.52}$$

$$Q_m(t) = \alpha(1+a)^t \sum_{t-m^{\star}(t)}^{t-1} (1+b)^v I_v \tag{23.53}$$

$$L(t) = [L(t-1)]^{1-\lambda} \left[\frac{\alpha}{\alpha'} \left(\frac{1+a}{1+a'} \right)^t \sum_{t-m(t)}^{t-1} \left(\frac{1+b}{1+b'} \right)^v I_v \right]^{\lambda} \tag{23.54}$$

where $m(t)$ is the effective extensive margin (i.e., the age of the oldest capital equipment in operation) and $m^{\star}(t)$ is the extensive margin of capacity (i.e., the age of the oldest profitable capital equipment). The determination of the variable is described in *Figure 23.1*.

In the French METRIC model [INSEE (1980)], the implicit production function is putty-clay. Hence, at each period of time, firms have to determine the convenient technology. Thereafter, the combination of inputs remains unchanged. The costminimization behavior of firms consists in minimizing $C = wL + rK$, given a production function such as $Q_a = e^{\lambda t} L^a I^b$, where Q_a is the supplement of possible added value. We have

$$k = \frac{I}{Q_a} = b e^{-\lambda t} \left(\frac{w}{r} \right)^a \tag{23.55}$$

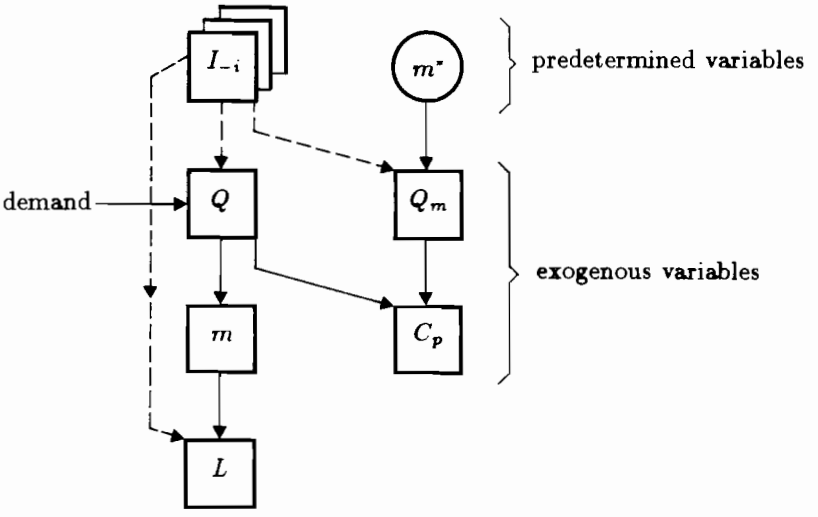

Figure 23.1: Process of the production functions in the DMS model.

Then, variable k enters the investment functions.

A *Solow-Cobb-Douglas production function* with age of capital, time variables, and individual firm effects is used by Mairesse (1978), who attempts to estimate the respective contribution of embodied and disembodied technical progress, simultaneously. The objective of the study is to explain which type of technical change best explains the difference in labor productivity between French manufacturing firms. The retained formulation is

$$\ln \frac{Q}{L} = \alpha \ln \frac{J}{L} + (a + \delta t) \tag{23.56}$$

where J is [just as defined in (23.42)] the stock of capital in efficiency units, and δ is the rate of disembodied progress.

In order to facilitate the estimation of parameters, Mairesse (1978) introduces several simplifications. Thus, capital J is approximated by

$$\ln J_t = \ln K_t + \lambda_K (t - A_t) \tag{23.57}$$

where A_t is the average age of capital. Thus, the model takes the following specification

$$\ln \frac{Q}{L} = \alpha \ln \frac{K}{L} - \beta A + \gamma t + a \tag{23.58}$$

In this specification, the elasticity of scale and the elasticity of substitution of capital for labor are both equal to one [see Mairesse (1978)]. Then, the respective rates of disembodied and embodied technical change are $\delta = \gamma - \beta$ and $\lambda_K = \beta/\alpha$.

Econometrically, when the estimates $\hat{\delta} = \hat{\gamma} - \hat{\beta}$ and $\hat{\lambda}_K = \hat{\beta}/\hat{\alpha}$ are derived from least squares techniques, the estimates $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ are consistent. We then approximate

$$\hat{\sigma}_{\hat{\delta}} = \sqrt{\hat{\sigma}_{\hat{\beta}}^2 + \hat{\sigma}_{\hat{\gamma}}^2} \quad \text{and} \quad \hat{\sigma}_{\hat{\lambda}_K} = \hat{\sigma}_{\hat{\beta}}/\hat{\alpha} \tag{23.59}$$

Since the model has been fitted to individual time series data, it concerns the firm i at time t, such as

$$\ln \frac{Q_{it}}{L_{it}} = \alpha \ln \frac{J_{it}}{L_{it}} - \beta A_{it} + \gamma t + a + (\varepsilon_i + \eta_{it}) \tag{23.60}$$

where the additive error term $\varepsilon_i + \eta_{it}$, in the log-linear form, is supposed to be composed of a constant individual error ε_i and a standard regression disturbance η_{it}.

Table 23.6 gives some of the results we may be interested in. The empirical results, achieved upon two different samples of firms, indicate that the rate of embodied technical change derived from between-estimates exceeds the rate achieved from within-estimates. The disembodied technical change registers the invers result.

The *translog functions*, which are estimated by Nelson (1984), also incorporate the effect of the average age of the plants in steam electric generation and of a time trend for the disembodied technical change. The translog cost function [Griliches and Ringstad (1971)] may be defined by

$$C = (\alpha_0 + \sum_i \alpha_i \ln w_i + \frac{1}{2} \sum_i \sum_j \gamma_{ij} \ln w_i \ln w_j) Q \tag{23.61}$$

where w_i $(i = 1, \ldots, n)$ are the wages of n inputs.

Welfe (1985) does not find any chance of *applying the vintage approach*, because of the lack of appropriate data on the age distribution of equipment. Moreover, the introduction of a time trend to capture the joint effects of embodied and disembodied technical change seems erroneous, since embodied technical progress has no reason to develop at constant growth rates. The construction of an indicator of embodied technical progress $(\tilde{K})$ is based on the assumption that the rate of growth of net output due to embodied technical change depends on the rate of growth of fixed capital. Hence, we have

$$\tilde{K}_t = \dot{K}_t + \tilde{K}_{t-1} \quad \text{where } \dot{K}_t = \frac{K_t - K_{t-1}}{K_{t-1}} \tag{23.62}$$

or

$$\tilde{K}_t = \sum_{i=1}^{t} \dot{K}_i + \tilde{K}_0 \quad \text{with } \tilde{K}_0 = 1 \tag{23.63}$$

Table 23.6: Estimates of the Solow production function for two French manufacturing firms over the period 1966–1975, according to Mairesse (1978). (The t-statistics are shown in parentheses.)

Method [a]	α	β	γ	a	*Technical change* δ	λ_K	R^2
Sample 1: SEDES / Caisse des Dépôts							
A	0.344	0.024	0.036	2.38	1.2	7.0	0.602
	(38.2)	(8.0)	(18.0)	(59.5)	(3.0)	(7.8)	
B	0.224	0.013	0.040	–	2.7	5.8	0.566
	(8.6)	(3.3)	(20.0)		(6.8)	(3.2)	
C	0.350	0.026	–	2.57	1.4	7.4	0.619
	(14.0)	(3.7)		(21.4)		(3.7)	
Sample 2: Crédit National							
A	0.355	0.041	0.030	2.56	−1.1	11.5	0.538
	(39.4)	(13.7)	(15.0)	(64.0)	(−2.8)	(14.4)	
B	0.165	0.000	0.043	–	4.3	0.0	0.494
	(9.2)		(21.5)		(10.8)		
C	0.376	0.057	–	2.76	−3.1	15.2	0.583
	(16.3)	(5.7)		(23.0)		(5.6)	

[a]The estimation is based on $q_{it} = \ln(Q_{it}/L_{it})$ (method A), on the deviations from the means $q_{it} - q_{i\cdot}$ (method B), and on the means $q_{i\cdot}$ (method C).

The production function takes the form

$$Q_t = Q(.)\tilde{K}_t^c \tag{23.64}$$

where c is a parameter. We have

$$\dot{Q}_t = c\dot{\tilde{K}}_t = c\,\frac{\sum_{i=1}^{t}\dot{K}_i - \sum_{i=0}^{t-1}\dot{K}_i}{\tilde{K}_{t-1}} = c\frac{\dot{K}_t}{\tilde{K}_{t-1}} \tag{23.65}$$

The application to Poland justifies including the effects of imported technology, which increases the efficiency of production. Welfe (1985) introduces an additional variable, which is defined as the lagged share of imported machinery equipment ($M7$) in the total investment outlays for machinery equipment (JV). The corresponding production function is

$$Q_t = AL_t^a K_t^b \tilde{K}_t^c \left(\frac{M7}{JV}\right)_{t-1} e^{dt} \tag{23.66}$$

23.3.3 Specification of production functions with endogenous technical change

Endogenization of technical change has been achieved in different ways, such as the introduction of specific relations on technical change, or specific variables in the production function, or the endogenization of the same parameters. Sato and Suzawa (1983) develop the following model

$$Q_t = T_t F(L_t, K_t) \tag{23.67}$$

$$\dot{T}_t = h(B_t, \theta_1) \tag{23.68}$$

$$\dot{B}_t = f(\theta_2) - \mu B_t \tag{23.69}$$

where T_t is the state of applied knowledge, B_t is the stock of basic knowledge, with the assumption that $h(0, \theta_1) = 0$, θ_1 is the current investment in applied research, and θ_2 is the current investment in basic research.

The introduction of specific variables of technical change among the inputs modifies the specification of a production function [Griliches (1979)] like

$$Q = F(L, K, \tilde{K}) \tag{23.70}$$

where $\tilde{K}$ is the current state of technical knowledge. Moreover, there may exist lags between $\tilde{K}$ and R&D expenditures such as

$$\tilde{K} = G(R_{-i}) \tag{23.71}$$

The variable R&D, which has been extensively introduced, will be examined later. For Sato and Nono (1982), a production function not only depends on current research, but also on a stock of technical knowledge that has been accumulated over several years. The learning process and education both affect the quality of labor. Brown and Conrad (1967) make the parameters of a CES production function dependent upon education. Their empirical results show significant effects.

R&D *expenditures* seem to have significant effects on growth. Griliches (1964) introduces public expenditures on agricultural research as an input variable in a Cobb-Douglas production function. The cross-section estimations give a significant elasticity of R&D on a growth rate of 0.05. Mansfield (1968) extends the approach to the manufacturing sector. Minasian (1969) regresses the total productivity growth on current and lagged R&D for a cross-section of chemical and pharmaceutical firms. In another attempt to estimate the contribution of R&D to productivity growth, Griliches (1979) introduces a sequence of past R&D expenditures and also takes the induced effects of R&D into account. In his FUGI gobal macroeconometric model, Onishi (1985) precisely specifies the direct impact that R&D have, not only on labor productivity, but also on non-housing fixed investment, and on trade relationships between Japan and the USA. To give an example, the productivity function that has been estimated for Japan over the period 1973–1984 is

$$\ln \frac{Q}{L} = \underset{(4.5)}{1.6} + \underset{(2.3)}{0.3} \ln\left(\frac{Q}{L}\right) + \underset{(1.8)}{0.4} \frac{\sum_{i=0}^{5} R_{-i}}{L} - \underset{(-2.3)}{0.04} \left(\frac{p_e}{p}\right)_{-1} \tag{23.72}$$

where R_{-i} is the R&D at constant price over the past five years, p_e is the average crude oil export price, and p is the deflator of the GDP.

Schott (1978) formalizes *the relationship between industrial* R&D *and factor demands* by private sector producers. The desired levels of factor demand are derived from the following cost minimisation problem:

$$\min C = LHw + Kc + \tilde{K}b \quad \text{s.t. } Q_m = kL^{\alpha_1} H^{\alpha_2} K^{\alpha_3} C_p^{\alpha_4} \tilde{K}^{\alpha_5} \tag{23.73}$$

where C are the total costs, L is the number of persons employed, H are the hours per person, w is the hourly wage rate, K is the capital stock, c is the user cost of capital, $\tilde{K}$ is the technical knowledge stock, b is the user cost of technical knowledge, Q is the potential industrial output, k is a constant, and C_p is the rate of capital utilization. Thus, the optimum long-run demand function for technical knowledge is [Schott (1978)]

$$\tilde{K}^{\star} = k_5 Q_m^{1/\rho}\, w^{(2\alpha_1-\alpha_2)/\rho}\, w'^{(\alpha_2-\alpha_1)/\rho}\, c^{(\alpha_3-\alpha_4)/\rho}\, C'^{\alpha_4/\rho}\, b^{-(\alpha_1+\alpha_3)/\rho} \tag{23.74}$$

with constants ρ, α_1, α_2, $\alpha_5 > 0$, and k_5, $C' = c \circ \delta(U) = \delta c/\partial U > 0$ is a composed function $c \circ \delta$ of U where δ is the depreciation rate, and $w' = \partial w/\partial H > 0$.

Griliches (1986) incorporates both disembodied and embodied technical knowledge in a Cobb-Douglas production function: $Q_t = Ae^{\lambda t}\tilde{K}_t^{\alpha}K_t^{\beta}L_t^{1-\beta}$ where Q_t is the output (sales or value added), [Griliches mentions difficulties with the measurement of Q in a research-intensive industry where the "quality may be rampant".] λ is the rate of disembodied "external" technical change, and $\tilde{K} = \sum_i w_i R_{t-i}$, a mean of accumulated and still productive research capital [Griliches (1986)] where R is the deflated gross investment in research.

Thus, $\tilde{K}$ measures the distributed lag effect of past investments on productivity. Lag effects occur since there exist lags between investment in research and the actual invention of a new technique, between invention and development and the acceptance of the new technique (or product) by the market [Griliches (1986)].

The elasticity of output with respect to R&D investment was about 0.07 in 1957–1965 on the basis of cross-section data on 883 US manufacturing corporations [Griliches (1980)]. Nadiri (1980) is also referring to a three-input Cobb-Douglas production function and estimates a "basic model" of labor productivity growth. [For Giersch and Wolter (1983), the correlation between the growth of labor productivity per employed person to the capital/labor ratio (in Cobb-Douglas accounting) or to the growth of real output, ceases to be statistically significant in 1973–1979, contrary to the period 1960–1973.]

$$\ln P = a_0 + a_1 \ln K/L + a_2 \ln U_t + a_3 \triangle \ln U_t + a_4 \ln R + a_5 t \tag{23.75}$$

where P is the level of ouput per manhour, K/L is the ratio of gross capital stock to manhours, U is the gap between the rates of growth of actual and normal output, R is a measure of aggregate stock of R&D (Kendrick's definition), and t is the time trend, which is a proxy for the disembodied technical change.

In the regression analysis, the time trend is dropped because of multicollinearity effects with the variable R&D in 1949-1978. The slowdown of R&D contributes for about one-fourth of the slowdown in productivity of US aggregate productivity growth. Distinctions may be found in the studies between federal and company R&D [Lichtenberg (1984)].

Giersch and Wolter (1983) test *the technology gap hypothesis* using international cross-section data from 1964 to 1973 and 1973 to 1979. For industry, the authors get

1. period 1964–1973: $\dot{P} = 25.33 - \underset{(-2.9)}{4.868 \ln GAP64}$, $\overline{R}^2 = 0.43$

2. period 1973–1979: $\dot{P} = 3.0 - \underset{(-0.01)}{0.019 \ln GAP63}$, $\overline{R}^2 = 0.00$

where $\dot{P}$ is the average annual growth of the real output per employee; and $GAP64$ and $GAP73$ are per capita income for an individual sample country in percentage of per capita incomes of the United States in 1964 or 1973, respectively, valued at purchasing power parities.

Lindbeck (1983) introduces a variant of technological gap, noting the changes in Denison-type residuals between the US and other countries. The author finds that the distinction between effects of capital accumulation and technological catch-up is arbitrary.

23.3.4 Specification of technical change in multisectoral and disaggregated models

The interindustry relationship between total factor productivity growth and R&D is reconfirmed in the study of Griliches and Lichtenberg (1984). Schmookler (1966) relates the improvements of a given firm's performance to the R&D expenditures made by other firms, as well as linking the incorporated R&D to purchased intermediate goods. Griliches and Lichtenberg (1984) compare their results to those of Scherer (1982), who proved the importance of "imported" R&D in the explanation of productivity growth. Three types of R&D are distinguished in this study: "own" process of R&D, "own" product R&D, and product R&D embodied in the inputs purchased from other industries. Some results of the regressions of total factor productivity growth rate on these three categories of R&D are presented in *Table 23.7* for two subperiods.

Input-output models receive an important revision by Sato and Ramachandran (1980) in their survey on input-output analysis. [An extensive description of input-output models and modeling can be found in Grassini and Smyshlyaev (1983)]. Technical change problems in multisectoral models are studied by Bacharach (1970), who examines biproportional changes in inputs coefficients and finds that product-oriented R&D within a firm is less influential on performance than either process R&D or R&D embodied in purchased inputs. Caravani (1981) proposes a more general formulation with a production innovation model. Craven (1983) derives conditions on the changes in input coefficients, so that the economy remains productive all the time, and defines Harrod-neutral change in a Leontief system. Under certain conditions, it is shown that Harrod neutrality increases output possibilities as fast as other forms of technical change. Special emphasis is given to technological change in the Finnish long-range model system [Forssell *et al.* (1983)]. The production functions in the price model (an extended input-output price model) are clay-clay vintage production functions with embodied and disembodied technical change. A discussion of structural changes with input-output models can be found in Smyshlyaev (1983).

Input-output coefficients may depend on a number of variables. In the model proposed by Arrow and Hoffenberg (1959), these variables are real disposable income, trend variations due to taste and technological changes, and learning effects depending on previous behavior. Ozaki (1976) considers the effects of technical change in economics through the variations of input-output coefficients. Recent experiences with changes in input-output coefficients are notably reported by Tomaszewicz (1983) for Poland and by Forssell (1983) for Finland. Forssell mentions the difficulties we may have in trying to separate the causes of changes in the specification

$$a_{ij}(t) = f_{ij}(K_j, r_j, M_j, \triangle Q_j) \tag{23.76}$$

Table 23.7: Regressions of total factor productivity on R&D: intensity variables [a] on cross-sectional industry data [b], according to Griliches and Lichtenberg (1984).

Period	*"Own" Process R&D*	*"Own" Product R&D*	*"Imported" Product R&D*	*Constant*
1959–63 to	0.762	0.211	0.289	0.093
1964–68	(2.8)	(2.5)	(0.6)	(4.5)
1964–68 to	0.578	0.040	0.687	0.005
1969–73	(2.8)	(0.6)	(1.9)	(0.3)
1969–73 to	0.384	0.299	0.465	−0.1
1974–78	(1.4)	(3.4)	(0.9)	(-4.9)

[a] Expenditure per unit of output.

[b] 193 US manufacturing industries.

where K_j is the technical development of industry j, r_j is the relative price input of industry j, M_j is the product mix, and $\triangle Q_j$ is the change of output.

Fontela and Pulido (1986) briefly survey the main results that have been achieved by more extensive studies on the variation of technical coefficients. In particular, they point out that it has been proved that changes in technical coefficients take place very slowly. The reason may be the delay necessary for the adoption of an innovation by an entire economic sector. Indeed, the process of technical change is only measurable in terms of new investment in equipment. The explanatory model remains a possible approach to the input-output coefficient. But Fontela and Pulido (1986) also consider other technological variables besides price effects, such as the stock of technological knowledge, the rythm of technological diffusion, and technological expenditures. In a disaggregated model, initiated by Krelle [see Kiy (1984)], the input-output coefficients depend on the price ratios, the capital/labor ratio, the degreee of capacity utilization, and other factors [see Krelle (1964) and Krelle *et al.* (1969)]. Hence, technical progress is only induced by higher capital/labor ratios, which in turn are induced by changes in the capital/labor cost ratio. In this model, the input coefficients are specified in a log-linear form. The input coefficients are determined by firms' cost minimization, given a Cobb-Douglas production function with constant return to scale, such as

$$\min_{X_{ij}} \sum_{i=1}^{n} p_i X_{ij} \quad \text{s.t.} \quad Q_j = \alpha_{0j} \tau_j \prod_{i=1}^{n} X_{ij}^{\alpha_{ij}} \tag{23.77}$$

with $\alpha_{ij} > 0$ and $\sum_{i=1}^{n} \alpha_{ij} = 1$, where X_{ij} stands for factor inputs, τ_j for the state of technology.

The specification of the domestic input coefficients in the model is

$$a_{ij} = c_{0j} (\overline{p_i^M / p_i})^{\alpha_{ij}^M} (\overline{K_j / L_j})^{\beta_j} \delta_j^{\gamma_j} \tag{23.78}$$

where $\overline{p_i^M/p_i}, \overline{K_j}$, and $\overline{L_j}$ are the mean values calculated at each time t on the 4 or 6 preceding years, M is an index for imports, and δ is the degree of capacity utilization.

Thus, for $a_{3,11}$ (3 = chemical industry, 11 = transport) we have

$$a_{3,11} = \overline{p_3^M/p_3}^{1.0027}(\overline{K_{11}}/\overline{L_{11}})^{1.5428}e^{-10.369} \tag{23.79}$$

the respective t-statistics of the estimates being 1.94, 9.1, and -12.7. The real technical coefficient $a_{3,11}$ then increases with induced technical progress, since the elasticity of the capital/labor ratio is 1.5428.

23.4 Estimation Techniques and Problems of Technical Change

23.4.1 Time series and cross-section data bias

Several sources of bias are common to time series and to cross-section data. In both cases, the stability estimates depend heavily on aggregation techniques. The grouping and weighting of individual observations may not remain unchanged over time, notably because of the modifications in technical characteristics. But the evaluation of capital inputs in both cases raises severe difficulties for both types of data, as already mentioned by Walters (1963a) and Nadiri (1970) in their surveys, respectively, on production function and total factor productivity. This is due to the various kinds of equipment. The imperfection ot time series and cross-section data generally imposes alternative measures and adjustments. Denison (1967) evaluates the value of capital in terms of cost, so that increases in quality will be reflected in technical progress, rather than in capital input. Several sources of bias are common to time series and cross-section evaluation of the estimated elasticity of substitution in a CES production function [Nadiri (1970)]:

1. The data read out in the estimation refer to the "average practice" factor proportion to input prices, while the marginal productivity condition supposes the "best practice" factor proportions.

2. The estimates heavily depend on interclass and intraclass elasticities of substitution, since the empirical production function may combine Cobb-Douglas and CES functions.

Several sources of bias differ as between time series and cross-section data. The deflation of time series that are observed in current value may raise severe difficulties. Thus, since there exists no obvious price of capital equipment for both interindustry and interfirm studies, there is no need of deflated data [see Walters (1963a)]. Of course, this is not the case for international cross-section studies. Moreover, the changing utilization of capital may be rather less important in cross-section studies than in time series ones. Generally, errors of measurement in the variables are more important for microdata.

The comparison between time series and cross-section estimates may be considered a test of "fitness" of the model specification. Indeed, the results achieved by Douglas in 1948 on both time series and cross-section data, for the US manufacturing industry, showed convergent results. Nerlove (1967) also considers both types of data to estimate a CES production function for the postwar period. The time series estimates of the elasticity of substitution are less than unity. The results vary among the different industries, but remain close to unity when they proceed from cross-section data. Other follow-on results, which are reported by Nadiri (1970), reach the sample conclusion with respect to the elasticity of substitution.

23.4.2 Methods of estimating technical change

Single-equation least squares estimates of a system described by a production function and marginal poductivity relations will be consistent and unbiased only under certain statistical circumstances. We have

$$Q = F(L, K)\, u_0 \tag{23.80}$$

$$\frac{\partial F}{\partial L} = \left(\frac{w}{p}\right) u_1 \tag{23.81}$$

$$\frac{\partial F}{\partial K} = \left(\frac{r}{p}\right) u_2 \tag{23.82}$$

where u_0, u_1, and u_2 are random variables, reflecting errors in the production process and relative factor prices. Under the normality assumption and independence of errors, as $E(u_0, u_1) = E(u_0, u_2) = 0$, simple least squares estimates will be consistent and unbiased. The model

$$Q_t = Ae^{\lambda t} L_t^{\alpha} K_t^{\beta} u_t \tag{23.83}$$

where technical change is supposed to be Hicks-neutral, may be treated as a log-linear relationship since

$$\ln Q_t = \ln A + \lambda t + \alpha \ln L_t + \beta \ln K_t + \ln u_t \tag{23.84}$$

where $\ln u_t$ is treated as an additive error term with zero mean. The CES production function, which is highly nonlinear, cannot be made linear by simple transformations.

The stepwise procedure of estimation is used, notably by Diwan (1963) and also by Bodkin and Klein (1967). The method is applied to Cobb-Douglas and CES production functions with disembodied technical change. The marginal productivities, which are used in the procedure, result from a cost minimization. Hence, the first step consists in estimating

$$\ln\left(\frac{r_t}{w_t}\right) = a_0 + b\ln\left(\frac{L_t}{K_t}\right) + \varepsilon_t \tag{23.85}$$

where

$$a_0 = \ln\frac{\beta}{\alpha} \quad \text{and} \quad b = 1 \tag{23.86}$$

in the Cobb-Douglas production function with unconstrained returns to scale; or

$$a_0 = \ln\frac{\delta}{1-\delta} \quad \text{and} \quad b = 1 + \rho \tag{23.87}$$

in the CES production function. In the second step, we replace the values of $(\frac{\hat{\beta}}{\alpha})$ obtained by (23.86) and of $\hat{\delta}, \hat{\rho}$ by (23.87) in the corresponding production functions, which are rewritten

$$\ln Q_t = \ln A + \lambda t + \alpha[\ln L_t + \left(\frac{\hat{\beta}}{\alpha}\right)\ln K_t] + \ln u_t \tag{23.88}$$

the Cobb-Douglas production function, where A, λ, and α are estimates of the second step,

$$\ln Q_t = \ln A + \lambda t - \frac{\mu}{\hat{\rho}}\ln[\hat{\delta}K_t^{-\hat{\rho}} + (1-\hat{\delta})L_t^{-\hat{\rho}}] + \ln v_t \tag{23.89}$$

the CES production function, where A, λ, and μ are the estimates of the second step.

A nonlinear maximum likelihood procedure is also recommended by Bodkin and Klein (1967). The system may be described by a Cobb-Douglas production function (23.90) and a marginal productivity condition (23.91), which is deduced from a cost minimization in competitive factor markets

$$Q_t = Ae^{\lambda t}L_t^{\alpha}K_t^{\beta} + u_t \tag{23.90}$$

$$\frac{r_t}{w_t} = \frac{\beta}{\alpha}\frac{L_t}{K_t}v_t \tag{23.91}$$

where u_t, v_t are random disturbances (v_t reflects an incomplete minimization).

Then, we form a joint likelihood function, stating that L_t and K_t are endogenous and Q_t, r_t, w_t exogenous. The likelihood function to be maximized with respect to the parameter values is

$$L(A, \alpha, \beta, \lambda, \Omega | r_t, w_t, Q_t) \tag{23.92}$$

where Ω is the population variance-covariance. Thereafter, the likelihood function will be solved iteratively until achievement of the lowest sum of squared errors. The convergence is very sensitive to the choice of initial conditions [Bodkin and Klein (1967)] [The authors also consider the use of a Cobb-Douglas production function with constrained returns to scale and a CES productivity function following the iterative procedure of Eisenpress and Greenstadt (1966)].

A full-information maximum likelihood method may be used for a constrained joint estimation problem, derived from the production function and factor demand determination; see Coen and Hickman (1970). The production function may be viewed as a "planning relation" between the expected output Q_m, and desired input labor $L^\star$ and capital stock $K^\star$. The Cobb-Douglas function with disembodied neutral technical progress is

$$Q_m = Ae^{\gamma t}(K^\star)^\alpha (L^\star)^\beta \tag{23.93}$$

Under competitive factor markets, costminimization behavior by firms will require the ratio of marginal product of labor and capital to be equal to the ratio of their expected prices. Then, we derive the input relations

$$L^\star = a_0 \left(\frac{r^\star}{w^\star}\right)^{a_1} (Q_m)^{a_2} e^{-a_3 t} \tag{23.94}$$

$$K^\star = b_0 \left(\frac{r^\star}{w^\star}\right)^{b_{-1}} (Q_m)^{b_2} e^{-b_3 t} \tag{23.95}$$

where $a_0 = [(\beta/\alpha)^\alpha A^{-1}]^{1/(\alpha+\beta)}$ and $b_0 = [(\alpha/\beta)^\beta A^{-1}]^{1/(\alpha+\beta)}$, $a_1 = \alpha/(\alpha+\beta)$ and $b_1 = \beta/(\alpha+\beta)$, $a_2 = b_2 = 1/(\alpha+\beta)$, and $a_3 = b_3 = \gamma/(\alpha+\beta)$. The singleness of the production function then prescribes the following restrictions: $\frac{a_0 a_1}{1-a_1} = b_0$, $a_1 + b_1 = 1$, and $a_2 = b_2$ $a_3 = b_3$. The estimation form of the system will be achieved with the determination processes of the expected values $r^\star/w^\star$ and Q_m and those of the desired inputs $L^\star$ and $K^\star$. Expected values $X^\star$ and desired values $Y^\star$ may be derived from

$$X_t^\star = \sum_{i=0}^{1} w_{ji} X_{t-i}, \quad j = 1,2 \tag{23.96}$$

where w_{ji} are weights associated to both equations (23.94) and (23.95); and

$$Y_t/Y_{t-1} = (Y_t^\star/Y_{t-1})^{\lambda_j} u_j, \quad j = 1,2 \tag{23.97}$$

where u_j are stochastic terms.

Then, the estimation form may be written as

$$\frac{L}{L_{-1}} = a_0^{\lambda_1} \left[\sum_{i=1}^{2} w_{1i} \left(\frac{r}{w}\right)_{-i+1}\right]^{\lambda_1 a_1} \left(\sum_{i=1}^{2} k_{1i} Q_{-i+1}\right)^{\lambda_1 a_2} e^{-\lambda_1 a_3 t} L_{-1}^{-\lambda_1} v_1 \tag{23.98}$$

$$\frac{K}{K_{-1}} = b_0^{\lambda_2}\left[\sum_{i=1}^{2} w_{2i}\left(\frac{r}{w}\right)_{-i+1}\right]^{-\lambda_2 b_1}\left(\sum_{i=1}^{2} k_{2i}Q_{-i+1}\right)^{\lambda_2 b_2} e^{-\lambda_2 b_3 t}K_{-1}^{-\lambda_2}v_2 \qquad (23.99)$$

where v_1 and v_2 are stochastic terms.

The estimation of (23.98)–(23.99) can be achieved by full-information maximum likelihood method, using the algorithm proposed by Eisenpress and Greenstadt for nonlinear systems. [The algorithm modifies the Newton-Raphson method.] A stepwise procedure may also be realized by successive estimations and identification of parameters, as proposed by Coen and Hickman (1970).

Iterative procedures among trial values for the parameters are extensively used in the estimation of vintage models [Benassy *et al.* (1975) and Vilares (1980)]. Thus, with equation (23.46) we had

$$Q_t = Ae^{(\gamma-\delta\beta)t}L_t^{\alpha}\left[\sum_{v=V}^{t}[1+(\delta+\lambda_K)]^v I_v\right]^{1-\alpha}$$

Trial values are assumed for α, δ, and γ in order to estimate λ_K by means of the iterative procedure. Berglas (1965) and Solow (1960) show that the estimates of λ_K are very sensitive to trial values.

The French DMS model has been described by equations (23.52)–(23.54). The estimation procedure kept for good by INSEE (1978) consists in different steps on an iterative process: parameters have been fixed *a priori*, such as b and m_t^*; trial values are taken for b' and extensive margins m_t are determined iteratively; and coefficients α, α', a, and a' are econometrically determined.

The error components model combines time series and cross-sectional data. The stochastic specification of the Solow production function in Mairesse's (1978) study considers the overall errors as composed of individual effects, u_i, specific to firms and of the standard regression error w_{it}. The model, which has already been presented [equation (23.60)] is

$$\ln\left(\frac{Q_{it}}{L_{it}}\right) = \alpha\ln\left(\frac{J_{it}}{L_{it}}\right) - \beta A_{it} + \gamma t + a + e_{it} \qquad (23.100)$$

with $e_{it} = \varepsilon_i + \eta_{it}$. Since the covariance matrix of errors has a convenient structure, we may compute the quasi-generalized least squares, which should be more efficient than the ordinary least squares when large samples are involved.

A modified ridge regression method [Maddala (1977)] is introduced by Lee (1983) in order to estimate a translog production function on the basis of farm records from a survey during the period 1955–1975 in Japan. [Another application of this method to Japan on a regional basis is proposed by Uno (1976).] Lee uses the translog production function

$$\ln q_f = \alpha_0 + h_r + z_t + \sum_i \alpha_{it} \ln x_{if} \frac{1}{2} \sum_i \sum_j \beta_{ij} \ln x_{if} \ln x_{jf} + u_f \tag{23.101}$$

with $f = 1, \ldots, N$ and $r = 2, \ldots, R$, where f refers to the sample farm, r to the prefecture, and i and j stand for production factors. Factor share functions are derivatives, provided that the production function is linear homogeneous and that the marginal productivities of factors inputs are equal to their prices. Thereafter, the factor share functions can be used to estimate the coefficients of the production function. The resulting estimation of the modified ridge regression method by Lee (1983) is

$$\hat{b} = \hat{\beta} + (X'X + kI)^{-1} R'[R(X'X + kI)^{-1} R']^{-1}(r - R\hat{\beta}) \tag{23.102}$$

where $\hat{\beta} = (X'X + kI)^{-1} X'y$, $r = R\hat{\beta}$; R and r stand for linear restrictions on β; and k is a nonnegative biasing factor.

23.4.3 Econometric problems

Technical progress estimation problems refer to a larger class of problems regarding the induction of production or productivity functions by means of econometric techniques applied to time series, cross-section, or panel data. A constant parameter model may be attached to a time trend in a production function, and thus will reflect neutral disembodied technical change. The parameters of a model may also be time-varying, since we are interested in reflecting nonneutral technical change. Moreover, other specific problems with technical change will arise in econometrics, as we attempt to distinguish between the effects of disembodied and embodied technical change. In any case, the econometric problems commonly linked to the estimation of a production function may necessarily have severe consequences in the evaluation of technical change, since it can be introduced in the equation as a variable (time trend, R&D expenditures, etc.) or considered a proxy of variations in global productivity.

Instability and inconsistency of the parameters may arise from incorrect measurement in the variables of the production function. The lack of direct evaluations, overdepreciation of equipment, and difficulties in adjusting the inputs for changes in utilization will introduce errors, whose consequences will generally be to underestimate the contribution of capital to growth and to overestimate the elasticity of labor because of incorrect weighting [Griliches (1963)].

An *omission of variables* may result in a traditional production function limited to labor and capital inputs. Nadiri (1970) indicates that the omission of materials in the production function will often lead to a positive bias in the estimation of the returns to scale as well as the elasticity of substitution between factors. Especially, the omission of R&D expenditures by the goverment may understate the growth rate of capital and, hence, overstate the rate of growth of global productivity.

Identification may be made difficult, owing to the lack of degrees of freedom as in time series data. Thus, it will be difficult to separate adequately the effects of parameters in

the production function [Jorgenson (1966)]. Simultaneity and multicollinearity problems that may arise will make the estimates very imprecise. Individual data may be the best way to reduce these problems.

The *specification of the production function* can give unbiased, efficient, and consistent coefficients. But the parameter estimates will lose the unbiased property, since nonlinear operations are involved in deriving them from the regression coefficients [Wallis (1973)]. Indeed, we have $E(1/b) \neq 1/E(b)$. The parameter estimates remain consistent.The Durbin-Watson statistic, usually associated with time series, may be a useful indication of misspecification in cross-section data.

Sensitivity analysis thus seems to be very helpful, since the results of production functions are sensitive to the measurement of variables, to the data, and to the alternative specification as well as to econometric methods.

23.5 Conclusions

This survey on econometric approaches to technical change aimed at giving us useful insights on this crucial problem of economic growth. Technical progress was early confined to a time trend in a production function on the aggregate level or reduced to the variations of the global productivity indexes. Since then, numerous studies have been devoted to measurement problems in the output and input factors. Errors in measurement and lack of data still raise great difficulties in empirical studies.

Improvements in the economic theory and practice of measuring technical change have certainly been stimulating, since new classifications and concepts have been introduced. Hence, the understanding of technical change has been greatly enhanced by aggregated and multisectoral models, as well as by input-output approaches. The interrelation of industries and firms and input-output models may help to clarify the process of technical change and further our understanding of structural change as well. Capital vintage models have helped to specify the process more precisely, since the heterogeneous nature of capital and firm's decisions regarding the allocation of factors may be taken into account. Endogenous technical progress is another class of technical change that has also been extensively studied. Thus, R&D expenditures of firms and governments tend to be introduced into models' production functions and other equations, as well as the investment of firms and sometimes international trade relations.

The econometric approach to technical change still raises numerous questions and difficulties, because of measurement errors in the variables, and the complexities of the specification of the equations and their estimation. The estimation problems have many origins, such as nonlinearities, interdependence of parameters, or the need for a more complete system (which may increase the estimation issues still further). However, the corroborative results tend to be encouraging, since different techniques and kinds of data have already been used.

References

Abramovitz, M. (1952), Economics of Growth, pp 132-178 in: B.Haley (ed.), *A Survey of Contemporary Economics*, Vol.2. Chicago: R.D.Irwin.

Ahmad, S. (1966), On the theory of induced invention. *Economic Journal*, **76** (302), 344-357.

Arrow, K.J. (1962), The economic implications of learning by doing. *Review of Economic Studies*, **29** (3), 155-173.

Arrow, K.J. and Hoffenberg, M. (1959), *A Time Series Analysis of Interindustry Demands*. Amsterdam: North-Holland.

Arrow, K.J., Chenery, H.B., Minhas, B.S., and Solow, R.M. (1961), Capital-labor substitution and economic efficiency. *Review of Economic and Statistics*, **43** (3), 225-250.

Bacharach, M. (1970), *Biproportional Matrices and Input-Output Change*. Cambridge, UK: Cambridge University Press.

Baily, M.N. (1984), Will productivity growth recover? Has it done so already? *American Economic Review*, **74** (2), 231-235.

Barger, H. (1976), Embodied versus disembodied improvements. *Review of Economics and Statistics*, **58** (3), 372-375.

Beckmann, M.J. and Sato, R. (1969), Aggregate production functions and types of technical progress: A statistical analysis. *American Economic Review*, **59** (1), 88-101.

Benassy, J.-P., Fouquet, D., and Malgrange, P. (1975), Estimation d'une fonction de production à génération de capital (Estimation of a vintage production function for the French industry — English summary). *Annales de l'INSEE*, **19**, 3-52.

Berglas, E. (1965), Investment and technological change. *Journal of Political Economy*, **73** (2), 173-180.

Berndt, E.R. and Christensen, L.R. (1974), Testing for the existence of a consistent aggregate index of labor inputs. *American Economic Review*, **64**, 391-464.

Berndt, E.R. and Khaled, M.S. (1979), Parametric productivity measurement and choice among flexible functional forms. *Journal of Political Economy*, **87** (6), 1220-1245.

Bodkin, R.G. and Klein, L.R. (1967), Nonlinear estimation of aggregate production functions. *Review of Economic and Statistics*, **49** (1), 28-44.

Brown, M. (1966), *On the Theory and Measurement of Technological Change*. New York: Cambridge University Press.

Brown, M. and Conrad, A. (1967), The influence of research and education on CES production relations, pp. 341-372 in: M. Brown (ed.), *The Theory and Empirical Analysis of Production*. New York: National Bureau for Economic Research.

Brown, M. and Popkin, J. (1962), A measure of technological change and returns to scale. *Review of Economics and Statistics*, **44** (4), 402-411.

Caravani, P. (1981), Technology and technical change in a dynamic production model. *Regional Science and Urban Economics*, **11** (3), 335-349.

Carter, A.P. (1970), *Structural Change in the American Economy*. Cambridge, MA: Harvard University Press.

Carter, A.P. (1974), Energy, environment, and economic growth. *Bell Journal of Economics*, **5** (2), 578-592.

Carter, A.P. (ed.) (1976), *Energy and Environment: A Structural Analysis*. Hanover, NH: University Press of New England for Brandeis University Press.

Chenery, H.B. (1949), Engineering production functions. *Quarterly Journal of Economics*, **63** (4), 507-531.

Chenery, H.B. (1953), Process and production functions from engineering data, pp. 297-325 in: W.Leontief (ed.), *Studies in the Structure of the American Economy*. Oxford: Oxford University Press.

Cobb, C.W. and Douglas, P.H. (1928), A theory of production. *American Economic*

Review, **18** (Suppl.), 139-165.
Coen, R.M. and Hickman, B.G. (1970), Constrained joint estimation of factor demand and production functions. *Review of Economics and Statistics*, **52** (3), 287-300.
Coen, R.M. and Hickman, B.G. (1980), Investment and growth in an econometric model of the United States. *American Economic Review*, **70** (2), 214-234.
Courbis, R. (1983), *Sur la fonction de production sous-jacente à la détermination de l'emploi industriel dans le modèle MOGLI*, GAMA Working Paper, University of Paris-Nanterre, 454.
Courbis, R., Fonteneau, A., Le Van, C., and Voisin, P. (1980), Le modèle MOGLI. *Prévision et Analyse Economique, Cahiers du GAMA*, **1** (2-3).
Craven, J. (1983), Input-output analysis and technical change. *Econometrica*, **51** (3), 585-598.
David, P.A. and Van de Klundert, T. (1965), Biased efficiency growth and capital-labor substitution in the U.S., 1899–1960. *American Economic Review*, **55** (3), 357-394.
Denison, E.F. (1967), *Why Growth Rates Differ: Post-war Experience in Nine Western Countries*. Washington, DC: Brookings Institution.
Denison, E.F. (1980), The contribution of capital to economic growth. *American Economic Review*, **70** (2), 220-224.
Denison, E.F. (1983), The interruption of productivity growth in the United States. *Economic Journal*, **93** (380), 56-77.
Dhrymes, P.J. (1963), A comparison of productivity behavior in manufacturing and service industries. *Review of Economics and Statistics*, **45** (1), 64-69.
Diwan, R.K. (1963), An empirical estimate of the constant elasticity of substitution production function, University of Birmingham (mimeo).
Domar, E.D. (1961), On the measurement of technological change. *Economic Journal*, **71** (284), 709-729.
Eisenpress, H. and Greenstadt, J. (1966), The estimation of nonlinear econometric systems. *Econometrica*, **34** (4), 851-861.
Fellner, W.(1969), Specific interpretations of learning by doing. *Journal of Economic Theory*, **1** (2), 119-140.
Ferguson, A.R. (1951), Empirical determination of a multidimensional marginal cost function. *Econometrica*, **18** (2), 217-235; **19** (1), 57-58.
Ferguson, C.E.(1969), *The Neoclassical Theory of Production and Distribution*. New York: Cambridge University Press.
Fontela, E. and Pulido, A. (1986), Input-output analysis and forecasting: New approaches. Sixth International Symposium on Forecasting, Paris, June 15–18 (mimeo).
Forssell, O. (1983), Experiences of studying changes in input-output coefficients in Finland, pp. 219-226 in: A. Smyshlyaev (ed.), *Proceedings of the Fourth IIASA Task Force Meeting on Input-Output Modeling*.
Forssell, O., Mäenpää, I., and Svento, R. (1983), The Finnish long-range model system, pp. 151-172 in: M. Grassini and A. Smyshlyaev (eds.), *Input-Output Modeling. Proceedings of the Third IIASA Task Force Meeting*, Collaborative Paper CP-83-S2. Laxenburg, Austria: International Institute for Applied Systems Analysis.
Førsund, F.R. and Hjalmarsson, L. (1983), Technical progress and structural change in the Swedish cement industry, 1955–1979. *Econometrica*, **51** (5), 1449-1467.
Førsund, F.R., and Jansen, E.S. (1983), Technical progress and structural change in the Norwegian primary aluminium industry. *Scandinavian Journal of Economics*, **85** (2), 113-126.
Freeman, C. (1974), *The Economics of Industrial Innovation*, 2nd Edition. London and Harmondsworth: Penguin.
Frisch, R. (1965), *Theory of Production*. Dordrecht: Reidel, and Chicago: Rand McNally.

Giersch, H. and Wolter, F. (1983), Towards an explanation of the productivity slowdown: An acceleration-deceleration hypothesis. *Economic Journal*, **93** (380), 35-55.

Gort, M. and Wall, R.A. (1986), The evolution of technologies and investment in innovation. *Economic Journal*, **96** (383), 741-757.

Grassini, M. and Smyshlyaev, A. (eds.) (1983), *Input-Output Modeling. Proceedings of the Third IIASA Task Force Meeting*, Collaborative Paper CP-83-S2. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Griliches, Z. (1963), The sources of measured productivity growth: United States agriculture, 1940–1960. *Journal of Political Economy*, **71** (4), 331-346.

Griliches, Z. (1964), Research expenditures, education and the aggregate agricultural production function. *American Economic Review*, **54** (6), 961-974.

Griliches, Z. (1967), Production functions in manufacturing: Some preliminary results, pp. 275-322 in: M. Brown (ed.), *The Theory and Empirical Analysis of Production*. New York: National Bureau of Economic Research.

Griliches, Z. (1979), Issues in assessing the contribution of R&D to productivity growth. *Bell Journal of Economics*, **10** (1), 92-116.

Griliches, Z. (1980), R&D and the productivity slowdown. *American Economic Review*, **70** (2), 343-348.

Griliches, Z. (ed.), (1984), *R&D, Patents, and Productivity. N.B.E.R. Conference Report*. Chicago and London: University of Chicago Press.

Griliches, Z. (1986), Productivity, R&D, and basic reasearch at the firm level in the 1970's. *American Economic Review*, **76** (1), 141-154.

Griliches, Z. and Hausman, J.A. (1985), Errors in variables in panel data. *Econometric Society Summer Workshop on Asymptotic Non-Linear Econometric Methods, Paris*, July 1–2.

Griliches, Z. and Lichtenberg, F.R. (1984), Interindustry technology flows and productivity growth: A re-examination. *Review of Economics and Statistics*, **66** (2), 324-329.

Griliches, Z. and Ringstad, V. (1971), *Economics of Scale and the Form of the Production Function*. Amsterdam: North-Holland.

Hahn, F.H. and Matthews, R.C.O. (1964), The theory of economic growth: A survey. *Economic Journal*, **74** (296), 779-902.

Hall, B.H., Griliches, Z., and Hausman, J.A. (1986), Patents and R and D: Is there a lag? *International Economic Review*, **27** (2), 265-283.

Haltmaier, J. (1984), Measuring technical change. *Economic Journal*, **94** (376), 924-930.

Hausman, J.A., Hall, B.H., and Griliches, Z. (1984), Econometric models for count data with an application to the patents-R&D relationship. *Econometrica*, **52** (4), 909-938.

Heady, E.D. and Dillon, J.L. (1962), *Agricultural Production Functions*. Ames: University of Iowa Press.

Hickman, B.G. and Coen, R.M. (1976), *An Annual Growth Model of the United States Economy*. Contributions to Economic Analysis, 100. Amsterdam: North-Holland.

Hicks, J.R. (1964), *The Theory of Wages*. London: Macmillan.

Hildebrand, G.H. and Liu, T.C. (1965), *Manufacturing Production Functions in the United States, 1957*. Ithaca, NY: New York School of Industrial Labor Relations.

INSEE (1978), DMS: Modèle dynamique multi-sectoriel (Multisectorial dynamic model DMS — English summary). *Collections de l'INSEE*, Série C, (64-65), 55-72.

INSEE (1980), METRIC: Une modèlisation de l'économie Française. Paris: INSEE, 81-98.

Intriligator, M.D. (1965), Embodied technical change and productivity in the United States, 1929–1958. *Review of Economics and Statistics*, **47** (1), 65-70.

Intriligator, M.D. (1978), *Econometric Models, Techniques and Applications*. Amsterdam and Oxford: North-Holland.

Istvan, R. (1974), Interindustry impacts of alternate utility investment strategies, pp.

129-140 in: M.M. Macrakis (ed.), *Energy: Demand, Conservation and Institutional Problems*. Cambridge, MA: MIT Press.

Johansen, L. (1959), Substitution versus fixed production coefficients in the theory of economic growth: A synthesis. *Econometrica*, **27** (2), 157-176.

Jorgenson, D.W. (1966), The embodiment hypothesis. *Journal of Political Economy*, **74** (1), 1-17.

Jorgenson, D.W. and Griliches, Z. (1967), The explanation of productivity change. *Review of Economic Studies*, **34** (3), 249-283.

Kaldor, N. (1957), A model of economic growth. *Economic Journal*, **67** (268), 591-624. Reprinted in Kaldor, N. (1960), *Essays on Economic Stability and Growth*, pp. 13-33. London: Duckworth.

Kaldor, N. and Mirrlees, J.A. (1962), A new model of economic growth. *Review of Economic Studies*, **29** (3), 174-192.

Kamien, M.I. and Schwartz, N.L. (1982), *Market Structure and Innovation*. Cambridge Surveys of Economic Literature Series. Cambridge, New York, and Sydney: Cambridge University Press.

Kendrick, J.W. (1961), *Productivity Trends in the United States*. Princeton, NJ: Princeton University Press.

Kendrick, J.W. (1981), International comparisons of recent productivity trends. In: *Essays in Contemporary Economic Problems*. Washington, DC: American Enterprise Institute for Public Policy Research.

Kendrick, J.W. and Sato, R. (1963), Factor prices, productivity and the theory of distribution. *Economic Journal*, **74** (295), 541-547.

Kennedy, C. (1964), Induced bias in innovation and the theory of distribution. *Economic Journal*, **74** (295), 541-547.

Kennedy, C. and Thirlwall, A.P. (1972), Surveys in applied economics: Technical progress. *Economic Journal*, **82** (325), 11-72.

Kiy, M. (1984), Ein disaggregiertes Prognosesystem für die Bundesrepublik Deutschland. In: M. Beckmann and W. Krelle (eds.), *Lecture Notes in Economics and Mathematical Systems 224*. Berlin, Heidelberg, New York, and Tokyo: Springer-Verlag.

Klein, L.R. (1953), *A Textbook of Econometrics*. Evanston, IL: Row, Peterson.

Kopp, R.J. and Smith, V.K. (1982), An evaluation of alternative indexes of technological change. Unpublished Working Paper, University of North Carolina, Chapel Hill.

Krelle, W. (1964), Ersetzung der Produktionsfunktion durch Preis- und Kapazitätsabhängige Produktionskoeffizienten. *Jahrbücher für Nationalökonomie und Statistik*, 176.

Krelle, W. (1984), The IIASA research program: Economic growth and structural change. *IIASA Conference on the Analysis and Forecasting of the Economics of Structural Change, Laxenburg (Austria)*, May 14–16.

Krelle, W., Beckerhoff, D., Langer, H.G., and Fuss, H. (1969). *Ein Prognosesystem für die Wirtschaftliche Entwicklung der Bundesrepublik Deutschland*. Meisenheim am Glan: A. Hain.

Lee, J.H. (1983), The measurement and sources of technological change biases with an application to postwar Japanese agriculture. *Economica*, **50** (198), 159-173.

Levhari, E., Kleiman, E., and Halevi, N. (1966), The relationship between two measures of total productivity. *Review of Economics and Statistics*, **48** (3), 345-347.

Lichtenberg, F.R. (1984), The relationship between federal contract R&D and company R&D. *American Economic Review*, **74** (2), 73-78.

Lindbeck, A. (1983), The recent slowdown of productivity growth. *Economic Journal*, **93** (380), 13-34.

Lucas, R.E. (1967), Tests of a capital-theoretic model of technological change. *Review of Economic Studies*, **34** (2), 175-189.

McCarthy, M.D. (1965), Embodied and disembodied technological progress in the CES production function. *Review of Economics and Statistics*, **47** (1), 71-75.

Maddala, G.S. (1977), *Econometrics*. New York: McGraw-Hill.

Maddala, G.S. and Kodane, J. (1967), Estimation of returns to scale and the elasticity of substitution. *Econometrica*, **35** (3), 419-423.

Mairesse, J. (1977), Deux essais d'estimation du taux moyen de progrès technique incorporé au capital [Two attempts to estimate the average rate of technical progress embodied in capital — English summary]. *Annales de l'INSEE*, **28**, 39-76.

Mairesse J. (1978), New estimates of embodied and disembodied technical progress. *Annales de l'INSEE*, **30-31**, 682-720.

Mansfield, E. (1968), *The Economics of Technological Change*. London: Longmans, Green.

Massell, B.F. (1960), Capital formation and technological change in United States manufacturing. *Review of Economics and Statistics*, **42** (2), 182-188.

Massell, B.F. (1962), Investment, innovation and growth. *Econometrica*, **30** (2), 239-252.

Mensch, G. (1975), *Das technologische Patt: Innovationen überwinden die Depression*. Frankfurt: Umschau Verlag Breidenstein.

Minasian, J. (1969), Research and development, production functions, and the rates of return. *American Economic Review*, **59** (2), 80-85.

Moore, F.T. (1959), Economics of scale: Some statistical evidence. *Quarterly Journal of Economics*, **73** (2), 232-245.

Morris, D.J. (1983), Comment on the paper by Professor Lindbeck. *Economic Journal*, **93** (380), 78-83.

Nadiri, M.I. (1970), Some approaches to the theory and measurement of total factor productivity. *Journal of Economic Literature*, **8** (4), 1137-1177.

Nadiri, M.I. (1980), Sectoral productivity slowdown. *American Economic Review*, **70** (2), 349-355.

Nelson, R.A. (1984). Regulation, capital vintage, and technical change in the electric utility industry. *Review of Economics and Statistics*, **66** (1), 59-69.

Nelson, R.R. (1964), Aggregate production functions and medium-range growth projections. *American Economic Review*, **54** (5), 576-605.

Nelson, R.R. (1965), The CES production function and economic growth projections. *Review of Economics and Statistics*, **47** (3), 326-328.

Nelson, R.R. (1968), Innovation, pp. 339-344 in: *International Encyclopedia of the Social Sciences*. New York: Macmillan.

Nelson, R.R. (1980), Production sets, technological knowledge, and R&D: Fragile and overworked constructs for analysis of productivity growth? *American Economic Review*, **70** (2), 62-71.

Nelson, R.R. (1981), Research on productivity growth and productivity differences: Dead ends and new departures. *Journal of Economic Literature*, **19** (3), 1029-1064.

Nerlove, M. (1967), Recent empirical studies of the CES and related production function, pp. 55-122 in: M. Brown (ed.), *The Theory and Empirical Analysis of Production*. New York: National Bureau of Economic Research.

Nordhaus, W. (1969), *Invention, Growth and Welfare: A Theoretical Treatment of Technological Change*. Cambridge, MA: MIT Press.

Ohta, M. (1974), A Note on the duality between production and cost functions: Rate of returns to scale and rate of technical progress. *Economic Studies Quarterly*, **25** (1), 63-65.

Onishi, A. (1985), Impacts of high-tech R&D on investment, trade and economic growth, 1985–2000. Paper prepared for the Madrid Meeting on Project LINK, September 2–6.

Ott, D.J., Ott, A.F., and Yoo, J.H. (1975), *Macroeconomic Theory*. New York: McGraw-Hill. (Kogakusha, Ltd. International Student Edition) 264-297.

Ozaki, I. (1976), The effects of technological change on economic growth of Japan, 1955–1970, pp. 93-111 in: K.R. Polenske and J.V. Skola (eds.), *Advances in Input-Output Analysis.* Cambridge, MA: Ballinger.

Pasinetti, L.L. (1985), *Structural Change and Economic Growth: A Theoretical Essay on the Dynamics of the Wealth of Nations.* Cambridge, London, New York, Melbourne, and Sydney: Cambridge University Press (first published in 1981).

Rosenberg, N. (ed.) (1971), *The Economics of Technological Change.* Harmondsworth, UK: Penguin Modern Economics Readings.

Rosenberg, N. (1974), Science, invention and economic growth. *Economic Journal*, **84** (333), 90-108.

Salter, W.E.G. and Reddaway, W.B. (1969), *Productivity and Technical Change*, 2nd edition. London and New York: Camgridge University Press.

Sato, R. (1975), *Production Functions and Aggregation.* Amsterdam and Oxford: North-Holland, and New York: American Elsevier.

Sato, R. and Nono, T. (1982), A theory of endogenous technical progress: Dynamic Böhm—Bawerk effect and optimal R&D policy. *Zeitschrift für Nationalökonomie*, **12** (1), 1-22.

Sato, R. and Ramachandran, R. (1980), Measuring the impact of technical progress on the demand for intermediate goods: A survey. *Journal of Economic Literature*, **18** (3), 1003-1024.

Sato, R. and Suzawa, G.S. (1983), *Research and Productivity: Endogenous Technical Change.* Boston: Auburn House.

Schankerman, M. and Pakes, A. (1984), *Estimates of the Value of Patents Rights in European Countries During the Post-1950 Period.* N.B.E.R. Working Paper 1650, New York, National Bureau of Economic Research.

Scherer, F.M. (1982), Inter-industry technology flows and productivity growth. *Review of Economics and Statistics*, **64** (4), 627-634.

Schmookler, J. (1952), The changing efficiency of the American economy, 1869–1938. *Review of Economics and Statistics*, **34** (3), 214-231.

Schmookler, J. (1966), *Invention and Economic Growth.* Cambridge, MA: Harvard University Press.

Schott, K. (1978), The relations between industrial research and development and factor demands. *Economic Journal*, **88** (349), 85-106.

Shell, K. (1966), Toward a theory of inventive activity and capital accumulation. *American Economic Review*, **56** (2), 62-68.

Smyshlyaev, A. (ed.) (1983), *Proceedings of the Fourth IIASA Task Force Meeting on Input-Output Modeling*, Collaborative Paper CP-83-S5. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Solow, R.M. (1957), Technical change and the aggregate production function. *Review of Economics and Statistics*, **39** (3), 312-320.

Solow, R.M. (1960), Investment and technical progress, pp. 89-104 in: K.Arrow, S.Karlin and P.Suppes (eds.), *Mathematical Methods in the Social Sciences, 1959.*. Palo Alto, CA: Stanford University Press.

Solow, R.M. (1962), Technical progress, capital formation, and economic growth. *American Economic Review*, **52** (3), 76-86.

Solow, R.M. (1967), Some recent developments in the theory of production. In: M. Brown (ed.), *The Theory and Empirical Analysis of Production.* New York: National Bureau of Economic Research.

Stiglitz, J. and Uzawa, H. (1969), *Readings in the Modern Theory of Economic Growth.* Cambridge, MA: MIT Press.

Stoneman, P.L. (1983). Patents and R&D: Searching for a Lag Structure — Comments. *Proceedings of the Conference on Quantitative Studies of Research and Development*

in Industry. ENSAE, Paris.

Taylor, C.T. and Silberston, A. (1973), *The Economic Impact of the Patent System.* Cambridge, UK: Cambridge University Press.

Tinbergen, J. (1942), Zur Theorie der langfristigen Wirtschaftsentwicklung. *Weltwirtschaftliches Archiv*, **55** (1), 511-547.

Tomaszewicz, L. (1983), Variations in input-ouput coefficients: The application of estimation and forecasting techniques for the case of Poland, pp. 207-218 in: A.Smyshlyaev (ed.), *Proceedings of the Fourth IIASA Task Force Meeting on Input-Output Modeling*, Collaborative Paper CP-83-SO5. Laxenburg, Austria: International Institute for Applied System Analysis.

Uno, K. (1986), Regional translog production functions with capital and labor inputs differentiated by educational attainment: The case of Japanese industry, 1968–1977. *Regional Science and Urban Economics*, **16** (3), 449-462.

Uzawa, H. (1960-1961), Neutral inventions and the stability of growth equilibrium. *Review of Economic Studies*, **28** (75, 76, 77), 117-124.

Vaccara, B.N. (1970), Changes over time in input-output coefficients for the United States, pp. 238-260 in: A.P. Carter and A. Brody (eds.), *Applications of Input-Output Analysis.* Amsterdam: North-Holland.

Valavanis-Vail, S. (1955), An econometric model of growth, USA 1869–1953. *American Economic Review*, **45** (2), 208-221.

Vilares, M.J. (1980), Fonctions de production à génération de capital: Théorie et éstimation (Capital vintage production functions: theory and estimation — English Summary). *Annales de l'INSEE*, **38-39**, 17-41.

Wallis, K.F. (1973) *Topics in Applied Econometrics.* Lectures in Economics 5. London: Gray-Mills.

Walters, A.A. (1963a) Production and cost functions: An econometric survey. *Econometrica*, **31** (1-2), 1-66.

Walters, A.A. (1963b) A note on economies of scale. *Review of Economics and Statistics*, **45** (4), 425-427.

Walters, A.A. (1968), Econometric studies of production and cost functions. In: *International Encyclopedia of the Social Sciences.* New York: Macmillan.

Welfe, W. (1985), *Econometric Macromodels of Unbalanced Growth.* Lodz: Prace Instytutu Ekonometrii i Statystyki, Uniwersytetu Lódzkiego.

Westfield, F. (1966), Technical progress and returns to scale. *Review of Economics and Statistics*, **48** (4), 432-441.

Wickens, M.R. (1970), Estimation of the vintage Cobb-Douglas production function for the United States, 1900–1960. *Review of Economics and Statistics*, **52** (2), 187-193.

CHAPTER 24

Local Autoregression Models for Detection of Changes in Causality

Wolfgang Polasek

Summary

Macroeconomic time series often exhibit various nonstationary influences, such as outliers, breaks or jumps in levels. These imply very sensitive estimation not only for univariate and multivariate autoregressive time series models, but also for a Wiener-Granger causality analysis. This chapter investigates the impact of nonstationary behaviour by estimating local stationary AR processes. Two types of local stationarity analysis are proposed: a so-called consecutive bisectrix method, where time series are repeatedly halved as long as reasonable estimation is possible; and a certain span or estimation window moving along the time axis. For both methods, the Geweke (1982) causality measures are derived by comparing univariate and multivariate AR models for the same time spans. An example involving Austrian interest rates for the 1970s demonstrates the two approaches. It is shown that causalities changed considerably in the 1970 decade and that at least three different causality periods can be detected.

24.1 Introduction

24.1.1 Purpose of this chapter

The assumption of constant relations between time series is a crucial one for multivariate time series, but it can hardly be justified for many economic processes. Besides extreme observations, which have a nonrobust influence on most time series estimates, we observe a series of changing regimes where the exact change points are often not known. This

implies that we have to analyze nonstationary time series, but it may be generally difficult to specify a particular class of models which are appropriate.

Local stationary models are a simple way of approximating nonstationary time series without specifying a certain class of models. The estimated stationary blocks can be considered as a first approximation to an unknown structure. Such an analysis is restricted only by the number of observations and the recording interval. Small recording intervals can be found quite frequently in the monetary sector of an economy, and therefore a local analysis seems to be appropriate: Many interest rates can be obtained on a monthly, weekly, or even daily basis. Unfortunately, a smaller recording interval does not always imply that the influence of nonstationarities decreases. Extreme observations can become more visible (e.g., "ultimo" problem in financial transactions), and other types of relationships (seasons) gain weight.

Nevertheless, we explore this changing field of time series relations through two types of modeling procedures. First, we estimate on the basis of AIC [Akaike's information criterion, see Akaike (1973)] local stationary models by the so-called bisectrix method. The time series are successively halved and tested to see whether the parts belong to one or two generating processes. Geweke causality measures can be easily calculated for such periods and indicate changes in the influence patterns. The second method employs the estimation of locally moving models with a fixed time span. By dropping the first and including a new observation, we can find out how the causality pattern changes over time. The associated relative Geweke causality measures are arranged in a so-called causality profile. These methods are demonstrated for five Austrian interest rates during 1970-1981: the Call Money and 3-Month Money Rates, the Deposit and Lending Rates, and the Bond Rate.

Section 24.2 discusses the results for univariate, Section 24.3 for multivariate, local stationary models. In Section 24.4 , we transform these AR estimation results into local (Wiener-Granger) causality measures using the Geweke (1982) approach. In a final section, we summarize our results.

24.1.2 Locally stationary AR models

Given the univariate time series $X_1, \ldots, X_N$, which is possibly nonstationary, we try to find locally stationary blocks by the following procedure. We divide the time interval $(1, N)$ into k blocks, the ith of the length n_i, $n_1 + n_2 + \ldots + n_k = N$. The following locally stationary AR model is fitted to the data

$$x_n = a_1^i x_{n-1} + \ldots + a_{m(i)}^i x_{n-m(i)} + \varepsilon_n^i \ , \quad n_{i-1} < n < n_i \ , \quad i = 1, \ldots, k \tag{24.1}$$

where the ε_n^i are Gaussian white noise with mean zero and variance σ_i^2.

Computing the likelihood of this model, we find that the AIC of our locally stationary model is then given by

$$AIC = \sum_{i=1}^{k} n_i \log \hat{\sigma}_i^2 + 2 \sum_{i=1}^{k} [m(i) + 2] \tag{24.2}$$

If the number of stationary blocks k is unknown, we have to find those k's and n_i's $(i = 1, \ldots, k)$ $m(i)$s that minimize the AIC.

Since such a procedure is very time-consuming, a simpler method is recommended for practical applications [see Kitagawa and Akaike (1978)]: We specify the basic span $n_i = n$ in advance and calculate the first AR model for the first data block, $x_1, \ldots, x_n$. For the second block of data, $x_{n+1}, \ldots, x_{2n}$, we have to choose between the following two models:

1. Compound model for block 1 and block 2: We fit an AR model of order M_0 to block 1 and an AR model of order M_1 to the second block. The AIC of the joined model is then given by the sum (assuming independence)

$$AIC_1 = n_0 \log \sigma_0^2 + n_1 \log \sigma_1^2 + 2(M_0 + M_1 + 4) \tag{24.3}$$

   where σ_0^2 and σ_1^2 are variances for the respective blocks with n_0 and n_1 observations.

2. Pooled data model: By pooling the data of block 1 and block 2, we fit a new AR model to the pooled data $x_1, \ldots, x_{2n}$ by the minimum AIC procedure. The AIC of this model is given by

$$AIC_\star = (n_0 + n_1) \log \sigma_\star^2 + 2(M_\star + 2) \tag{24.4}$$

   where $\sigma_\star^2$ is the variance and $M_\star$ is the order of the pooled AR model.

   If AIC_1 is less than $AIC_\star$ we switch to the new model of block 2, because the pooling of data did not yield a better model. If $AIC_\star \leq AIC_1$ we find that block 1 and block 2 can be considered homogeneous and the pooled model is accepted. This new model forms the basis for the comparison of AR models of block 3. Now the pooled model has an increased basic span; and the $AR_\star$ model takes the role of the AR_0 model in the compound model.

24.2 Univariate Locally Stationary Models

For the following analysis of the fitting of locally stationary AR models to the monthly Austrian interest rates, we choose a basic span of 18 months and the maximum lag length is set to be 10. Naturally, the choice of the span and the lag length determines the number of degrees of freedom. A ratio of 1/2 was considered to be appropriate. The main restrictions are the data length (91 observations) and the usefulness of the results for the multivarate nonstationary analysis.

Table 24.1: Deposite Rate 73.1–80.7

Span	*AIC*	*Variance*	*Coefficients*				*Periods*
11–28	29.3	0.0407	0.91				73.11–75.4
(11–46	32.9	0.0223)					
29–46	15.0	0.0036	0.95				75.5–76.10
(29–64	−11.7	0.0065)					
47–64	−12.0	0.0091	1.02				76.11–78.4
(47–91	67.3	0.3573)					
65–91	56.6	0.0518	1.03	−0.21	0.46	−0.52	78.5–80.7

Table 24.2: Corrected Lending Rate.

Span	*AIC*	*Variance*	*Coefficients*					*Periods*
11–28	−68.8	0.018	1.00					73.11–75.5
(11–46	−137.5	0.020)						
29–46	−138.6	0.006	0.40	0.27	0.55	0.04	0.23	75.5–76.10
			−0.47	0.23	0.04	0.10	−0.48	
(29–46	−130.6	0.019						
(47–64	−131.3	0.026	0.86					76.11–78.4)
47–91	−136.4	0.040	0.98	0.25	−0.33			76.11–80.7
(65–91	−135.4	0.048)						

Table 24.3: Lending Rate.

Span	*AIC*	*Variance*	*Coefficients*					*Periods*
11–28	−4.53	0.622	0.80					73.11–75.4
(11–46	−29.2	0.396)						
29–46	−39.9	0.065	1.00	−0.75	0.69	−0.46	0.38	75.5–76.10
			−0.17					
(29–46	−89.2	0.057)						
(47–64	−97.0	0.026	0.86					76.11–78.4)
47–91	−135.6	0.048						76.11–80.7
(65–91	−136.2	0.041	0.98	0.25	−0.32)			

Table 24.4: 3-Month Money Rate.

Span	*AIC*	*Variance*	*Coefficients*					*Periods*
11–28	−26.4	0.185	0.49					72.11–74.4
(11–46	−37.9	0.312)						
(29–46	−39.5	0.199	0.56	0.43	0.35	−1.03	0.96	74.5–75.11)
			0.09	−0.12				
29–64	−28.9	0.288	0.78	0.19	−0.22	0.15	0.29	74.5–77.4
			0.13	−0.44				
(47–64	−27.5	0.295)						
(47–82	−39.3	0.448)						
65–92	−42.1	0.344	1.37	−0.61				77.5–78.10
(65–116	−43.3	0.387)						
83–116	−43.6	0.363	1.02					78.11–81.9

Table 24.5: Call Money Rate.

Span	*AIC*	*Variance*	*Coefficients*			*Periods*
11–28	−13.0	0.389	0.44			72.11–74.4
(11–46	−34.8	0.340)				
(29–46	−45.0	0.121	1.42	−0.45		74.5–75.11)
(29–64	−46.6	0.220	1.39	−0.85	0.42	74.5–77.4)
(47–64	−45.1	0.310)				
29–82	−59.8	0.258	1.32	−0.68	0.31	74.5–78.10
(65–82	−58.7	0.168)				
(29–116	−65.1	0.407)				
83–116	−70.8	0.643	0.98			78.11–81.9

Table 24.6: Bond Rate.

Span	*AIC*	*Variance*	*Coefficients*					*Periods*
11–28	−92.2	0.0043	1.54	−0.62				75.11–77.4
(11–46	−173.2	0.0069)						
29–46	−174.2	0.0035	0.88	0.41	−0.07	0.03	0.01	77.5–78.10
			−0.42	−0.87	0.10	0.88		
(29–80	−177.3	0.0283)						
47–80	−187.3	0.0401	1.03					78.11–81.9

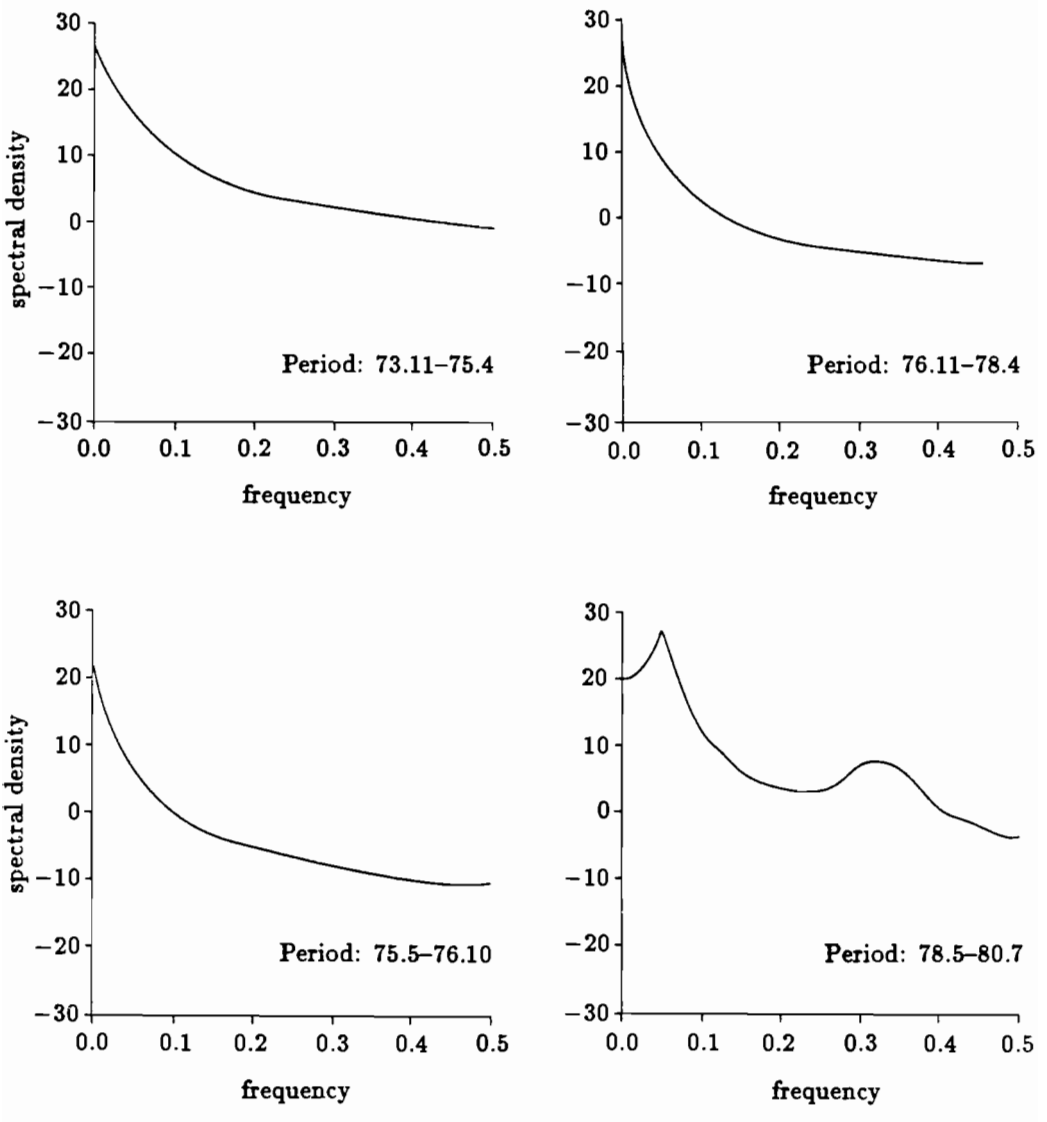

Figure 24.1: Deposit Rate.

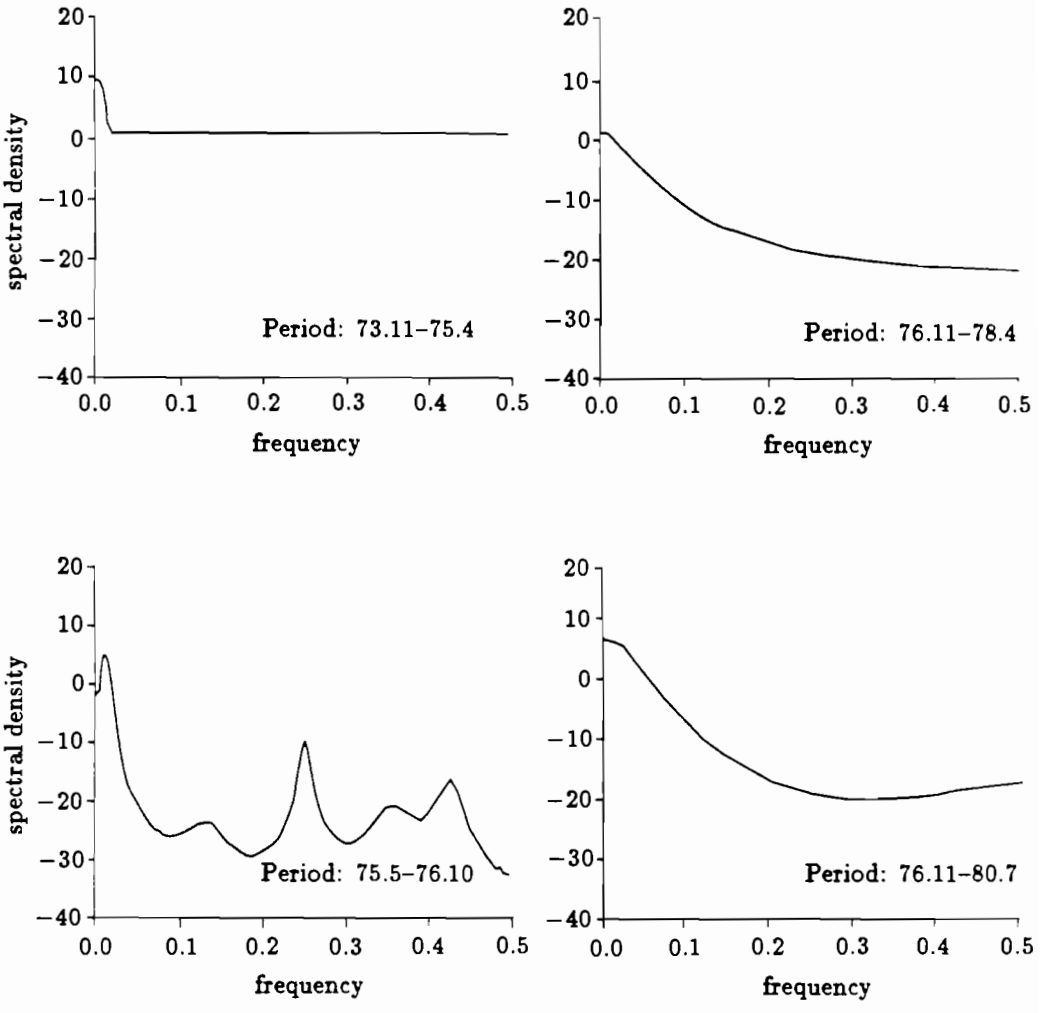

Figure 24.2: Corrected Lending Rate.

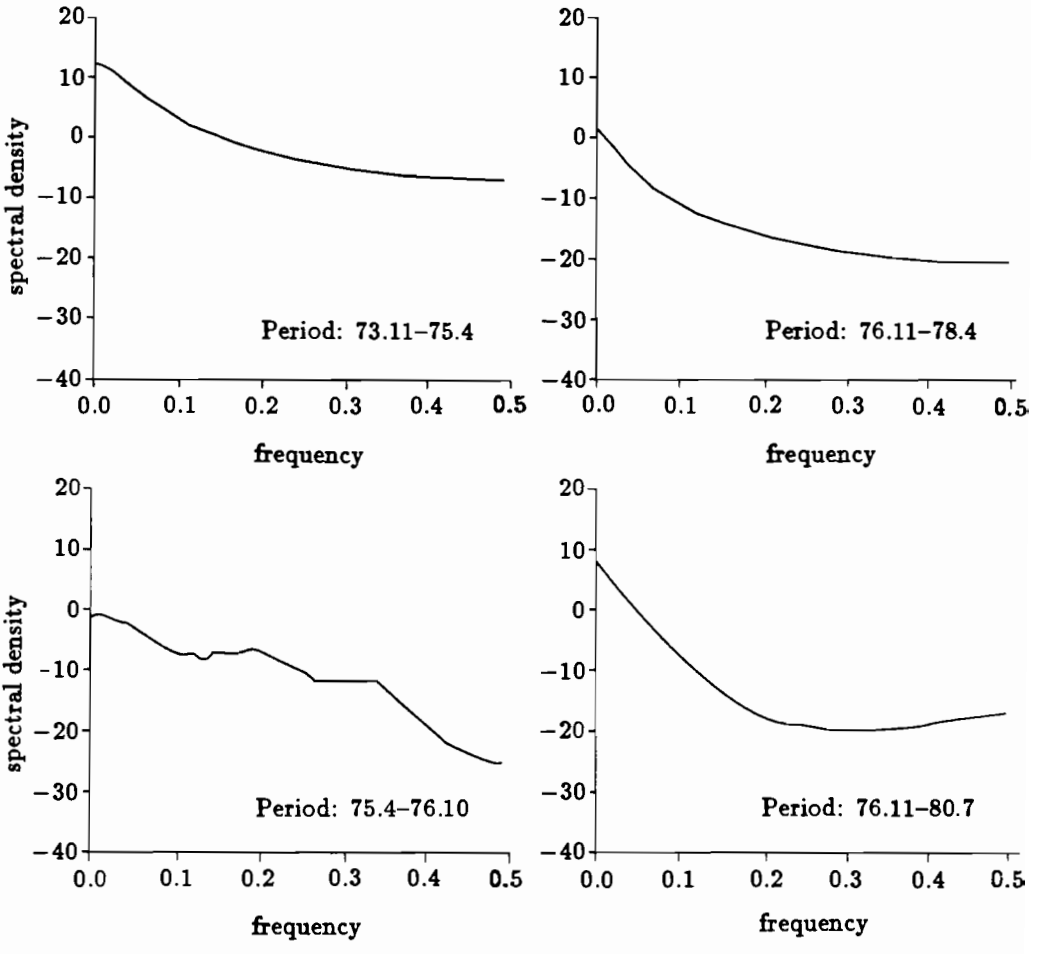

Figure 24.3: Lending Rate.

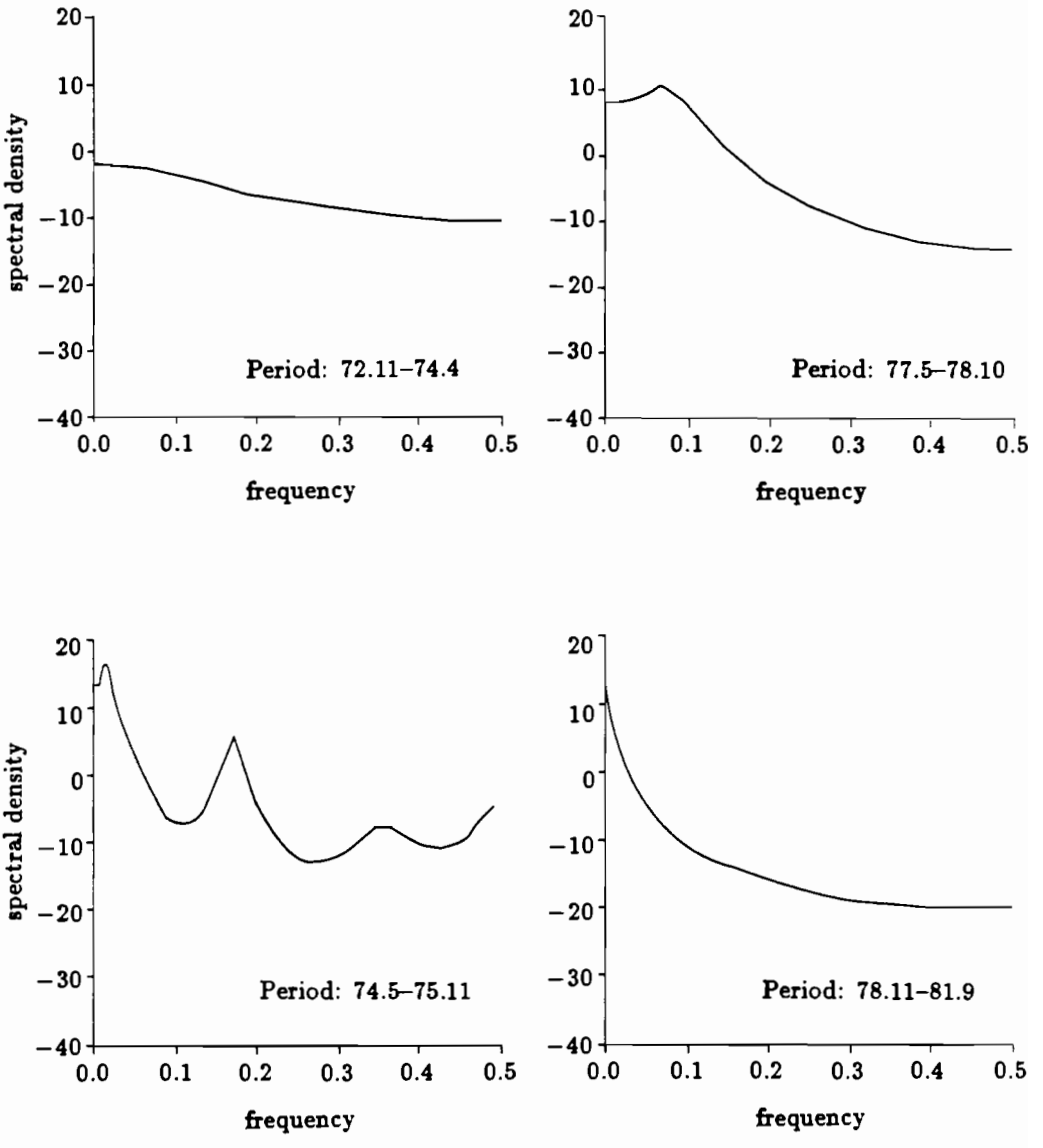

Figure 24.4: 3-Month Money Rate.

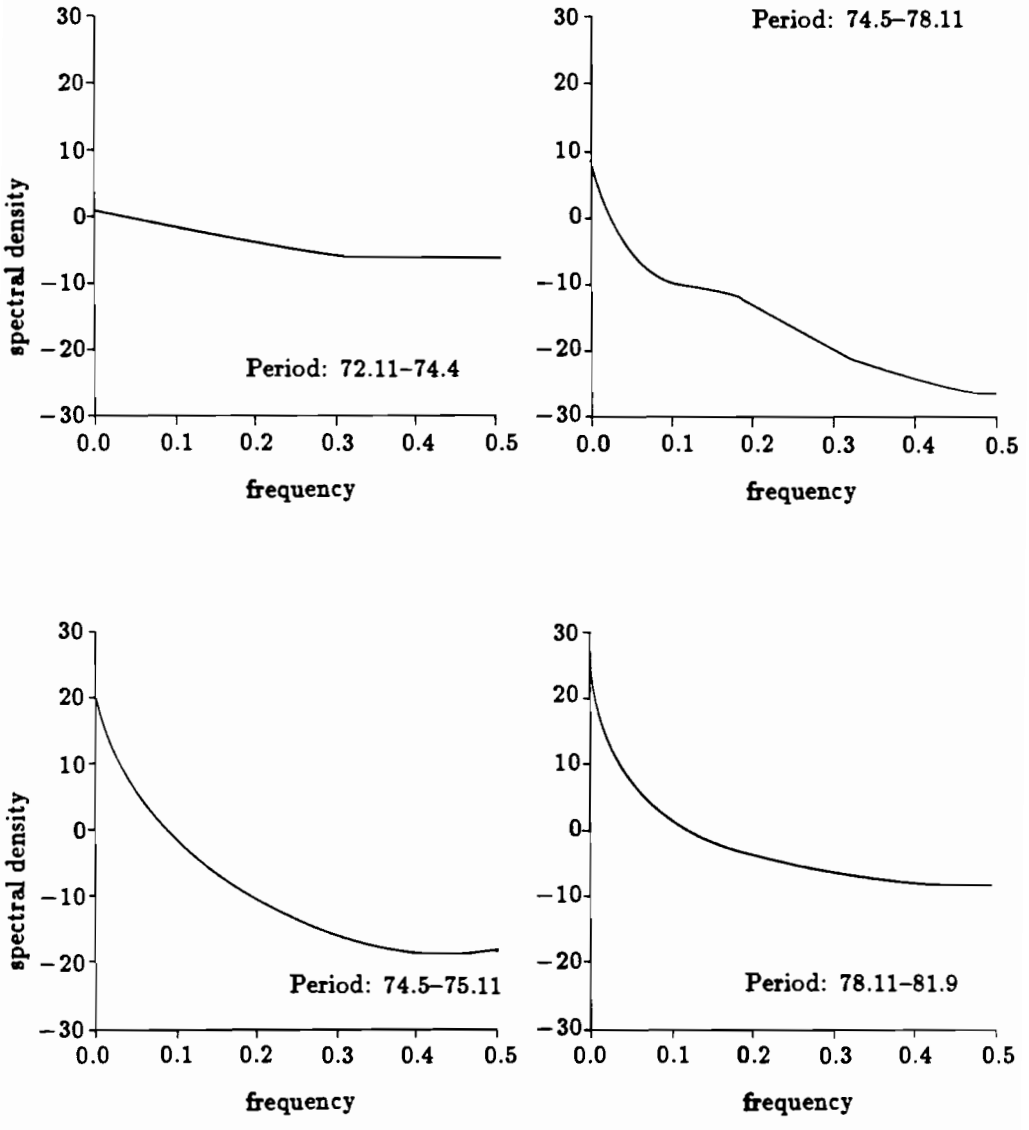

Figure 24.5: Call Money Rate.

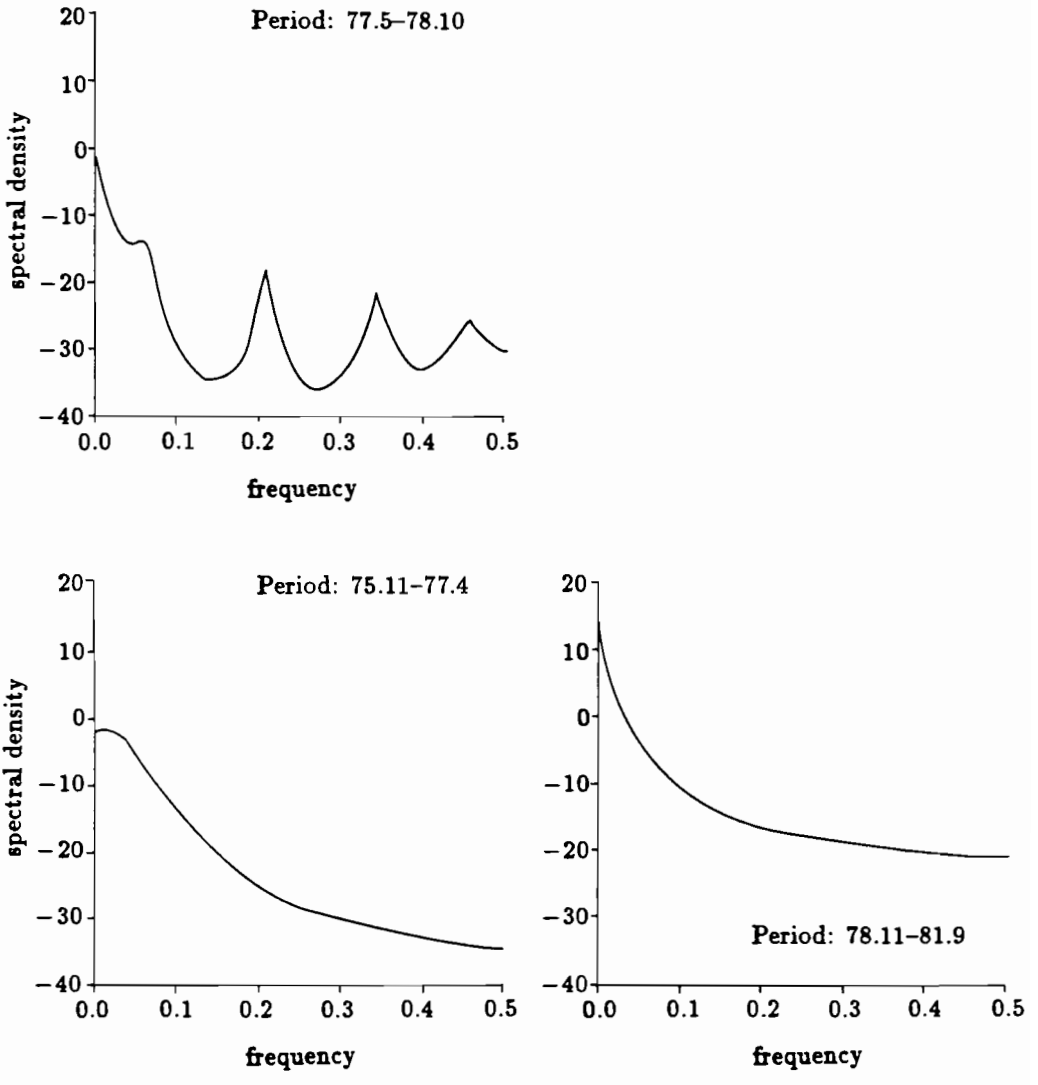

Figure 24.6: Bond Rate.

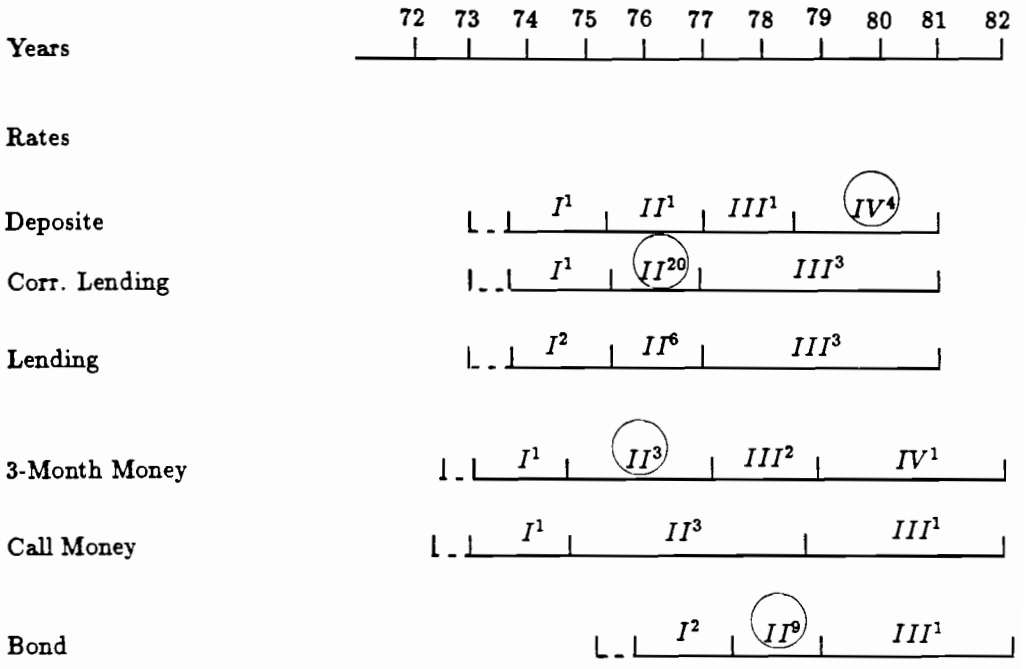

Figure 24.7: Summary: locally stationary models.

Table 24.7: Corrected Lending Rate (span = 12, order = 10).

Span	*AIC*	*Variance*	*Coefficients*					*Periods*
11–22	−44.0	0.013	0.43	1.22	−0.62			73.11–74.10
(11–34	96.8	0.015)						
23–34	−104.7	0.003	1.25	−0.74	0.47			74.11–75.10
(23–46	−92.8	0.008)						
35–46	−107.3	0.006	0.41	0.37	0.68	0.07	−0.04	75.11–76.10
			−0.65					
(35–58	−78.7	0.021)						
47–58	82.7	0.035	0.84					76.11–77.10
(47–70	−83.9	0.026)						
59–70	−90.2	0.002	0.77	0.04	−0.46	0.66	0.16	77.11–78.10
			0.64	0.42	−0.37	−0.61		
(59–91	−96.7	0.042)						
71–91	−107.0	0.067	0.96					78.11–80.7

Table 24.8: Corrected Lending Rate (span = 24, order = 10).

Span	*AIC*	*Variance*	*Coefficients*					*Periods*
11–34	−96.8	0.015	0.98					73.11–75.10
(11–58	−174.3	0.024)						
35–58	−175.7	0.021	0.62	0.29	0.41	−0.05	0.07	75.11–77.10
			−0.39					
35–91	−181.3	0.034	0.84	0.29	0.04	−0.03	−0.29	77.11–80.7
(59–91	−175.6	0.042)						

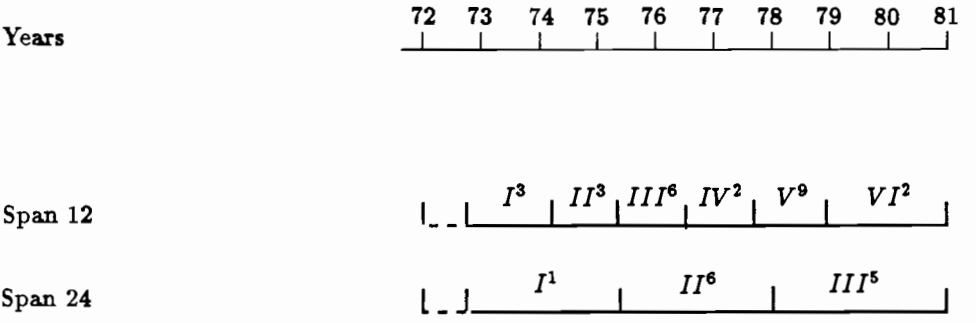

Figure 24.8: Summary: Corrected Lending Rate.

The results for every time series are displayed in *Tables 24.2* to *24.6* and *Figures 24.1* to *24.6*. A summary of the univariate analysis is given in *Figure 24.7*. There are at least three nonstationary periods for every time series.

For the Deposit and the Lending Rates, we find more nonstationary influences in the first half of the series than in the second half. The outliers in the Lending Rate affect only the first half of the series.

The 3-Month Money Rate and the Call Money Rate are also closely related in their nonstationary behavior. While both series show at both ends nonstationary influences, the behavior in the middle of the series (75–79) seems quite stable.

The Bond Rate is the shortest of all six series, but even for this series we find stationary intervals. Except for the Call Money Rate, almost all series show a pronounced cyclic behavior in at least one stationary period. It is interesting to note that these periods are different for every series. The only common feature is that the cyclic period was never the first period for all of the analyzed time series.

24.2.1 Deposit Rate

Four locally stationary models are found for the Deposit Rate. In the first three periods we find an AR(1) process close to a random walk but with different residual variance. The fourth period from 78.5–80.7 shows the highest variance together with a fourth order AR process. The spectrum in *Figure 24.1* shows clearly two peaks: one around 20–24 months, and one at the 3-month lag. In *Table 24.1*, the rows in parentheses show the competing models that have not been choosen.

24.2.2 Corrected Lending Rate (without outliers)

The Corrected Lending Rate exhibits 3 stationary periods. While the first 3 years can be decomposed into two very different periods, the last 3.5 years have more similar spectra (see *Figure 24.2*). The first period (i.e., the first 1.5 years) from 73.11–75.4 shows an almost flat spectrum, while the second period is characterized by a cyclic component. (The *AIC* attained the minimum at the maximum lag length 10). For this period, we find two peaks: one low-frequency peak at 6 months, and a pronounced short-term peak at 4 months. The last 1.5 years show no peak in the spectrum and the estimated AR(3) process exhibits the highest variance. (See *Table 24.2.*)

24.2.3 Lending Rate

Knowing the position of the outliers, we expect a behavior different from the original Lending Rate only for the first 2 periods. Also, we see from *Table 24.3* that the variances are approximately 10 times as high as for the Corrected Lending Rate. Especially for the second period, 75.5–76.10, we see that the pronounced peaks completely vanished. Only 2 bumps around 3 and 6 months can be seen.

24.2.4 3-Month Money Rate (72.1-81.9)

The nonstationarity procedure yields 4 different periods. The first period, from 72.11 to 74.4, has a flat spectrum and a small-valued AR(1) process. The second period extends for 3 years and exhibits three peaks in the spectrum: one long-term at 24 months, a half-year cycle (6 months), and a short-term peak at 3 months. The last two could be harmonics of the 2-year cycle. The corresponding AR process has an estimated lag of 7. The last two periods are AR(1) processes. The spectrum of the period from 77.5 to 78.10 shows a small peak around 18 months.

24.2.5 Call Money Rate

The Call Money Rate can be divided into three periods. The first period, from 72.11 to 74.4, shows a flat spectrum similar to the first period of the 3-Month Money Rate. The second period is 4.5 years long, extends from 74.5 to 78.10, and can be estimated by

an AR(3) process; but the spectrum shows hardly any cyclic behavior. The last period, 78.11–81.9, is again dominated by an AR(1) process.

24.2.6 Bond Rate

The Bond Rate can be also divided into three periods. The first one, from 75.11 to 77.4, is estimated by an AR(2) process without cyclical components in the spectrum. The second one, from 77.5 to 78.10, shows a rich cyclical behavior: 2-year cycle and pronounced peaks at 5 months and 3 months can be seen in *Figure 24.6*. The process in the last period, from 78.11 to 81.9, is close to a random walk.

24.2.7 Stability of the results

In order to check the stability of the nonstationary analysis of this section we have tried different block-lengths. *Tables 24.7* and *24.8* show the results for the Corrected Lending Rate for spans 12 and 24, respectively.

The detailed analysis for span 12 shows that the cyclical influence of period II in *Figure 24.7* is caused by a short cyclic period for 12 months in the years 76/77. The yearly analysis also shows that there has been another year with pronounced cyclical movement: 78/79. This explains why there are two cyclical periods for the 24-month span analysis. In the period from 78 to 80.6, the influence of the "cyclic year" 78 is large enough to be preserved for the whole period. In *Table 24.8* this is not the case: the nonstationary influences at the end of the series are dominating.

We see that these results are in concordance with the analysis of the 18-month span. But additionally we see that the Corrected Lending Rate has nonstationary influences almost every year. With a view to the multivariate analysis in the next section, the choice of an 18-month span seems to be an acceptable compromise.

24.3 Multivariate Locally Stationary Models

Using the program TIMSAC-78 [described in Akaike *et al.* (1979) and Kitagawa and Akaike (1981)], we can extend the method of nonstationary analysis to bivariate and trivariate models. Continuing the analysis of Polasek (1983), we estimate locally stationary models only for the following three multivariate processes: (1) Deposit Rate and Corrected Lending Rate, (2) 3-Month Money Rate and Call Money Rate, and (3) 3-Month Money Rate/Call Money Rate and Bond Rate.

Because the locally stationary models are estimated by a different method involving least squares and the Householder transformation [see Kitagawa and Akaike (1978)], we have to compare the locally stationary models with the same estimation technique for the whole series. The difference between the two estimation methods is the following: while the original method uses the multivariate autoregressive processes of increasing

order, including all parameters of the AR matrices, the Householder method starts with the univariate estimated series and selects only those components of the AR coefficient matrices that reduce the AIC value significantly. Consequently, this estimation procedures produces longer lag lenghts and "structural" zeroes in the AR coefficient matrices.

Table 24.9: Deposit Rate/Corrected Lending Rate: Multivariate AR model (all observations: Householder OLS).

AIC	*Variance matrix*	*Coefficient matrices*
−593.2	$\begin{pmatrix} 0.0303 & \\ 0.0148 & 0.0277 \end{pmatrix}$	D_t $\begin{pmatrix} 0.91 & 0.31 \\ 0.27 & 0.75 \end{pmatrix}$ $\begin{pmatrix} 0.0 & -0.25 \\ -0.10 & 0.15 \end{pmatrix}$
	$\lvert\Sigma\rvert = 62.027$	L_t^c $\begin{pmatrix} 0.0 & 0.0 \\ -0.10 & 0.0 \end{pmatrix}$ $\begin{pmatrix} 0.0 & 0.0 \\ -0.16 & 0.0 \end{pmatrix}$

Table 24.10: Call Money Rate/3-Month Money Rate: Multivariate AR model (all observations: Householder OLS).

AIC	*Variance matrix*	*Coefficient matrices*
−250.5	$\begin{pmatrix} 0.3723 & \\ 0.2215 & 0.3357 \end{pmatrix}$	R_t $\begin{pmatrix} 0.78 & 0.42 \\ 0.24 & 0.84 \end{pmatrix}$ $\begin{pmatrix} -0.09 & -0.36 \\ -0.06 & -0.22 \end{pmatrix}$
	$\lvert\Sigma\rvert = 0.07592$	$3R_t^c$ $\begin{pmatrix} 0.02 & 0.17 \\ 0.01 & 0.10 \end{pmatrix}$ $\begin{pmatrix} 0.34 & -0.22 \\ 0.20 & -0.13 \end{pmatrix}$ $\begin{pmatrix} 0.0 & -0.10 \\ 0.0 & -0.06 \end{pmatrix}$
		$\begin{pmatrix} 0.0 & 0.25 \\ 0.0 & 0.15 \end{pmatrix}$ $\begin{pmatrix} 0.0 & -0.32 \\ 0.0 & -0.19 \end{pmatrix}$

Table 24.11: 3-Month Money/Call Money/Bond Rate ($p_{\max} = 4$): Multivariate AR model (all observations: Householder OLS).

AIC	*Variance matrix*	*Coefficient matrices*
−457.6	$\begin{pmatrix} 0.3507 & & \\ 0.2571 & 0.4540 & \\ 0.0198 & 0.0350 & 0.0207 \end{pmatrix}$	$\begin{pmatrix} 0.59 & 0.39 & 0.79 \\ 0.21 & 0.86 & 0.42 \\ 0.02 & 0.01 & 1.01 \end{pmatrix}$ $\begin{pmatrix} 0.0 & -0.29 & 0.98 \\ 0.0 & -0.21 & 0.72 \\ 0 & -0.02 & 0.06 \end{pmatrix}$
		$\begin{pmatrix} 0.0 & 0.18 & -1.78 \\ 0.0 & 0.13 & -1.30 \\ 0.0 & 0.01 & -0.10 \end{pmatrix}$

The results are given in *Tables 24.9–24.11* and are generally different from the estimated models in the Polasek (1983) study. Now we can see more seasonal effects. Since stationary time series estimation methods are very sensitive to the estimation method and nonstationary effects, the comparison of these estimation methods is very difficult. Both

methods might be regarded as an approximation to the observed nonstationary process.

For the bivariate model of the Deposit Rate and the Lending Rate, the analysis shows at least three stationary periods. The first and the last sections of both time series are dominated by a strong autoregressive component at lag 1 and special seasonal interaction effects at lag 3 and 4. The model of the middle period from 75.6 to 77.5 also has autoregressive components at lag 1, but shows more interactions at lags 2 and 3.

The results of the locally bivariate Call Money/3-Month Money Rate analysis depend also very much on the specification of the key parameters — block-length and maximum order. The smaller the basic span, the more the time series is divided into stationary blocks. Roughly speaking, we find three different blocks where the model of the middle period, 75–78, of the time series shows the highest seasonal activity.

The trivariate local analysis is based on the smallest amount of available observations, since the Bond Rate was recorded only after 1978. Therefore, the results we found are more sensitive to changes in the key parameters than for the other models. In general, we also see two distinct periods. While the first period is characterized by an almost complete absence of any influence by the 3-Month Money Rate, the second period shows a strong unidirectional influence of the Bond Rate on the other two series.

Table 24.12: Bivariate models: Deposit Rate/Lending Rate (span 24, order 5).

Period	*Span*	*AIC*	*Variance matrix*	*Coefficient matrices*
73.6 75.5	6–29	−60.2	$\begin{pmatrix} 0.0277 & \\ 0.000 & 0.014 \end{pmatrix}$	$D_t \begin{pmatrix} 0.94 & 0.11 \\ 0 & 1.10 \end{pmatrix}$ L_t^c
			$\lvert\Sigma\rvert = 3.878 \times 10^{-4}$	$\begin{pmatrix} 0 & 0 \\ 0 & -0.30 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0.28 & 0.45 \end{pmatrix}$ $\begin{pmatrix} 0 & 0 \\ -0.65 & 0.27 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0.33 & 0 \end{pmatrix}$
	(6–53	−157.3)		
75.6 77.5	30–53	−167.5	$\begin{pmatrix} 0.0175 & \\ 0.000 & 0.014 \end{pmatrix}$	$D_t \begin{pmatrix} 0.44 & 0 \\ -0.09 & 0.68 \end{pmatrix}$ L_t^c
			$\lvert\Sigma\rvert = 2.45 \times 10^{-4}$	$\begin{pmatrix} 0.11 & 0 \\ -0.02 & 0.30 \end{pmatrix}$ $\begin{pmatrix} 0.33 & 0.07 \\ -0.07 & -0.01 \end{pmatrix}$
	(30–91	−156.4)		
77.6 80.7	54–91	−177.1	$\begin{pmatrix} 0.0307 & \\ 0.020 & 0.033 \end{pmatrix}$	$D_t \begin{pmatrix} 0.81 & 0.39 \\ 0.17 & 0.84 \end{pmatrix}$ L_t^c
			$\lvert\Sigma\rvert = 6.131 \times 10^{-4}$	$\begin{pmatrix} 0 & -0.29 \\ 0.05 & 0.11 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ -0.38 & 0 \end{pmatrix}$

Table 24.13: Bivariate models: Deposit Rate/Lending Rate (span 30, order 6).

Period	*Span*	*AIC*	*Variance matrix*	*Coefficient matrices*
73.7 75.12	7–36	−85.3	$\begin{pmatrix} 0.0230 & \\ 0.000 & 0.012 \end{pmatrix}$ $\lvert\Sigma\rvert = 2.76 \times 10^{-4}$	$D_t \begin{pmatrix} 0.93 & 0.10 \\ 0.23 & 1.11 \end{pmatrix}$ $L_t^c \begin{pmatrix} 0 & 0 \\ -0.32 & -0.49 \end{pmatrix}$ $\begin{pmatrix} 0 & 0 \\ 0.36 & 0.59 \end{pmatrix}$ $\begin{pmatrix} 0 & 0 \\ -0.65 & -0.26 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0.34 & 0 \end{pmatrix}$
	(7–66	−197.7)		
76.1 78.6	37–66	−208.8	$\begin{pmatrix} 0.0144 & \\ 0.0000 & 0.0179 \end{pmatrix}$ $\lvert\Sigma\rvert = 6.914 \times 10^{-4}$	$D_t \begin{pmatrix} 0.70 & 0.24 \\ 0.76 & 0.45 \end{pmatrix}$ $L_t^c \begin{pmatrix} 0.30 & -0.18 \\ -0.07 & 0.26 \end{pmatrix}$ $\begin{pmatrix} 0 & 0 \\ -0.11 & 0.33 \end{pmatrix}$ $\begin{pmatrix} 0 & 0 \\ -0.74 & -0.30 \end{pmatrix}$
	(37–91	130.1)		
78.7 80.7	67–91	159.4	$\begin{pmatrix} 0.0339 & 0.023 \\ 0.023 & 0.036 \end{pmatrix}$ $\lvert\Sigma\rvert = 6.914 \times 10^{-4}$	$D_t \begin{pmatrix} 0.68 & 0.20 \\ -0.04 & 0.81 \end{pmatrix}$ $L_t^c \begin{pmatrix} 0 & 0 \\ 0.04 & 0.29 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ -0.36 & 0 \end{pmatrix}$

Table 24.14: Bivariate models: Deposit Rate/Lending Rate (span 36, order 8).

Period	*Span*	*AIC*	*Variance matrix*	*Coefficient matrices*
73.9 76.8	9–44	−113.8	$\begin{pmatrix} 0.97 & \\ 0.000 & 0.016 \end{pmatrix}$ $\lvert\Sigma\rvert = 0.01552$	$D_t \begin{pmatrix} 0.92 & 0.10 \\ 0.25 & 0.94 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ -0.30 & 0 \end{pmatrix}$ L_t^c
	(9–91	−211.0)		
76.9 80.7	45–91	−215.3	$\begin{pmatrix} 0.0303 & 0.015 \\ 0.015 & 0.028 \end{pmatrix}$ $\lvert\Sigma\rvert = 6.234 \times 10^{-4}$	$D_t \begin{pmatrix} 0.86 & 0.36 \\ 0.21 & 0.81 \end{pmatrix}$ $L_t^c \begin{pmatrix} 0 & -0.30 \\ 0.01 & 0.13 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ -0.36 & 0 \end{pmatrix}$

Table 24.15: Locally bivariate models: Call Money Rate/3-Month Money Rate (span 24, order 5).

period	*Span*	*AIC*	*Variance matrix*	*Coefficient matrices*
72.6 74.5	6–29	−67.3	$\begin{pmatrix} 0.403 & \\ 0.208 & 0.206 \end{pmatrix}$ $\lvert\Sigma\rvert = 0.039754$	$R_t \begin{pmatrix} 0 & 1.26 \\ -0.31 & 1.34 \end{pmatrix}$ $3R_t \begin{pmatrix} 0 & -0.43 \\ 0 & -0.22 \end{pmatrix}$
	(6–53	−124.4)		
74.6 76.5	30–53	−130.1	$\begin{pmatrix} 0.363 & \\ 0.227 & 0.297 \end{pmatrix}$ $\lvert\Sigma\rvert = 0.056282$	$R_t \begin{pmatrix} 1.05 & 0 \\ 0.51 & 0.45 \end{pmatrix}$ $3R_t \begin{pmatrix} -0.07 & -0.39 \\ -0.04 & 0.19 \end{pmatrix}$ $\begin{pmatrix} 0.07 & -0.45 \\ 0.04 & -0.22 \end{pmatrix} \begin{pmatrix} 0.54 & 0.27 \\ 0.26 & 0.13 \end{pmatrix}$
	(30–77	−121.5)		
76.6 78.5	54–77	−132.5	$\begin{pmatrix} 0.278 & \\ 0.183 & 0.344 \end{pmatrix}$ $\lvert\Sigma\rvert = 0.062143$	$R_t \begin{pmatrix} 1.54 & -0.39 \\ 0.60 & 0.67 \end{pmatrix}$ $3R_t \begin{pmatrix} -0.98 & 0.18 \\ -0.53 & -0.06 \end{pmatrix}$ $\begin{pmatrix} 0.22 & 0.32 \\ -0.12 & -0.14 \end{pmatrix}$ $\begin{pmatrix} 0.43 & -1.10 \\ 0.72 & -0.75 \end{pmatrix} \begin{pmatrix} 0.51 & 0 \\ 0.35 & 0 \end{pmatrix}$
	(54–116	−133.3)		
78.6 81.6	78–116	−144.5	$\begin{pmatrix} 0.414 & \\ 0.203 & 0.305 \end{pmatrix}$ $\lvert\Sigma\rvert = 0.085061$	$R_t \begin{pmatrix} 0.58 & 0.49 \\ 0.10 & 0.90 \end{pmatrix}$ $3R_t$

Table 24.16: Locally bivariate models: Call Money Rate/3-Month Money Rate (span 30, order 6).

Period	*Span*	*AIC*	*Variance matrix*	*Coefficient matrices*
72.7 74.12	7–36	−86.7	$\begin{pmatrix} 0.319 & \\ 0.146 & 0.184 \end{pmatrix}$	$R_t \begin{pmatrix} 0.51 & 0.68 \\ -0.05 & 1.09 \end{pmatrix}$ $3R_t$
			$\lvert\Sigma\rvert = 0.03738$	$\begin{pmatrix} 0 & -0.46 \\ 0 & -0.21 \end{pmatrix}$
	(7–66	−159.2)		
75.1 77.6	37–66	−180.9	$\begin{pmatrix} 0.375 & \\ 0.189 & 0.265 \end{pmatrix}$	$R_t \begin{pmatrix} 0.98 & 0 \\ 0.60 & 0.28 \end{pmatrix}$ $3R_t$
			$\lvert\Sigma\rvert = 0.063654$	$\begin{pmatrix} -0.07 & -0.33 \\ -0.04 & -0.20 \end{pmatrix}$ $\begin{pmatrix} 0.43 & -0.49 \\ 0.27 & -0.30 \end{pmatrix} \begin{pmatrix} 0.40 & 0 \\ 0.25 & 0 \end{pmatrix}$ $\begin{pmatrix} 0.23 & 0 \\ 0.14 & 0 \end{pmatrix} \begin{pmatrix} -0.33 & 0 \\ -0.20 & 0 \end{pmatrix}$
	(37–116	−171.0)		
77.7 81.6	67–116	−187.0	$\begin{pmatrix} 0.393 & \\ 0.243 & 0.377 \end{pmatrix}$	$R_t \begin{pmatrix} 0.66 & 0.38 \\ 0.31 & 0.92 \end{pmatrix}$ $3R_t$
			$\lvert\Sigma\rvert = 0.089112$	$\begin{pmatrix} 0 & 0 \\ 0 & -0.29 \end{pmatrix}$

24.3.1 Results for the Deposit Rate and Corrected Lending Rate

In *Table 24.9* we find the estimation results of the VAR model for the Deposit Rate and the Corrected Lending Rate using the least squares method with the Householder transformation [see Kitagawa and Akaike (1978)]. The results differ considerably from the Polasek (1983) analysis. Now we obtain a fourth-order bivariate AR process with 3 "structural" zeros in the last two lag matrices. At lag 1 we find positive feedback, but at lag 2 negative feedback between the series. The Deposit Rate is influencing the Lending Rate negatively at lags 3 and 4.

Due to the nonstationary influences in both time series, the estimation results are not very stable. To allow comparisons with the bivariate nonstationary analysis, we have chosen a maximum lag length of order 8. There are some seasonal effects at lag 18; but because of the shortness of the time series, the estimation procedure failed to produce reliable results for larger lag lengths.

The results of the locally stationary analysis are given in *Tables 24.12–24.14*. We have chosen 3 types of spans — 24, 30, and 36 months — and the corresponding maximum

orders are 5, 6, and 8.

For the shortest span of 24 months, we obtain a fifth-order process for the period 73.6–75.5. This period is also characterized by a long-term influence of the Deposit Rate on the Lending Rate (until lag 5), but only a short-term influence for the reverse relationship. The second period, from 75.6 to 77.5, exhibits a shorter AR(3) process, and we observe small interactions between the time series at lags 2 and 3.

For the third period, from 77.6 to 80.7, we observe a high short-term interaction at the lag 1 coefficient matrix. There is a negative influence of the Deposit Rate at lag 3, while the Lending Rate has a negative effect at lag 2.

Table 24.17: Locally bivariate models: Call Money Rate/3-Month Money Rate (span 36, order 8).

Period	*Span*	*AIC*	*Variance matrix*	*Coefficient matrices*
72.9 75.8	9–44	−117.0	$\begin{pmatrix} 0.254 & \\ 0.123 & 0.168 \end{pmatrix}$	R_t $\begin{pmatrix} 0.69 & 0.55 \end{pmatrix}$ $3R_t$ $\begin{pmatrix} 0.08 & 1.00 \end{pmatrix}$
			$\lvert\Sigma\rvert = 0.027543$	$\begin{pmatrix} 0 & -0.56 \\ 0 & -0.27 \end{pmatrix}$
	(9–80	−181.9)		
75.8 78.7	45–80	−194.4	$\begin{pmatrix} 0.366 & \\ 0.236 & 0.337 \end{pmatrix}$	R_t $\begin{pmatrix} 1.17 & 0 \end{pmatrix}$ $3R_t$ $\begin{pmatrix} 0.62 & 0.45 \end{pmatrix}$
			$\lvert\Sigma\rvert = 0.067646$	$\begin{pmatrix} -0.57 & 0 \\ -0.43 & 0 \end{pmatrix}$ $\begin{pmatrix} 0.30 & 0 \\ 0.23 & 0 \end{pmatrix}$ $\begin{pmatrix} 0.04 & 0 \\ 0.03 & 0 \end{pmatrix}$ $\begin{pmatrix} 0.23 & 0 \\ 0.18 & 0 \end{pmatrix}$ $\begin{pmatrix} -0.40 & 0 \\ -0.30 & 0 \end{pmatrix}$ $\begin{pmatrix} 0.54 & -0.52 \\ 0.40 & -0.39 \end{pmatrix}$
75.8 81.6	45–116	−144.4	$\begin{pmatrix} 0.375 & \\ 0.230 & 0.378 \end{pmatrix}$	R_t $\begin{pmatrix} 0.76 & 0.38 \end{pmatrix}$ $3R_t$ $\begin{pmatrix} 0.32 & 0.82 \end{pmatrix}$
			$\lvert\Sigma\rvert = 0.08885$	$\begin{pmatrix} -0.06 & -0.35 \\ -0.04 & -0.35 \end{pmatrix}$ $\begin{pmatrix} 0.02 & 0.14 \\ 0.01 & 0.09 \end{pmatrix}$ $\begin{pmatrix} 0.54 & -0.27 \\ 0.33 & -0.17 \end{pmatrix}$ $\begin{pmatrix} 0 & -0.21 \\ 0 & -0.13 \end{pmatrix}$ $\begin{pmatrix} 0 & 0.32 \\ 0 & 0.20 \end{pmatrix}$ $\begin{pmatrix} 0 & -0.41 \\ 0 & -0.25 \end{pmatrix}$
	(81–116	−141.8)		

For span 30, we get a similar result for the first and the last span of the time series; but for the middle period, from 76.1–78.6, we can estimate an AR(4) process with some seasonal interactions. For span 36, we find two periods, and both AR processes are very

similar to the overall process of *Table 24.9*.

24.3.2 Results for the Call Money and the 3-Month Money Rates

For the whole observation period, we can estimate a bivariate AR(7) process for the pair of time series Call Money/3-Month Money Rate. Note that the lag matrices between order 5 and 7 have a particular structure. While the 3-Month Money Rate is influenced by the Call Money Rate and its own seasonal activities, there is no influence in the other direction. Up to order 4, we see a mutual interdependence between both time series.

Table 24.18: Locally bivariate models: Call Money Rate/3-Month Money Rate (span 36, order 4).

Period	*Span*	*AIC*	*Variance matrix*	*Coefficient matrices*
72.5 75.4	5–40	−111.9	$\begin{pmatrix} 0.306 & \\ 0.144 & 0.173 \end{pmatrix}$	R_t $\begin{pmatrix} 0.61 & 0.63 \\ 0.04 & 1.04 \end{pmatrix}$ $3R_t$
			$\lvert\Sigma\rvert = 0.032202$	$\begin{pmatrix} 0 & -0.45 \\ 0 & -0.21 \end{pmatrix}$
	(5–76	−183.9)		
75.5 78.4	41–76	−196.5	$\begin{pmatrix} 0.359 & \\ 0.219 & 0.317 \end{pmatrix}$	R_t $\begin{pmatrix} 1.01 & 0 \\ 0.63 & 0.29 \end{pmatrix}$ $3R_t$
			$\lvert\Sigma\rvert = 0.066714$	$\begin{pmatrix} -0.54 & 0 \\ -0.34 & 0 \end{pmatrix}$ $\begin{pmatrix} 0.48 & -0.38 \\ 0.30 & 0.24 \end{pmatrix}$ $\begin{pmatrix} 0.57 & -0.28 \\ 0.35 & -0.18 \end{pmatrix}$
	(41–116	−152.3)		
78.5 81.6	77–116	−161.8	$\begin{pmatrix} 0.487 & \\ 0.305 & 0.427 \end{pmatrix}$	R_t $\begin{pmatrix} 0.64 & 0.67 \\ 0.11 & 1.03 \end{pmatrix}$ $3R_t$
			$\lvert\Sigma\rvert = 0.114924$	$\begin{pmatrix} 0 & -0.28 \\ 0 & -0.15 \end{pmatrix}$

Looking at the locally stationary estimated models in *Tables 24.15–24.18*, we see that the seasonal activities are higher for the middle period (75–78) of the time series than for both ends. Also the differences for the *AIC* between the blocks are smaller for this pair of time series than for the Deposit/Lending Rate pair. Furthermore, the results depend very much on the choice of the maximum lag length. In *Table 24.17*, we have calculated local stationary models with a 3-year span and order 8, while in *Table 24.18* only for order 4. We see that except for the ends of the time series, where an AR(2) process dominates, the middle period is highly seasonal. Only for order 8 do we get a two-period result from this nonstationarity analysis. From *Table 24.17*, we also see, that essentially the period (76–81) dominates the structure of the whole series, since the estimates for this period are

very close to the overall estimates in *Table 24.10*.

24.3.3 Results for the trivariate $3R_t/R_t/B_t$ model

The overall model in *Table 24.11* was estimated by an AR(3) process. At lag 1 we see a complete interrelationship, while at lags 2 and 3 the 3-Month Money Rate produces no influence at all.

Table 24.19: Locally trivariate models: 3-Month Money/Call Money/Bond Rates (span = 36, order = 3).

Period	*Span*	*AIC*	*Variance matrix*	*Coefficient matrices*
75.4 78.3	4–39	−308.0	$\begin{pmatrix} 0.325 & & \\ 0.165 & 0.175 & \\ 0.013 & 0.014 & 0.0039 \end{pmatrix}$ $\lvert\Sigma\rvert = 0.86814 \times 10^{-4}$	$\begin{pmatrix} 0 & 0.84 & 0 \\ -0.24 & 1.37 & -2.75 \\ -0.02 & -0.01 & 1.46 \end{pmatrix}$ $\begin{pmatrix} 0 & 0 & 0 \\ 0 & -0.81 & 4.19 \\ 0 & -0.03 & -0.38 \end{pmatrix}$ $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0.59 & -2.3 \\ 0 & 0.04 & -0.18 \end{pmatrix}$
	(4–80	−465.5)		
78.4 81.6	40–80	−532.9	$\begin{pmatrix} 0.350 & & \\ 0.256 & 0.450 & \\ 0.020 & 0.035 & 0.021 \end{pmatrix}$ $\lvert\Sigma\rvert = 0.001681$	$3R_t$ R_t B_t $\begin{pmatrix} 0.88 & 0 & 1.18 \\ 0.58 & 0.41 & 0.78 \\ 0.03 & 0.02 & 0.96 \end{pmatrix}$ $\begin{pmatrix} 0 & 0 & 0.52 \\ 0 & 0 & 0.34 \\ 0 & 0 & 0.02 \end{pmatrix}$ $\begin{pmatrix} 0 & 0 & -1.59 \\ 0 & 0 & -1.05 \\ 0 & 0 & -0.05 \end{pmatrix}$

Looking at *Table 24.19*, we see that this pattern is a little more pronounced for both stationary periods. In the first period, from 75.4 to 78.3, the 3-Month Money Rate acts like a white noise process while the Call Money and the Bond Rate are very much interrelated up to lag 3. This result is a little bit surprising when compared with the univariate analysis in the previous section (*Table 24.4*). One explanation is certainly the drastic cut of the maximum lag length in order to enable the short-term estimation process. (Longer maximum lag lengths produce nonsense results).

For the second period, 78.4–81.6, we also estimated an AR(3) process. Now the 3-Month Money Rate has a 0.88 coefficient at lag 1 and shows feedback to the other variables. But the characteristic feature of the second period is the strong influence of the Bond Rate

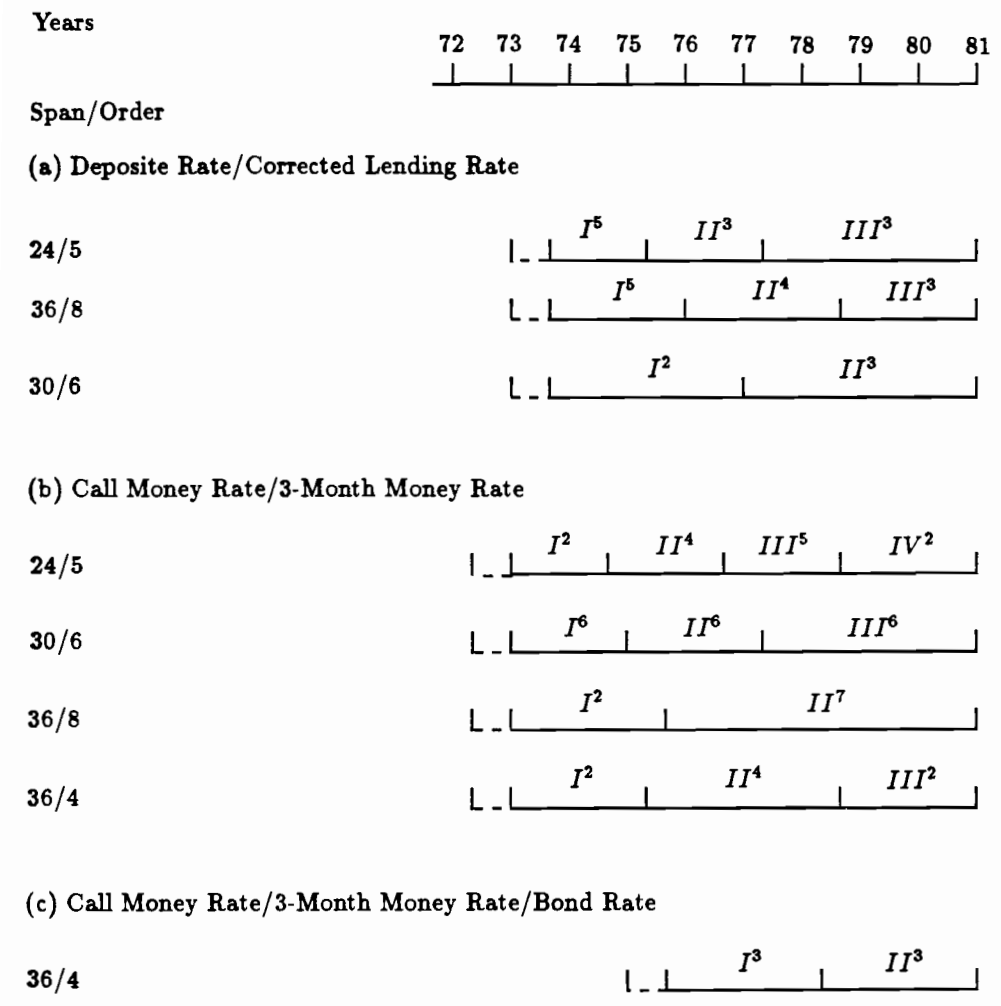

Figure 24.9: Summary: bivariate models.

on the other series throughout lag 1 to lag 3. The structural zeroes in the lag matrices 2 and 3 emphasize the strong unidirectional influence.

24.4 Local Causality Analysis

Based on the estimated vector autoregressive models of Sections 24.2 and 24.3, we can derive temporal (Wiener-Granger) causality measures. Using the concept of the generalized variance (determinant of the residual covariance matrix), Geweke defined three so-called "feedback" measures: unidirectional causality (= feedback) from X to Y ($X \rightarrow Y$), unidirectional causality from Y to X ($Y \rightarrow X$), and instantaneous causality ($X \star Y$). All these measures add up to a common so-called "measure of linear dependence", which can be interpreted as the information gain of a multivariate time series model to all (independent) univariate models.

These causality measures are defined as

$$\begin{aligned} F(X \rightarrow Y) &= \ln[Var(Y)/Var(Y|U)] \\ F(Y \rightarrow X) &= \ln[Var(X)/Var(X|U)] \\ F(X \star Y) &= \ln[Var(Y|U) \star Var(X|U)/Var(X,Y|U)] \end{aligned} \tag{24.5}$$

where "Var" stands for generalized residual variance, U denotes the chosen multivariate information set (past and present variables), and $(Y|U)$ and $(X|U)$ denote the Y and X block of the multivariate residual covariance matrix $(X, Y|U)$. More details on these measures can be found in Polasek (1983) and Geweke (1982). Because the three measures in (24.5) add up to one common measure, only their relative contribution are listed in *Figure 24.10.*

A better understanding of the changes in the relative sizes of the causality measures can be obtained, when we also consider the local mean and variances of the time series.

24.4.1 Deposit and Lending Rates

The local analysis for the pair Deposit (D)/Lending (L) Rate (only the Corrected Lending Rate is used in this section) is given in *Figure 24.10.* In part (b) of *Figure 24.10,* we find in the last row (label "all") the mean and the variance of the total time series: $\overline{D} = 5.42$ and $\overline{L} = 9.12$. This means that the interest differential in the period 73.7 to 80.7 is 3.7%, while the variances of both time series are approximately equal: $Var(D) = 0.316$ and $Var(L) = 0.305$. The means of the time series are written above the time axis, while the variances are listed below. (Note also the legends in the right column.)

The Geweke causality measures are given in part (a) of *Figure 24.10.* The overall measures are also found in the third row (label "all") and the three figures add to 100%. $F(L \rightarrow D) = 21.5\%$, $F(D \rightarrow L) = 17.5\%$ and $F(D \star L) = 61\%$. Instantaneous causality

is dominating, followed by an equal amount of feedback between the series. Proceeding to a 2-period local analysis (42 month = 3.5 years) in the second row of part (a) of *Figure 24.10*, we see that the instantaneous part is vanishing for the first period (73.7–77.1), but is more pronounced in the second period (77.1–80.7). This can be explained by the strong fluctuations of interest rates in the late 1970s and is also even more visible in the 2-year = 24-month analysis, given in the first row of part (a) of *Figure 24.10*. In the first of the 3 periods, the influence of the Deposit Rate on the Lending Rate was the strongest (82.8%), followed by two periods where the amount of feedback was approximately equal.

Economically, the strong ($D \rightarrow L$) causality can be explained by an inelastic demand for credits in the early 1970s in Austria. Banks were looking for money savings to satisfy this demand, which resulted in a markup of the lending rates. When the interest rates started to climb, this unidirectional influence was absorbed by high instantaneous causality.

The local mean and variance analysis exhibit some more details about the nonstationary behavior of both time series. The variances are the largest for the first period, because the Deposit Rate and Lending Rate started climbing about 1.5% within one year (74/75). The middle period shows the smallest amount of variation (but on a high level), while the last period has again a higher variation. This is because, after a drop of the interest rates during 1979, the figures started climbing again in 1980. This is also reflected in the movement of the differences of the local means to their overall mean, i.e., $\overline{D}_i - \overline{D}$ and $\overline{L}_i - \overline{L}$, which are listed above the time axis in the first and second rows of part (b) of *Figure 24.10*.

24.4.2 Call Money and 3-Month Money Rate

The local Geweke causality and the local mean/variance analysis of the pair Call Money (R) and 3-Month Money Rate ($3R$)can be found in the parts (a) and (b) of *Figure 24.11*, respectively. Because two more years of data were available, 4 periods could be used for the local analysis.

Looking at the third row of part (b) of *Figure 24.11*, we see that the overall mean of the Call Money Rate ($\overline{R} = 6.86$) was 1.1% less than the mean for the $3R$ Rate ($\overline{3R} = 7.95$). The variance of R is about one-third higher than that of $3R$. An increasing trend and fluctuations can be seen in both time series by the increase of the local mean and variances in the first and second rows of part (b) of *Figure 24.11*.

Part (a) of *Figure 24.11* shows that the instantaneous causality part was the biggest for both ends of the time series with almost no influence from $R \rightarrow 3R$ at the same time. The middle periods show the highest amount of feedback.

Summarizing, we can state the result that the influence directions have not been constant during the observation period. The changes are mainly reflected in a large or small amount of instantaneous causality. Both pairs of interest rates show an almost equal amount of feedback in all local periods, except the first and the last period of the pair $(R, 3R)$ where there is only influence from $R \rightarrow 3R$.

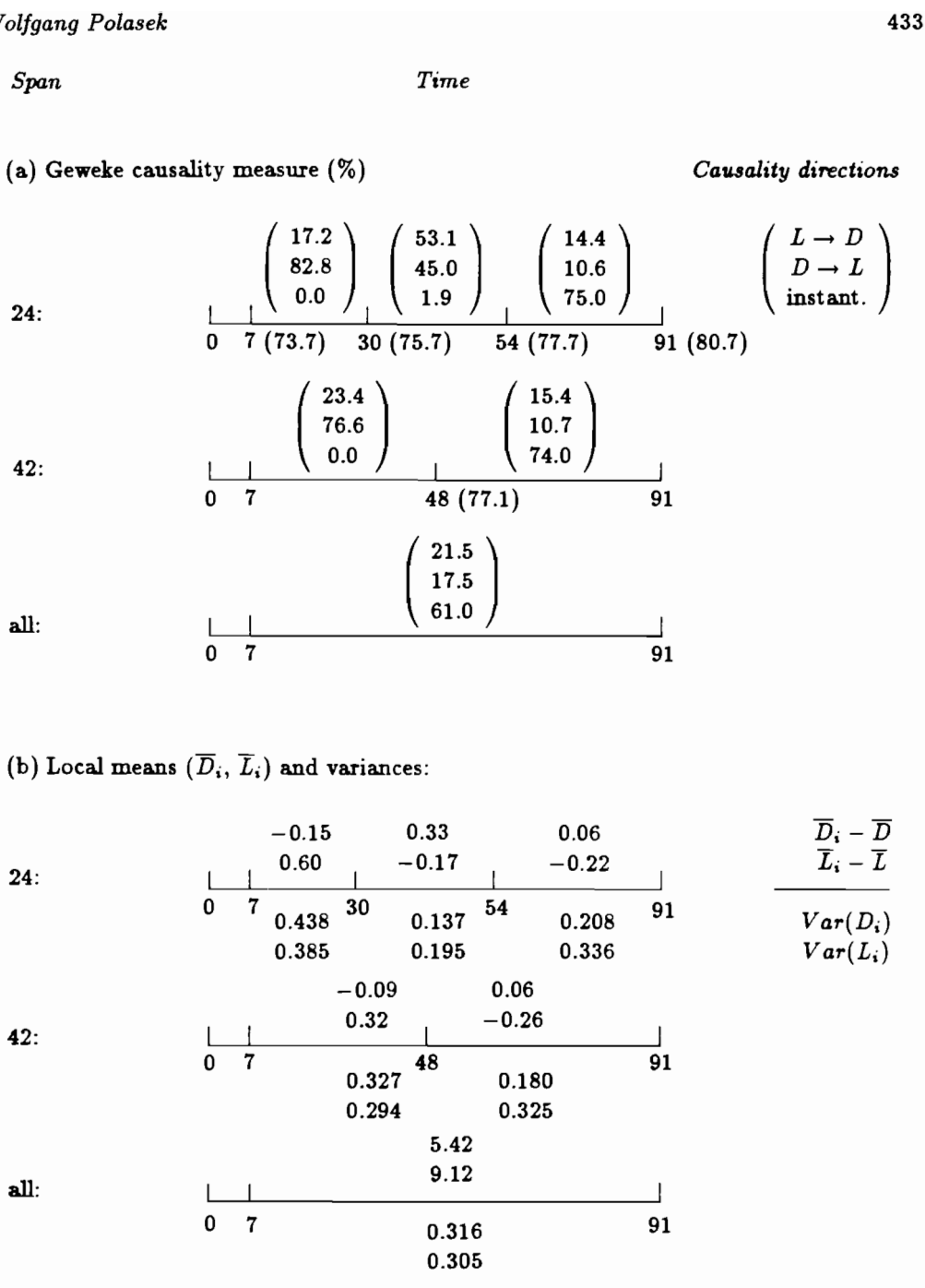

Figure 24.10: Local causality analysis for Austrian Deposit (D) and Lending (L) Rates: 73.1–80.7.

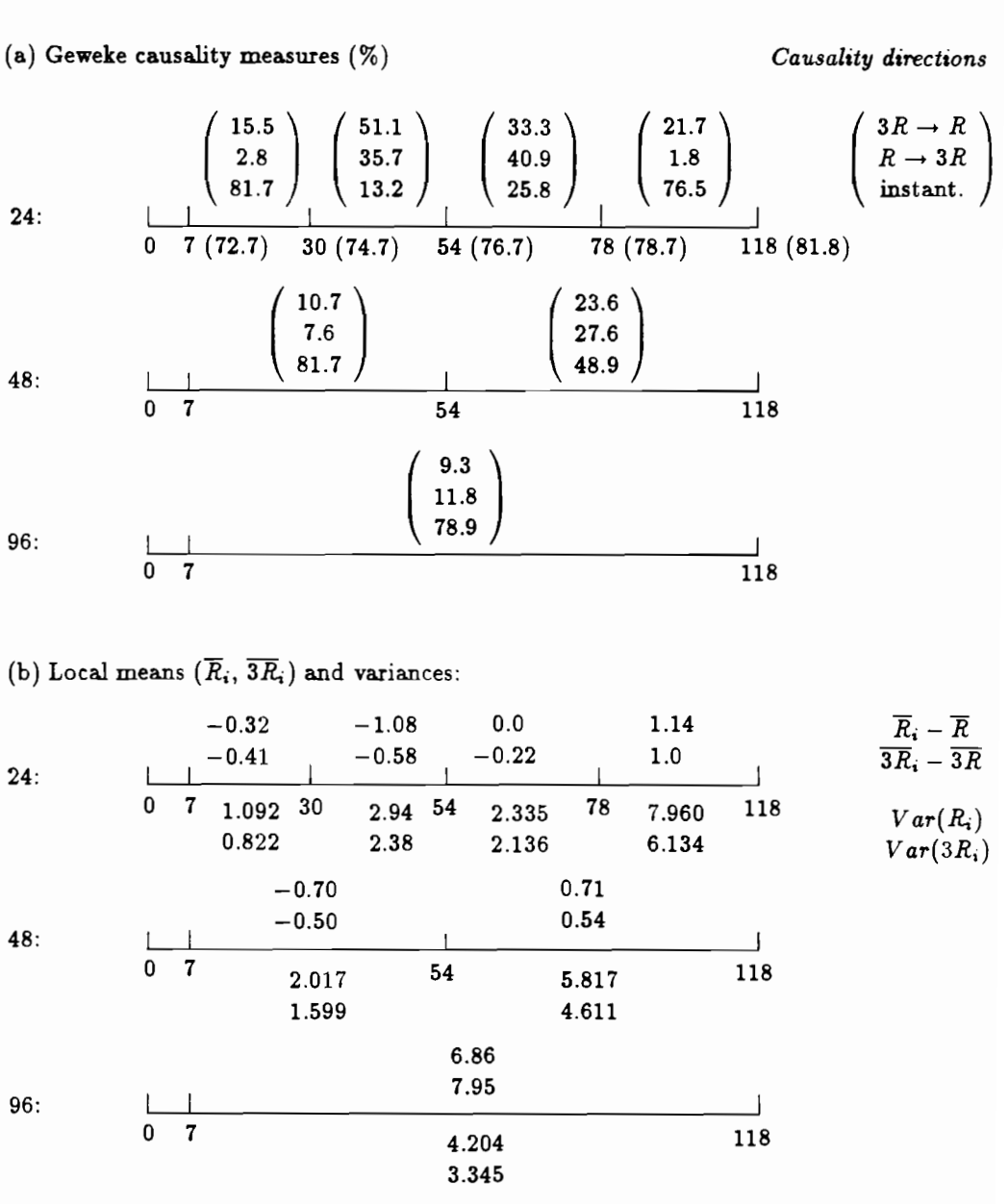

Figure 24.11: Local causality analysis for the Austrian Call Money (R) and the 3-Month (3R) Money Rate: 72.1–81.8.

For the pair (L, D) we find that the influence direction $D \rightarrow L$ is dominating in the first part of the observation period and is more balanced toward the end. Both results support the Polasek (1983) analysis on the same subject and give more insight into the dynamics of time series.

24.5 Locally moving AR-models

The third method we propose in this section is the analysis of the changing behavior of a time series by locally moving AR models. By prespecifing a fixed span n, we move a "window" of n consecutive points along the time axis. Again, univariate and multivariate models are estimated, and the Geweke measures are derived and assigned to the midpoint of the window-interval. Because every interval incorporates a new observation and drops an old one, the estimates are no longer independent. This is in contrast to the bisectrix approach, where the comparison of AIC values could be used for a likelihood ratio test.

The results of the local causality analysis is graphically summarized in a so-called causality profile, as in *Figure 24.12*. The standardized Geweke measures between two time series, X and Y, are ordered by the unidirectional influence from $Y \rightarrow X$, the instantaneous part $(X \star Y)$, and the other unidirectional part $X \rightarrow Y$. *Figure 24.12* shows the causality profile for the Call Money Rate and the 3-Month Money Rate (1972–1982) for a window span of 36 months. Note that the relative sizes of influence direction have obviously changed, but it is difficult to determine the exact location of the change points.

A simple way of getting a more clearcut profile is the application of so-called data smoothers, a method of the exploratory data analysis (EDA) field, to smooth the fluctuations of the causality profile. We suggest taking a running median of length 3 or 5 (i.e., $3R$, $5R$, or $53R$), because the smoothing of such series tends to produce plateaus (intervals with a constant smooth value). The transition periods from one plateau to another can be used to locate change points [see also Polasek (1983)].

The smoothed causality profile is shown in *Figure 24.12*. The window midpoints run from 73.11 to 80.3, and a window is centered on a midpoint, spreading out for ± 1.5 years. The causality profile shows that 3 or 5 intervals can be detected:

1. The first period with causality tripel (25, 66, 8) lasts from 73.11 to 74.11 for about one year. It is the starting period after the capital market was deregulated in 1973.
2. The second period (25, 25, 50) from 74.11 to 76.11, extends for two years.
3. A one year period follows with a 0-causality direction from $3R \rightarrow R(0, 80, 20)$ and interrupts the relatively homogenous period in the second half of the 1970s. This can be explained by the fact that 1977 was a year with balance-of-payments problems.
4. The next period returns to a more uniform causality distribution (33, 42, 25) lasting until 79.9. It reflects a consolidation phase after the 1977 balance-of-payment crisis year.

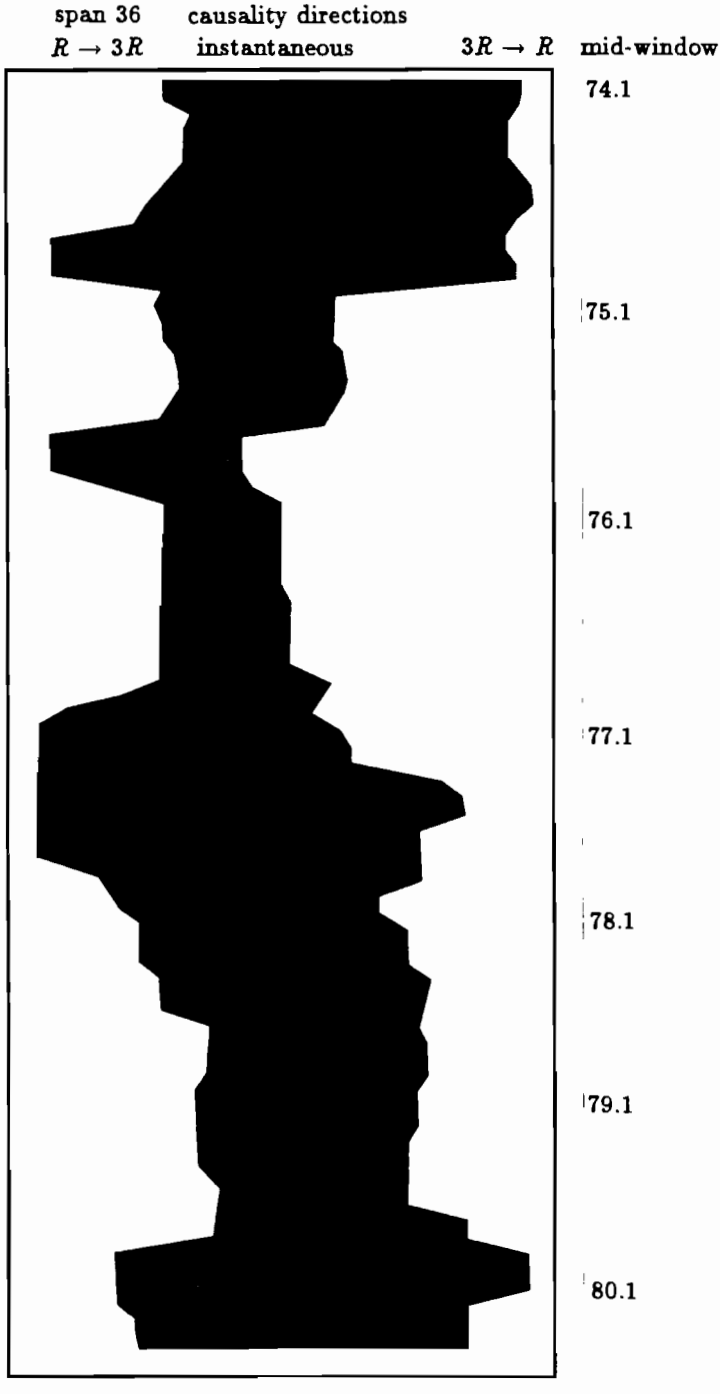

Figure 24.12: Smoothed causality profile of (R) the Call Money Rate, and ($3R$) the 3-Month Money Rate (span: 36 months).

5. The instantaneous causality part again increases in the last few months to 66% on account of unidirectional influences. This can be explained by the sharp increase of interest rates at the beginning of the 1980s.

Summarizing, we see that turbulent years tend to have a larger instantaneous causality share (up to 80%). This might be due to more alert behavior by all economic agents in such circumstances. Periods with stable developments reduce the instantaneous causality down to 20%, giving way to more influence from short-run to long-run (up to 50%) interest rates. Interestingly, the influence in the other direction can vanish completely for short periods.

The relation between the bisectrix method and the causality profile method is as follows: the causality measures of the bisetrix method can be explained as cuts in the causality profile, as is shown in *Figure 24.13.* Depending on the interval spans, the bisectrix technique is a summary of moving causality measures. We have to distinguish two cases:

1. When the bisectric span is smaller than the moving span, a point of the causality profile encompasses the bisectrix result. (In our case the moving span of 36 months is one year longer than the smallest bisectrix span.)

2. When the bisectrix span is larger than the moving span, the causality decomposition of the bisectrix model is a certain average of the moving models which fit into that window span of the causality profile.

It should be mentioned that a causality profile for an "evolving window" was also tried, starting from the lowest possible model up to the model for the total time series. The results of this method have not been satisfactory because one cannot get rid of bad observations. Bad observations have a very sensitive (nonrobust) impact on the estimation of time series models. High sensitivity to extreme observations is the disadvantage of the evolving method, while the disadvantage of the moving window method is its possible dependence on the window span. However, this example demonstrates that the combination of different approaches provides a good starting point to attack the difficult field of changing causality directions.

24.6 Conclusions

The analysis of 5 Austrian interest rates with local stationary models shows that homogeneous blocks of stationary developements are in general very short. Therefore, the assumption of a stationary monthly model for the whole period between 1972 and 1981 is not a very valid one. On the other side, forecasting properties can not be improved if the differencing procedure of Box-Jenkins modeling is applied [see also Polasek (1983)]. Therefore, we are faced with a specific nonstationary behavior, which cannot be solved by traditional methods of stationarity transformations.

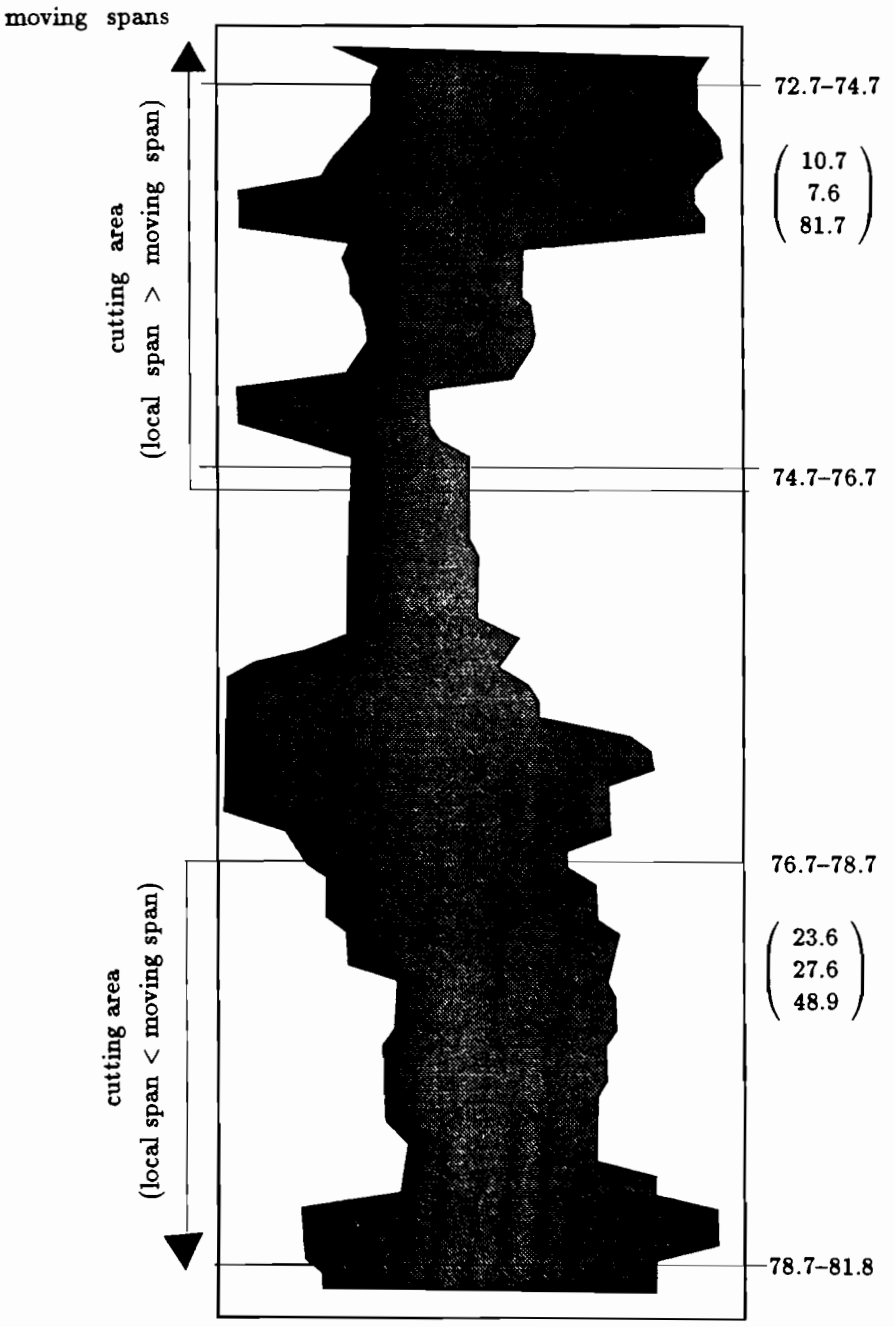

Figure 24.13: Local causality measures: relationships between moving and bisectrix measures for Austrian Call Money Rate (R), and 3-Month Money Rate ($3R$).

Despite the fact, that only a limited amount of monthly data are available for fitting local nonstationary models (note that the parameter/observation ratio becomes close to one), we find the following results:

1. In univariate models, we can find up to 4 different stationary blocks, if we choose a block length of 18 months and longer.
2. The longest stationary block was found for the Call Money Rate between 1974 and 1978.
3. Most of the blocked models (13 are AR(1) processes; 10 models are AR(2) or AR(3) processes.
4. Only 4 blocks of the Deposit, Corrected Lending, 3-Month Money, and Bond Rates show seasonal cyclic behaviors (see *Figure 24.7*).
5. If the block length is reduced from 18 to 12 months, more stationary intervals can be found (see *Figure 24.8*).
6. Results of bivariate blocked models are more sensitive to the chosen block length and maximum lag length than univariate models (see *Figure 24.8*).
7. In general, 3 blocks can be found for the pair Deposit/Corrected Lending Rate and Call Money/3-Month Money Rate.
8. Two blocks are found for the bivariate model Call Money/3-Month Money/Bond Rate.
9. Local causality results show that the instantaneous influence is dominating on both ends of the time series.
10. Most local periods show balanced feedback between the time series. The direction $D \rightarrow 3R$ is dominating in the first and the last local period; the direction $D \rightarrow L$ is dominating in the first half of the time series.

References

Akaike, H. (1973), Information theory and an extension of the maximum likelihood principle, in : B.N. Petrov and F. Csaki (eds.), Second International Symposium on Information Theory, Budapest: Akademiai Kiado.

Akaike, H. (1974), A new look at the statistical model identification. *IEEE Transaction on Automatic Control*, **999**, 716-723.

Akaike, H. (1978), A Bayesian analysis of the minimum AIC procedure. *Annals of the Institute of Statistical Mathematics*, **30-A**, 9-14.

Akaike, H. Kitagawa, G., Arahata, E., and Tada, F. (1979), *TIMSAC-78*, Computer Science Monographs, No.11. Tokyo: Institute of Statistical Mathematics.

Geweke, J. (1982), Measurements of linear dependence and feedback between time series. *Journal of the American Statistical Association*, **77**, 304-324.

Kitagawa, G. and Akaike, H. (1978), A procedure for the modeling of nonstationary time series. *Annals of the Institute of Statistical Mathematics*, **30-B**, 351-363.

Kitagawa, G. and Akaike, H. (1981), On TIMSAC-78, pp. 449-547 in: D. Findley (ed.), *Applied Time Series Analysis II*. New York: Academic Press.

Polasek, W. (1983), Multivariate time series models for Austrian interest rates. *Empirica*, **2**, 129-157.

CHAPTER 25

Investment, Taxation, and Econometric Policy Evaluation: Some Evidence on the Lucas Critique

Jean-Marie Dufour

Summary

The aggregate investment schedule may be used to study the impact of various policy measures, such as changes in corporate tax rates, depreciation allowances, and investment tax credits. Its parameters should be invariant with respect to the policy changes themselves, a point forcefully stressed by Lucas (1976). On the impact of investment tax credits, Lucas makes two predictions: first, if the model is implemented under an assumption of static expectations (versus rational expectations) and estimated from a period during which policy rules changed appreciably, it will exhibit parameter instability; second, the impact of tax credits is likely to be heavily underestimated. This chapter presents empirical evidence on both these effects by studying a version of the Hall-Jorgenson model estimated from US data (1956–1972). For this purpose, we use recursive stability analysis, an exploratory methodology that makes very weak assumptions on the form of the instability to be detected and provides indications on the direction of prediction errors. The main finding is a discontinuity associated with the first imposition of the tax credit (1964–1966); further, this shift led to underprediction of investment. The results thus support Lucas's hypothesis.

25.1 Introduction

The stability over time of the aggregate investment schedule has great importance for macroeconomic policy. In particular, one may use this relationship to study the impact of various policy measures, such as changes in nominal corporate tax rates, changes in depreciation allowances, investment tax credits, and the like. An ingenious formulation of an investment function making possible such studies is due to Hall and Jorgenson (1967). This model was employed, for example, by Gordon and Jorgenson (1976) to study the impact of investment tax credits in the United States over the period 1960–1985.

It is easy to understand that the model used for such policy simulations should exhibit a good stability over time. In particular, the parameters should be invariant with respect to the policy changes themselves, a point forcefully stressed by Lucas (1976). This author argues that parameters in econometric relationships reflect economic agents' decision rules: since the latter integrate knowledge about policies, changes in policy rules are likely to induce shifts in the parameters. Lucas describes three cases where such phenomena could be observed: the first one deals with income transfers and the aggregate consumption function, the second one concerns studies of the effect of investment tax credits with the help of the Hall-Jorgenson model of investment demand, while the third one is based on the Phillips curve. In this chapter, we provide empirical evidence on this issue by considering the second example.

In this case, Lucas argues that the effect of a change in the rate of an investment tax credit depends crucially on expectations concerning future changes in this rate: the impact of a change in the tax credit differs, depending on expectations about future changes of the tax credit. In other words, the response coefficient to a change in the rate of the tax credit depends on expectations about future changes of this rate. In particular, after developing a simple investment model, Lucas shows that the impact of a given change may be substantially bigger if it is viewed as transitory rather than permanent (once-and-for-all) [1]. Consequently, if an investigator assumes that changes in the investment tax credit are viewed as permanent by the relevant economic agents, while the latter in fact view it as transitory, he may appreciably underestimate the impact of the tax credit. Thus, to forecast accurately the effect of a proposed change in the tax credit, it is important

1. to make correct assumptions concerning expectations on future changes in the tax credit that will follow a proposed change;

2. to specify and estimate the model under correct expectational assumptions over the historical period used for estimation.

Note here that Hall and Jorgenson (1967), as well as Gordon and Jorgenson (1976), assumed that changes in tax rates were viewed as permanent.

To get evidence on the Lucas critique, we shall reexamine the same model and data as Gordon and Jorgenson (1976). Over the sampling period used for the estimation of their investment function (1956–1972), five major changes in the tax credit took place.

The tax credit was originally introduced in 1962 to stimulate investment. Then "the effectiveness of the tax credit was increased substantially in 1964 with the repeal of the Long Amendment [2]. The investment tax credit was suspended in 1966–1967 and repealed in 1969 in order to reduce the level of investment. The tax credit was re-enacted in 1971 to stimulate investment expenditures" [3]. These events suggest that policy regime changes took place over the period considered and, from Lucas' argument, we should observe parameter instability in the Gordon-Jorgenson model (unless expectations effectively obeyed the scheme implicitly assumed by Hall and Jorgenson). Further, since it is argued that the assumption of static expectations should lead to underestimating the impact of the tax credit, we also expect that the introduction of the investment tax credit be associated with underpredictions of investment expenditures.

To study such general effects, we need an exploratory methodology that is sensitive to a wide variety of possible structural changes and capable of providing information on the timing of structural change. Further, it should give indications on the direction of prediction errors associated with the use of a model. An attractive procedure of this type consists of estimating the model recursively (adding one observation at a time) and examining a number of resulting statistics. This approach was first formalized by Brown *et al.* (1975); a systematization as well as a number of extensions were provided by Dufour (1979, 1982, 1986). [For further work along those lines, see also Hackl (1980)]. Because it is especially well adapted to our objectives, this is the approach we will follow to study the Lucas effects.

In Section 25.2, we present the investment model that will be analyzed. In Section 25.3, we describe succinctly the methodology used and define the main statistics considered. In Section 25.4, we present the empirical results. In Section 25.5, we summarize our findings and conclusions.

25.2 The Model

The model studied by Gordon and Jorgenson (1976) is based on quarterly data and has the form

$$IPDE58_t = \alpha + \delta K_t + \sum_{i=0}^{6} \beta_i V_{t-i} + u_t \tag{25.1}$$

where $IPDE58_t$ is real investment (1958 dollars) in producers' durable equipment (during period t), K_t is gross beginning-of-period real capital stock of producers' durable equipment, V_t is a proxy for the desired capital stock defined as

$$V_t = (PGNP_{t-2})(GNP58_{t-1}/C_{t-2} \tag{25.2}$$

$GNP58_t$ is the real gross national product (1958 dollars), which, like $IPDE58_t$, is seasonally adjusted and measured at annual rates; $PGNP_t$ is the GNP price deflator, C_t is

the rental cost of capital, and u_t is a random disturbance. The cost of capital C_t is given by the expression

$$C_t = PIPDE_t[0.138 + R_t(1 - U_t)][1 - U_t Z_t - TC_t + Y_t Z_t TC_t U_t]/(1 - U_t) \quad (25.3)$$

where $PIPDE_t$ is the price deflator for investment in producers' durable equipment, 0.138 is the depreciation rate on producers' durable equipment as calculated by Christensen and Jorgenson (1969), U_t is the nominal corporate tax rate, R_t is the interest rate on new issues of high-grade corporate bonds, Z_t is the present discounted value of depreciation allowances, TC_t is the effective tax credit and Y_t equals 1.0 during those years in which the Long Amendment applied and zero otherwise.

In order to estimate this model, Gordon and Jorgenson (1976) used a second-degree Almon polynomial lag structure constrained to pass through zero after seven periods. This imposes the restrictions

$$\beta_i = a_0 - a_1 i - a_2 i^2 \ , \quad i = 0, 1, \ldots, 7 \quad (25.4)$$

with $\beta_7 = a_0 - 7a_1 - 49a_2 = 0$, so that there are effectively only two free parameters in the distributed lag on V_t. Under these conditions, the equation to be estimated is

$$IPDE58_t = \alpha + \delta K_t + a_1 W_{1t} + a_2 W_{2t} + u_t \quad (25.5)$$

where

$$W_{1t} = \sum_{i=0}^{6}(7 - i)V_{t-i} \quad \text{and} \quad W_{2t} = \sum_{i=0}^{6}(49 - i^2)V_{t-i} \ . \quad (25.6)$$

Furthermore, since the original Durbin-Watson statistic was 0.7554, a first-order autoregressive transformation was used (with $\hat{\rho} = 0.6223$, estimated by the Cochrane-Orcutt method). The following equation, based on the period 1956/I–1972/IV, was finally obtained:

$$\begin{aligned} IPDE58_t = {} & \underset{(1.522)}{-9.656} + \underset{(0.0163)}{0.0572}\ K_t + \underset{(0.00071)}{0.00181}\ V_t + \underset{(0.00033)}{0.00218}\ V_{t-1} \\ & + \underset{(0.00019)}{0.00233}\ V_{t-2} + \underset{(0.00031)}{0.00228}\ V_{t-3} + \underset{(0.00038)}{0.00202}\ V_{t-4} + \underset{(0.00036)}{0.00156}\ V_{t-5} \\ & + \underset{(0.00023)}{0.00088}\ V_{t-6}, \qquad R^2 = 0.9577, \quad DW = 1.9788, \quad SE = 1.0150. \end{aligned} \quad (25.7)$$

The sample 1956/I–1972/IV represents effective observations, not including those observations that are "lost" because of the presence of lagged explanatory variables and the

autoregressive transformation. The standard errors are given in parentheses. R^2 is the multiple correlation coefficient, DW is the Durbin-Watson statistic, and SE is the standard error of the regression (all for the transformed model).

This model is based on a static-expectations assumption [see equation (25.2)]. By contrast, in his theoretical argument, Lucas (1976) considers a tax credit that follows a Markovian scheme, which includes as special cases both a permanent credit (i.e., the probability that the tax credit will disappear is zero) and a frequently imposed but always transitory credit (i.e., the probability that the tax credit will disappear soon is high). Assuming rational expectations on the part of investors, he then shows that the impact of the tax credit can be much bigger if it is viewed as transitory rather than permanent. Indeed, under reasonable values of the model parameters, the ratio of the actual to predicted effect may be in the range of 4 to 7.

In this chapter, we study the stability over time of the above model. For this purpose, we use an "exploratory" methodology aimed at being sensitive to a wide variety of instability patterns. It is based on estimating recursively the model under study and considering associated paths of coefficient estimates and prediction errors. An especially interesting aspect of this approach for our problem is that it can give us information on the timing of parameter shifts and the direction of prediction errors, two issues for which Lucas's conjecture has implications. In the next section, we give a succinct description of the methodology employed.

On the basis of this approach, we shall present (in section 25.4) the results of three different recursive estimation experiments with the same data as Gordon and Jorgenson (1976). In the first one, we simply estimate equation (25.5) recursively by ordinary least squares. In the second one, we take into account the fact that Gordon and Jorgenson (1976) made a correction for "autocorrelation" (which, however, may only be an *ad hoc* response to a parameter instability problem) and we study how the conclusions are affected after making such a correction. We thus estimate recursively the transformed model

$$IPDE58_t(\hat{\rho}) = \alpha(1-\hat{\rho}) + \delta K_t(\hat{\rho}) + a_1 W_{1t}(\hat{\rho}) + a_2 W_{2t}(\hat{\rho}) + \varepsilon_t^\star \tag{25.8}$$

where $\hat{\rho} = 0.6223$, $IPDE58_t(\hat{\rho}) = IPDE58 - \hat{\rho}\, IPDE58_{t-1}$, $K_t(\hat{\rho}) = K_t - \hat{\rho} K_{t-1}$, etc. [See Dufour (1982, Section 2.5) for a discussion of this procedure. Note that ρ is not recursively estimated.]

Finally, in the third experiment, we try to deal with an extra difficulty: since the capital stock K_t is a function of past investment, K_t cannot, strictly speaking, be taken as independent of the disturbance vector. The regressor K_t may be viewed as a form of lagged dependent variable, and the tests performed in the two first experiments cannot be considered exact. A suggested in Dufour (1982, Section 2.5), we get rid of the troublesome regressor $K_t(\hat{\rho})$ by subtracting $\hat{\delta} K_t(\hat{\rho})$ on both sides of (25.6) where $\hat{\delta}$ is the estimate of δ based on the full sample. We thus consider the regression

$$\mathrm{IPDE58}_t(\hat{\rho}) - \hat{\delta} K_t(\hat{\rho}) = \alpha(1-\hat{\rho}) + a_1 W_{1t}(\hat{\rho}) + a_2 W_{2t}(\hat{\rho}) + \varepsilon_t^\star \tag{25.9}$$

where $\hat{\delta} = 0.0572$ and $\hat{\rho} = 0.6223$, and perform the recursive estimation experiment on the remaining coefficients. Of course, this third experiment loses some of the advantages of "recursivity" (δ is not estimated recursively), which may lead to a loss of power. But it appears necessary in the present circumstances as a way of cross-checking the results obtained without taking into account the presence of a lagged dependent variable.

25.3 Methodology

In this section, which draws heavily on Dufour (1986, Section 2.3), we sketch the main features of recursive stability analysis and define the main statistics used. For a detailed description and more complete bibliography, the reader is referred to Dufour (1982).

Let us consider the following regression model:

$$y_t = \boldsymbol{x}_t'\beta + u_t, \quad t = 1, \ldots, T \tag{25.10}$$

where y_t is an observation on the dependent variable, $\boldsymbol{x}_t$ is a $K \times 1$ column vector of nonstochastic regressors, β is a $K \times 1$ vector of regressor coefficients, u_t is a disturbance term that follows a normal distribution with mean zero and variance σ^2. Assume also that the disturbances $u_1, \ldots, u_T$ are independent.

We wish to investigate the constancy of the regression coefficient β over different observations. In other words, we consider the alternative model

$$y_t = \boldsymbol{x}_t'\beta_t + u_t, \quad t = 1, \ldots, T \tag{25.11}$$

and wish to test the null hypothesis H_0: $\beta_1 = \ldots = \beta_T \equiv \beta$.

When the data have a natural order (e.g., time), a simple way to investigate the stability of regression coefficients is to estimate the model recursively. Using the first K observations in the sample to get an initial estimate of β, we gradually enlarge the sample, adding one observation at a time, and reestimate β at each step. We get in this way the sequence of estimates

$$b_r = (X_r'X_r)^{-1}X_r'Y_r, \quad r = K, \ldots, T, \tag{25.12}$$

where $X_r' = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r)$ and $Y_r = (y_1, \ldots y_r)'$. We assume here that $\text{rank}(X_r) = K$, $r = K, \ldots, T$. A computationally efficient algorithm allows one to get this sequence easily [see Brown *et al.* (1975, p. 152)]. It is intuitively clear that the examination of this sequence of estimates can provide information on possible instabilities of the regression coefficients. We see also that two different permutations of the data usually yield different sequences of estimates. However, when the data are ordered (e.g., by time), it appears that the most easily interpretable results will be obtained by putting the data either in their original time order or in the reverse order. In the first case, one gets "forward recursive

estimates"; and in the second case, "backward recursive estimates". In this chapter we will concentrate our attention on forward recursive estimation.

Recursive estimates are a descriptive device reflecting the influence of different observations in a sequential updating process. However, recursive estimates are strongly autocorrelated, even under the null hypothesis of stability, and the analysis of their behavior remains delicate from a statistical point of view. One can show easily that recursive estimates follow a "heteroscedastic random walk"; see Dufour (1982, eq.24). Thus the observation of a "trend" must be interpreted with great care. Consequently, it is important to have statistics that are easier to interpret. For this purpose, we look at the associated sequences of prediction errors. Namely, we consider the sequences of prediction errors

$$v_{kr} = y_r - x_r' b_{r-k}, \quad r = K+k, \ldots, T \tag{25.13}$$

where $1 \leq k \leq T-K$. Since these have different variances, it is convenient to standardize them and to compute

$$w_{kr} = \frac{v_{kr}}{d_{kr}}, \quad r = K+k, \ldots, T \tag{25.14}$$

where $d_{kr} = [1 + x_r'(X_{r-k}' X_{r-k})^{-1} x_r]^{\frac{1}{2}}$. We call the sequence w_{kr}, $r = K+k, \ldots, T$, "k-steps-ahead recursive residuals" — or simply "k-step recursive residuals". Depending on whether the sample is in forward or backward order, we get "forward" or "backward recursive residuals". It is easy to verify that, under the null hypothesis of stability,

$$E(w_{kr}) = 0, \quad E(w_{kr}^2) = \sigma^2.$$

Further, when $k = 1$, one has

$$E(w_{1r} w_{1s}) = 0, \quad r \neq s$$

so that the sequence $w_{1r}, r = K+1, \ldots, T$, is a normal white noise. For $k \geq 2$, the sequence w_{kr}, $r = K+k, \ldots, T$, is dependent but only up to lag $k-1$ [see Dufour (1982, pp. 41–44)].

It is not difficult to determine how relatively simple forms of structural change will affect the behavior of prediction errors (or recursive residuals). For example, a sudden shift in the coefficients at some point t_0 will, in many circumstances, lead to an increase in the size of prediction errors and/or a tendency to either overpredict or underpredict the dependent variable (for $t \geq t_0$); a systematic drift in one or several of the coefficients will often lead to a systematic tendency to over- or underpredict; etc. Thus, we will first use the sequences of standardized prediction errors to perform an exploratory analysis and search for patterns indicative of possible structural shifts. For this purpose, it is especially useful to look at several "clues". The simple statistical properties of the one-step recursive residuals (forward or backward) designate them as the basic instrument of

analysis for that search. However, we will find instructive to compare these with the k-step recursive residuals ($k \geq 2$): since the latter are forecasts using estimates from a sample further away in time, they may exhibit better-defined and more recognizable patterns; they can also help to identify possible breakpoints.

When interpreting and comparing various sets of residuals, it is always useful to recall that all recursive residuals have the same standard error σ (under the null hypothesis). Interpretation will generally be easier if we scale the residuals with an estimate of σ. Since the most natural estimate is the one obtained from the full sample (i.e., the standard error of the regression), one computes

$$\tilde{w}_{kr} = \frac{w_{kr}}{\hat{\sigma}}, \quad r = K + k, \ldots, T \tag{25.15}$$

where

$$\hat{\sigma}^2 = \frac{(Y_T - X_T b_T)'(Y_T - X_T b_T)}{(T - K)} \quad .$$

This procedure can also help display the recursive residuals, for in most practical situations, it will bring all residuals in a convenient scale — not too close to zero and well inside the interval $(-10, +10)$ [4].

Though the first purpose of the instruments we defined is exploratory, it is important to assess the statistical significance of what is observed. Because one-step recursive residuals have such simple statistical properties, we will use these for this assessment. Roughly speaking, we expect two main types of effects to result from structural shifts: tendencies either to over- or underpredict the dependent variable and discontinuities in the size of the prediction errors. Consequently, we will compute a number of simple statistics aimed at being sensitive to these characteristics. Statistics especially sensitive to the sign of prediction errors include the CUSUM test originally suggested by Brown *et al.* (1975), a Student t-test and the corresponding Wilcoxon signed-rank test, run tests based on the number of runs and the length of longest run, and serial dependence tests. Statistics sensitive to discontinuities in the size of prediction errors include the CUSUM of squares test suggested by Brown *et al.* (1975) and heteroscedasticity tests. We now define succincty the various test statistics.

If we let $w_t \equiv w_{1t}$, $t = K + 1, \ldots, T$, the CUSUM test is based on considering the cumulative sums

$$W_r = \frac{1}{\hat{\sigma}} \sum_{j=K+1}^{r} w_j, \quad r = K + 1, \ldots, T \tag{25.16}$$

where $\hat{\sigma}^2 = S_T/(T - K)$ and $S_T = \sum_{t=K+1}^{T} w_t^2$. The null hypothesis H_0 is rejected at level α if $C \equiv \max_{K+1 \leq r \leq T} |\tilde{W}_r| \geq c_\alpha$, where

$$\tilde{W}_r = \frac{W_r}{\{\sqrt{T-K} + 2[(r-K)/\sqrt{T-K}]\}} \tag{25.17}$$

and c_α depends on the level of the test ($c_{0.01} = 1.143$, $c_{0.05} = 0.948$, $c_{0.10} = 0.850$). In other words, H_0 is rejected if the graph of W_r crosses either one of two straight lines determined by the level of the test. The t-test is based on the standard Student's t-statistic to test the hypothesis that the recursive residuals have zero mean against a systematic tendency to over- or underpredict. It is based on the statistic

$$\tilde{t} = \frac{\sqrt{T-K}\,\overline{w}}{s_w} \tag{25.18}$$

where

$$\overline{w} = \sum_{t=K+1}^{T} \frac{w_t}{T-K}, \qquad s_w^2 = \sum_{t=K+1}^{T} \frac{(w_t - \overline{w})^2}{T-K-1} \quad ; \tag{25.19}$$

under the null hypothesis, $\tilde{t}$ follows a Student's t-distribution with $T-K-1$ degrees of freedom. The Wilcoxon signed-rank test is based on the statistic

$$S = \sum_{t=K+1}^{T} u(w_t)\, R_t^+ \tag{25.20}$$

where

$$\begin{aligned} u(z) &= \begin{cases} 1, & \text{if } z \geq 0 \\ 0, & \text{if } z < 0 \end{cases} \\ R_t^+ &= \sum_{i=K+1}^{T} u(|w_t| - |w_i|) \quad . \end{aligned} \tag{25.21}$$

We may view it as a robust alternative to the t-test; its distribution (for $T \leq 50$) has been extensively tabulated by Wilcoxon *et al.* (1973, pp. 171–259). For $T > 50$, one can use the standardized form $X' = [S - E(S)]/\sigma(S)$, where $E(S) = n(n+1)/4$, $\sigma(S) = [n(n+1)(2n+1)/24]^{\frac{1}{2}}$ and $n = T - K$. Under H_0, S' is approximately $N(0,1)$.

An intuitively attractive way of looking at the sequence of the recursive residuals consists of observing runs of overpredictions (or underpredictions), as defined by the sequence $u(w_t)$, $t = K+1, \ldots, T$. Two simple tests are then obtained by considering the number of runs R in this sequence or by observing the length of the longest run in the sequence. The distribution of the number of runs R is obtained by noting that $R-1$ follows a binominal distribution $\text{Bi}(T-K-1, \frac{1}{2})$; a small number of runs suggests that the model has a tendency to overpredict or underpredict. Besides, an especially long run of negative or

positive errors of prediction suggests the presence of a shift. The probability of getting at least one run of a given length or greater may be computed from formulas (135) to (138) in Dufour (1982).

In a large number of cases, structural change leads to situations where the means of the cross products $Z_{kt} \equiv w_t w_{t+k}, t = K+1, \ldots, T-k$ (where $1 \leq k \leq T-K-1$) differ from zero [see Dufour (1982, pp. 52-55)]. This suggests testing whether Z_{kt} has mean zero: we can do this by using "serial dependence tests" not corrected for the mean (for such tests are more accurately viewed in this context as location tests rather than serial dependence tests). We will consider two types of statistics for doing this: the modified von Neumann ratio

$$VR = \frac{(n-1)^{-1} \sum_{t=K+1}^{T-1} (w_{t+1} - w_t)^2}{n^{-1} \sum_{t=K+1}^{T} w_t^2} \tag{25.22}$$

where $n = T - K$, and rank Wilcoxon-type serial dependence tests based on statistics of the form

$$S_k = \sum_{t=K+1}^{T-k} u(Z_{kt})\, R_{kt}^{+}, \qquad k = 1, 2, \ldots \tag{25.23}$$

where $R_{kt}^{+} = \sum_{i=K+1}^{T-k} u(|Z_{kt}| - |Z_{ki}|)$. VR provides an exact parametric test of the null hypothesis $E(w_t w_{t+1}) = 0, t = K+1, \ldots, T-1$; for a table, see Theil (1971, pp. 728–729). Each statistic S_k is distributed like the Wilcoxon signed-rank statistic to test the zero median hypothesis; it gives an exact test of the null hypothesis $E(w_t w_{t+k}) = 0$, $t = K+1, \ldots, T-k$, where $k \geq 1$ [see Dufour (1981)]. Further, under H_0, $E(S_k) = n_k(n_k+1)/4$, and $\sigma(S_k) = [n_k(n_k+1)(2n_k+1)/24]^{1/2}$, where $n_k = T - K - k$.

The CUSUM of squares test is based on considering the statistic

$$S_r = \frac{\sum_{j=K+1}^{r} w_j^2}{\sum_{j=K+1}^{T} w_j^2}, \qquad r = K+1, \ldots, T \tag{25.24}$$

The null hypothesis is rejected at level α if

$$S \equiv \max_{1 \leq j \leq T-K-1} |S_{K+j} - \frac{j}{T-K}| \geq d_\alpha \tag{25.25}$$

where d_α is obtained by entering Table 1 of Durbin (1969) at $\alpha/2$ and $n = (\frac{1}{2})(T-K) - 1$ if $T - K$ is even, or by interpolating linearly between $n = (\frac{1}{2})(T-K) - (3/2)$ and $n = (\frac{1}{2})(T-K) - (\frac{1}{2})$ if $T - K$ is odd. We do not use heteroscedasticity tests in this paper and so do not need to define them here.

Whenever possible we will report the marginal significance level (p-value) of each statistic. Of course, a test significant at a very low level provides stronger evidence against the null hypothesis. Note also that any of the tests suggested above can be applied to a subset of the one-step recursive residuals, provided this subset is suggested by *a priori* considerations (e.g., dates of policy changes).

25.4 Recursive Stability Analysis of Investment Demand

As indicated in Section 25.2, the first experiment consists of estimating equation (25.5) recursively by ordinary least squares (1956/I–1972/IV) [5]. The recursive estimates obtained are listed in *Table 25.1* and graphed in *Figure 25.1(a–d)*; the corresponding recursive residuals (1, 2, 3, 4, and 8 steps ahead) are listed in *Table 25.2*, with a number of test statistics in *Table 25.3*, and they are graphed in *Figure 25.2(a–d)* [6].

When we look at the recursive estimates, we distinguish four main phases:

1. The first phase (beginning to 1961/I) is characterized by relatively large fluctuations (including some "weird" values, especially at the very beginning, which is not surprising for, at the beginning, the estimations are based on few observations) and by a rough trend (upward for α and a_1, downward for δ and a_2).

2. The second phase (1961/II–1963/III) is one of relative stability exhibiting no clear trend, expect for δ which increases after 1962/IV.

3. The third phase (1963/IV–1966/IV) displays well-defined trends (downward for α and a_2, upward for δ and a_1) during which all coefficients change sign.

4. Finally, during the fourth phase (1967/I–1972/IV), a_1 and a_2 move in directions opposite to the ones they followed in the previous phase, while α and δ seem stable.

Thus, the fourth quarters of 1963 and 1966 appear to be breakpoints.

When we examine one-step-ahead recursive residuals, we find no systematic tendency to over- or underpredict over the full period (as indicated by the global location tests in *Table 25.3*). However, we can observe a run of 13 consecutive underpredictions from 1963/IV to 1966/IV, a very surprising outcome if the model is correct: under the null hypothesis of stability, the probability of getting at least one run of this length or more is 0.0065. The total number of runs of either over- or underpredictions (16) is extremely small in relation to the sample size, and there is strong evidence of serial dependence (at least up to a distance of 3 quarters). Indeed, the trajectory of the one-step recursive residuals has several striking features. The first period (beginning to 1963/III) exhibits a tendency to overpredict (negative residuals). This phenomenon is also indicated by the CUSUM test [see *Figure 25.2(f)*]. Note also that the CUSUM of squares test is not significant (at level 0.05). Next, we note a long run of 13 consecutive underpredictions (1963/IV–1966/IV), a "breakpoint" between 1966/IV and 1967/I, another run of 9 underpredictions (1967/IV–1969/IV), while the sequel of the series looks relatively "random". We can also observe that two-, three-, and four-steps-ahead recursive residuals display basically the same pattern. The form of the pattern is in fact clearer from the latter than from the one-step residuals.

It is interesting to compare the trajectory of the one-step recursive residuals with the movement of the effective investment tax credit [7]. The long run of underpredictions starts in 1963/IV, which roughly coincides with the repeal of the Long Amendment (1964/I),

Table 25.1: Gordon-Jorgenson model: forward recursive estimates (OLSQ): 1956/I–1972/IV.

Quarter [a]	α	δ	a_1	a_2
56.04	-326.102	1.6091	-0.0183925	0.0018862
57.01	-355.362	1.7910	-0.0199375	0.0020607
57.02	-200.678	0.9335	-0.0113759	0.0011276
57.03	-258.690	1.2004	-0.0145302	0.0014536
57.04	103.729	-0.2716	0.0044718	-0.0004420
58.01	51.225	-0.3009	0.0038135	-0.0004790
58.02	-201.634	0.5238	-0.0078796	0.0006041
58.03	-213.437	0.5234	-0.0076991	0.0005612
58.04	-133.600	0.1898	-0.0028328	0.0000649
59.01	-121.596	0.1442	-0.0021972	0.0000024
59.02	-127.719	0.1655	-0.0024678	0.0000277
59.03	-86.004	0.0339	-0.0010597	-0.0000916
59.04	-57.819	-0.0476	-0.0005529	-0.0001204
60.01	-40.011	-0.0949	-0.0004649	-0.0001113
60.02	-39.388	-0.0965	-0.0004630	-0.0001109
60.03	-27.361	-0.1262	-0.0004526	-0.0000987
60.04	-12.912	-0.1608	-0.0005411	-0.0000723
61.01	7.845	-0.2074	-0.0007922	-0.0000198
61.02	20.961	-0.2326	-0.0011519	0.0000372
61.03	27.251	-0.2398	-0.0011481	0.0000465
61.04	29.245	-0.2407	-0.0011379	0.0000489
62.01	29.312	-0.2407	-0.0011375	0.0000489
62.02	28.384	-0.2413	-0.0011478	0.0000481
62.03	28.268	-0.2420	-0.0011981	0.0000533
62.04	28.814	-0.2319	-0.0007962	0.0000129
63.01	30.011	-0.2123	-0.0007552	0.0000171
63.02	31.475	-0.1941	-0 0011388	0.0000686
63.03	31.800	-0.1914	-0.0012565	0.0000832
63.04	31.227	-0.1943	-0.0010485	0.0000580
64.01	30.255	-0.1965	-0.0007345	0.0000205
64.02	28.943	-0.1961	-0.0004098	-0.0000180
64.03	27.405	-0.1945	-0.0001861	-0.0000452
64.04	26.252	-0.1911	-0.0000746	-0.0000586
65.01	23.723	-0.1807	0.0002025	-0.0000908
65.02	21.435	-0.1694	0.0004589	-0.0001199
65.03	17.259	-0.1464	0.0008914	-0.0001683
65.04	12.713	-0.1186	0.0013065	-0.0002138
66.01	7.462	-0.0837	0.0016886	-0.0002547
66.02	2.236	-0.0461	0.0019594	-0.0002822
66.03	-2.754	-0.0090	0.0022482	-0.0003115
66.04	-6.225	0.0172	0.0025062	-0.0003381
67.01	-3.042	-0.0074	0.0022301	-0.0003095
67.02	-2.278	-0.0128	0.0020597	-0.0002909
67.03	-1.595	-0.0184	0.0019768	-0.0002822
67.04	-2.542	-0.0093	0.0019180	-0.0002744
68.01	-5.921	0.0244	0.0015286	-0.0002263
68.02	-6.344	0.0287	0.0014691	-0.0002190
68.03	-7.195	0.0375	0.0013446	-0.0002039
68.04	-7.587	0.0415	0.0012932	-0.0001976
69.01	-8.738	0.0531	0.0011811	-0.0001834
69.02	-8.968	0.0554	0.0011641	-0.0001812
69.03	-9.197	0.0578	0.0011459	-0.0001788
69.04	-9.346	0.0592	0.0011719	-0.0001815
70.01	-8.954	0.0553	0.0010426	-0.0001677
70.02	-8.798	0.0536	0.0009942	-0.0001625
70.03	-9.123	0.0576	0.0009975	-0.0001622
70.04	-7.915	0.0421	0.0012129	-0.0001890
71.01	-7.838	0.0410	0.0012349	-0.0001916
71.02	-8.278	0.0469	0.0010713	-0.0001723
71.03	-8.479	0.0495	0.0009754	-0.0001612
71.04	-9.088	0.0572	0.0006227	-0.0001205
72.01	-9.427	0.0612	0.0004061	-0.0000956
72.02	-9.405	0.0609	0.0004209	-0.0000973
72.03	-9.213	0.0593	0.0005416	-0.0001110
72.04	-9.204	0.0593	0.0005455	-0.0001114

[a] End-of-sample quarter. All samples start in 56.01 (1956/I).

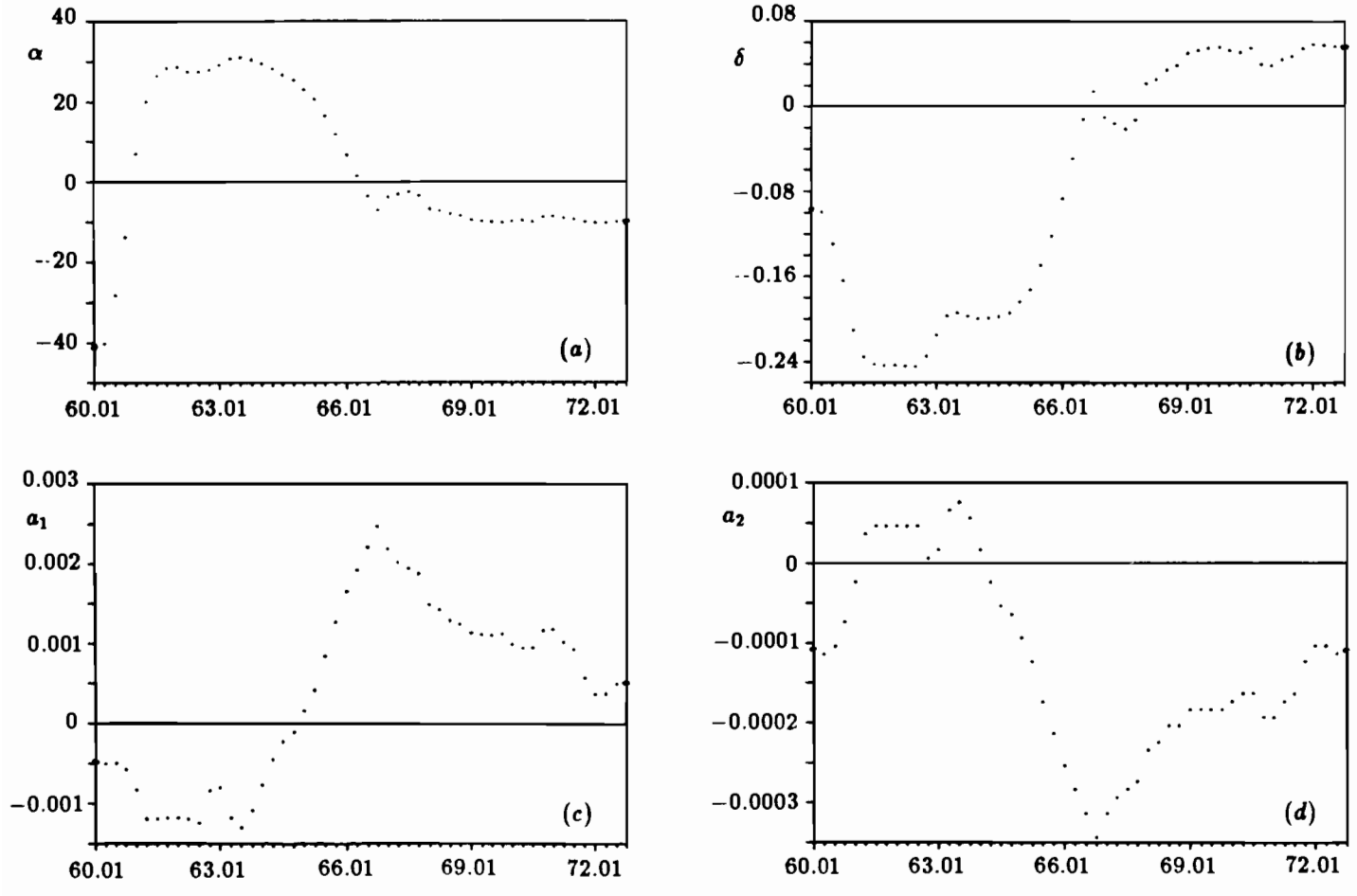

Figure 25.1: Gordon-Jorgenson model (OLSQ): recursive estimates.

Table 25.2: Gordon-Jorgenson model: forward recursive residuals (OLSQ): 1956/I–1972/IV

Quarter [a]	*1 step*	*2 steps*	*3 steps*	*4 steps*	*8 steps*
57.01	0.0352	-	-	-	-
57.02	-0.3443	-0.0826	-	-	-
57.03	0.1202	-0.2410	-0.0586	-	-
57.04	-0.9358	-0.3520	-0.4835	-0.1767	-
58.01	-2.0437	-1.9434	-0.7658	-0.6970	-
58.02	-0.9692	-1.6723	-1.4633	-1.1100	-
58.03	-0.7965	-1.2545	-2.1070	-2.3049	-
58.04	-0.8318	-1.0136	-0.6006	-1.4623	-0.2768
59.01	-0.1621	-0.7309	-0.8453	0.0888	-0.5647
59.02	0.1215	-0.0394	-0.6529	-0.7163	-0.3496
59.03	-1.0660	-0.6810	-0.5661	-0.9969	-0.6585
59.04	-1.1178	-1.5424	-1.0406	-0.8184	0.2308
60.01	-0.9078	-1.3542	-1.7216	-1.1940	0.2001
60.02	-0.0415	-0.4826	-1.0236	-1.4752	-1.1135
60.03	-1.0557	-0.9519	-1.2886	-1.6988	-1.3356
60.04	-1.6579	-1.9499	-1.7102	-1.9269	-1.2509
61.01	-3.0500	-3.4503	-3.5545	-3.0574	-2.0987
61.02	-2.4494	-3.3145	-3.6992	-3.7670	-3.0522
61.03	-1.5073	-2.1589	-3.1013	-3.5108	-3.0344
61.04	-0.5886	-1.0080	-1.7293	-2.7538	-2.7488
62.01	-0.0238	-0.1942	-0.6676	-1.4533	-2.5886
62.02	0.3981	0.3744	0.1668	-0.3661	-3.0006
62.03	0.1264	0.2424	0.2210	-0.0026	-3.0570
62.04	-1.2763	-1.0963	-0.8944	-0.8431	-3.4240
63.01	-1.8985	-2.2543	-1.9576	-1.6653	-2.9185
63.02	-1.5525	-2.1554	-2.4872	-2.1761	-2.2537
63.03	-0.2502	-0.8525	-1.5101	-1.8747	-1.3751
63.04	0.3969	0.2401	-0.4329	-1.1003	-0.9530
64.01	0.6896	0.7948	0.5781	-0.1357	-0.7935
64.02	0.9716	1.1661	1.2154	0.9388	-0.7791
64.03	1.6153	1.8299	1.9545	1.9152	-0.5585
64.04	1.0617	1.4188	1.6490	1.7873	-0.2733
65.01	1.8532	2.0483	2.3979	2.5852	1.0665
65.02	1.3797	1.7851	1.9928	2.3620	1.9435
65.03	2.3894	2.6599	3.0484	3.2200	3.2393
65.04	2.3603	2.8922	3.1595	3.5466	3.9159
66.01	2.5607	3.0896	3.6277	3.8720	4.6526
66.02	2.3817	2.9778	3.5237	4.0661	5.0238
66.03	2.2171	2.8001	3.4335	3.9931	5.1704
66.04	1.5839	2.2018	2.8415	3.5181	5.1804
67.01	-1.5789	-0.9150	-0.0800	0.7362	3.2344
67.02	-0.5451	-1.1402	-0.4052	0.4580	3.4158
67.03	-0.5047	-0.7066	-1.3239	-0.5287	2.7265
67.04	0.4544	0.2537	0.0078	-0.6620	2.5274
68.01	1.6167	1.6684	1.4435	1.1773	2.5613
68.02	0.2516	0.9135	1.0190	0.8428	1.4060
68.03	0.6852	0.7289	1.3712	1.4368	0.9470
68.04	0.4210	0.6192	0.6681	1.3293	0.2713
69.01	1.6063	1.6606	1.7926	1.7423	1.6934
69.02	0.3751	0.7129	0.7986	0.9801	1.4135
69.03	0.4108	0.4787	0.8364	0.9210	1.7631
69.04	0.3736	0.4438	0.5115	0.8600	1.7502
70.01	-1.2635	-1.1376	-1.0360	-0.9323	-0.0573
70.02	-0.5683	-0.8777	-0.7516	-0.6496	0.2055
70.03	0.9227	0.7655	0.4169	0.5001	1.1474
70.04	-2.7727	-2.5033	-2.5670	-2.7944	-1.6355
71.01	-0.1711	-0.8527	-0.6282	-0.7284	-0.5813
71.02	1.0118	0.9240	0.1605	0.3545	0.2081
71.03	0.4975	0.7913	0.7025	-0.0633	-0.0538
71.04	1.8427	1.8979	2.1286	1.9701	1.1664
72.01	1.3878	1.9639	2.0065	2.2322	1.3999
72.02	-0.1190	0.3230	0.9883	1.1032	0.6790
72.03	-1.3400	-1.3280	-0.8496	-0.1343	-0.3332
72.04	-0.0745	-0.2912	-0.3096	0.0560	0.9355

[a]Scaled recursive residuals are not reported: the standard error of the transformed regression is $\hat{\sigma} = 1.015$ (based on the sample 1956/I–1972/IV).

Table 25.3: Gordon-Jorgenson model: test statistics (OLSQ), based on the 64 one-step-ahead recursive residuals.

Type	*Indicator*				*Result*	*p-values* [a]
Global location tests [b]	t-test				0.619	0.9506
	Number of positive residuals				32	1.0000
	Wilcoxon test				1053	0.9307
Runs tests [c]	Number of runs				16	0.000019
	Length of longest run				13	0.0065
Serial correlation tests [d]	Modified von Neumann ratio				0.6779	< 0.002
	Rank tests					
	k	Signed-rank tests			Sign tests	
		S_k	S'_k	*p-value*	S_k	S'_k	*p-value*
	1	1735	4.977	0.00000065	48	4.158	0.000038
	2	1421	3.116	0.0018	41	2.540	0.0151
	3	1284	2.431	0.0150	37	1.664	0.1237
	4	1091	1.296	0.1951	33	0.7746	0.5190
	5	1041	1.177	0.2390	33	0.9113	0.4350
	6	1095	1.854	0.0637	35	1.576	0.1480
	7	1058	1.839	0.0659	35	1.722	0.1112
	8	983	1.509	0.1313	32	1.069	0.3497
	9	1015	2.053	0.0401	33	1.483	0.1770
	10	958	1.856	0.0635	32	1.361	0.2203
	11	877	1.430	0.1528	31	1.236	0.2717
	12	807	1.075	0.2825	30	1.109	0.3317

[a] Marginal significance levels.

[b] See Dufour (1982, Section 4.3). The tests are two-sided.

[c] See Dufour (1982, Section 4.5). The tests are one-sided: $P[R \leq 16] = 0.000019$ and $P[L \geq 16] = 0.0065$, where R = number of runs (of + 's or -'s) and L = length of the longest run.

[d] S_k is a rank statistic for testing serial dependence [see Dufour (1982, Section 4.6)], k being the lag used; test; $S'_k = (S_k - E_0(S_k))/\, Std_0(S_k)$. We consider here two-sided tests (against positive or negative serial dependence). For a more complete theory of these tests, see Dufour (1981).

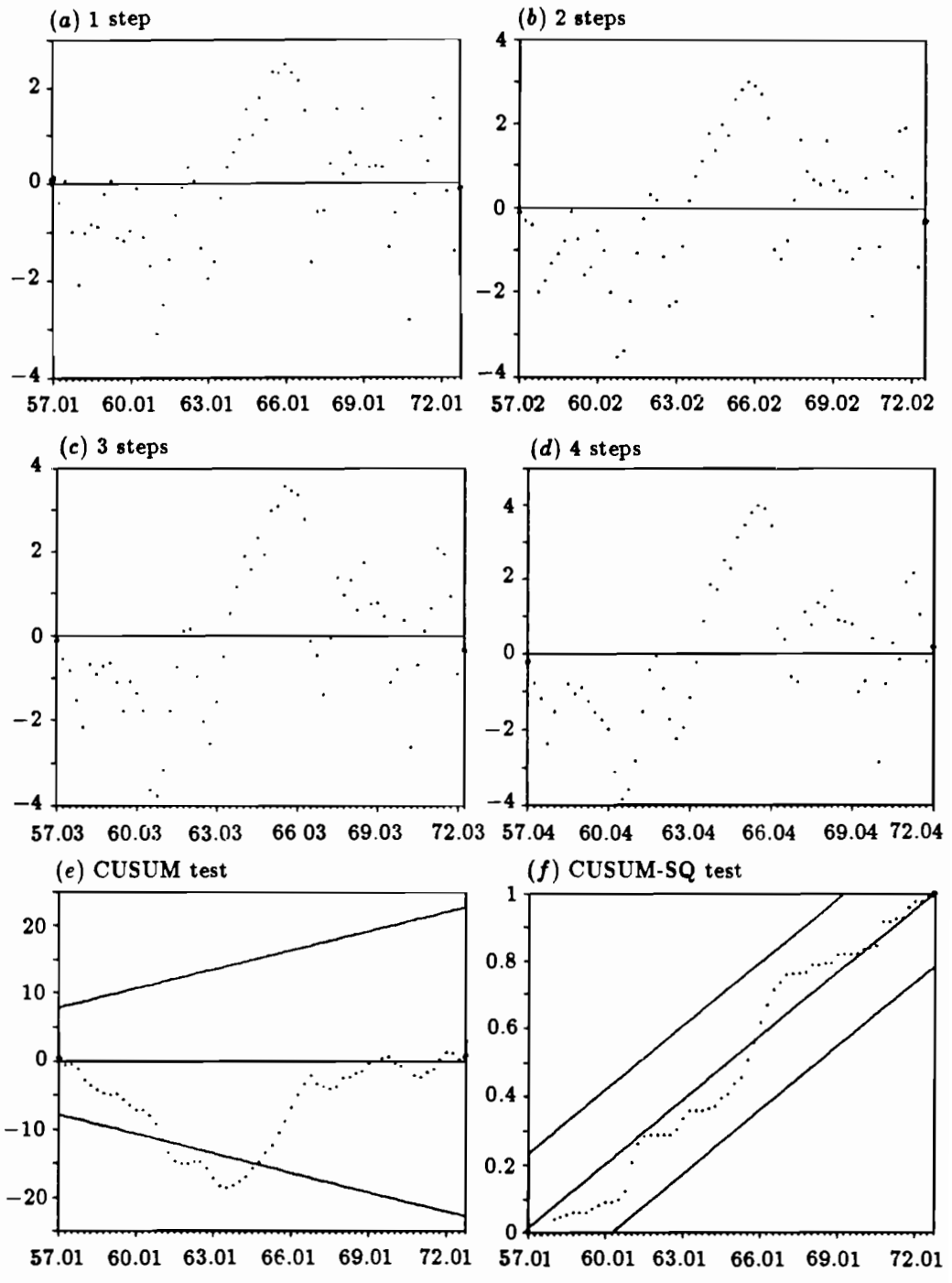

Figure 25.2: Gordon-Jorgenson model (OLSQ): recursive residuals and CUSUM tests. Significance boundaries in (*e*) and (*f*) correspond to tests of level 0.05.

Table 25.4: t-statistics for subperiods (OLSQ).

Period [a]	t	*p-value*
1962/I–1966/III	2.553	0.0200
1964/I–1966/III	8.834	0.00000251
1967/II–1969/I	1.724	0.128
1971/II–1972/IV	1.127	0.303
Remainder [b]	−3.790	0.000705

[a] 1962/I–1966/III corresponds to the first application of the tax credit; 1964/I–1966/III is the same period after the repeal of the Long Amendment; 1967/II–1969/I corresponds to the second application and 1971/II–1972/IV to the third one.

[b] 1957/I–1961/IV, 1966/IV–1967/I, 1969/II–1971/I.

and lasts as long as the effective tax credit is nonzero (up to 1966/IV). The suspension of the tax credit (1967/I) is associated with a discontinuity in the same series, while the following run of underpredictions (1967/IV–1969/IV) can be related to the reimposition of the tax credit (1967/II–1969/I).

On this issue, it is also instructive to compute t-statistics to test the null hypothesis of a zero mean (based on the one-step-ahead recursive residuals) for each of the subperiods corresponding to the different phases of the tax credit. This is justified by the fact that the (one-step-ahead) recursive residuals are i.i.d.$N(0, \sigma^2)$ under the null hypothesis [see Dufour (1982, Section 4.3)]. The results of these calculations are given in *Table 25.4*. From the latter, it is remarkable that each period where the effective tax credit is nonzero corresponds to a positive t-statistic (indicating a tendency to underpredict), while the period over which it does not apply yields a negative t-statistic. This effect is especially strong for the first application of the tax credit after the repeal of the Long Amendment.

Thus, if we estimate recursively equation (25.5) by ordinary least squares, we find several clues of instability. In particular, the results point to the presence of a substantial shift associated with the first imposition of the investment tax credit, especially after the repeal of the Long Amendment. Furthermore, this shift induced systematic underprediction of the level of investment expenditures over the corresponding period. On the other hand, the two other applications of the tax credit are not associated with statistically significant effects, even though the t-statistics are also positive.

Consider now the results of a similar experiment applied to equation (25.8), i.e., model (25.5) after correction for autocorrelation (using $\hat{\rho} = 0.6223$). The recursive estimates are listed in *Table 25.5* and graphed in *Figures 25.3(a–d)*; the recursive residuals are listed in *Table 25.6* with a number of test statistics in *Table 25.7*, and they are graphed in *Figures 25.4(a–d)* [8]. When we look at the recursive estimates, we can still observe the same basic phases: first (1957/I–1961/I), wide fluctuations with rough trends (upward for α and a_2, downward for δ and a_1); second (1961/II–1963/IV), a period of relative stability with no general trend (expect for δ which starts to increase near 1961/IV); third (1963/IV–1966/IV), well-defined trends for all coefficients (downward for α and a_2, upward

Table 25.5: Gordon-Jorgenson model: forward recursive estimates (data transformed with $\hat{\rho} = 0.6223$): 1956/I–1972/IV.

Quarter	α	δ	a_1	a_2
56.04	-363.656	-1.9131	0.0519153	-0.0071130
57.01	-271.063	1.2423	-0.0022213	0.0000800
57.02	-306.493	1.1325	0.0015707	-0.0004430
57.03	-313.996	1.4091	-0.0048950	0.0003493
57.04	255.175	-0.7712	0.0072497	-0.0006241
58.01	-391.011	1.1731	0.0056639	-0.0010445
58.02	-312.995	0.9128	0.0062321	-0.0010437
58.03	-252.689	0.7487	0.0023925	-0.0005566
58.04	-259.507	0.8018	-0.0006244	-0.0002177
59.01	-277.221	0.8733	-0.0013764	-0.0001448
59.02	-273.910	0.8607	-0.0013323	-0.0001475
59.03	-103.182	0.2313	-0.0007210	-0.0000952
59.04	-50.171	0.0343	-0.0014466	0.0000221
60.01	-38.357	-0.0089	-0.0016395	0.0000519
60.02	-53.803	0.0468	-0.0015769	0.0000337
60.03	-23.015	-0.0706	-0.0018095	0.0000797
60.04	-0.459	-0.1558	-0.0021968	0.0001378
61.01	33.432	-0.2782	-0.0028205	0.0002315
61.02	41.104	-0.3030	-0.0031011	0.0002691
61.03	42.370	-0.3038	-0.0030304	0.0002635
61.04	40.573	-0.3014	-0.0030468	0.0002626
62.01	38.703	-0.2985	-0.0030512	0.0002605
62.02	36.644	-0.2961	-0.0030638	0.0002588
62.03	35.078	-0.2750	-0.0023387	0.0001821
62.04	36.499	-0.2411	-0.0019230	0.0001498
63.01	40.351	-0.2223	-0.0024909	0.0002265
63.02	41.286	-0.2195	-0.0026724	0.0002494
63.03	37.019	-0.2210	-0.0018309	0.0001472
63.04	33.500	-0.2135	-0.0012767	0.0000814
64.01	30.380	-0.2017	-0.0009162	0.0000394
64.02	26.554	-0.1819	-0.0006697	0.0000114
64.03	23.462	-0.1666	-0.0008041	0.0000256
64.04	23.063	-0.1643	-0.0007949	0.0000246
65.01	15.427	-0.1174	-0.0003890	-0.0000193
65.02	13.657	-0.1057	-0.0002985	-0.0000288
65.03	4.147	-0.0414	0.0001009	-0.0000698
65.04	-1.534	-0.0006	0.0003055	-0.0000897
66.01	-8.201	0.0494	0.0004660	-0.0001035
66.02	-12.363	0.0820	0.0005219	-0.0001067
66.03	-16.481	0.1144	0.0006648	-0.0001197
66.04	-17.838	0.1252	0.0007345	-0.0001264
67.01	-5.792	0.0266	0.0000840	-0.0000639
67.02	-8.587	0.0465	0.0008241	-0.0001450
67.03	-7.345	0.0350	0.0008335	-0.0001475
67.04	-7.942	0.0409	0.0007639	-0.0001390
68.01	-11.173	0.0732	0.0003778	-0.0000914
68.02	-8.527	0.0467	0.0006748	-0.0001282
68.03	-9.310	0.0544	0.0006067	-0.0001196
68.04	-9.254	0.0539	0.0006092	-0.0001199
69.01	-10.988	0.0698	0.0006009	-0.0001170
69.02	-10.243	0.0626	0.0006111	-0.0001191
69.03	-10.426	0.0645	0.0006022	-0.0001178
69.04	-10.709	0.0673	0.0006918	-0.0001274
70.01	-9.662	0.0556	0.0004967	-0.0001075
70.02	-9.905	0.0586	0.0005126	-0.0001088
70.03	-11.118	0.0743	0.0003393	-0.0000866
70.04	-7.688	0.0284	0.0011028	-0.0001802
71.01	-8.888	0.0446	0.0007806	-0.0001413
71.02	-9.601	0.0541	0.0005443	-0.0001132
71.03	-9.385	0.0514	0.0006417	-0.0001245
71.04	-10.134	0.0596	0.0002408	-0.0000785
72.01	-10.252	0.0607	0.0001823	-0.0000719
72.02	-9.901	0.0584	0.0003461	-0.0000903
72.03	-9.456	0.0571	0.0004711	-0.0001039
72.04	-9.656	0.0572	0.0004686	-0.0001039

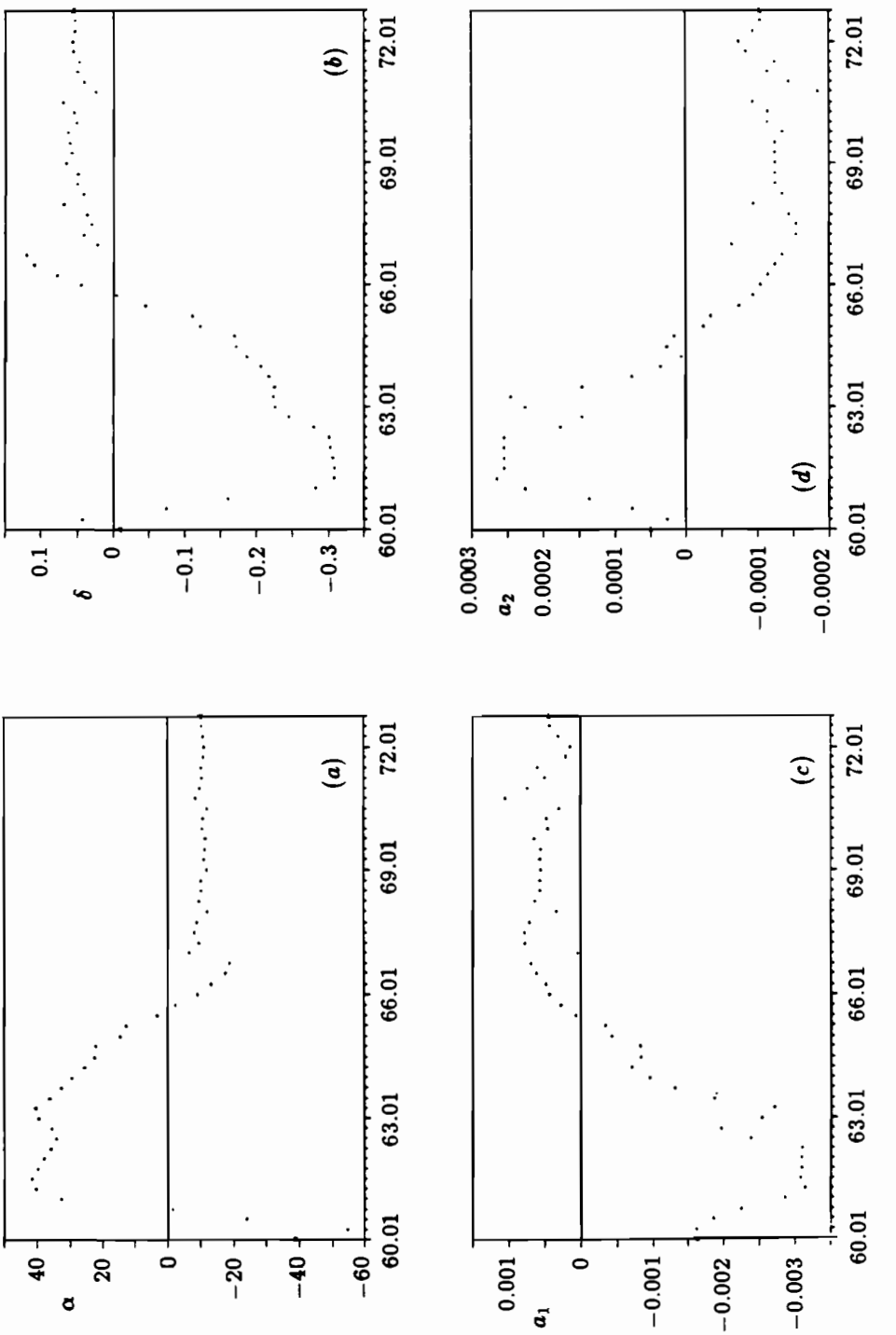

Figure 25.3: Gordon-Jorgenson model ($\hat{\rho} = 0.6223$): recursive estimates.

Table 25.6: Gordon-Jorgenson model: forward recursive residuals (data transformed with $\hat{\rho} = 0.6223$): 1956/I–1972/IV [a]

Quarter	*1 step*	*2 steps*	*3 steps*	*4 steps*	*8 steps*
57.01	0.8617	–	–	–	–
57.02	-0.2876	0.7745	–	–	–
57.03	0.5832	0.0882	0.8594	–	–
57.04	-1.1129	-0.4005	-0.4743	0.7274	–
58.01	-1.9398	-2.1876	-0.9812	-0.7629	–
58.02	0.1589	-1.0499	-0.4120	-0.1960	–
58.03	0.5489	0.5060	-0.7196	-0.6797	–
58.04	0.5562	0.7814	0.7972	0.0316	0.7833
59.01	0.2795	0.5693	0.7879	0.7170	-0.1665
59.02	-0.0302	0.1087	0.4062	0.6159	0.0048
59.03	-1.4740	-1.1866	-0.9079	-0.5339	-0.4487
59.04	-0.6864	-1.4794	-1.2503	-1.0678	0.4201
60.01	-0.2317	-0.5886	-1.4267	-1.2044	-0.6091
60.02	0.4567	0.2821	-0.1955	-1.1910	-0.7522
60.03	-1.1254	-0.8059	-0.8034	-1.0566	-1.1056
60.04	-1.0778	-1.4389	-1.0643	-1.0100	-1.3586
61.01	-2.1211	-2.3610	-2.6153	-2.0790	-1.9059
61.02	-0.6305	-1.1889	-1.4830	-1.8146	-2.0349
61.03	-0.1769	-0.3097	-0.9042	-1.2230	-1.2533
61.04	0.2439	0.1839	0.0216	-0.6485	-0.9146
62.01	0.2705	0.3254	0.2559	0.0762	-0.9703
62.02	0.3446	0.4043	0.4581	0.3720	-1.4314
62.03	-0.5286	-0.4090	-0.3113	-0.2141	-1.3980
62.04	-1.4905	-1.5515	-1.3492	-1.1775	-1.8167
63.01	-1.0779	-1.5502	-1.6256	-1.3894	-1.1316
63.02	-0.2264	-0.6454	-1.1347	-1.2500	-0.6952
63.03	0.8557	0.6718	0.1440	-0.3490	-0.0831
63.04	0.6493	0.9670	0.7573	0.2130	-0.0957
64.01	0.5692	0.7803	1.0892	0.8612	-0.1237
64.02	0.6716	0.8141	1.0016	1.2811	-0.1755
64.03	1.0488	1.1659	1.2715	1.4139	0.0405
64.04	0.0779	0.3026	0.4656	0.6190	0.2725
65.01	1.2389	1.2223	1.4180	1.5533	1.6100
65.02	0.2578	0.5865	0.5869	0.8015	1.5434
65.03	1.5437	1.5541	1.8458	1.7981	2.2194
65.04	0.8328	1.2251	1.2443	1.5695	1.9277
66.01	0.9921	1.1928	1.5967	1.5917	2.1479
66.02	0.6097	0.8805	1.1017	1.5327	1.9872
66.03	0.6486	0.8110	1.0981	1.3219	1.9889
66.04	0.2372	0.4550	0.6430	0.9605	1.9735
67.01	-2.4115	-2.1238	-1.6915	-1.3304	0.2121
67.02	1.0585	-0.0779	0.0356	0.3018	1.5654
67.03	-0.2880	0.2444	-0.8673	-0.6599	0.6505
67.04	0.1035	0.0075	0.1772	-0.6717	0.4132
68.01	0.8737	0.8043	0.6673	0.8197	0.5047
68.02	-1.0001	-0.5752	-0.4352	-0.5072	-0.6355
68.03	0.4017	0.0870	0.4180	0.4110	-0.3358
68.04	-0.0358	0.0552	-0.2357	0.1098	-0.7244
69.01	1.3803	1.3494	1.4046	1.0719	1.2657
69.02	-0.6068	-0.3531	-0.3529	-0.2463	-0.1877
69.03	0.1533	0.0395	0.2996	0.2842	0.3840
69.04	0.3481	0.3692	0.2434	0.5176	0.5684
70.01	-1.3258	-1.1676	-1.1161	-1.2200	-1.0088
70.02	0.3041	-0.0354	0.0668	0.0978	0.3577
70.03	1.2032	1.2403	0.9179	0.9743	1.0602
70.04	-3.4415	-3.0240	-2.8938	-3.1034	-2.4678
71.01	1.4386	0.4556	0.7541	0.7997	0.4413
71.02	0.9690	1.3068	0.2586	0.5726	0.4643
71.03	-0.3259	-0.0517	0.3485	-0.6952	-0.4223
71.04	1.4282	1.2599	1.4814	1.8241	0.9441
72.01	0.2908	0.6649	0.5394	0.7795	0.4317
72.02	-1.0287	-0.9480	-0.5379	-0.6084	-0.5938
72.03	-1.4314	-1.5655	-1.4939	-1.1499	-1.4101
72.04	0.6363	0.5052	0.4022	0.4314	0.9185

[a] Scaled recursive residuals are not reported: the standard error of the transformed regression is $\hat{\sigma} = 1.015$ (based on the sample 1956/I–1972/IV).

Table 25.7: Gordon-Jorgenson model ($\hat{\rho} = 0.6223$): test statistics. [a]

Type	*Indicator*	*Result*	*p-values*
Global location tests	t-test	−0.1203	0.9042
	Number of positive residuals	38	0.1686
	Wilcoxon test	1126	0.5652
Runs tests	Number of runs	29	0.2250
	Length of longest run	14	0.0032
Serial correlation tests	Modified von Neumann ratio	1.967	> 0.10
	Rank tests		

k	Signed-rank tests			Sign tests		
	S_k	S'_k	*p-value*	S_k	S'_k	*p-value*
1	1161	1.047	0.2949	35	0.8819	0.4500
2	1103	0.8869	0.3751	36	1.270	0.2529
3	1114	1.210	0.2262	36	1.408	0.2000
4	789	−0.9276	0.3536	26	−1.033	0.3663
5	897	0.0906	0.9278	32	0.6509	0.6029
6	1126	2.094	0.0362	36	1.838	0.0869
7	1092	2.109	0.0349	37	2.252	0.0331
8	787	−0.0897	0.9285	30	0.5345	0.6889
9	870	0.8379	0.4021	28	0.1348	1.0000
10	710	−0.2798	0.7796	26	−0.2720	0.8919
11	578	−1.217	0.2235	24	−0.6868	0.5831
12	696	0.0638	0.9492	25	−0.2774	0.8899

[a]Number of residuals: 64

Table 25.8: t-statistics for subperiods ($\hat{\rho} = 0.6223$).

Period	t	*p-value*
1962/I–1966/III	2.178	0.0429
1964/I–1966/III	6.066	0.0000812
1967/II–1969/I	1.130	0.256
1971/II–1972/IV	0.194	0.853
Remainder [a]	−1.839	0.0762

[a]1957/I–1961/IV, 1966/IV–1967/I, 1969/II–1971/I.

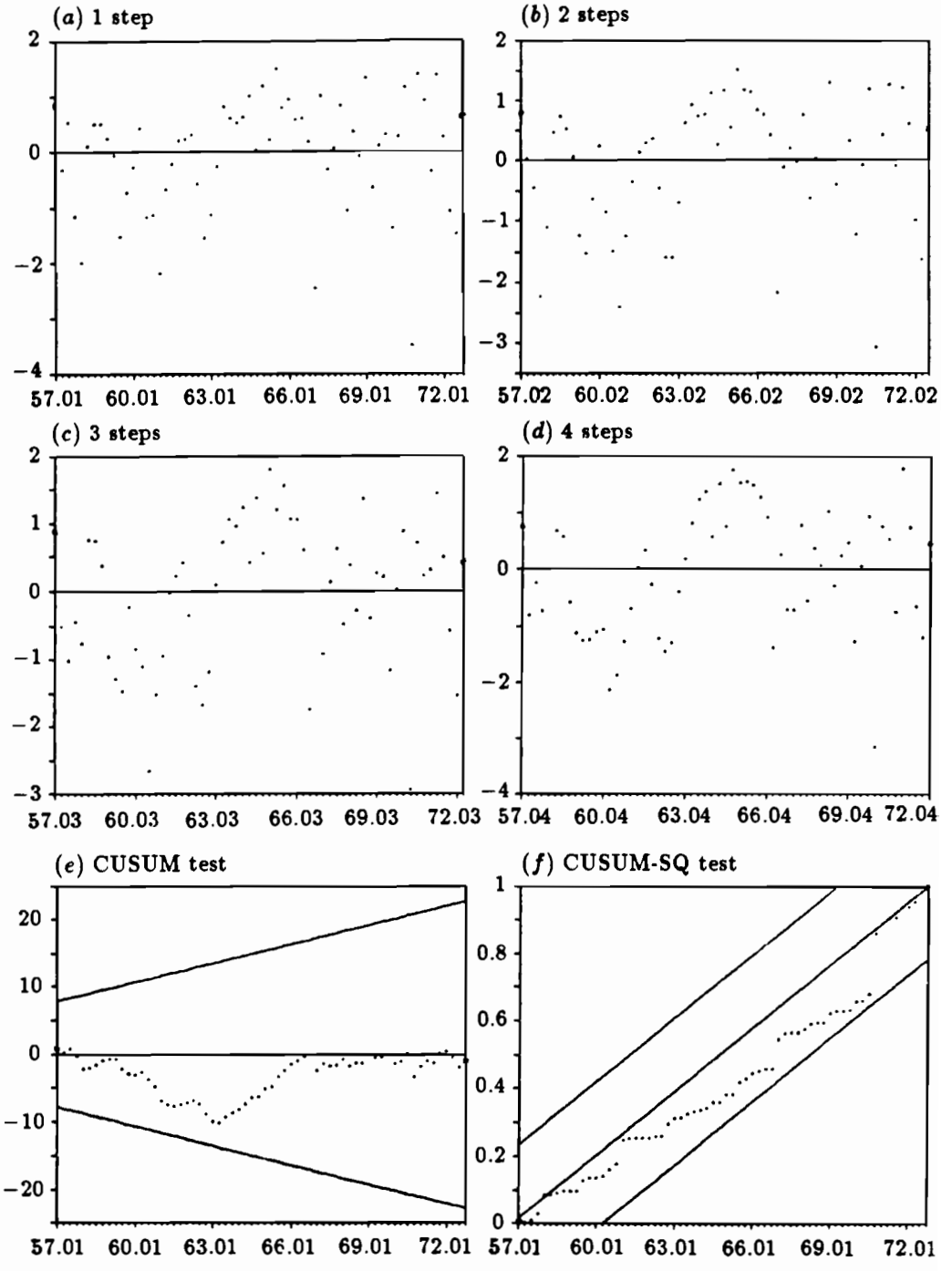

Figure 25.4: Gordon-Jorgenson model ($\hat{\rho} = 0.6223$): recursive residuals and CUSUM tests.

for δ and a_1) during which all coefficients change sign; fourth (1967/I–1972/IV), a period during which all coefficients seem to stabilize. On the other hand, the one-step recursive residuals [*Figure 25.4(a)*] appear more "random" than without the tranformation [compare *Figures 25.2(a)* and *25.4(a)*]. Global location tests and serial dependence tests are not significant at standard levels (say, 0.10). Nevertheless, we can still observe a tendency to overpredict in the earlier period (up to 1963/II) as well as a run of 14 consecutive underpredictions from 1963/III to 1966/IV followed by a sudden drop (1967/I) [9]. The (1967/IV–1969/IV) run of underpredictions disappears. These observations are confirmed when we look at several-steps-ahead recursive residuals [*Figures 25.4(b–d)*]. We thus continue to find signs of instability, especially in association with the first application of the tax credit (after the repeal of the Long Amendment).

The t-statistics for the separate subperiods corresponding to the different applications of the tax credit are reported in *Table 25.8*. As in the first experiment, the t-statistic for periods where the tax credit was in force are positive, while for the rest of the sample the t-statistic is negative. Moreover, the t-statistic for the first application period is significant (at level 0.04) and very strongly significant (at level 0.00008) if the period where the Long Amendment applied is excluded.

Finally, to take into account the fact that K_t is a form of lagged dependent variable, let us consider the result of estimating recursively equation (25.9). The recursive estimates are listed in *Table 25.9* and graphed in *Figures 25.5(a–c)*; the recursive residuals are listed in *Table 25.10*, with a number of test statistics in *Table 25.11*, and they are graphed in *Figures 25.6(a–d)*. From the recursive estimates, we still observe the same four phases: first (1956/IV–1961/I), wide fluctuations with rough trends (upwards for α and a_2, downward for a_1); second (1961/II–1963/II), a period of relative stability; third (1963/III–1966/IV), a clear trend (downward for α and a_2, upward for a_1); fourth (1967/I–1972/IV), a period where all coefficients seem to stabilize. On the basis of the one-step recursive residuals [*Figure 25.6(a)*], we find now that none of the test statistics in *Table 25.11* nor the CUSUM and CUSUM of squares tests [*Figures 25.6(e)* and (f)] are significant (at level 0.05). In particular, the longest-run test is not conclusive. [Two residuals in the middle of the longest run previously observed (1963/III–1966/IV) are now below the zero line.] Nevertheless, several-steps-ahead recursive residuals [*Figures 25.6(b–d)*] do not seem to be affected in the same way and exhibit basically the same pattern as in the previous experiment; in particular, two- and three-steps-ahead recursive residuals contain continuous runs of underpredictions covering the period 1963/III–1966/IV. Indeed, the similarity between *Figures 25.4(a)* and *25.6(a)* (showing one-step-ahead recursive residuals) is striking: we still note a tendency to overpredict up to 1963/II and a tendency to underpredict over the period 1963/III–1966/IV, while the rest looks relatively "random". If we compute t-statistics over the seperate subperiods corresponding to the seperate phases of the tax credit, we find results analogous to the ones obtained before (see *Table 25.12*). The t-statistic attached to 1962/I–1966/III (first application of the tax credit) is positive and significant at level 0.04 while, for the period 1964/I–1966/III (after the repeal of the Long Amendment), it is significant at level 0.00065. Note again the contrast between the application periods of the tax credit (which yield positive t-statistics) and the remainder of

Table 25.9: Gordon-Jorgenson model: forward recursive estimates (data transformed with $\hat{\rho} = 0.6223$, capital subtracted): 1956/I–1972/IV.

Quarter	α	a_1	a_2
56.03	-269.178	0.0255443	-0.0033785
56.04	-533.474	0.0279611	-0.0041437
57.01	44.434	0.0029566	-0.0002796
57.02	-9.307	0.0057092	-0.0006864
57.03	68.847	-0.0004134	0.0001406
57.04	4.559	0.0061358	-0.0007076
58.01	-63.294	0.0073405	-0.0009702
58.02	-81.517	0.0053041	-0.0007791
58.03	-75.324	0.0038777	-0.0006091
58.04	-67.076	0.0003639	-0.0002038
59.01	-65.305	-0.0011968	-0.0000274
59.02	-66.868	-0.0023877	0.0001015
59.03	-60.603	-0.0013351	-0.0000030
59.04	-55.332	-0.0013686	0.0000111
60.01	-51.703	-0.0014485	0.0000271
60.02	-55.690	-0.0015420	0.0000297
60.03	-43.791	-0.0013385	0.0000303
60.04	-31.604	-0.0014409	0.0000654
61.01	-11.526	-0.0016787	0.0001307
61.02	-4.316	-0.0020039	0.0001809
61.03	-2.846	-0.0019123	0.0001735
61.04	-3.729	-0.0019261	0.0001733
62.01	-4.591	-0.0019339	0.0001726
62.02	-5.986	-0.0019498	0.0001717
62.03	-5.767	-0.0018010	0.0001555
62.04	-0.979	-0.0014920	0.0001302
63.01	4.346	-0.0020012	0.0001969
63.02	5.020	-0.0020984	0.0002089
63.03	0.472	-0.0012382	0.0001047
63.04	-1.676	-0.0007826	0.0000500
64.01	-2.772	-0.0005481	0.0000218
64.02	-3.606	-0.0004066	0.0000045
64.03	-4.652	-0.0005419	0.0000177
64.04	-4.448	-0.0005549	0.0000195
65.01	-5.903	-0.0002692	-0.0000150
65.02	-5.859	-0.0002796	-0.0000137
65.03	-7.456	0.0000680	-0.0000553
65.04	-8.199	0.0002511	-0.0000771
66.01	-9.084	0.0004542	-0.0001013
66.02	-9.629	0.0005739	-0.0001156
66.03	-10.370	0.0008547	-0.0001482
66.04	-10.814	0.0010835	-0.0001745
67.01	-8.858	-0.0001366	-0.0000350
67.02	-9.637	0.0007289	-0.0001329
67.03	-9.487	0.0005972	-0.0001179
67.04	-9.458	0.0006046	-0.0001187
68.01	-9.734	0.0005181	-0.0001096
68.02	-9.439	0.0005908	-0.0001171
68.03	-9.544	0.0005848	-0.0001167
68.04	-9.523	0.0005818	-0.0001163
69.01	-9.996	0.0007138	-0.0001318
69.02	-9.832	0.0006639	-0.0001260
69.03	-9.892	0.0006779	-0.0001276
69.04	-9.988	0.0008207	-0.0001437
70.01	-9.773	0.0004730	-0.0001046
70.02	-9.809	0.0005361	-0.0001117
70.03	-9.967	0.0006536	-0.0001251
70.04	-9.564	0.0005583	-0.0001137
71.01	-9.678	0.0005456	-0.0001125
71.02	-9.787	0.0004912	-0.0001066
71.03	-9.701	0.0005681	-0.0001150
71.04	-10.012	0.0002613	-0.0000814
72.01	-10.088	0.0002035	-0.0000750
72.02	-9.845	0.0003521	-0.0000912
72.03	-9.461	0.0004706	-0.0001038
72.04	-9.656	0.0004686	-0.0001039

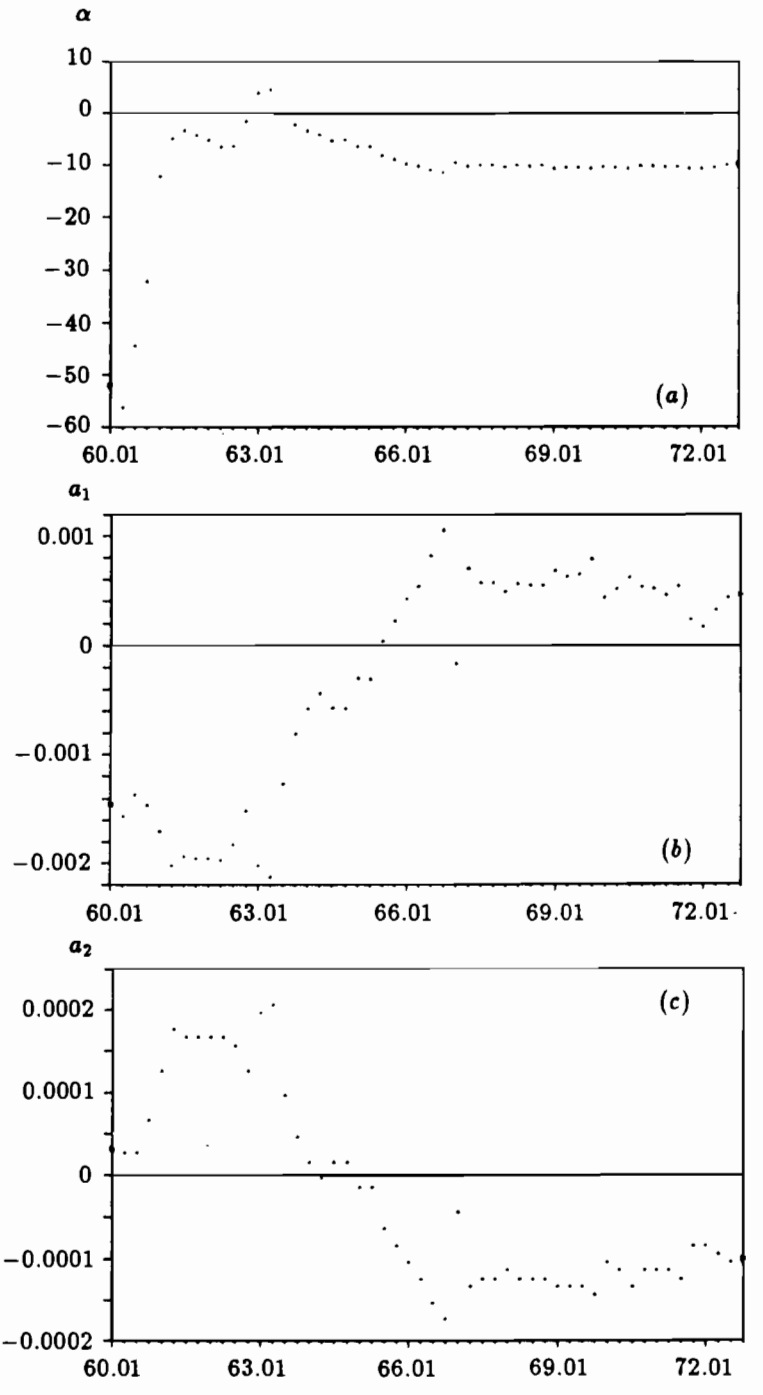

Figure 25.5: Gordon-Jorgenson model ($\hat{\rho} = 0.6223$, capital subtracted): recursive estimates.

Table 25.10: Gordon-Jorgenson model: forward recursive residuals (data transformed with $\hat{\rho} = 0.6223$, capital subtracted): 1956/I–1972/IV

Quarter	*1 step*	*2 steps*	*3 steps*	*4 steps*	*8 steps*
56.04	-0.4199	-	-	-	-
57.01	0.8541	0.1812	-	-	-
57.02	-0.2369	0.7132	0.1499	-	-
57.03	0.6522	0.1718	0.8587	0.3168	-
57.04	-1.0829	-0.2235	-0.3217	0.6768	-
58.01	-1.9935	-2.2629	-0.9657	-0.7152	-
58.02	-0.8332	-1.6755	-1.9926	-0.7974	-
58.03	0.1999	0.0907	-0.6741	-1.2200	0.2449
58.04	0.6683	0.6189	0.6749	0.1889	0.7389
59.01	0.6272	0.9089	0.8025	0.8915	0.0256
59.02	0.7184	0.9022	1.1223	0.9815	0.5414
59.03	-0.7690	-0.4986	-0.2149	0.2369	0.0179
59.04	-0.5613	-0.6971	-0.5808	-0.4781	0.1685
60.01	-0.3321	-0.4330	-0.5716	-0.4750	-0.0714
60.02	0.3859	0.3018	0.1742	-0.0535	0.4872
60.03	-1.2749	-1.1330	-1.1798	-1.2770	-0.8328
60.04	-1.4013	-1.6667	-1.5027	-1.5372	-1.4847
61.01	-2.5474	-2.8018	-3.0353	-2.8012	-2.7459
61.02	-1.0226	-1.5398	-1.8255	-2.0898	-2.1254
61.03	-0.2222	-0.4303	-1.0725	-1.4242	-1.6397
61.04	0.1430	0.0745	-0.1657	-0.8660	-1.3921
62.01	0.1549	0.1869	0.1104	-0.1498	-1.3990
62.02	0.2718	0.3031	0.3316	0.2379	-1.6604
62.03	-0.1166	-0.0304	0.0190	0.0664	-1.5225
62.04	-1.2664	-1.1632	-0.9974	-0.8782	-1.8958
63.01	-0.9177	-1.3088	-1.2424	-1.0481	-1.0927
63.02	-0.1178	-0.4758	-0.8837	-0.8700	-0.5227
63.03	0.8714	0.7339	0.2743	-0.1395	0.0708
63.04	0.5483	0.8849	0.7324	0.2609	0.0214
64.01	0.3913	0.5752	0.9175	0.7550	-0.0325
64.02	0.4291	0.5246	0.6941	1.0191	-0.1177
64.03	0.9357	1.0010	1.0654	1.1851	0.0794
64.04	-0.1538	0.0407	0.1380	0.2461	0.1134
65.01	0.9715	0.9208	1.0886	1.1629	1.3162
65.02	-0.0314	0.1821	0.1455	0.3315	1.0884
65.03	1.3211	1.2844	1.4728	1.4017	1.7198
65.04	0.6624	0.9118	0.8809	1.0904	1.3294
66.01	0.9107	1.0193	1.2699	1.2271	1.5876
66.02	0.6385	0.7855	0.8994	1.1596	1.4354
66.03	0.7928	0.8899	1.0479	1.1668	1.5086
66.04	0.4721	0.6282	0.7343	0.9060	1.4917
67.01	-2.3454	-2.1507	-1.9133	-1.7587	-0.9133
67.02	1.0167	0.1404	0.2867	0.4837	1.1923
67.03	-0.3686	0.1141	-0.6157	-0.4467	0.2265
67.04	-0.0860	-0.0986	-0.0645	-0.2878	0.1369
68.01	0.8886	0.8772	0.8678	0.8781	0.9148
68.02	-0.9809	-0.9008	-0.9049	-0.9173	-0.9136
68.03	0.3398	0.2651	0.3357	0.3271	0.1463
68.04	-0.0626	-0.0383	-0.1070	-0.0417	-0.3491
69.01	1.4165	1.4069	1.4279	1.3582	1.5590
69.02	-0.5310	-0.4060	-0.4097	-0.3839	-0.4400
69.03	0.2155	0.1735	0.2908	0.2849	0.2953
69.04	0.4084	0.4258	0.3655	0.5220	0.5205
70.01	-1.2921	-1.1593	-1.1369	-1.1864	-1.0270
70.02	0.3034	0.0739	0.1626	0.1793	0.2792
70.03	1.3505	1.3714	1.2230	1.2684	1.3395
70.04	-3.5859	-3.5319	-3.5113	-3.5892	-3.4442
71.01	1.0081	0.9063	0.9443	0.9528	0.9026
71.02	0.7913	0.8227	0.7326	0.7538	0.7647
71.03	-0.4334	-0.3792	-0.3406	-0.4102	-0.3727
71.04	1.3976	1.3088	1.3699	1.4116	1.4399
72.01	0.3310	0.5957	0.5165	0.5822	0.4998
72.02	-1.0090	-0.9369	-0.6372	-0.7017	-0.6821
72.03	-1.4293	-1.5536	-1.4869	-1.2252	-1.2691
72.04	0.6363	0.5065	0.4086	0.4396	0.6492

Table 25.11: Gordon-Jorgenson model ($\hat{\rho} = 0.6223$, capital subtracted): test statistics. [a]

Type	*Indicator*				*Result*	*p-values*
Global location tests	t-test				−0.4535	0.6502
	Number of positive residuals				35	0.6201
	Wilcoxon test				1112	0.7962
Runs tests	Number of runs				34	0.6460
	Length of longest run				7	0.3892
Serial correlation tests	Modified von Neumann ratio				1.974	> 0.10
	Rank tests					
	k	Signed-rank tests			Sign tests	
		S_k	S'_k	*p-value*	S_k	S'_k	*p-value*
	1	1101	0.4079	0.6833	31	−0.2500	0.9007
	2	1117	0.7462	0.4556	35	0.8819	0.4500
	3	1106	0.9079	0.3639	30	−0.2540	0.8991
	4	818	−0.9158	0.3598	25	−1.408	0.2000
	5	866	−0.3607	0.7183	25	−1.291	0.2451
	6	1100	1.623	0.1046	34	1.172	0.2976
	7	1092	1.831	0.0671	36	1.838	0.0868
	8	738	−0.7032	0.4820	25	−0.9272	0.4270
	9	973	1.427	0.1534	33	1.336	0.2288
	10	748	−0.1843	0.8538	26	−0.4045	0.7877
	11	639	−0.8912	0.3728	25	−0.5443	0.6835
	12	806	0.8012	0.4230	28	−0.4121	0.7838

[a]Number of residuals: 64.

Table 25.12: t-statistics for subperiods ($\hat{\rho} = 0.6223$, capital subtracted).

Period	t	*p-value*
1962/I–1966/III	2.197	0.0414
1964/I–1966/III	4.697	0.000653
1967/II–1969/I	0.957	0.370
1971/II–1972/IV	0.105	0.920
Remainder [a]	−1.944	0.0613

[a]1956/IV–1961/IV, 1966/IV–1967/I, 1969/II–1971/I.

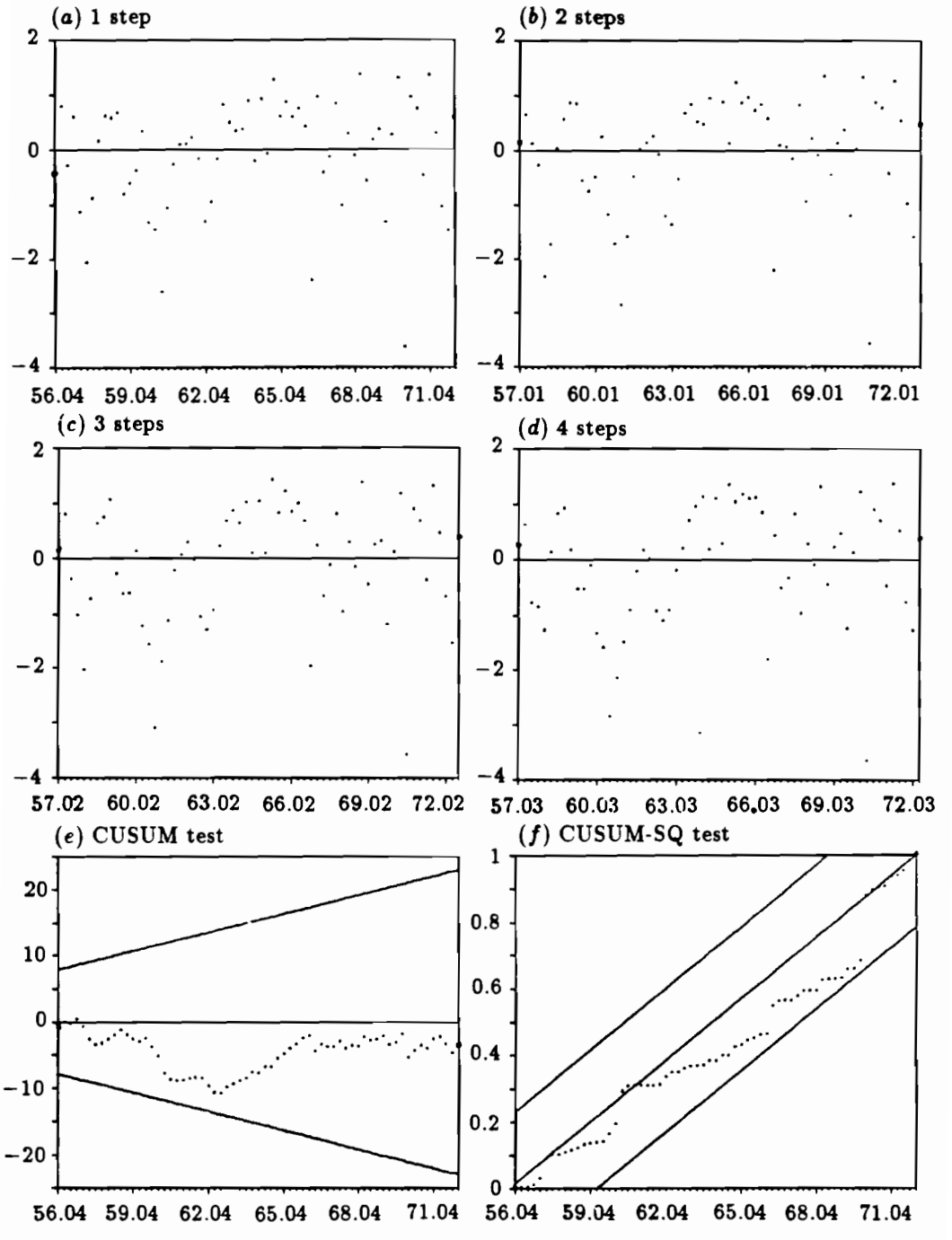

Figure 25.6: Gordon-Jorgenson model ($\hat{\rho} = 0.6223$, capital subtracted): recursive residuals and CUSUM tests.

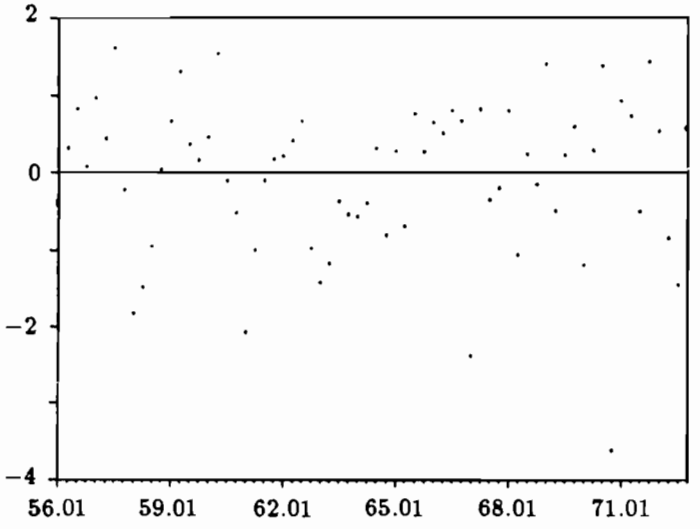

Figure 25.7: Gordon-Jorgenson model ($\hat{\rho} = 0.6223$): generalized least squares residuals.

the sample (which yields a negative t-statistic).

Although the evidence is less strong than for the two previous experiments, we continue to observe a phenomenon of underprediction associated with the first imposition of the tax credit, and this especially after the repeal of the Long Amendment. For the two other applications of the tax credit, we do not observe significant effects, although the corresponding t-statistics are positive and thus indicate a tendency to underpredict.

25.5 Conclusion

The results obtained in this recursive stability analysis are not as clear and definite as those we got, for example, for the demand for money during the German hyperinflation [Dufour (1986)]. They are confused, in particular, by the presence of a regressor (the capital stock) which contains lagged values of the dependent variable. Nevertheless, one feature remains constant throughout the three experiments performed: there appears to be a discontinuity associated with the introduction of the first investment tax credit (1962/I–1966/III), especially after the repeal of the Long Amendement (1964/I). Furthermore, the discontinuity is a type that leads to underprediction of investment, a behavior in contrast with the performance of the model before 1962 (where we find a tendency to overpredict). This phenomenon of underprediction is in agreement with Lucas's forecast. There is also some indication of a tendency to overpredict investment over the two other

Table 25.13: Effective investment tax credit (1961–1972).

Quarter	*Tax credit*	*Y*
61.01	0.0%	0
61.02	0.0%	0
61.03	0.0%	0
61.04	0.0%	0
62.01	3.1%	1
62.02	3.5%	1
62.03	3.9%	1
62.04	4.3%	1
63.01	4.7%	1
63.02	5.1%	1
63.03	5.5%	1
63.04	5.6%	1
64.01	5.6%	0
64.02	5.6%	0
64.03	5.6%	0
64.04	5.6%	0
65.01	5.6%	0
65.02	5.6%	0
65.03	5.6%	0
65.04	5.6%	0
66.01	5.6%	0
66.02	5.6%	0
66.03	5.6%	0
66.04	0.0%	0
67.01	0.0%	0
67.02	5.6%	0
67.03	5.6%	0
67.04	5.6%	0
68.01	5.6%	0
68.02	5.6%	0
68.03	5.6%	0
68.04	5.6%	0
69.01	5.6%	0
69.02	0.0%	0
69.03	0.0%	0
69.04	0.0%	0
70.01	0.0%	0
70.02	0.0%	0
70.03	0.0%	0
70.04	0.0%	0
71.01	0.0%	0
71.02	4.0%	0
71.03	5.0%	0
71.04	5.6%	0
72.01	5.6%	0
72.02	5.6%	0
72.03	5.6%	0
72.04	5.6%	0

periods where the tax credit was in force (1967/II–1969/I and 1971/II–1972/IV). This is suggested by the signs of the corresponding t-statistics, but the effects appear too small to be considered significant.

On the whole, the evidence we found is quite consistent with the type of instabilility suggested by Lucas (1976), even though it appears difficult to qualify this evidence as being very "strong". Of course, one could try to explain the instability detected by a misspecification other than the one pointed out by Lucas (e.g., the Almon lag scheme used may be wrong). In any event, whatever the "true" problem may be, it is certainly useful to know about its existence.

Acknowledgments

The author is Professor of Economics at the Unversité de Montréal and Research Fellow at C.R.D.E. (Centre de recherche et développement en économique).

The author wishes to thank Arnold Zellner, Nicholas Kiefer, Robert E. Lucas, Jr., Houston Stokes, and the member of the University of Chicago Econometrics and Statistics Colloquium for several helpful comments. I am also grateful to Denis Chénard and Sophie Mahseredjian for their superb research assistence. Special thanks go to Roger Gordon for supplying the data. This work was supported by grants from the Social Sciences and Humanities Research Council of Canada, the Natural Sciences and Engineering Research Council of Canada, and Fonds FCAR (Governement of Québec). It is a substantially revised version of Chapter 3 of the autor's Ph.D. dissertation at the University of Chicago (1979).

References

Brown, R.L., Durbin, J., and Evans, J.M. (1975), Techniques for testing the constancy of regression relationships over time (with discussion). *Journal of the Royal Statistical Society*, **B-37**, 149–192.

Christensen, L.R. and Jorgenson, D.W. (1969), The measurement of U.S. real capital input, 1929–1967. *Review of Income and Wealth*, **15**, 293–320.

Dufour, J.-M. (1979), *Methods for Specification Errors Analysis with Macroeconomic Applications*, Ph.D. dissertation, University of Chicago. .

Dufour, J.-M. (1981), Rank tests for serial dependence. *Journal of Time Series Analysis*, **2**, 117–128.

Dufour, J.-M. (1982), Recursive stability analysis of linear regression relationships: An exploratory methodology. *Journal of Econometrics*, **19**, 31–76.

Dufour, J.-M. (1986), Recursive stability analysis: The demand for money during the German hyperinflation, pp. 18–61 in: D.A. Belsely and E. Kuh (eds.), *Model Reliability*. Cambridge, MA: MIT Press.

Durbin, J. (1969), Tests for serial correlation in regression analysis based on the periodogram of least squares residuals. *Biometrika*, **56**, 1–15.

Gordon, R.H. and Jorgenson, D.W. (1976), The investment tax credit and counter-cyclical policy, pp. 275–314 in: O. Eckstein (ed.), *Parameters and Policies in the U.S. Economy*. Amsterdam: North-Holland.

Hackl, P. (1980), *Testing the Constancy of Regression Models over Time*. Göttingen: Vandenhoeck and Ruprecht.

Hall, R.E. and Jorgenson, D.W. (1967), Tax policy and investment behavior. *American Economic Review.* **59**, 391–414.

Lucas, R.E. (1976), Econometric policy evaluation: A critique, pp. 19–46 in K.Brunner and A.H. Meltzer (eds.), *The Phillips Curve and Labor Markets.* New York: Carnegie-Rochester Conference on Public Policy 1.

Theil, H. (1971), *Principles of Econometrics.* Amsterdam: North-Holland.

Wilcoxon, F., Katti, S.K., and Wilcox, R.A. (1973), Critical values and probability levels for the Wilcoxon rank sum test and the Wilcoxon signed rank test, pp. 171–259 in Vol. 1 of the *Selected Tables in Mathematical Statistics.* Providence (Rhode Island): The Institute of Mathematical Statistics and American Mathematical Society.

Notes

[1] Lucas (1976) assumes the tax credit follows a Markovian scheme (which includes as special cases both a permanent credit and a frequently imposed but always transitory credit) and shows that the impact of the tax credit on investment can be much bigger if it is viewed as transitory rather than permanent. Indeed, under reasonable values of the parameters, the ratio of the actual to predicted effect may be in the range of 4 to 7.

[2] The Long Amendment forbade firms to use for depreciation purposes the part of the cost of a capital asset financed by the tax credit.

[3] Gordon and Jorgenson (1976, p.278). We list in *Table 25.13* the "effective tax credit" (1961–1972) as measured by these authors. The "effective tax credit" could be nonzero longer than the nominal credit because, even after the credit was suspended, firms could still use a credit to which they were entitled but did not use when it was in force.

[4] Though scaled recursive residuals are similar to t-statistics, one can check easily that they do not generally follow Student t-distributions. Note also that $\tilde{\sigma}^2$ tends to overestimate σ^2 when structural change is present [see Dufour (1982, pp. 60–61)]: clearly, this can make a number of important residuals look "small" and should be discounted when interpreting the results.

[5] Of course, given that K_t is a form of lagged dependent variable and if disturbance are autocorrelated, least squares coefficient estimates could be inconsistent. Nevertheless, the appearance of "autocorrelation" may be a symptom of an instability problem and thus an experiment without such a correction seems indicated. In any case, this will allow us to illustrate how a misspecification can lead to a parameter instability phenomenon in a recursive estimation experiment.

[6] Eight-steps-ahead recursive residuals are not graphed. The test statistics in *Table 25.3*, as well as those in *Tables 25.7* and *25.11*, are based on forward one-step-ahead recursive residuals. We report systematically three categories of tests (general tests, runs tests, and serial dependence tests) that can be compared and cross-checked [see Dufour (1982, Section 4)].

[7] For a listing of the variables TC_t (effective tax credit rate) and U_t (dummy for Long Amendment) from 1961/I to 1972/IV, see *Table 25.13.*

[8] Recursive residuals obtained in this way do not enjoy exactly their convenient theoret-

ical properties (for the true value of ρ is unknown). However, if $\hat{\rho}$ is consistent estimate of ρ, we can still expect it will fall in the neighborhood of the true value of ρ and thus provide approximately valid test statistics. But this is not guaranteed. In view of this difficulty, we performed some sensitivity analysis by considering models transformed by different values of ρ inside a grid in the neighborhood of ρ. In all cases we obtained essentially the same conclusions. For further discussions of this problem, see Dufour (1982, Section 2.5).

[9] It is interesting to compare the residuals in *Figure 25.4(a)* (recursive) with the corresponding generalized least squares residuals in *Figure 25.7* and to see how more revealing recursive residuals are.

Author Index

Subject Index

List of Authors

Yngve Abrahamsen, University of St.Gallen, St.Gallen, Switzerland

Gordon J. Anderson, Economics Department, Toronto University, Toronto, Canada

Lyle D. Broemeling, Statistics Department, Virginia Polytechnic Institute and State University, Falls Church, USA

Francis X. Diebold, Board of Governors of the Federal Reserve System, Washington, USA

Jean-Marie Dufour, Centre de Recherche de Developpement en Economique (CRDE), University de Montreal, Canada

Jozef Dziechciarz, Oskar Lange Academy of Economics, Wroclaw, Poland

Valeri V. Fedorov, International Institute for Applied Systems Analysis, Laxenburg, Austria, now IEPSTP, Academy of Sciences of the USSR, Moscow, USSR

Peter Hackl, Institut f. Statistik, University of Economics, Vienna, Austria, now Department of Statistics and Actuarial Science, The University of Iowa, Iowa City, USA

Andrew C. Harvey, Department of Statistical and Mathematical Sciences, London School of Economics, London, UK

Marie Hušková, Department of Statistics and Probability, Charles University, Prague, CSSR

Walter Katzenbeisser, Institut f. Statistik, University of Economics, Vienna, Austria

André Keller, Group d'Analyse Macroeconomique Appliquee, CNRS, Nanterre, France

Jürgen Kleffe, Karl Weierstraß Institute of Mathematics, Academy of Sciences of the GDR, Berlin, GDR

Walter Krämer, Institut f.Quantitative Wirtschaftsforschung, University of Hannover, Hannover, FRG

Johannes Ledolter, Department of Statistics and Actuarial Science, The University of Iowa, Iowa City, USA

Yuri P. Lukashin, Academy of Sciences of the USSR, Institute of World Economy and International Relations, Moscow, USSR

Grayham E. Mizon, Economics Department, Southampton University, Southampton, UK

Tohru Ozaki, Instiute of Statistical Mathematics, Tokyo, Japan

Valerie H. Ozaki, Sophia University, Tokyo, Japan

Peter Pauly, Department of Economics, University of Pennsylvania, Philadelphia, USA

Garry D. A. Phillips, Department of Econometrics, University of Manchester, Manchester, UK

Werner Ploberger, Institut f. Ökonometrie, Technical University, Vienna, Austria

Wolfgang Polasek, Institut f. Statistik und Ökonometrie, University of Basel, Switzerland

Jaap Praagman, Fac.Bedrijfskunde, Eindhoven Univeristy of Technology, Eindhoven, The Netherlands

Peter M. Robinson, Department of Economics, London School of Economics, London, UK

Bernd Schips, University of St.Gallen, St.Gallen, Switzerland

Pranab K. Sen, Department of Biostatistics, University of North Carolina, Chapel Hill, USA

Howell Tong, Mathematical Institute, University of Kent, Canterbury, UK

Birgitta Törnkvist, Department of Statistics, University of Umeå, Umeå, Sweden

Hiroki Tsurumi, Department of Economics, State University of New Jersey, Rutgers, New Brunswick, New Jersey, USA

Zbigniew Wasilewski, Institute of Econometrics and Statistics, University of Lodz, Lodz, Poland

Anders H. Westlund, Department of Statistics, Stockholm School of Economics, Stockholm, Sweden

THE INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS

is a nongovernmental research institution, bringing together scientists from around the world to work on problems of common concern. Situated in Laxenburg, Austria, IIASA was founded in October 1972 by the academies of science and equivalent organizations of twelve countries. Its founders gave IIASA a unique position outside national, disciplinary, and institutional boundaries so that it might take the broadest possible view in pursuing its objectives:

To promote international cooperation in solving problems from social, economic, technological, and environmental change

To create a network of institutions in the national member organization countries and elsewhere for joint scientific research

To develop and formalize systems analysis and the sciences contributing to it, and promote the use of analytical techniques needed to evaluate and address complex problems

To inform policy advisors and decision makers about the potential application of the Institute's work to such problems

The Institute now has national member organizations in the following countries:

Austria
The Austrian Academy of Sciences

Bulgaria
The National Committee for Applied Systems Analysis and Management

Canada
The Canadian Committee for IIASA

Czechoslovakia
The Committee for IIASA of the Czechoslovak Socialist Republic

Finland
The Finnish Committee for IIASA

France
The French Association for the Development of Systems Analysis

German Democratic Republic
The Academy of Sciences of the German Democratic Republic

Federal Republic of Germany
Association for the Advancement of IIASA

Hungary
The Hungarian Committee for Applied Systems Analysis

Italy
The National Research Council

Japan
The Japan Committee for IIASA

Netherlands
The Foundation IIASA–Netherlands

Poland
The Polish Academy of Sciences

Sweden
The Swedish Council for Planning and Coordination of Research

Union of Soviet Socialist Republics
The Academy of Sciences of the Union of Soviet Socialist Republics

United States of America
The American Academy of Arts and Sciences

ISBN 3-540-51454-6
ISBN 0-387-51454-6