Lecture Notes in
Economics and
Mathematical Systems 407

Marius Ooms

Empirical Vector Autoregressive Modeling

Springer-Verlag

Lecture Notes in Economics and Mathematical Systems 407

Prof. Dr. G. Fandel
Fachbereich Wirtschaftswissenschaften
Fernuniversität Hagen
Feithstr. 140/AVZ II, D-58097 Hagen, FRG

Prof. Dr. W. Trockel
Institut für Mathematische Wirtschaftsforschung (IMW)
Universität Bielefeld
Universitätsstr. 25, D-33615 Bielefeld, FRG

Marius Ooms

Empirical Vector Autoregressive Modeling

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

Author

Dr. Marius Ooms
Department of Econometrics
Erasmus University Rotterdam
P. O. Box 1738
NL-3000 DR Rotterdam
The Netherlands

The Foundation for the Promotion of Research in Economic Sciences (ECOZOEK), which is part of the Netherlands Organization for Scientific Research (NWO), supported this study financially.

ISBN 3-540-57707-6 Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-57707-6 Springer-Verlag New York Berlin Heidelberg

Library of Congress Cataloging-in-Publication Data. Ooms, Marius, 1961- . Empirical vector autoregressive modeling / Marius. p. cm. – (Lecture notes in economics and mathematical systems; 407) ISBN 0-387-57707-6 (Springer-Verlag New York Berlin Heidelberg: acid-free paper) – ISBN 3-540-57707-6 (Springer-Verlag Berlin Heidelberg New York: acid-free paper). 1. Econometric models. 2. Autoregression (Statistics) I. Title. II. Series. HB141.055 1994 330'.01'5195–dc20 93-46882

Typesetting: Camera ready by author
SPIN: 10083814 42/3140-543210 - Printed on acid-free paper

O.q.[1]

'Segđu þat it nſunda, allz þik svinnan kveđa
ok þú, Vafþrúđnir, vitir:
hvađan vindr um kømr, sá er ferr vág yfir?
æ mađr hann siálfan um sér.'

V.q.

'Hræsvelgr heitir, er sitr á himins enda,
iǫtunn, í arnar ham;
af hans vængiom kveđa vind koma
alla menn yfir.'

---

[1]Source: Vafþrúđnismál, Edda (1927, pp. 49–50).

(O.q.: Óđinn kvađ) Óđinn spoke
'Say this ninth, if your sagacity is sufficient
and you, Vafþrúđnir, know:
where does the wind come from, which wanders over the waves?
never does a human see him himself.'

(V.q.: Vafþrúđnir kvađ) Vafþrúđnir spoke
'His name is Hræsvelgr, who sits at the end of heaven,
a giant, in an eagle's shape;
from his wings one says the wind comes
over all men.'

ACKNOWLEDGEMENTS

Several people deserve special credit for their assistance in the creation of this book. First I have to thank my supervisor Teun Kloek for formulating the research project which formed the basis of this study, for his advice and for his careful time-consuming reading effort. His sense for quality and perfection and his patience have been a great help.

I thank the members of the reading committee of my Ph.D. thesis, especially Franz Palm, for their participation and for their comments on the manuscript. Jan Kiviet gave me the idea to embark on this project.

I learned a lot from my colleagues at the econometrics department, who were so friendly to put up with me in this period. Herman van Dijk enthusiastically arranged numerous contacts with inspiring econometricians from all over the world. He was co-author of a paper which encompasses §6.5. Philip-Hans Franses introduced an effective sense for organization and management. Peter Schotman kept me in touch with fashions in theory and practice of empirical econometrics during many valuable conversations. He made a respectable career, which opened up my current job.

My computer skills significantly benefited from advice and tools of Jurgen Doornik, Gerrit Draisma and Arjen Merckens.

I am grateful to the Foundation for the Promotion of Research in Economic Sciences (ECOZOEK) for financial support during four years.

CONTENTS

1 INTRODUCTION

1.1 Integrating results

The empirical study of macroeconomic time series is interesting. It is also difficult and not immediately rewarding. Many statistical and economic issues are involved. The main problems is that these issues are so interrelated that it does not seem sensible to address them one at a time.

As soon as one sets about the making of a model of macroeconomic time series one has to choose which problems one will try to tackle oneself and which problems one will leave unresolved or to be solved by others. From a theoretic point of view it can be fruitful to concentrate oneself on only one problem. If one follows this strategy in empirical application one runs a serious risk of making a seemingly interesting model, that is just a corollary of some important mistake in the handling of other problems. Two well known examples of statistical artifacts are the finding of Kuznets "pseudo–waves" of about 20 years in economic activity (Sargent (1979, p. 248)) and the "spurious regression" of macroeconomic time series described in Granger and Newbold (1986, §6.4).

The easiest way to get away with possible mistakes is to admit they may be there in the first place, but that time constraints and unfamiliarity with the solution do not allow the researcher to do something about them. This can be a viable argument. Even in areas of research where it used to be possible to understand, or at least know about all relevant recent results, this has now become an impossible task. This is a conclusion that recent survey articles on different ranges of subjects share, see e.g. Pagan and Wickens (1990) on the broad subject of econometrics and Diebold and Nerlove (1990) on the narrower subject of unit roots in economic time series.

Pagan and Wickens described this phenomenon as the *fragmentation* of the subject. In every fragment one develops one's own language, which makes it difficult for outsiders to understand results and equally important to judge whether they are of interest for their own problems. A good example is the proliferation of acronyms in time series analysis (Granger (1982)) which has not stopped since then, take e.g. CHARMA, (Tsay (1987)), GARMA, (Gray et al. (1989)), RLARMA, (Chen and Tiao (1990)) only to name a few ending with ARMA.

Nordhaus (1984) aptly described a similar fragmentation in (future) macroeconomics as *balkanization*. Mankiw (1990) called the phenomenon *The Breakdown of the Consensus* in an update of this description.

The Babel caused by this fragmentation generates lots of irrelevant research, first because results are derived in one area, while they are already known and reasonably established in another one, cf. Breusch and Pagan (1979) with Cook and Weisberg (1983), second because many solutions to problems relevant for applied work simply do not reach the researchers involved, so they continue with inadequate methods.

Time series analysts have developed a habit to demonstrate the use of newly developed methods and models on a relatively small sample of canonical data sets, often from Box and Jenkins (1970). "Unit rooters" often used Nelson and Plosser's (1982) macroeconomic data set for the U.S.A. This provides a good method of communicating the relevance of the results, but is often insufficient to explain to relative outsiders why they came about.

In this study we try to contribute to the integration of results from different areas of econometrics, time series analysis and macroeconomics, rather than to their fragmentation. We try to confine ourselves to traditional, well understood models as much as possible and outline their merits and demerits with respect to the analysis of macroeconomic time series. This is only possible to a limited extent for inadequacies of traditional methods that have only recently come to the fore and the rather new solutions proposed to deal with them. Where published, well understood solutions are absent we make own contributions. Granger and Newbold (1986), Nerlove et al. (1979) and Lütkepohl (1991) are very good examples of books that combine results from time series analysis and econometrics to study different aspects of economic time series. This study may serve as a supplement.

1.2 Goal of the study

The basic idea behind this study is to combine results from the literature in the fields of econometrics, time series analysis and macroeconomics to develop, implement and apply a reasonably reliable method to uncover interesting linear time series relationships between macroeconomic variables. Although we hope that the method is adequate for a wider range of applications it is primarily developed to study quarterly, unadjusted multivariate time series of a relatively small dimension containing real variables like output. It is hard to define what constitutes "interesting"

relationships. It depends on the specific application of the model that is used to describe them, such as forecasting, tests of economic theories, policy evaluation or policy advice. Forecasting is not our first interest, although predictions from the model form an important aspect to judge model adequacy. The primary aim is to learn in a sufficiently sophisticated way what the data at hand can tell us. A wide range of diagnostic tests has to signal whether there are probably other important messages in the data that our method does not pick up.

Usually this reduced form analysis is not enough to give policy advice for the future. That would require a structural model with a firm belief in its identifying restrictions, not all of which can be tested empirically. The data based analysis within an adequate statistical framework should be useful to test the applicability of relevant economic theories and to evaluate effects of certain past policy interventions.

1.3 Data and measurement model

The data sets used in the applications are rather narrow. Detailed explanations of relationships can thus not be expected. Since the information conveyed by the observed variation in the series is already limited, we prefer to study the raw series with all their defects, without casual distortion by time averaging and the like. These series are normally not immediately useful to test the adequacy of models derived from economic theory. As Griliches (1985) put it: "The data in our information set are usually incomplete and imperfect." The hope that they will become satisfactory in any near future is very slight indeed. De Leeuw (1990a, 1990b) gave an overview of problems with two key macroeconomic variables, output and investment for the U.S. Problems like the measurement of price and quality of essential components like computer hardware and software compare favorably with the oncoming difficulties of the national products in European countries with the change and removal of their economic borders. Economic theory does not provide ready–to–use solutions either. Griliches (1985, 1986) again: "In practice one cannot expect theories to be specified to the last detail nor the data to be perfect or of the same quality in every context. Thus any serious data analysis has to consider at least two data generation components: the economic behavior model describing describing the stimulus–response behavior of economic actors and the measurement model, describing how and when this behavior was recorded and summarized. While it is usual to focus our analysis

on the former a complete analysis must consider them both."

This study is focusing on the latter and is thus more closely connected with areas like data analysis and time series analysis. Nowadays rather advanced time series analysis is also playing a role in the discussion of economic behavior models for aggregate data. The area of time series analysis seems a good choice as a basis to set up the model. It seems obvious that the analysis should be multivariate when applied to economic series. Quah (1990) even argued that in general "univariate characterizations of aggregate time series are simply not informative for economic theory." That is a pity because compared to the experience with univariate time series analysis of economic variables where a considerable amount of practical problems have been discovered, examined, come to be reasonably well understood, and dealt with, this is far from being the case for multivariate time series. There lie gaps in knowledge, methods and experience with applications that we try to help to fill.

An important part of the measurement process, which has to do with construction of aggregate data from surveys, the composition of price and quality indexes etc. is outside the area of data analysis and time series analysis. Other parts which used to be left to the statistical agencies like seasonal adjustment[1] and treatment of exceptional observations are well analyzed within the context of modern time series analysis and occasionally in the context of economic behavior models (Ghysels (1988)). It is known that these adjustments can influence the tests of subsequent statistical and economic models a great deal (Ghysels (1990a, 1990b), Ghysels and Perron (1993), Jäger and Kunst (1990)). Modeling of the seasonal and outlier component is thus an important topic of this study. Up to a few years ago this was a nearly untrodden area of research in the context of multivariate time series.

The seasonality and outliers bring us to the next important subject, namely the decomposition of multivariate time series in trend, cyclical, seasonal and outlier components, which has proven to be a fruitful idea in both statistical and economic analysis of time series. The study of the calendar component, which takes account of variation due to the changing number of trading days and holidays (Cleveland, (1983)) falls beyond the scope

[1] Hausman and Watson (1985) integrated part of a measurement model and a time series model.

of this study. Thury (1983) demonstrated that it can be present in quarterly macroeconomic data, but it is usually only important in data that are sampled more frequently.

1.4 Baseline model and methodology

The most important tool in our analysis is the VAR model, the use of which was advocated most strongly by Sims (1980). The methodology based on the ideas from that article has been extended and refined substantially over the last decade, especially in the field of the economic interpretations which can be extracted from this model. Blanchard (1989) is a good example of the progress in that field. Important properties of the model are its linearity and its flexibility, in the sense that one does not have to make many a priori restrictions to set it up, compared to other econometric models. The basic estimation procedure, least squares regression, is well understood, easy to apply and known to be quite robust to (near) (seasonal) unit root nonstationarities (Chan and Wei (1988)), which are now widely believed to be important characteristics of most macroeconomic time series. The superiority of least squares is also well established in the finite sample case (Tjøstheim and Paulsen (1983), Hannan and McDougall (1988)).

The question is whether this linear model is able to give a reliable description of the data. This is a point which received a fair amount of attention in Sims (1980, pp. 17–18) but has remained somewhat underdeveloped in the applications of the "VAR methodology" since. In this methodology the estimates of the unrestricted VAR are assumed to summarize the sample information about the joint process of the variables included. Since the VAR is merely a summary of all the variances and covariances of the variables and their lagged values, this implies that those must be sufficient statistics for all the parameters of interest. We are especially interested in parameters that are relatively stable over the sample period, i.e. not time dependent. It is therefore interesting to see whether this is a reasonable assumption.

A nice feature of the VAR model is that it can be seen as a multivariate linear regression model. A check for the stability of the autoregressive parameters can be performed by a range of tests developed for the (static) linear regression model, which have been shown to be applicable for dynamic regression models too (Krämer et al. (1988), Ploberger et al. (1989)). Lots of other tests on the specification can also be applied. Sensitivity analysis with respect to i.a. the choice of variables, sample period, lag length, can be done quite easily.

Although the VAR model is a finite parameter model, it is important to note that many parameters of interest are in fact functions (Parzen (1983)), e.g. the impulse responses or the spectrum of the series. This brings us to another advantage of the linearity, namely the relatively well known interpretation of the autocorrelation function and its Fourier transform, the spectrum of the series. The spectral analysis of economic time series models has recently become fruitful again (Singleton (1988), Quah (1990)). The VAR model cannot be taken as a true model of the data generating process. It is a finite parameter model which is used to approximate the truth, i.c. the multivariate spectrum and the linear lead–lag relationships between the variables in the sample. This can imply that the lag length needed in estimation is an increasing function of the length of the homogeneous time series available. In this opinion a true lag length does not exist. See Lütkepohl (1988, 1989) for statistical analysis in this spirit. Unfortunately this reduces the reliability of classical statistical analysis, but it may still serve as an intersubjective manner to communicate meaningful statistical results. In good practice the chosen significance levels are also functions of the sample size (see e.g. Arrow (1960)).

The advantages of the linear model can also be seen as disadvantages of nonlinear time series models. We present some of them in the sequel but mainly as alternatives to test our null against. Most alternative models are neither as well understood nor as easy to estimate, which makes sensitivity analysis difficult. Interpretation of autocorrelation function and spectrum is not always available. In summary: they are as yet unreliable. All in all we are very reluctant to give up our baseline model as a basic tool of analysis, but it has to be accompanied by reasonably powerful methods to detect its failures to capture important properties of the data. As Sims (1988) stated on forecasting applications of the model: "If these facts are not allowed for in analyzing models' historical forecasting records, results are likely to be anomalous or disappointing." One can strive to avoid anomalies and hope to reduce the occurrence of disappointments. Often relatively small modifications can be sufficient to resolve the deficiencies so that expeditions into the heart of unknown research territories can be postponed. The modifications, or rather supplements to the VAR model that we suggest are also data based and concern an adequate modeling of the trend, outlier and seasonal component to keep a covariance stationary process for the remaining cyclical component.

The ideas are inspired by literature on robust methods and the well

established literature on seasonal and calendar correction, where some basic methods, like the Census X–11 procedure are extremely long lived. Cleveland (1983) explained this: "Part of the X–11 reliability undoubtedly stems from the fact that the designers started with the data, knew the intricacies of the data, and built procedures to realistically face the data." This strand of the econometrics literature recognized the importance of robust methods at an early stage, not because of fancy mathematics but out of interest in the data. Another important characteristic was their interest in the development in new graphical methods. The message in the data is hard to summarize in a few statistics. A few pictures will help usually, first to find the origin of the statistics in the data, and second to show other characteristics not reflected in those statistics.

1.5 Outline of the study

In the second chapter we formally introduce the baseline version of the VAR model, that contains the basic statistical assumptions that we use in the sequel. One first has to check whether the variables one uses can be transformed to meet these assumptions. The first point is an analysis of the *univariate characteristics* of the series. The VAR model implies linear ARIMA models for the univariate series (Zellner and Palm (1974)). Important properties of these models can be examined in the sample: a bounded spectrum, the order of (seasonal) integration, linearity and normality after the appropriate transformation. The basic properties of univariate purely stochastic *integer unit root integration* are pointed out. Subsequently these properties are contrasted with the properties of stochastic *fractional integration.* We suggest data analytic tools to check the assumption of univariate unit root integration. In an appendix we give a detailed account of *unit root tests* for stochastic unit root nonstationarity versus deterministic nonstationarity at frequencies of interest.

The third chapter starts with a few sections on local and global *influence analysis*, which should point out the observations with the most notable impact on the estimates of location and covariance parameters. The results from this analysis can be helpful in spotting the sources of possible problems with the baseline model.

After the influence analysis we discuss the merits of different *statistical diagnostic tests* for the adequacy of the separate regression equations. After one has estimated the unrestricted VAR one should check some overall characteristics of the system. We present several suggestions how to do this.

The fourth chapter deals with common sources of misspecification stemming from problems with *seasonality* and *seasonal adjustment* in the multivariate model. We discuss a number of univariate unobserved components models for stochastic seasonality, which gives additional insight in the properties of models with unit root nonstationarity. We also suggest a modification of a simple but quite robust seasonal adjustment procedure. Some new data analytic tools are introduced to examine the seasonal component more closely. In appendix A4.1 we discuss the limitations of deterministic modeling of seasonality. In appendix A4.2 we discuss some aspects of backforecasting in models with nonstationarity in mean.

In the fifth chapter we introduce outlier models. We develop a testing procedure to direct and evaluate the treatment of exceptional observations in the VAR. We illustrate its application on an artificial data set that contains important characteristics of macroeconomic time series. The effect of the outliers and the effectiveness of the testing procedure is also analyzed on a four–variate set of quarterly French data, which exhibits cointegration. We compare some ready–to–use outlier correction methods in the last section.

The sixth chapter deals with restrictions on the VAR model. First we discuss a number of interesting reparameterizations of the VAR under unit root restrictions. The reparameterizations lead to different interpretations, which can help to assess the plausibility of empirical outcomes. We present some straightforward transformation formulae for a number of these parameterizations and show which assumptions are essential for the equivalence of these models. We illustrate this in simple numerical examples. Next we compare VAR based methods to estimate pushing trends and pulling equilibria in multivariate time series. The predictability approach of Box and Tiao receives special attention. Finally we discuss multivariate tests for unit roots.

The seventh chapter is devoted to an application of methods described in the previous chapters to analyze gross fixed capital investment in the Netherlands from 1961 to 1988 in a six–variate system. We discuss a number of economic approaches to model macroeconomic investment series. We list a number of problems in empirical applications of these models. In §7.3 we present empirically relevant aspects of the measurement model for macroeconomic investment.

In the next section we apply univariate techniques of chapters 2, 3, 4 and 5 on the investment series and five other macroeconomic variables which

can be expected to have a notable dynamic relationship with investment, viz. consumption, imports, exports, the terms of trade and German industrial production. The univariate analysis clearly shows the presence of nonstationary seasonal components in a number of the series. The model is extended with a structural break on the basis of results from the univariate analysis. The subsequent multivariate analysis confirms the need for a structural break in the model for the growth rates of the multivariate series. An empirically important equilibrium relation between investment, imports and exports is seen to remain stable over the entire sample period. The partial correlation of deviations from this equilibrium and growth rates of investment is large and stable.

The parameters in a comparatively parsimoniously parameterized VAR are reasonably stable after correction for seasonal movements, some notable additive outliers, and one structural break. We present the properties of this VAR through impulse responses and variance decompositions of orthogonalized innovations. Identification is based on the variance matrix of the residuals and on the long run variance matrix of the growth rates. The simulated effects of standard shocks to the terms of trade, German industrial production and exports do not change with the identification method. The innovations to these variables represent foreign shocks, which are not significantly influenced by domestic factors in our information set.

1.6 What is new?

The comparison of statistics to examine high order autocorrelations in §2.5 leads to a straightforward frequency by frequency analysis of fractional integration and a modification of adjusted range analysis, which we apply in §7.4.7.3. The simultaneous treatment of statistical unit root tests at different frequencies in appendix A2.2 leads to some new applications of unit root tests at seasonal frequencies in §7.4.5. The prior concern analysis of influence measures and diagnostic test statistics in §3.2.2, §3.3.6 and appendix A3.2 leads to a more insightful use of well known diagnostic tests for parameter stability in §7.5.3. The comparison of tests for parameter stability with unknown change point in §3.3.6 leads to a modification of the CUSUM test on OLS residuals to test the stability of the frequency ½ seasonal component in §7.4.7.7. We suggest a new overall multivariate test for parameter stability in §3.3.9. We introduce a graphical check on the number of unit roots in the VAR in §3.3.9. We apply it in §7.5.2. In §4.4 we suggest a data analysis of seasonal unit root variation by principal components, which

we apply in §7.4.7.4. We suggest a simple VAR modification of the Census X–11 method for seasonal correction in §4.5. We propose a multivariate outlier testing procedure in chapter 5. The application to longer forecast horizons of the nonpredictability approach to estimate pulling equilibria in §6.5.3 leads to a new graphical diagnostic for the number of linear equilibrium relations in a multivariate time series based on long run autocorrelations. We apply the diagnostic in §7.5.2. It shows clear evidence of a long run equilibrium in French macroeconomic data. The evidence in the Dutch data is less pronounced.

The new techniques are all developed because of their relevance for an appropriate VAR interpretation of the macroeconomic data in chapters 5 and 7. Chapter 7 contains the most comprehensive application. Most procedures are easy to implement in computer packages for time series analysis. We have written two programs, one in Sun Pascal (1988) with procedures from NAG (1988), and one in Turbo Pascal (1988). Most computations can be done by one of these programs and take a negligible amount of time for the sample sizes considered here. All figures can be made with these programs too.

2 THE UNRESTRICTED VAR AND ITS COMPONENTS

2.1 Introduction

This chapter deals with the first step in the preliminary analysis of the data in order to find out whether these can be analyzed usefully in the context of a VAR model. The main aim of the VAR model in this study is to discover meaningful linear lead–lag relationships between the variables, without using too much a priori information. The bulk of the information should therefore be extracted from the data.

Since an empirical VAR model is only based on the estimates of variances and covariances of the included variables, it is necessary to check whether one can interpret these estimates in a sound way. In the theory and practice of time series analysis of macroeconomic variables a number of interesting techniques exist that supply applied researchers with different interpretations of the variances and covariances, such as the spectrum and the "random shock" or impulse response representation. Econometric analysis of the rational distributed lag model has provided other interpretations and a statistical toolbox with slightly different contents, i.a. checks on model adequacy developed for relatively small samples.

Regression equations are the natural building blocks of a VAR model. In the context of regression analysis it has been argued that statistical analysis can only be meaningful within the framework of a flexible statistical model (Spanos (1986)) which has to be general enough to capture the major properties of the data. It is statistically unsound to attach meaning to outcomes of tests of specific hypotheses against a more general one if important aspects of the general model can be rejected a priori. Since we would rather not start with a disabled model, blind to the insights we can get from the statistical toolbox of the regression equation, we begin by setting up a slightly modified VAR model. It should serve as a useful maintained hypothesis for statistical analysis and thus be able to describe the well known features of interest of real quarterly macroeconomic data.

In the following section we introduce a formal representation of the modified VAR model. From this set–up we derive conditions on the univariate properties of the included variables which they have to fulfill in order to meaningfully relate them to each other in the multivariate model. In the third

section we present a derivation of the marginal univariate processes in the presence of unit root nonstationarity.

In the fourth section we give our definition of unit root processes. We then discuss the properties of these processes in more detail. In the last section we introduce alternative models which can describe nonstationarity, long memory and persistence. We advocate the use of data analytic tools geared towards detection of these alternatives. In appendix A2.2 we discuss statistical tests for unit root nonstationary. We apply these techniques in §7.4.

2.2 The model

The basic model reads:

$$\Phi(L)y_t = \varepsilon_t, \qquad t\in\mathbb{Z} \tag{2.1}$$

$$y_t = S(L)(w(x_t) - g(t) - h_t(L)\eta_t), \tag{2.2}$$

$$\Phi(L) = V(L)M(L)U(L), \tag{2.3}$$

where

x_t is an n–column vector of observed variables of interest,

y_t is a real, purely stochastic, zero mean n–vector, a transformation of interest of x_t, and

$\Phi(L) = \Phi_0 + \Phi_1 L + \Phi_2 L^2 + \ldots + \Phi_p L^p$ is an real $n\times n$–matrix lag polynomial ,with all roots of $\det(\Phi(z)) = 0$ on or outside the unit circle, and with $\Phi_0 = I_n$, the identity matrix of order n,

ε_t is an n–vector of serially uncorrelated disturbances, $\varepsilon_t \sim \mathrm{N}(0,\Sigma)$, $t = p, p+1, \ldots, T$. Normality is not essential. $\det(\Sigma) \neq 0$, so that identities and strict linear dependencies among the equations are excluded in (2.1).

$S(L)$ is a filter that preserves only the variation of interest in $(w(x_t) - g(t) - h_t(L)\eta_t)$,

$w(x_t)$ is a nonlinear function of x_t,

$g(t)$ is a deterministic or perfectly predictable function of time,

$V(L)$ and $U(L)$ are real matrix lag polynomials of orders p_v and p_u with the roots of $\det(V(z)) = 0$ and $\det(U(z)) = 0$ outside the unit circle,

$M(L)$ is a real diagonal matrix with the roots of $\det(M(z)) = 0$ on the unit circle,

L is the lag operator : $L^i y_t = y_{t-i}$, $i\in\mathbb{Z}$, and

η_t is an n–vector of disturbances different from ε_t, influencing y_t via $h_t(L)$.

Many observable variables x_t are not easily analyzed by the core VAR model alone. One should acknowledge that the core VAR model (2.1) probably is not able to explain all the variation in the data to such an extent that specification testing becomes meaningful. That is why we "extend" the model with (2.2). In practice only part of the variation in x_t, i.c. "the variation of interest" is analyzed in the VAR model. This can be done by an "instant" transformation of the variable, $w(.)$, time domain filtering, $S(L)$, or by correction for deterministic and other exogenous influences, $g(t)$ and $h_t(L)\eta_t$.

By this extension the VAR model essentially becomes a simple dynamic *unobserved components* model. In order to circumvent additional identification and interpretation problems arising from stochastic errors–in–variables specifications, (2.2) is a set of equations without a stochastic error term.

For the transformation in (2.2) to be interesting, y_t has to have an interpretable relation with the observed variable x_t as well as statistical characteristics that fit into the VAR model. For some variables it will be so difficult to find a transformation to meet these requirements that traditional VAR analysis becomes a waste of time. One could try to extend the model with an unobserved components element for the variance of the series in that case, see e.g. Harvey et al. (1992), but this is outside the scope of this study.

We provide empirically relevant choices for $S(L)$, w, $g(t)$ and $h_t(L)\eta_t$ in the following sections and chapters. In order to get a first idea one can think of them as scalar functions with $S(L)$ a moving average filter for seasonal adjustment (§4.5), w a logarithmic transformation , $g(t)$ a polynomial of time (§A4.2.4), and $h_t(L)\eta_t$ an occasional dummy variable for outlying observations (§5.2).

The (Smith–McMillan) factorization of $\Phi(L)$ plays a crucial role in multivariate unit root analysis, which can be helpful in extracting components of interest from the variables in the core model (2.1). We discuss this in sections 2.3, 3.3.9 and 6.2 below.

The aim of the preliminary analysis is to systematically check whether there exist well understood transformations for the data at hand. From (2.1) we can deduce the implied processes for the univariate y_{jt}s, $j=1,\ldots,n$. This is done by inverting (2.1) into its MA representation (Zellner and Palm (1974)), i.e. finding the particular solution of the difference equation (2.1). We abstract here from the deterministic parts of the solution, see §A2.2.2.3 for a short discussion on that.

We write the MA representation as

$$y_t = C(L)\varepsilon_t, \tag{2.4}$$

where $C(L) = I_n + C_1 L + C_2 L + \ldots$ is sometimes denoted as $\Phi(L)^{-1}$.

Suppose $M(L)$ equals the identity matrix I_n, i.e. $\Phi(z) = 0$ has no roots on the unit circle. Combined with appropriate starting conditions this implies that the univariate processes are mean and covariance stationary and that the polynomial expansion $\Phi(L)^{-1}$ dies out so quickly that the spectrum of y_t is bounded at all frequencies, i.e. the variation of the series in a sample grows at the same rate at all frequencies as the sample size increases.

Empirical analysis (Granger (1966)) has shown that this characteristic is not common among economic variables. Unbounded spectra seem to be the rule rather than the exception. Since some interpretations of autoregressive models only go through under the assumption of a bounded spectrum it was common practice before practical techniques for so–called cointegration analysis had been developed to analyze only univariate transformations with bounded spectra. Empirical analysis using the cointegration techniques and closely related canonical analysis based on Box and Tiao (1977) has made clear that valuable, and economically relevant information about linear relationships at the frequencies where the variables have unbounded spectrum may be recovered if the untransformed series are used.

It has been shown that model (2.1) with a matrix $M(L)$ with some so–called *unit roots* can be used to perform such an analysis. In the case of *unit root integration* the spectra are not only unbounded, but also nonstationary so that the ordinary definition of the spectrum cannot be applied. It is then better to talk of the *pseudo spectrum*, cf. Harvey (1989, p. 64) and §2.4.2. Below we use the terms *spectrum* and pseudo spectrum interchangeably.

Hylleberg et al. (1990) used the factorization of $\Phi(L)$. It finds its origin in the theory of *unimodular* matrices, see Kailath (1980, Chapter 6) and §3.3.9. Conditions for its existence are defined on the polynomial of the equivalent MA representation, which we derive in the next section. Different interesting parameterizations of this model exist. We discuss some of its explicit forms in chapter 6.

Consider the equations for the y_{jt}s in (2.4) separately. It is possible to show that the characterization of the unbounded spectra for the multivariate series can be checked to some extent using univariate analysis. In the literature most attention has been given to the spectrum of the series

near the zero frequency. Careful modeling of the low frequency components is especially important if one wants to use the model to search for theoretically appealing long run equilibrium relations in a particular data set (Engle and Granger, 1987). It may also be important for long range forecasting (Engle and Yoo (1987), LeSage (1990)). In the literature on seasonal models the seasonal frequencies received a lot of interest, see chapter 4. Hylleberg et al. (1990) set up a model for the analysis of all separate poles in the spectrum one at a time. In the next section we use their analysis to derive univariate characterizations for the y_{jt}s.

2.3 Univariate processes and unit roots

We start by making the following additional assumption on $\Phi(L)$, which simplifies the unit root analysis. The introduction of unit roots in the autoregressive representation of a process is one way to model the stylized fact of an unbounded spectrum.

Assumption 2.1

$\Phi(L)$ is of finite order p.

From this assumption it follows that $M(L)$ has a finite order, say p_m. This implies that the "unit" roots of $\det(M(z))=0$ belong to the set W with a finite number of elements w_1, w_2, w_3, defined by $\{e^{i2\pi j/S};\ S=1,2,\ldots,np_m;\ j=1,2,\ldots,S;\ i^2=-1\}$. Note that since the coefficients of $M(L)$ are real, $\det(M(w_j))=0 \Leftrightarrow \det(M(w_j^{-1}))=0$. All n diagonal elements $m_j(L)$ of $M(L)$ can thus be written as:

$$m_j(L)=\prod_l (1-w_{lj}L)^{d_{lj}},\ j=1,\ldots,n. \tag{2.5}$$

Define the appropriate scalar *difference operator* for the stationarity achieving transformation as

$$D(L)=\prod_k (1-w_kL)^{d_k}, \tag{2.6}$$

so that all roots w_k with multiplicity d_k of $D(z)=0$ are also root with multiplicity d_k for at least one equation $m_j(z)=0$, and all roots w_{lj} with multiplicity d_{lj} of $m_j(z)=0$ for $j=1,2,\ldots,n$ are a root of *at least* multiplicity d_{lj} for $D(z)=0$. The set of different roots of $D(z)=0$ is the *union* of the sets of different roots in $\det(M(z))=0$. The multiplicity of a root w_k^{-1} of $D(z)=0$ is the *maximum* multiplicity of this root in the polynomials $m_j(z)=0$. Now define $m_j^*(L)$ as $D(L)/m_j(L)$ and $M^*(L)$ as $\text{diag}(m_j^*(L))$ so that $D(L)I_n=M^*(L)M(L)$. Note that $M^*(L)$ is of an order less than or equal to that of

$D(L)$. We denote this order by p_m*. Substituting (2.3) into (2.1) and premultiplying by $V(L)^{-1}$, $M^*(L)$ and $U(L)^{-1}$ one obtains the so–called *vector moving average* (VMA) representation

$$D(L)y_t = U(L)^{-1}M^*(L)V(L)^{-1}\varepsilon_t. \tag{2.7}$$

A sufficient condition on the (rational) MA polynomial $C(L)$ for the existence of the factorization $C(L) = U(L)^{-1}M^*(L)V(L)^{-1}$ is that $C(z)$ is finite for all values of z on or within the unit circle, see Engle (1987). This implies that $D(L)y_{jt}$ has a finite spectrum for all j and t. Rewriting $U^{-1}(L)$ and $V^{-1}(L)$ as $\text{Adj}(U(L))/\det(U(L))$ and $\text{Adj}(V(L))/\det(V(L))$ one gets

$$D(L)\det((U(L)V(L))y_t = \text{Adj}(U(L))M^*(L)\text{Adj}((V(L))\varepsilon_t. \tag{2.8}$$

The y_{jt}s can be written as (non)stationary ARMA models of a type that is somewhat more general than the familiar ARIMA model. The AR parts of the univariate models are of maximum order np and the MA parts are of maximum order $n(p-1)$. In practice one can obtain lower orders of AR parts and MA parts, due to common factors in these components or other "coincidental situations", see Palm (1977). This is illustrated above for the factors with roots on the unit circle. One can sometimes interpret the cancellation of factors as a reduction in the number of underlying "inputs", "causes", or "trends", see e.g. §6.2.2.

The roots on the unit circle of the univariate AR parts are of special interest. We distinguish two cases, one where all the $m_j(L)$s are equal and an alternative where they are not. Granger developed the term *co–integration* for the latter situation. The AR component of a y_{jt} can have no more unit roots than $D(L)$. The order of $D(L)$, say p_d is lower than np_m whenever the $m_j(L)$s have roots in common. It can be as low as p_m when all the $m_j(L)$s are equal.

No cointegration

In the first case $M^*(L) = I_n$, and the number of unit roots in the AR part of the system, which equals n times the number of unit roots in $D(L)$ is equal to the sum of the numbers of unit roots for the univariate processes. This specification is the null hypotheses in *multivariate unit root tests* discussed in §6.6. The number of unit roots in their MA parts is then always equal to zero, so that the orders of the MA parts cannot be greater than $(n–1)(p_v+p_u)$.

Cointegration

This is no longer true if the $m_j(L)$s differ. Then the multivariate AR part and the MA part have common factors with unit roots, which cancel since

we only look at the purely stochastic particular solution of (2.1). Note that we arrange the deterministic parts of the solution under (2.2). This implies that the number of unit roots in the AR part of the system is smaller than the sum of the numbers of unit roots for the univariate processes.

The y_{jt}s can still have the same unit roots in the univariate representations, but some combinations, the so–called (polynomial) *cointegrating vectors* $e_j(U(L))y_t$, with e_j the j–th unit vector of order n, have fewer unit roots in the AR part of their univariate representations. For roots at the frequencies zero and ½ ($w_i=1$, $w_i=-1$) this effect can already occur when $U(L)$ has order zero.

In the case of different $m_j(L)$s it is also possible that $e_jy_t=y_{jt}$ already has fewer unit roots than $D(L)$, which we could call "trivial cointegration". Often it is preferable to begin with a *balanced system*, i.e. a system with equal (nonnegative) "orders of integration" (i.c. numbers of unit roots) for each y_{jt} at the frequencies of interest. It is a maintained assumption for some multivariate tests for unit root nonstationarity, see Chapter 6. In a balanced system the number of unit roots in the system is a linear function of the number of independent cointegrating vectors. In a cointegrated system the orders of the MA parts of the univariate processes are greater than or equal to $(n-1)(p_v+p_m)+p_{m*}$.

Examples

Here are two simple numerical examples of $V(L)M(L)U(L)y_t=\varepsilon_t$ for the second case with the different $m_j(L)$s:

Example 2.1 *Cointegration at frequency* 1/2

$$\begin{bmatrix}1-0.8L & 0\\ -0.1L & 1-0.7L\end{bmatrix}\begin{bmatrix}1-L^2 & 0\\ 0 & 1-L\end{bmatrix}\begin{bmatrix}1 & -1\\ -1 & 2\end{bmatrix}\begin{bmatrix}y_{1t}\\ y_{2t}\end{bmatrix}=\begin{bmatrix}\varepsilon_{1t}\\ \varepsilon_{2t}\end{bmatrix} \quad \Leftrightarrow$$

$$(1-L^2)(1-0.8L)(1-0.7L)\begin{bmatrix}y_{1t}\\ y_{2t}\end{bmatrix}=\begin{bmatrix}2 & 1\\ 1 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & 1+L\end{bmatrix}\begin{bmatrix}1-0.7L & 0\\ 0.1L & 1-0.8L\end{bmatrix}\begin{bmatrix}\varepsilon_{1t}\\ \varepsilon_{2t}\end{bmatrix} \quad (2.9)$$

The roots of $\det((M(z))=0$ are $e^{-i\pi}$, e^0 and e^0. The maximum multiplicity for both roots in an $m_j(L)$ is 1. So $D(L)$ is chosen as $(1-e^0L)(1-e^{i\pi}L)=(1-L^2)$. The AR part of both variables has order $n(p_v+p_u)+p_d=2(1+0)+2=4$. It follows that both y_{1t} and y_{2t} have unit roots at the frequencies zero and ½, if the MA part ε_t has a regular covariance matrix. The linear combination $2y_{2t}-y_{1t}$ has only a unit root at the zero frequency however. It does not have an infinite spectrum at the frequency ½ whereas the individual series have.

Change $U(L)$ to get

Example 2.2 *Trivial cointegration*

$$\begin{bmatrix} 1-0.8L & 0 \\ -0.1L & 1-0.7L \end{bmatrix}\begin{bmatrix} 1-L^2 & 0 \\ 0 & 1-L \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \quad \Leftrightarrow$$

$$(1-L^2)(1-0.8L)(1-0.7L)\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1+L \end{bmatrix}\begin{bmatrix} 1-0.7L & 0 \\ 0.1L & 1-0.8L \end{bmatrix}\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \quad (2.10)$$

In this case

$$(1-L)(1-0.8L)(1-0.7L)y_{2t} = 0.1L\varepsilon_{1t} + (1-0.8L)\varepsilon_{2t}.$$

The right hand side is written to make it a process with a finite spectrum bounded away from zero for all frequencies, including ½. We see that the original unit root part $m_2(L)$ for $e_2U(L)y_t = y_{2t}$ returns: a case of "trivial cointegration" at frequency ½. Note again that we only consider the purely stochastic part of the solution here. In appendix 4.2 we discuss deterministic parts of the solution for some cases of interest, for the multivariate case. §A2.2.2.2 contains some univariate examples.

Discussion

We have shown that the processes y_{jt} have to behave as stationary ARMA models of maximum orders $(np,(n-1)p)$ after appropriate "differencing" in order to be explained by model (2.1)–(2.3) under Assumption 1. Different orders of integration at different frequencies can be obtained. In systems where all univariate processes posses the same number of unit roots at the frequencies of interest, the total number of "unit roots" at these frequencies in the system can be smaller than the sum of the numbers of the univariate processes. In that case there can exist nontrivial linear combinations of the processes that have an order of integration lower than each of the separate components. Often these combinations are interesting from an economic theory point of view.

Several authors used VAR models to check whether certain types of this cointegration, predicted by economic theory could be traced back in the data (Campbell and Shiller (1987)). This procedure only works well if the univariate processes satisfy these special integration assumptions. In the next sections we discuss a range of useful methods to check these assumptions for the variables in a particular data set. Alternative models are discussed as well.

2.4 Integrated processes

2.4.1 Definitions and notation

We first introduce a terminology for the univariate processes that the unit root VARs can produce. It turns out that the zero frequency and the ½ frequency need a slightly different treatment than the other frequencies between zero and ½. The roots belonging to those frequencies are equal to their own inverse which entails that a first order moving average operator suffices to "remove" these unit roots.

Definition 2.1

A purely stochastic time series process is *integrated of order d* (at the zero frequency) if and only if the spectrum of the series is finite and bounded away from zero at the zero frequency after premultiplication of the series by the filter $(1-L)^d$. It is *integrated of order* $d_{1/2}$ *at frequency* ½ if and only if its spectrum is finite and bounded away from zero at frequency ½ after premultiplication of its series by the filter $(1+L)^{d_{1/2}}$.

Definition 2.2

A purely stochastic time series process is integrated of order d_λ, $\lambda \in \mathbb{Q}$, $0<\lambda<½$ at frequency λ, if and only if the spectrum of the series is finite and bounded away from zero at frequency λ after premultiplication of the series by the filter $(1-2\cos(2\pi\lambda)L+L^2)^{d_\lambda}$.

Notation

We denote these processes $I(d)$, $I_{1/2}(d_{1/2})$ and $I_\lambda(d_\lambda)$ respectively. This notation slightly generalizes that by Engle and Granger (1987) and Hylleberg et al. (1990).

Remark 2.1

It is implicitly assumed that *the* spectrum (or mutatis mutandis *the* autocovariance function) of the appropriately differenced series is not time dependent. In the time series literature this type of nonstationarity is called *homogeneous*, Box and Jenkins (1970), to distinguish it from other types of nonstationarity with time dependent spectra, cf. Hallin (1984).

Remark 2.2

Here we assume d is a nonnegative integer. The definition can be extended to rational d, see §2.5.2.

Discussion

In the practice of macroeconomic time series analysis the frequencies j/S $j=1,2,\ldots,S$, with S the number of observations per year, are the most likely

ones to exhibit an unboundedness of the (pseudo)spectrum, besides the zero frequency. They are interpreted as frequencies belonging to the yearly cycle and its harmonics. Suppose the process has a finite spectrum bounded a way from zero at the seasonal frequencies after applying the well known seasonal difference filter $(1-L^S)$. It is then readily shown that it is integrated of order 1 at all the frequencies j/S, $j=1,2,...,S-1$ and at the zero frequency. One can see this by using the equation

$$(1-L^S) = \prod_{j=1}^{S} (1-e^{i(j2\pi/S)}L) = (1-L)(1+L)^p \prod_{j=1}^{[(S-1)/2]} (1-2\cos(2\pi j/S)L+L^2), \qquad (2.11)$$

with $[\cdot]$ the entier function and p equal to 1 when S even and zero otherwise, see e.g. Harvey (1989, p. 21).

Integration at frequency j/S is equivalent to integration at frequency $(S-j)/S = 1-j/S$, since $\cos(2\pi(1-f)) = \cos(2\pi f)$. This may seem strange at first sight. It is not an a priori restriction to help estimation or interpretation of models for the underlying process later on. It is just a logical consequence of the measurement model. In practice we cannot make a distinction between those two frequencies, because the observed frequencies are the same. This is an example of the so–called *aliasing* phenomenon. Only more frequent sampling could help us make the distinction. Using quarterly data one cannot make a distinction between processes with period 4 quarters and 4/3 quarter because the last period is connected with a frequency greater than the *Nyquist frequency* of ½ cycle per quarter. With monthly observations the same periods belong to the frequencies 1/12 and ¼ which are both smaller than ½ cycle per month. The frequencies can then be discriminated.

At the zero frequency side of the spectrum we have similar problems that follow from our measurement model. We cannot discern variation at frequencies belonging to periods longer than the sample, and for somewhat higher frequencies for which there are not many nonoverlapping periods in the sample this is very difficult. Note that as S tends to infinity $I_{1/S}(1)$ processes come to resemble $I(2)$ ones, the "difference" operator tends to $(1-2L+L^2) = (1-L)^2$. It becomes impossible to discriminate between these types of integrations at the low frequencies. At the other side of the spectrum $I_{j/S}(1)$ processes come to resemble $I_{1/2}(2)$ ones when j/S tends to ½.

A useful graphical mnemonic is to picture the coefficients of an AR(2) process $1+\phi_1 L+\phi_2 L^2$ in the well known *triangle in the ϕ_1,ϕ_2 plane*. The lines $\phi_1-\phi_2=1$, $\phi_1+\phi_2=-1$, and $\phi_2=1$ enclose the area of $I(0)$ values. The lines represent $I(1)$ values with integration at frequencies ½, zero and other

frequencies λ_j respectively, except for the top intersections which represent the $I_{\frac{1}{2}}(2)$ and $I(2)$ values[1].

2.4.2 MA representation, autocorrelation and pseudo spectrum

Before the discussion of more general models a short recapitulation of the properties of univariate integrated processes is in order. By definition an integrated process y_t of order d at frequency λ can be written as:

$$(1-(2\cos 2\pi\lambda)L+L^2)^{dl(\lambda)}y_t=\varepsilon_t, \qquad 0\le\lambda\le\tfrac{1}{2},\ d\in\mathbb{N}, \tag{2.12}$$

with $l(0)=l(1)=\frac{1}{2}$ and $l(\lambda)=1$, $0<\lambda<\frac{1}{2}$ and ε_t a process with a spectrum $s_\varepsilon(\omega)$ so that $0<s_\varepsilon(\omega)<\infty$, $0<\omega\le 2\pi$. Again we use the fixed startup with $\varepsilon_t=0$, $t<0$. Here we discuss the simple integrated process where $d=1$ and ε_t is a zero mean white noise with variance σ_ε^2, $t\ge 0$. The MA representation of y_t can then written as follows, see appendix A2.1:

$$y_t=\textstyle\sum_0^\infty \theta_k\varepsilon_{t-k}, \tag{2.13}$$

with

$$\theta_k=\begin{cases}\sin(2\pi\lambda(k+1))/\sin(2\pi\lambda), & 0<\lambda<\frac{1}{2}\\ 1, & \lambda=0\\ (-1)^k, & \lambda=\frac{1}{2}\end{cases}.$$

It is clear that the series of partial sums $\sum\theta_k^2$ does not converge. The process for y_t is not covariance stationary. If one takes $\varepsilon_t=0$, $t<0$, the variance of y_t increases linearly with t, in particular $\mathrm{var}(y_{n/\lambda})=n\cdot\mathrm{var}(y_{1/\lambda})$, $\lambda>0$. If we observe the process y_t only at times $t=j/\lambda$, $j=1,2,\ldots$, it is equivalent to a simple *random walk*:

$$y_{[j/\lambda]}=y_{[(j-1)/\lambda]}+\nu_{[j/\lambda]}, \tag{2.14}$$

with $\nu_{[j/\lambda]}$ white noise with variance σ_ν^2 and

$$\sigma_\nu^2=\begin{cases}\sigma_\varepsilon^2\cdot\sum_1^{[1/\lambda]}\sin^2(2\pi\lambda(k+1))/\sin^2(2\pi\lambda), & 0<\lambda<\frac{1}{2}\\ \sigma_\varepsilon^2, & \lambda=\frac{1}{2}\end{cases}.$$

This implies that the autocorrelations at lags $m[j/\lambda]$, $m=1,2,\ldots$ asymptotically (i.e. as $t\to\infty$) approach unity, provided $m[j/\lambda]$ is small relative to t, cf. Granger and Newbold (1986, p. 42). Since $\theta_k\approx-\theta_{k-[j/2\lambda]+m[j/\lambda]}$, $m=1,2,\ldots$ the autocorrelations at lags $[j/2\lambda]+m[j/\lambda]$ approach -1 as $t\to\infty$:

$$y_{[j/2\lambda]}\approx -y_{[(j-1)/2\lambda]}+\nu_{[j/2\lambda]}, \tag{2.15}$$

---

[1] Box and Jenkins (1970) used $1-\phi_1L-\phi_2L^2$, which turns the triangle upside down.

with $\text{var}(\nu_{[j/2\lambda]}) \approx \frac{1}{2}\text{var}(\nu_{[j/\lambda]})$, where an equality sign holds when $j/2\lambda$ is integer. This is a second example of *aliasing*. The variation at the underlying frequency λ is observed at frequency zero. We employ this property in §4.4. The complete pattern of the oscillating nonconverging autocorrelation function is derived in appendix A2.1. The *pseudo autocovariance generating function* $\gamma(z) = \sum_{-\infty}^{\infty} \gamma_i z^i$ is given by

$$\gamma(z) = \lim_{\delta \uparrow 1} \sigma_\varepsilon^2 / \left| (1 - (2\delta\cos 2\pi\lambda)z + \delta^2 z^2)^{dl} \right|^{-2} = \lim_{\delta \uparrow 1} \sigma_\varepsilon^2 \left| \sum \delta^i \theta_i z^i \right|. \tag{2.16}$$

The implied variance tends to ∞, but the autocovariances γ_i increase at the same rate to get a finite limit for the autocorrelations.

The *pseudo spectrum*, $f_{yy}(\omega)$ for the process is defined as

$$f_{yy}(\omega) = (\sigma_\varepsilon^2/2\pi) \cdot ((1 - (2\cos 2\pi\lambda)e^{i\omega} + e^{2i\omega})(1 - (2\cos 2\pi\lambda)e^{-i\omega} + e^{-2i\omega}))^{-dl(\lambda)} =$$

$$(\sigma_\varepsilon^2/2\pi)(2\cos 2\pi\lambda - 2\cos\omega)^{-2dl(\lambda)}. \tag{2.17}$$

The concepts of pseudo spectrum and pseudo autocovariance generating function have found useful applications in UCARIMA models, see §4.3.3. The pseudo spectrum is finite for all values ω in the range $[-\pi,\pi]$ except those associated with frequency $2\pi\lambda$, but it increases very fast near those frequencies. One can decompose the variance in the frequency domain as follows:

$$2 \lim_{\delta \to 2\pi\lambda} \left({}_0\int^{\delta} s_y(\omega)d\omega + {}_{\delta}\int^{\pi} s_y(\omega)d\omega \right) = \lim_{t \to \infty} \text{var}(y_t) = \infty. \tag{2.18}$$

The variance of y_t goes to infinity but when we leave out two infinitesimal parts of the spectral decomposition of the variance it becomes finite. The variance in this small sector of the spectrum dominates the variance at the other frequencies. This extreme heteroskedasticity in the frequency domain is one of the most striking aspects of the time series properties of many macroeconomic variables.

Statistical tests under unit root nonstationarity

Appendix A2.2 contains a general discussion of univariate unit root nonstationarity and regression tests for models with this property. The outcomes of these tests can give us information that helps us to interpret the outcomes of the multivariate analysis. They also give indications how to transform the variables to make them meet the assumptions that are necessary in the multivariate analysis.

In a data based analysis we want to get as much information as possible from the observed variation in the variables. We see that by applying a

difference operator the variation of a variable with infinite second moment can become finite, which renders the outcomes of subsequent (multivariate) statistical analysis easier to interpret. One avoids unit root type problems where independent variables have correlations that are "seemingly significantly" different from zero if one incorrectly applies standard distribution theory.

The simplest example is the test for the statistical significance of the correlation between the first differences and the lagged levels a random walk series $y_t = y_{t-1} + \varepsilon_t$, $t = 1, \ldots, T$. For large numbers of observations one would normally reject the null of no correlation against negative correlation at the 5% level for estimates of the correlation coefficient smaller than $-1.65 \cdot T^{\frac{1}{2}}$. The "true" critical value at the 5% level is $-2.89 \cdot T^{\frac{1}{2}}$, however, see table A2.2.4 in the appendix. The nominal critical value corresponds to a significance level of about 45%, which gives an indication of the magnitude of the error one can make by incorrectly ignoring unit root characteristics.

On the other hand one must be careful not to apply the difference filters unnecessarily. Univariate unit root tests may provide guidance in this respect. Does this big loss of variation by the application of the difference operators also lead to a severe decrease in interesting information? Since we are interested in relationships between variables one cannot answer this question using univariate analysis only.

2.5 Alternative models for nonstationarity, long memory and persistence

2.5.1 Nonstationarity

In the discussion of our general model (2.1)–(2.3) we suggested that one should not use the VAR model (2.1) to describe all the important characteristics of the observed components x_{it}. Only the variation that can be related properly within the VAR should be left in. Cointegration theory (see e.g. Engle (1987) and Chapter 6) has established that unit root nonstationarity (particularly $I(1)$–ness) of the different variables can be related in an elegant way with interesting interpretations.

In the previous section we saw that the unit root restriction in the ARMA model provides a simple and potentially useful description of the nonstationarity of the mean and variance of an observed series. The outcomes of the unit root tests within the class of ARMA models defined in (A2.2.7) suggest ways to transform the observed univariate series to stationarity. They are important tools to help us interpret the the sample autocorrelation

function of a well transformed series. The tests are only appropriate when this class of models is adequate for the purpose. Robust versions of the tests have been developed to make them applicable for a wider range of models, allowing weaker assumptions on the process generating the disturbances ε_t^*, see Diebold and Nerlove (1990). Finite sample characteristics of these tests are not always very attractive, see e.g. Schwert (1987, 1989).

There are however still two questionable necessary conditions for adequate application, also in very large samples. The first is that the variance of the disturbances under the alternative of stationarity be finite. This condition is essential for most well known tests for serial correlation, see e.g. Phillips and Loretan (1991b). The second condition is that the spectrum of the stationary process be finite at the frequency zero. Certain *mixing conditions* (see e.g. Billingsley (1968, §20)), which restrict the memory of the process can guarantee this, see e.g. Phillips and Perron (1988).

These two conditions are precisely those that Mandelbrot (1969) thought were not fulfilled for the processes governing many typical economic variables. He stressed the practical implications in analyzing finite samples rather than asymptotic theoretical issues, so he did not "attach a deep significance to the difference between infinite and finite but very large." We don't either. He decided to publish much of his work in the hydrological literature on water resource management, where it seems to have become more widely known than in economics, see Mandelbrot (1977). In connection with series on river flow levels the infinite variance was called *Noah effect*. The property of infinite spectrum was called *Joseph effect*, Mandelbrot (1969). We do not consider the former effect relevant for the type of quarterly macroeconomic series that we are trying to analyze.

This infinite variance phenomenon, which manifests itself by occasional big outliers is more reasonably accounted for by other methods than the generalization of the distributional assumptions on the disturbances, see §3.3.5. That is not to say that data on speculative prices which are sampled at higher frequencies should be treated in the same way. In the fifth chapter we discuss the effect of outlying observations on unit root tests and measures of persistence.

2.5.2 Long memory, the variance time function and adjusted range analysis

Mandelbrot (1972, p. 268) stated that typical "economic time series cannot be spectral synthesized and their spectral analysis is purely a formal technique lacking concrete backing." This critique has not provoked a

revolution: most modern econometric and macroeconomic theory still uses the autocorrelation function of economic time series at the start of the analysis. The simple (continuous time) model of *fractional Gaussian noise*, which Mandelbrot (1972) suggested as a more relevant model for economic time series data, has later been shown to have a spectrum nearly equivalent to that of a simple *discrete fractionally integrated* model, presented by Hosking (1981) and Granger and Joyeux (1980). Jonas (1983, p. 83) compared the different fractional models. The continuous fractional noise has the property of *self-similarity*, i.e. the independence between the shape of the spectrum and the observation frequency of the series. Lawrance and Kottegada (1977) described this concept as *invariance* with respect to *change of time scale*. The discrete version of the model lacks this property. Another important difference is to be found in the phase functions. The discrete fractionally integrated model can be represented as follows:

$$(1-L)^d x_t = \varepsilon_t, \qquad -\tfrac{1}{2} < d < \tfrac{1}{2}, \tag{2.19}$$

with ε_t white noise i.i.d. N(0,1),
where $(1-L)^d$ is defined by its binomial expansion or Maclaurin series, i.e.

$$(1-L)^d = \sum_{j=0}^{\infty} \binom{d}{j} (-L)^j = 1 - dL - \tfrac{1}{2}d(d-1)L^2 - \tfrac{1}{6}d(1-d)(2-d)L^3 \ldots = \sum_{j=0}^{\infty} \frac{\Gamma(-d+j)}{\Gamma(-d)\Gamma(j+1)} L^j,$$

with

$$\Gamma(z) = \begin{cases} \int_0^{\infty} s^{z-1} e^{-s} \mathrm{d}s & \text{if } z > 0 \\ \infty & \text{if } z = 0, -1, -2, \ldots \end{cases},$$

which is defined for negative fractional z by $\Gamma(z) = \pi/(\sin(\pi z) \cdot (-z\Gamma(-z)))$.

It is a conceptually straightforward nonlinear alternative to the integer unit root model. Generalizations have been applied to price indexes (Geweke and Porter–Hudak (1983)) and interest rates (Shea (1989)). Using this model one can extend the definition of an integrated process to $d \in \mathbb{R}$. The operator $(1-L)^d$ is only properly defined asymptotically, which presents some problems for practical applications. One approach is to truncate the filter.

For $|d| > \tfrac{1}{2}$ the operator $(1-L)^d$ is defined through the convolution of the filter with the filters $(1-L)$ and its inverse $1+L+L^2+\ldots$. The latter filter is truncated in finite sample applications. Application of the truncated filter to generate $I(d)$ series from white noise errors is hazardous in moderately

sized samples[2], however.

In the hydrological literature and in parts of the economic literature on the test of market efficiency one uses the so–called *Hurst coefficient* $H = d + ½$, discussed below (eq. (2.29)); see Kaen en Rosenman (1986) for references. An interesting aspect of model (2.19) is the combination of an infinite spectrum and stationarity for $0 < d < ½$, which is impossible in the class of integer unit root models. Sowell (1990) showed that ordinary unit root tests applied to model (2.19) really test the hypothesis $d = 1$ against $d < 1$. Rejection of $d = 1$ does not necessarily lead to the choice for a stationary alternative, since the model is nonstationary for all $d \geq ½$. The power of the test is not impressive however. Stationarity regions for generalizations are derived in Gray et al. (1989). The infinite spectrum of stationary models is often characterized by the expression *long memory*. This definition (op cit.) does so formally:

Definition 2.3

A (stationary) discrete time series (defined over $t = 0, \pm 1, \ldots$) is said to be *long memory* if for some $\omega \in [0,\pi]$ the power spectrum $f(\omega)$ becomes unbounded.

It is clear from (2.19) that the simple fractionally integrated model is long memory at frequency zero for $d > 0$. Hosking (1981, Theorem 1) proved that the spectrum is given by[3]

$$f(\omega) = (2\pi)^{-1}(2\sin ½\omega)^{-2d}, \qquad -½ < d < ½. \tag{2.20}$$

He also proved that $2\pi f(\omega) \sim \omega^{-2d}$ as $\omega \to 0$ for the more general case where ε_t follows a stationary and invertible ARMA process as in (A2.2.7). Diebold and Nerlove (1990) called this model ARFIMA, Gray et al. (1989) used the term FARMA and introduced GARMA for fractional integration at other frequencies. Mandelbrot (1969) called this form of the spectral density near the zero frequency *H–spectrum*, with *H* standing for *Hyperbolic*. He showed the *H*–spectrum property to be a sufficient condition for the nonorthogonality of sample means, i.e. "the asymptotic dependence between the average of a process x_t over the future and its average over the past." Another expression to

---

[2] A better way to generate short *I*(d) series is by premultiplication of a vector of white noise residuals by the square root of the theoretical covariance matrix of an *I*(d) series. Hosking (1981) derived the autocovariance function. Jonas (1983) discussed numerical issues of the simulation of samples from models for fractional noise.

[3] We use the definition where the integral of the spectrum over $[0,2\pi)$ equals the variance, as in (2.17) and in Granger and Newbold (1986, p. 45), so that $f(0) = (2\pi)^{-1}\sum_{-\infty}^{\infty}\gamma_i$, with γ_i the autocovariance function at lag *i*. The variance of the process in (2.19) equals $\Gamma(-2d+1)/(\Gamma(-d))^2$.

characterize the implied behavior he introduced was "*nonperiodic cycles*, (with respect to any finite sample)", see Mandelbrot (1972).

Fractionally integrated processes have autocorrelation functions ρ_τ that decay hyperbolically at high lags, which is different from the geometric decay for stationary ARMA models

$$\rho_\tau \simeq c\tau^{2d-1}, \qquad d < \tfrac{1}{2}, \tau \to \infty,$$

as shown by Hosking (1981, Theorem 2). The *partial autocorrelation function* also dies out more slowly than for stationary ARMA models. This has important practical consequences if one confines oneself to analysis with the traditional ARMA model. In practice only the first sample autocorrelations are used in estimation and the procedure of testing for residual autocorrelation probably will not have substantial power to detect important information about the long run behavior of the variable. In order to find out more directly one can use all the available sample autocorrelations as presented in the sample periodogram of the variable, after the observations have been multiplied by a known filter to render the variable stationary.

The slow decay of the autocorrelation function for positively integrated processes has its mirror image in the hyperbolic decay of the *inverse autocorrelation* function ρi_τ for negatively integrated, or so–called *antipersistent* processes, see Hosking (1981, p. 168). The inverse autocorrelation function of a process is the autocorrelation function corresponding to the inverse spectrum. One can thus regard antipersistent processes as being "integrated in their MA part." We discuss persistence in §2.5.3.

Geweke and Porter–Hudak (1983) developed a simple regression procedure in the frequency domain to consistently estimate the fractional order of integration at the zero frequency, which was later generalized to estimate seasonal fractional integration parameters. Porter–Hudak (1990) published results on the latter type of integration for US monetary aggregates. Hassler (1993a) discussed the use of this procedure as an alternative for unit root tests discussed in appendix A2.2. Sowell (1992) noted defects of this procedure in small samples for series with autocorrelations functions similar to the autocorrelation function of the growth rates of real per capita GDP in the USA. He developed a maximum likelihood estimator for the ARFIMA model with Gaussian disturbances.

Variance time functions

A second way to look at the autocorrelations at lags of higher order is

to study the growth of the variance of partial sums. The sample variance of an $I(1)$ variable, which by definition can be written as an (infinite) partial sum of an $I(0)$ variable, is known to increase linearly with time T, as $T \to \infty$; see (2.13) as an example. Mandelbrot (1972) presented the following

Definition 2.4

The population *variance time function* V_T of a stationary random variable x_t, taken around population expectation is defined as the second moment of the sum of T successive values of x_t.

The variance time function can be written as:

$$V_T = \mathrm{Var}(S_T) = \mathrm{Var}(x_1 + x_2 + \ldots + x_{T-1}) = \tag{2.21}$$

$$T\sigma_x^2(1 + 2\sum_{\tau=1}^{T-1}\{(T-\tau)/T\}\rho_\tau), \tag{2.22}$$

with ρ_τ the autocorrelation function of x_t and σ_x^2 the variance of x_t, see e.g. Diebold and Rudebusch (1989). The importance of the high order autocorrelations is apparent in (2.22). Cochrane (1988) applied the *scaled variance time function*, defined by

$$\mathrm{Var}(S_T)/T\sigma_x^2, \tag{2.23}$$

to *the differences* of log per capita GNP for the US to check the $I(1)$ hypothesis for *the level* of this variable. The limit of the scaled variance time function for $T \to \infty$ is equal to $(2\pi/\sigma_x^2)f(0)$, with $f(0)$ the spectrum of the process x_t at the frequency zero, see Cochrane (1988, eq. (10)). Cochrane also discussed small sample issues, i.a. the effect of the choice of estimator for the time varying mean of the partial sum (as in (2.25) below). There is an asymptotically negligible difference between (2.21) and (2.22) if one employs the common estimators for the variances[4]. From (2.20) it follows that the spectrum at the zero frequency of an "overdifferenced" process ($d \leq 0$) is zero. This entails that the limit is zero for differences of $I(d)$ processes with $d \leq 1$, while it finite and greater than zero for $I(1)$ processes.

In practice the question is not only whether the series is $I(1)$ or $I(0)$. A differenced series may have a very small spectrum at zero, and could for many practical purposes still be considered $I(0)$ in levels. Campbell and Perron (1991, table I) illustrated this for forecasting purposes. The (limit

[4] Direct use of (2.21) downweights the observations at both ends of the sample. Recent research on estimation procedures in the frequency domain suggest that this so called *tapering* can have advantageous effects in small samples, see Harvey (1989). See also appendix 4.2.

of) the rate of increase in the (scaled) variance function gives an indication of the increase in the *uncertainty* of *out of sample forecasts*, which is property of immediate interest.

Cochrane used the limit of (2.23) as an indication of the *size of the random walk* (i.e simple $I(1)$ *component*) in the level of log per capita GNP in the USA. Cogley (1990) discussed problems of inference from this statistic and studied data for other countries as well. The limit of (2.23) (applied to the differenced data) clearly is 1 for a simple random walk. Statistical inference on this measure is not very precise for normal sample sizes. Without strong a priori assumptions on the adequacy of low dimensional parameterizations, which is not in the spirit of the VAR analysis of this study, confidence intervals will remain big, indicating the reasonable uncertainty about low frequency aspects of economic time series. Faust (1992) discussed finite sample behavior of so–called *variance ratio* tests in finite order AR models with normal disturbances in more detail.

The (limit of the) scaled variance time function can also be used to derive tests for an integer *moving average unit root*, i.e. an $I(-1)$ process. Saikkonen and Luukkonen (1990) used this statistic to derive a locally best invariant test for an MA(1) model. It essentially tests whether $f(0)$ is zero for the *series in levels*. Inference is nonstandard for processes with nonzero mean. Their test is based on the variance of the partial sum of residuals of a regression of the partial sum of observations of the $I(-1)$ process on a constant. They gave asymptotic results for more general error structures. Kwiatkowski et al. (1992) gave the statistic an interpretation in an LM test procedure for zero variance of the random walk component.

The test statistic also has an interpretation in a test for parameter stability. It is closely linked with the CUSUM test using OLS residuals, which uses the maximum of the absolute value of the partial sum of the OLS residuals, see §3.3.6.

For processes of orders of integration other than zero the asymptotic growth of the variance time function is of order $O(T^{1+2d})$, $T\to\infty$, see Diebold and Nerlove (1990). An explicit condition for the application of the (functional) central limit theorem used in the derivation of a range of unit root tests (e.g. Phillips and Perron (1988)) is

$$\lim_{T\to\infty} (1/T)(E(S_T^2)) = \sigma^2 > 0, \tag{2.24}$$

with $S_T = x_1 + x_2 + \ldots + x_T$, and x_i a stationary summand with mean zero. See e.g. Spanos (1986) as a reference in the econometric literature. The estimator of

the so-called *long run variance* σ^2 (of a stationary summand, i.e. the first differences) is a crucial element in almost all so-called unit root tests. In AR unit root tests it is the statistic which is used in corrections for short term autocorrelation, see §A2.2.2.4. We compare some estimation methods for σ^2 in the application in §7.4.7.2.

The conventional methods for ARMA models were developed to model flexibly, yet parsimoniously, the autocorrelation properties of short memory processes, i.e. processes with finite spectrum, or equivalently processes with absolutely summable autocorrelation functions. In finite samples they can also be used to reproduce long memory characteristics (Hipel and McLeod (1978a, 1978b)), which is easy to imagine for AR models with near unit roots. It can be treacherous to use them as a data analytic tool to examine long memory characteristics, however, see Cochrane (1988, 1991).

Adjusted range functions

Another way to study long memory aspects is to examine the *rescaled adjusted range function* of a process, introduced and propagated by Mandelbrot as *the* data analytic tool to study dependence structures in economic time series. The method was reviewed and appraised by McLeod and Hipel (1978). Davies and Harte (1987) derived some formal statistical tests based on the rescaled adjusted range for long memory serial correlation in some simple models. The rescaled adjusted range is derived as follows, see McLeod and Hipel (1978). Take first the partial sum adjusted for mean growth in a sample of size T $(t=1,\ldots,\ T)$

$$S_k = \sum_{j=1}^{k} x_j - k\bar{x}_T \tag{2.25}$$

with $S_0 = 0$ and $\bar{x}_T = T^{-1}\sum_{j=1}^{T} x_j$

The *adjusted range* is defined as

$$R_k = M_k - m_k \tag{2.26}$$

with $M_k = \max(0, S_1, S_2, \ldots, S_k)$, $m_k = \min(0, S_1, S_2, \ldots, S_k)$.

Definition 2.5

The *rescaled adjusted range* is given by

$$\bar{R}_k = R_k / D_k \tag{2.27}$$

with $D_k = T^{-1/2}\{\sum_{j=1}^{k}(x_t - \bar{x}_T)^2\}^{1/2}$

Interest in the RAR statistic $\bar{R}_T$ was inspired by Hurst who studied a wide range of time series, both geophysical and economic. Hydrologists applied the measure to examine long term water storage requirements. Direct applications in economics are not hard to come by. One can think of ruin problems, managing

systems for long term exchange rate mechanisms etc. For large T and underlying models for x_t with summable autocorrelation function[5] the following relation has been derived (McLeod and Hipel (1978a, eq. (30))):

$$E(\bar{R}_T) \simeq \pi(\sigma_x^{-2} f(0))^{\frac{1}{2}} T^{\frac{1}{2}}, \qquad T \to \infty. \tag{2.28}$$

Here we see that the scaled variation in the partial sums, now measured using the range, is a simple function of the spectrum at the frequency zero and a power of T.

For the simple fractionally integrated model we have the following relationship

$$E(\bar{R}_T) \simeq a T^H, \tag{2.29}$$

with $H = d + \frac{1}{2}$, the *Hurst coefficient*, also called *self-similarity parameter*, see e.g. Robinson (1991b, eq. (1.2)). McLeod and Hipel (1978) discussed some simple regression procedures to estimate H, and thus the order of integration d, from a sample. A simple one is the regression of the log of averaged values of $\bar{R}_{T'}$, calculated for different subsamples, on log T'. We use a modified version of the regression procedure of Davies and Harte (1987). In contrast to Davies and Harte we only use T' greater than 12 in order to concentrate on long memory autocorrelation.

Lo (1991) developed a *similar* (see §3.3.2) test for the absence of long memory autocorrelation which rejection region for a fixed size α does not depend on the short memory autocorrelation in the series. It combines the rescaled range and the limit of the variance time function to get the so-called *modified rescaled range* statistic

$$R_T^* = R_T / (\hat{\sigma} T^{\frac{1}{2}}), \tag{2.30}$$

with $\hat{\sigma}^2$, a consistent estimator of the *long run* variance of x_t, e.g. the Bartlett estimator of $2\pi f(0)$, see eq. (6.54) in Chapter 6. Under some often applied "fairly general" assumptions on the process for x_t R_T^* can be used to test the null of no long memory at the frequency zero. Asymptotically, R_T^* has the distribution of the *range* of a tied down Brownian motion. An R_T^* in the left critical region indicates antipersistence. Too high a value indicates

[5] Seasonally integrated models do not have a summable autocorrelation function. The variance time functions and adjusted ranges correlation can only be interpreted properly if integration at the seasonal frequencies has been removed first. See chapter 4 for a discussion on seasonal integration.

long memory, or *long range dependence*[6].

We suggest to use the three analytic tools presented in this section together, the sample spectral density of the appropriately differenced series, the variance time function and the rescaled adjusted range, to check the basic assumptions of the model. The outcomes of the statistical tests which are discussed in appendix A2.2 and in the next chapters can then be interpreted more fruitfully.

2.5.3 Persistence

A second interesting interpretation connected with the unit root hypothesis is the concept of *persistence* of the (unobserved) disturbances that affect the underlying ARMA process describing the data. The realized value of the random variables is seen as the cumulative effect of all past disturbances. The MA(∞) representation of an ARMA process as in (2.13) is then interpreted as an impulse response function with the MA coefficients θ_k (k=1,2,...) describing the "response" of variable y_t to a disturbance or "shock" k periods earlier. For integrated processes of at least order 1 $\lim_{k\to\infty} \theta_k$ is not equal to zero so that shocks have a persistent effect on the (seasonal) mean(s) of the process (2.13, A2.2.22). This aspect has received a lot of interest in the macroeconomic literature, because of its importance for time series models of the business cycle variation of macroeconomic variables, implicitly quantifying the scope for anti-cyclical policy. Diebold and Rudebusch (1989) presented a survey on this topic. One of the measures is the spectrum of the differenced series at the zero frequency, which appeared as a limit of the scaled variance time function above. In the economic literature the property of a zero spectrum of the differenced series at the zero frequency is called *mean reversion*. It follows that all series that are $I(d)$ with $d<1$ can be considered mean reverting, see Cheung and Lai (1993, p. 105). Modeling of persistence and business cycles is more interesting in a multivariate context.

---

[6] Lo (1991) presented a series expansion of the cumulative distribution function of this *range*. For $I(0)$-series the statistic has mean $(\frac{1}{2}\pi)^{0.5}$, which is consistent with (2.33). The distribution of the maximum of a *tied down Brownian motion* is best known as the asymptotic distribution for the *maximum absolute deviation* Kolmogorov test statistic for empirical distribution functions, see Billingsley (1968, p. 105). The tied down Brownian motion also appears as the underlying factor in derivations of the limit distributions of test statistics for parameter stability in regression and other tests for $I(0)$-ness, see §3.3.6.

Concluding remarks

One should always take statistical inference on long run characteristics with sufficiently many grains of salt. Empirical application of the products of this statistical theory suffers from at least one of two problems. The first problem is that inference is often based on comparatively short samples of the process, which invalidates the asymptotic approximations. There simply is no information about the long run in short samples. The second problem becomes more prominent when inference is based on longer samples of a process. Here "long" means "covering a long period in time", *not* "consisting of many observations". Long samples often cannot be considered to be homogeneous, which invalidates all inference based on this homogeneity assumption. Moreover, tests for heterogeneity will be more powerful, so that homogeneity can be rejected on statistical grounds as well.

As yet, comparatively little is known about combined tests for unit root nonstationarity and parameter stability. Tests for the null of stationarity, like tests for a moving average unit root in the first differences or nonparametric tests like the ones by Bierens and Guo (1993), have power against a number of nonstationary alternatives. Unit root nonstationarity is only one of them. In the next chapter we discuss some tools to check the homogeneity assumption.

A2.1 MA representation and autocorrelation integrated processes

This appendix contains the derivation of the MA representation of some integrated processes and their pseudo autocorrelation function.

A2.1.1 MA representations

The solution to (2.12) for $\lambda = 0, ½$ can easily be found by successive substitution, or polynomial division (Sargent (1979, Chapter IX)). For $0 < \lambda < ½$ we have the following

Problem

Find a particular solution to the difference equation

$$(1 - 2\cos 2\pi\lambda L + L^2) y_t = \varepsilon_t, \qquad \varepsilon_t = 0,\ t < 0,\ 0 < \lambda < ½. \tag{A2.1.1}$$

For $k \in \mathbb{N}$, $0 < \lambda < ½$, we prove the following

Lemma

$$D_{k/2}(4\pi\lambda) := \sum_{j=-k/2}^{k/2} e^{ij4\lambda\pi} = e^{i(-k/2)(4\lambda\pi)} + e^{i(-k/2+1)(4\lambda\pi)} + \ldots + e^{i(k/2)(4\lambda\pi)} =$$
$$\sin(2\pi\lambda(k+1))/\sin(2\pi\lambda). \tag{A2.1.2}$$

This expression is called *Dirichlet's Kernel* for $k = 0,2,4,..$, cf. Rudin (1964).

Proof.

$$(e^{i4\pi\lambda} - 1) D_{k/2}(4\pi\lambda) = e^{i(k/2+1)4\pi\lambda} - e^{-i((k/2))4\pi\lambda}. \tag{A2.1.3}$$

Multiply both sides of (A2.1.3) by $e^{-i2\pi\lambda}$ to obtain (A2.1.2). □

Proposition

The particular solution of (A2.1.1) is given by

$$y_t = \sum_{k=0}^{\infty} \theta_k \varepsilon_{t-k}, \tag{A2.1.4}$$

with $\theta_k = \sin(2\pi\lambda(k+1))/\sin(2\pi\lambda)$.

One can check the proposition by substitution and straightforward application of trigonometric identities as in Chan and Wei (1988). We present a constructive proof.

Proof

For $|z| < 1$ we have

$$(1 - 2\cos 2\pi\lambda\ z + z^2)^{-1} = (1 - e^{-i2\pi\lambda} z)^{-1} (1 - e^{i2\pi\lambda} z)^{-1} =$$
$$\left(\sum_{j=0}^{\infty} e^{-ij2\pi\lambda} z^j\right)\left(\sum_{h=0}^{\infty} e^{ih2\pi\lambda} z^h\right), \tag{A2.1.5}$$

where

θ_k is the coefficient belonging to z^k in the expansion (A2.1.5):

$$\theta_k = \sum_{j=0}^{k} e^{ij2\pi\lambda} e^{-i(k-j)2\pi\lambda} = \sum_{j=-k/2}^{k/2} e^{ij4\lambda\pi} = D_{k/2}(4\pi\lambda). \qquad \text{(A2.1.6)}$$

Applying the lemma one obtains (A2.1.4). □

In a similar way one can derive the MA representation for x_t with $x_t = (1-L)y_t$:

Following Rudin (1964) for $m \in \mathbb{N}$, $0 < \lambda < \frac{1}{2}$ one has

Lemma

$$\sum_{k=0}^{m} \theta_k = \{\ 1+\cos(2\pi\lambda) - \cos(m+1)2\pi\lambda - \cos(m+2)2\pi\lambda\ \}/(1-\cos 4\pi\lambda). \qquad \text{(A2.1.7)}$$

Proof

Define

$$K_m(2\pi\lambda) = \sum_{k=0}^{m} \theta_k = \sum_{k=0}^{m} D_{k/2}(4\pi\lambda). \qquad \text{(A2.1.8)}$$

By (A2.1.8) and (A2.1.3)

$$K_m(2\pi\lambda)(e^{i4\pi\lambda} - 1) = \sum_{k=0}^{m} e^{i(k/2+1)4\pi\lambda} - e^{-i((k/2))4\pi\lambda}. \qquad \text{(A2.1.9)}$$

Multiply both sides of (A2.1.9) by $(e^{-i4\pi\lambda} - 1)$ to obtain

$$K_m(2\pi\lambda)(e^{i4\pi\lambda} - 1)(e^{-i4\pi\lambda} - 1) = e^0 + e^{i2\pi\lambda} - e^{i(m+1)2\pi\lambda} - e^{i(m+2)2\pi\lambda} + e^0 + e^{-i2\pi\lambda} - e^{-i(m+1)2\pi\lambda} - e^{-i(m+2)2\pi\lambda} \qquad \text{(A2.1.10)}$$

Rearranging terms gives (A2.1.7). □

By (A2.1.7) it follows that the MA-representation of

$$(1 - 2\cos 2\pi\lambda\ L + L^2)(1-L)x_t = \varepsilon_t, \qquad \varepsilon_t = 0,\ t < 0, \qquad \text{(A2.1.11)}$$

can be written as

$$x_t = \sum_{0}^{\infty} \psi_m \varepsilon_{t-m},$$

with $\psi_m = \{\ 1+\cos(2\pi\lambda) - \cos(m+1)2\pi\lambda - \cos(m+2)2\pi\lambda\ \}/(1-\cos 4\pi\lambda)$. (A2.1.12)

A2.1.2 Pseudo autocorrelation function

The complete pattern of the oscillating nonconverging autocorrelation series of the integrated process in (2.12) can be derived as follows.

Proposition

The asymptotic autocorrelations ρ_τ, $\tau \in \mathbb{N}$ are given by

$$\rho_\tau = \begin{cases} 1 & , \quad \lambda = 0 \\ (-1)^\tau & , \quad \lambda = \frac{1}{2}. \\ \cos 2\pi\lambda\tau & , \quad 0 < \lambda < \frac{1}{2} \end{cases} \qquad \text{(A2.1.13)}$$

Proof

Consider the following process with roots outside the unit circle:

$$(1 - (2\delta\cos 2\pi\lambda)L + \delta^2 L^2)^{dl} y^*_t = \varepsilon_t, \qquad 0 \le \lambda \le \tfrac{1}{2},\ d \in \mathbb{N}, \qquad \text{(A2.1.14)}$$

which has MA representation

$$y_t^* = \sum_0^\infty \theta_k^* \varepsilon_{t-k},$$

with

$$\theta_k^* = \delta^k \theta_k \qquad 0 \le \lambda \le \tfrac{1}{2}.$$

The autocorrelations $\rho_\tau(\delta)$ are well defined for $|\delta| < 1$, $0 < \lambda < \tfrac{1}{2}$ by

$$\rho_0 = 1, \quad \rho_1(\delta) = (2\delta\cos 2\pi\lambda)/(1+\delta^2),$$

$$\rho_\tau(\delta) = (2\delta\cos 2\pi\lambda)\rho_{\tau-1}(\delta) - \delta^2\rho_{\tau-2}(\delta), \quad \tau = 2,3,\ldots \tag{A2.1.15}$$

so that

$$\lim_{\delta \uparrow 1} \rho_1(\delta) = \cos 2\pi\lambda. \tag{A2.1.16}$$

If we note that $\lim_{\delta \uparrow 1} \rho_\tau(\delta) = \rho_\tau$ it is clear that we have proved (2.16) for $\tau = 0,1$.

By induction it follows that (A2.1.13) holds for $\tau \in \mathbb{N}$:

$$\lim_{\delta \uparrow 1} \rho_{\tau+1}(\delta) = (2\cos 2\pi\lambda)\lim_{\delta \uparrow 1} \rho_\tau(\delta) - \lim_{\delta \uparrow 1} \rho_{\tau-1}(\delta) =$$

$$(2\cos 2\pi\lambda)\cos 2\pi\lambda\tau - \cos 2\pi\lambda(\tau-1) = \cos 2\pi\lambda(\tau+1). \quad \square \tag{A2.1.17}$$

A2.2 Univariate testing for unit root nonstationarity

First we discuss the univariate model with simultaneous unit root nonstationarity at different frequencies and present regression tests for specifications belonging to this class of models. A wide range of other tests is available, see Diebold and Nerlove (1990). We discuss a few. Campbell and Perron (1991) provided a more informal discussion of tests for zero frequency unit roots. Kiviet and Phillips (1992) presented exact similar tests for the zero frequency unit roots. We prefer the regression tests because of their comparatively reliable significance levels in situations that are relevant for empirical applications (Schwert (1989), Pantula (1991)), and because of their clear interpretation and computation, which allows for easy application of influence analysis and diagnostic checking, see §3.2. We discuss multivariate unit root tests in appendix 6.1. We present the models and tests with increasing levels of generality. By the simultaneous treatment of the different frequencies one can see how one extends (better known) techniques developed for the zero frequency to other frequencies.

A2.2.1 The pure unit root case without deterministic terms

A2.2.1.1 Notation and model

We want to test the following null hypothesis

$$H_0:\ D_0(L)D_{1/2}(L)\prod_{i=1}^{q} D_{\lambda_i}(L)D^*(L)y_t=\varepsilon_t,\qquad t=0,1,\ldots,T, \tag{A2.2.1}$$

with ε_t satisfying the conditions for ε_t in Chan and Wei (1988, Theorem 2.2.)[1].

Here we start with

$\{\ \varepsilon_t\ \}$ a sequence of i.i.d. $N(0,1)$ variables,

$D_0(L)=1-L,\ D_{1/2}(L)=1+L,$

$D_{\lambda_i}(L)=(1-2\cos 2\pi\lambda_i L+L^2),\ 0<\lambda_i<\tfrac{1}{2},\ \lambda_i$–values distinct[2],

$D^*(L)=\prod_{i=1}^{r} D_i^*(L),\ D_i^*(L)=1-d_iL,\ |d_i|<1,$ and

$y_0, y_1, \ldots, y_{p-1}$ fixed.

---

[1] Tsay and Tiao (1990) generalized the set-up to multivariate models.

[2] We do *not* generalize this part of the assumptions. We only work out null hypotheses where y_t is integrated of order 1 at some frequencies. These tests are not valid if y_t is integrated of order d, with d unequal to 0 or 1; see i.a. Dickey and Pantula (1987) for a discussion of practical consequences.

Let $D(L) = D_0(L)D_{1/2}(L)\prod_i D_{\lambda_i}(L)D^*(L) = D_I(L)D^*(L)$ be a lag polynomial of order $p = 2(q+1)+r$. The parameter r indicates the *order of augmentation* of the unit root test regression[3], q indicates the number of complex unit root frequencies.

Following Chan and Wei (1988), we define the following transformations:

$$\begin{aligned} &D(L) = D_I(L)D^*(L), \\ &y_{\mu_i t} = D_I(L)(D_{\mu_i}(L))^{-1}y_t, \ \mu_i = 0, \tfrac{1}{2}, \ \lambda_1, \lambda_2, \ldots, \lambda_q, \\ &D_I(L)(D_{\mu_i}(L))^{-1} = \bar{D}_{\mu_i}(L). \end{aligned} \tag{A2.2.2}$$

$D_I(L)$ is the minimal operator which make the series stationary, cf. (2.6), $\bar{D}_{\mu_i}(L)$ is the operator which makes the series stationary at all frequencies, except at μ_i.

First we assume $D^*(L) = 1$, i.e. the (distinct) roots of the autoregressive polynomial are all on the unit circle. This assumption is relaxed below. We set up the following auxiliary regression for the unit root tests:

$$D(L)y_t = \beta_1 L y_{\mu_1 t} + \beta_2 L y_{\mu_2 t} + \beta_3 L y_{\mu_3 t} + \beta_4 L^2 y_{\mu_3 t} + \ldots + \beta_{2q+1} L y_{\mu_{q+2} t} + \beta_{2q+2} L^2 y_{\mu_{q+2}} + \varepsilon_t. \tag{A2.2.3}$$

The integrated regressors belonging to different frequencies are asymptotically uncorrelated, Chan and Wei (1988, p. 379). Stationary regressors of the form $D^*(L)^{-1}\varepsilon_t$ are uncorrelated with the integrated regressors. This reduces the problem of the derivation of test statistics to a componentwise problem. Chan and Wei (1988) derived the limit distribution under this null hypothesis of the OLS estimator $b_T = (b_1, b_2, \ldots, b_p)_T$ of $\beta = (\beta_1, \ldots, \beta_p)$ using a growing sample size T (see Corollary 3.1.3, (3.2.5) and Corollary 3.3.8, and correct for an obvious typing error in (3.3.21)):

$$\begin{aligned} &Tb_1 \to_{\mathcal{L}} \nu_1 = \tfrac{1}{2}(Q_1^2 - 1)/\Gamma_1, \\ &Tb_2 \to_{\mathcal{L}} -\nu_1 = -\tfrac{1}{2}(Q_1^2 - 1)/\Gamma_1, \\ &Tb_4 \to_{\mathcal{L}} Tb_6 \to_{\mathcal{L}} \ldots Tb_p \to_{\mathcal{L}} -\nu_2 = (2 - Q_1^2 - Q_2^2)/(\Gamma_1 + \Gamma_2), \text{ and} \\ &b_1, b_2, b_4, b_6, \ldots, b_p \text{ asymptotically uncorrelated (Theorem 3.5.1),} \end{aligned} \tag{A2.2.4}$$

with

$Q_i = 2^{1/2} \sum_{j=1}^{\infty} \gamma_j \eta_{ij}, \ i = 1, 2,$

$\Gamma_i = \sum_{j=1}^{\infty} \gamma_j^2 \eta_{ij}^2, \ i = 1, 2,$

$\gamma_j = 2(-1)^j/[(2j-1)\pi],$

n_{ij} i.i.d. $N(0,1)$ random variables,

where $\to_{\mathcal{L}}$ denotes convergence in distribution/law.

---

[3] If either $D_0(L)$ or $D_{\frac{1}{2}}(L)$ or both factors are not present under the null, the lag order decreases to p-1 or p-2.

For later use we introduce[4]

$$W_i = 2^{1/2} \sum_{j=1}^{\infty} \gamma_j^2 \eta_{ij}, \quad i = 1,2, \tag{A2.2.5}$$

$$V_i = 2^{1/2} \sum_{j=1}^{\infty} (2\gamma_j^3 - \gamma_j^2)\eta_{ij}, \quad i = 1,2,$$

$$\tau_1 = \tfrac{1}{2}(Q_1^2 - 1)/(\Gamma_1)^{\frac{1}{2}} = \nu_1 \cdot (\Gamma_1)^{\frac{1}{2}},$$

$$\tau_2 = -1(2 - Q_1^2 - Q_2^2)/(\Gamma_1 + \Gamma_2)^{\frac{1}{2}} = \nu_2 \cdot (\Gamma_1 + \Gamma_2)^{\frac{1}{2}}.$$

A2.2.1.2 Discussion

Unfortunately the distributions ν_1 and $-\nu_2$ are nonstandard, but they can be approximated by truncating the infinite expansions for Q_i and Γ_i. Another way to generate them is by numerous replications of the regression itself for sufficiently large simulated samples. Small sample properties can be studied in a similar fashion. The asymptotic approximation of the finite sample distribution works better for the "unit root β–values" than for the β–values in a stationary autoregression. Note that the variance of the estimator tends to zero at rate $1/T$ whereas this rate is only $T^{-1/2}$ in the stationary case. This so–called "super–consistency" phenomenon is even more pronounced for higher orders of integration, see Chan and Wei (1988). Nabeya and Tanaka (1990) resorted to computer algebra packages to derive a "nonrandom" expression for the distribution function of ν_1 and some related limiting distributions we discuss below. Fuller (1976, pp. 371–372) and Dickey et al. (1984) presented tables with frequently used percentiles of the asymptotic and approximate finite sample distributions, cf. table A2.2.4.

They suggested to use the OLS t–statistics for the β's as well. Under the null these converge also to well behaved distributions, which are functionals of Brownian motion(s). Tables can be found in the same references, where it is also shown that $t_{b_1} \to_{\mathcal{L}} \tau_1$. MacKinnon (1991) estimated response surfaces relating critical values and (small) sample size of (i.a.) this statistic. The distributions for the t–statistic and Tb_1 at the zero frequency are known as the Dickey–Fuller distributions. Probably the best known notation is from Fuller (1976). He denoted Tb_1 by $n(\hat{\rho} - 1)$ and t_{b_1} by τ (from $y_t = \rho y_{t-1} + \varepsilon_t$, $t = 1,\ldots,n$). The negative of the limit distribution of Tb_4 is also known as the Dickey–Hasza–Fuller distribution. Dickey et al. (1984) denoted $-Tb_4$ by $n(\hat{\alpha}_2 - 1)$, which they used to test the hypothesis $\alpha_2 = 1$ in the equation $y_t = \alpha_2 y_{t-2} + \varepsilon_t$. Using the results of Chan and Wei for $\lambda_i = \frac{1}{4}$, it follows that $-\nu_2$ is the limiting distribution of Tb_4 for the model $y_t + y_{t-2} = \beta_4 y_{t-2} + \varepsilon_t$ with $\beta_4 = 0$. The distribution of $\hat{\alpha}_2$ for $\alpha_2 = -1$ is the mirror image of the distribution for

[4] We present corresponding notation in the form of functionals of Brownian motions in §A6.1, where we discuss multivariate generalizations.

$\alpha_2 = 1$, cf. Fuller (1976, pp. 370–372), which explains the similarity of the distributions. We denote the limit distribution of the t–statistic for $\alpha_2 = 1$ by τ_2, corresponding to notation in Dickey et al. (1984). An advantage of the testing approach of Chan and Wei (1988) and Hylleberg et al. (1990) over that of Dickey et al. (1984) is that roots of the equation $1 - z^S = 0$ can be tested separately. The limit distributions of the test statistics for the complex unit roots do not depend on the frequency, so that one can use the same tables for the quarterly and the monthly unit roots (i and $e^{i\pi/6}$). We present a selection of percentiles in table A2.2.4 below.

A2.2.2 Deterministic terms and unknown residual autocorrelation

A2.2.2.1 Generalization of the test regression

The set–up of the previous section is too narrow for practical applications, first because ε_t is assumed to be serially uncorrelated. One can make H_0 more general, by allowing ε_t to follow a stationary and invertible ARMA process and to allow for a drift term. The best known way to deal with serial correlation in ε_t is to assume that it can be approximated by extending the autoregressive process for $D_I(L)y_t$. We replace H_0 from (A2.2.1) by H_0^*, by putting $D^*(L) = 1$ and changing the assumption on the disturbances:

$$H_0^*\text{: } H_0 \text{ with } D^*(L) = 1 \text{ and } \phi(L)\varepsilon_t = \theta(L)\varepsilon_t^*, \tag{A2.2.6}$$

with $\phi(0) = \theta(0) = 1$, roots of $\phi(z) = 0$ and $\theta(z) = 0$ outside the unit circle[5] and ε_t^* i.i.d. $N(0,\sigma^2)$.

This modification of H_0 also covers the case with $D^*(L) \neq 1$ from the set–up of §A2.2.1: choose $\theta(L) = 1$ and $\phi(L) = D^*(L)$ and replace ε_t in (A2.2.1) by ε_t^* from the equation above. One then rewrites H_0^* using $\varepsilon_t = \varepsilon_t^* + (\theta(L)/\phi(L) - 1)\varepsilon_t^*$ and $\varepsilon_t^* = \{\phi(L)/\theta(L)\}\varepsilon_t$, as

$$H_0^*\text{: } D_I(L)y_t = \{(\theta(L) - \phi(L))/(\theta(L))\}\varepsilon_t + \varepsilon_t^*, \tag{A2.2.7}$$

with $\varepsilon_t = D_I(L)y_t$.

Anderson (1983) called this an ARUMA model. Note here that $\theta(0) - \phi(0) = 0$. It indicates that the test regression can be done in the "augmented" form:

---

[5] In chapter 4 we explain that often used seasonal adjustment filters like Census X-11 introduce seasonal unit roots in the MA parts of the ARMA representation of the series. This type of serial correlation does *not* fit the assumptions of many unit root tests. However, if one allows for a big enough augmentation r in the test regression, the distribution of the test statistics for integration at the zero frequency will not be affected too seriously; see Ghysels and Perron (1990).

$$D_I(L)y_t = \beta_1 L y_{\mu_1 t} + \ldots + \beta_{2q+2} L^2 y_{\mu_{q+2} t} + \sum_{i=2q+3}^{p} \beta_i L^{i-2q-2} D_I(L) y_t + \varepsilon_t^*, \qquad \text{(A2.2.8)}$$

with $\sum_{i=2q+3}^{p} \beta_i L^{i-2q-2}$ approximating $\{(\theta(L)-\phi(L))/(\theta(L))\}$.

Note that no lags are added to the $y_{\mu_i t}$s, the β_is are zero under the null. Using the simple null $(1-L)y_t = \varepsilon_t$, Said and Dickey (1984, Theorem 6.1) showed the limit distribution of t_{b_1} in the adequately augmented test equation to be equivalent to τ_1. They also proved that $c \cdot T b_1$ has limit distribution ν_1, with

$$c = \phi(1)^{-1}\theta(1). \qquad \text{(A2.2.9)}$$

Using the asymptotic orthogonality of the $y_{\mu_i t}$s a similar result should obtain for more general $D_I(L)$, as suggested by Hylleberg et al. (1990), Engle et al. (1991, 1993) and Beaulieu and Miron (1993). Under H_0^* c^2 is 2π times the spectral density of ε_t at frequency zero. One sees that the procedure fails if ε_t is $I(1)$. The test statistics for (at least) one unit root at one frequency have different distributions under null hypotheses with several unit roots at one frequency.

Alternatives to augmentation

Dickey, Hasza and Fuller (1984) and Said and Dickey (1985) presented easily interpretable alternatives to the augmentation procedure for the case where the errors follow an AR(MA) model of *known* order. They suggested to impose the unit roots first and get consistent estimates of the parameters of $\phi(L)$ and $\theta(L)$, e.g. first round estimates of an ARMA model for $D_I(L)y_t$. In the second step they used the following equation:

$$D_I(L)x_t = \beta_1 L x_{\mu_1 t} + \ldots + \beta_{2q+2} L^2 x_{\mu_{q+2} t} + \sum_{i=2q+3}^{p} \gamma_i L^{i-2q-2} D_I(L) y_t + \varepsilon_t^{\dagger}, \qquad \text{(A2.2.10)}$$

where

$x_t = \hat{\theta}(L)^{-1}\hat{\phi}(L)y_t$, and

$\varepsilon_t^{\dagger}$ an error term not necessarily equal to ε_t^*.

They showed that $Tb_1 \rightarrow_{\mathcal{L}} \nu_1$ in this second stage regression under H_0^*.

Another way to get rid of the nuisance parameter c was developed by Fountis and Dickey (1989). They considered the simple null hypotheses with only one unit root at the zero frequency, known order AR errors and no deterministic components, and proved (op cit., Corollary 1) that $T(\rho_n - 1) \rightarrow_{\mathcal{L}} \nu_1$, with ρ_n (the real part of) of the inverse root of the least squares estimate of the autoregressive polynomial for y_t, closest to one. Using a symmetry argument one can predict that $T(\rho_m + 1) \rightarrow_{\mathcal{L}} \nu_1$, with ρ_m the inverse root closest to -1 if y_t is $I_{\frac{1}{2}}(1)$. Finally one can conjecture that $T(\rho_0^2 - 1) \rightarrow_{\mathcal{L}} \nu_2$ with ρ_0 the inverse root closest to $e^{i2\pi\lambda_i}$, if y_t is $I_{\lambda_i}(1)$. Note that b_4 is the OLS

estimate of $1-\rho_0^2$ in the regression $(1-2\cos 2\pi\lambda_i L+L^2)y_t = b_3 y_{t-1}+b_4 y_{t-2}+\varepsilon_t$. We present some related Monte Carlo results in §7.4.5, cf. also (A2.2.15) below.

Adding deterministic terms

In practical applications the stationary alternative is not always the most interesting. H_1 can be made more general by allowing for deterministic nonstationarity. The "structural" form of the model that allows for deterministic nonstationarity under the alternative reads:

$$D_I(L)y_t^* = \beta_1 L y_{\mu_1 t}^* + \beta_2 L y_{\mu_2 t}^* + \beta_3 L y_{\mu_3 t}^* + \beta_4 L^2 y_{\mu_3 t}^* + \ldots + \beta_{2q+1} L y_{\mu_{q+2} t}^* + \beta_{2q+2} L^2 y_{\mu_{q+2} t}^* + \varepsilon_t,$$

with

$$y_t^* = y_t - g(t), \quad y_{\mu_1 t}^* = \bar{D}_{\mu_i}(L) y_t^*,$$

$$g(t) = \alpha_0 + \alpha_1(t-T/2) + \alpha_2(-1)^t + \alpha_3\cos(2\pi\lambda_1 t) + \alpha_4\sin(2\pi\lambda_1 t) + \ldots + \alpha_{2q+1}\cos(2\pi\lambda_q t) + \alpha_{2q+2}\sin(2\pi\lambda_q t). \tag{A2.2.11}$$

Under the alternative (with $\beta_i \neq 0$, $\forall i$) one has $E(y_t) = g(t)$, so that $y_t - g(t)$ is stationary. Model (A2.2.11) is not easy to estimate, because of nonlinearity in the parameters. Products of α_i–values and β_i–values occur. Nonidentifiability of the α_i–values under the null is another problem.

One can estimate and test more easily using the following "reduced form" augmented test equation (A2.2.12):

$$D_I(L)y_t = \beta_1 L y_{\mu_1 t} + \beta_2 L y_{\mu_2 t} + \beta_3 L y_{\mu_3 t} + \beta_4 L^2 y_{\mu_3 t} + \ldots + \beta_{2q+1} L y_{\mu_{q+2} t} + \beta_{2q+2} L^2 y_{\mu_{q+2} t} + \beta_{2q+3} D_I(L) L y_t + \ldots + \beta_p D_I(L) L^r y_t + \delta_0 + \delta_1(t-T/2) + \delta_2(-1)^t + \delta_3\cos(2\pi\lambda_1 t) + \delta_4\sin(2\pi\lambda_1 t) + \ldots + \delta_{2q+2}\sin(2\pi\lambda_q t) + \varepsilon_t^*. \tag{A2.2.12}$$

We discuss the interpretation of the α_i–values and δ_i–values and their relationship in §A2.2.2.3.

Unknown ARMA orders and nonnormality of error term

If the ε_t process contains an MA part (which cannot be modeled exactly by a finite order AR model), one should let the order r of augmentation increase with sample size. Said and Dickey (1984) assumed that $T^{-1/3}r \to 0$ and the existence of $s>0$, $k>0$ such that $sr > T^{1/k}$. Diebold and Nerlove (1990) suggested $r = T^{1/4}$ to be adequate.

The normality assumption on the residuals is not necessary. In this set–up the existence of the first two moments and the i.i.d. property are. Testing for unit root nonstationarity in models where the first two moments of the error term do not exist is useless in VAR analysis, since all interesting derivations are based on these assumptions. Zivot and Andrews (1992) discussed

distributions of unit root test statistics in cases where low order moments of the disturbances do not exist.

Phillips and Perron (1988) developed a test for the zero frequency unit root based on the first order autoregressive coefficient. They constructed a correction factor based on c and showed the appropriately corrected statistic to have limit distribution ν_1 under even more general assumptions on ε_t than mentioned above, which allowed for a slowly but not persistently changing conditional variance of ε_t. They used a nonparametric approach to estimate the spectral density of ε_t at frequency zero from the residuals of a first order autoregression. The finite sample performance of the test is not favorable if the MA part has roots close to one (see Schwert (1989)). See §A2.2.2.4 for explicit expressions and §7.4.7.2 for applications of estimation of c.

A2.2.2.2 Interesting null hypotheses, alternatives and tests

From now on we consider all hypotheses to be of the general type encompassed by (A2.2.12). The asterisks are deleted for simplicity. We consider tests of $I_\lambda(1)$ against $I_\lambda(0)$. Tests of $I_\lambda(i)$ against $I_\lambda(i-1)$ for $i>1$ can only be done in the same framework if one prefilters the data with $(D_\lambda(L))^{i-1}$. At each frequency there are four possibly interesting hypotheses:

1. H_{01}: unit root nonstationarity and deterministic nonstationarity,
2. H_{02}: "pure" unit root nonstationarity,
3. H_{11}: "pure" deterministic nonstationarity,
4. H_{22}: neither unit root nonstationarity nor deterministic nonstationarity.

Below we see that this framework is not appropriate to discriminate between the first two hypotheses. The parameters that could make the difference are not identified, abbreviated n.i.. Most tests are one sided. One easily derives the expected sign[6] under the alternative of stationarity in simple examples without augmentation and deterministic terms:

Example A2.1

Frequency 0:

H_0: $(1-L)y_t=\varepsilon_t$ against (A2.2.13)

H_1: $(1-\rho L)y_t=\varepsilon_t \Leftrightarrow (1-L)y_t=-(1-\rho)Ly_t+\varepsilon_t$, $0<\rho<1$,

expected sign $b_1<0$.

---

[6] Hylleberg et al. (1990) chose a parameterization with opposite signs for the regressors for frequencies $\frac{1}{2}$ and $\frac{1}{4}$.

Example A2.2

Frequency ½:

H_0: $(1+L)y_t=\varepsilon_t$ against (A2.2.14)

H_1: $(1+\rho L)y_t=\varepsilon_t \Leftrightarrow (1+L)y_t=(1-\rho)Ly_t+\varepsilon_t$, $0<\rho<1$,

expected sign $b_2>0$.

Example A2.3

Other frequencies λ:

H_0: $(1-2\cos 2\pi\lambda L+L^2)y_t=\varepsilon_t$ (A2.2.15)

H_1: $(1-2\rho\cos 2\pi\lambda L+\rho^2L^2)y_t=\varepsilon_t \Leftrightarrow$

$(1-2\cos 2\pi\lambda L+L^2)y_t=-2(1-\rho)\cos 2\pi\lambda Ly_t+(1-\rho^2)L^2y_t+\varepsilon_t$, $0<\rho<1$,

expected sign $b_{2j+2}>0$, $j=1,2,\ldots,$

expected sign $b_{2j+1}=-\text{sgn}(\cos 2\pi\lambda)$, $j=1,2,\ldots.$

The last example shows differences in the expected signs of β_{2j+1} and β_{2j+2} for the stationary alternatives for the complex unit root models. The estimate b_{2j+1} only gives extra information about (non)stationarity per se if a simple alternative like H_1 is considered, with an a priori specified relationship between β_{2j+1} and β_{2j+2}. Rejection of $\beta_{2j+1}=0$ does not in itself lead to rejection of nonstationarity per se, only to rejection of unit root nonstationarity at a specific frequency. Rejection of $\beta_{2j+2}=0$ can be directly interpreted as a rejection of unit–root nonstationarity.

Summary

Now we are able to summarize the results for the general test equation (A2.2.12). The hypotheses can be formulated as:

Frequency 0

H_{01}: $\beta_1=0,\ \delta_0\neq 0,\ \delta_1=0\ :\ (y_t-\alpha_0-\alpha_1(t-T/2))=I(1)$ and

H_{02}: $\beta_1=0,\ \delta_0=0,\ \delta_1=0\ :\ (y_t-\alpha_0)=I(1)$

against

H_{11}: $\beta_1<0,\ \delta_0\neq 0,\ \delta_1\neq 0\ :\ (y_t-\alpha_0-\alpha_1(t-T/2))=I(0)$ and

H_{12}: $\beta_1<0,\ \delta_0\neq 0,\ \delta_1=0\ :\ (y_t-\alpha_0)=I(0)$.

Frequency ½

H_{01}: $\beta_2=0,\ \delta_0\neq 0,\ \delta_2=0\ :\ (y_t-\alpha_0-\alpha_2(-1)^t)=I_{1/2}(1)$

against

H_{11}: $\beta_2>0,\ \delta_0\neq 0,\ \delta_2\neq 0\ :\ (y_t-\alpha_0-\alpha_2(-1)^t)=I_{1/2}(0)$

H_{12}: $\beta_2>0,\ \delta_0\neq 0,\ \delta_2=0\ :\ (y_t-\alpha_0)=I_{1/2}(0)$.

Frequency $\lambda_1 = ¼$

H_{01}: $\beta_3 = \beta_4 = 0, \delta_0 \neq 0, \delta_3 = 0, \delta_4 = 0$:

$$(y_t - \alpha_0 - \alpha_3 \cos(2\pi\lambda_1 t) - \alpha_4 \sin(2\pi\lambda_1 t)) = I_{\lambda_1}(1)$$

against

$$H_{11}: \beta_4 > 0, \delta_0 \neq 0, \delta_3 \neq 0, \delta_4 \neq 0: \ (y_t - \alpha_0 - \alpha_3 \cos(2\pi\lambda_1 t) - \alpha_4 \sin(2\pi\lambda_1 t)) = I_{\lambda_1}(0)$$

and

$$H_{12}: \beta_4 > 0, \delta_0 \neq 0, \delta_3 = 0, \delta_4 = 0: \ (y_t - \alpha_0) = I_{\lambda_1}(0).$$

Frequencies λ_i, $0 < \lambda_i < ¼$

H_{01}: $\beta_{1+2i} = \beta_{2+2i} = 0, \delta_0 \neq 0, \delta_{1+2i} = 0, \delta_{2+2i} = 0$:

$$(y_t - \alpha_0 - \alpha_{1+2i} \cos(2\pi\lambda_i t) - \alpha_{2+2i} \sin(2\pi\lambda_i t)) = I_{\lambda_i}(1)$$

against

H_{11}: $\beta_{1+2i} < 0, \beta_{2+2i} > 0, \delta_0 \neq 0, \delta_{1+2i} \neq 0, \delta_{2+2i} \neq 0$:

$$(y_t - \alpha_0 - \alpha_{1+2i} \cos(2\pi\lambda_i t) - \alpha_{2+2i} \sin(2\pi\lambda_i t)) = I_{\lambda_i}(0) \quad \text{and}$$

H_{12}: $\beta_{1+2i} < 0, \beta_{2+2i} > 0, \delta_0 \neq 0, \delta_{1+2i} = 0, \delta_{2+2i} = 0$: $(y_t - \alpha_0) = I_{\lambda_i}(0)$.

Frequencies λ_i, $¼ < \lambda_i < ½$

H_{01}: $\beta_{1+2i} = \beta_{2+2i} = 0, \delta_0 \neq 0, \ \delta_{1+2i} = 0, \delta_{2+2i} = 0$:

$$(y_t - \alpha_0 - \alpha_{1+2i} \cos(2\pi\lambda_i t) - \alpha_{2+2i} \sin(2\pi\lambda_i t)) = I_{\lambda_i}(1)$$

against

H_{11}: $\beta_{1+2i} > 0, \beta_{2+2i} > 0, \delta_0 \neq 0, \delta_{1+2i} \neq 0, \delta_{2+2i} \neq 0$:

$$(y_t - \alpha_0 - \alpha_{1+2i} \cos(2\pi\lambda_i t) - \alpha_{2+2i} \sin(2\pi\lambda_i t)) = I_{\lambda_i}(0) \text{ and}$$

H_{12}: $\beta_{1+2i} > 0, \beta_{2+2i} > 0, \delta_0 \neq 0, \delta_{1+2i} = 0, \delta_{2+2i} = 0$: $(y_t - \alpha_0) = I_{\lambda_i}(0)$.

We combine the tests for unit roots at the zero and ½ frequency in

Example A2.4

Frequencies 0 and ½

$$H_0: \ (1-L)(1+L)y_t = \varepsilon_t \qquad \text{(A2.2.16)}$$

$$H_1: \ (1-\rho_1 L)(1+\rho_2 L)y_t = \varepsilon_t, \ 0 < \rho_1 < 1, \ 0 < \rho_2 < 1 \Leftrightarrow$$

$$(1-L)(1+L)y_t = -½(1-\rho_1)(1+\rho_2)L(1+L)y_t + ½(1-\rho_2)(1+\rho_1)L(1-L)y_t + \varepsilon_t.$$

The expected signs of β_1 and β_2 under the alternative are the same as in the single frequency unit root case of examples A2.1 and A2.2. We conclude with the test for $D_I(L) = 1+L+L^2+L^3$, an often used moving average filter for quarterly data:

Example A2.5

Frequencies ½ and ¼

$$H_0\colon\ (1+L)(1+L^2)y_t=\varepsilon_t \tag{A2.2.17}$$

$$H_1\colon\ (1+\rho_2 L)(1+\rho_4 L^2)y_t=\varepsilon_t,\ 0<\rho_2<1,\ 0<\rho_4<1 \Leftrightarrow$$

$$(1+L)(1+L^2)y_t=\tfrac{1}{2}(1-\rho_2)(1+\rho_4)L(1+L^2)y_t+\tfrac{1}{2}(1-\rho_2)(1-\rho_4)L(1+L)y_t+$$
$$\tfrac{1}{2}(1+\rho_2)(1-\rho_4)L^2(1+L)y_t+\varepsilon_t.$$

All coefficients in the test regression are greater than zero under the alternatives considered here. Note that $\beta_3=\tfrac{1}{2}(1-\rho_2)(1-\rho_4)$ is equal to zero if either $\rho_2=1$ or $\rho_4=1$, so that the t–value of this coefficient cannot be used to test integration at frequency ¼ per se. Note also that $\beta_4=\tfrac{1}{2}(1+\rho_2)(1-\rho_4)$ is equal to zero if $\rho_2=-1$, independently of ρ_4: if one neglects this integration at the *zero* frequency, the test for integration at frequency ¼ using the t–value for β_4 will not be effective.

Multiple unit roots at a frequency

Higher order integration is only likely to be relevant at frequency zero. Dickey and Pantula (1987) provided convincing arguments for the strategy of first testing $I(i)$ against $I(i-1)$, with i the maximum order of integration considered and then testing downwards $I(i-1)$ against $I(i-2)$ and so on. This is against the practice of starting with the most general (unrestricted) model and subsequently testing more restricted models. Hasza and Fuller (1979) discussed tests for $I(2)$ against explosive and $I(0)$ alternatives as well.

A2.2.2.3 The parameters δ_i and α_i in (A2.2.11) and (A2.2.12)

Interpretation

The parameters δ_i and α_i are connected. The parameter vector $\alpha=(\alpha_0,\ldots,\alpha_{2q+2})$ determines the time dependent expectation of y_t under the alternative, which gives it a clear interpretation. The parameter vector $\delta=(\delta_0,\ldots,\delta_{2q+1})$ does not have a straightforward interpretation. One can call α a vector of "structural form" parameters and δ a vector of "reduced form" parameters. It is much easier to estimate δ than α, however, which makes the parameterization with δ instead of α useful. Under the alternative hypothesis the elements of α that determine nonstationarity at a certain frequency can be written as a function of the corresponding elements of δ and all the autoregressive parameters. The interpretations of both δ and α under the alternative do not go through under the null. See §4.2 and §4.3 for other "structural" parameterizations of time series models with unit roots. Kwiatkowski et al. (1992) applied such a parameterization to test the *null* of trend stationarity.

In our set-up we have chosen the deterministic regressors in the same way as the integrated regressors, so that each of them is nonstationary at only one frequency. In that way we can analyze the interpretation of the parameters also for one frequency at a time. It is convenient to assume no unit roots at the other frequencies first.

The parameter δ_0 determines the mean if the series is stationary. Under H_{01} at the zero frequency δ_0 determines the mean growth rate, since all other deterministic terms have mean zero by construction. Note that if the series is $I(1)$ we assume $\delta_1 = 0$. If one allows for $\delta_1 \neq 0$ if the series is $I(1)$, quadratic functions of time become part of the homogeneous solution of the equation. The parameter α_0 determines the mean of the series when it is stationary, otherwise it can only be related to the initial observations; see i.a. Said and Dickey (1985). We discuss simple starting conditions for nonstationary (seasonal) processes in §4.3.2. Here we assume that all the necessary starting values for ε_t are zero. The parameter δ_1 determines the mean rate of growth if the series is $I(0)$. For large t one has:

$$\delta_1 s = D(1)E(y_t - y_{t-s}) = D(1)\alpha_1 s,$$

with E the expectations operator and s the smallest common integer multiple of all periods $(1/\lambda_i)$. Usually $s = S$: the number of observation per year. All other deterministic terms have zero mean over time span s: $g(t) - g(t-s) = \alpha_1 s$. Note that $D(1) = D_I(1)D^*(1) \neq 0$.

The parameter δ_2 determines the size of the deterministic component at frequency ½ if the series is $I_{1/2}(0)$. It is the frequency ½ analog of δ_0 and determines the mean amplitude of the $(-1)^t$ component. One has:

$$D(L)\alpha_2(-1)^t = \delta_2(-1)^t \Rightarrow \delta_2 = \alpha_2 D(-1).$$

α_2 is not identified if y_t is $I_{1/2}(1)$. In that case $(1+L)(y_t + \alpha_2(-1)^t)$ does not depend on α_2 or δ_2: if one specifies a distribution function for the ε_ts and the initial values $y_0, \ldots, y_{p-1}$ the same likelihood results for all α_2s. The parameter α_2 is only related to the initial conditions if the series is $I_{1/2}(1)$; see the example below. It determines the deterministic frequency ½ component when the series is $I_{1/2}(0)$. One assumes $\delta_2 = 0$ under H_{01} to avoid periodic trends $t(-1)^t$ in y_t.

Explicit relationships

One can easily derive the explicit relationship between the α_is and δ_is for other frequencies as well. Let δ_{2j+1} and δ_{2j+2} be the parameters

determining the size of the deterministic component at frequency λ_j when the series is $I_{\lambda_j}(0)$. From

$$D(L)(\alpha_{2j+1}\cos 2\pi\lambda_j t+\alpha_{2j+2}\sin 2\pi\lambda_j t)=\delta_{2j+1}\cos 2\pi\lambda_j t+\delta_{2j+2}\sin 2\pi\lambda_j t \qquad \text{(A2.2.18)}$$

one gets the following relationship between $(\alpha_{2j+1}\ \alpha_{2j+2})$ and $(\delta_{2j+1}\ \delta_{2j+2})$:

$$\begin{bmatrix}\delta_{2j+1}\\ \delta_{2j+2}\end{bmatrix}=\begin{bmatrix}\mathrm{Re}(D(e^{i2\pi\lambda_j})) & -\mathrm{Im}(D(e^{i2\pi\lambda_j}))\\ \mathrm{Im}(D(e^{i2\pi\lambda_j})) & \mathrm{Re}(D(e^{i2\pi\lambda_j}))\end{bmatrix}\begin{bmatrix}\alpha_{2j+1}\\ \alpha_{2j+2}\end{bmatrix}. \qquad \text{(A2.2.19)}$$

The equality of (A2.2.19) and (A2.2.18) can be derived as follows. Use the equality $D(L)e^{i2\pi\lambda_j t}=D(e^{-i2\pi\lambda_j t})e^{i2\pi\lambda_j t}$. By equating the real and imaginary part of left and right hand side one gets the expressions for $D(L)\cos 2\pi\lambda_j t$ and $D(L)\sin 2\pi\lambda_j t$ needed to rewrite (A2.2.18) as (A2.2.19).

The determinant of the transformation matrix is the frequency response of the filter $D(L)$ at $2\pi\lambda_j$ which is nonzero by assumption. The phase of $\delta_{2j+1}\cos 2\pi\lambda_j t+\delta_{2j+2}\sin 2\pi\lambda_j t$ equals the phase of $\alpha_{2j+1}\cos 2\pi\lambda_j t+\alpha_{2j+2}\sin 2\pi\lambda_j t$ plus the principal argument of $D(e^{i2\pi\lambda_j})$, see e.g. Koopmans (1974, p. 84).

The parameters α_{2j+1} and α_{2j+2} are not identified when y_t is $I_{\lambda_j}(1)$ because $(1-2\cos 2\pi\lambda_j L+L^2)(y_t+\alpha_{2j+1}\cos 2\pi\lambda_j t+\alpha_{2j+2}\sin 2\pi\lambda_j t)$ does not depend on α_{2j+1} and α_{2j+2}. The parameters α_{2j+1} and α_{2j+2} determine the nonrandom frequency λ_j component when the series is $I_{\lambda_j}(0)$. They can only be related to the initial conditions when y_t is $I_{\lambda_j}(1)$. Formula (A2.2.19) can also be applied for $\lambda_j=0$ and $\lambda_j=\frac{1}{2}$. It provides the general solution.

Starting values and impulse responses

In order to give some insight in the interpretation of the α_i's in an augmented model we present an example for frequency $\frac{1}{2}$.

Example A2.6

We make the following simplifying assumptions: $y_0=x_0, y_1=x_1, \ldots, y_p=x_p$, $\varepsilon_t=0$, $t<0$, $x_t=0$, $t<0$ and $t>p$, $x_*=(x_1,\ldots,x_p)$ is a vector of fixed starting values. Let $x(z)$ denote the generating function of x_t: $x(z)=\sum_{-\infty}^{\infty}x_t z^t$, with z the usual complex dummy variable. The roots of $D(z)=0$ are all outside the unit circle, except one at -1, i.e.

$$D(z)=D^*(z)(1+z). \qquad \text{(A2.2.20)}$$

Define $h(t)$ as $h(t)=g(t)-\alpha_0-\alpha_2(-1)^t$,so that $h(t)$ is stationary at frequency $\frac{1}{2}$. The null hypothesis for frequency $\frac{1}{2}$

$$H_{01}: D^*(L)(1+L)(y_t-g(t))=\varepsilon_t$$

can by substitution of $1/((1+L)D^*(L)) = \sum_{j=0}^{\infty}(-L)^j/D^*(L)$ and the initial conditions be rewritten as

$$H_{01}: (y_t - \alpha_0 - \alpha_2(-1)^t) = h(t) + \sum_{j=0}^{t}(-1)^j \varepsilon_{t-j}/D^*(L) + \sum_{j=0}^{t}(-1)^{t-j} x_{t-j}/D^*(L), \text{(A2.2.21)}$$

with $t = 2,3,....$

One can rewrite the last term on the right hand side of (A2.2.21) as $x_t/((1+L)D^*(L))$, which tends to $(-1)^{t-j}x(-1)/D^*(-1)$ as $t \to \infty$. This can be shown using a partial fraction decomposition of $1/((1+L)D^*(L))$ in (A2.2.21), see e.g. Sargent (1979):

$$\frac{1}{(1+z)\prod_1^r D_i^*(z)} = \frac{1/D^*(-1)}{(1+z)} + \sum_{i=1}^{r} \frac{k_i}{(1-d_i z)}, \text{ the } k_i\text{s constants} \qquad \text{(A2.2.22)}$$

By setting α_2 in (A2.2.21) equal to $x(-1)/D^*(-1)$ one gets

$$H_{02}: (y_t - \alpha_0) = I_{1/2}(1).$$

The "deterministic part" of the nonstationarity becomes equivalent to a simple periodic transformation of the starting values. They differ only shortly after the beginning of the process and only when $D^*(L)$ is different from 1. In other words, the deterministic part of the nonstationarity can only be identified while the appropriately transformed variable, i.c. $(1+L)(y_t - h(t))$, still is nonstationary due to fixed start up conditions. Strictly speaking $(1+L)(y_t - h(t))$ is always nonstationary because of its time varying mean and (growing) variance. Diebold and Nerlove (1990) used the expression *asymptotic stationarity*. The strictly stationary case where the initial values are random is not analyzed in depth in the literature, where it is often assumed that $x(z) = 0$.

From this example one learns that the "deterministic periodic nonstationarity" at frequency ½ can be interpreted as a "byproduct" of unit root nonstationarity whenever the starting conditions satisfy $x(-1) \neq 0$. The parameterization also shows that disturbances to the $I_{1/2}(1)$ component will change the size of the periodic pattern permanently, but not its shape.

For other frequencies λ, with $0 < \lambda < ½$, one can derive equivalent properties. The corresponding partial fraction decomposition for $D(z) = D^*(z)(1 - 2\cos 2\pi\lambda z + z^2)$ reads:

$$\frac{1}{(1-e^{-2\pi\lambda i}z)(1-e^{2\pi\lambda i}z)\prod_j^r D_j^*(z)} = \frac{A+zB}{(1-2\cos 2\pi\lambda z+z^2)} + \sum_{j=1}^{r}\frac{k_j}{(1-d_j z)}, \quad \text{(A2.2.23)}$$

with

$A = \mathrm{Re}(1/(D^*(e^{2\pi\lambda i}))) - \cot(2\pi\lambda)\{\mathrm{Im}(1/(D^*(e^{2\pi\lambda i})))\}$ and

$B = \mathrm{Im}(1/(D^*(e^{2\pi\lambda i}))/\sin 2\pi\lambda$,

so that A and B reduce to $\mathrm{Re}(1/(D^*(i)))$ and $\mathrm{Im}(1/((D^*(i)))$ if $\lambda = ¼$.

One can also interpret δ_0 and δ_2 as parameters determining the (time varying) mean of the unobservable "shocks" $\eta_t = \varepsilon_t + \delta_0 + \delta_2(-1)^t$ to the system. Define $\eta(z) := \sum_1^T \eta_t z^t$. Series of shocks with $\eta(1) = T\delta_0 \neq 0$ lead to a deterministic trend in mean in y_t if $D(1) = 0$: Each shock has a permanent effect on the mean of later y_t-values, resulting in a mean trend if the mean shock is non zero. Shocks with $\eta(-1) = T\delta_2 \neq 0$ lead to a deterministic *periodic* trend $\alpha t(-1)^t$ in mean if $D(-1) = 0$: Each shock has a permanent effect on the periodic mean component $\alpha_2(-1)^t$ of later y_ts. This permanent cumulation of shocks into trends does not occur under the alternatives $D(1) < 0$ and $D(-1) > 0$.

The analysis with the partial fraction decomposition still goes through when $D^*(z) = 0$ has roots on other points of the unit circle. Exogenous shocks then have a permanent effect on all the deterministic terms with periodicities corresponding to each of the unit roots involved. In the case of multiple roots one gets "multiple" cumulation.

Multiple roots

If $D(z) = 0$ has two roots equal to 1, one time shocks lead to a trend in mean. Let x_t and $x(z)$ be defined as above. It follows that

$$(1-L)^{-2}x_t = \begin{cases} \sum_1^t jx_j & t \leq p \\ (t-p+1)\sum_1^p jx_j & t > p \end{cases}. \quad \text{(A2.2.24)}$$

The linear trend is the only deterministic term in $g(t)$ that can be regarded as an impulse response from an $I(2)$-process to a finite number of exogenous shocks. It is the "deterministic $I(2)$-process" that makes most sense for economic time series. Consider the corresponding $I_{1/2}(2)$ transformation:

$$(1+L)^{-2}x_t = \begin{cases} \sum_1^t j(-1)^j x_j & t \leq p \\ (t-p+1)(-1)^{t-p}\sum_1^p j(-1)^j x_j & t > p \end{cases}. \quad \text{(A2.2.25)}$$

Multiplicative seasonality in an additive model with trending variables may look to have this property. Instead of allowing for $I(2)$ processes one can then consider taking logs. This could also remove seasonal heteroskedasticity.

A2.2.2.4 Test statistics and distributions

When deterministic regressors are introduced in the test equation the limit distributions of the test statistics change. The regressors are no longer asymptotically uncorrelated as in the set–up of Chan and Wei, which complicates the nonstandard distribution theory. In particular one has to take account of the asymptotic correlation between the deterministic terms and the integrated regressors belonging to the same frequency. One can now extract results for a considerable number of potentially empirically relevant cases from the literature.

In §A2.2.2.2 we discussed parameter restrictions indicating unit root nonstationarity at different frequencies. The distributions of §A2.2.2.1 are only useful when both the initial values and the mean of the process under the alternative are assumed to be zero, a case which is rarely encountered in practice. It is advisable to include a constant in the regression in all other cases, since the other deterministic regressors in regression (A2.2.12) have mean zero by construction.

Why include regressors in the equation which are absent under the null?

In order to make the rejection regions of the relevant test statistics for a fixed size α independent from the initial conditions in small samples, extra regressors may have to be included, although their coefficients are zero under the null. In the Neyman–Pearson framework one calls this (desirable) property of a test *similarity*.

An even more important reason to include an extra regressor under the null is the preservation of *consistency*. This argument is also valid for large samples. The most (in)famous example is the test for H_{01} against H_{11} at the zero frequency. Although $\delta_1 = 0$ under H_0, a trend has to be included in the test equation in order to allow the probability limit of the OLS b_1 to be different from zero under the so–called "trend–stationary" alternative. It is not hard to imagine that the OLS estimate of b_1 in $\Delta y_t = b_1 y_{t-1} + \delta_0 + \varepsilon_t$ has probability limit 0 under H_{11}, so that the null will not be rejected under the alternative, even asymptotically. If one observes considerable mean growth, it does not seem wise to exclude this possibility under the alternative. Even if the alternative is stationary it is advisable to include the trend to preserve similarity with regard to the nuisance parameter δ_0 (Nankervis and Savin (1985)). See §3.3.1 for more general remarks on test selection.

Expressions of limit distributions and their notation

The limit distributions of the test statistics in test regressions with deterministic regressors are functionals of demeaned or detrended Brownian motions as tabulated in Fuller (1976, Tables 8.5.1, 8.5.2) and Dickey, Hasza

and Fuller (1984, Tables 6,7). Explicit expressions are (Dickey and Fuller (1979) and Dickey Hasza and Fuller (1984)):

$$\begin{aligned}
\nu_{\mu 1} &= -\tfrac{1}{2}(1-Q_1^2+2Q_1W_1)(\Gamma_1-W_1^2)^{-1}, \\
\tau_{\mu 1} &= -\tfrac{1}{2}(1-Q_1^2+2Q_1W_1)(\Gamma_1-W_1^2)^{-\frac{1}{2}}, \\
\nu_{\tau 1} &= -\tfrac{1}{2}(1-Q_1^2+2Q_1W_1+6Q_1V_1-12W_1V_1)(\Gamma_1-W_1^2-3V_1^2)^{-1}, \\
\tau_{\tau 1} &= -\tfrac{1}{2}(1-Q_1^2+2Q_1W_1+6Q_1V_1-12W_1V_1)(\Gamma_1-W_1^2-3V_1^2)^{-\frac{1}{2}}, \\
\nu_{\mu 2} &= -\tfrac{1}{2}(2-Q_1^2-Q_2^2+2Q_1W_2+2Q_2W_1)(\Gamma_1-W_1^2-W_2^2)^{-1}, \\
\tau_{\mu 2} &= -\tfrac{1}{2}(2-Q_1^2-Q_2^2+2Q_1W_2+2Q_2W_1)(\Gamma_1-W_1^2-W_2^2)^{-\frac{1}{2}},
\end{aligned} \tag{A2.2.26}$$

with $Q_1, Q_2, W_1, W_2, \Gamma_1, \Gamma_2, V_1, V_2$ defined above under (A2.2.4).

The subscripts 1 and 2 denote the number of independent Brownian motions that appear in the distributions. The subscripts μ and τ indicate whether these Brownian motions are demeaned or detrended (Dickey and Fuller (1979)).

Alternative tests

One can avoid the computation of c by preprocessing $D_I(L)y_t$ to make it white noise under the null. One can also use other corrections. The methods of Dickey et al. (1984) and Phillips and Perron (1988) that make the distribution of the normalized bias free of nuisance parameters can also be used for the null with real unit roots and nonzero mean or drift. Phillips and Perron (1988, p. 341) used the simple regression with $\beta_i = 0$, $i > 1$, $\delta_i = 0$, $i > 1$ (i.e. with trend), and showed $Tb_1 - c^*$ to converge to ν_1 under H_{01}. Their *additive* correction factor also involved an estimate of the spectrum of residuals $\hat{\varepsilon}_t$ at frequency zero, $2\pi \hat{f}_{\varepsilon\varepsilon}(0)$:

$$c^* = \tfrac{1}{2}(2\pi \hat{f}_{\varepsilon\varepsilon}(0) - \hat{\sigma}_\varepsilon^2)/M,$$

with

$$M = (1-T^{-2})m_{yy} - 12m_{ty}^2 + 12(1+T^{-1})m_{ty}m_y - (4+6T^{-1}+2T^{-2})m_y^2,$$

where

$$m_{yy} = T^{-2}\sum y_t^2, \quad m_{ty} = T^{-5/2}\sum t y_t, \quad m_y = T^{-3/2}\sum y_t \text{ and } \hat{\sigma}_\varepsilon^2 = (T-3)^{-1}\cdot\sum \hat{\varepsilon}_t^2.$$

They considered different estimates of $2\pi \hat{f}_{\varepsilon\varepsilon}(0)$, sometimes called the *long run variance*; see (6.54) as an example below. Dickey et al. (1984) extended the analysis also to a null with seasonal means.

Discussion of Tables

We consider the tests for one frequency at a time because of the asymptotic zero correlation across frequencies of the test regressors. This is immediately clear for the deterministic ones, whereas the zero correlation for the $y_{\mu_i t}$s can be established using the arguments in Chan and Wei (1988). Monte Carlo evidence for the finite sample distributions in the simultaneous equation with deterministic terms exists in Hylleberg et al. (1990).

Table A2.2.1 *Regression variables, test statistics, and limit distributions for unit root tests at frequency zero*

1. *Inclusion of deterministic variables in the regression*

Alternative\Null	H_{01}	H_{02}	H_{11}
H_{02}	$\delta_0,\beta_1=0$	una.	una.
H_{11}	δ_0,δ_1	δ_0,δ_1	una.
H_{12}	δ_0,δ_1	δ_0	δ_0,δ_1

2. *Useful test statistics*

Alternative\Null	H_{01}	H_{02}	H_{11}
H_{02}	t_δ	una.	una.
H_{11}	t_b , Tb_1	t_b , Tb_1	una.
H_{12}	t_{b1}, Tb_1	t_b , Tb_1	t_δ

3. *Limit distributions test statistics under alternative\null*

Alternative\Null	H_{01}	H_{02}	H_{11}
H_{02}	$t(\rho)\backslash t(0)$	una.	una.
H_{11}	$t(\rho)$, deg.$\backslash\tau_{\tau 1}, cv_{\tau 1}$	$t(\rho,1)$, deg.$\backslash\tau_{\tau 1}, cv_{\tau 1}$	una.
H_{12}	$t(\rho)$, deg.$\backslash\tau_{\tau 1}, cv_{\tau 1}$	$t(\rho,1)$, deg.$\backslash\tau_{\mu 1}, cv_{\mu 1}$	$t(\rho)\backslash t(0)$

NOTE: Source: Dickey and Fuller (1979).

Table A2.2.2 *Regression variables, test statistics, and limit distributions for unit root tests at frequency* 1/2

1. *Inclusion of deterministic variables in the regression*

Alternative\Null	H_{01}	H_{11}
H_{11}	δ_0,δ_2	una.
H_{12}	δ_0	δ_0,δ_1

2. *Useful test statistics in regression*

Alternative\Null	H_{01}	H_{11}
H_{11}	t_b, Tb_2	una.
H_{12}	t_b, Tb_2	t_δ

3. *Limit distributions test statistics under alternative\null*

Alternative\Null	H_{01}	H_{11}
H_{11}	$t(\rho)$, deg.$\backslash -\tau_{\mu 1}, -cv_1$	una.
H_{12}	$t(\rho)$, deg.$\backslash -\tau_1, -cv_1$	$t(\rho)\backslash t(0)$

NOTE Source: Dickey and Fuller(1979): $H_{01}\backslash H_{12}$, Hylleberg et al.(1990): $H_{01}\backslash H_{11}$

Table A2.2.3 *Regression variables, test statistics, and limit distributions for unit root tests at frequency* λ *,* $0<\lambda<1/2$

1. *Inclusion of deterministic variables in the regression*

	Null	
Alternative\Null	H_{01}	H_{11}
H_{11}	$\delta_0, \delta_{2j+1}, \delta_{2j+2}$	una.
H_{12}	δ_0	$\delta_0, \delta_{2j+1}, \delta_{2j+2}$

2. *Useful test statistics in regression*

	Null	
Alternative\Null	H_{01}	H_{11}
H_{11}	t_b, Tb_{2j+2}	una.
H_{12}	t_b, Tb_{2j+2}	t_δ, t_δ

3. *Limit distributions test statistics under alternative\null*

	Null	
Alternative\Null	H_{01}	H_{11}
H_{11}	$t(\rho), \text{deg.}\backslash -\tau_{\mu 2}, -c\nu_2$	una.
H_{12}	$t(\rho), \text{deg.}\backslash -\tau_2, -c\nu_2$	$t(\rho)\backslash t(0)$

NOTE: Source: Hylleberg et al. (1990) for $\lambda=\frac{1}{4}$. Freshly compiled by Monte Carlo for $\lambda=\frac{1}{3}$ and $\lambda=1/6$.

Each table consists of three parts. The first gives the set of frequency specific deterministic regressors to include in the test regression, denoted by their coefficient in (A2.2.12). Unavailable tests are indicated by una. A test in this context is not available if the null is not a restriction of the alternative. The reverse strategy to take the stationary model as the null and the unit root model as the alternative is considered elsewhere, cf. (2.30), §4.3.3. The second part of each table indicates the relevant normalized bias and regression t–statistics.

The third part of each table gives the limit distributions under null and alternative. We discussed the expected sign under the alternative above. In the tables we denote a noncentral t–distribution by $t(\rho)$, the central one by $t(0)$[7]. The parameter c is correction factor which is equal to one if there are only nonstationary regressors under the null, see e.g. (A2.2.9). It is a nuisance parameter for the "normalized bias" tests which use Tb_j. As far as we

[7] Converging to a normal distribution as the number of degrees of freedom tends to infinity.

know this correction factor has only been derived for the zero frequency. deg. denotes degenerate distributions, diverging to plus or minus infinity.

A2.2.2.5 Evaluation of methods

Critical values at nonstandard frequencies

Since the nonstandard distributions in the last table have not been established formally in the literature as far as we know, we conducted a small scale Monte Carlo Experiment (1000 replications of 100 observations) to check the results for $\lambda = 1/3$ and $\lambda = 1/6$ which are frequencies of interest for monthly data. For $\lambda = 1/3$ the Data Generating Process was $(1+L+L^2)y_t = \varepsilon_t$ with fixed zero starting values and ε_t normally distributed, and the test regression ε_t on y_{t-1}, y_{t-2}, a constant and the frequency specific terms $\cos(t\pi/2)$ and $\sin(t\pi/2)$. We generated the ε_t–values using the Box–Muller (1958) transform, see e.g. Knuth (1969), on random drawings from a uniform distribution in Turbo Pascal (1988). The last two regressors can be replaced by seasonal dummies as in Dickey et al. (1984). The results for the regression t–statistic of y_{t-2} corresponded closely to those for $-\tau_{\mu 2}$ in Table 7 in Dickey et al. (1984) reported below. For $\lambda = 1/6$ the DGP was $(1-L+L^2)y_t = \varepsilon_t$ and the test regression ε_t on y_{t-1}, y_{t-2}, a constant and the regressors $\cos(t\pi/3)$ and $\sin(t\pi/3)$. The same quantiles for the t–statistic for y_{t-2} emerged to a close approximation.

Corrections for stationary dynamics

In a second experiment we generated one thousand replications of a sample of 200 observations of the model $(1+L^2)(1+\rho L)y_t = \varepsilon_t$ to check the finite sample performance of the correction procedures for the serial correlation in the error term. We used the following values for ρ : 0.5, –0.5, 0.8, –0.8. Only the last 100 observations of each replication were used for the unit root tests. The two step procedure of Dickey, Hasza and Fuller (1984) as in equation (A2.2.10) seemed to work well. The critical values of $\tau_{\mu 2}$ were mimicked quite closely.

The straightforward application of t_{b4} in the augmented regression

$$(1+L^2)y_t = \beta_3 y_{t-1} + \beta_4 y_{t-2} + \beta_5 L(1+L^2)y_t + \delta_0 + \delta_3 \cos(t(2\pi/4)) + \delta_4 \sin(t(2\pi/4) + \varepsilon_t$$

did not produce the same critical values. Moreover, they did seem to depend on the value of the nuisance parameter ρ. This indicates that it helps to estimate the nuisance parameter of the "stationary part" of the dynamics under the null if the goal is to control the significance level of the test as best as possible. This extends the same findings in the Monte Carlo experiments of Schwert (1989) for the zero frequency unit root tests.

Percentiles of the nonstandard distributions are given in table A2.2.4.

Haldrup and Hylleberg (1991) gave a list of references with more detailed tables for zero frequency tests.

Dickey, Bell and Hillmer (1986) noted that Tb_1 "in higher order and mixed models is not as faithful to its nominal significance level as" t_{b_1}. This may lead one to suspect the same problems for the test of Fountis and Dickey, based on the roots of the AR polynomial closest to the unit circle. For classical testing purposes the t-statistics are therefore preferable, although they are not very reliable either for models with roots in the MA polynomial of the error process close to the unit circle.

Table A2.2.4 *Percentiles for nonstandard distributions*

	Probability of a Smaller Value					
	Sample size 100			*Limit distribution*		
Distribution	0.01	0.05	0.10	0.01	0.05	0.10
τ_1	−2.60	−1.95	−1.61	−2.58	−1.95	−1.62
τ_2	−2.58	−1.92	−1.59	−2.55	−1.92	−1.60
$\tau_{\mu 1}$	−3.51	−2.89	−2.58	−3.43	−2.86	−2.57
$\tau_{\tau 1}$	−3.99	−3.42	−3.13	−3.96	−3.41	−3.12
$\tau_{\mu 2}$	−4.00	−3.38	−3.07	−3.90	−3.34	−3.04
ν_1	−13.3	−7.9	− 5.6	−13.8	−8.1	−5.7
ν_2	−13.6	−8.2	− 5.9	−14.1	−8.5	−6.1
$\nu_{\mu 1}$	−16.3	−13.7	−11.0	−16.9	−14.1	−11.3
$\nu_{\tau 1}$	−27.4	−20.7	−17.5	−29.5	−21.8	−18.3
$\nu_{\mu 2}$	−24.3	−18.0	−15.0	−26.1	−19.0	−15.8

NOTE: See (A2.2.4), (A2.2.5), (A2.2.26) for definitions, §A2.2.4 for sources.

A2.2.2.6 Other approaches and some extensions

Reverse regressions

Two interesting related unit root tests have been suggested, which explore the property of stationary time series that correlation at lag i is equal to correlation at lag $-i$. We only discuss their simplest form. Augmentation and extension with deterministic variables can be done as in (A2.2.12). Dickey, Hasza and Fuller (1984) considered pooling information from regressions $y_t = (1+\beta_1)y_{t-1} + \delta_0 + \varepsilon_t$ and $y_{t-1} = (1+\beta_1^*)y_t + \delta_0^* + \varepsilon_t^*$ and derived the distribution of a "symmetric" estimator under the null. Johansen (1991a) used the multiple correlation coefficient R^2 of the regression of y_{t-1} on Δy_t and a constant, which can be interpreted as the product (or squared geometric mean) of the OLS estimate b_1 and the estimate b_1^* from the regression $y_{t-1} = b_1^* \Delta y_t + \delta_0^* + \varepsilon_t^*$. Simple calculation gives $R^2 = t_{b_1}^2/(t_{b_1}^2 + (T-2))$, so that $TR^2 \to_{\mathcal{L}} \tau_{\mu 1}^2$ under H_{02}.

The symmetric approach in Dickey et al. (1984) is no improvement over the ordinary method as they indicated in their article. The likelihood ratio approach of Johansen has the disadvantage that the information about the sign of the t–value is lost. The tests on $\beta_1 = 0$ are therefore two–sided. The (canonical) *correlation* approach has advantages over regression approaches if one extends the univariate unit root tests to multivariate unit root tests, however, see Chapter 6.

Cautionary notes on the assumption of parameter stability

All the analysis in this appendix is based on the assumption of parameter stability. If one extends the model to allow for (several one time) changes in the parameters, the unit root analysis changes fundamentally. New test statistics have to be used, which often have different distributions under the null (e.g. functions of Brownian motions adjusted for level–shifts at a certain point λT, see also appendix A6.1). Similarity and consistency of the test procedures have to be examined in the extended model as well.

In practice dummy variables with nonzero mean are sometimes added to test regressions to model a level shift under the null. In order to preserve consistency of the test one then has to introduce a step dummy which remains 1 after time t, to allow for a level shift under the alternative as well. This results in a change in the distribution of the test statistics. One can find a number of results in JBES (1992). If the timing of a one time level shift is not known a priori, different tests for unit root nonstationarity in both subsamples exist. Banerjee et al. (1992), Zivot and Andrews (1992) and Perron and Vogelsang (1991, 1992a, 1992b) elaborated on this point.

Tests on absence of deterministic terms

We did not list regression F–test statistics in the tables, like the test for $\beta_1 = 0$ *and* $\delta_1 = 0$ to test H_{01} against H_{11}, denoted by $\hat{\phi}_3$ in Dickey and Fuller (1981). The use of this extra restriction does not improve the power of the test. Haldrup (1990) listed distributions of test statistics for the nullity of the coefficients of deterministic terms. These are not easy to apply in unit root testing. Often they do not convey extra information, see e.g. Perron (1990b). A Monte Carlo experiment in §7.4.5 demonstrates that tests for the statistical significance of deterministic variables have nonstandard distributions in the presence of unit root dynamics in the stochastic part of the model. The asymptotic correlation of deterministic and integrated regressors works both ways: Introduction of deterministic terms influences the null distribution of the unit root test statistics. Introduction of integrated regressors influences the null distribution of the test statistics for the

absence of deterministic terms. The asymptotic zero correlation between the nonstationary regressors *at different frequencies* remains useful. It seems to apply to moderately sized samples as well.

3 DATA ANALYSIS BY VECTOR AUTOREGRESSION

3.1 Introduction

In the previous chapter we stressed the importance of preliminary univariate analysis of the data. This chapter deals with the multivariate analysis of the data information about "unrestricted" linear time series relationships between sets of variables of interest. "Unrestricted" should not be taken too literally. Some a priori restrictions on the "true" shape of the multivariate autocorrelation function should be appropriate in order to get some degree of precision for the analysis. At this stage we only consider so–called smoothness restrictions (in particular on the graphs of the impulse response representation). Decreasing the maximum order of the VAR can have such a smoothing effect. This order is determined by the number of observations and the number of observations per year. For quarterly data a natural choice for the minimum a priori lag length is 4. A choice for a higher a priori order depends on the number of variables of interest compared with the number of observations. Other restrictions are discussed in chapter 6.

The main aim of this type of multivariate time series analysis is the identification of disturbances that account for an interestingly big part of the variation of the variables of interest. Identification requires that the observable variables are related to these disturbances through invertible functions. The main tools for the analysis are the *impulse responses* of *orthogonalized innovations*, and *forecast error variance decompositions*. Sims (1980) introduced this procedure in macroeconometrics. It is also called *innovation accounting*, see Judge et al. (1988).

We first discuss methods to find out where the information that determines the results of the procedure is concentrated in the data. For stationary time series this information should be distributed quite evenly across the observations. The application of these methods forms the *influence analysis* (§3.2). Subsequently we address the question of misspecification (§3.3) by examining whether a sufficiently big part of the variation in the properly transformed variables can be analyzed by the VAR model (2.1).

3.2 Data-oriented measures of influence

This section addresses the detection of (components of) observations that have an important influence on the estimates of parameters of interest in our

analysis. In the introductory chapter (section 1.4) we expressed our preference for the least–squares estimator for these parameters on the grounds of robustness with regard to the integer order of integration of the process governing the outcomes of the variables. A second reason is its efficiency relative to other linear estimators for sample sizes often met in practice. Influence analysis for the least–squares estimates is well developed. It is not interesting to examine the influence of an observation per se. One has to state the primary goal of the analysis to say whether observations are influential. Johnson and Geisser (1985) did. They pointed out that "the relative influence of observations can differ widely depending on whether the focus is estimation or prediction." We employ their influence measures.

3.2.1 Goal of the influence analysis

An advantage of the decision–theoretic aspect of Bayesian analyses is the formulation a *loss function.* This function is defined on a parameter space "that reflects the possible states of nature relative to the unknown parameter vector β, and a set of all possible decisions or actions. The loss incurred depends on the outcome of an observable random variable y through a function f used to assign an action for a given y," (Judge et al. (1988, p. 808)). A comparable "classical" terminology exists in statistical decision theory, see Amemiya (1985, p. 46).

The formulation of both the parameter space and the action space are equally crucial in a statistical analysis. For our problem the parameter space is in principle infinitely dimensional. We restrict the number of variables and the number of lags for efficiency reasons, not because we think the distribution of all possible states of the economic variables can be represented by a finite parameter model.

The range of the loss function is one dimensional, so that outcomes can be ordered, and stochastic because of its dependence on a random variable. Usually it is an increasing function of a measure of the distance between the "true", but unknown value of the parameter vector and its estimate. The loss function formally reflects the focus of the analysis. Estimation errors for the parameters of interest are weighted heavily.

The decision maker minimizes expected loss by adjusting his decision rule. In economic theory agents are often assumed to behave as such decision makers. Econometric analysts do not often follow this scheme explicitly. Common practice of estimation of long run parameters in dynamic econometric models forms a good example; see Bewley and Fiebig (1990). One can "explain" this suboptimal behavior by assuming a different loss function. Efficient

estimation of the parameters of interest is not always the primary aim of applied econometric work.

The choice for a certain estimator is one of the crucial aspects of the decision rule. Oftentimes it is hard to make this rule explicit. Although we opt for the OLS estimator, we suspect that model (2.1) is not able to account appropriately for all possible states of our variables. One of the aims of misspecification testing discussed in §3.3 is to find out whether this indeed is so and to indicate ways to proceed.

The information we are after are impulse responses and variance decompositions. Reliable information helps us to decrease our loss: the shape of the impulse responses can be determined more precisely and the variance decompositions become more accurate as well. How can we track down the observations that convey most of this information? A straightforward idea is to compare parameter estimates with different subsamples left out; see e.g. Bruce and Martin (1989). This is a tedious exercise for dynamic models with as many parameters as the VAR. The parameters do not have a direct interpretation, furthermore there are usually too many of them to get a clear picture anyhow.

We suggest to study the influence of the observations separately for the different stages of the estimation we discuss below. In this way one is able to predict to some extent what happens to the final results of the analysis, whenever small modifications to the model and corresponding estimation procedure are made. More importantly, it makes the relation between the observations and the results more transparent.

In this study we basically use three estimation methods. The relation between estimates and (specific parts of) the data set in these methods is comparatively simple. The methods are therefore well suited for a data based approach. First we use (recursive) regression, where structure is imposed by the timing of the observations and the apparent exogeneity of deterministic regressors. Secondly we use (canonical) correlation analysis on residuals of the regression procedure. Thirdly we use some (nonparametric) essentially *nonadaptive*, i.e. not sample specific, smoothing procedures to remove less interesting variation that cannot be explained by the VAR.

3.2.2 Influence measures in regression

Our primary estimation method is (multivariate) regression, which is the optimal procedure for the normal linear model. A large literature on influence statistics for the normal linear model exists; see e.g. Cook and Weisberg (1982, chapters 3 and 4). Bayesian methods for sensitivity analysis also

provide useful statistics; see e.g. Polasek (1986).

Furthermore there is a vast literature on the subject of outliers, which has possible interest for this subject; see e.g. Barnett and Lewis (1984). The concepts of *outlier* and *influential observation* do not coincide. An outlier, or "surprising observation", may substantially alter an estimate of parameters of interest, but this is not necessary. Outliers are interesting in their own right, be they influential or not, see chapter 5.

Although influence analysis and outlier identification are two different things, the choice for a particular type of influence analysis is often based on a model for outliers. This outlier model reflects *prior concern* about the adequacy of the basic model. Leave–k–out statistics as in Bruce and Martin (1989) are based on the *gross–error type model.* This model states that there is a small probability that some (clusters) of observations have nothing to do with our basic model, e.g. due to registration failures. See also the section on assessment of influence measures below. More subtle ways of analyzing influence exist, which seem more appropriate for macroeconomic time series data.

Local influence analysis

Local influence analysis is one of the more delicate approaches. It examines the effect of a small perturbation of a data point, rather than its total deletion. It is a special form of *local sensitivity analysis*, advocated by Leamer (1978). One evaluates the stability of estimators with respect to small changes in the assumptions of the underlying model. Polasek (1984) presented a Bayesian interpretation.

We illustrate the concept of local influence by the example in Cook (1986). Consider the standard (one equation) linear model:

$$y = X\beta + \varepsilon, \tag{3.1}$$

where y is a T vector of observations of the endogenous variable, X is a $T \times k$ matrix of observations of nonstochastic explanatory variables, β is a k vector of parameters and the ε_i's are i.i.d. $N(0,\sigma^2)$, $i = 1,2,\ldots T$.

In the discussion we refer to our application of this analysis in the following obvious way, where one row of (3.1) is seen as an observation of one row of (2.1), where the latter equation is extended with deterministic regressors in the estimation phase. One row of (2.1) has np regressors.

Cook defined the *weighted influence statistic* as

$$D(w_i) = \| \hat{y} - \hat{y}_{w_i} \| / k\sigma^2, \tag{3.2}$$

where $\hat{y}_{w_i}$ is the vector of fitted values when the i'th observation has weight w_i in a weighted least squares procedure and the remaining observations (called *cases* by Cook) have weight one. It is clear that

$$\lim_{w_i \to 0} D(w_i) = D_i, \tag{3.3}$$

where D_i is the case deletion influence measure[1] proposed by Cook (1977), also known as Cook's distance. D_i is a *global influence statistic*, which measures (in the Euclidean metric) the displacement of the fitted values if we delete one observation i from the data set. The reason for the calculation of this measure may be the concern for gross errors in the data.

Cook also proposed the corresponding *local influence measure*

$$\partial D(w_i)/\partial w_i \big|_{w_i=1}. \tag{3.4}$$

It measures how a prediction would be affected by a slight perturbation of a case weight, or, from the sensitivity analysis point of view, a slight relaxation of the homoskedasticity assumption.

Why should one calculate this statistic? One can be concerned that a certain limited number of observations are not entirely specified by the basic model, but by a mixture of the basic model like (2.1) per se, and a disturbing model like (2.2). These marginal supplementary disturbances can have a substantial influence on functions of interest of the parameters in (2.1) if they are not taken into account appropriately. They may change the position (*location component*) and size (*covariation component*) of confidence intervals for estimates of impulse responses, values of test statistics, etc.

One can write the case weights w_i as a vector $w \in \mathbb{R}^T$ and compute the vector $\ell_{max} \in \mathbb{R}^T$ which indicates how to change w (i.e. how to perturb the postulated model) to obtain the greatest *local likelihood displacement*. §A3.1.2 contains explicit formulae for ℓ_{max}. The likelihood displacement is defined as follows: Let $L(\theta)$ be the log likelihood corresponding to a postulated model where θ is a k vector of unknown parameters. Let $\omega \in \Omega \subset \mathbb{R}^q$ be a q vector of perturbations (e.g. case weight perturbations) and let $L(\theta|\omega)$ be the log likelihood corresponding to the perturbed model given ω and let $\hat{\theta}$, $\hat{\theta}_\omega$ be the respective ML estimators. The likelihood displacement $LD(\omega)$ is under some regularity conditions defined as:

---

[1] Note $D_i = (\hat{\beta} - \hat{\beta}_{(i)})'(X'X)(\hat{\beta} - \hat{\beta}_{(i)})/[k\sigma^2]$, with $\hat{\beta}$ and $\hat{\beta}_{(i)}$ the OLS estimates using the full and the case deleted sample. D_i can be seen to measure the *displacement* of $\hat{\beta}$.

$$LD(\omega) = 2[L(\hat{\theta}) - L(\hat{\theta}_\omega)]. \tag{3.5}$$

One can choose θ according to the goal of the influence analysis. If one is only interested in β in (3.1), take $\theta = \beta$ (σ^2 known), if one is also interested in σ^2 take $\theta = (\beta', \sigma^2)$. In the context of (2.1)–(2.3) the impulse responses are mainly determined by β. For a lucid geometric interpretation of local influence introducing influence graphs and *normal curvatures* we refer to Cook (1986).

To get an overall measure of local problems for the estimation of β Cook introduced C_{max}, the *maximum* (*normal*) *curvature* (i.e. normal curvature in the direction ℓ_{max}). For simple mean (θ) estimation in random samples from a normal distribution C_{max} equals 2, which does not depend on the data. Our experience does not indicate (in disagreement with Cook's) that $C_{max} = 2$ is a useful rough guide when perturbing case weights in dynamic linear models like ours. Some simulations showed that values up to 5 do not have to be rejected as unlikely under the null hypothesis. C_{max} can also be computed if (β', σ^2) is the parameter of interest. Here values up to 10 do not seem to be unlikely. We advise to follow Cook's proposal to inspect ℓ_{max} regardless of the size of C_{max} since ℓ_{max} can identify more global problems that are not manifest locally.

Although the analysis does not use the serial dependence of the observations explicitly, the use of ℓ_{max} has the advantage that patches of influential observation can be spotted immediately, without having to estimate influence measures for ranges of multiple cases of different size, as is done for leave–k–out diagnostics.

It appears (Farebrother (1986)) that C_{max} is closely related to Breusch and Pagan's (1979) LM test statistic for heteroskedasticity, or Score test statistic (Cook and Weisberg (1983)). This points to another advantage of influence analysis. It can help the interpretation of outcomes of diagnostic tests, see remark A3.12 in appendix A3.2 for another example.

Global influence analysis

Apart from the local influence statistics we also recommend to compute the more widely known case and group (multiple observation) deletion measures. In econometrics one frequently applies the Studentized residual (Krasker et al. (1983)) and the predictive Chow (1960, §4) test. Crucial elements in all the influence measures for the normal linear model are the estimated residuals and the projection matrix, also known as *hat matrix*: $X(X'X)^{-1}X$. Observations with high diagonal elements for the hat matrix lie (in some metric) far removed from the center of the data; see e.g. Krasker et al. (1983, p. 662).

Most influence statistics can be computed at low cost once one has these crucial elements. Cook and Weisberg (1982, p. 30) suggested algorithms. The equations of an unrestricted VAR share the same set of regressors and thus the same hat matrix. Formulae for the measures are given in appendix 3.1.

Assessment of influence measures

It helps the interpretation of the outcomes of the analysis to put influence measures in a two–way table representing different combinations of goal of analysis and prior concern about the adequacy of the model used, see e.g. Cook and Weisberg (1982, p. 114). In the literature on robust statistics this prior concern is often formulated as a *perturbation* of the basic model. Influence analysis helps to assess the *stability* of the estimates, Barnett and Lewis (1984, p. 60).

Table 3.1 *Ordering of some regression influence measures*

	Prior Concern	
Goal of analysis	*possibility of gross errors*	*local deviations from homoskedasticity*
location of β	$DFBETA_i$	EIC_i
(co)variation of β	PF_i	$\ell_{max}(\sigma^2)_i$

Table 3.1 serves as an example. Suppose first that the primary interest of the analysis is in terms of β in (3.1) and that the prior concern is of the *gross–error type*. Krasker et al. (1983) formulated this concern. The data generating process may break down with positive probability η and produce an observation identically equal to some fixed (X_0, y_0), which has nothing to do with the hypothesized process of interest. In that situation one should pay most attention to the following *case deletion measure*:

$$DFBETA_i = \hat{\beta} - \hat{\beta}_{(i)} = (X'X)^{-1}x_i'e_i/(1-h_i), \tag{3.6}$$

where

e_i is the OLS residual of the i–th observation,

h_i is the i–th diagonal element of the hat matrix $H = X(X'X)^{-1}X'$,

x_i is the i–th row of X,

$\hat{\beta}$ is the OLS estimate for β using all observations, and

$\hat{\beta}_{(i)}$ is the OLS estimate for β with the i–th observation discarded.

The series $(T-1)DFBETA_i$ $(i=1,\ldots,T)$ is also known as the *sample influence curve* SIC_i which measures changes in $\hat{\beta}$. If we are only concerned that some

observation might be disturbed somewhat more than others the *case weight disturbing measure* should be more interesting, known as the *empirical influence curve* EIC_i (Cook and Weisberg, p. 110):

$$EIC_i = T\partial\hat{\beta}(w_i)/\partial w_i|_{w_i=1} = T(X'X)^{-1}x_i'e_i, \tag{3.7}$$

where

$\hat{\beta}(w_i) = (X'W_iX)^{-1}X'W_iy$, and

$W_i = \text{diag}(1,\ldots,1,w_i,1,\ldots,1)$, a diagonal matrix weighting case i.

Comparing (3.6) and (3.7) one sees that the hat matrix is less important if our concern is only in terms of a slight deviation in variance of a disturbance[2].

The Chow (1960, form. (13)) statistic for predictive failure in (3.8), here denoted PF_i can be compared with $DFBETA_i$ in (3.6), using lemma A3.2 from the appendix:

$$PF_i = (e_i^2/\hat{\sigma}_{(i)}^2)/(1-h_i), \tag{3.8}$$

where $\hat{\sigma}_{(i)}^2$ is the estimate for σ^2 with the i–th observation deleted. Both measures are meaningful under the gross–error prior concern. It is clear from (3.6) and (3.8) that the residuals play a more important (squared and standardized) role in PF_i than in $DFBETA_i$. PF_i focuses more on the constancy of σ^2 than on the constancy of β as is shown in Chow (1960). Therefore we are inclined to put it in the bottom row of table 3.1[3]. Note that replacing $\hat{\sigma}_{(i)}^2$ by $\hat{\sigma}^2$ in (3.8) does not change the influence order of the observations, see remark A3.1 in appendix 3.1. This Chow statistic is known in the statistics literature as a "multiple case analog of the externally Studentized residual", see Cook and Weisberg (1984, p. 30). It can also be used as a formal misspecification test for the basic model against a specific alternative if one has strong a priori idea on the timing of the break down point for the process of interest.

We also advise to calculate the (multiple case) deletion measures although they seem to be in conflict with our prior concern. First we want all the easily available data information we can get. We cannot assess its

---

[2] Cook and Weisberg (1982) also discussed $EIC_{(i)} = T\partial\hat{\beta}(w_i)/\partial w_i|_0 = T(X'X)^{-1}x_i'e_i/(1-h_i)^2$, where the basic model is a gross error at observation i and local influence of introducing observation i is computed. $DFBETA_i$ can be viewed as a compromise between EIC_i and $EIC_{(i)}$.

[3] The influence measure $DFFITS_i = ((h_i/(1-h_i))PF_i)^{\frac{1}{2}}$ has an intermediate position.

usefulness fully before we have tried. Furthermore there are reasons to believe that their outcomes can be helpful in the analysis, despite the presence of lagged dependent variables among the regressors, which does not fit the original assumptions under which they were derived. In contrast to the local influence statistics the global ones like the Chow statistic can be used to examine the influence of subsets of observations.

We compute the Chow statistics PF_i for series of nonoverlapping subsets of observations. We also compute corresponding "group deletion" measures called location component and covariation component in Johnson and Geisser (1983). They compared these two components by their contributions in a Kullback–Leibler divergence measure for predictive densities: "In this way subsets that affect mean vectors may be distinguished from those that affect covariance structures." They noted that it is not appropriate to employ asymptotic results for small sample sizes if one compares these two components.

3.2.3 Influence measures for dynamic and multiple equation models

Influence analysis becomes more interesting if one takes account of the dynamic properties of the model. Consider the simple univariate AR(1) model as an example:

$$y_t = \beta y_{t-1} + \varepsilon_t, \tag{3.9}$$

with

y_0 fixed, and ε_t i.i.d. $N(0,\sigma^2)$ $t = 1,\ldots,T$.

Prior concern about the homogeneity of the sample can take the serial dependence of the regressand and the regressor into account. Suppose one fears y_t to be drawn occasionally, say at period t', from another distribution than that specified by the simple model, which may entail that observation t' has to be treated separately. It seems reasonable to suppose that y_{t-1} as explanatory variable in the next period also might deserve special attention then.

Peña (1990) analyzed influence measures for univariate time series models. He discussed two global influence measures of the dynamic parameters of ARMA models, based on prior concern for two outlier types, the additive outlier and the innovation outlier. We discuss these outlier models in in more detail in chapter 5. Peña showed that global case deletion measures like Cook's distance can be interpreted as a reflection of the ***innovation outlier concern.***

In applied econometrics the introduction of occasional dummy variables

with value 1 at $t=i$ and zero elsewhere is rife. In static models these dummies have the effect of the deletion of the observation. In dynamic models the dummies can be interpreted as innovation outliers. Global analysis based on the innovation outlier concern examines the effect of introducing such dummy variables in models like (3.9) on the parameter estimates for β and/or σ^2.

The global influence analysis with *additive outlier concern* examines the effect of removing an observation y_i from the data set. In its most straightforward form one scrutinizes the effect of deleting both observations (y_{i-1}, y_i) and (y_i, y_{i+1}) on the parameter estimates in (3.9). Rather than deleting both observations one can use estimation procedures for models with missing observations and compare the normal estimates with those when y_i is treated as missing. Harvey (1989, §6.4) discussed some procedures and references. These procedures involve replacing the observation that is treated as missing by a prediction using all the observations t with $t<i$ and $t>i$. Instead of deleting (y_{i-1}, y_i) and (y_i, y_{i+1}), one then uses $(y_{i-1}, \hat{y}_i)$ and $(\hat{y}_i, y_{i+1})$, where $\hat{y}_i$ denotes the optimal forecast for y_i when it is treated as missing.

Introducing a dummy variable boils down to replacing only (y_{i-1}, y_i) by $(y_{i-1}, \hat{y}_i)$, with $\hat{y}_i$ the optimal forecast with y_i treated as affected by an innovation outlier.

Peña (1990) preferred influence measures based on the additive outlier concern. These measures can only be computed iteratively (and this for each set of observations of interest!). They are thus not so useful in the first stage of unrestricted data analysis. We agree however that additive outliers may present more serious problems for the estimation of dynamic models. In chapter 5 we introduce a related influence analysis and an estimation procedure for additive outliers.

Earlier Peña (1986) indicated that local influence analysis for regression models may be more appropriate in dynamic models. Peña only examined the influence of one observation at a time. Abraham and Chuang (1989) discussed the use of the multivariate internally studentized residual to study the influence of sets of observations in dynamic models[4].

Multivariate influence measures

Apart from being dynamic, the VAR model is also essentially multivariate, and we would like to take this into account in the influence analysis. Johnson and Geisser (1985) generalized their Bayesian influence analysis on the influence of subsets of observations on the estimates of β and σ^2 under a

[4] See appendix A3.2, remark A3.14.

gross–error concern to the multivariate general linear model (MGLM). We adapt their notation as follows:

$$Y = XB + E, \tag{3.10}$$

with

X a $T \times k$ matrix of observed predetermined variables,

Y a $T \times n$ matrix of observed dependent variables,

B a $k \times n$ matrix of unknown regression coefficients,

E a $T \times n$ matrix of disturbances, so that

$E = [E_1, \ldots, E_t, \ldots, E_T]'$, with E_t an $n \times 1$ vector, which is i.i.d. $N_n(0, \Sigma)$.

This analysis is especially interesting for the estimate of Σ. Decomposed versions of that matrix play a crucial role in forecast error variance decompositions. This analysis cannot be done equation by equation, whereas this is possible for the estimation of B.

Whereas the single equation analysis pointed to functions of transformed residuals (see appendix 3.1) as indicators of influence for subsets of observations, the multi equation analysis produces analogous functions of matrices of transformed residuals. The Lawley–Hotelling trace criterion is a well known example; see e.g. Spanos (1986, p. 577). In chapter 5 we show this criterion to be asymptotically equivalent with an LM test statistic for a multivariate innovation outlier at a certain known point in time.

All the influence measures derived by Johnson and Geisser solely depend on four matrices, which are easy to compute for interesting subsets of observations. These matrices are multivariate generalizations of the internally Studentized residuals, t_i, the diagonal elements of the hat matrix, h_i, Cook's Distance (3.3) measures using $(X'X)$ as a metric to measure the displacement of $\hat{\beta}$, $D_i(X'X)$, and lastly the equivalent distance measures with $(X_{(i)}'X_{(i)})$ used as a metric.

This analysis can bring influential observations for the crucial correlations between the disturbances of the different equations to the fore. Explicit formulas for the relevant measures are given in appendix 3.2.

Concluding remarks

In the influence analysis of regression equations the structure of the model helps to identify important observations. Influence analysis of dynamic models exploits the additional time series properties of the data, which may enable us to classify these observations into different categories of interest (Chapter 5). Although relatively little has been published in the context of multiple equation models, a considerable range of easily computable measures

is now available[5].

3.2.4 Other influence measures from multivariate analysis

Besides regression and time series influence analysis we also want to discuss influence in multivariate analysis; see e.g. Kendall (1975). The analysis of unrestricted covariances may seem unimportant for the analysis of the unrestricted VAR, but it is not. The interpretation of impulse responses and forecast error variance decompositions depends crucially on the unrestricted estimate of the covariance matrix of the residuals.

Different techniques from multivariate analysis have been suggested to identify "common trends" of the variables in the VAR (Stock and Watson (1988)) or to find underlying "equilibrium relations" (Box and Tiao (1977), Johansen (1991a)). Stock and Watson proposed the use of principal component analysis on the covariance matrix of the unfiltered data. Box and Tiao used the covariance matrices of the (vector of) levels of the variables and the one step ahead prediction errors to identify forecastable and unforecastable components. Johansen showed the partial canonical correlations of the lagged levels and first differences of the series to be the crucial elements in the maximum likelihood estimation of "equilibrium" cointegrating relationships in a finite order VAR with normal errors[6].

All these techniques require estimates of unrestricted covariance matrices of a vector of variables or residuals as input. It is not unlikely that functions of interest of the estimates of the covariances are dominated by a few observations.

Unfortunately the knowledge of econometrics, time series analysis and multivariate analysis is not often combined in applied studies of influence. Even more unfortunate is the apparent scarcity of well developed studies for the study of influence in multivariate analysis. As Critchley (1985) pointed out: "Nevertheless the use of the influence functions generally in a multivariate analysis remains a largely unexplored area." Critchley presented manageable principal component equivalents of the regression measures EIC_i, $EIC_{(i)}$, and SIC_i of Cook and Weisberg (1982, p. 110), which are reproduced in appendix A3.3.

Most expertise in data based multivariate influence analysis is typically with cross section data, where the alternative leave–k–out model reflecting

---

[5] Barnett and Lewis (1984) still used the title "Outliers in Time Series: A Little Explored Area" for their Chapter 11.

[6] These techniques are discussed in more detail in chapter 6.

the gross error prior concern is most appropriate. Graphical methods are harder to use because of the unstructured multidimensional character of the data.

The related study of outliers in multivariate analysis is also more difficult. An outlier can still be detected as a "surprising" observation, or as a discordant observation, which is "statistically unreasonable" in terms of some basic model. "But their expression is by no means as straightforward in more than one dimension", (Barnett and Lewis (1984, p. 243)). This can be noted from the residual based analysis of the previous section. "A multivariate outlier need not be an extreme in any of its components. Someone who is short and fat need not be the shortest or the fattest person around. But he can still be "too fat for comfort" and hence and "outlier"." (op cit.). Testing for outliers is an important aspect of specification testing discussed in the next section.

3.3 Diagnostic checking

After a careful analysis of the influence of the observations on the building blocks of our functions of interest, one may get the impression that the influence of some extreme observations should be downweighted in order to get a clearer picture of the impulse response pattern of the remaining observations. In economic statistics the downweighting of influential observations in the frequency domain is a well established tradition. Because of the extreme heteroskedasticity in the frequency domain (see §2.4) for many economic time series, the estimates of transfer functions relating one variable to the other are likely to be dominated completely by the low frequency (trend–) and seasonal components; see e.g. Sims (1974). One used to think it hard to relate the movement of different variables in these components to each other in an economically meaningful way. The variability in the corrected series, with trend– and seasonal components removed, was considered the variation of interest which one liked to describe by the multivariate model. Another well established tradition of downweighting the influence of observations is to introduce separate "models" for extreme ones by introducing dummy variables that affect these observations exclusively.

In the following sections we introduce and discuss some statistical criteria which can help us to assess whether the downweighting is necessary or whether it has been successful. In principle we would like to "explain" as much variation as possible by the basic model, because only variation in the data can help us identify interesting relationships. We also like to treat observations as equally as possible so that we can view the basic

relationships as approximately constant over time.

The detected relationships are best interpreted when model (2.1) can be viewed as a reliable statistical model that describes all the "salient features" of the data. Diagnostic checking of the statistical assumptions indicates whether one removed or downweighted the "unexplainable" variation to a satisfactory extent. After iterating between downweighting and diagnostic checking until the remaining variation can be assumed to fit the VAR, one has found the limits of the VAR for the description of one's data set. It then remains an open question whether the variation that is left in the data still constitutes an interesting part of the original. In the next subsections we discuss so–called misspecification tests that are geared to detecting deviations from normality and serial correlation in the disturbances and others that are used to detect significant nonconstancy of parameter estimates.

3.3.1 Choosing test statistics

The number of diagnostic test statistics that are of possible interest to test the adequacy of the VAR model (2.1)[7] is huge. Krämer and Sonnberger (1986), Kiviet (1987) and MacKinnon (1992) discussed interesting selections for the test of the adequacy of separate equations. Chong and Hendry (1986) presented some tests for the adequacy of small systems. Barnett and Lewis (1984) discussed 47 test statistics for the normality of a sample. It would not be hard to compile a similar number of tests for serial correlation, not to speak of the scores of tests for heteroskedasticity, functional form and parameter stability. The Journal of Econometrics (JoE, (1991)) published a volume titled Diagnostic Testing to celebrate 40 years of diagnostic testing in econometrics, surveying a number of approaches. The question of *test selection* has become a prominent one.

Lindley (1979) stated: "There are two broad types of approach to statistical inference .... In the first, one develops a series of techniques and investigates their properties. In the second one begins by asking what sort of properties are required, and then seeks ways of achieving them." We follow the second approach using results from others that pursued the first. The caveats from the introductory chapter apply. Our view necessarily suffers from a selectivity bias.

---

[7] For a number of practical reasons one can include deterministic regressors if one estimates (2.1), see appendix A4.2. In §3.3 the parameters of these regressors are assumed to be part of the dynamic parameters determining the mean. Sometimes we use the short hand notation B from (3.10) to denote these parameters. Equivalently one can extend y_t with deterministic components as in Sims, Stock and Watson (1990), and allow for a singular Σ.

3.3.2 Theoretical considerations for choosing tests

Testing is not a goal, but a means to get measures of goodness of fit of our statistical model, that are easy to communicate to others. In the Neyman–Pearson hypothesis testing framework one defines a test by the following components; see Cox and Hinkley (1974) or Spanos (1986, p. 392):

i the test statistic $\tau(y)$

ii the size α of the test

iii the distribution of $\tau(y)$ under the null H_0

iv the rejection (or acceptance) region

v the distribution of $\tau(y)$ under the alternative H_1

In order to stay in the Neyman–Pearson framework and apply it to a certain data set of interest, one has to define H_0 and H_1 a priori.

Preserving the probability base

As soon as one has done any (eye–ball or formal) test on the data set before the final choice of estimator the "firm probability base" is lost (Judge et al. (1980, p. 467)). In one's presentation one can act as if the estimator was chosen without pretesting. The resulting statistical model is then a *conditional* model. The estimation in the presence of outliers contains many examples of these so–called (Hogg (1974, §4)) *testimators*, see e.g. Perron's (1990a, p. 161) discussion. At this stage of analysis our goal is detective work, not inference, see Tukey (1979). The data are best viewed as a *training set*, see Efron (1983). We do not intend to give up training to preserve the probability base. According to *the strong repeated sampling principle* (Cox and Hinkley (1974, p. 45)) the selected model can only properly be tested using new data. Given the inherent nonstationarity of most macroeconomic processes it is hard to imagine if this ever becomes relevant for the kind of series we analyze, see the German economy and its world wide effects. Dewald et al. (1986) showed that very few empirical economic models in their study stood such a test.

One of the rules of contemporary econometric time series modeling is to present the outcomes of a whole range of test statistics for a preferred model. Unfortunately the rejection region for this multivariate statistic, which keeps a desired level α is extremely hard to derive, see Alt (1989) and Hillier (1991), even if one ignores the pretest problem mentioned above.

Statistical measures of fit

The fact that the Neyman–Pearson framework is hard to apply in practice, does not render it useless. It is better to use some standards of fit, than

none at all, see e.g. Sims (1988). "*P*–values in data analysis can be useful guidelines", Dempster and Gasko–Green (1981). An important attraction is the possibility of relatively straightforward generalization to multivariate measures of fit – especially when the normal distribution is used – which are much harder to develop graphically.

Choosing a null and an alternative

The five elements of the "classical test" mentioned above provide a good starting point for the choice of tests. One first has to choose an interesting null hypothesis. The null must not be made more restrictive than necessary, in order to avoid rejection of the model on the basis of uninteresting characteristics. In the parametric setting only parameters of interest should be restricted. The nuisance parameters, which provide the desired generality should not. For the analysis of the conditional mean of the series of interest (as in the impulse response analysis), one need not assume the symmetry of distribution of the disturbances to derive interesting tests. See Wooldridge (1991) for an analysis in this spirit.

After the choice for the possibly composite null one has to select an interesting alternative. In order to get some useful information for following action, were one to reject the null, the alternative must not be too general, compared to the null: the test has to be directed.

Subsequently one has to pick a size α of interest. As can be gathered from decision–theoretic frameworks (Arrow (1960)), 5 percent is not always optimal. Given the same parametric model it should be higher for small numbers of observations and lower for large numbers. This is nicely illustrated in the time series literature on selection criteria for the lag length in AR models where different purposes of the model lead to different criteria, most of which imply a number–of–observations dependent α; see Lütkepohl (1985) for a list of formulae and references.

Choosing a proper test statistic

Next one has to choose a test statistic with a contingent rejection region which has the preferred size α for every hypothesis which falls under the (composite) H_0. When H_0 is not specified to the last detail, i.e. a *non ideal* model according to Parzen (1979), this requirement of *similarity* of the test (Cox and Hinkley (1974, p. 134)) is usually too strong. One then has to select a test with a size close to, but certainly not greater than, size α for all hypotheses belonging to H_0. A good expression for the last property is *robustness of validity*; Barnett and Lewis (1984).

On the other hand one wants to choose the statistic and rejection region so that the power of the test is reasonable for every alternative of interest.

The minimum requirement is that the power be greater than the size. When this is *uniformly* so, i.e. for every alternative one can imagine, the test is said to be *unbiased* (Cox and Hinkley (1974, p. 105)). This is often seen as a necessary property for a useful test. Cox and Hinkley (1974) did not think so for two sided alternatives. Sometimes one is interested in *local power* (op cit., p. 106), i.e. power near the null in some direction into the set of alternative hypotheses. Lagrange Multiplier tests, an example of a derivation of which is given in the chapter 5, are optimal in this respect (op cit., p. 113), but power farther away from the null may be poor (op cit., p. 122). Schmidt and Phillips (1992) illustrated this for univariate unit root tests when they compared the power of their own LM test with Dickey and Fuller's LR test.

Evans and Savin (1984) provided a nice illustration of the Neyman–Pearson terminology and cumbersome[8] techniques to finite sample unit root tests under a very simple null with only one nuisance parameter.

As it as yet has been impossible to derive manageable results for many interesting tests in the finite sample case, a lot of attention is still focused on the asymptotic properties of tests. One desirable property is *consistency*. A test is called *consistent* when for any fixed level α and for a *fixed alternative* the power tends to one as the sample size goes to infinity (Cox and Hinkley, p. 317). Likelihood Ratio statistics have this property. A test which has reasonable power for a wide range of values for the nuisance parameters is said to have *robustness of efficiency*, see Barnett and Lewis (1984). Most times there is a trade off between robustness of validity and robustness of efficiency. Only quantifying a risk function in a decision theoretic approach can lead one to a formal derivation of an *optimal test*.

A final consideration in the choice of a test statistic is the *invariance* (Cox and Hinkley (1974, pp. 41–45)) of the outcome of the test under certain transformations of the sample space and associated transformations in the parameter space. In economic examples where price indexes are used, one would particularly like the outcome of the test to be independent of linear location and scale changes.

Testing strategy

Once one has figured out which tests to do and which to disregard, one has to develop a *testing strategy*, i.e. define an interesting sequence of tests and actions for every possible outcome of each test. Examples of such

---

[8] Perron (1991, footnote 2) discovered "errors due to numerical instability of the computations" only years later.

actions in our context are: to transform, add or delete a variable from the information set or to stop doing subsequent tests. As indicated above theory on multiple test procedures is still in its infancy. Many problems have been pointed out, but few practical solutions exist for problems of the generality we require here.

Mizon (1977) advocated the "logically consistent" *general to specific* procedure to start testing with the most general null and subsequently testing down to more specific null hypotheses to minimize the probability of overlooking the "true" model and to ensure robustness of validity of the testing procedure. This seems a sensible strategy from a data analysis point of view.

In strategies for testing the number of unit roots one generally prefers the *specific to general* strategy, see e.g. Dickey and Pantula (1987). In testing strategies for the number and timing of exogenous outliers it is also more convenient to start with a restricted model, cf. §3.3.7 and §5.4.

3.3.3 Practical considerations for choosing tests

It is clear that similarity is desirable from a practical point of view. The size must not depend on unknown nuisance parameters, which one does not want to model by assumption. From other empirical work on related data sets one can gather which nuisance parameters have to be taken into account.

Since we use diagnostic tests as an aid in data analysis, one has to know how statistically significant outcomes can be traced back in the data. Since we are confining ourselves to a linear model, this is not a severe problem. Many well known diagnostic tests have an *artificial regression* (Davidson and MacKinnon (1990), MacKinnon (1992)) or *variable addition* (Kiviet (1987)) interpretation, so that the influence analysis of the previous section can be applied immediately[9]. One should examine the influence on *untransformed* data as much as possible, since *transformations* might *smear* the influence of an observation in such a way that its direct influence may become hard to detect, see Welsch (1979).

For economic time series one can think of both univariate and multivariate transformations. An observation error in a price index series occurs twice in an inflation series, thereby increasing the danger of *masking*, i.e. "the tendency for the presence of extreme observations not declared as outliers to mask the discordancy of more extreme observations under investigation as outliers," Barnett and Lewis (1984, p. 114). The same error

[9] See remark A3.4 in §A3.1.1.

introduces an observation error in a volume series constructed by deflation of a nominal series by the price series.

The theoretical similarity and efficiency of a test is often based on asymptotic arguments. In small samples the asymptotic approximation may be very bad. Finite sample properties can sometimes be derived analytically, but only for specific sets of nuisance parameters.

Monte Carlo studies can provide interesting results, but are usually hard to generalize to one's own problem. Another problem is again the sheer magnitude of results available, which is very hard to summarize. It is usually dangerous to assume the results to hold for designs different from the one under which they were generated. Negative results which show that the finite samples size of asymptotically similar tests depends on easily detectable values of nuisance parameters help in the test selection problem. The existence of more evidence on small sample behavior can also provide arguments to prefer "old" (well tested) statistics to new ones in empirical analysis. If one really has no idea about the finite sample performance of a test one should do a small scale (with 1000 replications) Monte Carlo experiment oneself. For the linear techniques that we consider in this study, these exercises are not too computer time consuming.

Classical results often seem to apply under nonstandard situations. Some regression tests for parameter stability derived for nonstochastic regressors can be applicable in stationary models with lagged dependent variables as well, cf. Krämer et al. (1988). Asymptotically robust tests sometimes appear to be non–similar with regard to seemingly undisturbing nuisance parameters, cf. MacKinnon and White (1985).

A next consideration in the test choice is *computational ease*. This provides the most convincing argument against nonsimilar tests, where rejection regions have to be computed for a wide range of nuisance parameters. The alternative is to use tables of others with, say, 1800 different critical values as in Schmidt (1990). Tests using *bootstrap methods* can sometimes also be ruled out on these grounds, especially when there is no (empirical) proof that the bootstrap results are more reliable than easier–to–compute asymptotic ones. Runkle (1987) applied bootstrap methods to the VAR analysis. Griffiths and Lütkepohl (1990) compared bootstrap estimates and asymptotic estimates of the variance of estimated impulse responses of a VAR, which illustrated the premise of Huber (1964): "If an estimator is asymptotically normal, then the important central part of its distribution and confidence intervals for moderate confidence levels can better be approximated in terms of the asymptotic variance than in terms of the actual variance." The desire for

computational ease may also lead one to prefer tests based on well known distributions with "standard" rejection regions that are tabulated almost everywhere over tests with nonstandard rejection regions.

The last practical consideration for test choice is the existence of a successful action, were the test to reject the null. In principle an alternative may be quite interesting on theoretical grounds, but estimation procedures for the alternative model may be very unreliable, whereas other theoretically less appealing alternatives may be available with relatively straightforward well understood estimators. Under the null the resulting tests may not differ substantially and one may well only do the test with the easy alternative. Extreme observations, which are sometimes seen as a sign of *tail pathologies* (Parzen (1979)) are often more easily accounted for using a mixture of normal distributions (so called *slippage* alternatives, §3.3.5) than by relatively exotic heavy–tailed distributions.

We use the framework set out above to chose test statistics for the VAR. Model (2.1) is our primary null hypothesis of interest. Interesting alternatives can be written in the form (2.1)+(2.2). For the sake of exposition it is sometimes easier to consider a univariate process ($n=1$) first and then generalize the results to the multivariate case. Primary analysis of the univariate processes is also interesting in the light of the results of the previous chapter. From a pure diagnostic testing point of view it is often easier to consider one regression equation at a time. Three aspects of the null hypothesis require testing. First dynamic specification, second the distributional specification of the disturbances and third the assumption of parameter constancy, i.e. homogeneity of the sample.

3.3.4 Dynamic specification of the mean

First we consider the dynamic specification of the mean. An estimate of a time series model can only be considered as interesting "evidence" if it fits the sample spectrum reasonably well, explicitly (Milhøj (1981)) or implicitly (Box and Pierce (1970)): The residual variance has to be "significantly" lower than the original variation in the transformation of interest, and its sample spectrum must resemble that of white noise at the frequencies of interest. Our model (2.1) is only interesting if the spectrum of the appropriately differenced individual series can be approximated by ARMA models. In the time domain this entails that a certain interesting function of the sample autocorrelations has to be accounted for by the estimated model, and that a certain function of interest of the residual autocorrelations has to be close to zero, see Robinson (1991a). For the short run characteristics a function of

only the first few autocorrelations may be employed, for long run characteristics the use of higher order residual correlations is essential (op cit.).

Short term residual autocorrelation

Tests for the adequacy of the short run dynamics can be divided into two groups (Granger and Newbold (1986, p. 96)). The members of the first group test the significance of the improvement of fit of extra coefficients (like a variable addition test). The others test for serial correlation in the residuals. We prefer the first group of tests, given the homogeneity of the sample, measured through parameter constancy, discussed below. Their null hypothesis of a *mean innovation process* with respect to the information contained in all the (lagged) values of the explanatory variables is more interesting from an economic point of view (Hendry and Richard (1982, p. 11)) than the null hypothesis of no interesting serial correlation in the disturbances. For the class of ARMA processes with normal disturbances the robustness of validity and efficiency in small samples of the variable addition test clearly comes out on top, see e.g. Kiviet (1986). Once the sample distribution of the estimated residuals is approximated reasonably well by a normal distribution we prefer to use the so–called F–version of the LM test of Breusch and Godfrey (1981) which simply uses the F–statistic for the significance of lagged residuals and applies the classical tables to determine its rejection region. Poskitt and Tremayne (1982) showed the latter test to be an LM test to stationary ARMA alternatives, which guarantees local power optimality in that direction.

This test is also asymptotically robust in validity to unit root nonstationarity of the process. This has been proved formally by Chan and Wei (1988) and Tsay and Tiao (1990) for purely stochastic unit root nonstationarities at all frequencies, and by Sims, Stock and Watson (1990) for the zero frequency case with deterministic regressors. The test is not robust against nonnormality or heteroskedasticity, especially not when outliers with large leverage measures h_i can be found in the data. *Jackknife estimators* of the variance–covariance matrix of the residuals could then be employed to robustify the test to some extent, cf. MacKinnon and White (1985) and remark A3.5 below. Ali (1989) used the estimated "exact mean and covariance" of the autocorrelations to robustify residual based tests in small samples.

We prefer the data cleaning of the observations so as to make them satisfy the model restrictions of (2.1). It is easier to interpret in an unobserved components context than the application of White–type variance matrix estimators.

Another "classical" test on the dynamic specification of the mean is the check for serial correlation in the recursive residuals of Phillips and Harvey (1974). In our series time provides a natural ordering. Since it uses "unbiased" (i.e. unbiased in the classical regression setting) estimates of the disturbances, it may show more robustness of efficiency against certain types of heteroskedasticity or parameter variation than the ordinary tests for residual serial correlation, as Harvey (1975) indicated.

Long term residual autocorrelation

From a data analytic point of view, statistical tests on long run characteristics of the data cannot provide useful guidance for the relatively short series we have, simply because there is not enough information in most data sets. We discussed this already in §2.5 for the univariate case. In multivariate models this holds even stronger. If we decompose the sample information in the frequency domain to a sample spectrum or *periodogram*, only λT data points convey information on frequencies lower than λ. Fifty years of quarterly observations contain only ten "independent" data points directly containing information on frequencies lower than one cycle per five years: (1/20)·200. In the time domain this corresponds to ten nonoverlapping five year periods in the sample. In multivariate analysis the imaginary part of the Fourier transform of the sample autocovariances can also be used to describe low frequency relations, but for estimates of cross–correlations and partial correlations one obviously needs more degrees of freedom.

Testing is more useful in a later stage where different adequate powerful a priori smoothness restrictions are imposed on the (pseudo) spectral density, so that information in the higher frequencies can be used to do inference on low frequency properties.

3.3.5 Distribution of the disturbances

As far as the distribution of the disturbances is concerned, normality is certainly the preferred null. Normality can be made plausible by reference to the operation of a central limit theorem on contributory disturbance components, whereas we do not know of any other finite variance distribution for "continuous" non time series data that has this property. Variance decompositions only give a satisfactory description of the relations of variability between the variables if the disturbances are approximately normal.

Bera and Jarque (1982) interpreted the classical tests for normality on the third and fourth sample moment of the regression disturbances as LM tests with alternatives in the Pearson family of distributions. D'Agostino (1986)

discussed finite sample corrections which are relevant for samples sizes smaller than 100.

Classification of alternatives

Barnett and Lewis (1984) discussed more ready–to–use alternatives in an overview of the statistical literature on tests for discordant observations in one–dimensional samples. They would classify the Pearson family alternative as an *inherent alternative*. "In this view outliers are a sign of a greater degree of inherent variability of the whole sample. Under these alternatives all observations arise from a different distribution." (op cit.) We prefer their simple deterministic, *mixture or slippage alternatives*: "Deterministic alternatives cover the cases of outliers caused by obvious identifiable gross errors of measurement, recording and so on. The alternative is entirely specific to the actual data set and observed offending observations."

Under the *mixture alternative* "the sample is contaminated by a few members of a population other than represented by the basic model." This alternative has obvious small sample problems. "How are we to adjudge the propriety of the mixture alternative with no evidence other than the one (or two) outliers whose discordancy we are to assess using this alternative model?"

The *slippage alternative* is "by far the most common type of alternative hypothesis. All observations apart from some prescribed small number k (one or two, say) arise independently from the initial model F indexed by location and scale parameters, μ and σ^2, whilst the remaining k are independent observations from a modified version of F in which μ or σ^2 have been shifted in value (μ in either direction, σ^2 typically increased). In most published work F is a normal distribution."

Effect of regression on normality tests

Testing the distribution of estimated disturbances of a regression is not the same as testing the distribution in an ordinary location–scale model. Pierce and Kopecky (1979) provided regularity conditions under which other goodness–of–fit tests can be applied directly to the residuals of an ordinary regression equation, where the number of regressors is independent of the number of observations: "Essentially", a constant among the regressors and those conditions "needed to establish the asymptotic normality of $\hat{B}$ and $\hat{\Sigma}$ suffice."[10] Phillips and Durlauf (1986) showed that $\hat{B}$ can be asymptotically nonnormal when unit root nonstationarity is present (cf. also appendix A2.2),

[10] See Hannan and Kavalieris (1986, Th. 2.2) for restrictions on the number of parameters that ensures consistent estimation of the "true" disturbances in a time series context.

with $\hat{\Sigma}$ in general being asymptotically normal. The superconsistency of the OLS estimator of the identified parameters of B, however, may lead one to believe that direct analysis of the residuals still can be performed by classical methods in large enough samples, since one "essentially" wants $(\hat{B}-B)x_t$ to have an asymptotically negligible effect on the estimates of the residuals. Tests for constancy of the mean and variance of the disturbances, which can also be seen as normality tests are not valid when nonstationary regressors are present, however, see §3.3.6.

Our preferred class of alternatives to the normal distribution is closely related to some alternatives which are described in the econometrics literature as crashes, breaking trends (Perron (1989)) or *structural shifts* (Krämer et al. (1988)). A temporary shift in mean of the disturbances can have a permanent effect on the level of the series (the socalled persistence phenomenon, see §2.5.3).

Testing the distribution of the series directly

Lomnicki (1961) proved that normality tests on the third and fourth sample moment can also be used for stationary times series directly. This test is certainly not robust to (near) unit root nonstationarity. For appropriately differenced series it can provide a useful extra test on the adequacy of the combined normality and stationarity assumption, with both nonstationarity and nonnormality as interesting alternatives. The similarity of the test is not interesting, but the null hypothesis of normality is, since it provides a rationale to study the *sample quantile function*, (Parzen (1979)), i.e. the *empirical distribution function*, (Billingsley (1968), Pierce and Kopecky (1979)) immediately, which enables one to explore the isomorphism between sample cumulative periodograms and sample distribution functions as Parzen (1979) did in his application of autoregressive approximations to distribution functions. In the univariate case well known graphical methods from exploratory data analysis can be used as well, see e.g. Filiben (1975). Stephens (1986) presented an overview of formal tests based on the empirical distribution function.

3.3.6 Parameter constancy of dynamic and covariance parameters

The last important aspect of the model (2.1) that we want to examine by diagnostic tests is *parameter constancy*. Tests for parameter constancy are arms in the opposition against Lucas's (1976) widely cited critique on econometric policy evaluation. Until recently the Chow test PF_i used on two subsamples, recursively computed for a range of change points, was the favorite weapon in the counter–attack of Hendry and his co–authors, see

Doornik and Hendry (1992, chapters 3 and 5 and references therein) and JoPM (1992). Over the last decade one has developed a range of new tests. New central limit theorems for partial sums of weakly dependent stochastic variables were used in the derivations of *asymptotic* critical regions for the test statistics. Andrews (1990) provided an overview of tests of one–time changes of parameters with unknown change point in econometrics and statistics.

Parameter stability and unit root nonstationarity

Research on tests for parameter constancy in models with unit root nonstationarity is still developing (op cit., Hansen (1992b)). It is a complicated issue. Unit root tests do not have robustness of validity when simple alternative nonstationarities in the form of parameter changes are included in the null, see e.g. Perron (1990). Perron showed that the asymptotic distribution of an *extended* Dickey–Fuller t–statistic where one allows for a one–time change in mean in the middle of the sample equals $\tau_{\mu 2}$ instead of $\tau_{\mu 1}$ (see Tables A2.2.4 and 7.1). For an equivalent test with unknown timing of the change point the critical value at 5% significance shifts from −4.10 to −4.44, see Perron and Vogelsang (1992a). Note that extended does not mean *augmented* in the sense of the discussion in appendix A2.2. The unextended unit root tests discussed there certainly have no robustness of efficiency for one time parameter changes under the alternative (op cit.). We address this part of the issue in our discussion on outlier models in Chapter 5.

Simple tests on parameter constancy like the Chow test for predictive failure discussed in the appendix are not valid when unit roots are allowed under the null, see e.g. Kiviet (1986), Lütkepohl (1988) for Monte Carlo results. Lütkepohl (1988) provided a clue to a practical solution for the zero frequency unit roots. He noticed in simulation experiments, which mimicked the practical situation where the researcher has to select his approximating AR model from the data, that overdifferencing (introducing unit roots in the MA part of the system) does not affect the tests as badly as underdifferencing (leaving AR unit roots in the system). Provided one has cleaned the data first to avoid the effect of smearing dirty observations, this entails that parameter stability can be checked on appropriately differenced data. Whilst using AR unit root tests one should choose a low significance level, to minimize the risk of underdifferencing. Perron (1991) provided a formal illustration of the dependence of a parameter stability test on the the order of integration.

For prediction purposes, alternatives with changing parameters in the

last part of the sample are of primary interest. Goodness of fit is important for those observations because they are the main determinants of new out–of–sample predictions. Were the test to reject, it is difficult to decide whether one should adjust the model, the data or the initial significance level. Tests on parameter stability for other subsets of the sample may help.

Tests which do not need an a priori specification of the change point

For the study of relations between variables per se, focus on a specific part of the data set is not attractive. A range of tests based on the easy–to–compute recursive estimates exist to test stability over the whole sample against alternatives with an unknown point of permanent change in the parameters. One must know the strengths and weaknesses of these tests to apply them sensibly.

The *CUSUM test* of Brown et al. (1974) is most powerful to detect one–time parameter changes in B, not orthogonal to the "mean regressor", near the beginning of the sample. An example of a change orthogonal to the mean regressor is:

Example 3.1

$$y_t = \delta_0 + \delta_2(-1)^t + \delta_{22}(-1)^t D_t + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{3.11}$$

with

D_t a dummy variable equal to zero for $t \geq l$ and zero elsewhere.

Suppose one wants to test parameter stability under the assumption that $\delta_{22} = 0$. The mean regressor in this example is $[\delta_0\ 0]'$, while the parameter change under the alternative hypothesis equals $[0\ \delta_{22}]'$. The CUSUM test is not able to pick up this change in seasonal pattern.

The *fluctuation test* of Ploberger et al. (1989), which is based on the mean change in the recursive estimates of B is most powerful to detect changes in B near the end sample. Both tests are asymptotically valid for stationary models with lagged dependent variables (op cit.). Derivation of finite sample rejection regions for these tests, which are based on the maximum deviation of a series of dependent statistics from their mean is far from straightforward.

In static models, the empirical sizes of the recursive tests are usually lower than the nominal ones. Limit distributions are derived from crossing probabilities of continuous time processes. If the process is only measured at discrete time intervals, the crossing probability is lower. In dynamic models the story seems to be different. Because of many rejections of our seemingly reasonable null in practice, we simulated the null distribution of the fluctuation test in a VAR(1) of 200 observations and found rejection probabilities far above the theoretical ones for an AR(1) model with first

order autocorrelations as low as 0.4. The fluctuation test lacks invariance to changes in scale of the variables. The fluctuation test statistic and some modifications are described in detail in §3.3.8.

The CUSUM test on OLS residuals (Ploberger and Krämer (1992)) is entirely symmetric between 1 and T and is most powerful to detect changes in mean in the middle of the sample. One can interpret it as a test for stability of the parameter of the constant term, see Hansen (1992a). One can easily adapt this test to detect instability of δ_2 in (3.11), cf. §7.4.7.7.

The rejection region of tests based on cumulative residuals (or transformations as in Hansen (1992a)) depend (even asymptotically) on the number of parameters tested for stability and on the presence of nonstationary regressors in the test equation. The rejection region for the ordinary CUSUM test is even changed by mean adjustment of the data, cf. Ploberger and Krämer (1992, §3). The practical problems in the application of tests with unknown change point to dynamic models are considerable. They are comparable with the problems encountered in applied AR unit root testing. Note that variations of these statistics have been used to test for MA unit roots, see §2.5.

By combining the ordinary CUSUM test, the CUSUM test on OLS residuals and the fluctuation test one can develop a procedure with reasonable power against alternatives with changes in mean in the beginning, the middle and the end of the sample. The problems with these statistics in the Neyman–Pearson framework do not make them useless: The philosophy behind the CUSUM test of Brown et al. (1975) was basically that of "data analysis as expounded by Tukey (1962)."

Choice of significance level in sequences of tests with fixed change points

The statistical theory underlying this kind of tests provides guidelines to prevent one becoming unnecessarily worried about the quality of the approximation by the model, whenever one spots one "significant" residual. This could lead one to make every interpretable parameter stochastic or reduce the influence of all observations that contain valuable information far too much. In other words: it prevents one to follow the data too slavishly. An example on the test for outliers in a normal sample with unknown mean μ and *known* variance σ^2 from Barnett and Lewis (1984, p. 130) may make this clear[11].

Example 3.2

Suppose we have n identically (though not independently) distributed quantities $|T_i|$, each used as a test statistic for a *single* outlier at one of

[11] See op cit., §10.2.1 for applications in regression models. Note that critical values obtained in op cit. eq. (10.2.22), tabulated in Lund (1975) are for *internally* studentized residual ($\bar{t}_i^2$) in §A3.1.1.

a series of independent observations i. Denote the critical value by t. Define $T_{max} = \max|T_i|$. Let E_i denote the event $T_i > t$ and E the event $T_{max} > t$, so E is the union of the n events E_i. We are interested in the *significance probability* attached to t, defined by $SP(t) = P(T > t) = P(E)$ under our null of a homogeneous sample. Simple bounds for $P(E)$ can be derived from Bonferroni inequalities:

$$nP(E_1) - \tfrac{1}{2}\{nP(E_1)\}^2 < P(E) \leq nP(E_1). \tag{3.12}$$

The upper bound is well known, and is often judged "conservative". If one wants t so that $SP(t) \approx 0.05$ and $n = 20$ and one has $P(E_1) > 0.0025$ it follows that $0.04875 < P(E) < 0.05$. Tighter bounds have been derived, but seem trivial in light of the discussion above. The example indicates that the easily computed Bonferroni approximations can be quite accurate.

One sees that testing the null after scrutinizing the data for observations indicating misspecification must be done using (nonstandard) large deviate values corresponding to very small tail probabilities. The (1–0.05/2)–quantile of a t_{19}–distribution is 2.09, its (1–0.0025/2) quantile is 3.48 according to table 1 in Bailey (1977)[12]. It is still an open question how these results generalize to samples with dependent observations that we use.

Abraham and Chuang (1989) derived critical values for the maximum of the squared studentized residuals in simple stationary time series models. They used extreme value theory and checked their results by simulation in simple univariate AR models[13]. The results indicated that the supposed conservatism from Bonferroni values is bearable.

Instability of B or Σ?

Most tests for parameter constancy test both stability of B and Σ in (2.1). The Variance Ratio or Goldfeld–Quandt (1965) test is a rare exception. It is also valid when B is not constant between the subsamples. A range of LM tests of (2.1) against certain types of heteroskedasticity and random variation in the dynamic parameters use the same test statistic and rejection region (cf. Breusch and Pagan (1979), Weiss (1986) (ARCH), Tsay (1987) (CHARMA), Lee and King (1993) (GARCH)), so that one gets no clues in which direction the misspecification can be found, were the test to reject. For the sample sizes that we use we prefer the *noninherent alternatives* (§3.3.5), and

---

[12] It is dangerous to trust critical values from standard computer packages for very small p-values. See Press et al. (1986) for a discussion of the approximations used in numerical applications. Those work well here.

[13] For n=100 and an $F(1,98)$ for T_i, the Bonferroni 5 percent critical value for T_{max} is the (1-0.05/100)-quantile of an $F(1,98) \approx 11.5$, which lies between the 5 and 10 percent critical values of 12.4 and 10.7 found by Abraham and Chuang.

we would like to use the outcomes of the tests as indications for (a few one–time) changes in B or Σ. Once one knows which observations fail to fit the model, one can always try to find out whether this failure should be related to external (exogenous) or internal (endogenous) causes.

Changes in Σ make the estimates of variance decompositions for the full sample hard to interpret. WLS–type techniques could leave the impulse response analysis for the full sample still worth the effort. Changes in B however make full sample analysis treacherous. This type of problem is similar to that of the *Behrens–Fisher problem* for testing a constant mean across samples with unequal variances, where no easy solutions exist either, see Amemiya (1985, p. 36). Koschat and Weerahandi (1992) provided a solution using Monte Carlo integration for subsamples with full rank regressor matrices.

Using a prior concern analysis of §3.2.2 and considering the relationships between the influence statistics and well known test statistics formulated in table A3.1 (app. A3.2) one finds out which parameter stability tests are geared to detection of a change in Σ and which to a change in B.

The Chow test rejects for high values of PF_i if B changes and/or Σ "increases" provided the subsample of interest i is not too large. When the number of observations in the subsample increases faster than the number of regressors it converges to the variance ratio statistic VR_i. The variance ratio test has a two–sided rejection region, and is supposed to detect only changes in Σ, both "decreases" and "increases". We show the links between the different tests in an empirical example in §7.5.3. Figure 7.12 provides a graphical illustration.

The *analysis of covariance test* is meant to detect changes in B and is also most powerful for certain simple alternatives, see Chow (1960). The analysis–of–covariance test has since long been used in time series analysis to detect *heterogeneous* or *time dependent nonstationarity* as alternatives, both in the frequency and in the time domain, see Priestley and Rao (1969). In Monte Carlo experiments we found the test to lack robustness of validity in small samples (i.c. too many rejections under the null) if the test is applied to relatively large subsets in regressions with a constant term and substantial dynamics (in particular with unit root nonstationarity). These experiments confirmed results obtained by Campos et al. (1993, §4). In that respect the Chow test PF_i is superior. A vast number of *state–dependent alternatives*, that are nonlinear in the parameters have been proposed since then, see Priestley (1988). They are outside the scope of this study.

Common tests for *heteroskedasticity* (Breusch and Pagan (1979)) can be used to test directly against inherent alternatives for parameter stability,

e.g. variance of the disturbances increasing with the squared level of a variable of interest. Tsay (1988) presented a simple procedure to use the maximum and minimum of the variance ratios in order to detect one time changes in the variance of the disturbances at an unknown change point. Harvey (1985) suggested such a variance change in a model for GDP figures of the U.S.A. Hansen (1992a) suggested a simple test of homoskedasticity against a permanently changing residual variance based on the partial sums of squared residuals, a variant of the CUSUM of squares test, but here based on OLS residuals instead of recursive residuals. He applied it on more recently available historic GDP data and could not reject the null of homoskedasticity.

3.3.7 An alternative test for parameter stability

The analysis–of–covariance test can only be used for subsamples with a number of observations larger than the number of regressors. This is another severe handicap in unrestricted VARs with an interesting number of variables. We propose a *generalized number of outliers* criterion as an alternative test statistic. It can also be used to test overall parameter stability for smaller windows.

Compute $[T/m]$ outcomes of the Chow statistic PF_i which we discussed in the single equation version in §3.2.2, for nonoverlapping subsets of size m, denoted PF_{mi} with common critical value C_m corresponding to size α_m. The multivariate version $\bar{T}_{ii}$ of appendix 3.2 can be used as well. Let N_m denote the number of PF_{mi}–values greater than C_m. One can approximate the critical value C_N for a size α test of N_m by the α–rejection region of a statistic with a binomial distribution with parameters $[T/m]$ and α_m:

$$\sum_{i=0}^{C_N-1}\binom{[T/m]}{i}(\alpha_m)^i(1-\alpha_m)^{[T/m]-i} \leq (1-\alpha) < \sum_{i=0}^{C_N}\binom{[T/m]}{i}(\alpha_m)^i(1-\alpha_m)^{[T/m]-i}. \quad (3.13)$$

The test rejects for $N_m \geq C_N$. Note that the Bonferroni values the *overall level* α and the *partial level* $\alpha_m = \alpha/[T/m] - \varepsilon$ give a critical value of 1, for small positive ε and $[T/m]$ large: no outliers at size $\alpha_m - \varepsilon$ are allowed. The statistic can be computed for different values of m. The results from Abraham and Chuang (1989) indicate that the approximation is poor as soon as one uses overlapping subsets.

In the normal regression case with i.i.d. observations the Bonferroni critical values can be considered (slightly) *conservative*, i.e. one rejects the null in less than an α proportion of the cases. Schweder (1976, §6) described a sequential procedure with marginally increasing partial levels, which keeps the overall level α in this case. If one uses the test statistic

$$\min_m p\cdot[T/m]$$

with p the minimal *nominal* p–value for a Chow statistic for a subset of size m and a corresponding overall critical value of 0.05 one probably rejects the null more than 5% of the times in repeated samples. In practice we use a small number of values for m, like 1, 2, 4, 8, 16, 32. We do not select the borders between the larger subsets on the basis of data evidence. This test can also be viewed as a goodness–of–fit test for generalized residuals. For m small relative to T one can compare the goodness of fit of the F–statistics with the theoretical distribution under the null. For the multivariate Chow test only certain percentiles with known critical values can be used. A general easily computed formula for its distribution under the null has to be developed yet.

Dufour (1982) derived the distribution of a *generalized Chow test* statistic with an arbitrary number of nonoverlapping a priori fixed subsamples. He allowed for cases where some of these subsamples are *undersized*, and some are not. Cantrell et al. (1991) provided a number of predictive interpretations for this test. They rewrote (op cit., eq. (37)) the test statistic as a weighted sum of an analysis–of–covariance test statistic and an independent predictive Chow test statistic. In that view one can see Dufour's procedure as a test on the *mean* of the PF_{mi}–values.

We use this test for small subsamples together with the analysis–of–covariance test for larger subsamples. The only inputs needed to compute all these test statistics are the *residual sums of squares* for different subsamples, see Table A3.1 in appendix A3.2. The sums needed for the PF_i'*s* can be computed recursively in one forward run, see Brown et al. (1975). A backward run enables one to do the analysis–of–covariance test for a range of splits in two subsamples. The statistics for tests against the alternatives with unknown change point can be computed in the same runs.

3.3.8 Multivariate diagnostics

So far we have mainly discussed tests for single regression equations, which is interesting in its own right, since "satisfying the relevant single equation or small subblock criteria seems necessary, if not sufficient, for useful systems to emerge," see Chong and Hendry (1986). In multiple time series the system as a whole should be analyzed too. This is not often done in applied econometric work. In a comprehensive review article on applied diagnostic checking MacKinnon (1992) used only *one* reference for multivariate regression and more than 160 for other subjects.

Generalization of the F–tests for serial correlation is conceptually

straightforward. Testing for multivariate normality can be done by Mardia's (1970) test for multivariate excess skewness and kurtosis. This is essential for the validity of some tests like the multivariate version of the Chow test.

Small sample distributions of multivariate test statistics for parameter stability with a fixed change point, such as the Lawley–Hotelling trace statistic, have been tabulated for a range of interesting combinations of degrees of freedom, see appendix 3.2. Asymptotic chi square approximations are inadequate, bearing in mind the great loss of degrees of freedom due to the big number of unrestricted parameters in the VAR. The scale of the error (too many rejections under the null) one makes by applying the asymptotic distribution can be derived from the tables of the Lawley–Hotelling critical values. F–type approximations seem to fare better, see Woodland (1986). Simple degrees–of–freedom corrections help too, see §A5.2.1.

F–tests on linear restrictions are only "exact[14]" for tests of a number of linear restrictions in one equation or for tests of one linear restriction in a number of equations, cf. Anderson (1984). Note that one can interpret the predictive Chow tests as tests for the absence of a dummy for each observation for each equation to be forecast, cf. e.g. Pesaran et al. (1985). The F–type approximations have the advantage of easy computation of corresponding p–values, see e.g. Press et al. (1986). We use Theil's (1971, pp. 402–403) F–test in multivariate diagnostic variable addition tests and compute corresponding p–values for the overall test and for tests for the separate equations. For very small numbers of degrees of freedom (<20) this approximation may still be inadequate. In chapter 5 (e.g. table 5.1) we present examples of F–values in tests for residual autocorrelation. In chapter 7 we present p–values. See e.g. table 7.5. We use the degrees of freedom correction of the χ^2 test in §7.5.2. The corrections matter for the interpretation of test outcomes at conventional significance levels.

Many *asymptotic* results can be obtained by a proper vectorization of the set of equations and by imposing the necessary extra regularity restrictions. We develop a multivariate test in detail in chapter 5.

Tests based on recursive estimates can be generalized using central limit theorems for scaled partial sums of weakly dependent multivariate stochastic variables, see e.g. Phillips and Durlauf (1986). For tests that already employ this theory a generalization to a multi equation system merely requires a conventional vectorization and a corresponding increase of the dimensionality parameter for the null distribution.

---

[14] Ignoring the effect of lagged dependent variables and assuming normality.

A generalization of the fluctuation test, which uses the "maximum–norm" on changes in the estimates of B may serve as an example. We employ the following test statistic in a VAR which has the same stationary regressors in each equation:

$$F^{(T)} = \max_{t=K,\ldots,T} \|F_t^{(T)}\|_\infty, \tag{3.14}$$

where

$$F_t^{(T)} = \frac{t-K}{T-K}\Sigma^{-\frac{1}{2}} \otimes (X'X)^{\frac{1}{2}} \mathrm{vec}(\hat{B} - \hat{B}_{(t)}),$$

and $\hat{B}_{(i)}$ denotes the OLS estimate for B using the sample from 1 to time i. Σ is the full sample conditional maximum likelihood estimate of the covariance matrix of the disturbances, K is the number of regressors per equation, and n is the number of equations.

Under the null $F^{(T)}$ has the following limit distribution as $T \to \infty$:

$$\begin{aligned} P(F^{(T)} > b) &= 0 && b < 0 \\ &= [1 + 2\textstyle\sum_{i=1}^{\infty}(-1)^i \exp(-2i^2b^2)]^{nK}, && b \geq 0 \end{aligned}$$

so that empirical asymptotic p–values can be computed easily. It is derived from formulae for the crossing probability of the maximum maximorum of the absolute values of a so–called vector Brownian bridge W^0, cf. Billingsley (1968). The asymptotic critical values for the CUSUM test on OLS residuals is based on the same formula for $K = 1$. Lo (1991) and Hansen (1992a) employed other functions of the partial sums of the residuals and used corresponding other functions of the Brownian bridge process to derive asymptotic critical values for their test statistics, cf. §2.5.2 and §3.3.6.

Exact assumptions for the asymptotic validity of the fluctuation test can be found in Ploberger et al. (1989). The only extra condition for multivariate application is the assumption of a regular Σ. The use of $\hat{B}_{(t)}$ has the drawback that dummy variables that are constant, or indeed linearly dependent upon the other regressors, up to a certain point t, $K < t < T$, are not allowed in this set–up. The regressors have to be bounded asymptotically so that polynomial trends are not allowed either.

3.3.9 A diagnostic for multivariate unit roots

It is a good strategy to check the stationarity of the data before doing other diagnostic tests. Multivariate tests for unit roots may be a poor guide in the beginning of the selection process. It is hard to visualize the degree of nonstationarity in a system. One can make time series plots of all the series in levels and appropriate differences to check univariate stationarity. Differencing all the series may not be necessary to achieve stationarity of

the system. As Tsay and Tiao (1990) noted: "there is no convenient method currently available to identify the "genuine nonstationary components", of a vector process." The "true" nonstationary components are not interesting from a data analytic point of view, methods to identify them can be. Several proposals to identify these components in economic time series had been made in the econometrics literature, however[15].

Consistency of OLS estimators of unstable roots

Some of the properties of the estimated system can be illustrated by what we call the shape of the *implied typical log spectrum* of the (stochastic part) of the VAR. This is the 2π times the log of the squared gain of $\det[\Phi(L)(1-L)^{n-r}]$ in (2.1) where $(n-r)$ is the supposed number of unit roots at frequency 0 in the system, see (3.15). We adopt the convenience of Tsay and Tiao (1990) and refer to *inverses* of the zeros of $\det[\Phi(L)]=0$ as the *characteristic roots of* y_t. The "ordinary" roots 1 and –1 equal the corresponding characteristic roots. A convenient way to derive the determinant is to compute all the eigenvalues of the *companion matrix*, which is the AR(1) coefficient matrix where the n–variate AR(p) process is rewritten as an np–variate process, see e.g. (op cit, form (2.1)). These eigenvalues, which *are* the characteristic roots of y_t, are on or *inside* the unit circle under our null hypothesis. The OLS estimators of the AR parameters corresponding to the nonstationary characteristic roots in the AR part are consistent and converge at a rate faster than the usual $1/T^{\frac{1}{2}}$ under the general hypothesis of a purely stochastic multivariate ARMA process with finite MA order and *left coprimeness* of the AR and MA matrix polynomial (Tiao and Tsay (1990, Theorem 6.2)).

System theoretic definitions and some practical implications

Coprimeness is a common term in linear systems theory which now and then turns up in the theoretical econometrics and time series literature. Definitions can be found in Kailath (1980, pp. 369–370): Two polynomial matrices are *left coprime* if they only have *unimodular common left divisors*. A polynomial matrix $\Phi(z)$ is *unimodular* if and only if its determinant is a nonzero constant, independent of z. The polynomial matrix $D(z)$ is *nonsingular* if $\det[D(z)]$ is not identically zero. It is also a *common left divisor* of $\Phi(z)$ and $\Theta(z)$ if it can be chosen so that $\Phi(z)=D(z)\tilde{\Phi}(z)$ and $\Theta(z)=D(z)\tilde{\Theta}(z)$, with $\tilde{\Phi}(z)$ and $\tilde{\Theta}(z)$ polynomial matrices. A socalled *matrix fraction description*

[15] Judging from some persistent errors in the cross references there seemed to exist a disturbing lack of communication between different strands in literature that analyze comparable economic time series. Johansen and Juselius (1990) referred to Box and Tiao (1977) as Box and Tiao (1981). Tsay and Tiao (1990) referred to Phillips and Durlauf (1986) as Phillips and Durlauf (1985).

$C(z)=\Phi^{-1}(z)\Theta(z)$ is *irreducible* if $\Phi(z)$ and $\Theta(z)$ are left coprime. When we look at $C(L)$ as the matrix of impulse responses we see that it is unique "up to a matrix of degree zero in z". Translation of these conditions in practical situations is difficult. Bell (1987) showed for the univariate case with fixed initial conditions that the existence of a *nonstationary* common divisor can be translated in practice to the existence of deterministic, i.e. perfectly predictable components in the model, which indicates that common factors do not simply "cancel out". In a theoretical paper Hannan and Poskitt (1988) arrived at the conclusion that the number of unit roots in the MA part corresponds to the number of unit canonical correlations between future and past, and gave an example for the univariate MA(1) model.

Fountis and Dickey (1989) used the assumption of *controllability* in their derivation of a unit root test. In a VAR model this assumption excludes the possibility of perfectly predictable (i.e. deterministic) linear combinations of the variables.

The statistic

Since the largest characteristic roots should be close to their true unit root values[16], we can simply estimate the unrestricted VAR and delete the $n-r$ factors of the determinant closest to $(1-L)$. A similar analysis can be employed for the unit roots at the other frequencies. The formula for the *implied typical log spectrum* is, see e.g. Sargent (1979),

$$f(\omega,r)=-\log[2\pi\det(\Phi(e^{-i\omega}))\det(\Phi(e^{i\omega}))(1-e^{i\omega})^{-(n-r)}(1-e^{-i\omega})^{-(n-r)}]. \qquad (3.15)$$

As this is a trigonometric *even* (i.e. symmetrical around zero) function of period 2π, it suffices to plot the function for ω on the interval $[0,\pi]$, see e.g. fig. 7.7 in chapter seven. Since $\det[\Phi(L)(1-L)^{r-n}]$ is the AR component which all the series have in common, the well known typical spectral shape (Granger (1966)) should show up in $f(\omega,r)$ for a set of macroeconomic variables if the approximation by the model is adequate. Note that the typical spectral shape is a phenomenon of *detrended* series. Sowell (1987) called $f(\omega,n)$ the *common factor* of the spectral density of y_t, given by $(2\pi)^{-1}C(e^{i\omega})\Sigma C'(e^{i\omega})$, with $C(L)$ the polynomial matrix of the infinite MA representation for the appropriately differenced series. To see the direct connection with unit root tests on the sum of the coefficients in an AR approximation of the process note that $\lim_{\omega\to 0} f(\omega,n)=-2\log((2\pi)^{\frac{1}{2}}\det(\Phi(1))$. The plot provides a "semi-parametric" data based graphical check on outcomes of

[16] Fountis and Dickey (1989) gave asymptotic approximations of some distributions of the eigenvalues, see also chapter 6.

tests for the number of characteristic roots on the unit circle, which are discussed in Chapter 6. Furthermore it enables a direct look at possible (near) nonstationarities at all frequencies. The implied *individual* factors of the spectra of the different variables can be derived in a similar way, see Maravall and Mathis (1991).

3.3.10 Consequences of "rejection" of the model

If the model is rejected, there are several strategies to improve the results. The most obvious in our set–up, is to use other preprocessing formulae (2.2). When formulae are developed a priori we interpret this as a change in the choice of the variable of interest. A good example is the choice of a linear or a log–linear model, which indicates interest in absolute or relative growth. Another example is deflation of nominal series.

A different issue is the correction of series to reduce the influence of seasonal components or outliers to obtain parameter stability in the VAR for the corrected series. One must see to it that substantial variation is left in the data to identify interesting relationships, in order "not to throw the baby out with the bathwater". This risk exists if the outlier and seasonal components are estimated using *adaptive* smoothers, where the smoothness of the components is determined by the data. Without an estimate of the remaining "equivalent degrees of freedom[17]" subsequent statistical analysis is treacherous anyway.

On the other hand one must be careful not to introduce variation by the "corrections", so that it dominates the original variation, cf. the Kuznets story referred to in §1.1, see also Harvey (1982, p. 82). In this way one runs the risk of finding significant "error correction components" that one put in oneself in the first place.

One can take differencing as a simple example. The variance of an AR(1) process as in (3.9) with β bigger than 0.5 is reduced by differencing. For β–values smaller than 0.5 it is increased, so that the "remaining" variance is dominated by what maybe was meant a "safe" correction to analyze the original series more appropriately. One has used "polluted water to clean the baby."

Another relevant example is *deflation*. Deflation of several series by the same improper deflator can lead to *spurious correlation* problems as well, cf. e.g. Madansky (1964) and Griliches (1986).

We introduce our strategies for seasonal and outlier correction in the chapters 4 and 5. We discuss deflation problems in §7.3.

---

[17] Koopmans (1974, p. 273) defined this concept for spectral smoothers.

A3.1 Influence Measures for Normal Linear Model

This appendix gives single equation predictive influence measures discussed in §3.2.2 and discusses a connection with the Chow statistic for predictive failure.

A3.1.1 Global influence measures

First we present the global influence measures for the deletion of (subsets of) observations. We follow the notation of the model (3.1). The notation for the measures basically follows Johnson and Geisser (1983)[1].

- W — $X'X$, $k\times k$ moment matrix using all observations, with k the number of regressors,
- X_i — $m\times k$ submatrix of X containing m observations the influence of which is to be examined, $m<T-k$,
- $X_{(i)}$ — remaining part of X if X_i is deleted from (reordered) X, $(T-m)\times k$ matrix, $T-m>k$,
- $W_{(i)}$ — $X_{(i)}'X_{(i)}$,
- $y_{(i)}$ — remaining part of y if y_i is deleted from (reordered) y, $(T-m)\times 1$ vector,
- H_i — $X_iW^{-1}X_i'$, $m\times m$ diagonal block of the hat matrix,
- $\hat{\beta}$ — OLS estimate for β using full sample, $k\times 1$ vector,
- $\hat{\beta}_{(i)}$ — OLS estimate for β using subset deleted sample $(y_{(i)},X_{(i)})$,
- r_i — $m\times 1$ vector of OLS residuals of subset i,
- $\bar{r}_i$ — $y_i-X_i\hat{\beta}_{(i)}$ vector of OLS residuals using parameter estimate $\hat{\beta}_{(i)}$, also known as (multivariate generalization of) the *prediction error* or as *predicted residual*,
- $\hat{\sigma}^2$ — "Unbiased" OLS estimate of residual variance using full sample,
- s^2 — $\hat{\sigma}^2(T-k)$: residual sum of squares of full sample,
- $\hat{\sigma}^2_{(i)}$ — OLS estimate of residual variance with subset deleted sample,
- $s^2_{(i)}$ — $\hat{\sigma}^2_{(i)}(T-k-m)$: residual sum of squares of subset deleted sample,
- $D_i^2(W)$ — $r_i'(I-H_i)^{-1}H_i(I-H_i)^{-1}r_i$, the numerator of Cook's distance, see (3.3),
- t_i^2 — $s^{-2}r_i'(I-H_i)^{-1}r_i$,
- $\bar{t}_i^2$ — $\hat{\sigma}^{-2}r_i'(I-H_i)^{-1}r_i/m$, the (group analog of the) squared *internally Studentized residual*[2],

---

[1] Johnson and Geisser (1983) used p for the number of regressors, k for the size of the subset, S for the moment matrix of the regressors, V_i for the submatrix of H (in 3.6) and a^2 for the residual sum of squares. We use k, m, W, H_i and s^2.

[2] It is also called *standardized residual*, see e.g. Pfaffenberger and Dielman (1991).

T_i^2 $s_{(i)}^{-2} r_i'(I-H_i)^{-1} r_i = s_{(i)}^{-2} s^{-2} t_i^2 = (1-t_i^2)^{-1} t_i^2$,

$\bar{T}_i^2$ $\hat{\sigma}_{(i)}^{-2} r_i'(I-H_i)^{-1} r_i/m = [(T-k-m)/m] t_i^2 (1-t_i^2)^{-1}$, the (group analog of the) squared *externally Studentized residual*[3],

PF_i $\bar{r}_i'[I+X_i(X_{(i)}'X_{(i)})^{-1}X_i]^{-1}\bar{r}_i/(m\hat{\sigma}_{(i)}^2)$, the Chow (1960, formula (13)) statistic for predictive failure,

f_1 $[(T-m-2)/(T-k-m-2)]\cdot[1-t_i^2]$,

c_2 $(T-m-2)/(T-m)$, a constant,

$locPI_i$ $\frac{1}{4}\cdot c_2\cdot\hat{\sigma}^{-2}\cdot D_i^2(W)$, the location component of the influence of subset i on the predictive density for y,

$covPI_i$ $\frac{1}{4}\cdot f_1\cdot \mathrm{tr}[H_i(I-H_i)^{-1}] - \frac{1}{2}\ln|I+\frac{1}{2}[H_i(I-H_i)^{-1}]| + \frac{1}{2}T(f_1 - \ln|f_1| - 1)$, the covariation component of the influence of subset i on the predictive density for y.

Remark A3.1 $t_i^2 > t_j^2 \Leftrightarrow T_i^2 > T_j^2$: For the influence order of the subsets, it does not matter whether "Studentization" is done internally or externally. Note $t_i^2 < 1 \ \forall i$. Note that dividing the expression for T_i^2 by the one for t_i^2 gives

Lemma A3.1

$$s^{-2}s_{(i)}^2 = 1 - t_i^2, \tag{A3.1.1}$$

which can also be written as

$$s_{(i)}^{-2}s^2 = 1 + T_i^2. \tag{A3.1.2}$$

Equation (A3.1.1) follows from well known algebraic identities:
Applying the matrix inversion lemma, see e.g. Judge et al. (1985, p. 938),

$$(C+VDL)^{-1} = C^{-1} - C^{-1}V(D^{-1}+LC^{-1}V)^{-1}LC^{-1},$$

to evaluate $W_{(i)}^{-1}$ gives

$$W_{(i)}^{-1} = (W - X_i'X_i)^{-1} = W^{-1} + W^{-1}X_i'(I-H_i)^{-1}X_iW^{-1}. \tag{A3.1.4}$$

Substitution into

$$s_{(i)}^2 = y_{(i)}'y_{(i)} - y_{(i)}'X_{(i)}W_{(i)}^{-1}X_{(i)}'y_{(i)}, \tag{A3.1.5}$$

using

$$s^2 = y'y - y'XW^{-1}X'y, \tag{A3.1.6}$$

$$y'y = y_{(i)}'y_{(i)} + y_i'y_i,$$

and

$$y'X = y_{(i)}'X_{(i)} + y_i'X_i \tag{A3.1.7}$$

---

[3] It is also called *Studentized deleted residual* (op cit.).

gives $s^2 - s^2_{(i)} = r_i'(I-H_i)^{-1}r_i \Leftrightarrow t_i^2 = s^{-2}(s^2 - s^2_{(i)}) \Leftrightarrow$ (A3.1.1). □

Lemma A3.2 $\bar{T}_i^2 = PF_i$.

These statistics[4] are distributed as $F(m, T-k-m)$ under the null of (3.1). For $m=1$ it follows that $(\bar{T}_i^2)^{1/2}\text{sign}(r_i)$ has a Student $t_{(T-k-1)}$ distribution. This explains the name *squared Studentized residual* for $\bar{T}_i^2$.

Proof

The following well known identity, easily derived using (A3.1.4) and (A3.1.7), see e.g. Johnson and Geisser (1983, form (3.1)), or $DFBETA_i$ in (3.6)),

$$\hat{\beta}_{(i)} - \hat{\beta} = W^{-1}X_i(I-H_i)^{-1}r_i \tag{A3.1.8}$$

implies

$$\bar{r}_i = y_i - X_i\hat{\beta} + X_i(\hat{\beta} - \hat{\beta}_{(i)}) = (I + H_i(I-H_i)^{-1})[(I-H_i)(I-H_i)^{-1}]r_i = (I-H_i)^{-1}r_i. \tag{A3.1.9}$$

The next identity follows from the matrix inversion lemma:

$$[I + X_i(X_{(i)}'X_{(i)})^{-1}X_i]^{-1} = (1-H_i). \tag{A3.1.10}$$

Substitution of (A3.1.9) and (A3.1.10) in the expression for PF_i gives

$$PF_i = \bar{T}_i^2. \qquad \square$$

Pesaran et al. (1985) discussed other simple ways to write and interpret PF_i. Cantrell et al. (1991, §3) discussed *generalized* versions, i.e. versions for more than two subsets. They made extensive use of generalized inverses in their exposition. t_i^2 has a Beta distribution under the null, see (A3.2.12).

Remark A3.3 $(I-H_i)^{-1}$ does not exist for every set of regressors. From (A3.1.2) we see that the existence of $\hat{\beta}_{(i)}$ is a related problem. The nonexistence can occur in practice when dummies are introduced to model changing parameters for the observations of interest; see e.g. Cook and Weisberg (1982, p. 137).

Remark A3.4 The analysis can easily be applied to study the influence of subsets of observations on subsets of parameters of the vector β, say β_1, by using the well known *partial regression* technique. Regress y and the interesting components of X, say X_1 on the remaining components of X, say X_2 first. Call the fitted values $\hat{y}_2$ and $\hat{X}_{12}$. Then use the residuals of these

---

[4] For models with lagged dependent variables, this F-distribution is not exact, but merely an approximation in finite samples, but it can be expected to outperform asymptotic χ^2-approximations in retaining nominal size, see Kiviet (1985) for some Monte Carlo evidence.

regressions, $y-\hat{y}_2$ and $X_1-\hat{X}_{12}$, as inputs for the partial analysis of the subset of parameters β_1. A scatter plot of this regression is called *partial regression leverage plot.* Efficient computation may follow different lines, see e.g. Cook and Weisberg (1982, p. 45). We can use the analysis later in reparameterized VARs like the Vector Error Correction Model VECM (Chapter 6) where some coefficients of β get more attention than others.

A3.1.2 Local influence measures

We follow the same notation as in the previous section. Derivation and interpretation of the statistics can be found in Cook (1986).

E	$\mathrm{diag}(r_1, r_2, \ldots, r_T)$, a $T\times T$ diagonal matrix containing OLS residuals,
$R=M(\sigma^2)$	$\{ R_{ij} \}$ a $T\times T$ symmetric matrix with $R_{ij}=r_i^2 r_j^2/(2T\hat{\sigma}^2)$, $i,j=1,\ldots,T$,
H	$X(X'X)^{-1}X'$,
$M(\beta)$	EHE,
$M(\beta,\sigma^2)$	$EHE+R$,
$\ell_{max}(\beta)$	The eigenvector corresponding to the largest eigenvalue of $M(\beta)=EHE$, with $\|\ell_{max}(\beta)\|=1$, local perturbation of observation weights resulting in greatest likelihood displacement with respect to parameter vector β,
$\ell_{max}(\sigma^2)$	The eigenvector corresponding to the largest eigenvalue of $M(\sigma^2)=R$, with $\|\ell_{max}(\sigma^2)\|=1$, local perturbation of observation weights resulting in greatest likelihood displacement with respect to parameter σ^2. Note that R has rank 1, and that $\ell_{max}(\sigma^2)$ is proportional to the vector of squared residuals,
$\ell_{max}(\beta,\sigma^2)$	The eigenvector corresponding to the largest eigenvalue of $M(\beta,\sigma^2)$, with $\|\ell_{max}(\beta,\sigma^2)\|=1$, local perturbation of observation weights resulting in greatest likelihood displacement with respect to parameter vector $\delta=(\beta',\sigma^2)$,
$C_{max}(\beta)$	$\ell_{max}(\beta)'M(\beta)\ell_{max}(\beta)$, maximum normal curvature with respect to β. It equals the largest eigenvalue of $M(\beta)$.
$C_{max}(\sigma^2)$	$\ell_{max}(\sigma^2)'M(\sigma^2)\ell_{max}(\sigma^2)$, maximum normal curvature with respect to σ^2,
$C_{max}(\beta,\sigma^2)$	$\ell_{max}(\beta,\sigma^2)'M(\beta,\sigma^2)\ell_{max}(\beta,\sigma^2)$, maximum normal curvature with respect to $\delta=(\beta',\sigma^2)$.

Remark A3.5 $M(\beta)^2 = EHE^2HE = EX\Delta_0 X'E$, with $\Delta_0=(X'X)^{-1}(X'E^2X)(X'X)^{-1}$. Δ_0 is known in econometrics as White's (1980) heteroskedasticity–consistent

estimator of $\mathrm{Var}(\hat{\beta})$. $[T/(T-k)]\Delta_0$ is known as Hinkley's (1977) weighted *jackknife estimator* of $\mathrm{Var}(\hat{\beta})$. See MacKinnon and White (1985) for more details and alternative jackknife estimators. The degrees of freedom correction usually makes a big difference in VAR equations because of large k. In the alternatives of MacKinnon and White *studentized residuals* and *prediction errors* are used instead of ordinary residuals, so that remark A3.3 has to be taken into account, before the application of their formulae.

In the *jackknifing resampling procedure*, estimates of the model are recomputed T times, each time dropping one of the observations. The variability of the estimates is used as an estimate of the variability of the original estimator; Judge et al. (1985, p. 417). In the statistics literature the jackknife estimate is an example of an *internal* (distribution free) *scaling measure.* In econometrics one would call it a nonparametric measure. Scaling measures which interpretation is based on a statistical model, like the normal model are called *external*; (Cook and Weisberg (1982, p. 129). Algebraic identities like (A3.1.3) make computation for the linear model easy.

Remark A3.6 Direct computation of the measures involves the evaluation of the largest eigenvalue and eigenvector of a large (nonsparse) matrix. This may be costly for large T. Efficient algorithms exist; see e.g. Stewart (1973, p. 350).

Remark A3.7 See remark A3.4. The local influence analysis of case weight perturbation can also be applied to subsets of parameters of β. For single case analysis of one dimensional β_1 ℓ_{max_i} takes the simple form $r_i(X_1-\hat{X}_{12})_i$, where $\hat{X}_{12}$ is defined in remark A3.4.; see Cook (1986, §5). It follows that observations with $(X_1-\hat{X}_{12})_i=([I-X_2(X_2'X_2)^{-1}X_2']X_1)_i$ small, are not influential locally, whereas they may be globally (op cit.).

A3.2 Influence Measures for the Multivariate General Linear Model

This appendix gives the multivariate regression estimative influence measures discussed in §3.2.3. They are adapted from Johnson and Geisser (1985)[1]. We discuss the connection with the analysis of covariance test for parameter stability and the Chow statistic.

W	$X'X$, $k\times k$ moment matrix using all observations, with k the number of regressors,
X_i	$m\times k$ submatrix of X containing m observations the influence of which is to be examined, $m<T-k$,
W_i	$X_i'X_i$,
$X_{(i)}$	remaining part of X if X_i is deleted from (reordered) X, $(T-m)\times k$ matrix, $T-m>k$,
$W_{(i)}$	$X_{(i)}'X_{(i)}$,
Y_i	$m\times n$ submatrix of (reordered) Y corresponding to X_i.
$Y_{(i)}$	remaining part of Y if Y_i is deleted from (reordered) Y, $(T-m)\times n$ matrix,
H_i	$X_iW^{-1}X_i'$, $m\times m$ diagonal block of the hat matrix,
$\hat{B}$	OLS estimate for B using full sample, $k\times n$ matrix,
$\hat{B}_{(i)}$	OLS estimate for B using subset deleted sample $(Y_{(i)},X_{(i)})$,
$\hat{Y}$	$X\hat{B}$: $T\times n$ matrix of predicted values for Y using full sample,
$\hat{Y}_{(i)}$	$X\hat{B}_{(i)}$: $T\times n$ matrix of predicted values for Y using the subset deleted sample estimate of B,
R_i	$Y_i-X_i\hat{B}$: $m\times n$ matrix of OLS residuals of subset i,
$\bar{R}_i$	$Y_i-X_i\hat{B}_{(i)}$: $m\times n$ matrix of OLS residuals using parameter estimate $\hat{B}_{(i)}$, also known as (multivariate generalization of) the *prediction error* or as *predicted residual*,
A	$(Y-X\hat{B})'(Y-X\hat{B})$,
$A_{(i)}$	$(Y_{(i)}-X_{(i)}\hat{B}_{(i)})'(Y_{(i)}-X_{(i)}\hat{B}_{(i)})$,
$\hat{\Sigma}$	$A/(T-k)$: "Unbiased" OLS estimate of residual covariance matrix using full sample,
$\hat{\Sigma}_{(i)}$	$A_{(i)}/(T-m-k)$: OLS estimate of residual covariance with subset deleted sample,

---

[1] Johnson and Geisser (1985) used p for the number of equations (standard multivariate notation in Anderson (1984)), q for the number of regressors, k for the number of observations of interest and N for the total number of observations. We use n, k, m, and T. For n=1 we are back in appendix A3.1, see also remark A3.10 below.

d_{ii} $A^{-1}R_i'(I-H_i)^{-1}H_i(I-H_i)^{-1}R_i = A^{-1}(\hat{Y}-\hat{Y}_{(i)})'(\hat{Y}-\hat{Y}_{(i)})$ a generalization of Cook's distance, see (3.3), a measure of lack of fit,

e_{ii} $A^{-1}R_i'(I-H_i)^{-1}H_iR_i$,

t_{ii} $A^{-1}R_i'(I-H_i)^{-1}R_i$, the multivariate generalization of the (multiple observation) squared *internally studentized residual*,

T_{ii} $A_{(i)}^{-1}R_i'(I-H_i)^{-1}R_i = A_{(i)}^{-1}At_{ii} = (I-t_{ii})^{-1}t_{ii} = (I-t_{ii})^{-1} - I$,

L_i $(T-k-m)\mathrm{tr}T_{ii}$, the *Lawley–Hotelling trace criterion*, see e.g. Spanos (1986, p. 592) and Anderson (1984, Appendix B, C) for tables[2],

$\bar{T}_{ii}$ PF_i $:[(T-k-m)/m]\mathrm{tr}T_{ii}$, a multivariate generalization of the *Chow statistic for predictive failure*[3].

d_{ii}, e_{ii}, t_{ii} and T_{ii} all measure lack of fit, but relative to different metrics. Note $At_{ii} = Ae_{ii} + R_i'R_i$ by $(I-H_i)^{-1} = (I-H_i)^{-1}H_i + (I-H_i)^{-1}(I-H_i)$. For T_{ii} to exist one needs $I-t_{ii}$ to be of full rank: a sufficient number of observations in the subset deleted sample must remain.

The Bayesian estimative influence measures are defined as follows, see Johnson and Geisser (1985, form. (3.2), (3.4), and (4.1)):

$MEIM\Sigma_i$ $(T-k-m)[\ln|I-t_{ii}|^{-1} + \mathrm{tr}T_{ii}] - (T-k)[\mathrm{tr}t_{ii} - \ln|I-t_{ii}|] + c_1 + c_2$
Marginal Estimative Influence Measure relative to Σ^{-1}, with c_1 and c_2 constants, which are put to zero under computation, so that negative measures of this distance become possible,

$locB|\Sigma_i$ $(T-k)\mathrm{tr}\, e_{ii} + (T-k-m)\mathrm{tr}(I-t_{ii})^{-1}d_{ii}$,

$covB|\Sigma_i$ $n[\ln|I-H_i|^{-1} - \mathrm{tr}H_i] + n[\mathrm{tr}H_i(I-H_i)^{-1} - \ln|I+H_i(I-H_i)^{-1}|]$,

$CEIMB|\Sigma_i$ $locB|\Sigma_i + covB|\Sigma_i$,
Conditional Estimative Influence Measure relative to $B|\Sigma^{-1}$, decomposed into a location and a covariation component. That is, one component which measures the influence on the point estimates of B, and another one which measures the influence on the shape of the ellipsoidal posterior regions for the columns of B, i.e. the location parameters per equation.

---

[2] Anderson's notation (1984, Chapter 8, appendix C) corresponds to ours as follows. He used p, N, q, n, and m, where we use n, T, $k+m$, $T-k-m$ (the degrees of freedom parameter of the Lawley-Hotelling Statistic), and m. For models with lagged dependent variables, this Lawley-Hotelling distribution is not exact, but merely an approximation in finite samples, but it can be expected to outperform asymptotic χ^2-approximations in retaining nominal size. A small Monte Carlo experiment confirmed this expectation.

[3] For $n=1$ *or* $m=1$ its distribution under standard regression assumptions is $F(nm, T-k-\max(n,m))$. We prefer to call it a *multivariate Chow test* in order to avoid confusion with the *generalized Chow* test, which considers more than two subsamples. In the statistics literature the word *generalized* is often used as a synonym for multivariate, e.g. in *generalized variance*.

EIM_i $MEIM\Sigma_i + CEIMB\,|\,\Sigma_i$,
Estimative Influence Measure relative to (B,Σ^{-1}),

c $(T-m-k-2)/(T-k-2)$, a known constant degrees-of-freedom correction factor,

$locMEIMB_i$ $(T-k-m-2)\mathrm{tr}(I-t_{ii})^{-1}e_{ii}+(T-k-2)\mathrm{tr}d_{ii}$.

$covMEIMB_i$ $n[\ln|I-H_i|^{-1}-trH_i]+n[\mathrm{tr}H_i(I-H_i)^{-1}-\ln|I+H_i(I-H_i)^{-1}|]+$
$k[n(c-\ln c-1)+c\mathrm{tr}t_{ii}(I-t_{ii})^{-1}-\ln|I-t_{ii}|^{-1}]+$
$k[n(c^{-1}-\ln c^{-1}-1)+ln|I-t_{ii}|^{-1}-c^{-1}\mathrm{tr}t_{ii}]$
$-(trH_i)[c\mathrm{tr}t_{ii}(I-t_{ii})^{-1}+nc-n]-(trH_i(I-H_i)^{-1})[c^{-1}\mathrm{tr}t_{ii}-nc^{-1}+n]$,

$MEIMB_i$ $locMEIMB_i + covMEIMB_i$, the approximate Marginal Estimative Influence Measure relative to B, decomposed into a location and a covariation component.

Note the similarities and dissimilarities with the predictive influence measures of the previous section. Remember that H_i measures whether observations are *distantly observed* and t_{ii} measures whether observations exhibit *lack of fit*. It depends on the purpose of the estimation how much importance is to be attributed to each of these factors.

If $m>k$, i.e. when the subsample is no longer *undersized*, it is possible to use the following components of influence measures as well:

$\hat{B}_i$ $(X_i'X_i)^{-1}X_i'Y_i$,

$\tilde{R}_i$ $Y_i - X_i\hat{B}_i$,

A_i $(Y_i-X_i'\hat{B}_i)'(Y_i-X_i'\hat{B}_i)$,

s_{ii} $A^{-1}A_i$,

S_{ii} $A_{(i)}^{-1}A_i$,

q_{ii} $A^{-1}(\hat{B}_i-\hat{B})'(W_i^{-1}-W^{-1})^{-1}(\hat{B}_i-\hat{B})$,

Q_{ii} $A_{(i)}^{-1}(\hat{B}_i-\hat{B}_{(i)})'(W_i^{-1}+W_{(i)}^{-1})^{-1}(\hat{B}_i-\hat{B}_{(i)})$,

QT_i $\mathrm{tr}(Q_{ii})\cdot[(T-k-m)/m]$,

VR_i $\mathrm{tr}(S_{ii})\cdot[(T-m-k)/(m-k)]$, a generalization of the variance ratio statistic or Goldfeld–Quandt (1965) test statistic,

CH_i $\mathrm{tr}[(A_i+A_{(i)})^{-1}(A-A_i-A_{(i)})]\cdot[(T-2k)/k]$, a generalization of the Chow statistic for the parameter constancy of B or the "standard analysis of covariance test", see Chow (1960, form.(29) and (24)) for $n=1$, see remark A3.11 below[4].

Remark A3.8

In an unrestricted VAR the regressor matrix is the same as in the

[4] For the single equation case $n=1$ it is also known as the *Kullback–Rosenblatt F–ratio statistic*, see e.g. Cantrell et al. (1991).

univariate analysis. The multivariate character comes in by taking into account the cross correlations of the disturbances to the different equations. When these correlations are very small, the gains from the multivariate analysis will be low.

Remark A3.9

If $m=1$ and model (3.10) is true, we see that

$$L_i=(T-k-m)\mathrm{tr}A_{(i)}^{-1}R_i'(I-H_i)^{-1}R_i=\mathrm{tr}\hat{\Sigma}_{(i)}^{-1}R_i'(I-H_i)^{-1}R_i=R_i\hat{\Sigma}_{(i)}^{-1}(I-H_i)^{-1}R_i',$$

so that $L_i \to_{\mathcal{L}} \chi_n^2$, since $\underset{T\to\infty}{\mathrm{Plim}}\ \hat{\Sigma}_{(i)}(I-H_i)^{-1}=\Sigma$.

Remark A3.10

For $n=1$ the model collapses to the single equation of the previous section. The estimative influence measures can of course also be computed equation by equation for the interesting subsets of observations. The Chow PF_i is here equal to $L_i/m=\bar{T}_{ii}$, and still distributed as $F(m,T-k-m)$ under (3.10). Analogously with lemma A3.1 one can derive

$$A^{-1}A_{(i)}=(I+T_{ii})^{-1}=I-t_{ii}. \tag{A3.2.1}$$

Lemma A3.3

$$t_{ii}=s_{ii}+q_{ii}. \tag{A3.2.2}$$

Proof.

Note that by the matrix inversion lemma:

$$(W_i^{-1}-W^{-1})^{-1}=X_i'(I+X_iW_{(i)}^{-1}X_i')X_i. \tag{A3.2.3}$$

Rewrite (A3.2.2) as

$$(I-A^{-1}A_{(i)})=A^{-1}[A_i+(\hat{B}_i-\hat{B})'(W_i^{-1}-W^{-1})^{-1}(\hat{B}_i-\hat{B})]. \tag{A3.2.4}$$

Premultiply both sides by A and make the following subsequent substitutions: (A3.2.3), (A3.1.10), $X_i\hat{B}_i=X_iW_i^{-1}X_i'Y_i$ and $-X_i\hat{B}=R_i-Y_i$ to get

$$A-A_{(i)}=A_i+[R_i'+Y_i'(X_iW_i^{-1}X_i'-I)](I-H_i)^{-1}[(X_iW_i^{-1}X_i'-I)Y_i+R_i]. \tag{A3.2.5}$$

By (A3.1.10)

$$(I-H_i)^{-1}(X_iW_i^{-1}X_i'-I)=(X_iW_i^{-1}X_i'-I), \tag{A3.2.6}$$

and by using $\tilde{R}_i'X_i=0$ one obtains

$$Y_i'(X_iW_i^{-1}X_i'-I)R_i=-\tilde{R}_i'(Y_i-X_i\hat{B})=-\tilde{R}_i'\tilde{R}_i=-A_i=R_i'(X_iW_i^{-1}X_i'-I)Y_i$$

and

$$Y_i'(I-X_iW_i^{-1}X_i')Y_i=A_i.$$

Finally noting that the LHS of (A3.2.5) equals $R_i'(I-H_i)^{-1}R_i$ one has

$$R_i'(I-H_i)^{-1}R_i=A_i+R_i'(I-H_i)^{-1}R_i-A_i-A_i+A_i.\ \square$$

Analogously with lemma A3.3 one can derive (Johnson and Geisser(1985))

$$Q_{ii} = T_{ii} - S_{ii} = A_{(i)}^{-1}(A - A_{(i)} - A_i). \tag{A3.2.7}$$

Lemma A3.4

$$CH_i = [(I - q_{ii})^{-1}q_{ii}]\cdot[(T-2k)/k]. \tag{A3.2.8}$$

Proof.

Use $A_i = As_{ii}$, $A_{(i)} = A(I - t_{ii})$ and (A3.2.2) to get

$$(A_i + A_{(i)})^{-1}(A - A_i - A_{(i)}) = (A(s_{ii} + I - t_{ii}))^{-1}(A(I - s_{ii} - I + t_{ii})) = (1 - q_{ii})^{-1}q_{ii}.$$

□

In the one equation case $n = 1$ Brown et al. (1975, p. 154) noted that $A^{-1}A_i$ has a Beta distribution with mean $(m-k)/(T-k)$. See also table A3.1 below.

Remark A3.11

Analogously with well known results from analysis of covariance one has

$$(A - A_i - A_{(i)}) = (\hat{B}_i - \hat{B})'W_i(\hat{B}_i - \hat{B}) + (\hat{B}_{(i)} - \hat{B})'W_{(i)}(\hat{B}_{(i)} - \hat{B}), \tag{A3.2.9}$$

so that CH_i can be viewed as a generalization of the usual between groups over within groups ratio of mean squares when X is a vector of ones. In this sense CH_i tests the constancy of the mean of the error term.

Proof of (A3.2.9).

$$A - A_i - A_{(i)} = Y'Y - Y_i'Y_i - Y_{(i)}'Y_{(i)} - \hat{B}'W\hat{B} + \hat{B}_i'W_i\hat{B}_i + \hat{B}_{(i)}'W_{(i)}\hat{B}_{(i)} =$$

$$-\hat{B}'W_i\hat{B} - \hat{B}'W_{(i)}\hat{B} + \hat{B}_i'W_i\hat{B}_i + \hat{B}_{(i)}'W_{(i)}\hat{B}_{(i)}. \tag{A3.2.10}$$

Use

$$W_i\hat{B} + W_{(i)}\hat{B} = X'Y = W_i\hat{B}_i + W_{(i)}\hat{B}_{(i)}$$

in writing out (A3.2.9) to equate the right hand sides of (A3.2.9) and (A3.2.10). □

Remark A3.12

Johnson and Geisser (1985, p. 40) indicated that for $n = 1$, $VR_i \sim F(m-k, T-k-m)$ and $QT_i \sim F(k, T-k-m)$ when the model obtains. One does not often use QT_i in applied econometrics. The roles of the subsets can be changed for both VR_i and QT_i. For $n = 1$ one also has $CH_i \sim F(k, T-2k)$; see Chow (1960). For large subsets of interest $(m > k)$ CH_i is expected to be more powerful (Chow (1960), Dufour (1982)) than PF_i, when used as a test statistic for parameter constancy, see also Cantrell et al. (1991), but PF_i seems to keep the nominal significance level better in dynamic models, cf. §3.3.6. From the formula for

q_{ii} and (A3.2.8) we see that CH_i focuses primarily on the change in the parameter estimate for B from the subsample to the total sample. Note that under application with lagged dependent variables among the regressors, these distributions can only serve as approximations.

Note that by (A3.2.7) one can derive for $m > k$

$$PF_i = [k/m]QT_i + [(m-k)/m]VR_i, \tag{A3.2.11}$$

i.e. a decomposition of the Chow statistic for predictive failure into a location component $[k/m]QT_i$, which primarily measures the change in the estimate of B between two subsamples and a variance component, which primarily measures the change in the estimate of Σ. For $m = k$ the latter component disappears, but it is clear that it will dominate the first component when m, the number of observations in the subsample of interest grows faster than k, the number of regressors.

Summarizing in table A3.1 we have the following relations between well known test statistics and the influence measures, when $n = 1$ and $m > k$:

Table A3.1 *Test statistics expressed as a function of influence statistics (expression 1) and as a function of residual sums of squares (expression 2)*

Statistic	*expression* 1	*expression* 2	*Distribution*
PF_i	$c_1(1-t_{ii})^{-1}t_{ii}$	$c_1 A_{(i)}^{-1}(A - A_{(i)})$	$F(m, T-k-m)$
CH_i	$c_2(1-q_{ii})^{-1}q_{ii}$	$c_2(A_i + A_{(i)})^{-1}(A - A_i - A_{(i)})$	$F(k, T-2k)$
VR_i	$c_3 S_{ii}$	$c_3 A_{(i)}^{-1} A_i$	$F(m-k, T-k-m)$
QT_i	$c_4 Q_{ii}$	$c_4 A_{(i)}^{-1}(A - A_{(i)} - A_i)$	$F(k, T-k-m)$

NOTES: $c_1 = (T-k-m)/m$, $c_2 = (T-2k)/k$, $c_3 = (T-k-m)/(m-k)$, $c_4 = (T-k-m)/k$

Equivalent expressions for Beta distributed test statistic can be found in Anderson (1984, p. 301)[5]. One has

$$1-t_i^2 = A_{(i)}/A \sim B(\tfrac{1}{2}(T-k-m), \tfrac{1}{2}m), \tag{A3.2.12}$$

see (A3.1.7), and

$$(A_{(i)} + A_i)/A \sim B(\tfrac{1}{2}(T-2k), \tfrac{1}{2}k). \tag{A3.2.13}$$

Most statistical computer packages have options to compute p–values for Beta distributed statistics. These ratios lie between 0 and 1.

---

[5] Anderson's notation should be translated as follows. For the beta version of PF_i in (A3.2.12): $g_{11}=T^{-1}A_{(i)}$, $g_{11}+h_{11}=T^{-1}A$, $n=T-k-m$, $m=m$ and for the beta version of CH_i in (A3.2.13): $g_{11}=T^{-1}(A_{(i)}+A_i)$, $g_{11}+h_{11}=T^{-1}A$, $n=T-2k$, $m=k$.

Remark A3.13

When Σ is diagonal, the multivariate L_i is just the sum of the L_is for the single equations.

Remark A3.14

Abraham and Chuang (1989), referred to in §3.2.3, used the following notation (with $n=1$): $Q_k = At_{ii}$, $Q_{k1} = R_i'R_i$, $Q_{k2} = Ae_{ii}$. Q_{k2} can be used to detect additive outliers, Q_k and Q_{k1} can be used to detect both additive and innovation outliers. They noted (see also Cook and Weisberg (1982, p. 124)) that

$$Ae_{ii} = (\hat{B} - \hat{B}_{(i)})'W_{(i)}(\hat{B} - \hat{B}_{(i)}), \tag{A3.2.14}$$

which illustrates that e_{ii} measures the displacement of the fitted values of $Y_{(i)}$. The influence of H_i is bigger in e_{ii} than in t_{ii}. With an autoregressive model like (3.9) and an additive outlier at $t=i$, the regressor matrix at $t=i+1$ will surely be affected, whereas this is not so sure with an innovation outlier. An exceptionally high value for $e_{i+1,i+1}$ can therefore be used as an indication of an additive outlier at $t=i$, according to Abraham and Chuang (1989).

Remark A3.15

If one defines $Ae_{(ii)} = (\hat{B} - \hat{B}_i)'W_i(\hat{B} - \hat{B}_i)$, $R_{(i)}'R_{(i)} = A - R_i'R_i$, and $At_{(ii)} = Ae_{(ii)} + R_{(i)}'R_{(i)}$, so that $t_{(ii)}$ is proportional to the squared internally studentized residual when the remaining $T-m$ observations are viewed as subset of interest, one can rewrite q_{ii} as follows:

$$q_{ii} = t_{ii} + t_{(ii)} - I. \tag{A3.2.15}$$

Proof.

Substitute the left hand side of (A3.2.9) in $Aq_{ii} = A - A_i - A_{(i)}$ and use the new definitions. □

One sees that q_{ii}, and thus its monotone transformation CH_i, is dominated by $t_{(ii)}$ when $T-m$ is much greater than m.

A3.3 Influence Measures in Principal Components Analysis

Here we present the influence measures discussed in section 3.2.3. They are adapted from Critchley (1985).

Let F be a cumulative distribution function defined on $\mathbb{R}^n$, let $\mu(F)=\int x dF(x)$ be a vector of means, and let $\Omega(F)=\int\{x-\mu(F)\}\{x-\mu(F)\}'dF(x)$ denote its covariance matrix, which we assume to exist and have distinct eigenvalues $\lambda_1(F)>\ldots>\lambda_n(F)$, and corresponding orthonormal eigenvectors $\alpha_1(F),\ldots,\alpha_n(F)$ all of which are stacked into a single vector $T(F)\in\mathbb{R}^{n(n+1)}$. $T(F)$ forms the basis of the influence analysis of a full principal component analysis of a covariance matrix.

The perturbation scheme is of the gross–error type. Let δ_z denote the distribution giving unit mass to a point z in $\mathbb{R}^n$. Then for $0\leq\varepsilon\leq 1$ we have the identity

$$\Omega\{(1-\varepsilon)F+\varepsilon\delta_z\}=$$

$$\Omega(F)+\varepsilon[\{z-\mu(F)\}\{z-\mu(F)\}'-\Omega(F)]-\varepsilon^2\{z-\mu(F)\}\{z-\mu(F)\}', \qquad \text{(A3.3.1)}$$

which we regard as a perturbation of $\Omega(F)$, and which can be decomposed in a first order and a second order effect in ε.

The *empirical influence curve*, $\hat{I}(z,\lambda_j,F)$, at $z=x_i$ is defined in two components as:

$$\hat{I}_{ij}(\lambda)\equiv\hat{I}(x_i,\lambda_j,F)=y_{ij}^2-\hat{\lambda}_j=\hat{v}_{ij} \qquad \text{(A3.3.2)}$$

and

$$\hat{I}_{ij}(\alpha)\equiv\hat{I}(x_i,\alpha_j,F)=-y_{ij}\sum_{k\neq j}y_{ik}(\hat{\lambda}_k-\hat{\lambda}_j)^{-1}\hat{\alpha}_k=\hat{\beta}_{ij}, \qquad \text{(A3.3.3)}$$

with $\hat{\lambda}_j$ the j–th ordered eigenvalue with its eigenvector $\hat{\alpha}_j$ and y_{ij} the element (i,j) of the $(T\times n)$ matrix of principal component vectors y, all derived from the sample covariance matrix of $\Omega(F)=T^{-1}\sum(x_i-\bar{x})(x_i-\bar{x})'$.

The corresponding *empirical deleted influence curve* $\hat{I}_{(i)}(z,\lambda_j,F)$, at $z=x_i$, based on $\Omega(\hat{F}_{(i)})$ is defined as $\hat{I}_{(i)j}(\lambda)\equiv\hat{I}_{(i)}(x_i,\lambda_j,F)=\hat{v}_{(i)j}$ and $\hat{I}_{(i)j}(\alpha)\equiv\hat{I}_{(i)}(x_i,\alpha_j,F)=\hat{\beta}_{(i)j}$. It can be computed as follows.

Define

$$\begin{aligned}
b_{(j)}(r,s) &= \sum_{k\neq j}a_k^r(\lambda_k-\lambda_j)^{-s},\\
\pi_j &= -2a_j^2\{1+b_{(j)}(2,1)\},\ \text{with}\\
\gamma_j &= -a_j^2b_{(j)}(2,2)\alpha_j-2b_{(j)}(2,1)\beta_j-2a_j^3\sum_{k\neq j}a_k(\lambda_k-\lambda_j)^{-2}\alpha_k.
\end{aligned}$$

Take $a_j=y_{ij}$, and compute

$$\hat{I}_{(i)j}(\lambda) = \hat{v}_{ij} - (T-1)^{-1}(\hat{\pi}_{ij} - \hat{v}_{ij}), \tag{A3.3.4}$$

$$\hat{I}_{(i)j}(\alpha) = \hat{\beta}_{ij} - (T-1)^{-1}(\hat{\gamma}_{ij} - \hat{\beta}_{ij}). \tag{A3.3.5}$$

The corresponding *sample influence curve* is an average of the empirical influence curve and the empirical deleted influence curve:

$$\tilde{I}_{(i)j}(\lambda) = \hat{v}_{ij} - \tfrac{1}{2}(T-1)^{-1}(\hat{\pi}_{ij} - \hat{v}_{ij}), \tag{A3.3.6}$$

$$\tilde{I}_{(i)j}(\alpha) = \hat{\beta}_{ij} - \tfrac{1}{2}(T-1)^{-1}(\hat{\gamma}_{ij} - \hat{\beta}_{ij}). \tag{A3.3.7}$$

Critchley discussed the relation with *EIC*s and *SIC*s in the regression case, cf. 3.2.2. The influence measure for sets of observations is simply the sum of the influence measures for the separate observations. The three influence measures converge to each other asymptotically, but may give quite different influence orderings of the observations in finite samples.

4 SEASONALITY

4.1 Introduction

Seasonal variance can be a nuisance. One can get rid of this nuisance only if one knows the intricacies of the data and if one has a precise formulation of the problems that make one think seasonal variance to be troublesome.

The majority of economic models does not explain the seasonal patterns in macroeconomic variables. The difference between seasonal patterns in related variables and their gradual change over time are usually not accounted for either. This may underlie the preference of economists to use seasonally adjusted data in empirical research.

Econometricians are now barely able to cope with the consequences of the possibility of unit root nonstationarity at the zero frequency for the statistical interpretation of the sample variances and covariances. Digesting seasonal nonstationarity as well may be too much. The problem of testing for seasonal unit root nonstationarity versus deterministic nonstationarity has been attacked successfully for univariate series, see appendix A2.3. The analysis and treatment of seasonality in multivariate time series with feedback is still an open problem, see Ghysels (1990b) and Sims (1993).

Ghysels (1990b) gave an overview of different statistical models for seasonal data, each related with different formulations of the problems that economists and econometricians try to work out. He also edited an issue of the Journal of Econometrics on seasonality and econometric models (JoE (1993)). Nerlove et al. (1979) discussed the history and underlying ideas of the unobserved component approach for economic time series, and suggested several methods of implementation for the non–unit root case. Bell and Hillmer (1984) discussed a range of criteria and methods for univariate seasonal adjustment from an unobserved components with unit roots (UCARIMA) point of view and stressed the importance of checking the assumptions for this measurement model. Hylleberg (1986) analyzed a range of seasonal models from an econometric point of view. He compared an impressive number of diagnostic checks and estimation techniques in Monte Carlo studies and empirical econometric models. Dynamic regressions of stationary variables without feedback received a lot of attention in his study. Hylleberg (1992) contains

many relevant articles. Harvey (1989) gave an econometric interpretation to special univariate UCARIMA models, called them *structural time series models*, and developed a number of estimation techniques for them. These models were in first instance developed for forecasting purposes, but later also put to use in modeling dynamic relationships. Porter–Hudak (1990) discussed developments and applications of fractionally integrated time series models to seasonal series.

The analysis of seasonality cannot be successful without modeling the trend and irregular component of the series. The multivariate analysis of the trend components is subject of Chapter 6. The irregular component can contain outliers with a big impact on the estimates of the seasonal component and on the parameters of interest. Outliers are discussed in the next chapter.

4.2 Application of the idea of unobserved components

The treatment of seasonality as a separate aspect of a time series stems from the old idea of the decomposition of a time series. The additive version reads:

$$x_t = y_t + u_t. \tag{4.1}$$

One observes x_t, and by making a number of reasonable identifying assumptions on the component models one may resolve y_t and u_t. In this study we treat seasonality as one of the sources of error u_t in macroeconomic time series, adopting the view of Pierce (1981)[1]. The variation in x_t can then be decomposed into the variation of interest relating to x_t and the variation in the error term u_t. The main identifying assumptions are that variation in the seasonal error has most of its power in narrow bands around the seasonal frequencies and that this error is uncorrelated with the component of interest y_t. The model with constant seasonal dummies which figured as an alternative in the unit root tests in §A2.2.2 incorporates these assumptions. For quarterly observations it can be written as:

$$u_t = \alpha_0 + \alpha_2(-1)^t + \alpha_3\cos(\tfrac{1}{2}\pi t) + \alpha_4\sin(\tfrac{1}{2}\pi t) + \varepsilon_t, \tag{4.2}$$

with ε_t the nonseasonal part of the error term. One often incorporates this component implicitly by extending all equations of the VAR (2.1) with seasonal dummies. We discuss that model in appendix A4.2.

The model (4.2) allows for arbitrary changes in seasonal means within the

[1] This fits in the Dutch tradition of speaking of seasonal correction, rather than using the term seasonal adjustment, see Den Butter and Fase (1988).

year but assumes a constant seasonal pattern from year to year, which is an unreasonable approximation for many empirical macroeconomic time series of a considerable length.

The use of more flexible deterministic alternatives was advocated in the seventies, inspired by frequency domain methods of estimation. These boil down to the extension of (4.2) with trigonometric terms with variation in narrow bands around the seasonal frequencies (Sims (1974), Laroque (1977)). Although the resulting methods of estimation can be asymptotically justified for stochastic models of seasonality as well[2], see Bunzel and Hylleberg (1982), they are unsatisfactory when the seasonal pattern changes slowly and persistently.

Strong seasonality

In that case the model with *seasonal unit roots*, discussed in detail in Hylleberg et al. (1990), may be a better approximation. A good expression for this property is *strong seasonality*, which Granger (1984) used for series having infinite variance at the seasonal frequencies. Since we are mainly interested in the persistence of seasonal changes, we will formally call a series *strongly seasonal* when its spectrum is infinite at one of the seasonal frequencies, i.e. when there is a positive integer j smaller than or equal to half the number of observations per year S so that the series is $I_{j/S}(d)$ with $d>0$, see §2.4.1. This nomenclature fits in with the terminology of *strong correlation* for autocorrelation of positively fractionally integrated processes, see Robinson (1991a). The variance of an $I_{j/S}(d)$ process is infinite if $d \geq ½$, see Gray et al. (1989, Theorem 1). Gray et al. (1989) gave a natural extension of the definition of "seasonal" integration for j/S, $d \in \mathbb{Q}$. Their corresponding Gegenbauer ARMA models (GARMA) can be used to model persistent periodicity in the autocorrelation function at any frequency.

The properties of models for *seasonal integration* are most easily understood in the frequency domain. Identification methods for simple models of this class, so–called *seasonal Box–Jenkins models* have been in use in time series analysis for a long time, see Hylleberg (1986). Sharp peaks at the seasonal frequencies in the AR estimates of the spectrum of the univariate series using (3.15) can indicate the usefulness of the seasonal unit root approximation.

The test procedure in appendix A2.2 can also be applied to the unrestricted univariate series to check whether the seasonal unit roots model

---

[2] One can use the econometric interpretation of generalized least squares in the frequency domain, if u_t is either stationary or if it follows a circular process, see e.g. Harvey (1989, p. 377) and A4.1.

provides a better approximation than seasonal dummies. That procedure allows one to discriminate between integer unit roots at different seasonal frequencies, which is not so common in traditional time series analysis. In principle one can think of economic time series where the yearly cycle is only nonstationary within the year, while two– or three month cycles are nonstationary from year to year as well.

Interpretation

Classical determinants of the yearly cycle are climatological factors like air temperature, which may be considered stationary for long periods. Persistent changes in higher frequency cycles may be attributed to changes in average holiday periods, which in turn are due to demographic, technological or institutional changes. In quarterly series we can only observe half–yearly and yearly cycles.

Most of the changes in seasonal patterns in the series that we analyze in chapter 7 can be traced back to gradual shifts in the relative sizes of different sectors of the economy. Some of the factors determining these shifts may be correlated with movements in the "business cycle", which is typically considered to be a part of y_t. These remarks are based only on the observation of a few empirical facts, and have little to do with the analysis of real business cycles and sectoral shifts in two sector Robinson Crusoe economy models as discussed in, e.g., Cooper and Haltiwanger (1990) and its references, where the analysis is based on unobservable shocks. The only thing in common is the preference to use seasonally adjusted data in a VAR for the empirical analysis of economic relationships. Persistence of these changes indicates that "structural" changes must be the dominating factors, and that the independence assumption in (4.1) may not be too unreasonable from an economic point of view[3].

Up till now we have not observed interesting economic equilibrium relations between the sizes of different sectors of the economy with different seasonal investment, production and employment patterns which may indicate seasonal cointegration, see Hylleberg et al. (1990). Kunst (1990) found statistical evidence of such long run equilibrium relationships at seasonal frequencies using system estimation techniques. He did not use the traditional unobserved components interpretation. By likelihood ratio tests (cf. Lee (1992)) he found that $\Phi(e^{i2\pi j/S})$ in (2.1) had a reduced rank for $j=2$ and $j=4$,

[3] Mixed multiplicative additive seasonality can also give rise to correlation between seasonal variation and business cycle variation, but this effect cannot easily be analyzed using linear theory, see Harvey (1989, p. 174) for a mixed model.

and interpreted this "error correction at the seasonal frequencies" as a sign of seasonal cointegration. See §A4.2.4 and §6.2.3.

Barsky and Miron (1989) emphasized the importance of the deterministic part of the seasonal cycle in economic models. Given the construction process of our data, see §7.3, the deterministic part of the seasonal cycle cannot be considered relevant for any economic model for these data. Canova and Hansen (1992) discovered some relevant flaws in Barsky and Miron's statistical analysis of the seasonal movements.

We thus support the "general view that while seasonality in one variable may be related to the seasonality in other economic variables with which that variable interacts, ultimately the seasonal component reflects the effects of noneconomic factors that are exogenous to the economic system and uncontrollable," Wallis (1982). On lower levels of aggregation of economic time series data this view may lead one to overlook relevant empirical information about dynamic economic relationships between variables, Ghysels (1990b). One then moves the model for the seasonal variation from the measurement model to the economic behavior part of the model, see §1.3.

From an econometric point of view it is interesting to include special variables determining seasonal variation, such as the number of days of snow combined with severe frost etc. The curse of dimensionality combined with the high level of spatial, sectoral and temporal aggregation of our series prompt us not to follow this approach.

4.3 Application of linear filters to estimate unobserved components

For a quadratic loss function in the observation errors of the component of interest, the minimal mean squared error estimate of y_t is optimal. Given stationary short memory purely stochastic linear models for y_t and u_t and consequently for x_t one can derive the optimal so–called signal extraction filters for y_t.

There are several problems connected with this approach. First, it is essentially univariate and optimal univariate signal extraction does not imply optimal multivariate extraction. Second, reasonable empirical models for y_t and u_t are often not both short memory and purely stochastic. Third, one needs specific (preferably low dimensional) parametric models for y_t and u_t to estimate them. Fourth, the optimality results are only derived for doubly infinitely long samples. Fifth, the presence of outliers in series is often not adequately accounted for by linear models.

4.3.1 Optimal extraction in multivariate series

The first problem has not been successfully attacked yet. Using results

derived for simple stationary one–directional distributed lag models (see A4.1) one may infer that using the same linear filter for all series of interest does the least damage to the estimates of linear lead–lag relationships between variables. This filter should remove all seasonal variation, but it should leave the dynamic properties of the series at other frequencies, especially the low ones, intact.

A comparable issue occurred in the discussion (Pierce (1977)) on the effect of prewhitening on causality tests. Separate linear prefiltering of univariate series to "correct for autocorrelation" before doing subsequent causality tests is all right if one has strong a priori beliefs on the lack of a substantial linear relationship between the variables of interest. Separate prewhitening may be an unlucky choice if one intends to detect these relationships when they are there, see Sims (1977). Sims's basic critique was that one will not estimate a "good" univariate model for a series if one ignores the information of related series. "Bad" estimates arise due to omitted variable bias. The estimated residuals of univariate models will only have a strong correlation with the "true innovations" if the "true" univariate model is simple, selected correctly and estimated with a reasonable precision, which is an appropriate assumption if one has a homogeneous sample with thousands of observations. This is not very likely for our series.

Ghysels (1990b) provided a more in depth discussion on the subject of seasonal correction and time series relationships between economic variables. He did not discuss the situation where the use of seasonal adjustment filters is most needed in practice: namely when there is an indication of seasonal nonstationarity. Sims (1993, pp. 16–19) indicated that seasonal adjustment can be advantageous, especially when the seasonal pattern is highly predictable.

When there is no indication of any seasonal unit root in the system, seasonal adjustment may become a liability instead of an asset: "We recommend not adjusting series that follow ARIMA models without a seasonal difference because the seasonal pattern can change too rapidly," Bell and Hillmer (1984). The main concern that we have with leaving economically irrelevant seasonal components with a large variance in the system is that they may influence the estimates of impulse responses and variance decompositions in an undesirable way, as illustrated for zero frequency unit root components by Ohanian (1988) using data for the U.S. By using adjusted series, one possible source of spurious relationship in the sense of Granger is removed, see Bell and Hillmer (1984, p. 300).

4.3.2 Optimal extraction in nonstationary series

Bell (1984) solved part of the second problem and derived optimal univariate signal extraction formulae for models where either y_t and/or u_t have unit root nonstationarity. A remarkable result from this analysis is that the minimal mean squared error estimate of the nonseasonal component of interest is $I_{j/S}(-1)$ if the seasonal error component is $I_{j/S}(1)$. An $I_{j/S}(d)$ series has spectrum zero at frequency j/S for every $d<0$. Dips in the spectrum of a seasonally adjusted time series are not necessarily a sign of "overcorrection" as was pointed out by Grether and Nerlove (1970). A zero spectrum at the seasonal frequencies does not have to be either. In the time domain residual autocorrelation at seasonal lags will be a result. Since the resulting so-called *antipersistent* series which are $I_{j/S}(d)$ with $d<0$ cannot be approximated by AR models with a small number of lags the residuals of a VAR model of optimally adjusted strongly seasonal series cannot be expected to have a white noise spectrum.

Most officially adjusted series (e.g. data produced by the OECD), although not adjusted optimally, are also $I_{j/S}(d)$ with $d<0$, unless the raw series were $I_{j/S}(d)$, with $d\geq 2$. We don't know of any macroeconomic series that was approximated significantly better by a model with $d\geq 2$ than by one with $d<2$.

We thus infer that a dip in the spectrum of officially adjusted series is the rule, rather than the exception. The often advocated test serial correlation at lags of one or two years in the residuals for linear time series models relating these series should point this out too. A rejection of whiteness of the residual spectrum should not be interpreted as a bad sign of the model, but merely as a logical consequence of the measurement model used to extract the component of interest.

So a pursuit of "pure" white noise residuals in a model for the adjusted data is not useful. Bell and Hillmer (1984) introduced a "UCARIMA" definition of a *data consistent* model which application avoids genuine overcorrection. They emphasized the differences between the spectral characteristics of the underlying series and its estimate[4].

A class of UCARIMA models that can mimic a lot of the properties of quarterly macroeconomic series is the following:

$$x_t = y_t + u_t, \qquad t = 0, \pm 1, \pm 2, \ldots, \qquad (4.3a)$$

[4] Ansley and Wecker (1984) suggested a less accurate but spectrum preserving extraction formula for stationary series y_t and u_t.

$$D_1(L)y_t = \mu + \Phi(L)^{-1}\varepsilon_t, \tag{4.3b}$$

$$D_2(L)u_t = \Theta(L)\eta_t. \tag{4.3c}$$

$D_1(L)$ is the filter that makes the spectrum of y_t finite and bounded away from zero at the frequencies of interest. μ is a constant term. $D_2(L)$ is the filter that makes the spectrum of u_t finite at all frequencies. ε_t and η_t are mutually independent Gaussian white noise processes, with variances σ_ε^2 and σ_η^2. $\Phi(L)^{-1}$ and $\Theta(L)$ are filters that make the assumptions for ε_t and η_t reasonable for a particular observed series x_t. $\Phi(z)=0$ has all its roots outside the unit circle, $\Theta(z)=0$ has all its roots on or outside the circle. Consider

Example 4.1

Set $D_1(L)=(1-L)$ and $D_2(L)=(1+L+L^2+L^3)$. It follows that $x_t-\mu t$ is $I(1)$. If $\Theta(z)=0$ has all its roots outside the unit circle it follows that x_t is $I_{\frac{1}{2}}(1)$ and $I_{\frac{1}{4}}(1)$. Take $\Phi(L)=(1-0.4L)$ and $\Theta(L)=(1-\Theta_2 L)(1-\Theta_4 L^2)$, with $\Theta_2=0.7$ and $\Theta_4=0.9$ as "realistic" examples for the other polynomials and take $\sigma_\varepsilon^2=\sigma_\eta^2=1$. Since our main focus is the seasonal component one can take $\mu=0$.

Bell (1984) extensively discussed assumptions about starting values and generation schemes for the model if x_t is unit root nonstationary. We adopt the simplest assumptions, which he denoted by *A*. For example 4.1 it entails that the vector of *starting values* $x_*=(x_1,x_2,x_3,x_4)$ is generated independently of the doubly infinite realizations $\{u_t\}$, $\{y_t\}$ and $\{w_t\}$, where $\{w_t\}=\{(1-L^4)x_t\}$. Since $D_1(z)$ and $D_2(z)$ have no common roots, $y_*=(y_1)$ and $u_*=(u_1,u_2,u_3)$ can be uniquely derived from x_* (op cit.). Given x_* and $\{w_t\}$, the remaining x_t–values can be generated recursively for $t>4$ from $x_t=x_{t-4}+w_t$ and for $t\leq 0$ from $x_t=x_{t+4}-w_{t+4}$.

By straightforward application of Theorem 2 in Bell (1984) one finds the minimal mean squared error estimate of y_t, which equals its conditional expectation given all x_t–values:

$$E(y_t|\{x_t\}) = s(L)x_t, \tag{4.4a}$$

with

$$s(L) = D_2(L)D_2(L^{-1})\gamma_1(L)\left[D_2(L)D_2(L^{-1})\gamma_1(L)+D_1(L)D_1(L^{-1})\gamma_2(L)\right]^{-1}, \tag{4.4b}$$

where $\gamma_1(z)=\sigma_\varepsilon^2\left[\Phi(z)\Phi(z^{-1})\right]^{-1}$ and $\gamma_2(z)=\sigma_\eta^2\Theta(z)\Theta(z^{-1})$ are the autocovariance generating functions of $D_1(L)y_t$ and $D_2(L)u_t$. The *pseudo spectral density* $f_{xx}(\omega)$, which is a useful concept in UCARIMA analysis is derived from

$$f_{xx}(\omega) = f_{yy}(\omega)+f_{uu}(\omega),$$

with $f_{yy}(\omega) = |D_1(e^{i\omega})|^{-2}\gamma_1(e^{i\omega})$, $f_{uu}(\omega) = |D_2(e^{i\omega})|^{-2}\gamma_2(e^{i\omega})$. One can see from formula (4.4*b*) that it corresponds to the solution to the *historical filtering problem* for *stationary* series, see e.g. Koopmans (1974, p. 148):

$$s(z) = f_{yy}(z)\left(f_{xx}(z)\right)^{-1}.$$

The generating function $s(z)$ is real valued. This entails that $s(L)$ is symmetric. The information of future x_t–values is used when it is available, hence the name *historical filtering*. When a filter is not symmetric its application induces a phase shift at certain frequencies. This is most easily illustrated by application to a strictly periodic input as in (A2.2.19). This is one of the disadvantages of applying naive filters like $D_2(L)$. Turning points in $\hat{y}_t$ (which is usually seen as a mixture of the trend and the business cycle component) will then be dated too late.

Since $s(z)$ is a (pseudo) autocovariance generating function, the real roots of $s(z) = 0$ occur in reciprocal pairs and all complex roots in quartets of conjugates and reciprocals, see Nerlove et al. (1979, p. 128). From (4.4) it follows that the optimal estimate of y_t is $I_{1/j}(-1)$ if x_t contains a seasonal with an $I_{1/j}(1)$ component.

The extraction formula (4.4) shows that one cannot adjust in an optimal way if one does not take the model for the component of interest into account. This holds for traditional deterministic adjustment as well. A sequence of traditional preprocessing seasonal dummy regression estimates in a trending series will show an intra–year trend, which must not be mistaken for a part of the seasonal intra–year component. A flexible model for the trend and subsequent extension of the regressor set will produce more sensible estimates of the deterministic seasonal component, see Jorgenson (1967).

4.3.3 Specification of low dimensional univariate models

The third problem of the specification and identification of models for y_t and u_t consists of two stages. One first has to choose identifying restrictions which underpin the interpretation of the components. Not all of these identifying restrictions can be tested empirically. The main identifying restriction is the orthogonality of the seasonal component and the component of interest. These two components can always be constructed in such a way that the theoretical spectrum (or the autocorrelations) of x_t fits the empirical one. In order to obtain consistent estimates of the components one then has to choose some reasonable a priori smoothness restrictions for their (pseudo) spectra (or autocorrelations of the appropriately differenced series). These second stage restrictions can be tested on the data, either by

testing for serial correlation in the white noise components, or by actually comparing different parameterizations.

Some UCARIMA parameterizations

The model in example 4.1 is a comparatively large. Harvey (1989) set $\Theta(L)=1$ and called (4.3c) a *structural time series model* for the component with *dummy variable seasonality*. This *smoothness restriction* is commonly used as the mean of a prior distribution used in Bayesian forecasting models (op cit. p. 40). Additionally setting $D_2(L)=(1+L^2)$ gives an example of *trigonometric seasonality* (op cit. p. 510).

The main properties of the models change if the parameters come (close) to the boundaries of the parameter space. This induces problems (even asymptotically) in the regularity conditions for straightforward statistical interpretation of well known measures of fit like the likelihood ratio.

Continuous transition to deterministic seasonality

An interesting choice is $\Theta(L)=D_2(L)=(1+L+L^2+L^3)$. One then obtains a solution to (4.3c) of the form[5] $u_t=\delta_2(-1)^t+\delta_3\cos(\frac{1}{2}\pi t)+\delta_3\sin(\frac{1}{2}\pi t)+\eta_t$. The spectral density of u_t at $\frac{1}{2}\pi$ and π is no longer finite, not as a pole, but as a spike or mass point. One obtains *deterministic nonstationarity*. The same effect occurs obviously if one sets $\sigma_\eta^2=0$. Note that $D_2(L)$ tends to become a common factor of numerator and denominator in (4.4) when either $\Theta(z)$ tends to $D_2(z)$ or σ_η^2 tends to zero. The resulting dips in the spectrum of $\hat{y}_t$ become narrower. The inverse spectrum of $\hat{y}_t$ tends to a spike function as illustrated in fig. 2.4.1(c) in Harvey (1989, p. 62). In this way one can model a continuous transition from unit root seasonality to deterministic seasonality. In the context of this structural UCARIMA model for y_t and u_t non rejection of seasonal unit roots in the *reduced form* ARMA model for x_t indicates both that σ_η^2 is not negligible and that $\Theta(L)$ and $D_2(L)$ differ substantially, provided a sufficiently long sample is used.

Near stationarity

If one defines a model for $t>0$, where Θ_1 and Θ_2 in example 4.1 tend to one as the number of observations tends to infinity, e.g. by letting $\Theta_{iT}=(1-T^{-a})$, $0<a<\frac{1}{4}, i=1,2$ one gets so-called *nearly stationary models*. It is clear that it will be difficult to tell the difference between a model with deterministic nonstationarity and such a nearly stationary model in finite samples. A model with large different starting values for u_1, u_2 and u_3 (e.g. 2, −4, 4, $\sigma_\eta^2=1$) in our example is at first sight easily misjudged as being governed by deterministic nonstationarity only. Unit root tests do not have

---

[5] For $\eta_t=0$, $t=1,2,3,4$ one gets $\delta_2=\frac{1}{2}(u_1+u_3)$, $\delta_3=\frac{1}{2}(u_1-u_3)$, $\delta_4=\delta_2+u_2$.

augural properties. In finite samples they reject unit root nonstationarity far too often when Θ_1 and Θ_2 are close to one. Pantula (1991) provided evidence for zero frequency examples.

Univariate quarterly macroeconomic time series data will usually not contain enough information to discriminate between different specifications, unless these are very simple. This has led to the peaceful coexistence of a number of specifications. Hausman and Watson (1985) compared different UCARIMA models and chose the alternative with deterministic seasonal means. They also used a priori information on the measurement process to identify an extra component of the error term u_t. Kitagawa and Gersch (1984, eq. 3.5d) even considered $I_{j/S}(2)$ processes in their set–up. Simplicity (or parsimony) of the models is essential for efficient estimation of parameters of interest. It is hard to think of parameters of economic interest in a univariate specification of a macroeconomic time series. The persistence measures of §2.5.3 may be an exception, but these are usually not very well identified in the data.

In empirical applications of the direct estimation of UCARIMA models the estimated Θ_i–values are often found to be rather close to the unit circle, which indicates that the seasonal pattern changes slowly, see e.g. Hillmer and Tiao (1982), or Bell and Hillmer (1984). If one subsequently computes the coefficients for the implicit optimal filter $s(L)$ one finds that these only slowly go to zero at high seasonal lags. This in turn entails that the optimally seasonally adjusted data for a fixed point in calendar time will change substantially for a large number of years, even if one considers the parameters of (4.3) to be known with infinite precision. This is a nuisance for official statistical agencies, which have to produce seasonally adjusted data for policy analysis. Many use the Census X–11 method which involves a symmetric filter where only the last three years of adjusted observations are likely to undergo a change. These changes can be quite considerable, a fact that is "increasingly recognized". Burridge and Wallis (1985) calculated the resulting variance of the seasonal measurement error. Figure 3 of the change in Dutch GDP data for 1988–1990 in CBS (1991) provides a clear example of the relevance of this problem for Dutch macroeconomic data.

We are only interested in the frequency response of the symmetric filter $s(L)$, and want to obtain estimates of y_t that are robust to seasonal nonstationarity of u_t, so that subsequent correlations in a multivariate analysis are not due to slow changes in seasonal patterns of different univariate series. "The choice between filtering and component ARIMA modeling is largely one of practicality," Cleveland (1983, p. 53). We specify the filter directly which implies the specification of a seasonal component

model. This allows direct control of the amount of smoothness in the quarterly subseries of the seasonal component which can then be judged by graphical displays, like figure 7.6 below.

4.3.4 Optimal extraction in a finite sample

The fourth problem of having only a finite number of observations can be solved by extending the observation set with the optimal minimal mean squared error out–of–sample forecasts and backforecasts. Application of the filter $s(L)$ to this extended data set produces the expectation of y_t given the observations in the finite sample, see Bell (1984, p. 662), under the assumption that one knows the true parameters.

This so–called DeForest extension is already known for more than a century, see Cleveland (1983). Hausman and Watson (1985) once analyzed the extra uncertainty due to the estimation of the parameters and found it to be negligible, although their "point estimates of the parameters were not extremely precise."

It seems sensible to use a general VAR to produce these out of sample predictions, see appendix A4.2. The other variables in the VAR are expected to contain information about future values of the series of interest, otherwise they would not have been included. It is a small additional effort to prevent the domination of a priori restrictions on the final results of the analysis and it provides one with an extra insight in the dynamic properties of the multivariate series.

4.3.5 Optimal extraction in the presence of outliers

The fifth problem of outliers is not solved as easily. The signal extraction formula still provides the best linear estimate of the underlying component when the interesting shocks ε_t and seasonal shocks η_t are mildly nonGaussian, see Bell (1984). Extreme outliers deserve a separate treatment. The occurrence of exceptional shocks makes the statistical interpretation more difficult, but it can make the identification of the impulse responses easier. This prompts us not to classify the exceptional observations as part of the error term u_t. We discuss this in more detail in the next chapter. The relevant point here is that for optimal seasonal adjustment one needs a model for outliers as well.

4.4 Data analysis of the seasonal component

How can one quickly get an idea from the data whether the seasonal pattern changes slowly and persistently? The first thing one has to do to

obtain a clearer picture of the seasonal component is to remove other nonstationarities from the series, either by power transformations, log transformations, detrending, differencing or by combinations of these procedures. When there are indications of zero frequency unit roots, differencing is preferable, although this operation increases the ratio of the variance of the frequency ½ component to the variance of the frequency ¼ component by a factor[6] two. Often the variation in quarterly growth rates is more interesting from a political point of view than the variation in the level, so that this transformation gives a direct look at the variable of interest. One subtracts seasonal means to eliminate the deterministic component of seasonality.

For univariate models like the one in example 4.1, it follows that $w_t=(1-L)x_t$ is integrated at the seasonal frequencies. One can make the seasonal integration come out by changing the frequency of observation. Due to aliasing an $I_{j/S}(1)$ variable behaves as an $I(1)$ variable if one decreases the *sampling rate*, i.e. the number of observations per unit of time, by a factor $j/(S\cdot k)$, $k=1,2,\ldots$, see (2.14). For quarterly series one gets a fourdimensional column vector of observations for each year k $\{v_{ik}=w_{4k+i}$, $k=0,\pm1,\pm2,\ldots,\ i=1,2,3,4\}$. Let v_k be these vectors which contain quarterly subseries.

A time series plot of this vector series (simply connecting w_t and w_{t-4} instead of w_t and w_{t-1}) can give a first indication of remaining nonstationarities. The survey article by Cleveland (1983) and the references therein contain an extensive discussion of the graphical analysis of seasonal series. The series v_k is composed of both integrated and nonintegrated components. The yearly sum of the w_t–values (4.3) is stationary, so that $(1\ 1\ 1\ 1)'v_k$ is $I(0)$. Other linear combinations of the v_{ik}–values are not. When the yearly cycle is stationary (e.g. $\Theta_2=1$ in (4.3.*c*)) it follows that $(1\ 1\ 0\ 0)'v_i$ and $(1\ 0\ 0\ 1)'v_i$ are stationary as well. In that case w_t can be written as $w_t=(1+L)^{-1}\eta_t^*$, with η_t^* $I(0)$ at all frequencies, which entails that the pseudo spectrum of $s(L)w_t$ becomes finite for each moving average filter with $s(-1)=0$.

The $I_{\frac{1}{2}}(1)$ component shows most clearly in a time series plot of $(1\ -1\ 1\ -1)'v_k$. In fact, when the process is stationary at all other frequencies, that vector will be the only nonstationary component of v_k, since all other vectors can be written as a linear combination of the stationary processes $(1\ 1\ 1\ 1)'v_k$, $(1\ 0\ 0\ 1)'v_k$ and $(1\ 1\ 0\ 0)'v_k$. The integrated

[6] Apply (A2.2.19) with $D(z)=(1-z)$ for $z=-1$ and i.

components at other frequencies show most clearly in corresponding time series plots of the regressors $y^*_{\mu_q t}$ in the seasonal unit root test regression (A2.2.12), observed once a year.

Integrated processes w_t have infinite variances. The variance in the vector process v_k which encompasses the variance in w_t is infinite as well, but can be concentrated in one or two or three independent vectors. A straightforward idea to find a basis for the space of these vectors is to construct orthonormal linear combinations, with the first one have the largest variance, the second one having the next largest and the last one having the smallest variance. Implementing this procedure entails estimating the *principal components* of v_k. If the assumption of seasonal homoskedasticity holds approximately[7], the annoying problem of scale in principal components analysis, Kendall (1975, p. 23), is not so serious.

When even the largest principal component seems stationary, one can safely assume that bothering about seasonal unit roots is not necessary. One can even apply this set–up from Stock and Watson (1988) one step further and do a formal test for the number of integrated components. The sixth chapter contains a more detailed discussion of their procedure, which was originally developed for multivariate cointegration analysis at the zero frequency. Franses (1993) applied likelihood ratio unit root tests in the *vector model* (see e.g. Judge et al. (1980, §10.3.1)) for v_k where *all parameters* are seasonally varying. These so–called *periodic* models fall in the class of *state dependent* alternatives, which is outside the scope of this study. They do not have an unobserved components interpretation. Osborn and Smith (1989) applied periodic models on macroeconomic consumption in the U.K., see also Osborn (1991). If one thinks periodic models to be interesting alternatives one can use the procedures of §3.3 to test parameter stability across the different seasons, see also Lütkepohl (1991, §12.3).

The coefficients of the largest principal components can indicate whether integration dominates at frequency ½ (opposite coefficients for odd and even $v_{.i}$–values) or at frequency ¼ if the eigenvalues belonging to the components are sufficiently far apart. Their outcome can confirm the results from the spectral analysis. The largest principal component will show "trend–like" behavior and its time series plot will illustrate the property that "the stochastic part of the summer may become winter". To see whether the

[7] Seasonal heteroskedasticity can be important in some components of (per capita) macroeconomic series, as illustrated by Van Der Hoeven and Hundepool (1986) for Dutch live births and unemployment figures. It may be a consequence of mixed multiplicative additive seasonality, cf. footnote 3.

statistical summer really became winter one has to add the seasonal means before recomputing the principal components. We present an application of the principal component approach in §7.4.7.4.

Outliers in the levels of the series can have a disastrous effect on the analysis. An influence analysis as discussed in §3.2.4 is thus in order. Diagnostic tests for parameter stability, see §3.3.6, can also be put to good use in the analysis of (permanent) changes in the seasonal pattern. Some tests, like the CUSUM test on the OLS residuals, can be adapted quite easily, see §7.4.7.7. Canova and Hansen (1992) derived a number of these tests and showed them to be useful to test deterministic seasonality against unit root seasonality.

After the resolution of the series into its different interpretable components a wide range of graphical checks can be performed. By looking at the series for the separate quarters of the different components one can evaluate the consequences of the smoothness restrictions (Cleveland et al. (1982)) although one has to bear in mind that the optimal estimate of the component of interest is always smoother than the "true" or hypothesized underlying component, unless this component is deterministic, see Bell and Hillmer (1984). The same effect occurs in out–of–sample forecasting where the variation in the optimal forecasts is always smaller than the actual variation unless the model is deterministic. Figure 7.6 provides an example of graphical examination.

4.5 Application of the Census X-11 filter in a VAR[8]

The most widely used seasonal model is the one underlying the Census X–11 method, see Laroque (1977). Is this model consistent with our aim of estimating the relationships between economic variables? To answer this question it is best to look at the history of the method. It has been developed in the sixties mainly to model the seasonality in unemployment rates in the US. As the unemployment rate is an important business cycle indicator (see also Beveridge and Nelson (1981)) in the US it is important to have as accurate estimates of the deseasonalized unemployment figures as possible.

The method has been analyzed extensively and its limitations have been pointed out (see e.g. Wallis (1982)), but it still seems to work nicely for the aggregate unemployment series it originally was designed for, see Hausman and Watson (1985).

For quarterly data and a multiplicative seasonal component "the linear

[8] This section is based on material from Ooms (1988).

part of the method" comes down to the application of a symmetric 56–term moving average filter to the logged data, see Laroque (1977). It is impossible to apply this filter to the first 28 and the last 28 observations of the series. Since the last observations often are of primary interest, one has to correct them differently. The end–corrections of the Census method do not seem to be optimal, both in the sense that they do not estimate the seasonal component very efficiently (Hillmer and Tiao (1982)) and that they induce an undesirable phase shift, which is especially important when lead/lag relationships of variables are estimated. The first observations of the sample can contain important information about the relationships we examine, so it would be a pity to have to discard most of this information, simply to start up a proper estimation of the seasonal component.

Ideally we could capture the seasonality in our VAR model, but then we would need very long AR polynomials. Since we fear that lack of efficiency may tend to make us overlook important features in the data, we prefer to keep the ratio of the number of observations to the number of estimated parameters at a reasonably high level.

Since the shorter AR polynomials are not able to describe the seasonality of the type (4.3*c*) adequately, we have to extend our model in another way. It is important to have a model that allows for changing seasonality if it is to be consistent with the data, see Granger (1981). As Sims (1974) pointed out this becomes the more important the longer the period involved in the sample.

We are mainly interested in the estimation of distributed lag relations between variables. It seems therefore preferable to apply the same linear filter adjustment procedure to all series, because this does the least damage to these relationships. This view was not challenged by the advocates of "model based seasonal adjustment", see Bell and Hillmer (1984). The question is then which filter should be chosen. The Census filter is an obvious choice for data with seasonal integration. To see what the Census filter $s(L)$ does to the data one can take a look at its *squared gain* $|s(e^{i\omega})|^2$, see figure 4.1.

The graph of the function shows that the Census filter removes all the power of the series around the seasonal frequencies ¼ and ½. At the other frequencies it leaves the series essentially intact.

The question is whether the dips in the function are wide enough to capture the changing seasonality or whether they are too wide and that valuable information is lost unnecessarily. The shorter the sample period and the more homogeneous the aggregate variable the more constant the seasonality is and the narrower the dip in the frequency response function of the seasonal filter should be. With an infinite number of data one could tune the (very

long) filter to obtain the "optimal" frequency response, see e.g. Koopmans (1974, Chapter 6). With a finite number of data one has to compromise between *filter leakage* (the phenomenon that the power of certain frequencies is increased or diminished outside the region where the filter is meant to work, see the lobes, or *ripples*, between periods 4 and 2) and loss of data at beginning and end of the sample period. The filter has to be symmetric to prevent a phase shift, see (A2.2.19).

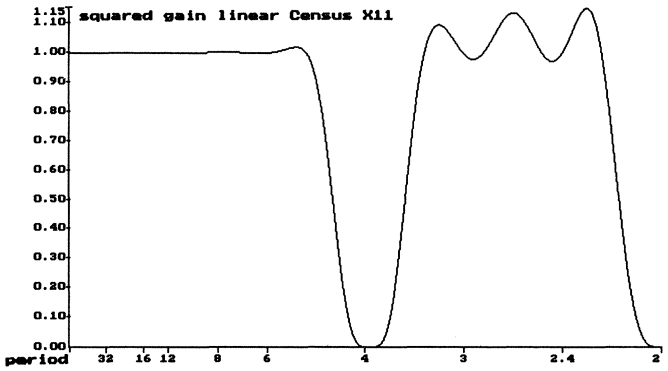

Figure 4.1 *The squared gain of the linear part of the Census X–11 filter*

The filter of the Census X–11 method (Shiskin et al. (1967)) seems to be a rather good compromise because it has survived for over 25 years, which cannot be said of many models for economic time series. It has also proven successful in the modeling of seasonal monetary variables and "performs quite well for an amazing number of series," Maravall (1984). Using forerunners of the Census X–11 method central banks were able to extract the (gradually changing) seasonal component of money demand. They adjusted money supply correspondingly, so that seasonal variation in interest rates largely disappeared (op cit.).

It is clear that variation of seasonality can differ widely across time series. Hence is it not surprising that the seasonal component is estimated more efficiently if one takes this into account. An easily implemented method with a clear interpretation in multivariate series does not exist to our knowledge.

We avoid the heavily criticized end–correction of the Census method by artificially extending our sample period with the best linear minimum mean square error estimates given model (2.1)–(2.3), i.e. we simply forecast as many future and initial values (see A4.2) of our endogenous variables as we

need to be able to apply the filter for the first and the last observation of our original sample, see §7.5.1. If one does not think the estimated seasonal component to be sufficiently smooth, one can increase its smoothness by increasing the length of the component filters that estimate the time varying seasonal component, as in Cleveland et al. (1982). One will then need more backforecasts and forecasts to adjust the series near the ends of the sample. In order to make our results easily interpretable, we choose the "standard options" for the linear part of the filter, as described in Laroque (1977).

Den Butter and Mourik (1990) applied the Census X–11 filter and Harvey's "Structural Time Series" method (see Harvey (1989) and §4.3.3) on a considerable number of Dutch macroeconomic series. The methods produced similar estimates of the seasonal components. It was not clear to them that the (sometimes computationally burdensome) results of Harvey's parametric unobserved components approach were much better where they differed[9]. The fact that most differences are small can be due in part to the seasonal unit root nonstationarity of the series. We present results on seasonal unit root nonstationarity in Dutch series in §7.4.7.6.

What should be the effect of prefiltering all our series on the estimate of the typical spectral shape $f(\omega,n)$? Let $s(L)$ denote the scalar Census filter and let x_t denote the (vector) observed data at time t. Since we do not want to make assumptions about the zero frequency components, we do not exclude the possibility that x_t is stationary around a trend. Appendix A4.2 contains a more detailed analysis of this model. Equations (4.5) and (4.6) illustrate the effect algebraically. Let $s(L)x_t$ be the approximation of the data variation of interest.

$$s(L)x_t = \frac{\mathrm{Adj}(\Phi(L))}{\det(\Phi(L))}(\delta_0+\delta_1 t+\varepsilon_t) \tag{4.5}$$

$$x_t = \frac{\mathrm{Adj}(\Phi(L))}{s^n(L)\det(\Phi(L))}(\delta_0+\delta_1 t+s^{n-1}(L))\varepsilon_t \tag{4.6}$$

One can see that (4.6) follows from (4.5), by noting that the cofactors of the adjoint matrix are multiplied by $s^{n-1}(L)$ and by using the well known

---

[9] If we do not consider the corrections towards the end of the series, the official Census filter is seen to perform poorly with respect to their criterion of "idempotency". This "negative" aspect applies to the use of *every* fixed filter, unless the series is entirely deterministic. If we apply a nontrivial filter twice, it will have a different effect on the series than if it is applied only once, cf. §3.3.10. This is obviously not true for adaptive procedures like regression and methods where a different model will be estimated if the procedure is applied for the second time on the same series.

facts that $s(1)=1$ and $s'(1)=0$, so that $s(L)(\delta_0+\delta_1 t)=\delta_0+\delta_1 t$. This can be seen at the zero frequency side of figure 4.1: $s^2(e^0)=1$, $2s(1)s'(1)=0$.

Let $f_{xx}(\omega,n)$ denote the typical log spectrum of the uncorrected data, cf. (3.15) above. Suppose that the pseudo spectrum of y_t in (2.1) really were equal to the pseudo spectrum of $s^{\frac{1}{2}}(L)x_t$, so that the Census filter were optimal, then one would get

$$f_{xx}(\omega,n)=f(\omega,n)-n\cdot 2\log\left|s(e^{i\omega})\right|. \tag{4.7}$$

In figure 4.1 it can also be seen that $\left|s(e^{i\omega})\right|^2=0$ for $\omega=\frac{1}{2}\pi$ and $\omega=\pi$. This implies that $f_{xx}(\omega,n)$ should have poles for those values of ω. The a priori assumption of the maximum number of seasonal unit roots for the VAR system of the raw data is per variable: one at i, one at[10] $-i$ and one at -1. Laroque (1977) presented the factorization of the Census X–11 filter, which is now easily calculated by numbers of standard computer packages. For n variables there can be at most n times as many roots.

In order to check whether this assumption is reasonable one can estimate a VAR system on the raw data using OLS and plot $f_{xx}(\omega,n)$. As we discussed in the previous section the spectrum of the optimally adjusted series should be zero at the seasonal frequencies when the seasonal noise is $I_{j/S}(1)$ at those frequencies. The same should hold if one applies the Census adjustment filter to all series integrated of orders lower than 2. This spectral dip thus becomes a common factor in the spectral density of $\hat{y}_t$ and should show up in $f(\omega,n)$ estimated on the adjusted data, provided the lag length of the VAR is sufficient. The application of the Census filter trades dangerous AR seasonal unit roots for more innocent MA seasonal unit roots, where danger is assessed in connection with the estimation of dynamic relationships at the frequencies of interest. See figure 7.7 as an example of the frequency domain effect of the seasonal correction.

An important part of the Census method is the correction for outliers. This makes the filter nonlinear in practice, although it is still linear in spirit. It is one of the aspects of the method that makes the statistical analysis of officially adjusted series difficult, Bell and Hillmer (1984, p. 344). Application of the linear part applied to one's own extended series produces approximately the same estimates in the middle of the sample and better ones toward the ends, provided the outliers were not very influential.

---

[10] Actually the linear part of the Census filter $s(z)$ has one root at i and one at (-0.03914+0.99923i), which we take to be equal here for practical purposes. The slight asymmetry shows in figure 4.1. Note again that roots in the AR part of the underlying process occur twice in the optimal filter adjustment filter.

A4.1 Trigonometric seasonal processes in regression

In this appendix we illustrate the remarks in Sims (1974) on the relationship between the speed of change of a seasonal pattern and the number of trigonometric regressors needed to estimate a dynamic regression relating two variables. This brings out the possibilities and limitations of regression procedures and corresponding frequency domain methods for seasonal adjustment.

A4.1.1 Notation and underlying model

$\tilde{a}$ is the Fourier transform of $a(s)$, $s = -\infty, \ldots, \infty$.

$x * a(t) = \sum_{s=-\infty}^{\infty} a(s)x(t-s)$.

One wants to filter two observed variables x_1 and x_2 by the same filter a. The variables x_1 and x_2 consist of the underlying variables of interest y_1 and y_2 contaminated by seasonal noises u_1 and u_2, which are orthogonal to y_1 and y_2 respectively. The idea is that the estimation of a transfer function from y_1 to y_2 using x_1 and x_2 would be dominated by the seasonal noises, whereas these seasonal noises have a time series relationship that is different from the time series relationship of interest between y_1 and y_2. A derivation of this effect for stationary series and a one–sided transfer function, i.e. with no feedback from y_2 to y_1, is given in Sims (1974)[1]. Consequently $|\tilde{a}|$ is chosen to have sharp dips at the seasonal frequencies.

A4.1.2 Zero correlation between seasonal patterns

Sims stated: "Ideally one would choose a so that $\tilde{a}$ is 0 in some band width δ about the seasonal frequencies and 1 elsewhere. The number δ would be chosen larger than the width of the largest seasonal peak in the spectral density of x'. With δ chosen equal to π/k, the implication is that correlation between annual seasonal patterns may become small within as little as $2k/S$ years, where S is the number of observations per year." Sims measured frequencies in radians per sample period. Consider the following

Lemma A4.1.1

$$\int_0^{1/\lambda} \cos(2\pi\lambda t)\cos(2\pi\lambda(t+\Delta t))dt = 0 \text{ for } \Delta t = \tfrac{1}{4} \bmod \tfrac{1}{2}. \qquad \text{(A4.1.1)}$$

---

[1] Sims (1993) discussed rational expectations models.

Proof.

$$\cos(2\pi\lambda(t+¼ \bmod 1)) = \sin(2\pi\lambda t), \tag{A4.1.2}$$

$$\cos(2\pi\lambda(t-¼ \bmod 1)) = -\sin(2\pi\lambda t). \tag{A4.1.3}$$

substitution in the integrand and obvious rearrangement gives:

$$\int_0^{1/\lambda} \cos(2\pi\lambda t)\sin(2\pi\lambda t)\mathrm{d}t = ½ \int_0^{1/\lambda} \sin(4\pi\lambda t)\mathrm{d}t = -(1/8\lambda)\cos(4\pi\lambda t)\Big]_0^{1/\lambda} = 0 =$$

$$= \int_0^{1/\lambda} \cos(2\pi\lambda t)(-\sin(2\pi\lambda t))\mathrm{d}t. \tag{A4.1.4}$$

Consider now functions $g(\lambda_i,t) = \cos(2\pi\lambda_i t)$ with frequencies differing $\Delta\lambda$, $\Delta\lambda << \lambda$. What will be the smallest interval $\Delta t = l{\cdot}S$ so that the correlation between $g_1(\lambda_1,t) = \cos(\lambda_1,t)$ and $g_2(\lambda_2,t) = (\cos(\lambda_1+\Delta\lambda, t+l{\cdot}S)$ is approximately zero? It follows from the lemma that this correlation is approximately zero if the phase difference between g_1 and g_2 is ¼ mod ½. The smallest Δt for which the phase difference (measured in fractions of a cycle of a periodic function of a seasonal frequency λ) equals ¼ is given by

$$\Delta t\Delta\lambda = 1/4. \tag{A4.1.5}$$

The smallest number of years l corresponds to

$$l = 1/(4\Delta\lambda S). \tag{A4.1.6}$$

When a removes a bandwidth δ of π/k centered around a seasonal frequency in radians $= 1/(4k)$, the maximum deviation from this frequency within the band is $\pi/2k = 2\pi/4k$ radians $= 1/(4k)$. This corresponds to a choice l in (A4.1.6) of $¼4k/S = k/S$.

Example A4.1

Suppose we have quarterly data for 32 years, which amounts to 128 observations in the time domain, so that the half open interval $(0,\pi]$ corresponds to 64 periodogram points in the frequency domain which differ $2\pi/128$ in frequency. We choose δ as $\pi/16$. This width in the spectrum corresponds to at most 5 equidistant data points: $[(\pi/16)/(2\pi/128)]+1$, with frequencies $n/S \pm i/128$, i=0,1,2. The corresponding values for the other parameters in (A4.1.5) and (A4.1.6) are then equal to:

$$S = 4,\ \Delta\lambda = 1/64,\ l = ¼{\cdot}64/4 = 4,\ \delta = 2{\cdot}2\pi/64 = \pi/16.$$

The correlation between seasonal patterns at yearly lags can become approximately zero within as little as k/S years ($=4$ in the example). This provides a lower bound which is tighter than Sims's. When we use the "ideal" deterministic seasonal correction its application corresponds to the inclusion of $3{\cdot}5 = 15$ (5 at each frequency $j/4$, j=1,2,3) trigonometric regressors in the (autoregressive distributed lag) regression of x_2 on x_1. This seems very

flexible. There is however always an annoying finite sample restriction on this seasonal model, which we discuss next.

A4.1.3 Circularity: Unit autocorrelation between seasonal patterns

It is clear that the correlation between $\cos(2\pi\lambda t)$ and $\cos(2\pi\lambda(t+1/\lambda))$ is one so that seasonal factors l years apart are nearly perfectly correlated for $l = j/\lambda S$, $j = 1,2,....$ Usually the seasonal trigonometric functions are chosen to have frequencies $\lambda = k/T$, where T is the number of observations and the k's are integers in the neighborhood of multiples of $[T/S]$. One certainly wants to include the functions with frequencies j/S, $j = 1,2,\ldots,S-1$, which produce the well known seasonal dummies. They are automatically included in this set-up when T is a multiple of[2] S.

Suppose we choose the smallest possible $\Delta\lambda = 1/T$. The frequencies λ of the seasonals are then $i/S + k/T$, $i = 1,2,\ldots,[S/2]$, $k = -1,0,1$. Using the equation above we find the correlation between seasonal factors will be approximately one for periods $l = j/((i/S + k/T)S)$ years apart. Choosing $j = i[T/S] + k$ which is integer by assumption we find $l = [T/S]$ years, which is just the sample period. So the correlation between all the seasonal factors at the beginning and end of the sample period will be close to one, whenever the total number of observations is a multiple of the number of observations per year. The same is true when we choose $\Delta\lambda = j/T$, $j = 2,3,....$

Although this implementation of the trigonometric seasonal model is more flexible than the seasonal dummy model it is still subject to the important and often unrealistic finite sample restriction of *circularity*: $w_t = w_{t+T}$ with T the sample period.

Reducing the effect of the circularity restriction

By choosing other frequencies for the trigonometric functions one can change the circularity to periods longer than the sample period, cf. Anderson (1971, §4.4). A sensible strategy seems hard to get. In time series analysis where frequency domain estimation is widely established one has developed a range of *tapering filters* of the data, which reduce the effect of the circularity restriction, and improve finite sample characteristics of the estimators, see e.g. Dahlhaus (1988). For models with slowly changing seasonals one can consider products of polynomials and trigonometric functions of time as in (A2.2.24). Hendry (1974) applied this technique. Reasonably defined seasonal adjustment filters will extract such seasonals as well.

---

[2] If one uses the standard periodogram estimates, one then first has to extend one's data set with zeros to an integer multiple of the number of observations per year (so-called *zero-padding*).

A4.2 Backforecasts and deterministic changes in mean

A4.2.1 Introduction

In this appendix we derive time dependent means and out–of–sample "backforecasting" formulae for univariate AR models with deterministic nonstationarities that are modeled by deterministic trends and/or seasonal dummies. These terms can be a useful contribution to the AR model when it is applied to relatively short macroeconomic time series.

For some applications in estimation, backforecasts of AR models are required. We mention the generation of extra observations for the application of linear symmetric filters near the beginning and the end of the sample as in the Statistics Canada Census X–11 method for seasonal correction (see Kuiper (1978)) and the generation of initial values for iterative least squares estimation of ARMA models as in Spliid (1983). See Nerlove et al. (1979, p. 122) for other applications in the signal extraction literature, and for an interpretation of the backward model.

One can also use the backforecasts to test parameter stability in the model. Forecasts play an important role in model validation, see Chong and Hendry (1986). Usually the prediction tests are only conducted in the straightforward way. However, it is statistically equally meaningful to test the model on the first observations given the parameter estimates based on the later ones as it is to test it on the last observations given the estimates using the first part of the sample. This is especially important if the first observations are influential. When the focus of the analysis is on testing of hypotheses concerning the parameters of the AR model, rather than on optimal out–of–sample forecasting one would not like to miss the information the outcome of the test can give. A simple look at the pattern of the backforecasts can already give extra insight in the plausibility of the used AR model.

For a univariate stationary AR model the backforecasts can be obtained by simply replacing the lag operator for the inverse lag operator in the forecasting formulas, see Box and Jenkins (1970). For the often applied AR model with seasonal dummies and deterministic trends this is not the case.

A4.2.2 Backforecasting and deterministic changes in mean with linear trends

First we consider the following VAR model which is stationary around a deterministic trend:

$$\Phi(L)x_t = m_0 + m_1 t + \varepsilon_t, \qquad t = 1,2,\ldots,T, \qquad \text{(A4.2.1)}$$

where

x_t is an n–vector of stochastic variables:$(x_{1t}, x_{2t}, \ldots, x_{nt})'$, which are partly deterministic,

ε_t is an n–vector of normally distributed serially uncorrelated white noise errors,

$\Phi(L) = \Phi_0 + \Phi_1 L + \ldots + \Phi_p L^p$ is an $(n \times n)$–matrix of lag polynomials of order p in the lag operator L, $\Phi(0) = I_n$, the identity matrix of order n, and

m_0 and m_1 are n–vectors of the parameters of the deterministic variables.

We assume first that $\Phi(z)$ is of full rank, $z = e^{i\omega} \in \mathbb{C}$, with all roots of $|\Phi(z)| = 0$ outside the unit circle. In this case x_t is only *nonstationary in mean* if $m_1 \neq 0$, which follows from

Lemma A4.1

x_t can be transformed to the zero mean process y_t:

$$y_t = x_t - Ex_t,$$

$$Ex_t = \Phi(1)^{-1} m_1 t + \Phi(1)^{-1} m_0 + \Phi(1)^{-1}\left(\Phi(1) - \Phi(L)\right)\Phi(1)^{-1} m_1 t. \qquad \text{(A4.2.2)}$$

Proof.

Using

$$E\Delta x_t = \Phi(1)^{-1} m_1 \qquad \text{(A4.2.3)}$$

and

$$\left(\Phi(1) - \Phi(L)\right)\Phi(1)^{-1} m_1 t = \sum_{i=1}^{p} i\Phi_i \Phi(1)^{-1} m_1, \qquad \text{(A4.2.4)}$$

which is a constant vector which we denote by m_{01}, one gets

$$E\Delta y_t = \Phi(1)^{-1} m_1 - \Phi(1)^{-1} m_1 = 0, \quad t = 1,2,\ldots, \qquad \text{(A4.2.5)}$$

which does not depend on t. It is easy to check that $\Phi(L) y_t = \varepsilon_t$, by noting that

$$\Phi(L)\Phi(1)^{-1} m_1 t = m_1 t - \left(\Phi(1) - \Phi(L)\right)\Phi(1)^{-1} m_1 t, \qquad \text{(A4.2.6)}$$

so that y_t is a normally distributed zero mean stationary AR process.

We know for $n = 1$ that the minimal mean squared error one step ahead prediction $\hat{y}_{t|t-1}$ for y_t is given by $\hat{y}_{t|t-1} = (I_n - \Phi(L))y_t$. Following Box and Jenkins (1970) we define the optimal, i.e. minimal mean squared error, one step backward predictor for univariate AR models as follows:

$$\hat{y}_t = \hat{y}_{t|t+1} = (I_n - \Phi(F))y_t, \tag{A4.2.7}$$

with $F = L^{-1}$, the forward shift operator: $F^k y_t = y_{t+k}$ $\quad k = 0,1,2,\ldots$.

Using the law of iterated projections we can easily extend this result to multi step backforecasts by using prediction formula (A4.2.7) for smaller t and substituting the predicted values for the necessary values of y_{t+i}. If one substitutes (A4.2.2) one can write the result for the optimal univariate backforecast $\hat{x}_t = Ex_t + (I - \Phi(F))(x_t - Ex_t)$ for x_t as follows:

$$\begin{aligned}
\hat{x}_t = {} & \Phi(1)^{-1}m_1 t + \Phi(1)^{-1}m_0 + \Phi(1)^{-1}(\Phi(1) - \Phi(L))\Phi(1)^{-1}m_1 t \\
& + (I - \Phi(F))(x_t - \Phi(1)^{-1}m_1 t - \Phi(1)^{-1}m_0 - \Phi(1)^{-1}(\Phi(1) - \Phi(L))\Phi(1)^{-1}m_1 t) \\
= {} & (I - \Phi(F))x_t + (\Phi(1)^{-1} - \Phi(1)^{-1} + \Phi(F)\Phi(1)^{-1})m_0 \\
& + (\Phi(1)^{-1} - \Phi(1)^{-1} + \Phi(F)\Phi(1)^{-1})(\Phi(1) - \Phi(L))\Phi(1)^{-1}m_1 t \\
& + (\Phi(1)^{-1} - \Phi(1)^{-1} + \Phi(F)\Phi(1)^{-1})m_1 t \\
= {} & (I - \Phi(F))x_t + m_0 + (\Phi(1) - \Phi(L))\Phi(1)^{-1}m_1 t \\
& - (\Phi(1) - \Phi(F))\Phi(1)^{-1}m_1 t + \Phi(1)\Phi(1)^{-1}m_1 t \\
= {} & (I - \Phi(F))x_t + m_0 + m_1 t + 2(\Phi(1) - \Phi(L))\Phi(1)^{-1}m_1 t \\
= {} & (I - \Phi(F))x_t + m_0 + m_1 t + 2m_{01}.
\end{aligned} \tag{A4.2.8}$$

One can see that the equalities hold by recognizing that

$$m_{01} = \sum_{i=1}^{p} i\Phi_i\Phi(1)^{-1}m_1 = -(\Phi(1) - \Phi(F))\Phi(1)^{-1}m_1 t, \tag{A4.2.9}$$

which does not depend on t.

A4.2.3 Backforecasting and deterministic changes in mean with seasonal dummies

Consider the following AR model with quarterly seasonal dummies:

$$\Phi(L)x_t = m_0 + m_{s1}Q_1 + m_{s2}Q_2 + m_{s3}Q_3 + \varepsilon_t \qquad t = 1,2,\ldots,T \tag{A4.2.10}$$

where

x_t is an n–vector of random endogenous variables: $(x_{1t}, x_{2t}, \ldots, x_{nt})'$,

ε_t is an n–vector of random white noise errors,

$\Phi(L)$ is an $(n \times n)$–matrix of lag polynomials of order p in the lag operator L, $\Phi(0) = I_n$, the identity matrix of order n,

m_0 and m_{si} are n–vectors of the parameters of the deterministic variables,

Q_i are seasonal dummy variables, that are equal to 1 if $(t \bmod 4) = i$ and equal to zero elsewhere.

We assume $\Phi(z)$ is of full rank, $z = e^{i\omega} \in \mathbb{C}$, with all roots of $|\Phi(z)| = 0$ outside the unit circle. In that case x_t is nonstationary in mean only if $m_{si} \neq 0$ for some i.

We transform the variable x_t to the stationary variable y_t by subtracting the time varying mean. We derive the time varying mean as follows.

For $t=5,9,13,\ldots$ one has: (A4.2.11)

$$E[x_t] = \Phi_1 E[x_{t-1}]+\Phi_2 E[x_{t-2}]+\Phi_3 E[x_{t-3}]+\Phi_4 E[x_{t-4}]+\ldots+\Phi_p E[x_{t-p}]+m_0+m_{s1}, \quad (a)$$

$$E[x_{t-1}] = \Phi_1 E[x_{t-2}]+\Phi_2 E[x_{t-3}]+\Phi_3 E[x_{t-4}]+\Phi_4 E[x_{t-5}]+\ldots+\Phi_p E[x_{t-p-1}]+m_0, \quad (b)$$

$$E[x_{t-2}] = \Phi_1 E[x_{t-3}]+\Phi_2 E[x_{t-4}]+\Phi_3 E[x_{t-5}]+\Phi_4 E[x_{t-6}]+\ldots+\Phi_p E[x_{t-p-2}]+m_0+m_{s3}, \quad (c)$$

$$E[x_{t-3}] = \Phi_1 E[x_{t-4}]+\Phi_2 E[x_{t-5}]+\Phi_3 E[x_{t-6}]+\Phi_4 E[x_{t-7}]+\ldots+\Phi_p E[x_{t-p-3}]+m_0+m_{s2}, \quad (d)$$

$$E[x_{t-4}] = E[x_t]. \quad (e)$$

We define:

$$\Phi_j^* = \sum_{i=0}^{[p/4]} \Phi_{4i+j}, \quad j=0,1,2,3 \quad \text{with } \Phi_0^* = -I_n,\ \Phi_j^* = 0,\ j>3, \text{ and} \quad (f)$$

$$\mu_j = E[x_{t-j}], \quad t=5,9,13,\ldots.\ j=0,1,2,3, \quad E[x_1]=\mu_0. \quad (g)$$

We can rewrite (A4.2.11) as (A4.2.12):

$$\Phi_0^*\mu_0+\Phi_1^*\mu_1+\Phi_2^*\mu_2+\Phi_3^*\mu_3 = -m_0-m_{s1}, \quad (A4.2.12a)$$

$$\Phi_3^*\mu_0+\Phi_0^*\mu_1+\Phi_1^*\mu_2+\Phi_2^*\mu_3 = -m_0, \quad (A4.2.12b)$$

$$\Phi_2^*\mu_0+\Phi_3^*\mu_1+\Phi_0^*\mu_2+\Phi_1^*\mu_3 = -m_0-m_{s3}, \quad (A4.2.12c)$$

$$\Phi_1^*\mu_0+\Phi_2^*\mu_1+\Phi_3^*\mu_2+\Phi_0^*\mu_3 = -m_0-m_{s2}. \quad (A4.2.12d)$$

In matrix notation:

$$\Phi^*\mu = -m: \quad \begin{bmatrix} \Phi_0^* & \Phi_1^* & \Phi_2^* & \Phi_3^* \\ \Phi_3^* & \Phi_0^* & \Phi_1^* & \Phi_2^* \\ \Phi_2^* & \Phi_3^* & \Phi_0^* & \Phi_1^* \\ \Phi_1^* & \Phi_2^* & \Phi_3^* & \Phi_0^* \end{bmatrix} \begin{bmatrix} \mu_0 \\ \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} -m_0-m_{s1} \\ -m_0 \\ -m_0-m_{s3} \\ -m_0-m_{s2} \end{bmatrix}. \quad (A4.2.13)$$

Note that

$$|\Phi^*| = |\Phi_0^*+\Phi_1^*+\Phi_2^*+\Phi_3^*|\,|\Phi_0^*-\Phi_1^*+\Phi_2^*-\Phi_3^*|\,(|\Phi_0^*-\Phi_2^*|^2+|\Phi_1^*-\Phi_3^*|^2) \neq 0,$$

by the assumption of full rank for $\Phi(z), z = 1,-1,i,-i$. If we solve for the μ_j's (i.e. compute $\mu=-(\Phi^*)^{-1}m$) we can define $y_t = x_t-\mu_j$ so that $E[y_t]=0$ for all t with $\Phi(L)y_t=\varepsilon_t$. Again we use that the minimal mean squared error prediction formula to backforecast a univariate y_t from values for y_{t+i}, $i>0$:

$$\hat{y}_t = (I_n-\Phi(F))y_t. \quad (A4.2.14)$$

This corresponds to the following backforecasting formula

$$\hat{x}_t = (I_n - \Phi(F))(x_t - E[x_t]) + E[x_t]$$
$$= (I_n - \Phi(F))x_t + (-I_n + \Phi(F) + I_n)E[x_t]$$
$$= (I_n - \Phi(F))x_t + m_0^* + m_{s1}^* Q_1 + m_{s2}^* Q_2 + m_{s3}^* Q_3. \qquad \text{(A4.2.15}a\text{)}$$

with

$$m_{s1}^* = \Phi(F)\mu_0 - \Phi(F)\mu_1, \qquad \text{(A4.2.15}b\text{)}$$

$$m_{s2}^* = \Phi(F)\mu_2 - \Phi(F)\mu_1, \qquad \text{(A4.2.15}c\text{)}$$

$$m_{s3}^* = \Phi(F)\mu_3 - \Phi(F)\mu_1, \qquad \text{(A4.2.15}d\text{)}$$

$$m_0^* = \Phi(F)\mu_1. \qquad \text{(A4.2.15}e\text{)}$$

In matrix notation:

$$m^* = \Phi^{*\prime}\mu : \begin{bmatrix} -m_0^* - m_{s1}^* \\ -m_0^* \\ -m_0^* - m_{s3}^* \\ -m_0^* - m_{s2}^* \end{bmatrix} = \begin{bmatrix} \Phi_0^* & \Phi_3^* & \Phi_2^* & \Phi_1^* \\ \Phi_1^* & \Phi_0^* & \Phi_3^* & \Phi_2^* \\ \Phi_2^* & \Phi_1^* & \Phi_0^* & \Phi_3^* \\ \Phi_3^* & \Phi_2^* & \Phi_1^* & \Phi_0^* \end{bmatrix} \begin{bmatrix} \mu_0 \\ \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}. \qquad \text{(A4.2.15}f\text{)}$$

One can extend these results to seasonality of period s greater than four, if one retains the structure of $\Phi^* = \{\Phi_{ij}^*\}$ where $\Phi_{ij}^* = \Phi_{((j-i) \bmod s)}^*$, $i, j = 1, 2, \ldots, s$.

Next we consider the encompassing AR model:

$$\Phi(L)x_t = m_0 + m_1 t + m_{s1}Q_1 + m_{s2}Q_2 + m_{s3}Q_3 + \varepsilon_t, \quad t = 1, 2, \ldots, T, \qquad \text{(A4.2.16)}$$

where the variables and parameters are defined as under (A4.2.1) and (A4.2.10). In this case we can define the corresponding mean stationary variable x_t^* as follows in (A4.2.17)

$$y_t = x_t - E[x_t], \qquad \text{(A4.2.17}a\text{)}$$

with

$$E[x_{t-j}] = \mu_j + \Phi(1)^{-1}(m_1 t + m_{01}), \quad t = 5, 9, 13, \ldots; j = 0, 1, 2, 3, \qquad \text{(A4.2.17}b\text{)}$$

$$E[x_1] = \mu_0 + \Phi(1)^{-1}m_{01}, \qquad \text{(A4.2.17}c\text{)}$$

where μ_j and m_{01} are defined as in (A4.2.11) and (A4.2.9) respectively. The backforecasting formula for the model in (A4.2.16) then becomes:

$$\hat{x}_t = (I_n - \Phi(F))x_t + m_0^* + m_{s1}^* Q_1 + m_{s2}^* Q_2 + m_{s3}^* Q_3 + m_1 t + 2m_{01}, \qquad \text{(A4.2.18)}$$

where m_0^*, m_{s1}^*, m_{s2}^* and m_{s3}^* are defined in (A4.2.15).

A4.2.4 Changes in mean in multivariate models with unit roots

The results of the preceding sections on the transformation to stationarity by subtraction of the time varying mean go through for vector AR models as well. The analysis shows that the seasonal dummies cannot simply be interpreted as indication of a periodically changing mean when $\Phi(z)$ does not have full rank at the seasonal frequencies, see (A4.2.13). This induces also problems for tests of the statistical significance of seasonal dummies and the fourier transformation of these variables introduced below. The critical values for t–statistics $t_{d\delta 2}$, $t_{d\delta 3}$, $t_{d\delta 4}$, in table 7.5, which one can obtain from a simple Monte Carlo experiment described in §7.4.5 illustrate that the (possible) presence of (seasonal) unit roots can make a large difference. We start the derivation of the changing mean when there are unit roots at the seasonal frequencies with a convenient representation for the deterministic inputs at the right hand side of (A4.2.16), which decomposes the seasonal dummies according to the variance at different frequencies

$$\delta_0+\delta_1 t+\delta_2(-1)^t+\delta_3\cos(\tfrac{1}{2}\pi t)+\delta_4\sin(\tfrac{1}{2}\pi t), \qquad \text{(A4.2.19)}$$

with

$$\text{vec}(\delta') = I_n\otimes\left\{\left[\begin{matrix}1&1&1&0&0\\1&2&0&1&0\\1&3&0&0&1\\1&4&0&0&0\\1&5&1&0&0\end{matrix}\right]^{-1}\left[\begin{matrix}1&1&1&0&1\\1&2&-1&-1&0\\1&3&1&0&-1\\1&4&-1&1&0\\1&5&1&0&1\end{matrix}\right]\right\}\cdot\text{vec}(m'), \qquad \text{(A4.2.20)}$$

with $\delta = \begin{bmatrix}\delta_0 & \delta_1 & \delta_2 & \delta_3 & \delta_4\end{bmatrix}$ and

$m = \begin{bmatrix}m_0 & m_1 & m_{s1} & m_{s2} & m_{s3}\end{bmatrix}$, $n\times 5$ matrices,

and I_n an identity matrix of dimension n.

By working out this basic Fourier transform one finds $\delta_0 = m_0 - m_{s1} - m_{s2}$, $\delta_1 = m_1$, $\delta_2 = 2m_{s1} - m_{s2} - m_{s3}$, $\delta_3 = -2m_{s2}$, $\delta_4 = 2m_{s1} - m_{s2} - m_{s3}$.

In general this system has products of linear trends and periodic terms as solutions of the homogeneous part of (A4.2.16) if $|\Phi(-1)| = 0$ and $\delta_2 \neq 0$ or if $|\text{Re}(\Phi(e^{\frac{1}{2}\pi i})|^2 + |\text{Im}(\Phi(e^{\frac{1}{2}\pi i})|^2 = 0$ (cf. (A2.2.19)) and either $\delta_3 \neq 0$ or $\delta_4 \neq 0$.

The exceptions at the frequency ½ can be found as follows. When $|\Phi(-1)| = 0$ and $\Phi(z)$ has full rank for every other z, there exists a moving average representation as in (2.7) with stationary stochastic part:

$$(1+L)x_t = C_{\frac{1}{2}}(L)\left(\varepsilon_t+\delta_0+\delta_1 t+\delta_2(-1)^t+\ \delta_3\cos(\tfrac{1}{2}\pi t)+\delta_4\sin(\tfrac{1}{2}\pi t)\right), \qquad \text{(A4.2.21)}$$

provided no linear combinations of the components of x_t are $I_{\frac{1}{2}}(2)$. $C_{\frac{1}{2}}(-1)$ would be zero if $|\Phi(-1)| \neq 0$. The deterministic part of x_t will have components

$t(-1)^t$ as in (A2.2.25) unless δ_2 lies in the right null space of $C_{\frac{1}{2}}(-1)$, i.e. the space of right eigenvectors of $C_{\frac{1}{2}}(-1)$ with eigenvalue zero. In general one can derive that $\text{rank}(C_{\frac{1}{2}}(-1)) = n - \text{rank}((\Phi(-1))$. This analysis is completely analogous to the analysis for the zero frequency which is done in more detail in Chapter 6. Using this transformation of the dummies one can safely interpret δ_0 as the important vector determining the mean growth when δ_1 is zero and $\Phi(z) = 0$ has some roots at $z = 1$. Johansen (1991a) advised to "center" the dummy variables Q_s to obtain the same interpretation of the vector of coefficients of the constant.

For frequency ¼ one derives that trending periodic functions will occur unless $(\delta_3 + i\delta_4)$ lies in the (complex) null space of $C_{\frac{1}{4}}(i)$, with $C_{\frac{1}{4}}(L)$ the corresponding matrix in the MA representation of $(1+L^2)x_t$, as presented in Hylleberg et al. (1990), see also Lee (1992). Use $\sin(½\pi t) = L\cos(½\pi t)$ and $L\sin(½\pi t) = -\cos(½\pi t)$ and evaluate

$$\left[\text{Re}[C_{\frac{1}{4}}(i)] + \text{Im}[C_{\frac{1}{4}}(i)]L\right]\left[\delta_3\cos(½\pi t) + \delta_4\sin(½\pi t)\right]. \qquad \text{(A4.2.22)}$$

where the first factor is from Hylleberg et al. (1990, form. (4.2)). This results in

$$\left\{\left[\text{Re}[C_{\frac{1}{4}}(i)\delta_3] - \text{Im}[C_{\frac{1}{4}}(i)\delta_4]\right] + \left[\text{Re}[C_{\frac{1}{4}}(i)\delta_4] + \text{Im}[C_{\frac{1}{4}}(i)\delta_3]\right]L\right\}\cos(½\pi t), \qquad \text{(A4.2.23)}$$

which is not equal to zero unless $C_{\frac{1}{4}}(i)(\delta_3 + i\delta_4) = 0$.

Backforecasting in a multivariate model with unit roots

The backforecasting formulae for multivariate AR models cannot be derived as straightforwardly as those of the univariate ones, basically because the autocorrelation generating function in no longer even, i.e. symmetric round zero. In *purely nonstationary* models, i.e. purely stochastic AR models with all their roots on the unit circle, the backward formulae for the AR parameters can be found by inversion of the matrix of the first order AR parameterization of the appropriately transformed *purely nonstationary series*, see Tsay and Tiao (1990, theorem 5.1). Tsay and Tiao proved the consistency of OLS estimators in the backward model for the AR parameters related with the nonstationary part of the series.

Inversion of the matrix of AR parameters for models that are nonstationary at frequencies zero or ½ is very special: the backward model has the same AR parameters as the forward one. This is not the case for frequency ¼ (roots i and $-i$) as is easily seen by comparing

$$y_t = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} y_{t-1} + \varepsilon_t \qquad \text{(A4.2.24)}$$

and

$$y_t = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} y_{t+1} + \eta_t, \qquad \text{(A4.2.25)}$$

where ε_t is the forward and η_t the "backward" innovation.

In the left hand side model y_{2t} lags behind y_{1t}, whereas y_{1t} "lags" in the backward model. By changing the frequency of observation or sampling rate one gets $y_t = -y_{t-2} + \varepsilon_{2t}$ and $y_t = -y_{t+2} + \eta_{2t}$ and regression equations similar to those used in the "symmetric approach" to test unit roots in Dickey et al. (1984), see §A2.2.2.4.

Bell (1984) gave references with detailed derivations of multivariate backward models. The results of Tsay and Tiao (1990) indicate that the backward least squares regression gives consistent estimates of the coefficients of the backforecasting formulae for purely stochastic AR models.

Backward prediction for purely stochastic models is thus fundamentally different from backforecasting in models with a mixed continuous/discrete spectral distribution. Interestingly the backward formulae for stochastic purely nonstationary models are sometimes equivalent to those for the purely deterministic case, as can be seen by putting the covariance matrix of ε_t and η_t equal to zero in the example above so that $y_t = -y_{t-2} = y_{t+2}$.

From a practical point of view the most easy way to get backforecasts is to generate them immediately from the backward regressions, including the deterministic terms. This method is too simple if there are outliers. Outliers are discussed in the next chapter. We apply backward regression on outlier corrected data in §7.5.1 with satisfactory results.

5 OUTLIERS

5.1 Introduction[1]

Why bother about outliers? One can give several reasons. Outliers can mess up the statistical analysis. A few outlying observations can change the estimates of parameters of interest and their standard errors considerably. Many testing procedures we discussed in Chapters 2 and 3 and other procedures that are discussed below are based on the assumption of a constant multivariate normal distribution for the disturbances. The outcomes of these tests are hard to interpret if the normality assumption does not hold. One can confine oneself to so–called robust tests, see e.g. MacKinnon and White (1985), but their use is often only justified for null hypotheses that still do not capture some interesting outlier types. Moreover they can often only be applied sensibly in large samples, so that their appeal is largely theoretical.

In the previous chapter we stressed that the use of linear filters for the estimation of unobserved seasonal components is most effective when there are no outliers in the sample. When they are present the "graduation of extreme values" becomes an indispensable component of successful seasonal correction.

In order to get more specific information about the presence and the character of the outliers we introduce a special testing procedure in the next sections. We generalize a method for the detection and identification of some well known outlier types in time series as in Abraham and Yatawara (1988) to the multivariate case. In our approach we use a simultaneous LM test procedure. It allows one to test for an outlier of unknown type before performing the identification of the type of outlier. In this simultaneous approach three types of outliers can be considered: additive outliers, innovative outliers and transient level changes in the terminology of Tsay (1988). The advantages of a simultaneous testing approach are spelled out in Bera and Jarque (1982) for other kinds of model specification tests. Especially for outliers where a large number of statistics can be computed it is nice to have simultaneous tests with (approximately) known distribution

[1] This chapter is a revised version of Ooms (1990).

under the null to avoid overcorrection.

It is attractive to have an objective measure to test for the existence, timing, source and identity of outliers. Oftentimes one is likely to have some a priori ideas on this issue, which one would like to see confirmed in the data. The outliers are by their exceptional, influential nature bound to have a considerable and sometimes decisive impact on the model selection where the order of a VAR, or the number of unit roots in the system are concerned. Testing for outliers is thus an essential part of the modeling process. A multivariate approach seems appropriate for economic time series.

Perron (1989) indicated both the theoretical and empirical relevance of occasional outliers in macroeconomic time series for the analysis of the unit root hypothesis. Chen and Tiao (1990) studied the effect of random level–shifts in ARMA models on integration analysis and forecasting. Ledolter (1989) examined the effect of additive outliers on forecasts of ARIMA models. All these authors came up with specific models for the outliers to study their effects. We present some multivariate examples. After identification of the outliers the question is always what to do about them. Tsay (1988) gave some ready to use recipes for the univariate case. We adapt his ideas to application in multivariate series.

There are several other approaches to deal with outlying observations like introducing dummy variables for innovation outliers, explicit ARMA modeling of transient outliers in an intervention analysis (Box and Tiao (1975), Abraham (1977)), robust filtering methods for additive outliers as in Kleiner, Martin and Thomson (1979) or even nonlinear smoothing algorithms if the outliers are a symptom of (nonstationary) sometimes long tailed densities (and thus non Gaussian) for the errors (Velleman (1980)). We suggest the use of a method based on robust filtering, presented in appendix A5.3.

The organization of the chapter is as follows. In the next section we introduce the outlier model. Section 3 contains a discussion on the effects of outliers for the purpose of this study. In section 4 we derive the test statistic. In section 5 the use of the test is illustrated on artificial examples. Section 6 contains a macroeconomic empirical example of the use of the test. In section 7 we illustrate the use of a robust filter and compare its effects with the conventional "John Wayne method" of firing a dummy at every outlier with a t–value bigger than 2.

5.2 The outlier model

Consider the following adaptation of model (2.1)–(2.3) which is assumed to be stationary in first differences of y_t but which may be nonstationary in

levels either because of unit roots in the characteristic determinantal equation or because of a deterministic drift term. The components of y_t are all $I(1)$ or $I(0)$, but can have a nonzero or even polynomially trending mean. Following Tsay (1988) it is assumed that the series starts at a fixed time point with given starting values if some of the roots of determinant$(\Phi(z))=0$ lie on the unit circle.

$$\Phi(L)y_t = \varepsilon_t + \delta_{t,l}O_2 + m_0 + tm_1, \qquad (5.1)$$
$$y_t^* = y_t + \delta_{t,l}O_1 + (1-\rho L)^{-1}\delta_{t,l}O_3, \qquad t = p, p+1, \ldots, T,$$

where

y_t^* is the observed n–vector time series,

y_t is an n–column vector of endogenous variables of interest,

$\Phi(L) = \Phi_0 + \Phi_1 L + \Phi_2 L^2 + \ldots + \Phi_p L^p$ is an $n \times n$–matrix lag polynomial of order p, with $\Phi_0 = I_n$: the identity matrix of order n, with all roots of $\det(\Phi(z))=0$ on or outside the unit circle,

ε_t is an n–vector of i.i.d. normally distributed disturbances: $\varepsilon_t \sim N(0,\Sigma)$, $t = p, p+1, \ldots, T$,

$\delta_{t,l}$ is the Kronecker delta($=1$ if $t=l$, 0 elsewhere),

l is the time where the outlier has its first impact,

m_0 and m_1 are n–vectors of coefficients for constant and trend,

O_1 is an n–vector resulting in an *additive outlier* (AO) in the observed time series,

O_2 is an n–vector of *innovation outliers* (IO) to the vector of endogenous variables,

O_3 is a "level–change" vector which causes a *transient outlier* (TO) in the level of the observed time series,

ρ is a scalar decay parameter for the transient level–change $(0<\rho<1)$.

The AO was called "aberrant observation" and the IO "aberrant innovation" by Abraham and Yatawara (1988). The AO is also known as *observational outlier*. The AO changes one observation and p residuals from $t=l$ on. One can look at O_1 as an "intervention vector" which changes the observed values irrespective of the underlying model. The IO changes one error term and all subsequent observations. One can describe it as a "shock" to the system which is propagated through the underlying VAR model. IOs can thus help to identify the parameters of $\Phi(L)$, provided they are correctly modeled as IOs. The TO changes a series of observations after it occurs. It is a long lasting series of interventions. Tsay (1988) called it "transient level–shift".

Limits of the outlier model

The assumption of transience $(0<\rho<1)$ for the effect of outlier O_3 is essential for the identification of the outlier under the alternative hypothesis where $O_3 \neq 0$. The identification is based on estimates of the residuals which are good estimates of the "true" disturbances only if the dynamic parameters of $\Phi(L)$ are estimated consistently. This is not the case if O_3 has a permanent effect $(\rho=1)$, see the example in §5.3.2. The LM–test which we derive has no robustness of efficiency (§3.3.2) in this respect. For the same reason we do not allow for a permanent innovation outlier in our testing procedure.

Likelihood ratio– and Wald tests preserve power, but require estimation under a wide range of alternatives if l is not known a priori, whereas the LM test only requires estimation under the null of no outliers. In the first stage of data analysis this is not an attractive option. Furthermore, the asymptotic distributions of tests for "permanent outliers" is known to depend on, i.a., the number of unit roots in $\Phi(L)$, see e.g. Perron (1991), whereas limit distributions of test statistics for the number of unit roots are known to depend on the presence of these "permanent outliers", see Perron (1990a). A Monte Carlo experiment in §7.5.3 illustrates these points. These tests do not have robustness of validity and are thus hard to interpret too.

The form of the effect of the transient outlier can be generalized to a form $\psi(L)^{-1}\Omega(L)\delta_{t,t}O_3$ following Box and Tiao (1975), under the assumption that the appropriate regularity conditions hold (no permanent effect). Abraham (1977, 1980) gave an example of such a model in a bivariate context. One could tune the form of the transient outlier effect also to detect single "patches of q outliers", e.g. by putting ψ(L) equal to I_n and $\Omega(L)$ to $1+L+\ldots+L^q$.

Limits of the dynamic model

One can generalize the model in another way without damaging the results that follow. Following the reasoning in Abraham and Yatawara (1988) the model may be generalized with MA errors as long as the roots of the determinantal equation for the matrix MA polynomial lie outside the unit circle and the determinants of the AR– and MA polynomial have no common factors. In that case the model can be rewritten in AR(∞) form and the formulas in §5.4 can be used for the corresponding reparameterization.

5.3 Some effects of outliers on VAR estimates

5.3.1 Permanent outliers and unit root tests

Perron (1989, 1990a) studied tests for the $I(1)$–hypothesis against the alternative that the process is trend–stationary. He allowed for the presence

of a one–time additive or innovative change in the level of the series or in the slope of the trend function under both null and alternative hypothesis. He developed tests to distinguish between the continuous nonstationarity of the unit root against the one–time sudden change nonstationarity of a "crash" or "change in growth". For yearly U.S. economic data of the last century he was able to reject the continuous nonstationarity against the one–time nonstationarity due to "exogenous shocks" of the 1929 crash and the 1973 oil price shock, see §7.4.5 for an application of this procedure.

One can extend the notion of cointegration as in Johansen (1988) where a stationary univariate process is looked upon as just a special "cointegrating vector" process of dimension one. In that light the unit root test is just a special (co)integration test. It is likely that phenomena described in Perron (1989, 1990a) are important for cointegration analysis in the multivariate case.

In the application we consider the effect on the likelihood ratio test for the number of cointegrating vectors. For these vectors, especially when logged real macroeconomic variables are concerned, significant one–time nonstationarities should not occur if the theory predicts a stable equilibrium. Short run equilibria are expected to shift, but long run relations are not supposed to. One would therefore expect (nearly) permanent level–shifts in the system to occur within a relatively short period in all the variables belonging to the cointegrating relation.

For this purpose one can extend the notion of cointegration to the one–period nonstationarities of level–shifts and the like, so that two variables are said to be cointegrated also if a linear combination of the variables does not show permanent level–shifts whereas the variables that make up the vector do individually. Although the mean of the cointegrating vector may change temporarily, due to large exogenous, maybe persistent shocks in the variables it is made of, its permanent component in the sense of Beveridge and Nelson (1981) does not change.

There are at least two ways to imagine how this could take place. First, if the stochastic component of the system also were nonstationary, due to unit roots in its AR part one could imagine that the level–shift occurred in one variable first, before being transmitted to the other variables of the system. The shift would then be identified as an innovation outlier. In §7.6 we find evidence of a similar phenomenon in Dutch data, where growth rates change permanently due to a permanent innovation outlier, while the mean of the equilibrium relation remains stable.

If the stochastic part of the system were stationary the only way to get

a permanent level–shift is via the transient outlier O_3 with ρ very close to one. Although the testing procedure is not especially designed to detect these shifts, it can be expected to have some power against these alternatives.

5.3.2 Effect of outliers on estimates of Φ

As Perron (1989) showed, one does not estimate the autoregressive parameters appropriately by least squares if one neglects important level–shifts. This is easy to imagine if one goes back to the univariate stationary AR(1) model and pictures a scatter diagram of y_t against y_{t-1} under a level–shift and the regression line one would estimate with the under normal circumstances consistent OLS procedure. An example is shown in figure 5.1. For the univariate AR(1) case with $\rho=1$ and O_3 proportional to the level of the series the probability limit of the OLS estimator of Φ_1 ($n=1, p=1$, under $O_1=O_2=O_3=0$) is equal to one even if $|\Phi_1|$ is very small, which indicates that a unit root test based on the autoregressive parameters, like the Dickey–Fuller t–test, detects indications of nonstationarity. If one is not alert on the presence of outliers one might interpret this sign in the wrong way. One can examine subsample (non)stationarity by allowing for a level–shift in the test regression, both under the null of a zero frequency unit root in $\Phi(L)$ and under the alternative of subsample stationarity, see Perron and Vogelsang (1992a, 1992b), who derived the appropriate limit distributions.

Under stochastic nonstationarity all shocks are thought to have a permanent effect, whereas under deterministic nonstationarity the permanent changes are due to only a few specific shocks. Other means of checking for nonstationarity are also decisively affected by level–shifts. Chen and Tiao (1990) illustrated the influence of a level–shift on the sample autocorrelation function of an ARMA process.

The ignorance about innovation or additive outliers can also have noticeable effects on the identification and estimation of the VAR model, see the example on air line passengers in Tsay (1988). All types of outliers affect tests for residual autocorrelation. Martin and Zeh (1977) suggested to use a scatter plot of the (spurious) residual first order autoregression to discriminate between AOs and IOs. They analyzed a simple univariate AR model.

The bias in the estimation of the parameters in the presence of outliers is a serious threat to our procedure, since the estimates matter in the identification of the outliers.

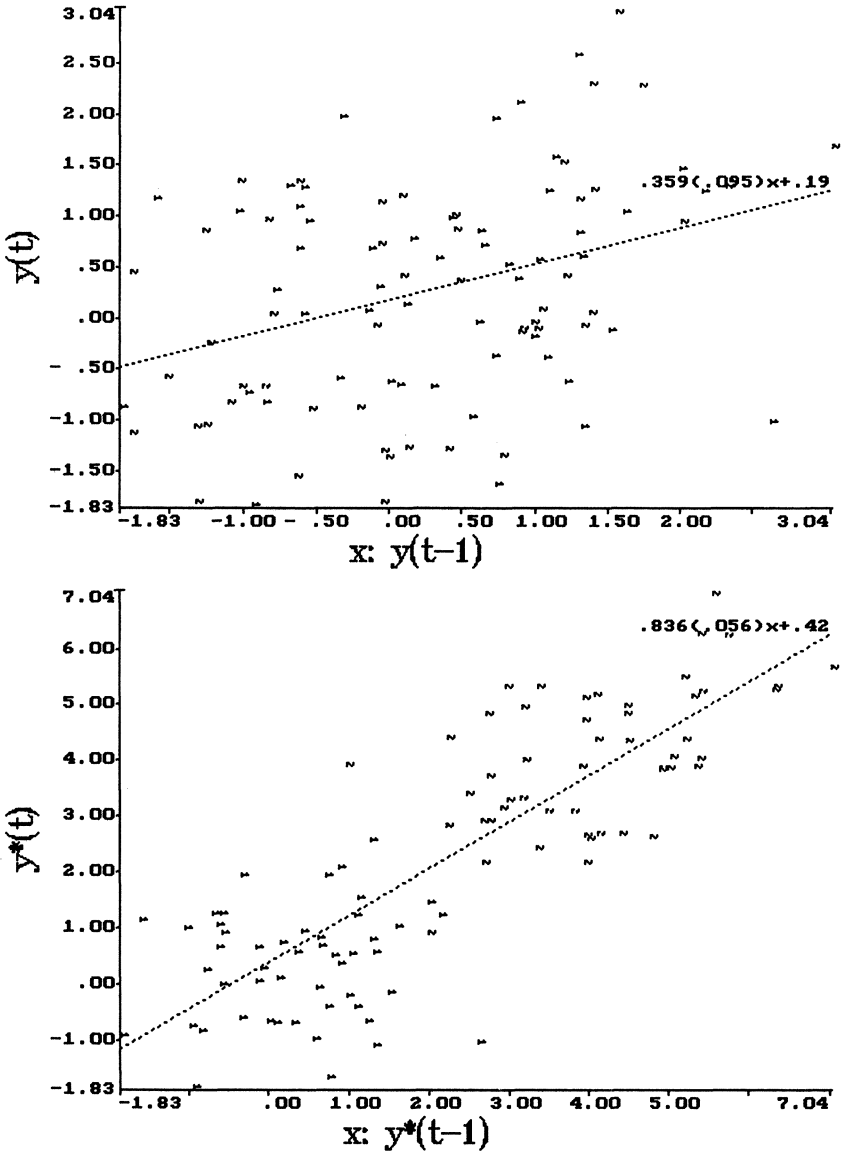

Figure 5.1 *Autoregression under a level-shift*
Both scatters are an artificial sample of 100 points from an AR(1) model with ϕ_1=0.3 and σ^2=1. The first 50 observations are denoted by ones and the rest by twos. The upper scatter is generated under the null of no level-shift. The second scatter illustrates the effect of a level-shift of 4σ after 50 observations on the slope of the (dotted) regression line of y_t^* on y_{t-1}^* and an intercept. A Dickey-Fuller test would accept (unit root) nonstationarity because of the one time change in mean.

Robust estimation methods of one of the types mentioned in the introduction could be used to resolve the problems of bias and spurious serial correlation of the disturbances to some extent. The distribution of a test statistic under the null would then be harder to derive. Furthermore our test could still be useful to indicate the character of the problematic observations and therefore help to design robust estimation methods. Multivariate influence analysis as in Polasek (1984) could help too.

5.4 Derivation of the LM-statistics

5.4.1 Case of known parameters and timing

First we assume that the Φ_i–values and Σ are known a priori. One suspects an outlier at time $t=l$, $p<l<T-p$. One easily violates the regularity conditions for a successful application if one uses the simultaneous test for $l>T-p$, see e.g. condition 5.3 below. The conditional log likelihood (i.e. conditional on the first p observations) given the observations $y_1^*, y_2^*, \ldots, y_T^*$ can be written as follows:

$$\ell(O \mid y_1,\ldots,y_T,\Sigma,\Phi) = c + (T/2)ln\left|\det(\Sigma^{-1})\right| - \tfrac{1}{2}tr\left[\Sigma^{-1}(E(O)'E(O))\right], \tag{5.2}$$

with

$$y_t = y_t^* - \delta_{t,l}O_1 - \delta_{t,l}(1-\rho L)^{-1}O_3,$$

c a constant,

$O=[O_1 : O_2 : O_3]$ an $n\times 3$–matrix of extra parameters whose nullity is tested,

$$E(O) = [Y(O) - X(O)B(O)] = [\varepsilon_{ti}] = [y_{ti} - (x_t'B)_i], i=1,\ldots,n; t=p+1,p+2,\ldots,T,$$

$$Y = \underset{(T-p)\times n}{\begin{bmatrix} y_{p+1}' \\ \vdots \\ y_T' \end{bmatrix}}, \qquad X = \underset{(T-p)\times(np+3)}{\begin{bmatrix} y_p' & y_{p-1}' & \cdots & y_1' & 1 & p & \delta_{p,l} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ y_{T-1}' & y_{T-2}' & \cdots & y_{T-p}' & 1 & T & \delta_{T,l} \end{bmatrix}} = \begin{bmatrix} x_{p+1}' \\ \vdots \\ x_T' \end{bmatrix},$$

$$B = \underset{(np+3)\times n}{\left[-\Phi_1 \; -\Phi_2 \; \ldots \; -\Phi_p \; m_0 \; m_1 \; O_2\right]'},$$

so that O_2 is written as the vector of regression coefficients of a one time dummy variable.

Under the null hypothesis H_{0AIT} that O_1, O_2 and O_3 all are equal to zero we can formulate the Lagrange Multiplier Test Statistic:

$$\left[\frac{\partial \ell}{\partial(\text{vec } O)'}\right] \cdot \left[E\left[\left[\frac{\partial \ell}{\partial(\text{vec } O)'}\right]'\left[\frac{\partial \ell}{\partial(\text{vec } O)'}\right]\right]\right]^{-1} \cdot \left[\frac{\partial \ell}{\partial(\text{vec } O)'}\right]' \Bigg|_{\text{vec } O=0}, \tag{5.3}$$

which is asymptotically distributed as a Chi square with $3n$ degrees of freedom under the appropriate regularity conditions (see below), with E the expectations operator.

We derive an expression for the score using matrix derivative properties in Magnus and Neudecker (1988, pp. 174–178, Ch.9), who use "good notation", where one *row* of a *Jacobian matrix* or *matrix of derivatives*, $DE(O)$, gives the derivatives of one element of the (vectorized (matrix)) "numerator" function E with respect to all the variables of the (vectorized (matrix)) "denominator" matrix of variables O: $DE(O) = \partial \text{vec}E(O)/\partial \text{vec}O'$.

The *score* is then defined as the *transposed derivative*:

$$\left[\frac{\partial \ell}{\partial(\text{vec } O)'}\right]' = \left[\frac{\partial -\tfrac{1}{2}\text{tr}\Sigma^{-1}E'E}{\partial(\text{ vec } O)'}\right]'. \tag{5.4}$$

In their notation for the differentials under consideration we get $\text{d tr } \Sigma^{-1}E'E = 2\text{tr } \Sigma^{-1}E'\text{d}E = 2(\text{vec}E\Sigma^{-1})'\text{d vec}E$ so that

$$\left[\frac{\partial \ell}{\partial(\text{vec } O)'}\right]' = \underset{1\times n(T-p)}{\left[-2\cdot\tfrac{1}{2}\cdot(\text{vec } E\Sigma^{-1})'\right]}\underset{n(T-p)\times 3n}{\left[\frac{\partial(\text{vec } E(O))}{\partial(\text{vec } O)'}\right]}'.$$

If we define Q_l as:

$$\left[\left[\frac{\partial \varepsilon_{tk}}{\partial(\text{vec } O)'}\right]'\right]_l = \left[q_{tk}\right]_l, \quad k=1,\ldots,n,\ t=p+1,p+2,\ldots,T,$$

so that Q_l is a $(T-p)\times n$–matrix of $3n$–vectors of transposed derivatives, with

$$q_{tk} = Ind_{[l,T]}(t)\begin{bmatrix} \Phi_{(t-l)k.}' \\ -\delta_{l,t}\ (I_n)_{k.}' \\ \Gamma_{(t-l)k.}' \end{bmatrix},$$

where

$Ind_{[l,T]}(t)$ is the indicator function equal to one for $l \le t \le T$, zero elsewhere,

$M_{.l}$ is the l^{th} column of elements of a matrix M,

$M_{k.}$ is the k^{th} row of elements of a matrix M and

Γ_r is the coefficient matrix of L^r in the matrix lag polynomial $\Gamma(L) = \Gamma_0 + \Gamma_1 L + \Gamma_2 L^2 + \ldots$, with $\Gamma(z) = (1-\rho z)^{-1}\Phi(z)$, so that

$$\left[\frac{\partial \ell}{\partial(\text{vec } O)'}\right]' = \sum_{k=1}^{n}\sum_{j=1}^{n}\sum_{t=l}^{T}\sigma^{kj}\varepsilon_{tj}q_{tk} = \sum_{t=l}^{T}\sum_{k=1}^{n}q_{tk}\Big(\sum_{j=1}^{n}\sigma^{kj}\varepsilon_{tj}\Big)$$

$$= \sum_{r=0}^{T-l}\begin{bmatrix} \Phi_r' \\ -\delta_{0,r}I_n \\ \Gamma_r' \end{bmatrix}\Sigma^{-1}\varepsilon_{l+r} = \begin{bmatrix} \Phi'(F) \\ -I \\ \Gamma'(F) \end{bmatrix}\Sigma^{-1}\varepsilon_l, \tag{5.5}$$

where $\Sigma^{-1} = \left[\sigma^{kj} \right]$, $k,j = 1,\ldots,n$,

$\varepsilon_l = (\varepsilon_{l.})'$, i.e. the transposed of the l–th row of E,

F is the inverse lag operator so that $F^i y_t = y_{t+i}$, $i = 0,1,2,\ldots$.

Defining $W = \sum_{r=1}^{p} \Phi_r' \Sigma^{-1} \Phi_r$, $C_l = \sum_{r=1}^{T-l} \Gamma_r' \Sigma^{-1} \Gamma_r$ which both are assumed to be nonsingular, defining $D = \sum_{r=1}^{p} \Gamma_r' \Sigma^{-1} \Phi_r$ and using the inversion rules for partitioned matrices one can derive an expression for the information matrix:

$$\left[\mathrm{E} \left[\left[\frac{\partial \ell}{\partial (\mathrm{vec}\ O)'} \right]' \left[\frac{\partial \ell}{\partial (\mathrm{vec}\ O)'} \right] \right] \right]^{-1} = \begin{bmatrix} \Sigma^{-1}+W & -\Sigma^{-1} & \Sigma^{-1}+D' \\ -\Sigma^{-1} & \Sigma^{-1} & -\Sigma^{-1} \\ \Sigma^{-1}+D & -\Sigma^{-1} & \Sigma^{-1}+C_l \end{bmatrix}^{-1} = \left[Z^{ij} \right],$$

where i and j run from 1 to 3.

If one defines the partitioned score vector as $\left[a_{lj} \right]$, $j = 1,2,3$ one can write down the expression for the simultaneous LM test statistic for the existence of either sort of outlier, be it additive, innovative, transient or a combination, which has a χ^2_{3n} limit distribution as

$$LM_{AIT} = \sum_{i=1}^{3} \sum_{j=1}^{3} a_i' Z^{ij} a_j, \tag{5.6}$$

with

$Z^{11} = (W - D'C^{-1}D)^{-1}$,

$Z^{21} = Z^{12'} = (I - C^{-1}D)Z^{11}$, $Z^{22} = \Sigma + C^{-1} + (I - C^{-1}D)Z^{11}(I - C^{-1}D)'$

$Z^{13} = Z^{31'} = -W^{-1}D'Z^{33}$, $Z^{23} = Z^{32'} = (I - W^{-1}D')Z^{33}$, $Z^{33} = (C - DW^{-1}D')^{-1}$,

$a_1 = \Phi'(F)\Sigma^{-1}\varepsilon_l$, $a_2 = -\Sigma^{-1}\varepsilon_l$, $a_3 = \Gamma'(F)\Sigma^{-1}\varepsilon_l$.

where we have dropped the subscripts l from C_l and a_l for notational simplicity. The expressions for the Z^{ij}'s are derived in appendix A5.1. There it is also shown that (5.6) can be simplified to:

$$\begin{aligned} LM_{AIT} = {} & \left[(\Phi'(F) - I)\Sigma^{-1}\varepsilon_l \right]' Z^{11} (\Phi'(F) - I)\Sigma^{-1}\varepsilon_l + \left[(\Gamma'(F) - I)\Sigma^{-1}\varepsilon_l \right]' Z^{33} (\Gamma'(F) - I)\Sigma^{-1}\varepsilon_l \\ & + \left[(\Gamma'(F) - I)\Sigma^{-1}\varepsilon_l \right]' Z^{31} (\Phi'(F) - I)\Sigma^{-1}\varepsilon_l + \left[(\Phi'(F) - I)\Sigma^{-1}\varepsilon_l \right]' Z^{13} (\Gamma'(F) - I)\Sigma^{-1}\varepsilon_l \\ & + \varepsilon_l' \Sigma^{-1} \varepsilon_l. \end{aligned} \tag{5.7}$$

Conditions for identification

From these formulae we can extract the regularity conditions for the appropriate interpretation of the LM statistic that follow from the regularity condition on the information matrix. The differences between the outlier types in their relation with the parameters of the VAR must be substantial in order to fulfill the regularity conditions.

Condition 5.1 *W is nonsingular.*

This implies that there must be some dynamics in the model, otherwise the difference between additive and innovation outliers ceases to be present. Not every kind of dynamics suffices. If condition 5.1 is not satisfied one should only use specific tests where the additive outlier does not play a role.

Condition 5.2 *C_l is nonsingular.*

I.e. $(1-\rho z)^{-1}\Phi(z)-\Gamma_0$ is nonsingular $\Leftrightarrow$ $\Phi(z)-(1-\rho z)I_n$ is nonsingular. This implies that the way in which the effect of the transient outlier dies out must be different from the way in which the effect of an innovation outlier disappears in the system.

Condition 5.3 *$(W-D'C_l^{-1}D)$ is nonsingular.*

One can look at this matrix as the sum of squared residual matrices

$$\sum_r \tilde{\Phi}_r'\tilde{\Phi}_r - \sum_r \tilde{\Phi}_r'\tilde{\Gamma}_r\left(\sum_r \tilde{\Gamma}_r'\tilde{\Gamma}_r\right)^{-1}\sum_r \tilde{\Gamma}_r'\tilde{\Phi}_r$$

of the "projection" of the $(T-p)$-vector of matrices $(\tilde{\Phi}_r)=(\Sigma^{-\frac{1}{2}}\Phi_r)$ on the analog vector of matrices $(\tilde{\Gamma}_r)$. The condition implies that $\Phi(z)-(1-\rho z)^{-1}\Phi(z)$ must be nonsingular which in turn implies that ρ must be different from zero, otherwise the transient outlier becomes observationally equivalent to an additive outlier. It is clear that these conditions are not fulfilled for the last observations. For $l=T-1$ one has, e.g., $\tilde{\Gamma}_1=\rho\tilde{\Phi}_1$ so that the condition is violated.

If W^{-1}, C_l^{-1} and $(W-D'C_l^{-1}D)^{-1}$ exist it follows from the matrix inversion lemma that $(C_l-DW^{-1}D')^{-1}$ exists too.

5.4.2 Case of estimated parameters and unknown timing

The LM test is asymptotically valid if the quantities involving AR–, drift– and disturbance covariance parameters are evaluated at the restricted ML estimates of the parameter vectors. This means that we have to replace the unknown parameters $\theta=\text{vec}(\Phi_1:\Phi_2:\ldots:\Phi_p:\Sigma:m_0:m_1)$ by their ML estimates. These are asymptotically equivalent to the MLE conditional on the first p observations, which boil down to OLS estimates conditional on the first p observations if θ is unrestricted. In the case of integrated processes for the regressors least squares is consistent whereas solution of the Yule–Walker equations is not feasible because of the asymptotic singularity of the moment matrix of regressors (Tsay (1987)).

In practice the exact position l of the outlier is often unknown. Abraham and Yatawara (1988) presented an asymptotic argument to approximate the

distribution of the maximum of all comparable m–dependent[2] χ_2–distributed LM statistics over the different periods in the sample. Let LM_{max} denote this statistic. They showed for $m=p$ that

$$\lim_{T\to\infty} P(LM_{max} \leq C_T(\xi)) = \exp(-\nu\xi), \tag{5.8}$$

where for a certain significance level α, $C_T(\xi)$ is equal to $2[\ln(T-2p) - \ln(-\ln(1-\alpha)) + \ln(\nu)]$. Using Monte Carlo results for LM_{AI} (see §5.4.3) they advised to use $\nu = 0.8$. ν is a so–called *extremal index*, which measures the effect of local dependence on the tail properties of the maximum, cf. Leadbetter et al. (1983, p67ff.). Actual computation of ν involves a large number of numerical integrations. For the innovation outlier test statistic LM_I of the next section, the extremal index is close to one, i.e. the case where one derives the distribution of the maximum of independent observations. For the transient outlier statistic LM_T in (5.11) based on a large number of subsequent residuals (Γ_i far from zero for large i) ν is much lower.

The χ^2_2 statistic applies in special cases, e.g. if the process is univariate and if we abstract from the possibility of transient outliers. §5.5 contains other examples with p–dependent LM statistics with χ^2_2 limit distributions under the null.

The extension to χ^2–variables of other degrees of freedom involves the inversion of the distribution function of the χ^2. For the specific test for innovation outliers, discussed below one can successfully use the Bonferroni values of example 3.1 to derive critical values of interest for LM_{max}.

5.4.3 Distinguishing between outlier types

For the analysis following the detection of outliers it can be important to know what types of outlier one is dealing with. In order to get insight in the nature of the outlier it is advisable to calculate the three specific LM statistics first:

A: The test at time l for an additive outlier only:
H_{0A}: $O_1 = 0$ Under the assumption that $O_2 = O_3 = 0$ one uses

$$LM_A = a_1'(W + \Sigma^{-1})^{-1}a_1, \tag{5.9}$$

which has a χ^2_n limit distribution.

---

[2] Let $\ldots,\xi_{-1},\xi_0,\xi_1,\ldots$ denote a strictly stationary sequence of random variables defined on a probability space (Ω, B, P). This sequence is said to be *m–dependent* if the random variables $(\xi_i,\ldots,\xi_k)$ and $(\xi_{k+n},\ldots,\xi_j)$ are independent whenever $n>m$, see Billingsley (1968, p. 167).

I: The test at time l for an innovation outlier only:

H_{0I}: $O_2 = 0$ Under the assumption that $O_1 = O_3 = 0$ one uses

$$LM_I = \varepsilon_l' \Sigma^{-1} \varepsilon_l, \tag{5.10}$$

which has a χ^2_n limit distribution.

T: The test at time l for a transient outlier only:

H_{0T}: $O_3 = 0$ Under the assumption that $O_1 = O_3 = 0$ one uses

$$LM_T = a_3' (C_l + \Sigma^{-1})^{-1} a_3, \tag{5.11}$$

which has a χ^2_n limit distribution as well.

Following Tsay (1988), Abraham and Yatawara (1988) and Muirhead (1986) one can form a decision rule to treat the outlier as type A, I or T depending on which statistic shows the largest value or smallest p–value.

The I test is asymptotically equivalent to the test based on the Lawley–Hotelling trace criterion L_i discussed in §A3.2. See appendix §A5.1.2 for a proof. One can conjecture that the use of the critical values of L_i is better suited for use in finite samples, since their approximation is based on "exact" finite sample distribution theory (albeit for fixed regressors). One needs estimates of the parameters for the "subset deleted sample" to compute L_i, viz. OLS estimates with an extra dummy variable in each equation. By application of the small sample correction of Anderson (1958, Theorem 8.6.2) on the likelihood ratio test of linear hypotheses in normal linear equation systems one has an alternative way to compute critical values for the test on innovation outliers. Kiviet (1986) provided encouraging Monte Carlo results on this correction for single equations. Monfort and Rabemananjara (1990) reproduced the formulae, see §A5.1.2.

In order to examine the possibility that an outlier is a combination of two types one can compute the test statistics for two out of three outlier types:

AI: The test at time l for an AO or IO only:

H_{0AI}: $O_1 = O_2 = 0$ Under the assumption that $O_3 = 0$ one uses

$$LM_{AI} = ((\Phi'(F) - I)\Sigma^{-1}\varepsilon_l)' W^{-1} (\Phi'(F) - I)\Sigma^{-1}\varepsilon_l + \varepsilon_l' \Sigma^{-1} \varepsilon_l, \tag{5.12}$$

which has a χ^2_{2n} limit distribution.

The derivation of this statistic is straightforward using the partitioned inverse Y^{-1} from the proof of (5.6) in §A5.1.1. This is a multivariate generalization of the statistic of Abraham and Yatawara (1988). Another example is:

IT: The test at time l for an IO or transient Level Shift only:
H_{0IT}: $O_2 = O_3 = 0$ Under the assumption that $O_1 = 0$ one uses

$$LM_{IT} = ((\Gamma'(F) - I)\varepsilon_l)' C_l^{-1} (\Gamma'(F) - I)\varepsilon_l + \varepsilon_l' \Sigma^{-1} \varepsilon_l, \tag{5.13}$$

which has a χ^2_{2n} limit distribution.

One can define and derive LM_{AT} analagously.

Summarizing one can compute seven test statistics, LM_{AIT}, LM_{AI}, LM_{AT}, LM_{IT}, LM_A, LM_I and LM_T, which all test the same null hypothesis. They are expected to have different powers for specific alternatives where outliers occur. In practice one would like to compute these statistics for a series of likely positions l of outliers. The behavior of the test in the presence of outliers is considered in §5.5.

5.4.4 Distinguishing between outliers in different equations

If one wants to know which equation was hit worst by the exogenous shock O, one should compute the test statistic for H_{0k}: $e_k O = 0$, with e_v an *n–row* vector of zeros except for its v–th element. An equation by equation analysis may be interesting if one uses the structural equation philosophy to build a VAR, see e.g. Monfort and Rabemananjara (1990).

The simultaneous LM test statistic for an outlier of either sort in equation k at time t can be derived easily if one notes the corresponding derivation of the score:

$$\left[q_{tk} \right]_l = \left[\left[\frac{\partial \varepsilon_{tk}}{\partial((e_v O)')} \right]' \right]_l = Ind_{[l,T]}(t) \begin{bmatrix} \Phi_{(t-l)kv} \\ -\delta_{0,t} I_n \\ \Gamma_{(t-l)kv} \end{bmatrix}, \tag{5.14}$$

$$\left[\frac{\partial \ell}{\partial((e_v O)')} \right]_l' = \sum_{r=0}^{T-l} \begin{bmatrix} e_v \Phi_r' \\ -\delta_{0,r} e_v I_n \\ e_v \Gamma_r' \end{bmatrix} \Sigma^{-1} \varepsilon_{l+r}. \tag{5.15}$$

The statistics for the null hypotheses H_{0k}: $e_v O = 0$; $v = 1, 2, \ldots, n$ become

$$LM_{vAIT} = \sum_{i=1}^{3} \sum_{j=1}^{3} a_{vi}' Z_v^{ij} a_{vj}, \tag{5.16}$$

with

$a_{vi} = e_v a_i$; $i = 1, 2, 3; k = 1, 2, \ldots, n$, and

$Z_v^{11} = 1/(e_v W e_v' - (e_v D' e_v')^2/(e_v C e_v'))$,

and the other Z_v^{ij} defined correspondingly using the v–th diagonal elements of the matrices W, D, C only. LM_{vAIT} is asymptotically distributed as a χ^2 with 3 degrees of freedom. Note that the equation by equation analysis still has an

important multivariate aspect, not only through the use of the lags of all variables in each equation but also by the standardization of the residuals with the matrix Σ, which is often not diagonal in macroeconomic applications.

5.5 An artificial example

In this section some of the tests introduced in section 3 are applied to an artificial data set. We generated 100 observations using the nonstationary VAR model with the following parameters:

Example 5.1

$$n=2,\ p=2,\ m_0=0,\ m_1=0,\ l=50,$$

$$\Phi_1=\begin{bmatrix}-2.30 & 0.00\\ -1.10 & -0.80\end{bmatrix},\ \Phi_2=\begin{bmatrix}1.82 & -0.48\\ 0.97 & -0.08\end{bmatrix},\ \Sigma=\begin{bmatrix}1.0 & 0.0\\ 0.0 & 1.0\end{bmatrix}. \tag{5.17}$$

The eigenvalues of this system are 1, $0.850\pm0.278i$ and 0.400 with $i^2=-1$. $\Phi(1)$ can be written as $\gamma\alpha'$ with[3] $\gamma=(0.04\ -0.01)'$ and $\alpha=(13\ -12)'$. If we define $\Phi^*(L)$ from the equation $\Phi^*(L)(1-L)=\Phi(L)-\Phi(1)L$ one can derive that $\gamma_\perp'\Phi^*(1)\alpha_\perp$ has rank 1. System (5.1) can then be interpreted as a Vector Error Correction Model (see §6.2.3): $\Phi^*(L)(1-L)y_t=-\gamma(\alpha'y_{t-1})+\varepsilon_t$, with $\Phi_0^*=\Phi_0$, $\Phi_1^*=-\Phi_2$, $\Phi_2=0$.

For a derivation of this approach of the Granger Representation Theorem see §6.3 and Johansen (1991a, §4). $\gamma_\perp$ is an $n\times(n-r)$ matrix of rank $n-r$ so that $\gamma_\perp'\gamma=0$, $\alpha_\perp$ is an $n\times(n-r)$ matrix of rank $n-r$ so that $\alpha_\perp'\alpha=0$.

For the simulation we used drawings generated by routine G05EZF from the NAG library (NAG (1988)), with seed 100 and starting values (2 2)' and (2.5 2.5)' for the first two observations of y_t.

As one can see from figure 5.2 the complex roots imply stochastic cyclical behavior for the first differences of the series. Besides the cointegration property this characteristic can be an interesting phenomenon of economic time series. The common factor in the spectral densities of $(1-L)y_{1t}$ and $(1-L)y_{2t}$ which stems from the squared determinant of $\Phi(e^{i\omega})$ divided by the one unit root $(1-e^{i\omega})$ $(0\le\omega<2\pi)$ has a maximum for a period of approximately 19.9 observations, which can be a reasonable guess for the length of business cycles in quarterly macroeconomic time series for the United States (Beveridge and Nelson (1981). Kunst (1989) found similar complex roots in his estimates for a VAR system of quarterly series of macroeconomic key variables of Western Germany (1960–1987), see §7.4.7.2 for other empirical evidence.

To illustrate the effect that the different kinds of outliers have on the system we introduce the following outlier models.

---

[3] We use the original notation of Engle and Granger (1987).

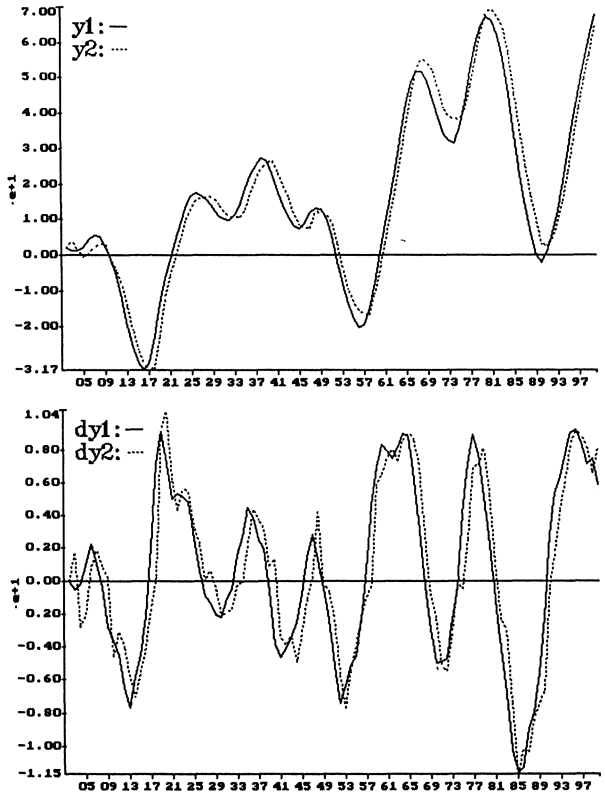

Figure 5.2 *Data from example* 5.1
Time series plots of 100 observations from an artificial sample of cointegrated system (5.17) in levels (y1 and y2) and first differences (dy1 and dy2).

Examples 5.2–5.4

5.2 *The AO model*: (2.1), (5.1) and $O_1 = (4\ 4)'$, $O_2 = 0$, $O_3 = 0$, $l = 50$. (5.18)

5.3 *The IO model*: (2.1), (5.1) and $O_1 = 0$, $O_2 = (4\ 4)'$, $O_3 = 0$, $l = 50$. (5.19)

5.4 *The TO model*: (2.1), (5.1) and $O_1 = 0$, $O_2 = 0$, $O_3 = (4\ 4)'$, $l = 50$, $\rho = .9$. (5.20)

Observed outliers

It is easy to note the different effects of the outliers from figure 5.3. The additive outlier has an effect for only one period on the level and for two periods on the first differences of the series.

Because of the unit root in the system, the innovation outlier has a permanent effect on the levels of the series.

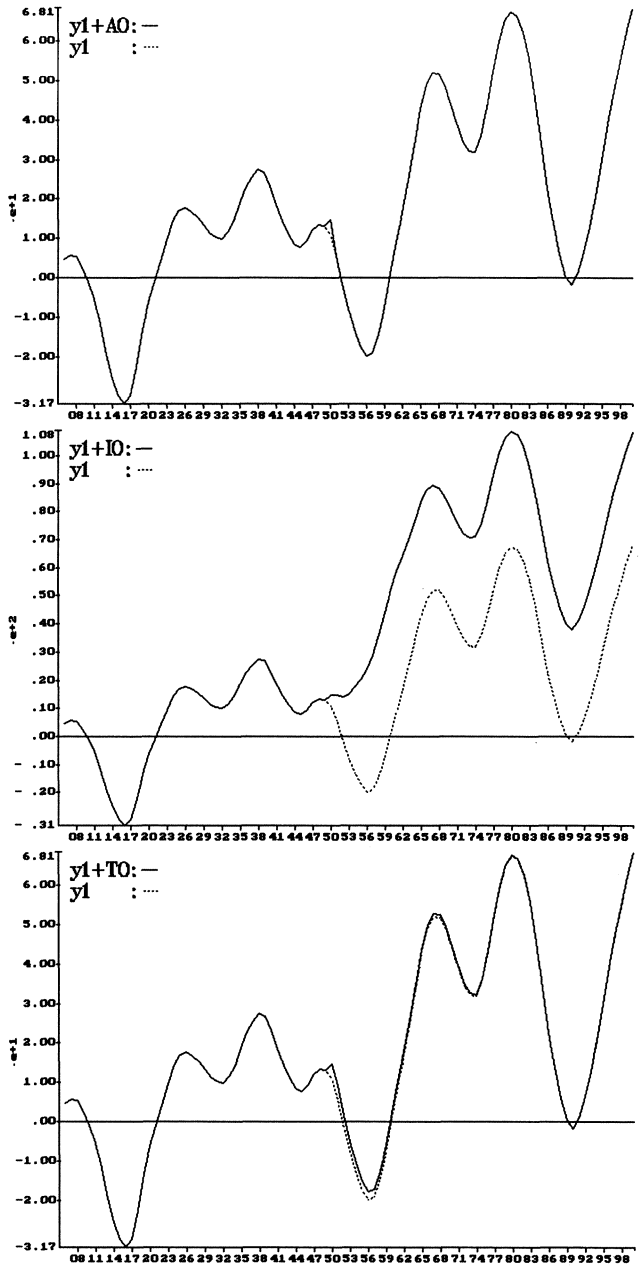

Figure 5.3a *Effect of different outliers on levels*
Time series plots (solid) of the first variable of a cointegrated system affected by an additive outlier (AO), innovation outlier (IO) and transient outlier (TO) at t=50 together with the unaffected series (dotted).

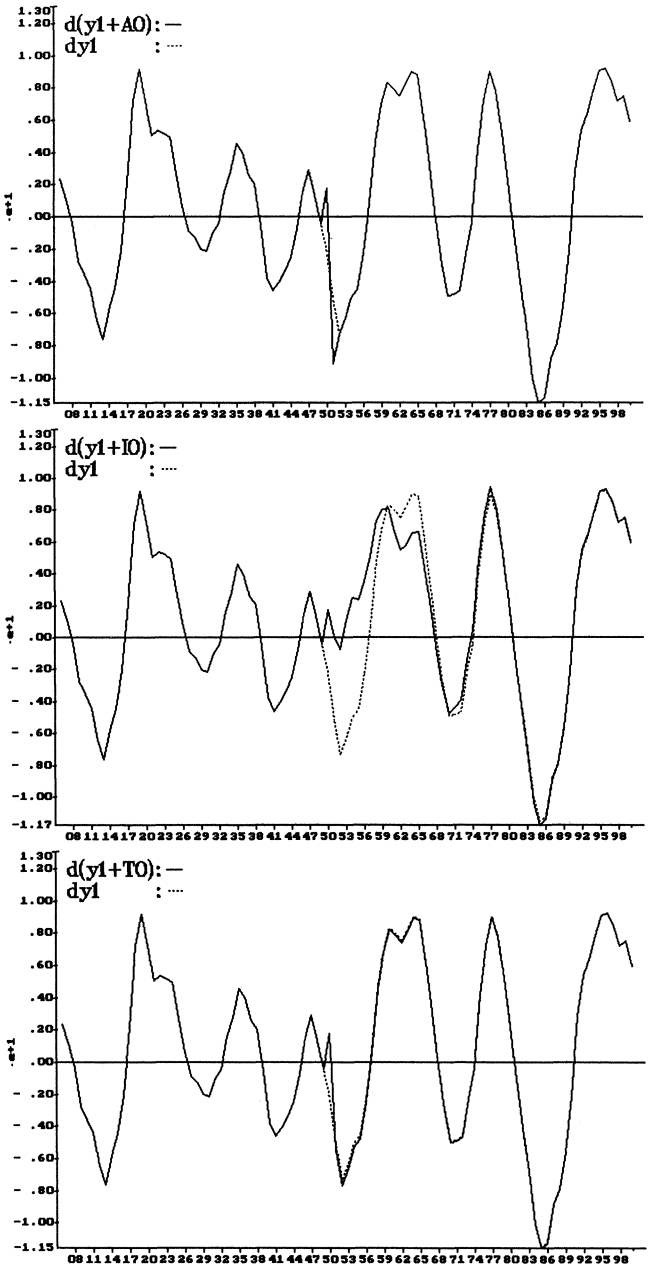

Figure 5.3b *Effect of different outliers on first differences*
Time series plots (solid) of the first variable of a cointegrated system affected by an additive outlier (AO), innovation outlier (IO) and transient outlier (TO) at t=50 together with the unaffected series (dotted), in first differences.

Table 5.1 *Influence of outliers on estimates in nonstationary VAR system*

		Model			
	True Process	Estimated	Estimate AO	Estimate IO	Estimate TO
$\hat{\lambda}$	$\begin{bmatrix} 1 \\ .85+.278i \\ .85-.278i \\ .40 \end{bmatrix}$	$\begin{bmatrix} .961 \\ .899+.323i \\ .899-.323i \\ .329 \end{bmatrix}$	$\begin{bmatrix} .947 \\ .876+.327i \\ .876-.327i \\ .287 \end{bmatrix}$	$\begin{bmatrix} .988 \\ .886+.321i \\ .886-.321i \\ .345 \end{bmatrix}$	$\begin{bmatrix} .958 \\ .890+.327i \\ .890-.327i \\ .300 \end{bmatrix}$
$Norm_1$	Expected:2	1.93	1855.60*	3.77	329.61*
$Norm_2$	Expected:2	0.55	378.17*	2.84	6.60*
NormS	Expected:4	2.11	204.44*	10.15*	40.56*
NormK	Expected:1	1.05	983.51*	2.67	168.57*
$AC(1)_1$	Expected:1	1.82	4.19*	0.46	0.64
$AC(1)_2$	Expected:1	0.28	4.92*	0.02	0.26
$AC(1)_v$	Expected:1	1.04	3.14*	0.23	0.40
$ARCH_1$	Expected:1	1.78	14.54*	2.59	28.04*
$ARCH_2$	Expected:1	1.58	12.36*	1.04	12.20*
$-2\ln(Q)$	$r \leq 1$	1.42	1.92	0.39	1.56
	$r = 0$	67.20*	46.86*	61.38*	50.52*
$\hat{D} = \det(\hat{\Sigma})$	1	1.02	3.13	1.26	1.88
$LR(D)$		1	111.82*	22.64*	61.86*
$\hat{\alpha}_1$	$\begin{bmatrix} 13/12 \approx 1.083 \\ -1.000 \end{bmatrix}$	$\begin{bmatrix} 1.096 \\ -1.000 \end{bmatrix}$	$\begin{bmatrix} 1.090 \\ -1.000 \end{bmatrix}$	$\begin{bmatrix} 1.095 \\ -1.000 \end{bmatrix}$	$\begin{bmatrix} 1.094 \\ -1.000 \end{bmatrix}$

NOTES: $\hat{\lambda}$: Eigenvalues of the system. $Norm_i$: LM test for normality of the residuals in equation i (Bera and Jarque (1982)), under the null asymptotically distributed as a χ^2 with two degrees of freedom. NormS and NormK: Tests for multivariate normality on Skewness and Kurtosis (Mardia (1970)), $\chi^2(n^2)$ and $\chi^2(1)$ under the null. $AC(1)_i$: F-version of test for (vector) first order serial correlation in the errors inflicting the i-th equation, under the null approximately distributed as $F(n,T-p-m)$, with m the number of regressors in the regression; Breusch and Godfrey (1981). $AC(1)_v$: Idem for VAR as a whole, $F(n^2, n(T-p-m))$; Poskitt and Tremayne (1982). $ARCH_i$: LM Test against multivariate first order serial correlation in conditional variance of the error term for equation i, Tsay (1987), $F(n,n(T-p))$. $-2\ln(Q)$: LR (trace) statistic for the rank r of $\phi(1)$ against the full rank alternative of $r=n$. LR(D): Likelihood ratio. $\hat{\alpha}_1$: ML estimate of the first two cointegrating vectors in the system so that $\phi(1)=\gamma\alpha_1'$ has rank 1; Johansen (1991a), §A6.1. C(1): sum of MA coefficients in Vector MA representation of Δy_t in restricted model; Johansen (1991a). An asterisk denotes significance at the 5%-level. A constant was included in all the regressions. A trend term only in the unrestricted VAR(2).

In this system the long run effect of the innovation outlier is (40 40·(13/12))', which can be derived using the $C(1)$ matrix of the Vector MA representation of (5.17), with $(1-L)y_t = C(L)\varepsilon_t$, where $C(1) = (1\ 13/12)'(2\ 8)$, see also §6.3.3 below. Such values can occur in macroeconomic empirical applications when there is high positive autocorrelation in the first differences of the series, which is true for

variables that are nearly $I(2)$, like some logged macroeconomic price levels.

The transient outlier with ρ=0.9 affects only one observation of the differenced series significantly. The influence on the first differences is given by $\delta_{t,i}(1-L)(1-0.9L)^{-1}(\ 4\ 4)'$. The level is affected for a longer period. The first couple of observations after the outlier are influenced in a way that mimics the effect of a level–shift. One can therefore expect the test for a transient outlier to have considerable power against a level–shift alternative as well.

Effect on parameter estimates

One can inspect the influence of the outliers on the estimated eigenvalues of the system, model specification tests and estimates of the cointegrating from table 5.1. We used only one random sample. Of course other parameter specifications will give different outcomes. The example is selected in such a way that some important well known features of macroeconomic VAR systems are captured. We abstract here from seasonality and deterministic drift terms.

The outliers nearly satisfy the cointegrating relation. Their size ($4\sigma_{11}$=$4\sigma_{22}$) is not too unrealistic. The occurrence of a vector of outliers of this size is unlikely under the null hypothesis though, which can be noted from the huge values for $LR(D)$ in table 5.1. The effect of the outliers shows thus quite clearly.

A Monte Carlo experiment would be appropriate in order to draw statistically meaningful conclusions about the power and true size of the tests and the influence of different combinations of outliers on the other statistics. Such an experiment is beyond the scope of this study. In macroeconometric practice one is not able to conduct such an experiment and one also has to confine oneself to one "drawing" of the series. One ought to look at this as an example of what could occur.

We see that the model specification test for normality of the error term is very sensitive to the additive outlier (AO) and transient outlier (TO). These outliers generate more "aberrant" residuals than does the innovation outlier (IO). For this parameter structure the AO introduces spurious indications of (negative) autocorrelation in the residuals. This is a serious problem. The introduction of extra lags as a wrong treatment of the defects of the model could worsen the power of the test statistic, thus preventing the right diagnosis. The same would hold for the introduction of extra parameters to capture spurious serial correlation in the residual variance (Tsay (1987)). On the other hand we see that the AO pulls the estimates of the characteristic roots of the system towards zero.

The misspecification tests are most badly affected by the AO and the least by the IO. The last one even has beneficial effect in improving the estimate of the largest eigenvalue of the system and in helping to identify the true cointegrating rank of the system more clearly. Here the IO can be interpreted as a deterministic input which improves the "signal–to–noise ratio" of the innovations. Statistical inference on the true rank r of $\Phi(1)$ is not decisively influenced by the outliers in this example at conventional significance levels.

Identification of outliers

Let us now turn to the application of the tests. We computed statistics (5.6), (5.9), (5.10), (5.11), (5.12) and (5.16) with Φ_r replaced by its OLS estimator and Σ replaced by the "unbiased" estimator $\sum \varepsilon_t \varepsilon_t' / (T - p - (np+1))$. Under the null the statistics seem to follow their theoretical null distribution fairly well as can be judged from figure 5.4. The mean corresponds to the theoretical one which equals the degrees of freedom which is the number of equations times the number of outlier types concerned as derived in the previous section. The number of observations beyond the 5% critical values[4] for the respective χ^2–distributions varies from 1 to 6, (LM_{1A}, LM_{1I}: 1; LM_I, LM_{AI}: 2; LM_{AIT}, LM_A: 3; LM_{2A}, LM_{2I}, LM_{2T}: 6). This seems to be on the conservative side. It may be due to the dependence of the test statistics. The values for LM_{Imax}, LM_{Amax}, LM_{Tmax} which are the maxima over l for the whole estimation period as in (5.8) are all well below the critical value of 14 derived in Abraham and Yatawara (1988) for the χ^2_2–distributed statistics.

Under the alternative hypotheses (5.18), (5.19) and (5.20) the simultaneous test is seen to work best for the AO and TO alternative, see figures 5.5 and 5.7. For the AO we find $LM_{AIT} = 62.7$; $LM_A = 59.1 > LM_T = 50.1 > LM_I = 45.7$. The TO model gives $LM_{AIT} = 45.4$; $LM_T = 41.0 > LM_I = 32.8 > LM_A = 32.3$.

The IO again does not seem to have such a big impact as the other alternatives. The identification of model (5.18) from the results plotted in figure 5.6 does not follow immediately. The test statistic for LM_I at $l = 50$, 16.3, is lower than that for LM_A at $l = 49$, 17.9. Interpreted as LM_{Amax} and LM_{Imax} they are both significant at the 5% level (critical value: 14, see section 3.2). The outlier could therefore be wrongly identified as an AO one period before the true IO. In order to make the IO distinguishable as such in the right period we increased O_2 to (7 7)' in (5.20) to get example 5.6.

---

[4] Critical values 3.84, 5.99, 7.82, 9.49, and 12.59 for degrees of freedom 1, 2, 3, 4, and 6.

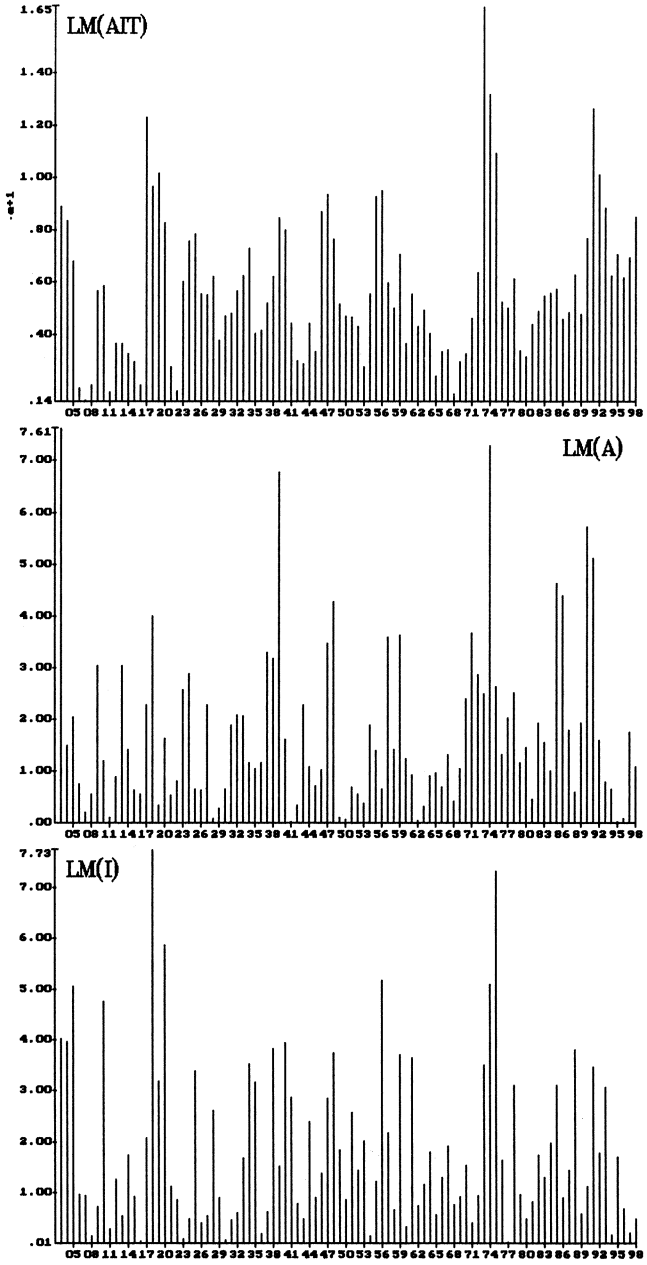

Figure 5.4a *LM tests under* H_0

Time series plots (l=3,4,...,98) of outcomes of simultaneous and specific test statistics for example 5.1 without outliers. Statistics LM_{AIT}, LM_{AI} and LM_{1T}. 5% asymptotic critical values if timing known: 12.59, 9.49 and 3.84.

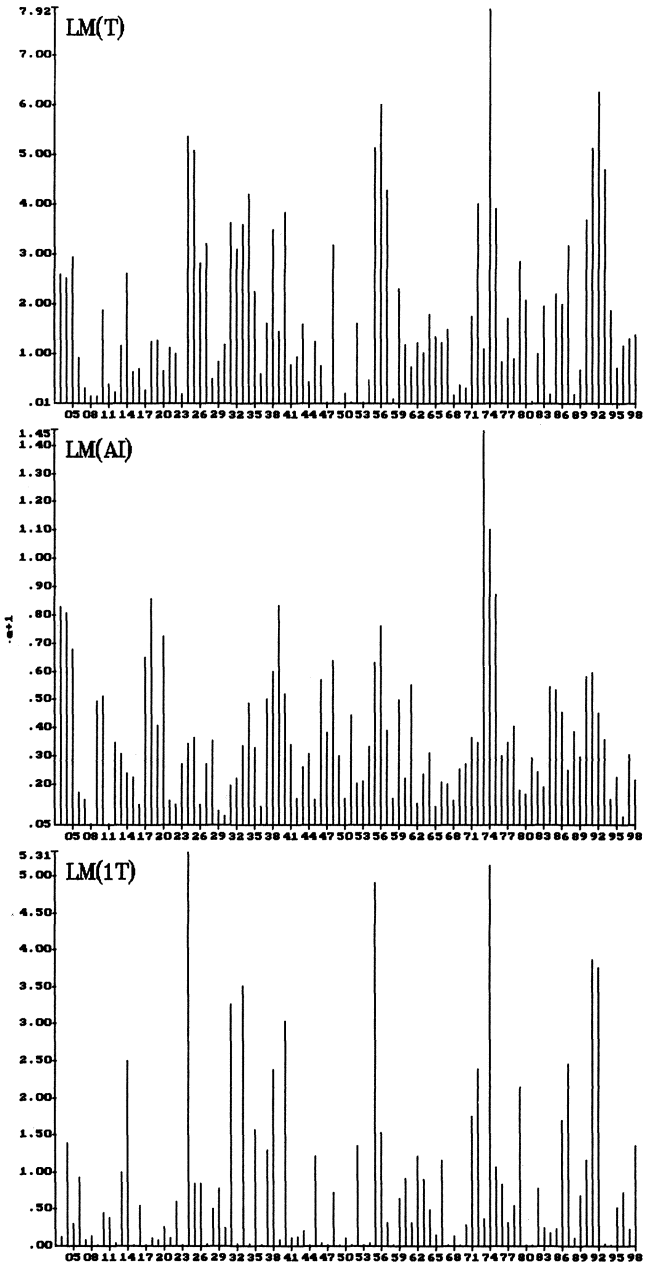

Figure 5.4b *LM tests under H_0 (continued)*
Time series plots (l=3,4,...,98) of outcomes of specific test statistics for example 5.1 without outliers. Statistics LM_A, LM_I, LM_T. 5% asymptotic critical value if timing known: 5.99.

Example 5.6

The large IO model: (2.1), (5.1) and $O_1 = 0$, $O_2 = (7\ 7)'$, $O_3 = 0$, $l = 50$. (5.21)

The simultaneous test comes out as large as with $O_3 = (4\ 4)'$ for the pure TO model as can be seen by comparing figures 5.7 and 5.8. The results of this experiment show that the simultaneous test is a good aid to test the null hypothesis, but when the null is rejected one should also compute the specific tests to get a clearer picture of the situation. Notice in figures 5.6a, 5.7a and 5.8a that the simultaneous test reaches a maximum before the period in which the outlier actually occurs. This alarm rings too early for the innovation and transient outlier.

From figure 5.5 one can see that by only using LM_I, which comes closest to a usual (eye–ball) test for exceptional residuals, one would tend to suspect an outlier one period later than the additive outlier occurred.

The example highlights the forward looking nature of the test. It takes at least p observations after the outlier occurs to identify its nature and timing correctly. In order to model the outliers it is therefore advisable to use the information of the later observations in the estimation of the parameters.

Masking

Straightforward recursive estimation makes it especially hard to identify outliers in the beginning of the sample. An iterative procedure as in Tsay (1988) and Kleiner et al. (1979) which uses the whole data set several times seems to be obligatory. This is also important, because it is rarely observed that an outlying observation in a multivariate data set comes on its own. When there are several outliers, one can mask the other e.g. by its influence on the nonrobust ML estimate of the variance/covariance matrix. We do not know of any easy solutions to this problem. Martin and Yohai (1986) remarked in the rejoinder to their survey paper on influence functions for time series: "Of course one cannot completely solve the *masking problem* with any degree of computational ease."

One could modify the tests and increase their power by using a more robust method to estimate this matrix. Tsay (1988) took such an approach, by estimating it under the alternative assumption with certain outliers. The caveats related to the first round analysis of the residuals using the LM tests must be borne in mind. Note e.g. the consistency problem of the OLS estimators under a level–shift, discussed in §5.3.2. We discuss an iterative "data cleaning" procedure in §A5.3. The next section provides an empirical example where the first round procedure seems to work already fairly well.

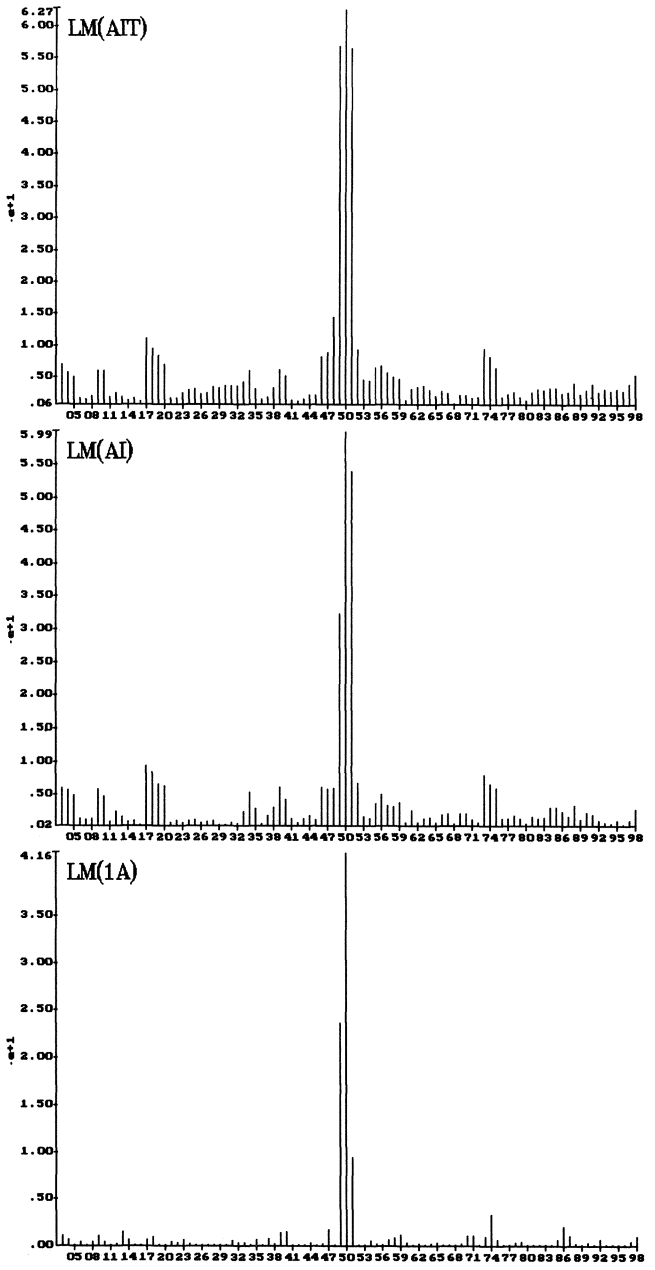

Figure 5.5a *LM tests with AO model*
Time series plots (l=3,...,98) of outcomes of simultaneous and specific tests for model (5.18) with additive outlier at l=50. The overall test detects the outlier at l=50. Statistics LM_{AIT}, LM_{AI}, LM_{1A}, 5% asymptotic critical values if timing known: 12.59, 9.49, and 3.84.

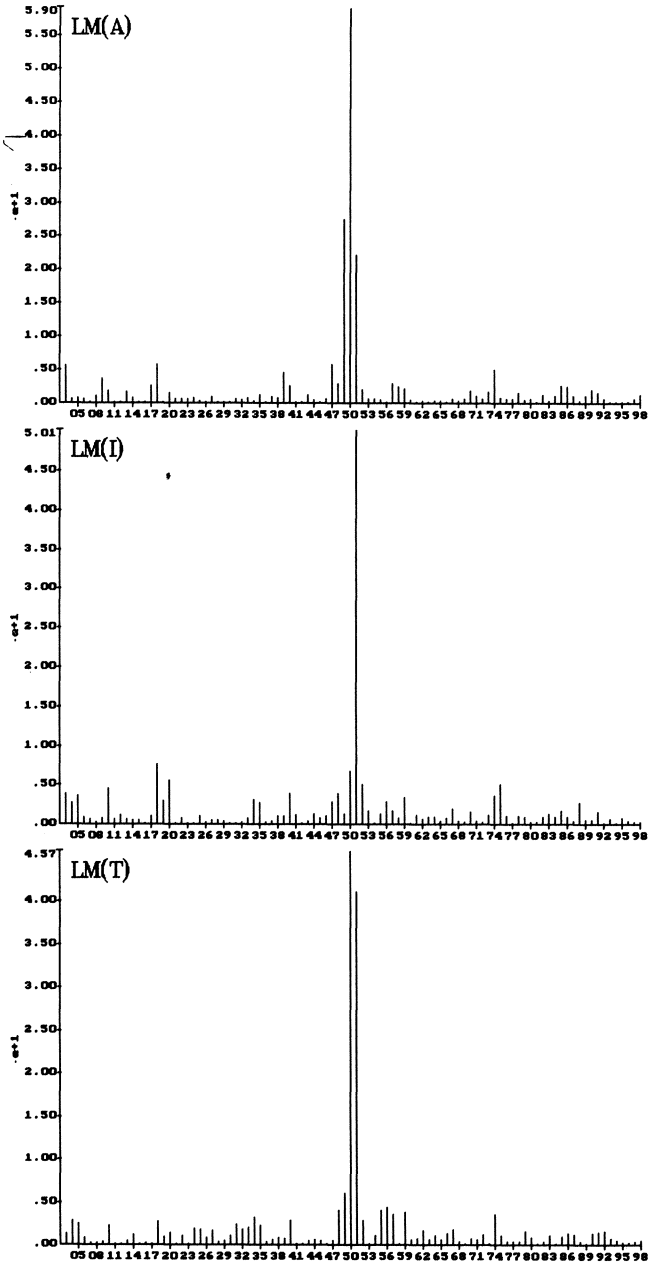

Figure 5.5b *LM tests with AO model (continued)*
Time series plots (l=3,...,98) of outcomes of specific tests for model (5.18) with additive outlier at l=50. An AO for l=50 is the most likely simple alternative. Statistics LM_A, LM_I and LM_T. 5% asymptotic critical values if timing known: 5.99.

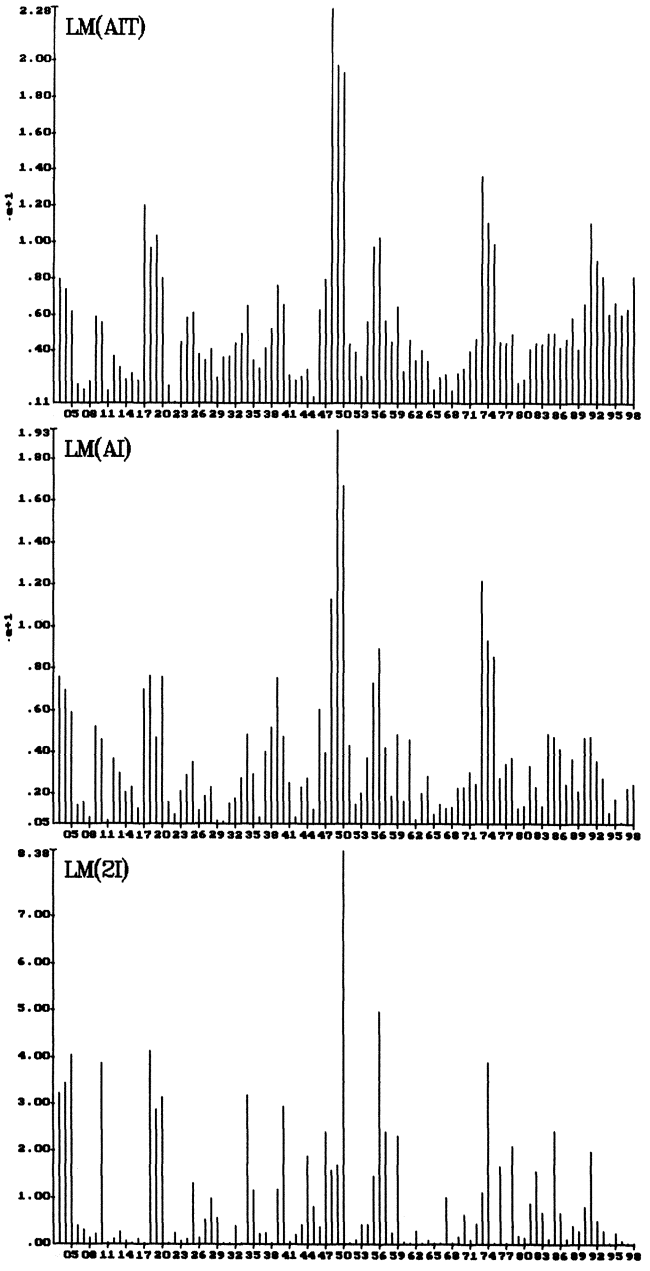

Figure 5.6a *LM tests with IO model*
Time series plots (l=3,...,98) of outcomes of simultaneous and specific tests for model (5.19) with innovation outlier at l=50. The overall test detects the outlier at l=48. Statistics LM_{AIT}, LM_{AI}, LM_{2I}. 5% asymptotic critical values if timing known: 12.59, 9.49, 3.84.

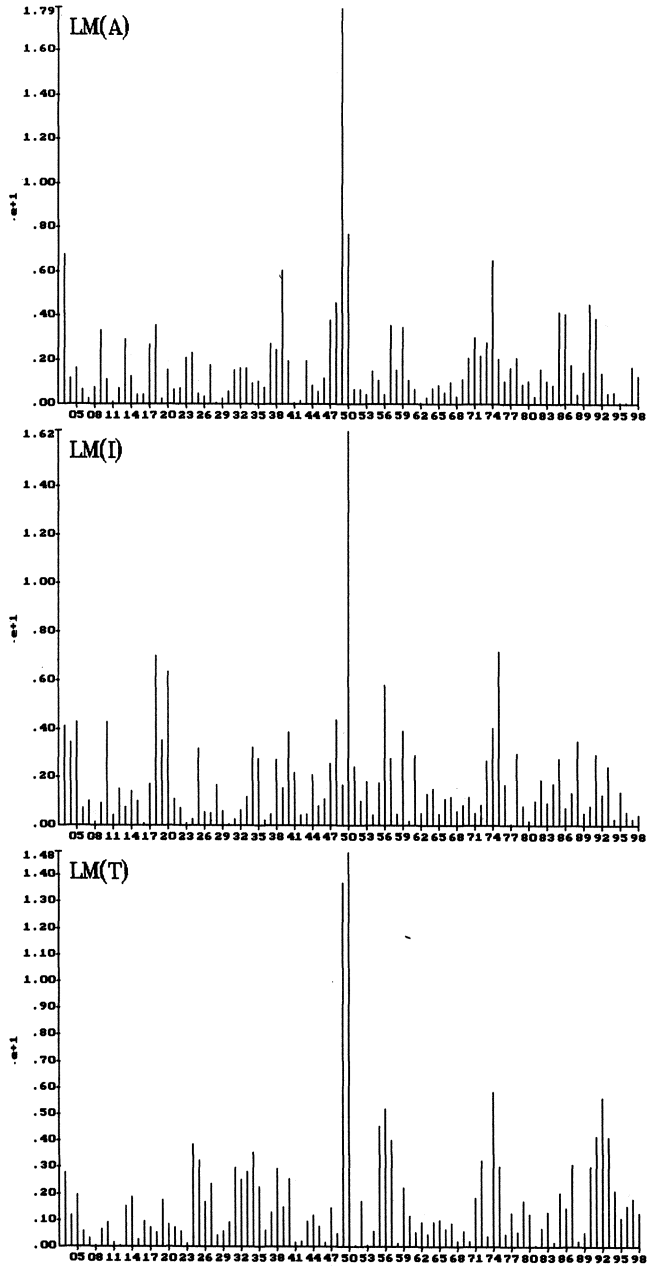

Figure 5.6b *LM tests with IO model (continued)*
Time series plots (l=3,...,98) of outcomes of specific tests for model (5.19) with innovation outlier at l=50. An AO for l=49 is the most likely simple alternative. Statistics LM_A, LM_I and LM_T. 5% asymptotic critical values if timing known: 5.99.

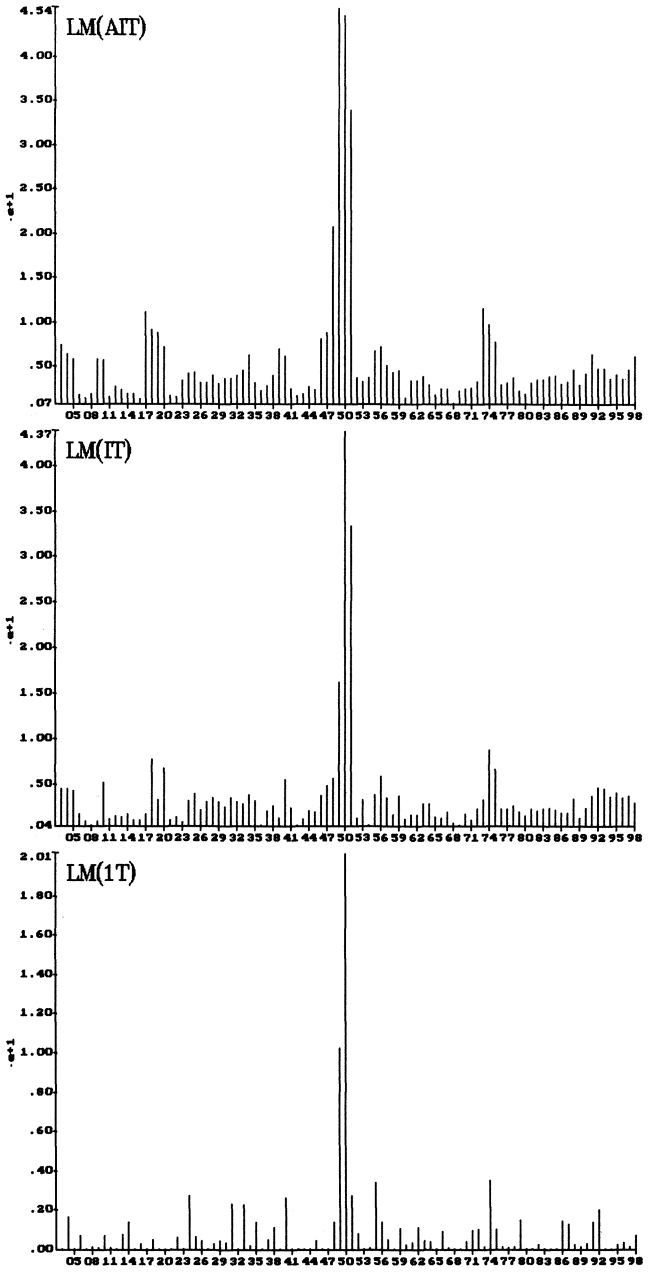

Figure 5.7a *LM tests with TO model*
Time series plots (l=3,...,98) of outcomes of simultaneous and specific tests for model (5.20) with transient outlier at l=50. The overall test detects the outlier at l=49. Statistics LM_{AIT}, LM_{AI}, LM_{1T}, 5% asymptotic critical values if timing known: 12.59, 9.49, 3.84.

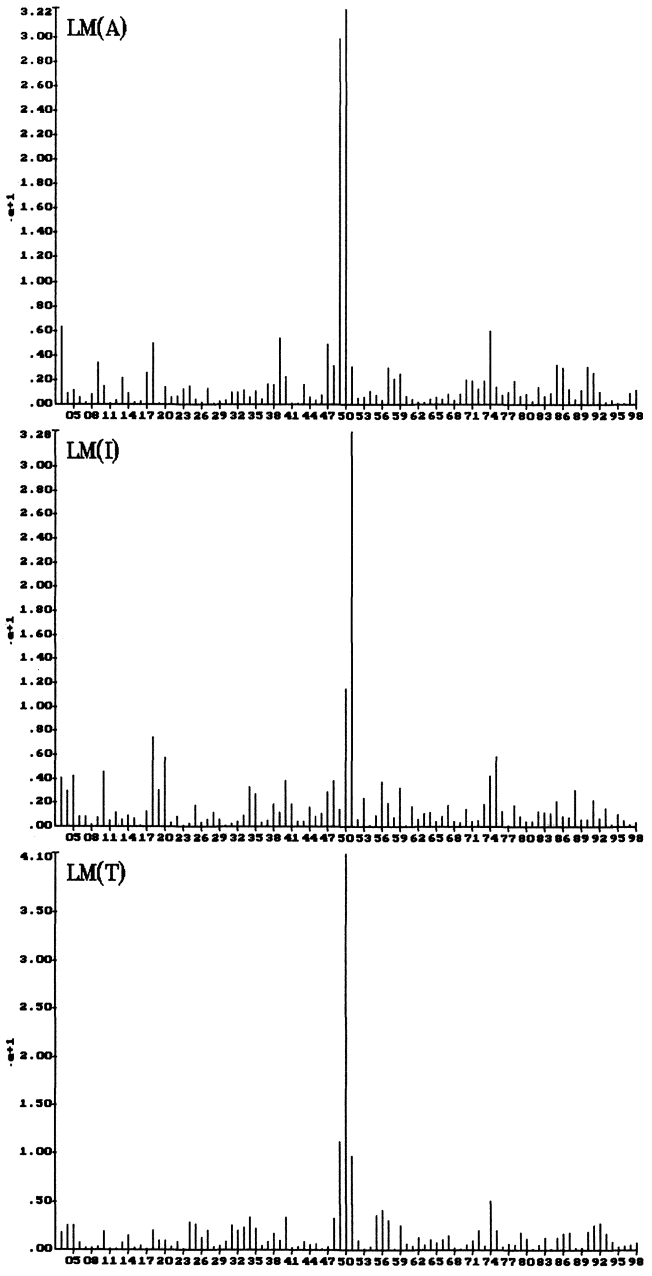

Figure 5.7b *LM tests with TO model (continued)*
Time series plots (l=3,...,98) of outcomes of specific tests for model (5.20) with transient outlier at l=50. A TO for l=50 is the most likely simple alternative. Statistics LM_A, LM_I and LM_T. 5% asymptotic critical values if timing known: 5.99.

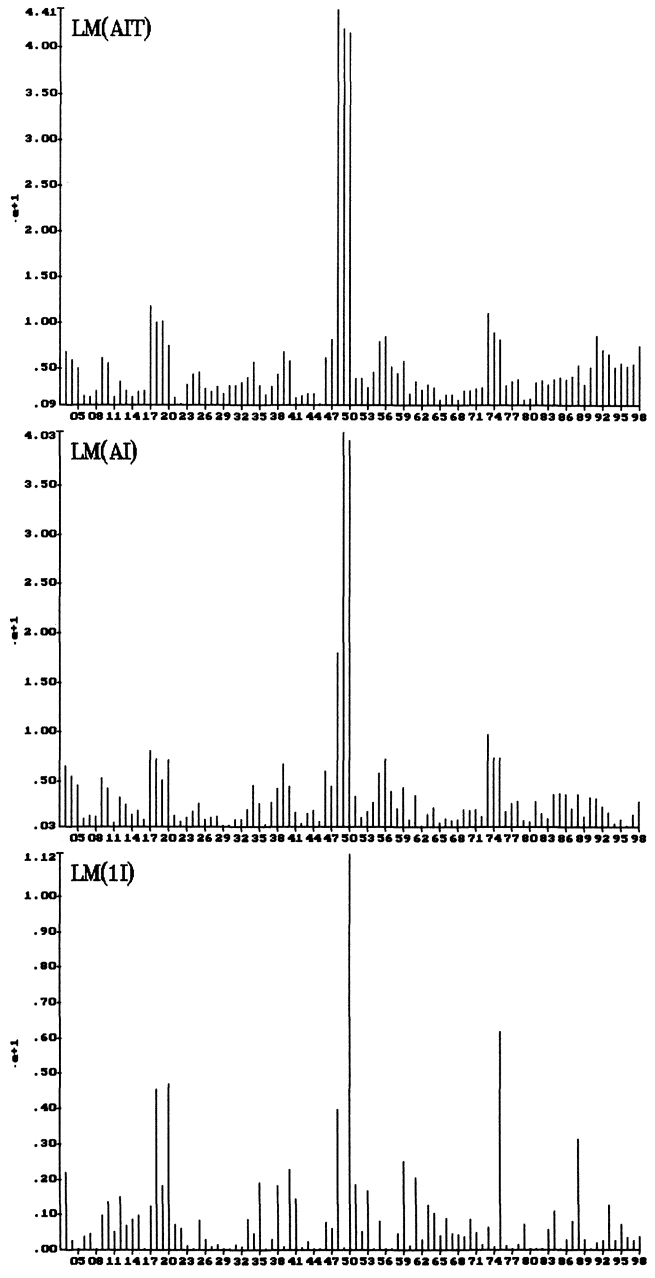

Figure 5.8a *LM tests with large IO model*
Time series plots (l=3,...,98) of outcomes of simultaneous and specific tests for model (5.21) with a large innovation outlier at l=50. The overall test detects the outlier at l=48. Statistics LM_{AIT}, LM_{AI}, LM_{1I}, 5% asymptotic critical values if timing known: 12.59, 9.49, 3.84.

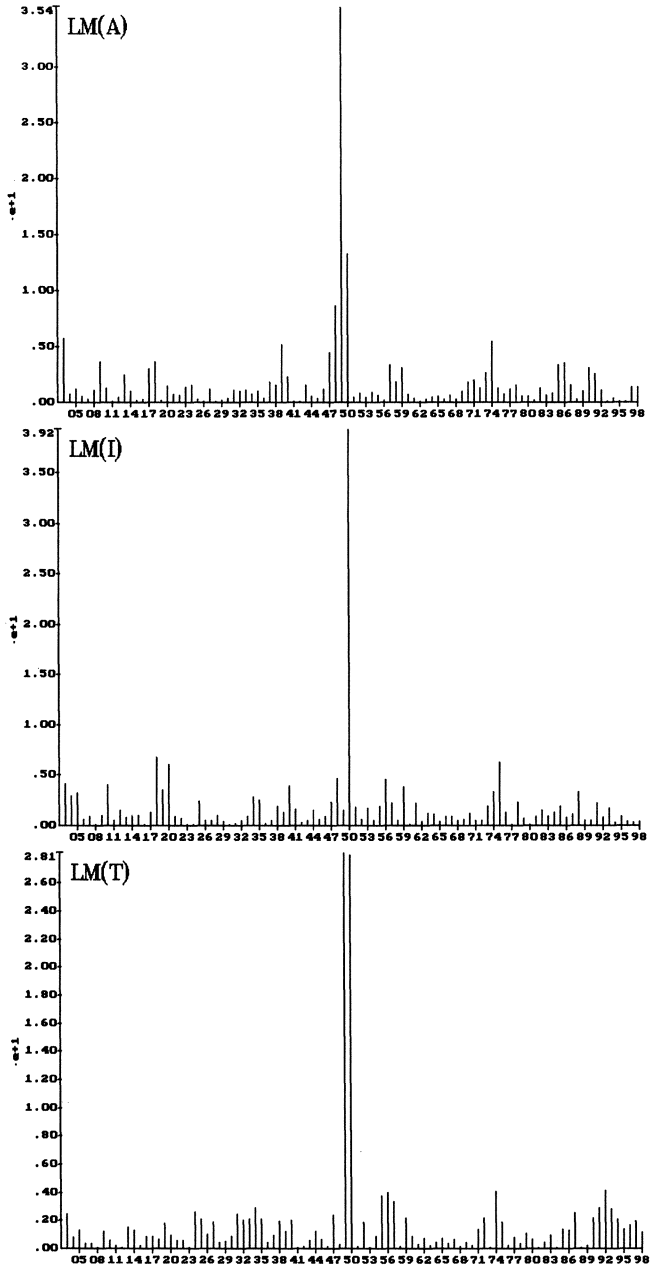

Figure 5.8b *LM tests with large IO model (continued)*
Time series plots (l=3,...,98) of outcomes of specific tests for model (5.21) with a large innovation outlier at l=50. An IO for l=50 is most likely simple alternative to the null. Statistics LM_A, LM_I and LM_T. 5% asymptotic critical values if timing known: 5.99.

5.6 Application to macroeconomic series

Introduction

The official institutions that collect the macroeconomic data apply various kind of corrections. One supposes that these corrections make the data easier to interpret in economic research. It depends on the aim of the study of the researcher whether he should be happy or not with these corrections. Correction for (series of) additive outliers or level–shifts can be performed quite easily. If the institution decides to correct for a transient outlier (e.g. a rise in the consumer price level due to a change in the VAT system (CPB (1970, pp. 76–83), OECD (1970)) this nonstationarity goes beyond notice of the researcher, who may well be especially interested in this phenomenon. In §7.3 we discuss specific information for Dutch data on correction for transient outliers in the volume data of national accounts.

Correction for additive outliers removes other information which may be important. Here one can think of a shift of consumption from one year to the year before because of an announced increase in prices, i.e. an additive outlier in the cumulative total consumption series since the start of the observation period. This could be important in modeling expectations formation. Without a specific (VAR) model a correction for innovative outliers is not feasible. Data information about long run effects is usually very limited, so that a priori information in the model selection procedure is crucial.

A basic question of the macroeconomic unit root literature (JME (1988)) is that of the persistence of shocks in the sense of O_2 in (2.1). Correction for innovation outliers and measurement of persistence should be done simultaneously. When a shock O_2 is interpreted as O_1 or O_3 and the data are corrected correspondingly, the unit root analysis may be badly affected. As Perron (1989) illustrated the nonstationarity may be confined to one or two observations. If the series is corrected the persistence or nonpersistence of important shocks may go unnoticed which may lead us to wrong conclusions.

IFS data for France

Here we apply the testing method to a quadrivariate series that still contains some definite outliers. The series are quarterly macroeconomic data for France[5] 1965.1–1988.1 obtained from a tape issued by International Financial Statistics in May 1988. The four variables are:

---

[5] Backus et al. (1992) used these data *without any outlier correction* in their study on "International Real Business Cycles".

y_1: y: log GDP in 1980 prices. (series 99br)
y_2: c: log private consumption (series 96f, deflated by consumer price index)
y_3: i: log gross fixed capital formation (series 93e,deflated by GDP deflator)
y_4: dp: first difference of log consumer price index (series 64)

The series are plotted in fig. 5.9 and their first differences in fig. 5.10. A VAR(2) including trends and constants was fitted to the data. Seasonality seemed to have been removed. The order was indicated by the Akaike criterion (also applicable in models with unit roots, Tsay (1987)). The outcomes of the different tests are plotted in figs. 5.11–5.13. Table 5.3 provides some descriptive statistics and outcomes of model specification– and cointegration tests. We discuss these tests in more detail in chapters 3 and 6.

Table 5.2 gives the results for the univariate unit root analysis. These do not seem to contradict the null hypothesis of $I(1)$ for all variables. The dangerous $I(2)$–hypothesis can be rejected when one does not use too many lags in the Dickey–Fuller regression. Judging from this information the null hypothesis of a unit root in the system is hard to reject. Innovation outliers must thus be considered to have a long lasting or permanent effect on the level of the series. The test statistics were computed from the OLS residuals of the VAR with $\rho = 0.9$.

Table 5.2 *Results of univariate Dickey–Fuller tests for French data*

	Estimated Model							
	DF1(4)	DF1(4)	DF1(8)	DF1(4)	DF2(4)	DF2(4)	DF2(8)	DF2(4)
period:	66–88	69–88	67–88	66–83	66–88	69–88	67–88	66–83
y	-1.09	-1.98	-1.35	-0.37	-3.37*	-3.33*	-2.93	-3.26*
c	-1.05	-1.14	-0.98	-0.13	-3.85*	-4.23*	-2.74	-3.96*
i	-2.74	-2.56	-2.45	-2.45	-3.80*	-3.64*	-3.24*	-3.66*
Δp	-1.56	-1.23	-1.60	-1.93	-4.95*	-4.82*	-2.86	-4.27*

NOTE: DF1(i): Augmented Dickey-Fuller test statistic with i lags, constant and trend included in the regression except for Δp which is not trending, used to test Null of I(1) against I(0), possibly with deterministic trend. DF2(i): idem for test of I(2) against I(1), only constant term in regression. An asterisk denotes asymptotic significance at the 5%-level.

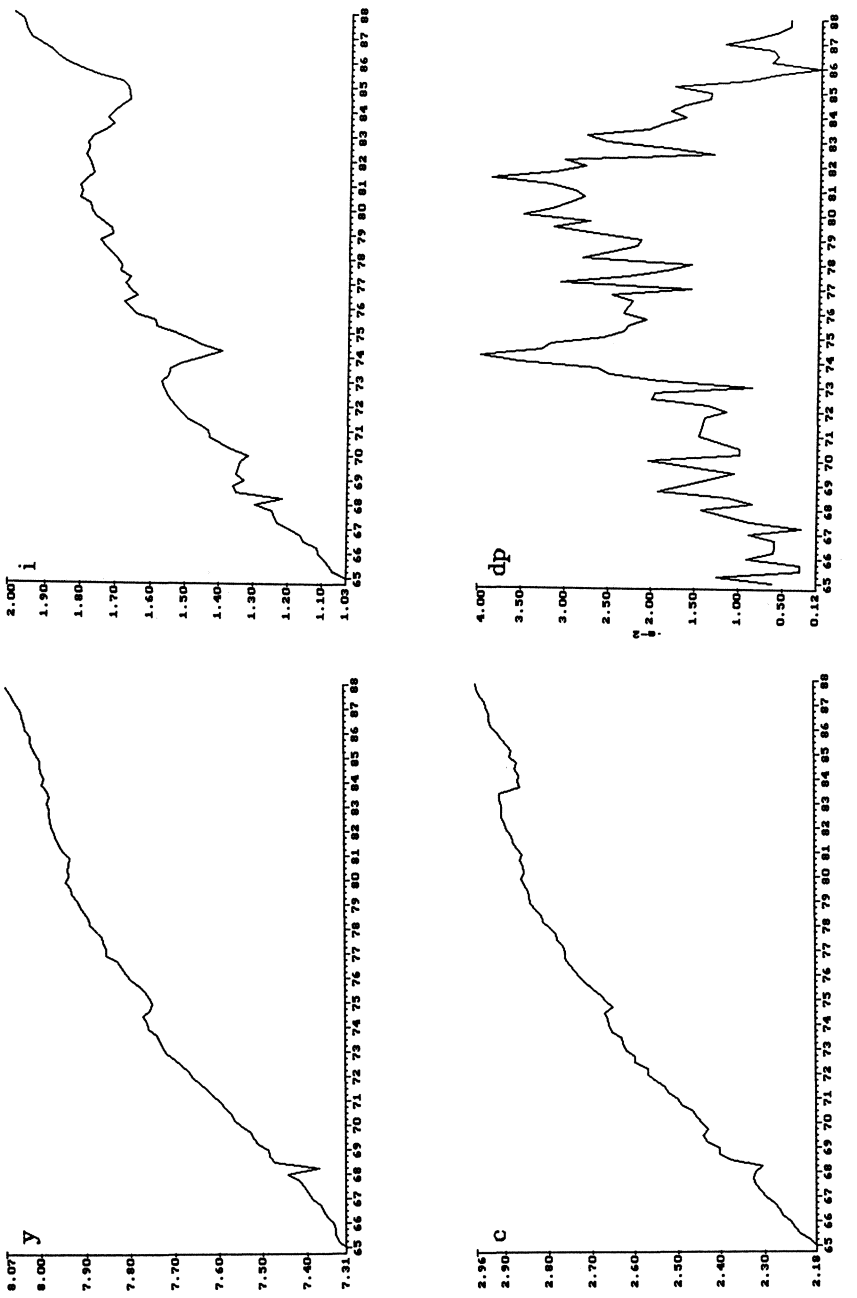

Figure 5.9 *Time series plots French macroeconomic data*
y: log GDP, *c*: log real consumption, *i*: log gross fixed capital formation, *dp*: consumer price inflation.

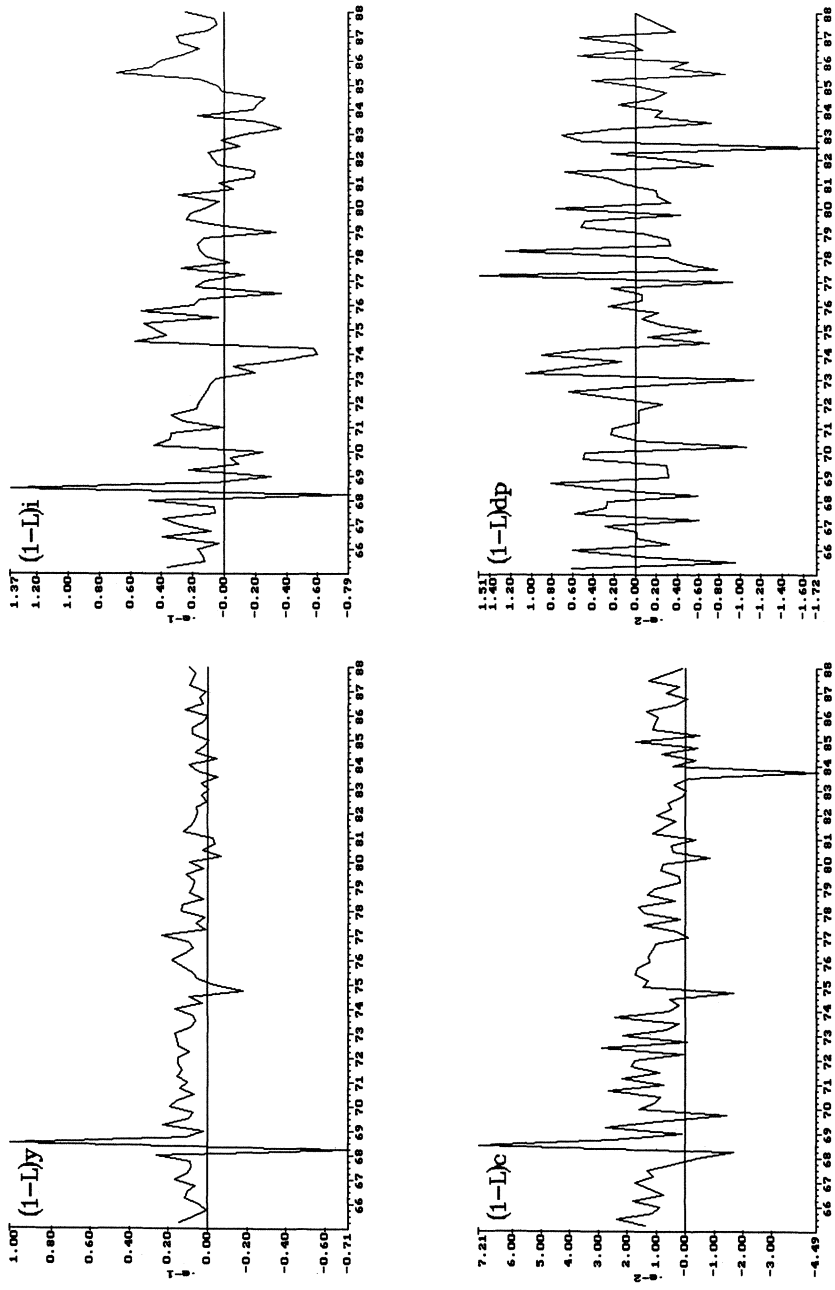

Figure 5.10 *Time series plots French macroeconomic data in first differences* (1-L)*y*: log GDP, (1-L)c: log real consumption, (1-L)*i*: log gross capital formation, (1-L)*dp*: consumer price inflation.

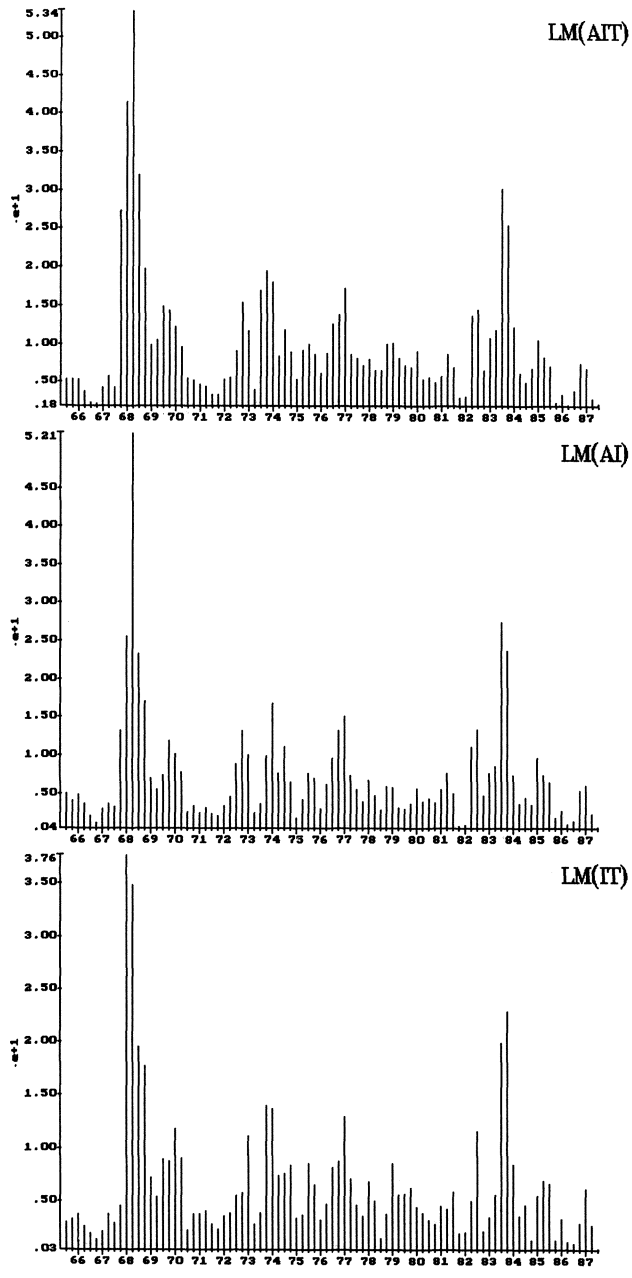

Figure 5.11a *System tests for French data*
Time series plots of outcomes of system tests; The simultaneous test detects outliers in 68.2 and 83.3. Statistics LM_{AIT}, LM_{AI} and LM_{IT}. 5% asymptotic critical values if timing known: 21.0, 15.5, and 15.5.

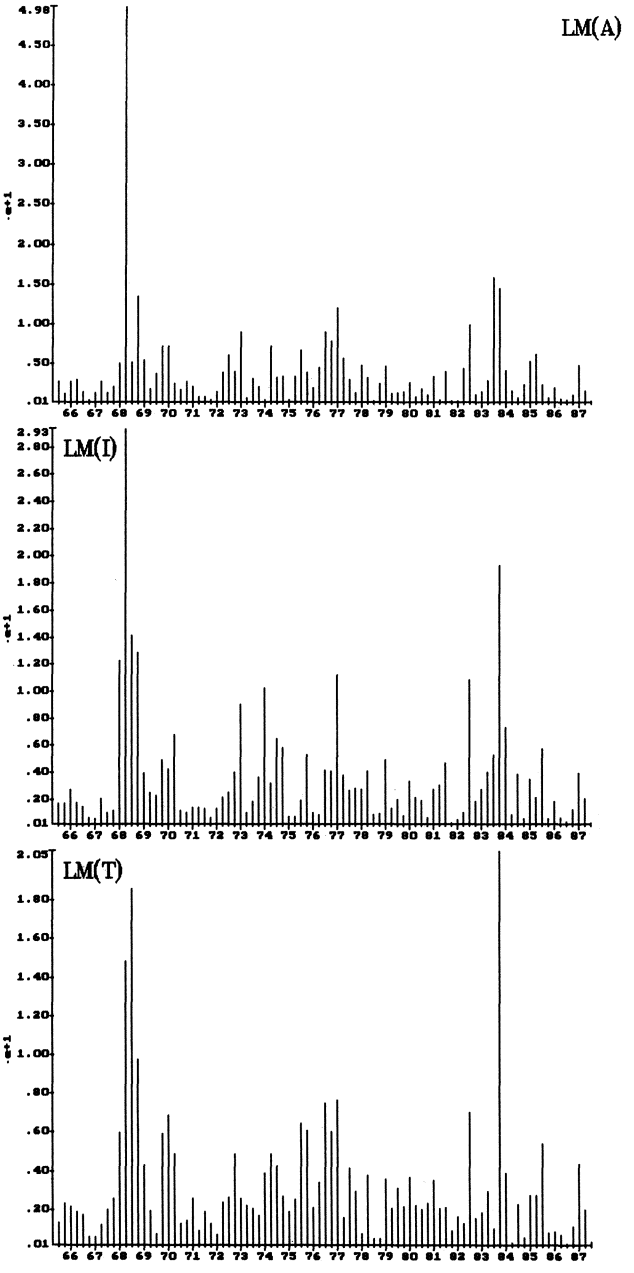

Figure 5.11b *System tests for French data (continued)*
Time series plots of outcomes of system tests; The specific tests indicate that the first overall outlier is an AO in 68.2 and the second one an IO or TO in 83.4. Statistics LM_A, LM_I, LM_T, 5% asymptotic critical values if timing known: 9.49.

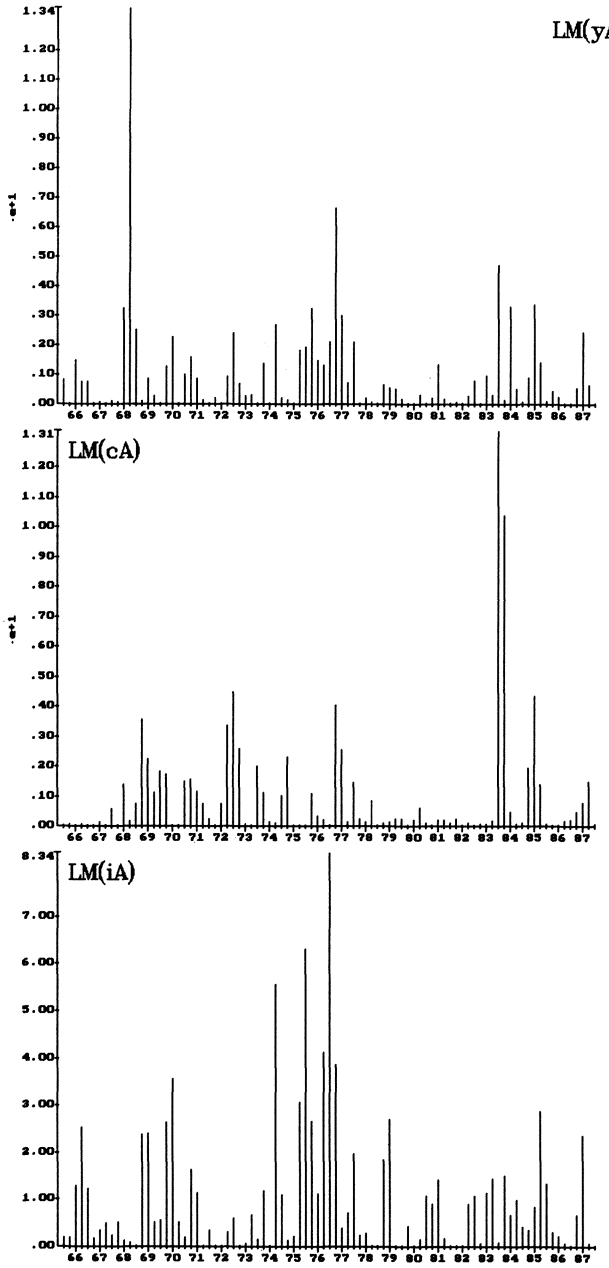

Figure 5.12a *Equation by equation tests for French data*
Time series plots of outcomes of specific equation tests. The AO of 68.2 had biggest impact on the y-equation. Statistics LM_{yA}, LM_{cA}, LM_{iA}. 5% asymptotic critical values if timing known: 3.84.

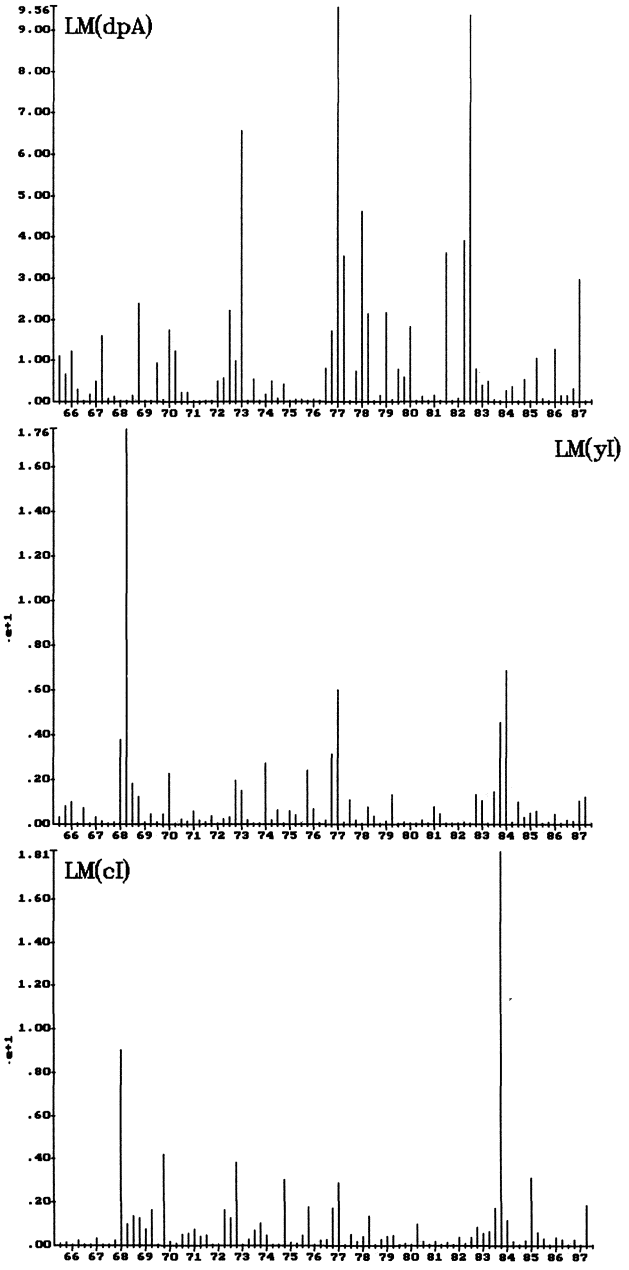

Figure 5.12b *Equation by equation tests for French data (continued)*
Time series plots of outcomes of specific equation tests. The IO of 83.4 hit c-equation worst. Statistics LM_{dpA}, LM_{yI} and LM_{cI}. 5% asymptotic critical values if timing known: 3.84.

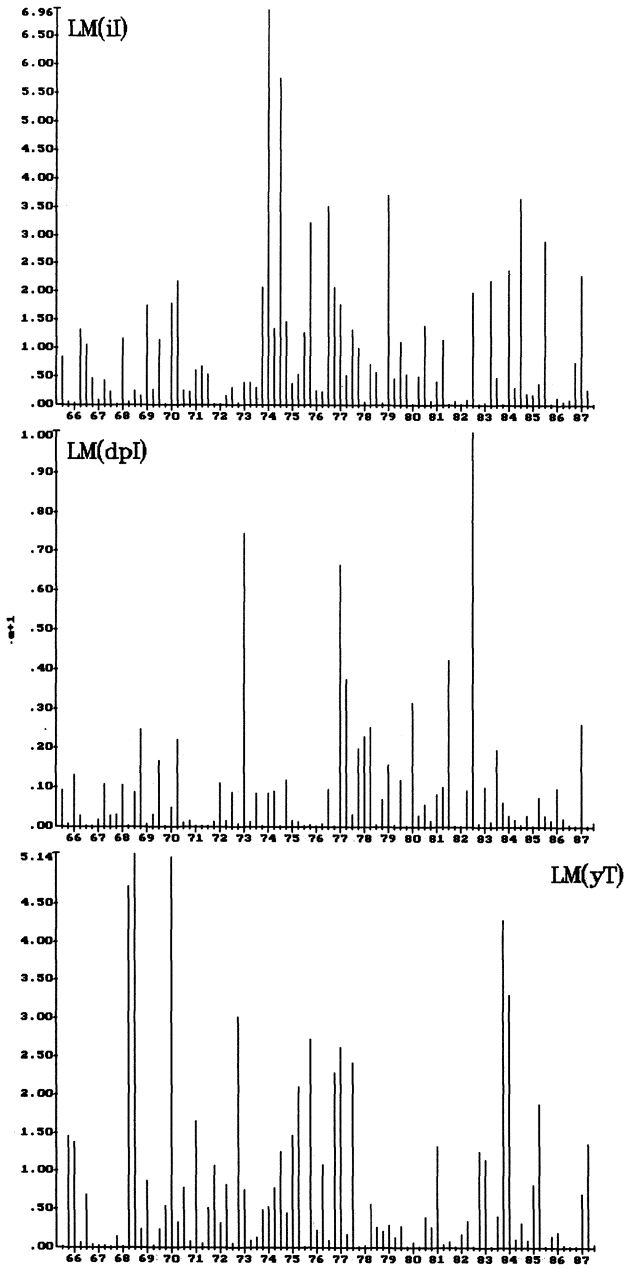

Figure 5.13a *Equation by equation tests for French data (continued)*
Time series plots of outcomes of specific equation tests. Statistics LM_{iI}, LM_{dpI}, LM_{yT}. 5% asymptotic critical values if timing known: 3.84.

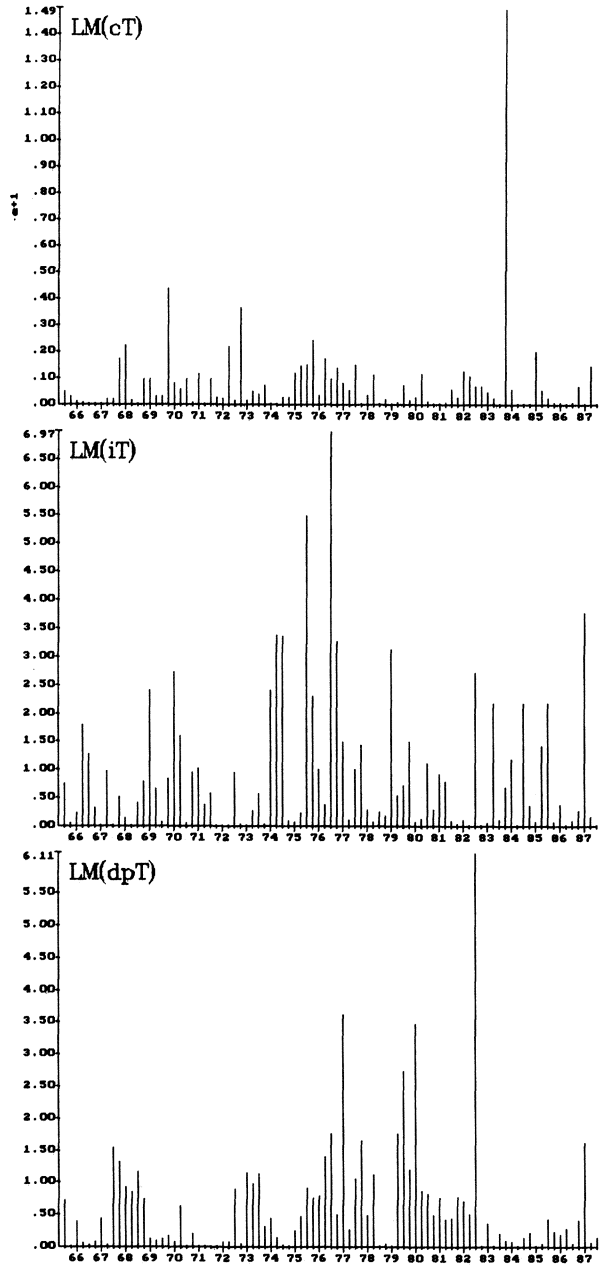

Figure 5.13b *Equation by equation tests for French data (continued)*
Time series plots of outcomes of specific equation tests. The TO of 83.4 hit c-equation worst. The dp-equation seems to be affected most seriously by a TO or IO in 82.3. Statistics LM_{cT}, LM_{iT} and LM_{dpT}. 5% asymptotic critical values if timing known: 3.84.

Table 5.3 *Results of estimation of VAR(2) system for French data* 1965.3–1988.1

	Estimated Model	
	Model (2.1) unrestricted	(2.1) with $m_1 = 0$ and $\text{rank}(\Phi(1)) = 2$
$\hat{\lambda}$	$\begin{bmatrix} 0.963 \\ 0.815 \\ 0.718 \\ .569+.132i \end{bmatrix}' \begin{bmatrix} .569-.132i \\ -0.483 \\ -0.125 \\ -0.075 \end{bmatrix}'$	$\begin{bmatrix} 1.000 \\ 1.000 \\ 0.853 \\ .572+.196i \end{bmatrix}' \begin{bmatrix} .572-.196i \\ -0.476 \\ -0.248 \\ -0.082 \end{bmatrix}'$
Norm_1:	543.249*	192.192*
Norm_2:	47.696*	87.135*
Norm_3:	0.908	2.456
Norm_4:	2.989	2.808
NormS:	86.680*	80.318*
NormK:	119.181*	120.689*
$\text{AC}(1)_1$:	2.378	2.334
$\text{AC}(1)_2$:	2.638*	2.409
$\text{AC}(1)_3$:	0.628	0.684
$\text{AC}(1)_4$:	1.386	1.941
$\text{AC}(1)_v$:	1.715	1.765*
$\text{ARCH}(1)_1$:	6.826*	7.111*
$\text{ARCH}(1)_2$:	11.060*	10.407*
$\text{ARCH}(1)_3$:	2.420	2.828*
$\text{ARCH}(1)_4$:	0.076	0.202
$\hat{D} = \det(\hat{\Sigma}) \cdot 10^{-16}$	1.017	1.093

$-2\ln(Q)$ s.t. $m_1 = 0$

$r \leq 3$	1.546
$r \leq 2$	8.165
$r \leq 1$	29.522
$r = 0$	55.843*

$\hat{\alpha}$: $\begin{bmatrix} 75.56 & -1.035 \\ -86.74 & 19.535 \\ 8.89 & -15.098 \\ 139.60 & -4.952 \end{bmatrix}$

$E(\Delta y) = \hat{C}(1)\hat{m}_0$: $\begin{bmatrix} 8.22 \\ 8.17 \\ 10.00 \\ 0.008 \end{bmatrix} \cdot 10^{-3}$ $\quad \hat{C}(1)$: $\begin{bmatrix} -0.036 & 0.043 & 0.624 & -0.181 \\ -0.484 & 0.279 & 0.844 & 0.769 \\ -0.543 & 0.315 & 1.009 & 0.836 \\ -0.246 & 0.130 & 0.122 & 0.522 \end{bmatrix}$

NOTES: See table 5.1 for an explanation of the upper part of this table. A constant was included in all the regressions. A trend term only in the unrestricted VAR(2), $-2\ln(Q)$ s.t. m_1=0: LR test on rank $\phi(1)$=r against full rank n, case *iii*, §A6.1. $\hat{\alpha}$: ML estimate of the first two cointegrating vectors in the system so that $\phi(1)=\gamma\hat{\alpha}'$ has rank 1; Johansen (1991a), §A6.1. $\hat{C}(1)$: sum of MA coefficients in Vector MA representation of Δy_t in restricted model; Johansen (1991a).

Outlier test results

The simultaneous test indicates two major outliers, one in 1968.2 (LM_{AIT}: 53.4) and one in 1983.3 (30.3). Using the criterion of the maximum outcome of the specific tests the first outlier can be interpreted as additive in 1968.2. It is most clearly indicated in the equation for y but from the plot of the first differences we see that the outlier occurred simultaneously in y, c, and i, which shows the possible gain for a multivariate approach of the analysis

of outlying observations. In this example the unrestricted VAR estimation gave rise to "significant" negative coefficients for Φ_{121} and Φ_{131} which may be due to the 1968.2 observation only and which may obscure the simultaneous outliers for consumption and investment in the testing procedure.

From the analysis of these data we may conclude either that an important observational error was made in 1968.2 for the three real variables (e.g. due to registration) or that the events in the Paris spring of 1968 only had a temporary effect on the French economy. Monfort and Rabemananjara (1990) presented a VAR model for the growth rates of wages and prices using quarterly series from 1964 to 1983 using a "structural equation methodology." They reduced the effect of the exceptional wage rise in 1968.2 and 1968.3 on the estimates of their VAR by introducing (IO model) dummy variables for those observations in the regression equations containing the wage variable.

A different story can be told for 1983. From the analysis of all equations together we conclude that we have to do with a TO in 1983.4. From the equation by equation testing and a glance at the series it seems clear that this second major outlier is due to an outlier with a (semi-)permanent effect on the consumption level, either an IO or a TO (which may be an indication of a level-shift). If c follows an independent random walk process, as predicted by some versions of the permanent income hypothesis, the difference between an innovation outlier and a one-time level-shift disappears. This decline in consumption was expected by the economic authorities, OECD (1984), because of a fairly sharp decline in disposable income, but one could not foresee that the impact would have such a lasting effect.

Of course one may also interpret this outlier as an observational error which has to be removed, because it has nothing to say about the effect of shocks on the level of consumption. In order to decide one has to have more detailed information on the data construction process. The test procedure does not use this information. A clue may be found in a comparison with OECD data (Main economic indicators, historical series code 14101511). Here the severe level-shift is not seen, which makes the measurement error hypothesis more likely.

Transient outliers are not important in the inflation equation. The inflationary shock in 1977.1 seems to be additive, whereas the exceptional decline in inflation in 1982.3 has a more permanent character. From these data we may conclude that the anti-inflationary policy by the French authorities in 1982 and 1983 was successful in having a long lasting effect. Monfort and Rabemananjara (1990) did not give 1982.3 a special treatment. The downward

shock occurred simultaneously in all their price series of interest, which probably helped them to identify the parameters in their simultaneous equation model. A more thorough historical analysis of these periods could reveal which measures or influences led to the temporary and which to the permanent shock. This could help in the evaluation of macroeconomic policy, e.g. in assessing the effectiveness of price regulations under indifferent regimes.

We stress that a first round analysis is treacherous, because of the impact that the outliers can have on the parameter estimates and diagnostic tests. This is illustrated by the outcomes of the analysis for shorter sample periods in tables 5.4, A5.1 and A5.2. The latter two tables are in the appendix. Note in table 5.2 that not only the number of lags but also the estimation period matters, and thus maybe the outlier, for the unit root analysis, especially for y.

In table 5.4 the period with the additive outlier in y is excluded. The outlier seems to matter for the cointegration analysis. The LR test now indicates that there are three instead of two unit roots in the system, cf. tables 5.3 and 5.4. The mean growth in the later period has gone down, but the ratios remain stable. Note also the huge influence upon the estimated determinant of the covariance matrix of the errors. The diagnostics come out better. Normality for the y–equation is restored, and (vector) ARCH effects are removed from the c–equation. The indication of ARCH effects in the y–equation remains.

By comparing tables 5.3 and A5.1, which leaves out the last five years of the sample period one can assess the impact of the more persistent outlier in c. The period of the last five years including the outlier is decisive for the unit root analysis. Without the information of the last five years one would prefer the stationary alternative with only one unit root. So here we seem to be dealing with a multivariate analog of the findings of Perron (1989), where the original indication of a unit root is based on the information contained in a small proportion of the data set. The estimate of the first "cointegrating vector" and the vector of mean growths still show the same pattern however. The last five years of observations do not change the outcome of the usual diagnostics in a decisive manner. The nonnormality of the error term for c is mitigated and ARCH effects in the investment equation become visible.

Table 5.4 *Results of estimation of VAR(2) system for French data* 1970.3–1988.1

	Estimated Model	
	Model (2.1) unrestricted	(2.1) with $m_1 = 0$ and $\text{rank}(\Phi(1)) = 2$
$\hat{\lambda}$	$\left[\begin{bmatrix} 0.917 \\ .735+.190i \\ .735-.190i \\ .622+.047i \end{bmatrix}' : \begin{bmatrix} .622-.047i \\ -0.226 \\ -.051+.089 \\ -.051-.089 \end{bmatrix}'\right]$	$\left[\begin{bmatrix} 1.000 \\ 1.000 \\ 0.827 \\ .585+.116i \end{bmatrix}' : \begin{bmatrix} .585-.116i \\ -0.271 \\ -0.120 \\ -0.026 \end{bmatrix}'\right]$
Norm_1:	0.505	0.214
Norm_2:	83.996*	105.023*
Norm_3:	1.627	11.053*
Norm_4:	4.278	2.594
NormS:	68.694*	77.678*
NormK:	19.201*	34.238*
$\text{AC}(1)_1$:	0.609	0.347
$\text{AC}(1)_2$:	1.874	1.828
$\text{AC}(1)_3$:	1.115	0.617
$\text{AC}(1)_4$:	0.686	1.181
$\text{AC}(1)_v$:	1.386	1.513
$\text{ARCH}(1)_1$:	7.152*	6.540*
$\text{ARCH}(1)_2$:	1.766	1.342
$\text{ARCH}(1)_3$:	0.516	0.107
$\text{ARCH}(1)_4$:	0.063	0.308
$\hat{D} = \det(\hat{\Sigma}) \cdot 10^{-16}$	0.224	0.241
$-2\ln(Q)$ s.t. $m_1 = 0$		
$r \leq 3$	0.920	
$r \leq 2$	7.244	
$r \leq 1$	22.348	
$r = 0$	57.777*	
$\hat{\alpha}$:	$\begin{bmatrix} 41.05 & -68.167 \\ -31.76 & 89.131 \\ -3.52 & -17.964 \\ 112.52 & -140.363 \end{bmatrix}$	
$E(\Delta y) = \hat{C}(1)\hat{m}_0$:	$\begin{bmatrix} 6.69 \\ 6.61 \\ 9.56 \\ -0.27 \end{bmatrix} \cdot 10^{-3}$	
$\hat{C}(1)$:	$\begin{bmatrix} 0.256 & -0.024 & 0.464 & -0.798 \\ -0.372 & 0.284 & 0.754 & 0.097 \\ -1.017 & 0.647 & 1.318 & 0.819 \\ -0.230 & 0.109 & 0.085 & 0.344 \end{bmatrix}$	

NOTE: See table 5.3 for an explanation of the symbols.

Table A5.2 shows the results for the shortest sample without the influence of the biggest outliers. The cointegration analysis has qualitatively the same results as in table A5.1. At the 10%–level there are no unit roots, at the 5% there are three. The number of regressors is large compared to the number of observations so that a degrees of freedom problem arises here. Now the abnormality of the errors in the investment and inflation equation in the oil price shock period become apparent. The large coefficient of inflation from the first estimated cointegrating vector is considerably smaller. The relatively big expectation for investment growth of the other periods is not seen as clearly in this subperiod.

The deletion of subperiods with outliers to reduce their influence is of course a very simple and all too radical solution. Either careful modelling of the outliers (intervention analysis in the case of innovation outliers) or the development of robust estimation methods should present more efficient ways to extract the relevant information from the data, without being misled by these exceptional observations.

Summary of application results

The example suggests that the analysis of the exceptional observations can give valuable insight in the way different kinds of measures and circumstances influence the variables in the system, especially about the question whether they have a long lasting effect or not. A sensitivity analysis on the outcomes for the cointegration tests with respect to the sample period makes clear that relatively short periods with devious observations can change statistical inference on the null hypothesis on a specific number of cointegration relations in both directions. The information about these observations from the test procedure can be valuable.

Monte Carlo Analysis for different parameter structures as well as application in other macroeconomic VAR systems can be performed to get a better insight in the merits and demerits of the method. The modeling of the outliers or the development of robust estimation methods for multivariate time series, along one of the lines suggested in the introduction is the next step in our analysis.

5.7 Two simple ways to study the influence of outliers

In econometrics the best known way to model outlying observations is to assume a sudden increase in the variance of the error term. This amounts to adding a dummy variable for each outlying observation if the supposed variance tends to infinity or to downweighting of the observations if the supposed variance is finite. As the single series of our model should be modeled as ARMA models where the order of the autoregressive part is at least 20, the addition of dummy variables or a downweighting scheme in such a way that the dependence of the observations is taken into account is not an attractive option. From a time series point of view the introduction of a dummy variable can be seen as the explicit modelling of an innovation outlier. The subsequent observations are not corrected since they are assumed to be generated by the model. We applied this common technique.

It is clear that the traditional methods are difficult to apply in multivariate dynamic models, cf. also §3.2.3. So what can we do with more recent methods? A nice view on the state of the art in the statistical

literature can be found in Kitagawa (1987) and the comments following this article. As we are primarily interested in the application of these techniques to multiple economic time series, it is important to note some differences between the "statistical literature series" and our economic time series. The outliers in the statistical series are often far more outspoken (e.g. the earthquake signal in Kitagawa (1987) and the ionosphere capacity in Kleiner et al. (1979) than those in our series. Furthermore outliers in macroeconomic time series can often be explained by known sample specific exogenous events both within an extended model (the ideal situation, when the supposed univariate outlier isn't an outlier after all) or outside the model.

The method outlined in appendix A5.3 tries to take these characteristics into account as much as possible. In order to estimate the parameters of interest as efficiently as possible it is important to use as much of the information we have in the data series. Therefore we do not discard entire observations if they cannot be explained fully by our model. Only parts of the observations that cannot be explained, and which often turn out to be rather small, are removed.

The method also provides us with estimates of the additive outlier component, so that we can see what information we removed. This is not the case in many other "robust" estimation methods. As with other methods that explicitly take account of outliers (Tsay (1988)) one has to choose (a series of) *tuning constants* (critical values, that imply a "degree" of cleaning). A limitation of our outlier model is the absence of an explicit distributional assumption for the exceptional observations. It is however doubtful whether this would be of great help for our goal of testing hypotheses about Φ. We applied our filter for example to the series C of Box and Jenkins (1976) and obtained virtually the same estimates of Φ and its standard error as Tsay (1987, §6) with the Box–Tiao (1975) method, without the disadvantage of having to identify the outlier a priori.

It has long been recognized (see e.g. Rosenblatt (1968)) that outliers can play an important disturbing role when the estimation of the spectrum of a time series is concerned. Our method was originally designed to reduce this role effectively, see Kleiner et al. (1979).

Results for French data

By comparing the outcomes reported in tables 5.5 and 5.6 one can detect the differences in the effects of the two approaches. Table 5.5 gives the results for the additive outlier correction procedure described in its most simple form with tuning constant 2. The procedure is discussed in more detail in the appendix A5.3. Table 5.6 gives the results for the extended model with

15 "significant" dummies. The dummies are added to each equation in order to avoid the influence of a priori ideas in the estimation of the residual covariance matrix. The casual introduction of dummies is a doubtful procedure as far as the subsequent testing of the model is concerned. It results in subsequent measures of fit of a dynamic model which are hard to interpret. We know that the introduction of dummy variables in a static regression model boils down to the actual deletion of the affected observations from the sample. In a static model one could interpret the procedure as a correction for observation errors, see the remarks to condition 5.1 in §5.3. This interpretation does not go through for dynamic models.

Table 5.5 *Results for VAR(2) French data* 1965.3–1988.1 *with AO correction*

	Estimated Model	
	Model (2.1) unrestricted	(2.1) with $m_1 = 0$ and $\text{rank}(\Phi(1)) = 2$
$\hat{\lambda}$	$\left\{\begin{bmatrix}0.935\\0.897\\0.744\\.538+.230i\end{bmatrix}' : \begin{bmatrix}.538-.230i\\-0.214\\-0.080\\-0.048\end{bmatrix}'\right\}$	$\left\{\begin{bmatrix}1.000\\1.000\\0.893\\.560+.271i\end{bmatrix}' : \begin{bmatrix}.560-.271i\\-0.220\\-0.171\\0.050\end{bmatrix}'\right\}$
Norm_1:	.189	.14
Norm_2:	2.410	1.65
Norm_3:	.273	.06
Norm_4:	1.126	.80
NormS:	38.950*	40.29*
NormK:	.597	1.55
$\text{AC}(1)_1$:	.246	.09
$\text{AC}(1)_2$:	1.356	1.28
$\text{AC}(1)_3$:	0.341	.05
$\text{AC}(1)_4$:	1.909	2.02
$\text{AC}(1)_v$:	1.055	1.10
$\text{ARCH}(1)_1$:	3.072*	2.40
$\text{ARCH}(1)_2$:	0.748	0.91
$\text{ARCH}(1)_3$:	4.337*	2.92*
$\text{ARCH}(1)_4$:	0.825	0.05
$\hat{D} = \det(\hat{\Sigma})\cdot 10^{-18}$	9.47	12.60

$-2\ln(Q)$ s.t. $m_1 = 0$

$r \le 3$	0.80
$r \le 2$	6.99
$r \le 1$	33.14*
$r = 0$	71.02*

$\hat{\alpha}$: $\begin{bmatrix}-15.66 & 74.84\\ -0.83 & -88.25\\ 13.33 & 10.57\\ -36.07 & 131.60\end{bmatrix}$

$E(\Delta y) = \hat{C}(1)\hat{m}_0$: $\begin{bmatrix}8.1\\8.0\\9.9\\-0.04\end{bmatrix}\cdot 10^{-3}$

$\hat{C}(1)$: $\begin{bmatrix}0.486 & -.460 & 0.655 & -0.384\\ 0.080 & -.334 & 0.879 & 0.634\\ -0.022 & -.377 & 1.159 & 1.093\\ -0.221 & 0.068 & 0.124 & 0.556\end{bmatrix}$

NOTE: See table 5.3 for an explanation of the symbols. Correction following Kleiner et al. (1979), with tuning constant: 2.00, see appendix A5.3.

Table 5.6 *Results for VAR(2) French data* 1965.3–1988.1 *with IO correction*

	Estimated Model	
	Model (2.1) unrestricted	(2.1) with $m_1 = 0$ and $\text{rank}(\Phi(1)) = 2$
$\hat{\lambda}$	$\left[\begin{bmatrix} .973+.022i \\ .973-.022i \\ 0.779 \\ .512+.158i \end{bmatrix}' : \begin{bmatrix} .512-.158i \\ -0.127 \\ -0.106 \\ 0.026 \end{bmatrix}'\right]$	$\left[\begin{bmatrix} 1.000 \\ 1.000 \\ 0.947 \\ .566+.225i \end{bmatrix}' : \begin{bmatrix} .566-.225i \\ -0.193 \\ -0.136 \\ -0.032 \end{bmatrix}'\right]$
Norm_1:	.52	.99
Norm_2:	1.03	1.42
Norm_3:	.99	3.78
Norm_4:	.45	1.86
NormS:	19.72	20.38
NormK:	4.95*	7.61*
$\text{AC}(1)_1$:	.30	.23
$\text{AC}(1)_2$:	2.78	1.88
$\text{AC}(1)_3$:	1.06	.99
$\text{AC}(1)_4$:	1.81	.87
$\text{AC}(1)_v$:	1.84*	1.13
$\text{ARCH}(1)_1$:	.48	.33
$\text{ARCH}(1)_2$:	1.31	1.17
$\text{ARCH}(1)_3$:	0.22	.34
$\text{ARCH}(1)_4$:	0.94	0.66
$\hat{D} = \det(\hat{\Sigma}) \cdot 10^{-18}$	1.24	1.64

$-2\ln(Q)$ s.t. $m_1 = 0$

$r \leq 3$	0.35
$r \leq 2$	6.72
$r \leq 1$	35.05*
$r = 0$	66.43*

$\hat{\alpha}$: $\begin{bmatrix} 70.05 & 13.21 \\ -75.57 & 3.90 \\ 2.67 & -11.84 \\ 149.14 & -8.27 \end{bmatrix}$

$E(\Delta y) = \hat{C}(1)\hat{m}_0$: $\begin{bmatrix} 7.2 \\ 7.9 \\ 10.4 \\ -0.4 \end{bmatrix} \cdot 10^{-3}$ $\quad \hat{C}(1)$: $\begin{bmatrix} .365 & -.219 & .680 & .731 \\ -.292 & .243 & .896 & 1.670 \\ .541 & -.327 & .972 & 1.027 \\ -.329 & .232 & .117 & .484 \end{bmatrix}$

Dummies significant according to single equation outlier test for:

	y	c	i	dp		y	c	i	dp
68.1	++	--	+		73.1				--
68.2	--	--	--		74.1	+		--	++
68.3	++	++	++		74.4	--	--		
69.3				-	77.1	++			--
69.4		--			82.3				--
70.1					83.4		--		
72.3		++		+	85.1		++		
72.4		-							

NOTE: See table 5.3 for an explanation of the symbols. Dummies introduced for nominally significant values of PF_i. ++/-- denotes significance at 5% level, +/- at 10%

Both approaches are successful in the removal of abnormality of the residuals of the separate equations. The additive outlier correction leaves multivariate skewness in the residuals, the innovation outlier model leaves multivariate kurtosis. Introduction of either outlier model has about the same effect on the dynamic parameters $\hat{\lambda}$ and $\hat{\alpha}$, representing the characteristic roots and the estimates of a basis of the space of cointegrating vectors.

Due to the level-shift correction in c with the IO model, the log of the "greater ratio" of consumption over GDP (see King et al. (1991)), C/Y, extended with an inflation effect, associated with the greater canonical correlation between growth rates and levels, comes out more clearly as the most "error correcting" vector (cf. test for $r=0$ and first column of $\hat{\alpha}$ in tables 5.3 and 5.6).

The estimates of the mean growth rates $\hat{C}(1)\hat{m}_0$ are clearly affected by the extension with innovation outliers which results in a "structural" downward trend in French inflation. Innovation outliers are not constrained to have mean zero so that the sample mean of the IO corrected data may deviate substantially form the sample mean of the original series.

The most striking difference between the additive and innovative outlier correction procedure is in the estimates of the residual covariance matrix. Its determinant has decreased by a factor 10 using the AO model, but by a factor 100 using the IO model. So one can expect considerable effects on subsequent impulse response analyses.

The downward bias of the IO model is easily explained. The dummy model effectively puts each observed large residual to zero, and produces a so-called *trimmed estimate*, see e.g. Amemiya (1985, p.71). The Kleiner-Martin implementation of the AO model puts large residuals to a predefined truncation point, and produces a *Winsorized estimate* (op cit, p. 73). One could modify the implementation of the IO model so as to produce Winsorized estimates as well. One would have to implement by correcting the data for the effect of the innovation outliers down to a certain truncation point. For a univariate model one could replace y_t^* by $\hat{y}_t = y_t - \Phi(L)^{-1}\delta_{l,t}\hat{O}_2$, with $\hat{O}_2 = \mathrm{sgn}(e_l)\cdot(|e_l| - c\tilde{\sigma})$, and $\tilde{\sigma}$ a robust estimate of the standard error, and c a the preferred *tuning constant*. It seems unattractive to apply such a method in multivariate models with unit roots. Since most innovation outliers have permanent effects in such a model, one would end up with a "corrected" series with a rather vague relation with the original series, which we do not want here. A last advantage of the AO correction is that it can be used for the "graduation of extreme values" which is a necessary prerequisite for a proper seasonal adjustment by linear filters.

All in all we prefer the additive outlier correction procedure for our unrestricted VAR analysis because it gives bigger weight to the data information relative to prior information in the subsequent results than does the dummy method. It allows one to have a "fresh" look at the data. One does not have to experiment with all kinds of dummy models, probably suggested by others, for a purpose different from one's own.

The LM test developed in this chapter can be used to evaluate the results of the procedure. To conclude this chapter on outliers we present figure 5.14 which points out how big the effect of the dummies in the explanation of changes of the variable of interest can be. The exact dates of the dummies are given in table 5.6. For comparison we have plotted the one period effect of the (significant) error correction terms in the same time series plots.

It is clear that a misspecification of the outlier component of the series may have a larger effect, than the improper treatment of error correction components.

Interpretation of results

From the sensitivity analysis as a whole we may conclude that the finding of two cointegrating vectors is quite robust. One cointegrating vector can be detected in subsamples as well. Another confirmation of the existence of a long run equilibrium relation can be found in the analysis of the long run autocorrelations presented in §7.5.2. The cointegrating vectors can be interpreted as linear combinations of the log balanced growth "greater ratios" C/Y and I/Y (King et al. (1991)), adjusted for a long run inflation effect. The innovations in the two *common stochastic trends* $C(1)(1-L)^{-1}e_t$ change the forecasts of "trend productivity" at all future dates. These innovations were considered constant in the steady state situation of the basic neoclassical model with deterministic trend growth in productivity (i.e. due to a constant growth rate in technology and labor supply). In modern stochastic real business cycle models where productivity is assumed to follow a random walk with drift these productivity innovations are assumed to behave as the outcomes of a stationary stochastic process. See §6.2.2 for a technical discussion of the common trends model and §6.3.4 for a specific example of the empirical implementation of an economic model along these lines.

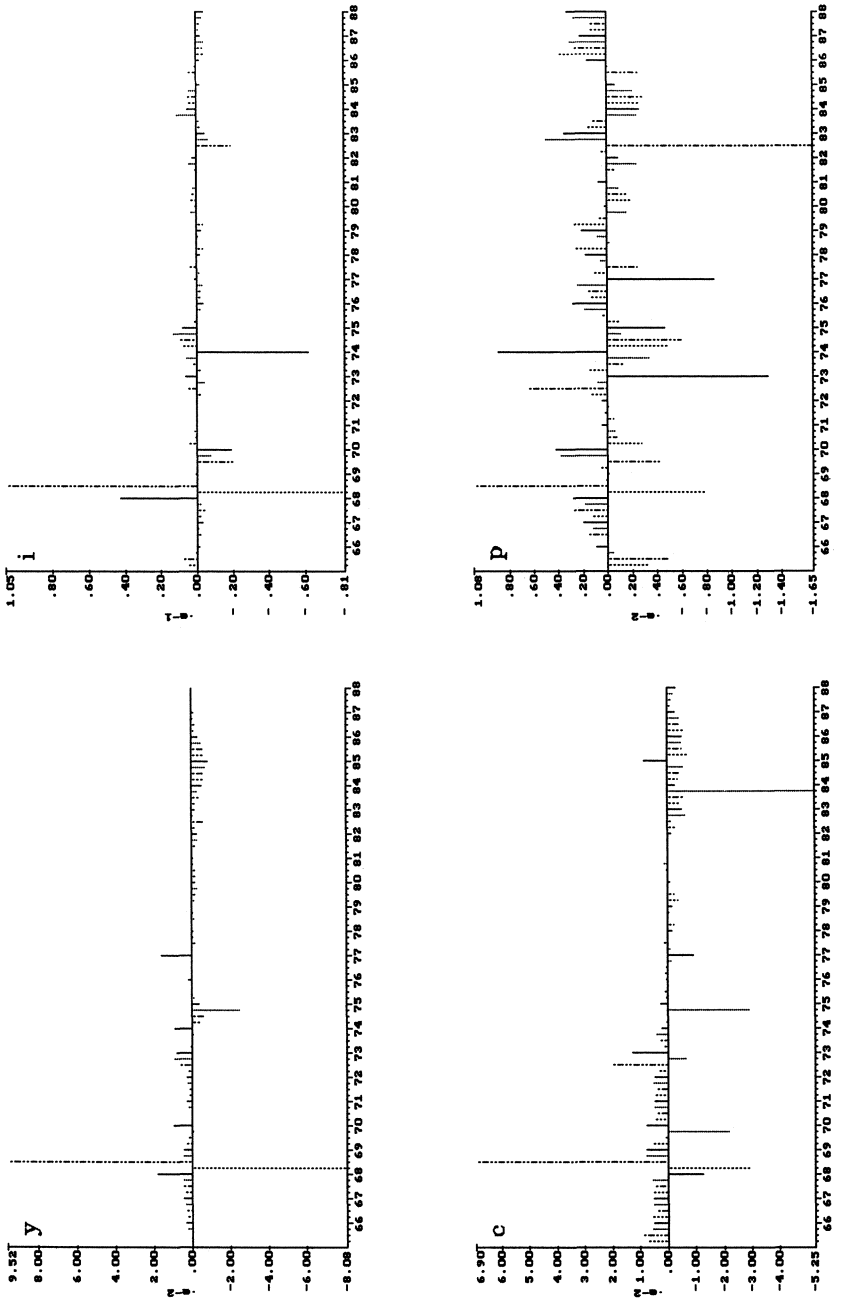

Figure 5.14 *Time series plots of the immediate impact dummy variables defined in table* 5.6 *plus the impact of the error correction terms lagged one period on the changes in French macroeconomic variables* y: log GDP, c: log real consumption, *i*: log gross capital formation, *p*: consumer price inflation.

In the French data we find a positive long run partial correlation of inflation with the first greater ratio and a negative long run partial correlation with the second. Consider inflation is the more exogenous variable, because it is mainly affected by a combination of foreign price shocks and domestic shocks from monetary policy (including exchange rate policy), which have their effect on the other variables mainly through changes in prices. The estimates indicate that consumption is not as badly affected by rising inflation as investment. The estimates of the size of the long run inflation effect differ widely across the different subperiods and outlier correction methods, cf. the last rows of $\hat{\alpha}$ in tables 5.3, 5.4, A5.1 and A5.2. The outlier model plays a crucial role in the assessment of the results on this effect.

Reimers (1991a) estimated a comparable model for Germany. Mellander et al. (1992) provided an elaboration of this model in an open economy context. Their model predicted a long run influence of the terms of trade on the first greater ratio (a rising share of consumption in income with a rising terms of trade). We applied their ideas in our application on Dutch data in chapter 7 and did not find such a price effect.

A5.1 Some proofs concerning outlier test statistics

In this appendix we present two proofs concerning outlier test statistics in §5.3 and §5.4. We also discuss some alternative tests.

A5.1.1 Proofs for derivation simultaneous test

Proposition[1]

The inverse of the following matrix,

$$\begin{pmatrix} \Sigma^{-1}+W & -\Sigma^{-1} & \Sigma^{-1}+D' \\ -\Sigma^{-1} & \Sigma^{-1} & -\Sigma^{-1} \\ \Sigma^{-1}+D & -\Sigma^{-1} & \Sigma^{-1}+C \end{pmatrix}^{-1} = \left(Z^{ij}\right) = Z, \qquad \text{(A5.1.1)}$$

with

$Z^{11} = (W - D'C^{-1}D)^{-1}$, $Z^{21} = Z^{12'} = (I - C^{-1}D)Z^{11}$,

$Z^{22} = \Sigma + C^{-1} + (I - C^{-1}D)Z^{11}(I - C^{-1}D)'$, $Z^{33} = (C - DW^{-1}D')^{-1}$,

$Z^{13} = Z^{31'} = -W^{-1}D'Z^{33}$, $Z^{23} = (I - W^{-1}D')Z^{33}$,

provided all the relevant inverses exist, which we assume in our regularity conditions. The only formulas we need are those for inversion of partitioned matrices and the matrix inversion lemma, see e.g. Magnus and Neudecker (1988, p. 11).

Proof.

Introduce auxiliary variables Y, X, and H as

$$Y = \begin{pmatrix} \Sigma^{-1}+W & -\Sigma^{-1} \\ -\Sigma^{-1} & \Sigma^{-1} \end{pmatrix}, \quad X = \begin{pmatrix} \Sigma^{-1}+D' \\ -\Sigma^{-1} \end{pmatrix}, \quad H = \Sigma^{-1}+C.$$

Do the first stage of the partitioned inversion:

$$\begin{pmatrix} Z^{11} & Z^{12} \\ Z^{21} & Z^{22} \end{pmatrix} = Y^{-1} + Y^{-1}X\left(H - X'Y^{-1}X\right)^{-1}X'Y^{-1}, \qquad \text{(A5.1.2)}$$

$$Y^{-1} = \begin{pmatrix} W^{-1} & W^{-1} \\ W^{-1} & \left(\Sigma^{-1} - \Sigma^{-1}(\Sigma^{-1}+W)\Sigma^{-1}\right)^{-1} \end{pmatrix} = \begin{pmatrix} W^{-1} & W^{-1} \\ W^{-1} & \Sigma + W^{-1} \end{pmatrix},$$

$$Y^{-1}X = \begin{pmatrix} W^{-1}\Sigma^{-1} + W^{-1}D' - W^{-1}\Sigma^{-1} \\ W^{-1}D' - I \end{pmatrix} = \begin{pmatrix} W^{-1}D' \\ W^{-1}D' - I \end{pmatrix}. \qquad \text{(A5.1.4)}$$

---

[1] See (5.6).

The last equality can be seen to hold by noting that

$$(\Sigma^{-1}+\Sigma^{-1}W^{-1}\Sigma^{-1})^{-1}=\Sigma-(\Sigma^{-1}+W)^{-1} \tag{A5.1.5}$$

and substituting the last term of the right hand side.

$$(Z^{33})^{-1}=H-X'Y^{-1}X=(C-DW^{-1}D'), \tag{A5.1.6}$$

$$Y^{-1}X(H-X'Y^{-1}X)^{-1}X'Y^{-1}=\begin{bmatrix} W^{-1}D' \\ W^{-1}D'-I \end{bmatrix}(C-DW^{-1}D')^{-1}\begin{bmatrix} DW^{-1} & DW^{-1}-I \end{bmatrix}. \tag{A5.1.7}$$

It follows that

$$Z^{11}=W^{-1}+W^{-1}D'(C-DW^{-1}D')^{-1}DW^{-1}=(W-D'C^{-1}D)^{-1}, \tag{A5.1.8}$$

which can also be seen using a symmetry argument applied to W and C.
The following elements are now easy to derive:

$$Z^{21}=Z^{11}-(C-DW^{-1}D')^{-1}DW^{-1}. \tag{A5.1.9}$$

Subtracting and adding $C^{-1}DZ^{11}$ and working out (A5.1.8) this is seen to be:

$$\begin{aligned} Z^{21}&=Z^{11}-C^{-1}D(W^{-1}+W^{-1}D'(C-DW^{-1}D')^{-1}DW^{-1}) \\ &\quad +C^{-1}DW^{-1}+(C^{-1}DW^{-1}D'-I)(C-DW^{-1}D')^{-1}DW^{-1} \\ &=Z^{11}-C^{-1}DZ^{11}+C^{-1}(I-(C-DW^{-1}D')(C-DW^{-1}D')^{-1})DW^{-1}=(I-C^{-1}D)Z^{11}. \end{aligned} \tag{A5.1.10}$$

$$Z^{22}=\Sigma+W^{-1}+(I-W^{-1}D')Z^{33}(I-DW^{-1}). \tag{A5.1.11}$$

$$\begin{bmatrix} Z^{13} \\ Z^{23} \end{bmatrix}=-Y^{-1}X(H-X'Y^{-1}X)^{-1}=-\begin{bmatrix} W^{-1}D' \\ W^{-1}D'-I \end{bmatrix}Z^{33}. \tag{A5.1.12}$$

Using symmetry arguments the expressions for the other elements of Z follow immediately. Z^{22} can also be written as $\Sigma+C^{-1}+(I-C^{-1}D)Z^{11}(I-C^{-1}D)'$. The formulas for Z^{13} and Z^{21} can also be derived from each other.

Proof of (5.7).
It is straightforward to show that (5.6) is equal to :

$$\begin{aligned} &(a_1'+a_2')Z^{11}(a_1+a_2)+(a_3'+a_2')Z^{33}(a_3+a_2) \\ &-(a_2'C^{-1}DZ^{11}+a_3'Z^{33}DW^{-1})(a_1+a_2)-(a_1'+a_2')(Z^{11}D'C^{-1}a_2+W^{-1}D'Z^{33}a_3) \\ &+a_2'(\Sigma+C^{-1}-Z^{33}+C^{-1}DZ^{11}D'C^{-1})a_2. \end{aligned} \tag{A5.1.13}$$

One can see that $C^{-1}DZ^{11}-Z^{33}DW^{-1}=0$ from equations (A5.1.9) an (A5.1.10), substituting this in (A5.1.13) and applying the matrix inversion lemma to Z^{33} in the last term gives (5.7). □

A5.1.2 Finite sample alternatives for I test procedure

This section discusses some alternative procedures for the test for a multivariate innovation outlier with known timing as in (5.10).

The Lawley–Hotelling test statistic L_l for discordancy of observation l can be written as (see remark A3.9, $m=1$, R_l is $(1\times n)$, I_m is (1×1)):

$$L_l = (I_m - H_l)^{-1} R_l \hat{\Sigma}_{(l)}^{-1} R_l'. \tag{A5.1.14}$$

By using the consistency of the OLS estimator under H_0 one obtains

$$R_l \to_{\mathcal{L}} e_l', \quad \operatorname*{Plim}_{T\to\infty} \hat{\Sigma}_{(l)}^{-1} = \Sigma^{-1},$$

and by combining the definitions from §A3.2 for H_l and from §5.4 for $X(O)$ (X' in §A3.2 $= X(O)$ in (5.10)) one finds

$$\operatorname*{Plim}_{T\to\infty}(I_m - H_l) = I_m = 1 \text{ and } LM_I = e_l' \Sigma^{-1} e_l,$$

so that

$$L_l \to_{\mathcal{L}} LM_I. \tag{A5.1.15}$$

A simple alternative for using L_l is the small sample correction of Anderson (1958, Theorem 8.6.2) of the likelihood ratio test:

$$P(T^{-1}[T - k - \tfrac{1}{2}(n - k_0 + 1)]\cdot\xi_T \le \lambda]) = P(\chi^2(nk_0) \le \lambda) + O(T^{-2}),$$

where ξ_T is the usual likelihood ratio statistic for testing a linear subhypothesis in a linear system of n equations; T is the sample size, k is the number of explanatory variables per equation in the unconstrained model; and k_0 the number of constraints per equation. In the example of §5.5 one would have $T=100$, $n=2$, $k=2\cdot 2+1+1=8$ (one for the constant and one for the dummy), and $k_0=1$.

Another option is to use the "cautious" system F–test suggested by Theil (1971, p. 402) to test the significance of the extra dummies. In Monte Carlo experiments we found this procedure to work comparatively well. We applied it in a from general to specific testing procedure to determine the order of a VAR. We used the set–up from Lütkepohl (1985) for this comparative test. We use this F–test in the variable addition version of the multivariate test for first to fourth order vector ARMA processes in the residuals, see also §3.3.4.

A5.2 Subsample analysis outlier influence

Table A5.1 *Results of estimation of VAR system for French data* 1965.3–1983.1

Estimated Model		
	Model (2.1) unrestricted	(2.1) with $m_1 = 0$ and rank$(\Phi(1)) = 2$
$\hat{\lambda}$	$\left[\begin{bmatrix} 0.982 \\ .621+.246i \\ .621-.246i \\ 0.653 \end{bmatrix}' : \begin{bmatrix} -0.513 \\ 0.489 \\ -0.114 \\ -0.078 \end{bmatrix}'\right]$	$\left[\begin{bmatrix} 1.000 \\ 1.000 \\ 0.689 \\ .606+.265i \end{bmatrix}' : \begin{bmatrix} .606-.265i \\ -0.502 \\ -.182+.083i \\ -.182-.083i \end{bmatrix}'\right]$
Norm_1:	151.526*	106.167*
Norm_2:	6.072*	5.845*
Norm_3:	2.609	0.636
Norm_4:	5.307	5.095
NormS:	75.410*	60.583*
NormK:	45.878*	37.041*
$\text{AC}(1)_1$:	3.671*	3.085*
$\text{AC}(1)_2$:	2.790*	2.587*
$\text{AC}(1)_3$:	0.465	0.354
$\text{AC}(1)_4$:	0.966	0.947
$\text{AC}(1)_v$:	2.066*	1.917*
$\text{ARCH}(1)_1$:	6.626*	7.635*
$\text{ARCH}(1)_2$:	15.162*	24.281*
$\text{ARCH}(1)_3$:	5.801*	5.556*
$\text{ARCH}(1)_4$:	0.109	0.095
$\hat{D} = \det(\hat{\Sigma}) * 10^{-16}$	0.917	1.073

$-2\ln(Q)$ s.t. $m_1 = 0$

$r \leq 3$	6.781*
$r \leq 2$	19.904*
$r \leq 1$	34.873*
$r = 0$	56.554*

$\hat{\alpha}$: $\begin{bmatrix} 46.56 & -52.52 \\ -40.64 & 81.25 \\ -7.58 & -25.18 \\ 116.55 & -100.27 \end{bmatrix}$

$E(\Delta y) = \hat{C}(1)\hat{m}_0$: $\begin{bmatrix} 9.02 \\ 9.39 \\ 10.16 \\ 0.33 \end{bmatrix} * 10^{-3}$

$\hat{C}(1)$: $\begin{bmatrix} 0.165 & 0.239 & 0.384 & -0.254 \\ -0.275 & 0.578 & 0.478 & 0.545 \\ -0.465 & 0.749 & 0.547 & 0.895 \\ -0.192 & 0.155 & 0.049 & 0.350 \end{bmatrix}$

NOTE: See table 5.3 for an explanation of the symbols. An asterisk denotes significance at the 5% level using table 1 in Osterwald-Lenum (1992).

Table A5.2 *Results of estimation of VAR system for French data* 1970.3–1983.1

	Estimated Model	
	Model (2.1) unrestricted	(2.1) with $m_1 = 0$ and $\text{rank}(\Phi(1)) = 2$
$\hat{\lambda}$	$\left\{\begin{bmatrix}0.924\\ .737+.303i\\ .737-.303i\\ .492+.081i\end{bmatrix}' : \begin{bmatrix}.492-.081i\\ -0.471\\ .050+.053i\\ .050-.053i\end{bmatrix}'\right\}$	$\left\{\begin{bmatrix}1.000\\ 1.000\\ .688+.169i\\ .688-.169i\end{bmatrix}' : \begin{bmatrix}0.511\\ -0.475\\ -0.184\\ -0.075\end{bmatrix}'\right\}$
Norm_1:	0.420	0.108
Norm_2:	1.705	1.475
Norm_3:	6.024*	3.600
Norm_4:	7.532*	6.729*
NormS:	45.341*	53.547*
NormK:	4.497*	12.154*
$\text{AC}(1)_1$:	0.901	0.920
$\text{AC}(1)_2$:	0.863	1.420
$\text{AC}(1)_3$:	1.532	0.802
$\text{AC}(1)_4$:	1.270	1.636
$\text{AC}(1)_v$:	1.917*	1.631
$\text{ARCH}(1)_1$:	4.327*	5.329*
$\text{ARCH}(1)_2$:	0.736	0.370
$\text{ARCH}(1)_3$:	5.072*	0.736
$\text{ARCH}(1)_4$:	0.156	0.331
$\hat{D} = \det(\hat{\Sigma}) * 10^{-16}$	0.123	0.135

$-2\ln(Q)$ s.t. $m_1 = 0$

$r \leq 3$	7.200**
$r \leq 2$	17.805**
$r \leq 1$	29.969*
$r = 0$	69.200**

$\hat{\alpha}$: $\begin{bmatrix}88.28 & -93.23\\ -68.89 & 69.34\\ -8.10 & -12.23\\ 33.01 & 212.06\end{bmatrix}$

$E(\Delta y) = \hat{C}(1)\hat{m}_0$: $\begin{bmatrix}7.14\\ 8.14\\ 8.59\\ 0.015\end{bmatrix} * 10^{-3}$

$\hat{C}(1)$: $\begin{bmatrix}-1.221 & 0.834 & 0.384 & -0.382\\ -1.549 & 1.170 & 0.478 & -0.556\\ -0.214 & -0.750 & 1.379 & 0.497\\ -0.018 & 0.027 & -0.015 & -0.015\end{bmatrix}$

NOTES: See Table 5.3 for an explanation of the symbols. One asterisk denotes asymptotic significance at the 5% level, two asterisks significance at the 2.5% level using table 1 in Osterwald-Lenum (1992).

A5.3 Robust estimation by extraction of additive outliers

The robust data cleaning procedure we discuss here is only designed to detect and remove additive outliers. The idea is the same as in §4.2. We do not accurately observe the latent variables we are interested in. In order to estimate the parameters of our model we have to remove the observation errors from the data. These errors occur because of certain sporadic exogenous influences which affect some observations of one or more of the variables of our model. In this set–up the other variables can remain unaffected by these influences.

Model

In the AO model we observe x_t instead of the variables of interest y_t.

$$x_t = y_t + u_t, \tag{A5.3.1}$$

where u_t has a mixture of two distributions so that for 90–95% of the sample y_t and observed x_t can be considered equal for all interesting purposes: $P(|u_t| < \varepsilon) > 0.90$, with ε small. In §5.2 one has $x_t = y_t^*$, $u_t = \delta_{t,t_1} O_1 + (1-\rho L)^{-1}\delta_{t,t_1} O_{3t}$.

We apply the AO model to a process specified in *levels* of y_t. Alternatively one can assume that the growth rate of a variable has been observed incorrectly only once. This would imply that all the subsequent observations of the level of the variable (a so–called *level outlier* (Tsay (1987)) have been suffering from the same error.

We do not specify an exact distribution function for the u_ts. Therefore no hard mathematical results for efficiency, consistency or bias of our following estimation procedure can be obtained. We only hope to develop a method with a better robustness against outliers of a type frequently occurring in economic time series, which can serve as an alternative to the ad hoc procedure of introducing exogenous dummy variables here and there.

Even in quarterly economic time series one can observe clustered or so–called *patchy* outliers which cannot be treated as the response of the system to an innovation outlier. In order to be able to account for this type of outliers we implicitly assume a slight intertemporal dependency of the outliers.

The basic idea is that the additive outlier is a response to an *intervention* in the sense of Box and Tiao (1975) which affects only one observed variable of our system significantly. This response can extend over

several periods. Its initial component is so large that it can be discriminated from the normal error term ε_t, and the subsequent components rapidly decrease in time, as opposed to the transient level change in §5.2.

In this way x_t can be affected by all effective interventions occurring before and at time t. Let u_t be the sum of the responses of x_t to all interventions at or before time t. Let the n–vector I_t (usually consisting of only zeros, sometimes containing one element unequal to zero) denote the *Intervention* at time t and let ℓ_t denote the *length* of the corresponding response. Then we can write u_t as follows:

$$u_t = \sum_{j=1}^{t-p} O_j(\, I_{t-j},\, \ell_{t-j}). \tag{A5.3.2}$$

The response (or observed outlier) function O_j is diagonal so that immediate cross influences are neglected. It is modeled as

$$O_j(I_{t-j},\ell_{t-j}) = \begin{bmatrix} o_{1j}(I_{1(t-j)},\ell_{1(t-j)}) \\ \vdots \\ o_{nj}(I_{n(t-j)},\ell_{n(t-j)}) \end{bmatrix}.$$

where the response at time t to an intervention at time $t-j$ is given by the following function: ($i=1,\ldots,n$ and $t=1,\ldots,T-1$):

$$o_{ij}(I_{i(t-j)},\ell_{i(t-j)}) = g(j,\ell_{i(t-j)})\cdot I_{i(t-j)}, \tag{A5.3.3}$$

with

$$g(j,\ell_{i(t-j)}) = \frac{h(\ell_{i(t-j)}-j)}{h(\ell_{i(t-j)})}, \qquad j=0,1,\ldots,\ell_{i(t-j)},$$

$$= 0, \qquad j>\ell_{i(t-j)},$$

with

$h(z_t) = z_t(z_t+1)$.

For $n=1$, $I_t=6$ and $\ell_t=3$ the response is given in fig. A5.3.1. It shows how fast the observed variable returns to the latent value after an intervention at time $t=0$. The growth rate adjustment is a linear, the level adjustment is a quadratic function of time. It is clear that an intervention which affects ℓ observations of x_t changes $\ell+1$ observations of Δx_t.

Robust filtering

How can one estimate the parameters $\Phi(L)$, m_0, m_1 in (5.1) if one observes

x_t instead of y_t? One can avoid the biases in the estimation by correcting the observed x_ts into the direction of the "true" y_ts of interest, the so–called *clean data*. This can be achieved by applying a so–called robust filter to the observed data. This is an iterative procedure which simultaneously "estimates" the parameters as well as the responses u_t.

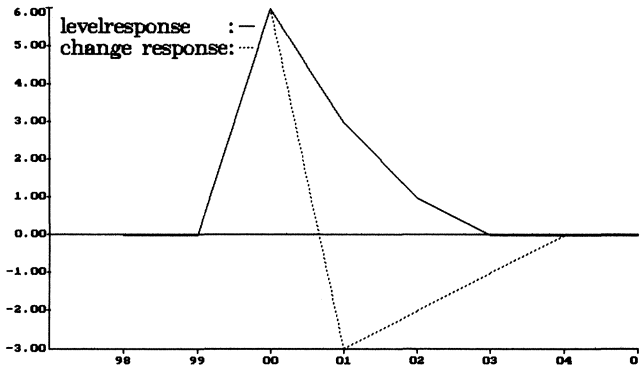

Figure A5.2.1 *Example of an outlier nonsystem impulse response*

Step one of this procedure, which is largely due to Kunst (1986) and based on ideas in Kleiner et al. (1979), is to estimate the parameters by OLS. Then one cleans the x_ts $(t=p+1,p+2,\ldots,T)$ by a filter (Step 2), i.e. one estimates u_t and calculates $\hat{y}_t$ using (A5.3.1). Note that in this filter the n–vectors $\hat{y}_t,\hat{y}_{t+1},\ldots,\hat{y}_{t+j}$ are calculated from the freshly corrected intermediate series $x^*_{t-1},x^*_{t-2},\ldots,x^*_{t-1+j}$. In this way cross influences between the variables are taken into account, which would not be possible if the filter was run separately for each variable.

Subsequently one reestimates $\Phi(L)$ and u_t using the cleaned data $\hat{y}_t$ and repeats Step 1 and Step 2 until the estimates for the parameters show only small relative changes. The filter is supposed to detect both the presence of the responses and the length of them. An observation is identified as being affected by the initial component of a response (outlier) if the estimated residual is improbably large. A response is identified as extending over several periods (patchy outlier) if the improbably large residual is followed by a sequence of relatively large residuals.

The filter computes recursively for $t=p+1,p+2,\ldots,T-1$:

$$\hat{y}_{t+j}=(I-\hat{\Phi}(L))x_t^*+m_0+m_1t+\Psi_j\left(\hat{e}_t,\hat{e}_{t+1},\ldots,\hat{e}_{t+\hat{\ell}_t}\right), \qquad j=0,\ldots,\hat{\ell}_t,$$

where

$$\hat{e}_t=x_t^*-((I-\hat{\Phi}(L))x_t^*+m_0+m_1t), \qquad \hat{\ell}_t=\max_i \hat{\ell}_{it}, \qquad \text{(A5.3.3}a\text{)}$$

with

$$\begin{aligned} &\ell_{it}=\max\{j,0\},\\ \text{s.t.}\quad &j\in\{0,1,\ldots,N-t\},\\ &\|\hat{e}_{it}\|>c_{i1},\\ &\|\hat{e}_{i(t+m)}\|>c_{i2}, \qquad m=1,\ldots,j, \end{aligned} \qquad \text{(A5.3.3}b\text{)}$$

and

$$\Psi_j\left(\hat{e}_t,\ \hat{e}_{t+1},\ldots,\hat{e}_{t+\hat{\ell}_t}\right)=\begin{bmatrix}\psi_{1j}(\hat{e}_{1t},\hat{e}_{1(t+j)})\\ \vdots\\ \psi_{nj}(\hat{e}_{nt},\hat{e}_{n(t+j)})\end{bmatrix}$$

is a diagonal function matrix so that for $i=1,\ldots,n$ and for $j=0$:

$$\psi_{i0}(\hat{e}_{it})=\begin{cases}\hat{e}_{it} & \text{if } \|\hat{e}_{it}\|\le c_{i1}\\ c_{i1} & \text{if } (\hat{e}_{it})> c_{i1},\\ -c_{i1} & \text{if } (\hat{e}_{it})<-c_{i1}\end{cases} \qquad \text{(A5.3.3}c\text{)}$$

and for $i=1,\ldots,n$ and $j>0$:

$$\psi_{ij}(\hat{e}_{it},\hat{e}_{i(t+j)})=\begin{cases}\hat{e}_{i(t+j)} & \text{if } j>\ell_{it}\\ \hat{e}_{i(t+j)}+g(j,\ell_{it})\cdot(\psi_{i0}(\hat{e}_{it})-\hat{e}_{it}) & \text{if } 0<j\le\ell_{it}\end{cases}. \qquad \text{(A5.3.3}d\text{)}$$

ψ_{i0} is a so-called *Huber function*, see e.g. Hogg (1974). It is one of the simplest functions to reduce the influence of outlying observations on the estimates for the parameters in a regression equation.

Choosing tuning constants

Note that an observation y_t is not corrected unless there is a variable i for which $(\max_j(\ell_{i(t-j)}-j)>1,\ j=1,2,\ldots,t)$ or for which $\|\hat{e}_{it}\|>c_{i1}$. The number of corrected observations increases with decreasing c_{i1} and c_{i2}. The c_is are measured in the well known robust estimates of the standard error of the residuals given by

$$s_i=\text{ sample median }\left(\|\hat{e}_{it}\|\right)/\ 0.67445.$$

The *tuning constant* c_{i1}^* is defined as the ratio of c_{i1} to s_i. The *scale estimate* s_i can be updated for every new choice of a lower tuning constant. c_{i2}^* is defined analogously. For the computations in §5.6 and §5.7 we set the c_{i1}^*s equal to 2 once and for all.

The length of a patchy outlier beginning at t is denoted by ℓ_{it}. The ℓ_{it}s are updated both within and between the iterations. A patchy outlier caused by an intervention is only detected as such if it gives rise to an initial prediction error $|\hat{e}_{it}|$ which is greater in absolute value than c_{i1} followed by a prediction error $\hat{e}_{i(t+2)}$ which is greater in absolute value than c_{i2}. A choice for the second tuning constant could be s_i. For a clean sequence of independent normally distributed prediction errors this would mean that the probability of an outlier being identified as patchy would be about 32%. This seems rather high, the expected costs in the form of improper correction of the observation following the outlier are not as large: The expected size of the correction of an "independent normal" observation following an outlier, denoted by $E(o_{t+1}|I_t>0)$ is less than one fifth of the initial correction I_t. It follows from (A5.3.3*b*) and (A5.3.3*d*) that

$$E(o_{t+1}|I_t>0) \approx \sum_{j=1}^{T-t}(1-2/(j+1))\cdot(0.32)^{(j+1)}\cdot I_t < 0.2\cdot I_t.$$

The suffix i has been deleted for simplicity.

In the French example in §5.6 and §5.7 we set the c_{i2}^*s equal to 10. In that way the bounds became ineffective, and we thus abstracted from the *adaptive* (i.e. identified from the data) patchy outlier alternative. For the Dutch data in chapter 7 we also only used the tuning constants c_{i1}.

We use a simple stopping criterion. Iteration stops when the Euclidean norm of the parameter change (a $(p+1)\times n^2+2n$ vector) comes below a certain threshold, e.g. 10^{-8}, depending on size and number of parameters.

Other modifications

Note that the observation following an initial response can be identified as an initial response to a second intervention as well, though with a smaller probability than the first one because of the initial a priori forward looking correction in (A5.3.3*d*). This can occur in the correction of a consumption series affected by a VAT increase. With more a priori information one can easily modify the filter by changing (A5.3.3*d*) to have a better resistance against this kind of anticipatory purchases which are offset in the next period(s) by a tantamount decline e.g.:

$$\psi_{ij}(\hat{e}_{it},\hat{e}_{i(t+j)}) = \hat{e}_{i(t+j)} + g(j)(\psi_{i0}(\hat{e}_{it}) - \hat{e}_{it}), \qquad \text{(A5.3.4)}$$

with $g(0)=1, \sum_{i=1}^{\infty} g(i) = -1$.

By using this filter one would also avoid the distortion of equilibrium relations in the cumulated levels of the variables, which could be of crucial importance in empirical applications of models with so–called *multicointegration*, a type of cointegration involving $I(2)$ variables, see Granger and Lee (1990) for an example. The outlier model itself is a simple (one–shot) example of so–called *integral correction*.

The French data (consumption 1968.2–1968.3, figure 5.14) and Dutch data (investment 1979.1–1979.2, figure 7.13) provide empirical examples of integral correction of the form $\ell_{it}=1$, $\hat{O}_{it}=\psi_{i0}(\hat{e}_{it})-\hat{e}_{it})<0$, $g(1)=-1$, $\hat{O}_{it+1}=-\hat{O}_{it}$. The effect on the *changes* Δx_t, Δx_{t+1} and Δx_{t+2} is $\hat{O}_{it}$, $-2\hat{O}_{it}$, and $\hat{O}_{it}$. The last example shows that the iterative estimation procedure may pick up this pattern anyway. The innovation outlier estimation procedure for the French data neglected the effect on Δx_{t+2}.

From the artificial example in §5.4 it is clear that the period with the biggest residual is not always the period where the additive outlier occurred first. Essentially we use a slightly modified robust univariate innovation outlier test here to detect the aberrant observations. If one starts the data cleaning one period too late, the data cleaning procedure may not be successful. In practice one will detect this by slow or non convergence. In principle one could generalize the method to detect surprising multivariate outliers using a robust version of the additive outlier test from §5.3. A simple initial robust estimate of the covariance matrix of the disturbances would then be required, which is harder to construct than the single equation robust estimate of the scale that we use.

Another tedious improvement would be to replace the filtering formula for $\hat{y}_{t+j}$ by a smoothing formula like (4.4) so that y_{t+j} would be replaced by its expected value given $\{x_t\}\backslash(x_t,x_{t+1},\ldots,x_{t+j})$. Peña (1990) suggested such a procedure for simple stationary univariate ARMA models.

6 RESTRICTIONS ON THE VAR

6.1 Introduction

We started by univariate analysis and an examination of influential observations, parameter stability, seasonality and outliers in an unrestricted VAR. After a subsequent "treatment" of the data it may be possible to start a reliable econometric analysis, which ultimately tries to come up with results about interpretable time series relationships between macroeconomic variables.

Incredible restrictions

The traditional vehicle used in this exploration is the "structural" *simultaneous equation model* (SEM). In this model one imposes restrictions derived from economic theory on the dynamic and covariance parameters of the VAR so that the remaining free parameters can be interpreted as multipliers, long run elasticities etc., see e.g. Monfort and Rabemananjara (1990).

Many of these structural assumptions can be formalized as zero restrictions on some coefficients. If these restrictions do not hold, inference based on these models becomes "incredible". Sims (1980) argued this to be a very serious problem, and suggested ways to circumvent the imposition of such precise restrictions. Especially forward looking behavior of economic agents, combined with the superior information of the agents relative to the information set used by the econometrician, can make seemingly reasonable exogeneity and Granger noncausality assumptions a hazardous basis for inference and interpretation of the results.

Campbell and Shiller (1987, footnote 7) provided a formal illustration of this point in a model where the long term interest rate is assumed to equal the present discounted value of expected future short term interest rates. They showed that Granger noncausality from the long term interest rate to the short term interest rate in a simple two–variable model must be rejected on theoretical grounds if the relevant information set of the econometrician only contains present and past values of the two variables. Their theoretical prediction was confirmed by U.S. data. Campbell (1987) showed in a related example that "economic" exogeneity (i.c. of labor income) must not be mixed with exogeneity in a statistical sense.

Modern "stochastic general equilibrium" representative agent models have serious defects as well. The restrictions are often nonlinear and hard to

implement and test. Derivation of the restrictions is usually only possible in very simple models which can be considered quite unfit for a number of reasons.

Both traditional and modern overidentifying restrictions can be tested, but proper inference can be conducted only if one has a reasonable idea how to transform the possibly integrated multiple time series to stationarity. Once one has established a proper way to do this, traditional tests can be applied, *conditional* on this model, see §3.3.2. If one has used and *adaptive* way to transform the series, i.e. by estimating some sample specific coefficients (e.g. coefficients in prior detrending) one has to take this into account in subsequent tests, see e.g. the effect of prior subtraction of the sample mean on tests for parameter stability (Ploberger and Krämer (1991)). The examinations of the influence of (im)proper detrending on tests of the permanent income hypothesis (see Campbell and Mankiw (1991) for references) showed that empirical economists should worry about this point.

Handling nonstationarity in mean

Finding the appropriate transformation in a multivariate model is rather more difficult than in a univariate model, although some ideas from the univariate case can be generalized quite easily. From a data based point of view it is attractive to use the information of all the variables to get the right transformation. A disadvantage is that graphical analysis becomes difficult for more than two variables. From an economic point of view multivariate analysis is of course far more interesting. Economic theory may suggest simple transformations to stationarity albeit with some unknown parameters (see e.g. Campbell and Shiller (1987)).

From an econometric point of view tests of the appropriateness of these transformations can thus be worth while. As one can see in the next section the transformation to stationarity can be linked with a multivariate unobserved components model, called the *common trends* model (Stock and Watson (1988)). The common trends specification is attractive since it can lead to a partly data based orthogonalization of the covariance matrix of the disturbances, so that subsequent impulse response analysis and variance decompositions can get an interpretation which does not rely exclusively on theoretic arguments.

Another reason for the interest for the right transformation to stationarity may be efficiency. Unrestricted estimators can be improved by incorporating (nearly) true unit root type restrictions, so that subsequent forecasts (Engle and Yoo (1987), LeSage (1990), Bradley and Lumpkin (1992)) or estimators of parameters of interest can become more precise. A final argument

in favor can be put forward by an appeal to Ockham's razor, reformulated in the statistics discipline as the Jeffreys–Winch simplicity postulate, see e.g. Zellner (1988).

In chapter 3 we remarked several times that the statistical framework associated with some diagnostic tests, such as those for parameter constancy, does not encompass unit root nonstationarity. After one has imposed the "correct number of unit roots" and after one has estimated the corresponding coefficients for the multivariate transformation to $I(0)$–ness a number of outcomes of diagnostic test statistics applied on the $I(0)$–system are somewhat easier to interpret. A rigorous statistical set–up, which combines manageable tests for unit root nonstationarity and time varying parameter nonstationarity in the VAR has to be developed yet.

If diagnostic checks on the appropriately transformed VAR do not show signs of misspecification on the aspects of interest, and if the variables of interest still show enough variation one can finally apply the traditional SEM framework or the rational representative present discounted value maximizing agent framework to test subsequent restrictions.

Outline of chapter

The remainder of this chapter deals with the important multivariate analysis of nonstationarity in the VAR model. In the next section we use the model and notation of chapter 2 to show how some empirically relevant unit root restrictions can be parameterized. This theoretical analysis does not immediately lead to closed form expressions which can be applied right away. In §3 we develop a special, empirically important case in more detail and incorporate deterministic terms in the model. In §4 we discuss extensions for the modeling of the deterministic trends in the VAR. In §5 we discuss several readily applicable methods to detect and estimate interesting unit root like restrictions, with associated ways to estimate pushing trends and pulling equilibria. In §6 we discuss several statistical tests.

6.2 Cointegration, common trends and the number of unit roots

Let us first recall the notation of chapter 2 where we introduced the factorized purely stochastic VAR model in (2.1)–(2.3)

$$V(L)M(L)U(L)y_t = \varepsilon_t, \tag{6.1}$$

with $V(L)$ and $U(L)$ stable matrix lag polynomials, $M(L)$ a *purely unstable* matrix lag polynomial and ε_t an n–vector of disturbances with a nonsingular covariance matrix Σ and zero mean. Using (2.7) one can write down the corresponding MA representation of $U(L)y_t$ as

$$D(L)U(L)y_t = M^*(L)V(L)^{-1}\varepsilon_t, \tag{6.2}$$

with $D(L)$ a scalar purely unstable lag polynomial so that $D(L)I = M^*(L)M(L)$, where I is an identity matrix of dimension n.

6.2.1 Cointegration and the number of unit roots

In this chapter we deal mainly with cointegration at the frequency zero. We therefore avoid using the term "at the frequency zero" from now on. Hylleberg et al. (1990) and Lee (1992) studied cointegration at frequencies ½ and ¼, see §A4.2.4.

If $M(L)$ has different diagonal elements, it follows from (6.2) that the components of $U(L)y_t$ are integrated of different orders. An important case of *cointegration of orders* (1,1), with *cointegrating rank* r in the sense of Engle and Granger (1987) can arise as follows.

One can obtain a cointegrated system in (6.1) by choosing $M(L)$ and $U(L)$ as

$$M(L) = \begin{bmatrix} (1-L)I_1 & 0 \\ 0 & I_2 \end{bmatrix}, \; U(L) = U_0 = \begin{bmatrix} I_1 & 0 \\ \alpha_{21} & \alpha_{22} \end{bmatrix}, \tag{6.3}$$

with I_1 a unit matrix of dimension k, I_2 a unit matrix of dimension r, α_{21} and α_{22} $r\times k$ and $r\times r$–matrices so that $\alpha' y_t = (\alpha_{21}\ \alpha_{22})y_t$ is $I(0)$, and so that α has full row rank r.

In order to guarantee the unit root nonstationarity of all the elements of y_t, i.e. to avoid "trivial cointegrating vectors", we assume that $\alpha_{22}^{-1}\alpha_{21}$ has full row rank. This implies that the last r rows of $M(L)$ do not lie in the space spanned by the rows of α'. Let H_{0k} denote the hypothesis that the model (6.1)–(6.3) is true. The corresponding MA representation of $U_0 y_t$ then reads

$$(1-L)U_0 y_t = C^+(L)\varepsilon_t, \tag{6.4}$$

with

$$C^+(L) = M^*(L)V^{-1}(L), \text{ and } M^*(L) = \begin{bmatrix} I_1 & 0 \\ 0 & (1-L)I_2 \end{bmatrix},$$

from which it is clear that the first k elements of y_t are $I(1)$: Both $(2\pi)^{-1}\Sigma$, the spectrum of the ε_t's at frequency zero, and $V^{-1}(e^{i0}) = V^{-1}(1)$ have full rank n. If $\alpha_{22}^{-1}\alpha_{21}$ has full rank r it is easy to establish that the last r

elements of y_t also are $I(1)$ so that all conditions[1] for the cointegration of orders(1,1) with rank r are fulfilled.

In this set-up where all components of y_t are $I(1)$ it is seen that *the number of unit roots* of $M(z)=0$, k, equals the number of components of y_t, n, minus the cointegrating rank r: $k=n-r$. After one has imposed the assumption of $I(1)$-ness for all the components of y_t a test for the cointegration rank r becomes equivalent to a test for a number of unit roots $n-r$ in $\det(\Phi(z))=\det(V(z)M(z)U(z))=0$. The shape of the implied typical log spectrum $f(\omega,r)$ near the frequency zero, which we introduced in (3.12), is based on this determinant. It can be used as an aid to check hypotheses about the cointegrating rank r. It should have a finite limit for $\omega\to 0$ under the null hypothesis. It diverges to infinity for $\omega\to 0$ if the rank is $r_1<r$ and it diverges to minus infinity for $\omega\to 0$ if the true rank is $r_2>r$.

6.2.2 Common trends

If one defines $C^*(L)$ from $C(L)=U_0^{-1}C^+(L)=C(1)+(1-L)C^*(L)$, so that $C_j^*=-\sum_{i=j+1}^{\infty}C_i$ and if one defines ξ_t as $(1-L)^{-1}\varepsilon_t=\sum_0^t\varepsilon_j$ one gets the specification which is known as the *common trends model*

$$y_t=C(1)\xi_t+C^*(L)\varepsilon_t. \tag{6.5}$$

$\text{Rank}(C(1))=\text{rank}(M^*(1))=k$. This implies that $C(1)$ has $n-k$ left eigenvectors with eigenvalues zero. $\alpha'y_t$ can only be $I(0)$ if $\alpha'C(1)$ is zero. This implies that under H_{0k} the left null space of $C(1)$ has dimension $n-k$. The number of independent integrated variables ξ_{it} that drives y_t is smaller than the dimension of y_t, because $C(1)$ does not have full rank. This explains the name of this parameterization. It also indicates a first way to obtain α from estimates of the unrestricted model, namely as a basis of the left null space of $C(1)$, which is also the left null space of $V=C(1)\Sigma C(1)'$ which in turn is proportional to the spectrum of $(1-L)y_t$ at frequency zero. Since $(1-L)y_t$ is stationary its spectrum is well defined. It does not have full rank at the zero frequency: the elements of $\alpha'(1-L)y_t$ have zero spectrum at the zero frequency. V is sometimes called the *long run covariance matrix of* Δy_t. We discuss its estimation in (6.57) below.

---

[1] Here we only consider *stochastic* cointegration in the sense of Engle and Granger. When deterministic terms are introduced, the notion of *deterministic cointegration* becomes relevant. The deterministic parts of the series are cointegrated at frequency zero if the component series have mean growth over a year unequal to zero, whereas nontrivial linear combinations have zero mean growth, see Campbell and Perron (1991, footnote 13).

The common trends model also leads to a *multivariate decomposition* of the series into a *trend component* and a *stationary* or *cyclical component*. The *trend component* is defined by

$$y_{pt} = C(1)\xi_t.$$

In the unit root literature y_{pt} is often called the *permanent component of* y_t. Innovations to this component have a permanent effect $C(1)$ on the level of y_t whereas innovations to the remaining stationary component naturally have not. King et al. (1987) were among the first to use the developments in cointegration theory to estimate the trend and cyclical component of US GDP. They performed forecast error variance decompositions which distinguished between the contribution of *permanent* and *temporary* innovations, see §6.3.3 below.

The interpretation of y_{pt} as the trend component of a series is due to Beveridge and Nelson (1981). They argued as follows. Suppose one wanted to forecast y into the far future, say at time $t+i$, given the observations up to t and suppose further that one knew the coefficients of $C(1)$ *and* $C^*(L)$ in (6.5). The optimal forecast in the mean squared error sense would be

$$\hat{y}_{t+i|t} = y_t + \sum_{j=0}^{\infty} C_j^* \left(\sum_{k=1}^{i} (\varepsilon_{t+k-j} - \varepsilon_{t+k-j-1})\right) = y_t + (-\sum_{j=0}^{i} C_j^* \varepsilon_{t-j}), \tag{6.6}$$

simply by adding y_t and the sum of optimal forecasts of the MA model for Δy_{t+j}, $\sum_1^i \Delta \hat{y}_{t+j|t}$, and putting all ε_j–values with $j > t$ equal to their (conditional) expectation zero. If one lets i tend to infinity the (negative of) the last term of this forecast can be labeled as "the sum of all forecastable changes in y_t" (beyond a deterministic rate of drift). Beveridge and Nelson claimed that "it is natural to regard this component as the transitory or cyclical portion of y_t." They called this component also "the momentum contained in y_t at time t." The residual "structural" part of the system was then labeled a "trend" term. Their trend is not identified using smoothness restrictions, as are many other definitions of trends in economic time series; see e.g. Kitagawa and Gersch (1984). In fact one can find many economic time series with positive serial correlation in the growth rates. If one approximates these processes by simple univariate AR models, one gets $C(1) > 1$ and a rather "ragged" and "overshooting" trend. The results that Beveridge and Nelson obtained by univariate ARMA techniques indicated strong correlation between the cyclical components of different macroeconomic time series in the United States.

An important difference with UCARIMA type trend extractions is the

correlation between innovations in the trend and the innovations in the stationary component. Whereas UCARIMA type models usually assume orthogonality of these innovations (cf. §4.3) to identify the different components, we see here that the innovations of the different components are both proportional to ε_t. They are thus perfectly correlated! Fernandez Macho et al. (1987) provided a UCARIMA type parameterization for the common trends model, which one could call a UC–VAR model. The models are often observationally equivalent so that statistical tests cannot be called upon to favor one interpretation or the other. Watson (1986) and Stock and Watson (1988b) discussed the effects of these very different, but both seemingly reasonable, identifying restrictions on the estimation and interpretation of economic models. Many of these structural models make a distinction between the effects on and the effects of the different components of an economic variable. The prime example where these assumptions matter is the ubiquitous life cycle model of consumption.

6.2.3 Vector error correction

Another interesting parameterization is the *vector error correction model* or VECM, introduced by Engle and Granger (1987):

$$\Phi^*(L)(1-L)y_t = \Phi(1)y_{t-1} + \varepsilon_t, \tag{6.7}$$

where $\Phi^*(L)$ is a stable lag polynomial.

Given (2.3), the conditions for cointegration and the regularity of α_{22} one can prove straightforwardly that a representation (6.7) exists. This is done below where we also discuss other versions and proofs of the so–called Granger representation theorem (see Engle and Granger (1987)). The rank of $\Phi(1)$ clearly equals $n-k$. Given appropriate identifying restrictions[2] for α one can factorize $\Phi(1)$: $\Phi(1)=\gamma\alpha'$. In this parameterization the stationary linear transformation $\alpha' y_t$ which lies in the space spanned by the *cointegrating vectors* $\alpha_{.i}' y_t$, $i=1,\ldots,r$ is labeled "error". The behavior of the growth variables $(1-L)y_t$ is seen to be partly determined by these "errors". In some economic theories these errors are interpreted as "disequilibrium" or "deviations from long term equilibrium relations". The $n\times r$ matrix γ is then interpreted as matrix of "feedback coefficients" or "loadings". Proposition 6.3 below leads to a clear interpretation in the context of impulse responses. Alogoskoufis and Smith (1991) summarized the history of different lines of

[2] Phillips (1990) chose $\gamma=(I_r\ 0)'$ in his *"triangular ECM"*. He chose not to parameterize the short run dynamics explicitly in that model. See also Perron and Campbell (1992).

theory and application of error correction models. We do not attempt to give an economic structural relation interpretation here.

The VECM is most suitable for a first estimation of the model. If one does not impose restrictions on the rank of $\Phi(1)$ it is obviously observationally equivalent with an unrestricted VAR in levels, but is has the advantage that regressors are somewhat more orthogonal. This makes the application of OLS numerically more stable, which is important if one considers the usually large number of regressors in a VAR. It also helps in the selection of significant regressors later on.

6.2.4 Other parameterizations

If one is not only interested in a model for the growth rates of the variables, the parameterization where one models the (a priori chosen) cointegrating vectors of interest directly may be more interesting:

$$M(L)U_0 y_t = (I - V(L))M(L)U_0 y_t + \varepsilon_t. \tag{6.8}$$

Campbell and Shiller (1987) were among the first to use parameterization (6.8) of a cointegrated system in empirical economic work. We will therefore refer to system (6.8) as the *CICS system*, (Co–Integration Campbell–Shiller). They estimated a VAR for $M(L)U(L)y_t = (\Delta r_t \; R_t - \theta r_t)$ with r_t the short term interest rate, and R_t the long term interest rate. R_t was assumed to be a linear function (whence the unknown parameter θ) of the present discounted value of r_t. The spread $R_t - \theta r_t$ was predicted to be stationary. Parameterization (6.8) is a good transformation for out–of–sample forecasting exercises. One only has to estimate a system for $M(L)U_0 y_t$. One can apply standard VAR forecasting from there, whereas forecasting with system (6.7) involves the recursive updating of the level variable y_{t-1}.

Hylleberg and Mizon (1989) proved the observational equivalence of several of the parameterizations of cointegrated systems of orders (1,1), i.c. the VAR representation, the common trends representation and a so–called *Bewley representation* in which $\alpha' y_t$ is expressed as a function of *present* and lagged first differences of all the components of y_t. Bewley (1979) introduced the latter parameterization to get direct (IV) estimators of long run elasticities in econometric models, which were later reinvented and advocated by Wickens and Breusch (1988). Bewley and Fiebig (1990) discussed ways to improve these estimators. Hylleberg and Mizon (1990) used the same Smith–McMillan factorization of $\Phi(L)$ which was introduced in econometrics by Salmon (1988) and by Yoo, see e.g. Engle (1987).

Extensions

Sims et al. (1990) extended the statistical analysis of cointegration to integration of arbitrary positive integer orders. They also considered nonzero means and polynomial trends in their set–up. They used the Jordan canonical form of the *companion form* of the VAR system to derive generalizations of the vector MA representation. An example of a companion form is given in (6.50) below. Tsay and Tiao (1990) extended the analysis further to cointegration of arbitrary positive integer order at arbitrary frequency. They did not include deterministic polynomials and trigonometric trends. Sowell (1987) considered the representation and estimation of cointegration models of fractionally integrated variables. Cheung and Lai (1993) found empirical evidence of a $CI(1,b)$ model $(0<b<1)$ in a purchasing power study of consumer price indexes.

Many derivations of these theoretical papers are not constructive: one cannot simply feed (estimates of) the (restricted) companion form into the computer to derive the desired Jordan form or Smith–McMillan decomposition. One often assumes the regularity of certain matrices if it is needed in a proof, without discussing its relevance. In the next section we present a straightforward constructive proof of the equivalence of the parameterizations introduced above. The results provide explicit formulae that can be used to translate estimates derived in one parameterization into those of another. We generalize the set–up by introducing constant terms that can lead to a nonzero mean and deterministic trends in y_t. It can be more efficient to extract the deterministic component in y_t by using the estimates of the VAR than by simple univariate regression. This depends on the efficiency one can achieve in the estimation of the dynamic parameters of the VAR which in turn depends on the degrees of freedom in estimation and the coefficients of the VAR, see also §6.5 below. Gradual changes in deterministic trends due to permanent innovation outliers (see §5.2) can only be estimated in the VAR.

6.3 Straightforward transformation formulae

The organization of §6.3 is as follows. First we introduce the notation. Second we derive the transformation from the Campbell–Shiller parameterization to the VECM model. Third we derive the opposite transformation. Fourth we give an expression for the $C(1)$–matrix of the MA representation. Fifth we derive the Common Trends Model and we present two examples. Sixth we compare our results with those of Johansen (1991a). Seventh we show in a few examples how the rank conditions needed in the derivations should be interpreted. Last we discuss some practical implications.

6.3.1 From Campbell-Shiller to vector error correction

The notation is as follows:

$k = n-r$ the number of independent common trends,

$\alpha = (\alpha_{.1} \dots \alpha_{.r})$ an $n\times r$–matrix of cointegrating vectors,

$S_1 = (I_1 : 0)$ a $k\times n$ row selection matrix, I_1: a $k\times k$ identity matrix,

$S_2 = (0 : I_2)$ an $r\times n$ row selection matrix, I_2: an $r\times r$ identity matrix,

$$G = \begin{pmatrix} S_1 \\ \alpha' \end{pmatrix} = \begin{pmatrix} I_1 & 0 \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \text{ an } n\times n \text{ transformation matrix,}$$

$$G^{-1} = \begin{pmatrix} I_1 & 0 \\ \beta_{21} & \beta_{22} \end{pmatrix}, \text{ i.e. } \beta_{21}+\beta_{22}\alpha_{21}=0;\ \beta_{22}\alpha_{22}=I_2;\ \alpha_{21}+\alpha_{22}\beta_{21}=\alpha' G^{-1}S_1' = 0,$$

$$\Delta x_t = \begin{pmatrix} (1-L)x_{1,t} \\ \vdots \\ (1-L)x_{n,t} \end{pmatrix} = \begin{pmatrix} x_{1,t}-x_{1,t-1} \\ \vdots \\ x_{n,t}-x_{n,t-1} \end{pmatrix},$$

$$A(L) = \begin{pmatrix} A_{11}(L) & A_{12}(L) \\ A_{21}(L) & A_{22}(L) \end{pmatrix},$$

where the matrix lag polynomials of the CICS model $A_{11}(L)$, $A_{21}(L)$, $A_{12}(L)$, $A_{22}(L)$ are $k\times k$, $r\times k$, $k\times r$ and $r\times r$ respectively. $A_{ij}(L) = A_{ij1} + A_{ij2}L + A_{ij3}L^2 + \dots + A_{ijp}L^{p-1}$ with p the maximum order of the AR polynomial $(1-LA(L))$. $B(L)$ is defined in the same way as $A(L)$ for the VECM model. c, c^* and m are n–vector of constants. c_1 and c_1^* are k–vectors of constants. c_2 and c_2^* are r–vectors of constants. ε_t is an n–vector of zero mean, serially independent normally distributed innovations, with $E(\varepsilon_t\varepsilon_t') = \Sigma$, a nonsingular matrix.

In this section we assume no longer that the endogenous variables have mean zero. We therefore change notation from y_t to x_t, cf. (2.2) and (A4.2.2). Note that we maintain the assumption of stationarity for the cointegrating vector. This assumption will be relaxed in §6.3.5. Here we maintain the following assumptions:

Assumptions 6.1

x_t is an n–vector of endogenous variables, all its components are $I(1)$–variables in the sense of Engle and Granger (1987), i.e. in first differenced form the stochastic parts of the components have a stationary ARMA representation with invertible MA part, and there are $r \le n-1$ independent vectors $\alpha_{.i}$ such that there exists a partitioning $\alpha' = (\alpha_{21}\ \alpha_{22})$ with α_{22} regular and so that $\alpha_{.i}'x_t$ is a mean stationary $I(0)$–variable, $i = 1,\dots,r$.

Campbell and Shiller used a stationary autoregressive system for $[(S_1\Delta x_t)'\ (\alpha' x_t)']$ to derive elegant tests of simple rational expectations models for the variable $\alpha' x_t$. Here we show explicitly how their model can be rewritten in the *vector error correction* form of Engle and Granger i.e. how to

transform a system of the form

$$(I-LA(L))\begin{bmatrix} S_1\Delta x_t \\ \alpha' x_t \end{bmatrix} = c+G\varepsilon_t, \tag{6.9}$$

where $c=Gm$, to a system of the following form

$$(I-LB(L))\Delta x_t = \gamma\alpha' x_{t-1}+m+\varepsilon_t. \tag{6.10}$$

Note that the stochastic part of (6.9) has form (6.1)–(6.3) with $V(L)=I-LA(L)$, $U(L)=U_0=G$, and $D(L)=(1-L)$, so that (6.9) really represents a cointegrated system, see also (6.8).

Proposition 6.1

Under assumptions 6.1, (6.9) implies (6.10)

Proof. First we prove the following under assumptions 6.1.

Lemma 6.1

$$\begin{bmatrix} S_1\Delta x_t \\ \alpha' x_t \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} A_{11}(L) & A_{12}(L) \\ A_{21}(L) & A_{22}(L) \end{bmatrix}\begin{bmatrix} S_1\Delta x_{t-1} \\ \alpha' x_{t-1} \end{bmatrix} + \begin{bmatrix} S_1\varepsilon_t \\ \alpha'\varepsilon_t \end{bmatrix} \Leftrightarrow \tag{6.11}$$

$$\begin{bmatrix} S_1\Delta x_t \\ S_2\Delta x_t \end{bmatrix} = \begin{bmatrix} c_1^* \\ c_2^* \end{bmatrix} + \begin{bmatrix} A_{11}^*(L) & A_{12}^*(L) \\ A_{21}^*(L) & A_{22}^*(L) \end{bmatrix}\begin{bmatrix} S_1\Delta x_{t-1} \\ \alpha' x_{t-1} \end{bmatrix} + \begin{bmatrix} S_1\varepsilon_t \\ S_2\varepsilon_t \end{bmatrix}, \tag{6.12}$$

where

$\beta_{21}=-(\alpha_{22})^{-1}\alpha_{21}$, $\beta_{22}=(\alpha_{22})^{-1}$,which is assumed to exist, $c_1^*=c_1, c_2^*=\beta_{21}c_1+\beta_{22}c_2$,
$A_{11}^*(L)=A_{11}(L)$, $A_{12}^*(L)=A_{12}(L)$, $A_{21}^*(L)=\beta_{21}A_{11}(L)+\beta_{22}A_{21}(L)$,
$A_{22}^*(L)=\beta_{21}A_{12}(L)+\beta_{22}A_{22}(L)-\beta_{22}=\beta_{21}A_{12}(L)-\beta_{22}\left[I-A_{22}(L)\right)$,
in short:

$$A^*(L)=G^{-1}(A(L)-S_2'S_2).$$

Proof. Substitute $G\Delta x_t+S_2'\alpha' x_{t-1}$ for the expression at the left hand side of (6.11), substitute $(S_1'S_1\Delta x_{t-1}+S_2'\alpha' x_{t-1})$ at the right hand side and premultiply both sides by G^{-1} to get

$$\Delta x_t+G^{-1}S_2'\alpha' x_{t-1}=G^{-1}c+G^{-1}(A(L)(S_1'S_1\Delta x_{t-1}+S_2'\alpha' x_{t-1}))+\varepsilon_t \Leftrightarrow$$

$$\Delta x_t=G^{-1}c+G^{-1}(A(L)-S_2'S_2)(S_1'S_1\Delta x_{t-1}+S_2'\alpha' x_{t-1})+\varepsilon_t.$$

Writing out the partitioned G, G^{-1}, α and $A(L)$ completes the proof lemma 6.1.□
By rearranging the lag polynomial operating on $\alpha' x_t$ in (6.12) one can show

Lemma 6.2

Equation (6.12) is equivalent with

$$\begin{bmatrix} S_1\Delta x_t \\ S_2\Delta x_t \end{bmatrix} = \begin{bmatrix} c_1^* \\ c_2^* \end{bmatrix} + \begin{bmatrix} B_{11}(L) & B_{12}(L) \\ B_{21}(L) & B_{22}(L) \end{bmatrix}\begin{bmatrix} S_1\Delta x_{t-1} \\ S_2\Delta x_{t-1} \end{bmatrix} + \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}\alpha' x_{t-1} + \begin{bmatrix} S_1\varepsilon_t \\ S_2\varepsilon_t \end{bmatrix}, \tag{6.13}$$

where $B_{11}(L)=A_{11}^*(L)-\tilde{A}_{12}^*(L)\alpha_{21}$, $B_{12}(L)=-\tilde{A}_{12}^*(L)\alpha_{22}$, $B_{21}(L)=A_{21}^*(L)-\tilde{A}_{22}^*(L)\alpha_{21}$,
$B_{22}(L)=-\tilde{A}_{22}^*(L)\alpha_{22}$, $\gamma_1=A_{12}^*(1)$, and $\gamma_2=A_{22}^*(1)$,

with

$$(1-L)\tilde{A}^*(L) = A^*(1) - A^*(L)$$

and

$$\tilde{A}^*(L) = \sum_i \tilde{A}_i^* L^i = \sum_i \Big(\sum_{j=i+1}^{\infty} A_j^* \Big) L^i = \sum_i \Big[A^*(1) - \sum_{j=1}^{i} A_j^* \Big] L^i. \tag{6.14}$$

Proof. Substitute $A^*(L) = A^*(1) - (1-L)\tilde{A}^*(L)$ for the part operating on $\alpha' x_{t-1}$ in (6.12) to get

$$\Delta x_t = c^* + A^*(L) S_1' S_1 \Delta x_{t-1} + (A^*(1) - (1-L)\tilde{A}^*(L)) S_2' \alpha' x_{t-1} + \varepsilon_t \Leftrightarrow$$

$$\Delta x_t = c^* + (A^*(L) S_1' S_1 - \tilde{A}^*(L) S_2' \alpha') \Delta x_{t-1} + A^*(1) S_2' \alpha' x_{t-1} + \varepsilon_t. \tag{6.15}$$

Writing out the partitioned α, $A^*(L)$ and $\tilde{A}^*(L)$ completes the proof of lemma 6.2. From lemma 6.1 it follows that $A^*(L) = G^{-1}(A(L) - S_2' S_2)$ so that lemma 6.2 completes the proof of proposition 6.1. □

Corollary 6.1

From (6.13) it follows that $B_{i2}(L) = -\tilde{A}_{i2}^*(L)\alpha_{22}$, $i = 1, 2$, so that the last r columns of the matrix B_p belonging to the highest lag of a VECM representation of a finite order CICS model *always* will be zero. A CICS representation of a VECM model will in general have one more lag on the first differences of the series, unless $B_p(L) S_2' = 0$, see also (6.18) below.

Corollary 6.2

The matrix $\gamma = A^*(1) S_2' = G^{-1}(A(1) - S_2' S_2) S_2'$ has the same rank as $((A(1) - I) S_2')$, which has full column rank r, since $(A(1) - I)$ has full rank n by the assumption of stability of the CICS system, see below (6.25). It thus follows that $\gamma\alpha'$ has rank r.

Note that both (6.9) and (6.10) have nonsingular covariance matrices for the innovations, so that ML estimation conditional on α and the initial observations can be done by OLS.

6.3.2 From vector error correction to Campbell-Shiller, mean growth

In this section we show directly that (6.10) implies (6.9).

Proposition 6.2

Under assumptions 6.1 and using the definitions from §6.3.1 one can rewrite (6.10), here presented as (6.16), as (6.17) – (6.18):

$$\begin{bmatrix} S_1 \Delta x_t \\ S_2 \Delta x_t \end{bmatrix} = \begin{bmatrix} B_{11}(L) & B_{12}(L) \\ B_{21}(L) & B_{22}(L) \end{bmatrix} \begin{bmatrix} S_1 \Delta x_{t-1} \\ S_2 \Delta x_{t-1} \end{bmatrix} + \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} \alpha' x_{t-1} + \begin{bmatrix} S_1(\varepsilon_t + m) \\ S_2(\varepsilon_t + m) \end{bmatrix} \Rightarrow \tag{6.16}$$

$$\begin{bmatrix} S_1 \Delta x_t \\ \alpha' x_t \end{bmatrix} = \begin{bmatrix} A_{11}(L) & A_{12}(L) \\ A_{21}(L) & A_{22}(L) \end{bmatrix} \begin{bmatrix} S_1 \Delta x_{t-1} \\ \alpha' x_{t-1} \end{bmatrix} + \begin{bmatrix} S_1(\varepsilon_t + m) \\ \alpha'(\varepsilon_t + m) \end{bmatrix}, \tag{6.17}$$

with

$A_{11}(L)=B_{11}(L)+B_{12}(L)\beta_{21}$, $A_{12}(L)=\gamma_1+B_{12}(L)(1-L)\beta_{22}$,

$A_{21}(L)=\alpha_{21}\left[B_{11}(L)+B_{12}(L)\beta_{21}\right]+\alpha_{22}\left[B_{21}(L)+B_{22}(L)\beta_{21}\right]$,

$A_{22}(L)=I_2+\alpha_{21}\left[\gamma_1+B_{12}(L)(1-L)\beta_{22}\right]+\alpha_{22}\left[\gamma_2+B_{22}(L)(1-L)\beta_{22}\right]$,

in short: $A(L)=\ (G\gamma+S_2')S_2+GB(L)G^{-1}-GB(L)G^{-1}S_2'S_2L.$ (6.18)

Proof. Write (6.16) compactly as

$$\Delta x_t=B(L)\Delta x_{t-1}+\gamma\alpha' x_{t-1}+\varepsilon_t+m.$$

Premultiply (6.16) by G and add $S_2'\alpha' x_{t-1}$ to both sides to get

$$G\Delta x_t+S_2'\alpha' x_{t-1}=GB(L)\Delta x_{t-1}+(G\gamma+S_2')\alpha' x_{t-1}+G(\varepsilon_t+m). \tag{6.19}$$

Write (6.17) compactly as

$$G\Delta x_t+S_2'\alpha' x_{t-1}=A(L)(G\Delta x_{t-1}+S_2'\alpha' x_{t-2})+G(\varepsilon_t+m). \tag{6.17a}$$

Use $G\Delta x_t+S_2'\alpha' x_{t-1}=S_1'S_1\Delta x_t+S_2'\alpha' x_t$, and note that $S_2S_1'=0$ to verify the equivalence of (6.19) and (6.17):

(6.19) $\Leftrightarrow$

$$G\Delta x_t+S_2'\alpha' x_{t-1}=(G\gamma+S_2')S_2(S_1'S_1\Delta x_{t-1}+S_2'\alpha' x_{t-1})+GB(L)G^{-1}(G\Delta x_{t-1}+S_2'\alpha' x_{t-2})$$

$$-GB(L)G^{-1}S_2'S_2(S_1'S_1\Delta x_{t-2}+S_2'\alpha' x_{t-2})+G(\varepsilon_t+m)\Leftrightarrow(6.17a).\ \square$$

Since (6.9) is a stable autoregressive process as in appendix 4.2 one can use it to derive an *explicit* formula for the computation of the mean growth rates of the variables for which a system in the VECM form was estimated:

$$E\begin{bmatrix} S_1\Delta x_t \\ \alpha' x_t \end{bmatrix}=\begin{bmatrix} I_1-A_{11}(1) & -A_{12}(1) \\ -A_{21}(1) & I_2-A_{22}(1) \end{bmatrix}^{-1}\begin{bmatrix} S_1 m \\ \alpha' m \end{bmatrix} \tag{6.20}$$

$$=\begin{bmatrix} I_1-B_{11}(1)-B_{12}(1)\beta_{21} & -\gamma_1 \\ -\alpha_{21}\left[B_{11}(1)+B_{12}(1)\beta_{21}\right]-\alpha_{22}\left[B_{21}(1)+B_{22}(1)\beta_{21}\right] & -\alpha'\gamma \end{bmatrix}^{-1}\begin{bmatrix} S_1 m \\ \alpha' m \end{bmatrix} \tag{6.21}$$

Define $M=S_1'+S_2'\beta_{21}=G^{-1}S_1'$. Then (6.21) is equivalent to

$$\begin{bmatrix} I_1-S_1B(1)M & -S_1\gamma \\ -\alpha' B(1)M & -\alpha'\gamma \end{bmatrix}^{-1}\begin{bmatrix} S_1 m \\ \alpha' m \end{bmatrix}=\left[\begin{bmatrix} I_1 & 0 \\ 0 & 0 \end{bmatrix}-\begin{bmatrix} S_1 \\ \alpha' \end{bmatrix}\begin{bmatrix} B(1)M & \gamma \end{bmatrix}\right]^{-1}\begin{bmatrix} S_1 \\ \alpha' \end{bmatrix} m. \tag{6.22}$$

The expected growth of S_2x_t follows now immediately.

$$E\left[\ S_2\Delta x_t\ \right]=\ \beta_{21}E\left[\ S_1\Delta x_t\ \right], \tag{6.23}$$

since $E(\alpha_{21}S_1\Delta x_t+\alpha_{22}S_2\Delta x_t)=0$, which is a necessary condition for the stationarity of $\alpha' x_t$, which in turn is implied by the assumption concerning (6.19). Using $G'=(\ S_1'\ \alpha)'$ we can rewrite this result in

$$E\Delta x_t=MS_1\left[\ S_1'S_1-G(\ B(1)M\ :\ \gamma\ )\ \right]^{-1}Gm, \tag{6.24}$$

or

$$E\Delta x_t = G^{-1}S_1'S_1\big(\ S_1'S_1 - G\big(B(1)G^{-1}S_1' \ :\ \gamma\big)\ \big)^{-1}Gm. \tag{6.25}$$

One proves easily that the inverse of the matrix between brackets in (6.25), $(I-A(1))$, exists under the cointegration restrictions. Suppose that it does not. Then the determinant of $(I-A(1))$ is zero. This implies that the characteristic equation of the CICS system in (6.9) contains a unit root, 1, in its AR part. This is contradicts the stationarity assumption: the matrix $(I-A(1))$ has to be regular.

Interpretation of constant terms[3] and matrix γ

Under assumptions 6.1 one can interpret $\mathrm{sp}(\gamma)$, the space spanned by the column vectors of γ, as the set of directions of shocks which have no permanent effect on the level of x_t. From (6.25) we see that the vector of constant terms m, which can be regarded as the nonzero mean of the disturbances in the VECM model, normally leads to linear trend components in x_t, except when $m \in \mathrm{sp}(\gamma)$. We prove this in

Proposition 6.3

$E\Delta x_t = 0 \Leftrightarrow m \in \mathrm{sp}(\gamma)$.

Proof. The matrix $(I-A(1))$ can be written as

$$H = \begin{bmatrix} H_1 \ :\ H_2 \end{bmatrix}, \text{ with } H_1 = \begin{bmatrix} H_{11} \\ H_{12} \end{bmatrix},\ H_2 = -G\gamma,$$

where H_1 is $n\times k$, H_2 $n\times r$, H_{11} $k\times k$ and H_{12} $r\times k$. Define the corresponding partitioned inverse $H^{-1} = J = \big(J_1' : J_2'\big)'$, where J_1 is $k\times n$ and $J_2 = r\times n$, so that $J_1H_1 = I_k$ and $J_1G\gamma = 0_{k\times r}$. From (6.25) it is seen that $E\Delta x_t = 0 \Leftrightarrow G^{-1}S_1'S_1JGm = 0_{n\times 1} \Leftrightarrow J_1Gm = 0_{k\times 1}$. Using $J_1G\gamma = 0_{k\times r}$ and the regularity of G it follows that $E\Delta x_t = 0 \Leftrightarrow m \in \mathrm{sp}(\gamma)$. □

In §A4.2.4 we list the corresponding situations where deterministic periodically changing means in ε_t lead to *periodically trending means* in x_t. A frequency ½ component in ε_t leads to a periodically trending mean unless its vector of coefficients lies in the right null space of $C_{\frac{1}{2}}(-1)$, with $C_{\frac{1}{2}}(L)$ the lag polynomial of the MA representation of $(1+L)x_t$. A frequency ¼ component in ε_t leads to a trending frequency ¼ component unless its vector of coefficients as defined in (A4.2.19) lies in the right null space of $C_{\frac{1}{4}}(i)$, with $C_{\frac{1}{4}}(L)$ defined analogously. Johansen (1991a) proved proposition 6.3 using a different parameterization, which we discuss below. In chapter 4 we advised to seasonally correct the data in such a way that no seasonal unit roots remain

[3] See §A2.2.2.3 for an analog interpretation of constant terms in the univariate case.

in the system of transformed variables. The proposed method removes all variance in the spectrum at the seasonal frequencies of all the variables so that the ranks of $C(-1)$ and $C(i)$, and thus the ranks of $C_{\frac{1}{2}}(-1)$ and $C_{\frac{1}{4}}(i)$ for an estimated system of the transformed variables all become zero. Remember that the spectral density of Δx_t is given by $(2\pi)^{-1}C(e^{i\omega})\Sigma C'(e^{i\omega})$ and that Σ is assumed to be full rank.

6.3.3 From vector error correction to common trends

Let us now turn to the MA representation for Δx_t. It is readily derived by inverting $(I-LA(L))$ in (6.9), which gives the MA representation for $\alpha' x_t$ and $S_1\Delta x_t$. The remaining r equations can be derived from $S_2\Delta x_t = S_2G^{-1}G\Delta x_t = (\alpha_{22})^{-1}(\alpha'\Delta x_t - \alpha_{21}S_1\Delta x_t)$. The key factor for the zero frequency analysis is $(I-A(1))$, which is regular by assumption. The MA representation with nonzero mean growth reads

$$\Delta x_t = C(L)(\varepsilon_t + m), \tag{6.26}$$

with

$$C(L) = G^{-1}\left[(S_1'S_1) + (1-L)(S_2'S_2)\right](I - LA(L))^{-1}G,$$

so that

$$C(1) = G^{-1}S_1'S_1\left[\, S_1'S_1 - G\left[B(1)G^{-1}S_1' \;:\; \gamma\right]\right]^{-1}G. \tag{6.27}$$

We write $\left[(S_1'S_1)+(1-L))(S_2'S_2)\right]$ for $M^*(L)$ in (6.2). $C(1)$ clearly has rank k, because $S_1'S_1$ has rank k and G is regular by assumption.

The $C(1)$–matrix in the MA representation for Δx_t is a matrix of interest. It represents the cumulative long run impulse responses and is thus interesting as a multivariate persistence measure. One can use the proof of proposition 6.3 to show that shocks in $\mathrm{sp}(\gamma)$ have persistence zero. Economic theories can be used to associate shocks with persistence zero to some kind of cyclical phenomenon, whereas shocks with nonzero persistence are associated with something "more structural", like "technology shocks." In some Real Business Cycle models like those by King et al. (1991), one exploited these ideas. The $C(1)$–matrix can be readily factorized as the product of an $n\times k$ matrix M and a $k\times n$ matrix F: $C(1) = MF$ to derive a *Common Trends model* used in King et al. (1987) and Stock and Watson (1988). We present an example below. If we make the additional

Assumptions 6.2

$\varepsilon_s = 0$ $(s \le 0)$ and x_0 has a nonrandom initial value,

and if we define $C^*(L)$ from $(1-L)C^*(L) = C(L) - C(1)$ analogously to (6.14) one can immediately write down a Common Trends specification

$$x_t = x_0 + C(1)(mt + \xi_t) + C^*(L)\varepsilon_t, \quad \xi_t = \sum_{i=1}^{t} \varepsilon_i, \tag{6.28}$$

$$x_t = x_0 + M\tau_t + C^*(L)\varepsilon_t, \quad \tau_t = \mu + \tau_{t-1} + \eta_t,$$

with

$$M = G^{-1}S_1', \quad \mu = Fm, \quad \eta_t = F\varepsilon_t, \quad F = S_1\left(S_1'S_1 - G\left(B(1)G^{-1}S_1' \; : \; \gamma\right)\right)^{-1}G = S_1C(1). \tag{6.29}$$

The matrix M, which clearly has rank k is the first factor of $C(1)$. It is clear that $\alpha'M = 0$ by the definition of G^{-1}. Given the initial value x_0 and given the estimates $\hat{\varepsilon}_t$ of ε_t using the parameter estimates of the VECM model obtained by the ML procedure of Johansen (1991a), the *permanent component* of x_t from (19) as defined in King et al. (1987), x_{pt} is then just

$$x_{pt} = x_0 + \hat{C(1)}\left(\sum_{i=1}^{t} \hat{\varepsilon}_t + mt\right). \tag{6.30}$$

The assumption implicit in assumptions 6.2, viz. that the *stationary component* $x_{st} = x_t - x_{pt} = C^*\hat{(L)}\hat{\varepsilon}_t$ is at its mean value at time $t = 0$ may not be too realistic in some macroeconomic applications. Especially European macroeconomic series that started just after the second world war can be considered to be somewhat under their mean trend line in their starting points. International trade series showed extreme growth rates (around 20 percent, see e.g. CBS (1987, §H) in the late forties and in the beginning of the fifties. Subsequently they returned to "normal" values. This can be interpreted as a "catching up" effect, related to the accumulation of large negative shocks or "innovation outliers" to trade in the years before.

Normalizations

Other common trend specifications can be generated by decomposing[4] $C(1)$ into M^*F^*, with $M^* = M\Pi$ and $F^* = \Pi^{-1}F$ and Π a regular $k \times k$ matrix. Variance decompositions as in King et al. (1987) are applied to orthogonalized innovations. The innovations $\eta_t = F\varepsilon_t$, which have a permanent effect on the variable x_t, can be orthogonalized into innovations $\eta_t^* = \Pi^*F\varepsilon_t = F^*\varepsilon_t$. One chooses Π as Π^*: the inverse of the unique lower triangular Choleski factor of $F\Sigma F'$ with positive diagonal elements, so that $\Pi^*\Pi^{*\prime} = F\Sigma F'$ and $E(\eta_t^*\eta_t^{*\prime})$ is diagonal.

One can further normalize Π^* to have ones as diagonal elements by putting Π^* equal to $\Pi^{**} = \Pi^*D^{*-1}$, with D^* a diagonal matrix with its (positive) diagonal elements equal to those of Π^*. These normalizations do not use economic theory. One may postmultiply M by a regular $k \times k$ matrix P to make its interpretation easier in the context of the favored theory. Orthogonalization

[4] King et al. (1991) used A for M, D for F.

of the innovations can then start from the factorization $C(1)=(MP)(P^{-1}F)$. The decomposition of the common trend factor $C(1)\varepsilon_t$ into interpretable orthogonalized innovations then looks like this:

$$C(1)\varepsilon_t = \tilde{M}\tilde{\eta}_t, \tag{6.31}$$

with $\tilde{M}=M\tilde{P}\tilde{\Pi}$, $\tilde{\Pi} = D^{*-1}\Pi^*$, $\Pi^*\Pi^{*\prime}=P^{-1}F\Sigma F'P^{-1\prime}$, $\tilde{F}=\tilde{\Pi}^{-1}P^{-1}F$, $\tilde{\eta}_t=\tilde{F}\varepsilon_t$ and D^* defined from Π^* as above, so that $E(\tilde{\eta}_t\tilde{\eta}_t')=\tilde{D}^2$.

Discussion of assumptions

The only extra condition, besides those necessary for cointegration, that we need for the proofs is the *regularity of* α_{22}. It implies that the r variables that are not introduced in first differences in the CICS model must enter the cointegrating relations with vectors of coefficients that span the $\mathbb{R}^r$–space. Campbell (1987) needed the *regularity* of G in his bivariate example in order to preserve regularity for the variance–covariance matrix $G\Sigma G'$ of the CICS disturbances. This is just the same condition[5].

In empirical examples with unrestricted ML estimation of the cointegrating vectors as in Johansen (1991a) this condition will be fulfilled almost surely. This is no reason for comfort. In order to prevent problems arising from near singularities one should be very careful not to introduce irrelevant random walks in the system, since this will mess up traditional VAR analyses completely. Prior regression on deterministic trends can make the situation even worse. Ohanian (1988, 1991) produced some Monte Carlo evidence on this in a VAR analysis of the effect monetary shocks on GDP in the U.S.A.. Introduction of irrelevant stationary variables is not as harmful.

Johansen (1991a) derived an alternative formula for the $C(1)$–matrix in a VECM model, using different regularity conditions in the proof. We discuss the interpretation of his assumption in our context below. King et al. (1987) derived the Common Trends Specification directly from the MA representation (6.26) using the Jordan Canonical form of $C(1)$. Warne (1991a,b) used an analysis similar to ours.

When restrictions are imposed on the cointegrating vectors the CICS representation can only be derived if all variables that do not enter the cointegrating relation appear in the first difference equations. If s variables enter all cointegrating relations with the same vector of coefficients (up to a scalar multiplication), at least s–1 of them must appear in the first difference equations. It is clear that care must be taken before

[5] King et al. (1991, p. 838) made a similar assumption (A3) of "no singularities in the structural model".

either one of the parameterizations can be assumed to represent the other.

Once one "knows" the cointegrating vectors α either by a priori reasoning from economic theory or by ML estimation it is easy to show that the corresponding ML estimates of $B(1)$ and γ in (6.24) can be obtained from one modified multivariate VECM regression (6.10):

$$\Delta\Delta x_t = (B(1) - I_n)\Delta x_{t-1} - B^{**}(L)\Delta\Delta x_{t-1} + \gamma(\alpha' x_{t-1}) + m + \varepsilon_t, \tag{6.32}$$

with $B(L) = -B^{**}(L)(1-L) + B(1)$, analogous to the decomposition of $A^*(L)$ in (6.14) and below. The set of regression equations (6.32) is just another way of writing (6.10). Johansen (1989) showed that the ML estimates of the coefficients in (6.16) conditional on α can be obtained by simple regression, so the same holds for (6.32).

6.3.4 Examples

We illustrate the use of the formulae in a numerical and in an economic example.

Example 6.1

Consider first the following VECM model. According to corollary 6.1, the last column of B_1 in (6.10) is chosen to be zero.

$$\begin{bmatrix}\Delta x_{1t}\\ \Delta x_{2t}\end{bmatrix} = \begin{bmatrix}0.50 & 0.00\\ 0.35 & 0.00\end{bmatrix}\begin{bmatrix}\Delta x_{1t-1}\\ \Delta x_{2t-1}\end{bmatrix} + \begin{bmatrix}-0.30\\ 0.15\end{bmatrix}\begin{bmatrix}1 & -2\end{bmatrix}\begin{bmatrix}x_{1t-1}\\ x_{2t-1}\end{bmatrix}\begin{bmatrix}\varepsilon_{1t}+0.40\\ \varepsilon_{2t}+1.00\end{bmatrix} \tag{6.33}$$

The corresponding CICS model (6.9) is:

$$\begin{bmatrix}\Delta x_{1t}\\ x_{1t} - 2x_{2t}\end{bmatrix} = \begin{bmatrix}0.50 & -0.30\\ -0.20 & 0.40\end{bmatrix}\begin{bmatrix}\Delta x_{1t-1}\\ x_{1t-1} - 2x_{2t-1}\end{bmatrix} + \begin{bmatrix}\varepsilon_{1t} & +0.4\\ \varepsilon_{1t} - 2\varepsilon_{2t} - 1.6\end{bmatrix}. \tag{6.34}$$

The matrices of interest in (6.27) read:

$$G = \begin{bmatrix}1 & 0\\ 1 & -2\end{bmatrix},\ G^{-1} = \begin{bmatrix}1 & 0\\ \frac{1}{2} & -\frac{1}{2}\end{bmatrix},\ B(1) = \begin{bmatrix}0.50 & 0.0\\ 0.35 & 0.0\end{bmatrix},\ C(1) = (5/8)\begin{bmatrix}2 & 4\\ 1 & 2\end{bmatrix},$$

$$E(\Delta x_{1t}) = 3,\ E(\Delta x_{2t}) = 1.50,\ E(x_{1t} - 2x_{2t-1}) = -11/3.$$

The corresponding Common Trends Model (6.28) reads:

$$\begin{bmatrix}x_{1t}\\ x_{2t}\end{bmatrix} = \begin{bmatrix}x_{10}\\ x_{20}\end{bmatrix} + \begin{bmatrix}1.00\\ 0.50\end{bmatrix}\tau_t + C^*(L)\begin{bmatrix}\varepsilon_{1t}\\ \varepsilon_{2t}\end{bmatrix},\ \tau_t = \tau_{t-1} + 3 + 1.25(\varepsilon_{1t} + 2\varepsilon_{2t}). \tag{6.35}$$

This example satisfies the condition mentioned in the previous section because $\alpha_{22} = -2 \neq 0$. The matrix $\left(S_1'S_1 - G\left(B(1)G^{-1}S_1' : \gamma\right)\right)^{-1}G$ in (6.25) is the common factor in the derivation of mean growth rates and the mean of the cointegrating vector. The inverse exists under the cointegration restrictions.

Example 6.2

Consider next the second VECM model of King et al. (1987). Their model contained five US variables in the vector $x_t' = (\ y,\ c,\ i,\ m,\ p)$, denoting the real variables output, consumption and investment and the monetary variables money stock and price level. All variables were logged. Their economic model for these variables suggested exactly three independent cointegrating vectors $\alpha_{.1}' = (\ 1\ -1\ 0\ 0\ 0\ )$, $\alpha_{.2}' = (\ 1\ 0\ -1\ 0\ 0\ )$ and $\alpha_{.3}' = (\ 1\ 0\ 0\ -1\ 1\ )$. The first two cointegrating vectors corresponded to the logs of the "greater ratios" $\ln(c/y)$ and $\ln(i/y)$. The third one established an equilibrium relation for the velocity of money. We can partition α' accordingly so that *assumption* 1 is met, by changing the order so that the monetary variables come first and the real variables come second: $x_t^{*\prime} = (\ m,\ p,\ y,\ c,\ i)$. In that case we have:

$$\alpha_{22} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 1 & 0 & 0 \end{bmatrix} \text{ in (6.11)}, M = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \end{bmatrix} \text{ in (6.28) for } x_t^*, M = \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ 1 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ for } x_t.$$

King et al. (april 1987)[6] chose $P = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ and $\Pi^* D^{\frac{1}{2}} = \begin{bmatrix} 1 & 0 \\ \pi & 1 \end{bmatrix}$ in (6.31).

Using the estimates of the VECM model of $B(L)$ and γ and α in (6.16) and an estimate of the variance covariance matrix of the residuals of this system, an estimate of $\Pi^* D^{\frac{1}{2}}$ in (6.31) and thus of π in this example is readily derived.

6.3.5 Conditions for VECM, $I(2)$-ness, and explosive systems

In this section we present the formulation of the cointegrated VAR in levels together with a translation of the regularity conditions in the proof of the Granger representation theorem in Johansen (1991a). The Vector Error Correction Model (6.10) can be written as a restricted version of a VAR in levels. Here we follow Johansen (1991a).

Consider a restricted VAR in levels:

$$\Phi(L)x_t = (I - L\Pi(L))x_t = \varepsilon_t + m \tag{6.36}$$

with $\Pi(L) = \Pi_1 + \Pi_2 L + \Pi_3 L^2 + \ldots + \Pi_p L^{p-1}$, x_t, ε_t, m defined as in (6.10), and the restriction that $-\Phi(1) = \Pi(1) - I_n = \gamma\alpha'$ has rank $r < n$ and that all roots of

---

[6] In this version of the model π was chosen to represent the "long run effect of a unit increase in the real component on the nominal variables p and M." The model was meant to contribute to a debate of the effect of nominal versus real shocks. As discussion in the real business cycle literature progressed, the empirical model and its interpretation changed considerably, see King et al. (1991).

$|\Phi(z)| = |1-z\Pi(z)| = 0$ are outside the unit circle or at $z=1$. Note that we abstract from seasonal unit roots.

One can rewrite (6.36) as system (6.10) if one defines $B(L)$ in the VECM system (6.10) from $B(L)(1-L) = \Pi(1) - \Pi(L)$ analogously to (6.14). Let $\Pi^*(1)$ denote $\sum_{i=1}^{p} i\Pi_i$. Then the condition of "balance" used by Johansen[7] to derive the MA representation (6.26) reads:

$$\gamma_\perp'\Pi^*(1)\alpha_\perp \text{ has rank } k. \tag{6.37}$$

with $\alpha_\perp$ and $\gamma_\perp$ $n\times k$–matrices of rank k such that $\alpha_\perp'\alpha = 0$ and $\gamma_\perp'\gamma = 0$. In our notation $\Pi^*(1)$ can be rewritten as $I_n - B(1) + \gamma\alpha'$ noting that

$$B(1) = \sum_{i=1}^{\infty} B_i = -\sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty}\Pi_i = -\sum_{i=2}^{\infty}(i-1)\Pi_i \Leftrightarrow \tag{6.38}$$

$$B(1) = -\sum_{i=1}^{\infty} i\Pi_i + \sum_{i=1}^{\infty}\Pi_i - I_n + I_n = -\Pi^*(1) + \gamma\alpha' + I_n, \tag{6.39}$$

so that (6.37) can be rewritten as:

$$\gamma_\perp'\left[I_n - B(1)\right]\alpha_\perp, \text{ has rank } k. \tag{6.40}$$

Johansen (1991a) considered yet another reparameterization of (6.9) and (6.10), namely a vector autoregressive system for $z_t' = ((\alpha_\perp^{+}\Delta x_t)'\ (\alpha^{+}x_t)')$, where α^{+} is the Moore–Penrose inverse of α, i.c. $(\alpha'\alpha)^{-1}\alpha'$ (see Rao and Mitra (1971)). Johansen needed condition (6.37) to avoid the possibility of unit roots in that system, whereas we need the stability of system (6.9) to prove (6.24). The connection between the results shows clearly for the degenerate case $r=0$. Rewrite (6.32) as

$$\Delta\Delta x_t = -B^{**}(L)\Delta\Delta x_{t-1} + (B(1) - I_n)\Delta x_{t-1} + \gamma\alpha' x_{t-1} + m + \varepsilon_t. \tag{6.41}$$

In the simple case where $\gamma\alpha' = 0_{n\times n}$, $\gamma_\perp = I_n$, and $\alpha_\perp = I_n$, rank condition (6.40) can be replaced by $\text{rank}(B(1) - I_n) = n$, which is just a familiar necessary condition for stationarity of the system for Δx_t in (6.41).

Johansen showed under (6.40) and the stationarity assumption for z_t that

$$C(1) = \alpha_\perp'\left[\gamma_\perp'\Pi^*(1)\alpha_\perp\right]^{-1}\gamma_\perp. \tag{6.42}$$

Stock and Watson (1988) also used the parameterization with $\alpha_\perp^{+}\Delta x_t$ and $\alpha^{+}x_t$ instead of $S_1\Delta x_t$ and $\alpha' x_t$ to rewrite the Vector MA representation (6.26). In our derivation we do not explicitly use restriction (6.37) on $\left[I_n - B(1)\right]$ and γ. From Johansen (1991a) it follows that these restrictions are necessary to guarantee the existence of a $CI(1,1)$–system in the first place, which is *implicit* in our *assumptions* 1. One can illustrate this in the example of the

---

[7] Johansen (1991a) denoted $\Sigma i\Pi_i$ in (6.36) by Ψ, α by β, γ by α, p by k and n by p in (6.10).

previous section. For $\alpha_\perp$ one can use $G^{-1}S_1'$ which clearly has rank k, and $\alpha' G^{-1}S_1' = 0$. From (6.38) we can extract restrictions on $B(1)$ and γ. Suppose that $\gamma \subset \mathrm{sp}((I_n - B(1))\alpha_\perp)$ so that $\gamma_\perp'\left[I_n - B(1)\right]\alpha_\perp = 0$ which contradicts (6.38). We adjust (6.33) accordingly:

Example 6.3

$$\begin{bmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{bmatrix} = \begin{bmatrix} 0.50 & 0.00 \\ 0.35 & 0.00 \end{bmatrix} \begin{bmatrix} \Delta x_{1t-1} \\ \Delta x_{2t-1} \end{bmatrix} + \begin{bmatrix} 0.50 \\ 0.15 \end{bmatrix} \begin{bmatrix} 1 & -2 \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \end{bmatrix} \begin{bmatrix} \varepsilon_{1t}+0.40 \\ \varepsilon_{2t}+1.00 \end{bmatrix} \quad (6.43)$$

In this case $\gamma_\perp'(I_n - B(1))\alpha_\perp = (0.15 \ -0.50)(0.50 \ 0.15)' = 0$, which implies $C(1) = 0$. The determinant of the corresponding VAR in levels (6.36) is now $(1-0.7L)(1-L)(1-L)$, and we have a $CI(2,1)$–system: Δx_{1t} and Δx_{2t} are $I(1)$ and $\Delta\alpha' x_t$ is $I(0)$. The mean growth has a trend, but $E\Delta\Delta x_t = 0.88/0.3$ and $E\Delta\Delta y_t = 0.44/0.3$ are constant. In fact Johansen (1992a) proved that a reduced rank of $s < k$ of the matrix in (6.40) is a necessary condition for the $I(2)$–ness of the components of the VAR. As far as estimation under this restriction is concerned Johansen (1992a) remarked that the two reduced rank conditions are not independent (compare the last terms in (6.41)), which "makes the analysis rather cumbersome". A simple strategy to avoid the cumbersome techniques is to test the $I(2)$–hypothesis for the component variables first (in the hope of rejecting!), before doing multivariate unit root tests. Influence analysis as suggested in chapter 3 can then be applied quite easily, whereas it would be costly for the iterative multivariate test for the $I(2)$–hypothesis. Example 6.3 illustrates that two unit roots in the determinant of the system cannot be interpreted as a sign of *noncointegration* per se. It does indicate here that no linear combination of the variables is stationary.

Next we consider a case where $\mathrm{sp}((I_n - B(1)) \subset \mathrm{sp}(\alpha)$ in

Example 6.4

$$\begin{bmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{bmatrix} = \begin{bmatrix} 0.00 & 2.00 \\ 0.50 & 0.00 \end{bmatrix} \begin{bmatrix} \Delta x_{1t-1} \\ \Delta x_{2t-1} \end{bmatrix} + \begin{bmatrix} -0.30 \\ 0.15 \end{bmatrix} \begin{bmatrix} 1 & -2 \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \end{bmatrix} \begin{bmatrix} \varepsilon_{1t}+0.40 \\ \varepsilon_{2t}+1.00 \end{bmatrix} \quad (6.44)$$

In this case $\gamma_\perp'\left[I_n - B(1)\right]\alpha_\perp = (0.15 \ 0.30)\left[(1 \ -0.50)'(1 \ -2)\right](1 \ 0.50)' = 0$. The determinant of the corresponding system (6.36) is now $(1+0.6L-L^2)(1-L)(1-L)$ which has one root inside, and two roots on the unit circle. The system is explosive. The last two examples clearly do not meet assumptions 6.1. To learn that (6.37) is not a sufficient condition for stability, modify γ in example 6.3 to get

Example 6.5

$$\begin{bmatrix}\Delta x_{1t}\\ \Delta x_{2t}\end{bmatrix}=\begin{bmatrix}0.50 & 0.00\\ 0.35 & 0.00\end{bmatrix}\begin{bmatrix}\Delta x_{1t-1}\\ \Delta x_{2t-1}\end{bmatrix}+\begin{bmatrix}-0.20\\ -0.50\end{bmatrix}\begin{bmatrix}1 & -2\end{bmatrix}\begin{bmatrix}x_{1t-1}\\ x_{2t-1}\end{bmatrix}\begin{bmatrix}\varepsilon_{1t}+0.40\\ \varepsilon_{2t}+1.00\end{bmatrix} \tag{6.45}$$

Here $\gamma_\perp'\left(I_n-B(1)\right)\alpha_\perp=(0.5\ \ -0.20)(0.5\ \ 0.15)'\neq 0$. Nevertheless the determinant of the corresponding system (6.36) is $(1-2.3L+0.86L^2)(1-L)$ which has one root inside the unit circle so that the system is still explosive. As we stated earlier in our discussion of univariate unit root tests in §A2.2.2.6 it is important to check the stability of the autoregression before the application of Johansen's LR test, see §6.6.1. Otherwise the observed high canonical correlation between first differences and levels in an explosive system could falsely be interpreted as a sign of cointegration. The same effect can occur if one neglects deterministic changes in mean growth, see Ooms and Van Dijk (1992).

6.4 Trend stationary processes and quadratic trends

An empirically relevant limitation of the set-up in the previous section, is that *trend stationary components*, i.e. components that are $I(0)$ around a deterministic trend, cannot be generated by (6.9). The stationarity of a component, say x_{it}, can only be established if $e_i'C(1)=0$, with e_i the $i-$th unit vector, see (6.28). The condition $e_i'C(1)=0$ implies also that $E\Delta x_{it}=e_i'C(1)m=0$ so that x_{it} cannot be both $I(0)$ and trending. We do not want to make the assumption of mean stationarity of the $I(0)$-combinations $\alpha_{.i}'x_t$ (as in assumptions 6.1) in this section. In order to generalize the set-up to include trend stationary components, or cointegrating vectors with a deterministic trend, one has to introduce a trending mean for the disturbances ε_t. If one does not impose restrictions on the coefficients of the trend, the variation in x_t can become dominated by quadratic trends.

Consider the following extended version of system (6.9):

$$(I-LA(L))\begin{bmatrix}S_1\Delta x_t\\ \alpha' x_t\end{bmatrix}=G(\varepsilon_t+m_0+m_1t). \tag{6.46}$$

This system is parameterized as a trend stationary VAR in the variables $S_1\Delta x_t$ and $\alpha'x_t$. It is thus equivalent to the system in (A4.2.1) where we derived the time varying mean of the components of such a VAR system. Note that (6.23) is not valid in this context. Using (A4.2.2) and the notation of this chapter[8] one obtains

---

[8] Substitute $((S_1\Delta x_t)'\ (\alpha'x_t)')$ for x_t, $(I\text{-}A(L))$ for $\phi(L)$, Gm_1, Gm_0 for m_1, m_0.

$$ES_1\Delta x_t = S_1\{(I-A(1))^{-1}Gm_1 t + (I-A(1))^{-1}(Gm_0 - \textstyle\sum_{i=1}^{p} iA_i(I-A(1))^{-1}Gm_1)\}$$

and

$$E\Delta\alpha' x_t = S_2(I-A(1))^{-1}Gm_1.$$

So instead of (6.23) one gets

$$ES_2\Delta x_t = \beta_{21}ES_1\Delta x_t + \beta_{22}E\Delta\alpha' x_t = \beta_{21}ES_1\Delta x_t + \beta_{22}S_2(I-A(1))^{-1}Gm_1,$$

so that by premultiplying with $G^{-1}S_1'$, noting (6.20) and rewriting (6.27) as

$$C(1) = G^{-1}S_1'S_1(I-A(1))^{-1}G,$$

one obtains

$$E\Delta x_t = C(1)(m_1 t + m_0) + G^{-1}S_1'S_1(I-A(1))^{-1}(-\textstyle\sum_{i=1}^{p} iA_i(I-A(1))^{-1}Gm_1)$$

$$+G^{-1}S_2'S_2(I-A(1))^{-1}Gm_1. \qquad (6.47)$$

One sees that x_t contains a quadratic trend if $C(1)m_1 \neq 0$. One can prove analogously with proposition 6.3 that $C(1)m_1 \neq 0 \Leftrightarrow m_1 \notin \mathrm{sp}(\gamma)$. This corresponds to the results in Johansen (1991b, eq. (2.3)). If we compare this analysis with the univariate model in §A2.2.2.3, one could call m_0 and m_1 reduced form parameters with no clear interpretation. Let $E\Delta x_t = \mu_0 + \mu_1 t$ be a short hand notation for (6.47), μ_0 and μ_1 can then be seen as structural parameters, where (6.47) establishes the rather complex relation between (m_0, m_1) and (μ_0, μ_1). The statistical analysis becomes complex as well, as one can see below. Johansen (1991b) discussed five cases. We reproduce these cases in table 6.1 for later use.

Table 6.1 *Deterministic terms and "structural" interpretation*

		Trends, means	
case	Restrictions on m_0, m_1	In components	In "cointegrating" vectors
i	$0 \neq m_0 \notin \mathrm{sp}(\gamma) \ \wedge 0 \neq m_1 \notin \mathrm{sp}(\gamma)$	quadratic	linear
ii	$0 \neq m_0 \notin \mathrm{sp}(\gamma) \ \wedge 0 \neq m_1 \in \mathrm{sp}(\gamma)$	linear	linear
iii	$0 \neq m_0 \notin \mathrm{sp}(\gamma) \ \wedge m_1 = 0$	linear	constant
iv	$0 \neq m_0 \in \mathrm{sp}(\gamma) \ \wedge m_1 = 0$	constant	constant
v	$m_0 = 0 \ \wedge m_1 = 0$	zero	zero

NOTES: quadratic and linear concern trends, constant and zero concern means. Model is (6.46), notation is from §6.3.1.

Johansen developed the statistical theory for all these models over a period of five years. In our simple framework where only cointegrated systems of orders (1,1) are considered one can already distinguish between $(n-1)\cdot 5$

types of restrictions on the VAR! All of these restricted models can be interesting, either as a null or as an alternative hypothesis. If one is only interested in the stochastic part of the model, the deterministic terms are nuisance parameters. Those are still important if one wants to do some statistical tests. Johansen (1992c) showed that LR tests on unit roots in the the stochastic part of the dynamics are (asymptotically) nonsimilar to the configuration of the deterministic terms. We discuss statistical tests below.

Example 6.5

Here is an example of a trend stationary cointegrating vector in a bivariate VAR, $m_1 \in \mathrm{sp}(\gamma)$ so that no component contains a quadratic trend. We are thus in case *ii*. The "cointegrating vector" is trivial. This makes computation of (6.47) simple. Since $\beta_{21}=0$, only the last line of (6.47) is needed to compute $E\Delta x_{2t}$. Component x_{1t} is $I(1)$, x_{2t} is $I(0)$ but nonstationary in mean.

$$\begin{bmatrix}\Delta x_{1t}\\ \Delta x_{2t}\end{bmatrix} = \begin{bmatrix}0.50 & 0.00\\ 0.35 & 0.00\end{bmatrix}\begin{bmatrix}\Delta x_{1t-1}\\ \Delta x_{2t-1}\end{bmatrix} + \begin{bmatrix}-0.30\\ 0.15\end{bmatrix}\begin{bmatrix}0 & 1.00\end{bmatrix}\begin{bmatrix}x_{1t-1}\\ x_{2t-1}\end{bmatrix} + \begin{bmatrix}\varepsilon_{1t}+0.40-0.20t\\ \varepsilon_{2t}+1.00+0.10t\end{bmatrix},$$

$$G=G^{-1}=I_2, B(1)=\begin{bmatrix}0.50 & 0.00\\ 0.35 & 0.00\end{bmatrix}, C(1)=\begin{bmatrix}-5.0 & -10.0\\ 0.0 & 0.0\end{bmatrix}, A(1)=\begin{bmatrix}0.50 & -0.30\\ 0.35 & 1.15\end{bmatrix},$$

$$E\begin{bmatrix}\Delta x_{1t}\\ x_{2t}\end{bmatrix} = \begin{bmatrix}-0.50 & -0.30\\ 0.35 & -0.15\end{bmatrix}^{-1}\cdot\left(\begin{bmatrix}-0.2t\\ 0.1t\end{bmatrix} + \begin{bmatrix}0.4\\ 1.0\end{bmatrix} - \begin{bmatrix}0.50 & -0.30\\ 0.35 & 1.15\end{bmatrix}\begin{bmatrix}-0.50 & -0.30\\ 0.35 & -0.15\end{bmatrix}^{-1}\begin{bmatrix}-0.2\\ 0.1\end{bmatrix}\right)$$

$$= \begin{bmatrix}0\\ -{}^2/_3 t\end{bmatrix} + \begin{bmatrix}-18^2/_3\\ 31^7/_9\end{bmatrix}, \text{ so that } E\Delta x_{2t} = -{}^2/_3.$$

6.5 Estimating pushing trends and pulling equilibria

6.5.1 Deterministic trends

The deterministic part of the common trends can be estimated by a simple multivariate static regression of x_t on a constant and a trend. The residual is then used as an estimate of the stationary zero mean variable y_t that can be used to analyze the stochastic part of the trend in a second stage. King et al. (1991) followed this popular approach. In small samples regression on a trend has marked influence on the residual autocorrelations (cf. Anderson (1971, p.614)), which may be undesirable.

A second approach is to perform univariate unit root tests for all variables as a first step. Next impose unit roots that are not rejected and estimate the implicit univariate trends by a simple static regression on a

constant or by computing the time varying mean as in §A4.2.2. Estimating a trend by regression of the first differences on a constant leads to the simple estimator $(y_T - y_1)/(T-1)$, which is obviously very robust to outliers in all but the first and the last observation. The influence of the different observations on the estimates in a regression of the levels on a constant and a trend is more evenly distributed. Still, observations far away from the sample mean have a bigger influence. If one wants to avoid the risks associated with this improper detrending in small samples, a rather low significance level of the unit root tests is in order. In subsequent (multivariate) analysis, notably Granger noncausality analysis, several authors showed the effects of (im)proper detrending to be rather important (Kang (1985), Ohanian (1991)).

A third approach is to apply multivariate unit root tests in a VAR with an appropriate number of deterministic regressors, to impose the unit roots and to estimate the implicit multivariate trend in (6.47). In this way the deterministic and the stochastic part of the trend are estimated simultaneously. In a simultaneous method, like the maximum likelihood procedure one can impose the restriction that the estimator of the cointegrating vector is not only $I(0)$, but mean stationary as well, i.e. not trending. Such a cointegrating vector removes the stochastic and the deterministic part of the trend. This extra restriction of *deterministic cointegration* makes sense in some economic applications, see e.g. Ogaki (1992, §3).

A fourth approach is to regress the first principal component of the multiple series x_t on a trend. Asymptotically the empirical variance of the deterministic part grows proportionally to T^2, whereas the variance of stochastic $I(1)$–part only grows proportionally to the sample size T. One expects that the first principal component is dominated by the deterministic part if the mean growth is not too small compared to the variance of the innovations in the permanent component of the series. In many real macroeconomic time series this mean growth is considerable.

In choosing among these four methods one has the usual problem of weighting bias and efficiency of the different procedures. The multivariate method is only preferable if one is convinced of "correct specification" and if the number of observations is sufficiently large.

6.5.2 Estimating the stochastic part of the trend

After one has removed the deterministic trend, the *stochastic trends* or $I(1)$–components are likely to dominate the variation in the series, provided

they are "big" in the sense of Cochrane (1988), i.e. if the $C(1)$ matrix in (6.28) is nonnegligible. In this section we assume this is the case.

Maximum variance approach

The components with the maximum variance in a principal component analysis will be highly correlated with the stochastic trends in the system. The principal component approach, used by Stock and Watson (1988) as part of their testing procedure does not take advantage of the information in the autocorrelations and is not *scale invariant.* The statistical interpretation of the orthogonal components is straightforward. Naturally the estimates are not sensitive to dynamic misspecification. When changes in the variables have the same comparable dimension, as is the case for e.g. growth rates, interest rates, or exchange rates, the *scale problem* might not be too serious, or easy to circumvent, see e.g. Koedijk and Schotman (1989). The application in the analysis of the different quarters of a seasonal series can also serve as an example, where scale problems are not so serious, see §4.4.

Predictability approach

A rather different approach, which takes account of possible autocorrelation and which avoids the scale problem is the "canonical" analysis of Box and Tiao (1977). In this approach independent linear combinations of the variables are selected and ordered according to their relative one step ahead forecast error variance. Let $f_i'y$ denote such linear combination. Box and Tiao defined the quantity λ_i as a measure of the *predictability*[9] of a linear combination of the levels of a series from the past by

$$\lambda_i = \hat{\sigma}^2_{f_i'y|-1} / \hat{\sigma}^2_{f_i'y}, \tag{6.48}$$

where $f_i'y|-1$ is the expectation of $f_i'y$ conditional on past history up to time $t-1$, and $\hat{\sigma}^2_z$ denotes the sample variance of z. In the VAR model they used λ_i simply corresponds to the multiple correlation coefficient R^2 of the implicit equation for $f_i'y$.

If the dynamic model is linear (and Gaussian), the expected value $f_i'y|-1$ is the minimal mean squared error one step ahead forecast of $f_i'y$. The predictability measures λ_i of the linear combinations $f_i'y$ are computed as the eigenvalues and corresponding n linearly independent eigenvectors of $\hat{\Gamma}_0(y)^{-1}\cdot(\hat{\Gamma}_{0|-1}(y))$, with $\hat{\Gamma}_0(y)$ the sample variance/covariance matrix of y_t and $\hat{\Gamma}_{0|-1}(y)$ the sample variance/covariance matrix of one step ahead predictions

[9] See Jewell and Bloomfield (1983) for standard definitions of this concept in the time series literature.

of y. If y follows an AR(1) process[10] $y_t = -\Phi_1 y_{t-1} + \varepsilon_t$, one has $\hat{\Gamma}_{0|-1}(y) = \hat{\Phi}_1 \hat{\Gamma}_0(y) \hat{\Phi}_1'$, and the λ_i–values are just the squared *canonical correlations*[11] of y_t and y_{t-1}, i.e. the eigenvalues of

$$Q_l = (Y'Y)^{-1}(Y'Y_{-l})(Y_{-l}'Y_{-l})^{-1}(Y_{-l}'Y), \tag{6.49}$$

where $l = 1$, $Y = [y_{1+l}\, y_{2+l} \cdots y_T]'$, and $Y_{-l} = [y_1\, y_2 \cdots y_{T-l}]'$.

The eigenvectors corresponding to the low eigenvalues yield the most *white noise like* linear combinations. The eigenvectors corresponding to the large eigenvalues yield the (nearly) nonstationary ones.

This analysis was originally developed for (nearly non) stationary series. The matrix Q_1 is just the product of the transposed coefficient matrices of the second and first multivariate regressions of y_t and y_{t-1}, $-\hat{\Phi}_1^{*\prime}$ and $-\hat{\Phi}_1'$: $Q_1 = \hat{\Phi}_1^{*\prime}\hat{\Phi}_1'$. This means one can use results derived for OLS estimators of AR parameters in nonstationary models here, see e.g. Tiao and Tsay (1989, p. 172). One can successfully apply the canonical correlation analysis on nonstationary series as well.

Fountis and Dickey (1989, §3) studied the case where the number of common trends varies from 0 to 1 in an n–variate autoregressive system. They suggested the one left eigenvector corresponding to the left null space of $\Phi(1)$, i.e. $I + \hat{\Phi}_1$ in an AR(1) process, as an estimate of what they called the "canonical unit root process". Fountis and Dickey also noted the asymptotic equivalence of the estimator of the eigenvector with eigenvalue closest to zero in their analysis and the largest eigenvalue of $T^{-2}Y'Y$, which is equivalent to the largest principal component in the maximum variance approach. Fountis and Dickey called their estimator "an extension of Box and Tiao's (1977) approach for the stationary case." Ahn and Reinsel (1990) also extended Box and Tiao's procedure for nonstationarities at frequency zero. Both papers supplied a test for unit root nonstationarity as well.

The predictability interpretation of the parameters of the VAR has a straightforward intuitive appeal. The interpretation does not crucially depend

---

[10] If one allows for a singular covariance matrix of the disturbances, each finite order AR(p) model can of course be rewritten as an AR(1) in companion form. One then obtains $n(p-1)$ unit canonical correlations belonging to the extra definition equations $y_{t-i} = y_{t-i}+0$, $i=2,3,\ldots,p$. See Fountis and Dickey (1989) for an application in the multivariate unit root literature. The λ_is are then taken as the R^2s of the equations with positive residual variance, see (6.50) below.

[11] In econometrics squared canonical correlations are used in definitions of *generalized correlation coefficients*, which measure the extent to which systematic relationships explain the fluctuation in the set of all jointly dependent variables, cf. Hooper (1959).

on the $I(1)$–assumption of the underlying series. This is an advantage over the Vector Error Correction and Common Trends interpretations, which need the $I(1)$–assumption.

In economic time series it is not only relevant to distinguish between stationary and nonstationary components, or to separate components with a large random walk and components with a small random walk. The distinction between predictable and unpredictable components may be just as interesting. One often examines predictability in the stock market. Predictability in excess returns is interpreted as a sign of either the "inefficiency" of markets or the "irrationality" of the agents, see, e.g, Baillie (1989).

It has been argued for 30 years that one should add a(n important) nonlinear component to the model in order to make the statistical analysis useful for this end. See Lo (1991) for a list of references. Using the test statistic in (2.30) Lo did not find clear statistical evidence of "long run nonlinearity in mean" in U.S. stock prices, though.

One sees easily that the limit of λ for $|\rho| \to 1$ of $f_i' y_t = e_t/(1-\rho L)$ equals 1, whereas the corresponding limit for $\rho \to 0$ is zero. Components for which $\lambda \to 1$, coincide with the *common trends*. Note that one can have $\lambda \to 1$ for periodic predictable series as well, take e.g. $\rho \to -1$. Well known U.S. economists sometimes neglected this fact in discussions on unit root like behavior, see Christiano and Eichenbaum (1990, p. 37, eq. (13)). It can be important in quarterly series with $I_{\frac{1}{2}}(1)$ or $I_{\frac{1}{4}}(1)$–like serial correlation, i.e. series with *seasonal integration*. Examples are $y_t = -y_{t-1} + \varepsilon_t$ and $y_t = -y_{t-2} + \varepsilon_t$. The Box–Tiao procedure can be used to find the most predictable seasonal trends at a particular seasonal frequency only if the nonstationarity at other frequencies has been removed first. Here we do not consider these *seasonal common trends*.

Components for which $\lambda \to 0$ coincide with relationships between the original variables of a "static" form. If the original series are (nearly) nonstationary these components are special cases of cointegrating vectors, or *equilibrium relations*. Note that this corresponds to an R^2 of zero for the last r equations in a CICS model[12] (6.17). In this respect the canonical analysis of Box and Tiao was "cointegration" analysis "avant la lettre."[13] If simple static forms of cointegrating relations exist, it follows that the parameterizations of the VAR can be simplified. Canonical correlation techniques are often applied in the search for so–called *parsimonious*

[12] In an AR(1), where $B(L)=0$, this can occur if $\alpha'\gamma = -I_2$.

[13] Their analysis also contained a forecast error variance decomposition, one of the key elements of modern time series analysis of macroeconomic data, see Sims (1980, §C).

parameterizations with fewer parameters than the unrestricted VAR. Tiao and Tsay (1989) contains a comprehensive discussion, cf. §7.5.4 below.

The canonical correlations derived from (6.49) can be seen as multivariate generalizations of the squared ordinary sample autocorrelations of a series. Gourieroux and Peaucelle (1989) used these correlations in the context of a VMA model to detect simplifying structures. They introduced the term *codependence* at lag l to indicate cases where $Q_l \neq 0$ *and* $\lambda_i = 0$ for some i at lag l: The autocorrelation for the corresponding linear combination of the original variables "cuts off" at an earlier point than the multivariate autocorrelation. Kugler and Schwendener (1990) derived tests for codependence using estimates of VAR models. Ahn and Reinsel (1988) used the concept of a *multivariate partial autocorrelation function.*

Tiao and Tsay (1989) presented a related analysis in VARMA models and introduced the term Scalar Component Models. They employed tests for $\lambda = 0$ which have standard χ^2 limit distributions. Below we discuss a test for $\lambda = 1$. As we know multivariate *unit root tests* often have nonstandard limit distributions.

Cointegration approaches

A third indirect method to compute the stochastic part of the trend uses cointegration theory. An estimate of the common trends representation (6.28) can be derived from the estimates of a VECM model and its disturbances, provided one first obtained estimates of the cointegrating vectors α. This implies that every method to estimate cointegrating vectors in a $CI(1,1)$ system simultaneously produces estimates of the common trends. These results do not always generalize easily to cointegrated systems of other orders. Gourieroux et al. (1987, corollary .23) showed for example that the $I(1)$–assumption is essential if one wants to estimate the "short run" parameters by OLS conditional on the first stage OLS estimate of α in so called *cointegrating regression*, introduced in §6.5.3. They proved that the OLS estimators of these parameters are only consistent if y_t is integrated of order d, with $3/4 < d < 5/4$. The choice of estimator for a cointegration vector must thus be done carefully. It is already difficult to check the assumption of $I(1)$ versus $I(0)$, let alone to check integration of order $d > 3/4$ versus $d < 3/4$.

6.5.3 Estimating pulling equilibria

In the previous section we saw that the estimation of trends and equilibria in the $CI(1,1)$ model often occur simultaneously. This holds for other methods as well.

Minimum variance approach, principal components

In the principal component approach the "last" principal components with the smallest variance can be considered as reasonable estimates of equilibrium relations.

Nonpredictability approach, another connection with cointegration

In the canonical decomposition method of Box and Tiao the least predictable, white noise like, components can be interpreted as equilibrium relations as well. One can extend their method to other forecast horizons, so as to make it even better comparable with cointegration analysis. By definition stationary combinations have a constant mean. The forecast of a stationary variable, conditional on a fixed information set converges to its mean as the forecast horizon recedes. This entails that the analysis of Box and Tiao of (infinitely) long series applied to longer forecast horizons will ultimately lead to components which have predictability zero at an infinite horizon. All cointegrating vectors have predictability zero, the common trends have predictability one at longer horizons.

There are several ways to generalize the analysis of Box and Tiao. Take the AR(1) case as an example. The generalization we suggest here is to do a canonical correlation analysis of y_t and y_{t-l} using Q_l. One then effectively uses a so–called "shifted" VAR model, $y_t = -\Phi_{1l} y_{t-1-l} + \varepsilon^*_{lt}$ by OLS. Kool (1989) studied the stationary version of this model in extensive detail. The coefficient matrices $(-\hat{\Phi}_1)^l$ correspond to the parameters of the MA representation of the process, which are at the heart of impulse response analysis and variance decompositions. In a shifted VAR model those coefficient matrices are estimated directly. It is known under quite general conditions that the largest eigenvalues of the OLS estimators $\hat{\Phi}_1$ are consistent estimators of the unit root eigenvalues of the system, see Tsay and Tiao (1990, Theorem 6.2). The most important necessary condition is that the order of the estimated AR model is large enough. The same holds for the OLS estimators of the shifted system or the reverse system[14] $y_t = -\Phi^*_1 y_{t+1} + \varepsilon^*_t$. In the AR(1) case the forecast of y_t into the far future, $\lim_{l\to\infty} y_{t+l|t}$, equals $\lim_{l\to\infty}(-\Phi_1)^l y_t = C(1)\xi_t$, the common trend term in the Beveridge–Nelson (1981) decomposition of y_t, see (6.6). By assumption the eigenvalues of Φ_1 are either on or inside the unit circle. Since we do not consider seasonal unit roots they are either *one* or inside the unit circle. The eigenvalues of $(\Phi_1)^{\infty}$ are thus only zero or one. The eigenvectors corresponding to the zero eigenvalues

[14] The nonstationary eigenvalues of the reverse system are the inverses of the corresponding eigenvalues of the standard system. For roots 1 they are equal.

of Q_∞' are equivalent to those of the vectors of the left null space of the long run variance $C(1)\Sigma C(1)'$ and thus to the vectors of the kernel of $C(1)'$. They are cointegrating vectors α of the system. The eigenvectors corresponding to the unit eigenvalues give the linear combinations of the components that generate random walks.

Here we do not consider trivial unit canonical correlations introduced by deterministic terms or by rewriting a higher order AR model (A4.2.1) as an AR(1) model in companion form as in

$$\begin{bmatrix} x_t \\ x_{t-1} \\ t \\ 1 \end{bmatrix} = \begin{bmatrix} -\Phi_1 & -\Phi_2 & m_1 & (m_1+m_0) \\ I & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ x_{t-2} \\ t-1 \\ 1 \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{6.50}$$

One can plot the λ_i–values, the ordered predictabilities of the linear combination of the components of y_t, as a function of the prediction horizon l. To paraphrase Cochrane (1988): "One can tell how big the noncointegration in the series is." From a data analytic point of view the use of (6.49) is interesting, since it brings higher order sample autocorrelations into play, which are neglected in other approaches to estimate equilibria. It seems odd not to use long run autocorrelations to estimate long run equilibria.

Another way to generalize the Box–Tiao method to estimate cointegrating vectors is to use *multi step ahead* predictions. For an AR(1) model this would require the computation of the eigenvalues and eigenvectors of

$$Q_l^* = (Y'Y)^{-1}(\hat{\Phi}_1^l Y_{-l}' Y_{-l} \hat{\Phi}_1^{l\prime}). \tag{6.51}$$

This is not a true canonical correlation matrix for $l>1$, so that predictabilities derived from (6.51) can turn out to be greater than one. For non explosive models one would expect the eigenvalues of Q_l^* to be close to or smaller than one for reasonably large l. One can conjecture that the distribution of the eigenvalues of Q_l^* is not as easy to use as the distribution of the eigenvalues of Q_l. The nonlinear transformation of the OLS matrix can lead to occasional very large eigenvalues for large l. We find the eigenvalues of (6.49) easier to interpret. We suggest a simple degrees of freedom adjustment in the application in §7.5.2.

Cointegration approaches and regression approaches

If the equilibria are of more immediate interest than the trends, one should try to estimate the cointegrating vectors directly. Application of *maximum likelihood* in the VAR model with Gaussian disturbances boils down to the maximization of the partial squared canonical correlation between levels

and first differences. Johansen (1992d) developed this approach in great detail and supplied the corresponding statistical analysis of the estimates as well. We discuss this approach in connection with unit root tests in §6.6.

Regression estimation, which one could call a "*directed*[15] *minimum variance approach*", has the advantage that it is applicable to a wider range of models than only the VAR. It has the important disadvantage that one should know beforehand that the regressors y_{2t} are not cointegrated, in particular that they are $I(1)$.

One can interpret the procedure in the context of the CICS model of §6.3. One normalizes the matrix of cointegrating vectors α with a hypothesized rank r so that $\alpha_{22}=I_2$. This can be achieved by rearranging the original variables in an appropriate way, as we showed in example 6.2. One subsequently carries out the multivariate regression of $(S_2y_t)'$ on $(S_1y_t)'$ (and a constant) to get an estimate of α_{21}, cf. (6.17):

$$y_t'S_2' = -y_t'S_1'\alpha_{21}' + m'\alpha + \zeta_t', \tag{6.52}$$

with $\zeta_t = S_2A(L)(S_1'S_1\Delta y_{t-1} + S_2'\alpha' y_t) + \alpha'\varepsilon_t$, which is $I(0)$ by assumption.

This estimator is (super)consistent as long as ζ_t is $I(d)$, with $d<1$, see Cheung and Lai (1993, p. 106). The choice of the regressands is not data based. The resulting estimates for the cointegrating space spanned by the column vectors of $(\hat{\alpha}_{21}\ -I_2)'$ depend on this choice. The identification of the parameters must be checked beforehand. One must choose the regressands so that α_{22} has full rank r, or, equivalently, so that the appropriately normalized moment matrix of the regressors, i.e. $\Sigma_{00} = \lim_{T\to\infty} T^{-2}\sum y_t y_t'$, has full rank k[16]. In other words: the regressors may not be cointegrated.

In traditional econometric models one can incorporate reasonable exogeneity assumptions about y_t such as the independence of S_1y_t and $S_2\varepsilon_t$ at different time lags. Correctly imposing these restrictions can improve the *asymptotic* efficiency of the estimators of α_{21}. Phillips (1991) provided an extensive discussion.

One way is to extend the right hand side of regression equations (6.52) with *present* and *lagged differences* of S_1y_t and S_2y_t, and estimate the equation by 2SLS with *lagged levels* of y_t as instrumental variables for the present differences of y_t. One then directly estimates the so–called Bewley

---

[15] Hillier (1990) analyzed *direction estimators* in an econometric context.

[16] In the notation of Engle and Granger (1987) we choose $q=\text{vec}(S_2')$, $Q=I_2\otimes S_1'$. Their identification condition then reads $Q'(I_2\otimes\Sigma_{00})Q=I_2\otimes S_1\Sigma_{00}S_1'$ has full rank $rk \Leftrightarrow S_1\Sigma_{00}S_1'$ has full rank k.

representation, see Hylleberg and Mizon (1989). The resulting estimators of α_{21} in the just identified equation are equal to the indirect least squares estimators for α_{21} in an equation where only lagged differences are added. The latter approach can be applied to $I(0)$–processes as well if one modifies the notion of equilibrium relation, see e.g. Alogoskoufis and Smith (1990). These simple estimators, which were often used to measure "long run equilibrium responses" in single equation analysis, unfortunately do not have *finite sample moments* in the standard case with Gaussian disturbances, see Mariano (1982). This may partly explain why estimates across comparable data sets are so disparate, see Bewley and Fiebig (1990).

If one only assumes instantaneous independence of S_1y_t and $S_2\varepsilon_t$ at lag 0, one can improve the efficiency of single equation estimators by adding *leads* of the differenced series as well, cf. Phillips and Loretan (1991a), King et al. (1991), Saikkonen (1991). Campbell and Perron (1991) reproduced some key formulae. These approaches were also generalized to produce confidence intervals for the normalized coefficients of the cointegrating vectors. Perron and Campbell (1992, §4.3) discussed the identification issues in the application of significance tests in likelihood ratio procedures of Johansen (1991a) and Ahn and Reinsel (1990).

Phillips and Loretan (1991a) reviewed a number of single equation methods to estimate equilibrium relations. They used asymptotic theory and Monte Carlo simulations in the evaluation. Some of these methods involve nonlinear optimization.

Boswijk (1992, chapters 2,3,6) discussed a number of results in this field in detail[17]. He concentrated on the case with Gaussian errors where $(S_2 - S_2\Sigma S_1'(S_1\Sigma S_1')^{-1}S_1)\varepsilon_t$ is an *innovation process* relative to the present and past of S_1y_t and $\gamma_1 = 0$, so that S_1y_t is *weakly exogenous* in the sense of Engle et al. (1983) for the parameter vectors γ_2 and α, as proved by Johansen (1992b). For the Gaussian VECM model this implies $S_2\Sigma S_1' = 0_{r\times k}$, i.e. independence of $S_1\varepsilon_t$ and $S_2\varepsilon_t$. For the CICS model this implies $A_{12}(1) = 0_{k\times r}$. The assumption on the disturbances is maintained and can be used to identify the shocks. The restriction on γ is overidentifying and can be tested.

Frequency domain approach

An alternative way to estimate the space of cointegrating vectors is to compute the left null space of the spectrum of the first differences of the series at the frequency zero, see §6.2.2. This approach also enables one to use the information in higher order sample autocorrelations. Many

[17] He used x_t, y_t, z_t, β, and α_2 for y_t, S_1y_t, S_2y_t, α, and γ_1.

semi–parametric alternatives exist to estimate the spectrum at the zero frequency. Phillips and Ouliaris (1988, 1990) used this approach, which also takes a more direct look at the long run correlations.

These methods are all interesting alternatives to estimate pulling equilibria next to the application of maximum likelihood in the Gaussian VAR model. Bewley et al. (1991) compared the finite sample efficiency of the Box–Tiao and Johansen approach in a first order VECM model. The Box–Tiao procedure seemed to come out on top.

An advantage of the minimum variance procedure of Stock and Watson and the minimum predictability procedure of Box and Tiao over the maximum likelihood procedure of Johansen[18] is the orthogonality of the estimated trend components. It gives a data based orthogonalization.

6.6 Multivariate tests for unit roots

Over the last five years the number of tests for noncointegration has grown steadily. Gregory (1991) surveyed fourteen of them in the context of a simple bivariate model with an exogenous variable. Other tests can be found in Boswijk (1992) and Fountis and Dickey (1989). A number of these tests are univariate *tests on residuals from cointegrating regressions* or other estimates of deviations from "equilibrium relations". The computation of these test statistics follows the lines of §A2.2.2.4.

Phillips and Ouliaris (1990, Appendix B) computed asymptotic similar critical regions for a number of these tests. They discussed cases where $r=0$ under the null and $r=1$ under the alternative. MacKinnon (1991) estimated number of observations dependent critical values for finite samples for three significance levels.

The cointegrating regression approach for testing multivariate unit roots is only of practical interest if there are $n-1$ exogenous variables which are uncorrelated with the disturbances at all lags, or if all cointegrating regressions are independent, each having a credible interpretation.

If one does not want to order the variables a priori in exogenous and endogenous variables of any kind, multivariate unit root tests are in order. Most univariate test statistics mentioned in §A2.2.4 have multivariate analogs.

6.6.1 Models with p=1 and zero mean

The most straightforward statistics to use in multivariate unit root

[18] If applied to an AR model of lag order p>1.

testing are the eigenvalues of the OLS estimate of the coefficient matrix $(-\Phi_1)$ in a multivariate AR(1) regression:

$$y_t = (-\Phi_1)y_{t-1} + v_t. \tag{6.53}$$

If the number of eigenvalues far from one is higher than the hypothesized number, the null hypothesis must be rejected. In the maximum likelihood approach proposed by Johansen one uses a multivariate analog of the Dickey–Fuller regression, see also §A2.2.2.6:

$$\Delta y_t = (-\Phi_1 - I)y_{t-1} + v_t, \tag{6.54}$$

combined with the second auxiliary regression

$$y_{t-1} = (-\Phi_1 - I)^* \Delta y_t + v_t^*. \tag{6.55}$$

The eigenvalues of the product of the OLS estimates of $(-\Phi_1 - I)^{*\prime}$ and $(-\Phi_1 - I)'$, i.e. the *squared canonical correlations* of Δy_t and y_{t-1}, are used to test the cointegrating rank r of the system. A number of significant canonical correlations higher than the hypothesized rank r indicates a rejection of the corresponding hypothesis of $n-r$ unit roots in the system.

Note that in this approach eigenvalues of $(-\Phi_1 - I) > 0$, which are usually seen as an indication that the estimated system is explosive, result in high canonical correlations as well! The tables belonging to the Stock–Watson test indicate that this possibility of an estimated explosive root is negligible if the variables are demeaned and all roots of the system are stable or one, see appendix A6.1. Johansen derived maximum likelihood estimators for the cointegrating vectors in a VECM model with Gaussian disturbances and fixed initial values. To obtain these vectors one computes the eigenvectors of the following generalized eigenvalue problem[19] to solve for λ in

$$\left|\lambda Y_{-1}'Y_{-1} - Y_{-1}'(Y - Y_{-1})((Y - Y_{-1})'(Y - Y_{-1}))^{-1}(Y - Y_{-1})'Y_{-1}\right| = 0. \tag{6.56}$$

Note that the squared canonical correlations and canonical variates are solutions of this problem. Johansen (1988) chose the canonical correlation normalization of the eigenvectors E so that $E'Y_{-1}'Y_{-1}E = I_n$.

The statistics of both methods are properly scaled by a multiplication by the sample size T, just as the "normalized bias" tests in the univariate unit root tests.

---

[19] Here the determinantal expression can be simplified by premultiplication with $(Y_{-1}'Y_{-1})^{-1}$. Asymptotically $(Y_{-1}'Y_{-1})$ does not have full rank. The principal component procedure of Stock and Watson uses this property by selecting eigenvectors of $Y'Y$ with small eigenvalues. In generalizations of the Johansen procedure Y_{-1} is sometimes extended with a constant or a trend, see §A6.1.1.

A third method is to examine the long run covariance matrix V of Δy_t, i.e. 2π times its spectrum at the frequency zero, for singularities. Phillips and Ouliaris suggested time domain (weighted autocovariance) methods to estimate V:

$$\hat{V} = T^{-1}\sum_1^T \hat{v}_t\hat{v}_t' + T^{-1}\sum_{s=1}^{l} w_{sl}\sum_{t=s+1}^{T}(\hat{v}_t\hat{v}_{t-s}' + \hat{v}_{t-s}\hat{v}_t'), \tag{6.57}$$

with $w_{sl} = 1 - s/(l+1)$, i.e. the Bartlett estimator.

Gregory (1991) discussed alternative estimators, like the Parzen estimator. The *trace test statistic* for which Phillips and Ouliaris developed asymptotic statistical theory reads

$$P_y = T\,\mathrm{tr}[\hat{V}T(Y'Y)^{-1}]. \tag{6.58}$$

P_y can also be seen as a multivariate analog of a *variance ratio*[20] *test*: For[21] $l = T-1$ and $v_t = \Delta y_t$, $T\hat{V}_T$ is the multivariate analog of the variance time function of Δy_t at T, which we defined in (2.21). It should not be surprising that one needs to use the residuals of a first order autoregression rather than the first differences themselves in the construction of P_y to get a *consistent* test. Otherwise P_y would become a complicated estimator of a unit matrix. Phillips and Ouliaris (1990) noted[22]: "$t\hat{V}$ is a direct estimate of the covariance matrix of y_t, while $T^{-1}(Y'Y)$ is simply the observed sample moment matrix." We apply univariate techniques to estimate scalar V's in §7.4.7.3.

The correspondence of Q_∞^* and P_y is striking. In an AR(1) model the term $T^{-1}\hat{\Phi}_1^l Y_{-1}'Y_{-1}\hat{\Phi}_1^{l\prime}$ should become proportional to $T^{-1}\sum_t C(1)\xi_t\xi_t'C(1)'$ as $l, T\to\infty$, and $l/T\to 0$. It is just another estimator of the variance matrix of the long run forecasts of y_t.

6.6.2 Deterministic terms and serial correlation in AR(1) residuals

The asymptotic distribution of the statistics of Stock and Watson and Johansen depend on the presence of a nonzero or possibly trending mean and on the presence of serial correlation in v_t. We discussed different univariate approaches to this problem in §A2.2.4. Most univariate extensions that make unit root tests similar to nuisance parameters characterizing the serial

---

[20] Homonymous with VR_i in §A3.2, which is used to test for parameter stability of the variance of the disturbances.

[21] Phillips and Ouliaris assumed $l=o(T)$. In order for a useful test to appear l must not be chosen too large. $l=[T^{0.9}]$ is most probably too large.

[22] They used a different notation.

correlation have been generalized to the multivariate case. Stock and Watson (1988) generalized both the normalized bias tests of Phillips and Perron (1988) and Dickey et al. (1984). This made their approach useful for ARMA models as well. We discuss the implementation of their "filtering" approach in more detail in appendix 6.1.

Johansen generalized the augmented likelihood ratio tests of Dickey and Fuller. For this test one simply extends the sets of regressors of (6.54) and (6.55) with a sufficient number of lags of Δy_t. Depending on the cases $i-v$ discussed in §6.4 one also extends the set of regressands in (6.55) y_{t-1} or the set of regressors with a constant and a trend. The asymptotic null distributions of the resulting test statistics differ from case to case, which makes it difficult to develop a sensible testing strategy. We list the test statistics and some corresponding asymptotic distributions in appendix 6.1. Johansen (1992c) discussed a "specific to general" test procedure to deal with the different distributions of the test statistics under null hypotheses of interest. Johansen developed the procedure for VAR models with known order p, but it can be expected to work for ARMA models as well if one extends the arguments of Said and Dickey (1984) to the multivariate case.

The generalization of the test of Phillips and Perron (1988) is asymptotically robust to some forms of heteroskedasticity. The trace test of Phillips and Ouliaris is as well. This robustness of validity depends crucially on the robustness of the estimator for the spectrum of v_t at the frequency zero. In finite samples the quality of the Bartlett estimator of (6.57), which use was propagated in econometrics by Newey and West (1987), is not always too impressive.

Finite sample validity

These robust estimators have already been shown to be inferior in robustness of validity to the augmentation or filtering approaches under null hypotheses with simple univariate ARMA processes for v_t. Those essentially provide parametric alternatives to estimate the nuisance parameter[23] $\hat{V}$. One cannot recommend these robust estimators in small samples.

The number of Monte Carlo studies on the finite sample performance of the different multivariate unit root tests is growing steadily. The development of degrees of freedom corrections, which have been so successful to improve the finite sample performance of many asymptotically valid variable addition tests, is still in its infancy. It is probably a sensible procedure to replace the normalization factor T by the asymptotically equivalent factor

[23] In §A2.2.2.1 and A2.2.2.4 V was approximated by $c^2\sigma_\varepsilon^2$. In (2.30) $\hat{\sigma}^2$ denoted $\hat{V}$.

$T^* = (T - n(p-1))$, see Reinsel and Ahn (1992). This results in asymptotically equivalent procedures if the lag length p is $o(T)$. In the Johansen procedure the normalization of the eigenvalues λ_i by T is preceded by a transformation $\ln(1+\lambda_i)$ in order to make the test statistic equivalent to the likelihood ratio test statistic, see the appendix for an example. For many critical values for λ_i this approximation is very poor even in relatively large samples. We therefore prefer the use of $T^* \cdot \lambda_i$ instead of $T \cdot ln(1+\lambda_i)$.

In §7.4.7.6 one can see in that these corrections matter in a "typical" macroeconomic application. The outcomes of the tests thus have to be interpreted with great care until a deeper understanding of the finite sample behavior of the tests has been established. Reimers (1991b) compared the three multivariate procedures in an extensive Monte Carlo simulation study with VAR designs. The degrees of freedom correction worked. Reimers' implementation[24] of the Phillips–Ouliaris (1988) test procedure did not.

First and foremost one has to check the crucial assumption of parameter stability in a serious way before one can interpret the outcomes of any unit root test, be it univariate or multivariate. Breaking trends or other modifications for changing growth influence the outcomes of the traditional tests. Allowing for these modifications under the null adds problems and new distributions of the test statistics to the testing procedure as a whole.

Testing other restrictions

In this chapter we considered procedures to interpret some functions of coefficients of the VAR. The use of a priori information was avoided as much as possible in the spirit of multivariate time series analysis. In traditional econometric models exogeneity restrictions are often used to improve the theoretical precision of estimators. Testing these restrictions using traditional methods to approximate the critical values at conventional significance levels in models with unit roots is only possible *conditional on* the imposition of the appropriate number of unit root restrictions. Testing for exogeneity in unrestricted models leads tot "nonstandard" inference in general, see i.a. Sims et al. (1990), Warne (1991a). Rare exceptions in $CI(1,1)$ models are cases where exogeneity restrictions do *not* restrict the space of cointegrating vectors.

---

[24] Reimers used a Daniell window to estimate V. The Daniell window operates in the frequency domain, see §7.4.7.3 for a univariate application in comparison with other estimators of V. V is proportional to the spectrum of Δy_t at zero. Phillips and Ouliaris (1988) based a test on the eigenvalues of this estimator of V. The estimator does not use an AR model approximation of the spectrum.

A6.1 Computation and distribution multivariate unit root test statistics

In the early years of cointegration analysis a confusion reigned among applied researchers, which tables to use, which nuisance parameters to take account of, et cetera. Now the development of appropriate asymptotic theory seems to have settled down. The univariate procedures of chapter 2 have all been generalized to the multivariate case. Stock and Watson (1988) generalized the filtering approach of Dickey et al. (1984) and the semi–parametric approach of Phillips and Perron (1988). Johansen (1988...1992d) generalized the Likelihood Ratio approach of Dickey and Fuller (1981).

A6.1.1 Computation

In all procedures discussed below one tests the null hypothesis of k unit roots against alternative hypotheses with k^* unit roots, where $0 \leq k^* < k \leq n$[1].

First we describe our implementation of a test of Stock and Watson (1988). They used a finite order VAR representation for $\alpha_\perp' \Delta x_t$, which does not fit into the framework of §6.3. Under the assumptions of that section $\alpha_\perp' \Delta x_t$ will in general follow a VARMA process for all $k<n$. It is therefore important to choose a satisfactory order for the approximating AR process for $\alpha_\perp' \Delta x_t$. This implementation of the Stock–Watson test is a straightforward generalization of the univariate two step procedure of Dickey et al. (1984), which we discussed as an alternative to the augmentation procedure of Dickey and Fuller in §A2.2.2.1.

The variables x_t can be demeaned (the standard procedure in principal components) or detrended by regression first. The asymptotic distribution under the null changes when these corrections are made. Correction should depend on the hypothesized mean growth rate of x_t under the null and the alternative. If the expected mean growth under either hypothesis is different from zero, the series should be detrended[2]. Incorrectly failing to do so results in either a nonsimilar or an inconsistent test. The problems are equivalent to the problems in the application of univariate Dickey–Fuller type tests. After the correction one determines[3] $\alpha_\perp' y_t$ and $\alpha' y_t$ as the first k and

---

[1] See Phillips and Ouliaris (1990, §6) for some critical remarks on tests for alternatives with $k^*>k$, i.e. tests for cointegration instead of test for noncointegration. It is hard to develop consistent versions of those tests. See Park (1990) for a reply. Some tests for parameter stability, like the CUSUM test on OLS residuals, can be interpreted as tests of stationarity versus unit root nonstationarity, see Canova and Hansen (1992) for examples.

[2] By detrending we mean taking residuals of regression on constant and trend.

[3] In the notation of §6.3 $\alpha_\perp$ can be written as $G^{-1}S_1$, since $\alpha' G^{-1} S_1 = 0$.

the last $n-k$ principal components of $y_t = x_t - (E\hat{x}_t)$. One subsequently estimates a VAR $\Pi^*(L)(\Delta\alpha_\perp' y_t) = v_t$, prefilters to get $y_t^* = \hat{\Pi}^*(L)(\alpha_\perp' y_t)$ and applies the equation (6.51) on the filtered data y_t^*, which can now be considered to have white noise residuals under the null.

Let $\hat{\lambda}_{fi}$, $\hat{\lambda}_{fi}^{\mu}$, and $\hat{\lambda}_{fi}^{\tau}$ denote the eigenvalues of the OLS estimator of the matrix of first order AR coefficients of y_t^* where the superscript depends on whether x_t has been demeaned or detrended, analogously to the notation of τ, τ_μ and τ_τ in §A2.2.4. In contrast to the univariate procedure of Dickey et al. (1984) there now exists the possibility that the estimated eigenvalues $\hat{\lambda}_{fi}$ are complex. Let p^* denote the chosen order of the autoregression to compute the appropriate filter for $\Delta\alpha_\perp' y_t$. If $k = n$, $p^* = p$. Stock and Watson (1988) provided low percentage and high percentage quantiles for the ordered test statistics $T(\mathrm{Re}(\hat{\lambda}_{fi})-1)$, and their corresponding demeaned and detrended versions, so that tests against explosive alternatives can be performed as well. We list a few below in table A6.1.

Johansen used squared canonical correlations, which are real, nonnegative and smaller than one by definition. In his procedure one has to consider all cases *i–v* of §6.4 separately and do correspondingly modified regressions (6.52) and (6.53). The extension implies that one computes *squared canonical partial correlations*, i.e. correlations holding the influence of the other independent variables, i.c. lagged differences and deterministic variables constant[4].

Here we only consider the "optimal" tests for each case. In practice one does not know whether one uses the optimal test since one does not know with certainty which case applies. This means that tables for "suboptimal" tests, that do not incorporate "true" restrictions on m_0 and m_1 have to be used as well. Johansen (1992c) advocated a "simple to general" test procedure, which follows the rule that a hypothesis is not rejected unless all its subhypotheses are rejected. So if one wants to reject $k = 1$, one has to compute test statistics for $k = 1 \wedge m_0 \in \mathrm{sp}(\gamma)$, $k = 1 \wedge m_0 \notin \mathrm{sp}(\gamma)$[5]. Subsequently one evaluates the outcomes of the tests for both subhypotheses. It is not easy to control the significance level of the procedure, but a lower bound can be derived.

---

[4] The maximum likelihood methods of Johansen and Ahn and Reinsel method differ in the extension to higher order models from the Box-Tiao method which uses (ordinary) squared correlations in higher order models as well. Hooper (1962) gave an econometric interpretation of squared canonical partial correlations.

[5] Johansen and Juselius (1990, Table A2) and Reinsel and Ahn (1992, Table 1) compiled tables for the case where $m_0 \in \mathrm{sp}(\gamma)$ where this restriction is not imposed in the test regressions and came up with 5% critical values for H_0: $k=1$ of 8.08 and 8.16.

The first issue in the application of Johansen's maximum likelihood procedure is the choice of a regressor set[6] of conditioning variables. In all cases one can extend the set of regressors in (6.52) and (6.53) with lags of Δy_t. Introducing too many lags reduces the efficiency of the test, introducing too few lags reduces its validity. As long as the way to choose the lag order is independent from the way to determine the cointegrating rank, the limit distributions of the unit root tests do not change, cf. Hall (1992). We discuss applied order selection in §7.5.1.

In the most restricted case *v* only *zero mean*[7] finite variance deterministic regressors can be added to make the residuals meet the Gaussian i.i.d. assumption. Periodic zero mean regressors to model seasonality can be used for this purpose. In cases *i–iii* one extends the regressor matrix with either a constant or a constant and a trend. In case *i* the regressor sets of (6.52) and (6.53) are extended with a constant and a trend, without any restriction. In cases *ii*, *iii*, only a constant is added to the regressors of (6.52) and (6.53).

In cases *ii* and *iv*, which involve restrictions on the coefficients of the deterministic terms as well, one extends the set of regressands in (6.53) as well. The extra dependent variable is then seen as belonging to y_{t-1}: $(-\Phi_1 - I)$ gets dimension $n\times(n+1)$, $(-\Phi_1 - I)^*$ gets dimension $(n+1)\times n$. In case *ii* one extends y_{t-1} with a trend variable t. In case *iv* one extends y_{t-1} with a constant.

Let λ_i^j denote the nonzero eigenvalues of the product of the coefficient matrices of the second and first multivariate regressions (6.53) and (6.52) in ascending order, where superscript j denotes the cases *i–v*.

In the procedure of Phillips and Ouliaris one uses the trace, i.e. the sum of the eigenvalues, of P_y. Johansen derived a corresponding trace statistic (i.e. the likelihood ratio statistic) for his procedure, using the partial sum of ordered eigenvalues of the canonical correlation matrix.

A6.1.2 Distribution

The multivariate unit root tests have nonstandard distributions. To our knowledge exact asymptotic quantiles of interest for these distributions,

---

[6] Johansen (1992b) considered the introduction of stochastic conditioning variables, which are assumed to be independent of the other innovations to the VAR at all lags, or "exogenous" in some other sense.

[7] Examples of *nonzero* mean deterministic regressors are permanent structural break variables $(1-L)^{-1}\delta_{t,l(T)}$ as in (5.1), with l a function of T, say $l=\lambda T$. Introduction of these regressors changes the asymptotic theory, cf. footnote 9

using characteristic function techniques as in Nabeya and Tanaka (1990), have not been computed yet. For dimensions bigger than one, critical values can be obtained by naive simulation. Ooms and Van Dijk (1992) contains a one page MATLAB program to do these simulations yourself. Considering other difficulties associated with the use of these tests, the uncertainty associated with these approximations is negligible[8].

Let $W(s)$ denote a k-dimensional Brownian motion or Wiener process on $C[0,1]^k$ with variance matrix I_k. $C[0,1]^k$ is the product space (see Billingsley (1968, p. 225) of k spaces $C[0,1]$ of continuous functions on [0,1], see Billingsley (1968, p. 54). Let W_μ denote a k-dimensional demeaned[9] standard Brownian motion of dimension k, and let W_τ denote a k-dimensional detrended Brownian motion of dimension k with the last component replaced by 1. Let W_c denote a k-dimensional standard Brownian motion extended with a constant component 1. Let $W_{\mu-t}$ denote a k-dimensional demeaned standard Brownian motion with the last component replaced by a trend component $t-½$. Let $W_{\mu t}$ denote a k dimensional demeaned Brownian motion. Finally, let $W_{\tau-t^2}$ denote a k dimensional detrended Brownian motion with the last component replaced by t^2. A k-dimensional standard Brownian motion can be also be written as an infinite sum of i.i.d. $N(0,I_k)$-variables Z_i, see e.g. Chan and Wei (1988, p. 382):

$$W(s)=\sum_{i=0}^{\infty}\frac{2\sqrt{2}}{(2n+1)\pi}\sin(n+½)\pi s Z_i.$$

Let $\lambda_i(X)$ denote the i-th ordered eigenvalue of a matrix X. Below we list the relevant asymptotic distributions. The assumptions needed for the *VECM* representation of a $CI(1,1)$ representation with normal disturbances are sufficient to derive these distributions. Some can be relaxed considerably. The existence of a finite and continuous spectrum for the disturbances is crucial, which entails that appropriate seasonal correction is essential.

---

[8] Cf. e.g. 5% critical values for k=1, case ℬ: 4.2 by Johansen (1988), 4.10 by Reinsel and Ahn (1992), who used most (30000) replications and a random walk series with the largest length (2000), 3.95 by Ouliaris (1991), and 3.84 by Osterwald-Lenum (1992), which seems to equal the (incorrect) chi-square critical value: a maximum difference of about 9% *in the critical values*. Differences for other null hypotheses are even smaller.

[9] Demeaning can be interpreted formally as the projection on the orthogonal complement of the constant function in an appropriately defined Hilbert space of functions in $L_2[0,1]^n$, cf. Phillips and Ouliaris (1990, p. 170, remark (c)). Generalization of this concept to projection on other deterministic variables is straightforward. See Schmidt and Phillips (1992) for similar generalizations of tied down Brownian motions. In Park and Phillips (1988) $W_0^*(s)$ and $W_1^*(s)$ denoted the demeaned and detrended Brownian motions.

Stock and Watson (1987, Th. 3.1, eq. (ii)) proved

$T\cdot(\mathrm{Re}(\lambda_{fi})-1)\to_{\mathcal{L}} \lambda_i((\int_0^1 W(\mathrm{d}W)')'(\int_0^1 WW'\mathrm{d}s)^{-1})$,

$T\cdot(\mathrm{Re}(\lambda^{\mu}_{fi})-1)\to_{\mathcal{L}} \lambda_i((\int_0^1 W_{\mu}\mathrm{d}W_{\mu}')'(\int_0^1 W_{\mu}W_{\mu}'\mathrm{d}s)^{-1})$,

$T\cdot(\mathrm{Re}(\lambda^{\tau}_{fi})-1)\to_{\mathcal{L}} \lambda_i((\int_0^1 W_{\tau}\mathrm{d}W_{\tau}')'(\int_0^1 W_{\tau}W_{\tau}'\mathrm{d}s)^{-1})$.

Johansen (1991a, Th. 4.1) proved[10]

$T\cdot\lambda_i^{v} \to_{\mathcal{L}} \lambda_i((\int_0^1(\mathrm{d}W)W_{\tau-t2})'(\int_0^1 W_{\tau-t2}W_{\tau-t2}'\mathrm{d}s)^{-1}(\int_0^1 W_{\tau-t2}(\mathrm{d}W)'))$,

$T\cdot\lambda_i^{iv} \to_{\mathcal{L}} \lambda_i((\int_0^1(\mathrm{d}W)W_{\mu t})'(\int_0^1 W_{\mu t}W_{\mu t}\mathrm{d}s)^{-1}(\int_0^1 W_{\mu t}(\mathrm{d}W)'))$,

$T\cdot\lambda_i^{iii} \to_{\mathcal{L}} \lambda_i((\int_0^1(\mathrm{d}W)W_{\mu-t})'(\int_0^1 W_{\mu-t}W_{\mu-t}'\mathrm{d}s)^{-1}(\int_0^1 W_{\mu-t}(\mathrm{d}W)'))$,

$T\cdot\lambda_i^{ii} \to_{\mathcal{L}} \lambda_i((\int_0^1(\mathrm{d}W)W_c)'(\int_0^1 W_cW_c'\mathrm{d}s)^{-1}(\int_0^1 W_c(\mathrm{d}W)'))$,

$T\cdot\lambda_i^{i} \to_{\mathcal{L}} \lambda_i((\int_0^1(\mathrm{d}W)W)'(\int_0^1 WW'\mathrm{d}s)^{-1}(\int_0^1 W(\mathrm{d}W)'))$.

Phillips and Ouliaris (1990, Th. 4.1) proved

$\mathrm{tr}(P_y)\to_{\mathcal{L}}\mathrm{tr}(\int_0^1 WW'\mathrm{d}s)$.

Below we list some selected percentiles in order to facilitate comparison with univariate results of Chapter 2 and to get an impression of order of magnitude of the differences from case to case. In applied work one probably would not need all the tables all the time. Furthermore, small sample corrections, which already have been developed, seem to matter more in many cases.

We list the critical values for the number of unit roots k equal to 2 and 1 as examples in table A6.1. Let λ_1 denote the eigenvalue with smaller real part and λ_2 denote the one with the bigger real part. In the Stock–Watson procedure one rejects the null of k unit roots against an explosive alternative for big λ–values. One rejects the unit roots against stationary alternatives for small λ–values. In the Johansen procedure one rejects the null for high values of λ, either against stationary or against explosive alternatives. In the Phillips–Ouliaris procedure one rejects the unit root hypothesis for high values of λ as well.

For $k=1$ one can link the results with the univariate analysis of appendix 2.2. One has $T\cdot(\lambda_{f1}-1)\to_{\mathcal{L}}\nu_1$, $T\cdot(\lambda^{\mu}_{f1}-1)\to_{\mathcal{L}}\nu_{\mu 1}$, $T\cdot(\lambda^{\tau}_{f1}-1)\to_{\mathcal{L}}\nu_{\tau 1}$, see Chan and Wei (1988, corrolary 3.1.3). Johansen and Juselius (1990, p. 180) showed $T\cdot\lambda_1^{v}\to_{\mathcal{L}}\tau_1^2$, and $T\cdot\lambda_1^{iii}\to_{\mathcal{L}}\chi_1^2$, a chi-square distribution with one degree of freedom, and approximate 95%–percentile 4.0. Hylleberg and Mizon (1989)

---

[10] We do not use the approximation $-T\ln(1-\hat{\lambda}_i)\simeq T\hat{\lambda}_i$, which Johansen and Juselius (1990, p. 179) made. The difference is substantial in their examples. For their Danish data they obtained (T=53, $\hat{\lambda}$=0.43) $-T\ln(1-\hat{\lambda}_4)$=30, while $T\hat{\lambda}_4$=23, with respective p-values of about 2.5 and 20 percent! This bias is towards rejection of unit roots. The small sample correction suggested by Reinsel and Ahn(1992) $-(T-np)\ln(1-\hat{\lambda}_i)$ (n=4, p=20) has a similar effect in this example.

discussed the empirical relevance of the last "normality" result for univariate series, and judged it to be minor. The test is only valid if the mean growth of x_t is "huge" compared to the variance of the deviations from the trend, see Schmidt (1990) for extensive tables. But if this is so, the stationary alternative is probably not very interesting. The same normality result holds for $T\cdot\lambda_i^i \rightarrow_{\mathcal{L}} \chi_1^2$. One can safely state that almost all unit root tests statistics of interest have nonstandard asymptotic distributions.

Fountis and Dickey (1989) developed a test for $k=1$ and arbitrary n, which has a direct intuitive appeal. It employs the biggest characteristic root of a purely stochastic zero mean finite order VAR denoted by $\hat{\lambda}_n$. They proved (op cit., Theorem 1):

$$T(\hat{\lambda}_n - 1) \rightarrow_{\mathcal{L}} \nu_1.$$

It equals test statistic of Stock and Watson for $p^*=1$, $k=1$. It can also be used to test against explosive alternatives.

Table A6.1 *Percentiles multivariate test statistics under appropriate null*

	$k=2$		$k=1$		
Statistic	5%	95%	5%	95%	
$T\cdot(\mathrm{Re}(\lambda_{f2})-1)$	-3.8	1.5			(1)
$T\cdot(\mathrm{Re}(\lambda_{f1})-1)$	-17.5	-.6	-8.0	1.3	(2)
$T\cdot(\mathrm{Re}(\lambda_{f2}^{\mu})-1)$	-8.8	.3			(3)
$T\cdot(\mathrm{Re}(\lambda_{f1}^{\mu})-1)$	-23.0	3.1	-14.1	-.1	(4)
$T\cdot(\mathrm{Re}(\lambda_{f2}^{\tau})-1)$	-14.9	-2.0			(5)
$T\cdot(\mathrm{Re}(\lambda_{f1}^{\tau})-1)$	-30.8	-6.3	-21.7	-2.7	(6)
$T\cdot\lambda_2^i$		16.7			(7)
$T\cdot\sum\lambda_j^i$		18.1		3.8	(8)
$T\cdot\lambda_2^{i\,i}$		19.2			(9)
$T\cdot\sum\lambda_j^{i\,i}$		25.4		12.5	(10)
$T\cdot\lambda_2^{iii}$		14.0			(11)
$T\cdot\sum\lambda_j^{iii}$		15.2		4.0	(12)
$T\cdot\lambda_2^{iv}$		15.8			(13)
$T\cdot\sum\lambda_j^{iv}$		20.2		9.1	(14)
$T\cdot\lambda_2^{v}$		11.4			(15)
$T\cdot\sum\lambda_j^{v}$		12.3		4.1	(16)
$\mathrm{tr}(P_y)$(standard)		71.3		40.8	(17)
$\mathrm{tr}(P_y)$(demeaned)		89.8		55.2	(18)
$\mathrm{tr}(P_y)$(detrended)		124.4		81.4	(19)

NOTES: k: number of unit roots under the null. Subscript j denotes order of eigenvalues λ_j. Superscripts $i,\ldots,v$ denote configuration deterministic terms, see table 6.1. Source: rows (1)-(6): Stock and Watson (1988, Tables 1, 2, 3), rows (7)-(10) Johansen (1991b, Tables 2, 3), rows (11)-(14): Johansen and Juselius (1990, Tables A1, A3), row (15): Osterwald-Lenum (1992, Table 0), row (16): Reinsel and Ahn (1992, Table 1). rows(17)-(19): Phillips and Ouliaris (1990, Tables IVa, IVb, IVc).

7 APPLIED VAR ANALYSIS FOR AGGREGATE INVESTMENT

7.1 Introduction

The proof of the pudding is in the eating. In the previous chapters we discussed a number of techniques that have proved useful ingredients of the VAR analysis of quarterly macroeconomic variables. As a final course of this study we present an application, which may serve as a suggestion how to use a tasteful mix of these techniques.

During the last years the price of the application of these techniques has fallen dramatically. Programming time has fallen, both by increased ability of programmers, and the availability of new cheap computers and software. The cost of getting all relevant information on the reliability of both software and hardware has risen with the number of products arriving on the market, especially if one takes the first sentence of this chapter seriously.

Although the implementation of many techniques has absorbed much time in the past, it probably won't in the future. In the course of this study we have chosen to implement procedures in a way to our own liking. Most procedures are also available, or easy to program, in specially developed statistical and econometric software packages, which are frequently reviewed in scientific journals, like Computational Statistics and Data Analysis, The American Statistician and the Journal of Applied Econometrics. A cross check on the virtue of the different procedures can reveal important errors and omissions. A periodic reevaluation as new versions of the programs arrive is in order. An advantage of our restriction to linear estimation methods is that one does not get involved in all kinds of issues in numerical mathematics, as long as one uses 8 byte precision and robust eigenvalue and inversion procedures. The computation time is negligible compared to programming time.

In the second section of this chapter we discuss our variable of interest and its theoretical relationships with other variables. We also discuss empirical models for investment. In section three we focus on the measurement model. In section four we present results from the univariate analysis of some candidate variables in our VAR. In the fifth section we discuss the results of the multivariate analysis.

7.2 The variable of interest and some of its supposed relationships

The observations of interest in this application are quarterly data on aggregate investment in fixed capital in the Netherlands in their relation to other macroeconomic variables. The investment variable exhibits a substantial variation and serial correlation. A VAR model should be able to pick up a reasonable proportion of this variation. A big part of the variation in the investment series is concentrated at the zero frequency and the seasonal frequencies, so that unit root analysis becomes interesting. The series also exhibits some variation at the business cycle frequencies, and is often used as an indicator in measurements of the business cycle.

7.2.1 Theoretical relationships

Microeconomic considerations

Other business cycle indicators are often assumed to "determine" a substantial part of the variation in investment. Quite reasonable assumptions exist under which a certain increase in expected demand leads the management of a majority of the firms to invest extra in order to increase the "present discounted value" of their well being, while others may be more inclined to start a new business. The reverse effect can be expected when expected sales fall. According to the figures actual aggregate "disinvestment" has not occurred. It is not measured on the firm level either, see Broer and Van Leeuwen (1991). Only after 1969 sales of fixed capital abroad are registered as such, see C.B.S. (1988, p. 208).

One should not neglect supply side effects. A rise in expected output has a substantial effect on fixed investment only if it increases the expected profits of the firm relative to the future revenues of other investments. Increasing costs of the purchase or rental of fixed capital and complementary production factors bring down the investment's profitability. In the Dutch situation labor costs have received a lot of attention in macroeconomic modeling. Labor was considered complementary in the short run, and a substitute in the longer run. Unfortunately there is not enough information in macroeconomic data to discern these effects up to now in empirical models. Rising expected profits on other investments decrease the inclination to invest in fixed capital within the country, be it by domestic or by foreign investors. Examples of relevant alternatives are investments in inventories at the wake of a price increase or as a buffer to absorb productivity shocks (Christiano (1988)), or investments in shares and bonds of other firms and

institutions in the country of interest and abroad. Share prices for comparable companies do not differ too much from country to country in efficiently operating capital markets. Reproduction costs of capital differ more in general.

A crucial additional factor is the relative uncertainty of the expected profits on the alternative investments. Malinvaud (1987) discussed, among other important issues, the influence of an increase in uncertainty about future demand on the expected profitability of investment in a firm. Similar analyses can be made for the effect of uncertainty about future energy prices, tax rules, investment subsidies, exchange rates etc..

In traditional empirical analyses one examines functions of empirical covariance matrices, based on some explicit formulation of an economic theory. A range of restrictive assumptions must be made, in order to keep these models tractable. Nickell (1978) discussed many assumptions and implications for the observed correlations of investment and other variables of interest. He then studied the effect of relaxations of many of these restrictions which are often used in applied equation by equation analysis. A slight relaxation of more than one restriction already results in an intractable model for investment.

Lag structure

Especially as far as the all important lag structure is concerned, it is not difficult to "explain" all kinds of covariances. A wide range of reasonable assumptions on the specification of e.g. adjustment costs on the supply and demand side of the market for investment goods (Warner (1978), Galeotti (1987)), irreversibility of investment (vintage technologies like putty–clay), delivery lags, gestation lags, level of aggregation (Nickell (1984)) can "explain" most lag structures in effective empirical factor demand equations. Chirinko (1988) presented a comprehensive survey of assumptions and results for neoclassical models. Many "classical" tractable explicit theories are based on the behavior of single representative agents. The effect of relaxation of the homogeneity assumption for model consistent behavior for the aggregate is hard to foretell.

Theoretical VAR models

The VAR technology has often been applied to test implications of Real Business Cycle models, more aptly called stochastic dynamic general equilibrium models. In these models one tries to avoid making some of the "incredible restrictions" in the sense of Sims (1980), and make other "more credible" restrictions derived from "rational" forward looking behavior and

analytically tractable intertemporal utility functions of representative agents. The model by King et al. (1987), which we referred to in chapter 6 is a simple example. From an empirical point of view one can hardly take the tractable versions of these models seriously, however. Authors admitted this. Plosser (1991, p. 963) stated about Long and Plosser (1983): "Unfortunately, the model exhibited some unusual characteristics that made it difficult to treat it seriously as a model of fluctuations. Not the least of which was that employment was constant.....Another feature was that we assumed 100% depreciation." In principle supply and demand are not modeled separately. Shocks affect demand and supply simultaneously. "In these models shocks occur either to tastes, technologies/opportunities, or endowments." (op cit, p. 965). But "most real business cycle models to date treat technology shocks as the driving force." Even with the strong assumptions of rational expectations and dynamic general equilibrium one has only obtained tractable models which still exhibit some important "unusual characteristics". Palm et al. (1992) showed that accounting for time–to–build in stochastic partial equilibrium factor demand models (notably demand for plants) leads to VARMA specifications for the factor demand variables, even if the variables that are not Granger caused in the system, i.c. real factor prices, are assumed to follow VAR processes.

Macroeconomic considerations

Disequilibrium models, see e.g. Lambert (1990), which explicitly take account of the aggregation of different submarkets, some of which may be in disequilibrium, are also hardly tractable as soon as the forward looking behavior of the agents is taken into account properly. A way to circumvent this problem, is to include extra variables derived from business surveys. These can be used to construct aggregate data for expectations, which are not based on the model one is trying to estimate. They can also be used to get model independent estimates of the proportions of capacity and demand constrained firms. In this study we are only interested in correlations between observable variables from the real world which indicate a certain behavior. We do not consider the expectations per se.

7.2.2 Empirical models

Disaggregation

Preoccupation with the unreality of the assumption of the homogeneity of capital led Dutch macroeconomists in the seventies to study the behavior of

investment in rather disaggregated models, see e.g. Van Den Berg et al. (1983), as far as the capital stock is concerned. This disaggregation has remained ever since, see e.g. Den Butter (1991). One distinguished fixed capital goods in six main asset categories: equipment, nonresidential buildings, dwellings, cattle, ships and planes, and mining (exhaustible resources). During the post war period investment in the last five categories has been regulated to a big extent by the government. Nonresidential buildings, roads, and dwellings were built as far as the number of permits allowed, see i.a. Bakker (1962). Investment in cattle was often determined by special arrangements in the EC. The government also regulated the investment and scrap of capital in the mining industry as well as other investments in the energy sector. Because of these separate influences it makes sense in an equation by equation approach to model those components separately. Investments in buildings and equipment were treated separately because of the big difference in the lifetime of those capital goods, and because of the different tax facilities for these two groups of investment goods.

In VAR methods one does not make exogeneity assumptions about the probability laws governing policy behavior and the public's reaction, see Sims (1985, 1986). Government behavior is endogenous. In principle we assume the relevant conditional probability law of the behavior of the policy makers to be stationary. Important effects of changes in the relevant reaction functions can be traced back as indications of nonlinearities in the diagnostic checking of the model. A prime example is the extravagant behavior of world interest rates in the beginning of the eighties after an important change in U.S. monetary policy, which is very hard to explain using linear time series models estimated on earlier data.

One applied a further disaggregation in models which emphasized the *irreversibility* of investment. One distinguished effects on replacement investment and expansion investment. One hypothesized that an amount of capital goods with finite maintenance costs were scrapped because of their inferior profitability relative to new capital goods. Large increases in the price of labor and energy relative to the production factor capital could have caused such investment in the Netherlands, mainly in the sixties and seventies. In such circumstances the lifetime distribution of the capital stock and corresponding distribution of productivities can become an extra factor in the decision to invest. This effect was supposed to be important for the investment in equipment and much smaller for the investment in buildings. Endogenising the depreciation complicated the model considerably.

Capital stock specification

The specification for the current capital stock is a crucial ingredient for all empirical analyses of investment. A range of models, such as the Q–model choose the ratio of current investment over the existing capital stock as their variable of interest. Sensenbrenner (1991) compared different measures for the aggregate capital stock for a number of countries, and showed them to be quite different. Many other models specify a desired level for the capital stock, see e.g. Nickell (1978, Ch. 11), so that the difference between the actual and desired level of the capital stock becomes the main determinant of investment. Clearly, misspecification in the model for the capital stock could then cause trouble in the interpretation of the estimates of parameters of structural equations.

Sensible model independent estimates of the capital stock are very hard to obtain. If one does not have independent estimates one can construct capital stock data by assuming a scrapping and "evaporation" process for the past investments, see e.g. Nickell (1978, Ch. 11).

As soon as one specifies a model in which the scrapping decision becomes dependent on the lifetime distribution of the existing capital stock and a number of other variables, one easily runs out of degrees of freedom to estimate any parameter of interest. Model based estimates of the aggregate capital stock are thus either strongly dependent on parameters of a simple unrealistic model, or on the parameters of a more sophisticated model that one can barely estimate. If one does not take this into account one runs the risk of reestimating the model used to generate the capital stock data. Chirinko (1988) gave an example.

Investment price specification

A second crucial ingredient in the specification of the variable of interest is a measurement model for the price of aggregate investment. De Leeuw (1990a) addressed this problem. Do we evaluate investment in computers in bits of memory or in bits of computing capacity? Do we evaluate software differently? Does externally purchased software belong to the capital stock? What about internally developed software? Most users do not even know. How can one agree on an economically meaningful definition for a price index of investment goods? A sensible strategy is to pursue a sensitivity analysis for the effect of the choice of a price of aggregate investment, and the choice of a model for the capital stock. Sensenbrenner (1991) presented a sensitivity analysis in this spirit. He found that the main conclusions about the relative

adequacy of the models in his study were not influenced too much.

Specification of related variables

The specification of the related variables of interest presents similar problems. How does one specify a meaningful aggregate user cost of capital? Wages are expected to play a role as far as labor is a complementary production factor. It is very hard to obtain sensible figures on the costs and returns of labor without specifying an explicit model which introduces a set of production possibilities. Such a model may make sense on the micro or industry level, but it is very hard to defend it as reasonable for a macroeconomy. We do not explicitly include wages in the empirical model. Wages are an important factor in the price of non–residential buildings, which is used in the deflation of the investment series. Wages also play a role in the price of exports. Real wage costs have to take account of the quality of labor, i.e. productivity, which depends on capital/output and labor/output ratios which are not measured directly on a quarterly basis. Changes in these ratios are mitigated and lagged in the measurement model, see §7.3.2. Nominal wages change significantly only once or twice a year, which introduces seasonal heteroskedasticity. Yearly data seem to be a more convenient choice to study the effects of wages in an empirical economic behavior model.

How does one specify an aggregate opportunity cost? How does one specify the associated risks? Most attempts are admittedly crude. Even parts of models for capital costs with many known parameters, such as the components describing the effects of taxes and subsidies have to incorporate many unrealistic simplifications. Chirinko (1988) analyzed the US situation. Gérard (1990) analyzed the European situation. Ter Rele (1989) compared Dutch parameters with foreign ones.

Scope of a VAR analysis

A VAR model is of course far too crude to pick up all the effects we discussed above. Only historically quantitatively important relationships can show up. The crucial difference between the application of a VAR methodology and the use of traditional equation by equation analysis is that one tries to avoid making "incredible" a priori restrictions in the former approach, see §6.1. This seems especially relevant for investment models where all variables are assumed to be related to expected future economic activity and capital costs as well as a range of opportunity costs.

Effects of expected changes in risks are usually modeled via inherently nonlinear relationships between expected covariances and means of different variables, see e.g. Engle et al. (1987). A VAR model is not meant to pick up

these effects. If those effects are big enough they could show up during the specification testing of the model.

The approach of looking at all covariances at different lags simultaneously, severely limits the number of variables that can be analyzed at one time. In a VAR model for Dutch investment one would like to include variables strongly related to future economic activity at home and abroad, variables strongly related to expected capital costs and variables strongly related to relevant opportunity costs. One then has to transform the variables in such a way that an interestingly big part of their variability can be satisfactorily explained by a VAR model. Examples include power transformations.

One can only achieve this if the empirical distribution and autocovariance functions of the transformed variables can be approximated by corresponding functions for generalized ARIMA models with finite variance innovations. Checking the adequacy of the model is simplified if one can assume these innovations to be reasonably Gaussian, possibly around a well behaved acceptable deterministic function of time with a small number of interventions.

Nonlinear alternatives

Dutch macroeconomic data exhibit special types of variation. This led macroeconomists to use slightly a–typical complicated nonlinear vintage models. It can also lead one to specify complicated nonlinear time series models, see e.g. Lin and Teräsvirta (1991). One pays a price for the introduction of non–linearities. The relationship between the observations and the parameter estimates, in so far it can be discerned, becomes vague. Influence analysis becomes costly. The improvement in "fit" induced by extra parameters to model "nonlinearity" does not automatically lead to a "better" description of the data.

Once one knows how to transform the data to stationarity one can always model the purely stochastic part and the purely deterministic part linearly using the well known *Wold decomposition*, see e.g. Hannan and Deistler (1988, p. 21). We follow this strategy in sections 7.4 and 7.5.

Many techniques in the previous chapters were discussed because they seemed to be relevant for Dutch macroeconomic data which exhibit seemingly nonstationary growth, slowly, but persistently changing seasonal patterns, and a number of extraordinary interventions, some of which may have had a permanent, and some of which may have had a temporary effect on the series.

According to the different macroeconomic theories hundreds of variables

could be tested for a relationship with aggregate investment. In one VAR model one can only test a few.

7.3 Measurement model

7.3.1 Investment in the national accounts

In this section we give some necessary background information on the construction process of Dutch macroeconomic investment data. We discuss statistical definitions of value and quantity indexes. We concentrate on national accounts data. National accounts data are the most reliable from a statistical point of view. The simultaneous computation of macroeconomic values from the input and output side[1] provides an extra check on the reliability of many figures, and leads to substantial corrections over the years. The data are constructed so that macroeconomic (budget) identities are fulfilled both in value terms and in real terms for each year.

Deflation

Comparison across years of price indexes and figures in real terms presents some problems however. National accounts are not set up to check economic theories. The influence of changing relative prices, which plays such a major role in neo–classical theories, is usually mitigated in national accounts data, see Al et al. (1985). The deflators of macroeconomic entities are merely a byproduct. On the aggregate level one can decompose a value change into a price component, a quantity component and a structural change component. The last component reflects the changes in the composition of the aggregate, i.e. relative quantities of the products, due to the appearance of new products or changes in relative prices. The structural change does not have its own index, whereas price changes and quantity changes have. National accounts deflators pass Fisher's factor reversal test which requires

$$P \cdot Q = V, \tag{7.1}$$

i.e. the product of price index and quantity index equals the value index. The structural change effect is distributed among the price change and the quantity change. Al et al. (1985) stated: "The aim is not to isolate *pure* price components or *pure* quantity components. The primary aim when decomposing the value change in aggregates in the national accounts and input–output tables is to reflect the real (volume) changes in the transaction aggregate."

[1] This is not the case for quarterly investment figures, see below.

In practice one combines Laspeyres volume indexes, $Q_L = p_0'q_1/p_0'q_0$, and Paasche price indexes, $P_P = p_1'q_1/p_0'q_1$, where p_i and q_i denote price and quantity (column) vectors in period i. One is mainly interested in yearly changes. p_1 and q_1 denote prices and quantities in the current year, p_0 and q_0 denote those of (the same period) last year.

From both price indexes and volume indexes one can construct chain indexes for comparison across longer periods. The use of chain indexes is very much in line with statistical practice, since it simplifies the treatment of disappearing goods or production. Tables in constant prices are not available for long periods. A severe problem of the chain index is *additive inconsistency*: Tables produced by multiplying the basket of the base period by the volume chain index for an observation moment separated from the base moment by more than one period exhibit aggregation discrepancies, see Al et al. (1985). The volume index of GDP is not equal to the basic weighted sum of the volume indexes of its components. This may be one of the reasons for the difficulty of finding long run equilibrium relationships between real macroeconomic entities. The purpose of the measurement model is clearly not in line with this type of research. Al et al. (1985) stated: "The chain index is preferable since users of such information are primarily interested in the way in which the changes in the aggregates come about. It is clearly only a minority of cases that these data are used to make a comparison between two moments separated by a long time interval." Valliant and Miller (1989) suggested an index estimator which is a compromise between a fixed base index and a chain index. Lichtenberg and Griliches (1989) discussed the accuracy of different estimators for long term output changes and price changes in the U.S.A.

7.3.2 Definition of investment

Investment is neither defined according to the economic purpose of the good, nor according to the use of the good. The statistical office decides on the nature of the good whether it belongs to investment or consumption. Statistically important examples of investment goods are buildings, equipment and cattle. Private cars and government military expenditure are part of macroeconomic consumption. Land is not considered an investment good.

The value of macroeconomic investment is computed in a commodity–flow system (CBS (1986, p. 28)). For each investment good j one computes the value of investment expenditure I_j as

$$I_j = Y_j + M_j - X_j + \Delta S_j, \tag{7.2}$$

with Y_j domestic product, M_j imports, X_j exports and ΔS_j change in stocks. Aggregate investment is then obtained by summation. Up till 1986 the Central Bureau of Statistics did not produce quarterly national accounts, because of poor quarterly estimates of ΔS_j. They reconstructed data from 1977.1 onwards. Quarterly data from 1970.1 have been reconstructed by the OECD (1991). De Nederlandsche Bank, DNB (1986), produced figures for simplified national accounts for the period 1957.1–1984.4, and compared their figures to those of the quarterly national accounts of the years 1977.1–1984.4. Figures up to 1969 include indirect taxes in production figures of separate industries. Data from 1969 on do not. The Value Added Tax system was introduced in 1969.

Production measurement

Domestic product Y_j is not measured directly. One estimates Y_j from data on the value of inputs used by the investment goods producing industries. For the construction of quarterly data for the building industry, only labor input figures are used. One employs last year's input–output coefficients to compute production figures. Input–output coefficients are updated only once a year. Quarterly data thus do not contain information on structural changes in the production process that is not already present in yearly data. In the construction of yearly data, price information is sometimes taken into account for the allocation of roughly specified items into categories, but *gradual* changes in input–output coefficients are preferred, see De Boer and Broesterhuizen (1986): "This conservatism, in the sense that the statistician will usually assume a gradual, rather than an extreme structural change when faced with lack of information, is justified because he must pursue minimization of errors in his estimates. As he does not know in such cases in which direction the effects will take place, he opts for an average."

Quarterly data are made consistent with the yearly data so that both value and quantity figures add up to the yearly ones. Data are deflated using average prices per quarter. Yearly data are deflated using average prices per year. Seasonally unadjusted data are only computed as yearly growth rates. This leaves the relative position of the different quarters undetermined. Quarterly values and volumes were only computed for the base year 1977. Separate quarterly series move independently from there, except for the adding up restriction in connection with the yearly data. This may be an explanation for the apparent seasonal nonstationarity in Dutch macroeconomic time series, see §7.5 below. A multiplicative seasonal adjustment filter like Census X–11

removes the arbitrary initial seasonal multiplication factors. Quarterly growth rates are only published in seasonally adjusted form.

The Netherlands have been in a customs union with Belgium and Luxembourg in the sample period. Imports and exports of investment goods to and from those countries have to be estimated as well. Because of the opening up of "the free European market" this problem for the construction of investment figures will aggravate in the near future. No component of quarterly aggregate investment will be measured directly any more.

7.3.3 Other macroeconomic price indexes

Traditionally other price indexes for macroeconomic variables have been used in purchasing power studies. The consumer price index is the best known. In principle this is a Laspeyres index with fixed base year, so that prices across years are comparable. In longer time series the basis for the index is changed at least every ten years, so that prices across longer periods are not comparable either. Dutch internationally issued trade figures are deflated using so-called *unit values*, see e.g. the OECD's *Main Economic Indicators*, *historical statistics*, or the IMF's *International Financial Statistics*. These agencies use a chained (ideal) Fisher price and volume index, i.e. a chain index of geometric averages of Laspeyres and Paasche indexes.

Although the national account figures may be the most reliable, there is no convincing reason to exclude other deflators and production indicators from the information set, except for restrictions on the total number of variables that can be handled. Since many economic theories stress the importance of expectations, it makes sense to include other figures as well, especially if they are available sooner and more widely publicized. Unfortunately, information on their construction and reliability is not as well known.

All in all there are many reasons to believe that structural permanent changes occur in the measurement model, which cannot be explained by the economic behavior model. A priori restrictions derived from economic theory which are said to be applicable for a certain set of macroeconomic data always have to be judged against this information.

7.4 Univariate analysis

7.4.1 The variables

We examine the following six variables, viz.

c: log volume total private consumption in the Netherlands,

i: log volume total investment in fixed capital in the Netherlands,

m: log volume imports of goods and services into the Netherlands,

x: log volume exports of goods and services from the Netherlands,

y^*: log volume of total industrial production in the Federal Republic of Germany, and

px/pm: log ratio of export prices over import prices (=log terms of trade).

Appendix 7.1 contains a detailed account of the construction and sources of the series. The choice of these variables is based on real business cycle VAR models (King et al. (1991)), where consumption, investment and GDP form the core. Mellander et al. (1992) adapted such a model to a small open economy situation and introduced the terms of trade. Unfortunately there are no reliable, long quarterly series for Dutch real GDP, because of poor quarterly registration of the changes in stocks. Therefore we choose to model the remaining factors[1] determining GDP separately. This enables us to study the dynamics of the trade balance in real and nominal terms as well. In this way important factors determining expected demand in the economy are included. The price of labor is not included. It is hard to construct model independent measurements of effective costs and returns on capital and labor. The terms of trade can be an important factor in the real costs of Dutch investment.

7.4.2 Graphs and influence analysis

In this day and age of cheap computers nobody would start an empirical analysis without making some time series plots of the data first. Although this has been the main step in statistical analysis for many years (Beniger and Robyn (1978)), little scientific research has been done to investigate the power of graphical methods, see Cleveland and McGill (1984).

We present the data in levels and in four quarterly subseries of first differences in figures 7.1a–7.1d. These subseries are used in a principal component analysis as described in §4.4. A first visual influence analysis can be confirmed by Critchley's influence analysis described in §A3.3.

---

[1] The main factor that we exclude is government consumption, which constitutes 20 to 25 % of total consumption. Private consumption is considered a more interesting macroeconomic indicator.

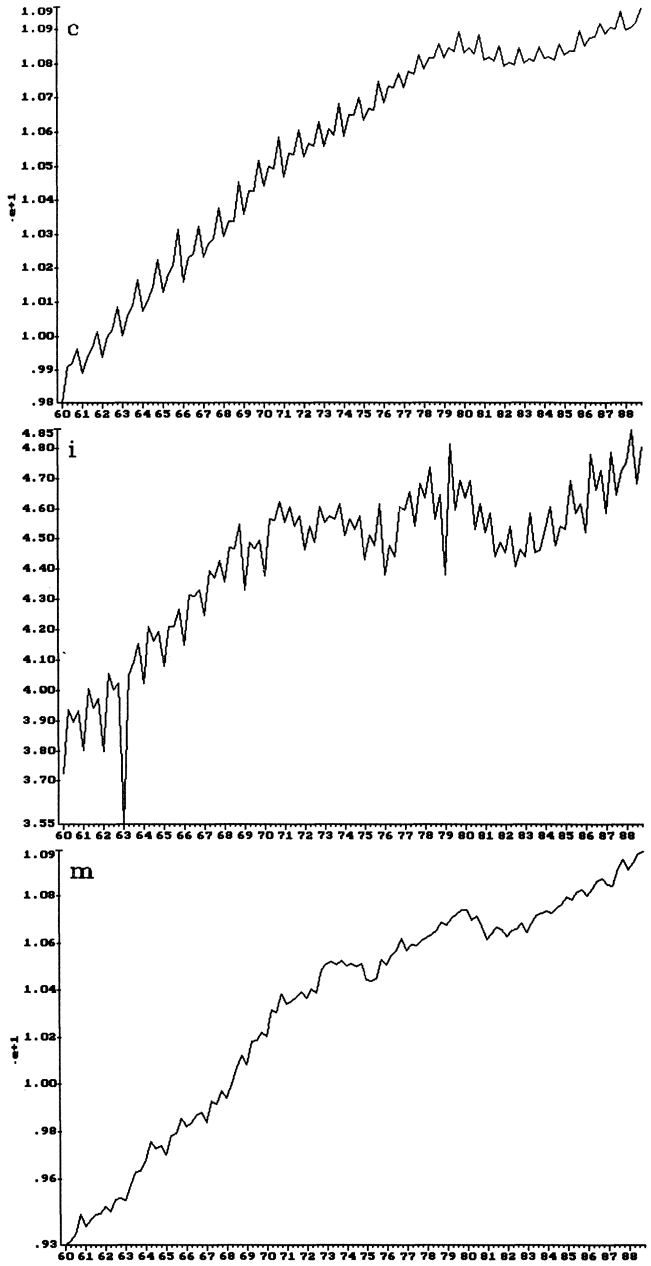

Figure 7.1a *Time series plots of macroeconomic data in log levels* 1960–1988
c: consumption, *i*: investment, *m*: imports.

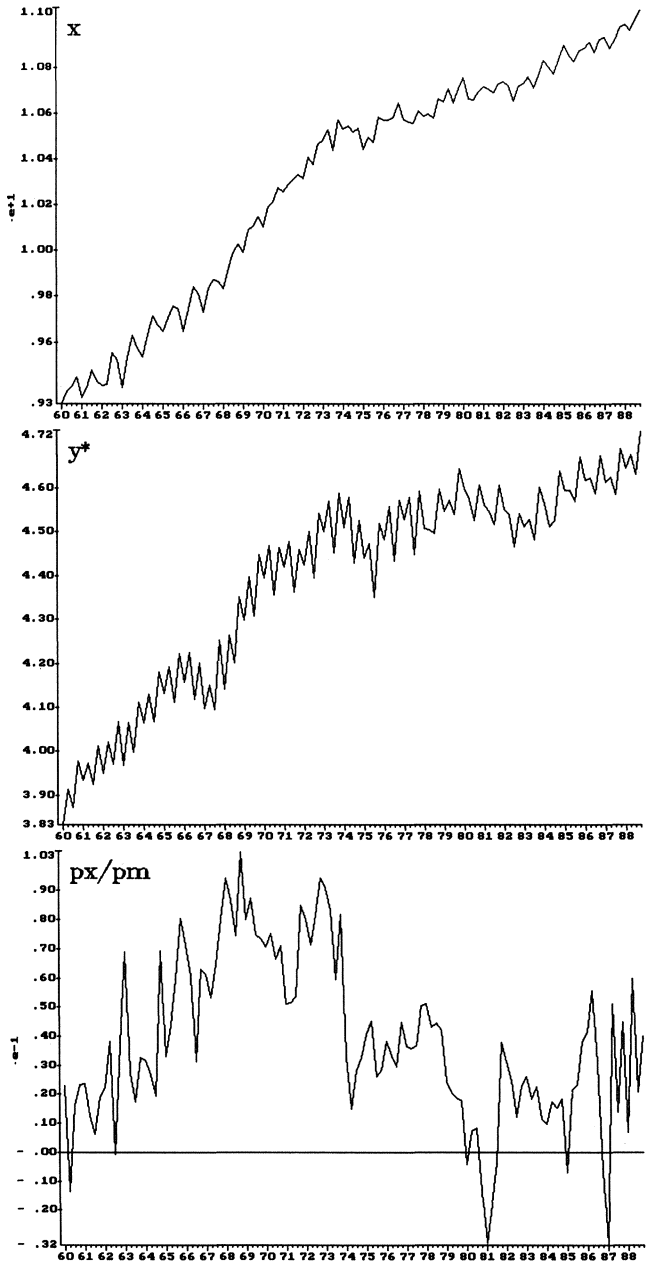

Figure 7.1b *Time series plots of macroeconomic data in log levels (continued)*
x: exports, *y**: German industrial production, *px/pm*: terms of trade.

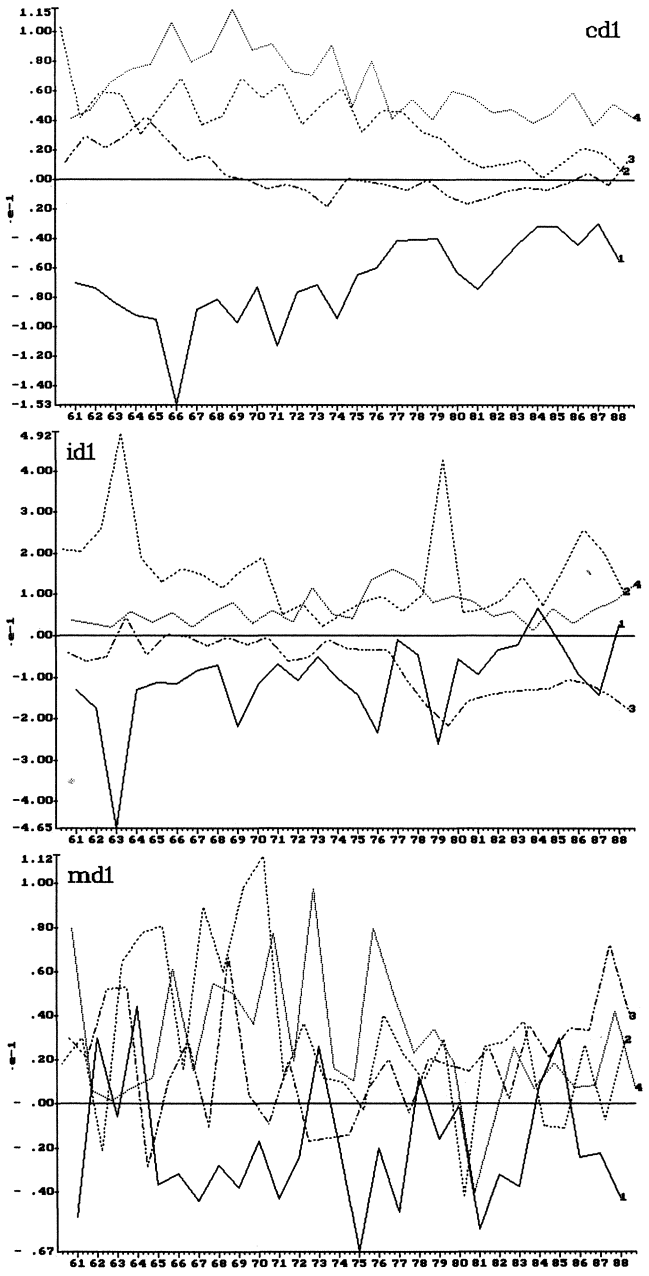

Figure 7.1c *Time series plots of macroeconomic data in first differences of log levels* 1960.2–1988.4; *cd1*: consumption, *id1*: investment, *md1*:imports Numbers at end of plots indicate quarter of year. Plots connect growth rate value in a specific quarter with growth rate in same quarter one year before.

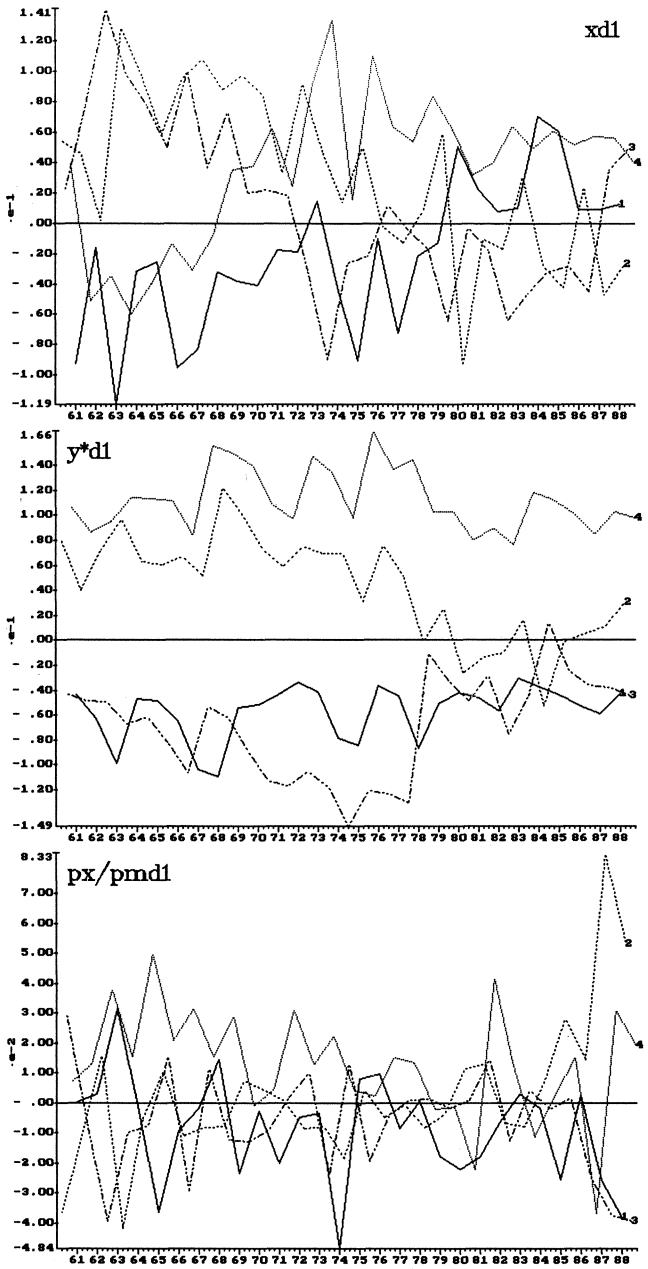

Figure 7.1d *Time series plots of macroeconomic data in first differences of log levels (continued);* *xd1*: exports, *y*d1*: German industrial production, *px/pmd1*: terms of trade. Numbers at end of plots indicate quarter of year. Plots connect growth rate value in a specific quarter with growth rate in same quarter one year before.

This first inspection may lead one to do the outlier tests of chapter 5 on the univariate series first, where the univariate ARMA models are approximated by long AR models for convenience.

Time series plots of statistics from the influence analysis can be used to examine the need for some sort of downweighting of influential observations. One can use the additive outlier correction described in appendix 5.3 to achieve this "whitening" straightforwardly. Some of the recursive estimation techniques and associated tests from §3.3.6 can indicate significant changes in the lag structure towards the beginning and end of the sample. One can direct the focus of the influence analysis to unit root tests by examining partial regressions (remark A3.4) in a transformed autoregression (A2.2.12).

7.4.3 Representations of the autocorrelation function

If the data are considered clean enough it makes sense to estimate the (inverse[2]) autocorrelation function of the appropriately differenced series, probably adjusted for some deterministic functions of time. The deterministic components can take account of a trend in mean, seasonal shifts in mean and (some) structural breaks. An estimate of the *extended sample autocorrelation function* of Tsay and Tiao (1984) can be used to get an easy estimate of the AR order which does not depend on assumption of appropriate differencing[3].

Subsequently one presents the data (or its autocorrelation function) in the frequency domain. We compare different estimates of the spectrum at the frequencies of interest. This is useful if one wants to explain differences between outcomes of competing unit root tests. All unit root tests need some estimate of the spectrum of the appropriately differenced series at the frequencies of interest. In the employment of Dickey–Fuller tests one uses an AR(MA) model approximation of the spectrum[4]. The Phillips–Perron tests offer a wide choice of spectral estimators. It makes sense to compare a few in a sensitivity analysis.

---

[2] The *inverse autocorrelation function* is the autocorrelation function corresponding to the inverse of the spectrum. The roles of AR- and MA parameters are interchanged. The inverse autocorrelation function "cuts" off for finite order AR models, see Cleveland (1972).

[3] One can easily improve the estimators of the statistical significance of the *residual* autocorrelations in their extended regressions (application of the rule of thumb critical values $\pm 2T^{-\frac{1}{2}}$) using Godfrey's (1978) test.

[4] Selection of ARMA orders and estimation methods for the ARMA model is still an issue of debate in the time series literature. Some members of this class of approaches are labeled *maximum entropy*, see e.g. Gray and Woodward (1986).

The periodogram estimates are used to examine the long memory properties of the data. One simply regresses the log periodogram estimates in the neighborhood of $2\pi\lambda$ on the logarithm of the theoretical spectrum of an $I_\lambda(1)$ series given in (2.17, last line) and a constant. The slope coefficient is a consistent estimator of the fractional order[5] of integration d for $-½ < d < ¼$. In practice only a very few points can be used. Only large absolute values of the coefficient serve as a notable indication. A negative sign indicates that the series probably is not long memory at that frequency. A positive one indicates that it probably is not antipersistent.

For quarterly data the frequencies of interest λ are 0, ¼ and ½. It makes sense to round the number of observations to an integer multiple of four so that these frequencies belong to the standard ones for the periodogram estimates. The frequencies of interest themselves are not included in the regression. If the series is adjusted for deterministic seasonality those periodogram estimates are already zero by construction.

If one is convinced that the series is not long memory at frequencies other than zero –application of any symmetric seasonal adjustment filter will do that trick as explained in §4.3.2– one can apply the variance time function techniques and adjusted range techniques to view the long memory characteristics at the zero frequency.

The limit of the scaled variance time function defined in (2.23) gives an impression of "the size of the random walk component of the series". If this function tends to zero the series probably is not "very integrated".

7.4.4 Adjusted range techniques

The slope coefficient of a regression of the values of the log of the rescaled adjusted range on the logs of the corresponding subsample sizes gives another estimate of the order of integration at the frequency zero. A value (much) bigger than ½ indicates long memory properties, a value (much) smaller than ½ indicates antipersistence. One can use the modified rescaled range statistic defined in (2.30) in a formal test of the hypothesis that the series is $I(0)$. Low p–values of this test lead to rejection against $I(d)$–alternatives with $d<0$, high p–values, i.e. low tail probabilities for the right hand side of the distribution, lead to rejection against alternatives with $d>0$. This

---

[5] Porter-Hudak (1990) and Hassler (1993b) discussed theoretical developments. They did not use the frequency by frequency approach we suggest here. Inspired by the results of Chan and Wei (1988), discussed in appendix A2.2 one can guess that fractionally integrated components at different frequencies are asymptotically independent as well, so that the frequency by frequency approach is valid for $0<d<\frac{1}{4}$, although the spectrum is not finite in that case.

test also needs an ordinary estimate of the spectrum at the frequency zero.

With the results of this part of the data analysis in hand one is well equipped to do a first round of regression unit root tests as presented in appendix A2.2. Probably some cycles where different types of conditioning on deterministic variables, different ways of differencing and different corrections for outliers are employed, may prove to be necessary before one obtains a sufficiently intimate knowledge of the important univariate memory characteristics.

7.4.5 Allowing for a break in regression unit root tests

It seems sensible to allow for a change in both the mean growth and the seasonal means in the data at hand. Let $D_{l,t}$ be a dummy variable equal to one for $t \geq l$, and zero elsewhere. We apply test equation (A2.2.12) for $D_I(L) = (1-L^4)$ with each deterministic variable from tables A2.2.2, A2.2.3 and A2.2.4 and the products of these deterministic variables and $D_{l,t}$. These dummy variables can be viewed as structural series of innovation outliers, or as a *permanent innovation outlier*, in the context of the model of chapter 5. Put $O_2 = 1$ for $l = t+1, t+2, \ldots, T$ in (5.1) to get a *structural break at frequency zero.* Note that we do not consider tests for these types of outliers in chapter 5. Ordinary t–tests on these dummy variables have nonstandard distributions under the null, see e.g. Perron (1991).

Perron ((1989), (1990a), Perron and Vogelsang (1992b)) derived the asymptotic distribution of the zero frequency unit root test statistics from this extended regression[6]. The critical values are larger in absolute value. The absolute deviation of the critical values of these modified test statistics from the unit root critical values from table A2.2.4 is a decreasing function of $\text{abs}((l-0.5T)/T)$. We use the 5% critical values for $l = 0.5$ from Perron (1989, 1990a). The values for frequency ½ are the just the negatives of those for the zero frequency. Perron did not compile the critical values for the complex unit root test statistics in the test regression with a structural break.

The critical values for some of the statistics that we intend to use are not published as far as we know. Fortunately it is easy to estimate the

---

[6] Perron and Vogelsang (1991a, footnote 4, 1992b) noted that introduction of a single one-time dummy (=1 for $t=l+1$, 0 elsewhere) in the test regression is necessary if one wants to apply the same asymptotic theory to permanent *additive* outlier models. Perron's claims in (1990a) were false in this respect. One may suspect that our additive outlier correction picks up a substantial part of the effect of the introduction of this extra dummy.

critical values in a small Monte Carlo experiment[7].

We estimated the 5% asymptotic critical values for the test statistics for the complex unit root hypotheses in 10,000 replications of the test in samples of size 500. The data generating process was $y_t - y_{t-4} = \varepsilon_t \sim$ n.i.d.(0,1). Extra deterministic structural change variables were included as indicated above.

Table 7.1 *Percentiles nonstandard distributions allowing for a halfway break*

	Probability of a Smaller Value					
	Sample size 500			Limit distribution		
Distribution	0.01	0.05	0.10	0.01	0.05	0.10
(1) $\tilde{\tau}_{\mu 1}$	-3.89	-3.33	-3.04	-3.90	-3.34	-3.04
(2) $\tilde{\tau}_{\tau 1}$	-4.80	-4.20	-3.92	-4.90	-4.24	-3.96
(3) $\tilde{\tau}_{\mu 2}$	-4.72	-4.10	-3.77			
(4) $\tilde{\nu}_{\mu 1}$	-25.2	-18.3	-15.3	-26.1	-18.9	-15.8
(5) $\tilde{\nu}_{\tau 1}$	-40.8	-31.5	-27.5	-44.1	-33.8	-29.4
(6) $\tilde{\nu}_{\mu 2}$	-35.3	-27.3	-23.5			
(7) $\tilde{\nu}_{\mu 2}^*$	-36.2	-27.7	-24.0			

NOTES: Left hand side panel freshly compiled in 10,000 simultaneous test regressions, with a (shifting) trend included only for rows (2) and (5). Right hand side panel from Perron (1989, Tables VI.A, VI.B); 1990a Table 3); Right hand side figures correspond closely to values for sample size 200 available in Perron (1990a). Last row gives critical values based on Fountis and Dickey's maximum eigenvalue approach, which probably gives a better approximation of the limit distribution. $\tilde{\nu}$ critical values seem to decrease (substantially) with sample size. Haug (1992) gave corresponding results for the Phillips-Ouliaris test.

We also employed this experiment to examine the validity of the test of Fountis and Dickey (1989) applied to frequencies ½ and ¼ as described in §A2.2.2.1. Let ρ_1 and ρ_2 denote the inverse roots of the OLS estimator of the AR(4) polynomial[8] with real parts closest to 1 and –1. $T(1-\rho_1)$ and $T(1-\rho_2)$ do seem to have the same limit distributions as $-Tb_1$ and Tb_2 in regression equation A2.2.12. Let ρ_4^2 denote the square of the inverse root with imaginary part closest to i (i.e. $(-1)^{\frac{1}{2}}$). $T(1-\rho_4^2)$ has the same limit distribution as Tb_4. In the simulation the finite sample critical values for the Fountis–Dickey approach ended up between the Dickey–Fuller numbers and the

[7] Note that *bootstrapping* the distribution of this test statistic on the basis of the least squares estimator of the autoregressive parameters is not a valid procedure under the null of unit roots, cf. Harris (1992).

[8] Equivalently one can speak of the eigenvalues or characteristic roots of the companion matrix of the AR model with real part closest to 1, –1, see §3.3.9. One can call the Fountis-Dickey method the *maximum eigenvalue* approach.

values of the corresponding limit distributions given in Perron (1989, 1990a).

Table 7.1 contains a selection of critical values for $l = 0.5T$. The critical values of the unit root test statistics in a simultaneous test regression do not tend as fast to the limit values as the critical values in a test regression for a unit root at only one frequency. This follows both from a comparison of table 7.1 with the tables in Perron (1990a) and from a comparison of Table A2.2.4 with Table 1 in Hylleberg et al. (1990). The asymptotic independence of the regressors in test regression (A2.2.12) is apparently not reached in small samples. Compared to other approximation errors this difference can be considered minor, however.

7.4.6 Application

In our application we look at the univariate dynamics in four stages. We difference the data to get approximations of growth rates[9]. First we consider these growth rates corrected for additive outliers, see §A5.2. Then we look at the rates adjusted for both additive outliers and for (permanent additive) quarterly deterministic seasonality (constant, $(-1)^t$, $\cos\frac{1}{2}\pi t$, and $\sin\frac{1}{2}\pi t$). Subsequently we examine the "whitened" rates adjusted for deterministic seasonality with a break at 1976.4/1977.1.

The break is part of the measurement model. We use data from 1977.1 onwards from the national accounts. The models for the OECD data in our set seem to need this break even more, see the statistical analysis below. Most probably this is a measurement issue as well. Looking at the data it seems obvious that one has to allow for a change in the mean growth rates. Conditional on a break the series can be considered $I(1)$ whereas one would suspect it to be $I(2)$ without one. Statistical criteria like the maximum F-test (Andrews (1990)) lead to dates for the change points that differ between the series. Growth in private consumption slowed down significantly after 1980 mainly because of the first time post war decreases in total labor income[10]. Mean growth in exports already declined after 1973. The choice for one break at 1977.1 is a compromise and avoids the "testimator" problems mentioned in §3.3.2. Acting "as if" one has no a priori ideas about the timing of the break has severe repercussions for the statistical analysis, most notably in the important unit root testing area, see e.g. Zivot et al. (1992)

---

[9] The "delta log" transformation and relative growth rates differ only for large relative changes.

[10] The distribution of income between capital and labor plays an important part in the Dutch macroeconomic modeling tradition. Measurements of changes in the income shares are significantly influenced by changes in revenues from sales of natural gas.

and Banerjee et al. (1992).

The final transformation of the univariate series we consider is the differenced series after correction for additive outliers and seasonality. By applying the Census X–11 filter one removes $I_{\frac{1}{2}}(2)$ and $I_{\frac{1}{4}}(2)$ components. It is an alternative method to adjust for a changing seasonal pattern. It smooths sudden changes. This corresponds more closely to the prior beliefs of many about macroeconomic seasonality. Details of this *univariate seasonal correction* are in the notes of table 7.2 below. Outlier tests and unit root tests are done using a "permanent innovation outlier model" for the breaks in seasonal pattern and mean growth.

7.4.7 Results

This section contains the empirical results of the univariate analysis.

7.4.7.1 Outliers

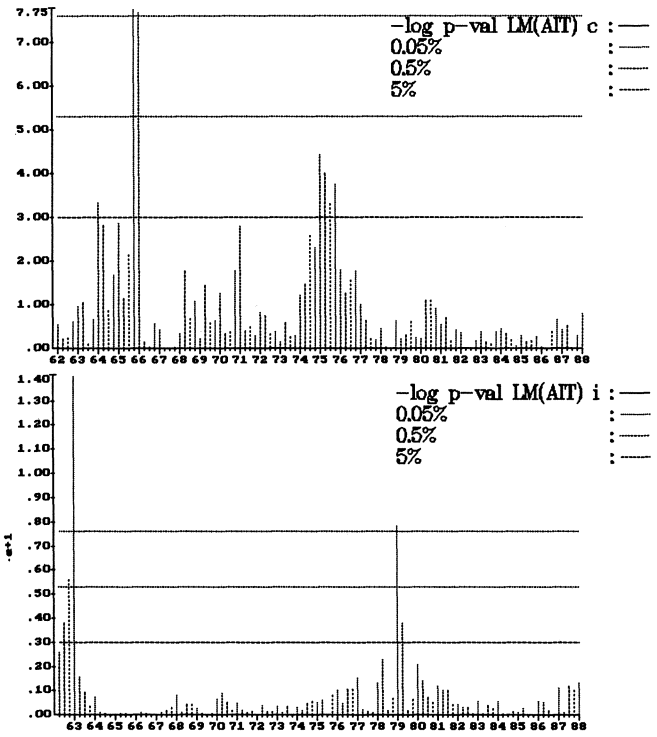

Figure 7.2 *-log p-values of simultaneous univariate LM test for outliers in the log levels* of consumption (c) and investment (i). Based on asymptotic χ^2_3 approximation.

The time series show a number of outliers. Outlier tests confirm this. Even using a significance level of $(0.05/T)$ as suggested in §3.3.6 and §3.3.7

we find several outliers. The simultaneous test of chapter 5, applied to the *levels* with $\rho = 0.95$ rejects the null for the c series in 1965.4 and 1966.1 and for the i–series in 1963.1 and 1979.1, see figure 7.2.

Normality of the residuals is rejected at small significance levels as well. In the principal component analysis of seasonality the same observations have a huge influence. The empirical influence curve $\hat{I}_{t1}$ (see A3.3.2) for the largest eigenvalue λ_1 (i.e. the variance of the first principal component) of the covariance matrix of the first differences indicates this too. Values for $\hat{I}_{t1}/\lambda_1$ are 5.9 for Δc in 1966 and 23.8 for Δi in 1963! Some form of outlier correction is essential. We employ the Kleiner–Martin method introduced in §A5.3 with a tuning constant of 2.0. Its application renders residuals and influence curves normal. So we discuss some results of the next transformation.

7.4.7.2 Autocorrelations

We examine the (extended) sample autocorrelation function (up to ARMA(20,10) models), the partial autocorrelation function (up to lag 30) , the periodogram and four spectral estimators for the "cleaned" growth rates, adjusted for deterministic seasonality. From the partial autocorrelation function we find that AR models of order 10 can give reasonable approximations of the correlation structure of the series. We check this in the frequency domain by comparing four different spectral estimators, the first of which is the one based on the AR model. The other ones are Parzen estimates, based on a weighted autocovariance function with window widths 40 and 8 and a Daniell estimate based on a weighted periodogram with window width 7. See Koopmans (1974) for a detailed discussion of these estimators.

The resulting AR, Parzen(40) and Daniell estimates are quite similar in shape. Differences in the log spectrum are in the order of 0.10. Differences up to 0.50 occur at the zero frequency anyhow. Those deviations are well in the range of the 90% confidence interval of the Parzen(40) estimate, which we computed using techniques from Koopmans (1974). The Parzen(8) estimator seems to smooth too much. This makes applications of unit root tests which use even fewer autocovariances in their estimates of the zero frequency doubtful. The differences between the estimates also indicate that data information about unit root characteristics is rather diffuse. Note that the spectral density at frequency zero (not its log!) is a multiplicative correction factor in some unit root tests, see §A2.2. Figure 7.3 contains an example of the plots used in a graphical analysis of the autocorrelation function.

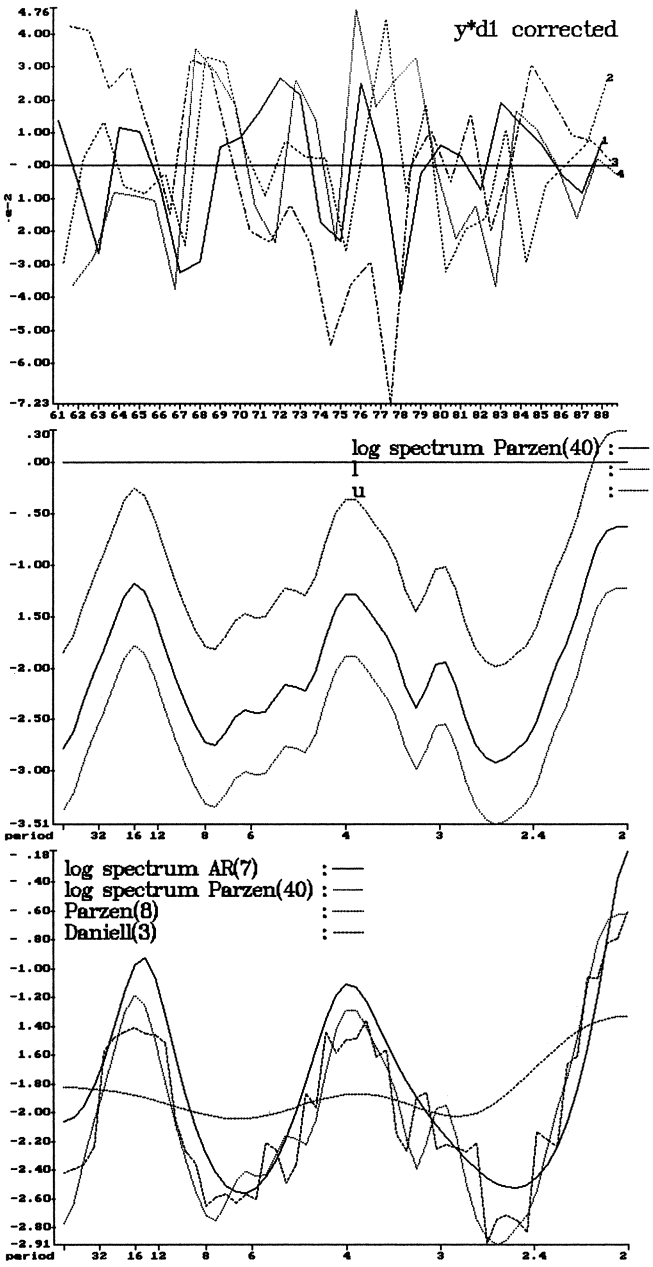

Figure 7.3a *Univariate representations of German industrial production growth* *y*d1*: Time series of growth rates adjusted for temporary additive outliers, and for deterministic seasonality with one permanent shift in 1977.1, cf. unadjusted series in fig. 7.1d. Parzen(40): log spectral density using the Parzen weighted autocovariance estimator with 90% confidence interval. Next four different estimates of the spectral density, discussed in text.

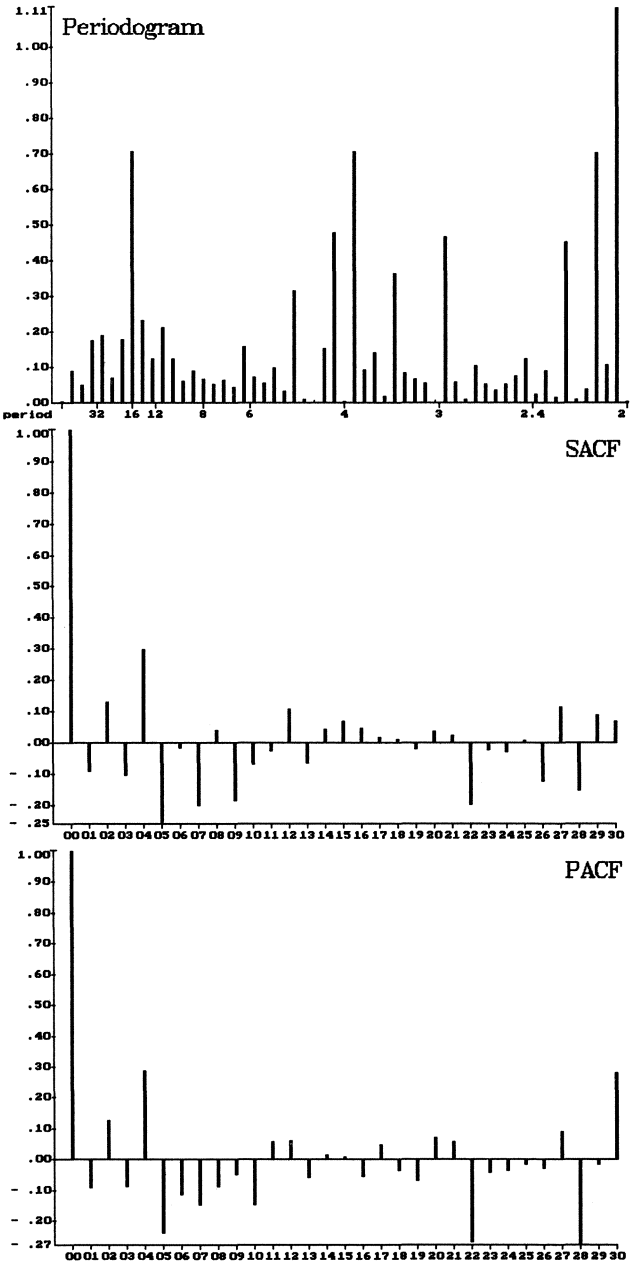

Figure 7.3b *Univariate representations of German industrial production growth (continued)* Time series characteristics of y*dl (fig. 7.3a): Periodogram, SACF: autocorrelations up to lag 30, PACF: Partial autocorrelations up to lag 30.

7.4.7.3 Long memory analysis

The estimates of fractional integration parameters given in the upper panel of table 7.2 do not indicate severe antipersistence of any of the growth rates. The negative sign of $\hat{d}_0$ for Δy^* is due to the poor resolution of this simple estimator and a business cycle peak in the spectrum at frequency 1/16, see figure 7.3. The negative estimate for $\Delta(px/pm)$ seems to be more reliable in this respect. All variables show signs of positive integration at the seasonal frequencies. Δc and Δx seem to be nonstationary at frequency ¼, Δy^* seems to be nonstationary at frequency ½.

Table 7.2 *Regression estimates alternative long memory models*

	Δc	Δi	Δm	Δx	Δy^*	$\Delta(px/pm)$
$\hat{d}_0$	0.8	0.8	0.4	0.4	-0.3	-0.3
$\hat{d}_{\frac{1}{2}}$	0.9	0.9	0.8	0.6	1.4	0.3
$\hat{d}_{\frac{1}{4}}$	1.2	0.6	0.0	1.6	0.5	0.2
$\hat{H}$	0.5	0.4	0.6	0.6	0.3	0.3
Lo	0.66	0.93	0.46	0.70	0.97	0.95
	$\Delta s(L)c$	$\Delta s(L)i$	$\Delta s(L)m$	$\Delta s(L)x$	$\Delta s(L)y^*$	$\Delta s(L)(px/pm)$
$\hat{d}_0$	0.8	0.8	0.4	0.4	-0.3	-0.3
$\hat{d}_{\frac{1}{2}}$	-0.3	-0.2	-0.9	-1.0	-0.1	-0.6
$\hat{d}_{\frac{1}{4}}$	-0.8	-0.8	-0.7	-0.3	-0.8	-0.9
$\hat{H}$	1.0	0.7	0.7	0.8	0.6	0.4
Lo	0.18	0.89	0.39	0.42	0.63	0.87

NOTES: $\hat{d}_\lambda$: regression estimate of slope coefficient log periodogram in neighborhood frequency $2\pi\lambda$ on theoretical log (pseudo) spectrum $I_\lambda(1)$ variable, given in (2.17). Time domain sample: 1961.1-1988.4. Window widths frequency domain estimates: 8, 8 and 11 for $\hat{d}_0$, $\hat{d}_{\frac{1}{2}}$ and $\hat{d}_{\frac{1}{4}}$. Frequencies 0, π, and $\frac{1}{2}\pi$ excluded. Variables corrected for additive outliers as in Table 7.5. $\hat{H}$: estimate of Hurst coefficient from regression estimate of slope coefficient of logs of means of nonoverlapping subsample rescaled adjusted ranges on logs of the (sub)sample sizes 13, 15, 18, 22, 28, 37, 56, 112. Alternative estimate of $\hat{d}_0+\frac{1}{2}$. Lo: Lo statistic's asymptotic p-value for the test of the $I(0)$ hypothesis against anti persistence based on the modified rescaled range for full sample, see (2.30). High p-values are an indication of positive integration. Spectrum at frequency zero estimated using Bartlett weighted covariance estimator with window width 32. Reject null hypothesis $d_0=0$ for high p-values against $d_0>0$. Reject against $d_0<0$ for low p-values. s(L): linear part of the Census X-11 filter as described in §4.5. Out-of-sample forecasts and backforecasts generated using equations from Table 7.5, estimation period 1963.1/2/3-1988.4, i.e. 4 observations shorter, in order to use the same information as upper panel, and in order to avoid the influential first observation in Δc, which cannot be adjusted using our outlier correction method.

The rescaled adjusted range analysis does not point to positive integration for the growth rates. The last two Hurst parameter estimates smaller than 0.5 correspond nicely with the negative estimates for $\hat{d}_0$. The

parameters do not indicate severe "overdifferencing" at the zero frequency anyway. The Lo statistics do not either.

The theoretical effect of the application of the linear part of the Census X–11 filter shows clearly in the lower panel of table 7.2. The order of integration at the zero frequency is barely affected. The order of integration at the seasonal frequencies should be decreased by two. Although the properties of these estimators are not known outside ranges of integration [–½,¼] a decrease in the order of integration of at least one is seen in most entries.

The seasonal adjustment has a marked influence on the Hurst coefficients. All estimates increase. The outcomes for the seasonally adjusted data are more closely in line with the fractional integration parameter estimates. Lo's statistics go down mainly because of changes in the numerator of expression (2.30). For consumption growth which has a relatively big variance at the seasonal frequencies, the adjusted range goes down by about 20 percent. The numerator increases by about 10 percent. This affects the p–values considerably. Table 7.2 shows the dangers of the straightforward application of adjusted range techniques on seasonally unadjusted data.

Practically we can conclude from the long memory analysis that the $I(1)$ hypothesis is reasonable for the zero frequency. The first four series may be integrated of a somewhat higher order. The $I(0)$ hypothesis for the seasonal frequencies must be rejected for c, x and y^*.

7.4.7.4 Data analysis seasonal components

We can draw similar conclusions from the results of a principal component analysis of the quarterly subseries of the growth rates presented in table 7.3. The eigenvector α_1 of the largest eigenvalue of the covariance matrix is not parallel to (1 1 1 1): variation at other frequencies dominates the variance at the zero frequency for all variables. Nonstationarity of Δx at frequency ¼ shows up clearly again: α_1 is a virtually exact linear combination of (1 0 –1 0) and (0 1 0 –1). The first component explains about two thirds of the total variation. Similar effects occur for Δy^* at frequency ½, whereas Δc seems to be nonstationary at both seasonal frequencies, although there this effect seem to change periodically. The Δc series is thus the only one for which the usual yearly moving average filter $(1+L)(1+L^2)$ seems to be reasonable at this point of the analysis.

Time series plots of the first principal components of each variable (not reproduced here) show random walk like behavior for Δc, Δi and Δx, in agreement with the aliasing phenomenon discussed in §2.4.2. The first

principal component of Δm seems to be stationary, whereas the series for Δy^* and $\Delta(px/pm)$ show marked one time permanent shifts in mean in 1978 and 1987[11] respectively.

Table 7.3 *Principal component analysis of quarterly subseries*

	Δc	Δi	Δm	Δx	Δy^*	$\Delta(px/pm)$
$\lambda_1/\mathrm{tr}(V'V)$	.69	.53	.48	.66	.66	.45
r_0^2	.08	.01	.17	.00	.01	.01
$r_{\frac{1}{2}}^2$	.55	.10	.80	.00	.75	.69
$r_{\frac{1}{4}}^2$	.38	.89	.03	.99	.24	.30
R_{11}^2	.85	.82	.16	.67	.07	.59
R_{21}^2	.75	.45	.81	.57	.83	.89
R_{31}^2	.20	.48	.11	.74	.78	.07
R_{41}^2	.68	.04	.58	.68	.31	.03
max $\hat{I}_{t1}$	1984	1963	1980	1963	1984	1987/88

NOTES: Application on quarterly subseries v_k, k=1, 2, 3, 4 as introduced in §4.4. $V'V$: Empirical covariance matrix. Sample 1961.1-1988.4. Series corrected for additive outliers as in Table 7.5. λ_1: largest eigenvalue of $V'V$. r_i^2: squared (multiple) correlation of eigenvector α_1 belonging to λ_1, with (1 1 1 1) (r_0^2), (1 -1 1 -1) ($r_{\frac{1}{2}}^2$) and both (1 0 -1 0) and (0 1 0 -1) ($r_{\frac{1}{4}}^2$) respectively. R_{kj}^2: Proportion of variance of v_k explained by principal component j. max $\hat{I}_{t1}$: observation for which empirical influence curve (A3.3.2) for first principal component is maximum.

7.4.7.5 Variance time functions

As a last data analytic tool we use the variance time function V_N/N (not reproduced here), see (2.21)–(2.23). For the first five series *in seasonally corrected levels* this function increases up to $N=36$ and clearly has a positive limit, although there are too few observations to say whether this limit is statistically significantly different from zero. This result is in line with the findings of Cogley (1990) for century long per capita GNP series of European countries like France and Italy, which have substantially bigger random walk components than the corresponding series for the U.S.A. The limit for consumption is extremely high, the scaled variance time function[12] $V_N/(NV_1)$ reaches values around 10. The function decreases for px/pm up till $N=16$, but

[11] "Permanent" also in light of post sample evidence. We discuss an influence analysis below.

[12] The univariate outlier correction decreases $\hat{V}_1=\hat{\sigma}_x^2$ substantially, whereas the long run variance is not influenced as much. The long run variance is influenced significantly by correction for permanent outliers (structural break dummies), which we introduce below.

levels off at a rather high value: the random walk component is not very small.

For the seasonally adjusted *growth rates* of the series all variance time functions decrease. They are not significantly different from zero for values of $N>20$, if one employs the crude standard errors given in Cochrane (1988) to test the significance. The variance time function analysis again leads to the conclusion that one time differencing is a good idea.

7.4.7.6 Statistical unit root analysis

The statistical unit root analysis first concentrates on the $I(2)$ hypothesis at the zero frequency. We present relevant results in table 7.4. The $I(2)$ hypothesis cannot be rejected for c and x. The t_{b1}–values for these variables can be brought down to levels below –3.4 by changing the date of the break to 1980.1 for consumption and to 1974.1 for exports. Rejection of the $I(2)$ hypothesis is harder in 9 year subsamples. One cannot reject the $I(2)$ hypothesis either if one increases p to values of, say, 20. More lags per observation reduce the power of unit root tests considerably. The additive outlier correction does not have a decisive impact on the outcome of this unit root test.

Table 7.4 *Regression tests for $I(2)$ hypothesis*

	Δc	Δi	Δm	Δx	Δy^*	$\Delta px/pm$	5%–c.v.
p	8	9	8	10	8	8	
t_{b1}	-3.19	-3.44*	-3.49*	-3.15	-4.52*	-4.56*	-3.4
$\hat{\sigma}$ (·01)	1.05	4.90	2.79	3.22	1.98	1.62	
AC(4)	0.13	0.08	0.26	0.35	0.67	0.55	
Norm	2.9	0.53	1.48	1.63	1.88	1.07	5.99
$t_{d\delta 0}$	-2.12	-0.67	-2.00	2.06	-1.36	-0.29	±2.9
N_4	0	1	2	3	2	2	4
N_8	0	1	1	0	1	2	3
N_{16}	0	1	0	0	1	1	2
N_{36}	0	0	1	0	0	0	2

NOTES: Regression (A2.2.12) for $D_I(L)=(1-L^4)(1-L)$, extended with terms $d\delta_0 D_{77.1}$, $d\delta_2 D_{77.1}(-1)^t$, $d\delta_3 D_{77.1}\cos(\frac{1}{2}\pi t)$, and $d\delta_4 D_{77.1}\cos(\frac{1}{2}\pi t)$ and restriction $\delta_1=0$. Data corrected for one time additive outliers *in levels* using method of Kleiner and Martin (1979), as in table 7.5. p: order of AR model. t_{b1}: t-statistic to test $I(2)$-hypothesis. $\hat{\sigma}$: OLS estimate standard error. AC(4): Breusch-Godfrey test against fourth order serial correlation in the residuals, p-value. Norm: Bera-Jarque test for normality residuals. N_i: number of outliers of size i at 5%-level in regression *without* "break" dummies, see (3.13). Only subsamples of $\Delta(px/pm)$, which include 1987.1–1988.4 have p-values<0.01. Critical values for unit root tests in last column simulated in 10,000 replications of sample size 116, see table 7.1.

Table 7.5 *Regression tests for $I(1)$ hypothesis*

	c	i	m	x	y^*	px/pm	5%–c.v.
p	8	9	8	10	8	8	
t_{b1}	-3.29	-0.92	-1.79	-2.07	-2.00	-2.13	-4.3
t_{b2}	3.11	2.90	3.31	3.01	3.76*	2.76	3.4
t_{b4}	2.37	3.86	3.73	1.57	3.26	4.46*	4.2
$\min t_{b1}$	-3.01	-2.47	-2.34	-1.73	-3.13	-2.66	-4.3
in year	71.4	88.4	72.3	88.4	71.4	86.4	
$\max t_{b2}$	2.59	2.88	2.44	3.11	2.72	5.33	
in year	76.3	88.2	88.2	87.1	78.2	86.4	
$\max t_{b4}$	2.05	2.24	3.68	1.69	3.12	3.83	
in year	88.4	76.3	87.1	88.4	88.4	86.3	
$T\rho_1$	-8.3	-4.6*	-17.1*	-9.6*	-21.7*	-15.1*	-31.6
$T\rho_2$	-19.0	-32.9*	-28.1*	-26.8*	-12.0	-42.0*	-19.3
$T\rho_4^2$	-16.4	-42.8*	-55.7*	-6.5	-48.2*	-65.3*	-27.4
$\hat{\sigma}$ (·01)	0.99	4.78	2.74	3.14	1.94	1.58	
AC(4)	0.17	0.09	0.12	0.62	0.64	0.17	
Norm	2.9	0.53	1.48	1.63	1.88	1.07	5.99
$t_{d\delta 0}$	2.71	-0.74	0.85	1.22	1.16	-0.99	±2.8
$t_{d\delta 1}$	-2.88	0.69	-1.12	-1.48	-1.34	0.61	±3.6
$t_{d\delta 2}$	-3.04*	0.67	-2.29	-1.56	-4.00*	-0.54	±2.9
$t_{d\delta 3}$	2.46	-0.72	0.75	0.81	2.08	1.62	±3.0
$t_{d\delta 4}$	0.69	-3.83*	1.01	-0.68	1.97	2.52	±3.0
N_4	2	1	2	3	2	4	4
N_8	1	1	0	1	1	2	3
N_{16}	1	0	0	0	0	1	2
N_{36}	0	0	1	0	0	1	2

NOTES: Regression (A2.2.12) for $D_I(L)=(1-L^4)$ extended with terms $d\delta_0 D_{77.1}$, $d\delta_1 D_{77.1}(t-T/2)$, $d\delta_2 D_{77.1}(-1)^t$, $d\delta_3 D_{77.1}\cos(\frac{1}{2}\pi t)$, and $d\delta_4 D_{77.1}\cos(\frac{1}{2}\pi t)$. Data corrected for one time additive outliers using method of Kleiner and Martin (1979), see §A5.3, with tuning constant 2.00. Sample period 1962.1–1988.4 for c, m, y^*, and px/pm. Sample for i starts in 1962.2. Sample for x starts in 1962.3. $\min t_{b1}$: minimum (in: period) of recursive estimates of t_{b1} 1971.1–1988.4, *shift dummies not included*, see Banerjee et al. (1992). $\max t_{b2}$, and $\max t_{b4}$ defined analogously. Distribution not simulated. $\hat{\sigma}$: OLS estimate standard error disturbances. AC(4): p-value Breusch-Godfrey test against fourth order serial correlation in the disturbances. Norm: Bera-Jarque test for normality residuals. N_i: number of outliers of size i at 5%-level in regression *without* "break" dummies, see (3.13). True level may be somewhat higher do to presence of integrated regressors, see §3.3.6. 5% critical values for unit root tests in last column simulated in 10,000 replications of sample size 116, allowing for halfway break. DGP: $y_t-y_{t-4}=\varepsilon_t \sim \text{nid}(0,1)$. Critical values for t-values of breaking mean, trend and seasonal mean pattern, $t_{d\delta i}$, computed from same simulations. Negative values indicate a decrease in mean ($t_{d\delta 0}$) mean growth ($t_{d\delta 1}$), amplitude frequency $\frac{1}{2}$ component ($t_{d\delta 2}$) and amplitude frequency $\frac{1}{4}$ components ($t_{d\delta 3}$ and $t_{d\delta 4}$). Robustness of validity of latter statistics largely unknown. Critical values merely indicate the nonnormality of the tails of their distributions under the null. 1977.1 is somewhat away from the center of the sample, so "true" critical values for unit root tests may be a little smaller in absolute value.

In the next stage of the unit root testing we consider the $I(1)$ hypothesis at three frequencies, presented in table 7.5. At the zero frequency we cannot reject the $I(1)$ hypothesis at the 5% level. Banerjee et al. (1992) tabulated the distribution of the minimum of the recursive estimates of t_{b1} in the regression without break dummies. One does not need a priori information on the timing of the break to compute this unit root test.

For the other frequencies the results are mixed. Using the Dickey–Fuller tests we can only reject the $I_{\frac{1}{2}}(1)$–hypothesis for y^* and the $I_{\frac{1}{4}}(1)$–hypothesis for px/pm. Using the Fountis–Dickey tests, $T\rho_2$ and $T\rho_4^2$, we can reject the $I_{\frac{1}{2}}(1)$–hypothesis for i, m, x, and px/pm and we reject the $I_{\frac{1}{4}}(1)$–hypothesis for i, m, y^* and px/pm. The latter results correspond to the results for the long memory models and the principal component analysis where only c and y^* seemed to be nonstationary at frequency ½, whereas c and x were nonstationary at frequency ¼.

The reverse outcome between the Dickey–Fuller and the Fountis–Dickey test on the $I_{\frac{1}{2}}(1)$ hypothesis is remarkable. From a plot of the estimated spectrum of Δy^* corrected for the significant break after 1977 in fig. 7.3a above we find confirmation of the Fountis–Dickey test result. The overall results suggest a bigger power of the Fountis–Dickey test. The robustness of validity of the test has to be examined more precisely in order to judge the practical relevance of this suggestion.

The choice of lag length has a considerable influence on the outcomes of the Dickey–Fuller tests. More lags decrease the number of rejections. Only the adjusted partial autocorrelation function (not reproduced here) of Δm indicates that fewer lags could be used. From the autocorrelation functions we may conclude that Δm adjusted for additive outliers and structural breaks is very close to white noise. Lowering p to 4 gives values of 5.8 for t_{b2} and 7.1 for t_{b4}, without introducing serial correlation in the residuals. Δm *cannot* be considered white noise without allowance for the breaks, in particular removal of the break in the frequency ½ component introduces serial correlation.

7.4.7.7 Parameter stability and influence analysis

The results for the "raw" data are not reproduced here. We find the additive outlier correction to matter decisively only for i. The $I_{\frac{1}{4}}(1)$ hypothesis can be rejected without the additive outlier correction, whereas the deterministic change in the frequency ¼ component becomes significant at the 5% level only after the additive outlier correction.

The evidence of breaks, see $t_{d\delta 4}$ and $t_{d\delta 2}$, is statistically significant at even lower levels for the frequency ¼ component in i and the frequency ½

component in y^*. The break in the investment series seems to be a consequence of the change of measurement model for the OECD series from 1977.1.

The sensitivity of the Dickey–Fuller test to the end point of the sample can be assessed from the extreme values for the recursive estimates reported in table 7.5. The change in the test statistic for the $I_{\frac{1}{2}}(1)$ hypothesis for $\Delta(px/pm)$ is remarkable, see $\max t_{b2}$.

A spectral analysis of the *standardized recursive residuals* (not reproduced here) shows significant peaks at frequency $2\pi/2$ for y^* and px/pm. These are also due to the change in seasonal patterns after 1976 in y^* and after 1986 in px/pm, see figure 7.1 and table 7.3. As discussed in chapter §3.3.6 ordinary CUSUM tests do not pick up this change in seasonal pattern.

It is easy to adapt the test to detect changes in parameters which model the frequency ½ component. The CUSUM–OLS test applied to the partial sum series of *alternating* OLS residuals of the y^* equation, $\sum(-1)^t e_t$, gives[13] a p–value of 8%, whereas the CUSUM–OLS test gives a p–value of 66%. The change also shows clearly in time series plots of the recursive estimates of b_2 and δ_2 (not reproduced here). The change is (shortly) after the middle of the sample.

A similar adaptation of the CUSUM–test on partial sums of alternating recursive residuals does not have as much power. This analysis shows how one can improve the diagnostic testing procedure by using the right test for the right problem. Just doing one test for parameter stability is usually not enough.

The outlying character of the subset 1987.1–1988.4 for $\Delta(px/pm)$ is also pointed out by an *influence analysis* of the test equation for integration of $\Delta(px/pm)$ without "break" dummies, used for table 7.4. Note that the regressors of that equation do not exhibit unit root nonstationarity. The location component of the influence of this subset on the predictive density of $\Delta\Delta_4(px/pm)$, *locPI* (see appendix 3.1, §A3.1.1) is 4.8, whereas the components of the other subsets are all smaller than 0.7. The estimative influence measure relative to σ^{-2}, $MEIM\Sigma$ (see appendix 3.2) equals 4.3 compared to a maximum of –0.1 for the other nonoverlapping subsets of size 8.

The Chow test for predictive failure PF_i (see appendix 3.1, §A3.1.1) which is sensitive to changes in β and σ has a p–value of 0.00030. This is

---

[13] Note that this sum over the whole sample equals zero, because of the inclusion of $(-1)^t$ as a regressor. The asymptotic distributions of the extremes of $\sigma t^{-\frac{1}{2}}\Sigma_1^t e_i$ (given in Ploberger and Krämer (1992)) and $\sigma t^{-\frac{1}{2}}\Sigma_1^t(-1)^i e_i$ are the same under the null, because of the absence of autocorrelation and the symmetry of the distribution of e_t. See Canova and Hansen (1992) and §4.2 for a comparable test.

significant at an overall level of 5%. The total number of nonoverlapping subsets of size 8 is 13, $0.05/13 = 0.00038$, see §3.6 and §3.7. For subsets of size 4 the minimum p–value is 0.015. This indicates again that influence analysis should be done for different subsizes.

The test equations for the $I(1)$ hypotheses are also used to make the univariate out of sample (back)forecasts which are necessary for the application of the Census X–11 filter, which mitigates the effect of the changes in seasonal pattern.

7.4.7.8 Summary of univariate results

All series are $I(d_0)$, with $d_0 > 0.5$. $d_0 < 2$ for i, m, y^* and px/pm. c and y^* are $I_{\frac{1}{2}}(d_{\frac{1}{2}})$ with $d_{\frac{1}{2}} > 0.5$, whereas $d_{\frac{1}{2}} < 1$ for i, m, x, and px/pm. c and x are $I_{\frac{1}{4}}(d_{\frac{1}{4}})$ with $d_{\frac{1}{4}} > 0.5$. Both temporary and permanent outliers have a significant impact on the analysis and need to be taken into account. Δm is approximately white noise after correction for deterministic components. The other series are not. The spectral density of Δy^* has a significant peak at the business cycle frequency of 1/16. The other series have not.

7.5 Multivariate analysis

7.5.1 Predictions and seasonality in the unrestricted VAR

First we apply the multivariate outlier tests of chapter 5. A selection of the results is presented in figure 7.4. The multivariate tests give greater p–values than the univariate tests. The outcomes again indicate the need for outlier correction. We make first round models to generate (back)forecasts which are needed for the application of the seasonal adjustment filter. In these first round models we abstain from imposing unit roots. Instead we introduce some extra dummy variables to take account of permanent changes in mean growth and seasonal means, see (7.3). We use a similar model to estimate the additive outlier component. We employ a rather strict tuning constant of 2 in order to get fairly smooth estimates of the seasonal components.

$$y_t = (I - \Phi(L))y_t + m_0 + m_{0d}D_t + (m_1 + m_{1d}D_t)t + (m_2 + m_{2d}D_t)(-1)^t + (m_3 + m_{3d}D_t)\cos(\tfrac{1}{2}\pi t) + (m_4 + m_{4d}D_t)\sin(\tfrac{1}{2}\pi t) + \varepsilon_t, \tag{7.3}$$

$$y_t = x_t - u_t,$$

with

x_t the observed series,

u_t the additive outlier component,

$\Phi(L)$ an AR matrix polynomial, and

D_t a dummy variable equal to one for t after 1976.4 and zero elsewhere.

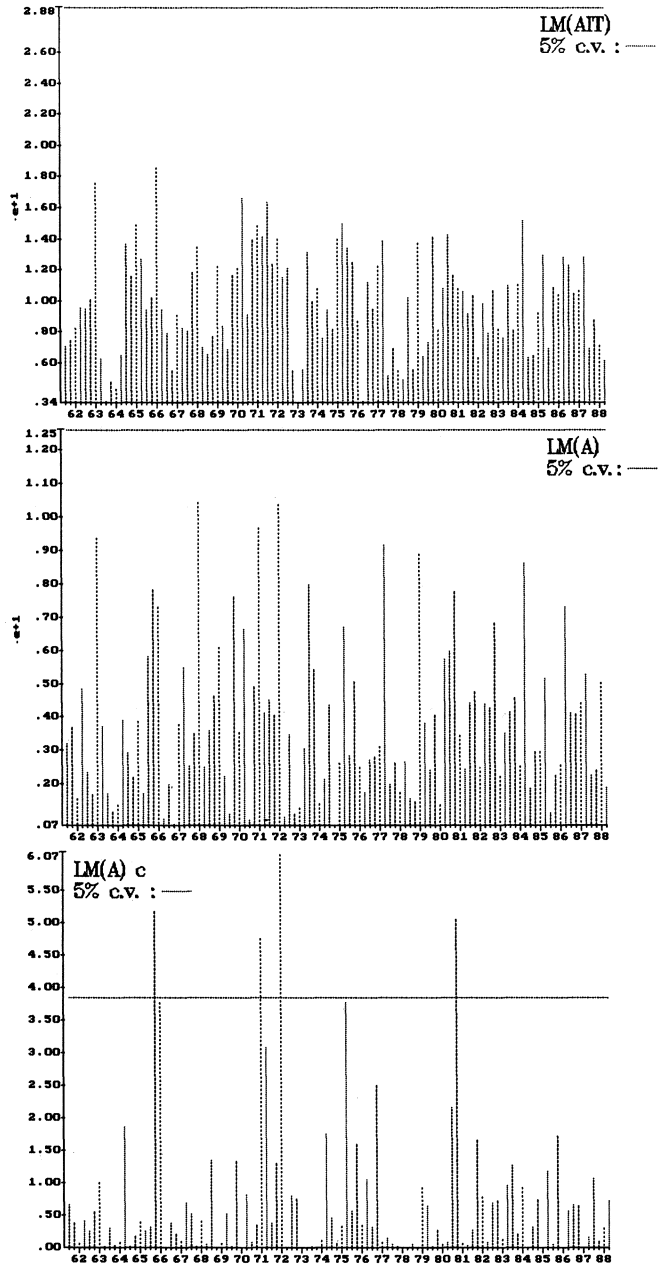

Figure 7.4a *Results of multivariate outlier tests Dutch data*
See §5.4.1 for LM(AIT) and LM(A) defined in (5.6) and (5.9). See §5.4.4 for outlier tests in different equations. 5% critical values for case with known timing.

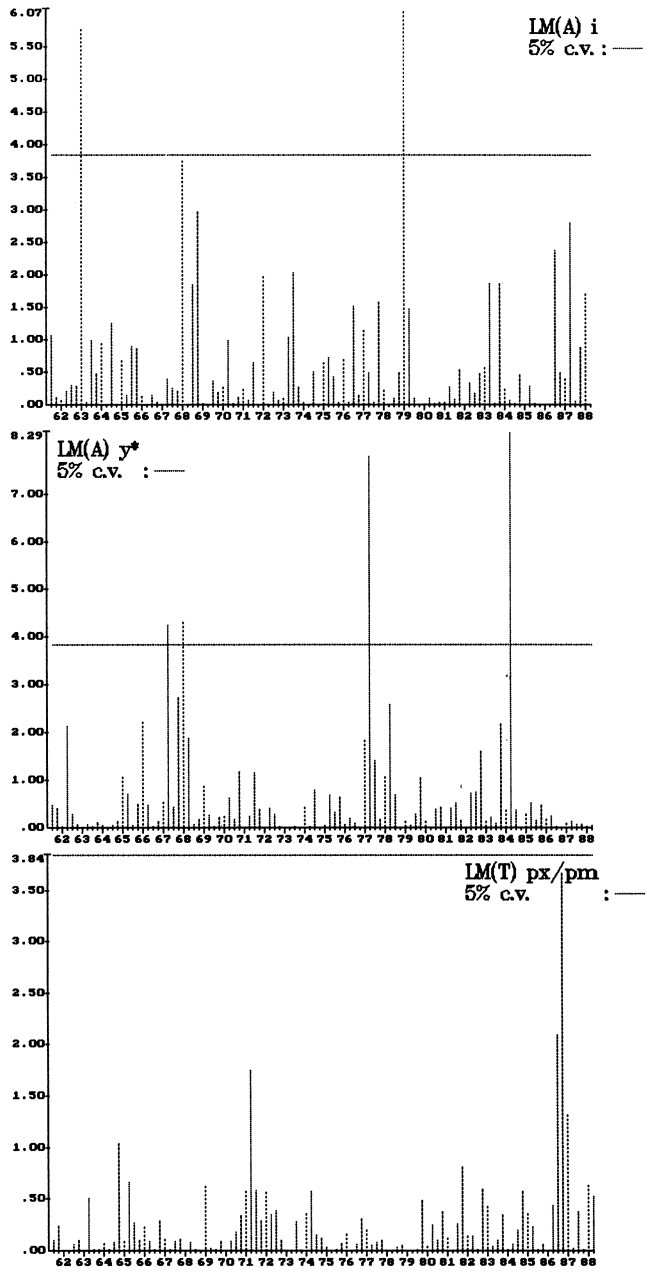

Figure 7.4b *Results of multivariate outlier tests Dutch data (continued)*
See §5.4.4 for outlier tests in different equations. 5% critical values for case with known timing

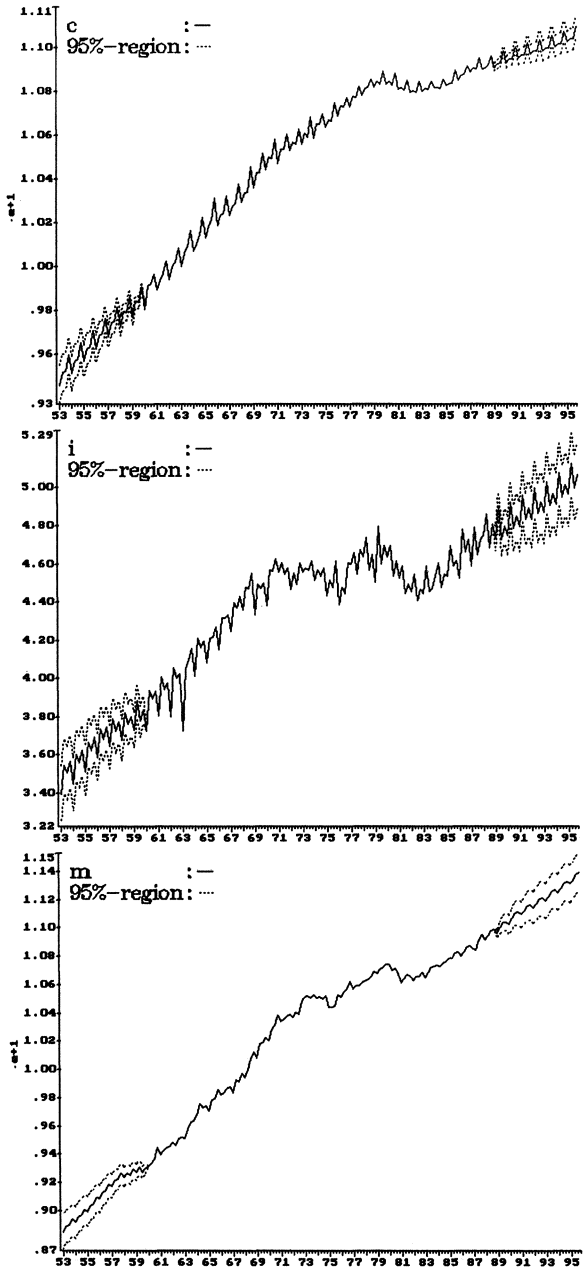

Figure 7.5a *Forecasts* (1989–1996) *and backforecasts* (1953–1960) *from unrestricted VAR, with standard* 95% *confidence region*; c: consumption, i: investment, m: imports.

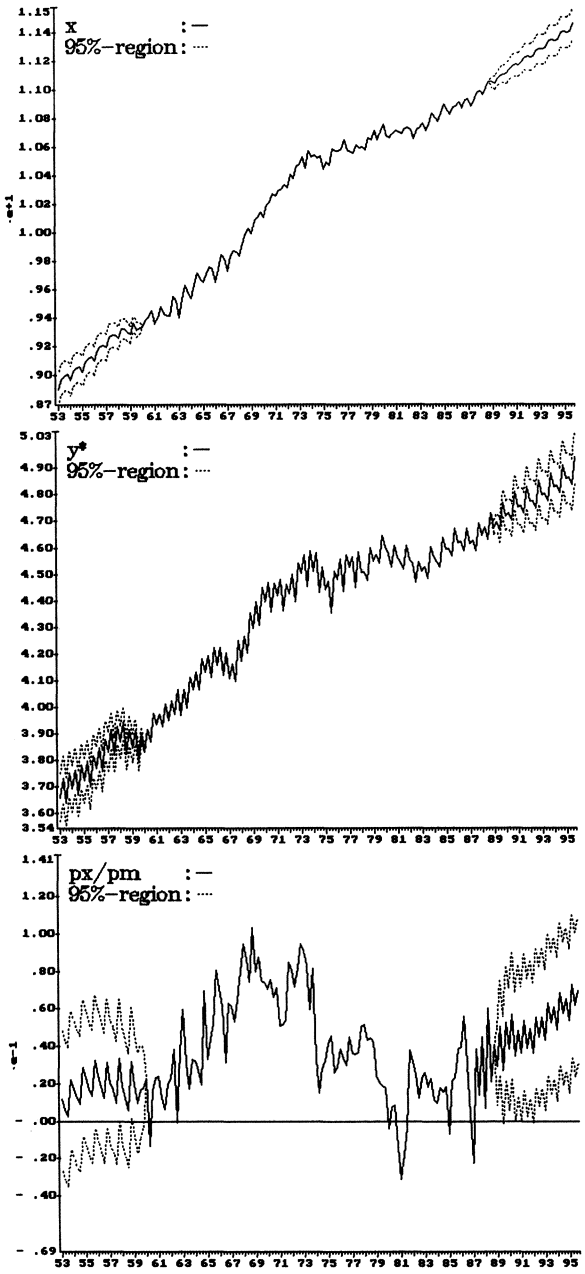

Figure 7.5b *Forecasts* (1989–1996) *and backforecasts* (1953–1960) *from unrestricted VAR, with standard* 95% *confidence region* (*continued*); x: exports, y*: German industrial production, *px/pm*: terms of trade.

The forecasts and backforecasts are plotted in figure 7.5. They do not look nonsensical, although they are based on unrestricted models with a large number of parameters. It is obvious that a VAR model cannot predict the comparatively high aggregate level of industrial production in the old German empire (before 1945) and the new German republic (after 1989) on the basis of our sample information.

Here we use an order of 5 for $\Phi(L)$. In the forecasting model we use sample period 1965.2–1988.4, in the (reverse regression) backforecasting model we employ sample period 1960.1–1983.3. The confidence intervals are the standard ones based on maximum likelihood theory, which are easily obtained from standard computer packages. They give a first idea of the forecast uncertainty.

The "true" confidence intervals are bigger. No degrees of freedom corrections were used in the estimate of the residual variance matrix. The probability of additive outliers occurring in the near future is put to zero. The uncertainty in the parameter estimates is not taken into account, which is especially important for m_1 and m_{1d}. No unit roots are imposed. Since the OLS estimator of the eigenvalues of the system is biased towards stable values in small samples, this leads to an underestimate of the forecast error variance if some unit roots are present. This is not entirely unlikely given the evidence of the univariate analysis. Last but not least, the VAR model is not the "true" model.

Order selection

The traditional AR order selection criteria are not a big help in our situation with approximately 110 observations (depending on the AR order) and 30 regressors per equation. If one computes the implicit "optimal significance level" for an F–test for 6 extra regressors following the analysis in Amemiya (1985, §2.1.6), one gets a value bigger than 20% for the AIC criterion and a value smaller than 0.5% for the BIC criterion[14]. Neither value seems to be attractive here. Kavalieris (1989) derived easily understandable loss functions for these criteria. The AIC criterion is asymptotically equivalent with so called cross–validation, i.e. minimizing the sum of squared prediction errors (defined as $\bar{r}_i$ in §A3.1.1). The application of BIC is asymptotically equivalent with minimizing the sum of squared recursive (unscaled) residuals. Neither of these loss functions is attractive here. If one applies

[14] For a fixed number of 109 observations and 1, 7, 19, 20, 25 and 31 regressors under the null and 6 extra regressors under the alternative one has series of optimal p–values for AIC of 0.076, 0.095, 0.119, 0.149, 0.185 and 0.227 and for BIC of 0.00015, 0.00030, 0.00058, 0.00112, 0.00213 and 0.00401.

conventional significance levels one finds that the equations for consumption and imports need a VAR order of at least 6. The outlier and seasonal adjustment generally seem to reduce the statistical need for more lags in the VAR: the p–values of tests against higher order models go down.

Seasonal component

The seasonal component is estimated as $(I-s(L))y_t$ as described in §4.5. This filter significantly smooths sudden changes in the seasonal component over a period of approximately six years. The seasonal components of the first differences are plotted in figure 7.6. The smoothing effect is clearly visible in the series for the German industrial production. The most striking changes in seasonal pattern occur in the export series, where the relative growth rate of the fourth quarter increased during the late sixties and the beginning of the seventies. The availability of cheap natural gas may be one of the reasons. A similar change in seasonal pattern can be observed in the investment series, although not so pronounced. In the German industrial production series the increase in the relative growth rate of the first quarter at the end of the seventies is remarkable. This may just be a measurement issue. The quite sudden permanent change in the seasonal pattern of the terms of trade towards the end of the sample seems to be a measurement issue as well. The outlier testing procedure points to a transient outlier, see figure 7.4, which can be interpreted as an indication of a permanent (seasonal) level shift. After the removal of the seasonal component s_t we estimate the additive outlier component u_t for a second time. The estimates for u_t do not change very much.

Influence analysis

Next we focus our attention on multivariate unit root analysis and the corresponding estimation of equilibrium relations between the seasonally adjusted series. It appears that the additive outliers do not have a significant impact on the statistical analysis of the number of unit roots. One can easily analyze the influence on the estimates of the equilibrium relation in the minimum variance procedure of Stock and Watson (1988) by Critchley's influence statistics (see §A3.3.2) of the "least principal" component. For the seasonally adjusted data we find that $\hat{I}_{t6}/\lambda_6$ has a maximum value for t=1964.3 which varies from 9.3 to 11.3 from the dirty data to the data without the additive outliers. For the next to last principal component, which has a bigger component of the investment series we find corresponding values $\hat{I}_{t5}/\lambda_5$ for 1963.1 of 10.1 and 6.2. Here the additive outlier had a bigger influence. A graph of the separate components of the cointegrating vector $\alpha_i y_{it}$ is a useful additional data analytic tool.

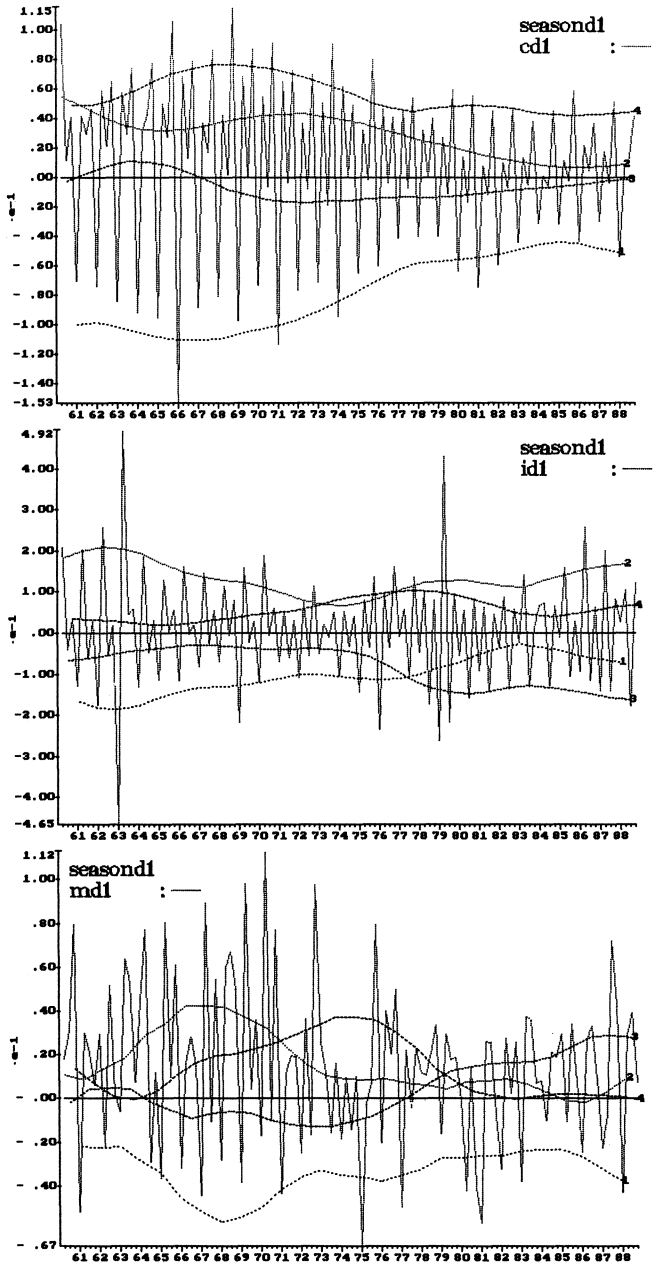

Figure 7.6a *Time series plots of growth rates and estimated seasonal components*; *seasond1* connects seasonal component quarterly growth rate with same component one year before. Numbers at ends of lines indicates number of quarter. *cd1*: consumption, *id1*: investment, *md1*: imports.

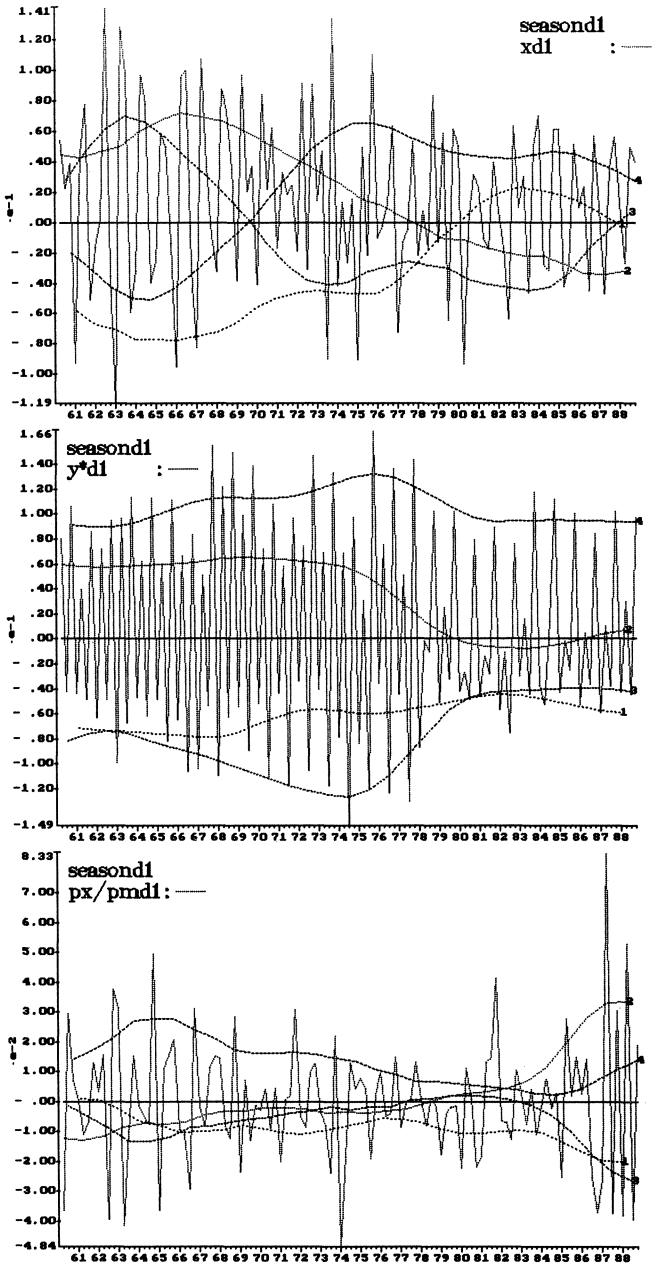

Figure 7.6b *Time series plots of growth rates and estimated seasonal components (continued)*; *seasond1* connects seasonal component quarterly growth rate with same component one year before. Numbers at ends of lines indicates number of quarter. *xd1*: exports, *y*d1*: German industrial production, *px/pmd1*: terms of trade.

It gives an idea of the empirical significance of the separate coefficients. Formal testing procedures of the statistical significance of the coefficients depend very much on asymptotic approximations that are not sure to work very well for the number of parameters and the sample size that we use here.

7.5.2 Unit root analysis

Evidence from short run autocorrelations

Figure 7.7 contains the graphs of the implied typical log spectrum $f(\omega,r)$ (see (3.15)) for different numbers of cointegrating ranks r, with $r=n-k$, and k the supposed number of unit roots in the system. The corresponding supposedly stationary equilibrium relations were estimated using three of the methods of §6.5.3 viz. Maximum Likelihood (ML), Minimum Variance (MV) and Minimum Predictability (MP). In the ML method we used an AR(6) and D_t and a constant as conditioning regression variables, see §A6.1.1. In the MV method we applied an ordinary principal component analysis using the covariance matrix of the levels of the variables. In the MP method we first subtracted the means before the application of the standard Box–Tiao method in an AR(6)[15]. Subsequently we estimated the parameters of the restricted VAR in a Vector Error Correction specification (see (6.10)), with the same conditioning variables as in the ML estimation method, so that all resulting models allow for a (gradual) change in mean after 1976.4. The shape of this gradual change is shown below, see figure 7.14.

The basic model in this second round is thus

$$\Phi(L)y_t = m_0 + m_0 D_t + \varepsilon_t, \qquad (7.4)$$
$$y_t = x_t - u_t - s_t,$$

with

x_t the observed series,

u_t the additive outlier component, and

D_t a permanent innovative outlier:

$D_t=1$ for t after 1976.4, D_t is zero elsewhere.

s_t is the seasonal component and

$\Phi(L)$ is an AR matrix polynomial with at least one unit root.

The general pattern of $f(\omega,r)$ is the same for these three estimation methods, with the biggest differences occurring at frequencies smaller than 1/12. Judging from this graph the hypothesis with $r=1$ seems reasonable: $f(\omega,1)$ clearly does not diverge to infinity near the zero frequency. The graph also

[15] Correction for additive shifts in mean or mean growth in the first stage of the MV and MP method did not produce satisfactory results.

shows that under the assumption of $r=1$ most of the common variation left in the series is at high frequencies between ¼ and ½ (or equivalently between ½ and 3/4). The variation at the seasonal frequencies is removed by the application of the seasonal adjustment filter. The theoretical spectrum at the seasonal frequencies is close to zero. The log spectrum is thus very negative.

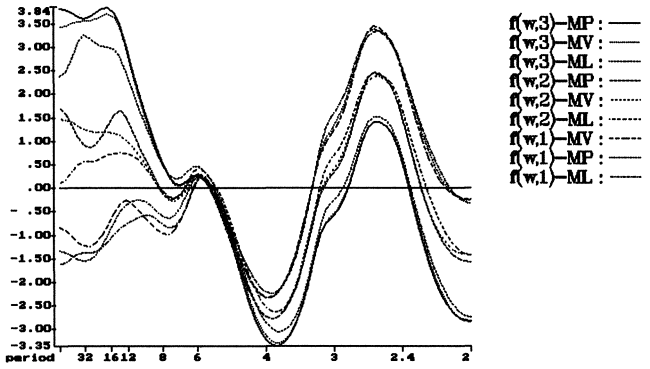

Figure 7.7 *Implied typical log spectra for different numbers of unit roots and different estimation methods* f(w,r): $f(\omega,r)$ in (3.15), with $w=2\pi/\omega$ (on x-axis). MP: Minimum predictability, MV: Minimum variance, ML: Maximum likelihood. Series in legend ordered according to zero frequency values.

Pulling equilibrium

From our point of view the least predictable component is most interesting here, since it has the biggest partial correlation with investment growth rates. Table 7.6 contains the estimates of the ML–, MV– and MP methods of the equilibrium relation $\alpha' y_t$ (§6.5.3), together with the implied mean growth rates before and after the shift. All mean growth rates are reasonably close to their sample counterparts.

Next we consider the MV and MP estimates as restrictions in a classical hypothesis testing situation. We test the hypotheses $\alpha=\alpha_{MV}$ and $\alpha=\alpha_{MP}$. We find similar outcomes for both restrictions. The likelihood ratio test statistic on the MP restriction, which has the conventional asymptotic χ^2-distribution if we suppose that we already imposed the correct number of unit roots under the null, equals 2·(1882.7–1873.5)=18.4. There are five linearly independent restrictions. If we apply the finite sample correction of Anderson (1958) (see §A5.2) we get a test value of 12.4, which corresponds to a p–value of about 3% for a χ^2_5 distribution. These results indicate that the estimates of the parameters of the equilibrium relations are not very precise in a statistical sense, since the α–vectors are rather disparate, cf. α_{ML}, α_{MV} and α_{MP} in table 7.6. The separate coefficients are not significantly different from zero,

since five of the variables are highly collinear, cf. fig. 7.8.

If we want to include the obvious trade balance variables m and x, we need to put i in the cointegrating relation as well to make it insignificantly different from the unrestricted maximum likelihood estimate at the 5% level, cf. α_{rest1} in table 7.6.

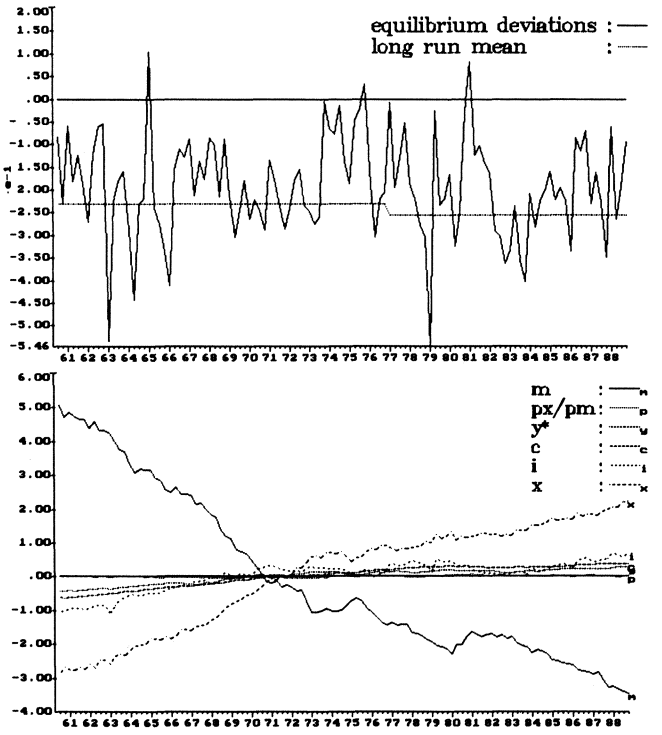

Figure 7.8 *Least predictable linear combination* Time series plot of $\alpha_{MP}'y_t$, see table 7.6. Lower figure gives decomposition this series. *c*: consumption, *i*: investment, *m*: imports, x: exports, y*: German industrial production, *px/pm*: terms of trade.

The plausibility of the hypothesis that $r=1$ can also be assessed from a univariate analysis of the estimates of the cointegrating vectors. According to a univariate analysis similar to that of §7.4, the deviations from the MP estimate of the pulling equilibrium seem to be closest to stationarity. This fits in with the purpose of the MP method to find the most "white noise like" linear combination of the variables. Figure 7.8 contains a time series plot of this series with the coefficients normalized as in Table 7.6. The long run

mean of this cointegrating vector as defined in the last r rows of[16] (6.20) changes only slightly after 1976. The components of imports and exports are clearly dominating in the companion plot of the decomposition of the cointegrating vector. Investment also plays a role of interest. This is in line with analysis of the statistical significance of these components discussed above.

Table 7.6 *Estimates of equilibrium relations and persistence*

	c	i	m	x	y^*	$(px-pm)$	logL	c.v. 5%
MG 60–76	1.3	1.1	1.9	1.9	1.0	0.03		
MG 77–88	0.4	0.5	0.8	0.9	0.3	−.01		
MG 60–88	0.9	0.8	1.4	1.4	0.7	0.02		
α_{ML}'	1.0	0.3	−2.5	1.4	0.4	1.7	1882.75	
$C(1)_{ii}$	0.98	0.89	0.38	0.66	0.14	0.48		
$E\Delta x_i$ A77	1.3	0.9	1.9	1.9	0.9	0.01		
$E\Delta x_i$ P76	0.4	0.4	0.8	0.9	0.3	0.04		
test $\alpha_i=0$	0.20	1.13	0.39	1.24	1.10	0.09		3.85
α_{rest1}'	0	1.0	−3.1	2.3	0	0	1880.11	
α_{rest2}'	1.0	−1.0	0	0	0	5.7	1872.20	
α_{MV}'	1.0	1.9	−5.0	4.9	−4.4	1.2	1873.23	
$C(1)_{ii}$	1.63	0.62	0.78	0.43	0.89	0.48		
$E\Delta x_i$ A77	1.2	0.8	1.7	1.9	0.8	0.01		
$E\Delta x_i$ P76	0.4	0.5	0.8	0.9	0.4	0.02		
α_{MP}'	1.0	1.9	−5.3	3.1	0.9	−1.0	1873.50	
$C(1)_{ii}$	1.59	0.53	0.65	0.44	0.95	0.51		
$E\Delta x_i$ A77	1.2	0.8	1.7	1.9	0.8	0.01		
$E\Delta x_i$ P76	0.4	0.5	0.8	0.9	0.4	0.02		

NOTES: logL: Log likelihood computed for VECM model. Methods: ML: Maximum Likelihood, MV: Minimum Variance, MP: Minimum Predictability. MG j_1–j_2: "Mean growth" from year j_1 to year j_2. α: normalized estimate of pulling equilibrium. $C(1)_{ii}$: diagonal elements of $C(1)$ (or common persistence) matrix, see §6.2.2, computed in VECM model from (6.27). $E\Delta x_i$: expected growth rates computed from (6.25), A77: before 1977, P76: after 1976, constant term m differs between these periods (m_0 and m_0+m_{0d} in (7.4)). test α_i=0: likelihood ratio test statistics, with small sample correction from §A5.2. α_{rest1}: restricted estimate of α with $\alpha_1=\alpha_5=\alpha_6=0$. α_{rest2}: restricted estimate of α with $\alpha_3=\alpha_4=\alpha_5=0$, $\alpha_1=-\alpha_2$, discussed in appendix 7.3.

Pushing trends and business cycle

Subsequently one can take a look at the permanent component (see (6.30)) and the stationary component of the series to see whether they make sense. In this case they do.

[16] Given α, one can easily estimate $B(1)$ and γ using regression (6.32).

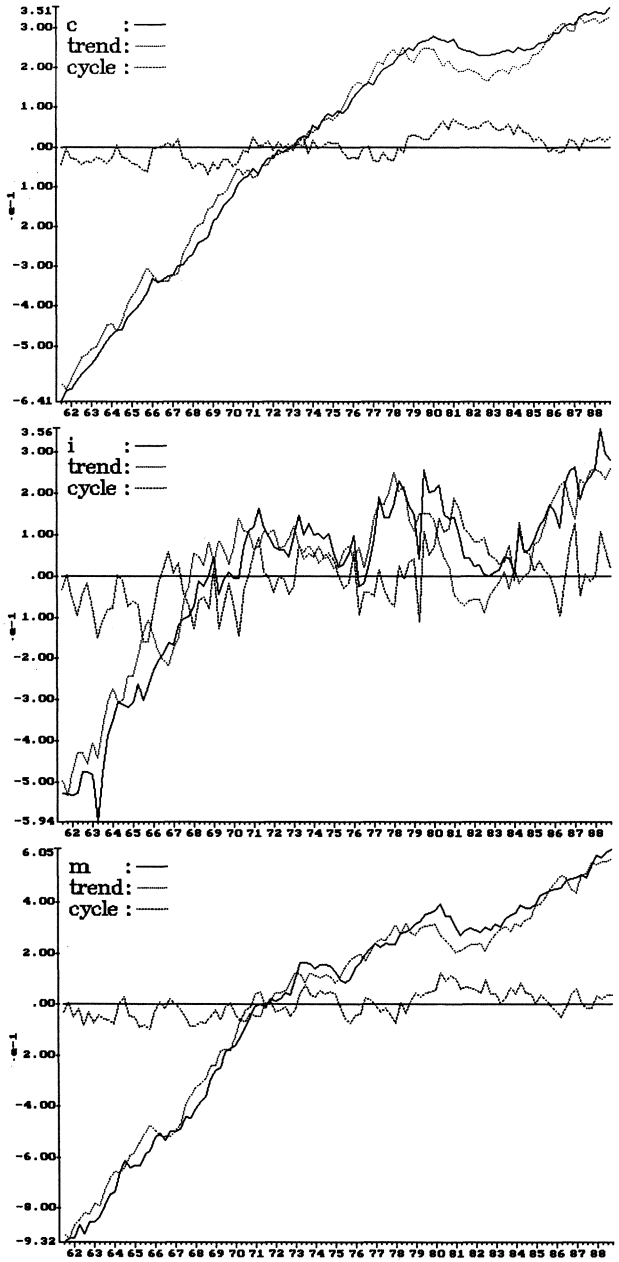

Figure 7.9a *Trend cycle decomposition time series plots* Based on model with one long run equilibrium relation estimated using the minimum predictability method. Data corrected for additive outliers and seasonally adjusted. *c*: consumption, *i*: investment, *m*: imports.

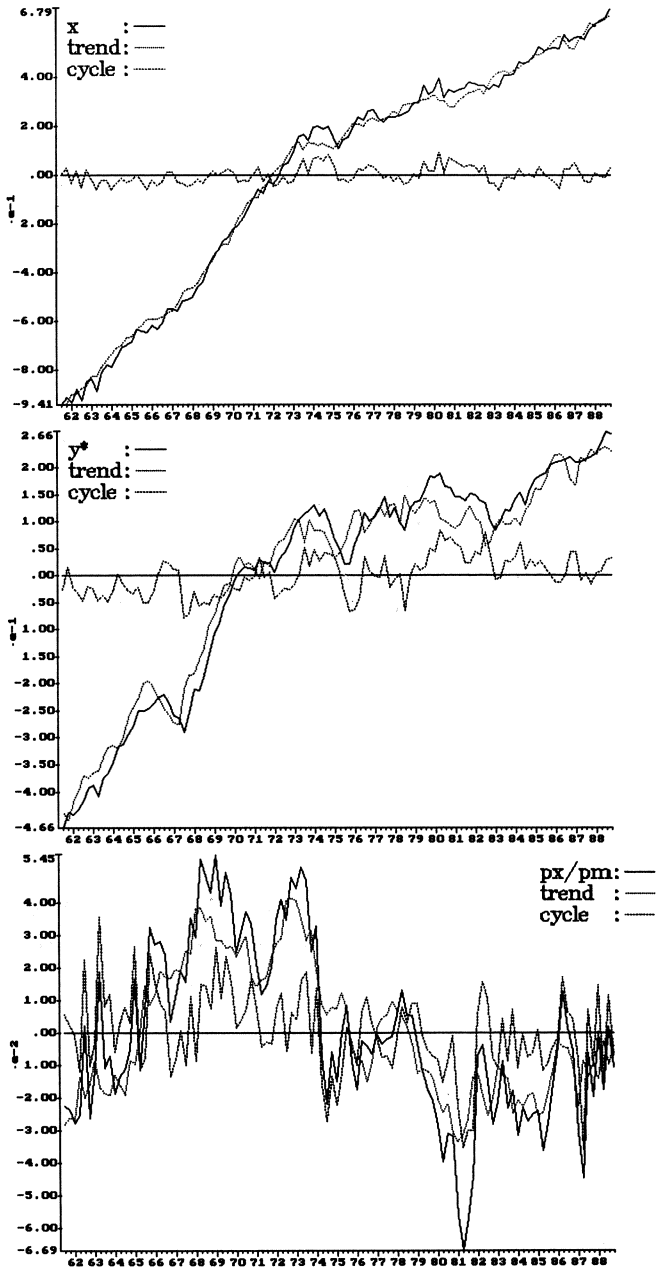

Figure 7.9b *Trend cycle decomposition time series plots* Based on model with one long run equilibrium relation estimated using the minimum predictability method. Data corrected for additive outliers and seasonally adjusted. *x*: exports, *y**: German industrial production, *px/pm*: terms of trade.

According to Beveridge and Nelson (1981), the stationary components should be related to common sense measurements of the "business cycle", see §6.2.2. Figure 7.9 contains the plots of the trend components which results from the Box–Tiao estimates of α. Both trend and cyclical components are presented as deviates from their sample mean.

The volatility in the remaining cyclical component is biggest for the investment series, see Table 7.7. It varies from –20% (1967) to +20% (1981). The sum of all forecastable changes (beyond the mean rate of growth) tends to have the same direction for the first five variables, whereas the sum of forecastable changes in the terms of trade tends to be in the other direction. In the period around 1981–1982 the stationary component was high for all the real variables, which entails that the sum of all forecastable changes was at a low. This was also the period where aggregate investment went through its steepest decline.

Table 7.7 *Variation and covariation stationary components*

	c	i	m	x	y^*	$(px-pm)$
var. $y_{it}-y_{ipt}$	.0011	.0052	.0025	.0011	.0013	.0002
corr. i_t-i_{pt}	.80	1.00	.85	.59	.88	–.40

NOTES: var.: variance of stationary component, see §6.2.2, computed from VECM with one cointegrating vector estimated using MP method (see table 7.6). corr.: correlation with stationary component investment.

Evidence from long run autocorrelations

In order to see whether the predictability of the cointegrating vector(s) really goes to zero for long forecast horizons one can plot the six predictabilities for the corresponding shifted VARs (see §6.5.2 and 6.5.3):

$$y_t = (I - \Phi_h(L))y_{t-h} + \varepsilon_{ht}, \qquad h = 0,1,\ldots \tag{7.5}$$

Since we are working with a great number of regressors it makes sense to adjust the predictabilities for degrees of freedom in a way similar to Theil's (1961, p. 213) adjustment of the multiple correlation coefficient in regression. The adjusted predictabilities $\bar{\lambda}_{li}$ are related to the λ_{li}'s introduced above (6.48) as follows:

$$\bar{\lambda}_{li} = \lambda_{li}T_l/(T_l-k) - k/(T_l-k), \tag{7.6}$$

with

λ_{li} the i–th predictability at forecast horizon l, i.e. the i–th squared canonical correlation between y_t and $\hat{y}_{t|t-h}$, the forecast of y_t using

parameter estimates of the shifted VAR and observations up to $t-h$,
T_l the effective number of observations for the estimation of $\Phi_h(L)$ in (7.5),
k the number of estimated parameters per equation in (7.5).

The $\bar{\lambda}_{li}$–values lie between $-k/(T_l-k)$ and 1. Figure 7.10 presents the adjusted predictabilities in descending order together with their absolute lower bound. We adjusted the data for seasonality and additive outliers and subtracted the mean before estimation of the shifted VARs[17] . Only the last predictability (l1 in figure 7.10) points to the existence of a stationary linear combination. It is declining gradually up to a forecast horizon of approximately 5 years. For longer horizons the last predictability goes up again. The evidence in the higher order autocorrelations does not support the cointegration hypothesis. There is a serious lack of degrees of freedom problem in this "long run" evidence, however. The lower bound (lb_l in figure 7.10) declines to values below –1 for a forecast horizon of 10 years. This indicates that the number of parameters is greater than the degrees of freedom. Classical error bounds around these predictabilities at greater forecast horizons would just be very large, virtually covering the whole set of possible values.

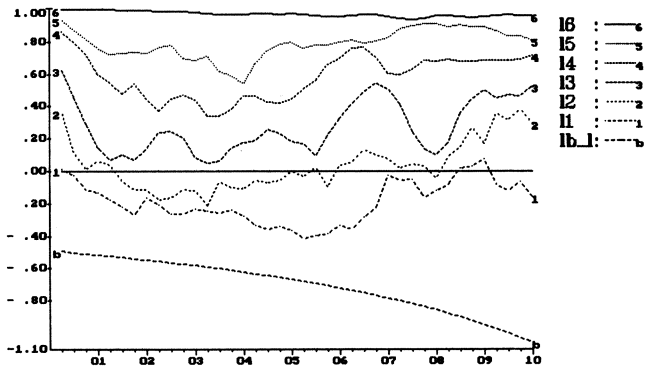

Figure 7.10 *Predictabilities from shifted VAR(5) of the Dutch macroeconomic variables* These variables are presented in figure 7.1a. They were corrected for additive outliers and seasonally adjusted. Forecast horizons (x-axis) vary from one quarter to ten years. lb denotes lower bound.

Another example shows that the procedure can give more convincing results in favor of the cointegration assumption. The long run autocorrelations of the

[17] Additional correction for additive (breaking) trends led to lower λ_{l1}s and higher λ_{li}s, $i>1$, i.e. less clear cut evidence of cointegration.

French data (corrected for additive outliers) of §5.6 contain pretty clear evidence of the existence of a cointegrating vector, see figure 7.11. The last adjusted predictability is close to the lower bound for forecast horizons longer than two years and stays there. As can be seen from the lower bound the lack of degrees of freedom problem is not as serious in the model for the French data. The ratio of the number of parameters and the degrees of freedom per equation is only 0.19 for the 10 year forecast horizon.

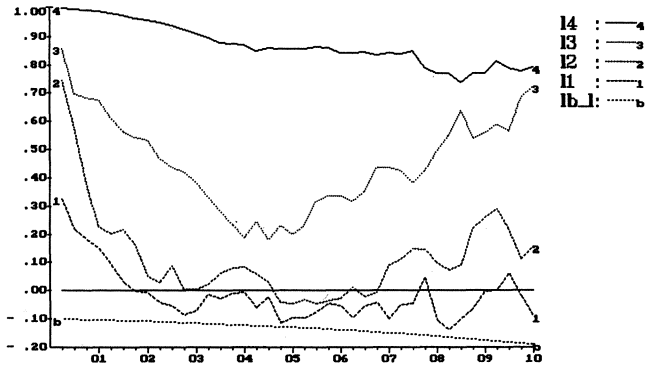

Figure 7.11 *Predictabilities from shifted VAR(2) of the French macroeconomic variables* These variables are presented in figure 5.9. They were corrected for additive outliers. Forecast horizons (x-axis) vary from one quarter to ten years. lb_1 denotes lower bound.

Hypothesis tests

Small sample adjustments also play an important role if one interprets the outcomes of some multivariate unit root tests reported in table 7.8. The evidence for the existence of a cointegrating vector is rather weak. The Stock–Watson test procedure does not lead to the rejection of the null hypothesis of the presence of six unit roots in the system at any conventional significance level. The additive adjustment for (breaking) trends in a regression prior to the estimation of the VAR may be a poor procedure in our sample, however.

Application of Johansen's asymptotic tests (assuming Gaussian disturbances in the adjusted data), allowing for one innovative change in mean growth after 1976, leads to the rejection of the hypothesis of six unit root against the alternative with five at the 5% level. If one applies the degrees of freedom adjustment suggested by Ahn and Reinsel (row (6) in the table) the null hypothesis of 6 unit roots cannot be rejected at the 5% level. The choice of the lag order and the precise timing of the innovative outlier do not influence the outcomes too much. The value of the likelihood ratio test

statistic is greater if the innovative outlier is introduced earlier or later, or not at all. The estimate of the maximum likelihood equilibrium relation shifts considerably. The maximum squared canonical correlation between growth rates and lagged levels, λ_6^{iii}, is always between 0.3 and 0.35. Reducing the AR order increases λ_6^{iii}, increasing the AR order decreases λ_6^{iii}. A similar effect of decreasing evidence of mean reversion occurs in the univariate analysis.

Table 7.8 *Tests for 6 unit roots against 5 or less*

Statistic	order AR	Det.terms	value	c.v. 5%	
$T(\mathrm{Re}(\lambda_{f6}^{T})-1)$	6	m_0, m_{0d}, m_0, m_{1d}	−36.3	-62.4^{*}	(1)
$(T-np-4)(\mathrm{Re}(\lambda_{f6}^{T})-1)$	6	m_0, m_{0d}, m_0, m_{1d}	−34.4	-62.4^{*}	(2)
$-T(\ln(1-\lambda_6^{iii}))$	6	m_0, m_{0d}	41.6	39.4	(3)
$-T(\sum\ln(1-\lambda_i^{i\,i}))$	6	m_0, m_{0d}	113.4	94.2	(4)
$T(\sum\lambda_i^{iii})$	6	m_0, m_{0d}	99.0	94.2	(5)
$(T-np-2)(\sum\lambda_i^{iii})$	6	m_0, m_{0d}	81.9	94.2	(6)

NOTES: Det. terms: extra explanatory variables in regressions, additive in Stock-Watson procedure, innovative in Johansen procedure. c.v. 5%: asymptotic critical value of statistic at 5% level. Sources: (1): Stock and Watson (1988), (3)-(5): Osterwald-Lenum (1992). Rows(1)-(2): Stock-Watson procedure (see §A6.1.1) extended with (additive) change in mean and mean growth term in 1977.1 in "a priori regression. (1) No degrees of freedom correction, (2): degrees of freedom correction. Critical value derived without change in trend, "true" critical value even larger in absolute value. Rows (3)-(6) Johansen's procedure, (see §A6.1.1).(3)-(4): no correction. (5): correction of log term. Row (6): Degrees of freedom correction.

7.5.3 Detecting a structural break

The first tools in our analysis of parameter stability are the recursive tests applied to the VECM model of the data adjusted for seasonals and additive outliers. Both adjustments improve stability of the VAR estimates. The recursive test can only be applied if one does not include the structural change dummy[18] D_t. One could adjust the data for the effect of the permanent innovation outlier first, but here we want to see whether the parameter stability tests indicate the need for such a "structural break" in the model. Table 7.9 contains some results. The nominal p–values for the fluctuation tests are low, but this may also due to the generally poor performance of robustness in size (cf. §3.3.2) of this test in truly dynamic models. The p–value for the backward application of the test on the investment equation points to a serious break. The CUSUM test only points to a significant break in the consumption equation after 1980.

In figure 7.12 we present some time series plots of minus the logs of the p–values of the tests of Table A3.1 in §A3.2. The test statistics CH_i, VR_i and

[18] m_{0d} is not estimable in the early subsamples.

QT_i can only be computed for the period 1969.2–1980.3. In this example the subset[19] (X_i, y_i) contains the observations for the period l–1988.4. The subset $(X_{(i)}, y_{(i)})$ contains the data between 1961.3 and $(l-1)$. A_i and $A_{(i)}$ denote the residual sums of squares for the regressions on these subsets. We also computed the test with the roles of A_i and $A_{(i)}$ interchanged. This gives different results for the variance ratio test and the QT test for the stability of the other parameters. The ANOVA test CH_i does not change. In the first way the variance ratio rejects if the variance has increased too much in the course of time. In the second way it rejects if the variance has decreased too much.

Table 7.9 *Parameter stability test asymptotic p–values*

Test	c	i	m	x	y^*	$(px–pm)$	system
Fluc_f	0.003*	0.03	0.035	0.005*	0.035	0.42	0.007*
Fluc_b	0.85	<10^{-4}*	0.307	0.264	0.0035*	0.003*	0.0003*
CUSUM_f	<0.05*	>0.05	>0.05	>0.05	>0.05	>0.05	
CUSUM_b	>0.05	>0.05	>0.05	>0.05	>0.05	>0.05	
CUSUM_{OLS}	0.67	0.88	0.98	0.75	0.98	0.39	

NOTES: Fluc_f: asymptotic p-values fluctuation test (see §3.3.8) in VECM equations with one cointegrating vector estimated using Box-Tiao method, see table 7.6. Fluc_b: fluctuation test applied to observations in reverse order, first row of X-matrix becomes last. CUSUM_f: CUSUM test (§3.3.6), CUSUM_b CUSUM test applied to observations in reverse order. CUSUM_{OLS}: CUSUM test on OLS residuals (§3.3.6).

Figure 7.12 also shows most serious problems for the investment equation, but this seems largely due to the instability of the variance of the disturbances. According to the analysis of Andrews (1990), the maximum of the F–test (i.e. CH_i, the analysis of variance test) over the center 70% of the sample has a 5% critical value which corresponds to a nominal significance level of 0.3 to 1%[20] if one applied the test "as if" one knew the timing before hand. His analysis gives a rough indication that only nominal *p–values* for the CH_i below 1% should be taken seriously if the true significance level of interest is 5%. The nominal p–values under 5% of the CH_i statistic for the consumption (1970–1972), investment (1975), and terms of trade (1973) equation do not indicate serious misspecification.

[19] Here X contains the observations for lagged differences, the constant term and the lagged deviation from the equilibrium relation.

[20] Andrews only computed tables for regressions with up to 20 regressors in static regressions. Note that the decrease in the nominal p-value that should be used is much smaller than in example 3.1, where the maximum of test statistics for the congruence of *nonoverlapping* subsets was used.

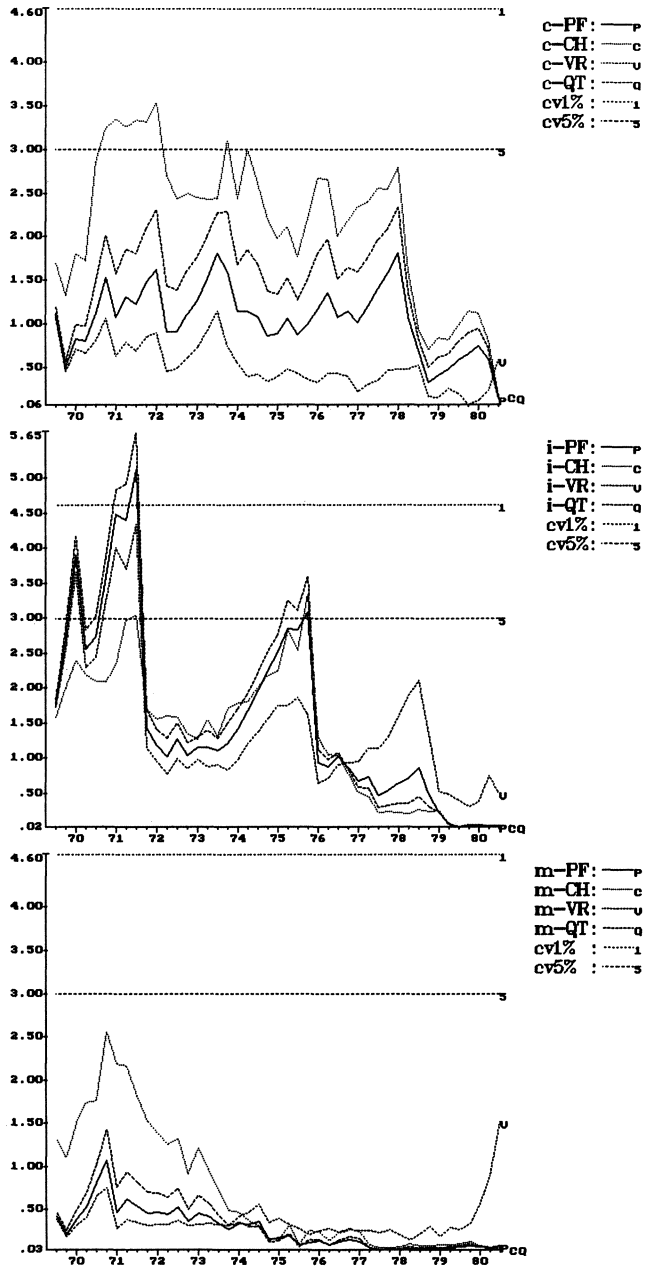

Figure 7.12a *Nominal –log p–values of parameter stability test statistics*
x-axis gives date of where sample is split in two to apply the tests. Expressions are given in table A3.1. cv5%: -log(0.05), cv1%: -log(0.01). *c–*: consumption, *i–*: investment, *m–*: imports.

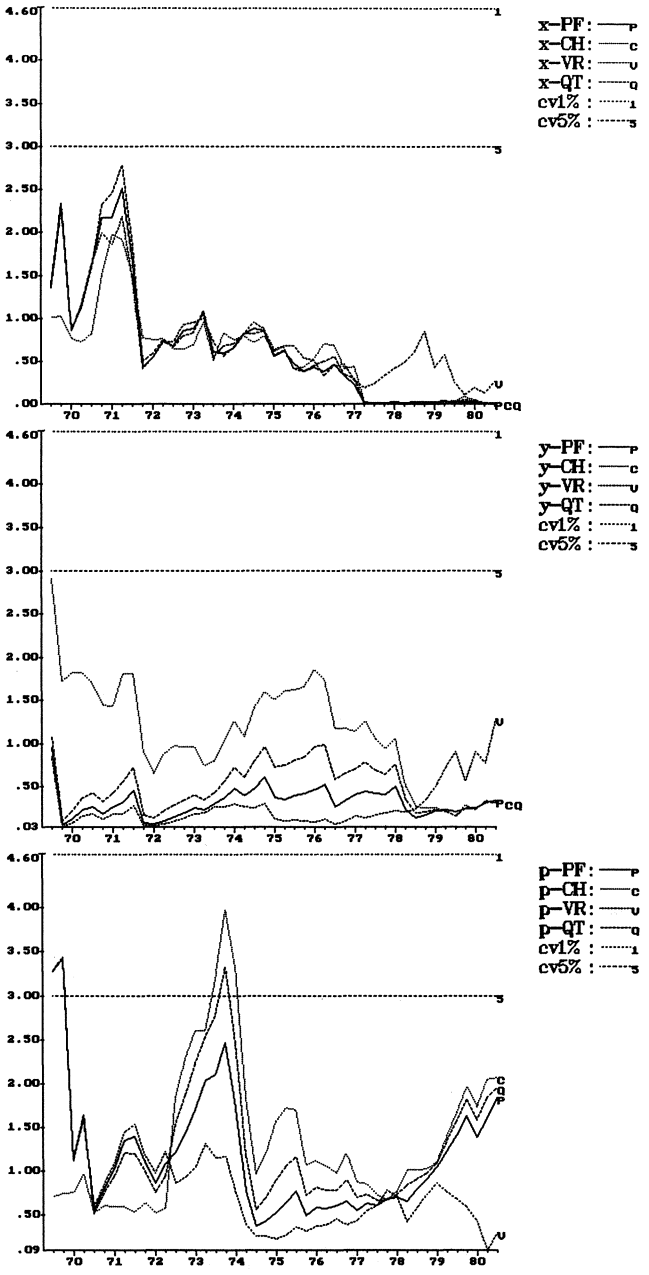

Figure 7.12b *Nominal –log p–values of parameter stability test statistics (continued)* x-axis gives date of where sample is split in two to apply the tests. Expressions are given in table A3.1. cv5%: -log(0.05), cv1%: -log(0.01). *x–*: exports, *y–*: German industrial production, *p–*: terms of trade.

One can also see from the figures that not all test statistics lead to the same conclusion about the adequacy of the model. For l close to 1969.3 the CH test stands out. For l close to 1980.3 the VR test does. There the Chow test for predictive failure PF_i is nearly equal to the QT_i test as can be derived from (A3.2.11). The weight of the VR_i test in the PF_i test is very low: (33–32)/33. Near 1972.3 the respective weights are 32/64 and 32/64: the p–value for PF_i ends up in the middle between those for QT_i and VR_i.

The multivariate outlier criterion defined in §3.3.7 uses the PF_i statistics for nonoverlapping subsets of equal size. It can also be applied to the model with the structural break. We discuss results below.

One can also check parameter stability via an autocorrelation test of the recursive residuals. One can expect dips in the spectra of these residuals at the seasonal frequencies because of the seasonal adjustment. On the whole the autocorrelations do not point to a serious misspecification. A spectral analysis of the recursive residuals of the export equation indicates a considerable degree of antipersistence (negative degree of integration) however. The CUSUM analysis of this equation shows a consistent downward trend of the partial sum of the recursive residual (not reproduced here).

Let us summarize the evidence of the parameter stability tests of the system without a structural break. We find that the CUSUM test rejects parameter stability of the consumption equation. The outcomes of the other tests are no reason for comfort. The introduction of an extra variable in order to allow for a change in mean growth seems especially warranted for the consumption and the exports equation.

7.5.4 The final model

Data based restriction of the parameter space

The unrestricted VECM model with structural break dummies contains 198 location parameters and 21 covariance parameters. There is some scope for simplification. We identified the nonzero location parameters in the VECM model using the OLS t–values of the separate coefficients. Since the regressors in the VECM model are reasonably orthogonal, especially if one compares them with the regressors in the unrestricted VAR in levels, this is not a hopeless task. A degree of "arbitrariness" remains, though.

Other methods exist. Koreisha and Pukkila (1987) suggested ways to remove this arbitrariness by using an automatic identification procedure. Penm and Terrell (1984) suggested an automatic procedure to select optimal so–called *subset* AR models in a Granger causality study. Tsay (1989b) discussed two identification methods to obtain parsimonious (VARMA) models for stationary

time series. There the contemporaneous relationships are selected on data evidence as well. The first method is based on Kronecker indexes[21]. The second method is based on Scalar Component Models, which is an extension of the Box–Tiao (1977) method. In our case the most white noise like component, $\alpha_{MP}'y_t$, depicted in fig. 7.8 is still far from white noise.

We employ the first step of the Scalar Component modeling technique only to model the longer run autocorrelations. The VECM representation allows direct identification of partial correlations between growth rates of the different variables, which are the main transformations of interest in our case. The resulting impulse responses of our method are far from smooth. Reduction of the order of the VAR smooths the impulse responses considerably, the imposition of our data based restrictions does not.

We test the overall restriction in the multiple regression model, and check the restricted model for residual autocorrelation and parameter stability. By imposing the correct restrictions one can increase the precision of the parameter estimators. The estimators of the disturbances improve correspondingly. The number of outliers criterion indicates that a further whitening of the data is necessary. We employ a tuning constant of 2.4 (cf. A5.3) on the data which result from the first stage whitening. This last whitening does not have a significant impact on the estimation of the pulling equilibrium.

Final VECM model

Finally one can estimate the model using a SUR estimator[22]. Since there is considerable correlation between the disturbances of the different equations, this improves the efficiency of the estimates even further. Table A7.2 in the appendix contains the main estimation results. The terms of trade is not Granger (1969) caused by any of the other variables at positive lags. Its disturbances have a contemporaneous negative partial correlation[23] with exports. This may partly be due to the common factor of export prices, which is used to obtain the real export figures. The model picks up some linear dynamic relations between the other variables of interest.

The disturbances in German industrial production are contemporaneously

---

[21] Luetkepohl (1991) discussed a number of identification methods in detail.

[22] Penm and Terrell (1984, p. 328) postmultiplied X_t by K^{-1}, with K regular and $KK'=\Sigma$. This does not lead to a SUR estimator. They used Yule-Walker equations to fit a VAR for detrended differences. Their method is now outdated.

[23] Here we apply the usual rule of thumb critical value of $2T^{-\frac{1}{2}}$. We did not compute the ML variance covariance matrix of the variance covariance parameters of the disturbances.

correlated with disturbances to Dutch consumption, imports and exports. The equations for these variables do no indicate a strong relationship with Dutch investment.

The other disturbances of the four Dutch variables are strongly interrelated. The contemporaneous correlation between import disturbances and export shocks is especially large. Changes in exports are partially correlated with lagged changes in German industrial production. The hypothesis of non Granger causality of deviations from the equilibrium relation to changes in exports can be rejected at rather low significance levels. Changes in imports definitely have a partial positive correlation with lagged changes in consumption and the terms of trade.

The statistically most interesting explanatory variable in the investment equation is the error correction variable which measures the deviations of (most notably) imports, exports and investment from their "extended trade balance" equilibrium relation. Rescaling the estimates of table 7.6 one gets

$$\Delta i_t = z_t' b_z + (1.14 m_{t-1} - 0.66 x_{t-1} - 0.40 i_{t-1} - 0.21 c_{t-1}) + \varepsilon_{it}, \qquad (7.7)$$

with

$z_t' b_z$ the other explanatory variables in the regression.

Recursive analysis of this error correction (not reproduced here) shows that it has been in effect over the entire sample period. A decline in imports relative to exports and investment significantly decreases the growth potential of both investment and exports. The role of consumption in this equilibrium relation is not well identified in the data. Changes in consumption have negative partial correlations with lagged changes in the trade balance and positive partial correlations with lagged changes in the terms of trade.

Some figures from the Dutch input–output table for the year 1984 in table 7.10 below indicate the size of the simultaneity in the measurement model. Dietzenbacher (1992) provided a detailed analysis of inter industry linkages in the Netherlands. The deviations from equilibrium in 1984 were minimal, see figure 7.8. One does not construct complete input–output tables on a quarterly basis.

On the basis of this information one would expect the biggest contemporaneous correlation in the disturbances between consumption and imports. This is not the case in our sample. We may infer that the disturbances are not entirely dominated by measurement errors and attach some meaning to impulse response analyses and variance decompositions. First we

evaluate the outlier component, i.e. the variation in the data that could not be used in the VAR model.

Table 7.10 *Extract from Dutch input–output table* 1984

	building	oil	intermediate	c	i	x	total
building	9345	464	17684	1178	30407	1739	51008
oil	795	6192	15754	5877	100	59699	63218
m	8066	48531	150343	38845	18336	20231	227755
total	51008	63218	762346	236749	74318	248563	

Source: Nationale Rekeningen 1986, Aanhangsel 2.1. Figures in 10^6 guilders of 1984. Totals from entire input-output table. Building and oil(=oil and chemicals) industry part of intermediate goods and services. c: private consumption, i: investment, x: exports, m: imports.

Evaluation of additive outlier correction

The normality tests and multivariate outlier tests are used to check whether the outlier adjustment has been satisfactory. Many outliers follow the one–period AO model in the levels, used in the estimation procedure described in appendix A5.3. Figure 7.13 presents the estimates of the outlier component together with the seasonally adjusted series, both in first differences.

The outlier correction is most urgently needed in the investment series, and seems to work well there. The negative additive outlier in 1963.1 was due to the most severe winter in Western Europe of this century. A compensating rise in 1963.2 did not occur. The negative outlier in 1979.1 (also due to a severe winter) was compensated to a great extent in the following quarter. This winter ended somewhat earlier and the economy was not working at full capacity in this period which made it easier to realize this compensation.

The outlier adjustment did not work perfectly for the consumption series. It is clear that the first observation is influential, but without previous observations one cannot tell whether it is an outlier. The 1966.1 drop in consumption is a consequence of the anticipatory purchases in the quarter before. Consumers apparently expected an important rise in the prices of consumer goods in the beginning of 1966. This price rise did not occur. The beginning of the trouble for the core VAR model thus occurred in 1965.4. The (univariate) test for additive outliers picked this up. The outlier correction method did not. One could use a priori information to fine tune the method on this point. Another way to improve the method could be to use additive outlier test statistics instead of normalized residuals (see §A5.3) in the assessment of the need for adjustment.

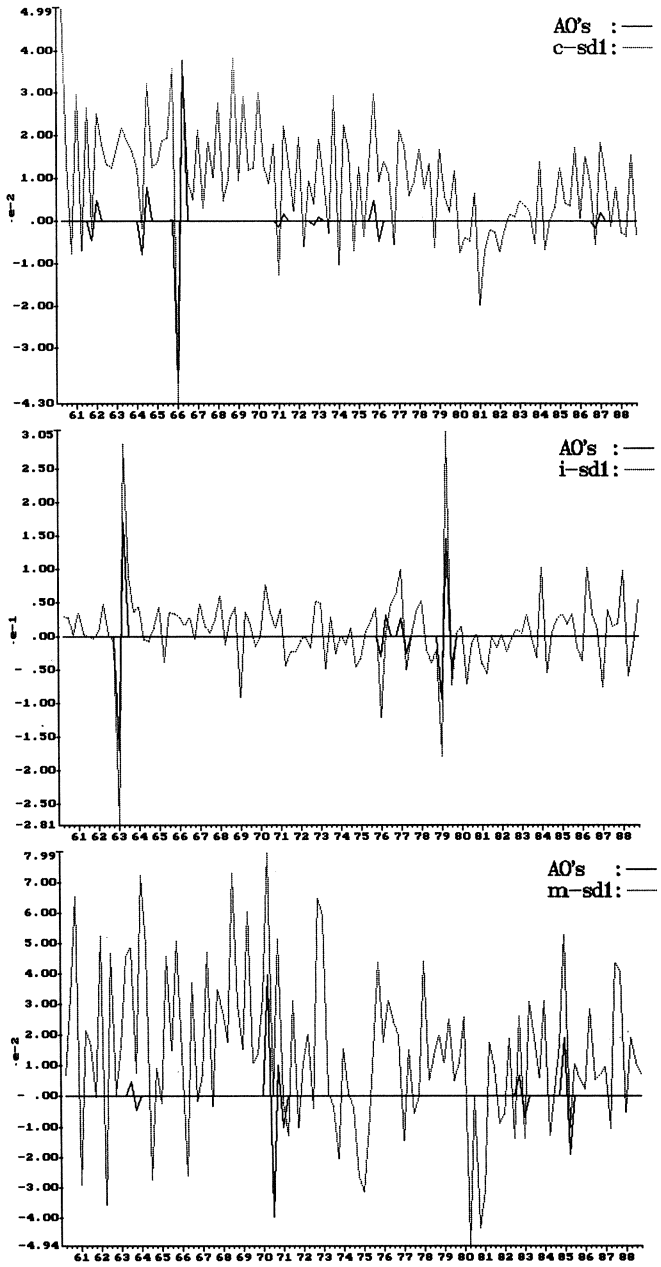

Figure 7.13a *Final estimates additive outlier component and seasonally adjusted growth rates* Additive outlier component estimated in restricted VECM model for seasonally adjusted data. *c-sd1*: consumption, *i-sd1*: investment, *m-sd1*: imports.

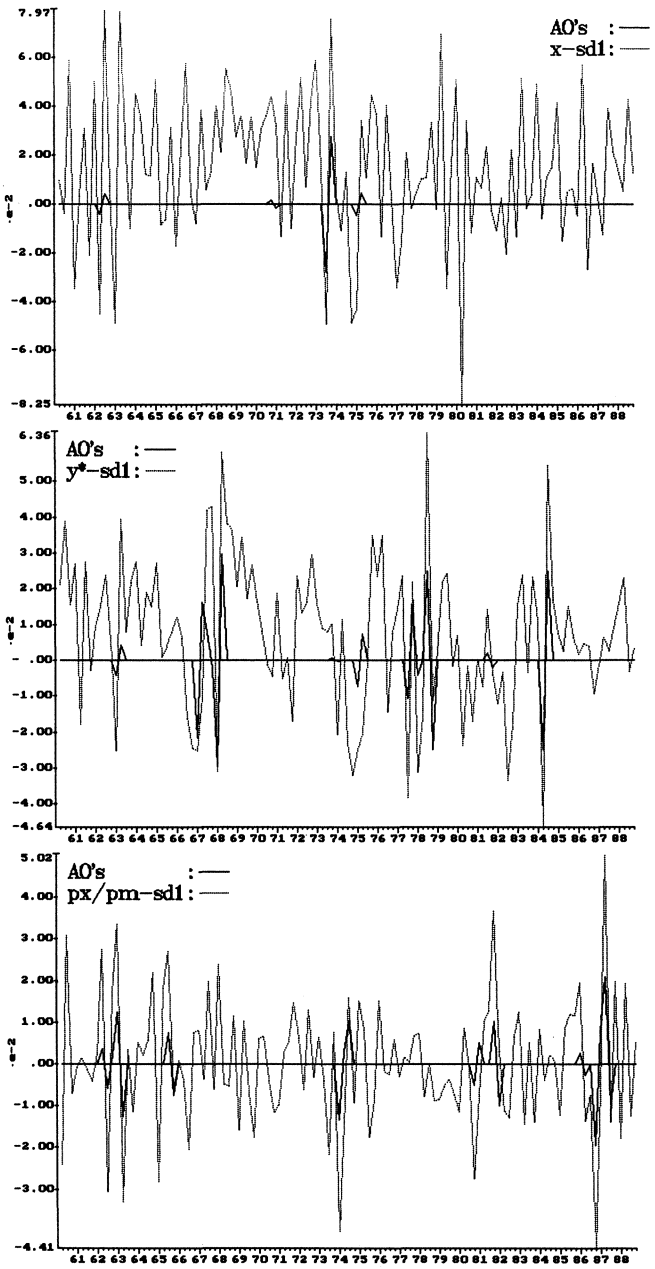

Figure 7.13b *Final estimates additive outlier component and seasonally adjusted growth rates (continued)* Additive outlier component estimated in restricted VECM model for seasonally adjusted data. *x-sd1*: exports, *y*-sd1*: German industrial production, *px/pm-sd1*: terms of trade.

The movement in industrial production figures in Germany in 1967–1968 is not well explained by the standard model. Apparently some important determining factors are missing. The other important outliers in the data are basically one time additive outliers, which show up as a $O_1, -O_1$ at $t = l, l+1$ in the time series plots of first differences.

Evaluation of correction for structural break

The plausibility of the permanent innovation outlier model can be judged from figures 7.14 and 7.15. The permanent effect of this outlier on the change in mean growth is largest for imports and exports. The mean growth in consumption is also affected considerably. The mean growth rates of investment and German industrial production are not as badly affected. The small mean growth rate in the terms of trade is reduced by the innovation outlier. The aim of the adjustment for outliers or structural breaks is to obtain data with stationary growth rates. Figure 7.15 shows that the adjustment for one structural break removes the major part of the apparent nonstationarity of the growth rates.

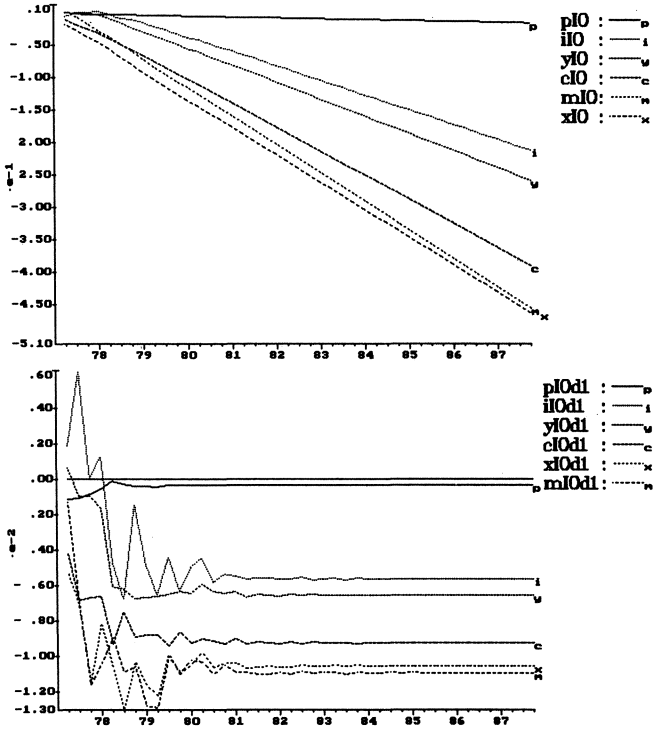

Figure 7.14 *Effect of permanent innovation outlier in* 1977.1
Upper figure (IO) gives effect on levels. Lower figure (IOd1) gives effect on growth rates. See also figure 7.15.

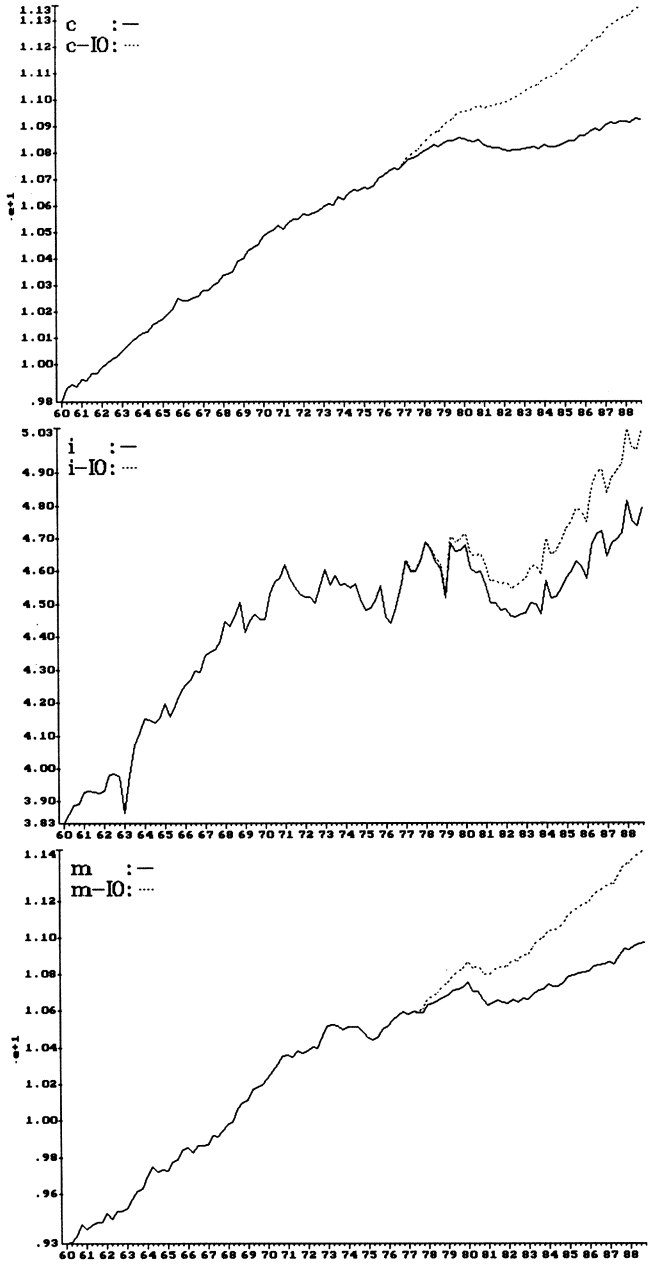

Figure 7.15a *Adjustment for breaking trends* Dotted lines (-IO) give seasonally adjusted series corrected for additive outliers minus the (negative) effect depicted in the upper part of figure 7.14. *c*: consumption, *i*: investment, *m*: imports.

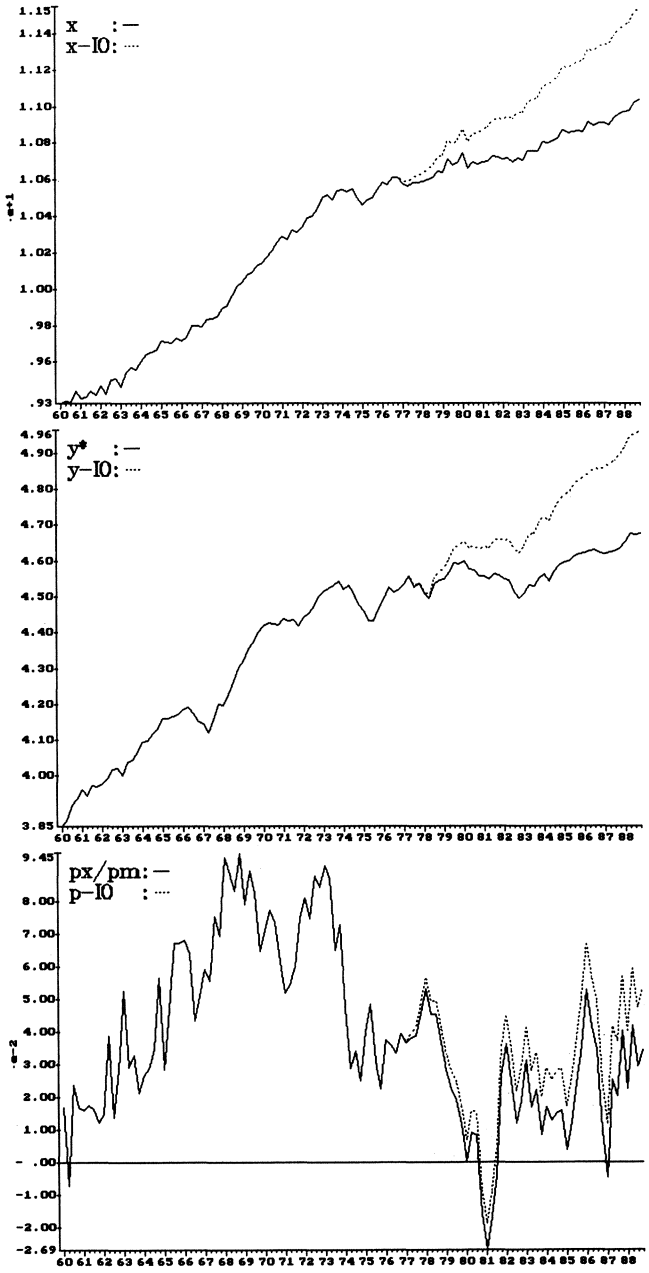

Figure 7.15b *Adjustment for breaking trends (continued)* Dotted lines (-IO) give seasonally adjusted series corrected for additive outliers minus the (negative) effect depicted in the upper part of figure 7.14. x: exports, y*: German industrial production, *px/pm*: terms of trade.

Diagnostic tests

We test the model using the OLS estimates per equation on the data corrected for additive outliers. We check the assumption of a mean innovation process for the disturbances by joint F–tests on the significance of the omitted regressors and lagged disturbances of all the equations up to order 4. We obtain no p–values under 5%. Multivariate and single equation normality of the disturbances is not rejected at the 5%–level either.

Recursive tests are applied on the data corrected for the structural break, since those are the data which should have stable growth rates. The results for the parameter stability tests improve considerably. The number of outliers criterion is also used on the equations estimated by OLS and gives no indication of serious parameter instability, see table 7.11.

Table 7.11 *Parameter stability tests restricted structural break model*

Test	c	i	m	x	y^*	$(px-pm)$c.v.5%	$[T/m]$·5%	
Fluc_f	.49	.91	.98	.91	.99	.77		(1)
Fluc_b	.11	.18	.41	.65	.98	.95		(2)
CUSUM_f	>.05	>.05	>.05	>.05	>.05	>.05		(3)
CUSUM_b	>.05	>.05	>.05	>.05	>.05	>.05		(4)
CUSUM_{OLS}	.73	.81	.93	.93	.84	.75		(5)
N_1	5	4	6	7	5	9	10	(6)
N_2	3	3	3	3	1	1	7	(7)
N_8	1	1	0	0	2	0	3	(8)
N_{16}	0	1	0	1	0	0	2	(9)
N_{32}	0	1	0	1	0	0	2	(10)
p_{1min}	.015	.009	.010	.013	.020	.037	.00045	(11)
p_{2min}	.010	.002	.024	.019	.016	.013	.0009	(12)
p_{8min}	.027	.021	>.05	>.05	.020	>.05	.0038	(13)
p_{16min}	>.05	.008*	>.05	.046	>.05	>.05	.0083	(14)
p_{32min}	>.05	.017	>.05	.024	>.05	>.05	.0167	(15)

NOTES: See also table 7.9. Rows (1)-(5): p-values. The restrictions are given in appendix A7.2. The effect of the permanent innovation outlier is presented in figures 7.14 and 7.15. The p-values do not take the conditioning on the nonstationary structural break regressor into account. Rows (6)-(10): N_m: number of Chow test statistics on nonoverlapping subsets of size m greater than 5% critical value. Column headed by C.V. 5% denotes critical value C_N, see §3.3.7. Rows (11)-(16): $p_{m\,min}$ denotes minimal p-value for Chow statistics on subsets of size m. Column headed by $[T/m]$·5% denotes approximate overall 5% critical value for these statistics, see §3.3.7. Asterisk denotes only significant value for investment equation in period 1977–1980 (16 quarters).

Short run identification impulse responses and variance decompositions

The first way to orthogonalize the shocks is to use the estimates of the covariance matrix Σ of the disturbances. One needs a short run model to identify shocks for this end. We call this *short run identification*. Sims (1986) proposed to estimate a simultaneous equations like model to achieve this identification. We use a short run recursive ordering. We attribute the

foreign price shocks to the disturbances in the terms of trade. We allocate the real foreign shocks to the components of the disturbances in exports and the German industrial production which are unrelated to the foreign price shocks.

Apparent orderings

There is strong evidence that the terms of trade can be considered exogenous. See appendix 7.2 and figures 7.16–7.19 below. The disturbances of imports and exports are very closely related. As the estimates of the VECM model in appendix 7.2 show, exports adjust to "longer run" deviations from the extended trade balance equilibrium only in the longer run, whereas changes in imports react faster to past changes in exports. The evidence of the reaction of imports to deviations from the equilibrium is not as clear. In the short run the export disturbances move somewhat more independently than the import disturbances. It seems sensible to attribute these disturbances to foreign shocks. It is hard to label the component of the import disturbances which cannot be explained by the disturbances in exports and foreign prices and production.

The remaining part of the disturbances in consumption and investment which cannot be attributed to foreign shocks can be identified as domestic. The correlation between the consumption and investment disturbances is comparatively low, so that changing their order does not influence the results of the analysis so much.

For the short run identification we use a recursive structure of the disturbances in the order px/pm, y^*, x, m, c, i.

Long run identification impulse responses and variance decompositions

One can also orthogonalize the shocks on the basis of the long run covariance matrix $C(1)\Sigma C(1)'$, which is proportional to the spectral density of the growth rates at the zero frequency. We call this *long run identification*, see example 6.2. The impulse responses in our model get very close their long run levels within a period of five to six years. It is thus somewhat more realistic to talk about *medium run identification*. As the analysis of the predictabilities in figure 7.10 and the statistical analysis of the cointegrating vector in table 7.6 shows, the evidence about real long run equilibria is rather weak.

Temporary responses and permanent effects

In the long term identification we first examine the effect of the (business cycle) shocks which have no permanent effect on the levels of the series. In our case with one cointegrating vector the coefficient vector which determines these shocks is parallel to the vector of factor loadings γ of the

error correction variable, see proposition 6.3. These shocks only have important effects on investment and exports. The variation in these shocks is largely dominated by the investment disturbances. The letter "i" denotes the effect of this shock in figures 7.17 and 7.19.

One can identify the other five shocks, which push the common trends, according to a similar recursive structure as we use for the short run identification. We employ the recursive structure in the order px/pm, y^*, x, m, c. The first three of these five shocks can be considered foreign. Only the first shock has a permanent effect on the terms of trade. Only the the first and the second shock have a permanent effect on the terms of trade and German industrial production and so on.

Empirical results

The results of the impulse response analysis are presented in figures 7.16 to 7.19. Figures 7.16 and 7.18 present the results from the short run identification. The responses to one standard error shocks are measured in percentage points of the levels of the original variables. The standard errors are derived from the ML estimate of the covariance matrix of the disturbances. The ML estimates are known to be downward biased in small samples. The responses are thus no "overestimates". The forecast error variance components are measured in percentage points of the total forecast error variance.

Note the recursive structure from the left hand sides of the graphs. The figures which represent the effect on the terms of trade indicate that only foreign price shocks have a notable influence. The same effect arises in figures 7.17 and 7.19 where we present results from the long run identification. Here the recursive structure can be noted from the right hand sides of the graphs. The plot of the impulse responses *on* i does not show six nonzero effects at the right hand side. The effect *of* i denotes the effect of the business cycle shock which has long run effect zero on all variables by construction.

Effects of foreign and domestic disturbances

The German industrial production series can also be considered rather exogenous in this information set, see figures 7.18 and 7.19. Negative shocks in the terms of trade (a rise of Dutch foreign prices, e.g. due to oil price changes in the seventies) have a negative effect on German industrial production. Surprisingly the dynamic relationship between German industrial production and domestic consumption is more pronounced than its relationship with domestic investment, see the effect of y *on* c and y *on* i in figures 7.18 and 7.19. This can also be seen directly from the plots of the raw data in figure 7.1.

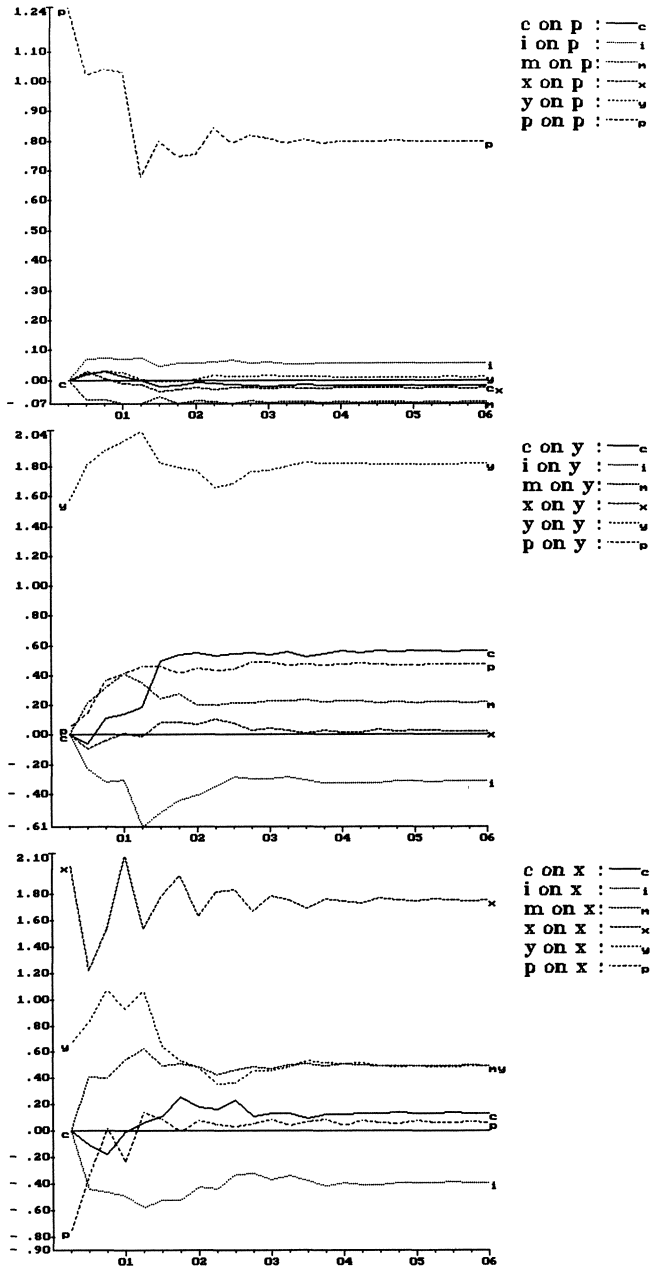

Figure 7.16a *Impulse responses identified from short run covariance matrix*
Recursive model with order *p*: terms of trade, *y*: German industrial production, *x*: exports, *m*: imports, *c*: consumption, *i*: investment. Effect of one standard shock measured in percentage points of the level of a variable (y-axis) after n years (x-axis).

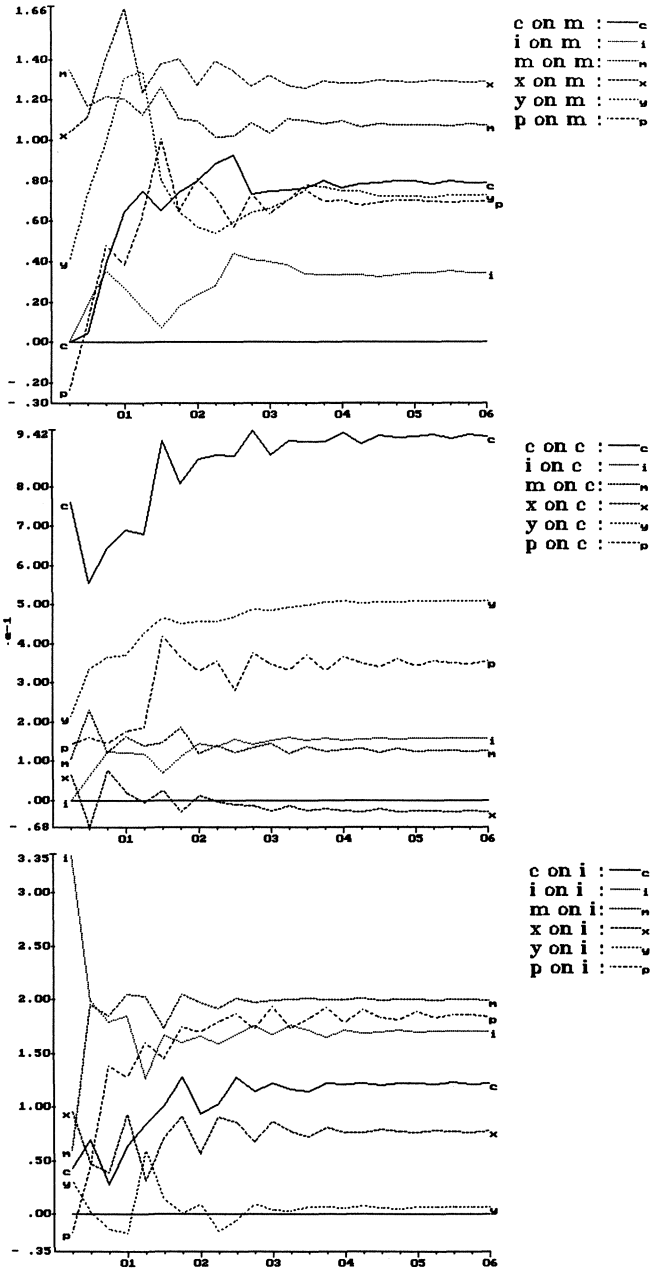

Figure 7.16b *Impulse responses identified from short run covariance matrix (continued)* Recursive model with order *p*: terms of trade, *y*: German industrial production, *x*: exports, *m*: imports, *c*: consumption, *i*: investment. Effect of one standard shock measured in percentage points of the level of a variable (y-axis) after n years (x-axis).

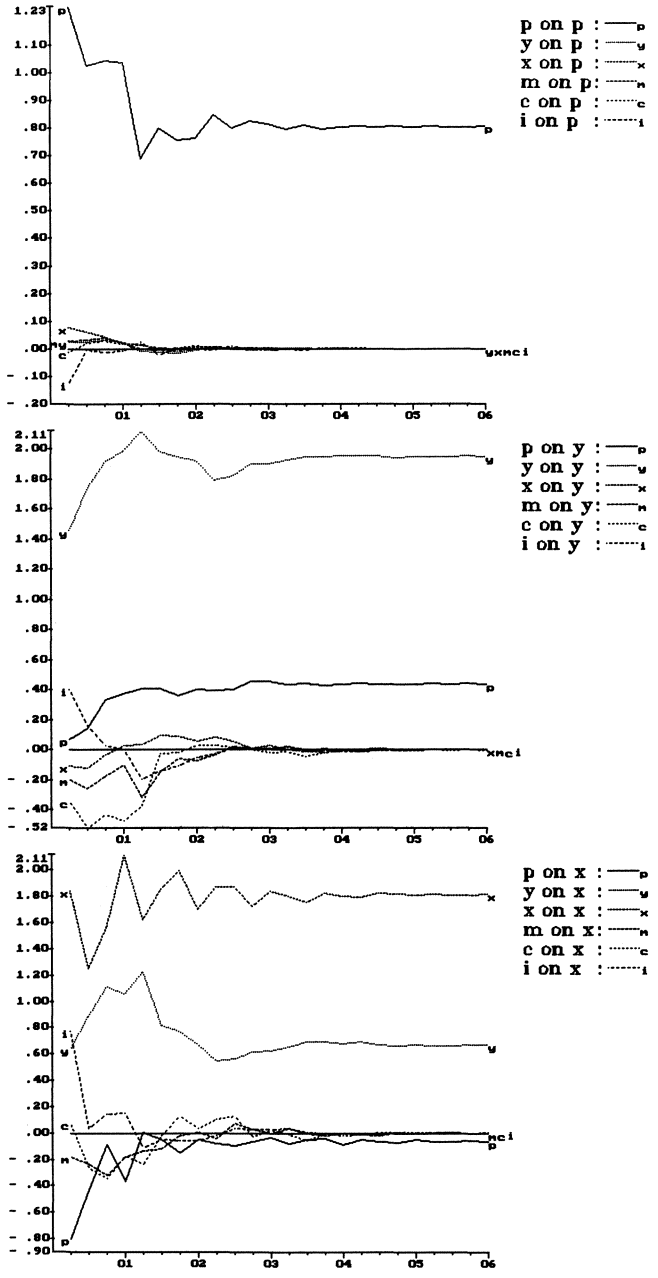

Figure 7.17a *Impulse responses identified from long run covariance matrix*
Recursive model with order *p*: terms of trade, *y*: German industrial production, *x*: exports, *m*: imports, *c*: consumption, *i on*: business cycle shock. *on i*: on investment. Effect of one standard shock measured in percentage points of the level of a variable (y-axis) after n years (x-axis).

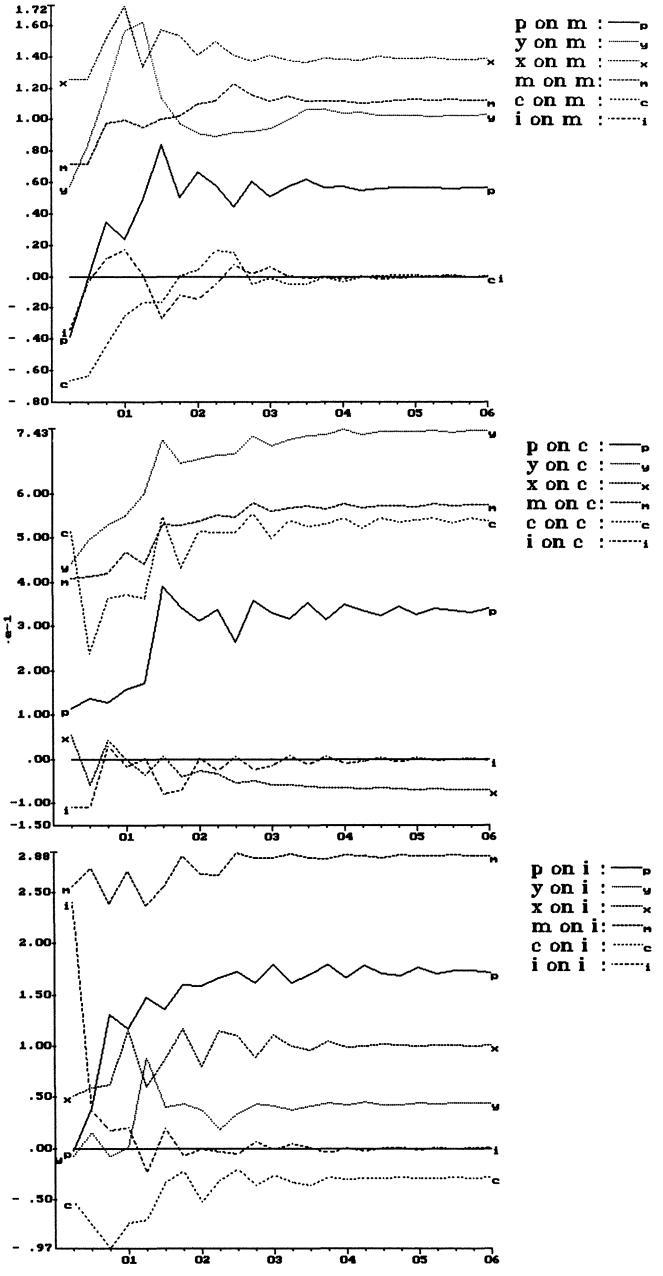

Figure 7.17b *Impulse responses identified from long run covariance matrix (continued)* Recursive model with order *p*: terms of trade, *y*: German industrial production, *x*: exports, *m*: imports, *c*: consumption, *i on*: business cycle shock. *on i*: on investment. Effect of one standard shock measured in percentage points of the level of a variable (y-axis) after n years (x-axis).

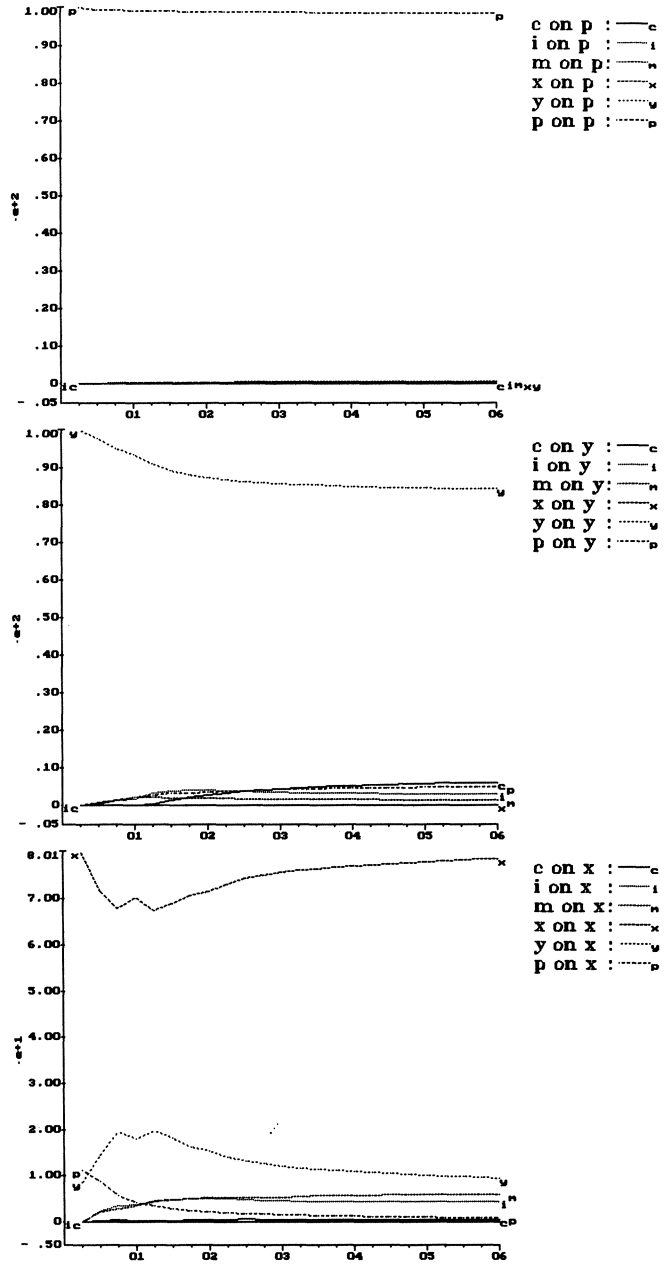

Figure 7.18a *Forecast error variance decompositions identified from short run covariance matrix* Recursive model with order *p*: terms of trade, *y*: German industrial production, *x*: exports, *m*: imports, *c*: consumption, *i*: investment.Components measured in percentage points of the total forecast error variance (y-axis) for an n year horizon (x-axis).

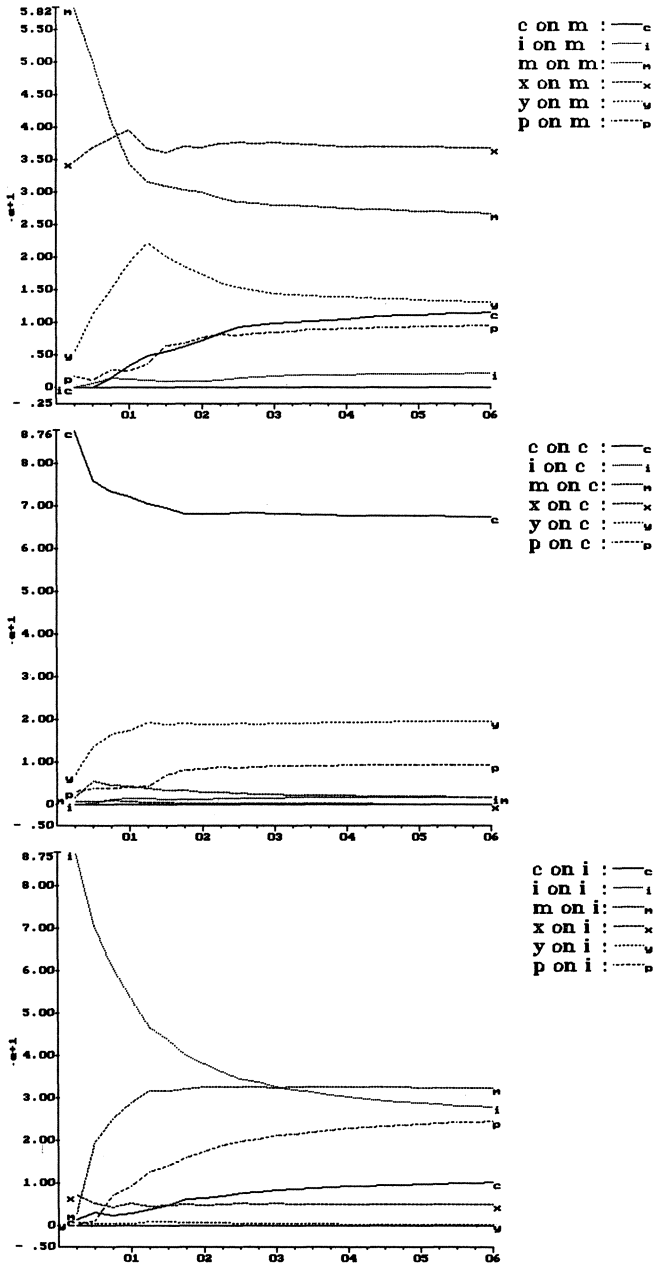

Figure 7.18b *Forecast error variance decompositions identified from short run covariance matrix (continued)* Recursive model with order *p*: terms of trade, *y*: German industrial production, *x*: exports, *m*: imports, *c*: consumption, *i*: investment. Components measured in percentage points of the total forecast error variance (y-axis) for an n year horizon (x-axis).

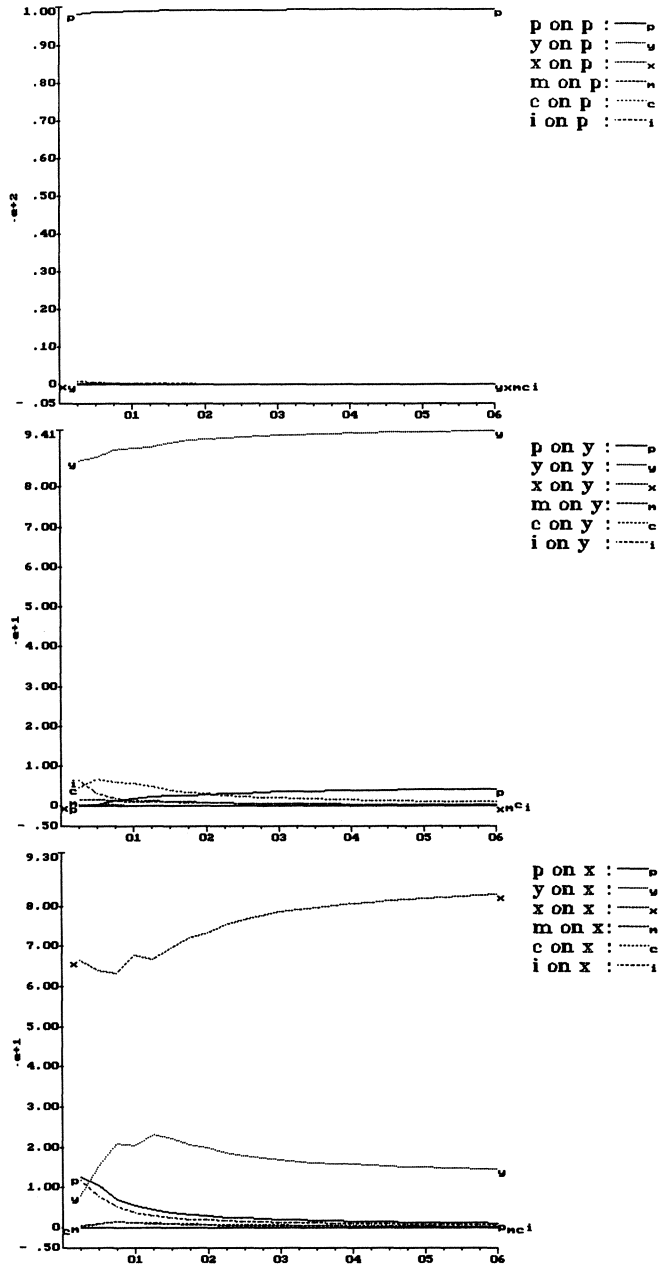

Figure 7.19a *Forecast error variance decompositions identified from long run covariance matrix* Recursive model with order *p*: terms of trade, *y*: German industrial production, *x*: exports, *m*: imports, *c*: consumption, *i on*: business cycle shock. *on i*: on investment. Components measured in percentage points of the total forecast error variance (y-axis) for an n year horizon (x-axis).

Exports only show a short run negative reaction to changes in the terms of trade. In the longer run positive real foreign shocks have a lasting positive impact, see the effects of *p on x*, *y on x* and *x on x* in figs. 7.16 and 7.17.

The results for the effects on the variables *m* and *c* differ widely depending on whether the identification is done on the short run or the long run covariances. The short run identification produces the more sensible results. Domestic shocks have a long run effect on imports which should not be assumed away. Imports show noticeable long lasting responses to both foreign and domestic shocks. The domestic investment disturbances have the smallest effect, especially if one measures them according to their contribution in the forecast error variance. If one combines the results from the exports and imports equation one sees that standard positive shocks in the terms of trade (*p on x* and *p on m*) and domestic consumption (*c on x* and *c on m*) have a comparable effect on the trade balance in the longer run. They have virtually no effect on exports and can both be used to explain about 10% of the long run forecast error variation of imports, see figures 7.16 and 7.18.

The most remarkable responses of consumption are those to foreign real shocks as represented by the German industrial production, which can explain about 20% of the long run forecast error variance.

Effects on investment

The variable of interest investment is the only variable which is significantly influenced by the business cycle shocks denoted by *i on* in figures 7.17 and 7.19. Only 15% of the forecast error variance at a one year horizon can be explained by these shocks. The shocks with permanent effect almost totally dominate the forecast error variance at further horizons. About 20% of the long run forecast error variance can be attributed to foreign price shocks. Only a quarter of the long run variance is explained by the "truly" foreign disturbances. (*p*, *y*, and *x*), see fig. 7.19. The truly domestic shocks from the short run identification (*c* and *i*) account for only 40% of the forecast error variance at the three year horizon, see fig. 7.18.

The most striking contribution to the forecast error variance comes from the import shocks, which are mixtures of domestic and foreign shocks. This dynamic relationship between investment and imports shows most clearly in the data in the partial correlation between investment growth rates and lagged deviations from the extended trade balance equilibrium (see (7.7)), which is statistically significant at very low significance levels (<0.01%). It is clear that imports figures can help in the prediction of investment at longer forecast horizons.

Interpretation

One could try to find a cause for this phenomenon in the measurement model. Investment good producing industries may anticipate future investment on the basis of information outside our VAR model, and import production factors in advance. Table 7.10 gives some figures for the building industry which supplied 40% of Dutch demand for investment goods in 1984. Only 3.5% of the total imports is used by the building industry. Industries which use a bigger amount of total import like the oil– and chemical industry (20% of imports) do not produce notable amounts of investment goods. The anticipated change in demand for investment goods cannot be the dominating reason for rises in total imports. Figure 7.8 shows that imports were comparatively low in the major economic troughs around 1975 and 1981. This can partly be explained by changes in the terms of trade and real foreign shocks. Other factors outside our information set may be able to account for these movements in domestic demand for foreign goods, which are more closely related with investment than obvious candidate variables like private consumption and exports. In appendix 7.3 we discuss the results in the context of the economic model of Mellander et al. (1992).

Conclusion

One can successfully analyze empirical dynamic relationships between six key macroeconomic variables for the Netherlands in the period 1960.1–1988.4 using the methods suggested in the previous chapters. The standard VAR model with some unit roots, extended with models for additive outliers, seasonality and a structural break can be used to explain a significant amount of the variation of the data set under scrutiny. The parameters of the core model are reasonably stable.

We find a strong dynamic relation between investment and imports, which is not easily explained by reference to aspects of the measurement model. We present empirical evidence of an "extended trade balance" equilibrium, which is mean stationary over the entire sample period. Lagged deviations from this equilibrium have a large partial correlation with investment growth rates. Comparatively high import levels tend to be rapidly followed by relatively high growth rates for investment. Changes in the terms of trade can be considered exogenous in our information set. Exports are only temporarily affected by changes in the terms of trade. Imports show a longer lasting reaction.

The methods are still developing. We encourage the reader to employ them critically. Modification will be necessary to obtain a good insight in linear dynamic relationships among other macroeconomic time series.

A7.1 Data sources and construction

This appendix contains detailed information about the construction and source of the data. Only the common sample 1960.1–1988.4 is used in the empirical analysis.

Table A7.1 *Data sources and construction*

Series		*Period*	*Construction*	
Total consumption by households	volume	57.1–69.4	(1980 prices)·0.923	(1)
		70.1–78.3		(2)
		78.3–91.2	Constructed from annual growth rates	(3)
Total investment: dwellings + other fixed capital + government investment stocks excluded	value	60.1–62.4	Index(1975=100)·.62·44.5/47.25	
		63.1–70.4	Index(1980=100)·44.5/47.25	
		71.1–89.4	Index(1985=100)	
	price	57.1–70.4	(Total value/Volume 1980 prices) ·.46110/.49011	(4)
		71.1–89.4	Value in market prices(1980=100) volume (1985=100)	
	volume	60.1–89.4	Value/price	(5)
Imports (c.i.f.) and Exports (f.o.b.) Goods and Services	volume	57.1–69.4		(6)
		70.1–78.1		(7)
		78.1–91.2	Constructed from annual growth rates	(8)
	value	57.1–78.1		(9)
		78.1–91.2		(10)
	price	57.1–91.2	Value/volume	
German industrial production	volume	60.1–89.1	Index (1985=100)	(11)

NOTE: Sources: (1): DNB (1986), (2): DNB (1986), (3): CBS (1991), (4): DNB (1986), (5): OECD (1991), (6): DNB (1986), (7): DNB (1986), (8): CBS (1991), (9) DNB (1991), (10) CBS(1991), (11): OECD (1991).

A7.2 Estimation results final model

Table A7.2 *Results of final VECM model SUR estimation*

Equation		Dc	Di	Dm	Dx	Dy*	Dpx/pm
Regressor	lag						
D c	1	-0.291	0.796				
		(-3.82)*	(2.39)*				
Dm	1	0.159					
		(3.45)*					
D x	1	-0.152			-0.308		
		(-4.15)*			(-4.70)*		
Dy *	1	0.126		0.167	0.331	0.212	
		(2.69)*		(1.66)*	(2.80)*	(2.53)*	
Dp x /pm	1			0.329			-0.158
				(3.03)*			(-1.84)*
D c	2			0.386		0.299	
				(2.63)*		(2.11)*	
D i	2			0.042			
				(1.22)			
Dm	2		-0.274	-0.008			
			(-1.90)*	(-0.10)			
D x	2			0.141			
				(2.01)*			
Dy *	2				0.169		
					(1.74)*		
Dp x /pm	2		0.420	0.426	0.270	0.079	
			(1.61)*	(3.13)*	(1.72)*	(0.70)	
D c	3			0.451			
				(3.01)*			
Dm	3			-0.129			
				(-1.55)*			
D x	3			0.287	0.246		
				(3.94)*	(3.19)*		
Dy *	3				-0.206		
					(-2.02)*		
D c	4			0.349			
				(2.41)*			
D i	4		-0.138			-0.096	
			(-1.73)*			(-3.01)*	
Dm	4			-0.201			
				(-3.11)*			
D x	4		-0.236		-0.131		
			(-1.86)*		(-1.73)*		
Dy *	4		0.364				
			(1.90)*				
Dp x /pm	4						-0.277
							(-3.31)*
D c	5	0.310				0.355	
		(4.44)*				(2.58)*	
Dm	5			0.056			
				(0.91)			
D x	5				-0.088		
					(-1.23)		
Dy *	5		-0.383	-0.359	-0.259	-0.223	
			(-2.01)*	(-3.50)*	(-2.16)*	(-2.69)*	
Dp x /pm	5	0.160		0.447			
		(2.78)*		(4.09)*			

continued on next page

Equation		Dc	Di	Dm	Dx	Dy*	Dpx/pm
Regressor	lag						
cointvec1	1	-0.0650	1.4257	-0.2012	0.4589	0.2372	-0.0747
		(-1.198)	(6.593)*	(-1.758)*	(3.333)*	(2.605)*	(-1.018)
constant	0	0.0132	-0.0293	0.0074	0.0116	-0.0048	0.0026
		(5.781)*	(-3.399)*	(1.515)*	(2.191)*	(-1.207)	(1.011)
sdum77.1	0	-0.0079	-0.0083	0.0014	-0.0144	-0.0017	-2.95e-4
		(-4.298)*	(-1.083)	(0.329)	(-3.102)*	(-0.501)	(-0.123)
Adjusted R^2		0.418	0.322	0.362	0.331	0.178	0.064
sigmahat		8.47e-3	3.76e-2	1.93e-2	2.39e-2	1.63e-2	1.28e-2
Durbin-Watson		2.053	1.983	2.013	2.072	2.099	2.012

log-likelihood: $1.81677 \cdot 10^3$

ML estimate covariance matrix disturbances $\Sigma \cdot 10^5$:

	Dc	Di	Dm	Dx	Dy*	Dpx/pm
D c	6.590	5.017	2.645	1.718	3.426*	1.759
D i		128.7	19.78 *	22.79 *	4.849	-2.135
Dm			31.22	25.44 *	6.240*	-2.912
D x				50.73	9.976*	-9.364*
D y *					24.37	.631
D p x/pm						15.51

Correlations disturbances

	Dc	Di	Dm	Dx	Dy*	Dpx/pm
D c	1.00	0.17	0.18	0.09	0.27	0.17
D i	0.17	1.00	0.31	0.28	0.09	-0.05
Dm	0.18	0.31	1.00	0.64	0.23	-0.13
D x	0.09	0.28	0.64	1.00	0.28	-0.33
D y *	0.27	0.09	0.23	0.28	1.00	0.03
D p x/pm	0.17	-0.05	-0.13	-0.33	0.03	1.00

$C(1)$-matrix

	c	i	m	x	y*	px/pm
c	1.1951	0.0471	-0.0206	-0.0658	0.1862	0.1008
i	1.3186	0.5084	1.1553	-0.4985	-0.3286	1.3389
m	0.9813	0.1028	0.6763	0.2085	0.0461	0.7135
x	0.2417	-0.1162	0.4033	0.7074	-0.0985	0.5177
y *	0.7949	-0.0933	0.1422	-0.0432	1.0570	0.2348
p x/pm	-0.0331	0.0171	-0.0599	0.0108	0.0190	0.6426

cointegrating vector cointvec: α'

c	i	m	x	y*	px/pm
-0.1507	-0.2845	0.8001	-0.4614	-0.1425	0.1519

Long run covariance matrix: $C(1)\Sigma C(1)' \cdot 10^2$, subdiagonal triangle: correlations.

	c	i	m	x	y*	px/pm
c	0.0129	0.0232	0.0150	0.0035	0.0159	0.0027
i	0.58	0.1245	0.0605	0.0201	0.0162	0.0139
m	0.62	0.80	0.0458	0.0317	0.0226	0.0046
x	0.16	0.29	0.77	0.0373	0.0128	-0.0005
y *	0.70	0.23	0.53	0.33	0.0400	0.0035
p x/pm	0.30	0.49	0.27	-0.03	0.22	0.0065

NOTES: Each column contains results of one equation. t-values in parentheses (asterisk for values greater than 1.4 in absolute value). Prefix D denotes first difference. All variables are logged. *c*: consumption, *i*: investment, *m*: imports, *x*: exports, *y**: German industrial production, *px/pm*: terms of trade. Asterisks in covariance matrix disturbances indicate values greater than $2T^{-\frac{1}{2}}$.

A7.3 Open economy dynamic stochastic general equilibrium models

If we view the results of §7.6 in the light of the open economy real business cycle model of Mellander et al. (1992), we find little support for the applicability of their model to our data. Their model predicts the stationarity of the first "greater ratio" I/Y, with I domestic investment and Y domestic product, whereas the log of the second greater ratio C/Y should be cointegrated with the log of the terms of trade. If one substitutes out domestic product one should find cointegration among log investment, i, log consumption, c, and log terms of trade, px/pm in the following manner:

$$i - c + (1-\beta)(px/pm) \tag{A7.1}$$

with β the share of goods produced at home in total consumption.
The figures in Table 7.10 indicate that β is far from one. The estimate α_{rest2} in table 7.6, which incorporates the four restrictions on the cointegrating vector does not give a sensible estimate of β, i.c. 6.7. If we only examine the likelihood of an equilibrium relation between these three variables we find a p–value of this restriction of 0.7% even if we apply the small sample correction for the evaluation of the likelihood ratio test for these restrictions. The positive long run relationship between the terms of trade and the share of consumption in national income is not seen in the data. We find that investment reacts more strongly to changes in the terms of trade than consumption. The key assumption in the model of Mellander et al. which is not fulfilled in the Dutch situation is the assumed equality of the terms of trade to the ratio of the price of tradable goods at home and the price of foreign goods not produced at home. Dutch export prices adjust very quickly to foreign prices. This manifested itself most clearly in the seventies where yearly increases in import prices were matched by increases in export prices far beyond the increases in the prices of home goods.

Backus et al. (1992) attempted to interpret "stylized facts" in a real business cycle context and found many empirical high frequency correlations between macroeconomic variables hard to interpret. They explained the negative correlation between investment and the trade balance as resulting from country specific technology shocks.

SUMMARY

Means and goal

In this study we integrate results from the literature in the fields of econometrics, time series analysis and macroeconomics which concern linear multivariate modeling of macroeconomic time series. The aim of the study is to develop, implement and apply a comparatively reliable method to detect interesting linear dynamic relationships between a limited number of macroeconomic time series. The method is primarily data based.

We treat seasonal variation in the data as an unobserved underlying component. The method takes the possibility of a limited number of nonsystematic outliers into account. The method also presents a multivariate decomposition of the observed series into trend– and cyclical components. The cyclical component consists of all predictable changes in deviation from the mean rate of growth.

The vector autoregressive model is the most important tool in the analysis. We represent the properties of this model in impulse responses to artificial mutually independent shocks. We are also interested in the corresponding variance decompositions of the variables, like Sims (1980). Chapter one treats the goal and set–up of the study and the originality of our contribution.

Univariate analysis

In chapter two we introduce the basic vector autoregressive model. From this model we derive some univariate properties of the variables in it. The specific memory characteristics of the time series observations of these variables get special attention. We discuss a wider class of model with long memory properties. Within this model class one can test specific characteristics of candidate variables for the vector autoregressive model.

We propose three methods to do this. The first is based on variance ratios of partial sums of observations. The second uses spectral estimates. This method is also readily applicable for a specific analysis of the memory characteristics of the seasonal component. The third method uses ratios of empirical ranges of partial sums and (long–run) variances. In appendix 2.2 we discuss regression methods for univariate unit root testing. One can apply these methods both to the trend component of the yearly mean as well as to the seasonal component.

Influence analysis and diagnostic tests

In chapter three we first discuss influence analysis. We order statistical influence measures on the basis of a number of characteristics. We look at the kind of perturbations in the model specifications considered. We also distinguish the methods according to the goal of the model which one applies the influence analysis to. Furthermore we look at the size of the perturbations which are taken into account. In appendices 3.1 and 3.2 we discuss univariate and multivariate influence measures for the location and (co)variation of predictions and estimates of location and covariance parameters.

Next we discuss the use of diagnostic tests in the vector autoregressive model. We propose some new tests too. The influence analysis and application of diagnostic tests serves two goals. The first is the detection of interesting information in the data. The second goal is to obtain an intersubjective quantitative judgment of the fitness of a specific vector autoregressive approximation as a representation of the sample information.

Seasonal components

In chapter four we discuss the application of the idea of decomposition of a series into unobserved components. We give special attention to the memory characteristics of the seasonal component. Next we discuss the five aspects of the estimation procedure of this component. We propose a method of exploratory data analysis which improves the understanding of the seasonal component's behavior. Last we discuss a simple implementation of a reliable method of seasonal adjustment. Out of sample backforecasting forms an important part. We treat this aspect in more detail in appendix 4.2

Outliers

In chapter five we discuss models for temporary and (nearly) permanent outliers in vector autoregressive models. We treat some consequences of uncareful modeling of outliers. We derive a number of Lagrange multiplier test statistics on the absence of different types of outliers and propose a testing strategy. After an artificial application we use the proposed strategy successfully on the analysis of French macroeconomic data. It appears that outliers from 1968 and the seventies only had a temporary character. The surprising shocks of the eighties had a longer lasting effect.

Acceptable restrictions, pushing trends and pulling equilibria

In chapter six we discuss the use of acceptable restrictions in the vector autoregressive model which can help the estimation and interpretation of impulse responses and variance decompositions. We concentrate on restrictions on long–term aspects. We derive straightforward relations between

different parameterizations with interesting interpretations. These parameterizations, the common trends model, the vector error correction model and a generalization of the Campbell–Shiller (1987) model, which explicitly describes the deviations from long–term equilibrium, are only equivalent to a certain degree.

In §6.5 we discuss alternative vector autoregressive methods to detect the number and form of the pushing trends and pulling equilibria. The notion of predictability, which one defines as the variance ratio of the forecast errors and the variable to be predicted, plays a crucial role. In the so–called cointegration analysis one is primarily interested in the existence of linear combinations of variables which are not predictable in the (very) long run. Most samples contain little information about this. One can usually say more for short(er) forecast horizons. We propose a simple way to depict the predictability of the linear combinations of the series. Finally we discuss the use of different unit root tests in multivariate time series.

Empirical application

Chapter seven contains a comprehensive application of the methods discussed in previous chapters. The main focus is on aggregate investment in fixed capital in the Netherlands in the period 1960–1988. First we discuss a number of theoretically appealing relationships of investment with other macroeconomic variables. Most theoretic models could only have been applied successfully if one would be able to obtain a data set much more informative than ours.

Next we discuss a number of empirical models. Many empirical models incorporate a substantial amount of information from economic theory. Economic theory is particularly used in the construction of key variables like the net return on alternative investments, the capital stock and the effective rental price of capital and other production factors.

Our choice of variables corresponds closely to modern macro models in which vector autoregressive models form the empirical basis. In §7.3 we discuss aspects of the measurement model for the macroeconomic variables in our application. In the next section we treat the results of the univariate analysis of Dutch investment, private consumption, imports, exports and terms of trade and German industrial production. The analysis shows the presence of outliers, nonstationary seasonal components and at least one structural change in the growth rates of these variables. We adjust the application of unit root tests to this situation.

In the following multivariate analysis we confirm the presence of outliers and a structural break. The most interesting partial correlation

concerning investment growth rates is the one with lagged deviations from a medium–run equilibrium between levels of imports, exports and investment itself. This partial correlation is large, stable and remarkably significant throughout the entire sample.

Finally we present impulse responses and variance decompositions based on two different normalizations. The first normalization is based on short–run correlations between the innovations, the second one uses long–run correlations between the growth rates. The impulse responses of the terms of trade, the German industrial production and exports do not differ much across the normalizations. The innovations of these variables mainly describe foreign shocks, which are not remarkably influenced by the domestic factors in our information set. Exports appear to be influenced by innovations in the terms of trade in the short run only. The effect on imports is longer lasting.

Concluding remarks

The proposed combination of methods can be applied successfully. The selected vector autoregressive model has stable parameters. It gives a systematic description of an interestingly large part of the observed variation in our information set. The presented partial correlations can, according to recent econometric standards, not be labeled as "spurious". One can easily interpret the economic meaning of a number of partial correlations. It appears that variables determining the trade balance in goods and services are valuable in predicting investment growth rates in the Netherlands.

REFERENCES

Abraham, B. (1977), "Intervention in Multiple Time Series," pp. 611-615 in *Proceedings of the Business and Economic Statistics Section of the American Statistical Association*, American Statistical Association, Washington D.C., USA.

Abraham, B. (1980), "Intervention Analysis and Multiple Time Series," *Biometrika*, 67, 73-78.

Abraham, B., and Chuang, A. (1989), "Outlier Detection and Time Series Modeling," *Technometrics*, 31, 241-248.

Abraham, B., and Yatawara N. (1988), "A Score Test for Detection of Time Series Outliers," *Journal of Time Series Analysis*, 9, 109-119.

Ahn, S.K., and Reinsel, G.C. (1988), "Nested Reduced Rank Autoregressive Models for Multiple Time Series," *Journal of the American Statistical Association*, 83, 849-856.

Ahn, S.K., and Reinsel, G.C. (1990), "Estimation for Partionally Nonstationary Multivariate Autoregressive Models," *Journal of the American Statistical Association*, 85, 813-823.

Al, P.G., Balk, B.M., De Boer, S., and Den Bakker, G.P. (1985), The use of chain indices for deflating the national accounts, no. NA-005, National accounts, Netherlands Central Bureau of Statistics, Voorburg, The Netherlands.

Ali, M.M. (1989), "Tests for Autocorrelation and Randomness in Multiple Time Series," *Journal of the American Statistical Association*, 84, 533-540.

Alogoskoufis, G., and Smith, R. (1991), "On Error Correction Models: Specification, Interpretation, Estimation," *Journal of Economic Surveys*, 5, 97-128.

Alt, R. (1989), "Hierarchical Test Problems and the Closure Principle," pp. 162-176 in *Multiple Hypothesenprüfung*, eds. Bauer, P., Hommel, P., and Sonneman, E., Springer Verlag, Berlin, Germany.

Amemiya, T. (1985), *Advanced Econometrics*, Basil Blackwell, Oxford, UK.

Anderson, O.D. (1983), "A Simplified Theory for the Serial Covariance Structure of ARMA, ARIMA and ARUMA models," pp. 291-301 in *Time Series Analysis: Theory and Practice* 2, ed. Anderson, O.D., North-Holland, Amsterdam, The Netherlands.

Anderson, T.W. (1958), *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, New York, USA.

Anderson, T.W. (1971), *The Statistical Analysis of Time Series*, John Wiley and sons, New York, NY, USA.

Anderson, T.W. (1984), *An Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, New York, USA, Second edition.

Andrews, D.W.K. (1990), Tests for Parameter Instability and Structural Change with Unknown Change Point, Discussion Paper, no. 943, Cowles Foundation of Research in Economics at Yale University, Yale University, New Haven, Connecticut, USA.

Ansley, C.F., and Wecker, W.E. (1984), "Comment On Dips in the Spectrum of Seasonally Adjusted Time Series," *Journal of Business and Economic Statistics*, 2, 323-324.

Arrow, K. (1960), "Decision Theory and the Choice for a Level of Significance for the t-Test," pp. 70-78 in *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, eds. Olkin, I., Ghurye, S.G., Hoeffding, W., Madow, W.G., and Mann, H.B., Stanford University Press, Stanford, USA.

Backus, D.K., Kehoe, P.J., and Kydland, F.E. (1992), "International Real Business Cycles," *Journal of Political Economy*, 100, 745-775.

Bailey, B.J.R. (1977), "Tables of Bonferroni statistics," *Journal of the American Statistical Association*, 72, 469-478.

Baillie, R.T. (1989), "Econometric Tests of Rationality and Market Efficiency," *Econometric Reviews*, 8, 151-186.

Bakker, E. (1962), "De voorspellingskwaliteit van het bouwprogramma," *De Economist*, 110, 739-751.

Banerjee, A., Lumsdaine, R.L., and Stock, J.H. (1992), "Recursive and Sequential Tests of the Unit-Root and Trend-Break Hypotheses: Theory and International Evidence," *Journal of Business and Economic Statistics*, 10, 271-287.

Barnett, V., and Lewis, T. (1984), *Outliers in Statistical Data*, John Wiley, New York, NY, USA, second edition.

Barsky, R.B., and Miron, J.A. (1989), "The Seasonal Cycle and the Business Cycle," *Journal of Political Economy*, 97, 503-534.

Beaulieu, J.J., and Miron, J.A. (1993), "Seasonal unit roots in aggregate U.S. data," *Journal of Econometrics*, 55, 305-328.

Bell, W. (1984), "Signal Extraction for Nonstationary Time Series," *Annals of Statistics*, 12, 646-664.

Bell, W.R. (1987), "A Note on Overdifferencing and the Equivalence of Seasonal Time Series Models With Monthly Means and Models With $(0,1,1)_{12}$ Seasonal Parts When $\theta = 1$," *Journal of Business and Economic Statistics*, 5, 383-387.

Bell, W.R., and Hillmer, S.C. (1984), "Issues Involved With the Seasonal Adjustment of Economic Time Series," *Journal of Business and Economic Statistics*, 2, 291-320.

Beniger, J.R., and Robyn, D.L. (1978), "Quantitative Graphics in Statistics: A Brief History," *The American Statistician*, 32, 1-11.

Bera, A.K., and Jarque, C.M. (1982), "Model Specification Tests, A Simultaneous Approach," *Journal of Econometrics*, 20, 59-82.

Beveridge, S., and Nelson, C.R. (1981), "A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the "Business Cycle"," *Journal of Monetary Economics*, 7, 151-174.

Bewley, R.A. (1979), "The direct estimation of the equilibrium response in a linear model," *Economics Letters*, 3, 24-26.

Bewley, R., and Fiebig, D.G. (1990), "Why are Long-Run Parameter Estimates So Disparate?," *Review of Economics and Statistics*, 72, 345-349.

Bewley, R., Orden, D., Yang, M., and Fisher, L.A. (1993), Comparison of Box-Tiao and Johansen Canonical Estimators of Cointegrating Vectors in VEC(1) Models, University of New South Wales, PO Box 1, Kensington NSW Australia 2033, forthcoming *Journal of Econometrics*.

Bierens, H.J., and Guo, S. (1993), "Testing Stationarity and Trend Stationarity against the Unit Root Hypothesis," *Econometric Reviews*, 12, 1-32.

Billingsley, P. (1968), *Convergence of Probability Measures*, John Wiley and Sons, New York, USA.

Blanchard, O.J. (1989), "A Traditional Interpretation of Macroeconomic Fluctuations," *American Economic Review*, 79, 1146-1164.

Boswijk, H.P. (1992), *Cointegration, Identification and Exogeneity*, Thesis/Tinbergen Institute, Amsterdam, The Netherlands.

Box, G.E.P., and Jenkins, G.M. (1970), *Time Series Analysis, Forecasting and Control*, Holden-Day, San-Fransisco, CA, USA.

Box. G.E.P., and Jenkins, G.M. (1976), *Time Series Analysis, Forecasting and Control*, Holden-Day, San-Fransisco, CA, USA, second edition.

Box, G.E.P., and Muller, M.E. (1958), "A Note on the Generation of Random Normal Deviates," *Annals of Mathematical Statistics*, 29, 610-611.

Box, G.E.P., and Pierce, D.A. (1970), "Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models," *Journal of the American Statistical Association*, 65, 1509-1526.

Box, G.E.P., and Tiao, G.C. (1975), "Intervention Analysis with Applications to Economic and Environmental Problems," *Journal of the American Statistical Association*, 70, 70-79.

Box, G.E.P., and Tiao, G.C. (1977), "A Canonical Analysis of Multiple Time Series," *Biometrika*, 64, 355-365.

Bradley, M.G., and Lumpkin, S.A. (1992), "The Treasury Yield Curve as a Cointegrated System," *Journal of Financial and Quantitative Analysis*, 27, 449-463.

Breusch, T.S., and Godfrey, L. (1981), "A Review of Recent Work on Testing for Auto-Correlation in Dynamic Simultaneous Models," pp. 63-110 in *Macroeconomic analysis: essays in macroeconomics and econometrics*, eds. Currie, D., Nobay, R., and Peel, D., Croom Helm, London, UK.

Breusch, T.S., and Pagan, A.R. (1979), "A Simple Test for Heteroscedasticity and Random Coefficient Variation," *Econometrica*, 47, 1287-1294.

Broer, D.P., and Van Leeuwen, G. (1991), Investment Behaviour of Dutch Industrial Firms, no. 9116/G, Institute for Economic Research, Erasmus University Rotterdam, The Netherlands.

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 Association, Series B*, 37, 149–163.

Bruce, A.G., and Martin, R.D. (1989), "Leave-k-out Diagnostics for Time Series," *Journal of the Royal Statistical Society, Series B*, 51, 363–424.

Bunzel, H., and Hylleberg, S. (1982), "Seasonality in Dynamic Regression Models," *Journal of Econometrics*, 19, 345–366.

Burridge, P., and Wallis, K.F. (1985), "Calculating the Variance of Seasonally Adjusted Series," *Journal of the American Statistical Association*, 80, 541–552.

Campbell, J.Y. (1987), "Does Saving Anticipate Declining Labor Income? An Alternative Test of The Permanent Income Hypothesis," *Econometrica*, 55, 1249–1273.

Campbell, J.Y., and Mankiw, N.G. (1991), "The response of consumption to income. A cross-country investigation," *European Economic Review*, 35, 723–767.

Campbell, J.Y., and Perron, P. (1991), "Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots," pp. 141–201 in *NBER Macroeconomics Annual* 1991, eds. Blanchard, O.J., and Fisher, S., MIT Press.

Campbell, J.Y, and Shiller, R.J. (1987), "Cointegration and Tests of Present Value Models," *Journal of Political Economy*, 95, 1062–1088.

Campos, J., Ericsson, N.R., and Hendry, D.F. (1993), Cointegration Tests in the Presence of Structural Breaks, International Finance Discussion Papers, no. 440, Board of Governors of the Federal Reserve System.

Canova, F., and Hansen, B.E. (1992), Are Seasonal Patterns Constant Over Time? A Test for Seasonal Stability, European University Institute, San Domenico di Fiesole (FI), Italy.

Cantrell, R.S., Burrows, P.M., and Vuong, Q.H. (1991), "Interpretation and use of generalized Chow tests," *International Economic Review*, 32, 725–741.

CBS, Centraal Bureau voor de Statistiek (1986), *Kwartaalrekeningen, Methodebeschrijving en Resultaten* 1977–1982, Staatsuitgeverij/cbs-publikaties, The Hague, The Netherlands.

CBS, Centraal Bureau voor de Statistiek (1987), *Nationale rekeningen* 1986, Staatsuitgeverij/cbs-publikaties, The Hague, The Netherlands.

CBS, Centraal Bureau voor de Statistiek (1988), *Nationale rekeningen* 1987, Staatsuitgeverij/cbs-publikaties, The Hague, The Netherlands.

CBS, Centraal Bureau voor de Statistiek (1991), *Kwartaalrekeningen*, 1991-*II*, vol. 7, sdu/uitgeverij/cbs-publikaties, The Hague, The Netherlands.

Chan, N.H., and Wei, C.Z. (1988), "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes," *Annals of Statistics*, 16, 367–401.

Chen, C., and Tiao, G.C. (1990), "Random Level-Shift Time Series Models, ARIMA Approximations, and Level-Shift Detection," *Journal of Business and Economic Statistics*, 8, 83–97.

Cheung, Y.-W., and Lai, K.S. (1993), "A Fractional Cointegration Analysis of Purchasing Power Parity," *Journal of Business and Economic Statistics*, 11, 103–112.

Chirinko, R.S. (1988), "Will "The" Neoclassical Theory of Investment Please Rise?: The General Structure of Investment Models and Their Implications of Tax Policy," in *The Impact of Taxation on Business Investment*, eds. Mintz, J.H., and Purvis, D.D., presented at the meeting of the Association d'Économetrie Appliquée on the Econometrics of Investment in Lille, France, December 1987.

Chong, Y.Y., and Hendry, D.F. (1986), "Econometric Evaluation of Linear Macro-Economic 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.

Christiano, L.J. (1988), "Why does Inventory Investment Fluctuate So Much?," *Journal of Monetary Economics*, 21, 247–280.

Christiano, L.J., and Eichenbaum, M. (1990), "Unit Roots in Real GNP: Do We Know, and Do We Care?," pp. 7–62 in *Unit Roots, Investment Measures and Other Essays* ed. Meltzer, A.H., Carnegie-Rochester Conference Series on Public Policy, Elsevier

Science Publishers (North-Holland), Amsterdam, The Netherlands.
Cleveland, W.S. (1983), "Seasonal and Calendar Adjustment," pp. 39-72 in *Handbook of Statistics*, *Vol.* 3, eds. Brillinger, D.R., and Krishnaiah, P.R., North Holland, Amsterdam, The Netherlands.
Cleveland, W.S., Devlin, S.J., and Perpenning, I.J. (1982), "The SABL Seasonal and Calendar Adjustment Procedures," pp. 539-564 in *Time Series Analysis*: *Theory and Practice* 1, ed. Anderson, O.D., North Holland, Amsterdam, The Netherlands.
Cleveland, W.S., and McGill, R. (1984), "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods," *Journal of the American Statistical Association*, 79, 531-554.
Cochrane, J.H. (1988), "How Big is the Random Walk in GNP?," *Journal of Political Economy*, 96, 893-919.
Cochrane, J.H. (1991), "A critique of the application of unit root tests," *Journal of Economic Dynamics and Control*, 15, 275-284.
Cogley, T. (1990), "International Evidence on the Size of the Random Walk in Output," *Journal of Political Economy*, 98, 501-518.
Cook, R.D. (1977), "Detection of influential observations in linear regression," *Technometrics*, 17, 15-18.
Cook, R.D. (1986), "Assessment of Local Influence," *Journal of the Royal Statistical Society*, *Series B*, 48, 133-155.
Cook, R.D., and Weisberg, S. (1982), *Residuals and Influence in Regression*, Chapman and Hall, New York, NY, USA.
Cook, R.D., and Weisberg, S. (1983), "Diagnostics for Heteroscedasticity in Regression," *Biometrika*, 70, 1-10.
Cooper, R., and Haltiwanger, J. (1990), "Inventories and the Propagation of Sectoral Shocks," *American Economic Review*, 80, 170-190.
Cox, D.R., and Hinkley, D.V. (1974), *Theoretical Statistics*, Chapman and Hall, London, UK.
CPB, Centraal Planbureau (1970), *Centraal Economisch Plan*, Staatsuitgeverij, The Hague, The Netherlands.
Critchley, F. (1985), "Influence in Principal Components Analysis," *Biometrika*, 72, 627-636.
D'Agostino, R.B. (1986), "Tests for the Normal Distribution," pp. 367-419, eds. D'Agostino, R.B., and Stephens, M.A., Marcel Dekker, inc. , New York, USA.
Dahlhaus, R. (1988), "Small sample effects in time series analysis: a new asymptotic theory and a new estimate," *Annals of Statistics*, 16, 808-841.
Davidson, R., and MacKinnon, J.G. (1990), "Specification Tests Based on Artificial Regressions," *Journal of the American Statistical Association*, 85, 220-227.
Davies, R.B., and Harte, D.S. (1987), "Tests for Hurst effect," *Biometrika*, 74, 95-101.
De Boer, S., and Broesterhuizen, G. (1986), The simultaneous compilation of current price and deflated input-output tables, no. NA-013, National accounts, netherlands central bureau of statistics, Voorburg, The Netherlands.
De Leeuw, F. (1990a), "Interpreting Investment-to-Output Ratios: Nominal/Real, Net/Gross, Stock/Flow, Narrow/Broad," pp. 83-120 in *Unit Roots*, *Investment Measures and Other Essays* ed. Meltzer, A.H., Carnegie-Rochester Conference Series on Public Policy, Elsevier Science Publishers (North-Holland), Amsterdam, The Netherlands.
De Leeuw, F. (1990b), "The Reliability of U.S. Gross National Product," *Journal of Business and Economic Statistics*, 8, 191-203.
Dempster, A.P., and Gasko-Green, M. (1981), "New Tools for Residual Analysis," *Annals of Statistics*, 9, 945-959.
Den Butter, F.A.G. (1991), "Macroeconomic modelling and the policy of restraint in the Netherlands," *Economic Modelling*, 8, 16-33.
Den Butter, F.A.G, and Fase, M.M.G. (1988), *Seizoenanalyse en Beleidsdiagnose*, no. 8 of Monetaire Monografieën, De Nederlandsche Bank, Kluwer, Deventer, The Netherlands.
Den Butter, F.A.G., and Mourik, T.J. (1990), "Seasonal Adjustment Using Structural Time Series Models: An application and a comparison with the Census X-11 method," *Journal of Business and Economics Statistics*, 8, 385-394.
Dewald, W., Hurtsby, J., and Anderson, R. (1986), "Replication in Empirical

Economics," *American Economic Review*, 76, 587-603.
Dickey, D.A., Bell, W.R., and Miller, R.B. (1986), "Unit Roots in Time Series Models: Tests and Implications," *The American Statistician*, 40, 12-26.
Dickey, D.A., and Fuller, W.A. (1979), "Distribution of the Estimators for Autoregressive Time Series With a Unit Root," *Journal of the American Statistical Association*, 74, 427-431.
Dickey, D.A., and Fuller, W.A. (1981), "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica*, 49, 1058-1072.
Dickey, D.A., Hasza, H.P., and Fuller, W.A. (1984), "Testing for Unit Roots in Seasonal Time Series," *Journal of the American Statistical Association*, 79, 427-431.
Dickey, D.A., and Pantula, S.G. (1987), "Determining the Order of Differencing in Autoregressive Processes," *Journal of Business and Economic Statistics*, 5, 455-461.
Diebold, F.X., and Nerlove, M. (1990), "Unit Roots in Economic Time Series: A Selective Survey," pp. 3-69 in *Advances in Econometrics, Vol. 8, Co-Integration, Spurious Regressions, and Unit Roots*, eds. Fomby, T.B., and Rhodes, Jr., G.F., JAI Press Inc., Greenwich, Connecticut, USA.
Diebold, F.X., and Rudebusch, G. (1989), "Long Memory and Persistence in Aggregate Output," *Journal of Monetary Economics*, 24, 189-209.
Dietzenbacher, E. (1992), "The measurement of inter industry linkages, Key sectors in the Netherlands," *Economic Modelling*, 9, 419-437.
DNB, De Nederlandsche Bank n.v., *Kwartaalconfrontatie van middelen en bestedingen* 1957-1984, Kluwer, Deventer, The Netherlands.
Doornik, J.A., and Hendry, D.F. (1992), *PcGIVE version 7, An Interactive Econometric Modelling System*, Institute of Economics and Statistics, University of Oxford, UK.
Dufour, J.-M. (1982), "Recursive Stability Analysis of Linear Regression Relationships: An Exploratory Methodology," *Journal of Econometrics*, 19, 31-76.
Edda (1927), Neckel, G. ed., *Edda, die lieder des codex regius nebst verwandten denkmälern*, Carl Winters Universitätsbuchhandlung, Heidelberg, Germany.
Efron, B. (1983), "Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation," *Journal of the American Statistical Association*, 78, 316-331.
Engle, R.F. (1987), On the Theory of Cointegrated Economic Time Series, Paper presented at the European Meeting of the Econometric Society in Copenhagen, August 24-28, 1987.
Engle, R.F., and Granger, C.W.J. (1987), "Co-Integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55, 251-276.
Engle, R.F., Granger, C.W.J., Hylleberg, S., and Lee, H.S. (1991), Seasonal Cointegration: The Japanese Consumption Function, University of California at San Diego.
Engle, R.F., Granger, C.W.J., Hylleberg, S., and Lee, H.S. (1993), "Seasonal cointegration, The Japanese consumption function," *Journal of Econometrics*, 55, 275-298.
Engle, R.F., Hendry, D.F., and Richard, J.F. (1983), "Exogeneity," *Econometrica*, 51, 277-304.
Engle, R.F., Lilien, D.M., and Robins, R.P. (1987), "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model," *Econometrica*, 55, 391-407.
Engle, R.F., and Yoo, B.S. (1987), "Forecasting and Testing in Co-Integrated Systems," *Journal of Econometrics*, 35, 143-159.
Engle, R.F., and Yoo, B.S. (1991), "Cointegrated Economic Time Series: An Overview with New Results," pp. 237-266 in *Long-Run Economic Relationships, Readings in Cointegration*, eds. Engle, R.F., and Granger, C.W.J., Oxford University Press, Oxford, UK.
Evans, G.B.A., and Savin, N.E. (1984), "Testing for Unit Roots: 2," *Econometrica*, 52, 1241-1269.
Farebrother, R.W. (1986), "Discussion of Professor Cook's Paper," *Journal of the Royal Statistical Society, Series B*, 48, 162.
Faust, J. (1992), "When are variance ratio tests for serial dependence optimal?," *Econometrica*, 60, 1215-1226.
Fernandez Macho, F.J., Harvey, A.C., and Stock, J.H. (1987), "Forecasting and Interpolation Using Vector Autoregressions with Common Trends," *Annales de l'INSEE*,

6/7, 279-287.
Filiben, J.J. (1975), "The probability plot correlation coefficient test for normality," *Technometrics*, 17.
Fountis, N.G., and Dickey, D.A. (1989), "Testing for a Unit Root Nonstationarity in Multivariate Autoregressive Time Series," *Annals of Statistics*, 17, 419-428.
Franses, P.H. (1993), "A Multivariate Approach to Modeling Univariate Seasonal Time Series," *Journal of Econometrics*, ??, ??, forthcoming.
Fuller, W.A. (1976), *Introduction to Statistical Time Series*, John Wiley, New York, NY, USA.
Galeotti, M. (1987), Specification of the Technology for Neoclassical Investment Theory: Testing the Adjustment Cost Approach, no. 87-23, C.V. Starr Center for Applied Economics, New York University, USA.
Gérard, M. (1990), Marginal Effective Tax Rate in a Transnational Setting, Faculté Universitaire Catholique de Mons, Mons, Belgium.
Geweke, J.F., and Porter-Hudak, S. (1983), "The Estimation and Application of Long Memory Time Series Models," *Journal of Time Series Analysis*, 4, 221-238.
Ghysels, E. (1988), "A Study Toward a Dynamic Theory of Seasonality for Economic Time Series," *Journal of the American Statistical Association*, 83, 168-172.
Ghysels, E. (1990a), "Unit-Root Tests and the Statistical Pitfalls of Seasonal Adjustment: The Case of U.S. Postwar Real Gross National Product," *Journal of Business and Economic Statistics*, 8, 145-152.
Ghysels, E. (1990b), On the economics and econometrics of seasonality, Centre de recherche et développement en économique and Département de sciences économiques, Université de Montréal, P.O. Box 6128, Station A, Montréal, Canada.
Ghysels, E., and Perron, P. (1993), "The Effect of Seasonal Adjustment Filters on Tests for a Unit Root," *Journal of Econometrics*, 55, 57-98.
Goldfeld, S.M., and Quandt, R.E. (1965), "Some Tests for Homoscedasticity," *Journal of the American Statistical Association*, 60, 539-547.
Gourieroux, C., Maurel, F., and Monfort, A. (1987), Regression and non stationarity, Document de Travail, no. 8708, Institut National de la Statistique et des Études Économiques, Unité de Recherche, Paris, France.
Gourieroux, C., and Peaucelle, I. (1989), Detecting a Long Run Relationship, no. 8902, Centre d'Études Prospectives d'Économie Mathématique Appliquées a la Planification, Paris, France.
Granger, C.W.J. (1966), "The Typical Spectral Shape of an Economic Variable," *Econometrica*, 34, 150-161.
Granger, C.W.J. (1969), "Investigating Causal Relations By Econometric Models and Cross-Spectral Methods," *Econometrica*, 37, 424-438.
Granger, C.W.J. (1981), "Some Properties of Time Series Data and Their Use in Econometric Model Specification," *Journal of Econometrics*, 16, 121-130.
Granger, C.W.J. (1982), "Acronyms in Time Series Analysis (ATSA)," *Journal of Time Series Analysis*, 3, 103-107.
Granger, C.W.J. (1984), "Comment," *Journal of Business and Economic Statistics*, 2, 335-336.
Granger, C.W.J., and Joyeux, R. (1980), "An Introduction to Long-Memory Time Series Models and Fractional Differencing," *Journal of Time Series Analysis*, 1, 15-29.
Granger, C.W.J., and Lee, T.-H. (1990), "Multicointegration," pp. 71-84 in *Advances in Econometrics, Vol. 8, Co-Integration, Spurious Regressions, and Unit Roots*, eds. Fomby, T.B., and Rhodes, Jr., G.F., JAI Press Inc., Greenwich, Connecticut, USA.
Granger, C.W.J., and Newbold, P. (1976), "Forecasting Transformed Series," *Journal of the Royal Statistical Society, Series B*, 38, 189-203.
Granger, C.W.J., and Newbold, P. (1986), *Forecasting Economic Time Series*, Academic Press, Harcourt Brace Jovanovich, New York, NY, USA, second edition.
Gray, H.L., and Woodward, W.A. (1986), "A New ARMA Spectral Estimator," *Journal of the American Statistical Association*, 81, 1100-1108.
Gray, H.L., Zhang, N.-F., and Woodward, W.A. (1989), "On Generalized Fractional Processes," *Journal of Time Series Analysis*, 10, 233-257.
Gregory, A.W. (1991), Testing for Cointegration in Linear Quadratic Models, Discussion Paper, no. 811, Department of Economics, Queen's University, Kingston, Canada.

Grether, D.M., and Nerlove, M. (1970), "Some Properties of Optimal Seasonal Adjustment," *Econometrica*, 38, 682-703.

Griffiths, W., and Lütkepohl, H. (1990), Confidence intervals for impulse responses from VAR models: a comparison of asymptotic theory and simulation approaches, Econometrics department, University of New England, Armidale, Australia.

Griliches, Z. (1985), "Data and Econometricians- The Uneasy Alliance," *American Economic Association Papers and Proceedings*, 75, 196-200.

Griliches, Z. (1986), "Economic Data Issues," pp. 1465-1514 in *Handbook of Econometrics, Volume III*, eds. Griliches, Z., and Intriligator, M.D., Elsevier Science Publishers BV, Amsterdam, The Netherlands.

Haldrup, N., and Hylleberg, S. (1991), Integration, Near-Integration and Deterministic Trends, no. 1991-15, Økonomisk institut, Århus universitet, Denmark.

Hall, A. (1992), "Testing for a unit root in time series using instrumental variable estimators with pretest data based model selection," *Journal of Econometrics*, 54, 223-250.

Hallin, M. (1984), "Spectral Factorization of Nonstationary Moving Average Processes," *Annals of Statistics*, 12, 172-192.

Hannan, E.J., and Deistler, M., *The Statistical Theory of Linear Systems*, John Wiley & Sons, New York, USA.

Hannan, E.J., and McDougall, A.J. (1988), "Regression Procedures for ARMA Estimation," *Journal of the American Statistical Association*, 83, 490-498.

Hannan, E.J., and Poskitt, D.S. (1988), "Unit Canonical Correlation Between Future and Past," *Annals of Statistics*, 16, 784-790.

Hansen, B.E. (1992a), "Testing for Parameter Instability in Linear Models," *Journal of Policy Modeling*, 14, 517-533.

Hansen, B.E. (1992b), "Tests for Parameter Instability in Regressions with I(1) Processes," *Journal of Business and Economic Statistics*, 10, 321-335.

Harris, R.I.D. (1992), "Small Sample Testing For Unit Roots," *Oxford Bulletin of Economics and Statistics*, 54, 615-625.

Harvey, A.C. (1975), "Discussion of the Paper by Brown, Durbin and Evans," *Journal of the Royal Statistical Society, Series B*, 37, 179-180.

Harvey, A.C. (1985), "Trends and Cycles in Macroeconomic Time Series," *Journal of Business and Economic Statistics*, 3, 216-227.

Harvey, A.C. (1989), *Forecasting, structural time series models and the Kalman Filter*, Cambridge University Press, Cambridge, UK.

Harvey, A.C., Ruiz, E., and Shephard, N. (1992), Multivariate Stochastic Variance Models, Department of Statistical and Mathematical Sciences, London School of Economics.

Hassler, U. (1993a), "Unit root tests: the autoregressive approach in comparison with the periodogram regression," *Statistical Papers*, 34, 67-82.

Hassler, U. (1993b), "Regression of Spectral Estimators with Fractionally Integrated Time Series," *Journal of Time Series Analysis*, 14, 369-380 and 549.

Hasza, D.P., and Fuller, W.A. (1979), "Estimation for Autoregressive Processes with Unit Roots," *Annals of Statistics*, 7, 1106-1120.

Haug, A.A. (1992), "Critical values for the $\hat{Z}_\alpha$-Phillips-Ouliaris Test for Cointegration," *Oxford Bulletin of Economics and Statistics*, 54, 473-480.

Hausman, J.A., and Watson, M.W. (1985), "Errors in Variables and Seasonal Adjustment Procedures," *Journal of the American Statistical Association*, 80, 531-540.

Hendry, D.F. (1974), "Stochastic Specification in an Aggregate Demand Model of the United Kingdom," *Econometrica*, 42, 559-578, also in Hendry (1993, pp. 32-51).

Hendry, D.F. (1993), *Econometrics, alchemy or science*, Blackwell, Oxford, UK.

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.

Hillier, G.H. (1991), "On multiple diagnostic procedures for the linear model," *Journal of Econometrics*, 47, 47-66.

Hillmer, S.C., and Tiao, G.C. (1982), "An ARIMA model based approach to seasonal adjustment," *Journal of the American Statistical Association*, 77, 63-70.

Hinkley, D.V. (1977), "Jackknifing in Unbalanced Situations," *Technometrics*, 19, 285-292.

Hipel, K.W., and McLeod, A.I. (1978a), "Preservation of the Rescaled Adjusted Range. 2. Simulation Studies Using Box-Jenkins Models," *Water Resources Research*, 14, 509-516.

Hipel, K.W., and McLeod, A.I. (1978b), "Preservation of the Rescaled Adjusted Range. 3. Fractional Gaussian Noise Algorithms," *Water Resources Research*, 14, 517-518.

Hogg, R.V. (1974), "Adaptive Robust Procedures: A Partial Review and Some Suggestions for Future Applications and Theory," *Journal of the American Statistical Association*, 69, 909-923.

Hooper, J.W. (1959), "Simultaneous Equations and Canonical Correlation Theory," *Econometrica*, 27, 245-256.

Hooper, J.W. (1962), "Partial Trace Correlations," *Econometrica*, 30, 324-331.

Hosking, J.R.M. (1981), "Fractional Differencing," *Biometrika*, 68, 165-176.

Huber, P.J. (1964), "Robust Estimation of a Location Parameter," *Annals of Mathematical Statistics*, 34, 598-611.

Hylleberg, S. (1986), *Seasonality in Regression*, Academic Press, Harcourt Brace Jovanovich, New York, NY, USA.

Hylleberg, S., ed. (1992), *Modelling Seasonality*, Oxford University Press, Oxford, UK.

Hylleberg, S., Engle, R.F., Granger, C.W.J., and Yoo, B.S. (1990), "Seasonal Integration and Cointegration," *Journal of Econometrics*, 44, 215-238.

Hylleberg, S., and Mizon, G.E. (1989), "A Note on the Distribution of the Least Squares Estimator of a Random Walk With Drift," *Economics Letters*, 29, 225-230.

Hylleberg, S., and Mizon, G.E. (1989), "Cointegration and Error Correction Mechanisms," *Economic Journal*, 99, 113-125.

Jäger, A., and Kunst, R.M. (1990), "Seasonal Adjustment and Measuring Persistence in Output," *Journal of Applied Econometrics*, 5, 47-58.

JBES, Journal of Business and Economic Statistics (1992), *Breakpoints and Unit Roots*, vol. 10, number 3.

Jewell, N.P., and Bloomfield, P. (1983), "Canonical correlations of past and future for time series: definitions and theory," *Annals of Statistics*, 11, 837-847.

JME, Journal of Monetary Economics (1988), *Proceedings of the Conference on Real Business Cycles*, vol. 21, North-Holland, Amsterdam, The Netherlands.

JoE, Journal of Econometrics, Hillier, G.H, and King, M.L., eds. (1991), *Diagnostic Testing*, vol. 47, pp. 1-194, North-Holland, Amsterdam, The Netherlands.

JoE, Journal of Econometrics, Ghysels, E., ed. (1993), *Seasonality and Econometric Models*, vol. 55, pp. 1-357, North-Holland, Amsterdam, The Netherlands.

Johansen, S. (1988), "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231-254.

Johansen, S. (1989), Likelihood Based Inference on Cointegration, Lecture Notes, Centro Interuniversitario di Econometria, Bagni di Lucca, Italy, c/o Dipartimento di Scienze Economiche, 40125 Bologna, Italy.

Johansen, S. (1991a), "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59, 1551-1580.

Johansen, S. (1991b), The role of the constant term in cointegration analysis of I(1) variables, to appear in Econometric Reviews, Institute of Mathematical Statistics, University of Copenhagen, Denmark.

Johansen, S. (1992a), "A representation of vector autoregressive processes integrated of order 2," *Econometric Theory*, 8, 188-202.

Johansen, S. (1992b), "Cointegration in Partial Systems and the Efficiency of Single-equation Analysis," *Journal of Econometrics*, 53, 389-402.

Johansen, S. (1992c), "Determination of Cointegrating Rank in the Presence of a Linear Trend," *Oxford Bulletin of Economics and Statistics*, 54, 383-397.

Johansen, S. (1992d), Estimating Systems of Trending Variables, Institute of Mathemathical Statistics, University of Copenhagen, Denmark, to appear in Econometric Reviews.

Johansen, S., and Juselius, K. (1990), "Maximum Likelihood Estimation and Inference on Cointegration - With Applications to the Demand for Money," *Oxford Bulletin of Economics and Statistics*, 52, 169-210.

Johnson, W., and Geisser, S. (1983), "A Predictive View of the Detection and Characterization of Influential Observations in Regression Analysis," *Journal of the American Statistical Association*, 78, 137-143.

Johnson, W., and Geisser, S. (1985), "Estimative Influence Measures for the Multivariate General Linear Model," *Journal of Statistical Planning and Inference*, 11, 33-56.

Jonas, A.J. (1983), *Persistent Memory Random Processes*, Department of Statistics, Harvard University, Cambridge, MA, USA.

JoPM, Journal of Policy Modeling, Ericsson, N.E. ed. (1992), *Cointegration, Exogeneity, and Policy Analysis, Part II. Inflation and Parameter Constancy*, vol. 14, pp. 395-560, Elsevier Science Publishing Co., New York, NY, USA..

Jorgenson, D.W. (1967), "Seasonal adjustment of data for econometric analysis," *Journal of the American Statistical Association*, 59, 681-724.

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

Judge, G.G., Griffiths, W.E., Hill, R.C., Lütkepohl, H., and Lee, T.-C. (1988), *Introduction to the Theory and Practice of Econometrics*, John Wiley & Sons, New York, USA.

Kaen, F.R., and Rosenman, R.E. (1986), "Predictable Behavior in Financial Markets: Some Evidence in Support of Heiner's Hypothesis," *American Economic Review*, 76, 212-220.

Kailath, T. (1980), *Linear Systems*, Prentice Hall, London, UK.

Kang, H. (1985), "The Effects of Detrending on Granger Causality Tests," *Journal of Business and Economic Statistics*, 3, 344-349.

Kavalieris, L. (1989), "The Estimation of the Order of an Autoregression Using Recursive Residuals and Cross-Validation," *Journal of Time Series Analysis*, 10, 271-281.

Kendall, M.G. (1975), *Multivariate Analysis*, Charles Griffin & Co. Ltd., High Wycombe, UK.

King, R., Plosser, C., Stock, J., and Watson, M. (1987), Stochastic Trends and Economic Fluctuations, University of Rochester, USA.

King, R., Plosser, C., Stock, J., and Watson, M. (1991), "Stochastic Trends and Economic Fluctuations," *American Economic Review*, 81, 819-840.

Kitagawa, G. (1987), "Non-Gaussian State-Space Modeling of Nonstationary Time Series," *Journal of the American Statistical Association*, 82, 1032-1041.

Kitagawa, G., and Gersch, W. (1984), "A Smoothness Priors-State Space Modeling of Time Series With Trend and Seasonality," *Journal of the American Statistical Association*, 79, 378-389.

Kiviet, J.F. (1986), "On the Rigour of Some Misspecification Tests for Modelling Dynamic Relationships," *Review of Economic Studies*, 53, 241-261.

Kiviet, J.F. (1987), *Testing Linear Econometric Models*, Faculty of Actuarial Science & Econometrics, University of Amsterdam, The Netherlands.

Kiviet, J.F., and Phillips, G.D.A. (1992), "Exact Similar Tests for Unit Roots and Cointegration," *Oxford Bulletin of Economics and Statistics*, 54, 349-367.

Kleiner, B., Martin, R.D., and Thomson, D.J. (1979), "Robust Estimation of Power Spectra," *Journal of the Royal Statistical Society, Series B*, 41, 313-351.

Knuth, D.E. (1969), *The Art of Computer Programming, Vol. 2, Seminumerical Algorithms*, Addison-Wesley Publishing Company.

Koedijk, K., and Schotman, P. (1989), "Dominant Real Exchange Rate Movements," *Journal of International Money and Finance*, 8, 517-531.

Kool, H. (1989), *Multi-Step Prediction Models for Stationary Multivariate Time Series*, Free University Press, Amsterdam, The Netherlands.

Koopmans, L.H. (1974), *The Spectral Analysis of Time Series*, Academic Press, Harcourt Brace Jovanovich, New York, NY, USA.

Koreisha, S.G., and Pukkila, T.M. (1987), "Identification of Nonzero Elements in the Polynomial Matrices of Mixed VARMA Processes," *Journal of the Royal Statistical Society, Series B*, 49, 112-126.

Koschat, M.A., and Weerahandi, S. (1992), "Chow-type Tests Under Heteroscedasticity," *Journal of Business and Economic Statistics*, 10, 221-228.

Krämer, W., Ploberger, W., and Alt, R. (1988), "Testing for Structural Change in Dynamic Models," *Econometrica*, 56, 1355-1369.

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

Krasker, W.S., Kuh, E., and Welsch, R.E. (1983), "Estimation for Dirty Data and Flawed Models," pp. 651-698 in *Handbook of Econometrics, Volume I*, eds. Griliches, Z., and Intriligator, M.D., North Holland, Amsterdam, The Netherlands.

Kugler, P., and Schwendener, P. (1990), Codependence in a VAR Framework, Volkswirtschaftliches Institut Universität Bern, Switzerland.

Kuiper, J. (1978), "A survey and comparative analysis of various methods of seasonal adjustment," pp. 55-76 in *Seasonal Analysis of Economic Time Series, Proceedings of the Conference on the Seasonal Analysis of Economic Time Series, Washington D.C. September* 9-10, 1976, ed. Zellner, A., U.S. Department of Commerce, Bureau of the Census, Washington D.C., USA.

Kunst, R. (1986), Multivariate Robust Filtering, no. 226, Institute for Advanced Studies, Vienna, Austria.

Kunst, R.M. (1989), Cointegration in Macroeconomic Systems: Seasonality and Explosive Roots, RM no. 255, Institute for Advanced Studies, Vienna, Austria.

Kunst, R.M. (1990), Cointegration in Macroeconomic Systems: Case Studies for Small and Large European Countries, Institute for Advanced Studies, Vienna, Austria, revised in Kunst (1993), *Review of Economics and Statistics*, 75, 325-330.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., and Shin, Y. (1992), "Testing the null hypothesis of stationarity against the alternative of a unit root; How sure are we that economic time series have a unit root?," *Journal of Econometrics*, 54, 159-178.

Lambert, J.-P. (1990), "The French Unemployment Problem, Lessons from a Rationing Model Relying on Business Survey Information," *European Economic Review*, 34, 423-433.

Laroque, G. (1977), "Analyse d'une Méthode de désaisonnalisation: le programme X 11, du US Bureau of Census, version trimestrielle," *Annales de l'INSEE*, 28, 105-127.

Lawrance, A.J., and Kottegada, N.T. (1977), "Stochastic Modelling of Riverflow Time Series," *Journal of the Royal Statistical Society, Series A*, 140, 1-31.

Leadbetter, M.R., Lindgren, G., and Rootzén, H. (1983), *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag, New York, etc.

Leamer, E.E. (1978), *Specification Searches*, John Wiley, New York, USA.

Ledolter, J. (1989), "The Effect of Additive Outliers on the Forecasts from ARIMA models," *International Journal of Forecasting*, 5, 241-247.

Lee, H.S. (1992), "Maximum likelihood inference on cointegration and seasonal cointegration," *Journal of Econometrics*, 54, 1-47.

Lee, J.H.H., and King, M.L. (1993), "A Locally Most Mean Powerful Based Score Test For ARCH and GARCH Regression Disturbances," *Journal of Business and Economic Statistics*, 11, 17-27.

LeSage, J.P. (1990), "A Comparison of the Forecasting Ability of ECM and VAR models," *Review of Economics and Statistics*, 72, 664-671.

Lichtenberg, F.R., and Griliches, Z. (1989), "Errors of Measurement in Output Deflators," *Journal of Business and Economic Statistics*, 7, 1-9.

Lin, C.J., and Teräsvirta, T. (1991), Testing the Constancy of Regression Parameters against Continuous Structural Change, Department of Economics, University of California, San Diego , CA. USA.

Lindley, D.V. (1979), "Comment," *Journal of the American Statistical Association*, 74, 127.

Lo, A.W. (1991), "Long-Term Memory in Stock Market Prices," *Econometrica*, 59, 1279-1313.

Lomnicki, Z.A. (1961), "Tests for departure from normality in the case of linear stochastic processes," *Metrika*, 4, 27-62.

Long, J.B., and Plosser, C.I. (1983), "Real business cycles," *Journal of Political Economy*, 91, 36-69.

Lucas, R.E. (1976), "Econometric Policy Evaluation: A Critique," pp. 19-46 in *The Phillips Curve and Labor Markets*, eds. Brunner, K., and Meltzer, A.H., North-Holland, Amsterdam, The Netherlands.

Lund, R.E. (1975), "Tables for an approximate test for outliers in linear regressions," *Technometrics*, 17, 473-476.

Lütkepohl, H. (1988), "Prediction Tests for Structural Stability," *Journal of Econometrics*, 39, 267-296.

Lütkepohl, H. (1989), "Prediction Tests for Structural Stability of Multiple Time Series," *Journal of Business and Economic Statistics*, 7, 129-135.

Lütkepohl, H. (1991), *Introduction to multiple time series analysis*, Springer Verlag, Berlin, Germany.

MacKinnon, J.G. (1992), "Model Specification Tests and Artificial Regressions," *Journal of Economic Literature*, 30, 102-146.

MacKinnon, J.G., and White, H. (1985), "Some Heteroskedastic Consistent Covariance Estimators with Improved Finite Sample Properties," *Journal of Econometrics*, 29, 305-325.

Madansky, A. (1964), "Spurious Correlation due to Deflating Variables," *Econometrica*, 32, 652-655.

Magnus, J.R., and Neudecker, H. (1988), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley & Sons, Chichester, UK.

Malinvaud, E. (1987), "Capital productif, incertitudes et profitabilité," *Annales d'Économie et de Statistique*, 5, 1-34.

Mandelbrot, B.B. (1969), "Long-run Linearity, Locally Gaussian Processes, H-Spectra and Infinite Variances," *International Economic Review*, 10, 82-111.

Mandelbrot, B.B. (1972), "Statistical Methodology for Nonperiodic Cycles: From the Covariance to R/S Analysis," *Annals of Economic and Social Measurement*, 1/3, 259-290.

Mandelbrot, B.B. (1977), "Discussion on the Paper by Drs Lawrance and Kottegada," *Journal of the Royal Statistical Society, Series A*, 140, 39-40.

Mankiw, N.G. (1990), "A Quick Refresher Course in Macroeconomics," *Journal of Economic Literature*, 28, 1645-1660.

Maravall, A. (1984), "Comment," *Journal of Business and Economic Statistics*, 2, 337-339.

Maravall, A., and Mathis, A. (1991), Encompassing Univariate Models in Multivariate Time Series: A Case Study, Department of Economics, European University Institute, San Domenico di Fiesole, Italy.

Mardia, K.V. (1970), "Measures of Multivariate Skewness and Kurtosis with Applications," *Biometrika*, 57, 519-530.

Mariano, R.S. (1982), "Analytical Small-Sample Distribution Theory in Econometrics: The Simultaneous-Equations Case," *International Economic Review*, 23, 503-533.

Martin, R.D., and Yohai, V.R. (1986), "Influence Functionals for Time Series," *Annals of Statistics*, 14, 781-855.

Martin, R.D., and Zeh, J.E. (1977), "Determining the Character of Time Series Outliers," pp. 818-823 in *Proceedings of the Business and Economic Statistics Section of the American Statistical Association*, American Statistical Association, 806 15th street, N.W. Washington D.C. 20005, USA.

McLeod, A.I., and Hipel, K.W. (1978), "Preservation of the Rescaled Adjusted Range. 1. A Reassessment of the Hurst Phenomenon," *Water Resources Research*, 14, 491-508.

Mellander, E., Vredin, A., and Warne, A. (1992), "Stochastic Trends and Economic Fluctuations in a Small Open Economy," *Journal of Applied Econometrics*, 7, 369-394.

Milhøj, A. (1981), "A test of fit in time series models," *Biometrika*, 68, 177-187.

Mizon, G. (1977), "Inferential Procedures in Nonlinear Models: An Application in a UK Industrial Cross Section Study of Factor Substitution and Returns to Scale," *Econometrica*, 45, 1221-1242.

Monfort, A., and Rabemananjara, R. (1990), "From a VAR Model to a Structural Model, With an Application to the Wage-Price Spiral," *Journal of Applied Econometrics*, 5, 203-227.

Muirhead, C.R. (1986), "Distinguishing Outlier Types in Time Series," *Journal of the Royal Statistical Society, Series B*, 48, 39-47.

Nabeya, S., and Tanaka, K. (1990), "A General Approach to the Limiting Distribution for Estimators in Time Series Regression with Nonstable Autoregressive Errors,"

Econometrica, 58, 145–163.

NAG (1988), *The NAG Fortran Library Manual - Mark* 13, The Numerical Algorithms Group Limited, Oxford, UK.

Nankervis, J.C., and Savin, N.E. (1985), "Testing the Autoregressive Parameter with the t-Statistic," *Journal of Econometrics*, 27, 143–161.

Nelson, C.R., and Kang, H. (1981), "Spurious Periodicity in Inappropriately Detrended Time Series," *Econometrica*, 49, 741–751.

Nelson, C.R., and Plosser, C.I. (1982), "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications," *Journal of Monetary Economics*, 10, 139–162.

Nerlove, M., Grether, D.M., and Carvalho, J.L. (1979), *Analysis of Economic Time Series; A Synthesis*, Academic Press, New York, NY, USA.

Newey, W.K., and West, K.D. (1987), "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Nickell, S.J. (1978), *The Investment Decision of Firms*, James Nisbet & Co. Ltd, Digswell Place Welwyn/ Cambridge University Press, UK.

Nickell, S. (1984), "An Investigation of the Determinants of Manufacturing Employment in the United Kingdom," *Review of Economic Studies*, 51, 529–557.

Nordhaus, W.D. (1992), "Discussion," *American Economic Association Papers and Proceedings*, 74, 419–421.

OECD, Organization for Economic Cooperation and Development (1971), *Economic Surveys* 1969-1970, *Netherlands*, Author.

OECD, Organization for Economic Cooperation and Development (1984), *Economic Surveys* 1983-1984, *France*, Author.

OECD, Organization for Economic Cooperation and Development (1991), *Main Economic Indicators, Historical Supplement*, 26, Author.

Ogaki, M. (1992), "Engel's Law and Cointegration," *Journal of Political Economy*, 100, 1027–1046.

Ohanian, L.E. (1988), "The Spurious Effect of Unit Roots on Vector Autoregressions, A Monte Carlo Study," *Journal of Econometrics*, 39, 251–266.

Ohanian, L.E. (1991), "A Note on Spurious Inference in A linearly Detrended Vector Autoregression," *Review of Economics and Statistics*, 73, 568–571.

Ooms, M. (1988), Decomposing Multiple Economic Time Series, an integrating view, no. 8834/a, Econometric Institute, Erasmus University Rotterdam, The Netherlands.

Ooms, M. (1990), Outliers in Macroeconomic VAR Systems, Simultaneous Testing, and Cointegration Analysis, Econometric Institute, Erasmus University Rotterdam, presented at the 6th World Congress of the Econometric Society, Barcelona, 1990.

Ooms, M., and Van Dijk, H.K. (1992), Estimating Pushing Trends and Pulling Equilibria, no. 9271/a, Econometric Institute, Erasmus University Rotterdam, The Netherlands, forthcoming in Econometric Reviews.

Osborn, D.R. (1991), "The implications of periodically varying coefficients for seasonal time-series processes," *Journal of Econometrics*, 48, 373–384.

Osborn, D.R., and Smith, J.P. (1989), "The Performance of Periodic Autoregressive Models in Forecasting Seasonal UK Consumption," *Journal of Business and Economic Statistics*, 7, 117–127.

Osterwald-Lenum, M. (1992), "A Note with Fractiles of the Asymptotic Distribution of the Likelihood Cointegration Rank Test Statistics: Four Cases," *Oxford Bulletin of Economics and Statistics*, 54, 461–471.

Ouliaris, S. (1991), COINT, GAUSS Procedures for Co-integrated Regression Models, version 1.0, Author, #-03-170 Bukit Merah Lane 1, Republic of Singapore.

Pagan, A.R., and Wickens, M.R. (1989), "Econometrics: A Survey," *Economic Journal*, 1–113.

Palm, F. (1977), "On Univariate Time Series Methods and Simultaneous Equation Econometric Models," *Journal of Econometrics*, 5, 379–388.

Palm, F.C., Peeters, H.M.M., and Pfann, G.A. (1992), Adjustment Costs and Time-to-Build in Factor Demand in the U.S. Manufacturing Industry, RM 92-012, Faculty of Economics, Limburg University, Maastricht, The Netherlands.

Pantula, S.G. (1991), "Asymptotic Distributions of Unit-Root Tests When the Process is

Nearly Stationary," *Journal of Business and Economic Statistics*, 9, 63-96.

Park, J.Y. (1990), "Testing for Unit Roots and Cointegration by Variable Addition," pp. 107-134 in *Advances in Econometrics, Vol. 8, Co-Integration, Spurious Regressions, and Unit Roots*, eds. Fomby, T.B., and Rhodes, Jr., G.F., JAI Press Inc., Greenwich, Connecticut, USA.

Park, J.Y., and Phillips, P.C.B. (1988), "Statistical Inference in Regressions with Integrated Processes: Part 1," *Econometric Theory*, 4, 468-497.

Parzen, E. (1979), "Nonparametric Statistical Data Modeling," *Journal of the American Statistical Association*, 74, 105-121.

Parzen, E. (1983), "Autoregressive Spectral Estimation," pp. 221-247 in *Handbook of Statistics, Vol. 3*, eds. Brillinger, D.R., and Krishnaiah, P.R., North Holland, Amsterdam, The Netherlands.

Peña, D. (1986), "Discussion of Professor Cook's Paper," *Journal of the Royal Statistical Society, Series B*, 48, 164-165.

Peña, D. (1990), "Influential Observations in Time Series," *Journal of Business and Economic Statistics*, 8, 235-241.

Penm, J.H.W., and Terrell, R.D. (1984), "Multivariate Subset Autoregressive Modelling with Zero Constraints for Detecting Overall Causality," *Journal of Econometrics*, 24, 311-330.

Perron, P. (1989), "The Great Crash, The Oil Price Shock and The Unit Root Hypothesis," *Econometrica*, 57, 1361-1401.

Perron, P. (1990a), "Testing for a Unit Root in a Time Series With a Changing Mean," *Journal of Business and Economic Statistics*, 8, 153-162.

Perron, P. (1990b), "Tests of Joint Hypotheses for Time Series Regression with a Unit Root," pp. 135-160 in *Advances in Econometrics, Vol. 8, Co-Integration, Spurious Regressions, and Unit Roots*, eds. Fomby, T.B., and Rhodes, Jr., G.F., JAI Press Inc., Greenwich, Connecticut, USA.

Perron, P. (1991), A Test for Changes in a Polynomial Trend Function for a Dynamic Time Series, Econometric Research Program, Princeton University, Princeton, USA.

Perron, P., and Campbell, J.Y. (1992), "Racines unitaires en macroéconomie: Le cas multidimensionnel," *Annales d'Économie et de Statistique*, 27, 1-50.

Perron, P., and Vogelsang, T.J. (1991), Nonstationarity and Level Shifts with an Application to Purchasing Power Parity, Research Memorandum, no. 359, Econometric Research Program, Princeton University, Princeton, NJ, USA.

Perron, P., and Vogelsang, T.J. (1992a), "Nonstationarity and Level Shifts With an Application to Purchasing Power Parity," *Journal of Business and Economic Statistics*, 10, 301-320.

Perron, P., and Vogelsang, T.J. (1992b), "Testing for a Unit Root in a Time Series With a Changing Mean: Corrections and Extensions," *Journal of Business and Economic Statistics*, 10, 467-470.

Pesaran, M.H., Smith, R.P., and Yeo, J.S. (1985), "Testing for Structural Stability and Predictive Failure," *Manchester School*, 53, 280-295.

Pfaffenberger, R.C., and Dielman, T.E. (1991), "Testing normality of regression disturbances, A Monte Carlo Study of the Filiben Test," *Computational Statistics and Data Analysis*, 11, 265-273.

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

Phillips, P.C.B. (1991), "Optimal Inference in Cointegrated Systems," *Econometrica*, 59, 283-306.

Phillips, P.C.B., and Durlauf, S.N. (1986), "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies*, 53, 473-495.

Phillips, P.C.B., and Loretan, M. (1991a), "Estimating Long-run Economic Equilibria," *Review of Economic Studies*, 58, 407-436.

Phillips, P.C.B., and Loretan, M. (1991b), "The Durbin-Watson Ratio under Infinite-Variance Errors," *Journal of Econometrics*, 47, 85-114.

Phillips, P.C.B., and Ouliaris, S. (1988), "Testing for Cointegration Using Principal Component Methods," *Journal of Economic Dynamics and Control*, 205-230.

Phillips, P.C.B., and Ouliaris, S. (1990), "Asymptotic Properties of Residual Based Tests for Cointegration," *Econometrica*, 58, 165-193.

Phillips, P.C.B., and Perron, P. (1988), "Testing for a Unit Root in Time Series Regression," *Biometrika*, 55, 335-346.

Pierce, D.A. (1977), "Relationships- and the Lack Thereof- Between Economic Time Series, with Special Reference to Money and Interest Rates," *Journal of the American Statistical Association*, 73, 11-22, 24-26.

Pierce, D.A. (1981), "Sources of Error in Economic Time Series," *Journal of Econometrics*, 17, 305-321.

Pierce, D.A., and Kopecky, K.J. (1979), "Testing goodness of fit for the distribution of errors in regression models," *Biometrika*, 66, 1-5.

Ploberger, W., and Krämer, W. (1992), "The CUSUM test with OLS residuals," *Econometrica*, 60, 271-285.

Ploberger, W., Krämer, W., and Kontrus, K. (1989), "A New Test for Structural Stability in the Linear Regression Model," *Journal of Econometrics*, 40, 307-318.

Plosser, C.I. (1991), "Comments," *European Economic Review*, 35, 961-966.

Polasek, W. (1984), "Regression Diagnostics for General Linear Regression Models," *Journal of the American Statistical Association*, 79, 336-340.

Polasek, W. (1986), "Local Sensitivity Analysis and Bayesian Regression Diagnostics," pp. 375-387 in *Bayesian Inference and Decision Techniques*, eds. Goel, P., and Zellner, A., Elseviers Science, Amsterdam, The Netherlands.

Porter-Hudak, S. (1990), "An Application of the Seasonal Fractionally Differenced Model to the Monetary Aggregates," *Journal of the American Statistical Association*, 85, 338-344.

Poskitt, D.S., and Tremayne, A.R. (1981), "An Approach to Testing Linear Time Series Models," *Annals of Statistics*, 9, 974-986.

Poskitt, D.S., and Tremayne, A.R. (1982), "Diagnostic Tests for Multiple Time Series Models," *Annals of Statistics*, 10, 590-604.

Press, W.H., Flannery B.P., Teukolski, S.A., and Vetterling, W.T. (1986), *Numerical Recipes: the Art of Scientific Computing*, 381-449, Cambridge University Press, Cambridge, New York.

Priestley, M.B. (1988), "Current Developments in Time Series Modelling," *Journal of Econometrics*, 37, 67-86.

Priestley, M.B., and Rao, T.S. (1969), "A Test for Non-stationarity of Time-Series," *Journal of the Royal Statistical Society, Series B*, 31, 140-149.

Quah, D. (1990), "Permanent and Transitory Movements in Labor Income: An explanation for "Excess Smoothness" in Consumption," *Journal of Political Economy*, 98, 449-475.

Rao, C.D., and Mitra, S.K. (1971), *Generalized Inverse of Matrices and its Applications*, Wiley & Sons, New York, NY, USA.

Reimers, H.-E. (1991a), *Analyse kointegrierter Variablen mittels vektorautoregressiver Modelle*, Physica-Verlag, Berlin 31, Germany.

Reimers, H.-E. (1991b), Comparisons of Tests for Multivariate Cointegration, Deutsche Bundesbank, Frankfurt am Main, Germany, forthcoming in Statistical Papers.

Reinsel, G.C., and Ahn, S.K. (1992), "Vector Autoregressive Models with Unit Roots and Reduced Rank Structure: Estimation, Likelihood Ration Test, and Forecasting," *Journal of Time Series Analysis*, 13, 353-375.

Robinson, P.M. (1991a), "Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression," *Journal of Econometrics*, 47, 67-84.

Robinson, P.M. (1991b), Time Series with Strong Dependence, London School of Economics.

Rosenblatt, H.M. (1968), "Spectral Evaluation of BLS and Census Revised Seasonal Adjustment Procedures," *Journal of the American Statistical Association*, 63, 472-501.

Rudin, W. (1964), *Principles of Mathematical Analysis, Second Edition*, McGraw-Hill Book Company, Kogakusha Company, Ltd., New York, USA, Tokyo, Japan.

Runkle, D.E. (1987), "Vector Autoregressions and Reality," *Journal of Business and Economic Statistics*, 5, 437-442.

Said, S.E., and Dickey, D.A. (1984), "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," *Biometrika*, 71, 599-607.

Said, S.E., and Dickey, D.A. (1985), "Hypothesis Testing in ARIMA(p,1,q) models,"

Journal of the American Statistical Association, 80, 369-374.
Saikkonen, P. (1991), "Asymptotically Efficient Estimation of Cointegration Regressions," *Econometric Theory*, 7, 1-21.
Saikkonen, P., and Luukkonen, R. (1990), Testing for a Moving Average Unit Root, Department of Statistics, University of Helsinki, Helsinki, Finland.
Salmon, M. (1988), "Error Correction Models, Co-integration and the Internal Model Principle," *Journal of Economic Dynamics and Control*, 12, 523-529.
Sargent, T.J. (1979), *Macroeconomic Theory*, 214-297, Academic Press, New York, USA.
Schmidt, P. (1990), "Dickey-Fuller Tests With Drift," pp. 161-200 in *Advances in Econometrics, Vol. 8, Co-Integration, Spurious Regressions, and Unit Roots*, eds. Fomby, T.B., and Rhodes, Jr., G.F., JAI Press Inc., Greenwich, Connecticut, USA.
Schmidt, P., and Phillips, P.C.B. (1992), "LM tests for a Unit Root in the Presences of Deterministic Trends," *Oxford Bulletin of Economics and Statistics*, 54, 257-287.
Schweder, T. (1976), "Some "Optimal" Methods to Detect Structural Shifts or Outliers in Regression," *Journal of the American Statistical Association*, 71, 491-354.
Schwert, G.W. (1987), "Effects of Model Specification on Tests for Unit Roots in Macroeconomic Data," *Journal of Monetary Economics*, 20, 73-103.
Schwert, G.W. (1989), "Tests for Unit Roots: A Monte Carlo Investigation," *Journal of Business and Economic Statistics*, 7, 147-159.
Sensenbrenner, G. (1991), "Aggregate investment, the stock market, and the Q-model," *European Economic Review*, 35, 769-832.
Shea, G.S. (1989), Uncertainty and implied variance bounds in long-memory models of the interest rate term structure, Pennsylvania State University, Pennsylvania.
Shiskin, J., Young, A.H., and Musgrave, J.C. (1967), The X-11 variant of the Census method II seasonal adjustment program, Technical Paper no. 15, Bureau of the Census, U.S. Department of Commerce, Washington D.C., USA.
Sims, C.A. (1974), "Seasonality in Regression," *Journal of the American Statistical Association*, 69, 618-626.
Sims, C.A. (1977), "Comment," *Journal of the American Statistical Association*, 72, 23-24.
Sims, C.A. (1980), "Macroeconomics and Reality," *Econometrica*, 48, 1-48.
Sims, C.A. (1985), A rational expectations framework for policy analysis, no. 223, Minnesota Center for Economic Research Discussion Paper, Minneapolis, USA, also in the proceedings of the Second Texas Symposium on Economics, Austin, Texas, 1985.
Sims, C.A. (1986), "Are Forecasting Models Usable for Policy Analysis?," *Federal Reserve Bank of Minneapolis, Quarterly Review*, 2-16.
Sims, C.A. (1988), "Uncertainty Across Models," *American Economic Association Papers and Proceedings*, 78, 167-168.
Sims, C. (1993), "Rational expectations modeling with seasonally adjusted data," *Journal of Econometrics*, 55, 9-19.
Sims, C.A., Stock, J.H., and Watson, M.W. (1990), "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica*, 58, 113-144.
Singleton, K.J. (1988), "Econometric Issues in the Analysis of Equilibrium Business Cycle Models," *Journal of Monetary Economics*, 21, 362-386.
Sowell, F. (1987), Maximum Likelihood Estimation of Fractionally Integrated Time Series Models, no. 87-07, Department of Economics, Duke University.
Sowell, F. (1990), "The Fractional Unit Root Distribution," *Econometrica*, 58, 495-505.
Sowell, F. (1992), "Modeling long-run behavior with the fractional ARIMA model," *Journal of Monetary Economics*, 29, 277-302.
Spanos, A. (1986), *Statistical foundations of econometric modelling*, Cambridge University Press, Cambridge, UK.
Spliid, H. (1983), "A Fast Estimation Method for the Vector Autoregressive Moving Average Model With Exogenous Variables," *Journal of the American Statistical Association*, 78, 843-849.
Stephens, M.A. (1986), "Tests based on EDF statistics," pp. 97-193 in *Goodness-of-Fit Techniques*, eds. D'Agostino, R.B., and Stephens, M.A., Marcel Dekker, inc., New York, USA.
Stewart, G.W. (1973), *Introduction to matrix computation*, Academic Press, Orlando, USA.

Stock, J.H., and Watson, M.W. (1987), Testing for Common Trends, Discussion Paper, no. E-87-2, The Hoover Institution, USA.

Stock, J.H., and Watson, M.W. (1988), "Testing for Common Trends," *Journal of the American Statistical Association*, 83, 1097-1107.

Sun Pascal (1988), *Revision A of May*, Sun Microsystems, USA.

Ter Rele, H.J.M. (1989), Het fiscale klimaat om te sparen en investeren in een aantal landen, Onderzoeksmemorandum, no. 58, Centraal Planbureau, The Hague, The Netherlands.

Theil, H. (1961), *Economic Forecasts and Policy*, North-Holland Publishing, Amsterdam, The Netherlands, 2nd edition.

Theil, H. (1971), *Principles of Econometrics*, North-Holland, Amsterdam, The Netherlands.

Thury, G. (1982), "Modelling Private Consumer Expenditure in Austria by Intervention Analysis," pp. 307-322 in *Time Series Analysis: Theory and Practice* 1, ed. Anderson, O.D., North Holland, Amsterdam, The Netherlands.

Tiao, G.C., and Tsay, R.S. (1989), "Model Specification in Multivariate Time Series (with Discussion)," *Journal of the Royal Statistical Society, Series B*, 51, 157-195 and 208-213.

Tjøstheim, D., and Paulsen, J. (1983), "Bias of some commonly-used time series estimates," *Biometrika*, 70, 389-399.

Tsay, R.S. (1987), "Conditional Heteroscedastic Time Series Models," *Journal of the American Statistical Association*, 82, 590-604.

Tsay, R.S. (1988), "Outliers, Level Shifts, and Variance Changes in Time Series," *Journal of Forecasting*, 7, 1-20.

Tsay, R.S. (1989a), "Identifying Multivariate Time Series Models," *Journal of Time Series Analysis*, 10, 357-372.

Tsay, R.S. (1989b), "Parsimonious Parameterization of Vector Autoregressive Moving Average Models," *Journal of Business and Economic Statistics*, 7, 327-341.

Tsay, R.S., and Tiao, G.C. (1984), "Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Nonstationary ARMA Models," *Journal of the American Statistical Association*, 79, 84-96.

Tsay, R.S., and Tiao, G.C. (1990), "Asymptotic Properties of Multivariate Nonstationary Processes with Applications to Autoregressions," *Annals of Statistics*, 18, 220-250.

Tukey, J.W. (1962), "The future of data analysis," *Annals of Mathematical Statistics*, 33, 1-67.

Tukey, J.W. (1979), "Comment," *Journal of the American Statistical Association*, 74, 121-122.

Turbo Pascal, (1988), *Reference Guide Version* 5.0, Borland International, Scotts Valley, CA, USA.

Valliant, R., and Miller, S.M. (1989), "A Class of Multiplicative Estimators of Laspeyres Price Indexes," *Journal of Business and Economic Statistics*, 7, 387-394.

Van Den Berg, P.J.C.M., Don, F.J.H., and Sandee, J., (1983), *Kompas, Kwartaalmodel voor Prognose, Analyse en Simulatie*, vol. 26 of Centraal Planbureau Monografieën, Staatsuitgeverij, The Hague, Netherlands.

Van Der Hoeven, H., and Hundepool, A.J. (1986), "A method for Seasonally Adjusting Time Series With Variation in the Seasonal Amplitude," *Journal of Business and Economic Statistics*, 4, 455-471.

Velleman, P.F. (1980), "Definition and Comparison of Robust Nonlinear Data Smoothing Algorithms," *Journal of the American Statistical Association*, 75, 609-615.

Wallis, K.F. (1974), "Seasonal Adjustment and Relations Between Variables," *Journal of the American Statistical Association*, 69, 18-31.

Wallis, K.F. (1982), "Seasonal Adjustment and Revision of Current Data: Linear Filters for the X-11 Method," *Journal of the Royal Statistical Society, Series A*, 145, 74-85.

Warne, A. (1991a), Inference in Vector Autoregressions with Integrated Time Series: Part 1, Working Paper, no. 31, Department of Finance, Stockholm School of Economics, Sweden.

Warne, A. (1991b), A Common Trends Model: Identification, Estimation and Asymptotics,

Working Paper, Department of Finance, Stockholm School of Economics, Sweden.
Warner, J.C. (1978), "Unfulfilled Long-Term Interest Rate Expectations and Changes in Business Fixed Investment," *American Economic Review*, 68, 339–347.
Watson, M.W. (1986), "Univariate Detrending Methods with Stochastic Trends," *Journal of Monetary Economics*, 18, 49–75.
Weiss, A.A. (1986), "ARCH and Bilinear Time Series Models: Comparison and Combination," *Journal of Business and Economic Statistics*, 4, 59–70.
Welsch, R.E. (1979), "Comment," *Journal of the American Statistical Association*, 74, 123–124.
White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838.
Wickens, M.R., and Breusch, T.S. (1988), "Dynamic Specification, The Long Run and the Estimation of Transformed Regression Models," *Economic Journal*, 98, 189–205.
Woodland, A.D. (1986), "An Aspect of the Wald Test for Linear Restrictions in the Seemingly Unrelated Regressions Regressions Model," *Economics Letters*, 20, 165–169.
Wooldridge, J.M. (1991), "On the application of robust, regression-based diagnostics to models of conditional means and conditional variances," *Journal of Econometrics*, 47, 5–46.
Zellner, A. (1988), "Bayesian Analysis in Econometrics," *Journal of Econometrics, Annals*, 37, 27–50.
Zellner, A., and Palm, F. (1974), "Time Series Analysis and Simultaneous Equation Econometric Models," *Journal of Econometrics*, 2, 17–54.
Zivot, E., and Andrews, D.W.K. (1992), "Further evidence on the Great Crash, the Oil-Price Shock and the Unit-Root Hypothesis," *Journal of Business and Economic Statistics*, 10, 251–270.

NAME INDEX

SUBJECT INDEX

D

F

P

Q

R

S

U

V

W

LIST OF FIGURES

LIST OF TABLES

Printing: Weihert-Druck GmbH, Darmstadt
Binding: Buchbinderei Schäffer, Grünstadt

Lecture Notes in Economics and Mathematical Systems

For information about Vols. 1–229
please contact your bookseller or Springer-Verlag

Vol. 230: Macro-Economic Planning with Conflicting Goals. Proceedings, 1982. Edited by M. Despontin, P. Nijkamp and J. Spronk. VI, 297 pages. 1984.

Vol. 231: G.F. Newell, The M/M/8 Service System with Ranked Servers in Heavy Traffic. XI, 126 pages. 1984.

Vol. 232: L. Bauwens, Bayesian Full Information Analysis of Simultaneous Equation Models Using Integration by Monte Carlo. VI, 114 pages. 1984.

Vol. 233: G. Wagenhals, The World Copper Market. XI, 190 pages. 1984.

Vol. 234: B.C. Eaves, A Course in Triangulations for Solving Equations with Deformations. III, 302 pages. 1984.

Vol. 235: Stochastic Models in Reliability Theory Proceedings, 1984. Edited by S. Osaki and Y. Hatoyama. VII, 212 pages. 1984.

Vol. 236: G. Gandolfo, P.C. Padoan, A Disequilibrium Model of Real and Financial Accumulation in an Open Economy. VI, 172 pages. 1984.

Vol. 237: Misspecification Analysis. Proceedings, 1983. Edited by T.K. Dijkstra. V, 129 pages. 1984.

Vol. 238: W. Domschke, A. Drexl, Location and Layout Planning. IV, 134 pages. 1985.

Vol. 239: Microeconomic Models of Housing Markets. Edited by K. Stahl. VII, 197 pages. 1985.

Vol. 240: Contributions to Operations Research. Proceedings, 1984. Edited by K. Neumann and D. Pallaschke. V, 190 pages. 1985.

Vol. 241: U. Wittmann, Das Konzept rationaler Preiserwartungen. XI, 310 Seiten. 1985.

Vol. 242: Decision Making with Multiple Objectives. Proceedings, 1984. Edited by Y.Y. Haimes and V. Chankong. XI, 571 pages. 1985.

Vol. 243: Integer Programming and Related Areas. A Classified Bibliography 1981–1984. Edited by R. von Randow. XX, 386 pages. 1985.

Vol. 244: Advances in Equilibrium Theory. Proceedings, 1984. Edited by C.D. Aliprantis, O. Burkinshaw and N.J. Rothman. II, 235 pages. 1985.

Vol. 245: J.E.M. Wilhelm, Arbitrage Theory. VII, 114 pages. 1985.

Vol. 246: P.W. Otter, Dynamic Feature Space Modelling, Filtering and Self-Tuning Control of Stochastic Systems. XIV, 177 pages.1985.

Vol. 247: Optimization and Discrete Choice in Urban Systems. Proceedings, 1983. Edited by B.G. Hutchinson, P. Nijkamp and M. Batty VI, 371 pages. 1985.

Vol. 248: Pural Rationality and Interactive Decision Processes. Proceedings, 1984. Edited by M. Grauer, M. Thompson and A.P. Wierzbicki. VI, 354 pages. 1985.

Vol. 249: Spatial Price Equilibrium: Advances in Theory, Computation and Application. Proceedings, 1984. Edited by P.T. Harker. VII, 277 pages. 1985.

Vol. 250: M. Roubens, Ph. Vincke, Preference Modelling. VIII, 94 pages. 1985.

Vol. 251: Input-Output Modeling. Proceedings, 1984. Edited by A. Smyshlyaev. VI, 261 pages. 1985.

Vol. 252: A. Birolini, On the Use of Stochastic Processes in Modeling Reliability Problems. VI, 105 pages. 1985.

Vol. 253: C. Withagen, Economic Theory and International Trade in Natural Exhaustible Resources. VI, 172 pages. 1985.

Vol. 254: S. Müller, Arbitrage Pricing of Contingent Claims. VIII, 151 pages. 1985.

Vol. 255: Nondifferentiable Optimization: Motivations and Applications. Proceedings, 1984. Edited by V.F. Demyanov and D. Pallaschke. VI, 350 pages. 1985.

Vol. 256: Convexity and Duality in Optimization. Proceedings, 1984. Edited by J. Ponstein. V, 142 pages. 1985.

Vol. 257: Dynamics of Macrosystems. Proceedings, 1984. Edited by J.-P. Aubin, D. Saari and K. Sigmund. VI, 280 pages. 1985.

Vol. 258: H. Funke, Eine allgemeine Theorie der Polypol- und Oligopolpreisbildung. III, 237 pages. 1985.

Vol. 259: Infinite Programming. Proceedings, 1984. Edited by E.J. Anderson and A.B. Philpott. XIV, 244 pages. 1985.

Vol. 260: H.-J. Kruse, Degeneracy Graphs and the Neighbourhood Problem. VIII, 128 pages. 1986.

Vol. 261: Th.R. Gulledge, Jr., N.K. Womer, The Economics of Made-to-Order Production. VI, 134 pages. 1986.

Vol. 262: H.U. Buhl, A Neo-Classical Theory of Distribution and Wealth. V, 146 pages. 1986.

Vol. 263: M. Schäfer, Resource Extraction and Market Struucture. XI, 154 pages. 1986.

Vol. 264: Models of Economic Dynamics. Proceedings, 1983. Edited by H.F. Sonnenschein. VII, 212 pages. 1986.

Vol. 265: Dynamic Games and Applications in Economics. Edited by T. Basar. IX, 288 pages. 1986.

Vol. 266: Multi-Stage Production Planning and Inventory Control. Edited by S. Axsäter, Ch. Schneeweiss and E. Silver. V, 264 pages.1986.

Vol. 267: R. Bemelmans, The Capacity Aspect of Inventories. IX, 165 pages. 1986.

Vol. 268: V. Firchau, Information Evaluation in Capital Markets. VII, 103 pages. 1986.

Vol. 269: A. Borglin, H. Keiding, Optimality in Infinite Horizon Economies. VI, 180 pages. 1986.

Vol. 270: Technological Change, Employment and Spatial Dynamics. Proceedings, 1985. Edited by P. Nijkamp. VII, 466 pages. 1986.

Vol. 271: C. Hildreth, The Cowles Commission in Chicago, 1939–1955. V, 176 pages. 1986.

Vol. 272: G. Clemenz, Credit Markets with Asymmetric Information. VIII,212 pages. 1986.

Vol. 273: Large-Scale Modelling and Interactive Decision Analysis. Proceedings, 1985. Edited by G. Fandel, M. Grauer, A. Kurzhanski and A.P. Wierzbicki. VII, 363 pages. 1986.

Vol. 274: W.K. Klein Haneveld, Duality in Stochastic Linear and Dynamic Programming. VII, 295 pages. 1986.

Vol. 275: Competition, Instability, and Nonlinear Cycles. Proceedings, 1985. Edited by W. Semmler. XII, 340 pages. 1986.

Vol. 276: M.R. Baye, D.A. Black, Consumer Behavior, Cost of Living Measures, and the Income Tax. VII, 119 pages. 1986.

Vol. 277: Studies in Austrian Capital Theory, Investment and Time. Edited by M. Faber. VI, 317 pages. 1986.

Vol. 278: W.E. Diewert, The Measurement of the Economic Benefits of Infrastructure Services. V, 202 pages. 1986.

Vol. 279: H.-J. Büttler, G. Frei and B. Schips, Estimation of Disequilibrium Modes. VI, 114 pages. 1986.

Vol. 280: H.T. Lau, Combinatorial Heuristic Algorithms with FORTRAN. VII, 126 pages. 1986.

Vol. 281: Ch.-L. Hwang, M.-J. Lin, Group Decision Making under Multiple Criteria. XI, 400 pages. 1987.

Vol. 282: K. Schittkowski, More Test Examples for Nonlinear Programming Codes. V, 261 pages. 1987.

Vol. 283: G. Gabisch, H.-W. Lorenz, Business Cycle Theory. VII, 229 pages. 1987.

Vol. 284: H. Lütkepohl, Forecasting Aggregated Vector ARMA Processes. X, 323 pages. 1987.

Vol. 285: Toward Interactive and Intelligent Decision Support Systems. Volume 1. Proceedings, 1986. Edited by Y. Sawaragi, K. Inoue and H. Nakayama. XII, 445 pages. 1987.

Vol. 286: Toward Interactive and Intelligent Decision Support Systems. Volume 2. Proceedings, 1986. Edited by Y. Sawaragi, K. Inoue and H. Nakayama. XII, 450 pages. 1987.

Vol. 287: Dynamical Systems. Proceedings, 1985. Edited by A.B. Kurzhanski and K. Sigmund. VI, 215 pages. 1987.

Vol. 288: G.D. Rudebusch, The Estimation of Macroeconomic Disequilibrium Models with Regime Classification Information. VII,128 pages. 1987.

Vol. 289: B.R. Meijboom, Planning in Decentralized Firms. X, 168 pages. 1987.

Vol. 290: D.A. Carlson, A. Haurie, Infinite Horizon Optimal Control. XI, 254 pages. 1987.

Vol. 291: N. Takahashi, Design of Adaptive Organizations. VI, 140 pages. 1987.

Vol. 292: I. Tchijov, L. Tomaszewicz (Eds.), Input-Output Modeling. Proceedings, 1985. VI, 195 pages. 1987.

Vol. 293: D. Batten, J. Casti, B. Johansson (Eds.), Economic Evolution and Structural Adjustment. Proceedings, 1985. VI, 382 pages.

Vol. 294: J. Jahn, W. Knabs (Eds.), Recent Advances and Historical Development of Vector Optimization. VII, 405 pages. 1987.

Vol. 295. H. Meister, The Purification Problem for Constrained Games with Incomplete Information. X, 127 pages. 1987.

Vol. 296: A. Börsch-Supan, Econometric Analysis of Discrete Choice. VIII, 211 pages. 1987.

Vol. 297: V. Fedorov, H. Läuter (Eds.), Model-Oriented Data Analysis. Proceedings, 1987. VI, 239 pages. 1988.

Vol. 298: S.H. Chew, Q. Zheng, Integral Global Optimization. VII, 179 pages. 1988.

Vol. 299: K. Marti, Descent Directions and Efficient Solutions in Discretely Distributed Stochastic Programs. XIV, 178 pages. 1988.

Vol. 300: U. Derigs, Programming in Networks and Graphs. XI, 315 pages. 1988.

Vol. 301: J. Kacprzyk, M. Roubens (Eds.), Non-Conventional Preference Relations in Decision Making. VII, 155 pages. 1988.

Vol. 302: H.A. Eiselt, G. Pederzoli (Eds.), Advances in Optimization and Control. Proceedings, 1986. VIII, 372 pages. 1988.

Vol. 303: F.X. Diebold, Empirical Modeling of Exchange Rate Dynamics. VII, 143 pages. 1988.

Vol. 304: A. Kurzhanski, K. Neumann, D. Pallaschke (Eds.), Optimization, Parallel Processing and Applications. Proceedings, 1987. VI, 292 pages. 1988.

Vol. 305: G.-J.C.Th. van Schijndel, Dynamic Firm and Investor Behaviour under Progressive Personal Taxation. X, 215 pages.1988.

Vol. 306: Ch. Klein, A Static Microeconomic Model of Pure Competition. VIII, 139 pages. 1988.

Vol. 307: T.K. Dijkstra (Ed.), On Model Uncertainty and its Statistical Implications. VII, 138 pages. 1988.

Vol. 308: J.R. Daduna, A. Wren (Eds.), Computer-Aided Transit Scheduling. VIII, 339 pages. 1988.

Vol. 309: G. Ricci, K. Velupillai (Eds.), Growth Cycles and Multisectoral Economics: the Goodwin Tradition. III, 126 pages. 1988.

Vol. 310: J. Kacprzyk, M. Fedrizzi (Eds.), Combining Fuzzy Imprecision with Probabilistic Uncertainty in Decision Making. IX, 399 pages. 1988.

Vol. 311: R. Färe, Fundamentals of Production Theory. IX, 163 pages. 1988.

Vol. 312: J. Krishnakumar, Estimation of Simultaneous Equation Models with Error Components Structure. X, 357 pages. 1988.

Vol. 313: W. Jammernegg, Sequential Binary Investment Decisions. VI, 156 pages. 1988.

Vol. 314: R. Tietz, W. Albers, R. Selten (Eds.), Bounded Rational Behavior in Experimental Games and Markets. VI, 368 pages. 1988.

Vol. 315: I. Orishimo, G.J.D. Hewings, P. Nijkamp (Eds), Information Technology: Social and Spatial Perspectives. Proceedings 1986. VI, 268 pages. 1988.

Vol. 316: R.L. Basmann, D.J. Slottje, K. Hayes, J.D. Johnson, D.J. Molina, The Generalized Fechner-Thurstone Direct Utility Function and Some of its Uses. VIII, 159 pages. 1988.

Vol. 317: L. Bianco, A. La Bella (Eds.), Freight Transport Planning and Logistics. Proceedings, 1987. X, 568 pages. 1988.

Vol. 318: T. Doup, Simplicial Algorithms on the Simplotope. VIII, 262 pages. 1988.

Vol. 319: D.T. Luc, Theory of Vector Optimization. VIII, 173 pages. 1989.

Vol. 320: D. van der Wijst, Financial Structure in Small Business. VII, 181 pages. 1989.

Vol. 321: M. Di Matteo, R.M. Goodwin, A. Vercelli (Eds.), Technological and Social Factors in Long Term Fluctuations. Proceedings. IX, 442 pages. 1989.

Vol. 322: T. Kollintzas (Ed.), The Rational Expectations Equilibrium Inventory Model. XI, 269 pages. 1989.

Vol. 323: M.B.M. de Koster, Capacity Oriented Analysis and Design of Production Systems. XII, 245 pages. 1989.

Vol. 324: I.M. Bomze, B.M. Pötscher, Game Theoretical Foundations of Evolutionary Stability. VI, 145 pages. 1989.

Vol. 325: P. Ferri, E. Greenberg, The Labor Market and Business Cycle Theories. X, 183 pages. 1989.

Vol. 326: Ch. Sauer, Alternative Theories of Output, Unemployment, and Inflation in Germany: 1960–1985. XIII, 206 pages. 1989.

Vol. 327: M. Tawada, Production Structure and International Trade. V, 132 pages. 1989.

Vol. 328: W. Güth, B. Kalkofen, Unique Solutions for Strategic Games. VII, 200 pages. 1989.

Vol. 329: G. Tillmann, Equity, Incentives, and Taxation. VI, 132 pages. 1989.

Vol. 330: P.M. Kort, Optimal Dynamic Investment Policies of a Value Maximizing Firm. VII, 185 pages. 1989.

Vol. 331: A. Lewandowski, A.P. Wierzbicki (Eds.), Aspiration Based Decision Support Systems. X, 400 pages. 1989.

Vol. 332: T.R. Gulledge, Jr., L.A. Litteral (Eds.), Cost Analysis Applications of Economics and Operations Research. Proceedings. VII, 422 pages. 1989.

Vol. 333: N. Dellaert, Production to Order. VII, 158 pages. 1989.

Vol. 334: H.-W. Lorenz, Nonlinear Dynamical Economics and Chaotic Motion. XI, 248 pages. 1989.

Vol. 335: A.G. Lockett, G. Islei (Eds.), Improving Decision Making in Organisations. Proceedings. IX, 606 pages. 1989.

Vol. 336: T. Puu, Nonlinear Economic Dynamics. VII, 119 pages. 1989.

Vol. 337: A. Lewandowski, I. Stanchev (Eds.), Methodology and Software for Interactive Decision Support. VIII, 309 pages. 1989.

Vol. 338: J.K. Ho, R.P. Sundarraj, DECOMP: an Implementation of Dantzig-Wolfe Decomposition for Linear Programming. VI, 206 pages.

Vol. 339: J. Terceiro Lomba, Estimation of Dynamic Econometric Models with Errors in Variables. VIII, 116 pages. 1990.

Vol. 340: T. Vasko, R. Ayres, L. Fontvieille (Eds.), Life Cycles and Long Waves. XIV, 293 pages. 1990.

Vol. 341: G.R. Uhlich, Descriptive Theories of Bargaining. IX, 165 pages. 1990.

Vol. 342: K. Okuguchi, F. Szidarovszky, The Theory of Oligopoly with Multi-Product Firms. V, 167 pages. 1990.

Vol. 343: C. Chiarella, The Elements of a Nonlinear Theory of Economic Dynamics. IX, 149 pages. 1990.

Vol. 344: K. Neumann, Stochastic Project Networks. XI, 237 pages. 1990.

Vol. 345: A. Cambini, E. Castagnoli, L. Martein, P Mazzoleni, S. Schaible (Eds.), Generalized Convexity and Fractional Programming with Economic Applications. Proceedings, 1988. VII, 361 pages. 1990.

Vol. 346: R. von Randow (Ed.), Integer Programming and Related Areas. A Classified Bibliography 1984–1987. XIII, 514 pages. 1990.

Vol. 347: D. Ríos Insua, Sensitivity Analysis in Multiobjective Decision Making. XI, 193 pages. 1990.

Vol. 348: H. Störmer, Binary Functions and their Applications. VIII, 151 pages. 1990.

Vol. 349: G.A. Pfann, Dynamic Modelling of Stochastic Demand for Manufacturing Employment. VI, 158 pages. 1990.

Vol. 350: W.-B. Zhang, Economic Dynamics. X, 232 pages. 1990.

Vol. 351: A. Lewandowski, V. Volkovich (Eds.), Multiobjective Problems of Mathematical Programming. Proceedings, 1988. VII, 315 pages. 1991.

Vol. 352: O. van Hilten, Optimal Firm Behaviour in the Context of Technological Progress and a Business Cycle. XII, 229 pages. 1991.

Vol. 353: G. Ricci (Ed.), Decision Processes in Economics. Proceedings, 1989. III, 209 pages 1991.

Vol. 354: M. Ivaldi, A Structural Analysis of Expectation Formation. XII, 230 pages. 1991.

Vol. 355: M. Salomon. Deterministic Lotsizing Models for Production Planning. VII, 158 pages. 1991.

Vol. 356: P. Korhonen, A. Lewandowski, J . Wallenius (Eds.), Multiple Criteria Decision Support. Proceedings, 1989. XII, 393 pages. 1991.

Vol. 357: P. Zörnig, Degeneracy Graphs and Simplex Cycling. XV, 194 pages. 1991.

Vol. 358: P. Knottnerus, Linear Models with Correlated Disturbances. VIII, 196 pages. 1991.

Vol. 359: E. de Jong, Exchange Rate Determination and Optimal Economic Policy Under Various Exchange Rate Regimes. VII, 270 pages. 1991.

Vol. 360: P. Stalder, Regime Translations, Spillovers and Buffer Stocks. VI, 193 pages . 1991.

Vol. 361: C. F. Daganzo, Logistics Systems Analysis. X, 321 pages. 1991.

Vol. 362: F. Gehrels, Essays In Macroeconomics of an Open Economy. VII, 183 pages. 1991.

Vol. 363: C. Puppe, Distorted Probabilities and Choice under Risk. VIII, 100 pages . 1991

Vol. 364: B. Horvath, Are Policy Variables Exogenous? XII, 162 pages. 1991.

Vol. 365: G. A. Heuer, U. Leopold-Wildburger. Balanced Silverman Games on General Discrete Sets. V, 140 pages. 1991.

Vol. 366: J. Gruber (Ed.), Econometric Decision Models. Proceedings, 1989. VIII, 636 pages. 1991.

Vol. 367: M. Grauer, D. B. Pressmar (Eds.), Parallel Computing and Mathematical Optimization. Proceedings. V, 208 pages. 1991.

Vol. 368: M. Fedrizzi, J. Kacprzyk, M. Roubens (Eds.), Interactive Fuzzy Optimization. VII, 216 pages. 1991.

Vol. 369: R. Koblo, The Visible Hand. VIII, 131 pages.1991.

Vol. 370: M. J. Beckmann, M. N. Gopalan, R. Subramanian (Eds.), Stochastic Processes and their Applications. Proceedings, 1990. XLI, 292 pages. 1991.

Vol. 371: A. Schmutzler, Flexibility and Adjustment to Information in Sequential Decision Problems. VIII, 198 pages. 1991.

Vol. 372: J. Esteban, The Social Viability of Money. X, 202 pages. 1991.

Vol. 373: A. Billot, Economic Theory of Fuzzy Equilibria. XIII, 164 pages. 1992.

Vol. 374: G. Pflug, U. Dieter (Eds.), Simulation and Optimization. Proceedings, 1990. X, 162 pages. 1992.

Vol. 375: S.-J. Chen, Ch.-L. Hwang, Fuzzy Multiple Attribute Decision Making. XII, 536 pages. 1992.

Vol. 376: K.-H. Jöckel, G. Rothe, W. Sendler (Eds.), Bootstrapping and Related Techniques. Proceedings, 1990. VIII, 247 pages. 1992.

Vol. 377: A. Villar, Operator Theorems with Applications to Distributive Problems and Equilibrium Models. XVI, 160 pages. 1992.

Vol. 378: W. Krabs, J. Zowe (Eds.), Modern Methods of Optimization. Proceedings, 1990. VIII, 348 pages. 1992.

Vol. 379: K. Marti (Ed.), Stochastic Optimization. Proceedings, 1990. VII, 182 pages. 1992.

Vol. 380: J. Odelstad, Invariance and Structural Dependence. XII, 245 pages. 1992.

Vol. 381: C. Giannini, Topics in Structural VAR Econometrics. XI, 131 pages. 1992.

Vol. 382: W. Oettli, D. Pallaschke (Eds.), Advances in Optimization. Proceedings, 1991. X, 527 pages. 1992.

Vol. 383: J. Vartiainen, Capital Accumulation in a Corporatist Economy. VII, 177 pages. 1992.

Vol. 384: A. Martina, Lectures on the Economic Theory of Taxation. XII, 313 pages. 1992.

Vol. 385: J. Gardeazabal, M. Regúlez, The Monetary Model of Exchange Rates and Cointegration. X, 194 pages. 1992.

Vol. 386: M. Desrochers, J.-M. Rousseau (Eds.), Computer-Aided Transit Scheduling. Proceedings, 1990. XIII, 432 pages. 1992.

Vol. 387: W. Gaertner, M. Klemisch-Ahlert, Social Choice and Bargaining Perspectives on Distributive Justice. VIII, 131 pages. 1992.

Vol. 388: D. Bartmann, M. J. Beckmann, Inventory Control. XV, 252 pages. 1992.

Vol. 389: B. Dutta, D. Mookherjee, T. Parthasarathy, T. Raghavan, D. Ray, S. Tijs (Eds.), Game Theory and Economic Applications. Proceedings, 1990. ??, ?? pages. 1992.

Vol. 390: G. Sorger, Minimum Impatience Theorem for Recursive Economic Models. X, 162 pages. 1992.

Vol. 391: C. Keser, Experimental Duopoly Markets with Demand Inertia. X, 150 pages. 1992.

Vol. 392: K. Frauendorfer, Stochastic Two-Stage Programming. VIII, 228 pages. 1992.

Vol. 393: B. Lucke, Price Stabilization on World Agricultural Markets. XI, 274 pages. 1992.

Vol. 394: Y.-J. Lai, C.-L. Hwang, Fuzzy Mathematical Programming. XIII, 301 pages. 1992.

Vol. 395: G. Haag, U. Mueller, K. G. Troitzsch (Eds.), Economic Evolution and Demographic Change. XVI, 409 pages. 1992.

Vol. 396: R. V. V. Vidal (Ed.), Applied Simulated Annealing. VIII, 358 pages. 1992.

Vol. 397: J. Wessels, A. P. Wierzbicki (Eds.), User-Oriented Methodology and Techniques of Decision Analysis and Support. Proceedings, 1991. XII, 295 pages. 1993.

Vol. 398: J.-P. Urbain, Exogeneity in Error Correction Models. XI, 189 pages. 1993.

Vol. 399: F. Gori, L. Geronazzo, M. Galeotti (Eds.), Nonlinear Dynamics in Economics and Social Sciences. Proceedings, 1991. VIII, 367 pages. 1993.

Vol. 400: H. Tanizaki, Nonlinear Filters. XII, 203 pages. 1993.

Vol. 401: K. Mosler, M. Scarsini, Stochastic Orders and Applications. V, 379 pages. 1993.

Vol. 402: A. van den Elzen, Adjustment Processes for Exchange Economies and Noncooperative Games. VII, 146 pages. 1993.

Vol. 403: G. Brennscheidt, Predictive Behavior. VI, 227 pages. 1993.

Vol. 404: Y.-J. Lai, Ch.-L. Hwang, Fuzzy Multiple Objective Decision Making. XIV, 475 pages. 1994.

Vol. 405: S. Komlósi, T. Rapcsák, S. Schaible (Eds.), Generalized Convexity. Proceedings, 1992. VIII, 404 pages. 1994.

Vol. 406: N. M. Hung, N. Van Quyen, Dynamic Timing Decisions Under Uncertainty. X, 194 pages. 1994.

Vol. 407: M. Ooms, Empirical Vector Autoregressive Modeling. XIII, 380 pages. 1994.

Lecture Notes in Economics and Mathematical Systems

This series reports on new developments in mathematical economics, economic theory, econometrics, operations research, and mathematical systems, research and teaching-quickly, informally and at a high level. The type of material considered for publication includes

1. Research monographs
2. Lectures on a new field or presentations of a new angle in a classical field
3. Seminars on topics of current research
4. Reports of meetings, provided they are
   a) of exceptional interest and
   b) devoted to a single topic.

Texts which are out of print but still in demand may also be considered if they fall within these categories.

In the case of a research monograph, or of seminar notes, the timeliness of a manuscript may be more important than its form, which may be preliminary or tentative. A subject index should be included.

Manuscripts

Manuscripts should be no less than 150 and preferably no more than 500 pages in length. They should be submitted in camera-ready form according to Springer-Verlag's specifications. Technical instructions will be sent on request. Manuscripts should be sent directly to Springer-Verlag Heidelberg.

ISBN 3-540-57707-6
ISBN 0-387-57707-6