Essays on Nonlinear Time Series Modelling and Hypothesis Testing

Birgit Strikholm

AKADEMISK AVHANDLING

Som för avläggande av ekonomie doktorsexamen vid Handelshögskolan i Stockholm framläggs för offentlig granskning den 13 December 2004, kl 10.15 i sal KAW, Handelshögskolan,
Sveavägen 65
Essays on Nonlinear Time Series Modelling and Hypothesis Testing
EFI Mission
EFI, the Economic Research Institute at the Stockholm School of Economics, is a scientific institution which works independently of economic, political and sectional interests. It conducts theoretical and empirical research in the management and economic sciences, including selected related disciplines. The Institute encourages and assists in the publication and distribution of its research findings and is also involved in the doctoral education at the Stockholm School of Economics. At EFI, the researchers select their projects based on the need for theoretical or practical development of a research domain, on methodological interests, and on the generality of a problem.

Research Organization
The research activities at the Institute are organized in 21 Research Centers within eight Research Areas. Center Directors are professors at the Stockholm School of Economics.

ORGANIZATION AND MANAGEMENT
Management and Organisation (A) Prof Sven-Erik Sjöstrand
Center for Ethics and Economics (CEE) Adj Prof Hans de Geer
Center for Entrepreneurship and Business Creation (E) Prof Carin Holmquist
Public Management (F) Prof Nils Brunsson
Information Management (I) Prof Mats Lundeborg
Center for People and Organization (PMO) Prof Jan Löwstedt
Center for Innovation and Operations Management (T) Prof Christer Karlsson

ECONOMIC PSYCHOLOGY
Center for Risk Research (CFR) Prof Lennart Sjöberg
Economic Psychology (P) Prof Guje Sevon

MARKETING
Center for Consumer Marketing (CCM) Acting Prof Magnus Söderlund
Center for Information and Communication Research (CIC) Adj Prof Bertil Thorngren
Marketing, Distribution and Industry Dynamics (D) Prof Björn Axelson

ACCOUNTING, CONTROL AND CORPORATE FINANCE
Accounting and Managerial Finance (B) Prof Johnny Lind
Center for Financial Analysis and Managerial Economics in Accounting (BFAC) Prof Kent Skogsvik

FINANCE
Finance (FI) Prof Clas Bergström

ECONOMICS
Center for Health Economics (CHE) Prof Bengt Jönnsson
International Economics and Geography (IEG) Prof Mats Lundahl
Economics (S) Prof Lars Bergman

ECONOMIC STATISTICS
Economic Statistics (ES) Prof Anders Westlund

LAW
Law (RV) Prof Erik Nerep
Center for Tax Law Prof Bertil Wiman

Chair of the Board: Professor Carin Holmquist Director: Associate Professor Filip Wijkström

Address
EFI, Box 6501, SE-113 83 Stockholm, Sweden • Homepage: www.hhs.se/efi/
Telephone: +46(0)8-736 90 00 • Fax: +46(0)8-31 62 70 • E-mail efi@hhs.se
Essays on Nonlinear Time Series Modelling and Hypothesis Testing

Birgit Strikholm

STOCKHOLM SCHOOL OF ECONOMICS
HANDELSHÖGSKOLAN I STOCKHOLM

EFI, The Economic Research Institute
To my family
Acknowledgments

I Summary of Thesis

II The Chapters

1 Determining the Number of Regimes in a Threshold Autoregressive Model Using Smooth Transition Autoregressions 15
  1.1 Introduction ................................................. 17
  1.2 Smooth transition regression model ............................. 19
  1.3 Smooth transition approach ...................................... 21
  1.4 Simulation study ............................................ 23
    1.4.1 Estimating the empirical size ............................. 24
    1.4.2 Simulating TAR models ............................... 25
  1.5 Application ............................................... 31
  1.6 Final remarks .............................................. 32
Appendix A: Properties of threshold estimates .................. 35
References .................................................... 39
Tables ......................................................... 41

2 Determining the Number of Breaks in a Piecewise Linear Regression Model Using Multiple Smooth Transition Regression 49
  2.1 Introduction ................................................. 51
  2.2 Smooth transition regression model ............................. 52
  2.3 Smooth transition-based break detection ........................ 55
  2.4 Simulation study ............................................ 58
    2.4.1 Estimating the empirical size ............................. 59
    2.4.2 Simulating structural break models ........................ 60
2.5 Conclusions .............................................................. 66
References ................................................................. 69
Tables ................................................................. 71

3 The Effects of Institutional and Technological Change and Business Cycle Fluctuations on Seasonal Patterns in Quarterly Industrial Production Series 81
3.1 Introduction ................................................................. 83
3.2 Preliminaries .............................................................. 85
  3.2.1 Data ................................................................. 85
  3.2.2 Deterministic and stochastic seasonality ......................... 86
3.3 The TV-STAR model ......................................................... 87
3.4 Changes in the seasonal pattern and their causes ......................... 90
  3.4.1 Testing linearity and parameter constancy in the TV-STAR framework ......................................................... 90
  3.4.2 Testing hypotheses of interest .................................. 92
  3.4.3 Results ............................................................ 93
3.5 Modelling changing seasonal patterns by TV-STAR models .......... 96
  3.5.1 A TV-STAR model for UK industrial production .......... 96
  3.5.2 Other countries .................................................... 99
3.6 Final remarks .......................................................... 108
References ................................................................. 111
Figures ................................................................. 113
Tables ................................................................. 131

4 Testing the Granger Noncausality Hypothesis in Stationary Nonlinear Models of Unknown Functional Form 139
4.1 Introduction ................................................................. 141
4.2 Tests of the Granger noncausality hypothesis ......................... 142
  4.2.1 Standard linear Granger noncausality test ...................... 142
  4.2.2 Framework for the general test ................................ 143
  4.2.3 Noncausality tests based on a Taylor series approximation 144
  4.2.4 Discussion ........................................................ 148
4.3 Monte Carlo experiments ............................................... 149
  4.3.1 Simulation design ................................................. 149
  4.3.2 Simulation results ................................................ 150
4.4 Conclusions ............................................................. 156
References ................................................................. 157
Figures ................................................................. 159
Acknowledgments

There are a number of people without whose help this thesis would not be complete today. First of all, I would like to express my sincere gratitude to my main supervisor, Timo Teräsvirta. He has offered me his full support since my very first attempts at converting from a statistician to an econometrician. I have greatly benefited from our weekly discussions - the well of new ideas, comments, explanations and good humour seems endless. All this combined with his patience has made the completion of this thesis possible.

Secondly, I would like to thank my secondary supervisor, Dick van Dijk, and not only for being the co-author of my first published paper. One cannot help but brush up ones definitions of “work morale” and “efficiency” when working with Dick. Thanks for the reality check!

I would also like to thank Professors Pentti Saikkonen, Eilev Jansen and Anne Péguiin-Feissolle for discussion and useful remarks. I am grateful to all members of the Department of Economic Statistics for creating a stimulating and pleasant working environment. The value of all their comments cannot be underestimated. I am indebted to our secretaries Pernilla Watson and Carina Morton-Fincham for their help with administrative matters.

Warm thanks go to Brigitta Hultblad. Your friendship and care smoothed my path into the Swedish society. The whole PhD-journey would not have been half the fun it has without the “downstairs gang” - my dear office-mate Rickard Sandberg, and the next-doors neighbours Annastiina Silvennoinen and Mika Meitz. You guys officially rock! I want to thank the rest of my friends around the world, and Kairi Kasemets in particular, for providing the laughs and talking me through the tough times. Last but not least, I would like to thank my family for their constant love and support.

Financial support from the Jan Wallander’s and Tom Hedelius’ Foundation for Social Research Grants No. J99/37 and J02-35 is gratefully acknowledged.

Stockholm, October 2004
Birgit Strikholm
Part I

Summary of Thesis
Introduction and summary of the chapters

There seems to be a common understanding nowadays that the economy is nonlinear. Economic theory suggests models that should be able to accommodate nonlinear production functions, buffer stocks and asymmetric responses to negative and positive news (shocks), to name a few. Over the last twenty years a solid body of empirical evidence of nonlinearities in economic time series has also been gathered. A thorough treatment of the subject - modelling nonlinear time series in economics and finance - can be found in Granger and Teräsvirta (1993) and Franses and van Dijk (2000), for example.

Nonlinearities in stationary macroeconomic series are often modelled with regime-switching or state-dependent models. In their simplest forms, these models only have two linear regimes and the switching between them is abrupt. The estimation of such models was already considered four decades ago. Quandt (1958, 1960) introduced a switching regression model in which the state variable is time. He also proposed a test for a structural break at an unknown point within that framework. The univariate counterparts of the switching regression model, the threshold autoregressive (TAR) model by Tong (1978) and the structural break model, have gained popularity and have often been the competing alternatives to the linear model.

Switching regression models can be generalized in such a way that the switching between regimes becomes smooth and one can think of the resulting model as of one with an infinite number of regimes. One example of these models is the Smooth Transition Regression (STR) model by Teräsvirta (1994, 1998). In STR models the transition between the regimes is characterized by a bounded and continuous (often logistic) function of a certain transition variable. Classical smooth transition autoregressive (STAR) models and their extensions have been successfully applied to account for nonlinear features in various macroeconomic series. See Teräsvirta and Anderson (1992) and Skalin
Introduction and summary of the chapters


When the transition variable in a STAR model is a lagged value of the dependent variable and the transition as a function of this variable is sufficiently rapid, the model can be thought of as an approximation to the Self-Exciting TAR model. A Time-Varying Autoregression (TV-AR) model, see Lundbergh, Teräsvirta, and van Dijk (2003), can, by the same token, approximate a univariate autoregressive model with structural breaks. A clear advantage of an STR model compared to threshold and structural break models is that the log-likelihood of an STR model is continuous, which makes inference and model evaluation easier than it is in the former models.

In this thesis we use the smooth transition model as a tool for modelling and hypothesis testing. The first two chapters are concerned with model selection issues within switching (auto)regression models. In most economic applications of the TAR model, economic theory is not specific about the complete structure of the model. In particular, most often the number of regimes in the model cannot be assumed known a priori and it is the task of the modeller to determine the number of regimes from the available time series. In the first chapter we approximate the threshold autoregressive model by a smooth transition one and use the STR modelling framework to determine the number of regimes in a TAR model. This includes testing linearity against TAR. In the second chapter we employ a similar idea and approximate the structural break model with a time-varying autoregressive model. This is again done in the STR modelling framework which is used for sequentially determining the number of breaks, including the possibility of no breaks. In the third chapter we use time-varying smooth transition autoregressive models to model the (possible) changes in seasonal patterns and shed light on the hypothesis that institutional and technological changes (proxied by time) may have a stronger effect on seasonal patterns than business cycle. The application to the index of industrial production series in G7 and two Scandinavian countries discussed in the chapter seems to lend support to this hypothesis.

Causality has been an important concept for philosophers. Aristotle already devoted attention to causality and it has been given a central role in the recent film The Matrix Reloaded. Causality has also practical significance in economics as different economic theories postulate causal relationships between economic variables. Nonlinear time series analysis has become
Introduction and summary of the chapters

crucial in understanding patterns of interaction between different variables. There exist various formal definitions of causality, of which the concept of Granger causality, see Granger (1969), is probably most easily tested for and therefore also most widely used. Testing has usually been carried out in a linear framework. Geweke (1984) and Lütkepohl (1993) provide comprehensive surveys of the literature. As discussed above, the economic laws are often nonlinear and the empirical notion of linear Granger causality is probably too simple. In the last chapter of the thesis we study the possibility of testing for Granger causality in nonlinear systems when the exact form of the nonlinear relationship between variables is not known. The idea is to globally approximate the nonlinear system by one that is linear in parameters and get round the difficulty caused by the unknown functional form of the relationship under investigation.

A more detailed summary of each of the chapters is given below.

Chapter 1. Determining the number of regimes in a Threshold Autoregressive model using Smooth Transition Autoregressions

As already mentioned, in applications the number of regimes in a threshold autoregressive model is typically not known in advance. In this chapter we propose a solution to this problem. It consists of a sequential model selection procedure using the smooth transition autoregressions as a tool. Tests available for testing the adequacy of a smooth transition autoregressive model are applied sequentially to determine the number of regimes. The main characteristic of this method is that standard statistical inference is used, as opposed to nonstandard inference, information criteria or bootstrap-based methods. As the smooth transition model is just an approximation to the threshold autoregressive one, no asymptotic properties are claimed for the proposed method. At first sight this may appear as a drawback: however, assuming that the switching regression or threshold autoregressive model under consideration has a fixed number of thresholds, the model selection problem at hand is a finite-sample problem. From the practical point of view it is therefore sufficient to require that the procedure performs in a satisfactory fashion in small and moderate samples. A simulation study is carried out in order to

\footnote{I am grateful to Jesús Gonzalo and Jean-Yves Pitarakis who gracefully allowed me to use their GAUSS code written for selecting the number of regimes using model selection criteria.}
find out the finite-sample properties of the procedure and to compare it with two other techniques of regime determination available in the literature, see Gonzalo and Pitarakis (2002) and Hansen (1999). We find that our smooth transition-based method works reasonably well in comparison, both for single and multiple threshold models. Overall, it may be somewhat conservative, but its performance in selecting the TAR model with a correct number of regimes can be deemed acceptable. The technique is computationally simple, and it also performs well when a true threshold lies outside the [0.1,0.9] interquantile range of the observed series. This is important because the other techniques are unable to locate thresholds so far out in the tails of the empirical distribution of the observations. Besides, as a by-product, it yields quite accurate estimates for the threshold parameters which is a strong advantage as well.

Chapter 2. Determining the number of breaks in a piecewise linear regression model using Multiple Smooth Transition Regression

A model builder who holds open the possibility of structural breaks in an otherwise linear model faces several problems: first (s)he has to test for structural change in the parameters of the model and then estimate both the number and the location of breaks. Most of the work in this area of research has been related to the case of detecting and estimating a single break. The questions related to multiple structural changes have received somewhat less attention. In a seminal paper Bai and Perron (1998) derived tests for multiple structural changes and proved consistency of the estimators of the break dates. Last, but not least, they also proposed several methods (one of which is purely sequential) for determining the number of breaks and efficient algorithms for computing the estimates. Since all the tests they considered have nonstandard distributions, Bai and Perron (2003) also provided asymptotic critical values for a number of possible testing situations. Naturally, if the model one is interested in does not agree with those specifications, it would be the task of the investigator to obtain the necessary critical values by simulation.

In this chapter we propose a sequential method for determining the number of breaks in piecewise linear models using smooth transition regression as a tool. This method has similarities with the technique for determining the number of regimes in a TAR model, which was the topic of Chapter 1. An

\footnote{I would like to thank Jushan Bai and Pierre Perron for making their original GAUSS code available to me.}
advantage of the method is that it is based on standard statistical inference and no restrictions regarding the number of parameters that are allowed to change are being made. The time-varying smooth transition model is just an approximation to the structural break one and no asymptotic properties are claimed for the proposed method. Analogously to Chapter 1, if the structural break model under consideration has a fixed number of breaks and the sample size is fixed, the model selection problem is a finite-sample problem. The idea is to sequentially apply tests developed for testing the adequacy of a smooth transition regression model to the problem of determining the number of regimes in the structural break model. The simulations show that our procedure is able to choose the linear model with the expected frequency when there is no break, and works very well in comparison in small and moderate samples, for both single and multiple break cases. The results indicate that our STR-based method is a useful complement to the approach of Bai and Perron, as neither of the two methods dominates the other.

Chapter 3. The effects of institutional and technological change and business cycle fluctuations on seasonal patterns in quarterly industrial production series

Seasonal fluctuations are an important source of variation in many macroeconomic time series. When monthly or quarterly series are modelled or when time series are seasonally adjusted by standard techniques, it is often assumed that the seasonal pattern of the series is constant over time. In that case it may be characterized by incorporating seasonal dummy variables into the model. Despite its widespread use, the dummy approach to seasonality neglects two important facts. First, the traditional separation of seasonal and business cycles is not an attribute of modern dynamic business cycle theory. Second, it has been known for a long time that seasonal patterns of many (macroeconomic) series appear to evolve over time. A number of authors link changes in the seasonal patterns of many macroeconomic time series to the phase of the business cycle. Despite these findings, business cycle fluctuations may not be the only possible reason for changes in the seasonal pattern of macroeconomic time series such as output or employment series. In particular, technological change and changes in institutions and habits may cause changes

---

3This chapter is based on joint work with Timo Teräsvirta and Dick van Dijk. A shorter version has appeared in Econometrics Journal, 6 79 - 98 (2003).
in seasonality as well. Such changes are difficult to quantify directly, but when we allow for the possibility that the aggregate change is smooth and continuous, we can simply use time as a proxy variable for it. It follows that combining these two possible types of change is also of interest.

Our aim in this chapter is to study the seasonal patterns of industrial production series in a number of countries and compare the effects of gradual institutional and technological change with the effects attributable to the business cycle. There is a coherent way of incorporating these two types of changes into a single model. This can be achieved by using the TV-STAR model of Lundbergh et al. (2003), because it is capable of describing business cycle nonlinearity (seasonal pattern is allowed to change as a function of business cycle indicator) and structural change (seasonal pattern is allowed to change as a smooth function of time) in a stationary time series variable simultaneously. By imposing appropriate restrictions on the TV-STAR structure, models of interest can be defined and the relative importance of different types of changes studied. We examine the quarterly volume index of industrial production series of the G7 and two Scandinavian countries, nine series in all. These series display strong seasonal patterns and also contain the business cycle fluctuations. The empirical results of the chapter suggest that seasonal patterns in these nine quarterly industrial production series have been changing over time and, furthermore, that the business cycle fluctuations do not seem to be the main cause for this change. The question of how the current seasonal adjustment methods cope with series with time-varying seasonality, and what the consequences of such variation are on using seasonally adjusted series in macroeconomic modelling, is left for future research.

Chapter 4. Testing the Granger noncausality hypothesis in stationary nonlinear models of unknown functional form

In a seminal paper, Granger (1969) provided an operational definition of causality between two variables that has since formed a starting-point for testing the null hypothesis of one variable (linearly) not causing the other. In many cases, relationships between the variables may be best described by nonlinear models. Several authors have tried to generalize the noncausality test such that it would allow for nonlinear relationships between variables. Examples include Baek and Brock (1992) and Hiemstra and Jones (1994). Most of the solutions provided are nonparametric (based on correlation integrals) and rather computer intensive. Skalin and Teräsvirta (1999) proposed a
parametric test based on the smooth transition regression model. That test is easy to compute, but it still makes assumptions about the specific functional form of the causality. In the empirical literature the presence of nonlinear Granger causality has been suggested for money-output relationship in Rothman, van Dijk, and Franses (2001), for the price volatility - trading volume relation in Hiemstra and Jones (1994), and for relative money supply - exchange rates in Ma and Kanas (2000), to name a few.

In this chapter we propose a general method for testing the Granger non-causality hypothesis in stationary nonlinear models of unknown functional form. These tests are based on a Taylor expansion of the nonlinear model around a given point in a sample space. Our framework nests both the linear case and the test proposed in Skalin and Teräsvirta (1999).

We study the small-sample performance of our tests by a Monte Carlo experiment and compare these to the most widely used linear test. The size simulations we have performed indicate that the theory is applicable in small samples, i.e., our tests appear to be well-sized. The right balance between the number of lags, the order of the Taylor expansion and the number of observations is important, however and needs further investigation. Our tests have reasonably good power and are useful in investigating potential Granger causality between variables. It becomes obvious that the more we know about the functional form, the more we gain in terms of power. On the other hand, if the true causal relationship is nonlinear but we apply tests designed for linear relationships, the loss of power may be substantial.
References


Introduction and summary of the chapters


Part II

The Chapters
Chapter 1

Determining the Number of Regimes in a Threshold Autoregressive Model Using Smooth Transition Autoregressions
“Choice. The problem is choice.”

– Neo, The Matrix Reloaded
1.1 Introduction

The switching regression model (Quandt, 1958) and its univariate counterpart, the threshold autoregressive (TAR) model (Tong, 1978) are popular nonlinear models. The TAR model in particular has generated a wide range of papers covering both theoretical and empirical issues. An overview can be found in Tong (1990); see also, for example, Caner and Hansen (2001), Hansen (1996, 1999a, 2000), Kapetanios (2003), Koop and Potter (1999), Medeiros et al. (2002), among others.

In most economic applications of the TAR model, economic theory is not specific about the complete structure of the model. In particular, most often the number of regimes in the model cannot be assumed known a priori. Furthermore, the switch variable or, in the TAR case, the delay determining the threshold variable is often unknown as well. Some work exists on how to select the number of regimes in TAR models. Tsay (1989) suggested a graphical approach for locating the values of thresholds. He used scatterplots of standardized predictive residuals (in arranged autoregression) and recursive t-ratios of an AR coefficient versus the threshold variable to detect the number and locations of the thresholds. Hansen (1996) considered inference and testing for linearity in situations when a nuisance parameter is not identified under the null hypothesis. He provided a general framework using weighted average and supremum LM tests and gave the asymptotic theory for inference. Hansen (1999a) suggested a sequential testing approach to the regime selection problem. This meant starting with a linear model and adding thresholds until the first acceptance of a null hypothesis. A statistical complication is that the parameters of the TAR model are only identified under the alternative, that is, when the larger model is true. He suggested a likelihood ratio-type test and showed how inference can be conducted using an empirical null distribution of the test statistic generated by the bootstrap. We shall investigate how such a sequential procedure works in practice.

More recently, Gonzalo and Pitarakis (2002), henceforth GP, suggested choosing the number of regimes or thresholds sequentially starting from a linear model (a single regime) and using model selection criteria for choosing between models with \(m\) and \(m + 1\) thresholds. Their argument was that the procedure is easy to use, and as opposed to statistical tests, there is no need to choose significance levels. The work of Gonzalo and Pitarakis was inspired by the results in Bai (1997) and Bai and Perron (1998) who showed that one can estimate break-points in a multiple break model consistently even when

\[1\) The threshold parameters constitute the nuisance parameters in the TAR case.
the number of breaks estimated is smaller than the actual number of breaks.

Applying model selection criteria or sequential likelihood ratio testing to the present problem requires estimation of models with both \( m \) and \( m + 1 \) thresholds. This may not be considered desirable because the larger model is not identified when the smaller model is true. Another potential difficulty with the approach based on information criterion is that implied significance level of the test of testing the model with \( m \) against one with \( m + 1 \) thresholds (a comparison with two nested models using a model selection criterion is equivalent to a likelihood ratio test) may vary substantially with the size of the smaller model. On the other hand, the user of sequential likelihood ratio tests is, at least in theory, in full control of the significance level of each test in the sequence. A potential disadvantage of Hansen's tests compared to the GP approach is that they require a substantial computational effort. Besides, GP argue that it is not clear whether or not the sequential approach using these tests can be extended to models with more than two regimes. Some simulation results in this chapter illustrate this concern.

The purpose of this chapter is to propose a sequential model selection procedure consisting of a sequence of misspecification tests in which a model with \( m \) thresholds is tested against one with \( m + 1 \) thresholds. Important features of this method are that standard statistical inference is used in the sequential selection of the number of thresholds and that the modeller has a reasonable if not full control of the significance level of each test. If the true model is a switching regression or a threshold autoregressive one, no claims about asymptotic properties of our tests can be made. Nevertheless, we do claim to have an approximate idea of what the significance levels of the tests in finite samples are. Assuming that the switching regression or threshold autoregressive model under consideration has a fixed number of thresholds, the model selection problem at hand is a finite-sample problem. Therefore, it is sufficient to require that the procedure works in a satisfactory fashion in small and moderate samples. Our simulation experiments suggest that this is indeed the case. Another advantage of our procedure is that it is computationally simple and, as a by-product, yields accurate estimates of the threshold parameters of the TAR model. At each stage only the smaller model is estimated, so that the complication of estimating at least one model that is too large is minimized.

The chapter is organized as follows. Section 1.2 provides the motivation for our procedure and contains a brief overview of smooth transition regression (STR) models on which our technique is based. The technique itself is presented in Section 1.3. Section 1.4 contains a simulation study in which our
procedure is compared both with the approach of Hansen (1999a) and the one in GP. An empirical application based on the sunspots numbers series can be found in Section 1.5, and Section 1.6 contains final remarks.

### 1.2 Smooth transition regression model

The general idea underlying our procedure is quite old. Goldfeld and Quandt (1972, pp. 263–264; 1973) considered the estimation of parameters in the switching regression model and pointed out that discontinuity of the log-likelihood complicates the estimation. Their suggestion was to replace the sudden switch or threshold by a smooth transition. This removes the discontinuity, and the parameters of the resulting smooth transition regression model can be estimated by conditional maximum likelihood, using an appropriate iterative algorithm.

In this chapter we will apply the same idea - approximation of sudden switches by smooth transitions - to the regime selection problem. That allows us to use standard inference in determining the number of regimes in a TAR model.

A classical logistic STR (LSTR) model for $y_t$ is defined as follows:

$$y_t = x_t' \beta_0 + x_t' \beta_1 G_{1t} + \epsilon_t, \quad t = 1, \ldots, T, \quad (1.1)$$

where $x_t = (1, x_{1t}, x_{2t}, \ldots, x_{kt})'$ is a $((k+1) \times 1)$ vector of explanatory variables, $\beta_0$ and $\beta_1$ are $((k+1) \times 1)$ parameter vectors and $\{\epsilon_t\}$ is a sequence of independent, identically distributed normal errors with zero mean and variance $\sigma^2$. The transition function $G_{1t}$ in (1.1) is defined as follows:

$$G_{1t} = G_1(s_t; \gamma_1, c_1) = (1 + \exp\{-\gamma_1(s_t - c_1)\})^{-1}, \quad \gamma_1 > 0. \quad (1.2)$$

As $\gamma_1 \to \infty$ in (1.2), the logistic $G_{1t}$ function approaches the indicator function $I[s_t > c_1]$ and the LSTR model becomes a switching regression (SR) or, in the univariate case, a TAR model with two regimes. The parameter $c_1$ is then the switch or threshold parameter. Thus the STR model (1.1) with (1.2) is a reasonable approximation to the SR model when $\gamma_1$ is sufficiently large.

Analogously, we can approximate a multiple-threshold model with a Multiple LSTR (MLSTR) model. For example, an MLSTR model with two transitions has the form

$$y_t = x_t' \beta_0^* + x_t' \beta_1^* G_{1t} + x_t' \beta_2^* G_{2t} + \epsilon_t, \quad (1.3)$$

where the transition function $G_{2t} = G_2(s_t; \gamma_2, c_2)$ is again defined as in (1.2). For the purposes of this chapter we set $\gamma_1 = \gamma_2 = \gamma$. 
To illustrate how MLSTR model (1.3) mimics the three-regime TAR model, we reparameterize (1.3) as follows:

\[ y_t = x_t' \beta_1 (1 - G_{1t}) + x_t' \beta_2 (G_{1t} - G_{2t}) + x_t' \beta_3 G_{2t} + \varepsilon_t. \]  

(1.4)

Letting \( \gamma \to \infty \) we get a piecewise linear form. Figure 1.1 depicts the three regimes created by \( G_{1t} \) and \( G_{2t} \) in (1.4), when \( \gamma = 200, c_1 = 0.3, c_2 = 0.6 \) and \( s_t = t/T \).

![Figure 1.1: Three regimes](image)

Rearranging the terms in (1.4) we obtain

\[ y_t = x_t' \beta_1 + x_t' (\beta_2 - \beta_1) G_{1t} + x_t' (\beta_3 - \beta_2) G_{2t} + \varepsilon_t, \]  

or more generally, in case of \( (m + 1) \) regimes:

\[
\begin{align*}
  y_t &= x_t' \beta_1 + \sum_{j=2}^{m+1} x_t' (\beta_j - \beta_{j-1}) G_{j-1,t} + \varepsilon_t \\
  &= x_t' \beta_1 (1 - G_{1t}) + \sum_{j=2}^{m} x_t' \beta_j (G_{j-1,t} - G_{jt}) + x_t' \beta_{m+1} G_{mt} + \varepsilon_t.
\end{align*}
\]

Suppose now that the true model is a TAR model with two thresholds. We can approximate this model by the STAR model (1.5) where \( \gamma \) is large and known. Suppose, however, that we estimate (1.1) with \( \gamma \) large and known using maximum likelihood. How does this misspecification affect our threshold parameter estimate? Analogously to GP we argue that underspecification of the number of regimes affects the estimates of slope coefficients \( \beta_i \), but hardly those of \( c_i \). In other words, in our three-regime example \( c_1 \) can be estimated
1.3 Smooth transition approach

reasonably accurately even when the number of regimes is misspecified by ignoring $G_2t$ in (1.5).

In the Appendix we show that the average Hessian used as an estimate of the covariance matrix of the average score function, is nearly block-diagonal when $\gamma$ is large. This means that location parameters can be estimated practically independently of each other, which is necessary for our procedure to work. We also provide simulation evidence from three different three-regime TAR models, showing that when estimating only a two-regime model, the $c_1$ estimate will be (very close to) one of the true thresholds.

1.3 Smooth transition approach

In this section we follow GP and consider the univariate TAR model. Our strategy is, however, applicable to switching regression models as well. The starting-point is that the true model is either a linear model or a TAR model (but possibly with just one threshold), so the first choice is between $m = 0$ (linearity) and $m = 1$ (two regimes). As a whole, the procedure works as follows:

1. Test linearity of (1.1) (i.e. \( \gamma = 0 \) in \( G_{1t}(y_{t-d}; \gamma, c_1) \)), where \( x_t = (1, \bar{x}_t)' = (1, y_{t-1}, \ldots, y_{t-k})' \). In order to circumvent the identification problem approximate the transition function by its Taylor expansion around $\gamma = 0$. The first-order approximation can be written as $T_1 = \delta_0 + \delta_1 s_t + R_1(\gamma, c; s_t)$, where $R_1$ is the remainder and $\delta_0$ and $\delta_1$ are constants. Substituting $T_1$ for $G_{1t}$ in (1.1) and reparameterizing yields

\[
y_t = x_t' \theta_0 + (x_t s_t)' \theta_1 + \varepsilon_t^*,
\]

where $\varepsilon_t^* = \varepsilon_t + (x_t' \beta_1) R_1(\gamma, c; s_t)$. The parameter vector $\theta_1 = \gamma \tilde{\theta}_1$, where $\theta_1 \neq 0$, and thus our null hypothesis of linearity in (1.1) implies $H_0' : \theta_1 = 0$ in (1.6). Under $H_0' : \varepsilon_t^* = \varepsilon_t$. For further discussion of the test, see, for example, Luukkonen, Saikkonen, and Teräsvirta (1988) or Teräsvirta (1998). The resulting test has power against STAR but also against TAR ($\gamma \to \infty$) models. Under the null hypothesis and the assumption $Ey_t^4 < \infty$, the test statistic has an asymptotic $\chi^2$-distribution with $k + 1$ degrees of freedom, and following the suggestions in earlier papers an $F$-approximation to it is recommended. The test can be carried out in three stages using just linear regressions:
(a) Regress \( y_t \) on \( x_t \) and compute the residual sum of squares
\[
SSR_0 = \frac{1}{T} \sum_{t=1}^{T} \tilde{\varepsilon}^2_t.
\]

(b) Regress \( \tilde{\varepsilon}_t \) (or \( y_t \)) on \( x_t \) and \( \tilde{x}_t s_t \), and compute the residual sum of
squares \( SSR_1 = \frac{1}{T} \sum_{t=1}^{T} \tilde{\varepsilon}^2_t \).

(c) Compute
\[
F = \frac{(SSR_0 - SSR_1)/k}{SSR_1/(T - 2k - 1)}
\]
that is approximately \( F_{k,T-2k-1} \) distributed under the null of
linearity.

2. If the null hypothesis is rejected at a predetermined significance level \( \alpha \),
estimate the parameters of (1.1) by nonlinear least squares fixing \( \gamma \) at a
sufficiently high but finite value. Then the STAR model approximates
a TAR model with \( m = 1 \) and threshold value \( c_1 \) while the transition
function still retains its smooth character (as a result the likelihood
function is well-behaved).

3. If LSTAR model (1.1) with fixed \( \gamma \) is accepted, test it against a Multiple
LSTAR model (1.5) with transition function \( G_{2t} \). This is done by making
use of the first-order Taylor expansion of the transition function \( G_{2t} \),
see, for example, Eitrheim and Teräsvirta (1996) or Teräsvirta (1998).
Accept (1.5) if the null hypothesis is rejected at significance level \( \tau \alpha \),
\( 0 < \tau < 1 \). Reducing the significance level compared to the preceding
test favours parsimonious models. Choosing \( \tau \) is left to the modeller: in
the simulations we set \( \tau = 1/2 \). Starting-values for the estimation may
be obtained by using the estimates of \( \beta_1, (\beta_2 - \beta_1) \) and \( c_1 \). The starting-
value for \( c_2 \) is obtained by a one-dimensional grid search over a possible
set of candidates. This also yields an initial value for \( \beta_3 \). The estimated
model is then tested for another regime. The sequential estimation and
testing is continued until the first acceptance of a null hypothesis. This
yields the specification for the final model.

4. Estimate consistently the parameters of the final TAR model by con-
ditional least squares (Chan, 1993) or using a dynamic programming
algorithm, see Bai and Perron (2003), to estimate the thresholds con-
sistently before estimating the remaining parameters by least squares.
1.4 Simulation study

The test can also be constructed using the third-order Taylor approximation of $G_t$. That variant of the test should be more powerful in cases where the process is returning back to its original level after the second threshold, for example. For our STAR-approximation procedure to work in the univariate case we need to assume that $\epsilon_t$ are iid, the transition variable is weakly stationary, and the $2(n+1)$-th moment, where $n$ is the order of Taylor expansion, of $y_t$ exist. It may also be mentioned that the tests can be robustified against heteroskedasticity following Wooldridge (1990).

It should be noted that when a TAR model with $m$ thresholds is tested against one with $m + 1$ thresholds, $m \geq 1$, using our test, the asymptotic significance level of the test is unknown. This is the case because the null model is a smooth transition approximation to the null threshold autoregressive model. In testing linearity, however, the asymptotic significance level is known because in that case the null model is not an approximation.

Lack of asymptotic inference may be viewed as a disadvantage, but then, the model selection problem is always a finite-sample problem. Finite sample properties of our technique will be investigated by simulation. The advantages of the STAR-approach are that the tests are computationally simple and that one obtains remarkably accurate values for the threshold parameters even when some of them lie near the smallest or largest observation in the sample.

1.4 Simulation study

In this section, the small sample performance of the three strategies will be compared by simulation. Choosing between two nested models using an appropriate model selection criterion is equivalent to carrying out the likelihood ratio test, and in some situations the significance level of the model selection criterion based test can be worked out; see, for instance, Teräsvirta and Mellin (1986). In the present case that is not possible even asymptotically because of the identification problem previously mentioned. It is, however, possible to obtain an idea of the empirical size of these tests by simulation. In what follows we shall investigate both the size of these procedures and their success in finding the correct number of regimes.

In all experiments the true (maximum) lag length of the TAR model is assumed known. In practice one would have to determine the appropriate lag length either simultaneously or before determining the number of regimes. Quite often the lag length is selected prior to building a nonlinear model, using a suitable information criterion.
1.4.1 Estimating the empirical size

Following GP we simulate univariate autoregressive models, so the alternative to the linear model is a TAR model. We adopt the AR(1) model considered in GP that has the form

$$y_t = \rho y_{t-1} + \varepsilon_t$$

with $\rho = (0.5, 0.7, 0.9, 1.0)$, where $\{\varepsilon_t\} \sim \text{nid}(0, 1)$. In order to check the effect of the number of lags on the empirical size of the model selection criteria we also simulate a number of AR(4) models

$$\varepsilon_t = (1 - \rho L)(1 - 0.2L)(1 - 0.25L)(1 - 0.34L)y_t$$

$$= (1 - \rho L)(1 - 0.79L + 0.203L^2 - 0.017L^3)y_t$$

with $\rho = (0.5, 0.7, 0.9, 1.0)$, where $\{\varepsilon_t\} \sim \text{nid}(0, 1)$ and $L$ is the lag operator. The idea with (1.8) is to vary the value of the dominant root and, in particular, see what happens when it approaches unity. For $\rho = 1$, the asymptotic distribution theory for testing $\theta_1 = 0$ in (1.6) is no longer valid.

In the following tables AIC, BIC, BIC2 and BIC3 refer, as in GP, to the model selection criteria

$$Q_T(m) = \max_{\varepsilon_1, \ldots, \varepsilon_m} \log \left[ \frac{\hat{\sigma}^2}{\hat{\sigma}^2(c_1, \ldots, c_m)} \right] - \frac{\lambda_T}{T} Km$$

with penalty terms $\lambda_T = 2$, $\lambda_T = \log(T)$, $\lambda_T = 2\log(T)$ and $\lambda_T = 3\log(T)$, respectively. In (1.9), $\hat{\sigma}^2$ is the residual variance in the linear model, $\hat{\sigma}^2(c_1, \ldots, c_m)$ the residual variance of the TAR model with $m$ thresholds, $T$ is the operative number of observations and $K$ is the number of parameters in every regime.

We use, following GP, three different sample sizes $(T = 200, 400, 600)^2$, and three different nominal sizes $\alpha = 0.1, 0.05$ and 0.01, respectively. For each DGP and for every sample size, 2000 Monte Carlo replications are carried out.

The results for the AR(1) model (1.7) appear in Table 1.1. The threshold or transition variable is assumed to be $y_{t-1}$. In (1.7) the intercept is zero, but in practice one would most probably at least tentatively include an intercept in the model. For this reason we assume the intercept to be unknown and a parameter to be estimated from the data. BIC seems to be the only model selection criterion that selects the linear model 4 - 10% of the time, except when $T = 200$ and $\rho = 1$. Both BIC2 and BIC3 point to the correct

\^For every sample size we actually generate $T + 200 + k$ observations and discard the first 200 observations from each sample to minimize the impact of starting-values, and use the $k$ extra observations to construct the autoregressive lags of $y_t$. 

---

\(\hat{\sigma}^2\) for every sample size we actually generate $T + 200 + k$ observations and discard the first 200 observations from each sample to minimize the impact of starting-values, and use the $k$ extra observations to construct the autoregressive lags of $y_t$. 

---
(linear) model with an empirical probability very close to one and thus have an empirical size close to zero. AIC, as GP also stress, does not work well in this set-up, but its performance is reported here for the sake of comparison.

On the contrary, the STAR-approach\textsuperscript{3} has reasonable size properties in the sense that the empirical sizes are rather close to the ones determined from the $F$-distribution unless the root of the lag polynomial is close to unity. The linearity test as a whole is seen to be somewhat conservative in small samples. The asymptotic distribution theory of the test is not valid if the AR process is non-stationary, ($\rho = 1$), which explains the increasing size distortion when $\rho \rightarrow 1$. Hansen's bootstrap-based test has good size properties already at $T = 200$. Even when the AR process contains a unit root, the empirical size of the test is not too far from the nominal one.

The results for the AR(4) model using $y_{t-1}$ as the threshold variable\textsuperscript{4} in Table 1.2 are very different from the ones in Table 1.1 when the BIC-type model selection criteria are concerned. The increase in the penalty term due to the increased lag length has a remarkable effect on the empirical size. It is practically zero already at $T = 200$. From this we can conjecture that the empirical size of the GP procedure for any AR model with an even longer lag would be practically zero for these criteria at the sample sizes GP considered. AIC is still heavily oversized. The linearity test based on the STAR approximation tends to be slightly undersized, but at some parameter combinations it competes with Hansen's bootstrap-based method that is well-sized already in small samples.

### 1.4.2 Simulating TAR models

In order to consider the performance of the three procedures when the true model is a genuine TAR model we simulate two models also included in the simulation study of GP. One of them has two regimes ($m = 1$) and the other one three ($m = 2$). Furthermore, we complete the experiment with yet another TAR model with $m = 1$. This is done to better demonstrate differences among properties of the three regime-selection procedures. These experiments could be called power simulations except for the fact that the empirical sizes of the three procedures differ substantially from each other.

\textsuperscript{3}Throughout Sections 1.4 and 1.5 we report the results for test sequences where the test statistics are based on the first-order Taylor approximation. For DGP-s used in this study the discrepancies between the first-order and third-order Taylor approximation approaches were minor.

\textsuperscript{4}The results using any other lag $y_{t-d}$, $d = 2, 3, 4$, as the threshold variable are very similar to the ones reported here.
The error terms in these simulations are constructed to be standard normal variates. We use three different sample sizes \((T = 200, 400, 800)\), and three different nominal test size sequences \((\alpha, \alpha \tau, \alpha \tau^2, \ldots)\), where \(\alpha = 0.1, 0.05\) and 0.01, respectively, and \(\tau = 1/2\). Our method seems to be robust\(^{5}\) with respect to the choice of \(\tau\). For each DGP and for every sample size, 2000 Monte Carlo replications are carried out.

We begin our STAR-based procedure by testing linearity against (1.1), assuming the transition variable to be known\(^{6}\). If linearity is rejected we proceed to estimate an LSTAR model, fixing the slope parameter \(\gamma = 200\). The approach is robust to the choice of \(\gamma\)\(^{7}\), as long as the logistic function does not deviate much from a step function and the log-likelihood is still well-behaved.

Choosing good starting-values for the optimization algorithm is crucial. We therefore run a grid search over the \([0.1, 0.9]\) interquantile range of the transition variable. This accords with the notion that each regime should contain at least 10% of the total number of observations (see Hansen (1996), Bai and Perron (1998) and GP). After estimating the LSTAR model we look for the second threshold, that is, we test (1.1) against (1.5) as discussed in Section 1.3. If the presence of only a single threshold is rejected, we run another grid search to find a good starting-value for the second location parameter, estimate the corresponding MLSTAR model, and proceed until the first acceptance of null hypothesis.

The GP procedure is applied as in the original paper. The required regime size is 10% of the whole sample and thresholds are estimated sequentially, using the Bai (1997) repartition technique. That means re-estimating the threshold parameters conditionally on the initially estimated ones so that each refined estimate is obtained without an underlying neglected regime. In two threshold case, for instance, the first threshold \(r^{(1)}\) is re-estimated taking the second threshold estimate \(\hat{r}^{(2)}\) as given and \(\hat{r}^{(2)}\) re-estimated taking the refined estimate of \(r^{(1)}\) as given.

When using Hansen’s bootstrap-based method we reduce the significance level \(\alpha\) as in the STAR-based procedure. Because simulating the likelihood ratio statistics in the sequence can be computationally rather burdensome,

---

\(^{5}\)We let \(\tau\) change between 0.1, \ldots, 1.0. The power loss with respect to the highest-power case was about 0.5 – 1 percentage points and never greater than 2.8 percentage points (two thresholds, \(T = 200\)).

\(^{6}\)It is also possible to define a set of potential transition variables, test against each of them and choose the one giving the strongest rejection (lowest \(p\)-value) of linearity.

\(^{7}\)We let \(\gamma = 100, 200, \ldots, 1000\). The largest power loss relative to the maximum, about 2%, occurred at \(T = 800, \gamma = 100\). On the average the loss was about 0.6%.
1.4. Simulation study

we use only 199 model-based bootstrap replications in the application of Hansen’s technique. For finding out the power loss that this implies, we refer to Davidson and MacKinnon (2000) who considered the problem of choosing the number of bootstrap replications in bootstrap-based tests. For the test at the 0.10 level the implied power loss should be less than 1%, for a test at 0.01 level the loss should not be greater than 2.5% – 3%.

DGP1: a single threshold model

We begin by considering a TAR model with a single threshold. The data are generated from the following model in GP:

\[ y_t = \begin{cases} 
-3 + 0.5y_{t-1} - 0.9y_{t-2} + \varepsilon_t & y_{t-2} \leq 1.5 \\
2 + 0.3y_{t-1} + 0.2y_{t-2} + \varepsilon_t & y_{t-2} > 1.5. 
\end{cases} \]  

(1.10)

In Table 1.3 we report the selection frequencies for DGP1 using GP-procedure, i.e. adjusted numbers for Table 6 in GP (page 340). The high frequency for choosing a three-regime model instead of a two-regime model in their original table is due to a slight error in their computer code related to applying the 10% minimum regime size rule mentioned above. The second threshold is often found so close to the first one that there are not sufficiently many observations within the thresholds to make a genuine regime. When the 10% rule is properly applied, the results improve, and in large samples a correct decision is made in over 97% of the occasions.

Results for DGP1 using Hansen’s bootstrap and STAR-approach are reported in Table 1.4. The bootstrap procedure performs about as well as the information criterion based ones. The results for STAR-approach show that the linear model is chosen surprisingly often, about 9% of the time even for \( T = 400 \). The reason is that the \( \hat{c}_1 \) obtained by numerical optimization sometimes falls outside the [.1, .9] interquantile range and is ignored. Picking a “good” starting-value inside this range does not help when the actual true threshold value lies out in either of the tails of the empirical density of the threshold variable.

This situation is worth a further comment. GP remarked that DGP1 generates realizations that on the average have approximately the same number of observations in each regime. The true threshold value in our experiment is indeed close to the median of the samples (the average quantile of the threshold value over the replications for any of the three sample sizes is about 0.53). At the same time, in a single sample the true threshold value 1.5 can be very far out in the tails of the empirical distribution, in small samples in particular. Figure 1.2 shows the frequencies with which the observed deciles of the
empirical distribution cover the true threshold value. Decile "0" contains the cases where 1.5 is less than the value of the smallest observation in the sample and decile "11" the cases where the true threshold value exceeds the largest observed value in the sample.

![Graphs showing the distribution of observed deciles for T=200, T=400, and T=800.](image)

**Figure 1.2:** The frequencies with which the observed deciles of the empirical distribution of the threshold variable cover the true threshold value; for \(T = 200\), \(T = 400\) and \(T = 800\).

Consider first the case \(T = 200\). In about 4.5% of the realizations the threshold value 1.5 lies outside the range of the simulated series. Thus, at least for these cases a linear model should be selected. In addition to that, in 22% of the cases the true value falls into the first or the last decile. Whenever our location estimate \(\hat{c}_i\) (even if it happens to be close to 1.5) falls outside the \([1.1, 9]\) interquantile range the decision has been that it does not signal a genuine threshold. The 10% regime size rule thus explains the high frequencies for selecting \(\hat{m} = 0\) with the STAR-approach. Based on this example we can conjecture that the GP as well as Hansen's procedure might therefore be picking up the second or third best option for the threshold value (from the \([1.1, 9]\) range they are restricted to), given that TAR model is preferred to the linear specification. For \(T = 400\), the true value is contained in the first decile or is outside the range about 3.5% and in the last one about 6% of the time. The results are quite similar to the previous case, as the STAR approximation selects the linear model in about 9% of the cases as opposed to 2% for the model selection approach of GP.

The effect of the 10% rule is shown in Table 1.5 where we report the results of the same experiment after relaxing the regime size restriction for the first threshold. We thus allow its value to belong to the first or the last decile of the observed threshold variable, but we still apply the rule to the next thresholds. Now the linear model is chosen less frequently and the majority of the wrong decisions consists of erroneously detecting a second threshold, except for the
smallest sample size. The results are now as good as the ones obtained using Hansen’s procedure and signal another advantage of the STAR-approach: the 10% minimum regime size requirement is not necessary when this technique is applied.

We should also mention a difficulty encountered in generating series by the bootstrap for Hansen’s procedure. When the optimal threshold value is selected from the [.1, .9] interquantile range and it is not close to the true value, the parameter estimates of the two AR models are (sometimes) far from their true values as well. In that case a number of series generated from the estimated model by bootstrap are explosive. In this experiment, such realizations were discarded and new bootstrap samples generated until the number of valid realizations reached 199. As an example, for sample size $T = 400$, we needed to generate extra bootstrap samples in 5% of the cases. The number of explosive bootstrap series varied between 77 and 2045. We also imposed a “maximum 5000 explosive bootstraps allowed” rule. For DGP1 this rule was flexible enough allowing us to obtain 199 valid bootstrap replicates for every Monte Carlo replication at sample sizes $T = 200, 400$. That was no longer the case for $T = 800$, because it was difficult to generate long non-explosive series. There were 9 cases for which 5000 additional bootstraps were not enough and in the worst case only 24 valid bootstrap series were generated. For these 9 cases the empirical distribution of the $F$ statistic was completed by imputing the missing values with the average of existing bootstrapped statistics.

This difficulty may actually be anticipated. Hansen (1999a, pp. 571), when discussing bootstrapping the distribution for the TAR($m = 1$) vs TAR($m = 2$) test statistic, writes: “We do this with some caution, because there has not yet been a demonstration that a bootstrap procedure can properly approximate the sampling distribution of $F_{23}$ under the SETAR(2) null hypothesis.”

In practice, an exploding realization may be taken as a sign of something being wrong with the null model.

**DGP2: multiple threshold model**

The second DGP, also from GP, is a TAR model with three regimes. It has the following form:

$$y_t = \begin{cases} 
2.7 + 0.8y_{t-1} - 0.2y_{t-2} + \varepsilon_t, & y_{t-2} \leq 5 \\
6 + 1.9y_{t-1} - 1.2y_{t-2} + \varepsilon_t, & 5 < y_{t-2} \leq 12 \\
1 + 0.7y_{t-1} - 0.3y_{t-2} + \varepsilon_t, & y_{t-2} > 12,
\end{cases}$$

\[ (1.11) \]

\footnote{In Hansen’s notation SETAR($n$) denotes a model with $n$ regimes, i.e. with $n - 1$ thresholds. Furthermore $F_{23}$ denotes the test statistic for testing a 2-regime model against a 3-regime specification i.e. one threshold vs two thresholds.}
where \( \{\varepsilon_t\} \sim \text{nid}(0, 1) \). In Table 1.6 we report the corrected selection frequencies of Table 7 in GP (page 341). Main tendencies are the same as before in that the number of incorrect decisions is small. The thresholds themselves are estimated with reasonable accuracy; see the Appendix.

The results we obtain by applying the STAR-approach to this double-threshold case are quite similar to results from the experiment with one threshold. Power is good even in moderate samples. It does not, however, seem to increase with the sample size. This is due to the fact that the increasing information about the DGP makes the STAR-approximation with a constant \( \gamma \) become less accurate. This disadvantage can be remedied by making the slope parameter \( \gamma \) an increasing function of the sample size.

In this experiment, the problem of explosive realizations when applying the likelihood ratio test sequentially became very severe. When \( T = 200 \), and when two regimes were tested against three and the "maximum 5000 extra bootstraps" rule was not applied, it took 19252 extra realizations on the average to obtain an empirical distribution based on 199 bootstrap realizations. The maximum number was 490710. The reason for this was that even if one of the thresholds was estimated consistently, merging the two other regimes of the DGP into one (the null model in Hansen's model-based bootstrap) very often led to a highly explosive two-regime model.

The results for \( T = 200 \) can be found in Table 1.7. It appears that the sequential likelihood ratio test procedure does not perform as well as the STAR-approach. Simulating the procedure for \( T > 200 \) is out of the question because of the amount of computations needed to obtain sufficiently many non-explosive realizations. As a whole, one may conclude that the sequential likelihood ratio test procedure may run into problems when the data have been generated by a TAR model with more than two regimes. They can be avoided by making use of the STAR-approximation to the TAR model.

A complementary experiment

GP conclude that overall the BIC criterion displays desirable large sample properties and a reasonably good finite sample behavior. They rightly point out, however, that one should interpret any experimental results with caution since the performance of the criterion depends on the data-generating process. In order to emphasize this feature we complement the experiments in GP by a "real-world" one. The observations are generated by the TAR(2; 10, 2) model in Tong (1990, p. 421), estimated for Wolf's sunspot numbers 1700 - 1979
transformed as in Ghaddar and Tong (1981). The DGP is

\[ y_t = \begin{cases} 
1.89 + 0.86y_{t-1} + 0.08y_{t-2} - 0.32y_{t-3} + 0.16y_{t-4} & \text{if } y_{t-8} \leq 11.93 \\
-0.21y_{t-5} - 0.0005y_{t-6} + 0.19y_{t-7} - 0.28y_{t-8} + 0.20y_{t-9} + 0.01y_{t-10} + \varepsilon_t & \text{if } y_{t-8} > 11.93 \\
4.53 + 1.41y_{t-1} - 0.78y_{t-2} + \varepsilon_t & \text{if } y_{t-8} > 11.93 
\end{cases} \] (1.12)

where \( \{\varepsilon_t\} \sim \text{nid}(0,3.734) \). The variance is a “pooled variance”; see Tong (1990, p. 421).

In this experiment our starting-point is an AR(10) model, which implies that the alternative model is a TAR model with ten lags in every regime. An interesting question arises: should one after rejecting the null hypothesis against the TAR model with two regimes determine the lag length in them before proceeding further (see equation (1.12) where the second regime only contains two lags), but it is not addressed here.

Results for all three procedures can be found in Table 1.8. As may be expected from the size simulations, the BIC-type criteria BIC2 and BIC3 strongly favour the linear AR(10) model. Even BIC does that unless the sample size is large \( T = 800 \). We also report the results when using AIC, for the sake of comparison in such an extreme case. This criterion works better than any BIC for \( T = 200 \), but a question arises: which one of these criteria should one use and when? It can be concluded that Hansen’s procedure is the best one of the three for this DGP. The STAR-approach is less powerful than Hansen’s technique for \( T \leq 400 \) but performs better than the model selection criteria. In this experiment it overestimates the number of regimes less frequently than Hansen’s approach.

1.5 Application

As an empirical example we consider the original time series of Wolf’s sunspot numbers from 1700 – 1979, transformed as in Ghaddar and Tong (1981). The series with 280 observations is depicted in Figure 1.3 and exhibits asymmetric cyclical behaviour. It is a very clear-cut example of a nonlinear time series.

When building a TAR model for the series, the autoregressive lag length \( k \) for every regime is unknown. It is selected from the linear autoregressive model such that there is no error autocorrelation left in the residuals. We apply the Breusch-Godfrey LM test sequentially: \( k \) is increased until the null hypothesis of no error autocorrelation can no longer be rejected at the 5% significance level. This results in \( \hat{k} = 10 \).
Using the STAR-approximation we test linearity of the AR(10) model against all ten lags one at a time. Linearity is rejected in eight cases out of ten at 1% level and the lag 8 as the transition variable gives the strongest rejection. From Table 1.9 it is seen that the sequential procedure suggested in Section 1.3 leads to one threshold. Using Hansen’s procedure with 2000 bootstrap replications we find one or two thresholds, depending on the initial significance level. To apply the information criterion-based procedure of GP we use delay $d = 8$ found previously. The two information criteria with largest penalty terms, BIC2 and BIC3, prefer the linear model, and only BIC1 is able to detect one threshold.

We also consider lagged first differences of $y_t$, $\Delta y_{t-d}$, $d = 1, \ldots, 10$, as possible threshold variables. All methods choose $\Delta y_{t-1}$ to be the threshold variable. The results of selecting the number of regimes can be found in Table 1.9. One threshold is found to be present, with the exception that BIC2 and BIC3 favour a linear model. All estimation methods yield a threshold value close to zero ($\hat{c}_1 \approx 0.8$), which suggests separate regimes for years with positive and ones with negative growth in sunspot intensity.

### 1.6 Final remarks

In this chapter we have developed a simple and computationally feasible method for selecting the number of regimes in a switching regression or threshold autoregressive model.

---

9From Hansen (1999a) it is known that with a homoskedastic model-based bootstrap one would reject the null of a two-regime model, and with a heteroskedastic bootstrap one would not do that.
1.6. Final remarks

As already pointed out the tests in the STAR-approach can be robustified against heteroskedasticity and thus we only have to assume the independence of errors for the procedure to work. In order to apply Hansen’s technique \( \varepsilon_t \) has to be assumed a uniformly square-integrable martingale difference sequence with respect to the natural filtration, the Borel sigma-field \( \mathcal{I}_{t-1} = \sigma(y_{t-1}, y_{t-2}, y_{t-3}, \ldots) \), and \( E\varepsilon_t^2 < \infty \). For the bootstrap one also has to assume that the errors are independent. Gonzalo and Pitarakis make quite general assumptions, requiring \( \varepsilon_t \) to be a real-valued martingale difference sequence with respect to some increasing sequence of sigma fields \( \mathcal{F}_t \) generated by \( \{(x_{j+1}, z_{j+1}, \varepsilon_j), j \leq t\} \), where \( z \) is the threshold variable, and with \( E|\varepsilon_t|^{4r} < \infty \) for some \( r > 1 \). To obtain the limiting distributions of the estimators they make some additional high-level (LLN and FCLT-type) assumptions that exclude integrated processes. GP note that \( T \) times the first component in the right-hand side of (1.9) is the likelihood ratio statistic for testing linearity against a model with \( m \) thresholds. Thus their method can, in principle, accommodate the presence of heteroskedasticity through the use of heteroskedasticity-robust versions of this LR or Wald-type statistic. Obviously the method can be generalized such that it will simultaneously allow for selecting the threshold variable as well.

An obvious conclusion from our simulation experiments is that the results of the sequential approach based on model selection criteria are crucially dependent on the number of lags in the TAR model. Admittedly, the users of this approach do not have to choose significance levels for their tests. But then, they face an equivalent problem in the case of GP, which is the one of choosing an appropriate information criterion.

Hansen’s bootstrap-based LR-type test can be recommended if it is known that the true number of regimes in the TAR or switching regression model does not exceed two and if computational resources are not a problem. If the existence of more than two regimes cannot be excluded \textit{a priori}, the sequential likelihood ratio test approach may not always work properly. Although the threshold parameters in the model are estimated accurately even when the number of regimes is assumed too small, the estimates of the other parameters in such a model may, due to this misspecification, cause difficulties when it comes to constructing the empirical distribution of the next test statistic by bootstrap.

The STAR-approach works well in comparison with the other two approaches. It is somewhat conservative, but its performance in selecting the correct TAR model can be deemed acceptable also when the sequential likelihood ratio test procedure excels, that is, when the true model is either
linear or has two regimes. The technique is computationally simple, and it performs remarkably well even when a true threshold lies outside the [.1, .9] interquantile range of the observed series. One can relax the minimum regime-size requirement and still estimate the threshold parameters quite accurately.

The discussion in this chapter has been restricted to the univariate TAR model, but our technique can also be applied to switching regression models. Besides, it appears that it can be used for determining the number of regimes in the panel threshold regression (PTR) model of Hansen (1999b). This would be done by approximating Hansen’s model by the panel smooth transition regression model introduced in González, Teräsvirta, and van Dijk (2004) and using tests described in that paper to determine the number of regimes in the PTR model.

It also seems possible to apply the procedure to detecting the number of breaks in a linear model. This can be in principle done by letting time be the transition variable in the STR model instead of a random transition variable. This possibility is considered in Chapter 2.
Appendix A: Properties of threshold estimates

The purpose of this appendix is to give an explanation to the outcome that the threshold parameters can be estimated sequentially with reasonable accuracy from smooth transition approximations to the threshold autoregressive model. Because the STAR model is an approximation to the data-generating process, the arguments are merely suggestive and not based on any asymptotic theory. It suffices to study the block corresponding to the location parameters in the average Hessian and show that it is approximately diagonal. From this it follows that sequential estimation of threshold values yields quite accurate estimates because the estimators of the thresholds are approximately independent. This will be demonstrated using the MLSTAR model (1.3) that contains two transitions.

Assume that \( \{\varepsilon_t\}, \ t = 1, \ldots, T, \) is a sequence of identically normally distributed random variables with mean zero and variance \( \sigma^2. \) Then the log-likelihood of the STAR model with two transitions for observation \( t \) is

\[
l_t = a - \frac{1}{2} \ln \sigma^2 - \frac{\varepsilon_t^2}{2\sigma^2}, \tag{1.13}
\]

where \( a \) is a constant, and \( \varepsilon_t = y_t - x'_t \beta_0^* - x'_t \beta_1^* G_{1t} - x'_t \beta_2^* G_{2t} \) with \( G_{it} = (1 + e^{-\gamma(s_t - c_i)})^{-1}. \) Let

\[
I(|s_t - c_i| < \varepsilon_\gamma), \quad i = 1, 2, \tag{1.14}
\]

where \( I(A)=1 \) when \( A \) is true and zero otherwise. Thus

\[
\frac{\partial G_{it}}{\partial c_i} = \gamma \left(1 + e^{-\gamma(s_t - c_i)}\right)^{-2} e^{-\gamma(s_t - c_i)} = \gamma G_{it}(1 - G_{it}). \tag{1.15}
\]

For sufficiently large \( \gamma, \) derivative (1.15) only takes values greater than an arbitrarily small positive constant in a small neighbourhood described by the argument of the indicator function (1.14). In particular, \( \gamma G_{it}(1 - G_{it})|_{s_t = c_i} = \gamma/4. \)

Now, assume \( |c_1 - c_2| > \delta_\gamma, \) where \( \delta_\gamma > 0 \) is such that if \( |s_t - c_1| < \varepsilon_\gamma, \) then \( |s_t - c_2| > \varepsilon_\gamma \) and vice versa, where \( \varepsilon_\gamma > 0. \) Setting \( L_T = \sum_{t=1}^{T} l_t, \) the elements of the block of interest in the average Hessian are

\[
\frac{1}{T} \frac{\partial^2 L_T}{\partial c_i^2} = \frac{1}{\sigma^2} \left( \frac{1}{T} \sum_{t=1}^{T} \left(x'_t \beta_i^*\right)^2 \left(\frac{\partial G_{it}}{\partial c_i}\right)^2 + \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t (x'_t \beta_i^*) \frac{\partial^2 G_{it}}{\partial c_i^2} \right) \tag{1.16}
\]

\[
\approx \frac{1}{\sigma^2} \left( \frac{\gamma^2}{16T} \sum_{t=1}^{T} \left(x'_t \beta_i^*\right)^2 I(|s_t - c_i| < \varepsilon_\gamma) \right) + o(1), \quad i = 1, 2
\]
and

\[ \frac{1}{T} \frac{\partial^2 L_T}{\partial c_1 \partial c_2} = \frac{1}{\sigma^2} \left( \frac{1}{T} \sum_{t=1}^{T} (x_t' \beta_1^*) (x_t' \beta_2^*) \frac{\partial G_{1t}}{\partial c_1} \frac{\partial G_{2t}}{\partial c_2} \right) \]

\[ \approx \frac{1}{\sigma^2} \left( \frac{\gamma^2}{T} \sum_{t=1}^{T} (x_t' \beta_1^*) (x_t' \beta_2^*) I(|s_t - c_1| < \varepsilon_\gamma) I(|s_t - c_2| < \varepsilon_\gamma) \right) \]

\[ = 0 \quad (1.17) \]

because \( I(|s_t - c_1| < \varepsilon_\gamma) I(|s_t - c_2| < \varepsilon_\gamma) = 0 \). As a consequence, the expression (1.16) is of larger order of magnitude than (1.17), and the relevant block of the Hessian is approximately diagonal.

**Simulation evidence**

To verify that our estimates of \( c \) are reasonably accurate when the true number of thresholds is greater than the number of thresholds estimated, consider the DGP2 in our study,

\[
y_t = \begin{cases} 
2.7 + 0.8y_{t-1} - 0.2y_{t-2} + \varepsilon_t, & y_{t-2} \leq 5 \\
6 + 1.9y_{t-1} - 1.2y_{t-2} + \varepsilon_t, & 5 < y_{t-2} \leq 12 \\
1 + 0.7y_{t-1} - 0.3y_{t-2} + \varepsilon_t, & y_{t-2} > 12. \end{cases} \quad (1.18) 
\]

When estimating a model with one threshold, the estimates are distributed as follows:

![Figure 1.4: The first threshold estimate distributions for \( T = 200, T = 400 \) and \( T = 800 \).](image)

The estimates are centered around the true value 12 and the spread of the estimates diminishes when the sample size grows. The same seems to hold for
Appendix A: Properties of threshold estimates

a case where the outer regimes are identical:

\[
y_t = \begin{cases} 
2.7 + 0.8y_{t-1} - 0.2y_{t-2} + \varepsilon_t, & \text{if } y_{t-2} \leq 5 \\
6 + 1.9y_{t-1} - 1.2y_{t-2} + \varepsilon_t, & \text{if } 5 < y_{t-2} \leq 12 \\
2.7 + 0.8y_{t-1} - 0.2y_{t-2} + \varepsilon_t, & \text{if } y_{t-2} > 12.
\end{cases}
\]

(1.19)

Figure 1.5: The first threshold estimate distributions for \( T = 200 \), \( T = 400 \) and \( T = 800 \).

Or alternatively:

\[
y_t = \begin{cases} 
2.7 + 0.8y_{t-1} - 0.2y_{t-2} + \varepsilon_t, & \text{if } y_{t-2} \leq 3 \\
1 + 0.7y_{t-1} - 0.3y_{t-2} + \varepsilon_t, & \text{if } 3 < y_{t-2} \leq 6 \\
2.7 + 0.8y_{t-1} - 0.2y_{t-2} + \varepsilon_t, & \text{if } y_{t-2} > 6.
\end{cases}
\]

(1.20)

Figure 1.6: The first threshold estimate distributions for \( T = 200 \), \( T = 400 \) and \( T = 800 \).
References


Table 1.1: GP-procedure, STAR-approach and Hansen’s bootstrap: The empirical size in per cent based on 2000 replications from model (1.7), using 2000 model-based bootstrap replications in Hansen’s procedure.

<table>
<thead>
<tr>
<th></th>
<th>GP</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AIC BIC BIC2 BIC3 10% 5% 1%</td>
<td>10% 5% 1%</td>
<td>10% 5% 1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T = 200$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho = 0.5$</td>
<td>80.6 8.75 0.10 0.00</td>
<td>8.75 4.20 0.90</td>
<td>10.00 4.75 0.95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho = 0.7$</td>
<td>81.3 9.15 0.05 0.00</td>
<td>7.65 3.35 0.75</td>
<td>10.10 5.10 1.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho = 0.9$</td>
<td>80.9 9.50 0.15 0.00</td>
<td>5.00 2.25 0.25</td>
<td>9.60 4.90 0.95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho = 1.0$</td>
<td>88.8 16.35 0.45 1.05</td>
<td>5.65 2.20 0.30</td>
<td>12.30 6.65 1.60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T = 400$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho = 0.5$</td>
<td>83.1 4.70 0.00 0.00</td>
<td>9.35 3.85 0.70</td>
<td>9.25 4.20 0.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho = 0.7$</td>
<td>82.4 5.60 0.05 0.00</td>
<td>8.00 3.65 0.55</td>
<td>9.80 4.95 1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho = 0.9$</td>
<td>81.8 6.35 0.00 0.00</td>
<td>6.60 3.20 0.10</td>
<td>10.35 5.05 1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho = 1.0$</td>
<td>89.8 9.55 0.20 0.00</td>
<td>5.20 2.45 0.45</td>
<td>10.75 5.15 0.95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T = 600$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho = 0.5$</td>
<td>83.8 4.15 0.00 0.00</td>
<td>9.50 4.80 0.85</td>
<td>9.75 4.75 0.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho = 0.7$</td>
<td>83.3 3.80 0.00 0.00</td>
<td>8.90 4.00 0.70</td>
<td>10.35 4.35 0.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho = 0.9$</td>
<td>83.8 4.65 0.00 0.00</td>
<td>7.40 3.60 0.40</td>
<td>10.75 5.15 0.90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho = 1.0$</td>
<td>89.4 8.35 0.10 0.00</td>
<td>5.25 2.15 0.25</td>
<td>12.10 5.90 1.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1.2: GP-procedure, STAR-approach and Hansen’s bootstrap: The empirical size in per cent based on 2000 replications from AR(4), using 2000 model-based bootstrap replications.

<table>
<thead>
<tr>
<th></th>
<th>GP</th>
<th>10%</th>
<th>5%</th>
<th>1%</th>
<th>10%</th>
<th>5%</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AIC</td>
<td>BIC</td>
<td>BIC2</td>
<td>BIC3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T = 200$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho = 0.5$</td>
<td>61.3</td>
<td>0.15</td>
<td>0.00</td>
<td>0.00</td>
<td>8.55</td>
<td>4.25</td>
<td>0.75</td>
</tr>
<tr>
<td>$\rho = 0.7$</td>
<td>57.3</td>
<td>0.20</td>
<td>0.00</td>
<td>0.00</td>
<td>7.80</td>
<td>4.05</td>
<td>1.20</td>
</tr>
<tr>
<td>$\rho = 0.9$</td>
<td>59.9</td>
<td>0.30</td>
<td>0.00</td>
<td>0.00</td>
<td>6.80</td>
<td>3.10</td>
<td>0.60</td>
</tr>
<tr>
<td>$\rho = 1.0$</td>
<td>66.9</td>
<td>0.20</td>
<td>0.05</td>
<td>0.00</td>
<td>7.00</td>
<td>3.55</td>
<td>0.85</td>
</tr>
<tr>
<td>$T = 400$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho = 0.5$</td>
<td>60.7</td>
<td>0.05</td>
<td>0.00</td>
<td>0.00</td>
<td>8.70</td>
<td>4.95</td>
<td>0.85</td>
</tr>
<tr>
<td>$\rho = 0.7$</td>
<td>59.9</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>9.10</td>
<td>4.50</td>
<td>0.90</td>
</tr>
<tr>
<td>$\rho = 0.9$</td>
<td>61.8</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>7.90</td>
<td>3.80</td>
<td>0.85</td>
</tr>
<tr>
<td>$\rho = 1.0$</td>
<td>67.6</td>
<td>0.10</td>
<td>0.00</td>
<td>0.00</td>
<td>7.35</td>
<td>3.30</td>
<td>0.60</td>
</tr>
<tr>
<td>$T = 600$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho = 0.5$</td>
<td>62.0</td>
<td>0.05</td>
<td>0.00</td>
<td>0.00</td>
<td>10.00</td>
<td>4.85</td>
<td>0.70</td>
</tr>
<tr>
<td>$\rho = 0.7$</td>
<td>62.4</td>
<td>0.05</td>
<td>0.00</td>
<td>0.00</td>
<td>9.60</td>
<td>4.60</td>
<td>0.75</td>
</tr>
<tr>
<td>$\rho = 0.9$</td>
<td>62.8</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>9.25</td>
<td>4.30</td>
<td>0.95</td>
</tr>
<tr>
<td>$\rho = 1.0$</td>
<td>69.0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>6.55</td>
<td>2.70</td>
<td>0.50</td>
</tr>
</tbody>
</table>
Table 1.3: Adjusted (10% rule applied properly) Table 6 of GP: Selection frequencies for DGP1, \( m = 1 \).

<table>
<thead>
<tr>
<th>( T = 200 )</th>
<th>( \hat{m} = 0 )</th>
<th>( \hat{m} = 1 )</th>
<th>( \hat{m} \geq 2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIC</td>
<td>7.75</td>
<td>90.70</td>
<td>1.55</td>
</tr>
<tr>
<td>BIC2</td>
<td>9.60</td>
<td>90.40</td>
<td>0.00</td>
</tr>
<tr>
<td>BIC3</td>
<td>10.40</td>
<td>89.60</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( T = 400 )</th>
<th>( \hat{m} = 0 )</th>
<th>( \hat{m} = 1 )</th>
<th>( \hat{m} \geq 2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIC</td>
<td>1.30</td>
<td>97.95</td>
<td>0.75</td>
</tr>
<tr>
<td>BIC2</td>
<td>1.75</td>
<td>98.25</td>
<td>0.00</td>
</tr>
<tr>
<td>BIC3</td>
<td>2.30</td>
<td>97.70</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( T = 800 )</th>
<th>( \hat{m} = 0 )</th>
<th>( \hat{m} = 1 )</th>
<th>( \hat{m} \geq 2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIC</td>
<td>0.00</td>
<td>99.70</td>
<td>0.30</td>
</tr>
<tr>
<td>BIC2</td>
<td>0.05</td>
<td>99.95</td>
<td>0.00</td>
</tr>
<tr>
<td>BIC3</td>
<td>0.05</td>
<td>99.95</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 1.4: STAR-approach and Hansen’s bootstrap: Selection frequencies for DGP1, \( m = 1 \).

<table>
<thead>
<tr>
<th>( T = 200 )</th>
<th>STAR</th>
<th>BOOTSTRAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha = 0.10 )</td>
<td>21.10</td>
<td>76.30</td>
</tr>
<tr>
<td>( \alpha = 0.05 )</td>
<td>22.00</td>
<td>76.65</td>
</tr>
<tr>
<td>( \alpha = 0.01 )</td>
<td>23.50</td>
<td>76.05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( T = 400 )</th>
<th>STAR</th>
<th>BOOTSTRAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha = 0.10 )</td>
<td>8.65</td>
<td>89.25</td>
</tr>
<tr>
<td>( \alpha = 0.05 )</td>
<td>8.80</td>
<td>90.25</td>
</tr>
<tr>
<td>( \alpha = 0.01 )</td>
<td>9.10</td>
<td>90.75</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( T = 800 )</th>
<th>STAR</th>
<th>BOOTSTRAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha = 0.10 )</td>
<td>1.20</td>
<td>97.70</td>
</tr>
<tr>
<td>( \alpha = 0.05 )</td>
<td>1.20</td>
<td>98.30</td>
</tr>
<tr>
<td>( \alpha = 0.01 )</td>
<td>1.20</td>
<td>98.75</td>
</tr>
</tbody>
</table>
Table 1.5: STAR-approach: Selection frequencies for DGP1 when not applying the 10% regime-size rule.

<table>
<thead>
<tr>
<th>$T = 200$</th>
<th>$\hat{m} = 0$</th>
<th>$\hat{m} = 1$</th>
<th>$\hat{m} \geq 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha = 0.10$</td>
<td>8.65</td>
<td>87.20</td>
<td>4.15</td>
</tr>
<tr>
<td>$\alpha = 0.05$</td>
<td>10.05</td>
<td>87.80</td>
<td>2.15</td>
</tr>
<tr>
<td>$\alpha = 0.01$</td>
<td>12.05</td>
<td>87.35</td>
<td>0.60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$T = 400$</th>
<th>$\hat{m} = 0$</th>
<th>$\hat{m} = 1$</th>
<th>$\hat{m} \geq 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha = 0.10$</td>
<td>1.75</td>
<td>94.20</td>
<td>4.05</td>
</tr>
<tr>
<td>$\alpha = 0.05$</td>
<td>1.95</td>
<td>96.10</td>
<td>1.95</td>
</tr>
<tr>
<td>$\alpha = 0.01$</td>
<td>2.45</td>
<td>97.15</td>
<td>0.40</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$T = 800$</th>
<th>$\hat{m} = 0$</th>
<th>$\hat{m} = 1$</th>
<th>$\hat{m} \geq 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha = 0.10$</td>
<td>0.00</td>
<td>96.30</td>
<td>3.70</td>
</tr>
<tr>
<td>$\alpha = 0.05$</td>
<td>0.00</td>
<td>98.15</td>
<td>1.85</td>
</tr>
<tr>
<td>$\alpha = 0.01$</td>
<td>0.05</td>
<td>99.45</td>
<td>0.50</td>
</tr>
</tbody>
</table>
**Table 1.6**: Adjusted (10% rule applied properly) Table 7 of GP: Selection frequencies for DGP2, $m = 2$.

<table>
<thead>
<tr>
<th>$T = 200$</th>
<th>$\hat{m} \leq 1$</th>
<th>$\hat{m} = 2$</th>
<th>$\hat{m} \geq 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIC</td>
<td>0.00</td>
<td><strong>94.20</strong></td>
<td>5.80</td>
</tr>
<tr>
<td>BIC2</td>
<td>0.00</td>
<td><strong>100.00</strong></td>
<td>0.00</td>
</tr>
<tr>
<td>BIC3</td>
<td>0.00</td>
<td><strong>100.00</strong></td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$T = 400$</th>
<th>$\hat{m} \leq 1$</th>
<th>$\hat{m} = 2$</th>
<th>$\hat{m} \geq 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIC</td>
<td>0.00</td>
<td><strong>97.55</strong></td>
<td>2.45</td>
</tr>
<tr>
<td>BIC2</td>
<td>0.00</td>
<td><strong>100.00</strong></td>
<td>0.00</td>
</tr>
<tr>
<td>BIC3</td>
<td>0.00</td>
<td><strong>100.00</strong></td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$T = 800$</th>
<th>$\hat{m} \leq 1$</th>
<th>$\hat{m} = 2$</th>
<th>$\hat{m} \geq 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIC</td>
<td>0.00</td>
<td><strong>98.70</strong></td>
<td>1.30</td>
</tr>
<tr>
<td>BIC2</td>
<td>0.00</td>
<td><strong>100.00</strong></td>
<td>0.00</td>
</tr>
<tr>
<td>BIC3</td>
<td>0.00</td>
<td><strong>100.00</strong></td>
<td>0.00</td>
</tr>
</tbody>
</table>

**Table 1.7**: Selection frequencies for DGP2, $m = 2$, STAR-approach and Hansen’s bootstrap.

<table>
<thead>
<tr>
<th>$T = 200$</th>
<th>$\hat{m} \leq 1$</th>
<th>$\hat{m} = 2$</th>
<th>$\hat{m} \geq 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha = 0.10$</td>
<td>0.40</td>
<td><strong>97.80</strong></td>
<td>1.80</td>
</tr>
<tr>
<td>$\alpha = 0.05$</td>
<td>0.90</td>
<td><strong>98.05</strong></td>
<td>1.05</td>
</tr>
<tr>
<td>$\alpha = 0.01$</td>
<td>4.90</td>
<td><strong>94.95</strong></td>
<td>0.15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$T = 400$</th>
<th>$\hat{m} \leq 1$</th>
<th>$\hat{m} = 2$</th>
<th>$\hat{m} \geq 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha = 0.10$</td>
<td>0.00</td>
<td><strong>97.55</strong></td>
<td>2.45</td>
</tr>
<tr>
<td>$\alpha = 0.05$</td>
<td>0.00</td>
<td><strong>99.00</strong></td>
<td>1.00</td>
</tr>
<tr>
<td>$\alpha = 0.01$</td>
<td>0.00</td>
<td><strong>99.75</strong></td>
<td>0.25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$T = 800$</th>
<th>$\hat{m} \leq 1$</th>
<th>$\hat{m} = 2$</th>
<th>$\hat{m} \geq 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha = 0.10$</td>
<td>0.00</td>
<td><strong>97.40</strong></td>
<td>2.60</td>
</tr>
<tr>
<td>$\alpha = 0.05$</td>
<td>0.00</td>
<td><strong>98.45</strong></td>
<td>1.55</td>
</tr>
<tr>
<td>$\alpha = 0.01$</td>
<td>0.00</td>
<td><strong>99.75</strong></td>
<td>0.25</td>
</tr>
</tbody>
</table>
Table 1.8: Selection frequencies for model (1.12), \( m = 1 \), for four information criterion-based methods, for the STAR-based approach and for the homoskedastic model-based bootstrap (denoted by subscript \( H \)), using starting-significance levels \( \alpha = 0.10 \), \( \alpha = 0.05 \), \( \alpha = 0.01 \).

<table>
<thead>
<tr>
<th>( T )</th>
<th>( \hat{m} = 0 )</th>
<th>( \hat{m} = 1 )</th>
<th>( \hat{m} \geq 2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>0.60</td>
<td><strong>63.00</strong></td>
<td>36.40</td>
</tr>
<tr>
<td>BIC</td>
<td>71.10</td>
<td><strong>28.90</strong></td>
<td>0.00</td>
</tr>
<tr>
<td>BIC2</td>
<td>100.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>BIC3</td>
<td>100.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>( \alpha = 0.10 )</td>
<td>26.90</td>
<td><strong>70.00</strong></td>
<td>3.10</td>
</tr>
<tr>
<td>( \alpha = 0.05 )</td>
<td>34.30</td>
<td><strong>64.75</strong></td>
<td>0.95</td>
</tr>
<tr>
<td>( \alpha = 0.01 )</td>
<td>51.75</td>
<td><strong>48.05</strong></td>
<td>0.20</td>
</tr>
<tr>
<td>( \alpha_H = 0.10 )</td>
<td>14.35</td>
<td><strong>80.90</strong></td>
<td>4.75</td>
</tr>
<tr>
<td>( \alpha_H = 0.05 )</td>
<td>17.45</td>
<td><strong>80.25</strong></td>
<td>2.30</td>
</tr>
<tr>
<td>( \alpha_H = 0.01 )</td>
<td>26.50</td>
<td><strong>73.05</strong></td>
<td>0.45</td>
</tr>
<tr>
<td>( T = 400 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>0.05</td>
<td><strong>64.55</strong></td>
<td>35.40</td>
</tr>
<tr>
<td>BIC</td>
<td>18.80</td>
<td><strong>81.20</strong></td>
<td>0.00</td>
</tr>
<tr>
<td>BIC2</td>
<td>95.80</td>
<td>4.20</td>
<td>0.00</td>
</tr>
<tr>
<td>BIC3</td>
<td>100.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>( \alpha = 0.10 )</td>
<td>9.15</td>
<td><strong>87.05</strong></td>
<td>3.80</td>
</tr>
<tr>
<td>( \alpha = 0.05 )</td>
<td>10.35</td>
<td><strong>87.55</strong></td>
<td>2.55</td>
</tr>
<tr>
<td>( \alpha = 0.01 )</td>
<td>14.20</td>
<td><strong>85.30</strong></td>
<td>0.50</td>
</tr>
<tr>
<td>( \alpha_H = 0.10 )</td>
<td>1.85</td>
<td><strong>93.05</strong></td>
<td>5.10</td>
</tr>
<tr>
<td>( \alpha_H = 0.05 )</td>
<td>2.65</td>
<td><strong>95.00</strong></td>
<td>2.35</td>
</tr>
<tr>
<td>( \alpha_H = 0.01 )</td>
<td>4.45</td>
<td><strong>94.95</strong></td>
<td>0.60</td>
</tr>
<tr>
<td>( T = 800 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>0.00</td>
<td><strong>65.80</strong></td>
<td>34.20</td>
</tr>
<tr>
<td>BIC</td>
<td>1.60</td>
<td><strong>98.40</strong></td>
<td>0.00</td>
</tr>
<tr>
<td>BIC2</td>
<td>21.05</td>
<td><strong>78.95</strong></td>
<td>0.00</td>
</tr>
<tr>
<td>BIC3</td>
<td>91.45</td>
<td>8.55</td>
<td>0.00</td>
</tr>
<tr>
<td>( \alpha = 0.10 )</td>
<td>5.00</td>
<td><strong>90.70</strong></td>
<td>4.30</td>
</tr>
<tr>
<td>( \alpha = 0.05 )</td>
<td>5.05</td>
<td><strong>92.45</strong></td>
<td>2.50</td>
</tr>
<tr>
<td>( \alpha = 0.01 )</td>
<td>5.20</td>
<td><strong>94.30</strong></td>
<td>0.50</td>
</tr>
<tr>
<td>( \alpha_H = 0.10 )</td>
<td>0.00</td>
<td><strong>94.70</strong></td>
<td>5.30</td>
</tr>
<tr>
<td>( \alpha_H = 0.05 )</td>
<td>0.00</td>
<td><strong>97.45</strong></td>
<td>2.55</td>
</tr>
<tr>
<td>( \alpha_H = 0.01 )</td>
<td>0.00</td>
<td><strong>99.35</strong></td>
<td>0.65</td>
</tr>
</tbody>
</table>
Table 1.9: Results of sequential model selection procedures. Here $\hat{d}$ denotes the estimated delay defining the threshold variable and $\hat{m}$ is the estimated number of thresholds. Asterisk (*) indicates the cases where the threshold variable was assumed known in advance.

<table>
<thead>
<tr>
<th></th>
<th>$y_{t-d}$</th>
<th>$\Delta y_{t-d}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{d}$</td>
<td>$\hat{m}$</td>
</tr>
<tr>
<td>BIC1</td>
<td>8*</td>
<td>1</td>
</tr>
<tr>
<td>BIC2</td>
<td>8*</td>
<td>0</td>
</tr>
<tr>
<td>BIC3</td>
<td>8*</td>
<td>0</td>
</tr>
<tr>
<td>$\alpha = 0.10$</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>$\alpha = 0.05$</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>$\alpha = 0.01$</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>$\alpha_H = 0.10$</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>$\alpha_H = 0.05$</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>$\alpha_H = 0.01$</td>
<td>8</td>
<td>1</td>
</tr>
</tbody>
</table>
Chapter 2

Determining the Number of Breaks in a Piecewise Linear Regression Model Using Multiple Smooth Transition Regression
2.1 Introduction

Models with structural breaks (SB) have been of interest to many researchers for at least the last four decades. The model builder faces several problems: first one has to test for a structural change in the parameters of the model and then estimate both the number and the location of breaks.

Most of the work in this area of research has been related to the case of detecting and estimating a single break. See Chow (1960), Andrews (1993), and Bai, Lumsdaine, and Stock (1998), among others. The questions related to multiple structural changes have received somewhat less attention. Early works include Yao (1988) and Liu, Wu, and Zidek (1997) who advocated the use of the (modified) Bayesian Information Criterion and showed that the number of breaks can be estimated consistently (at least for a normal sequence of random variables with shifts in mean). In Bai (1997) it was shown that one can consistently estimate break-points, one-by-one, in a multiple break model even when the number of breaks estimated is smaller than the actual number of breaks. He also proposed a simple sequential procedure for consistently estimating the number of breaks. In a seminal paper Bai and Perron (1998) proved consistency of the estimators of the break dates, provided tests for multiple structural changes and constructed confidence intervals for the break dates. Last, but not least, they also proposed several methods (one of which is purely sequential) for determining the number of breaks and efficient algorithms for computing the estimates. In two companion papers, see Bai and Perron (2001, 2003a), the authors considered practical matters related to the methods proposed in Bai and Perron (1998): such as the behaviour of estimators and tests in finite samples, and comparisons between different methods for determining the number of breaks. Since all the tests considered have nonstandard distributions, Bai and Perron (2003b) also provided asymptotic critical values for a set of possible specifications (nominal level $\alpha = \{0.10, 0.05, 0.025, 0.01\}$, the minimum relative regime size $\epsilon_R = \{0.05, 0.10, 0.15, 0.20, 0.25\}$ and the number of regressors whose parameters are allowed to vary across regimes $q = 1, \ldots, 10$). In coming sections, we refer to papers by Bai and Perron providing the theory and simulations results, and to the methodology in general, as BP.

More recently, Prodan (2003) proposed a new procedure, including a restricted version designed to detect trend reversions, for choosing the number of breaks. This method is based on a sequence of likelihood ratio-type tests of Bai (1999) for which the critical values have to be bootstrapped.

The main limitation of the current implementation of Bai and Perron sequential procedure is that critical values exist for a restricted number of
combinations only (four significance levels, five regime sizes, up to ten regressors). If one wants to test for breaks, say, in 11 monthly dummies, then one would have to simulate the critical values. Furthermore, as documented in Prodan (2003), the asymptotic critical values obtained under the null of independent and identically distributed errors might be inadequate for relatively short (in her simulations $T = 125$) but persistent series, which in turn can cause severe size distortions. The main limitation of the procedure proposed by Prodan is that it is rather time-consuming since the critical values have to be bootstrapped for every test in the sequence.

In this chapter we propose an alternative sequential procedure for determining the number of breaks in a structural break model. The technique itself is based on a sequence of misspecification tests of (multiple) Smooth Transition Regression (STR) models where a model with $m$ breaks (transitions) is tested against one with $m + 1$ breaks. Its advantages include the fact that standard statistical inference applies and the modeller has approximate control over the significance level of each test. No claims about the asymptotic properties of our procedure can be made, however, because in the present context the STR model is merely an approximation to the true switching regression data-generating process. Nevertheless, in practice the model (or break) selection problem is a finite-sample problem. Therefore it will be sufficient to show that the procedure works in a satisfactory fashion in small and moderate samples. Furthermore, our technique is easy to implement, it imposes no restrictions on the number of regressors, (approximate) significance levels of individual tests or minimum regime size. It can be applied to situations where all parameters are assumed to change over time (pure structural change model) and to the ones in which just a subset of parameters is subject to change (partial structural change model). The procedure is analogous to the one suggested in Chapter 1 for finding the number of regimes in a threshold autoregressive model.

The plan of the chapter is as follows. Section 2.2 gives a short overview of STR models. Section 2.3 describes our method step-by-step. Section 2.4 contains the results of a simulation study. The size and power properties of our method are discussed and compared to results in Bai and Perron (2001). Section 2.5 concludes.

2.2 Smooth transition regression model

The general idea underlying our procedure is quite old. Goldfeld and Quandt (1972, pp. 263–264) considered the estimation of parameters in the switching
regression model and pointed out that discontinuity of the log-likelihood complicates the estimation. Their suggestion was to replace the sudden switch by a smooth transition. This removes the discontinuity, and the parameters of the resulting smooth transition regression model can be estimated by conditional maximum likelihood, using an appropriate iterative algorithm.

In this chapter we will apply the idea of approximating sudden changes with smooth transitions to the regime selection problem. That, in turn, allows us to use standard inference in determining the number of regimes in a (multiple) structural break model.

A classical logistic STR (LSTR) model for a univariate time series $y_t$ is given by

$$y_t = x'_t \beta_0 + x'_t \beta_1 G_{1t} + \varepsilon_t, \quad t = 1, \ldots, T, \quad (2.1)$$

where $x_t = (1, x_{1t}, x_{2t}, \ldots, x_{kt})'$, $y_t = (1, y_{t-1}, \ldots, y_{t-p}, w_{1t}, \ldots, w_{nt})'$, with $k = p + n$ is a $((k + 1) \times 1)$ vector of explanatory variables, $\beta_0$ and $\beta_1$ are $((k + 1) \times 1)$ parameter vectors and $\{\varepsilon_t\}$ is a sequence of independent, normally distributed errors with zero mean and variance $\sigma^2$. The transition function $G_{1t}$ in (2.1) is defined as follows:

$$G_{1t} = G_1(s_t; \gamma_1, c_1) = (1 + \exp\{-\gamma_1(s_t - c_1)\})^{-1}, \quad \gamma_1 > 0. \quad (2.2)$$

As $\gamma_1 \to \infty$ in (2.2), the logistic function $G_{1t}$ approaches the indicator function $I[s_t > c_1]$ and the LSTR model becomes a switching regression (SR) model. The parameter $c_1$ is then the switch or breakpoint parameter. Thus the STR model (2.1) with (2.2) is a reasonable approximation to the SR model when $\gamma_1$ is sufficiently large. Letting the variable $s_t = t$ (or, rescaling time to be between 0 and 1, $s_t = t^* = t/T$) and $\gamma \to \infty$, we obtain a single structural break model.

Analogously, we can approximate a multiple structural change model with a Multiple LSTR (MLSTR) model. For example, an MLSTR model with two transitions has the form

$$y_t = x'_t \beta_0^* + x'_t \beta_1^* G_{1t} + x'_t \beta_2^* G_{2t} + \varepsilon_t, \quad (2.3)$$

where the transition function $G_{2t} = G_2(t^*; \gamma_2, c_2)$ is defined as in (2.2). For the purposes of this chapter we set $\gamma_1 = \gamma_2 = \gamma$.

To illustrate how MLSTR model (2.3) mimics a structural break model with three regimes, we reparameterize (2.3) as follows:

$$y_t = x'_t \beta_1(1 - G_{1t}) + x'_t \beta_2(G_{1t} - G_{2t}) + x'_t \beta_3 G_{2t} + \varepsilon_t. \quad (2.4)$$
Determining the number of structural breaks

Letting $\gamma \to \infty$, we get a piecewise linear form. Figure 2.1 depicts the approximation to the three regimes created by $G_{1t}$ and $G_{2t}$ in (2.4), when $\gamma = 200$, $c_1 = 0.3$, $c_2 = 0.6$ and $s_t = t/T$.

Rearranging the terms in (2.4), we obtain

$$y_t = x_t' \beta_1 + x_t' (\beta_2 - \beta_1) G_{1t} + x_t' (\beta_3 - \beta_2) G_{2t} + \varepsilon_t,$$

or more generally, in case of $(m+1)$ regimes:

$$y_t = x_t' \beta_1 + \sum_{j=2}^{m+1} x_t' (\beta_j - \beta_{j-1}) G_{j-1,t} + \varepsilon_t$$

Suppose now that the true model is a structural change model with two breaks. We can approximate this model by the STR model (2.5) where $\gamma$ is large and known. Suppose, however, that we estimate (2.1) with $\gamma$ large and known using maximum likelihood. How does this misspecification affect our break date estimates? Analogously to BP, we argue that underspecification of the number of regimes hardly affects the estimates of $c_i$, i.e. our breakpoint parameter estimates. In other words, in our three-regime example, $c_1$ can be estimated reasonably accurately even when the number of regimes is misspecified by ignoring $G_{2t}$ in (2.5). As already mentioned, no asymptotic properties are claimed for this method because equation (2.5) is just an approximation to the three-regime structural break model.

Bai (1997) proved that consistent estimation of breaks is possible even when the breakpoints are estimated sequentially. In the appendix of Chapter 1
2.3 Smooth transition-based break detection

In this section we describe our procedure for determining the number of breakpoints in the piecewise linear structural break model. Let $m$ denote the number of breaks. The starting-point of our procedure is that the true model is a linear model, so that the first choice is between $m = 0$ (linearity) and $m = 1$ (two regimes). As a whole, the procedure works as follows:

1. Test linearity of (2.1) (i.e. $\gamma = 0$ in $G_{ll}(t*; \gamma, c_1)$). In order to circumvent the identification problem, approximate the transition function by its Taylor expansion around $\gamma = 0$. The third-order approximation can be written as $T_3 = \delta_0 + \delta_1 t^* + \delta_2 t^{*2} + \delta_3 t^{*3} + R_3(\gamma, c; t^*)$ where $R_3$ is the remainder and $\delta_0, \delta_1, \delta_2$ and $\delta_3$ are constants. Substituting $T_3$ for $G_{ll}$ in (2.1) and reparameterizing we obtain

$$y_t = x'_t \theta_0 + (x_t t^*)' \theta_1 + (x_t t^{*2})' \theta_2 + (x_t t^{*3})' \theta_3 + \varepsilon_t^*, \quad (2.6)$$

where $\varepsilon_t^* = \varepsilon_t + (x'_t \theta_1)R_3(\gamma, c; t^*)$. The parameter vectors $\theta_j = \gamma \tilde{\theta}_j$, where $\tilde{\theta}_j \neq 0$, and thus our null hypothesis of linearity (parameter constancy) in (2.1) implies $H'_0 : \theta_j = 0, j = 1, 2, 3$, in (2.6). Since the auxiliary regression (2.6) is linear in parameters and $\varepsilon_t^* = \varepsilon_t$ under $H'_0$, one can test this null hypothesis by a straightforward Lagrange Multiplier (LM)-type test

$$\chi^2_{LM} = \hat{\sigma}^{-2} \left( \sum_{t=1}^{T} \hat{u}_t w_t \right)' \left( \hat{M}_{11} - \hat{M}_{10} \hat{M}_{00}^{-1} \hat{M}_{01} \right)^{-1} \left( \sum_{t=1}^{T} w_t \hat{u}_t \right), \quad (2.7)$$

where $\hat{M}_{00} = \sum_{t=1}^{T} z_t z'_t$, $\hat{M}_{01} = \hat{M}_{10}' = \sum_{t=1}^{T} z_t w'_t$, $\hat{M}_{11} = \sum_{t=1}^{T} w_t w'_t$, $\hat{\sigma}^2 = 1/T \sum_{t=1}^{T} \hat{u}_t^2$. Here $\hat{u}_t$ is the residual estimated under the null hypothesis, $z_t = x_t$ and $w_t = (x_t t^*, x_t t^{*2}, x_t t^{*3})$. Under the null hypothesis, the test statistic has an asymptotic $\chi^2$-distribution with $3(k+1)$

---

It was shown that block corresponding to the location parameters of the MLSTR model in the average Hessian used as an estimate of the covariance matrix of the average score function is approximately diagonal when $\gamma$ is large. The same result holds when $s_t = t^*$, which is the case here. This means that even in this situation the location parameters can be estimated practically independently of each other, which is necessary for our sequential procedure to work.
degrees of freedom. This result requires the existence of all the moments implied by (2.7).

The test can also be based on the first-order Taylor approximation of $G_{1t}$. In that case $T_1 = \delta_0 + \delta_1 t^* + R_1(\gamma, c; t^*)$, where $R_1$ is the remainder and $\delta_0$ and $\delta_1$ are constants, and one substitutes $T_1$ for $G_1$ in (2.1). The null of linearity is now $\theta_1 = 0$ in (2.6) whereas $\theta_2 = \theta_3 = 0$ by definition. This variant of the test is less powerful than the test of $H_0$ in (2.6) in cases where the process is returning back to its original level after the second break. It has only trivial power when the only breaking parameter of the model is the intercept. We return to these issues in Section 2.4 where we compare the performance of different tests by simulation. For further discussion of the LM-type test, see, for example, Luukkonen, Saikkonen, and Teräsvirta (1988) or Teräsvirta (1998).

Following the recommendation in many papers, we use an $F$-approximation to the $\chi^2_{LM}$ statistic. The test can be carried out in three stages using just linear regressions:

(a) Regress $y_t$ on $x_t$ and compute the residual sum of squares
$$SSR_0 = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_t^2.$$  

(b) Regress $\hat{u}_t$ (or $y_t$) on $x_t$, $x_t t^*$, $x_t t^{*2}$, and $x_t t^{*3}$, and compute the residual sum of squares $SSR_1 = \frac{1}{T} \sum_{t=1}^{T} \hat{v}_t^2$.

(c) Compute
$$F = \frac{(SSR_0 - SSR_1)/(3(k + 1))}{SSR_1/(T - 4k - 4)}.$$  

This statistic is approximately $F_{3(k+1), T-4k-4}$ distributed under the null of linearity.

2. If the null hypothesis is rejected at a predetermined (approximate) significance level $\alpha$, estimate the parameters of (2.1) by nonlinear least squares fixing $\gamma$ at a sufficiently high but finite value. Then the STR model approximates an SR model with one break ($m = 1$) at the time fraction $t^* = c_1$, while the transition function still retains its smooth character (as a result, the likelihood function is well-behaved).

3. If LSTAR model (2.1) with fixed $\gamma$ is accepted, test it against a Multiple LSTAR model (2.5) with transition functions $G_{1t}$ and $G_{2t}$. This is done
by making use of the Taylor expansion of the transition function $G_{2t}$, see, for example, Eitrheim and Teräsvirta (1996) or Teräsvirta (1998). Accept (2.5) if the null hypothesis is rejected at significance level $\tau \alpha$, $0 < \tau < 1$. Reducing the significance level compared to the preceding test favours parsimonious models. Choosing $\tau$ is left to the modeller: in the simulations we set $\tau = 1/2$. Starting-values for the estimation may be obtained by using the estimates of $\beta_1, (\beta_2 - \beta_1)$ and $c_1$. The starting-value for $c_2$ is obtained by a one-dimensional grid search over a possible set of candidates. This also yields an initial value for $\beta_3$. The estimated model is then tested for another regime. The sequential estimation and testing is continued until the first acceptance of a null hypothesis. This determines the number of breaks in the SR model.

4. Estimate consistently the parameters of the final structural break model by the dynamic programming algorithm, see Bai and Perron (2003a), using the global minimization procedure with the given number of breaks.

It is easy to incorporate testing for partial structural breaks into this framework. The parameter constancy tests can be carried out for any subset of parameters. This is done by setting some elements (the ones we assume to be constant) $\beta_{1i} = 0$ in (2.1) \textit{a priori}. This in turn means that the same elements in auxiliary regressions are assumed to be equal to zero as well.

For our STR approximation procedure to work in the multivariate case, we need to assume that $\varepsilon_t$ are iid and that $x_{1t}, \ldots, x_{kt}$ are jointly stationary. In addition we require that all the cross-moments $E w_{it} w_{jt}$ and $E y_{t-i} w_{jt}$, exist (given that the coefficients of all explanatory variables among $x_t$ are changing). Finally, the errors are assumed uncorrelated with $x_t$.

It should be noted that when an SB model with $m$ breaks is tested against one with $m + 1$ breaks, $m \geq 1$, the asymptotic significance level of the test is unknown. This is the case because the null model is then a smooth transition approximation to the smaller structural change model. The critical value of the $F$-statistic is determined by assuming that the unknown null distribution is the $F$-distribution in question. In testing linearity or the no-break hypothesis, however, the asymptotic significance level is known because then the null model is not an approximation.

The lack of asymptotic inference may be viewed as a disadvantage from the purely theoretical point of view, but the model selection problem is a finite-sample one and the test is computationally simple. Another advantage is that our modelling strategy yields rather accurate values for the break dates (time fractions) even when some of them lie at either end of the sample.
2.4 Simulation study

In this section we investigate the small-sample behaviour of our model selection procedure by simulation. This also allows us to compare the proposed technique with the one Bai and Perron developed.

Several versions of the Bai and Perron testing sequence can be constructed (and are supported in their GAUSS code) depending on the assumptions on the distribution of the covariates and the errors across segments: the errors can be serially correlated or uncorrelated, regressors are either identically distributed or are allowed to have heterogenous distributions across segments, and finally also heteroskedasticity of residuals can be permitted. When serial correlation and/or heteroskedasticity is present, BP use a heteroskedasticity and serial correlation consistent (HAC) estimator of the parameter covariance matrix and allow for prewhitening. The authors do not, however, give clear guidelines for when prewhitening should be applied. For the simulation study, they use prewhitening only when the data generating process involves serially correlated errors. They do not discuss the size/power properties of their procedure when prewhitening is applied to uncorrelated series or not applied to serially correlated series. Our results suggest that this issue should be addressed, because the power of the BP procedure can vary by 15 percentage points when prewhitening is erroneously applied to uncorrelated series, see the simulation study below. Bai and Perron also note that the correction for possible serial correlation can be made, allowing the distribution of the regressors and errors to differ across regimes. In the construction of the tests, they do not consider imposing the restriction that the distribution of the regressors $z_t$ be the same across segments even if they are. This means that they explicitly allow the regressors to have heterogenous distributions. Some information about the size distortion implied by robustification, when the corresponding features are not present in the data, can be found in Bai and Perron (2001).

We, on the other hand, do not make use of these above-mentioned nonparametric techniques. To make our procedure comparable with the one Bai and Perron suggest, we use “parametric correction” for possible serial correlation. This amounts to first determining the AR order of the linear model using BIC (maximum lag length considered is $p = 5$) and setting up the testing sequence as we do when testing for a partial structural change and letting only the parameters of interest to change (the non-AR parameters in this study, if not noted otherwise). This is how most practitioners would cope with serial correlation. To be fair to both procedures, we report the results for both uncorrected and corrected versions (even if the data generated contain no features that have to be corrected for). The columns in Tables 2.1 - 2.7 labelled “LM1” and “LM3”
correspond to the STR-based procedure, making use of the first-order and third-order Taylor expansion, respectively. “Corrected” test sequences involve the nonparametric correction technique of BP and “parametric correction” in STR-based tests. The subscript “\(\text{PW}\)” denotes that prewhitening has been used in BP’s sequence.

### 2.4.1 Estimating the empirical size

Following Bai and Perron (2001) we simulate a number of univariate models with no structural changes and study how often the methods actually select the alternative of no breaks. The models are as follows:

<table>
<thead>
<tr>
<th>Generate data from</th>
<th>Estimate ( y_t = z_t \delta_j + u_t, j = 1, \ldots, m + 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) ( y_t = \varepsilon_t )</td>
<td>( z_t = {1} )</td>
</tr>
<tr>
<td>(b) ( y_t = \varepsilon_t + \Psi_t )</td>
<td>( z_t = {1, \Psi_t} )</td>
</tr>
<tr>
<td>(c) ( y_t = 0.5y_{t-1} + \varepsilon_t )</td>
<td>( z_t = {1, y_{t-1}} )</td>
</tr>
<tr>
<td>(d) ( y_t = 0.5y_{t-1} + \varepsilon_t )</td>
<td>( z_t = {1} )</td>
</tr>
<tr>
<td>(e) ( y_t = 0.5\varepsilon_{t-1} + \varepsilon_t )</td>
<td>( z_t = {1} )</td>
</tr>
<tr>
<td>(f) ( y_t = -0.3\varepsilon_{t-1} + \varepsilon_t )</td>
<td>( z_t = {1} )</td>
</tr>
</tbody>
</table>

where \( \{\varepsilon_t\} \sim \text{nid}(0, 1), \{\Psi_t\} \sim \text{nid}(1, 1) \) and uncorrelated with \( \{\varepsilon_t\} \), and \( z_t \) denotes the vector of covariates whose coefficients are allowed to change. For each DGP, we generate 2000 Monte Carlo replications with \( T = 120 \)\(^1\). Because the size of the BP sequential procedure is somewhat affected by the size of the trimming \( \varepsilon_R \) (minimum relative regime size), we report, following their recommendations, the results using \( \varepsilon_R = 0.05 \) for cases with no serial correlation correction, and \( \varepsilon_R = 0.20 \) for cases with serially correlated errors (if not otherwise noted). The nominal test size is \( \alpha = 0.05 \). For the STR-based procedure, we choose \( \tau = 0.5 \), that is we halve the level of the test at each consecutive step. This has no effect on size simulations, because the first linearity test still has the correct nominal size. If we choose not to reduce the level of the test at every step, that is \( \tau = 1 \), the only differences would appear in probabilities \( P(m = a), a \neq 0 \), where \( m \) denotes the number of breaks.

Table 2.1 contains the results for the six DGPs listed above. The size distortion when correcting for non-existing serial correlation for our procedure is minor, see panels (a) and (b). The size of the BP sequential procedure depends heavily on whether prewhitening is used or not. Applying the prewhitening for the DGP in panel (b) can cause a size distortion as large as

---

\(^1\)We first generate 200 initial observations that will be removed, to minimize the possible effect of the starting-values.
Determining the number of structural breaks

10 percentage points. In the presence of serial correlation, one should try to correct for it, because ignoring its presence may lead to serious size distortions, see the results for the uncorrected versions of the tests in panels (d) and (e). Both procedures are well sized when the serial correlation in the errors is accounted for, although the BP procedure appears more oversized than our technique when the nonzero autocorrelations are positive. Prewhitening does affect the size of the BP sequential procedure even here, and not applying it when it would be necessary to do so can cause noticeable size distortion, see panel (d), for example.

A thorough discussion of size distortion in the Bai and Perron sequential procedure can be found in Prodan (2003). Simple simulations with an AR(1) model show that the size distortion becomes severe the closer we get to the unit root. Our simulations support her results. Our “corrected” procedure displays somewhat less size distortion than the BP sequential procedure, but is still oversized when the autoregressive coefficient approaches one. On nominal 5% level with $\rho = 0.9$ we reject LM$_1$ test in about 11% and the LM$_3$ test in about 21% of the cases (compared to about 22% for BP).

2.4.2 Simulating structural break models

To study the power of the procedures, we again replicate the experiments in Bai and Perron (2001). Even here, $\{\Psi_t\} \sim \text{nid}(1, 1)$ and $\{e_t\} \sim \text{nid}(0, 1)$, and these sequences are mutually uncorrelated. The minimum relative regime size for cases with no error autocorrelation is $\epsilon_R = 0.05$, and for cases with serial correlation $\epsilon_R = 0.2$, that is 20% of the length of the series.

A single break

First we look at a battery of data generating processes with a single break in the middle of the series. The model has a general form

$$
\begin{align*}
y_t &= \mu_1 + \nu_1 \Psi_t + e_t, & \text{if} & \quad t \leq [0.5T] \\
y_t &= \mu_2 + \nu_2 \Psi_t + e_t, & \text{if} & \quad t > [0.5T],
\end{align*}
$$

and we are testing for the break in both parameters, i.e. $z_t = \{1, \Psi_t\}$. The results appear in Table 2.2. In the case of a single break, the power of our procedure (either LM$_1$ or LM$_3$) is generally somewhat higher than that of the Bai-Perron sequential procedure. The test based on the first-order Taylor approximation performs a bit better than the one based on the third-order approximation because in the case of a single break, the higher-order auxiliary terms do not carry helpful extra information. Correcting for non-existing serial
2.4. Simulation study

correlation does not seem to have a large effect on the power of our procedure. The BP procedure, on the other hand, can lose as much as 15 percentage points of its power when prewhitening is applied. When prewhitening is not applied, the results are similar to the ones of our LM1 test, see panels (c), (f), (g) and columns labelled “BP” and “BP\textsubscript{pw}” in Table 2.2.

Second, we consider two different DGPs (and two sample sizes) with one break and serially correlated errors. The model has the following general form:

\begin{equation}
\begin{align*}
y_t &= \mu_1 + \nu_t, \quad \text{if} \quad t \leq [0.5T] \\
y_t &= \mu_2 + \nu_t, \quad \text{if} \quad t > [0.5T],
\end{align*}
\end{equation}

where \(\nu_t = 0.5\nu_{t-1} + \epsilon_t\) and we are testing for the break in the intercept, i.e. \(z_t = \{1\}\). The results appear in Table 2.3. Again, our LM1 test does about as well as the BP sequential procedure. It is easier to detect small breaks with procedures without serial correlation correction than with procedures with correction. A problem arises because the correction tends to partially absorb the break. This seems to be true for both methods. If the jump (the break is in the intercept) is large enough, then accounting for serial correlation pays off (panel (k)), otherwise the effect is rather the opposite. In three cases out of four, the sequential method of BP has an advantage of a few percentage points in power, but the differences are not large. Differences become somewhat larger if one did (erroneously) not apply the prewhitening technique when using BP’s procedure, especially for smaller breaks and sample sizes.

Two breaks

To see how the procedures compare to each other in the presence of multiple structural breaks, we simulate data from the following model:

\begin{equation}
\begin{align*}
y_t &= \mu_1 + \nu_1 \Psi_t + \epsilon_t, \quad \text{if} \quad 1 < t \leq [1/3T] \\
y_t &= \mu_2 + \nu_2 \Psi_t + \epsilon_t, \quad \text{if} \quad [1/3T] < t \leq [2/3T] \\
y_t &= \mu_3 + \nu_3 \Psi_t + \epsilon_t, \quad \text{if} \quad [2/3T] < t < T.
\end{align*}
\end{equation}

In model (2.10) there are two equally-spaced breaks and all parameters of the model are potentially subject to change \((z_t = \{1, \Psi_t\})\), and the errors are serially uncorrelated and homoskedastic. The results are presented in Table 2.4. If there are only breaks in the intercept, however large, the power loss for our procedure is substantial when correcting for non-existing serial correlation. As an example, the power may drop from 68% to 8%, see panel (g) in Table 2.4. This suggests that a break in the intercept can be “explained” by adding dynamic structure to the model. This, however is common practice in cases
Determining the number of structural breaks

where there is no prior information about the dynamic behaviour of \( y_t \). The loss in power is only a few percentage points when the other parameters change as well. BP for some reason often gain from correcting for non-existing serial correlation (and often even more so when prewhitening the data) and their sequential procedure with (erroneous) correction is working better than ours.

In the presence of multiple breaks of opposite direction, the test based on the third-order Taylor expansion (LM3) works better than the one based on first-order approximation (LM1). Differences in the performance can be large, see panels (g) and (i), for example. This can be explained by the added flexibility in the third-order approximation that allows for nonmonotonic and asymmetric parameter constancy. The sequence based on LM1 still works better than the one based on LM3 when the change in the parameters is gradual, see panels (j) – (l). In that case the first break is often estimated in the middle of the sample and the second break becomes difficult to detect, even with all the extra terms in the auxiliary regression.

To further study the properties of the tests, we simulate data with intercept shifts and serially correlated errors. That is,

\[
\begin{align*}
Y_t &= \mu_1 + v_t, & \text{if } 1 < t \leq \lfloor 1/3T \rfloor, \\
Y_t &= \mu_2 + v_t, & \text{if } \lfloor 1/3T \rfloor < t \leq \lfloor 2/3T \rfloor, \\
Y_t &= \mu_3 + v_t, & \text{if } \lfloor 2/3T \rfloor < t < T,
\end{align*}
\]

where \( v_t = 0.5v_{t-1} + e_t \) with \( \{e_t\} \sim \text{nid}(0,1) \). We focus on cases where the mean returns to its old value at the second break, i.e. \( \mu_1 = \mu_3 = 0 \). The results can be found in Table 2.5. When no correction is carried out for the serial correlation present in the data, our procedure clearly dominates and the procedure of BP selects models with \( m \geq 3 \) more often than parsimonious models. Our procedure also dominates when the autocorrelation is accounted for, except when shifts are small and samples short. As one would expect, the test sequence based on the first-order Taylor approximation has no power at all, but the one based on the third-order approximation performs very well. In BP case, prewhitening improves the power when breaks are large and samples long.

We replicate two more experiments from Bai and Perron (2001). The data is generated from equation (2.10) but allowing the distribution of errors and variables to change across segments. That is, we use:

\[
\begin{align*}
\Psi_t^* &\sim \text{nid}(\varsigma_1, 1), & \text{if } 1 < t \leq \lfloor 1/3T \rfloor, \\
\Psi_t^* &\sim \text{nid}(\varsigma_2, 1), & \text{if } \lfloor 1/3T \rfloor < t \leq \lfloor 2/3T \rfloor, \\
\Psi_t^* &\sim \text{nid}(\varsigma_3, 1), & \text{if } \lfloor 2/3T \rfloor < t < T.
\end{align*}
\]
2.4. Simulation study

and

\[ e_t \sim \text{nid}(0, \sigma_1^2), \quad \text{if} \quad 1 < t \leq [1/3T] \]
\[ e_t \sim \text{nid}(0, \sigma_2^2), \quad \text{if} \quad [1/3T] < t \leq [2/3T] \] (2.13)
\[ e_t \sim \text{nid}(0, \sigma_3^2), \quad \text{if} \quad [2/3T] < t < T \]

in (2.10).

In Table 2.6 we report the results of one of the experiments where the fixed parameters (regression coefficients) are set as follows: \( \nu_1 = 1, \nu_2 = 1.5, \nu_3 = 0.5 \) and \( \mu_1 = 0, \mu_2 = 0.5, \mu_3 = -0.5 \). The minimum relative regime size is set to \( \epsilon_R = 0.15 \). The changes in error variance and in the mean of \( \Psi_t \) are of the same type: starting from value 1, jumping to a higher level, and then returning to value 1 again. That is, for every DGP in this experiment \( \sigma_1^2 = \sigma_2^2 = 1 \) and \( \varsigma_1 = \varsigma_3 = 1 \).

The first few columns in Table 2.6 concern the case where no correction for serial correlation and/or heteroskedasticity is made. Columns 6–9 refer to the case most likely to be encountered in practice, which is correcting for serial correlation but not accounting for changes in error variance. The last column covers the results for the correct variant of BP’s test sequence. The test statistics used there account for heteroskedasticity and do not allow for serial correlation correction. That test performs exceptionally well and requires no further comment. Currently there does not exist an STR counterpart to it.

When no correction is undertaken for heteroskedasticity or serial correlation, the BP procedure excels. That is to be expected, as their test accommodates the possibility of heterogenous regressors and, allowing for that possibility, is highly recommended by BP. Although the best of the STR-based tests is somewhat less powerful than the sequential procedure of BP, it is able to choose \( m = 2 \) frequently enough, except for cases when the sample size is small and either \( \varsigma_2 = 4 \) or \( \sigma_2^2 = 4 \) or both. In situations like that, a model with a single break is selected more often than a model with two breaks. If one corrects for the serial correlation, then results are mixed: no test clearly dominates the other and the power of both methods decreases. BP can lose even as much as 76 percentage points (see panel (d)) when prewhitening is used, and even more when data are not prewhitened. At least in half of the cases, the correction (partly) absorbs the changes, and models that are too parsimonious are selected most frequently.

Table 2.7 reports the results of the other experiment. The DGPs are such that the intercept and slope parameters increase at breakpoints gradually, i.e. \( \nu_1 = 1, \nu_2 = 1.5, \nu_3 = 2 \) and \( \mu_1 = 0, \mu_2 = 0.5, \mu_3 = 1 \). Even here \( \sigma_1^2 = \sigma_3^2 = 1 \) and \( \varsigma_1 = \varsigma_3 = 1 \) in all cases, whereas the second segment mean
Determining the number of structural breaks

and variance differ from the values above. Minimum regime size is set to $\epsilon_R = 0.20$ to ensure tests with adequate sizes for BP procedure. Again, the test with exactly correct setup has the best performance, see the last column in Table 2.7. The uncorrected version of BP performs somewhat better than ours, because it has the advantage of explicitly allowing for heterogenous regressors. For larger samples, our procedure is able to select $m = 2$ reasonably frequently. In this case the serial correlation correction has a rather devastating effect on the procedure of BP. Our procedure loses some power as well but not nearly as much as BP and has superior power in five cases out of eight and about equal power in two more.

The results of our technique depend somewhat on the choice of the discount coefficient $\tau$. It is clear that increasing $\tau$ makes the strategy less parsimonious. When $\tau = 1$, we retain the same nominal level for each test in the sequence and are more likely to choose less parsimonious models than if we choose $\tau < 1$. When setting $\tau = 1$, our STR-based procedure can lose up to 5 percentage points in power for the current DGPs with one break compared to the case $\tau = 0.5$. The hypothesis of only one break is rejected somewhat more frequently and some probability mass is shifting from $P(\hat{m} = 1)$ to $P(\hat{m} = 2)$. On the other hand, for models with two breaks we can gain up to 15 percentage points in power. This is true for configurations where it was previously difficult to detect the second break. We may also lose a little in the cases where the number of breaks was estimated precisely, since now models with more than two breaks have a chance to be selected. Setting $\tau < 0.5$ has the opposite effect - parsimony is strongly preferred and finding the second break becomes more difficult.

We can also conclude that it is not necessary to set the minimum regime size equal to 15–20% for the “corrected” cases when the STR-based procedure is applied. One could easily set the minimum regime length to be 5% of the total sample size, without much affecting the power of our test. Depending on the DGP, the average change would be less than one percentage point and maximum gains and losses about three percentage points.

Dynamic models and breaks

It is not completely clear, however, how one would in practice handle the problem of detecting breaks in the presence of autocorrelation. The design of the simulations just discussed implicitly suggests that the model builder is primarily prepared for finding a break in the intercept. The conditional mean of the simulated models has a very simple structure, and the error process is assumed to be autocorrelated. An interesting question is what would happen if
the breaks were of more general character. In order to illuminate this situation, we simulated data from an AR(2) model with two structural breaks in the dynamic behaviour of the process:

\[
\begin{align*}
    y_t &= \begin{cases} 
    2.7 + 0.8y_{t-1} - 0.2y_{t-2} + \varepsilon_t & \text{if } 1 < t \leq [1/3T] \\
    0.3 - 0.2y_{t-1} + 0.5y_{t-2} + \varepsilon_t & \text{if } [1/3T] < t \leq [2/3T] \\
    1 + 0.7y_{t-1} - 0.3y_{t-2} + \varepsilon_t & \text{if } [2/3T] < t < T. 
    \end{cases}
\end{align*}
\] (2.14)

In (2.14) each segment is covariance stationary and an example of a generated series can be found in Figure 2.2.

![Figure 2.2](image)

We consider the following three strategies for proceeding that are also supported in the GAUSS code of BP:

- **Strategy (1):** One makes the assumption that there are breaks in the overall unconditional mean of the process but that the dynamics are not changing over time. In that case, when using BP’s sequential method, one would only test for breaks in the intercept and correct for possible serial correlation in the errors using nonparametric methods.

- **Strategy (2):** One assumes that both the mean and the dynamics of the process are changing over time, but that the distribution of regressors should be the same in every segment. This would mean that the AR order of the model is first selected by an appropriate information criterion and the model is then tested for structural breaks using the homogenous version of BPs sequential procedure.

- **Strategy (3):** One proceeds as in Strategy (2) but follows the recommendations of BP and relaxes the homogeneity assumption.
In Table 2.8 we report the model selection frequencies of the sequential procedure of BP for those three strategies and the model selection frequencies of the "practitioner's strategy" (similar to the second strategy above) when using our STR-based method. It is clear from the table that just prewhitening and using the HAC estimator is not sufficient. The frequency of choosing the correct number of breaks in the column corresponding to the first strategy is relatively low. Comparing the results in columns (2) and (3), it is obvious why Bai and Perron strongly recommend one allows for different variances of regressors across segments, or in this case, a different variance for $y$ in each regime. It appears that this assumption has a large positive effect on power.

It is rather striking how much the results in columns (2) and (3) can differ from each other. STR-based procedures perform as expected, $LM_3$ is more powerful than $LM_1$ because there are two breaks in the simulated model. This small experiment indicates that model uncertainty is a serious issue and that results depend on the modelling approach used in the study.

### 2.5 Conclusions

In this chapter we show how a smooth transition regression approximation to a piecewise linear structural break model is useful in determining the number of breaks in the latter model when it is not known in advance. The approach proposed and simulated in the chapter is based on sequential hypothesis testing and is simple to apply in practice. The whole procedure is based on standard inference and the user can (approximately) control the overall significance level of the tests in the sequence. In addition, no restrictions are imposed on the number of changing parameters. The simulations show that our procedure is well-sized and works very well in comparison with the sequential procedure suggested by Bai and Perron. Neither of the alternatives dominates the other in small and moderate samples.

The examples discussed above show that the results of both our and Bai and Perron's approaches depend on the way the error autocorrelation is being taken care of. Adding lags to the model is a common practice, but sometimes small breaks can get absorbed by the extra dynamics. Then again, in practice a firm knowledge of the presence of a break is rather an exception than rule and a casual modeller would add lags to the model. One has to be careful when applying Bai and Perron's technique as well. The results can depend heavily on the assumptions one is or is not willing to make about the error term and series at hand. Allowing for different distributions of covariates in different segments helps a great deal, but an unnecessary prewhitening can
considerably weaken the procedure's ability to detect breaks.

Overall, our STR-based method can be considered a complement to the classical approach of Bai and Perron. Our procedure may be extended to accommodate heteroskedasticity by making the error variance change over time at the same points as the mean. This extension is, however, left for further research.
References


Determining the number of structural breaks


### Table 2.1: Selection frequencies of Bai and Perron sequential procedure (BP) and STR-based procedures (LM$_1$ and LM$_3$). Data are generated with no breaks, i.e. $m = 0$.

<table>
<thead>
<tr>
<th>Model</th>
<th>Choice</th>
<th>Uncorrected</th>
<th>Corrected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>BP</td>
<td>LM$_1$</td>
</tr>
<tr>
<td></td>
<td>m = 0</td>
<td>95.15</td>
<td>94.15</td>
</tr>
<tr>
<td>(a)</td>
<td>m = 1</td>
<td>4.70</td>
<td>5.85</td>
</tr>
<tr>
<td></td>
<td>m = 2</td>
<td>0.15</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>m = 0</td>
<td>95.95</td>
<td>94.45</td>
</tr>
<tr>
<td>(b)</td>
<td>m = 1</td>
<td>3.90</td>
<td>5.55</td>
</tr>
<tr>
<td></td>
<td>m = 2</td>
<td>0.15</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>m = 0</td>
<td>95.15</td>
<td>95.35</td>
</tr>
<tr>
<td>(c)</td>
<td>m = 1</td>
<td>4.50</td>
<td>4.60</td>
</tr>
<tr>
<td></td>
<td>m = 2</td>
<td>0.30</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>m = 0</td>
<td>46.90</td>
<td>73.30</td>
</tr>
<tr>
<td>(d)</td>
<td>m = 1</td>
<td>28.15</td>
<td>25.00</td>
</tr>
<tr>
<td></td>
<td>m = 2</td>
<td>16.85</td>
<td>1.70</td>
</tr>
<tr>
<td></td>
<td>m = 0</td>
<td>73.75</td>
<td>84.35</td>
</tr>
<tr>
<td>(e)</td>
<td>m = 1</td>
<td>20.70</td>
<td>15.50</td>
</tr>
<tr>
<td></td>
<td>m = 2</td>
<td>4.95</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>m = 0</td>
<td>99.85</td>
<td>99.70</td>
</tr>
<tr>
<td>(f)</td>
<td>m = 1</td>
<td>0.15</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>m = 2</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

**Notes:** The table contains selection frequencies in per cent based on 2000 Monte Carlo replications. Columns labelled "Uncorrected" contain the results when not correcting for serial correlation, and columns labelled "Corrected" correspond to tests that correct for serial correlation either non-parametrically or parametrically. The column labelled "BP$_{PW}$" contains the results of the BP sequential test when prewhitening is applied before estimating the long-run covariance matrix. The columns labelled "LM$_1$" and "LM$_3$" correspond to the STR-based tests that make use of the first-order and third-order Taylor expansions, respectively.
Table 2.2: Model selection frequencies of the uncorrected and corrected versions of the BP and STR-based procedures. Data are generated from equation (2.8), i.e. errors are uncorrelated and $m = 1$.

<table>
<thead>
<tr>
<th>Model Choice</th>
<th>Uncorrected</th>
<th>Corrected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BP</td>
<td>LM$_1$</td>
</tr>
<tr>
<td><strong>Change in the intercept only: $\nu_1 = \nu_2 = 1$</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_1 = 0$</td>
<td>$m = 0$</td>
<td>55.75</td>
</tr>
<tr>
<td>(a) $\mu_2 = 0.5$</td>
<td>$m = 1$</td>
<td>42.60</td>
</tr>
<tr>
<td>$T = 120$</td>
<td>$m = 2$</td>
<td>1.60</td>
</tr>
<tr>
<td>$\mu_1 = 0$</td>
<td>$m = 0$</td>
<td>20.20</td>
</tr>
<tr>
<td>(b) $\mu_2 = 0.5$</td>
<td>$m = 1$</td>
<td>77.85</td>
</tr>
<tr>
<td>$T = 240$</td>
<td>$m = 2$</td>
<td>1.95</td>
</tr>
<tr>
<td>$\nu_1 = 0$</td>
<td>$m = 0$</td>
<td>0.90</td>
</tr>
<tr>
<td>(c) $\mu_2 = 1$</td>
<td>$m = 1$</td>
<td>95.00</td>
</tr>
<tr>
<td>$T = 120$</td>
<td>$m = 2$</td>
<td>4.10</td>
</tr>
<tr>
<td><strong>Change in the slope only: $\mu_1 = \mu_2 = 0$</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\nu_1 = 1$</td>
<td>$m = 0$</td>
<td>21.85</td>
</tr>
<tr>
<td>(d) $\nu_2 = 1.5$</td>
<td>$m = 1$</td>
<td>75.45</td>
</tr>
<tr>
<td>$T = 120$</td>
<td>$m = 2$</td>
<td>2.65</td>
</tr>
<tr>
<td>$\nu_1 = 1$</td>
<td>$m = 0$</td>
<td>0.30</td>
</tr>
<tr>
<td>(e) $\nu_2 = 1.5$</td>
<td>$m = 1$</td>
<td>95.75</td>
</tr>
<tr>
<td>$T = 240$</td>
<td>$m = 2$</td>
<td>3.90</td>
</tr>
<tr>
<td><strong>Change in all parameters: $\nu_1 = 1, \mu_1 = 0$</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\nu_2 = 1.5$</td>
<td>$m = 0$</td>
<td>0.05</td>
</tr>
<tr>
<td>(f) $\mu_2 = 0.5$</td>
<td>$m = 1$</td>
<td>96.20</td>
</tr>
<tr>
<td>$T = 120$</td>
<td>$m = 2$</td>
<td>3.65</td>
</tr>
<tr>
<td>$\nu_2 = 2$</td>
<td>$m = 0$</td>
<td>0.00</td>
</tr>
<tr>
<td>(g) $\mu_2 = 1$</td>
<td>$m = 1$</td>
<td>95.40</td>
</tr>
<tr>
<td>$T = 120$</td>
<td>$m = 2$</td>
<td>4.55</td>
</tr>
</tbody>
</table>

**Notes:** The table contains selection frequencies in per cent based on 2000 Monte Carlo replications. Columns labelled “Uncorrected” contain the results when not correcting for serial correlation, and columns labelled “Corrected” correspond to tests that correct for serial correlation either nonparametrically or parametrically. The column labelled “BP$_{pw}$” contains the results of the BP sequential test when pre-whitening is applied before estimating the long-run covariance matrix. The columns labelled “LM$_1$” and “LM$_3$” correspond to the STR-based tests that make use of the first-order and third-order Taylor expansions, respectively.
Table 2.3: Model selection frequencies of the uncorrected and corrected versions of the BP and STR-based procedures. Data are generated from equation (2.9), i.e. errors are serially correlated and $m = 1$.

<table>
<thead>
<tr>
<th>Model</th>
<th>Uncorrected</th>
<th>Corrected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Choice BP</td>
<td>LM₁</td>
</tr>
<tr>
<td>$\mu_1 = 0$</td>
<td>$\mu_2 = 0.5$</td>
<td>$\mu_2 = 0.5$</td>
</tr>
<tr>
<td>$T = 120$</td>
<td>$T = 120$</td>
<td>$T = 240$</td>
</tr>
<tr>
<td>$m = 0$</td>
<td>$m = 1$</td>
<td>$m = 1$</td>
</tr>
<tr>
<td>$m = 2$</td>
<td>$m = 2$</td>
<td>$m = 2$</td>
</tr>
<tr>
<td>$m = 0$</td>
<td>$m = 1$</td>
<td>$m = 1$</td>
</tr>
<tr>
<td>$m = 1$</td>
<td>$m = 2$</td>
<td>$m = 2$</td>
</tr>
<tr>
<td>$m = 0$</td>
<td>$m = 1$</td>
<td>$m = 1$</td>
</tr>
<tr>
<td>$m = 1$</td>
<td>$m = 2$</td>
<td>$m = 2$</td>
</tr>
<tr>
<td>$m = 0$</td>
<td>$m = 1$</td>
<td>$m = 1$</td>
</tr>
<tr>
<td>$m = 1$</td>
<td>$m = 2$</td>
<td>$m = 2$</td>
</tr>
</tbody>
</table>

Notes: The table contains selection frequencies in per cent based on 2000 Monte Carlo replications. Columns labelled “Uncorrected” contain the results when not correcting for serial correlation, and columns labelled “Corrected” correspond to tests that correct for serial correlation either nonparametrically or parametrically. The column labelled “BPₚₖ” contains the results of the BP sequential test when prewhitening is applied before estimating the long-run covariance matrix. The columns labelled “LM₁” and “LM₃” correspond to the STR-based tests that make use of the first-order and third-order Taylor expansions, respectively.
Table 2.4: Model selection frequencies of the uncorrected and corrected versions of the BP and STR-based procedures. Data are generated from equation (2.10), \( m = 2 \).

<table>
<thead>
<tr>
<th>Model</th>
<th>Choice</th>
<th>Uncorrected</th>
<th>Corrected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>BP</td>
<td>LM(_1)</td>
</tr>
<tr>
<td>( \mu_1 = 0 )</td>
<td>( m = 0 )</td>
<td>88.90</td>
<td>95.10</td>
</tr>
<tr>
<td></td>
<td>( m = 1 )</td>
<td>9.60</td>
<td>4.60</td>
</tr>
<tr>
<td></td>
<td>( m = 2 )</td>
<td>1.50</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>( m = 3 )</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>( \mu_1 = 0 )</td>
<td>( m = 0 )</td>
<td>57.60</td>
<td>97.15</td>
</tr>
<tr>
<td></td>
<td>( m = 1 )</td>
<td>8.90</td>
<td>1.70</td>
</tr>
<tr>
<td></td>
<td>( m = 2 )</td>
<td>31.30</td>
<td>1.15</td>
</tr>
<tr>
<td></td>
<td>( m = 3 )</td>
<td>2.15</td>
<td>0.00</td>
</tr>
<tr>
<td>( \mu_1 = 0 )</td>
<td>( m = 0 )</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>( m = 1 )</td>
<td>41.50</td>
<td>42.90</td>
</tr>
<tr>
<td></td>
<td>( m = 2 )</td>
<td>55.10</td>
<td>55.95</td>
</tr>
<tr>
<td></td>
<td>( m = 3 )</td>
<td>3.20</td>
<td>1.15</td>
</tr>
<tr>
<td>( \mu_1 = 0 )</td>
<td>( m = 0 )</td>
<td>0.00</td>
<td>10.40</td>
</tr>
<tr>
<td></td>
<td>( m = 1 )</td>
<td>14.80</td>
<td>31.00</td>
</tr>
<tr>
<td></td>
<td>( m = 2 )</td>
<td>81.80</td>
<td>64.90</td>
</tr>
<tr>
<td></td>
<td>( m = 3 )</td>
<td>3.30</td>
<td>3.85</td>
</tr>
<tr>
<td>( \mu_1 = 0 )</td>
<td>( m = 0 )</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>( m = 1 )</td>
<td>14.80</td>
<td>31.00</td>
</tr>
<tr>
<td></td>
<td>( m = 2 )</td>
<td>81.80</td>
<td>65.15</td>
</tr>
<tr>
<td></td>
<td>( m = 3 )</td>
<td>3.30</td>
<td>3.85</td>
</tr>
<tr>
<td>( \mu_1 = 0 )</td>
<td>( m = 0 )</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>( m = 1 )</td>
<td>14.25</td>
<td>17.60</td>
</tr>
<tr>
<td></td>
<td>( m = 2 )</td>
<td>80.95</td>
<td>80.10</td>
</tr>
<tr>
<td></td>
<td>( m = 3 )</td>
<td>4.60</td>
<td>2.30</td>
</tr>
<tr>
<td>( \mu_1 = 0 )</td>
<td>( m = 0 )</td>
<td>0.00</td>
<td>29.25</td>
</tr>
<tr>
<td></td>
<td>( m = 1 )</td>
<td>0.00</td>
<td>1.55</td>
</tr>
<tr>
<td></td>
<td>( m = 2 )</td>
<td>93.95</td>
<td>67.50</td>
</tr>
<tr>
<td></td>
<td>( m = 3 )</td>
<td>5.85</td>
<td>1.70</td>
</tr>
</tbody>
</table>

Notes: See the Notes of Table 2.3.
Table 2.4: Model selection frequencies of the uncorrected and corrected versions of the BP and STR-based procedures. Data are generated from equation (2.10), \( m = 2 \), (cont.).

<table>
<thead>
<tr>
<th>Model</th>
<th>Uncorrected</th>
<th>Corrected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Choice BP</td>
<td>LM</td>
</tr>
<tr>
<td>( \nu_1 = 1 )</td>
<td>( m = 0 )</td>
<td>78.05</td>
</tr>
<tr>
<td>( \nu_2 = 1.5 )</td>
<td>( m = 1 )</td>
<td>14.35</td>
</tr>
<tr>
<td>( \nu_3 = 1 )</td>
<td>( m = 2 )</td>
<td>7.25</td>
</tr>
<tr>
<td></td>
<td>( m = 3 )</td>
<td>0.35</td>
</tr>
<tr>
<td>( \nu_1 = 1 )</td>
<td>( m = 0 )</td>
<td>20.05</td>
</tr>
<tr>
<td>( \nu_2 = 2 )</td>
<td>( m = 1 )</td>
<td>1.25</td>
</tr>
<tr>
<td>( \nu_3 = 1 )</td>
<td>( m = 2 )</td>
<td>72.20</td>
</tr>
<tr>
<td></td>
<td>( m = 3 )</td>
<td>6.35</td>
</tr>
<tr>
<td>( \nu_1 = 1 )</td>
<td>( m = 0 )</td>
<td>0.25</td>
</tr>
<tr>
<td>( \nu_2 = 1.5 )</td>
<td>( m = 1 )</td>
<td>86.70</td>
</tr>
<tr>
<td>( \nu_3 = 2 )</td>
<td>( m = 2 )</td>
<td>12.55</td>
</tr>
<tr>
<td></td>
<td>( m = 3 )</td>
<td>72.20</td>
</tr>
<tr>
<td>( \nu_1 = 1 )</td>
<td>( m = 0 )</td>
<td>0.00</td>
</tr>
<tr>
<td>( \nu_2 = 2 )</td>
<td>( m = 1 )</td>
<td>2.35</td>
</tr>
<tr>
<td>( \nu_3 = 3 )</td>
<td>( m = 2 )</td>
<td>89.50</td>
</tr>
<tr>
<td></td>
<td>( m = 3 )</td>
<td>7.75</td>
</tr>
<tr>
<td>( \nu_1 = 1 )</td>
<td>( m = 0 )</td>
<td>0.00</td>
</tr>
<tr>
<td>( \nu_2 = 0.5 )</td>
<td>( m = 1 )</td>
<td>60.95</td>
</tr>
<tr>
<td>( \nu_3 = -0.5 )</td>
<td>( m = 2 )</td>
<td>37.50</td>
</tr>
<tr>
<td></td>
<td>( m = 3 )</td>
<td>1.55</td>
</tr>
</tbody>
</table>

Change in all parameters

| \( \mu_1 = 0, \nu_1 = 1 \) | \( m = 0 \) | 87.85 | 94.40 | 79.40 | 76.20 | 82.05 | 94.60 | 79.55 |
| \( \mu_2 = 0.5, \nu_2 = 0.5 \) | \( m = 1 \) | 10.55 | 5.30 | 16.80 | 15.85 | 14.70 | 5.10 | 15.70 |
| \( \mu_3 = 0, \nu_3 = 1 \) | \( m = 2 \) | 1.55 | 0.30 | 3.65 | 7.60 | 3.20 | 0.30 | 4.65 |
| \( m = 3 \) | 0.05 | 0.00 | 0.15 | 0.35 | 0.05 | 0.00 | 0.10 |
| \( \mu_1 = 0, \nu_1 = 1 \) | \( m = 0 \) | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| \( \mu_2 = 1, \nu_2 = 1.5 \) | \( m = 1 \) | 13.75 | 18.80 | 28.30 | 13.50 | 9.65 | 22.20 | 32.60 |
| \( \mu_3 = 2, \nu_3 = 2 \) | \( m = 2 \) | 83.70 | 78.55 | 69.15 | 86.15 | 90.35 | 77.80 | 67.40 |
| \( m = 3 \) | 2.55 | 2.65 | 2.55 | 0.35 | 0.00 | 0.00 | 0.00 |
| \( \mu_1 = 0, \nu_1 = 1 \) | \( m = 0 \) | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| \( \mu_2 = 1, \nu_2 = 2 \) | \( m = 1 \) | 16.20 | 27.45 | 32.40 | 20.15 | 12.20 | 31.30 | 37.55 |
| \( \mu_3 = 2, \nu_3 = 1 \) | \( m = 2 \) | 80.10 | 70.80 | 65.85 | 78.90 | 87.70 | 64.95 | 62.15 |
| \( m = 3 \) | 3.55 | 1.75 | 1.75 | 0.00 | 0.10 | 0.00 | 0.00 |

Notes: See the Notes of Table 2.3.
Table 2.5: Model selection frequencies of the uncorrected and corrected versions of the BP and STR-based procedures. Data are generated from equation (2.11), i.e. errors are serially correlated and \( m = 2 \).

<table>
<thead>
<tr>
<th>Model</th>
<th>Uncorrected</th>
<th>Corrected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Choice</td>
<td>BP</td>
</tr>
<tr>
<td>(a) ( \mu_2 = 0.5 )</td>
<td>( m = 0 )</td>
<td>37.65</td>
</tr>
<tr>
<td></td>
<td>( m = 1 )</td>
<td>25.20</td>
</tr>
<tr>
<td></td>
<td>( m = 2 )</td>
<td>22.45</td>
</tr>
<tr>
<td></td>
<td>( m = 3 )</td>
<td>10.75</td>
</tr>
<tr>
<td>(b) ( \mu_2 = 0.5 )</td>
<td>( m = 0 )</td>
<td>27.45</td>
</tr>
<tr>
<td></td>
<td>( m = 1 )</td>
<td>22.25</td>
</tr>
<tr>
<td></td>
<td>( m = 2 )</td>
<td>27.00</td>
</tr>
<tr>
<td></td>
<td>( m = 3 )</td>
<td>16.45</td>
</tr>
<tr>
<td>(c) ( \mu_2 = 1 )</td>
<td>( m = 0 )</td>
<td>18.75</td>
</tr>
<tr>
<td></td>
<td>( m = 1 )</td>
<td>14.15</td>
</tr>
<tr>
<td></td>
<td>( m = 2 )</td>
<td>32.80</td>
</tr>
<tr>
<td></td>
<td>( m = 3 )</td>
<td>24.45</td>
</tr>
<tr>
<td>(d) ( \mu_2 = 1 )</td>
<td>( m = 0 )</td>
<td>4.20</td>
</tr>
<tr>
<td></td>
<td>( m = 1 )</td>
<td>5.20</td>
</tr>
<tr>
<td></td>
<td>( m = 2 )</td>
<td>39.55</td>
</tr>
<tr>
<td></td>
<td>( m = 3 )</td>
<td>33.10</td>
</tr>
<tr>
<td>(e) ( \mu_2 = 2 )</td>
<td>( m = 0 )</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>( m = 1 )</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>( m = 2 )</td>
<td>39.05</td>
</tr>
<tr>
<td></td>
<td>( m = 3 )</td>
<td>38.50</td>
</tr>
<tr>
<td>(f) ( \mu_2 = 2 )</td>
<td>( m = 0 )</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>( m = 1 )</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>( m = 2 )</td>
<td>35.60</td>
</tr>
<tr>
<td></td>
<td>( m = 3 )</td>
<td>37.55</td>
</tr>
<tr>
<td>(g) ( \mu_2 = 4 )</td>
<td>( m = 0 )</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>( m = 1 )</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>( m = 2 )</td>
<td>37.50</td>
</tr>
<tr>
<td></td>
<td>( m = 3 )</td>
<td>38.90</td>
</tr>
<tr>
<td>(h) ( \mu_2 = 4 )</td>
<td>( m = 0 )</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>( m = 1 )</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>( m = 2 )</td>
<td>36.35</td>
</tr>
<tr>
<td></td>
<td>( m = 3 )</td>
<td>37.55</td>
</tr>
</tbody>
</table>

Notes: See the Notes of Table 2.3.
Table 2.6: Model selection frequencies of the uncorrected and corrected versions of the BP and STR-based procedures. Uncorrelated but heterogenous data and errors across segments, $m = 2$, $\epsilon_R = 0.15$.

<table>
<thead>
<tr>
<th>Model</th>
<th>Uncorrected</th>
<th>AC-Correct</th>
<th>$\sigma(u)^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Choice</td>
<td>BP</td>
<td>LM$_1$</td>
</tr>
<tr>
<td>$\sigma^2_2 = 2$</td>
<td>$\sigma^2_2 = 2$</td>
<td>$\sigma^2_2 = 2$</td>
<td>$\sigma^2_2 = 2$</td>
</tr>
<tr>
<td></td>
<td>$c_2 = 2$</td>
<td>$c_2 = 2$</td>
<td>$c_2 = 2$</td>
</tr>
<tr>
<td></td>
<td>$T = 120$</td>
<td>$T = 120$</td>
<td>$T = 120$</td>
</tr>
<tr>
<td>$\sigma^2_2 = 2$</td>
<td>$m = 0$</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>$m = 1$</td>
<td>8.40</td>
<td>26.00</td>
</tr>
<tr>
<td></td>
<td>$m = 2$</td>
<td><strong>90.10</strong></td>
<td><strong>52.85</strong></td>
</tr>
<tr>
<td>$\sigma^2_2 = 2$</td>
<td>$m = 0$</td>
<td>0.00</td>
<td>1.10</td>
</tr>
<tr>
<td></td>
<td>$m = 1$</td>
<td>0.05</td>
<td>4.70</td>
</tr>
<tr>
<td></td>
<td>$m = 2$</td>
<td><strong>98.80</strong></td>
<td><strong>93.85</strong></td>
</tr>
<tr>
<td>$\sigma^2_2 = 2$</td>
<td>$m = 0$</td>
<td>0.00</td>
<td>24.30</td>
</tr>
<tr>
<td></td>
<td>$m = 1$</td>
<td>19.55</td>
<td>48.75</td>
</tr>
<tr>
<td></td>
<td>$m = 2$</td>
<td><strong>78.80</strong></td>
<td><strong>26.10</strong></td>
</tr>
<tr>
<td>$\sigma^2_2 = 2$</td>
<td>$m = 0$</td>
<td>0.00</td>
<td>2.20</td>
</tr>
<tr>
<td></td>
<td>$m = 1$</td>
<td>0.35</td>
<td>26.40</td>
</tr>
<tr>
<td></td>
<td>$m = 2$</td>
<td><strong>98.30</strong></td>
<td><strong>69.75</strong></td>
</tr>
<tr>
<td>$\sigma^2_2 = 4$</td>
<td>$m = 0$</td>
<td>0.00</td>
<td>34.70</td>
</tr>
<tr>
<td></td>
<td>$m = 1$</td>
<td>27.55</td>
<td>35.05</td>
</tr>
<tr>
<td></td>
<td>$m = 2$</td>
<td><strong>71.35</strong></td>
<td><strong>30.10</strong></td>
</tr>
<tr>
<td>$\sigma^2_2 = 4$</td>
<td>$m = 0$</td>
<td>0.00</td>
<td>3.40</td>
</tr>
<tr>
<td></td>
<td>$m = 1$</td>
<td>2.85</td>
<td>13.95</td>
</tr>
<tr>
<td></td>
<td>$m = 2$</td>
<td><strong>95.95</strong></td>
<td><strong>82.50</strong></td>
</tr>
<tr>
<td>$\sigma^2_2 = 4$</td>
<td>$m = 0$</td>
<td>0.00</td>
<td>36.05</td>
</tr>
<tr>
<td></td>
<td>$m = 1$</td>
<td>44.35</td>
<td>50.85</td>
</tr>
<tr>
<td></td>
<td>$m = 2$</td>
<td><strong>54.25</strong></td>
<td><strong>12.85</strong></td>
</tr>
<tr>
<td>$\sigma^2_2 = 4$</td>
<td>$m = 0$</td>
<td>0.00</td>
<td>4.80</td>
</tr>
<tr>
<td></td>
<td>$m = 1$</td>
<td>6.50</td>
<td>42.25</td>
</tr>
<tr>
<td></td>
<td>$m = 2$</td>
<td><strong>91.80</strong></td>
<td><strong>51.25</strong></td>
</tr>
</tbody>
</table>

Notes: The table contains selection frequencies in per cent based on 2000 Monte Carlo replications. Columns labelled “Uncorrected” contain the results when not correcting for serial correlation, and columns labelled “AC-Correct” correspond to tests that correct for serial correlation either nonparametrically or parametrically. The column labelled “BP$_{PW}$” contains the results of the BP sequential test when prewhitening is applied before estimating the long-run covariance matrix. The column labelled “$\sigma(u)^2$” contains the results of the BP sequence when the correct specification of the test is applied. The columns labelled “LM$_1$” and “LM$_3$” correspond to the STR-based tests that make use of the first-order and third-order Taylor expansions, respectively.
Table 2.7: Model selection frequencies of the uncorrected and corrected versions of the BP and STR-based procedures. Uncorrelated but heterogenous data and errors across segments, \( m = 2, \epsilon_R = 0.20 \).

<table>
<thead>
<tr>
<th>Model</th>
<th>Uncorrected</th>
<th>AC-Correct</th>
<th>( \sigma^2(u) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma^2 = 2 )</td>
<td>Choice</td>
<td>BP</td>
<td>LM_1</td>
</tr>
<tr>
<td>(a) ( \sigma^2 = 2 )</td>
<td>( m = 0 )</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>( T = 120 )</td>
<td>( m = 1 )</td>
<td>26.75</td>
<td>36.95</td>
</tr>
<tr>
<td>(b) ( \sigma^2 = 2 )</td>
<td>( m = 0 )</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>( T = 240 )</td>
<td>( m = 1 )</td>
<td>0.35</td>
<td>3.95</td>
</tr>
<tr>
<td>(c) ( \sigma^2 = 2 )</td>
<td>( m = 0 )</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>( T = 120 )</td>
<td>( m = 1 )</td>
<td>36.90</td>
<td>60.25</td>
</tr>
<tr>
<td>(d) ( \sigma^2 = 2 )</td>
<td>( m = 0 )</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>( T = 240 )</td>
<td>( m = 1 )</td>
<td>0.90</td>
<td>19.25</td>
</tr>
<tr>
<td>(e) ( \sigma^2 = 4 )</td>
<td>( m = 0 )</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>( T = 120 )</td>
<td>( m = 1 )</td>
<td>71.30</td>
<td>70.25</td>
</tr>
<tr>
<td>(f) ( \sigma^2 = 4 )</td>
<td>( m = 0 )</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>( T = 240 )</td>
<td>( m = 1 )</td>
<td>12.75</td>
<td>20.55</td>
</tr>
<tr>
<td>(g) ( \sigma^2 = 4 )</td>
<td>( m = 0 )</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>( T = 120 )</td>
<td>( m = 1 )</td>
<td>73.90</td>
<td>79.50</td>
</tr>
<tr>
<td>(h) ( \sigma^2 = 4 )</td>
<td>( m = 0 )</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>( T = 240 )</td>
<td>( m = 1 )</td>
<td>15.30</td>
<td>43.15</td>
</tr>
</tbody>
</table>

Notes: See the Notes of Table 2.6.
Table 2.8: Model selection frequencies at 5% nominal level, data generated from (2.14).

<table>
<thead>
<tr>
<th>Choice</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>LM₁</th>
<th>LM₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>m = 0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.10</td>
<td>29.80</td>
<td>1.10</td>
</tr>
<tr>
<td>m = 1</td>
<td>24.40</td>
<td>2.60</td>
<td>4.70</td>
<td>10.20</td>
<td>22.30</td>
</tr>
<tr>
<td>m = 2</td>
<td><strong>42.20</strong></td>
<td><strong>26.00</strong></td>
<td><strong>88.30</strong></td>
<td><strong>58.60</strong></td>
<td><strong>75.60</strong></td>
</tr>
<tr>
<td>m = 3</td>
<td>33.40</td>
<td>27.50</td>
<td>6.20</td>
<td>1.40</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*Notes:* The columns labelled "(i)", i = 1, 2, 3, refer to the Strategy (i) on page 65. The columns labelled "LM₁" and "LM₃" correspond to the STR-based tests that make use of the first-order and third-order Taylor expansions, respectively.
Chapter 3

The Effects of Institutional and Technological Change and Business Cycle Fluctuations on Seasonal Patterns in Quarterly Industrial Production Series
3.1 Introduction

Seasonal fluctuations are an important source of variation in many macroeconomic time series. When monthly or quarterly series are modelled, it is often assumed that the seasonal pattern of the series is constant over time, in which case it may be characterized by seasonal dummy variables, see Miron (1996) and Miron and Beaulieu (1996), among others. On the other hand, it has been known for a long time that seasonality in a series may change. As Kuznets (1932) remarked:

"For a number of years statisticians have been concerned with the problem of measuring changes in the seasonal behaviour of time series."

The possible causes for such changes have also been a longstanding object of interest. After examining a number of employment series from various countries and regions, Gjermoe (1931) wrote (in Norwegian):

"The strength of seasonal fluctuations has to do with the level of business activity. A month in a year of low employment is more affected by seasonality than the same month in a year of high employment." (emphasis in original)

In fact, Gjermoe (1926) already made a similar point.

The possibility that seasonality is affected by the business cycle has been reconsidered in the more recent literature. For example, Canova and Ghysels (1994) investigated this issue using quarterly US output. Their autoregressive model with seasonal dummies for the first difference of the logarithmic series contained a set of extra seasonal dummies that obtained nonzero values only if the NBER business cycle dating defined the corresponding quarter to be a recession quarter. Whether the business cycle influences the seasonal cycle was examined by testing the null hypothesis that the coefficients of the extra dummies were zero. Franses (1996, pp. 86-87) later argued that a more appropriate null hypothesis to test was that these coefficients were equal but possibly non-zero. In that case, the autoregressive model contains a business cycle influenced intercept under the null hypothesis. The empirical results of these authors suggest that the seasonal pattern in the quarterly US output series is indeed affected by the business cycle. In particular, it is found that the summer slowdown is less pronounced around business cycle peaks. A similar conclusion is reached by Cecchetti and Kashyap (1996) and Matas-Mir and Osborn (2004), using an international data set of monthly production series at the industry-level. Cecchetti, Kashyap, and Wilcox (1997) and Krane and Wascher (1999) document the same effect of the business cycle on seasonal
Changes in seasonality

patterns in US production, inventories and employment, which is attributed to the fact that during a boom the presence of capacity constraints forces firms to produce a larger fraction of output in off-peak seasons.

Business cycle fluctuations are not the only possible reason for changes in the seasonal pattern of output or employment series. In particular, technological change and changes in institutions and habits may cause changes in seasonality as well. As an example of the former, in the construction industry it has become possible to keep a construction site going year-round in countries where, for a few decades ago, work was interrupted for the winter months. As to the latter type of change, the increase in paid leisure over the last few decades has gradually changed people’s vacation habits. At least in some Scandinavian countries it has become customary to spend a week of the annual holiday in the winter. Yet another example may be the increasingly efficient use of capital and just-in-time production techniques. Many factories in Europe no longer close down for the summer vacation but keep the production process running without interruption. In all these examples, the result may have been that the seasonal pattern of, for example, output and consumption series has changed over time.

Our aim is to compare the effects on seasonality of gradual institutional and technological change with the effects attributable to the business cycle. As for the former, there do not seem to exist reliable aggregate measures for these changes. We allow for the possibility that the aggregate change is steady and continuous and simply use time as a proxy variable for it. This means that we in fact contrast “Kuznets-type” unspecified changes in seasonality with “Gjermoe-type” changes caused by fluctuations in economic activity. The main question we ask is: which of the two types is more prominent in practice, if any? We shall investigate the problem using quarterly industrial production series of the world’s leading market economies, the G7 countries, and of two Scandinavian countries.

In this work, the logarithmic time series are differenced once in order to achieve stationarity, and the differenced series are used for modelling. For most countries, these first differences, or quarterly growth rates, are dominated by seasonal variation that almost completely inundates the other features of the series, see Figures 3.1–3.9. The models we are going to build are constructed strictly to investigate the main research question formulated above about reasons for time-variation in seasonal patterns of industrial production series. Because most of the variation captured by the parameters will be seasonal variation, models based on first differences are not likely to be appropriate tools in forecasting industrial production several quarters ahead. A typical forecaster
would rather begin his analysis by transforming the original quarterly series into annual differences which emphasize the low-frequency variation he is most likely interested in, build a linear or nonlinear model for them and use that for forecasting.

The plan of the chapter is as follows. In Section 3.2, we describe the output series, focusing on the properties of their seasonal cycles. In Section 3.3, we present our main statistical tool, the time-varying smooth transition autoregressive (TV-STAR) model. In Section 3.4, we use Lagrange Multiplier (LM) tests derived from the TV-STAR framework for addressing the question of whether the changes in the seasonal patterns in the output series are due to the effects of business cycle fluctuations or to technological and institutional change or both. For all series except the US, we find convincing evidence that "Kuznets-type" unspecified change is much more important than "Gjerme-type" business cycle-induced change. In contrast to previous research, our analysis suggests that for the US seasonality does not vary over the business cycle nor has it changed over time. In Section 3.5, we specify and estimate TV-STAR models to gain further insight into when and how seasonality in the quarterly output series has changed. Section 3.6 contains final remarks.

### 3.2 Preliminaries

#### 3.2.1 Data

Our data set consists of quarterly seasonally unadjusted industrial production volume indexes for the G7 countries and for two Scandinavian countries, Finland and Sweden, taken from the OECD Main Economic Indicators. The sample period runs from 1960:1 until 2001:3, except for Canada for which the series is available only from 1961:1 and for the Scandinavian countries for which the sample runs until 2001:4. Obvious outliers in 1963:1 and 1968:2 for France and in 1969:4 for Italy are replaced by the average of the index values in the same quarter of the previous and the following year.

The importance of seasonal variation in the industrial production series may be illustrated by regressing the first differences of the logarithmic series on a set of four seasonal dummy variables. The coefficient of determination from this regression lies between 0.79 and 0.97 for the six European series and is equal to 0.68 for Canada, while it is appreciably lower only for Japan (0.21) and the United States (0.06). Incidentally, the correlation between the seven standard deviations of the fitted values (the "seasonal cycle") and the corresponding residual standard deviations (the "business cycle" plus noise) equals 0.88. This accords well with the finding of Beaulieu, MacKie-Mason,
Changes in seasonality

and Miron (1992) that countries with large seasonal cycles tend to have large business cycles as well.

We may also inspect the seasonal patterns in the series visually. Figures 3.1–3.9 show graphs of the level, the first difference per quarter, and the seasonal difference of the log industrial production series. The differenced and seasonally differenced series are multiplied by 100 to express the changes in percentage points.

A common feature for Canada and five European countries, France, Germany, Finland, Sweden and the UK, is that the seasonal variation in the industrial output series appears to have dampened over time, see panels (b). In particular, the drop in output in the third quarter and the fourth-quarter peak have become less pronounced. This is not true, however, for the remaining European country, Italy, where rather the opposite has occurred. This corresponds with results in Canova and Hansen (1995), who test for structural change in the seasonal patterns of the four European series (over the period 1960 – 1989) and find that rejections of the null hypothesis of constancy are concentrated in the third and fourth quarters. The Japanese and the US series do not show a third-quarter summer holiday slack in production. In the US, the quarterly growth in the 1990s is actually highest in the third quarter and lowest in the fourth quarter. There is no visible tendency in the amplitude of seasonal fluctuations in these countries.

3.2.2 Deterministic and stochastic seasonality

In the case of nonstationary time series, time-varying seasonal patterns may often be conveniently characterized by seasonal unit roots, see Hylleberg (1994). Autoregressive models of seasonally differenced data are capable of generating series in which the seasonal pattern evolves over time. For example, in realizations from such models “summer may become winter” or, in general, seasons may “trade places”. Structural time series models offer another way of modelling stochastically time-varying seasonality; see Harvey (1989, Chapter 6). In this approach, the time series is divided into components, of which the seasonal one is represented by a linear combination of trigonometric functions with stochastic coefficients. If these coefficients have zero variance, seasonality is deterministic.

Neither one of these alternatives, seasonal differencing or decomposing the time series, is directly applicable to our situation. The reason is that we intend to consider two types of time-varying seasonality, variation due to technological and institutional change (“unspecified change” i.e. “Kuznets-type” changes) and variation induced by cyclical fluctuations in the economic
activity ("Gjermoe-type" changes), simultaneously. This requires a model within which we can distinguish these two different sources of variation on the seasonal pattern from each other and thus compare their relative importance. The Time-Varying Smooth Transition AutoRegressive (TV-STAR) model introduced by Lundbergh, Teräsvirta, and van Dijk (2003) seems ideally suited for this purpose. In the next section we discuss this model and some of its properties. The paper that comes closest to this study as far as the modelling approach is concerned is Matas-Mir and Osborn (2004). These authors use a threshold autoregressive model in which the seasonal pattern, characterized by seasonal dummy variables, switches according to a business cycle indicator. Structural changes in seasonality are also accounted for by allowing linear trends in the coefficients of the seasonal dummy variables. An important aspect in which our approach differs from the one in Matas-Mir and Osborn (2004) is that the TV-STAR model allows for more flexible nonlinear trends in seasonality, as will become clear below.

### 3.3 The TV-STAR model

We use the TV-STAR model to investigate the source of changes in seasonal patterns of the output series, because it is capable of describing business cycle nonlinearity and structural change in the characteristics of a time series variable simultaneously. To suit our purposes, we augment the model by seasonal dummies, such that for our quarterly time series it has the following form:

\[
\Delta y_t = \left[ (\phi'_1 x_t + \delta'_1 D_t)(1 - G_1(w_t)) + (\phi'_2 x_t + \delta'_2 D_t)G_1(w_t) \right] [1 - G_2(t^*)] \\
+ \left[ (\phi'_3 x_t + \delta'_3 D_t)(1 - G_1(w_t)) + (\phi'_4 x_t + \delta'_4 D_t)G_1(w_t) \right] G_2(t^*) + \varepsilon_t,
\]

where \( y_t \) is the log-level of the industrial production index, \( w_t \) a stochastic transition variable, \( \Delta \) denotes the first differencing operator, defined by \( \Delta_k y_t \equiv y_t - y_{t-k} \) for all \( k \neq 0 \) and \( \Delta \equiv \Delta_1 \), \( x_t = (1, \bar{x}_t)' \), \( \bar{x}_t = (\Delta y_{t-1}, \ldots, \Delta y_{t-p})' \), \( D_t = (D^*_1 t, D^*_2 t, D^*_3 t)' \), \( D^*_i t \equiv D_{i,t} - D_{4,t} \), \( i = 1, 2, 3 \), \( D_{s,t}, s = 1, \ldots, 4 \) are seasonal dummy variables, with \( D_{s,t} = 1 \) when time \( t \) corresponds with season \( s \) and \( D_{s,t} = 0 \) otherwise, and \( t^* = t/T \) with \( T \) denoting the sample size. The transition functions \( G_j(s_t; \gamma_j, c_j) \), \( j = 1, 2 \), are assumed to be given by the logistic function

\[
G_j(s_t; \gamma_j, c_j) = \left( 1 + \exp \left\{ -\gamma_j (s_t - c_j)/\sigma_{s_t} \right\} \right)^{-1}, \quad \gamma_j > 0,
\]

where the transition variable \( s_t = w_t \) (\( j = 1 \)) or \( s_t = t^* \) (\( j = 2 \)), and \( \sigma_{s_t} = [\text{var}(s_{t})]^{1/2} \) makes \( \gamma_j \) scale-free. As \( s_t \) increases, the logistic function
Changes in seasonality changes monotonically from 0 to 1, with the change being symmetric around the location parameter $c_j$, as $G_j(c_j - z; \gamma_j, c_j) = 1 - G_j(c_j + z; \gamma_j, c_j)$ for all $z$. The slope parameter $\gamma_j$ determines the smoothness of the change in the value of the logistic function. As $\gamma_j \to \infty$, the logistic function $G_j(s_t; \gamma_j, c_j)$ approaches the indicator function $I[s_t > c_j]$ and, consequently, the change of $G_j(s_t; \gamma_j, c_j)$ from 0 to 1 becomes instantaneous at $s_t = c_j$. When $\gamma_j \to 0$, $G_j(s_t; \gamma_j, c_j) \to 0.5$ for all values of $s_t$.

The TV-STAR model distinguishes four regimes corresponding with combinations of $G_1(w_t)$ and $G_2(t^*)$ being equal to 0 or 1. The transition variable $w_t$ in (3.1) is assumed to be a lagged seasonal difference, $w_t = \Delta_4 y_{t-d}$, for certain $d > 0$. As this variable tracks the business cycle quite closely for our quarterly industrial production series (see panels (c) of Figures 3.1–3.9), and because the logistic function $G_j(s_t)$ is a monotonic transformation of $s_t$, the regimes associated with $G_1(\Delta_4 y_{t-d}) = 0$ and 1 will roughly correspond with recessions and expansions, respectively. Thus, using $\Delta_4 y_{t-d}$ as transition variable ensures that the TV-STAR model allows for “Gjermoe-type” change in the seasonal pattern of $y_t$.1 On the other hand, the function $G_2(t^*)$ enables the model to describe “Kuznets-type” unspecified change as well.

The reason for defining the elements of $D_t$ as $D_{s,t}^* = D_{s,t} - D_{4,t}$, $s = 1, 2, 3$, is that it effectively separates the deterministic seasonal fluctuations from the overall intercept. For example, the coefficients in $\delta_1 = (\delta_{11}, \delta_{12}, \delta_{13})'$ measure the difference between the intercept in the first three quarters of the year and the overall intercept, given by the first element of $\varphi_1$, in the regime $G_1(\Delta_4 y_{t-d}) = 0$ and $G_2(t^*) = 0$. The difference for the fourth quarter $\delta_{14}$ is obtained as $\delta_{14} = -\sum_{s=1}^{3} \delta_{1s}$. This parameterization makes it easy, for example, to test constant seasonality while allowing for a business cycle influenced intercept under the null hypothesis, cf. Franses (1996, pp. 86–87).

The general TV-STAR model in (3.1) allows both the dynamics and the seasonal properties of the growth rate of industrial production to vary both over the business cycle and over time. By imposing appropriate restrictions on either the autoregressive parameters or the seasonal dummy parameters or on both, more restrictive models can be obtained. Of particular interest here are models in which seasonality only varies either over time or over the business cycle. A model in which seasonality is constant over time is obtained if $\delta_1 = \delta_3$ and $\delta_2 = \delta_4$ in (3.1). Similarly, a model in which seasonality is constant

---

1It may be argued, however, that GNP is a more representative and more commonly used indicator of the business cycle than the industrial production. In fact, we repeated our tests described in Section 4 using lagged seasonal differences of GNP instead of $\Delta_4 y_{t-d}$ as the transition variable. The results were very similar to the ones obtained by using $\Delta_4 y_{t-d}$ and therefore are omitted.
over the business cycle is obtained by setting $\delta_1 = \delta_2$ and $\delta_3 = \delta_4$. When $\delta_1 = \delta_2 = \delta_3 = \delta_4$, seasonality is linear and constant over time. Imposing analogous restrictions on $\phi_i$, $i = 1, \ldots, 4$ results in models with constant but nonlinear, linear but time-varying, and linear and constant autoregressive dynamics, respectively. If both the seasonal patterns and the autoregressive dynamic structure are constant either over time or over the business cycle, the TV-STAR model reduces to a STAR or TV-AR model, respectively. All these restrictions are testable, as will be discussed in the next section. Often a useful restricted TV-STAR model is an additive one, containing a nonlinear and a time-varying component. For example, a model in which the seasonal dummy coefficients vary over time and the autoregressive parameters enter nonlinearly can be written as

$$\Delta y_t = \phi_1^{*'} x_t + \delta_1^{*'} D_t + \phi_2^{*'} x_t G_1(w_t) + \delta_2^{*'} D_t G_2(t^*) + \epsilon_t. \quad (3.3)$$

In Section 3.5 we will use this form for the models for the industrial production series.

On the other hand, the TV-STAR model (3.1) is restrictive in the sense that it requires any nonlinearity or structural change to be common across the autoregressive dynamics and seasonal dummies. For example, if the coefficients of both the lagged growth rates and the seasonal dummies are time-varying, the structural change is centered at the same point in time, $t^* = c_2$, and occurs at the same speed, as determined by $\gamma_2$. It is straightforward to extend or modify the model to allow for different timing and/or speed of the structural change of the two sets of coefficients. But then, models of type (3.3) do not contain that restriction because $G_1$ only controls the lag structure and $G_2$ the seasonal component.

Finally, a potential limitation of the models in (3.1) and (3.3) is that they only allow for a single pattern change in seasonality over time. The model can be generalized in a straightforward fashion to accommodate multiple changes by including additional time-varying components. For example, an additive model in which the autoregressive parameters enter nonlinearly and the seasonal dummy coefficients change over time following a mixture of two patterns is given by

$$\Delta y_t = \phi_1^{*'} x_t + \delta_1^{*'} D_t + \phi_2^{*'} x_t G_1(w_t) + \delta_2^{*'} D_t G_2(t^*) + \delta_3^{*'} D_t G_3(t^*) + \epsilon_t. \quad (3.4)$$
3.4 Changes in the seasonal pattern and their causes

3.4.1 Testing linearity and parameter constancy in the TV-STAR framework

The question posed in the introduction of this chapter about the causes of fluctuations in the seasonal pattern is addressed within the framework of the TV-STAR model (3.1), in particular by testing hypotheses about the coefficients of the model. In the previous section, it was emphasized that linearity or parameter constancy in the TV-STAR model (3.1) may be achieved by imposing equality restrictions on certain coefficient vectors $\delta_i$ and/or $\phi_i$. Note however that linearity or parameter constancy of both the seasonal pattern and the dynamic autoregressive structure also results if the smoothness parameter $\gamma_j$ in the corresponding transition function $G_j$ is set equal to zero. This is an indication of an identification problem present in the model: the TV-STAR model is only identified under the alternative, not under the null hypothesis. For a general discussion, see Hansen (1996). We follow the approach of Lundbergh et al. (2003) and circumvent the identification problem by approximating the transition functions by their first-order Taylor expansions, see also Luukkonen, Saikkonen, and Teräsvirta (1988).

Let the null hypothesis of interest be $H_0 : \gamma_1 = \gamma_2 = 0$, which is to be tested against the alternative hypothesis $H_1 : \gamma_1 > 0$ and/or $\gamma_2 > 0$. Under $H_0$, model (3.1) reduces to a seasonality-augmented linear autoregressive model, which we assume to be stationary and ergodic. Furthermore, we assume that the moment condition $E((\Delta y_t)^2(\Delta^4 y_t)^2) < \infty$ is satisfied, which is necessary for the asymptotic inference to be valid. In testing $H_0$, we only assume that $d \in \{1, \ldots, r\}$ in (3.1), that is, the true delay is unknown but assumed to be not greater than $r$. A convenient way to parameterize this assumption is by setting $w_t = \sum_{i=1}^r a_i \Delta^4 y_{t-i}$, with $a_d = 1$ and $a_i = 0$ for all $i \neq d$; see Luukkonen et al. (1988) for further details. In this case, after rearranging terms the first-order Taylor expansion of (3.1) around $H_0$ becomes

$$
\Delta y_t = \phi_1' x_t + \delta_1' D_t + (\phi_2' x_t + \delta_2' D_t)t^* + \sum_{i=1}^r (\phi_{3,i}' x_t + \delta_{3,i}' D_t) \Delta^4 y_{t-i} \\
+ \sum_{i=1}^r (\phi_{4,i}' x_t + \delta_{4,i}' D_t)t^* \Delta^4 y_{t-i} + R(\gamma_1, \gamma_2) + \varepsilon_t, \quad (3.5)
$$

where $R(\gamma_1, \gamma_2)$ is a remainder from the two Taylor expansions. Under the null hypothesis of linearity and parameter constancy, $R(\gamma_1, \gamma_2) \equiv 0$, such that this remainder does not affect the distribution theory.
Equation (3.5) is linear in parameters. Furthermore, and this is crucial, the parameter vectors \( \phi_2^* = \gamma_2 \delta_2^*(\theta) \) and \( \delta_2^* = \gamma_2 \delta_2^*(\theta) \), \( \phi_3^* = (\phi_{3,1}^*, \ldots, \phi_{3,r}^*, \ldots, \phi_{3,1}^*, \ldots, \phi_{3,r}^*)' = \gamma_1 \phi_3^*(\theta) \) and \( \delta_3^* = (\delta_{3,1}^*, \ldots, \delta_{3,r}^*, \ldots, \delta_{3,1}^*, \ldots, \delta_{3,r}^*)' = \gamma_1 \delta_3^*(\theta) \), and \( \phi_4^* = (\phi_{4,1}^*, \ldots, \phi_{4,4}^*, \ldots, \phi_{4,1}^*, \ldots, \phi_{4,4}^*)' = \gamma_1 \gamma_2 \phi_4^*(\theta) \) and \( \delta_4^* = (\delta_{4,1}^*, \ldots, \delta_{4,r}^*, \ldots, \delta_{4,1}^*, \ldots, \delta_{4,r}^*)' = \gamma_1 \gamma_2 \delta_4^*(\theta) \), where \( \delta_j^*(\theta) = \delta_j^*(\theta), j = 2, 3, 4, \) are non-zero functions of the parameters \( \theta = (\phi_1', \ldots, \phi_4', \delta_1', \ldots, \delta_4')' \).

In view of this, the original null hypothesis becomes

\[
H'_0 : \phi_2^* = \phi_{3,i}^* = \phi_{4,i}^* = 0, \quad \delta_2^* = \delta_{3,i}^* = \delta_{4,i}^* = 0, \quad i = 1, \ldots, r
\]

in the transformed equation (3.5). The standard Lagrange Multiplier statistic for testing \( H'_0 \) has an asymptotic \( \chi^2 \) distribution with \( (p + 4)(1 + 2r) \) degrees of freedom under the null hypothesis. In practice, an \( F \)-version of the test is recommended because its size properties in small and moderate samples are much better than those of the \( \chi^2 \)-based test statistic, especially when the number of parameters tested becomes large relative to the sample size. It should be noted that, depending on the values of \( p \) and \( r \), certain terms \( \phi_{3,i,0} \Delta_4 y_{t-i} \) and \( \phi_{3,i,j} \Delta y_{t-j} \Delta_4 y_{t-i} \) should be excluded from (3.5) to avoid perfect multicollinearity.

In order to keep the notation simple, we so far have discussed the case of the standard logistic function (3.2) being the transition function. It is useful to generalize this slightly as follows. Let

\[
G_j(s_t; \gamma_j, c_j) = \left( 1 + \exp \left\{ -\frac{\gamma_j}{\sigma_j} \prod_{i=1}^{k} (s_t - c_{ji}) \right\} \right)^{-1}, \quad \gamma > 0, c_{j1} \leq \ldots \leq c_{jk}.
\]

This function allows more flexibility in the transition. When we test linearity against the TV-STAR model (3.1) with (3.6), a first-order Taylor expansion of (3.6) leads to terms with higher powers of \( \Delta_4 y_{t-j} \) and \( t^* \) in equation (3.5); see, for example, Luukkonen et al. (1988), Granger and Teräsvirta (1993, Chapter 6) or Lundbergh et al. (2003). The dimension of the null hypothesis increases linearly in \( k \), which implies that for small sample sizes such as the one available here, the tests for \( k > 1 \) can only be computed for fairly small values of \( p \) and \( r \). Below we report results for \( k = 1 \) and, whenever possible, for \( k = 3 \). The corresponding statistics are denoted as \( \text{LM}_k \).

Finally, it should be pointed out that the lag length \( p \) in (3.1) is unknown. It is selected from the linear seasonality-augmented autoregressive model using BIC with the maximum order set equal to \( p_{\text{max}} = 12 \). As remaining residual autocorrelation may be mistaken for nonlinearity, we apply the Breusch-Godfrey LM test to examine the joint significance of the
first 12 residual autocorrelations in the model that is preferred by the BIC. If necessary, the lag length $p$ is increased until the null hypothesis of no error autocorrelation can no longer be rejected at the 5% significance level. Testing is carried out conditionally on the selected lag length $\hat{p}$.

3.4.2 Testing hypotheses of interest

The test just described is a general linearity test within our maintained TV-STAR model (3.1). The main interest in this work, however, lies in testing subhypotheses that place restrictions on the seasonal dummy variables. We may also set certain parameter vectors to zero (null vectors) \textit{a priori}. This leads to a maintained model that is a submodel of (3.1). In particular, we are interested in testing constant seasonality against the alternative that the seasonal pattern changes smoothly over time, conditional on the assumption that seasonality is not affected by the business cycle and that the autoregressive structure does not change. In terms of the parameters in (3.5), the corresponding null hypothesis is

$$H_0^{TV-AR,D_s,t}: \delta_2^* = 0 \text{ assuming } \phi_2^* = \phi_3^* = \phi_4^* = 0, \delta_3^* = \delta_4^* = 0, i = 1, \ldots, r.$$ 

Another hypothesis of interest is testing constant seasonality against the alternative that the seasonal pattern is affected by the business cycle only:

$$H_0^{STAR,D_s,t}: \delta_{3,i}^* = 0 \text{ assuming } \phi_2^* = \phi_3^* = \phi_4^* = 0, \delta_2^* = \delta_4^* = 0, i = 1, \ldots, r.$$ 

A test against the joint alternative of smooth change and fluctuations ascribed to the business cycle may be formed accordingly. The corresponding null hypothesis is denoted as $H_0^{TV-STAR,D_s,t}$.

These tests are based on the assumption of linearity of the dynamic structure of the time series $\Delta y_t$. But then, the first difference of the volume of industrial production may be a nonlinear or time-varying process. One way of accounting for this possibility is to relax the zero restrictions on $\phi_2^*, \phi_3^*, \phi_4^*$ in the above tests. This variant of $H_0^{TV-AR,D_s,t}$ becomes

$$H_0^{TV-AR^*,D_s,t}: \delta_2^* = 0 \text{ assuming } \delta_{3,i}^* = \delta_{4,i}^* = 0, i = 1, \ldots, r.$$ 

Similarly, we have

$$H_0^{STAR^*,D_s,t}: \delta_{3,i}^* = 0 \text{ assuming } \delta_2^* = \delta_{4,i}^* = 0, i = 1, \ldots, r.$$ 

While testing the resulting hypotheses is not difficult in practice, this may not be an optimal way to proceed. Instead it may be better to test our two
3.4. Changes in the seasonal pattern and their causes

competing hypotheses concerning seasonality within a model which explicitly models the changes in the autoregressive structure, either as a function of time (TV-AR) or as a function of the business cycle (STAR). In that case, we may begin by testing linearity against STAR and TV-AR. The relevant null hypotheses (assuming constant seasonality and unknown delay $d$) are

$$H_{0}^{STAR, \Delta y_{t-j}}: \phi_{3,i}^* = 0 \text{ assuming } \phi_2^* = \phi_{4,i}^* = 0, \delta_2^* = \delta_{3,i}^* = \delta_{4,i}^* = 0, i = 1, \ldots, r,$$
and

$$H_{0}^{TV-AR, \Delta y_{t-j}}: \phi_{3,i}^* = 0 \text{ assuming } \phi_2^* = \phi_{4,i}^* = 0, \delta_2^* = \delta_{3,i}^* = \delta_{4,i}^* = 0, i = 1, \ldots, r,$$

respectively. Assume for a moment that $H_{0}^{STAR, \Delta y_{t-j}}$ is rejected and $H_{0}^{TV-AR, \Delta y_{t-j}}$ is not. This implies that the dynamic behaviour of the process, excluding seasonality, may be adequately characterized by a STAR model. We subsequently specify, estimate and evaluate a STAR model for $\Delta y_t$. The issue is now the constancy of the coefficients of the seasonal dummy variables in the STAR model. The maintained model may be written as follows:

$$\Delta y_t = \phi_1' x_t + \phi_2' x_t G_1(s_t) + \{ \delta_1 + \delta_2 G_2(\Delta_4 y_{t-l}) + \delta_3 G_3(t^*) + \delta_4 G_2(\Delta_4 y_{t-l})G_3(t^*) \}' D_t + \varepsilon_t,$$  \hfill (3.7)

where the transition functions $G_2(\Delta_4 y_{t-l})$, $l > 0$, and $G_3(t^*)$ are logistic functions as in (3.6). Note that in principle we can choose either $s_t = \Delta_4 y_{t-d}$ or $s_t = t^*$ in (3.7) given the type of previous rejections. The relevant parameter constancy hypotheses can now be formulated within equation (3.7) in terms of the slope parameters in the transition functions $G_2(\Delta_4 y_{t-l})$, $l > 0$, and $G_3(t^*)$ or in terms of the coefficient vectors $\delta_2$, $\delta_3$ and $\delta_4$. Asymptotic theory for inference requires the assumption that the null model, (3.7) with $\delta_2 = \delta_3 = \delta_4 = 0$, is stationary and ergodic. Testing is based on the first-order Taylor approximation of $G_2(\Delta_4 y_{t-l})$ and $G_3(t^*)$ as described in Section 3.4.1; for a general account of STAR model misspecification tests, see, for example, Teräsvirta (1998).

### 3.4.3 Results

Table 3.1 reports $p$-values of the $F$-statistics for testing $H_{0}^{TV-AR, \Delta y_{t-j}}$, $H_{0}^{STAR, \Delta y_{t-j}}$ and $H_{0}^{TV-STAR, \Delta y_{t-j}}$ based on a linear null model. The column headings LM1 and LM3 correspond to tests based on the first-order Taylor expansion of the transition function (3.6) with $k = 1$ and 3, respectively. The row headings $D_{s,t}$ and $\Delta y_{t-j}$ correspond to tests involving the seasonal pattern only and the autoregressive
coefficients only, respectively. All tests are computed with the maximum value of the unknown delay \( r \) set equal to 4. A broad classification of the results by the magnitude of the \( p \)-values appears in the upper panel of Table 3.2.

Plenty of evidence is found to support the argument that seasonality is changing for unspecified reasons, including institutional and technological change and actions by the statistics producer, proxied by the time variable. The results for the \( \text{LM}_1 \) statistic are mixed, but \( \text{LM}_3 \) rejects the null hypothesis \( H_0^{\text{TV-AR,}D_{s,t}} \) at the 0.01 level for all series except the US. On the other hand, there is much less evidence to support the notion that seasonality varies with the business cycle, as the \( p \)-values for the tests corresponding to \( H_0^{\text{STAR,}D_{s,t}} \) are considerably larger for all nine countries. The only occasion in which a \( p \)-value lies below 0.01 is \( \text{LM}_3 \) for Japan. For Japan, there is in fact substantial evidence of both nonlinearity and parameter nonconstancy in the series. For the other eight countries, it seems that business cycle fluctuations are not a major cause for changes in the seasonal pattern. Finally, another fact obvious from Table 3.1 and the upper panel of Table 3.2 worth mentioning is that testing against both types of changes in seasonality jointly has an adverse effect on power. More information is gained by looking at the two alternatives separately.

Two objections may be made at this point. First, seasonality may not be fully explained by the seasonal dummy variables. A part of the seasonal variation may be absorbed in (or explained through) the autoregressive dynamic structure of model (3.1). Pierce (1978) discussed this possibility in connection with seasonal adjustment of economic time series. This variation may be related to the business cycle. Second, results on testing linearity against STAR in Table 3.1 (cells \( \Delta y_{t-j}, \text{STAR} \)), suggest that the dynamic behaviour of some of the industrial production series may be nonlinear. For other series a case can be made for a TV-AR process, that is, the dynamic behaviour may be time-varying because of phenomena proxied by time. It may therefore be argued that the results just presented are affected by misspecification of the null model and that in order to avoid this, it should already accommodate non-seasonal nonlinearity.

This possibility can be considered by first modelling nonseasonal nonlinearity and carrying out the tests of constancy of the seasonal parameters within the nonlinear model, as discussed in Section 3.4.2. We proceed as

\[ \text{An obvious modification of the tests presented in Table 3.1 would be to allow the overall intercept to be affected by the business cycle when testing for "Gjermoe"-type changes in seasonality, and to allow for the intercept to be time-varying when testing for "Kuznets"-type changes in seasonality, cf. Franses (1996, pp. 86-87). Results from these tests are very similar to the ones shown in Table 3.1 and are therefore omitted.} \]
3.4. Changes in the seasonal pattern and their causes

follows. Consider the row of linearity tests ($\Delta y_{t-j}$) in Table 3.1. We choose the type of transition variable, either a lag of $\Delta_y$ or $t^*$, for each series by comparing the $p$-values of the tests against STAR and against TV-AR. If the $p$-value of the test against STAR is smaller than the one for TV-AR, we choose a STAR model, otherwise we proceed with a TV-AR model. In this case, both $LM_k$ statistics yield the same result for France, Japan, Sweden, the UK (TV-AR), Germany and the US (STAR). For Canada, Finland and Italy, we obtain conflicting results but, considering the two statistics jointly, it appears that the evidence for STAR is somewhat stronger than the evidence for TV-AR for Canada, while the reverse holds for Italy. Next, we specify and estimate a STAR or TV-AR model for $\Delta_y$ with the seasonal dummies only entering linearly as given in (3.7), following the modelling strategy described in, for example, van Dijk, Teräsvirta, and Franses (2002). We then test the constancy of the coefficients of the seasonal dummy variables within this model. The $p$-values of the appropriate test statistics can be found in Table 3.3, where the first column of the table gives the transition variable used in the null model. The lower panel of Table 3.2 gives a summary of the results.

It is seen that the basic message is still the same, although the $p$-values are somewhat higher than before. Obviously, some of the seasonal variation has been absorbed by the re-specified dynamic structure of the model. Nevertheless it seems that the seasonal parameters are still changing over time for unspecified reasons rather than as a function of the cyclical fluctuations in the economy. At the 5% level, $H_{t}^{\text{STAR},D,z,t}$ can only be rejected for one of the tests for Japan and the UK. Rejections against TV-AR are still the rule, in particular when the LM$_3$ test is considered. Allowing for time-varying or nonlinear autoregressive dynamics eliminates the evidence for "Kuznets-type" change in the seasonal pattern only for the UK and the US. In those cases, the null model thus influences our view of the situation.

Putting all this together, our general conclusion is that the institutional, technological and other changes proxied by time are the main cause for changes in seasonal pattern in the output series of Sweden, Finland and G7 countries. The results of this section also illustrate the fact that our conclusions to some extent depend on the model used for carrying out the relevant tests. This may not be surprising, and mentioning this fact may even sound trivial. But then again, we may also argue that our general conclusion seems remarkably robust to the choice of the null model.
3.5 Modelling changing seasonal patterns by TV-STAR models

Our test results in the previous section clearly show that seasonal patterns in the output series are not constant over time. In this section, our aim is to characterize this change with a parametric model, instead of just demonstrating its existence through a number of hypothesis tests. We will attempt to build an adequate TV-STAR model for each of the series and focus on the components related to seasonal variation.

As the TV-STAR model is a rather flexible nonlinear model, we need a coherent modelling strategy or modelling cycle in order to arrive at an acceptable parameterization. We choose the "specific-to-general" strategy of Lundbergh et al. (2003). The main features of this modelling cycle are the following. First, starting with a seasonality-augmented linear autoregressive model, test linearity against STAR ($\Delta_4 y_{t-d}$ being the transition variable, where the value of $d$ is varied to determine the appropriate value of the delay parameter) and TV-AR ($t^*$ being the transition variable). Choose the submodel against which the rejection is strongest (if it is strong enough, otherwise accept the linear model). Estimate the chosen model; this involves repeated estimation while reducing the size of the model through imposing exclusion and equality restrictions on parameters. Evaluate the estimated model by subjecting it to a number of misspecification tests. The results may either indicate that the estimated model is adequate or they may point at the necessity of extending the model further, for example towards a full TV-STAR model. A detailed account of the modelling strategy can be found in Lundbergh et al. (2003). Below, we first illustrate the various steps in the modelling cycle for the UK industrial production and then provide a brief summary of the results obtained for the other countries.

3.5.1 A TV-STAR model for UK industrial production

The test results for the UK in Table 3.1 indicate that both linearity and constancy can be rejected for both the seasonal dummy parameters and the lagged autoregressive terms at conventional significance levels. See also the upper panel of Table 3.4, which reports results from linearity tests against STAR with $\Delta_4 y_{t-d}$ for $d = 1, \ldots, 4$ as transition variable. As the $p$-value of the parameter constancy test applied to the seasonal dummies is the smallest, we start with specifying a TV-AR model where only the seasonal pattern is allowed to change over time. The second panel of Table 3.4 contains results from misspecification tests of parameter constancy and no remaining
nonlinearity in this model. It can be seen that linearity of the autoregressive parameters is rejected, where the tests indicate that $\Delta^4 y_{t-1}$ is the appropriate transition variable. In the resulting model, we find evidence for additional time-variation in the seasonal pattern, see the misspecification tests in the third panel of Table 3.4. Hence, we specify a model which includes two TV-components operating on the seasonal dummy coefficients and a single STAR-component (with $\Delta^4 y_{t-1}$ as transition variable) operating on the coefficients of the lagged first differences. After deleting insignificant lagged first differences and intercepts, the final specification is:

$$\Delta y_t = 1.12 + 1.65D_{1,t}^* - 2.68D_{2,t}^* - 10.2D_{3,t}^* + 0.37\Delta y_{t-2} + 0.20\Delta y_{t-4}$$

$$+ (- 6.57D_{1,t}^* + 2.68D_{2,t}^* + 10.2D_{3,t}^*) \times G_1(t^*; \gamma_1, c_1) + (- 0.69 - 0.37\Delta y_{t-2}$$

$$+ 0.13\Delta y_{t-3} - 0.27\Delta y_{t-4} - 0.13\Delta y_{t-5} - 0.15\Delta y_{t-7}) \times G_2(\Delta^4 y_{t-1}; \gamma_2, c_2)$$

$$+ (4.95D_{1,t}^* - 5.68D_{2,t}^* - 2.37D_{3,t}^*) \times G_3(t^*; \gamma_3, c_3) + \tilde{\epsilon}_t,$$

(3.8)

$$G_1(t^*; \gamma_1, c_1) = (1 + \exp\{-2.02(t^* - 0.44)/\sigma_{t^*}\})^{-1},$$

(3.9)

$$G_2(\Delta^4 y_{t-1}; \gamma_2, c_2) = (1 + \exp\{-19.5(\Delta^4 y_{t-1} + 0.37)/\sigma_{\Delta^4 y_{t-1}}\})^{-1},$$

(3.10)

$$G_3(t^*; \gamma_3, c_3) = (1 + \exp\{-500(t^* - 0.31)/\sigma_{t^*}\})^{-1},$$

(3.11)

\[\hat{\sigma}_e = 1.65, \hat{\sigma}_{TV-STAR/AR} = 0.76, SK = 0.34 (0.042), EK = 1.29 (5.5E - 4),\]

\[JB = 13.7 (1.0E - 3), \text{LMsc}(1) = 0.24 (0.62), \text{LMsc}(4) = 0.59 (0.67),\]

\[\text{LMsc}(12) = 0.69 (0.76), \text{ARCH}(1) = 0.19 (0.66), \text{ARCH}(4) = 16.4 (2.5E - 3),\]

\[\text{AIC}_{TV-STAR/AR} = -0.39, \text{AIC}_{TV-STAR/AR} = 0.15,\]

where OLS standard errors are given in parentheses below the parameter estimates, \(\tilde{\epsilon}_t\) denotes the regression residual at time \(t\), \(\hat{\sigma}_e\) is the residual standard deviation, \(\hat{\sigma}_{TV-STAR/AR}\) is the ratio of the residual standard deviations in the estimated TV-STAR model (3.8) and the best fitting subset AR model, SK is skewness, EK excess kurtosis, JB the Jarque-Bera test of normality of the residuals, LMsc\((j)\) is the LM test for no residual autocorrelation up to and including lag \(j\), ARCH\((q)\) is the LM test of no ARCH effects up to order \(q\), and AIC\(_{TV-STAR/AR}\) and BIC\(_{TV-STAR/AR}\) are differences between the Akaike and Schwarz Information Criteria, respectively, of the estimated TV-STAR and the AR models. The numbers in parentheses following the test statistics are \(p\)-values. The misspecification tests in the bottom panel of Table
Changes in seasonality

3.4 indicate that the model is adequate, at least in the sense that parameter constancy and no remaining nonlinearity are not rejected.

The two structural changes in the seasonal pattern have quite different characteristics. On the one hand, an almost instantaneous change has occurred in 1975, which involves a strengthening of the seasonal pattern mainly in the first and second quarters: whereas initially the average growth rate for the first (second) quarter was larger (smaller) than the overall mean by only 1.65% (2.68%), after the change in 1975 this equals 6.59% (8.36%). On the other hand, the model indicates the presence of a very smooth structural change, which takes the entire 40-year sample period to be completed. This change effectively eliminates the seasonal effect in the first quarter in the sense that when $G_1 = G_3 = 1$ the coefficient for $D_{1,t}$ is very close to zero. It also involves a strong reduction in the amplitude of the seasonal pattern for the third and fourth quarters. The (implied) deviation of the mean during the third (fourth) quarter changes from $-10.23$ (11.27) when $G_1 = G_3 = 0$ via $-12.60$ (14.38) when $G_1 = 0$ and $G_3 = 1$ to $-2.37$ (8.02) when $G_1 = G_3 = 1$.

Finally, the STAR component in the model effectively captures the business cycle nonlinearity in the series. From panel (d) in Figure 3.17, it can be seen that periods when $G_2 = 0$ roughly coincide with recessionary periods in the UK. Apart from a lower mean growth rate, the dynamic properties of the series also are different during recessions. Note that a ± restriction is imposed on the coefficients of $\Delta y_{t-2}$ such that it equals zero when $G_2 = 1$, while the effective coefficient on $\Delta y_{t-4}$ is close to zero in this case as well.

Table 3.5 contains some summary statistics for the estimated AR, TV-AR and TV-STAR models. These illustrate that adding time-varying and nonlinear components do not improve the properties of the residuals, in the sense that residuals in the TV-AR model and TV-STAR models suffer from significant skewness and excess kurtosis, such that normality is strongly rejected. Both the Akaike and the Schwarz Information Criteria prefer the elaborate TV-STAR model.
3.5. Modelling changing seasonal patterns by TV-STAR models

3.5.2 Other countries

In this section we present the estimated models for the remaining countries in full detail together with a brief account of the most important modelling events or decisions made during the modelling cycle. All models reported here are estimated over the sample period 1963:2 - 2001:3 (154 observations), except for Canada for which the estimation period is 1964:2 - 2001:3 and two Scandinavian countries that have also the observation for 2001:4 available. Misspecification tests are given in Table 3.6. To illustrate the main implications of the models for the seasonal patterns in industrial production in these countries, panel (b) of Figures 3.10-3.18 shows the seasonal component in the estimated TV-STAR model and the deterministic components of the seasonality-augmented linear AR models. A striking feature apparent in all figures is that the latter have a much smaller amplitude than the corresponding components in the nonlinear models. Obviously, in a linear model, the parameters of the dynamic structure have to explain a greater part of the seasonality as the deterministic structure is assumed constant over time. Also, in linear models a part of seasonal variation may remain unexplained. In this respect, we note that the residual standard deviation of the nonlinear model is in all cases about 70 - 80% of that of the corresponding linear AR model. Other panels of Figures 3.10-3.18 depict the series itself, the residuals from the TV-STAR model and the best fitting subset AR model, and the transition functions.

Canada

In the seasonality-augmented linear AR model for Canada, linearity of the seasonal dummy coefficients is rejected most strongly against “Kuznets”-type unspecified change; see Table 3.1. However, a TV-AR model with the standard logistic transition function (3.2) does not satisfactorily describe the variation in the seasonal pattern. This is not surprising given the way the seasonal pattern evolves, as shown in panel (b) of Figure 3.1. A TV-AR model with a generalized logistic function (3.6) with \( k = 2 \) also is inadequate, because the decline in the amplitude of seasonal fluctuations after 1978 is different (both in terms of magnitude and speed) from the increase during the first part of the sample, as shown in panel (c) of Figure 3.1. We therefore use an additive TV-AR model with two standard logistic functions. After sequentially deleting insignificant coefficients and increasing the maximum lag order to 11 to capture remaining autocorrelation in the residuals, we obtain the model
\[ \Delta y_t = 1.50 - 1.44 D_{1,t}^* - 1.44 D_{3,t}^* + 0.29 \Delta y_{t-1} \]
\[ + 0.37 \Delta y_{t-2} - 0.15 \Delta y_{t-3} - 0.24 \Delta y_{t-5} - 0.18 \Delta y_{t-11} \]
\[ + (-1.01 + 5.59 D_{1,t}^* - 7.37 D_{2,t}^* - 3.83 D_{3,t}^*) \times G_1(t^*; \gamma_1, \epsilon_1) \]
\[ + (-1.40 D_{1,t}^* + 3.43 D_{2,t}^* + 1.48 D_{3,t}^*) \times G_2(t^*; \gamma_2, \epsilon_2) + \tilde{\varepsilon}_t, \]

\[ G_1(t^*; \gamma_1, \epsilon_1) = (1 + \exp\{-3.38 (t^* - 0.24) / \sigma_t\})^{-1}, \]
\[ G_2(t^*; \gamma_2, \epsilon_2) = (1 + \exp\{-30.2 (t^* - 0.68) / \sigma_t\})^{-1}, \]

The deterministic component obtained from the model for Canadian output accords with the visual information in Figure 3.1: the amplitude of the seasonal pattern first slowly increases until a rapid decrease takes place in the late 1980s. The parameter estimates in (3.12) and the graphs in Figure 3.10 show that the model captures the observed changes in the seasonal pattern. The misspecification tests of no remaining nonlinearity and parameter constancy reported in Table 3.6 do not lead to rejections of their respective null hypotheses.

**Finland**

The strongest rejection of linearity in the linear model for Finland is obtained when the alternative is that the coefficients of the seasonal dummy variables are time-varying. A TV-AR model of this type is estimated. The tests of no additive nonlinearity indicate that both the parameters of autoregressive terms and seasonal components are varying with the business cycle indicator and thus this model should be augmented by a STAR component. Fitting the corresponding TV-STAR model to the data results in a model which is still inadequate: it appears that the specification of the dummy variables is not satisfactory (the p-value of the parameter constancy test LM1 equals 0.022). Adding another time-varying component for the seasonal dummies finally results in the following model:
3.5. Modelling changing seasonal patterns by TV-STAR models

\[
\Delta y_t = 0.96 + 2.21 D_{1,t} + 0.11 \Delta y_{t-1} + 0.14 \Delta y_{t-2} + 0.44 \Delta y_{t-4} - 0.17 \Delta y_{t-6} +
\]
\[
+ (-7.41 D_{2,t} + 6.43 D_{3,t} + 0.55 \Delta y_{t-1} - 0.38 \Delta y_{t-2} - 0.34 \Delta y_{t-5} +
\]
\[
+ 0.52 \Delta y_{t-6} - 0.28 \Delta y_{t-7}) \times G_1(\Delta_4 y_{t-2}; \gamma_1, \gamma_2) - 12.8 D_{3,t} \times G_3(t^*; \gamma_3, \gamma_4) +
\]
\[
+ (-6.57 D_{1,t} + 3.71 D_{2,t} + 11.2 D_{3,t}) \times G_2(t^*; \gamma_2, \gamma_3) + \xi_t,
\]

where \( \xi_t \) is the error term.

The estimate of the location parameter in the transition function \( G_1 \) is high: \( \hat{\gamma}_1 = 7.85 \). This means that the corresponding nonlinear component only contributes to \( \Delta y_t \) when the industrial production is growing very strongly. High growth causes a shift in the seasonal pattern: during those periods the typical slack in the third quarter is less pronounced than it normally is. This effect is clearly visible from Figure 3.11(b): see, for example, the periods around 1965, 1970 and 1980, and the year 1973. Clearly, during the expansion, the existing production capacity has been utilized more fully during the holiday season when the demand has been high. This corresponds to the findings of Cecchetti and Kashyap and Matas-Mir and Osborn. On the other hand, it appears that the production of the second quarter has been less strong than normal; see the negative coefficient \(-6.57\) of \( D_{2,t} \). This, however, is most likely due to the fact that the production capacity has been in full use. Owing to high growth on the second half of the 1990s, the seasonal component has been rather irregular during that period. The estimates of the transition function \( G_2 \) show that the "holiday slack" has been reduced over time, obviously due to the tendency of the firms to reduce inventories and produce to order. The estimated transition function is very
Changes in seasonality

smooth and the amplitude of the seasonal pattern has steadily decreased since
the end of the 1970s. The remaining nonlinear component with transition
function $G_3$ characterizes the early increase in the amplitude of seasonality
related to the third quarter. That is most likely due to an institutional change
in the form of increased amount of paid leisure for the workers.

**France**

In the linear model for France, parameter constancy is rejected for both
the seasonal dummies and the lagged first differences, although rejection is
stronger for the deterministic terms. In a TV-AR model in which only the
seasonal pattern is allowed to change, the misspecification tests still reject
parameter constancy for the seasonal dummy coefficients rather strongly, while
linearity of the autoregressive parameters also seems untenable. Sequentially
including a second time-varying component for the dummies and a STAR
component for the lagged dependent variables, we arrive at the following spe­
cification after recursively deleting insignificant coefficients:

\[
\Delta y_t = -4.81 D^*_{1,t} + 3.66 D^*_{2,t} - 21.8 D^*_{3,t} + 0.92 \Delta y_{t-1} - 1.83 \Delta y_{t-4} - 0.69 \Delta y_{t-5} \\
+ 1.86 \Delta y_{t-8} + (-1.02 + 5.79 D^*_{1,t} - 3.32 D^*_{2,t}) \times G_1(t*; \gamma_1, c_1) \\
+ (-5.98 D^*_{1,t} + 11.4 D^*_{2,t}) \times G_2(t*; \gamma_2, c_2) + (1.50 - 0.92 \Delta y_{t-1} + 1.83 \Delta y_{t-4} \\
+ 0.69 \Delta y_{t-5} - 0.15 \Delta y_{t-6} - 1.86 \Delta y_{t-8}) \times G_3(\Delta y_{t-1}; \gamma_3, c_3) + \bar{\varepsilon}_t, \\
(3.19)
\]

\[
G_1(t*; \gamma_1, c_1) = (1 + \exp\{-5.55 (t* - 0.30) / \sigma_{\Delta 1}\})^{-1}, \\
(2.46) \\
G_2(t*; \gamma_2, c_2) = (1 + \exp\{-2.57 (t* - 0.58) / \sigma_{\Delta 2}\})^{-1}, \\
(1.21) \\
G_3(\Delta y_{t-1}; \gamma_3, c_3) = (1 + \exp\{-500(\Delta y_{t-1} + 2.53) / \sigma_{\Delta 4 y_{t-1}}\})^{-1}, \\
(0.02)
\]

\[
\hat{\sigma}_t = 1.43, \quad \hat{\sigma}_{\text{TV-STAR/AR}} = 0.74, \quad SK = -0.10 (0.31), \quad EK = 1.00 (2.9E - 3), \\
JB = 7.88 (0.019), \quad \text{LM}_{\text{SC}}(1) = 0.72 (0.40), \quad \text{LM}_{\text{SC}}(4) = 0.40 (0.81), \\
\text{LM}_{\text{SC}}(12) = 1.37 (0.19), \quad \text{ARCH}(1) = 0.25 (0.61), \quad \text{ARCH}(4) = 3.21 (0.52), \\
\text{AIC}_{\text{TV-STAR/AR}} = -0.46, \quad \text{BIC}_{\text{TV-STAR/AR}} = -0.24.
\]

The model is made more parsimonious by imposing ± restrictions on the coefficients of $\Delta y_{t-1}$, $\Delta y_{t-4}$, $\Delta y_{t-5}$, and $\Delta y_{t-8}$, which are supported by the data.
3.5. Modelling changing seasonal patterns by TV-STAR models

Because of these restrictions, the model does not contain any autoregressive dynamics when \( G_3 = 1 \), except for \( \Delta y_{t-6} \) which enters with a small coefficient. This means that the industrial production growth rate almost behaves like a white noise series during most of the time - note from panel (d) in Figure 3.12 that \( G_3 \) only becomes zero during the most severe recessions that hit France during the sample period. As shown by the estimated intercepts and the coefficients on the seasonal dummies in (3.19), the first structural change involves a decline of the mean growth rate by 1.0% and affects the seasonal pattern only in the first and second quarter. By contrast, the second time-varying component captures changes in the first and third quarters which imply a large reduction of the amplitude of the seasonal cycle. See also Figure 3.12, that shows a slowly changing seasonal pattern with decreasing amplitude.

**Germany**

For Germany, the results from the LM-type misspecification tests in the linear model indicate that the seasonal dummy coefficients may be varying for unspecified reasons and the autoregressive dynamics may be varying with the business cycle. The evidence for the latter disappears, however, once we allow the seasonal dummies to vary over time. To capture the variation in the seasonal pattern, we find that three TV components with standard logistic functions are required. The final specification is:

\[
\Delta y_t = 1.67 - 8.95 D^*_1,t + 7.02 D^*_2,t - 8.25 D^*_3,t \\
+ 0.16 \Delta y_{t-1} + 0.12 \Delta y_{t-2} - 0.13 \Delta y_{t-7} - 0.18 \Delta y_{t-8} \\
+ (-1.32 + 1.97 D^*_1,t - 2.56 D^*_2,t - 6.26 D^*_3,t) \times G_1(t^*; \gamma_1, c_1) \\
+ (-6.24 D^*_1,t + 12.0 D^*_3,t) \times G_2(t^*; \gamma_2, c_2) \\
+ (4.52 D^*_2,t + 1.98 D^*_3,t) \times G_3(t^*; \gamma_3, c_3) + \tilde{\epsilon}_t, \\
(3.23)
\]

\[
G_1(t^*; \gamma_1, c_1) = (1 + \exp\{-5.89 (t^* - 0.20)/\sigma_{1*}\})^{-1}, \\
G_2(t^*; \gamma_2, c_2) = (1 + \exp\{-500 (t^* - 0.40)/\sigma_{2*}\})^{-1}, \\
G_3(t^*; \gamma_3, c_3) = (1 + \exp\{-5.58 (t^* - 0.83)/\sigma_{3*}\})^{-1}, \\
(3.24) \\
(3.25) \\
(3.26)
\]
Changes in seasonality

\[ \hat{\sigma}_\varepsilon = 1.60, \hat{\sigma}_{TV-STAR/AR} = 0.70, SK = -0.43 (0.015), EK = 0.77 (0.026), JB = 8.55 (0.014), LM_{SC}(1) = 0.38 (0.54), LM_{SC}(4) = 0.37 (0.83), LM_{SC}(12) = 1.17 (0.31), ARCH(1) = 8.04 (4.5E - 3), ARCH(4) = 8.66 (0.070), AIC_{TV-STAR/AR} = -0.51, BIC_{TV-STAR/AR} = -0.22. \]

The amplitude of the seasonal pattern is also ultimately decreasing for Germany. The first change, which is centered around 1970:1, implies a substantial decline of the overall mean growth rate by 1.3%. In addition, the seasonal effects during the third and fourth quarters are amplified, as can be seen from panel (b) in Figure 3.13. This is partially reversed during the instantaneous change in the seasonal pattern that occurred in 1978, which also captures a change in the seasonal effect in the second quarter from positive \((7.02 - 2.56 = 4.46)\) to negative \((4.46 - 6.24 = -1.78)\). The latter is reversed again during the last change, the German unification in 1989, which appears to have initiated a more gradual decline in the amplitude of the seasonal pattern.

**Italy**

For Italy, the strongest rejection occurs when constancy of the seasonal dummy parameters is tested against unspecified change, see Table 3.1. Misspecification tests of a TV-AR model in which only the seasonal dummy coefficients are time-varying still reject constancy of the dummy coefficients. Adding a second TV-AR component, we find strong evidence against parameter constancy of the coefficients of the lagged first differences. Allowing these to be time-varying as well, we cannot reject the hypothesis that the change in the autoregressive dynamics and the first change in the seasonal pattern occur simultaneously. Furthermore, we find that exclusion restrictions on the autoregressive parameters implying that the combined parameter equals zero when \(G_1 = 1\) cannot be rejected. This means that the industrial production growth rate is a white noise series with seasonal means after the first smooth transition has been completed. The estimated model has the form

\[ \Delta y_t = 2.01 + 1.99 D_{1t}^* + 4.75 D_{2t}^* - 13.1 D_{3t}^* + 0.18 \Delta y_{t-2} - 0.48 \Delta y_{t-5} (0.36) (0.46) (1.05) (0.59) (0.083) (0.093) - 0.35 \Delta y_{t-6} + (-1.59 - 3.49 D_{2t}^* - 8.77 D_{3t}^* - 0.18 \Delta y_{t-2} + 0.48 \Delta y_{t-5} (0.094) (0.43) (1.08) (0.81) (0.083) (0.093) + 0.35 \Delta y_{t-6} \times G_1(t^*; \gamma_1, c_1) + (-2.04 D_{1t}^* + 3.90 D_{3t}^*) \times G_2(t^*; \gamma_2, c_2) + \hat{\varepsilon}_t, (0.094) (0.77) (0.80) \]

\[ (3.27) \]
3.5. Modelling changing seasonal patterns by TV-STAR models

\[ G_1(t^*; \gamma_1, c_1) = (1 + \exp\{-9.43 (t^* - 0.39)/\sigma_{t^*}\})^{-1}, \quad (3.28) \]
\[ G_2(t^*; \gamma_2, c_2) = (1 + \exp\{-101 (t^* - 0.81)/\sigma_{t^*}\})^{-1}, \quad (3.29) \]

\( \hat{\sigma}_\epsilon = 2.05, \hat{\sigma}_{TV\text{-}STAR/AR} = 0.80, \ SK = 0.14 (0.24), EK = 0.38 (0.17), \)
\( JB = 1.41 (0.49), \ LM_{SC}(1) = 0.30 (0.58), LM_{SC}(4) = 0.55 (0.70), \)
\( LM_{SC}(12) = 0.83 (0.61), ARCH(1) = 13.9 (1.9E-4), ARCH(4) = 16.9 (2.0E-3), \)
\( AIC_{TV\text{-}STAR/AR} = -0.35, BIC_{TV\text{-}STAR/AR} = -0.21. \)

The estimates of the seasonal dummy coefficients in (3.27) and the graphs in Figure 3.14 clearly show that the deviations from the mean growth rate in the first and second quarters have declined during the second half of the 1970s, whereas the seasonal pattern in the third and fourth quarters has been amplified. Overall, the seasonality became more pronounced during the second half of the 1970s, followed by a swift (but relatively small) decline in the amplitude in 1994.

Japan

In the linear model for the Japanese industrial production series, both parameter constancy and linearity are forcefully rejected for both the lagged autoregressive parameters and the seasonal dummy coefficients. As parameter constancy of the seasonal dummy coefficients is rejected most convincingly, we start with a TV-AR model with time-varying seasonal dummy coefficients only. In the resulting TV-AR model, linearity and parameter constancy of the seasonal dummies still are strongly rejected by the diagnostic tests. Accounting for this by sequentially including a nonlinear component with \( \Delta_4 y_{t-1} \) as transition variable and a second TV-AR component and recursively deleting insignificant coefficients, we finally obtain the specification:

\[
\Delta y_t = 2.29 - 1.77 D_{1,t}^* + 5.72 D_{2,t}^* + 0.51 \Delta y_{t-1} + 0.17 \Delta y_{t-2} - 0.36 \Delta y_{t-5} - 0.17 \Delta y_{t-10} + 0.29 \Delta y_{t-11} - 0.18 \Delta y_{t-12} + (-1.61 - 4.41 D_{1,t}^* + 3.34 D_{2,t}^*) \times G_1(t^*; \gamma_1, c_1) + (-1.04 + 7.91 D_{1,t}^* - 11.3 D_{2,t}^* + 5.85 D_{3,t}^*) \times G_2(t^*; \gamma_2, c_2) + (-5.66 D_{4,t}^* - 1.27 D_{5,t}^*) \times G_3(\Delta_4 y_{t-1}; \gamma_3, c_3) + \hat{\epsilon}_t, \quad (3.30)
\]
Changes in seasonality

\[ G_1(t^*; \gamma_1, c_1) = (1 + \exp\{-8.92 (t^* - 0.23)/\sigma_{t^*}\})^{-1}, \quad (3.31) \]

\[ G_2(t^*; \gamma_2, c_2) = (1 + \exp\{-3.47 (t^* - 0.79)/\sigma_{t^*}\})^{-1}, \quad (3.32) \]

\[ G_3(\Delta_4 y_{t-1}; \gamma_3, c_3) = (1 + \exp\{-1.60 (\Delta_4 y_{t-1} + 8.56)/\sigma_{\Delta_4 y_{t-1}}\})^{-1}, \quad (3.33) \]

\[ \hat{\sigma}_e = 1.30, \quad \hat{\sigma}_{\text{TV-STAR/AR}} = 0.72, \quad \text{SK} = -0.41(0.020), \quad \text{EK} = 0.48(0.11), \]

\[ \text{JB} = 5.70(0.058), \quad \text{LM}_{\text{SC}}(1) = 0.32(0.57), \quad \text{LM}_{\text{SC}}(4) = 0.46(0.76), \]

\[ \text{LM}_{\text{SC}}(12) = 1.13(0.34), \quad \text{ARCH}(1) = 1.48(0.22), \quad \text{ARCH}(4) = 5.30(0.26), \]

\[ \text{AIC}_{\text{TV-STAR/AR}} = -0.46, \quad \text{BIC}_{\text{TV-STAR/AR}} = -0.17. \]

The model contains two relatively smooth changes in the seasonal pattern, see also Figure 3.15. The first, which occurred during the first half of the 1970s, considerably amplified the seasonal pattern, especially during the first and second quarter. The seasonal pattern is changed completely during the second transition, which started around 1985 and was almost completed at the end of the sample period: the deviations of the mean during the first, second and third quarters change from -6.18, 3.40 and -1.27, respectively, when \( G_1 = G_3 = 1 \) and \( G_2 = 0 \) to 1.73, -7.94, and 4.58 when \( G_1 = G_2 = G_3 = 1 \). Note that the structural changes also involve a substantial reduction of the average growth rate, from 2.29% via 0.68% to -0.37%.

Sweden

Parameter constancy of the seasonality-augmented linear AR model is strongly rejected. Estimating a TV-AR model for \( \Delta y_t \) takes care of most of the non-constancy, but LM1 statistic applied to the coefficients of the seasonal dummy variables still indicates model inadequacy. A TV-AR model with two nonlinear components is almost adequate. The only \( p \)-values less than 0.05 have to do with testing the hypothesis of no additive nonlinearity. Including an additive STAR component and reducing the size of the model, results in the following final specification:

\[ \Delta y_t = 0.68 - 20.58 D_{5,t}^* + 0.22 \Delta y_{t-2} + 0.22 \Delta y_{t-4} + 0.11 \Delta y_{t-5} \]

\[ + (-1.42 + 0.095 \Delta y_{t-7}) \times G_1(\Delta_4 y_{t-4}; \gamma_1, c_1) \]

\[ + (2.38 D_{5,t}^* + 5.18 D_{6,t}^*) \times G_2(t^*; \gamma_2, c_2) - 3.25 D_{1,t}^* \times G_3(t^*; \gamma_3, c_3) + \hat{\varepsilon}_t, \quad (3.34) \]
3.5. Modelling changing seasonal patterns by TV-STAR models

\[ G_1(\Delta_4 y_{t-4}; \gamma_1, c_1) = (1 + \exp\{-16.0 (\Delta_4 y_{t-4} - 6.28)/\sigma_{\Delta_4 y_{t-4}}\})^{-1}, \quad (3.35) \]
\[ G_2(t*; \gamma_2, c_2) = (1 + \exp\{-23.4 (t* - 0.73)/\sigma_{t*}\})^{-1}, \quad (3.36) \]
\[ G_3(t*; \gamma_3, c_3) = (1 + \exp\{-7.51 (t* + 0.38)/\sigma_{t*}\})^{-1}, \quad (3.37) \]

\[ \hat{\sigma}_t = 2.17, \hat{\sigma}_{TV-STAR/AR} = 0.84, \ SK = -0.056(0.39), \ EK = -0.07(0.43), \]
\[ JB = 0.11(0.95), \ LM_{SC}(1) = 2.01(0.16), \ LM_{SC}(4) = 0.74(0.57), \]
\[ LM_{SC}(12) = 0.50(0.91), \ ARCH(1) = 1.58(0.21), \ ARCH(4) = 16.5(2.4E-3), \]
\[ AIC_{TV-STAR/AR} = -0.25, \ BIC_{TV-STAR/AR} = -0.11. \]

Model (3.34) passes most of the misspecification tests; see Table 3.6, but the hypothesis of no autoregressive conditional heteroskedasticity is rejected. This is most likely due to a couple of outliers in the residuals. The ones in 1989:4 and 1990:4; see Figure 3.16(c), are the likely culprits (note that the null hypothesis is only rejected against ARCH(4), not against ARCH(1)).

The interpretation of the estimated model is rather straightforward. The most dominant feature in seasonality is the decrease in the amplitude in the first half of the 1990s. The decrease is obviously forced by the recession visible in Figure 3.7(a). It is parameterized through transition function \( G_2 \) and the related seasonal dummy variables. Transition function \( G_1 \) describes a diminishing relative contribution from the first quarter to the industrial production in the 1970s (\( \hat{\gamma}_3 = 0.38; \) see also Figure 3.16(d)). It represents an institutional change as it can be attributed to an increase in paid leisure allowing a winter vacation. As to business-cycle induced change, there is no evidence of it in the quarterly Swedish series.

**United States**

For the US, constancy of the seasonal dummy coefficients and linearity of the autoregressive parameters are rejected. Starting with a STAR model allowing for nonlinearity in the coefficients of the lagged first differences only and using \( \Delta_4 y_{t-2} \) as transition variable (rejection of linearity is strongest in this case), the hypothesis of constancy of the dummy coefficients is no longer rejected at conventional significance levels. Hence, we arrive at the STAR model.
\[
\Delta y_t = 0.44 \Delta y_{t-1} - 0.41 \Delta y_{t-2} - 0.68 \Delta y_{t-5} - 0.46 \Delta y_{t-7},
\]
\[
+ (0.49 + 0.41 \Delta y_{t-2} + 0.25 \Delta y_{t-6}) (0.26) (0.095) (0.071)
\]
\[
+ 0.54 \Delta y_{t-7}) \times G_1(\Delta y_{t-2}; \gamma_1, c_1) + \tilde{\epsilon}_t,
\]
\[
G_1(\Delta y_{t-2}; \gamma_1, c_1) = (1 + \exp\{ -59.0 (\Delta y_{t-2} - 0.95)/\sigma_{\Delta y_{t-2}} \})^{-1},
\]
Note that a \pm restriction on the coefficients of \( \Delta y_{t-2} \) is imposed. A similar restriction on the coefficients of \( \Delta y_{t-7} \) is rejected. The skewness and excess kurtosis of the residuals are caused entirely by the observations for 1975:1 and 1980:2, where large negative residuals occur.

**3.6 Final remarks**

The results of this chapter suggest that seasonal patterns in quarterly industrial production series for the G7 and two Scandinavian countries have been changing over time. On the other hand, business cycle fluctuations do not seem to be the main cause for this change. Our findings are in contrast with Canova and Ghysels (1994) and Franses (1996), who considered US output and concluded that the business cycle influences the seasonal cycle. Similarly, Cecchetti et al. (1997) found that in the US seasonal fluctuations in production and inventories vary with the state of the business cycle. There are at least two reasons for differences between our results and the ones of the above authors. First, they only considered US series and include the GNP and inventories. The second, and perhaps the more important, reason is that those authors did not consider other causes than business cycle fluctuations. Less restrictive considerations appear to lead to rather different conclusions.

It seems possible to reconcile our results with the findings of Matas-Mir and Osborn (2004). An important detail is that they used monthly series, whereas our data are quarterly. As the authors explain, a business-cycle induced change in summer months, which is visible in monthly series, can
3.6. Final remarks

be substantially masked at a quarterly frequency. Another reason for the differences in results is that Matas-Mir and Osborn (2004) implicitly give a preference to business-cycle induced pattern shifts, because other types of change are only described by linear trends in seasonal dummy coefficients. This may be too rigid a solution, see, for example, Figures 3.10 – 3.18, and a more flexible parameterization, offered by the TV-STAR model, is needed to fully assess the role and significance of institutional and technological change in seasonal patterns of the series considered here. Thus the differences in results between Matas-Mir and Osborn (2004) and our work may to a large extent be ascribed to differences in the emphasis, reflected both in the frequency of the series and the choice of model.

As the "Kuznets-type" unspecified change in seasonal patterns is in our work proxied by time, we cannot give a definite answer to the question of what kind of change, technological, institutional, or "other", has been important in the industrial output series we have investigated. The importance of our results lies in the fact that they make us aware of changes such as the gradual decrease in amplitude many series are showing. Nevertheless, some speculation about the reasons for this may be allowed. There is evidence for changes in inventory management affecting the seasonal pattern of industrial output. Carpenter and Levy (1998) showed that inventory investment and output are highly correlated not only at business cycle frequencies but also at seasonal frequencies. Given the importance of inventories for (changes in) fluctuations in output (see Sichel (1994), and McConnell and Perez-Quiros (2000), among others), it may well be that changes in inventory management such as the use of "just-in-time" techniques have dampened the seasonal cycle in inventory investment and thereby changed the seasonal cycle in production. This would seem a plausible explanation or a part of it in cases where the amplitude of the seasonal pattern has decreased over time. On the other hand, very abrupt changes, such as the one in the German industrial output series in 1978, may perhaps best be ascribed to the agency producing the data, unless other information about the nature of the change is available. In general, it may sometimes be relatively easy to suggest individual causes for shifts in the seasonal pattern at the industry level. Because of aggregation this becomes more difficult when the volume of the total industrial output is concerned.

The models estimated for seasonally unadjusted first differences in this work cannot be expected to be useful in forecasting the volume of industrial production. Models that enhance and explain the low-frequency fluctuations in the series are better suited for that purpose. Our results give rise to the question of how the current seasonal adjustment methods cope with series with
Changes in seasonality, and what the consequences of such variation are on using seasonally adjusted series in macroeconomic modelling. Investigating this question in the present context, however, must be left for future work.
References


Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 413–430.


Figures

Figure 3.1: Industrial production Canada

(a) Level

(b) First difference per quarter

(c) Seasonal difference
Figure 3.2: Industrial production Finland

(a) Level

(b) First difference per quarter

(c) Seasonal difference
Figure 3.3: Industrial production France

(a) Level

(b) First difference per quarter

(c) Seasonal difference
Changes in seasonality

Figure 3.4: Industrial production Germany

(a) Level

(b) First difference per quarter

(c) Seasonal difference
Figure 3.5: Industrial production Italy

(a) Level

(b) First difference per quarter

(c) Seasonal difference
Figure 3.6: Industrial production Japan

(a) Level

(b) First difference per quarter

(c) Seasonal difference
Figure 3.7: Industrial production Sweden

(a) Level

(b) First difference per quarter

(c) Seasonal difference
Figure 3.8: Industrial production United Kingdom
Figure 3.9: Industrial production United States
Figure 3.10: Characteristics of TV-STAR model for quarterly industrial production growth rates in Canada in (3.12)
Figure 3.11: Characteristics of TV-STAR model for quarterly industrial production growth rates in Finland in (3.15)

(a) First difference

(b) Seasonal intercepts AR (dashed line) and TV-STAR model (solid line)

(c) Residuals from AR (dashed line) and TV-STAR model (solid line)

(d) Transition functions in TV-STAR model
Figure 3.12: Characteristics of TV-STAR model for quarterly industrial production growth rates in France in (3.19)
Figure 3.13: Characteristics of TV-STAR model for quarterly industrial production growth rates in Germany in (3.23)

(a) First difference

(b) Seasonal intercepts AR (dashed line) and TV-STAR model (solid line)

(c) Residuals from AR (dashed line) and TV-STAR model (solid line)

(d) Transition functions in TV-STAR model
Figure 3.14: Characteristics of TV-STAR model for quarterly industrial production growth rates in Italy in (3.27)
Figure 3.15: Characteristics of TV-STAR model for quarterly industrial production growth rates in Japan in (3.30)
Figure 3.16: Characteristics of TV-STAR model for quarterly industrial production growth rates in Sweden in (3.34)
Figure 3.17: Characteristics of TV-STAR model for quarterly industrial production growth rates in United Kingdom in (3.8)

(a) First difference

(b) Seasonal intercepts AR (dashed line) and TV-STAR model (solid line)

(c) Residuals from AR (dashed line) and TV-STAR model (solid line)

(d) Transition functions in TV-STAR model
Figure 3.18: Characteristics of TV-STAR model for quarterly industrial production growth rates in United States in (3.38)
### Table 3.1: Testing linearity and parameter constancy

<table>
<thead>
<tr>
<th>Parameters tested</th>
<th>STAR</th>
<th>TV-AR</th>
<th>TV-STAR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LM$_1$</td>
<td>LM$_3$</td>
<td>LM$_1$</td>
</tr>
<tr>
<td><strong>Canada ($\hat{p} = 8$)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D_{s,t}$</td>
<td>0.21</td>
<td>0.72</td>
<td>0.28</td>
</tr>
<tr>
<td>$\Delta y_{t-j}$</td>
<td>0.068</td>
<td>0.35</td>
<td>0.37</td>
</tr>
<tr>
<td><strong>Finland ($\hat{p} = 7$)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D_{s,t}$</td>
<td>0.15</td>
<td>0.11</td>
<td>8.8E-3</td>
</tr>
<tr>
<td>$\Delta y_{t-j}$</td>
<td>0.097</td>
<td>0.17</td>
<td>0.17</td>
</tr>
<tr>
<td><strong>France ($\hat{p} = 8$)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D_{s,t}$</td>
<td>0.34</td>
<td>0.035</td>
<td>1.0E-5</td>
</tr>
<tr>
<td>$\Delta y_{t-j}$</td>
<td>0.062</td>
<td>0.12</td>
<td>4.9E-5</td>
</tr>
<tr>
<td><strong>Germany ($\hat{p} = 5$)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D_{s,t}$</td>
<td>0.11</td>
<td>0.51</td>
<td>0.012</td>
</tr>
<tr>
<td>$\Delta y_{t-j}$</td>
<td>3.8E-3</td>
<td>8.3E-3</td>
<td>0.015</td>
</tr>
<tr>
<td><strong>Italy ($\hat{p} = 6$)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D_{s,t}$</td>
<td>0.17</td>
<td>0.21</td>
<td>0.061</td>
</tr>
<tr>
<td>$\Delta y_{t-j}$</td>
<td>0.094</td>
<td>0.12</td>
<td>0.11</td>
</tr>
<tr>
<td><strong>Japan ($\hat{p} = 5$)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D_{s,t}$</td>
<td>0.066</td>
<td>8.2E-4</td>
<td>0.019</td>
</tr>
<tr>
<td>$\Delta y_{t-j}$</td>
<td>0.024</td>
<td>8.1E-3</td>
<td>3.7E-3</td>
</tr>
<tr>
<td><strong>Sweden ($\hat{p} = 7$)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D_{s,t}$</td>
<td>0.10</td>
<td>0.21</td>
<td>9.1E-4</td>
</tr>
<tr>
<td>$\Delta y_{t-j}$</td>
<td>0.16</td>
<td>0.086</td>
<td>0.014</td>
</tr>
<tr>
<td><strong>United Kingdom ($\hat{p} = 9$)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D_{s,t}$</td>
<td>0.030</td>
<td>0.071</td>
<td>2.1E-3</td>
</tr>
<tr>
<td>$\Delta y_{t-j}$</td>
<td>0.039</td>
<td>0.075</td>
<td>5.1E-3</td>
</tr>
<tr>
<td><strong>United States ($\hat{p} = 7$)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D_{s,t}$</td>
<td>0.034</td>
<td>0.11</td>
<td>0.020</td>
</tr>
<tr>
<td>$\Delta y_{t-j}$</td>
<td>6.4E-3</td>
<td>0.014</td>
<td>0.49</td>
</tr>
</tbody>
</table>

**Notes:** The table contains $p$-values of $F$-variants of the $	ext{LM}_k$, $k = 1, 3$, tests of linearity and parameter constancy within the TV-STAR model (3.1). The null hypotheses of the different tests are linearity conditional on parameter constancy [STAR], constancy conditional on linearity [TV-AR], and linearity and constancy [TV-STAR]. Rows labelled $D_{s,t}$ and $\Delta y_{t-j}$ contain results for testing the seasonal dummies and the lagged growth rates, respectively. All tests are performed conditional on assuming that the remaining parameters enter linearly and with constant parameters. A dash indicates that the test could not be computed due to a shortage in degrees of freedom.
### Table 3.2: Summary of test results

<table>
<thead>
<tr>
<th>Parameters tested</th>
<th>p-value</th>
<th>Linear null model</th>
<th>Nonlinear null model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>STAR LM</td>
<td>LM₃</td>
</tr>
<tr>
<td>Dₛₜ</td>
<td>&gt; 0.05</td>
<td>7 7</td>
<td>2 0</td>
</tr>
<tr>
<td></td>
<td>0.01–0.05</td>
<td>2 1</td>
<td>3 1</td>
</tr>
<tr>
<td></td>
<td>&lt; 0.01</td>
<td>0 1</td>
<td>4 8</td>
</tr>
<tr>
<td>Δᵧₜ₋ₗ</td>
<td>&gt; 0.05</td>
<td>5 6</td>
<td>4 3</td>
</tr>
<tr>
<td></td>
<td>0.01–0.05</td>
<td>2 1</td>
<td>2 1</td>
</tr>
<tr>
<td></td>
<td>&lt; 0.01</td>
<td>2 2</td>
<td>3 2</td>
</tr>
<tr>
<td>Dₛₜ</td>
<td>&gt; 0.05</td>
<td>8 8</td>
<td>5 3</td>
</tr>
<tr>
<td></td>
<td>0.01–0.05</td>
<td>1 1</td>
<td>2 1</td>
</tr>
<tr>
<td></td>
<td>&lt; 0.01</td>
<td>0 0</td>
<td>2 5</td>
</tr>
</tbody>
</table>

**Notes:** The table reports the number of series for which the p-values of the tests of linearity and parameter constancy, shown in Tables 3.1 and 3.3, fall within the indicated category. The null hypotheses of the different tests are linearity conditional on parameter constancy [STAR], constancy conditional on linearity [TV-AR], and linearity and constancy [TV-STAR].
Table 3.3: Testing linearity and parameter constancy in STAR/TV-AR models

<table>
<thead>
<tr>
<th>Transition variable</th>
<th>STAR</th>
<th></th>
<th>TV-AR</th>
<th></th>
<th>TV-STAR</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LM₃</td>
<td>LM₃</td>
<td>LM₁</td>
<td>LM₁</td>
<td>LM₁</td>
<td>LM₃</td>
</tr>
<tr>
<td>Canada</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s_t = \Delta_4 y_{t-3}$</td>
<td>0.13</td>
<td>0.62</td>
<td>0.23</td>
<td>1.4E-4</td>
<td>0.34</td>
<td>0.40</td>
</tr>
<tr>
<td>Finland</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s_t = \Delta_4 y_{t-2}$</td>
<td>0.20</td>
<td>0.19</td>
<td>2.0E-3</td>
<td>1.0E-5</td>
<td>0.022</td>
<td>0.085</td>
</tr>
<tr>
<td>France</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s_t = t^*$</td>
<td>0.46</td>
<td>0.20</td>
<td>0.38</td>
<td>0.015</td>
<td>0.48</td>
<td>0.073</td>
</tr>
<tr>
<td>Germany</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s_t = \Delta_4 y_{t-1}$</td>
<td>0.51</td>
<td>0.61</td>
<td>7.6E-3</td>
<td>9.4E-4</td>
<td>0.42</td>
<td>0.27</td>
</tr>
<tr>
<td>Italy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s_t = t^*$</td>
<td>0.47</td>
<td>0.58</td>
<td>0.26</td>
<td>8.8E-3</td>
<td>0.64</td>
<td>0.26</td>
</tr>
<tr>
<td>Japan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s_t = \Delta_4 y_{t-1}$</td>
<td>0.17</td>
<td>0.046</td>
<td>0.024</td>
<td>2.4E-6</td>
<td>0.033</td>
<td>0.11</td>
</tr>
<tr>
<td>Sweden</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s_t = t^*$</td>
<td>0.36</td>
<td>0.44</td>
<td>0.019</td>
<td>0.095</td>
<td>0.35</td>
<td>0.72</td>
</tr>
<tr>
<td>United Kingdom</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s_t = t^*$</td>
<td>0.041</td>
<td>0.063</td>
<td>0.21</td>
<td>0.32</td>
<td>0.081</td>
<td>0.056</td>
</tr>
<tr>
<td>United States</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s_t = \Delta_4 y_{t-2}$</td>
<td>0.064</td>
<td>0.10</td>
<td>0.11</td>
<td>0.32</td>
<td>0.057</td>
<td>0.51</td>
</tr>
</tbody>
</table>

Notes: The table contains p-values of F-variants of the LMₖ, $k = 1, 3$, tests of linearity of the seasonal pattern within the STAR model (3.7) for quarterly industrial production growth rates, where the delay parameter $e$ is assumed unknown but restricted to be less than or equal to 4. The null hypotheses of the different tests are linearity conditional on parameter constancy [STAR], constancy conditional on linearity [TV-AR], and linearity and constancy [TV-STAR].
Table 3.4: Diagnostic tests of parameter constancy and no remaining nonlinearity in TV-STAR models for quarterly growth rates in UK industrial production

<table>
<thead>
<tr>
<th>Transition variable</th>
<th>$D_{s,t}$</th>
<th>$\Delta y_{t-j}$</th>
<th>$\sigma^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LM$_1$</td>
<td>LM$_3$</td>
<td>LM$_1$</td>
</tr>
<tr>
<td>AR model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t$</td>
<td>2.1E-3</td>
<td>2.1E-3</td>
<td>5.1E-3</td>
</tr>
<tr>
<td>$\Delta_4 y_{t-1}$</td>
<td>0.20</td>
<td>0.092</td>
<td>0.22</td>
</tr>
<tr>
<td>$\Delta_4 y_{t-2}$</td>
<td>0.55</td>
<td>0.28</td>
<td>0.65</td>
</tr>
<tr>
<td>$\Delta_4 y_{t-3}$</td>
<td>0.12</td>
<td>0.060</td>
<td>0.064</td>
</tr>
<tr>
<td>$\Delta_4 y_{t-4}$</td>
<td>0.27</td>
<td>0.30</td>
<td>0.34</td>
</tr>
<tr>
<td>TV-AR model with $D_{s,t}^*$, time-varying</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t$</td>
<td>0.034</td>
<td>0.12</td>
<td>0.43</td>
</tr>
<tr>
<td>$\Delta_4 y_{t-1}$</td>
<td>0.49</td>
<td>0.62</td>
<td>0.013</td>
</tr>
<tr>
<td>$\Delta_4 y_{t-2}$</td>
<td>0.95</td>
<td>0.88</td>
<td>0.81</td>
</tr>
<tr>
<td>$\Delta_4 y_{t-3}$</td>
<td>0.82</td>
<td>0.99</td>
<td>0.20</td>
</tr>
<tr>
<td>$\Delta_4 y_{t-4}$</td>
<td>0.63</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td>TV-STAR model with $D_{s,t}^*$, time-varying and $\Delta y_{t-j}$ nonlinear ($s_t = \Delta_4 y_{t-1}$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t$</td>
<td>0.041</td>
<td>0.030</td>
<td>0.51</td>
</tr>
<tr>
<td>$\Delta_4 y_{t-1}$</td>
<td>0.40</td>
<td>0.54</td>
<td>0.096</td>
</tr>
<tr>
<td>$\Delta_4 y_{t-2}$</td>
<td>0.56</td>
<td>0.34</td>
<td>0.61</td>
</tr>
<tr>
<td>$\Delta_4 y_{t-3}$</td>
<td>0.62</td>
<td>0.71</td>
<td>0.78</td>
</tr>
<tr>
<td>$\Delta_4 y_{t-4}$</td>
<td>0.15</td>
<td>0.12</td>
<td>0.30</td>
</tr>
<tr>
<td>TV-STAR model with 2 time-varying components for $D_{s,t}^*$ and with $\Delta y_{t-j}$ nonlinear ($s_t = \Delta_4 y_{t-1}$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t$</td>
<td>0.21</td>
<td>0.087</td>
<td>0.65</td>
</tr>
<tr>
<td>$\Delta_4 y_{t-1}$</td>
<td>0.46</td>
<td>0.54</td>
<td>0.89</td>
</tr>
<tr>
<td>$\Delta_4 y_{t-2}$</td>
<td>0.92</td>
<td>0.35</td>
<td>0.50</td>
</tr>
<tr>
<td>$\Delta_4 y_{t-3}$</td>
<td>0.69</td>
<td>0.44</td>
<td>0.21</td>
</tr>
<tr>
<td>$\Delta_4 y_{t-4}$</td>
<td>0.62</td>
<td>0.29</td>
<td>0.45</td>
</tr>
</tbody>
</table>

Notes: The table contains p-values of F-variants of LM diagnostic tests of parameter constancy (rows labelled $t$) and no remaining nonlinearity (rows labelled $\Delta y_{t-l}$, with $l = 1, \ldots, 4$) of seasonal dummy coefficients (columns headed $D_{s,t}$), autoregressive parameters (columns headed $\Delta y_{t-j}$), and residual variance (columns headed $\sigma^2$) in estimated AR, TV-AR and TV-STAR models for UK quarterly industrial production.
Table 3.5: Summary statistics of estimated models for quarterly growth rates in UK industrial production

<table>
<thead>
<tr>
<th></th>
<th>AR</th>
<th>TV-AR</th>
<th>TV-STAR</th>
<th>TV-STAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\gamma}_1$</td>
<td>-</td>
<td>9.99</td>
<td>214</td>
<td>2.03</td>
</tr>
<tr>
<td>$\hat{\gamma}_2$</td>
<td>-</td>
<td>0.43</td>
<td>0.43</td>
<td>0.44</td>
</tr>
<tr>
<td>$\hat{\gamma}_3$</td>
<td>-</td>
<td>-</td>
<td>30.5</td>
<td>19.5</td>
</tr>
<tr>
<td>$\hat{\gamma}_4$</td>
<td>-</td>
<td>-</td>
<td>-1.91</td>
<td>-0.37</td>
</tr>
<tr>
<td>$\hat{\gamma}_5$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>500</td>
</tr>
<tr>
<td>$\hat{\gamma}_6$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.31</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>2.15</td>
<td>2.04</td>
<td>1.82</td>
<td>1.65</td>
</tr>
<tr>
<td>AIC</td>
<td>1.66</td>
<td>1.55</td>
<td>1.40</td>
<td>1.28</td>
</tr>
<tr>
<td>BIC</td>
<td>1.84</td>
<td>1.72</td>
<td>1.72</td>
<td>1.69</td>
</tr>
<tr>
<td>SK</td>
<td>-0.12</td>
<td>-0.32</td>
<td>-0.48</td>
<td>0.34</td>
</tr>
<tr>
<td>(0.27)</td>
<td>(0.050)</td>
<td>(7.8E-3)</td>
<td>(0.096)</td>
<td></td>
</tr>
<tr>
<td>EK</td>
<td>0.63</td>
<td>1.89</td>
<td>1.15</td>
<td>1.29</td>
</tr>
<tr>
<td>(0.056)</td>
<td>(8.9E-7)</td>
<td>(1.8E-3)</td>
<td>(5.5E-4)</td>
<td></td>
</tr>
<tr>
<td>JB</td>
<td>2.89</td>
<td>25.5</td>
<td>14.4</td>
<td>13.7</td>
</tr>
<tr>
<td>(0.24)</td>
<td>(2.9E-6)</td>
<td>(7.6E-4)</td>
<td>(1.1E-3)</td>
<td></td>
</tr>
<tr>
<td>ARCH(1)</td>
<td>1.13</td>
<td>6.05</td>
<td>0.44</td>
<td>0.19</td>
</tr>
<tr>
<td>(0.29)</td>
<td>(0.014)</td>
<td>(0.51)</td>
<td>(0.66)</td>
<td></td>
</tr>
<tr>
<td>ARCH(4)</td>
<td>10.2</td>
<td>10.6</td>
<td>7.85</td>
<td>16.4</td>
</tr>
<tr>
<td>(0.037)</td>
<td>(0.032)</td>
<td>(0.097)</td>
<td>(2.5E-3)</td>
<td></td>
</tr>
<tr>
<td>LMSC(1)</td>
<td>0.88</td>
<td>0.094</td>
<td>0.63</td>
<td>0.24</td>
</tr>
<tr>
<td>(0.35)</td>
<td>(0.76)</td>
<td>(0.43)</td>
<td>(0.62)</td>
<td></td>
</tr>
<tr>
<td>LMSC(4)</td>
<td>1.34</td>
<td>0.62</td>
<td>0.78</td>
<td>0.59</td>
</tr>
<tr>
<td>(0.26)</td>
<td>(0.65)</td>
<td>(0.54)</td>
<td>(0.67)</td>
<td></td>
</tr>
<tr>
<td>LMSC(12)</td>
<td>1.44</td>
<td>1.04</td>
<td>1.06</td>
<td>0.69</td>
</tr>
<tr>
<td>(0.16)</td>
<td>(0.42)</td>
<td>(0.40)</td>
<td>(0.76)</td>
<td></td>
</tr>
</tbody>
</table>

Notes: The table presents selected parameter estimates and misspecification tests for the estimated AR, TV-AR and TV-STAR models for quarterly growth rates in UK industrial production. $\sigma^2$ denotes the residual variance, SK is skewness, EK excess kurtosis, JB the Jarque-Bera test of normality of the residuals, ARCH(r) is the LM test of no Autoregressive Conditional Heteroscedasticity (ARCH) up to order r, and LMsc(q) denotes the (F variant of the) LM test of no serial correlation in the residuals up to and including order q. The numbers in parentheses below parameter estimates and test statistics are standard errors and p-values, respectively.
### Table 3.6: Diagnostic tests of parameter constancy and no remaining nonlinearity in TV-STAR models

<table>
<thead>
<tr>
<th>Transition variable</th>
<th>$D_{s,t}$</th>
<th>$\Delta_{yt-j}$</th>
<th>$\sigma^2_{\varepsilon}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LM$_1$</td>
<td>LM$_3$</td>
<td>LM$_1$</td>
</tr>
<tr>
<td>Canada</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t$</td>
<td>0.47</td>
<td>0.34</td>
<td>0.58</td>
</tr>
<tr>
<td>$\Delta_{4yt-1}$</td>
<td>0.39</td>
<td>0.88</td>
<td>0.60</td>
</tr>
<tr>
<td>$\Delta_{4yt-2}$</td>
<td>0.61</td>
<td>0.96</td>
<td>0.43</td>
</tr>
<tr>
<td>$\Delta_{4yt-3}$</td>
<td>0.88</td>
<td>0.93</td>
<td>0.76</td>
</tr>
<tr>
<td>$\Delta_{4yt-4}$</td>
<td>0.83</td>
<td>0.90</td>
<td>0.53</td>
</tr>
<tr>
<td>Finland</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t$</td>
<td>0.35</td>
<td>0.70</td>
<td>0.66</td>
</tr>
<tr>
<td>$\Delta_{4yt-1}$</td>
<td>0.11</td>
<td>0.11</td>
<td>0.40</td>
</tr>
<tr>
<td>$\Delta_{4yt-2}$</td>
<td>0.14</td>
<td>0.16</td>
<td>0.65</td>
</tr>
<tr>
<td>$\Delta_{4yt-3}$</td>
<td>0.75</td>
<td>0.19</td>
<td>0.79</td>
</tr>
<tr>
<td>$\Delta_{4yt-4}$</td>
<td>0.46</td>
<td>0.13</td>
<td>0.53</td>
</tr>
<tr>
<td>France</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t$</td>
<td>0.67</td>
<td>0.18</td>
<td>0.23</td>
</tr>
<tr>
<td>$\Delta_{4yt-1}$</td>
<td>0.69</td>
<td>0.77</td>
<td>0.47</td>
</tr>
<tr>
<td>$\Delta_{4yt-2}$</td>
<td>0.74</td>
<td>0.72</td>
<td>0.83</td>
</tr>
<tr>
<td>$\Delta_{4yt-3}$</td>
<td>0.91</td>
<td>0.75</td>
<td>0.65</td>
</tr>
<tr>
<td>$\Delta_{4yt-4}$</td>
<td>0.80</td>
<td>0.48</td>
<td>0.93</td>
</tr>
<tr>
<td>Germany</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t$</td>
<td>0.91</td>
<td>0.83</td>
<td>0.61</td>
</tr>
<tr>
<td>$\Delta_{4yt-1}$</td>
<td>0.54</td>
<td>0.24</td>
<td>0.22</td>
</tr>
<tr>
<td>$\Delta_{4yt-2}$</td>
<td>0.29</td>
<td>0.11</td>
<td>0.89</td>
</tr>
<tr>
<td>$\Delta_{4yt-3}$</td>
<td>0.70</td>
<td>0.90</td>
<td>0.83</td>
</tr>
<tr>
<td>$\Delta_{4yt-4}$</td>
<td>0.41</td>
<td>0.60</td>
<td>0.40</td>
</tr>
<tr>
<td>Italy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t$</td>
<td>0.71</td>
<td>0.90</td>
<td>0.39</td>
</tr>
<tr>
<td>$\Delta_{4yt-1}$</td>
<td>0.85</td>
<td>0.88</td>
<td>0.17</td>
</tr>
<tr>
<td>$\Delta_{4yt-2}$</td>
<td>0.86</td>
<td>0.89</td>
<td>0.39</td>
</tr>
<tr>
<td>$\Delta_{4yt-3}$</td>
<td>0.51</td>
<td>0.93</td>
<td>0.27</td>
</tr>
<tr>
<td>$\Delta_{4yt-4}$</td>
<td>0.55</td>
<td>0.89</td>
<td>0.50</td>
</tr>
</tbody>
</table>

**Notes:** The table contains $p$-values of $F$-variants of LM diagnostic tests of parameter constancy (rows labelled $t$) and no remaining nonlinearity (rows labelled $\Delta_{yt-l}$, with $l = 1, \ldots, 4$) of seasonal dummy coefficients (columns headed $D_{s,t}$), autoregressive parameters (columns headed $\Delta_{yt-j}$), and residual variance (columns headed $\sigma^2_{\varepsilon}$) in estimated TV-STAR models for quarterly industrial production growth rates.
### Table 3.6: Diagnostic tests of parameter constancy and no remaining nonlinearity in TV-STAR models, (cont.)

<table>
<thead>
<tr>
<th>Transition variable</th>
<th>$D_{s,t}$</th>
<th>$\Delta y_{t-j}$</th>
<th>$\sigma^2_\epsilon$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$LM_1$</td>
<td>$LM_3$</td>
<td>$LM_1$</td>
</tr>
<tr>
<td><strong>Japan</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t$</td>
<td>0.34 0.047</td>
<td>0.87 0.35</td>
<td>0.68 0.65</td>
</tr>
<tr>
<td>$\Delta_4y_{t-1}$</td>
<td>0.21 0.025</td>
<td>0.49 0.21</td>
<td>0.28 0.41</td>
</tr>
<tr>
<td>$\Delta_4y_{t-2}$</td>
<td>0.35 0.58</td>
<td>0.73 0.31</td>
<td>0.89 0.99</td>
</tr>
<tr>
<td>$\Delta_4y_{t-3}$</td>
<td>0.59 0.44</td>
<td>0.87 0.95</td>
<td>0.41 0.83</td>
</tr>
<tr>
<td>$\Delta_4y_{t-4}$</td>
<td>0.56 0.78</td>
<td>0.70 0.95</td>
<td>0.14 0.52</td>
</tr>
<tr>
<td><strong>Sweden</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t$</td>
<td>0.76 0.20</td>
<td>0.58 0.42</td>
<td>0.57 0.33</td>
</tr>
<tr>
<td>$\Delta_4y_{t-1}$</td>
<td>0.99 0.20</td>
<td>0.76 0.89</td>
<td>0.24 0.49</td>
</tr>
<tr>
<td>$\Delta_4y_{t-2}$</td>
<td>0.90 0.94</td>
<td>0.85 0.44</td>
<td>0.38 0.51</td>
</tr>
<tr>
<td>$\Delta_4y_{t-3}$</td>
<td>0.36 0.46</td>
<td>0.31 0.72</td>
<td>0.56 0.69</td>
</tr>
<tr>
<td>$\Delta_4y_{t-4}$</td>
<td>0.48 0.57</td>
<td>0.34 0.57</td>
<td>0.31 0.43</td>
</tr>
<tr>
<td><strong>United States</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t$</td>
<td>0.43 0.62</td>
<td>0.52 0.81</td>
<td>0.27 0.038</td>
</tr>
<tr>
<td>$\Delta_4y_{t-1}$</td>
<td>0.83 0.25</td>
<td>0.60 0.91</td>
<td>0.11 0.26</td>
</tr>
<tr>
<td>$\Delta_4y_{t-2}$</td>
<td>0.70 0.84</td>
<td>0.75 0.60</td>
<td>0.34 0.75</td>
</tr>
<tr>
<td>$\Delta_4y_{t-3}$</td>
<td>0.98 0.18</td>
<td>0.78 0.58</td>
<td>0.41 0.75</td>
</tr>
<tr>
<td>$\Delta_4y_{t-4}$</td>
<td>0.64 0.038</td>
<td>0.80 0.60</td>
<td>0.55 0.84</td>
</tr>
</tbody>
</table>

**Notes:** The table contains $p$-values of $F$-variants of LM diagnostic tests of parameter constancy (rows labelled $t$) and no remaining nonlinearity (rows labelled $\Delta y_{t-l}$, with $l = 1, \ldots, 4$) of seasonal dummy coefficients (columns headed $D_{s,t}$), autoregressive parameters (columns headed $\Delta y_{t-j}$), and residual variance (columns headed $\sigma^2_\epsilon$) in estimated TV-STAR models for quarterly industrial production growth rates.
Chapter 4

Testing the Granger Noncausality Hypothesis in Stationary Nonlinear Models of Unknown Functional Form
“You see, there is only one constant, one universal, it is the only real truth: causality. Action. Reaction. Cause and effect.”

– Merovingian, The Matrix Reloaded
4.1 Introduction

In a seminal paper, Granger (1969) introduced an operational definition of causality between two variables. In particular, if the variance of the prediction error of the first variable is reduced by including measurements from the second variable, then the second variable is said to have a causal influence on the first variable. This definition has since formed the starting-point for testing the null hypothesis of one variable not causing the other. Note, however, that prediction in the original definition has in practice come to mean in-sample, not necessarily out-of-sample, prediction. The testing has most often been carried out in the linear framework. Geweke (1984) and Lütkepohl (1993) provide comprehensive surveys of the literature. For an example of a genuine out-of-sample application, see Ashley, Granger, and Schmalensee (1980).

During the last decade there has been growing interest in generalizing the test to allow for nonlinear relationships between variables. Baek and Brock (1992) suggested a generalization based on the BDS test described in Brock, Dechert, Scheinkman, and LeBaron (1996); Hiemstra and Jones (1994) proposed another version of this test, relaxing the iid assumption. In 1996 Bell, Kay, and Malley proposed a procedure for causality testing between two univariate time series using non-parametric regression ("generalized" additive models). The above-mentioned tests are all nonparametric and rather computer-intensive. Skalin and Teräsvirta (1999) proposed a parametric test based on the smooth transition regression model and applied that to a set of long Swedish macroeconomic series. That test is easy to compute, but it relies on assumptions about the specific functional form of the causal relationship. Chen, Rangarjan, Feng, and Ding (2004) extended Granger's idea to nonlinear situations by proposing a procedure based on local linear approximation of the nonlinear function. Apparently, no asymptotic distribution theory is available for inference in this framework and the results are only descriptive.

In this chapter, we suggest two new tests that require little knowledge of the functional relationship between the two variables. The idea is to globally approximate the potential causal relationship between the variables by a Taylor series expansion, which can be seen as a way of linearizing the testing problem. In that sense, noncausality tests based on a single linear regression form a special case in which the Taylor series expansion approximating the actual relationship is of order one. In other words, our framework nests the linear case. Compared to nonparametric procedures, the tests introduced in this chapter are very easy to compute. They are also available in large samples where the computational burden of nonparametric techniques becomes prohibitive. Rech, Teräsvirta, and Tschernig (2001) recently applied this idea to
nonlinear variable selection.

The chapter is organized as follows. Section 4.2 contains a description of our noncausality tests. Section 4.3 reports results of a simulation study: both the size and the power of these tests are investigated by Monte Carlo experiments. Section 4.4 concludes.

4.2 Tests of the Granger noncausality hypothesis

4.2.1 Standard linear Granger noncausality test

Before turning out to Taylor expansion-based causality testing, we first recall the standard way for testing the linear Granger noncausality hypothesis. In that framework, a series $x_t$ is defined not to (linearly) Granger cause another series $y_t$ $x$ NGC $y$ if the null hypothesis

$$H_0 : \beta_1 = \ldots = \beta_q = 0$$

holds in

$$y_t = \theta_0 + \theta_1 y_{t-1} + \ldots + \theta_p y_{t-p} + \beta_1 x_{t-1} + \ldots + \beta_q x_{t-q} + \epsilon_t.$$  \hspace{1cm} (4.2)

We make the following assumptions:

A1. $\{\varepsilon_t\}$ is a sequence of normal, independent $(0, \sigma^2)$ errors.

A2. $\{y_t\}$ is stationary and ergodic under (4.1), that is, the roots of the lag polynomial $1 - \sum_{j=1}^{q} \theta_j L^j$ lie outside the unit circle.

A3. $\{x_t\}$ is a weakly stationary and ergodic sequence.

If $\{x_t\}$ is a linear autoregressive-moving average process, then the process is stationary if and only if the roots of the autoregressive lag polynomial lie outside the unit circle. In the nonlinear case, probabilistic properties, such as stationarity and ergodicity, do not seem to be available except in a few special cases, see Stensholt and Tjostheim (1987), for example.

Assumption A1 is made to allow maximum likelihood-based inference. Robustifying the inference against non-normal errors is possible, however, but is not considered here. Assumptions A2 and A3 guarantee the existence of the second moments needed for the standard distribution theory to be valid.

Under these assumptions one can test the noncausality hypothesis in the single equation setup (4.2) using an LM-type test statistic (denoted by the
subscript \( SE \)). Following the recommendation in many earlier papers, we use an \( F \)-approximation to the asymptotically \( \chi^2 \)-distributed statistic:

\[
\text{Linear}_{SE} = \frac{(SSR_0 - SSR_1)/q}{SSR_1/(T - p - q - 1)} \quad \approx \quad F_{q, T-p-q-1},
\]

where \( SSR_0 \) and \( SSR_1 \) are sums of squared residuals from regressions under the null and the alternative hypotheses, respectively, and \( T \) is the number of observations. The test for testing the null of \( y_t \) not Granger causing \( x_t \) (\( y \text{ NGC } x \)) can be defined analogously.

Testing the noncausality hypothesis within (4.2) contains the implicit assumption that \( y_t \) does not Granger cause \( x_t \). If this assumption is not valid, testing has to, in principle, be carried out within a bivariate system. Testing the hypothesis of \( x_t \) not causing \( y_t \) amounts to testing

\[
H_0 : \beta_{11} = \ldots = \beta_{1q_1} = 0
\]

in the system:

\[
\begin{align*}
 y_t &= \theta_{10} + \theta_{11}y_{t-1} + \ldots + \theta_{1p_1}y_{t-p_1} + \beta_{11}x_{t-1} + \ldots + \beta_{1q_1}x_{t-q_1} + \varepsilon_{1t} \\
 x_t &= \theta_{20} + \theta_{21}y_{t-1} + \ldots + \theta_{2p_2}y_{t-p_2} + \beta_{21}x_{t-1} + \ldots + \beta_{2q_2}x_{t-q_2} + \varepsilon_{2t},
\end{align*}
\]

where \( \varepsilon_{it} \) are assumed white noise with a variance-covariance matrix

\[
\Sigma = \begin{pmatrix}
\sigma_{11}^2 & \sigma_{12}^2 \\
\sigma_{21}^2 & \sigma_{22}^2
\end{pmatrix}
\]

under \( H_0 \).

The \( F \)-version of the LM-test, see Bewley (1986), for testing (4.4) within the equation system with feedback (4.5), denoted by the subscript \( FB \), can then be computed as

\[
\text{Linear}_{FB} = \frac{T}{q_1} \left( m - \text{tr} (\widehat{\Omega}_1 \widehat{\Omega}_0^{-1}) \right) \quad \approx \quad F_{q_1, T},
\]

where \( m \) is the number of equations in the system. Matrices \( \widehat{\Omega}_0 = \widehat{E}_0' \widehat{E}_0 \) and \( \widehat{\Omega}_1 = \widehat{E}_1' \widehat{E}_1 \) are the cross-product matrices of the residuals from estimating the model under the null and under the alternative, respectively. More specifically, \( \widehat{E}_0 = (\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_T)' \) and \( \widehat{E}_1 = (\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_T)' \) where \( \hat{\varepsilon}_i \) and \( \hat{\varepsilon}_t \), \( t = 1, \ldots, T \) are the \((m \times 1)\) vectors of restricted and unrestricted estimated residuals, respectively. Analogously, the hypothesis \( y \text{ NGC } x \) corresponds to \( H_0 : \theta_{21} = \ldots = \theta_{2p_2} = 0 \) in (4.5).

### 4.2.2 Framework for the general test

Suppose now that we have two weakly stationary and ergodic time series \( \{x_t\} \) and \( \{y_t\} \). The functional form of the relationship between the two is unknown
but it is assumed that the possible causal relationship between \( y \) and \( x \) is adequately represented by the following equation:

\[
y_t = f_y(y_{t-1}, \ldots, y_{t-p}, x_{t-1}, \ldots, x_{t-q}; \theta) + \varepsilon_{1t}, \tag{4.7}
\]

where \( \theta \) is a parameter vector and \( \varepsilon_{1t} \sim \text{nid}(0, \sigma_1^2) \). In this framework, \( x \) does not Granger cause \( y \) if

\[
f_y(y_{t-1}, \ldots, y_{t-p}, x_{t-1}, \ldots, x_{t-q}; \theta) = f^*(y_{t-1}, \ldots, y_{t-p}; \theta^*). \tag{4.8}
\]

This means that the conditional mean of \( y_t \) given the past values of \( x_t \) and \( y_t \) is not a function of the past values of \( x_t \).

If the possibility of causality from \( y \) to \( x \) cannot be excluded a priori, one has to assume that there exists a reduced form of the relationship between the two variables. Its precise form is unknown but we assume that it is represented by the following bivariate system:

\[
y_t = f_y(y_{t-1}, \ldots, y_{t-p_1}, x_{t-1}, \ldots, x_{t-q_1}; \theta_1) + \varepsilon_{1t} \\
x_t = f_x(y_{t-1}, \ldots, y_{t-p_2}, x_{t-1}, \ldots, x_{t-q_2}; \theta_2) + \varepsilon_{2t}, \tag{4.9}
\]

where \( \theta_i, i = 1, 2 \), are parameter vectors and \( \varepsilon_{it} \sim \text{nid}(0, \sigma_i^2) \) and \( E\varepsilon_{it}\varepsilon_{js} = 0 \) for all \( t, s \). In this framework, \( x \) NGC \( y \) if

\[
f_y(y_{t-1}, \ldots, y_{t-p_1}, x_{t-1}, \ldots, x_{t-q_1}; \theta_1) = f^*(y_{t-1}, \ldots, y_{t-p_1}; \theta_1^*) \tag{4.10}
\]

in (4.9). Analogously, \( y \) NGC \( x \) if

\[
f_x(y_{t-1}, \ldots, y_{t-p_2}, x_{t-1}, \ldots, x_{t-q_2}; \theta_2) = f^{**}(x_{t-1}, \ldots, x_{t-q_2}; \theta_2^*) \tag{4.11}
\]

in (4.9). The question is how to test noncausality hypothesis within (4.9).

### 4.2.3 Noncausality tests based on a Taylor series approximation

In order to test either (4.10) or (4.11) in (4.9) when \( f_y \) and \( f_x \) are unknown, we assume that Assumption A1 is valid for \( \{\varepsilon_{1t}, \varepsilon_{2t}\} \). Furthermore, to make our test operational we make the following assumption:

\[ A4. \text{ Functions } f_y \text{ and } f_x \text{ have convergent Taylor expansions at any arbitrary point of the sample space for every } \theta \in \Theta \text{ (the parameter space).} \]
4.2. Tests of the Granger noncausality hypothesis

The test will be based on a finite-order Taylor expansion. Assumption A4 is required to ensure that the remainder of the Taylor expansion converges to zero when the order of the Taylor expansion increases. If this is not the case, our test cannot be expected to work well.

The null hypothesis of no Granger causality can be tested as follows. First, linearize \( f_y \) and \( f_x \) by approximating them with general polynomials of sufficiently high order (this can be done arbitrarily accurately by the Stone-Weierstrass theorem). After merging terms and reparameterizing, the \( k \)th-order Taylor approximation of \( f_y \) has the following form:

\[
\begin{align*}
y_t &= \beta_0 + \sum_{j=1}^{p_1} \beta_j y_{t-j} + \sum_{j=1}^{q_1} \gamma_j x_{t-j} + \\
&+ \sum_{j_1=1}^{p_1} \sum_{j_2=1}^{p_1} \beta_{j_1j_2} y_{t-j_1} y_{t-j_2} + \sum_{j_1=1}^{p_1} \sum_{j_2=1}^{q_1} \delta_{j_1j_2} y_{t-j_1} x_{t-j_2} + \\
&+ \sum_{j_1=1}^{q_1} \sum_{j_2=1}^{q_1} \gamma_{j_1j_2} x_{t-j_1} x_{t-j_2} + \cdots + \\
&+ \sum_{j_1=1}^{p_1} \sum_{j_2=1}^{p_1} \cdots \sum_{j_k=1}^{p_1} \beta_{j_1\cdots j_k} y_{t-j_1} \cdots y_{t-j_k} + \cdots + \\
&+ \sum_{j_1=1}^{q_1} \sum_{j_2=1}^{q_1} \cdots \sum_{j_k=1}^{q_1} \gamma_{j_1\cdots j_k} x_{t-j_1} \cdots x_{t-j_k} + \epsilon_t,
\end{align*}
\]

where the error term \( \epsilon_t \) includes the approximation error \( R_t^{(k)} = f_y - T_y^k(y, x) \), and \( q_1 \leq k \) and \( p_1 \leq k \) for notational convenience. Expansion (4.12) contains all possible combinations of lagged values of \( y_t \) and lagged values of \( x_t \) up to order \( k \). A similar expression can be defined for \( x_t \), and the testing is carried out within the system

\[
\begin{align*}
y_t &= T_y^k(y, x) + \epsilon_{1t} \\
x_t &= T_x^k(x, y) + \epsilon_{2t}.
\end{align*}
\]

**General test**

Owing to the approximation (4.12), the testing problem is straightforward as it has been returned to the problem of testing a linear hypothesis in a bivariate
system that is linear in parameters. The assumption that \( x_t \) does not Granger cause \( y_t \) means that all terms involving functions of lagged values of \( x_t \) in (4.12) must have zero coefficients. In the most general case\(^1\), the null hypothesis of \( x_t \) not Granger causing \( y_t \) can be written as

\[
H_{02} : \begin{cases} 
\gamma_j = 0, & j = 1, \ldots, q_1 \\
\delta_{j_1, j_2} = 0, & j_1 = 1, \ldots, p_1, \ j_2 = 1, \ldots, q_1 \\
\gamma_{j_1, j_2} = 0, & j_1 = 1, \ldots, q_1, \ j_2 = j_1, \ldots, q_1 \\
\vdots \\
\gamma_{j_1 \ldots j_k} = 0, & j_1 = 1, \ldots, q_1, \ j_2 = j_1, \ldots, q_1, \ldots, j_k = j_{k-1}, \ldots, q_1.
\end{cases}
\]

In practice, this test can be carried out in ordinary fashion by using an \( F \)-version of the LM test. Thus,

\[
\text{General}_{FB} = \frac{T}{N_1} \left( m - \text{tr} (\hat{\Omega}_1 \hat{\Omega}_0^{-1}) \right) \overset{\text{approx}}{\sim} F_{N_1, T},
\]

where the matrix \( \hat{\Omega}_0 = \hat{E}_0' \hat{E}_0 \) is obtained from regression (4.13) under the null hypothesis and \( \hat{\Omega}_1 = \hat{E}_1' \hat{E}_1 \) from the full auxiliary regression (4.13). Furthermore, \( T \) is the number of observations and \( N_1 \) the number of parameters in (4.12) to be tested under the null hypothesis. The latter quantity is defined as follows:

\[
N_1 = N - N_2 = \left( 1 + \sum_{r=1}^{k} \left( \frac{p_1 + q_1 + r - 1}{r} \right) \right) - \left( 1 + \sum_{r=1}^{k} \left( \frac{p_1 + r - 1}{r} \right) \right),
\]

where \( N \) is the total number of parameters and \( N_2 \) the number of parameters not under test.

As already mentioned, this testing problem is straightforward in the sense that one tests a linear hypothesis in a linearized system. Note, however, that the error term also contains the remainder term, \( R_t^{(k)} \). To avoid size distortion, \( R_t^{(k)} \) has to be small, that is, \( k \) has to be sufficiently large. Asymptotically, however, the test does not have its assumed nominal size for any fixed \( k \) unless the \( k \)-th order Taylor expansion of function \( f_y \) is an exact characterization of this function. On the other hand, the approximation error may be made arbitrarily small by increasing \( k \) and thus the number of parameters \( N \), say, in (4.12) as \( T \to \infty \). Showing that standard asymptotic theory works when

\(^1\)We are only going to present here the example related to the bivariate case. Extensions to higher-dimensional systems are straightforward.
the approximation error converges to zero \((N \to \infty)\), such that \(N/T \to 0\) as \(T \to \infty\), may be difficult. At any rate, the present testing problem is essentially a small-sample problem, and the finite-sample behaviour of the standard distribution theory will be investigated by simulation.

There are two practical difficulties related to equation (4.13). One is numerical whereas the other one has to do with the amount of information. Numerical problems may arise because the regressors in (4.13) tend to be highly collinear if \(k, p_1\) and \(q_1\) (and also \(p_2, q_2\)) are large. The other difficulty is that the number of regressors increases rapidly with \(k\), so the dimension of the null hypothesis may become rather large. For instance, when \(p_1 = 2\) and \(q_1 = 3\) then \(N_1 = 46\) for \(k = 3\), and \(N_1 = 231\) when \(k = 5\). A practical solution to both problems is to replace some matrices by their largest principal components as follows. First, divide the regressors in the auxiliary test-equation equation, say, (4.12), into two groups: those being functions of lags of \(y_t\) only and the remaining ones. Replace the second group of regressors by their first \(p^*\) principal components. The null hypothesis now is that the principal components have zero coefficients. This yields the following test statistic:

\[
\text{General}_{FB}^* = \frac{T}{p^*} \left( m - \text{tr} \left( \tilde{\Omega}_1 \tilde{\Omega}_0^{-1} \right) \right) \approx \frac{F_{p^*, T}}{\Omega_0}.
\]

The "remainder" now also includes the approximation error due to the omitted principal components related to the smallest eigenvalues.

**Semi-additive test**

In some cases it may be reasonable to assume that the general model is "semi-additive". That means that the model has the following form:

\[
y_t = g_y(y_{t-1}, \ldots, y_{t-p_1}; \theta_{11}) + f_y(x_{t-1}, \ldots, x_{t-q_1}; \theta_{12}) + \varepsilon_{1t} \\
x_t = g_x(y_{t-1}, \ldots, y_{t-p_2}; \theta_{21}) + f_x(x_{t-1}, \ldots, x_{t-q_2}; \theta_{22}) + \varepsilon_{2t}
\]

(4.15)

Here we assume again that \(\{x_t\}\) and \(\{y_t\}\) are weakly stationary and ergodic, and that functions \(f_y, f_x, g_y\) and \(g_x\) satisfy the same assumptions as before.

We state again that \(x_t\) does not cause \(y_t\) if the past values of \(x_t\) contain no information about \(y_t\) that is not already contained in the past values of \(y_t\). When this is the case \(f_y(x_{t-1}, \ldots, x_{t-q_1}; \theta_{12}) \equiv \text{constant}\). The functions \(g_y, g_x, f_y\) and \(f_x\) are now separately expanded into \(k\)th-order Taylor series. For example, linearizing \(g_y\) and \(f_y\) in (4.15) by expanding both functions into a \(k\)th-order Taylor series around arbitrary points in the sample space, merging
Granger Causality

\[ y_t = \beta_0 + \sum_{j=1}^{p_1} \beta_j y_{t-j} + \sum_{j=1}^{q_1} \gamma_j x_{t-j} + \sum_{j_1=1}^{p_1} \sum_{j_2=1}^{p_1} \beta_{j_1,j_2} y_{t-j_1} y_{t-j_2} + \]
\[ \sum_{j_1=1}^{q_1} \sum_{j_2=1}^{q_1} \gamma_{j_1,j_2} x_{t-j_1} x_{t-j_2} + \cdots + \sum_{j_1=1}^{p_1} \sum_{j_2=1}^{p_1} \cdots \sum_{j_k=1}^{p_1} \beta_{j_1\cdots j_k} y_{t-j_1\cdots j_k} + \epsilon_t, \]
\[ (4.16) \]

where \( q_1 \leq k \) and \( p_1 \leq k \) for notational convenience. Expansion (4.16) contains all possible combinations of \( y_{t-j} \) and \( x_{t-i} \) up to order \( k \), but no cross-terms. Therefore, the hypothesis \( H \) becomes:

\[ H_{03} : \left\{ \begin{array}{l}
\gamma_j = 0, \ j = 1, \ldots, q_1 \\
\gamma_{j_1,j_2} = 0, \ j_1 = 1, \ldots, q_1, \ j_2 = j_1, \ldots, q_1 \\
\vdots
\gamma_{j_1\cdots j_k} = 0, \ j_1 = 1, \ldots, q_1, \ j_2 = j_1, \ldots, q_1, \ldots, j_k = j_{k-1}, \ldots, q_1.
\end{array} \right. \]

The number of parameters to be tested under the null hypothesis is

\[ N_{11} = \sum_{r=1}^{k} \binom{q_1 + r - 1}{r}. \]

Although the number of parameters for any fixed \( k \) is smaller than in the unrestricted nonadditive case, the problems of collinearity and the large dimension of the null hypothesis may still be present. The previous solution still applies: the regressors are replaced by \( p^* \) principal components of corresponding observation matrix. Again an LM-type test can be used, and the resulting test statistic is called \( Additive_{FB}^* \). Under the null hypothesis, \( Additive_{FB}^* \) has approximately an \( F \)-distribution with \( p^* \) and \( T \) degrees of freedom. Note that approximating \( f_y \) through principal components may only affect the power of the test, not its size.

If equation (4.15) is valid, then the corresponding test can be expected to be more powerful than the ones based on equation (4.9). Conversely, applying \( Additive_{FB}^* \) when the relationship is not semi-additive as in (4.15) may result in a substantial loss of power compared to the power of \( General_{FB}^* \).

4.2.4 Discussion

As discussed in an earlier section, our approach is based on global approximation of the unknown nonlinear function. The starting-point for the local
linear approximation of Chen et al. (2004) is the standard delay coordinate embedding reconstruction of the phase space attractors, see Boccaletti, Valladares, Pecora, Geffert, and Carroll (2002). A full description of a given attractor requires a nonlinear set of equations, but it is possible to locally approximate the dynamics linearly by a vector autoregressive model. Chen et al. then test for Granger causality at each local neighbourhood, average the resulting statistical quantities over the entire attractor and compute the so-called Extended Granger Causality Index. There are a number of decisions one has to make when using their method: one has to determine the embedding dimension and time delay. Determining the optimal neighbourhood size is also a nontrivial issue. It appears that no asymptotic distribution theory is available for inference in this framework, so the results are obviously bound to be rather descriptive. It may be guessed that an application of this procedure requires rather long time series unless the nonlinear relationship is nearly linear.

4.3 Monte Carlo experiments

4.3.1 Simulation design

In this section we shall investigate the small-sample performance of the proposed noncausality tests. We compare the tests with the standard linear testing procedure because that is what practitioners generally use in their work. Moreover, it is often the case that the researcher chooses to ignore the possible presence of feedback (causality in the other direction) and conducts the analysis within a single equation. One may then ask: does it matter whether the presence of feedback is explicitly acknowledged or not?

We report the results for all tests based both on the bivariate equation system (denoted with subscript FB) and on the single equation only (denoted with subscript SE), and compare the results. The tests included in our comparison are:

- \( \text{Linear}_{SE} \) and \( \text{Linear}_{FB} \) defined in section (4.2.1), formulas (4.3), (4.6)
- \( \text{General}^*_{SE} \) and \( \text{General}^*_{FB} \) defined in section (4.2.3), formula (4.14)
- \( \text{Additive}^*_{SE} \) and \( \text{Additive}^*_{FB} \) defined in section (4.2.3).

We present our size and power results for all tests graphically as Davidson and MacKinnon (1998) have recommended. Their graphs are easier to interpret than the conventional tables typically used for reporting such results.
The basis of these graphs is the empirical distribution function (EDF) of the $p$-values of the simulated realizations $\tau_j$, $j = 1, \ldots, N$, of some test statistic $\tau$. Let $p_j$ be the $p$-value associated to $\tau_j$, i.e., $p_j = p(\tau_j) = P(\tau > \tau_j)$, the probability of observing a value of $\tau$ greater than $\tau_j$ of the statistic. The EDF of the $p_j$’s is defined by:

$$\hat{F}(\xi_i) = \frac{1}{N} \sum_{j=1}^{N} I(p_j \leq \xi_i), \quad (4.17)$$

where $I$ is an indicator function given by:

$$I(p_j \leq \xi_i) = \begin{cases} 1 & \text{if } p_j \leq \xi_i \\ 0 & \text{otherwise} \end{cases}$$

and $\xi_i$ is a point in the $[0,1]$ interval. Following Davidson and MacKinnon (1998), a parsimonious set of values $\xi_i$, $i = 1, \ldots, m$, is

$$\xi_i = 0.002, 0.004, \ldots, 0.01, 0.02, \ldots, 0.99, 0.992, \ldots, 0.998 \quad (m = 107). \quad (4.18)$$

Concerning the size of the tests, we present the simple plot of $\hat{F}(\xi_i) - \xi_i$ against $\xi_i$ for each test. We know that if the distribution of $\tau$ is the one assumed under the null hypothesis, the $p_j$’s should be a sample from a uniform $[0,1]$ distribution. In that case, the plot of $\hat{F}(\xi_i) - \xi_i$ should be close to the $45^\circ$ line, whereas $\hat{F}(\xi_i) - \xi_i$ should fluctuate around zero as a function of $\xi_i$.

The results of the power comparisons are reported using simple power curves, instead of the size-corrected size-power curves advocated by Davidson and MacKinnon. The reason for this is that in practice we would not know how to size-correct the results, and our aim is to study the tests from the practitioners point of view. Therefore, we graph the loci of points $((\xi_i), \hat{F}^*(\xi_i))$ where the values of the $\xi_i$’s are given by (4.18), and $\hat{F}^*(\xi_i)$ is the EDF generated by a process belonging to the alternative hypothesis. In other words, we record the $p$-values for every Monte Carlo replication and just plot the curves corresponding to rejection rates at given nominal levels.

### 4.3.2 Simulation results

For all the experiments, the number of replications $N_R = 1000$ and the number of observations$^2$ $T = 150$. The innovations $\varepsilon_{it} \sim \text{nid}(0,1)$, $i = 1, 2$,

$^2$We let the data-generating process run for a while to get rid of the possible initial effects, i.e. we discard first 500 observations and use only the last 150.
4.3. Monte Carlo experiments

$t = 1, \ldots, T$, and sequences $\{\varepsilon_{1t}\}$ and $\{\varepsilon_{2t}\}$ are mutually independent in all experiments. We make use of the second-order Taylor approximation of $f_y, g_y, f_x$ and $g_x$ where $p_1 = 3; q_1 = 3; q_2 = 3; p_2 = 3$. The number of principal components is determined separately for each case. Only the largest principal components that together explain at least 90% of the variation in the matrix of observations are used. The system consists of unrelated equations and is estimated equation by equation by least squares.

For every DGP we present two graphs: panel (a) contains the results of the test of $x \text{ NGC } y$, and panel (b) the results of the test of $y \text{ NGC } x$. In every panel the performance of the three causality tests Linear, General* and Additive*, is reported, both for the single equation (SE) and the system (FB) framework. The corresponding lines on graphs are labelled Linear_SE, General*_SE, Additive*_SE, Linear_FB, General*_FB and Additive*_FB, respectively.

Empirical size of the tests

To illustrate the behaviour of the tests under the null hypothesis, we simulated six different systems. The data-generating processes are presented together with the $p$-values of the linearity test of Harvill and Ray (1999), denoted by $HRp$. These $p$-values are reported in order to give the reader an indication of how nonlinear the systems are.

The first system is a first-order vector autoregressive model ($HRp = 0.59$):

\begin{align}
  y_t &= 1 + 0.3y_{t-1} + 0.1\varepsilon_{1t} \\
  x_t &= 0.4 - 0.63x_{t-1} + 0.2\varepsilon_{2t}.
\end{align}

(4.19)

The second experiment involves a nonlinear system where $y_t$ is generated by a logistic smooth transition autoregressive (STAR) model and $x_t$ follows a bilinear model ($HRp = 8 \times 10^{-10}$):

\begin{align}
  y_t &= (0.02 - 0.9y_{t-1} + 0.795y_{t-2})/(1 + \exp(-25(y_{t-1} - 0.02))) + 0.1\varepsilon_{1t} \\
  x_t &= 0.8 - 0.6x_{t-1} + 0.1\varepsilon_{2,t-1}x_{t-1} + 0.3\varepsilon_{2t}.
\end{align}

(4.20)

In the third system $y_t$ is ratio-polynomial and $x_t$ is generated by an exponential STAR model ($HRp = 0.054$):

\begin{align}
  y_t &= -0.8 + 0.9/(1 + y_{t-1}^2 + y_{t-2}^2) + 0.1\varepsilon_{1t} \\
  x_t &= (0.2 - 0.6x_{t-1} + 0.45x_{t-2})(1 - \exp(-10(x_{t-1})^2)) + 0.1\varepsilon_{2t}.
\end{align}

(4.21)
The fourth system consists of two self-exciting threshold autoregressive (SETAR) models as follows $(HRp = 2 \times 10^{-30})$:

\[
\begin{align*}
y_t &= \begin{cases} 
0.1y_{t-1} + \varepsilon_{1t} & y_{t-1} \leq 0 \\
-0.5y_{t-1} + \varepsilon_{1t} & y_{t-1} > 0
\end{cases} \\
x_t &= \begin{cases} 
-0.5 + 0.5x_{t-1} - 0.7x_{t-2} + \varepsilon_{2t} & x_{t-1} \leq 0 \\
0.5 - 0.3x_{t-1} + 0.2x_{t-2} + \varepsilon_{2t} & x_{t-1} > 0
\end{cases}
\end{align*}
\] (4.22)

The fifth system is linear with fifth-order autoregression such that causality is only running in one direction, from $y$ to $x$ $(HRp = 0.55)$:

\[
\begin{align*}
y_t &= 1.41y_{t-1} - 1.38y_{t-2} + 1.0813y_{t-3} - 0.23015y_{t-4} + 0.0182y_{t-5} + \varepsilon_{1t} \\
x_t &= 1 - 0.55x_{t-1} + 0.16x_{t-2} - 0.4y_{t-4} - 0.3y_{t-5} + \varepsilon_{2t}.
\end{align*}
\] (4.23)

The final system is a bivariate nonlinear MA model $(HRp = 2 \times 10^{-12})$:

\[
\begin{align*}
y_t &= \varepsilon_{1t} - 0.4\varepsilon_{1,t-1} + 0.3\varepsilon_{1,t-2} + 0.4\varepsilon_{1,t-3} \\
x_t &= \varepsilon_{2t} + 0.55\varepsilon_{2,t-1} - 0.3\varepsilon_{2,t-2} - 0.2\varepsilon_{2,t-3}.
\end{align*}
\] (4.24)

The results appear in Figures 4.1 - 4.6. They show the $p$-value discrepancy plots, i.e. the graphs of $\hat{F}(\xi_i) - \xi_i$ against $\xi_i$. These figures are reproduced for small nominal sizes that are of practical interest.

The size distortions seem minor at low levels of significance. The single equation-based tests seem somewhat less size distorted than the system-based ones. Also the Linear test seems better-sized than the General or Additive tests, unless there is feedback, as in Equation (4.23) is which case the Linear test is grossly oversized. Size distortions seen in Figure 4.5 occur partly because of the misspecified lag length under the null hypothesis: only three lags are used in Taylor expansions. But, there is feedback from $y$ to $x$ through fourth and fifth lag, and when lags of $x$ enter the auxiliary test-equation, they are found to be helpful in explaining $y$. The same explanation - too few lags used in the approximation - is valid when explaining the size distortions for the nonlinear moving average model. These size distortions can partly be removed by using more lags when approximating the possibly nonlinear causal relationship. The size distortions of test statistics General and Additive are much smaller for simple linear systems when the order of Taylor expansions is lowered, i.e. the approximation becomes closer to the true DGP and there are fewer nuisance auxiliary terms in the test equations. Naturally, the size distortions diminish when the error variance is reduced and when the true lag length is used in the Taylor approximation. We recommend first testing the
linearity of the system as in Harvill and Ray (1999), and if linearity is not rejected, using the Linear test should suffice.

**Empirical power of the tests**

In this subsection we consider a number of cases where (nonlinear) Granger causality is present between the variables. More precisely, the power-curve figures correspond to the following systems:

- **Figure 4.7** ($x \rightarrow y$ bilinear; $y \rightarrow x$ linear, $HRp = 7 \times 10^{-7}$):
  \[
  y_t = 0.5 + 0.1y_{t-1} + 0.5\varepsilon_{1,t-1}x_{t-1} + \varepsilon_{1t}
  
  x_t = 0.22 - 0.39x_{t-1} + 0.46y_{t-2} + \varepsilon_{2t}
  \]

- **Figure 4.8** ($x \rightarrow y$ (long)linear$^3$; $y \rightarrow x$ linear, $HRp = 0.847$):
  \[
  y_t = 1.1y_{t-1} - 0.56y_{t-2} + 0.1591y_{t-3} - 0.0119y_{t-4} + 0.55x_{t-4} + 0.1\varepsilon_{1t}
  
  x_t = 0.5 - 0.1566x_{t-1} + 0.2083x_{t-2} - 0.4y_{t-2} + 0.3\varepsilon_{2t}
  \]

- **Figure 4.9** ($y \rightarrow x$ (long)linear, $HRp = 0.575$):
  \[
  y_t = 1.41y_{t-1} - 1.38y_{t-2} + 1.08y_{t-3} - 0.23y_{t-4} + 0.02y_{t-5} + 0.5\varepsilon_{1t}
  
  x_t = 1 - 0.55x_{t-1} + 0.16x_{t-2} - 0.4y_{t-4} - 0.3y_{t-5} + 0.5\varepsilon_{2t}
  \]

- **Figure 4.10** ($x \rightarrow y$ semi-additive, $HRp = 0.243$):
  \[
  y_t = 0.1 + 0.4y_{t-2} + (0.5 - 0.8x_{t-1})/(1 + \exp(-5(x_{t-3} - 0.2))) + 0.5\varepsilon_{1t}
  
  x_t = 0.22 + 0.39x_{t-1} - 0.55x_{t-2} + 0.3\varepsilon_{2t}
  \]

- **Figure 4.11** ($x \rightarrow y$ semi-additive; $y \rightarrow x$ (long)linear, $HRp = 6 \times 10^{-30}$):
  \[
  y_t = 0.5y_{t-1} - 0.6x_{t-1}^2/(1 + x_{t-2}^2) + \varepsilon_{1t}
  
  x_t = 1 + 0.4x_{t-1} - 0.3x_{t-2} + 0.4y_{t-4} - 0.3y_{t-5} + 0.2y_{t-6} + \varepsilon_{2t}
  \]

- **Figure 4.12** ($x \rightarrow y$ general, $HRp = 8 \times 10^{-23}$):
  \[
  y_t = \begin{cases} 
  -1 + 0.5y_{t-1} - 0.9y_{t-2} + \varepsilon_{1t} & x_{t-1} \leq -0.2 \\
  2 + 0.3y_{t-1} + 0.2y_{t-2} + \varepsilon_{1t} & x_{t-1} > -0.2 
  \end{cases}
  
  x_t = 0.2 - 0.56x_{t-1} + \varepsilon_{2t}
  \]

$^3$By "(long)" we denote the situation where the lag length in the actual DGP is longer than the one used in approximations.
• Figure 4.13 \((x \rightarrow y\) general, \(HRp = 0.00883\)):

\[
y_t = 0.1 + 0.4y_{t-2} + (1 - 0.8y_{t-2})/(1 + \exp(-9x_{t-1}^2)) + 0.15\varepsilon_{1t}
\]
\[
x_t = 0.22 + 0.39x_{t-1} - 0.55x_{t-2} + 0.3\varepsilon_{2t}
\]

(4.31)

• Figure 4.14 \((x \rightarrow y\) general; \(y \rightarrow x\) linear, \(HRp = 1 \times 10^{-5}\)):

\[
y_t = 1 - 0.2y_{t-1} + (-1 + 0.5y_{t-2})(1 - \exp(-10x_{t-1}^2)) + 0.3\varepsilon_{1t}
\]
\[
x_t = 0.5x_{t-1} + 0.3y_{t-1} + 0.5\varepsilon_{2t}
\]

(4.32)

• Figure 4.15 \((x \rightarrow y\) bilinear; \(y \rightarrow x\) general, \(HRp = 9 \times 10^{-24}\)):

\[
y_t = 0.1y_{t-1} - 0.5y_{t-2} + 0.4x_{t-1}^2y_{t-1} + 0.5x_{t-2}^2y_{t-2} + 0.3\varepsilon_{1t}
\]
\[
x_t = 1 + 0.3x_{t-1} - 0.5x_{t-2} + (2 + 0.4x_{t-1} - 0.3x_{t-2} - 0.15x_{t-3})
\]
\[(1 - \exp(-10y_{t-2}^2)) + 0.1\varepsilon_{2t}
\]

(4.33)

• Figure 4.16 \((x \rightarrow y\) general; \(y \rightarrow x\) semi-additive, \(HRp = 8 \times 10^{-5}\)):

\[
y_t = \begin{cases} 
0.1y_{t-1} + 0.9x_{t-1}^2 + 0.4\varepsilon_{1t} & y_{t-1} \leq 0 \\
-0.5y_{t-1} + 0.4\varepsilon_{1t} & y_{t-1} > 0
\end{cases}
\]
\[
x_t = 0.3x_{t-1} + \frac{0.9}{(1 + x_{t-1}^2 + x_{t-2}^2)} - \frac{0.5}{(1 + \exp(-2y_{t-1}))} + 0.25\varepsilon_{2t}
\]

(4.34)

• Figure 4.17 \((x \rightarrow y\) general; \(y \rightarrow x\) general, \(HRp = 8 \times 10^{-8}\)):

\[
y_t = \begin{cases} 
0.1y_{t-1} + 0.3x_{t-1}^2 - 0.5x_{t-2}^2 + 0.2\varepsilon_{1t} & y_{t-1} \leq 0 \\
-0.3y_{t-1} - 0.5x_{t-1}^2 + 0.7x_{t-2}^2 + 0.2\varepsilon_{1t} & y_{t-1} > 0
\end{cases}
\]
\[
x_t = 0.3x_{t-1} + \frac{0.9}{(1 + x_{t-1}^2 + x_{t-2}^2)} + \frac{(-0.5x_{t-1} + 0.3x_{t-2})}{(1 + \exp(-30y_{t-1}))} + 0.24\varepsilon_{2t}
\]

(4.35)

• Figure 4.18 \((x \rightarrow y\) general; \(y \rightarrow x\) general, \(HRp = 3 \times 10^{-28}\)):

\[
y_t = \begin{cases} 
0.1y_{t-1} + \frac{(2 - 0.45x_{t-1})}{(1 + \exp(-5x_{t-1}))} + 0.4\varepsilon_{1t} & y_{t-1} \leq 0 \\
-0.5y_{t-1} + \frac{(1 - 0.3x_{t-1} + 0.45x_{t-2})}{(1 + \exp(-5x_{t-1}))} + 0.4\varepsilon_{1t} & y_{t-1} > 0
\end{cases}
\]
\[
x_t = 1 + 0.3x_{t-1}y_{t-2} - 0.15x_{t-2}y_{t-3} + \varepsilon_{2t}
\]

(4.36)
4.3. Monte Carlo experiments

The results of the *Linear* causality test do not offer any great surprises. It is clear that the test works best when the true causal relationship is linear (Figures 4.7(b), 4.8(b), 4.14(b)) but it may only have weak power when this is no longer the case. The *Additive* test as well as the *General* one both suffer somewhat from overparameterization when applied to linear systems. The *Linear* test also seems to perform well when it comes to detecting slowly evolving logistic STAR-type causal relationships, see Figures 4.10(a), 4.16(b). It also works surprisingly well for a case when the causing variable is the threshold variable in a two-regime TAR model, 4.12(a). *Linear* test also seems to be able to detect a (linear) causal relationship when the lags contributing to the explanation of the other variable are outside the range of the lags included in Taylor expansion (and thus used in the test), see Figure 4.11(b), 4.9(b).

At small nominal sizes, the *Additive* test is the best performer of these three tests in Figure 4.16(a), where the corresponding model actually is semi-additive. It is often more powerful than the *General* test even when the true model is no longer semi-additive, see Figures 4.13(a), 4.14(a), for example. Figures 4.14(a) and 4.15(b) illustrate the behaviour of the tests in the case where the causality is represented by an exponential smooth transition regression function and the causing variable is the transition variable. The nonlinearity in those models is of *General*-type, but the semi-additive approximation seems to capture most of the relationship. Consequently the *Additive* test is the most powerful one. From the low power of the *Linear* test it can be inferred that in this case excluding the higher order terms from the auxiliary regression is not a good idea.

Figures 4.7(a) and 4.15(a) correspond to systems with a bilinear equation and in those cases the *General* test strongly dominates the other test procedures. This may be expected as the relationship is no longer semi-additive, and making that assumption implies a loss of power. From Figures 4.17(a) and 4.18(a), it is seen that the *General* test also seems to perform well in a case where the causing variable enters through a regime or the regimes of a SETAR model.

When interpreting the results, one should be aware of the fact that the power of the tests depends on the variance of the error term $ce_t$ which controls the signal-to-noise ratio. Even some of the performance rankings indicated above may be changed by varying $c$. Also, for the ESTR-type models, the ability of the tests to detect causality depends quite heavily on the presence of the intercept in the nonlinear part. When there is only change in the amplitude of the fluctuations and no clear shift in the mean, the power of all tests is extremely low.
There seems to be no big difference between the single equation and system-based tests. This indicates that not much power is lost by ignoring the possible feedback. This is obviously due to the fact that the restrictions imposed by the null hypothesis are not cross-equation restrictions, so little is gained by including the unrestricted equation in the considerations.

4.4 Conclusions

The noncausality tests introduced in this chapter are based on standard statistical distribution theory. The size simulations indicate that the idea of polynomial approximation of unknown nonlinear functions is applicable in small samples. The right balance between the number of lags, the order of the Taylor expansion, the degree of nonlinearity and the number of observations is important, however, and has yet to be studied carefully. The power simulations suggest that the tests are indeed useful in discovering potential Granger causality between variables. They also demonstrate the obvious fact that the more we know about the functional form, the more we gain in terms of power. If the true causal relationship is nonlinear whereas testing is carried out under the assumption of linearity, the ensuing loss of power may be substantial. It is therefore advisable to test the Granger noncausality hypothesis both in the linear and the nonlinear framework to ensure that existing causal relationships between the variables are found as efficiently as possible. Because our tests are based on the idea of linearizing the unknown relationship, they are not computationally more difficult to carry out than traditional linear tests. However, the length of the time series may restrict the applicability of our technique. Given a sufficient amount of data, our tests should be a useful addition to the toolbox of both applied economists interested in investigating Granger causality and time series econometricians.
References


Figures

Figure 4.1: Size discrepancy plots, data generated from system (4.19).

Figure 4.2: Size discrepancy plots, data generated from system (4.20).
Granger Causality

Figure 4.3: Size discrepancy plots, data generated from system (4.21).

Figure 4.4: Size discrepancy plots, data generated from system (4.22).
Figures

Figure 4.5: Size discrepancy plots, data generated from system (4.23).

Figure 4.6: Size discrepancy plots, data generated from system (4.24).
Granger Causality

Figure 4.7: Power-curves, data generated from system (4.25).

Figure 4.8: Power-curves, data generated from system (4.26).
Figure 4.9: Power-curves, data generated from system (4.27).

Figure 4.10: Power-curves, data generated from system (4.28).
Figure 4.11: Power-curves, data generated from system (4.29).

Figure 4.12: Power-curves, data generated from system (4.30).
**Figure 4.13:** Power-curves, data generated from system (4.31).

**Figure 4.14:** Power-curves, data generated from system (4.32).
Figure 4.15: Power-curves, data generated from system (4.33).

Figure 4.16: Power-curves, data generated from system (4.34).
Figure 4.17: Power-curves, data generated from system (4.35).

Figure 4.18: Power-curves, data generated from system (4.36).
EFI
The Economic Research Institute

Reports since 1999
A complete publication list can be found at www.hhs.se/efi

Published in the language indicated by the title

2004
Andersson, A., Essays in Behavioral Finance.
Jutterström, M., Att påverka beslut – företag i EU:s regelsättande.
Larsson, P., Förändringens villkor. En studie av organisatoriskt lärande och förändring inom skolan.
Lagerwall, B., Empirical Studies of Portfolio Choice and Asset Prices.
Salabasis, M., Bayesian Time Series and Panel Models – Unit Roots, Dynamics and Random Effects.
Skallsjö, S., Essays on Term Structure and Monetary Policy.
Söderström, J., Från Produkt till Tjänst. Utveckling av affärs- och miljöstrategier i produktionerade företag.
Talia, K., The Scandinavian Currency Union, 1873-1924 - Studies in Monetary Integration and Disintegration

2003
Andersson, H., Valuation and Hedging of Long-Term Asset-Linked Contracts.
Bergman, M., Essays on Human Capital and Wage Formation.
Damsgaard, N., Deregulation and Regulation of Electricity Markets.
Eklund, B., Four Contributions to Statistical Inference in Econometrics.
Exploring Patterns in Information Management. Concepts and Perspectives for Understanding IT-Related Change. Sundgren, B, Mårtensson, P., Mähring, M., Nilsson, K., (editors)
Globalization and its Enemies, Lundahl, M (editor).
Hakkala, K., Essays on Restructuring and Production Decisions in Multi-Plant Firms.
Holgersson, C., Rekrytering av företagsledare. En studie i homosocialitet.
Lange, F., Brand Choice in Goal-derived Categories - What are the Determinants?
Le Coq, C., Quantity Choices and Market Power in Electricity Market.
Meisiek, S., Beyond the Emotional Work Event Social Sharing of Emotion in Organizations.
Nilsson, C., Processorientering och styrning – Regler, mål eller värderingar?
Sturluson, J.T., Topics in the Industrial Organization of Electricity Markets.
Tillberg, U., Ledarskap och samarbete - En jämförande fallstudie i tre skolor.
Waldenström, D., Essays in Historical Finance.
Wallén, U., Effektivitet i grundskolan i anslutning till en stadsdelsnämndsreform.

2002
Barinaga, E., Levelling Vagueness - A study of cultural diversity in an international project group.
Berglund, J., De otillräckliga - En studie av personalspecialisternas kamp för erkännande och status.
Bolander, P., Anställningsbilder och rekryteringsbeslut.
Damjanovic, T., Essays in Public Finance.
Företagserikan - Om kvinnor och entreprenörskap. Holmquist, C. och Sundin, E (red)
Heyman, F., Empirical Studies on Wages, Firm Performance and Job Turnover.
Kallifatides, M., Modern företagsledning och omoderna företagsledare.
Kaplan, M., Acquisition of Electronic Commerce Capability - The Cases of Compaq and Dell in Sweden.
Mähring, M., IT Project Governance.
Rekrytering av koncernstyrelsen - Nomineringsförfaranden och styrelsesammansättning med focus på kvinnors ställning och möjligheter.
Sjöstrand, S-E. och Petrelius, P.,(red)
Scener ur ett företag - Organiseringssteori för kunskaps samhället. Löwstedt, J.
Szymne, B.,(red).
Schenkel, A., Communities of Practice or Communities of Discipline - Managing Deviations at the Øresund Bridge.
Schuster, W., Företagets Valutarisk - En studie av horisontella och vertikala styreprocesser.
Skogsvik, S., Redovisningsmätt, värderelavans och informationseffektivitet.
Sundén, D., The Dynamics of Pension Reform.
Tullberg, J., Reciprocitet - Etiska normer och praktiskt samarbete.
Westling, G., Balancing Innovation and Control – The Role of Face-to-face Meetings in Complex Product Development Projects.
Viklund, M., Risk Policy – Trust, Risk Perception, and Attitudes.

2001
Adolfson, M., Monetary Policy and Exchange Rates – Breakthrough of Pass-Through.
Andersson, P., Expertise in Credit Granting: Studies on Judgment and Decision-Making behavior.
Björklund, C., Work Motivation - Studies of its Determinants and Outcomes.
Charpentier, C., Uppföljning av kultur- och fritidsförvaltningen efter stadsdelsnämndsreformen.
Dahlén, M., Marketing on the Web - Empirical Studies of Advertising and Promotion Effectiveness.
Eckerlund, I., Essays on the Economics of Medical Practice Variations.
Ekelund, M., Competition and Innovation in the Swedish Pharmaceutical Market.
Engström, S., Success Factors in Asset Management.
Ericsson, D., Kreativitetsmysteriet – Ledtrådar till arbetslivets kreativisering och skrivandets metafysik.
Eriksson, R., Price Responses to Changes in Costs and Demand.
Frisell, L., Information and Politics.
Giordani, P., Essays in Monetary Economics and Applied Econometrics.
Gustavsson, P., Essays on Trade, Growth and Applied Econometrics.
Hedlund, A., Konsumentens erfarenhet – och dess inverkan på livsmedelsinköp på Internet.
Hvenmark, J., Varför slocknar elden? Om utbriindhet bland chefer i ideella organisationer.
Hägglund, P.B., Företaget som investeringsobjekt – Hur placerare och analytiker arbetar med att ta fram ett investeringsobjekt.
Höök, P., Stridspiloter i vida kjolar, om ledarutveckling och jämställdhet.
Johansson, C., Styrning för samordning.
Josephson, J., Evolution and Learning in Games.
Kjellberg, H., Organising Distribution - Hakonbolaget and the efforts to rationalise food distribution, 1940-1960.
Lange, F. och Wahlund, R., Category Management – När konsumenten är manager.
Liljenberg, A., Customer-geared competition – A socio-Austrian explanation of Tertius Gaudens.
Lindkvist, B., Konsumentens erfarenhet – och dess inverkan på livsmedelsinköp på Internet.
Ljunggren, U., Nyckeltal i grundskolan i Stockholms stad före och efter stadsdelsnämndsreformen.
Löf, M., On Seasonality and Cointegration.
Martensen, K., Essays on Entry Externatilities and Market Segmentation.
Matros, A., Stochastic Stability and Equilibrium Selection in Games.
Mårtensson, P., Management Processes – An Information Perspective on Managerial Work.
Nilsson, A., Market Transparency.
Norberg, P., Finansmarknadsens amoralitet och det kalvinska kyrkorummet – En studie i ekonomisk mentalitet och etik.
Persson, B., Essays on Altruism and Health Care Markets.
Rech, G., Modelling and Forecasting Economic Time Series with Single Hidden-layer Feedforward Autoregressive Artificial Neural Networks.
Skoglund, J., Essays on Random Effects Models and GARCH.
Strand, N., Empirical Studies of Pricing.
Thorén, B., Stadsdelsnämndsreformen och det ekonomiska styrsystemet - Om budgetavvikelsor.

2000
Brodin, B., Lundkvist, L., Sjöstrand, S-E., Östman, L., Koncernchefen och ägarna.
Bornefalk, A., Essays on Social Conflict and Reform.
Edman, J., Information Use and Decision Making in Groups.
Emling, E., Svenskt familjeföretagande.
Ericson, M., Strategi, kalkyl, känsla.
Hyll, M., Essays on the Term Structure of Interest Rates.
Håkansson, P., Beyond Private Label – The Strategic View on Distributor Own Brands.
I huvudet på kunden. Söderlund, M (red). EFI och Liber Förlag.
Karlsson Stider, A., Familien och firman.
Ljunggren, U., Styning av grundskolan i Stockholms stad före och efter stadsdelsnärmändsreformen – Resultat från intervjuer och enkät.
Ludvigsen, J., The International Networking between European Logistical Operators.
Nittmar, H., Produktutveckling i samarbete – Strukturförändring vid införande av nya informationssystem.
Stenström, E., Konstiga företag.
Sweet, S., Industrial Change Towards Environmental Sustainability – The Case of Replacing Chlorofluorocarbons.
Tamm Hallström, K., Kampen för auktoritet – standardiseringsorganisationer i arbete.

1999
Adler, N., Managing Complex Product Development.
Allgulin, M., Supervision and Monetary Incentives.
Andersson, P., Experto Credite: Three Papers on Experienced Decision Makers.
Ekman, G., Från text till batong – Om poliser, busar och svennar.
Eliasson, A-C., Smooth Transitions in Macroeconomic Relationships.
Hamrefors, S., Spontaneous Environmental Scanning.
Helgesson, C-F., Making a Natural Monopoly: The Configuration of a Techno-Economic Order in Swedish Telecommunications.
Lindé, J., Essays on the Effects of Fiscal and Monetary Policy.
Ljunggren, U., En utvärdering av metoder för att mäta produktivitet och effektivitet i skolan – Med tillämpning i Stockholms stads grundskolor.
Lundbergh, S., Modelling Economic High-Frequency Time Series.
Nilsson, K., Ledtider för ledningsinformation.
Sandström, M., Evaluating the Benefits and Effectiveness of Public Policy.
Skalin, J., Modelling Macroeconomic Time Series with Smooth Transition Autoregressions.
Strauss, T., Governance and Structural Adjustment Programs: Effects on Investment, Growth and Income Distribution.
Söderström, U., Monetary Policy under Uncertainty.
Wijkström, F., Svenskt organisationsliv – Framväxten av en ideell sektor.