Smooth Transition Models: Extensions and Outlier Robust Inference

This book is no. **200** of the Tinbergen Institute Research Series. This series is established through cooperation between Thesis Publishers and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Smooth Transition Models: Extensions and Outlier Robust Inference

(Geleidelijke overgangs-modellen:
uitbreidingen en uitschieter-robuuste analyse)

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van de Rector Magnificus
Prof. dr. P.W.C. Akkermans M.A.
en volgens besluit van het College voor promoties

De openbare verdediging zal plaatsvinden op
donderdag 16 september 1999 om 13.30 uur
door
Dirk Jacobus Cornelis van Dijk
geboren te Leiden

Promotiecommissie

Promotor:	Prof. dr. P.H.B.F. Franses
Overige leden:	Prof. dr. H.K. van Dijk Prof. C.W.J. Granger Prof. dr. T. Kloek

Voor Judith, Gert-Jan en Matthijs

Acknowledgements

"We act as though comfort and luxury were the chief requirements of life, while all that we really need is something to be enthusiastic about."
Albert Einstein

When I started my PhD project four years ago, I wasn't too sure whether this was something to be enthusiastic about. Looking back I can say it certainly was. A decisive factor in bringing about this change of mind has been my supervisor Philip Hans Franses. He has been a great support and inspiration from the very first until the very last day of the past four years. His inexhaustible supply of ideas, his never-failing optimism, and his ability to know exactly when to stimulate me and when to slow me down have been (and still are!) quite incredible. Working together with someone like Philip Hans, it simply is impossible not to become enthusiastic about econometric research. I can only hope that we can continue to co-operate in the future.

I thank the members of the small committee, Herman van Dijk, Clive Granger and Teun Kloek, for their insightful comments on an earlier draft of this thesis. Unfortunately, I lacked the time to incorporate most suggestions in the final manuscript, but I am convinced that I will benefit from them in my future research.

One aspect of being a PhD which certainly aroused my enthusiasm was the opportunity to travel abroad to participate in conferences or to stay at other universities. I would like to thank NWO, NACEE and the Tinbergen Institute for their financial support, and the Departamento de Estadística y Econometría of the Universidad Carlos III de Madrid and the Graduate School of Business of the University of Chicago for their hospitality and for providing a very stimulating research environment. In particular, I would like to thank Alvaro Escribano, Ruey Tsay and Jeff Russell for their willingness to host me and to co-operate with me on research projects.

Getting to know other people by doing research together is at least as interesting as the research itself. In this respect, I would like to thank my co-authors Robert-Paul Berben, Charles Bos, Rob Eisinga, Alvaro Escribano, André Lucas, Stefan Lundbergh, Philip Rothman, Jeff Russell, Nick Taylor, Robert Taylor, Timo Teräsvirta, Ruey Tsay and Albert-Willem Veenstra. A special word of thanks to André Lucas, for introducing me to the subject of robust statistics, for showing me how to do research in an efficient, yet accurate and thorough manner, and for essential contributions to several chapters of this thesis.

I thank my colleague PhD students at the Tinbergen Institute, especially those who had the courage to share an office with me. In addition, I am grateful to the secretarial staff of the Tinbergen Institute for their support.

Of course there are things in life to be more enthusiastic about than writing a thesis - although at times I have found it necessary that other people pointed this out to me. I would like to thank my friends Arjan Baaijen, Herjan Barnhard, Ellen and Henk Binnendijk, Petra and Rian Binnendijk, Christa van Egmond, Arend Guijt, Linda and Erik Heemskerk, Hanke Kleij, Caroline Ravensbergen, Mariëlle and Peter Ribbers, Gerben Roordink, Arine and Wilfred van Rijn, Silvia and Machiel Siezenga, Ineke and André Siezenga, Cornelis and Henja Visser, Marlijn Visser, Anne Vletter, Arjan Zuidhof, Bram van Zuylen, and all others who I forgot to mention by name for reminding me of (or even arousing) my enthusiasm for hiking, strong coffee, spicy meals, and skiing.

Finally, I would like to thank my parents and Gerrit and Carolien for their neverfailing moral support. I hope that this thesis is a more appropriate answer than the usual "good" or "could be better" to their question how my research was proceeding. I dedicate this thesis to my niece Judith and nephews Gert-Jan and Matthijs. Watching them grow up makes me realize more than anything else that comfort and luxury are helpful, but certainly not the chief requirements in life.

Dick
Rotterdam, July 1999

Contents

Chapter 1

Introduction and Summary

It has long been recognized that both the dynamic behaviour of economic variables as well as many relationships between economic variables are inherently nonlinear. Both theoretical and empirical researchers have stressed the importance of such nonlinearities – an illustrative example being provided by the business cycle literature. Keynes (1936, p. 314) already noted a certain type of asymmetry in the dynamic patterns of macro-economic variables over the business cycle by observing that

> "...the substitution of a downward for an upward tendency often takes place suddenly and violently, whereas there is, as a rule, no such sharp turning point when an upward is substituted for a downward tendency."

Subsequently, the large-scale empirical investigation of Burns and Mitchell (1946) demonstrated that many (US) macro-economic variables behave differently in different phases of the business cycle. An often used example is unemployment, which tends to rise faster during recessions than decline during booms. At the same time, Kaldor (1940) and Goodwin (1951) developed theoretical nonlinear business cycle models, followed by many others. See Mullineux and Peng (1993) and Granger and Teräsvirta (1993, Sec. 3.1) for recent surveys.

Until recently, (economic) time series analysis was dominated by what one might call the 'linearity paradigm', that is, in practice one mainly considered the use of *linear* models, despite the evidence for the presence of possible nonlinearities. At least two reasons can be given for the predominant use of linear models. First, in many cases linear models (are assumed to) provide reasonable approximations to the true nonlinear relationships, the precise form of which often is unknown anyway. Second, it was practically impossible to specify and estimate complex nonlinear models due to a lack of computing power. Over the last fifteen years, say, the interest in the analysis and actual implementation of nonlinear time series models has been steadily increasing again. This might be attributed to a large extent to the improvements in computer technology, which have more or less eliminated the second reason why one would want to avoid using such models.

The problem one immediately faces when considering the use of (parametric) nonlinear time series models is the vast, if not unlimited, number of possible models, see for example the surveys by Tong (1990) and Granger and Teräsvirta (1993).

Sometimes economic theory is helpful in choosing a particular model, but more often it is not. Hence, there seems to be a need for criteria which facilitate specification of nonlinear models and which facilitate choosing between alternative, competing models in particular. An additional complication in this respect is that criteria which have been developed for specifying linear models can be of little or no use in a nonlinear context, see Granger and Teräsvirta (1999) for an illustrative example.

A natural approach to modeling economic time series with nonlinear models seems to be to define different *states* of the world or *regimes*, and to allow for the possibility that the dynamic behaviour of economic variables depends on the regime which occurs at any given point in time. By 'state-dependent dynamic behaviour' of a time series it is meant that certain properties of the time series such as its mean, variance and/or autocorrelations are different under different circumstances.

State-dependent behaviour can be modeled in various ways, as discussed in more detail in Section 1.1 below. In this thesis, I focus on the smooth transition model, which is one of the parametric time series models that recently have been developed for this purpose. The next section also discusses how this model is related to other approaches that have been pursued to capture regime-switching behaviour. Section 1.2 outlines the contents of the other chapters of this thesis and summarizes the main conclusions.

1.1 Modeling state-dependent behaviour

A (univariate) time series model that often is used in practice to describe the observations on a variable y_t at times $t = 1, \ldots, T$, is the autoregressive [AR] model

$$y_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t, \tag{1.1}$$

where $\phi_0, \ldots, \phi_p$ are fixed parameters and ε_t is a noise process, in the sense that ε_t is independent of its own past and past observations $y_{t-1}, y_{t-2}, \ldots$, and the expected value of ε_t is equal to 0. The AR model might be called linear, as it describes the value of y at time t as a (fixed) linear combination of its p most recent values plus an exogenous shock ε_t.

A straightforward generalization of the linear AR model which allows for state-dependent behaviour is given by

$$y_t = \phi_0(s_t) + \phi_1(s_t) y_{t-1} + \cdots + \phi_p(s_t) y_{t-p} + \varepsilon_t, \tag{1.2}$$

where $\phi_0(s_t), \phi_1(s_t), \ldots, \phi_p(s_t)$ now are functions of the regime or state s_t. This state-dependent model usually is attributed to Priestley (1980,1988).

Several approaches have been considered to make the general state-dependent model operational, by specifying how the state s_t or, equivalently, the parameters $\phi_0(s_t), \phi_1(s_t), \ldots, \phi_p(s_t)$ evolve over time. For example, by assuming that the $\phi_j(s_t)$, $j = 0, \ldots, p$, are stochastic processes, independent of ε_t, one obtains the class of doubly stochastic or random coefficient models, see Tjøstheim (1986) and Nicholls and Quinn (1982), respectively. Alternatively, one can assume that the regime s_t can be characterized by a vector of state-variables x_t such as $x_t = (y_{t-1}, \dots, y_{t-p}, z_{1t}, \dots, z_{kt})'$, where $z_{1t}, \dots, z_{kt}$ are (observable) exogenous variables. This implies that (1.2) can be written as

$$y_t = \phi_0(x_t) + \phi_1(x_t)y_{t-1} + \cdots + \phi_p(x_t)y_{t-p} + \varepsilon_t, \tag{1.3}$$

and $\phi_j(x_t)$, $j = 0, 1, \dots, p$ can be estimated using nonparametric estimation methods, see Chen (1998), for example.

These approaches do not really focus on the interpretation of the state-dependent model (1.2) as a means to describe regime-switching behaviour, but consider the model as a locally linear approximation to the underlying nonlinear relationship. This is especially clear in the nonparametric approach because, if (1.3) is augmented with additional terms $\phi_{p+1}z_{1,t}, \dots, \phi_{p+k}z_{k,t}$, the resulting model can be regarded as a first-order Taylor approximation to the nonlinear model

$$y_t = f(y_{t-1}, \dots, y_{t-p}, z_{1,t}, \dots, z_{k,t}) + e_t, \tag{1.4}$$

where $f(\cdot)$ is a nonlinear function. Alternatively, it might be said that these methods allow for an infinite number of regimes, as they allow the parameters in the model to be general functions of time or the vector of state-variables x_t. In economics however, it is more common to keep the number of regimes limited. For example, in the literature on business cycles, it is common practice to distinguish only two regimes, associated with expansions and contractions.

In recent years a number of parametric time series models have been proposed which formalize the idea of the existence of a limited number of different regimes, generated by a stochastic process. Conceptually, these regime-switching models can be thought of as a set of linear models, with each of the models corresponding to a particular regime. Hence, at each point in time, only one (or a linear combination) of the models in the set is active to describe the behaviour of a time series, where the activity depends on the regime at that particular moment. The available regime-switching models differ in the way the regime evolves over time. Roughly speaking, two main classes of models can be distinguished. The models in the first class assume that the regimes can be characterized (or determined) by observable variables x_t. These models thus can be written as in (1.3), but now $\phi_j(x_t)$, $j = 0, 1, \dots, p$, are parametric functions of the variables x_t. Note that in this case the regimes which have occurred in the past and present are known with certainty (although they have to be found by statistical techniques, of course). The models in the second class assume that the regime cannot actually be observed but is determined by an underlying unobservable stochastic process s_t. This implies that one can never be certain that a particular regime has occurred at a particular point in time, but can only assign probabilities to the occurrence of the different regimes. A prominent member of this second class of models is the Markov-Switching model, popularized by Hamilton (1989), which assumes that s_t follows a low-order Markov chain process with a small number of possible states.

In this thesis I focus on a particular member of the first class of regime-switching models, the smooth transition model. The basic smooth transition model allows for two distinct regimes, where the transition from one regime to the other is smooth, in the sense that the parameters $\phi_j(x_t)$, $j = 0, 1, \ldots, p$ in (1.3) are continuous functions of the variables in x_t and hence change gradually.

1.2 Summary and conclusions

This thesis can be divided into three parts. The first part, consisting of this and the next chapter, provides an introduction and an overview of previous research on the smooth transition model. This allows the topics discussed in the second and third parts to be put in perspective. The second part, consisting of Chapters 3, 4 and 5, discusses extensions of the smooth transition model to allow for multiple regimes, to allow for time-varying properties in addition to regime-switching behaviour, and to a multivariate context. The third and final part, comprising Chapters 6 and 7, focuses on outlier robust inference in smooth transition models.

The basic smooth transition model

The purpose of Chapter 2 is to provide a concise overview of different aspects of the basic smooth transition model. The chapter is based on Franses and van Dijk (1999, Chapter 3). The discussion is cast in terms of a specification procedure for smooth transition models, developed in Teräsvirta (1994) and Eitrheim and Teräsvirta (1996). A similar set up is used in several other chapters. Certain issues that are reviewed, such as forecasting and impulse response analysis, apply more generally to other nonlinear time series models as well.

Extensions of the smooth transition model

The smooth transition model, as it is most frequently applied, allows for only two regimes. Even though this might be sufficient for most practical purposes, sometimes it may be of interest to consider the possibility of more than two regimes. For example, most research on business cycle asymmetry traditionally has focused on the differences in behaviour of macro-economic variables during expansions and recessions – hence, distinguishing two regimes only. Recent evidence suggests that such a two-regime characterization of the business cycle is not fully adequate. Chapter 3 considers an extension of the smooth transition model to allow for multiple regimes. This chapter is based on van Dijk and Franses (1999).

It is shown that a multiple-regime smooth transition model can be obtained from the two-regime model in a simple yet elegant way. A specific-to-general approach for specification of models involving multiple regimes is discussed, which builds upon the specification procedure for the two-regime model. An application of the multiple-regime smooth transition model to characterize the behaviour of the growth rate of post-war US real GNP provides evidence in favor of the existence of multiple business cycle phases.

Another prominent characteristic of economic variables when observed over prolonged periods of time is structural instability. As the economic system as a whole evolves, properties of variables such as industrial output or the unemployment rate change. For example, Stock and Watson (1996) report an overwhelming amount of evidence for instability in both univariate and multivariate models for a large number of US post-war macro-economic time series. Traditionally, nonlinearity and structural instability have been investigated in isolation. In Chapter 4, a model based on the principle of smooth transition is considered that can be used to capture both phenomena simultaneously. The resultant time-varying smooth transition model arises naturally as a special case of the multiple-regime smooth transition model discussed in Chapter 3. This chapter is based on Lundbergh, Teräsvirta and van Dijk (1999).

One of the reasons for the usual separate analysis of structural instability and regime-switching probably is the fact that the two can be observationally equivalent. Regime-switching models can be parameterized such that the resultant time series closely resemble time series that are subject to occasional structural changes, and vice versa. The statistical tests that are used in the specification procedure for the time-varying smooth transition model are shown to be capable of distinguishing between structural instability and nonlinearity. The application in this chapter to the growth in industrial production for the UK demonstrates that both regime-switching characteristics and structural instability seem to be present in this series.

The smooth transition model essentially is a univariate model, which has as its main purpose to describe and forecast a single variable. Sometimes it may be worthwhile to model several time series jointly, to exploit possible linkages that exist between them. In the context of empirical macro-economics, such models might be useful to examine whether the relationship between variables displays regime-switching characteristics, where the regimes may be linked with different phases of the business cycle. In Chapter 5 the smooth transition model is generalized to a multivariate setting. This chapter is based on van Dijk and Franses (1998) and Taylor, van Dijk, Franses and Lucas (1999).

Representation and specification of a multivariate smooth transition model are considered at a quite general level. In that sense, this chapter complements Krolzig (1997) and Tsay (1998), who treat similar issues for multivariate Markov-switching and threshold models, respectively. Second, so-called smooth transition equilibrium correction models are explored in somewhat more detail. These models can describe situations where certain variables are linked by a linear (long-run) equilibrium relation, whereas adjustment toward this equilibrium is nonlinear. The application in this chapter demonstrates that smooth transition equilibrium correction models are useful to characterize the behaviour of intra-day spot and future prices of the FTSE100 index.

Outliers and robust inference in nonlinear models

The smooth transition model can be parameterized in such a way that it generates very asymmetric realizations, in the sense that the observations are unevenly distributed across the two regimes. The resultant time series thus may resemble a linear time series with some outliers. Conversely, a linear time series that is contaminated with a few aberrant observations may suggest the presence of a regime-switching structure. This observational equivalence between time series generated from a nonlinear model and time series generated from a linear model with outlier contamination suggests that these might be easily mistaken in practice, at least when standard specification procedures are used. The third part of the thesis therefore examines whether alternative methods can be developed that are capable of distinguishing between these two types of time series. Two issues are addressed in particular: testing for nonlinearity and estimation of nonlinear models.

In Chapter 6, I use outlier-robust estimation techniques for linear time series models to develop test statistics for smooth transition nonlinearity that are resistant to aberrant observations. Analytic arguments, supplemented with simulation evidence, demonstrate that the resultant test statistics are able to recognize whether a time series is generated from a highly asymmetric nonlinear model or from a linear model with outlier contamination, at least more often than the standard testing procedures. The chapter is based on van Dijk *et al.* (1999a) and Escribano, Franses and van Dijk (1998).

Chapter 7 is motivated by the consideration that outliers and nonlinearity also can occur simultaneously. The effects of outliers on parameter estimates and inference in linear models has been studied quite extensively in the literature. For nonlinear models, this is not the case. This is partly due to the fact that many of the concepts that are routinely used to assess the effects of outliers in linear models appear problematic or even impossible to apply to nonlinear models. This chapter therefore takes a more pragmatic approach, and uses Monte Carlo simulation to examine several different approaches that might be useful to obtain reasonable estimates of the parameters in a smooth transition model in the presence of outliers. It appears that existing methods only are useful when the outliers are not very large and do not occur very frequently and, therefore, alternative estimation methods need to be developed.

Chapter 2

The Smooth Transition Model

The purpose of this chapter is to introduce the smooth transition model and discuss aspects of the model that are relevant for subsequent chapters. The discussion is framed in terms of an empirical specification procedure for the smooth transition model, elements of which are discussed in Teräsvirta (1994) and Eitrheim and Teräsvirta (1996). A review of the smooth transition model similar in spirit to this chapter is given by Teräsvirta (1998).

The plan of this chapter is as follows. In Section 2.1, representation of the smooth transition model, interpretation of the model parameters, and aspects such as stability and stationarity of the model are discussed. The empirical specification procedure for smooth transition models is outlined in Section 2.1.2. The various steps in this specification procedure are examined in more detail in subsequent sections. Section 2.2 deals with testing for the presence of smooth transition effects. Estimation of the model parameters is the subject of Section 2.3. Evaluation of estimated smooth transition models by means of diagnostic tests and impulse response analysis are addressed in Sections 2.4 and 2.6, respectively. As the latter makes use of forecasts from the smooth transition model, out-of-sample forecasting is discussed first in Section 2.5.

2.1 Representation

The model that is central in this thesis is the smooth transition model for a univariate time series y_t, which is observed at times $t = 1-p, -p, \ldots, -1, 0, 1, \ldots, T-1, T$, given by

$$y_t = \phi_1' x_t (1 - G(s_t; \gamma, c)) + \phi_2' x_t G(s_t; \gamma, c) + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{2.1}$$

where x_t is a vector consisting of lagged endogenous and exogenous variables, $x_t = (1, \tilde{x}_t')'$ with $\tilde{x}_t = (y_{t-1}, \ldots, y_{t-p}, z_{1t}, \ldots, z_{kt})'$ and $\phi_i = (\phi_{i,0}, \phi_{i,1}, \ldots, \phi_{i,m})'$, $i = 1, 2$, with $m = p + k$. The ε_t's are assumed to be a martingale difference sequence with respect to the history of the time series, which is denoted as $\Omega_{t-1} = \{y_{t-1}, y_{t-2}, \ldots, y_{1-(p-1)}, y_{1-p}\}$, that is, $\mathrm{E}[\varepsilon_t | \Omega_{t-1}] = 0$. For simplicity, I also assume that the conditional variance of ε_t is constant, $\mathrm{E}[\varepsilon_t^2 | \Omega_{t-1}] = \sigma^2$. The transition function $G(s_t;\gamma,c)$ is a continuous function, which usually is bounded between 0 to 1. The transition variable s_t can be a lagged endogenous variable ($s_t = y_{t-d}$ for certain integer $d > 0$), an exogenous variable ($s_t = z_t$), or a (possibly nonlinear) function of lagged endogenous and exogenous variables ($s_t = h(\tilde{x}_t)$ for some function $h(\cdot)$). Another possibility is to take s_t equal to a (function of a) linear time trend ($s_t = t$), which gives rise to a model with smoothly changing parameters, see Lin and Teräsvirta (1994). Most of the discussion in this and following chapters is cast in terms of the smooth transition autoregressive [STAR] model, which is obtained from (2.1) by restricting the vector of regressors $\tilde{x}_t$ to contain lagged dependent variables only[1]. Written out in more detail, the STAR model thus is given by

$$\begin{aligned} y_t = (\phi_{1,0} + \phi_{1,1}y_{t-1} + \cdots + \phi_{1,p}y_{t-p})\,(1 - G(s_t;\gamma,c)) \\ + (\phi_{2,0} + \phi_{2,1}y_{t-1} + \cdots + \phi_{2,p}y_{t-p})\,G(s_t;\gamma,c) + \varepsilon_t. \end{aligned} \quad (2.2)$$

As noted in the previous chapter, the STAR model can be interpreted as a regime-switching model that allows for two regimes, associated with the extreme values of the transition function, $G(s_t;\gamma,c) = 0$ and $G(s_t;\gamma,c) = 1$, whereas the transition from one regime to the other is gradual. The regime that occurs at time t can be determined by the observable variable s_t and the associated value of $G(s_t;\gamma,c)$.

A popular choice for $G(s_t;\gamma,c)$ is the logistic function

$$G(s_t;\gamma,c) = \frac{1}{1+\exp\{-\gamma(s_t - c)\}}, \qquad \gamma > 0, \quad (2.3)$$

and the resultant model is called the logistic STAR [LSTAR] model. The parameter c in (2.3) can be interpreted as the threshold between the two regimes corresponding to $G(s_t;\gamma,c) = 0$ and $G(s_t;\gamma,c) = 1$, in the sense that the logistic function changes monotonically from 0 to 1 as s_t increases, while $G(c;\gamma,c) = .5$. The parameter γ determines the smoothness of the change in the value of the logistic function and, thus, the smoothness of the transition from one regime to the other. Figure 2.1 shows some examples of the logistic function for various different values of the smoothness parameter γ.

From this figure it is seen that as γ becomes very large, the logistic function $G(s_t;\gamma,c)$ approaches the indicator function $I[s_t > c]$, defined as $I[A] = 1$ if A is true and $I[A] = 0$ otherwise, and, consequently, the change of $G(s_t;\gamma,c)$ from 0 to 1 becomes almost instantaneous at $s_t = c$. Hence, the LSTAR model (2.2) with (2.3) nests a two-regime threshold autoregressive [TAR] model as a special case. In case $s_t = y_{t-d}$ this model is called a self-exciting TAR [SETAR] model. An extensive discussion of (SE)TAR models can be found in Tong (1990). When $\gamma \to 0$, the logistic function becomes equal to a constant (equal to 0.5) and when $\gamma = 0$, the LSTAR model reduces to a linear model.

In the LSTAR model, the two regimes are associated with small and large values of the transition variable s_t relative to the threshold c. This type of regime-switching

[1] In the STAR model as discussed by Teräsvirta (1994), the transition variable is assumed to be a lagged dependent variable as well, that is, $s_t = y_{t-d}$. I will not make this assumption here and leave s_t unspecified most of the time.

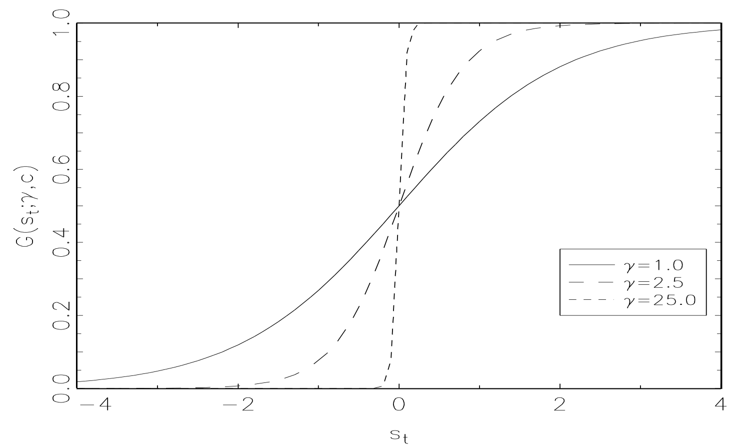


Figure 2.1: Examples of the logistic function $G(s_t; \gamma, c)$ as given in (2.3) for various values of the smoothness parameter γ and threshold $c = 0$.

can be convenient for modeling, for example, business cycle asymmetry to distinguish expansions and recessions. If y_t represents the growth rate of output, the LSTAR model with $s_t = y_{t-d}$ can be used to describe different dynamics during periods of positive and negative growth.

It is however easy to think of examples where other types of regime-switching behaviour are more appropriate. For example, it can be argued that the behaviour of the real exchange rate depends on the size of the deviation from purchasing power parity [PPP]. The presence of transaction costs, such as costs of transportation and storage of goods, leads to the notion of different regimes in real exchange rates. In particular, the profits from commodity arbitrage, which is generally thought to be the ultimate force behind maintaining PPP, do not make up for the costs involved in the necessary transactions for small deviations from the equilibrium real exchange rate. This implies the existence of a band around the equilibrium rate in which there is no tendency of the real exchange rate to revert to its equilibrium value. Outside this band, commodity arbitrage becomes profitable, which forces the real exchange rate back towards the band. See Dumas (1992) and O'Connell and Wei (1997) for analytic models that incorporate effects of transaction costs as described above.

If regime-switching of the form described in this example is to be captured by a STAR model, with y_t the real exchange rate and $s_t = y_{t-d}$, it appears more appropriate to specify the transition function such that the regimes are associated with small and large absolute values of s_t. This can be achieved by using, for example, the exponential function

$$G(s_t; \gamma, c) = 1 - \exp\{-\gamma(s_t - c)^2\}, \qquad \gamma > 0. \tag{2.4}$$

The resultant exponential STAR [ESTAR] model has been applied to real exchange rates by Michael, Nobay and Peel (1997) and Baum, Caglayan and Barkoulas (1998).

A drawback of the exponential function (2.4) is that for either $\gamma \to 0$ or $\gamma \to \infty$, the function collapses to a constant, as can be seen in Figure 2.2. Hence, the model becomes linear in both cases and the ESTAR model does not nest a SETAR model

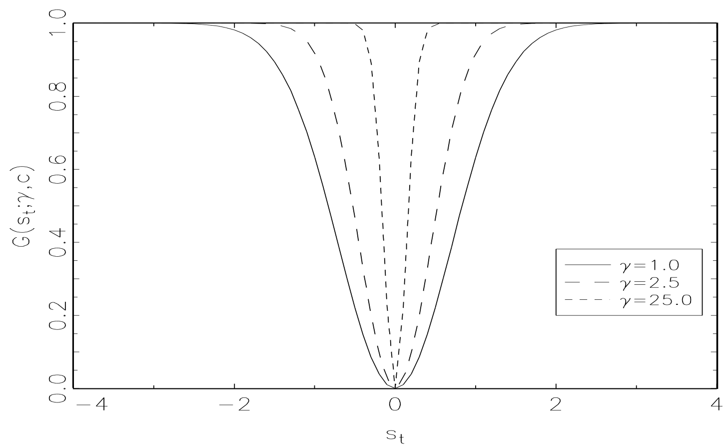


Figure 2.2: Examples of the exponential function $G(s_t;\gamma,c)$ as given in (2.4) for various values of the smoothness parameter γ and threshold $c=0$.

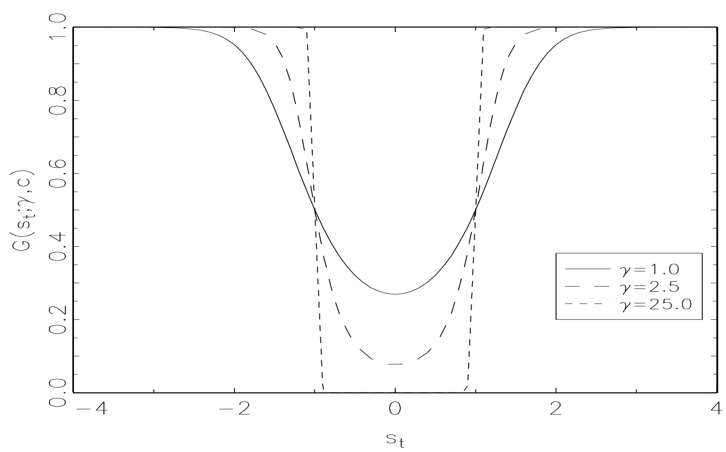


Figure 2.3: Examples of the quadratic logistic function $G(s_t;\gamma,c)$ as given in (2.5) for various values of the smoothness parameter γ and $c_1=-1$ and $c_2=1$.

as a special case. This can be remedied by using the quadratic logistic function

$$G(s_t;\gamma,c)=\frac{1}{1+\exp\{-\gamma(s_t-c_1)(s_t-c_2)\}},\qquad c_1\le c_2,\gamma>0, \tag{2.5}$$

where now $c=(c_1,c_2)'$, as proposed by Jansen and Teräsvirta (1996). In this case, if $\gamma\to 0$, the model becomes linear, whereas if $\gamma\to\infty$, the function $G(s_t;\gamma,c)$ is equal to 1 for $s_t<c_1$ and $s_t>c_2$ and equal to 0 in between. Hence, the STAR model with this particular transition function nests a three-regime (SE)TAR model. Note that for moderate values of γ, the minimum value taken by the function (2.5), which is attained for $s_t=(c_1+c_2)/2$, is not equal to zero; see Figure 2.3. This has to be kept in mind when interpreting estimates from models with this particular transition function.

2.1.1 Properties of the STAR model

In this subsection I examine several properties of the STAR model. The discussion will be rather informal and intuitive - for a more formal treatment, the interested reader is referred to Tong (1990, Chapters 2 and 4). Throughout I concentrate on the LSTAR model (2.2) and (2.3) with $p = 1$ and $s_t = y_{t-1}$, that is,

$$y_t = (\phi_{1,0} + \phi_{1,1}y_{t-1})(1 - G(y_{t-1}; \gamma, c)) + (\phi_{2,0} + \phi_{2,1}y_{t-1})G(y_{t-1}; \gamma, c) + \varepsilon_t. \quad (2.6)$$

The first thing to notice is that quite a large variety of dynamic patterns can be generated by this simple model by choosing the parameters appropriately. To give some impression of the possibilities, Figure 2.4 shows realizations of $T = 200$ observations from (2.6) with $\phi_{1,1} = -0.5$, $\phi_{2,1} = 0.5$, and the parameters in the logistic function (2.3) set equal to $\gamma = 2.5$ and $c = 0.5$. The shocks ε_t, $t = 1, \ldots, T$, are drawn independently from a standard normal distribution, which will be denoted as $\varepsilon_t \sim$ i.i.d. $N(0, 1)$ throughout. All series are started with $y_0 = 0$, while the same values for the shocks $\varepsilon_t, t = 1, \ldots, T$, are used to generate subsequent observations. The intercepts $\phi_{1,0}$ and $\phi_{2,0}$ are varied to generate different behaviour. The scale on the vertical axis in the different panels of Figure 2.4 is taken to be the same to facilitate comparison of the different time series.

Figures 2.5 and 2.6 contain scatter diagrams of y_t versus y_{t-1} for some of the series shown in Figure 2.4, and some additional series generated from (2.6) with slightly different parameter configurations. In Figures 2.5 and 2.6, the deterministic part of the model

$$F(y_{t-1}; \theta) \equiv (\phi_{1,0} + \phi_{1,1}y_{t-1})(1 - G(y_{t-1}; \gamma, c)) + (\phi_{2,0} + \phi_{2,1}y_{t-1})G(y_{t-1}; \gamma, c), \quad (2.7)$$

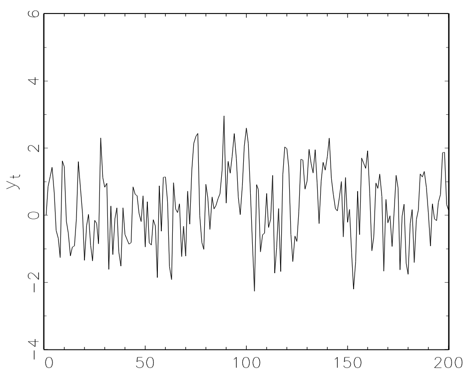
with θ the vector containing all parameters in the model, that is, $\theta = (\phi_{1,0}, \phi_{1,1}, \phi_{2,0}, \phi_{2,1}, \gamma, c)'$, also is shown. Notice that $F(y_{t-1}; \theta)$ is the conditional expectation of y_t at time $t - 1$. This deterministic and predictable part of the model commonly is referred to as the *skeleton* of the model, a concept introduced by Chan and Tong (1985). Much can be learned about the properties of time series generated from nonlinear models by analyzing the properties of the skeleton or the associated difference equation

$$y_t = F(y_{t-1}; \theta). \quad (2.8)$$

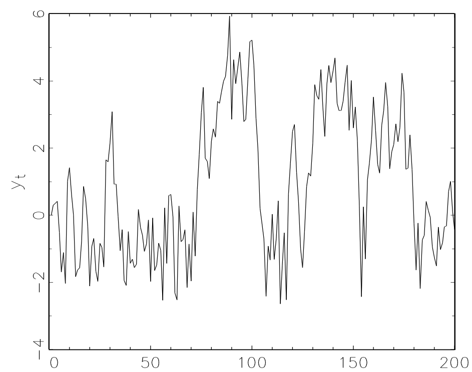
A useful way of interpreting this equation is that it represents the STAR model (2.6) with the noise ε_t turned off, meaning to say that ε_t is set equal to zero.

The first difference equation (2.8) is said to have an *equilibrium* at y^* if $y^* = F(y^*; \theta)$, that is, if y^* is a *fixed point* of the skeleton. If the skeleton only depends on the first-order lag y_{t-1}, as in the example considered here, an easy method to determine whether equilibria exist is to look for intersection points of the skeleton with the 45°-line in the scatter of y_t versus y_{t-1}. An equilibrium is called *locally stable* if the sequence $y_0, y_1, y_2, \ldots$, generated from (2.8) converges to y^* for values of y_0 close to y^*. An equilibrium is *globally stable* if the series $y_0, y_1, y_2, \ldots,$ converges to y^* for all initial values y_0. For example, the linear difference equation

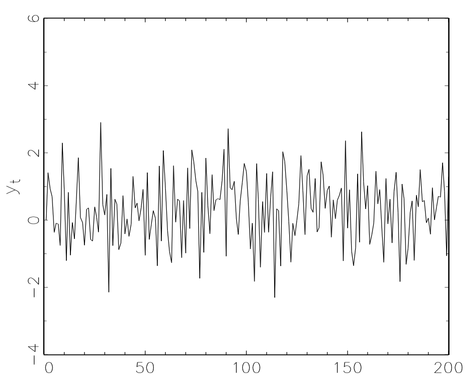
$$y_t = \phi_0 + \phi_1 y_{t-1}, \quad (2.9)$$



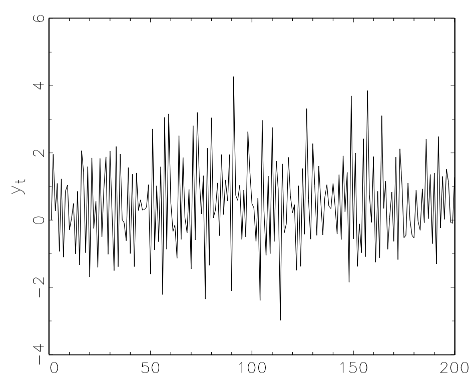
(a) $\phi_{1,0} = -0.5$, $\phi_{2,0} = 0.5$



(b) $\phi_{1,0} = -1.5$, $\phi_{2,0} = 1.5$



(c) $\phi_{1,0} = 0.5$, $\phi_{2,0} = -0.5$



(d) $\phi_{1,0} = 1.5$, $\phi_{2,0} = -1.5$

Figure 2.4: Example time series generated from the LSTAR model (2.6) with (2.3), with $\phi_{1,1} = -0.5$, $\phi_{2,1} = 0.5$, $\gamma = 2.5$, $c = 0.5$, and $\varepsilon_t \sim$ i.i.d. $N(0, 1)$.

has a unique and globally stable equilibrium if and only if $|\phi_1| < 1$. If this condition is satisfied, the equilibrium is given by $y^* = \phi_0/(1-\phi_1)$. Notice that this is equal to the mean of the associated time series y_t, generated from (2.9) augmented with a noise process ε_t.

Nonlinear difference equations can have a single (stable or unstable) equilibrium, multiple equilibria, or no equilibrium at all. Furthermore, even if the equilibrium is unique and stable, in general it is not equal to the mean of the corresponding time series. A necessary and sufficient condition for an equilibrium of (2.8) to be (locally) stable is

$$\left|\frac{\partial F(y^*;\theta)}{\partial y}\right| < 1.$$

In that case, for points y_t close to y^*

$$\begin{aligned} y_{t+1} - y^* &= F(y_t;\theta) - F(y^*;\theta) \\ &\approx \frac{\partial F(y^*;\theta)}{\partial y}(y_t - y^*), \end{aligned}$$

and hence

$$|y_{t+1} - y^*| < |y_t - y^*| \qquad \Leftrightarrow \qquad \left|\frac{\partial F(y^*;\theta)}{\partial y}\right| < 1.$$

In words, y_{t+1} will be closer to y^* than y_t if $F(y;\theta)$ is a contraction in a neighborhood of $y = y^*$. For the skeleton of the LSTAR model given in (2.7) it is straightforward to derive that

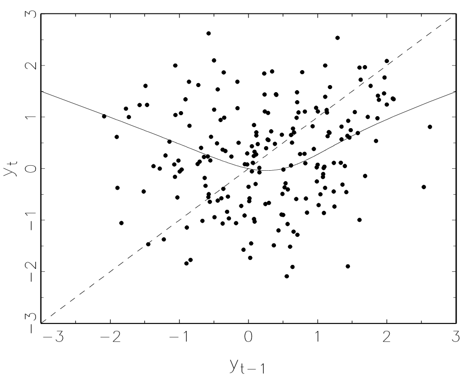
$$\begin{aligned} \frac{\partial F(y^*;\theta)}{\partial y} &= \gamma[(\phi_{2,0} - \phi_{1,0}) + (\phi_{2,1} - \phi_{1,1})y^*]G(y^*;\gamma,c)[1 - G(y^*;\gamma,c)] \\ &\quad + \phi_{1,1}[1 - G(y^*;\gamma,c)] + \phi_{2,1}G(y^*;\gamma,c). \quad (2.10) \end{aligned}$$

A stable equilibrium is also called an *attractor*, which stems from the fact that in the absence of shocks the time series is *attracted* by the stable equilibrium. Given that a nonlinear time series can have multiple stable equilibria, it follows that it can also have several attractors. That is, y^* is the attractor for $\bar{y}$ if $y_t = \bar{y}$ and

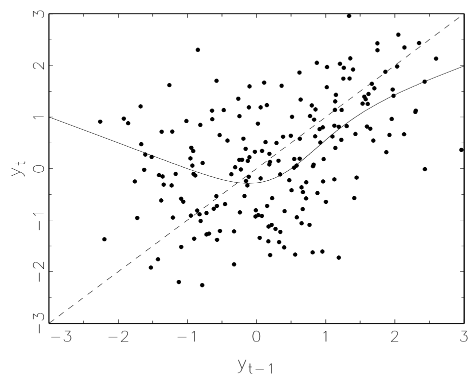
$$y_{t+n} \to y^* \quad \text{as } n \to \infty \qquad \text{if } \varepsilon_{t+j} = 0 \text{ for all } j > 0.$$

A different way to express this is to say that $\bar{y}$ is in the *domain of attraction* of y^*. As will be seen below, a stable equilibrium is not the only possible form of attractor of a nonlinear time series.

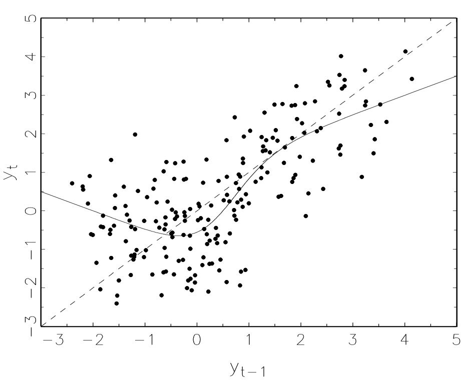
Panel (a) of Figure 2.5 shows the scatterplot of the series generated from (2.6) with both constants $\phi_{1,0}$ and $\phi_{2,0}$ set equal to 0, which implies that the means of the AR(1) models in the two regimes are equal to 0. In this case, the equilibrium is unique and stable and also equal to 0. However, the mean of the time series y_t is *not* equal to 0. This can be understood by noting that because $\phi_{1,1}$ is negative, the series has a tendency to leave the lower regime $y_{t-1} < 0.5$ very quickly. In fact, in the absence of a shock ε_t the series reverts to the upper regime immediately for large negative values of y_{t-1}, as $\mathrm{E}[y_t|y_{t-1}] = \phi_{1,1}y_{t-1} > 0.5$ if $y_{t-1} < -1$. Because



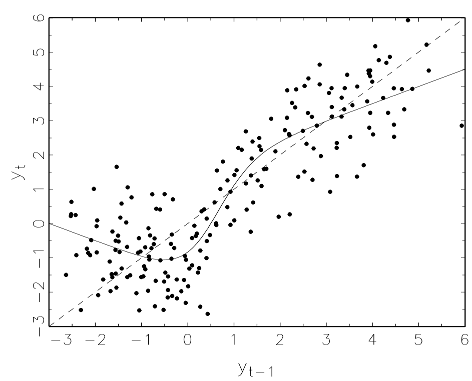
(a) $\phi_{1,0} = 0.0$, $\phi_{2,0} = 0.0$



(b) $\phi_{1,0} = -0.5$, $\phi_{2,0} = 0.5$



(c) $\phi_{1,0} = -1.0$, $\phi_{2,0} = 1.0$



(d) $\phi_{1,0} = -1.5$, $\phi_{2,0} = 1.5$

Figure 2.5: Scatterplots of example time series generated from the LSTAR model (2.6) with (2.3), with $\phi_{1,1} = -0.5$, $\phi_{2,1} = 0.5$, $\gamma = 2.5$, $c = 0.5$ and $\varepsilon_t \sim$ i.i.d. $N(0,1)$. The solid line is the skeleton of the model, the dashed line is the 45°-line.

$\phi_{2,1}$ is positive, the series is expected to remain in the upper regime for large positive values of y_{t-1} (although it will be pulled towards the threshold $c = 0.5$ as $\phi_{2,0} = 0$ and $\phi_{2,1} < 1$). This suggests that the time series will be positive on average - hence, the mean of y_t will be larger than 0.

The scatters in panels (b) through (d) of Figure 2.5 demonstrate that by increasing the absolute value of the constants in the two regimes, a model with multiple equilibria is obtained. The model shown in panel (d) has three points of intersection with the 45°-line, located at $y_1^* = -0.966$, $y_2^* = 0.899$ and $y_3^* = 2.975$. The equilibria y_1^* and y_3^* are stable whereas y_2^* is unstable, as can be readily seen from the graph. Alternatively, evaluating (2.10) with $y^* = y_2^*$ shows that $|\partial F(y_2^*)/\partial y)| > 1$. The point y_2^* also is the boundary between the domains of attraction of y_1^* and y_3^*. For values of $y_t < y_2^*$, the sequence $y_{t+1}, y_{t+2}, \ldots$ obtained from (2.8) converges to y_1^*, whereas for $y_t > y_2^*$ the resultant series $y_{t+1}, y_{t+2}, \ldots$ is attracted by y_3^*. From panel (d) of Figure 2.4, which shows an example of how series generated from this model evolve over time, it is seen y_t has a tendency to stay close to one of the two stable equilibria. For example, if y_t is in the domain of attraction of y_3^* (y_1^*), only a large negative (positive) shock causes a transition of the series to the neighborhood of y_1^* (y_3^*).

The skeleton of the LSTAR model as given in (2.7) is a continuous function of y_{t-1} and, using the stability/stationarity conditions discussed below, it can be shown that the model always has at least one equilibrium y^*. This equilibrium however need not be stable. This is illustrated by Figure 2.6, which contains scatters for series generated by the LSTAR model (2.6) with $\phi_{1,0} = 1.5$, $\phi_{1,1} = -0.5$, $\phi_{2,0} = -1.5$, $\phi_{2,1} = 0.5$ and $c = 0.5$ and various values of γ. For $\gamma = 1$, the skeleton of the model contains an equilibrium at $y^* = 0.205$, which is stable as $\partial F(y^*)/\partial y = -0.757$. As γ increases, the equilibrium moves closer towards c, such that effectively the value of the derivative of the skeleton with respect to y evaluated at y^*, see (2.10), decreases. For $\gamma > 1.35$, $\partial F(y^*)/\partial y < -1$, and the equilibrium is unstable.

An alternative way to understand that the equilibrium becomes unstable for larger values of γ, recall that in case $\gamma \to \infty$, the logistic function $G(y_{t-1}; \gamma, c)$ approaches the indicator function $I[y_{t-1} > c]$ and the STAR model reduces to a SETAR model. The skeleton of this SETAR model is not continuous as shown in panel (d) of Figure 2.6, but contains a jump at c, as

$$\lim_{y_{t-1} \uparrow c} F(y_{t-1}; \theta) = 1.25 \qquad \text{and} \qquad \lim_{y_{t-1} \downarrow c} F(y_{t-1}; \theta) = -1.25.$$

Furthermore, the skeleton has no fixed points and, therefore, the SETAR model has no equilibrium in this case. Also note that the means of the AR(1) models in the two regimes both are in the other regime. Intuitively, this suggests that the series has no point at which it could come to 'rest'. If it is in the upper regime it is pulled towards the lower regime and vice versa. Still, the model *does* have an attractor. In fact, the model contains what is called a *limit cycle*. A k-period limit cycle is defined as a set of points $y_1^*, \ldots, y_k^*$, such that $y_j^* = F(y_{j-1}^*)$ for $j = 2, \ldots, k$, and $y_1^* = F(y_k^*)$. That is, if the time series started in one of the points y_j^*, $j = 1, \ldots, k$, and no shocks occurred, the series would 'cycle' among the k-points $y_1^*, \ldots, y_k^*$. In

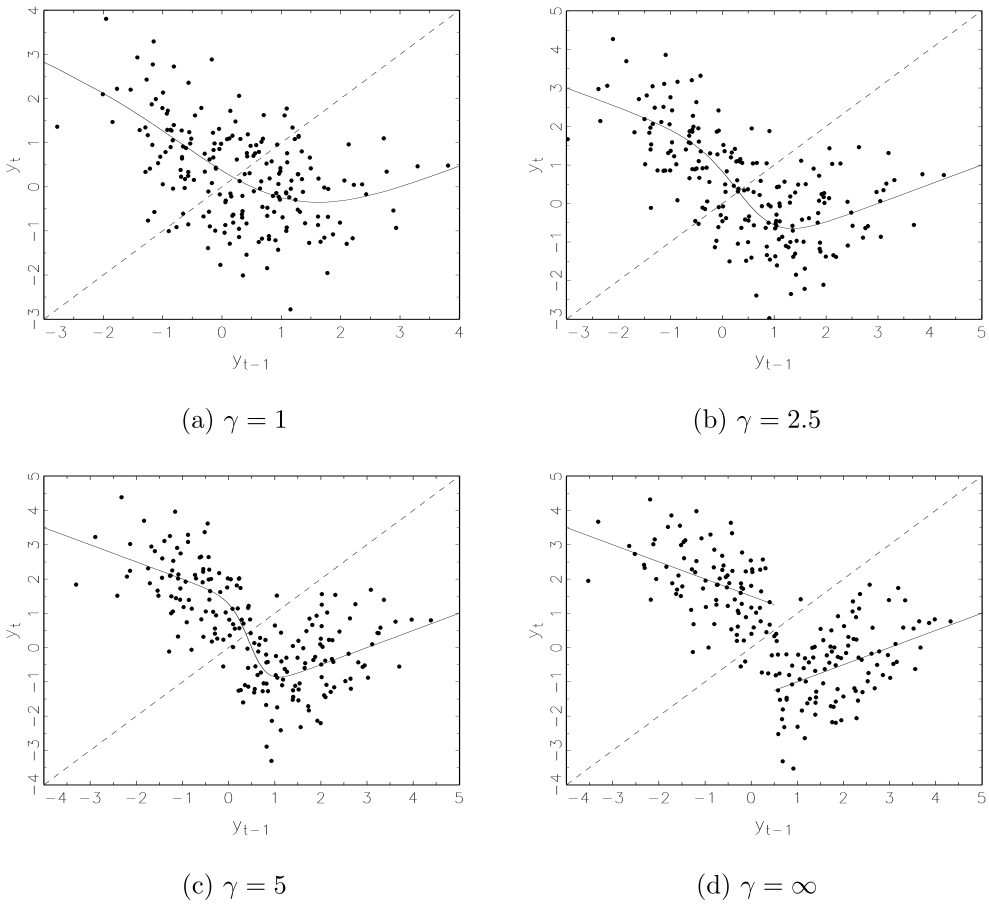


Figure 2.6: Scatter diagrams of example time series generated from the LSTAR model (2.6) with (2.3), with $\phi_{1,0} = 1.5$, $\phi_{1,1} = -0.5$, $\phi_{2,0} = -1.5$, $\phi_{2,1} = 0.5$, $c = 0.5$ and $\varepsilon_t \sim$ i.i.d. $N(0, 1)$. The solid lines are the skeletons of the model, the dashed line is the 45°-line.

the SETAR model, the limit cycle consists of two points, $y_1^* = -0.60$ and $y_2^* = 1.80$. It can also be shown that the limit cycle is a global attractor, in the sense that the series $y_{t+1}, y_{t+2}, \ldots$, would converge to the cycle if the noise were turned off after time t, regardless of the value of y_t. The STAR model with $\gamma > 1.35$ also contains a limit cycle of length 2, although the exact values of y_1^* and y_2^* depend on the value of γ. For $\gamma > 10$, say, the logistic function is virtually indistinguishable from the indicator function $I[y_{t-1} > c]$, and the limit cycle in the STAR model is the same as the limit cycle in the SETAR model.

This last example demonstrates that nonlinear models can contain *endogenous dynamics*, which means to say that even in the absence of shocks y_t fluctuates. This is in contrast with linear time series, for which the fluctuations are caused entirely by the exogenous shocks ε_t. The debate whether observed dynamics in time series are endogenous or exogenous has a long history, also in the business cycle literature; see Mullineux and Peng (1993). A general discussion on nonlinear time series models, endogenous dynamics and the related concept of chaos is given in Tong (1995). See also Chappell and Peel (1998), who demonstrate that the ESTAR model can have multiple equilibria and can exhibit chaotic dynamics.

As a final remark, notice that the models in the example above only differ in the values taken by the intercepts in the two regimes, $\phi_{1,0}$ and $\phi_{2,0}$, whereas the autoregressive parameters $\phi_{1,1}$ and $\phi_{2,1}$ are kept the same. The fact that the models nevertheless generate series with quite different behaviour illustrates the important role which is played by constants in nonlinear time series models.

For models with higher order AR dynamics in the regimes, it can be quite difficult to establish the existence of equilibria, attractors and/or limit cycles analytically. A pragmatic way to investigate the properties of the skeleton of a higher-order model is to use what might be called 'deterministic simulation'. That is, given starting values $y_0, \ldots, y_{1-p}$, one computes the values taken by $y_1, y_2, \ldots$, while setting all $\varepsilon_t, t = 1, 2, \ldots$, equal to zero. Doing this for many different starting values gives some impression about the characteristics of the (skeleton of the) model, see Teräsvirta and Anderson (1992) and Peel and Speight (1996) for applications of this procedure.

Stationarity

Little is known about the conditions under which STAR models generate time series that are stationary. Such conditions have only been established for the first-order SETAR model, which is obtained from (2.6) and (2.3) by allowing $\gamma \to \infty$. As shown by Chan and Tong (1985), a sufficient condition for stationarity is $\max(|\phi_{1,1}|, |\phi_{2,1}|) < 1$, which is equivalent to the requirement that the AR(1) models in the two regimes are stationary. Chan, Petrucelli, Tong and Woolford (1985) show that stationarity of the first order model actually holds under less restrictive conditions. In particular, the SETAR model is stationary if and only if one of the following conditions is satisfied:

(i) $\phi_{1,1} < 1, \phi_{2,1} < 1, \phi_{1,1}\phi_{2,1} < 1$;

(ii) $\phi_{1,1} = 1, \phi_{2,1} < 1, \phi_{1,0} > 0$;

(iii) $\phi_{1,1} < 1, \phi_{2,1} = 1, \phi_{2,0} < 0$;

(iv) $\phi_{1,1} = 1, \phi_{2,1} = 1, \phi_{2,0} < 0 < \phi_{1,0}$;

(v) $\phi_{1,1}\phi_{2,1} = 1, \phi_{1,1} < 0, \phi_{2,0} + \phi_{2,1}\phi_{1,0} > 0$.

Condition (i) corresponds with the sufficient condition of Chan and Tong (1985), although it should be noted that (i) allows one of the AR parameters to become smaller than -1. Conditions (ii)-(iv) show that the AR model in one or even both regimes may contain a unit root. In such cases, the time series is 'locally nonstationary'. The conditions on the intercepts $\phi_{1,0}$ and $\phi_{2,0}$ are such that the time series has a tendency to revert to the stationary regime and, hence, the time series is 'globally' stationary. Testing for unit roots in SETAR models is discussed in Caner and Hansen (1998), Enders and Granger (1998) and Berben and van Dijk (1999).

A rough-and-ready check for stationarity of nonlinear time series models in general is to determine whether or not the skeleton is stable, using deterministic simulation as described previously. Intuitively, if the skeleton is such that the series tends to explode for certain starting values, the series is nonstationary.

Even less is known about the stationary distributions of time series generated from S(E)TAR models. Anděl (1989) discusses some analytic results for a special case of the first-order SETAR model in which $\phi_{1,0} = \phi_{2,0} = c = 0$, $\phi_{1,1} = -\phi_{2,1}$ and $\phi_{1,1} \in (0, 1)$. In general, one has to resort to numerical procedures to evaluate the stationary distribution of y_t. Some of the methods which can be applied are discussed in Moeanaddin and Tong (1990) and Tong (1990, Sec. 4.2).

2.1.2 Empirical specification procedure

A commonly applied specification strategy in empirical research is to start with a simple or restricted model and to proceed to more complicated or general ones only if diagnostic tests indicate that the maintained model is inadequate. Granger (1993) strongly recommends to employ such a specific-to-general procedure when considering the use of nonlinear time series models to describe the features of a particular variable. The empirical specification procedure for STAR models put forward by Teräsvirta (1994) follows this approach and consists of the following steps.

1. Specify an appropriate linear AR model of order p [AR(p)] for the time series under investigation;
2. Test the null hypothesis of linearity against the alternative of STAR-type nonlinearity. If linearity is rejected, select the appropriate transition variable s_t and the form of the transition function $G(s_t; \gamma, c)$;
3. Estimate the parameters in the selected STAR model;
4. Evaluate the model using diagnostic tests;
5. Modify the model if necessary;
6. Use the model for descriptive or forecasting purposes.

Steps 2-6 in this specification procedure are discussed in detail in the following sections. To conclude this section, some remarks are made on the specification of a linear model in the first step.

The main element involved in specifying an AR(p) for y_t, that is,

$$y_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t, \tag{2.11}$$

is the choice of the lag order p. This lag order should be such that the corresponding residuals are approximately white noise, as the tests for nonlinearity that are used in the next step are sensitive to residual autocorrelation. The order of the AR model can be selected by conventional methods, such as the Akaike Information Criterion [AIC]

$$\text{AIC} = T \ln \hat{\sigma}^2 + 2k, \tag{2.12}$$

or the Schwarz Information Criterion [BIC]

$$\text{BIC} = T \ln \hat{\sigma}^2 + k \ln T, \tag{2.13}$$

where $\hat{\sigma}^2 = \sum_{t=1}^{T} \hat{\varepsilon}_t^2$ with $\hat{\varepsilon}_t$ the residuals from the estimated AR(p) model, and $k = p + 1$ the number of parameters in the model. The value of $p \in \{0, 1, \ldots, \bar{p}\}$ that minimizes the AIC or BIC is selected as the appropriate lag order, where $\bar{p}$ is a pre-specified maximum order. Alternatively, one can test for residual autocorrelation directly by using the Ljung-Box statistic

$$\text{LB}(m) = T(T+2) \sum_{k=1}^{m} (T-k)^{-1} r_k^2(\hat{\varepsilon}), \tag{2.14}$$

where $r_k(\hat{\varepsilon})$ is the k-th autocorrelation of the residuals, given by

$$r_k(\hat{\varepsilon}) = \frac{\sum_{t=k+1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_{t-k}}{\sum_{t=1}^{T} \hat{\varepsilon}_t^2}.$$

Under the null hypothesis of no residual autocorrelation at lags 1 to m, the LB test has an asymptotic χ^2 distribution with $(m - p)$ degrees of freedom.

The purpose of the different criteria is to select a model that adequately captures the linear (autocorrelation) properties of a time series. When these criteria are used in a specification procedure for a nonlinear model, it should be kept in mind that nonlinear time series can have very misleading linear properties. For example, the time series generated by the LSTAR model (2.6) with (2.3) shown in panel (b) of Figure 2.4 is seen to resemble a (linear) time series with occasional level shifts. Analytic expressions for the autocorrelations of y_t generated from this model are not available. Estimates obtained from simulations suggest that they are quite substantial, even at long lags, at least for small to moderate sample sizes. In fact, the autocorrelations decline very slowly, and closely resemble the hyperbolic decay characteristic of long-memory processes. Hence, when an AR(p) model is considered for such series, the selected lag order may easily become very large. See Granger and Teräsvirta (1999) for a similar example.

To illustrate the caveat noted above, I consider the following simulation experiment. Time series are generated from the LSTAR model given by (2.6) with (2.3), with $\phi_{1,1} = -0.5$, $\phi_{2,1} = 0.5$, $\gamma = 2.5$, $c = 0.5$, and $\varepsilon_t \sim$ i.i.d. $N(0, 1)$. The sample size is taken to be $T = 150$ or 300 observations. An AR(p) model is specified for those series, where p is set equal to the lag length minimizing AIC or BIC with maximum order $\bar{p} = 8$, or to the minimum lag length for which the LB statistic with $m = 15$ is not statistically significant at the 5% level (results for other choices of m are similar). Table 2.1 shows the number of replications, out of a total of 1000, for which different values of p are selected as the appropriate AR order. Clearly, the

Table 2.1: AR order selection

		p								p							
$\phi_{1,0}$	$\phi_{2,0}$	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
		AIC, $T = 150$								AIC, $T = 300$							
-0.50	0.50	669	151	68	33	30	22	12	15	560	236	69	37	32	26	22	18
-1.00	1.00	428	353	85	49	34	22	11	18	174	519	128	52	39	34	30	24
-1.50	1.50	312	467	156	56	50	24	18	22	74	497	221	89	49	35	22	30
0.50	-0.50	564	149	57	45	31	20	17	12	655	156	55	34	31	16	20	14
1.00	-1.00	615	191	65	47	31	20	18	13	574	236	74	40	20	22	17	17
1.50	-1.50	609	175	82	47	29	24	17	17	530	216	109	54	33	25	14	19
		BIC, $T = 300$								BIC, $T = 300$							
-0.50	0.50	947	48	4	1	0	0	0	0	922	71	6	1	0	0	0	0
-1.00	1.00	772	206	20	2	0	0	0	0	590	392	18	0	0	0	0	0
-1.50	1.50	512	416	60	8	3	1	0	0	228	680	87	2	3	0	0	0
0.50	-0.50	958	40	1	1	0	0	0	0	970	27	3	0	0	0	0	0
1.00	-1.00	938	56	4	2	0	0	0	0	918	79	2	1	0	0	0	0
1.50	-1.50	944	50	4	2	0	0	0	0	921	72	6	1	0	0	0	0
		LB(m), $m = 15$, $T = 150$								LB(m), $m = 15$, $T = 300$							
-0.50	0.50	553	50	20	22	16	16	17	306	589	95	24	12	19	15	13	233
-1.00	1.00	458	131	33	24	31	12	17	294	392	262	36	13	21	19	18	239
-1.50	1.50	323	211	55	27	19	13	22	330	196	406	75	28	24	16	19	236
0.50	-0.50	544	53	23	21	21	14	14	310	676	38	13	18	11	11	15	218
1.00	-1.00	540	61	23	29	15	13	11	308	643	63	29	10	10	18	9	218
1.50	-1.50	533	52	26	35	15	22	11	306	630	52	32	17	18	20	15	216

Frequencies of lag length selection in AR(p) models estimated on series generated from the STAR model (2.6), with $\phi_{1,1} = -0.5$, $\phi_{2,1} = 0.5$, $\gamma = 2.5$, $c = 0.5$ and $\varepsilon_t \sim$ i.i.d. $N(0, 1)$.

order of the linear model that is required to capture the autocorrelation properties of the time series can be much larger than the lag order in the STAR model from which the series were generated, especially when AIC or the LB statistic are used.

In itself, the fact that the order of the approximating AR model can be much larger than the true nonlinear model is not really problematic. The bottom-line of the above is that, if linearity is rejected by the tests in the second step of the specification procedure, ideally the lag order of the STAR model should be specified from scratch before proceeding with estimation of the model. This, however, is not commonly done in practice - usually one simply retains the lag order that was found in the first step.

2.2 Testing for smooth transition nonlinearity

Having specified an AR(p) model for y_t, the next step in the specification procedure consists of testing linearity against the alternative of STAR-type nonlinearity.

To facilitate the notation in this section, rewrite the STAR model given in (2.2) in the format of (2.1), that is,

$$y_t = \phi_1' x_t(1 - G(s_t; \gamma, c)) + \phi_2' x_t G(s_t; \gamma, c) + \varepsilon_t, \tag{2.15}$$

where $x_t = (1, \tilde{x}_t')'$, $\tilde{x}_t = (y_{t-1}, \ldots, y_{t-p})'$. The null hypothesis of linearity can be expressed as equality of the autoregressive parameters in the two regimes, that is, $H_0 : \phi_1 = \phi_2$, which is to be tested against the alternative hypothesis $H_1 : \phi_{1,j} \neq \phi_{2,j}$ for at least one $j \in \{0, \ldots, p\}$.

The testing problem is complicated by the presence of unidentified nuisance parameters under the null hypothesis. Informally, the STAR model contains parameters which are not restricted by the null hypothesis, but which nevertheless are no longer present in the model when the null hypothesis holds true. For example, the null hypothesis as given above does not restrict the parameters in the transition function, γ and c, but when $\phi_1 = \phi_2$, $G(s_t; \gamma, c)$ and, hence, γ and c drop out of the model.

An alternative way to illustrate the presence of unidentified nuisance parameters in this case is to note that the null hypothesis of linearity can be expressed in several different ways. Besides equality of the AR parameters in the two regimes, $H_0 : \phi_1 = \phi_2$, the alternative null hypothesis $H_0' : \gamma = 0$ also gives rise to a linear model. For example, if $\gamma = 0$ the logistic function (2.3) is equal to 0.5 for all values of s_t, and the STAR model (2.15) reduces to an AR model with parameters $(\phi_1 + \phi_2)/2$. In case H_0' is used, the threshold c and the parameters ϕ_1 and ϕ_2 are the unidentified parameters. Under H_0', ϕ_1 and ϕ_2 can take any value as long as their average remains the same.

The problem of unidentified nuisance parameters under the null hypothesis was first considered in some depth by Davies (1977, 1987) and occurs in many testing problems, see Hansen (1996) for a recent account. The main consequence of the presence of such nuisance parameters is that the conventional statistical theory cannot be applied to obtain the (asymptotic) distribution of test statistics. Instead, the test statistics tend to have non-standard distributions for which an analytic expression often is not available. This implies that critical values have to be determined by means of simulation methods.

The problem of testing linearity against STAR-type alternatives was addressed in Luukkonen *et al.* (1988). Their proposed solution amounts to replacing the transition function $G(s_t; \gamma, c)$ by a suitable Taylor approximation. In the reparametrized model, the identification problem is no longer present, and linearity can be tested by means of a Lagrange Multiplier [LM] statistic which has a standard asymptotic χ^2-distribution under the null hypothesis. The main advantages of this approach are that first, using the LM principle avoids estimating the model under the alternative hypothesis and second, the ability to use the conventional asymptotic distribution avoids the use of simulation methods to assess the significance of test statistics.

When testing linearity against the alternative of a STAR model, based on an AR(p) model under the null hypothesis, it is useful to distinguish three situations depending on the nature of the transition variable s_t:

1. s_t is a lagged endogenous variable y_{t-d}, with $1 \leq d \leq p$;

2. s_t is a lagged endogenous variable y_{t-d} with $d > p$, or an exogenous variable z_t;

3. s_t is a linear combination of $y_{t-1}, \ldots, y_{t-p}$, that is $\alpha' \tilde{x}_t$, with α unknown.

The first two cases test linearity against STAR with a specific transition variable, which turns out to be useful at a later stage in this part of the specification procedure, as discussed below. The test statistics differ slightly because in the first case, s_t is contained as a regressor in the model under the null hypothesis, whereas in the second case it is not. Even though the test statistics that result from the third case can also be considered as tests against STAR with a specific transition variable, it is more common to interpret them as general tests against STAR, leaving s_t unspecified (except that it is assumed that s_t is a linear combination of $y_{t-1}, \ldots, y_{t-p}$) or even as general nonlinearity tests, without explicit reference to the STAR alternative. Below I first present an explicit derivation of the test statistics that are used in the first case, where $s_t = y_{t-d}$, for certain $1 \leq d \leq p$, followed by some remarks on the differences that arise in the second and third cases.

Tests against LSTAR

Consider again the LSTAR model as given in (2.15) with $s_t = y_{t-d}$ for certain $1 \leq d \leq p$, and rewrite this as

$$y_t = \phi_1' x_t + (\phi_2 - \phi_1)' x_t G(y_{t-d}; \gamma, c) + \varepsilon_t. \tag{2.16}$$

Luukkonen *et al.* (1988) suggest to approximate the function $G(y_{t-d}; \gamma, c)$ with a first order Taylor approximation around $\gamma = 0$, that is,

$$\begin{aligned} T_1(y_{t-d}; \gamma, c) &= G(y_{t-d}; 0, c) + \gamma \left. \frac{\partial G(y_{t-d}; \gamma, c)}{\partial \gamma} \right|_{\gamma=0} + R_1(y_{t-d}; \gamma, c) \\ &= \frac{1}{2} + \frac{1}{4}\gamma(y_{t-d} - c) + R_1(y_{t-d}; \gamma, c), \end{aligned} \tag{2.17}$$

where $R_1(y_{t-d}; \gamma, c)$ is a remainder term. Substituting $T_1(\cdot)$ for $G(\cdot)$ in (2.16) and rearranging terms yields the auxiliary model

$$y_t = \beta_{0,0} + \beta_0' \tilde{x}_t + \beta_1' \tilde{x}_t y_{t-d} + e_t, \tag{2.18}$$

where $e_t = \varepsilon_t + (\phi_2 - \phi_1)' x_t R_1(y_{t-d}; \gamma, c)$. Notice that under the null hypothesis, $R_1(y_{t-d}; \gamma, c) = 0$ and $e_t = \varepsilon_t$. Consequently, the remainder term of the Taylor approximation does not affect the properties of the residuals under the null hypothesis and hence the distribution theory for the test statistics. The relationships between the parameters $\beta_i = (\beta_{i,1}, \ldots, \beta_{i,p})'$, $i = 0, 1$, in the auxiliary regression model (2.18) and the parameters in the STAR model (2.16) are given by

$$\beta_{0,0} = \frac{1}{2}(\phi_{1,0} + \phi_{2,0}) - \frac{1}{4}\gamma c(\phi_{2,0} - \phi_{1,0}), \tag{2.19}$$

$$\beta_{0,d} = \frac{1}{2}(\phi_{1,d} + \phi_{2,d}) - \frac{1}{4}\gamma(c(\phi_{2,d} - \phi_{1,d}) - (\phi_{2,0} - \phi_{1,0})), \tag{2.20}$$

$$\beta_{0,j} = \frac{1}{2}(\phi_{1,j} + \phi_{2,j}) - \frac{1}{4}\gamma c(\phi_{2,j} - \phi_{1,j}), \qquad j = 1, \ldots, p,\ j \neq d, \tag{2.21}$$

$$\beta_{1,j} = \frac{1}{4}\gamma c(\phi_{2,j} - \phi_{1,j}), \qquad j = 1, \ldots, p. \tag{2.22}$$

The above equations demonstrate that the restrictions $\phi_1 = \phi_2$ or $\gamma = 0$ imply $\beta_{1,j} = 0$ for $j = 1, \ldots, p$. Hence testing the null hypothesis $H_0 : \phi_1 = \phi_2$ or $H_0' : \gamma = 0$ in (2.16) is equivalent to testing the null hypothesis $H_0'' : \beta_1 = 0$ in (2.18). This null hypothesis can be tested by a standard variable addition test in a straightforward manner. The test statistic, to be denoted as LM_1, has an asymptotic χ^2 distribution with p degrees of freedom under the null hypothesis of linearity. See Tsay (1986a) and Saikkonen and Luukkonen (1988) for discussion of the regularity conditions which have to be satisfied. As the LM_1 statistic does not test the original null hypothesis $H_0' : \gamma = 0$ but rather the auxiliary null hypothesis $H_0'' : \beta_1 = 0$, this test is usually referred to as an LM-*type* statistic[2].

As noted by Luukkonen *et al.* (1988), the above test statistic does not have power in situations where only the intercept is different across regimes, that is, when $\phi_{1,0} \neq \phi_{2,0}$ but $\phi_{1,j} = \phi_{2,j}$ for $j = 1, \ldots, p$. This is seen immediately from (2.22) which shows that $\beta_{1,j} = 0,\ j = 1, \ldots, p$, under these assumptions concerning the alternative. Luukkonen *et al.* (1988) suggest to remedy his deficiency by replacing the transition function $G(y_{t-d}; \gamma, c)$ by a third-order Taylor approximation instead, that is,

$$\begin{aligned} T_3(y_{t-d}; \gamma, c) &= G(y_{t-d}; 0, c) + \gamma \left. \frac{\partial G(y_{t-d}; \gamma, c)}{\partial \gamma} \right|_{\gamma=0} \\ &\quad + \frac{1}{6}\gamma^3 \left. \frac{\partial^3 G(y_{t-d}; \gamma, c)}{\partial \gamma^3} \right|_{\gamma=0} + R_3(y_{t-d}; \gamma, c) \\ &= \frac{1}{2} + \frac{1}{4}\gamma(y_{t-d} - c) + \frac{1}{48}\gamma^3(y_{t-d} - c)^3 + R_3(y_{t-d}; \gamma, c), \end{aligned} \tag{2.23}$$

where use has been made of the fact the second derivative of $G(y_{t-d}; \gamma, c)$ with respect to γ evaluated at $\gamma = 0$ equals zero. Replacing the transition function $G(\cdot)$ with this approximation yields the auxiliary model

$$y_t = \beta_{0,0} + \beta_0'\tilde{x}_t + \beta_1'\tilde{x}_t y_{t-d} + \beta_2'\tilde{x}_t y_{t-d}^2 + \beta_3'\tilde{x}_t y_{t-d}^3 + e_t, \tag{2.24}$$

where $e_t = \varepsilon_t + (\phi_2 - \phi_1)' x_t R_3(y_{t-d}; \gamma, c)$, and $\beta_{0,0}$ and the $\beta_i,\ i = 1, 2, 3$, again are functions of the parameters ϕ_1, ϕ_2, γ and c. Inspection of the exact relationships

[2] The test statistic can also be developed from first principles as a genuine LM statistic, see Granger and Teräsvirta (1993, pp.71-72). It can be shown that the statistic is in fact the supremum of the pointwise statistics for fixed $\phi_2 - \phi_1$ and c and, hence, is similar in spirit to the test statistic that is commonly applied to test for TAR-type nonlinearity, see Hansen (1997).

demonstrates that the null hypothesis $H_0' : \gamma = 0$ now corresponds to $H_0'' : \beta_1 = \beta_2 = \beta_3 = 0$, which again can be tested by a standard LM-type test. Under the null hypothesis of linearity, the test statistic, to be denoted as LM_3, has an asymptotic χ^2 distribution with $3p$ degrees of freedom.

The expressions of $\beta_i,\ i = 1, 2, 3$, in terms of ϕ_1, ϕ_2, γ and c also reveal that the only parameters that depend on the constants $\phi_{1,0}$ and $\phi_{2,0}$ are $\beta_{2,d}$ and $\beta_{3,d}$. Hence, a parsimonious, or 'economy', version of the LM_3 statistic can be obtained by augmenting the auxiliary model (2.18) with regressors y_{t-d}^3 and y_{t-d}^4, that is,

$$y_t = \beta_{0,0} + \beta_0' \tilde{x}_t + \beta_1' \tilde{x}_t y_{t-d} + \beta_{2,d} y_{t-d}^3 + \beta_{3,d} y_{t-d}^4 + e_t, \tag{2.25}$$

and testing the null hypothesis $H_0'' : \beta_1 = 0$ and $\beta_{2,d} = \beta_{3,d} = 0$. The resultant test statistic, denoted LM_3^e, has an asymptotic χ^2 distribution with $p+2$ degrees of freedom. The usual motivation for considering the LM_3^e statistic is that it requires considerably less degrees of freedom than the LM_3 statistic. In the present case, where s_t is a single variable, this argument may not be all that relevant. However, in case s_t is an unspecified linear combination of $y_{t-1}, \ldots, y_{t-p}$, saving degrees of freedom can be important, as discussed below.

Tests against ESTAR

Granger and Teräsvirta (1993) suggest that linearity might be tested against an ESTAR alternative, given by (2.16) with (2.4), by replacing the exponential transition function with a first-order Taylor approximation around $\gamma = 0$

$$\begin{aligned} T_1(y_{t-d}; \gamma, c) &= G(y_{t-d}; 0, c) + \gamma \left. \frac{\partial G(y_{t-d}; \gamma, c)}{\partial \gamma} \right|_{\gamma=0} + R_1(y_{t-d}; \gamma, c) \\ &= \gamma (y_{t-d} - c)^2 + R_1(y_{t-d}; \gamma, c), \end{aligned} \tag{2.26}$$

which gives rise to the auxiliary model

$$y_t = \beta_{0,0} + \beta_0' \tilde{x}_t + \beta_1' \tilde{x}_t y_{t-d} + \beta_2' \tilde{x}_t y_{t-d}^2 + e_t, \tag{2.27}$$

where $e_t = \varepsilon_t + (\phi_2 - \phi_1)' x_t R_1(y_{t-d}; \gamma, c)$. The expressions for $\beta_{0,0}$ and $\beta_i,\ i = 0, 1, 2$, show that the restriction $\gamma = 0$ corresponds with $\beta_1 = \beta_2 = 0$ in (2.27). The LM_2 statistic which tests this null hypothesis has an asymptotic χ^2 distribution with $2p$ degrees of freedom.

Escribano and Jordá (1999) argue that a first-order approximation for the exponential function is not sufficient to capture its distinguishing characteristics, in particular the two inflexion points of this function, see also Figure 2.2. Hence, they conclude that a second-order Taylor approximation is necessary, that is,

$$\begin{aligned} T_2(y_{t-d}; \gamma, c) &= G(y_{t-d}; 0, c) + \gamma \left. \frac{\partial G(y_{t-d}; \gamma, c)}{\partial \gamma} \right|_{\gamma=0} \\ &\quad + \frac{1}{2} \gamma^2 \left. \frac{\partial^2 G(y_{t-d}; \gamma, c)}{\partial \gamma^2} \right|_{\gamma=0} + R_2(y_{t-d}; \gamma, c) \\ &= \gamma (y_{t-d} - c)^2 - \frac{1}{2} \gamma^2 (y_{t-d} - c)^4 + R_2(y_{t-d}; \gamma, c), \end{aligned} \tag{2.28}$$

yielding the auxiliary regression,

$$y_t = \beta_{0,0} + \beta_0'\tilde{x}_t + \beta_1'\tilde{x}_t y_{t-d} + \beta_2'\tilde{x}_t y_{t-d}^2 + \beta_3'\tilde{x}_t y_{t-d}^3 + \beta_4'\tilde{x}_t y_{t-d}^4 + e_t. \tag{2.29}$$

The null hypothesis to be tested now is $H_0' : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$. The resulting LM-type test statistic, denoted LM_4, has an asymptotic χ^2 distribution with $4p$ degrees of freedom under the null hypothesis.

When s_t is a lagged endogenous variable y_{t-d} with $d > p$ or an exogenous variable z_t, the resultant test statistics are very similar to the ones derived above. The only difference is that additional regressors s_t^i, $i = 1, 2, \ldots$, enter the auxiliary model. For example, the auxiliary model (2.24) based on the third-order Taylor approximation of the logistic function now becomes

$$y_t = \beta_{0,0} + \beta_0'\tilde{x}_t + \beta_{1,0}s_t + \beta_1'\tilde{x}_t s_t + \beta_{2,0}s_t^2 + \beta_2'\tilde{x}_t s_t^2 + \beta_{3,0}s_t^3 + \beta_3'\tilde{x}_t s_t^3 + e_t.$$

In case linearity is tested against an alternative with $s_t = \alpha'\tilde{x}_t$, the number of auxiliary regressors in the reparameterized model increases very rapidly when the parameter vector α, which defines the linear combination of $y_{t-1}, \ldots, y_{t-p}$ that is used as transition variable, is left completely unspecified. For example, Teräsvirta, Lin and Granger (1993) derive a test statistic against the alternative of an LSTAR model in which only the constants in the two regimes are different, that is, $\phi_{1,0} \neq \phi_{2,0}$ but $\phi_{1,j} = \phi_{2,j}$ for $j = 1, \ldots, p$, with $s_t = \alpha'\tilde{x}_t$ for general α. The reparameterized model that is obtained from replacing the logistic function $G(\alpha'\tilde{x}_t; \gamma, c)$ by a third-order Taylor approximation is given by

$$y_t = \beta_{0,0} + \beta_0'\tilde{x}_t + \sum_{i=1}^{p}\sum_{j=i}^{p}\xi_{ij}y_{t-i}y_{t-j} + \sum_{i=1}^{p}\sum_{j=i}^{p}\sum_{k=j}^{p}\psi_{ijk}y_{t-i}y_{t-j}y_{t-k} + e_t, \tag{2.30}$$

where β_0, ξ_{ij} and ψ_{ijk} are such that the null hypothesis $H_0' : \gamma = 0$, corresponds with $H_0'' : \xi_{ij} = 0$, $\psi_{ijk} = 0$, $i = 1, \ldots, p$, $j = i, \ldots, p$ and $k = j, \ldots, p$. The number of restrictions in this alternative null hypothesis or, equivalently, the number of degrees of freedom required by the statistic that is used to test this hypothesis, is equal to $p(p+1)/2 + p(p+1)(p+2)/6$. Obviously, this becomes large very quickly when p increases. To be able to compute the test in practice, p needs to be set fairly small or the length of the time series has to be sufficiently large.

The number of auxiliary regressors can be limited by imposing structure on the parameter vector α. For example, Luukkonen *et al.* (1988) assume that $\alpha = (0, \ldots, 0, 1, 0, \ldots, 0)'$, where the 1 appears in the d-th position, but d is left unspecified. Hence, the resultant test statistics can be interpreted as testing linearity against the alternative of a STAR model with $s_t = y_{t-d}$ for unknown d. Under this assumption about α it follows that $(\alpha\tilde{x}_t)^i = \alpha\tilde{x}_t^i$, which reduces the number of auxiliary regressors in the reparametrized model considerably. For example, the test statistic against the LSTAR alternative based on a third-order Taylor approximation now is based on the auxiliary model

$$y_t = \beta_{0,0} + \beta_0'\tilde{x}_t + \sum_{i=1}^{p}\sum_{j=i}^{p}\xi_{ij}y_{t-i}y_{t-j} + \sum_{i=1}^{p}\sum_{j=1}^{p}\psi_{ij}y_{t-i}y_{t-j}^2 + \sum_{i=1}^{p}\sum_{j=1}^{p}\zeta_{ij}y_{t-i}y_{t-j}^3 + e_t, \tag{2.31}$$

Table 2.2: Auxiliary regressors in LM-type tests for STAR nonlinearity

Test	y_{t-i}^2	$y_{t-i}y_{t-j}$ $i<j$	y_{t-i}^3	$y_{t-i}y_{t-j}^2$ $i\neq j$	y_{t-i}^4	$y_{t-i}y_{t-j}^3$ $i\neq j$	y_{t-i}^5	$y_{t-i}y_{t-j}^4$ $i\neq j$	df
LM_1	×	×							$p(p+1)/2$
LM_2	×	×	×	×					$p(p+1)/2+p^2$
LM_3	×	×	×	×	×	×			$p(p+1)/2+2p^2$
LM_3^e	×	×	×						$p(p+1)/2+p$
LM_4	×	×	×	×	×	×	×	×	$p(p+1)/2+3p^2$

Auxiliary regressors used by LM-type statistics to test linearity against the alternative of a STAR model with $s_t = y_{t-d}$ with d left unspecified. df = degrees of freedom in asymptotic χ^2 distribution.

and requires only $p(p+1)/2 + 2p^2$ degrees of freedom. Table 2.2 summarizes the auxiliary regressors that are used by the respective test statistics discussed before when applied in the present case.

The LM_1 statistic with $s_t = \alpha'\tilde{x}_t$ is identical to the linearity test of Tsay (1986b), which explains the remark made above that the LM-type tests with this choice of transition variable can be interpreted as general nonlinearity tests, without explicit reference to the alternative of a STAR model.

In small samples, the usual recommendation is to use F-versions of the LM test statistics because these have better size and power properties than the χ^2 variants. As an example, the F-version of the LM_3 statistic based on (2.24) can be computed as follows:

1. Estimate the model under the null hypothesis of linearity by regressing y_t on x_t. Compute the residuals $\hat{\varepsilon}_t$ and the sum of squared residuals $SSR_0 = \sum_{t=1}^{T} \hat{\varepsilon}_t^2$.

2. Estimate the auxiliary regression of $\hat{\varepsilon}_t$ on x_t and $\tilde{x}_t y_{t-1}^i$, $i = 1, 2, 3$, and compute the sum of squared residuals from this regression SSR_1.

3. The LM_3 statistic can now be computed as

$$LM_3 = \frac{(SSR_0 - SSR_1)/3p}{SSR_1/(T - 4p - 1)}, \qquad (2.32)$$

which under the null hypothesis is approximately F distributed with $3p$ and $T - 4p - 1$ degrees of freedom.

Selecting the transition variable

Two additional objectives of this stage of the specification procedure are to select the appropriate transition variable in the STAR model and the most suitable form of the transition function. As noted by Teräsvirta (1994), even though the LM_3 statistic was developed as a test against the LSTAR alternative, it should have power against ESTAR alternatives as well. An intuitive way to understand this is to compare the auxiliary models (2.27) and (2.24) which are used for computing the the LM_2 and LM_3 statistics, respectively. It is seen that all auxiliary regressors in (2.27) are contained in (2.24). For that reason, the LM_3 test might be expected to have power against ESTAR alternatives as well. The appropriate transition variable in the STAR model then can be determined first, without specifying the form of the transition function, by computing the LM_3 statistic for several candidate transition variables $s_{1t}, \ldots, s_{mt}$, say, and selecting the one for which the p-value of the test is smallest. The rationale behind this procedure is that the test should have maximum power in case the alternative model is correctly specified, that is, if the correct transition variable is used[3]. Simulation results in Teräsvirta (1994) suggest that this approach works quite well, at least in a univariate setting.

Selecting the transition function

When linearity is rejected in favor of STAR-type nonlinearity and the appropriate transition variable has been selected, the final decision to be made concerns the form of the transition function $G(s_t; \gamma, c)$. Roughly speaking, the choice is between the logistic function (2.3) on the one hand and the exponential function (2.4) or the quadratic logistic function (2.5) on the other (or similar functions which have the same properties). Teräsvirta (1994) suggests to use a decision rule based upon a sequence of tests nested within the null hypothesis corresponding to LM_3. In particular, he proposes to test the hypotheses

$$\begin{aligned} H_{03}: &\ \beta_3 = 0, \\ H_{02}: &\ \beta_2 = 0 |\ \beta_3 = 0, \\ H_{01}: &\ \beta_1 = 0 |\ \beta_3 = \beta_2 = 0, \end{aligned}$$

in (2.24) by means of LM-type tests. Closer inspection of the expressions of the auxiliary parameters β_1, β_2 and β_3 in terms of parameters of the original STAR model reveals that (i) β_3 is nonzero only if the model is an LSTAR model[4], (ii) β_2 is zero if the model is an LSTAR model with $\phi_{1,0} = \phi_{2,0}$ and $c = 0$ but is always nonzero if the model is an ESTAR model, and (iii) that β_1 is zero if the model is an

[3] It might be argued that it is not appropriate to choose the transition variable by comparing p-values as suggested above, because the models with different choices for s_t are non-nested. An alternative way to interpret and motivate this decision rule is the following. If the choice of the transition variable is made endogenous, one could estimate (L)STAR models for various choices of s_t and select the model which minimizes the residual variance (assuming that the lag order p of the AR models in the two regimes is fixed). An obvious drawback of this procedure is the necessary estimation of nonlinear (L)STAR models, which may be time-consuming. However, if the auxiliary regression model that is used in calculating the LM_3 statistic is considered to approximate the (L)STAR model to a certain degree of accuracy, estimation of nonlinear models can be avoided by selecting s_t as the choice which minimizes the residual variance of this auxiliary model. Since the LM-type test is a monotonic transformation of the residual variance, this is equivalent to selecting s_t as the variable which maximizes the LM-type statistic. This is exactly what the minimum p-value rule employed here does, see also Caner and Hansen (1998).

[4] Note that this is only true under the assumption that a first-order Taylor expansion is sufficient for the exponential function.

ESTAR model with $\phi_{1,0} = \phi_{2,0}$ and $c = 0$ but is always nonzero if the model is an LSTAR model. Combining these three properties of the auxiliary parameters leads to the following decision rule: if the p-value corresponding to H_{02} is the smallest, an ESTAR model should be selected, while in all other cases an LSTAR model is to be the preferred choice.

Escribano and Jordá (1999) propose an alternative procedure which makes use of LM_4 as a test for general STAR-type nonlinearity. Their decision rule to choose between the LSTAR and ESTAR alternatives is based on the observation that, assuming $\phi_{1,0} = \phi_{2,0}$ and $c = 0$ in (2.16), the properties of β_1 and β_2 given above also apply to β_3 and β_4 in (2.29), respectively. Therefore they suggest to test the hypotheses

$$H_{0E}: \ \beta_2 = \beta_4 = 0,$$
$$H_{0L}: \ \beta_1 = \beta_3 = 0,$$

in (2.29) and to select an LSTAR (ESTAR) model if the minimum p-value is obtained for H_{0L} (H_{0E}).

Escribano and Jordá (1999) present extensive simulation evidence on the relative performance of the linearity tests LM_3 and LM_4 and the two decision rules given above. Their main findings can be summarized as follows. In case the true data generating process [DGP] is an LSTAR model, the power of the LM_3 test in general is higher than the power of the LM_4 test, while the reverse holds if the DGP is an ESTAR model. This makes sense intuitively, as the p additional auxiliary regressors $\beta_4' \tilde{x}_t y_{t-d}^4$ in (2.29) are redundant in case of an LSTAR model, and the use of p extra degrees of freedom by the LM_4 statistic causes a loss in power. In case of an ESTAR model, these extra terms contain vital information which more than compensates the use of additional degrees of freedom. Concerning the two decision rules to choose between LSTAR and ESTAR models, in general the performance of the Escribano-Jordá procedure appears superior to the procedure of Teräsvirta (1994).

Heteroskedasticity and tests for STAR nonlinearity

Neglected heteroskedasticity has similar effects on tests for nonlinearity as residual autocorrelation, in that it may lead to spurious rejection of the null hypothesis. Davidson and MacKinnon (1985) and Wooldridge (1990, 1991) have developed specification tests which can be used in the presence of heteroskedasticity, without the need to specify the form the heteroskedasticity (which often is unknown) explicitly. Their procedures may be readily applied to robustify the tests against STAR-type nonlinearity, see also Granger and Teräsvirta (1993, pp.69-70).

For example, a heteroskedasticity-consistent [HCC] variant of the LM_3 statistic based upon (2.24) can be computed as follows

(i) Regress y_t on x_t and obtain the residuals $\hat{\varepsilon}_t$;

(ii) Regress the auxiliary regressors $\tilde{x}_t y_{t-d}^i$, $i = 1, 2, 3$, on x_t and compute the residuals $\hat{r}_t$;

(iii) Weight the residuals $\hat{r}_t$ from the regression in (ii) with the residuals $\hat{\varepsilon}_t$ obtained in (i) and regress 1 on $\hat{\varepsilon}_t \hat{r}_t$. The explained sum of squares from this regression is the LM-type statistic.

2.3 Estimation

When the transition variable s_t and the transition function $G(s_t; \gamma, c)$ have been selected, one can proceed to the next stage of the specification procedure, that is, estimating the parameters in the STAR model. This is the subject of this section.

Estimation of the parameters in the STAR model (2.2) is a relatively straightforward application of nonlinear least squares [NLS], that is, the parameters $\theta = (\phi_1', \phi_2', \gamma, c)'$ can be estimated as

$$\hat{\theta} = \underset{\theta}{\text{argmin}} \; Q_T(\theta) = \underset{\theta}{\text{argmin}} \; \sum_{t=1}^{T} (y_t - F(x_t; \theta))^2 , \tag{2.33}$$

where $F(x_t; \theta)$ is the skeleton of the model

$$F(x_t; \theta) = \phi_1' x_t (1 - G(s_t; \gamma, c)) + \phi_2' x_t G(s_t; \gamma, c). \tag{2.34}$$

Under the additional assumption that the errors ε_t are normally distributed, NLS is equivalent to maximum likelihood. Otherwise, the NLS estimates can be interpreted as quasi maximum likelihood estimates. Under certain regularity conditions, which are discussed in White and Domowitz (1984), Gallant (1987) and Pötscher and Prucha (1997), among others, the NLS estimates are consistent and asymptotically normal, that is,

$$\sqrt{T}(\hat{\theta} - \theta_0) \rightarrow N(0, C), \tag{2.35}$$

where θ_0 denotes the true parameter values. The asymptotic covariance-matrix C of $\hat{\theta}$ can be estimated consistently as $\widehat{A}_T^{-1}\widehat{B}_T\widehat{A}_T^{-1}$, where $\widehat{A}_T$ is the Hessian evaluated at $\hat{\theta}$

$$\widehat{A}_T = -\frac{1}{T}\sum_{t=1}^{T}\nabla^2 q_t(\hat{\theta}) = \frac{1}{T}\sum_{t=1}^{T}\left(\nabla F(x_t;\hat{\theta})\nabla F(x_t;\hat{\theta})' - \nabla^2 F(x_t;\hat{\theta})\hat{\varepsilon}_t\right), \tag{2.36}$$

with $q_t(\hat{\theta}) = (y_t - F(x_t;\hat{\theta}))^2$, $\nabla F(x_t;\hat{\theta}) = \partial F(x_t;\hat{\theta})/\partial\theta$, and $\widehat{B}_T$ is the the outer product of the gradient

$$\widehat{B}_T = \frac{1}{T}\sum_{t=1}^{T}\nabla q_t(\hat{\theta})\nabla q_t(\hat{\theta})' = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^2\nabla F(x_t;\hat{\theta})\nabla F(x_t;\hat{\theta})'. \tag{2.37}$$

The estimation can be performed using any conventional nonlinear optimization procedure, see Quandt (1983), Hamilton (1994, Sec. 5.7) and Hendry (1995, Appendix A5) for surveys. Issues that deserve particular attention are the choice of starting values for the optimization algorithm, concentrating the sum of squares function and the estimate of the smoothness parameter γ in the transition function.

Starting values

Obviously, the burden put on the optimization algorithm can be alleviated by using good starting values. For fixed values of the parameters in the transition function, γ and c, the STAR model is linear in the autoregressive parameters ϕ_1 and ϕ_2. Thus, conditional upon γ and c, estimates of $\phi = (\phi_1', \phi_2')'$ can be obtained by ordinary least squares [OLS] as

$$\hat{\phi}(\gamma, c) = \left(\sum_{t=1}^{T} x_t(\gamma, c)x_t(\gamma, c)'\right)^{-1}\left(\sum_{t=1}^{T} x_t(\gamma, c)y_t\right), \tag{2.38}$$

where $x_t(\gamma, c) = (x_t'(1 - G(s_t;\gamma, c)), x_t'G(s_t;\gamma, c))'$ and the notation $\phi(\gamma, c)$ is used to indicate that the estimate of ϕ is conditional upon γ and c. The corresponding residuals can be computed as $\hat{\varepsilon}_t = y_t - \hat{\phi}(\gamma, c)'x_t(\gamma, c)$ with associated variance $\hat{\sigma}^2(\gamma, c) = T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t^2(\gamma, c)$. A convenient method to obtain sensible starting values for the nonlinear optimization algorithm then is to perform a two-dimensional grid search over γ and c and select those parameter estimates which render the smallest estimate for the residual variance $\hat{\sigma}^2(\gamma, c)$.

Concentrating the sum of squares function

As suggested by Leybourne, Newbold and Vougas (1998), another way to simplify the estimation problem is to concentrate the sum of squares function. Due to the fact that the STAR model is linear in the autoregressive parameters for given values of γ and c, the sum of squares function $Q_T(\theta)$ can be concentrated with respect to

ϕ_1 and ϕ_2 as

$$Q_T(\gamma, c) = \sum_{t=1}^{T} (y_t - \phi(\gamma, c)' x_t(\gamma, c))^2. \tag{2.39}$$

This reduces the dimensionality of the NLS estimation problem considerably, as the sum of squares function as given in (2.39) needs to be minimized with respect to the two parameters γ and c only. This can be especially important in extensions of the two regime model that are to be discussed in the second part of this thesis.

The estimate of γ

It turns out to be notoriously difficult to obtain a precise estimate of the smoothness parameter γ. One reason for this is that for large values of γ, the shape of the logistic function (2.3) changes only little. Hence, to obtain an accurate estimate of γ one needs many observations in the immediate neighborhood of the threshold c. As this is typically not the case, the estimate of γ is rather imprecise in general and often appears to be insignificant when judged by its t-statistic. This estimation problem is discussed in a more general context in Bates and Watts (1988, p.87). The main point to be taken is that insignificance of the estimate of γ should not be interpreted as evidence against the presence of STAR-type nonlinearity. This should be assessed by means of different diagnostics, some of which are discussed in the next section.

To gauge the finite sample properties of the NLS estimates, the following simulation experiment is performed. Time series are generated from the LSTAR model given by (2.6) with (2.3), with $\phi_{1,1} = -0.5$, $\phi_{2,1} = 0.5$, $\gamma = 2.5$, $c = 0.5$ and $\varepsilon_t \sim$ i.i.d. $N(0, 1)$. The sample size is taken to be $T = 150$ or 300 observations. The parameters in the STAR model, with the lag orders set at their true values and the correct transition function and transition variable, are estimated by NLS as described above. Table 2.3 shows the medians and median absolute deviations [MAD] of the estimates based on 1000 replications. The MAD of a variable z_t, $t = 1, \ldots, T$, is computed as

$$\text{MAD}(z_t) = \text{median}(|z_t - \text{median}(z_t)|), \tag{2.40}$$

and can be interpreted as a measure of scale of z_t that is robust to outliers. The reason for reporting these robust measures of the location and scale of the finite sample distribution of $\hat{\theta}$ is that for some replications non-sensical estimates are obtained, due to the fact that $\hat{\gamma}$ approaches zero, or $\hat{c}$ drifts outside the range of the transition variable. The results in Table 2.3 suggest that when $T = 300$ fairly precise estimates can be obtained except for γ, where still substantial deviations from the true value can be observed.

Table 2.3: Median and MAD (in parentheses) of NLS estimates of parameters in STAR model

$\phi_{1,0}$	$\phi_{2,0}$	$\hat{\phi}_{1,0}$	$\hat{\phi}_{1,1}$	$\hat{\phi}_{2,0}$	$\hat{\phi}_{2,1}$	$\hat{\gamma}$	$\hat{c}$
				$T = 150$			
−0.50	0.50	−0.33	−0.39	0.28	0.52	7.82	0.49
		(0.32)	(0.26)	(0.61)	(0.29)	(7.01)	(0.60)
−1.00	1.00	−0.94	−0.48	1.14	0.42	3.03	0.56
		(0.35)	(0.23)	(0.74)	(0.25)	(1.49)	(0.23)
−1.50	1.50	−1.49	−0.52	1.58	0.47	2.60	0.51
		(0.39)	(0.24)	(0.42)	(0.12)	(0.85)	(0.17)
0.50	−0.50	0.48	−0.44	−0.22	0.36	32.90	0.38
		(0.23)	(0.25)	(0.57)	(0.30)	(32.87)	(0.77)
1.00	−1.00	0.83	−0.56	−0.73	0.39	7.02	0.45
		(0.32)	(0.29)	(0.60)	(0.26)	(6.09)	(0.54)
1.50	−1.50	1.44	−0.51	−1.45	0.49	3.38	0.46
		(0.54)	(0.34)	(0.48)	(0.19)	(1.84)	(0.25)
				$T = 300$			
−0.50	0.50	−0.41	−0.44	0.41	0.51	3.55	0.58
		(0.25)	(0.19)	(0.66)	(0.27)	(2.14)	(0.34)
−1.00	1.00	−1.00	−0.51	1.13	0.45	2.58	0.54
		(0.24)	(0.16)	(0.49)	(0.16)	(0.75)	(0.15)
−1.50	1.50	−1.51	−0.52	1.56	0.48	2.54	0.51
		(0.27)	(0.16)	(0.30)	(0.08)	(0.57)	(0.11)
0.50	−0.50	0.44	−0.48	−0.36	0.42	5.10	0.39
		(0.21)	(0.21)	(0.47)	(0.23)	(4.36)	(0.77)
1.00	−1.00	0.90	−0.55	−0.91	0.47	3.53	0.43
		(0.34)	(0.26)	(0.47)	(0.21)	(2.23)	(0.32)
1.50	−1.50	1.48	−0.51	−1.52	0.51	2.70	0.48
		(0.36)	(0.20)	(0.38)	(0.14)	(0.87)	(0.16)

Median and median absolute deviation [MAD] of NLS estimates of the parameters in the LSTAR model given by (2.6) with (2.3), with $\phi_{1,1} = -0.5$, $\phi_{2,1} = 0.5$, $\gamma = 2.5$, $c = 0.5$ and $\varepsilon_t \sim$ i.i.d. $N(0, 1)$. The Table is based on 1000 replications.

2.4 Diagnostic checking

After estimating the parameters in a STAR model, the next step in the specification procedure is to evaluate the properties of the fitted model. Next to 'common sense' diagnostics, such as examining the properties of the skeleton and inspecting the regimes that are implied by the model, one might subject the residuals to more formal diagnostic tests, comparable to the usual practice in the Box-Jenkins approach in linear time series modeling. It turns out, however, that not all test statistics which have been developed in the context of ARMA models are applicable to the residuals from nonlinear models. The Jarque-Bera test for normality of the residuals is an example of a test which remains valid, while the Ljung-Box test statistic (2.14) is an example of a test which does not, see Eitrheim and Teräsvirta (1996). The LM approach to testing for serial correlation can still be used however, as shown by Eitrheim and Teräsvirta (1996) and discussed in some detail below. Out-of-sample forecasting and impulse response analysis are other methods to evaluate the properties of estimated STAR models. These topics are addressed in the following sections. In the remainder of this section I review the diagnostic checks for residual autocorrelation, remaining nonlinearity and parameter constancy, developed by Eitrheim and Teräsvirta (1996).

Testing for residual autocorrelation

Consider the STAR model of order p,

$$y_t = F(x_t;\theta) + \varepsilon_t, \tag{2.41}$$

where $x_t = (1, \tilde{x}_t')'$, $\tilde{x}_t = (y_{t-1}, \ldots, y_{t-p})'$ as before and the skeleton $F(x_t;\theta)$ is given in (2.34). An LM-test for q-th order serial dependence in ε_t can be obtained as nR^2, where R^2 is the coefficient of determination from the regression of $\hat{\varepsilon}_t$ on $\partial F(x_t;\hat{\theta})/\partial\theta$ and q lagged residuals $\hat{\varepsilon}_{t-1}, \ldots, \hat{\varepsilon}_{t-q}$. Hats indicate that the relevant quantities are estimates under the null hypothesis of serial independence of ε_t. The resultant test statistic, to be denoted as $\text{LM}_{\text{SI}}(q)$, is χ^2 distributed with q degrees of freedom asymptotically.

This test statistic is in fact a generalization of the LM-test for serial correlation in an AR(p) model of Breusch and Pagan (1979), which is based on the auxiliary regression

$$\hat{\varepsilon}_t = \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} + \beta_1 \hat{\varepsilon}_{t-1} + \cdots + \beta_q \hat{\varepsilon}_{t-q} + \nu_t, \tag{2.42}$$

where $\hat{\varepsilon}_t$ are the residuals of the AR(p) model. To understand why this is the case, note that for a linear AR(p) model (without an intercept) $F(x_t;\theta) = \sum_{i=1}^{p} \phi_i y_{t-i}$ and

$$\frac{\partial F(x_t;\hat{\theta})}{\partial\theta} = (y_{t-1}, \ldots, y_{t-p})'.$$

In case of a STAR model, the skeleton is given by $F(x_t;\theta) = \phi_1' x_t(1 - G(s_t;\gamma,c)) + \phi_2' x_t G(s_t;\gamma,c)$. Hence, in this case $\theta = (\phi_1, \phi_2, \gamma, c)$ and the relevant partial derivatives $\partial F(x_t;\hat{\theta})/\partial\theta$ can be obtained in a straightforward manner, see Eitrheim and Teräsvirta (1996) for details.

Testing for remaining nonlinearity

An important question when using nonlinear time series models is whether the estimated model adequately captures the nonlinear features of the time series under investigation. An obvious way to examine this is to apply a test for no remaining nonlinearity to an estimated model. In case of regime-switching models such as the STAR model, a natural approach is to specify the alternative hypothesis of remaining nonlinearity as the presence of an additional regime. For example, one might want to test the null hypothesis that a two-regime model is adequate against the alternative that a third regime is necessary.

Eitrheim and Teräsvirta (1996) develop an LM statistic to test a two-regime STAR model against the alternative of an additive STAR model, which can be written as

$$y_t = \phi_1' x_t + (\phi_2 - \phi_1)' x_t G_1(s_{1t}; \gamma_1, c_1) + (\phi_3 - \phi_1)' x_t G_2(s_{2t}; \gamma_2, c_2) + \varepsilon_t. \tag{2.43}$$

The two-regime model that has been estimated is assumed to have $G_1(\cdot)$ as transition function. Hence, the hypothesis to be tested concerns the question whether extending the model with the terms $(\phi_3 - \phi_1)' x_t G_2(\cdot)$ is appropriate. The null hypothesis of a two-regime model can be expressed as either $H_0 : \phi_3 = \phi_1$ or $H_0' : \gamma_2 = 0$. Evidently, this testing problem suffers from a similar identification problem as the problem of testing the null hypothesis of linearity against the alternative of a two-regime STAR model discussed in Section 2.2. The solution here is the same as well, in that the transition function $G_2(s_{2t}; \gamma_2, c_2)$ is replaced by a Taylor approximation around the point $\gamma_2 = 0$. In case of a third-order approximation, the resultant auxiliary model is given by

$$y_t = \beta_0' x_t + (\phi_2 - \phi_1)' x_t G_1(s_{1t}; \gamma_1, c_1) + \beta_1' \tilde{x}_t s_{2t} + \beta_2' \tilde{x}_t s_{2t}^2 + \beta_3' \tilde{x}_t s_{2t}^3 + e_t, \tag{2.44}$$

where the parameters $\beta_i,\ i = 0, 1, 2, 3$, are functions of the parameters ϕ_1, ϕ_2, γ_2 and c_2. The null hypothesis $H_0' : \gamma_2 = 0$ in (2.43) translates into $H_0'' : \beta_1 = \beta_2 = \beta_3 = 0$ in (2.44). The test statistic can be computed as nR^2 from the auxiliary regression of the residuals obtained from estimating the model under the null hypothesis $\hat{\varepsilon}_t$ on the partial derivates of the regression function with respect to the parameters in the two-regime model, ϕ_1, ϕ_2, γ_1 and c_1, evaluated under the null hypothesis, and the auxiliary regressors $\tilde{x}_t s_{2t}^i$, $i = 1, 2, 3$. The resultant test statistic LM_{AMR} has an asymptotic χ^2 distribution with $3p$ degrees of freedom. The additive STAR model as given in (2.43) is discussed in somewhat more detail in the next chapter.

Testing parameter constancy

Parameter constancy also is an important prerequisite of an empirically relevant model. Consider the following representation

$$y_t = \phi_1(t)' x_t (1 - G_1(s_t; \gamma_1, c_1)) + \phi_2(t)' x_t G_1(s_t; \gamma_1, c_1) + \varepsilon_t, \tag{2.45}$$

with

$$\phi_1(t) = \phi_1[1 - G_2(t;\gamma_2, c_2)] + \phi_3 G_2(t;\gamma_2, c_2), \tag{2.46}$$
$$\phi_2(t) = \phi_2[1 - G_2(t;\gamma_2, c_2)] + \phi_4 G_2(t;\gamma_2, c_2), \tag{2.47}$$

and

$$G_2(t;\gamma_2, c_2) = \frac{1}{1 + \exp\{-\gamma_2(t - c_2)\}}, \qquad \gamma_2 > 0. \tag{2.48}$$

Substituting (2.46) and (2.47) in (2.45), the model can be rewritten as

$$y_t = [\phi_1' x_t(1 - G_1(s_t)) + \phi_2' x_t G_1(s_t)][1 - G_2(t)] + [\phi_3' x_t(1 - G_1(s_t)) + \phi_4' x_t G_1(s_t)]G_2(t) + \varepsilon_t, \tag{2.49}$$

which is a so-called time-varying STAR [TV-STAR] model, allowing for both nonlinear dynamics of the STAR-type and time-varying parameters. This model is discussed in detail in Chapter 4. The point of interest here is that by testing the hypothesis $\gamma_2 = 0$, one can test for parameter constancy in the two-regime STAR model (2.41) with (2.34), against the alternative of smoothly changing parameters as in (2.45). LM-type test statistics, to be denoted as $\mathrm{LM_C}$, can be constructed in a straightforward manner, see also Section 3.2.1.

2.5 Forecasting with smooth transition models

Nonlinear time series models may be considered for various purposes. Sometimes the main objective merely is obtaining an adequate description of the dynamic patterns that are present in a particular variable. Very often, however, an additional goal is to employ the model for forecasting future values of the time series. Furthermore, out-of-sample forecasting also can be considered as a way to evaluate estimated models. Especially comparison of the forecasts from nonlinear models with those from a benchmark linear model might enable one to determine the added value of the nonlinear features of the model.

In this section I discuss several issues related to out-of-sample forecasting with nonlinear time series models. It is well-known that forecasting with nonlinear models is considerably more involved than forecasting with linear models. For that reason, various techniques to construct point and interval forecasts from nonlinear models are dealt with in Sections 2.5.1 and 2.5.2, respectively. In Section 2.5.3 the questions of how to evaluate forecasts from nonlinear models and how to compare forecasts from linear and nonlinear models in particular are addressed. As the issues that are addressed here are not restricted to the smooth transition model but apply to nonlinear time series models in general, the discussion is held in general terms, although the STAR model is used as an illustrative example at some points.

Clements and Hendry (1998) give an in-depth treatment of forecasting with linear models. For alternative reviews of forecasting with nonlinear models see Tong (1990, Chapter 6) and Granger and Teräsvirta (1993, Section 8.1).

2.5.1 Point forecasts

Let $\hat{y}_{t+h|t}$ denote a forecast of y_{t+h} made at time t, with associated forecast or prediction error $e_{t+h|t}$,

$$e_{t+h|t} = y_{t+h} - \hat{y}_{t+h|t}. \tag{2.50}$$

Here I restrict attention to forecasting with a quadratic loss function, that is, the forecast $\hat{y}_{t+h|t}$ minimizes the (conditional) squared prediction error [SPE]

$$\text{SPE}(h) \equiv \text{E}[e^2_{t+h|t}] = \text{E}[(y_{t+h} - \hat{y}_{t+h|t})^2|\Omega_t], \tag{2.51}$$

where Ω_t denotes the history of the time series up to and including time t. The forecast which minimizes (2.51) is the conditional expectation of y_{t+h} at time t, that is

$$\hat{y}_{t+h|t} = \text{E}[y_{t+h}|\Omega_t], \tag{2.52}$$

see Box and Jenkins (1970).

Point forecasts from linear models can be obtained very easily. For example, for the AR(1) model

$$y_t = \phi_1 y_{t-1} + \varepsilon_t, \tag{2.53}$$

it follows that

$$\hat{y}_{t+h|t} = \text{E}[\phi_1 y_{t+h-1} + \varepsilon_{t+h}|\Omega_t] = \phi_1 \hat{y}_{t+h-1|t}, \tag{2.54}$$

with $\hat{y}_{t+h-1|t} = y_t$ in case $h = 1$. This recursive relationship allows multiple-step ahead forecasts to be obtained with minimum effort. Generalizing (2.54) to AR(p) models with $p > 1$ is straightforward.

Computing point forecasts from nonlinear models is considerably more involved than computing forecasts from linear models. Consider the case where y_t is described by the general nonlinear autoregressive model of order 1,

$$y_t = F(y_{t-1}; \theta) + \varepsilon_t, \tag{2.55}$$

for some nonlinear function $F(y_{t-1}; \theta)$. Recall that for the STAR model with $s_t = y_{t-1}$

$$F(y_{t-1}; \theta) = (\phi_{1,0} + \phi_{1,1} y_{t-1})(1 - G(y_{t-1}; \gamma, c)) + (\phi_{2,0} + \phi_{2,1} y_{t-1}) G(y_{t-1}; \gamma, c). \tag{2.56}$$

Using the fact that $\text{E}[\varepsilon_{t+1}|\Omega_t] = 0$, the optimal 1-step ahead forecast of y_{t+1} is easily obtained as

$$\hat{y}_{t+1|t} = \text{E}[y_{t+1}|\Omega_t] = F(y_t; \theta). \tag{2.57}$$

which is equivalent to the optimal 1-step ahead forecast in case the model $F(y_{t-1}; \theta)$ is linear. When the forecast horizon is larger than 1 period, things become more complicated however. For example, the optimal 2-step ahead forecast follows from (2.55) as

$$\hat{y}_{t+2|t} = \text{E}[y_{t+2}|\Omega_t] = \text{E}[F(y_{t+1}; \theta)|\Omega_t]. \tag{2.58}$$

In general, the *linear* conditional expectation operator E can not be interchanged with the *nonlinear* operator F, that is

$$\text{E}[F(\cdot)] \neq F(\text{E}[\cdot]).$$

Put differently, the expected value of a nonlinear function is not equal to the function evaluated at the expected value of its argument. Hence,

$$\mathrm{E}[F(y_{t+1};\theta)|\Omega_t] \neq F(\mathrm{E}[y_{t+1}|\Omega_t];\theta) = F(y_{t+1|t};\theta). \tag{2.59}$$

Rather, the relation between the 1- and 2-step ahead forecasts is given by

$$\begin{aligned}\hat{y}_{t+2|t} &= \mathrm{E}[F(F(y_t;\theta)+\varepsilon_{t+1};\theta)|\Omega_t] \\ &= \mathrm{E}[F(y_{t+1|t}+\varepsilon_{t+1};\theta)|\Omega_t].\end{aligned} \tag{2.60}$$

The above demonstrates that a simple recursive relationship between forecasts at different horizons, which could be used to obtain multiple-step ahead forecasts in an easy fashion analogous to (2.54), does not exist for nonlinear models in general. Of course, a 2-step ahead forecast might still be constructed as

$$\hat{y}^{(n)}_{t+2|t} = F(\hat{y}_{t+1|t};\theta). \tag{2.61}$$

Brown and Mariano (1989) show that this 'naïve' approach, which lends it name from the fact that it effectively boils down to setting $\varepsilon_{t+1}=0$ in (2.60) (or interchanging E and F in (2.58)), renders biased forecasts. Over the years, several methods have been developed to obtain more accurate multiple-step ahead forecasts, some of which are discussed below.

First, one might attempt to obtain the conditional expectation (2.60) directly by computing

$$\hat{y}^{(c)}_{t+2|t} = \int_{-\infty}^{\infty} F(y_{t+1|t}+\varepsilon;\theta)f(\varepsilon)d\varepsilon, \tag{2.62}$$

where f denotes the density of ε_{t+1}. Brown and Mariano (1989) refer to this forecast as the 'closed form' forecast - hence the superscript (c). An alternative way to express this integral follows from (2.58) as

$$\begin{aligned}\hat{y}^{(c)}_{t+2|t} &= \int_{-\infty}^{\infty} F(y_{t+1};\theta)g(y_{t+1}|\Omega_t)dy_{t+1} \\ &= \int_{-\infty}^{\infty} \mathrm{E}[y_{t+2}|y_{t+1}]g(y_{t+1}|\Omega_t)dy_{t+1},\end{aligned} \tag{2.63}$$

where $g(y_{t+1}|\Omega_t)$ is the distribution of y_{t+1} conditional upon Ω_t. This conditional distribution is in fact equal to the distribution $f(\cdot)$ of the shocks ε_{t+1} with mean equal to $F(y_t;\theta)$, that is, $g(y_{t+1}|\Omega_t) = f(y_{t+1}-F(y_t;\theta))$. As an analytic expression for the integral (2.62) (or (2.63)) is not available in general, it needs to be approximated using numerical techniques. An additional complication is the fact that the distribution of ε_{t+1} is never known with certainty. Usual practice is to assume normality of ε_{t+1}.

The closed form forecast becomes quite tedious to compute for forecasts more than two periods ahead. To see why, consider the Chapman-Kolgomorov relation

$$g(y_{t+h}|\Omega_t) = \int_{-\infty}^{\infty} g(y_{t+h}|y_{t+h-1})g(y_{t+h-1}|\Omega_t)dy_{t+h-1}. \tag{2.64}$$

where $g(y_{t+h}|y_{t+h-1})$ is the distribution of y_{t+h} conditional upon y_{t+h-1}. By taking conditional expectations on both sides of (2.64) it follows that

$$\mathrm{E}[y_{t+h}|\Omega_t] = \int_{-\infty}^{\infty} \mathrm{E}[y_{t+h}|y_{t+h-1}]g(y_{t+h-1}|\Omega_t)dy_{t+h-1}, \tag{2.65}$$

which can be recognized as a generalization of (2.63). To evaluate this integral to obtain the h-step ahead exact forecast, one needs the conditional distribution $g(y_{t+h-1}|\Omega_t)$. In principle this distribution can be obtained recursively from (2.64), by observing that $g(y_{t+1}|y_{t+h-1})$ again is equal to the distribution of the shocks ε_{t+1} with its mean shifted to $F(y_{t+h-1};\theta)$. The recursion can be started for $h = 2$ by using the fact that $g(y_{t+1}|\Omega_t) = f(y_{t+1} - F(y_t;\theta))$ as noted above. To obtain the conditional distribution $g(y_{t+h-1}|\Omega_t)$ for $h > 2$ involves repeated numerical integration, which may become rather time-consuming, in particular if a large number of forecasts is to be made.

An alternative is to assume that the $(h-1)$-step ahead forecast error $e_{t+h-1|t} = y_{t+h-1} - \hat{y}_{t+h-1|t}$ is normally distributed with mean zero and variance σ^2_{h-1}. In that case, $g(y_{t+h-1}|\Omega_t)$ is normal with mean equal to the $(h-1)$-step ahead point forecast $\hat{y}_{t+h-1|t}$ and variance σ^2_{h-1}. This so-called normal forecast error [NFE] method was developed by Pemberton (1987) for general nonlinear autoregressive models, and applied by Al-Qassam and Lane (1989) to exponential autoregressive models (which are closely related to the ESTAR model) and by de Gooijer and de Bruin (1998) to SETAR models. For the two-regime SETAR model, given by (2.55) with (2.56), with $\gamma = \infty$, h-step ahead NFE forecasts can be computed from the recursion

$$\begin{aligned}\hat{y}^{(nfe)}_{t+h|t} = {} & \Phi(z_{t+h-1|t})(\phi_{1,0} + \phi_{1,1}\hat{y}_{t+h-1|t}) + \\ & \Phi(-z_{t+h-1|t})(\phi_{2,0} + \phi_{2,1}\hat{y}_{t+h-1|t}) + \phi(z_{t+h-1|t})(\phi_{2,1} - \phi_{1,1})\sigma_{h-1},\end{aligned} \tag{2.66}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal distribution and density functions, respectively, σ^2_{h-1} is the variance of the $(h-1)$-step ahead forecast error $e_{t+h-1|t}$ and $z_{t+h-1|t} = (c - \hat{y}_{t+h-1|t})/\sigma_{h-1}$. Observe that (2.66) essentially is a weighted average of the optimal forecasts from the two regimes, with weights equal to the probability of being in the particular regime at time $t+h-1$ under normality, plus an additional correction factor. A similar recursion for the variance of the forecast error, σ^2_{h-1}, is also available, see de Gooijer and de Bruin (1998).

An alternative approach to computing multiple-step ahead forecasts is to use Monte Carlo or bootstrap methods to approximate the conditional expectation (2.60). The 2-step ahead Monte Carlo forecast is given by

$$\hat{y}^{(mc)}_{t+2|t} = \frac{1}{k}\sum_{i=1}^{k} F(y_{t+1|t} + \varepsilon_i;\theta), \tag{2.67}$$

where k is some large number and the ε_i are drawn from the presumed distribution of ε_t. The bootstrap forecast is very similar, the only difference being that the residuals from the estimated model, $\hat{\varepsilon}_t, t = 1, \ldots, T$, are used,

$$\hat{y}^{(b)}_{t+2|t} = \frac{1}{T}\sum_{i=1}^{T} F(y_{t+1|t} + \hat{\varepsilon}_i;\theta). \tag{2.68}$$

The advantage of the bootstrap over the Monte Carlo method is that no assumptions need to be made concerning the distribution of ε_t.

Lin and Granger (1994) and Clements and Smith (1997) compare various methods to obtain multiple-step ahead forecasts for STAR and SETAR models, respectively. Their main findings are that the Monte Carlo and bootstrap methods compare favorably to the other methods.

2.5.2 Interval forecasts

In addition to point forecasts one may also be interested in confidence intervals for these point forecasts. For forecasts obtained from linear models, the usual forecast confidence region is taken to be an interval symmetric around the point forecast. This is based upon the fact that (under the assumption of normally distributed innovations ε_t), the conditional distribution $g(y_{t+h}|\Omega_t)$ of a linear time series is normal with mean equal to $\hat{y}_{t+h|t}$. For example, for the AR(1) model (2.53), the h-step ahead forecast error is[5]

$$\begin{aligned} e_{t+h|t} = y_{t+h} - \hat{y}_{t+h|t} &= \phi_1 y_{t+h-1} + \varepsilon_{t+h} - \phi_1^h y_t \\ &= \phi_1^2 y_{t+h-2} + \varepsilon_{t+h} + \phi_1 \varepsilon_{t+h-1} - \phi_1^h y_t \\ &= \dots \\ &= \phi_1^h y_t + \sum_{i=1}^{h} \phi_1^{h-i} \varepsilon_{t+i} - \phi_1^h y_t \\ &= \sum_{i=1}^{h} \phi_1^{h-i} \varepsilon_{t+i}, \end{aligned} \tag{2.69}$$

with associated SPE

$$\mathrm{SPE}(h) \equiv \mathrm{E}[e_{t+h|t}^2|\Omega_t] = \sum_{i=1}^{h} \phi_1^{2(h-i)} \sigma^2. \tag{2.70}$$

Assuming normality, a 95% forecast confidence interval for y_{t+h} is bounded by $\hat{y}_{t+h|t} - 1.96 \cdot \sqrt{\mathrm{SPE}(h)}$ and $\hat{y}_{t+h|t} + 1.96 \cdot \sqrt{\mathrm{SPE}(h)}$.

For nonlinear models, the conditional distribution $g(y_{t+h}|\Omega_t)$ need not be normal. In fact, the conditional distribution can be asymmetric and even can contain multiple modes. Whether a symmetric interval around the mean or, equivalently, the point forecast is the most appropriate forecast confidence region in this case can be questioned. This topic is discussed in detail in Hyndman (1995). He argues that there are three methods to construct a $100 \cdot (1 - \alpha)\%$ forecast region:

[5] Throughout this section I assume that the parameters in the model are known. In practice the parameters of course have to be estimated, which leads to an additional forecast error. See Clements and Hendry (1998) for a taxonomy of forecast errors and the relative importance of the different sources of forecast uncertainty in linear time series models.

1. An interval symmetric around the point forecast
$$S_\alpha = (\hat{y}_{t+h|t} - w, \hat{y}_{t+h|t} + w),$$
where w is such that $P(y_{t+h} \in S_\alpha|\Omega_t) = 1 - \alpha$.

2. The interval between the $\alpha/2$ and $(1-\alpha/2)$ quantiles of the forecast distribution, denoted $q_{\alpha/2}$ and $q_{1-\alpha/2}$, respectively,
$$Q_\alpha = (q_{\alpha/2}, q_{1-\alpha/2}).$$

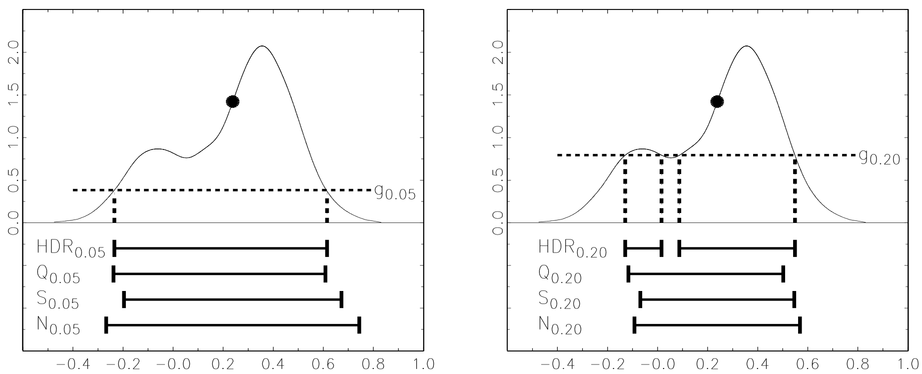
3. The highest-density region [HDR]
$$HDR_\alpha = \{y|g(y_{t+h}|\Omega_t) \geq g_\alpha\}, \tag{2.71}$$
where g_α is such that $P(y_{t+h} \in HDR_\alpha|\Omega_t) = 1 - \alpha$.

For symmetric and unimodal distributions, these three regions are identical. For asymmetric or multimodal distributions they are not. Hyndman (1995) argues that the HDR is the most natural choice. The reasons for this claim are that first, HDR_α is the smallest of all possible $100 \cdot (1-\alpha)\%$ forecast regions and, second, every point inside the HDR has conditional density $g(y_{t+h}|\Omega_t)$ at least as large as every point outside the region. Furthermore, only the HDR will reveal features such as asymmetry or multimodality of the conditional distribution $g(y_{t+h}|\Omega_t)$. HDRs are straightforward to compute when the Monte Carlo or bootstrap methods described previously are used to compute the point forecast $\hat{y}_{t+h|t}$. Let $y^i_{t+h|t}$, $i = 1, 2, \ldots$, denote the i-th element used in computing the Monte Carlo forecast (2.67) or bootstrap forecast (2.67), that is, $y^i_{t+h|t} = F(\hat{y}_{t+h-1|t} + \varepsilon_i; \theta)$ or $y^i_{t+h|t} = F(\hat{y}_{t+h-1|t} + \hat{\varepsilon}_i; \theta)$. Note that the $y^i_{t+h|t}$ can be thought of as being realizations drawn from the conditional distribution of interest $g(y_{t+h}|\Omega_t)$. Estimates $g_i \equiv g(y^i_{t+h|t}|\Omega_t)$, $i = 1, \ldots, k$, then can be obtained by using a standard kernel density estimator, that is

$$g_i = \frac{1}{k} \sum_{j=1}^{k} K([y^i_{t+h|t} - y^j_{t+h|t}]/b), \tag{2.72}$$

where $K(\cdot)$ is a kernel function such as the Gaussian density and $b > 0$ is the bandwidth, and $k = T$ in case the bootstrap forecast is used. An estimate of g_α in (2.71) is given by $\hat{g}_\alpha = g_{(\lfloor \alpha k \rfloor)}$, where the $g_{(i)}$ are the ordered g_i and $\lfloor \cdot \rfloor$ denotes integer part. See Hyndman (1996) for more details and some suggestions about the display of HDRs.

As an example, consider again the STAR model (2.6), with $\phi_{1,1} = -0.5$ and $\phi_{2,1} = 0.5$. By setting $\phi_{1,0} = 0.3$, $\phi_{1,0} = -0.1$, $\gamma = 25$, $c = 0$, the resultant model has a limit cycle consisting of three points, $y_1^* = -0.06667$, $y_2^* = 0.06667$ and $y_3^* = 0.33333$. Assuming that $\varepsilon_t \sim$ i.i.d. $N(0, .125^2)$, the optimal 2-step ahead forecasts of y_{t+2} given $y_t = y_2^*$ is approximately equal to $\mathrm{E}[y_{t+2}|y_t = y_2^*] = 0.238$. This can be verified using the recursive NFE-forecast (2.66), as this is identical to



(a) $g(y_{t+2}|y_t = y_2^*)$ with 95% confidence regions

(b) $g(y_{t+2}|y_t = y_2^*)$ with 80% confidence regions

Figure 2.7: Two-step ahead conditional distributions for for the STAR model (2.6), with $\phi_{1,0} = 0.3$, $\phi_{1,1} = -0.5$, $\phi_{2,0} = -0.1$, $\phi_{2,1} = 0.5$, $\gamma = 25$, $c = 0$ and $\varepsilon_t \sim \text{NID}(0, 0.125^2)$, together with confidence regions for the 2-step ahead forecast.

the exact 2-steps ahead forecast in case the errors are normally distributed and, for the large value of γ that is chosen, the STAR model effectively reduces to a SETAR model. The corresponding 2-step ahead forecast error variance is equal to 0.258^2. The conditional distribution $g(y_{t+2}|y_t = y_2^*)$ is given in Figure 2.7, and is seen to be bimodal. Intuitively, if $y_t = y_2^*$, it is very likely that y_{t+2} will be close to y_1^* as the time series iterates among the three points of the limit cycle in case no shocks occur. This corresponds with the largest mode of the conditional distribution. There is however a small probability that the time series will 'linger' around either y_2^* or y_3^*, giving rise to the smaller mode. The optimal point forecast $\hat{y}_{t+2|t}$ is shown as a solid circle.

Below the conditional densities, 95% and 80% confidence regions have been drawn in the left and right panels, respectively. For the 95% confidence regions, the HDR is almost identical to the region $Q_{0.05}$ based on the quantiles of the conditional distribution. The interval symmetric around the point forecast, $S_{0.05}$ is shifted somewhat to the right. Also shown is a region $N_{0.05}$, which is the confidence interval obtained when the conditional distribution is assumed to be normal, and the confidence interval is constructed in the usual manner as $(\hat{y}_{t+2|t} - 1.96\sigma_2, \hat{y}_{t+2|t} + 1.96\sigma_2)$. Clearly, this renders an interval which is too wide and has more than 95% coverage. The 80% confidence regions shown in the right panel show that the HDR need not be a continuous interval, but can consist of several disjoint segments.

2.5.3 Evaluating forecasts

It is good practice to evaluate the quality of forecasts from a time series model. Relative forecast performance can also be used as a model selection criterion, as an alternative or complement to an in-sample comparison of different models. In this section, I discuss a few dimensions along which forecasts from a (set of) model(s) can be evaluated (compared). As of yet, there is no agreement on the most adequate measure to use. Hence I will consider only a few rather simple ones which are easy to compute. An extensive evaluation of more advanced criteria, such as those proposed in Christoffersen (1998), Diebold, Gunther and Tay (1998), and Clements and Smith (1998a) is left for further research. For simplicity, I assume that m 1-step ahead forecasts $\hat{y}_{T+j|T+j-1}$, $j = 1, \ldots, m$, and the corresponding realizations $y_{T+1}, \ldots, y_{T+m}$ are available. It should be remarked that the various evaluation criteria discussed below can be applied to multiple-step ahead forecasts as well.

Evaluating forecasts based on squared and absolute forecast errors

Traditional forecast evaluation criteria are the mean squared prediction error [MSPE],

$$\text{MSPE} = \frac{1}{m}\sum_{j=1}^{m}(\hat{y}_{T+j|T+j-1} - y_{T+j})^2, \tag{2.73}$$

and the mean absolute prediction error [MAPE]

$$\text{MAPE} = \frac{1}{m}\sum_{j=1}^{m}|\hat{y}_{T+j|T+j-1} - y_{T+j}|. \tag{2.74}$$

Models with smaller MSPE and/or MAPE have a better forecast performance.

Diebold and Mariano (1995) discuss several statistics that can be used to examine whether the MSPEs or MAPEs of two alternative models A and B are significantly different. Define the loss differential for the j-th forecast, denoted as d_j, as

$$d_j = |e_{T+j|T+j-1,A}|^k - |e_{T+j|T+j-1,B}|^k \qquad j = 1, 2, \ldots, m,$$

with $e_{T+j|T+j-1,A}$ and $e_{T+j|T+j-1,B}$ the forecast errors at time $T+j$ associated with the forecasts from from models A and B, respectively, and k is equal to 2 and 1 if the goal is to compare the squared and absolute prediction errors, respectively. The null hypothesis of equal forecast accuracy of models A and B is $\text{E}[d_j] = 0$ for all j, which can be tested by examining the average loss differential $\bar{d} = \frac{1}{m}\sum_{j=1}^{m} d_j$. The relevant test statistic is given by

$$\text{DM} = \frac{\bar{d}}{\sqrt{\omega}}, \tag{2.75}$$

where ω is the asymptotic variance of $\bar{d}$. Diebold and Mariano (1995) suggest to estimate ω by an unweighted sum of the autocovariances of d_j, denoted $\hat{\gamma}_i(d)$, as

$$\hat{\omega} = \sum_{i=-(h-1)}^{h-1} \hat{\gamma}_i(d), \tag{2.76}$$

where h is the forecast horizon for which the prediction errors are compared. Notice that in case $h = 1$, it follows from (2.76) that $\hat{\omega}$ is simply the variance of d_j,

$\hat{\gamma}_0(d)$. The DM statistic has a standard normal distribution asymptotically. See also Harvey, Leybourne and Newbold (1997) for a refinement of this statistic.

Even though traditional criteria such as the MSPE are applicable to forecasts from nonlinear models, they might not do the nonlinear model justice. As noted by Tong (1995), 'how well we can forecast depends on where we are.' In case of regime-switching models for example, it might very well be that the forecastability of the time series is very different in different regimes. One therefore might evaluate the forecasts for each regime separately to investigate whether the nonlinear model is especially useful to obtain forecasts in a particular regime or state, see Tiao and Tsay (1994), Clements and Smith (1999) and the empirical example in Section 2.7.

Evaluating forecasts based on correct regime prediction [CRP]

An alternative way to evaluate out-of-sample forecasts from regime-switching models, such as the STAR model, is to examine their ability to predict future regimes. Pesaran and Timmermann (1992) develop a test statistic to assess the accuracy of predicted directions of change in a variable. It is straightforward to adapt this idea to construct a test for the accuracy of predicted regimes, as shown below. See Clements and Smith (1999) for related measures.

Define

$$\hat{r}_{T+j+h|T+j} = \begin{cases} 1 & \text{if } G(\hat{s}_{T+j+h|T+j}; \gamma, c) > 0.5, \\ -1 & \text{if } G(\hat{s}_{T+j+h|T+j}; \gamma, c) < 0.5, \end{cases} \tag{2.77}$$

where $\hat{s}_{T+j+h|T+j}$ is the h-step ahead forecast of the transition variable at $t = T + j + h$. The variable $\hat{r}_{T+j+h|T+j}$ can be interpreted as the h-step ahead forecast of the regime that will be realized test at $T + j + h$. Define the success ratio [SR]

$$\text{SR} = \frac{1}{m} \sum_{j=1}^{m} I[r_{T+j+h} \cdot \hat{r}_{T+j+h|T+j} > 0], \tag{2.78}$$

where r_{T+j+h} is defined as in (2.77), but using the realized values of the transition variable s_{T+j+h} instead, and $I[\cdot]$ is the usual indicator function. Notice that SR is the fraction of time the prevailing regime is predicted correctly. It is straightforward to test whether the value of SR differs significantly from the success ratio that would be obtained in case r_{T+j+h} and $\hat{r}_{T+h+j|T+j}$ are independent, that is, if the ability of the model to predict the regime correctly were no better than a random guess. Define

$$P = \frac{1}{m} \sum_{j=1}^{m} I[r_{T+j+h} = 1],$$

and

$$\hat{P} = \frac{1}{m} \sum_{j=1}^{m} I[\hat{r}_{T+j+h|T+j} = 1].$$

The success ratio in case of independence [SRI] of r_{T+j+h} and $\hat{r}_{T+j+h|T+j}$ can be computed as

$$\text{SRI} = P\hat{P} + (1 - P)(1 - \hat{P}), \tag{2.79}$$

which has variance given by

$$\mathrm{var}(\mathrm{SRI}) = \frac{1}{m}[(2\hat{P}-1)^2 P(1-P) + (2P-1)^2\hat{P}(1-\hat{P}) + \frac{4}{m}P\hat{P}(1-P)(1-\hat{P})]. \quad (2.80)$$

The variance of the success ratio SR in (2.78) is equal to

$$\mathrm{var}(\mathrm{SR}) = \frac{1}{m}\mathrm{SRI}(1-\mathrm{SRI}). \quad (2.81)$$

The regime accuracy [RA] test is now calculated as

$$\mathrm{RA} = \frac{\mathrm{SR}-\mathrm{SRI}}{\sqrt{\mathrm{var}(\mathrm{SR})-\mathrm{var}(\mathrm{SRI})}}, \quad (2.82)$$

which has an asymptotic standard normal distribution under the null hypothesis that r_{T+j+h} and $\hat{r}_{T+j+h|T+j}$ are independently distributed.

Notice that the statistic derived above is closely related to the standard χ^2 test of independence in a 2×2 contingency table of predicted and realized regimes, see also Pesaran and Timmermann (1994). It is straightforward to extend the test to allow for more than 2 regimes.

Discussion

In general, the fact that a particular model describes the features of a time series within the estimation sample better than other models is no guarantee that this model also renders better out-of-sample forecasts. Clements and Hendry (1998) discuss various reasons why a model with a superior in-sample fit may nevertheless yield inferior out-of-sample forecasts. The above seems particularly relevant for nonlinear time series models. It is found quite often that, even though a nonlinear model appears to describe certain characteristics of the time series at hand much better than a linear model, the forecasting performance of the linear model is no worse than that of the nonlinear model, see de Gooijer and Kumar (1992) among others. A lot of reasons can be brought up why this might be the case, see also Diebold and Nason (1990). For example, the nonlinearity may be 'spurious', in the sense that other features of the time series, such as heteroskedasticity, structural breaks or outliers, suggest the presence of nonlinearity. Even though one might successfully estimate a nonlinear model for such a series, it is very unlikely that this will result in improved forecasts.

Another possible cause for the poor forecast performance of nonlinear models is that the nonlinearity does not show up during the forecast period. In case of regime-switching models, for example, it might be that only one of the regimes is realized during the forecast period. Hence, empirical forecasts do not always allow to assess the forecasting quality of the nonlinear model completely. A potential solution to this problem is to perform a simulation experiment in which one uses an estimated regime-switching model to generate artificial time series and to perform an out-of-sample forecasting exercise on each of those series. In this controlled environment one can make sure that forecasts in each of the regimes are involved. See Clements and Smith (1998b,1999) for applications of this approach. This simulation approach can also be applied to compare the forecast performance of alternative nonlinear models by using each of the alternatives as DGP in turn, see Clements and Krolzig (1998).

2.6 Impulse response functions

Another way to evaluate the properties of estimated regime-switching models is to examine the effects of the shocks ε_t on the evolution of the time series y_t. Impulse response functions are a convenient tool to carry out such an analysis.

Impulse response functions are meant to provide a measure of the response of y_{t+h} to a shock or impulse δ at time t. The impulse response measure which is commonly used in the analysis of linear models is defined as the difference between two realizations of y_{t+h} which start from identical histories of the time series up to time $t-1$, denoted as ω_{t-1}. In one realization, the process is 'hit' by a shock of size δ at time t, while in the other realization no shock occurs at time t. All shocks in intermediate periods between t and $t+h$ are set equal to zero in both realizations. That is, the traditional impulse response function [TIRF] is given by

$$\begin{aligned}\mathrm{TIRF}_y(h,\delta,\omega_{t-1}) = {}& \mathrm{E}[y_{t+h}|\varepsilon_t=\delta,\varepsilon_{t+1}=\ldots=\varepsilon_{t+h}=0,\omega_{t-1}]- \\ & \mathrm{E}[y_{t+h}|\varepsilon_t=0,\varepsilon_{t+1}=\ldots=\varepsilon_{t+h}=0,\omega_{t-1}], \quad (2.83)\end{aligned}$$

for $h=0,1,2,\ldots$. The second conditional expectation usually is called the benchmark profile.

The traditional impulse response function as defined above has some characteristic properties in case the model is linear. First, the TIRF then is *symmetric*, in the sense that a shock of $-\delta$ has exactly the opposite effect as a shock of size $+\delta$. Furthermore, it might be called *linear*, as the impulse response is proportional to the size of the shock. Finally, the impulse response is *history independent* as it does not depend on the particular history ω_{t-1}. For example, in the AR(1) model (2.53), it follows easily that $\mathrm{TIRF}_y(h,\delta,\omega_{t-1})=\phi^h\delta$, which clearly demonstrates the aforementioned properties of the impulse response function.

These properties do not carry over to nonlinear models. In nonlinear models, the impact of a shock depends on the sign and the size of the shock, as well as on the history of the process. Furthermore, if the effect of a shock on the time series $h>1$ periods ahead is to be analyzed, the assumption that no shocks occur in intermediate periods might give rise to quite misleading inference concerning the propagation mechanism of the model.

To illustrate these points, consider the SETAR model

$$y_t = \begin{cases}\phi_{1,1}y_{t-1}+\varepsilon_t & \text{if } y_{t-1}\le 0,\\ \phi_{2,1}y_{t-1}+\varepsilon_t & \text{if } y_{t-1}>0.\end{cases} \quad (2.84)$$

The traditional impulse response 1 period ahead in this case is equal to

$$\text{TIRF}_y(1, \delta, \omega_{t-1}) = \begin{cases} \phi_{1,1}\delta & \text{if } y_{t-1} + \delta \leq 0 \text{ and } y_{t-1} \leq 0, \\ \phi_{1,1}\delta + \phi_{2,1}(\phi_{1,1} - \phi_{2,1})y_{t-1} & \text{if } y_{t-1} + \delta \leq 0 \text{ and } y_{t-1} > 0 \\ \\ \phi_{2,1}\delta + \phi_{1,1}(\phi_{2,1} - \phi_{1,1})y_{t-1} & \text{if } y_{t-1} + \delta > 0 \text{ and } y_{t-1} \leq 0, \\ \phi_{2,1}\delta & \text{if } y_{t-1} + \delta > 0 \text{ and } y_{t-1} > 0. \end{cases}$$

This example makes clear that the impulse response depends on the combined magnitude of the history y_{t-1} and the shock δ (relative to the threshold $c = 0$). Hence, the impulse response is not symmetric, as it might happen that $y_{t-1} + \delta > 0$ while $y_{t-1} - \delta \leq 0$, nor is it linear or history independent.

To illustrate the consequence of assuming no shocks occurring after time t, assume that $y_{t-1} = 0$ and the shock δ is negative. As no more shocks enter the system, the process remains in the lower regime after time t, and the effect of the shock δ decays geometrically with rate $\phi_{1,1}$. However, in practice, regime-switches are quite likely to occur due to subsequent shocks, which changes the dynamics of the process and, hence, the persistence of the shock δ. Thus, it might be very misleading to consider only the response which occurs when all shocks in intermediate periods are equal to zero.

In fact, the assumption of zero shocks in intermediate periods can be justified for linear AR(p) models by the fact that such models can be written in terms of the Wold representation

$$y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}, \tag{2.85}$$

where the parameters ψ_i, $i = 0, 1, 2, \ldots$, are functions of the AR parameters $\phi_1, \ldots, \phi_p$. The representation in (2.85) shows that shocks in different periods do not interact and, hence, any assumptions concerning the shocks $\varepsilon_{t+1}, \ldots, \varepsilon_{t+h}$ do not affect the value of the impulse response $\text{TIRF}_y(h, \delta, \omega_{t-1})$. Nonlinear time series models do not have a Wold representation however. They can be rewritten in terms of (past and present) shocks by means of the Volterra expansion,

$$y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} + \sum_{i=0}^{\infty} \sum_{i=j}^{\infty} \xi_{ij} \varepsilon_{t-i} \varepsilon_{t-j} + \sum_{i=0}^{\infty} \sum_{i=j}^{\infty} \sum_{k=j}^{\infty} \zeta_{ij} \varepsilon_{t-i} \varepsilon_{t-j} \varepsilon_{t-k} + \cdots, \tag{2.86}$$

see Priestley (1988). This expression shows that the effect of the shock ε_t on y_{t+h} depends on the shocks $\varepsilon_{t+1}, \ldots, \varepsilon_{t+h}$, as well as on past shocks $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots$, which constitute the history ω_{t-1}.

The Generalized Impulse Response Function [GIRF], introduced by Koop, Pesaran and Potter (1996) provides a natural solution to the problems involved in defining impulse responses in nonlinear models. The GIRF for an arbitrary shock $\varepsilon_t = \delta$ and history ω_{t-1} is defined as

$$\text{GIRF}_y(h, \delta, \omega_{t-1}) = \text{E}[y_{t+h} | \varepsilon_t = \delta, \omega_{t-1}] - \text{E}[y_{t+h} | \omega_{t-1}], \tag{2.87}$$

for $h = 0, 1, 2, \ldots$. In the GIRF, the expectations of y_{t+h} are conditioned only on the history and/or on the shock. Put differently, the problem of dealing with shocks occurring in intermediate time periods is dealt with by averaging them out. Given this choice, the natural benchmark profile for the impulse response is the expectation of y_{t+h} conditional only on the history of the process ω_{t-1}. Thus, in the benchmark profile the current shock is averaged out as well. It is easily seen that for linear models the GIRF in (2.87) is equivalent to the TIRF in (2.83).

The GIRF is a function of δ and ω_{t-1}, which are realizations of the random variables ε_t and Ω_{t-1}. Koop *et al.* (1996) stress that, hence, the GIRF as defined in (2.87) itself is a realization of a random variable given by

$$\text{GIRF}_y(h, \varepsilon_t, \Omega_{t-1}) = \text{E}[y_{t+h}|\varepsilon_t, \Omega_{t-1}] - \text{E}[y_{t+h}|\Omega_{t-1}]. \tag{2.88}$$

Using this interpretation of the GIRF as a random variable, various conditional versions can be defined which are of potential interest. For example, one might consider only a particular history ω_{t-1} and treat the GIRF as a random variable in terms of ε_t, that is,

$$\text{GIRF}_y(h, \varepsilon_t, \omega_{t-1}) = \text{E}[y_{t+h}|\varepsilon_t, \omega_{t-1}] - \text{E}[y_{t+h}|\omega_{t-1}]. \tag{2.89}$$

Alternatively, one could reverse the role of the shock and the history by fixing the shock at $\varepsilon_t = \delta$ and consider the GIRF as a random variable in terms of the history Ω_{t-1}. In general, one might compute the GIRF conditional on particular subsets A and B of shocks and histories respectively, that is, $\text{GIRF}_y(h, A, B)$. For example, one might condition on all histories in a particular regime and consider only negative shocks.

One possible use of the GIRF is to assess the significance of asymmetric effects over time. Potter (1994) defines a measure of asymmetric response to a particular shock $\varepsilon_t = \delta$, given a particular history ω_{t-1} as the difference between the GIRF for this particular shock and the GIRF for the shock of the same magnitude but with opposite sign, that is,

$$\text{ASY}_y(h, \delta, \omega_{t-1}) = \text{GIRF}_y(h, \delta, \omega_{t-1}) - \text{GIRF}_y(h, -\delta, \omega_{t-1}). \tag{2.90}$$

Alternatively, one could average across all possible histories to obtain

$$\begin{aligned}\text{ASY}_y(h, \delta) = \text{E}[\text{GIRF}_y(h, \delta, \omega_{t-1})] - \text{E}[\text{GIRF}_y(h, -\delta, \omega_{t-1})] = \\ \text{E}[y_{t+h}|\varepsilon_t = \delta] - \text{E}[y_{t+h}|\varepsilon_t = -\delta].\end{aligned} \tag{2.91}$$

Koop *et al.* (1996) discuss in great detail how the GIRF can be used to examine the persistence of shocks, see also Potter (1995a). It is intuitively clear that if a nonlinear model is stationary, the effect of a particular shock on the time series eventually becomes zero for all possible histories of the process. Hence, $\text{GIRF}_y(h, \delta, \omega_{t-1})$ defined in (2.87) becomes equal to zero as the horizon goes to infinity. From this it follows that the dispersion of the distribution of $\text{GIRF}_y(h, \varepsilon, \Omega_{t-1})$ defined in (2.88) at finite horizons can be interpreted as a measure of persistence of shocks. Conditional versions of the GIRF are particularly suited to assess the persistence of shocks. For example, one might compare the dispersion of the distributions of GIRFs conditional on positive and negative shocks to determine whether negative shocks are more persistent than positive, or vice versa. A potential problem with this approach is that no unambiguous measure of dispersion exists, although, as noted by Koop *et al.* (1996), the notion of second-order stochastic dominance might be useful in the context of GIRFs.

Notice that the second conditional expectation in the right-hand side of (2.87) is the optimal point forecast of y_{t+h} at time $t-1$, whereas the first conditional expectation can be interpreted as the optimal forecast of y_{t+h} at time t in case $\varepsilon_t = \delta$. Therefore the GIRF can be interpreted as the change in forecast of y_{t+h} at time t relative to time $t-1$, given that a shock δ occurs at time t. This also suggests that if the distribution of the conditional GIRF (2.89) (or other versions of the GIRF) effectively is a spike at zero for certain $h \geq m$, the nonlinear model is not useful for forecasting more than m periods ahead.

Because for general nonlinear models analytic expressions for the conditional expectations involved in the GIRF in (2.88) (and subsequent conditional versions) are not available, the Monte Carlo methods discussed in the previous subsection can be used to obtain estimates of the relevant impulse response measure. Koop *et al.* (1996) suggest to use the same realizations of the shocks in intermediate time periods for computing the two components of the GIRF, in order to reduce the Monte Carlo error.

2.7 An empirical illustration

The notion of business cycle asymmetry has been around for quite some time. For example, Keynes (1936, p. 314) already observed that 'the substitution of a downward for an upward tendency often takes place suddenly and violently, whereas there is, as a rule, no such sharp turning point when an upward is substituted for a downward tendency'. Following Burns and Mitchell (1946), conventional wisdom has long held that 'contractions are shorter and more violent than expansions'. Starting with Neftçi (1984), interest in the subject of business cycle asymmetry has revived and over the last fifteen years many macroeconomic variables have been examined for asymmetry - output and unemployment rates in particular. Applications of regime-switching models to output series are discussed in detail in the next chapter. In this section I use a US unemployment rate to illustrate the specification procedure for STAR models discussed in this chapter.

The statistical procedures that have been employed to test for business cycle asymmetry can be divided into two main categories, see also Mittnik and Niu (1994) for a comprehensive overview. First, various nonparametric techniques have been used. For example, Neftçi (1984), Falk (1986), Sichel (1989), Rothman (1991), and McQueen and Thorley (1993), among many others, test for asymmetry between expansions and contractions using Markov chain methods to examine whether the transition probabilities from one regime to the other differ.

In general, the evidence for asymmetry in the unemployment rate has been quite convincing. Neftçi (1984) suggests that increases in the aggregate unemployment rate are steeper than decreases. Sichel (1989) identifies a mistake in Neftçi's analysis, and is not able to reject symmetry with a corrected procedure. Rothman (1991) considers industrial sector unemployment rates and again finds indications of steepness, whereas Neftçi (1993) shows that conventional linear models are able to replicate the observed patterns in the unemployment rate only with very small probability. Escribano and Jordá (1999) also reject linearity for sectoral unemployment rates using tests against STAR-type nonlinearity. Peel and Speight (1996) succesfully estimate SETAR models for (logistically transformed) unemployment rates. Rothman (1998) estimates several nonlinear models for the aggregate unemployment rate and examines their usefulness for (long-term) forecasting and finds that several nonlinear models perform superior to a linear model, see also Montgomery, Zarnowitz, Tsay and Tiao (1998) and Parker and Rothman (1997).

The unemployment rate that I consider in this section represents the seasonally adjusted[6] unemployment rate among US males aged 20 and over[7]. The series is sampled at monthly frequency and covers the period January 1954 until December 1998 (540 observations). The same series is analyzed by Hansen (1997) and Caner and Hansen (1998) using SETAR models, albeit they consider a different sample period and different transformation (see below). The series is shown in Figure 2.8, where circles indicate individual peaks and troughs as dated by the NBER[8].

The cyclical behaviour of the unemployment rate can be characterized as steep increases during recessions, followed by slow(er) declines during expansions. It is also clear that, especially after 1970, the unemployment rate did not return to previous lows during expansions. In fact, this change in level or, more generally, the long-run properties of the unemployment rate have received much more attention than its asymmetric properties. The two competing viewpoints are the 'natural rate' hypothesis and the hysteresis hypothesis of Blanchard and Summers (1987). The natural rate hypothesis states that there exists a unique equilibrium unemployment level[9] and shocks to the unemployment rate only lead to temporary deviations. In contrast, the hysteresis hypothesis states that shocks to the unemployment rate have permanent effects and, hence, equilibrium rates do not exist at all. Put differently, under the natural rate hypothesis, the unemployment rate is mean-reverting, whereas it is non-stationary under the hysteresis hypothesis. Thus, the two hypothe-

[6]The use of seasonally adjusted data might be 'sub-optimal', because seasonal adjustment procedures may influence the nonlinear properties of a time series. The question whether seasonal adjustment actually masks (or even completely removes) or introduces nonlinearity or perhaps changes the kind of nonlinearity in a time series is a subject of ongoing research, see Ghysels, Granger and Siklos (1996) and Franses and Paap (1999).

[7]The series is constructed by taking the ratio of the unemployment level and civilian labor force of this population group. Data have been obtained from the *Bureau of Labor Statistics*.

[8]These peaks and troughs differ from the reference business cycle turning points, as the unemployment rate is, on average, leading at peaks and lagging at troughs.

[9]Although the equilibrium level, or natural rate, may be time-varying, see for example Staiger, Stock and Watson (1997) and other papers in the same issue of the *Journal of Economic Perspectives* for some recent viewpoints.

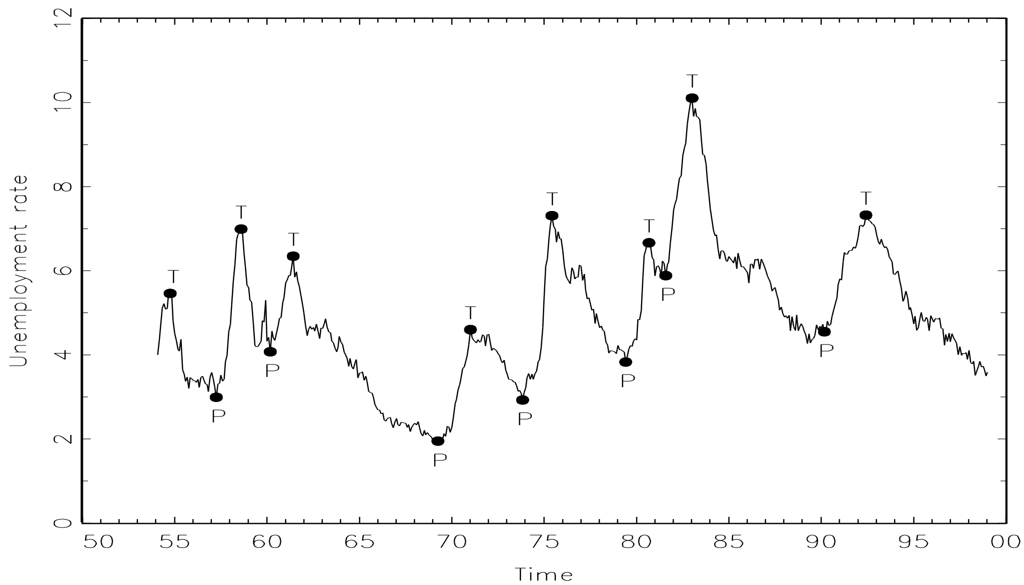


Figure 2.8: Monthly US unemployment rate, males aged 20 and above, January 1954-December 1998. Solid circles indicate NBER-dated unemployment peaks (P) and troughs (T).

ses imply that different transformations (levels and first differences, respectively) of the unemployment rate are appropriate or, in other words, that different methods of detrending should be applied. Obviously, the method of detrending might influence any subsequent analysis, but this point is beyond the scope of this section[10]. Here I use the raw data, and consider models that are similar in spirit to Skalin and Teräsvirta (1998) and Bianchi and Zoega (1998).

I use the first 35 years of data (January 1954 - December 1988) for estimation and testing, and reserve the final 10 years for out-of-sample forecasting. Following the specification procedure as outline in Section 2.1.2, I start with specifying a linear model for the series. Both AIC and SIC indicate that an AR(p) model with $p = 5$ is appropriate, which is estimated as

$$\begin{array}{rllllll} y_t = & \underset{\substack{(0.03)\\ [0.03]}}{0.08} & + \underset{\substack{(0.05)\\ [0.06]}}{1.06}\, y_{t-1} & + \underset{\substack{(0.07)\\ [0.07]}}{0.20}\, y_{t-2} & - \underset{\substack{(0.07)\\ [0.08]}}{0.13}\, y_{t-3} & - \underset{\substack{(0.07)\\ [0.09]}}{0.05}\, y_{t-4} & - \underset{\substack{(0.05)\\ [0.05]}}{0.10}\, y_{t-5} + \hat{\varepsilon}_t, \end{array} \quad (2.92)$$

$\hat{\sigma}_\varepsilon = 0.20$, SK $= -0.08(0.26)$, EK $= 1.41(0.00)$, JB $= 98.64(0.00)$, ARCH(1) $= 3.38(0.07)$, ARCH(4) $= 51.23(0.00)$, LB(8) $= 1.26(1.00)$, LB(12) $= 19.04(0.09)$, AIC $= -3.190$, BIC $= -3.130$,

where OLS and HCC standard errors are given in parentheses and brackets below the parameter estimates, $\hat{\varepsilon}_t$ denotes the regression residual at time t, $\hat{\sigma}_\varepsilon$ is the

[10]See Canova (1994) and Gordon (1997) for more discussion on the influence of detrending procedures on the analysis of the cyclical component in macro-economic time series.

Table 2.4: p-values for LM-type test for STAR nonlinearity for monthly US unemployment rate

Transition	Standard tests				HCC tests			
variable s_t	LM_1	LM_2	LM_3	LM_4	LM_1	LM_2	LM_3	LM_4
$\Delta_1 y_{t-1}$	0.054	0.008	0.087	0.012	0.130	0.144	0.203	0.270
$\Delta_2 y_{t-1}$	0.111	0.014	0.177	0.019	0.191	0.146	0.278	0.259
$\Delta_3 y_{t-1}$	0.061	0.006	0.023	0.002	0.199	0.111	0.120	0.246
$\Delta_4 y_{t-1}$	0.077	0.015	0.076	0.006	0.146	0.162	0.144	0.282
$\Delta_5 y_{t-1}$	0.087	0.001	0.100	0.002	0.192	0.064	0.136	0.176
$\Delta_6 y_{t-1}$	0.100	0.003	0.033	0.019	0.206	0.081	0.101	0.221
$\Delta_7 y_{t-1}$	0.040	0.002	0.005	0.017	0.110	0.075	0.022	0.105
$\Delta_8 y_{t-1}$	0.116	0.006	0.010	0.042	0.174	0.091	0.048	0.203
$\Delta_9 y_{t-1}$	0.036	0.000	0.004	0.005	0.081	0.069	0.017	0.077
$\Delta_{10} y_{t-1}$	0.054	0.000	0.005	0.006	0.127	0.079	0.026	0.124
$\Delta_{11} y_{t-1}$	0.038	0.000	0.002	0.002	0.131	0.041	0.025	0.096
$\Delta_{12} y_{t-1}$	0.028	0.000	0.001	0.001	0.089	0.043	0.020	0.107
t	0.032	0.029	0.057	0.017	0.177	0.296	0.266	0.165

p-values of F variants of the LM-type tests for STAR nonlinearity, applied to monthly US unemployment rates, January 1954-December 1988. The tests are based on an AR(5) model.

residual standard deviation, SK is skewness, EK excess kurtosis, JB the Jarque-Bera test of normality of the residuals, and ARCH is the LM test of no AutoRegressive Conditional Heteroscedasticity [ARCH]. The figures in parentheses following the test statistics are p-values. See Franses (1998) for an explanation of these statistics.

The model appears to show several kinds of shortcomings, as the residuals suffer from excess kurtosis, heteroskedasticity and serial correlation. It appears that these deficiencies can not be fixed by simply increasing the lag order. Thus, I proceed with the next step in the specification procedure and test against STAR nonlinearity, using the LM-type statistics discussed in Section 2.2. As I am concerned with the behaviour of the unemployment rate over the business cycle, I am interested in medium-term movements. The month-to-month unemployment rate exhibits considerable short-term fluctuations, especially in the last months of expansions, as shown by Figure 2.8. This makes the monthly rate unsuitable as an indicator of the business cycle regime, see Birchenhall, Jessen and Osborn (1996) and Neftçi (1984) for more elaborate discussions of this point. For that reason, I concentrate on the use of long differences of the unemployment rate as a potential transition variable[11], that is, $s_t = \Delta_d y_{t-1} \equiv y_{t-1} - y_{t-d-1}$. Table 2.4 displays p-values of the standard and HCC LM-type tests with $\Delta_d y_{t-1}$, $d = 1, \ldots, 12$, as transition variable. LM-type tests against the alternative of smoothly changing parameters, where $s_t = t$, are given as well. The p-values of the standard tests indicate that linearity can be rejected quite convincingly. The HCC test results, however, suggest that the evidence for nonlinearity might be due to neglected heteroskedasticity, especially for transition

[11] The same approach is used by Hansen (1997) and Caner and Hansen (1998).

Table 2.5: p-values for LM-type test for STAR nonlinearity for monthly US unemployment rate

Transition	Teräsvirta			Escribano-Jorda	
variable s_t	LM_{H1}	LM_{H2}	LM_{H3}	LM_{HL}	LM_{HE}
$\Delta_7 y_{t-1}$	0.109	0.013	0.040	0.147	0.342
$\Delta_8 y_{t-1}$	0.065	0.024	0.116	0.380	0.264
$\Delta_9 y_{t-1}$	0.005	0.046	0.036	0.076	0.130
$\Delta_{10} y_{t-1}$	0.002	0.088	0.054	0.112	0.142
$\Delta_{11} y_{t-1}$	0.001	0.090	0.038	0.007	0.238
$\Delta_{12} y_{t-1}$	0.001	0.046	0.028	0.003	0.193

p-values of F variants of the LM-type tests for STAR nonlinearity, applied to monthly US unemployment rates, January 1954-December 1988. The tests are based on an AR(5) model. The hypothesis H_1, H_2, H_3, H_L and H_E are given in Section 2.2.

variables $\Delta_d y_{t-1}$ with $d \leq 6$. For that reason I focus on long-term differences $\Delta_d y_{t-1}$ with $6 < d < 12$. Table 2.5 presents p-values of the various LM-type statistics which test the sub-hypotheses in the specification procedures of Teräsvirta (1994) and Escribano and Jordá (1999) for those transition variables.

Based on the decision rule of the procedure of Teräsvirta (1994) discussed in Section 2.2, the test sequence suggests that an ESTAR model is most appropriate when $\Delta_7 y_{t-1}$ or $\Delta_8 y_{t-1}$ are used as transition variables, whereas an LSTAR model is indicated for the other transition variables. The results from the two statistics used in the Escribano-Jorda procedure do not allow to make a clear decision in case $d = 7$,8 or 10, whereas they confirm the suggestion obtained from the other decision rule in case $d = 9$, 11 or 12.

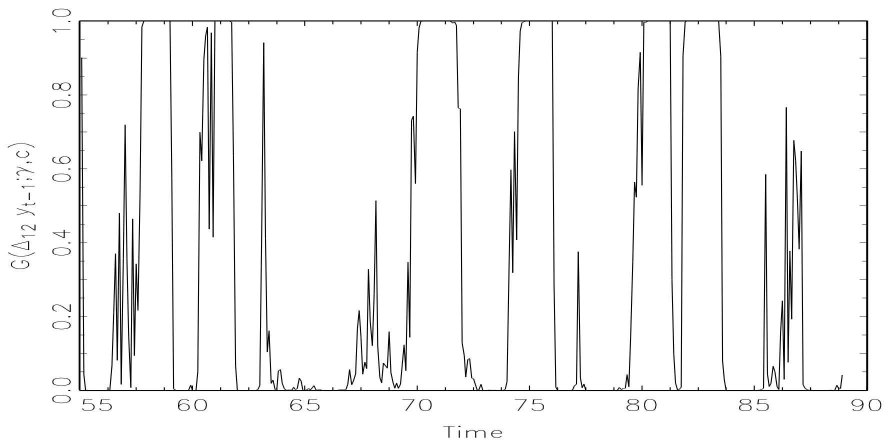
Based on the combined evidence in Tables 2.4 and 2.5 I select the 12-month change in the unemployment rate as transition variable in an LSTAR model. The model is estimated as

$$\begin{aligned} y_t = {} & [\underset{\substack{(0.04)\\ [0.03]}}{0.05} + \underset{\substack{(0.08)\\ [0.11]}}{0.77}\, y_{t-1} + \underset{\substack{(0.09)\\ [0.10]}}{0.29}\, y_{t-2} - \underset{\substack{(0.09)\\ [0.11]}}{0.05}\, y_{t-3} + \underset{\substack{(0.09)\\ [0.11]}}{0.07}\, y_{t-4} - \underset{\substack{(0.07)\\ [0.07]}}{0.10}\, y_{t-5}] \times [1 - G(\Delta_{12} y_{t-1})] \\ & + [-\underset{\substack{(0.06)\\ [0.06]}}{0.15} + \underset{\substack{(0.08)\\ [0.09]}}{1.27}\, y_{t-1} + \underset{\substack{(0.12)\\ [0.12]}}{0.04}\, y_{t-2} - \underset{\substack{(0.12)\\ [0.13]}}{0.19}\, y_{t-3} - \underset{\substack{(0.12)\\ [0.13]}}{0.17}\, y_{t-4} + \underset{\substack{(0.08)\\ [0.08]}}{0.03}\, y_{t-5}] \times G(\Delta_{12} y_{t-1}) \end{aligned} \tag{2.93}$$

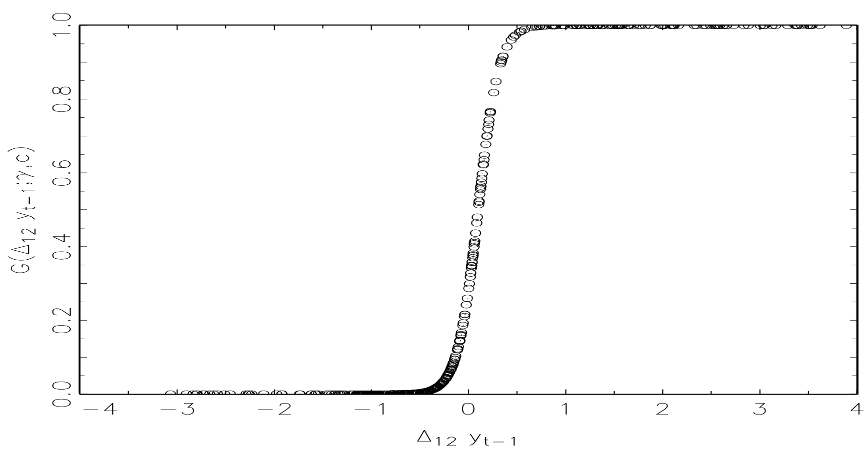
$$G(\Delta_{12} y_{t-1}; \gamma, c) = (1 + \exp[-\underset{(-)}{12.11}(\Delta_{12} y_{t-1} - \underset{\substack{(0.11)\\ [0.11]}}{0.10})/\sigma_{\Delta_{12} y_{t-1}}])^{-1} \tag{2.94}$$

$\hat{\sigma}_\varepsilon = 0.19$, SK $= 0.04(0.36)$, EK $= 1.44(0.00)$, JB $= 35.43(0.00)$, ARCH(1) $= 6.92(0.01)$, ARCH(4) $= 47.58(0.00)$, AIC $= -3.244$, BIC $= -3.106$.

The transition function $G(\Delta_{12} y_{t-1}; \gamma, c)$ is shown in Figure 2.9, both over time and against the transition variable $\Delta_{12} y_{t-1}$. The estimate of the threshold c is not



(a) Transition function versus time



(b) Transition function versus $\Delta_{12}y_{t-1}$

Figure 2.9: Transition function in STAR model for monthly US unemployment rate.

significantly different from 0, indicating that the two regimes in the LSTAR model can be characterized roughly by positive and negative changes in the unemployment rate over the past twelve months. Comparing the top panel of Figure 2.9 with Figure 2.8 shows that the two regimes correspond reasonably close with the contractions and expansions as identified by the NBER turning points. As the transition variable is the change in the unemployment rate over the previous year, the switches between the regimes do not coincide exactly with the peaks and troughs of the unemployment rate; that is, the transition from one regime to the other is not instantaneous at turning points. The estimate of the parameter γ suggests that the transition between the two regimes is fairly rapid. From the bottom panel of Figure 2.9 it is seen that $G(\Delta_{12}y_{t-1};\gamma,c)$ changes from 0 to 1 as $\Delta_{12}y_{t-1}$ changes from -0.5 to 0.5, approximately.

Figure 2.10 shows the negative of the sum of squares function $Q_T(\gamma,c)$ in the neighborhood of the NLS estimate $(\hat{\gamma},\hat{c}) = (12.11, 0.10)$. The negative of Q_T is

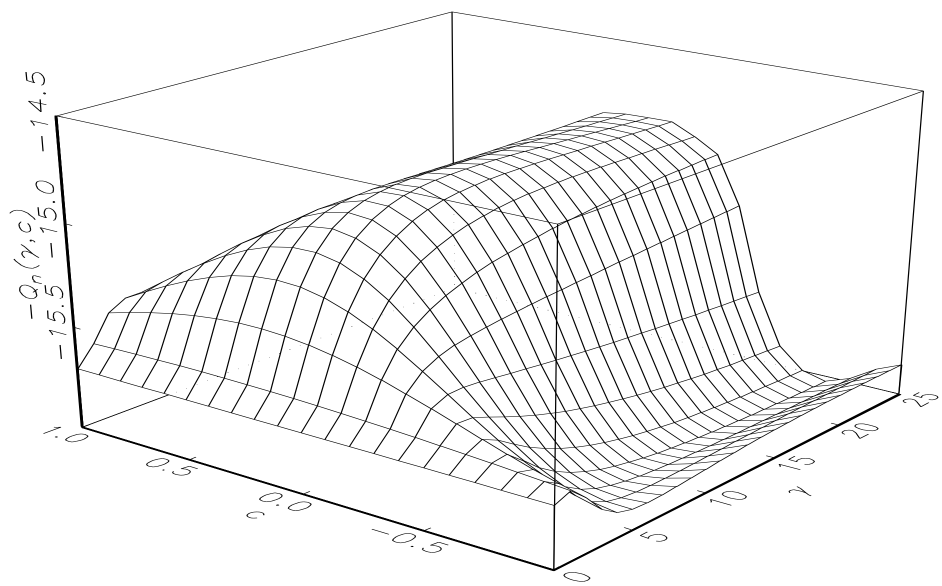


Figure 2.10: Negative of the sum of squares function $Q_T(\gamma, c)$ in the neighborhood of the NLS estimate $(\hat{\gamma}, \hat{c}) = (12.11, 0.10)$.

shown to visualize its properties more clearly. It is seen that for $\gamma \geq 5$, say, the sum of squares function is essentially flat in the direction of γ for fixed values of c. This illustrates the point raised in Section 2.3 concerning the NLS estimate of γ: for large values of γ, the logistic transition function (2.3) does not change shape and, hence, the fit of the model does not change.

The estimated LSTAR model is evaluated further by applying the diagnostic tests for no serial correlation, no remaining nonlinearity, and parameter constancy discussed in Section 2.4. Table 2.6 contains p-values for the various test statistics. The results from the standard tests suggest that the model is not entirely adequate. In particular, there is considerable evidence of serial correlation and remaining nonlinearity. The HCC tests on the other hand again indicate that the apparent misspecification may be due for a large part to neglected heteroskedasticity. Therefore, the model is not refined or extended any further.

Some insight into the properties of the skeleton of the model can be gained by performing deterministic simulation. Starting from a certain point in the history of the time series future shocks are set to zero. Repeating this for all possible histories in the estimation sample should give a reasonably complete picture of the properties of the skeleton. Doing this for the estimated LSTAR model reveals that the model contains an unique limit cycle of 135 months. The cycle is shown in Figure 2.11, starting from December 1988, the last month in the estimation sample. It is seen that the cycle reflects the properties of the series quite well, in the sense that it consists of a relatively short period in which the series increases rapidly, followed by a longer period of gradual decline. The cycle is unique in the sense that the skeleton converges to this cycle irrespective of the starting point[12].

[12]Although, of course, the time required before the cycle is attained depends on the starting point

Table 2.6: Diagnostic tests of STAR model estimated for monthly US unemployment rate

	Tests for q-th order serial correlation					
	Standard tests			HCC tests		
q	4	8	12	4	8	12
p-value	0.097	0.276	0.003	0.177	0.464	0.010
	Tests for parameter constancy					
	Standard tests			HCC tests		
	$\text{LM}_{\text{C},1}$	$\text{LM}_{\text{C},2}$	$\text{LM}_{\text{C},3}$	$\text{LM}_{\text{C},1}$	$L\text{LM}_{\text{C},2}$	$\text{LM}_{\text{C},3}$
p-value	0.083	0.131	0.423	0.478	0.468	0.835
	Tests for remaining nonlinearity					
Transition	Standard tests			HCC tests		
variable s_{2t}	$\text{LM}_{\text{AMR},1}$	$\text{LM}_{\text{AMR},3}$		$\text{LM}_{\text{AMR},1}$	$\text{LM}_{\text{AMR},3}$	
$\Delta_1 y_{t-1}$	0.650	0.181		0.906	0.685	
$\Delta_2 y_{t-1}$	0.790	0.357		0.823	0.216	
$\Delta_3 y_{t-1}$	0.270	0.178		0.454	0.192	
$\Delta_4 y_{t-1}$	0.537	0.201		0.482	0.129	
$\Delta_5 y_{t-1}$	0.436	0.042		0.602	0.242	
$\Delta_6 y_{t-1}$	0.250	0.044		0.508	0.279	
$\Delta_7 y_{t-1}$	0.086	0.026		0.339	0.205	
$\Delta_8 y_{t-1}$	0.036	0.027		0.193	0.147	
$\Delta_9 y_{t-1}$	0.052	0.008		0.120	0.112	
$\Delta_{10} y_{t-1}$	0.056	0.007		0.153	0.123	
$\Delta_{11} y_{t-1}$	0.140	0.022		0.246	0.228	
$\Delta_{12} y_{t-1}$	0.013	0.022		0.079	0.389	

Diagnostic tests for estimated STAR model for monthly US unemployment rates. $\text{LM}_{\text{C},i}$, $i = 1, 2, 3$, denote the LM-type test for parameter constancy based on an i-th order Taylor approximation of the transition function. $\text{LM}_{\text{AMR},1}$ and $\text{LM}_{\text{AMR},3}$ denote the tests for no remaining nonlinearity based upon a first- and third-order Taylor expansion, respectively.

The final 10 years of data, from January 1989 until December 1998, are used to evaluate the forecast performance of the estimated LSTAR model. For each point in this forecast period, I compute 1 to 12-steps ahead forecasts of the unemployment rate from the LSTAR model as given in (2.93)-(2.94) and the AR model given in (2.92). To obtain the forecasts from the LSTAR model I use the bootstrap method outlined in Section 2.5.1. I thus obtain 120 1-step ahead forecasts, 119 2-steps ahead forecasts, etc. The parameters are not updated as new observations become available. Table 2.7 contains several forecast evaluation criteria, based upon the entire forecast period. The most striking figures in this table are the ratios of the MSPE for the LSTAR model relative to the AR model. It is seen that the

of the extrapolation. For example, for histories in the 1980s, with relatively high unemployment rates, it takes about 5 years before the series 'locks into' the cycle. For starting points inside the range of the cycle (3-6%) the cycle occurs much faster.

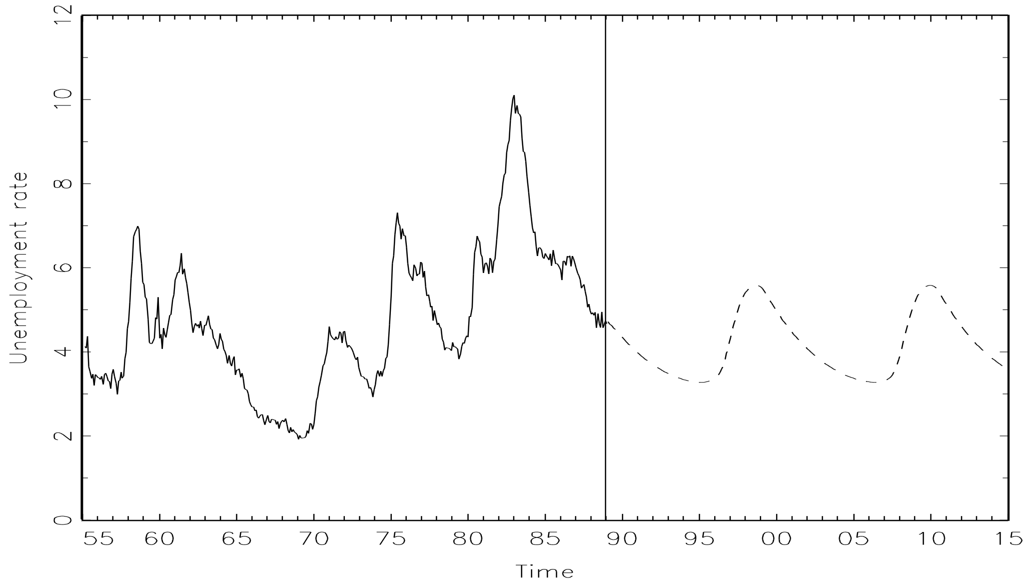


Figure 2.11: Deterministic extrapolation of LSTAR model estimated for monthly US unemployment rate, males aged 20 and above, January 1954-December 1988.

nonlinear model offers improved forecast performance especially at longer forecast horizons, where a reduction of 30% in the MSPE is attained. This is confirmed by the success ratio's SR, which indicate the fraction of observations for which the regime is predicted correctly. At short horizons the percentage of correct regime predictions is approximately the same for both models, whereas they differ considerably at longer horizons. The fact that the SR is close to 1 for very short horizons should not come as a surprise. Given that the regime is determined by the change in the unemployment rate over the last 12 months, the 1-step ahead forecast, for example, has to be very different from the realized value to render an incorrect regime prediction.

Table 2.8 contains the same forecast evaluation criteria, where the forecasts $y_{t+h|t}$ are grouped depending on whether the transition function $G(\Delta_{12}y_{t-1};\gamma,c)$ in (2.94) is smaller or larger than 0.5, that is, conditional upon the regime that is realized at the forecast origin. Interestingly, MPE and MedPE suggest that both models render biased forecasts in both regimes, although the STAR model yields less biased forecasts. When the unemployment rate is declining, that is, during periods of expansions, both models are overly pessimistic and predict too high unemployment rates on average. The opposite occurs when the unemployment rate is increasing, that is, during recessions, when the unemployment rate is consistently under-predicted. The MSPE ratios show that the gains in forecast accuracy for the nonlinear model are larger in the recessionary regime, especially at short and medium-term forecast horizons.

Table 2.7: Unconditional forecast evaluation of AR and LSTAR models for monthly US unemployment rates

Forecast	AR model			LSTAR model			MSPE	
horizon	MPE	MedPE	SR	MPE	MedPE	SR	LSTAR/AR	DM
1	0.00	0.00	–	0.01	0.01	–	0.89	2.58
2	0.01	−0.00	0.97	0.02	0.04	0.98	0.87	2.37
3	0.01	−0.02	0.93	0.04	0.04	0.96	0.79	3.36
4	0.02	−0.02	0.88	0.05	0.03	0.95	0.74	4.11
5	0.03	0.00	0.84	0.06	0.02	0.88	0.72	4.57
6	0.04	−0.03	0.77	0.08	0.04	0.85	0.72	4.84
7	0.06	0.00	0.80	0.09	0.03	0.86	0.70	4.73
8	0.07	0.02	0.78	0.10	0.06	0.83	0.71	4.79
9	0.07	0.00	0.71	0.10	0.03	0.80	0.71	4.84
10	0.08	0.00	0.71	0.11	0.03	0.81	0.70	4.77
11	0.09	−0.04	0.63	0.12	0.03	0.73	0.70	4.86
12	0.10	−0.07	0.59	0.13	0.00	0.69	0.72	4.92

Unconditional forecast evaluation of AR and LSTAR models for monthly US unemployment rates. The forecast period runs from January 1989 until December 1998. MPE = Mean Prediction Error, MedPE = median Prediction Error, SR = success ratio of regime prediction, MSPE = Mean Squared Prediction Error. The column headed DM contains the forecast comparison statistic of Diebold and Mariano (1995) as given in (2.75), based on squared prediction errors.

Table 2.8: Conditional forecast evaluation of AR and LSTAR models for monthly US unemployment rates

Forecast	AR model			LSTAR model			MSPE	
horizon	MPE	MedPE	SR	MPE	MedPE	SR	LSTAR/AR	DM
	Lower regime $G(\Delta_{12}y_{t-1};\gamma,c) < 0.5$ (82 obs.)							
1	−0.01	−0.02	–	0.01	0.01	–	0.84	2.90
2	−0.03	−0.04	0.96	0.01	−0.01	0.99	0.75	3.54
3	−0.04	−0.06	0.94	0.01	0.00	0.95	0.69	3.79
4	−0.05	−0.05	0.90	0.01	0.02	0.96	0.65	3.95
5	−0.07	−0.10	0.90	0.00	−0.02	0.94	0.62	3.81
6	−0.08	−0.10	0.82	0.00	−0.01	0.92	0.62	3.86
7	−0.10	−0.08	0.85	−0.02	−0.02	0.92	0.63	3.71
8	−0.12	−0.11	0.81	−0.03	−0.02	0.91	0.62	3.83
9	−0.15	−0.22	0.75	−0.05	−0.12	0.85	0.66	3.78
10	−0.17	−0.20	0.75	−0.06	−0.11	0.88	0.69	3.30
11	−0.19	−0.28	0.66	−0.08	−0.12	0.73	0.72	3.47
12	−0.20	−0.26	0.63	−0.08	−0.12	0.71	0.72	3.45
	Upper regime $G(\Delta_{12}y_{t-1};\gamma,c) > 0.5$ (38 obs.)							
1	0.03	0.02	–	0.02	0.02	–	1.02	−0.24
2	0.07	0.06	0.97	0.06	0.08	0.97	1.10	−1.16
3	0.12	0.12	0.92	0.10	0.15	0.97	0.97	0.36
4	0.17	0.20	0.85	0.15	0.18	0.92	0.86	1.55
5	0.22	0.24	0.72	0.18	0.13	0.77	0.82	2.53
6	0.28	0.31	0.67	0.23	0.20	0.72	0.78	3.06
7	0.36	0.34	0.69	0.29	0.31	0.74	0.75	3.24
8	0.41	0.39	0.72	0.33	0.41	0.69	0.75	3.35
9	0.48	0.50	0.62	0.38	0.32	0.72	0.73	3.71
10	0.54	0.55	0.64	0.43	0.45	0.69	0.71	3.96
11	0.60	0.70	0.56	0.48	0.49	0.72	0.70	4.03
12	0.64	0.77	0.51	0.51	0.55	0.64	0.72	4.03

Conditional forecast evaluation of AR and LSTAR models for monthly US unemployment rates. The forecast period runs from January 1989 until December 1998. MPE = Mean Prediction Error, MedPE = median Prediction Error, SR = succes ratio for regime prediction, MSPE = Mean Squared Prediction Error. The column headed DM contains the forecast comparison statistic of Diebold and Mariano (1995) as given in (2.75), based on squared prediction errors.

Chapter 3

Multiple-Regime Smooth Transition Models

The smooth transition model as discussed in the previous chapter essentially is a univariate regime-switching model that distinguishes two regimes only. Even though this might be sufficient in many applications, it sometimes can be necessary to consider a more elaborate model to accommodate more flexible dynamic patterns. The second part of this thesis, comprising Chapters 3, 4 and 5, deals with possible extensions of the basic model. The aim of this chapter is to generalize the STAR model to allow for more than two regimes.

A multiple-regime model potentially is useful for describing the behaviour of macro-economic variables in different phases of the business cycle. Most research on business cycle asymmetry has focused on the difference between expansions and contractions, which by now is well documented. Recent evidence suggests that such a two-phase characterization of the business cycle might be too restrictive. Sichel (1993) and Ramsey and Rothman (1996) discuss concepts such as 'deepness', 'steepness' and 'sharpness', which relate to different aspects of asymmetry. A cyclical time series is said to exhibit steepness if the slope of the expansion phase differs from the slope of the contraction phase. The unemployment rate considered in Section 2.7 is an example of a series that exhibits steepness, as the increase during recessions is much faster than the decline during expansions. Deepness occurs when the distance from the mean of the cycle to the peak is not equal to the distance from the mean to the trough. Common belief is that contractions are more 'violent', that is, deeper, than expansions. Sharpness focuses on the relative curvature around peaks and troughs. Peaks generally are thought to be 'rounder' than troughs, see Emery and Koenig (1992) and McQueen and Thorley (1993).

Sichel (1993) argues that most research on business cycle asymmetry has focused exclusively on the possibility of steepness, neglecting other forms of asymmetry. The evidence presented by Sichel (1993) suggests, however, that deepness might be a more important characteristic of macro-economic variables. This is confirmed by the analysis by Verbrugge (1997), which demonstrates that depth is a feature of numerous economic time series, whereas steepness is a feature of (un)employment related variables but is largely absent from real GDP and aggregate industrial production.

Intuitively, if a macro-economic variable exhibits different types of asymmetry simultaneously, the distinction between expansions and contractions might not be sufficient to characterize its behaviour over the business cycle completely. Sichel (1994) observes that real GNP tends to grow faster immediately following a trough than during the rest of the expansion phase. Wynne and Balke (1992) and Emery and Koenig (1992) present additional evidence in favor of this 'bounce-back' effect. This suggests the possibility of three business cycle phases: a contraction phase, a high-growth recovery immediately following the trough of the cycle, and a subsequent moderate growth phase. Similarly, Sichel (1993) shows that the US unemployment rate exhibits both steepness and deepness characteristics, which might also be taken as an indication of the possible existence of multiple regimes.

The outline of this chapter is as follows. In Section 3.1, I argue that the basic STAR model essentially allows for two regimes only, irrespective of the particular transition function that is used. To overcome this limitation, the class of Multiple Regime STAR [MRSTAR] models is introduced. I demonstrate that a multiple-regime STAR [MRSTAR] model can be obtained from the two-regime model in a simple yet elegant way. The main features of the MRSTAR model are illustrated by means of a simple example. In Section 3.2, I outline an empirical specification procedure for MRSTAR models. The discussion in this section is rather sketchy, because the proposed procedure is a straightforward extension of the specification procedure for the two-regime STAR model that was discussed at length in the previous chapter. The main new element is a Lagrange Multiplier-type test statistic that can be used to test a two-regime model against a multiple-regime alternative, which is developed in Section 3.2.1. Simulations are used to examine its empirical performance in Section 3.2.2. In Section 3.3, I discuss previous research on modeling business cycle asymmetry in Us output variables in more detail, emphasizing attempts to refine the basic distinction between recessions and expansions. An application of the MRSTAR model to characterize the behaviour of the growth rate of post-war US real GNP provides evidence in favor of the existence of multiple business cycle phases. Finally, Section 3.4 contains some concluding remarks.

3.1 Representation

Consider again the STAR model for a univariate time series y_t,

$$y_t = \phi_1' x_t (1 - G(s_t; \gamma, c)) + \phi_2' x_t G(s_t; \gamma, c) + \varepsilon_t, \tag{3.1}$$

where $x_t = (1, \tilde{x}_t')'$, $\tilde{x}_t = (y_{t-1}, \ldots, y_{t-p})'$, $\phi_i = (\phi_{i,0}, \phi_{i,1}, \ldots, \phi_{i,p})'$, $i = 1, 2$, and ε_t is a white noise error process with mean zero and variance σ^2. As before, the transition function $G(s_t; \gamma, c)$ is a continuous function bounded between zero and one, whereas the transition variable s_t can be a lagged endogenous variable ($s_t = y_{t-d}$ for certain $d > 0$), an exogenous variable ($s_t = z_t$), or a (possibly nonlinear) function of lagged endogenous and exogenous variables ($s_t = h(\tilde{x}_t)$ for some function $h(\cdot)$ with $\tilde{x}_t = (y_{t-1}, \ldots, y_{t-p}, z_{1t}, \ldots, z_{kt})'$).

The way the model is written in (3.1) highlights the basic characteristic of the STAR model, which is that at any given point in time, the evolution of y_t is determined by a weighted average of two different linear autoregressive [AR] models. The weights assigned to the two models depend on the value taken by the transition function $G(s_t; \gamma, c)$, which in turn depends on the value of the transition variable s_t. As an example, suppose $G(s_t; \gamma, c)$ is taken to be the logistic function

$$G(s_t; \gamma, c) = \frac{1}{1 + \exp\{-\gamma(s_t - c)\}}, \qquad \gamma > 0, \tag{3.2}$$

resulting in the logistic STAR [LSTAR] model. In that case, for small (large) values of s_t, $G(s_t; \gamma, c)$ is approximately equal to zero (one) and, consequently, almost all weight is put on the first (second) model. The parameter γ determines the speed at which these weights change as s_t increases: the higher γ, the faster this change is. If $\gamma \to 0$, the weights become constant (and equal to 0.5) and the model becomes linear, whereas, if $\gamma \to \infty$, the logistic function approaches the indicator function $I[s_t > c]$, taking the value 0 for $s_t < c$ and 1 for $s_t > c$. In that case, the LSTAR model reduces to a two-regime (SE)TAR model.

The LSTAR model seems particularly well suited to describe asymmetry of the type that frequently is encountered in macro-economic time series. For example, the model has been successfully applied by Teräsvirta and Anderson (1992) and Teräsvirta, Tjøstheim and Granger (1994) to characterize the different dynamics of industrial production indexes in a number of OECD countries during expansions and recessions, see also Granger, Teräsvirta and Anderson (1993) and Teräsvirta (1995). In fact, the LSTAR model can capture different types of asymmetry for suitable choices of the transition variable s_t in (3.2). For example, if y_t is the growth rate of an output variable, and if the transition variable is the growth rate in the previous period, $s_t = y_{t-1}$ and $c \approx 0$, the model distinguishes between periods of positive and negative growth, that is, between expansions and contractions, and can describe steepness characteristics. Alternatively, if s_t is a lagged change in the growth rate, $s_t = \Delta y_{t-1}$ with Δ the first difference operator, and again $c \approx 0$, the LSTAR model can capture sharpness, as it allows for different dynamics around peaks (where $\Delta y_{t-1} < 0$) and troughs (where $\Delta y_{t-1} < 0$). Notice however, that the model can handle only one type of asymmetry at a time.

The notation in (3.1) shows that the set of linear AR models of which the STAR model is composed contains only two elements. Hence, it is immediately clear that the STAR model cannot accommodate more than two regimes, irrespective of what form the transition function takes. At first glance, it might be thought that a three-regime model is obtained by using the exponential function

$$G(s_t, \gamma, c) = 1 - \exp\{-\gamma(s_t - c)^2\}, \qquad \gamma > 0, \tag{3.3}$$

as transition function in (3.1). Again, if y_t represents the growth rate of output, if $s_t = y_{t-1}$ and c is close to zero, the exponential STAR [ESTAR] model allows contractions ($y_{t-1} \ll 0$), turning points ($y_{t-1} \approx 0$) and expansions ($y_{t-1} \gg 0$) to have different dynamics, see also Teräsvirta and Anderson (1992). However, this

ESTAR model assumes that the dynamics surrounding peaks and throughs are the same, which runs against the common perception that these are very different (see also the discussion in Section 3.3). The same applies to the dynamics in contractions and expansions. Furthermore, the ESTAR model also consists of two AR models only, so that effectively there still are only two distinct regimes

To obtain a STAR model that allows for more than two genuinely different regimes, it is useful to distinguish two cases, depending on whether the regimes are characterized by a single transition variable s_t or by a combination of several variables $s_{1t}, \ldots, s_{mt}$, say. In case the prevailing regime is determined by a single variable, start with the LSTAR model (3.1) with (3.2), rewritten as

$$y_t = \phi_1' x_t + (\phi_2 - \phi_1)' x_t G_1(s_t; \gamma_1, c_1) + \varepsilon_t, \tag{3.4}$$

where a subscript 1 has been added to the logistic transition function and the parameters contained therein for reasons that will become clear shortly. A three-regime model can be obtained by simply adding a second nonlinear component to give

$$y_t = \phi_1' x_t + (\phi_2 - \phi_1)' x_t G_1(s_t; \gamma_1, c_1) + (\phi_3 - \phi_2)' x_t G_2(s_t; \gamma_2, c_2) + \varepsilon_t. \tag{3.5}$$

If, without loss of generality, it is assumed that $c_1 < c_2$, the autoregressive parameters in this model change smoothly from ϕ_1 via ϕ_2 to ϕ_3 for increasing values of s_t, as first the function G_1 changes from 0 to 1, followed by a similar change of G_2. More generally, one can define a set of $m-1$ smoothness parameters $\gamma_1, \ldots, \gamma_{m-1}$, and a set of $m-1$ thresholds $c_1, \ldots, c_{m-1}$, to arrive at a STAR model with m regimes as

$$\begin{aligned} y_t = \phi_1' x_t + (\phi_2 - \phi_1)' x_t G_1(s_t) + (\phi_3 - \phi_2)' x_t G_2(s_t) + \cdots + \\ (\phi_m - \phi_{m-1})' x_t G_{m-1}(s_t) + \varepsilon_t, \end{aligned} \tag{3.6}$$

where the $G_j(s_t) \equiv G_j(s_t; \gamma_j, c_j), j = 1, \ldots, m-1$, are logistic functions as in (3.2) with smoothness parameter γ_j and threshold c_j.

The 'additive' model as given in (3.5) is very similar to the multiple-regime model that is (implicitly) used in the diagnostic test for no remaining nonlinearity developed by Eitrheim and Teräsvirta (1996) and discussed in Section 2.4. The only difference between the model in (3.5) and the model that is supposed to hold under the alternative hypothesis of this test, as given in (2.43), is that in the latter the transition variables in $G_1(\cdot)$ and $G_2(\cdot)$ are not restricted to be the same. It then is possible that the two transition functions vary more or less independently from each other, thus giving rise to four distinct regimes, each corresponding to a particular combination of extreme values of the transition functions. From (2.43) it also is apparent that the set of linear models comprising this STAR model consists only of three elements - hence, the dynamics in the four regimes are not independent.

An alternative way to extend the basic STAR model in case the regimes are determined by different variables is to build upon the notation used in (3.1). A 4-regime model is obtained by 'encapsulating' two different two-regime LSTAR models as follows:

$$\begin{aligned} y_t = [\phi_1' x_t (1 - G_1(s_{1t}; \gamma_1, c_1)) + \phi_2' x_t G_1(s_{1t}; \gamma_1, c_1)][1 - G_2(s_{2t}; \gamma_2, c_2)] \\ + [\phi_3' x_t (1 - G_1(s_{1t}; \gamma_1, c_1)) + \phi_4' x_t G_1(s_{1t}; \gamma_1, c_1)] G_2(s_{2t}; \gamma_2, c_2) + \varepsilon_t. \end{aligned} \tag{3.7}$$

Clearly, the models in the four regimes associated with the possible combinations of $G_1 = 0, 1$, and $G_2 = 0, 1$, now are independent. For values of G_1 and G_2 between 0 and 1, the effective relationship between y_t and its lagged values is given by a linear combination of those four models. I call the model given in (3.7) the Multiple Regime STAR [MRSTAR] model.

The MRSTAR model as given in (3.7) allows for a maximum of four different regimes, but it will be obvious that by applying the principle of encapsulating repeatedly, the model can be extended straightforwardly to contain 2^m regimes with $m > 2$, at least conceptually. A three-regime model can be obtained from (3.7) by imposing appropriate restrictions on the parameters of the autoregressive models that prevail in the different regimes.

It should be emphasized again that the MRSTAR model as given in (3.7) is only appropriate when the transition variables s_{1t} and s_{2t} are different. If $s_{1t} = s_{2t} \equiv s_t$, that is, if a single variable governs the transitions between all regimes, and if $c_1 < c_2$, G_1 changes from zero to one prior to G_2 for increasing values of s_t and, consequently, the product $(1 - G_1)G_2$ will be equal to zero for almost all values of s_t, especially if γ_1 and γ_2 are large. Hence, it makes sense to exclude the model corresponding to this particular regime by imposing the restriction $\phi_3 = 0$. Because in that case $G_1G_2 \approx G_2$, the resultant model may be rewritten in the additive form given in (3.5).

The MRSTAR model is a very flexible model that can accommodate a wide variety of regime-switching dynamics in y_t. This potential of the MRSTAR can be understood by noting that the model nests the (single hidden layer) artificial neural network [ANN] model as a special case. For example, an ANN is obtained from (3.7) by assuming that the transition variables s_{1t} and s_{2t} are linear combinations of lagged dependent variables, that is, $s_{it} = \alpha_i'\tilde{x}_t$, $i = 1, 2$, and imposing the restrictions $\phi_{i,j} = 0,\ i = 1, \ldots, 4,\ j = 1, \ldots, p$, and $\phi_{4,0} = \phi_{2,0} + \phi_{3,0} - \phi_{1,0}$. The last restriction ensures that the interaction term $\phi_{4,0}^* G_1 G_2$, where $\phi_{4,0}^* = \phi_{1,0} - \phi_{2,0} - \phi_{3,0} + \phi_{4,0}$ drops out of the model, which now can be rewritten as

$$y_t = \phi_{0,0}^* + \phi_{1,0}^* G_1(\alpha_1'\tilde{x}_t; \gamma_1, c_1) + \phi_{2,0}^* G_2(\alpha_2'\tilde{x}_t; \gamma_2, c_2) + \varepsilon_t, \tag{3.8}$$

where $\phi_{0,0}^* = \phi_{1,0}$, $\phi_{1,0}^* = \phi_{2,0} - \phi_{1,0}$ and $\phi_{2,0}^* = \phi_{3,0} - \phi_{1,0}$. It is well-known that by incorporating additional nonlinear components or so-called hidden units $\phi_{i,0}^* G_i(\alpha_i'\tilde{x}_t; \gamma_i, c_i)$, $i = 3, 4, \ldots$, in (3.8), the ANN can approximate any continuous function to any desired degree of accuracy, see Cybenko (1989), Carroll and Dickinson (1989), Funabashi (1989) and Hornik, Stinchcombe and White (1989, 1990). It follows that the same holds true for the MRSTAR model.

An important conceptual difference between the MRSTAR model and ANNs is that the former may be designed explicitly to model the multiple-regime structure in the behaviour of a time series by specifying the transition variables s_{1t} and s_{2t} in (3.7) appropriately. This of course also is possible with the ANN (3.8) by specifying the parameter vectors α_1 in advance, but this is rarely done in practice[1]. In applications of ANNs α_1 and α_2 usually are left unspecified and are parameters to be estimated,

[1]See Eisinga, Franses and van Dijk (1998) for an example.

as the interest is more in approximating the (unknown) relationship between y_t and x_t than in answering the question whether this relationship can be summarized or characterized by a particular (interpretable) regime-structure. See Kuan and White (1994) and Franses and van Dijk (1999, Chapter 5) for introductions to ANNs from an econometric perspective.

Finally, the MRSTAR model reduces to a (SE)TAR model with multiple regimes determined by multiple sources in case the smoothness parameters γ_1 and γ_2 become arbitrarily largely, such that the logistic functions G_1 and G_2 approach indicator functions $I[s_{1t} > c_1]$ and $I[s_{2t} > c_2]$, respectively. The resultant Nested TAR [NeTAR] model is discussed in Astatkie, Watts and Watt (1997). The name *nested* TAR model stems from the fact that the time series y_t can be thought of as being described by a two-regime SETAR model with regimes defined by s_{1t}, and within each of those regimes by a two-regime SETAR model with regimes defined by s_{2t}, or vice versa.

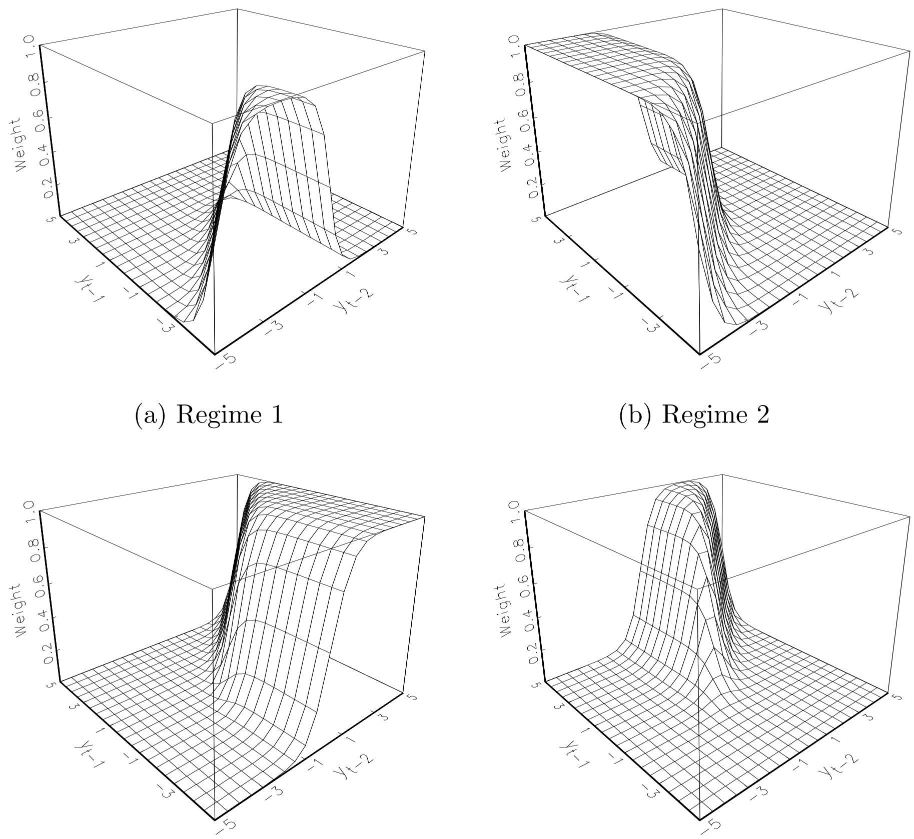
A simple example

To conclude this section, I use a simple example of the four-regime MRSTAR model (3.7) to highlight some features of the model. I set $p = 1$, require all intercepts to be equal to zero, and take $s_{1t} = \Delta y_{t-1}$ and $s_{2t} = y_{t-2}$. The model thus is given by

$$\begin{aligned} y_t = {} & [\phi_1 y_{t-1}(1 - G_1(\Delta y_{t-1})) + \phi_2 y_{t-1} G_1(\Delta y_{t-1})][1 - G_2(y_{t-2})] \\ & + [\phi_3 y_{t-1}(1 - G_1(\Delta y_{t-1})) + \phi_4 y_{t-1} G_1(\Delta y_{t-1})] G_2(y_{t-2}) + \varepsilon_t. \end{aligned} \quad (3.9)$$

For each combination of the transition variables $(\Delta y_{t-1}, y_{t-2})$, the model is a weighted average of the four AR(1) models associated with the four extreme regimes. Figure 3.1 shows the weights given to each of these four models in the (y_{t-1}, y_{t-2}) plane, with $\gamma_1 = \gamma_2 = 2.5$ and $c_1 = c_2 = 0$, where the subscripts of the autoregressive parameters are used to identify the regime number. For $(\Delta y_{t-1}, y_{t-2}) = (0, 0)$ or, equivalently, $(y_{t-1}, y_{t-2}) = (0, 0)$, all models are given equal weight. Along the lines $y_{t-1} = y_{t-2}$ and $y_{t-2} = 0$, which might be interpreted as representing the borders between the different regimes, the models receive equal weight pairwise. For example, along $y_{t-2} = 0$, the models in the first and third regimes receive equal weight, and the same holds for the models in the second and fourth regimes. Moving into a particular regime increases the weight of the corresponding model.

To illustrate the possible dynamics that can be generated by the MRSTAR model, Figure 3.2 shows some time series generated by the sample model (3.9). Two hundred pseudo-random numbers are drawn from the standard normal distribution to obtain a sequence of errors ε_t, while the necessary initial values y_{-1} and y_0 are set equal to zero. The thresholds c_1, c_2 and the parameters γ_1 and γ_2 are set equal to the values given above. In all panels of Figure 3.2, a realization of an AR(1) model $y_t = \phi y_{t-1} + \varepsilon_t$ with autoregressive parameter $\phi = 0.6$, using the same errors ε_t, also is plotted for comparison. In the upper panel of Figure 3.2, the autoregressive parameters are set as $\phi_1 = \phi_2 = 0.3$ and $\phi_3 = \phi_4 = 0.9$. Hence, the model reduces to a two-regime LSTAR model (3.1) with y_{t-2} as transition variable. Although the



(a) Regime 1 (b) Regime 2

(c) Regime 3 (d) Regime 4

Figure 3.1: Weights assigned to AR models in the different regimes in the example MRSTAR model (3.9).

time series generated by the LSTAR model has the same average autoregressive parameter as the linear AR(1) model, the behaviour is markedly different: for positive values of y_{t-2}, the tendency of the series to return to its attractor (which is equal to zero) is much smaller than for negative values of the transition variable. The middle panel of Figure 3.2 shows the AR(1) series together with a realization of the MRSTAR model with $\phi_1 = \phi_3 = 0.3$ and $\phi_2 = \phi_4 = 0.9$. Following Enders and Granger (1998), the resultant model might be called a momentum STAR [M-STAR] model, as the regime is determined by the direction in which the time series is moving, that is, by its momentum. In this example, the memory of the series is longer for upward than for downward movements. The main difference between the AR and M-STAR models occurs in the peaks, the upward (downward) peaks being more (less) pronounced in the nonlinear model. Finally, the lower panel of Figure 3.2 shows the AR(1) series together with a realization of the MRSTAR model (3.9), with the autoregressive parameters taken to be the averages of the parameters in the LSTAR and M-STAR models, that is, ϕ_1 through ϕ_4 are set equal to 0.3, 0.6, 0.6, and 0.9, respectively. Obviously, this time series combines the properties of the LSTAR and M-STAR models: persistence is strongest for positive and increasing values, intermediate for positive and decreasing values and negative and increasing values, and smallest for negative and decreasing values of the time series.

3.2 Specification of MRSTAR models

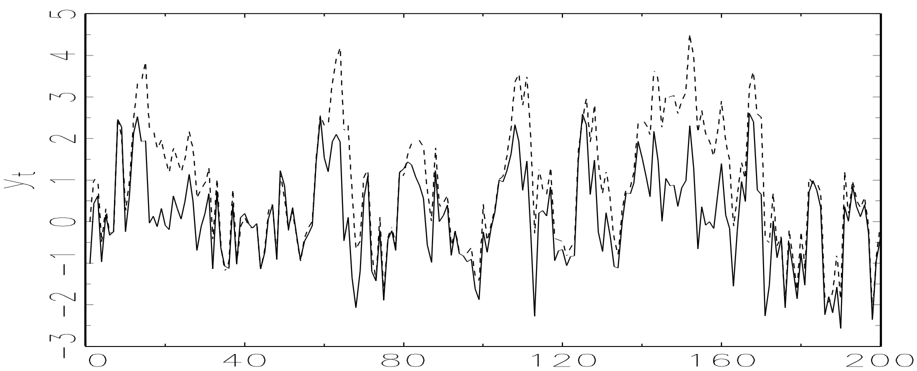
A specific-to-general approach, in which the number of regimes is increased gradually by iterating between testing for the desirability of additional regimes and estimating multiple-regime models, seems the most natural way to specify MRSTAR models[2]. The first part of a possible specification procedure for MRSTAR models then consists of specifying and estimating a two-regime LSTAR model (3.1), using, for example, the approach of Teräsvirta (1994) as outlined in Section 2.1.2. Next, the two-regime model should be tested against the alternative of a general MRSTAR as given in (3.7). In principle, the test for remaining nonlinearity of Eitrheim and Teräsvirta (1996) based upon (2.44) can be used for this purpose, although, as noted before, this statistic actually tests the two-regime model against the additive multiple-regime model as given in (2.43). Recall, however, that such a restricted specification is appropriate only if the transition variables s_{1t} and s_{2t} are in fact the same, as in (3.5). In Section 3.2.1, I therefore derive an alternative LM-type statistic that explicitly tests the null hypothesis of a two-regime model against a multiple-regime alternative of the 'encapsulated' form as given in (3.7). It might be expected that this test has better power properties than the diagnostic test of Eitrheim and Teräsvirta (1996). This is investigated by means of simulation experiments in Section 3.2.2.

If the two-regime model is rejected in favor of the four-regime alternative, one might proceed with estimation of the alternative model by nonlinear least squares [NLS]. Sensible starting values can be obtained by noting that the model is linear once the parameters in the transition functions $G(s_{1t}; \gamma_1, c_1)$ and $G(s_{2t}; \gamma_2, c_2)$ are fixed. Hence, a grid search over γ_1, c_1, γ_2 and c_2 might be performed to obtain starting values for the NLS algorithm. Although this might be quite time consuming given the dimensionality of the grid, it is likely that this will render starting values that are reasonably close to the optimum, hence reducing the burden on the nonlinear optimization considerably.

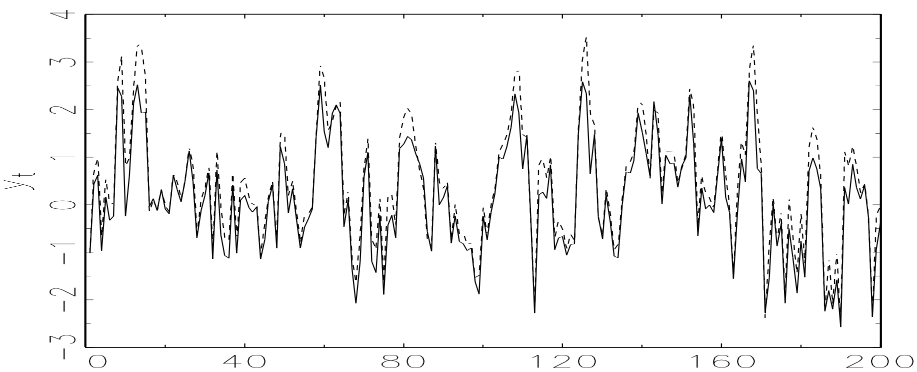
The MRSTAR model as given in (3.7) contains $4(p+1)+2+2$ parameters, which can be quite substantial when the order of the AR models in the different regimes is large. Concentrating the sum of squares function with respect to $\phi_1, \ldots, \phi_4$ as described in Section 2.3 for the two-regime model reduces the dimensionality of the nonlinear optimization problem to 4. Obviously this can be a big advantage for the numerical properties and speed of convergence of the estimates.

The remarks made in Section 2.3 concerning the parameter estimates of STAR models apply to the MRSTAR model as well. In particular, given the difficulty of accurately estimating the smoothness of the transitions between the different regimes, as characterized by γ_1 and γ_2, one is likely to find large standard errors for these parameters. This should not be taken as evidence of insignificance of the

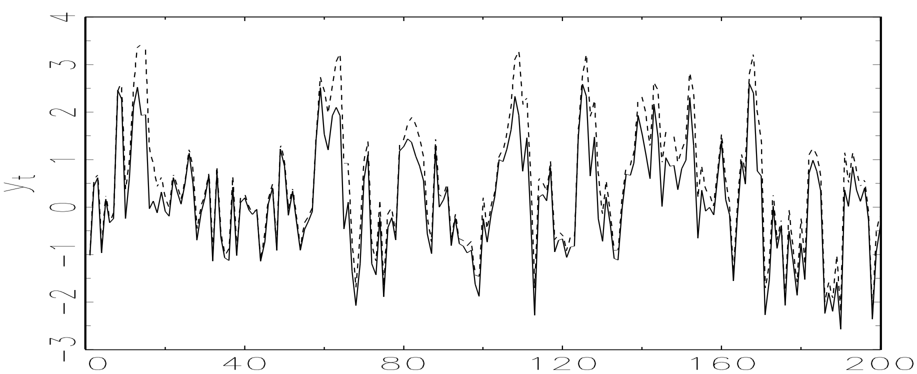
[2] See Teräsvirta and Lin (1993) for a similar approach to determine the appropriate number of hidden units in ANN models.



(a) AR versus STAR



(b) AR versus M-STAR



(c) AR versus MRSTAR

Figure 3.2: The dashed line is a realization from the sample MRSTAR model (3.9) with $\gamma_1 = \gamma_2 = 2.5$, $c_1 = c_2 = 0$, $\varepsilon_t \sim$ i.i.d. $N(0,1)$ and $\phi_1 = \phi_2 = 0.3$ and $\phi_3 = \phi_4 = 0.9$ (panel (a)), or $\phi_1 = \phi_3 = 0.3$ and $\phi_2 = \phi_4 = 0.9$ (panel (b)), or $\phi_1 = 0.3$, $\phi_2 = 0.6$, $\phi_3 = 0.6$, and $\phi_4 = 0.9$ (panel (c)). The solid line is a realization from an AR(1) process with autoregressive parameter 0.6, using the same errors ε_t.

nonlinearity however. One reason for the large standard errors is that the transition between the two regimes is very rapid for large values of γ_1 (γ_2), and therefore a large change in γ_1 (γ_2) has a relatively small impact on the shape of the logistic function. Another reason is that, especially for large values of γ_1 (γ_2), there can be only few observations situated in-between the two regimes.

As far as evaluation of estimated MRSTAR models is concerned, extending the misspecification tests of Eitrheim and Teräsvirta (1996) to obtain diagnostic checks for residual serial correlation, additional nonlinearity and parameter non-constancy in the MRSTAR model is straightforward. In practice, computing the latter two statistics may be complicated due to a lack of degrees of freedom if the time series is short and all parameters are examined simultaneously. This problem may be avoided by testing for remaining nonlinearity or time-variation in a subset of the parameters only.

3.2.1 An LM-type test against MRSTAR

For the purpose of deriving an LM-type test of the two-regime LSTAR model against the MRSTAR alternative it is convenient to rewrite the model in (3.7) as

$$\begin{aligned} y_t = \phi_1^{*\prime} x_t + \phi_2^{*\prime} x_t G_1(s_{1t};\gamma_1,c_1) + \phi_3^{*\prime} x_t G_2(s_{2t};\gamma_2,c_2) \\ + \phi_4^{*\prime} x_t G_1(s_{1t};\gamma_1,c_1) G_2(s_{2t};\gamma_2,c_2) + \varepsilon_t, \quad (3.10) \end{aligned}$$

where $\phi_1^* = \phi_1$, $\phi_2^* = \phi_2 - \phi_1$, $\phi_3^* = \phi_3 - \phi_1$, and $\phi_4^* = \phi_1 - \phi_2 - \phi_3 + \phi_4$. The two-regime model that has been estimated is assumed to have $G_1(\cdot)$ as transition function. Hence, the hypothesis to be tested concerns the question whether the addition of the regimes determined by $G_2(\cdot)$ is appropriate. From (3.7), it is easily seen that the null hypothesis can be expressed as $H_0 : \phi_1 = \phi_3$ and $\phi_2 = \phi_4$ or, in terms of the parameters in (3.10), $H_0 : \phi_3^* = \phi_4^* = 0$. Evidently, this testing problem suffers from the same problem as testing linearity against the alternative of a two-regime STAR model discussed in Section 2.2, in that the alternative model contains nuisance parameters that are not identified under the null hypothesis, in this case γ_2 and c_2. Again the idea of Luukkonen *et al.* (1988) to approximate the transition function with a Taylor expansion can be used to solve this identification problem. Replacing the transition function $G_2(s_{2t};\gamma_2,c_2)$ in (3.10) with a third-order Taylor expansion[3] as given in (2.23) and rearranging terms, the model becomes

$$\begin{aligned} y_t = \theta_1' x_t + \theta_2' x_t G_1(s_{1t};\gamma_1,c_1) + \beta_1' \tilde{x}_t s_{2t} + \beta_2' \tilde{x}_t s_{2t}^2 + \beta_3' \tilde{x}_t s_{2t}^3 \\ + (\beta_4' \tilde{x}_t s_{2t} + \beta_5' \tilde{x}_t s_{2t}^2 + \beta_6' \tilde{x}_t s_{2t}^3) G_1(s_{1t};\gamma_1,c_1) + e_t, \quad (3.11) \end{aligned}$$

where $e_t = \varepsilon_t + R_3(s_{2t};\gamma_2,c_2)$, with $R_3(s_{2t};\gamma_2,c_2)$ the approximation error in the Taylor expansion. In going from (3.10) to (3.11), I have implicitly assumed that s_{2t} is an element of $\tilde{x}_t$. If this is not the case, additional terms $\beta_{i,0} s_{2t}^i$, $i = 1, 2, 3$, and

[3] Because I restrict attention to logistic transition functions, a first-order Taylor expansion would suffice. However, there might be certain alternatives against which the test statistic based on this expansion has very little or no power, see Section 2.2.

$\beta_{i,0}s_{2t}^{i-3}G_1(s_{1t};\gamma_1,c_1)$, $i=4,5,6$, should be added to (3.11). The parameter vectors $\beta_i=(\beta_{i,1},\ldots,\beta_{i,p})'$, $i=1,\ldots,6$, in (3.11) are defined in terms of ϕ_i^*, $i=1,\ldots,4$, γ_2 and c_2, such that the null hypothesis can be reformulated as $H_0':\beta_i=0$, $i=1,\ldots,6$. Note that, under the null hypothesis, $\theta_1=\phi_1^*=\phi_1$, $\theta_2=\phi_2^*=\phi_2-\phi_1$, and $e_t=\varepsilon_t$. Assuming the errors to be normally distributed, it follows that the conditional log-likelihood for observation t is given by

$$l_t=-\frac{1}{2}\ln 2\pi-\frac{1}{2}\ln\sigma^2-\frac{e_t^2}{2\sigma^2}. \tag{3.12}$$

Because the information matrix is block diagonal, the error variance σ^2 can be assumed to be fixed. The remaining partial derivatives evaluated under the null hypothesis are given by

$$\left.\frac{\partial l_t}{\partial\theta_1}\right|_{H_0'}=\frac{1}{\sigma^2}\hat{\varepsilon}_t x_t, \tag{3.13}$$

$$\left.\frac{\partial l_t}{\partial\theta_2}\right|_{H_0'}=\frac{1}{\sigma^2}\hat{\varepsilon}_t x_t G_1(s_{1t};\hat{\gamma}_1,\hat{c}_1), \tag{3.14}$$

$$\left.\frac{\partial l_t}{\partial\beta_i}\right|_{H_0'}=\frac{1}{\sigma^2}\hat{\varepsilon}_t\tilde{x}_t s_{2t}^i, \qquad i=1,2,3, \tag{3.15}$$

$$\left.\frac{\partial l_t}{\partial\beta_i}\right|_{H_0'}=\frac{1}{\sigma^2}\hat{\varepsilon}_t\tilde{x}_t G_1(s_{1t};\hat{\gamma}_1,\hat{c}_1)s_{2t}^{i-3}, \qquad i=4,5,6, \tag{3.16}$$

$$\left.\frac{\partial l_t}{\partial\gamma_1}\right|_{H_0'}=\frac{1}{\sigma^2}\hat{\varepsilon}_t\hat{\theta}_2' x_t\frac{\partial G_1(s_{1t};\hat{\gamma}_1,\hat{c}_1)}{\partial\gamma_1}, \tag{3.17}$$

$$\left.\frac{\partial l_t}{\partial c_1}\right|_{H_0'}=\frac{1}{\sigma^2}\hat{\varepsilon}_t\hat{\theta}_2' x_t\frac{\partial G_1(s_{1t};\hat{\gamma}_1,\hat{c}_1)}{\partial c_1}, \tag{3.18}$$

where $\hat{\varepsilon}_t$ are the residuals obtained from the two-regime LSTAR model under the null hypothesis, and

$$\begin{aligned}\frac{\partial G_1(s_{1t};\hat{\gamma}_1,\hat{c}_1)}{\partial\gamma_1}&=(1+\exp\{-\hat{\gamma}_1(s_{1t}-\hat{c}_1)\})^{-2}\exp\{-\hat{\gamma}_1(s_{1t}-\hat{c}_1)\}(s_{1t}-\hat{c}_1)\\&=G_1(s_{1t};\hat{\gamma}_1,\hat{c}_1)(1-G_1(s_{1t};\hat{\gamma}_1,\hat{c}_1))(s_{1t}-\hat{c}_1),\end{aligned} \tag{3.19}$$

$$\begin{aligned}\frac{\partial G_1(s_{1t};\hat{\gamma}_1,\hat{c}_1)}{\partial c_1}&=\hat{\gamma}_1(1+\exp\{-\hat{\gamma}_1(s_{1t}-\hat{c}_1)\})^{-2}\exp\{-\hat{\gamma}_1(s_{1t}-\hat{c}_1)\}\\&=\hat{\gamma}_1 G_1(s_{1t};\hat{\gamma}_1,\hat{c}_1)(1-G_1(s_{1t};\hat{\gamma}_1,\hat{c}_1)).\end{aligned} \tag{3.20}$$

The partial derivatives (3.19) and (3.20) are denoted as $\widehat{G}_{\gamma_1}(s_{1t})$ and $\widehat{G}_{c_1}(s_{1t})$, respectively. I also use the shorthand notation $\widehat{G}_1(s_{1t})$ to denote $G_1(s_{1t};\hat{\gamma}_1,\hat{c}_1)$.

The LM-type test statistic to test H_0' is given by LM $=\hat{l}_\beta'\text{cov}(\hat{l}_\beta)^{-1}\hat{l}_\beta$, where $l_\beta=(l_{\beta,1},\ldots,l_{\beta,T})'$ and $l_{\beta,t}=(\partial l_t/\partial\beta_1,\ldots,\partial l_t/\partial\beta_6)'$ and hats indicate that all elements of l_β should be evaluated under the null hypothesis. The expressions for the partial derivatives of the log-likelihood given above suggest that the statistic can be computed in a few steps as follows:

1. Estimate the two-regime LSTAR model (3.1) with (3.2) by nonlinear least squares, obtain the residuals $\hat{\varepsilon}_t \equiv y_t - \hat{\phi}_1 x_t(1 - \widehat{G}_1(s_{1t})) - \hat{\phi}_2 x_t \widehat{G}_1(s_{1t})$, and compute the sum of squared residuals under the null hypothesis, $\text{SSR}_0 = \sum_{t=1}^{T} \hat{\varepsilon}_t^2$.

2. Regress the residuals $\hat{\varepsilon}_t$ on x_t, $x_t\widehat{G}_1(s_{1t}), \hat{\theta}_2' x_t \widehat{G}_{\gamma_1}(s_{1t})$, $\hat{\theta}_2' x_t \widehat{G}_{c_1}(s_{1t})$ and the auxiliary regressors $\tilde{x}_t s_{2t}^i$ and $\tilde{x}_t \widehat{G}_1(s_{1t}) s_{2t}^i$, $i = 1, 2, 3$, and compute the sum of squared residuals under the alternative, SSR_1.

3. Compute the LM-type test statistic as

$$\text{LM}_{\text{EMR}} = \frac{(\text{SSR}_0 - \text{SSR}_1)/6p}{\text{SSR}_1/(T - 6p - 2(p+1))}, \tag{3.21}$$

   where T denotes the sample size.

In step 2, the estimates of the autoregressive parameters in the LSTAR model are used to obtain an estimate of θ_2, that is, $\hat{\theta}_2 = \hat{\phi}_2 - \hat{\phi}_1$, which is consistent under the null hypothesis. Under the null hypothesis, the statistic LM_{EMR} is F distributed with $6p$ and $T - 6p - 2(p+1)$ degrees of freedom. As usual, the F version of the test statistic is preferable to the χ^2 variant in small samples because its size and power properties are better.

The remarks made by Eitrheim and Teräsvirta (1996) concerning potential numerical problems are relevant for the test in (3.21) as well. If $\hat{\gamma}_1$ is very large, such that the transition between the two regimes in the model under the null hypothesis is fast, the partial derivatives of the transition function G_1 with respect to γ_1 and c_1, as given in (3.19) and (3.20), approach zero functions (except for $\widehat{G}_{c_1}(s_{1t})$ at the point $s_{1t} = \hat{c}_1$). Hence, the moment matrix of the regressors in the auxiliary regression becomes near-singular. However, because the terms in the auxiliary regression involving these partial derivatives are likely to be very small for all $t = 1, \ldots, T$, they contain very little information. It is therefore suggested that these terms simply be omitted under such circumstances, as this will not harm the power properties of the the test statistic. Furthermore, the residuals $\hat{\varepsilon}_t$ obtained from estimating the two-regime LSTAR model may not be exactly orthogonal to the gradient matrix, due to the fact that the two-regime model is estimated with a numerical optimization algorithm[4]. Following Eitrheim and Teräsvirta (1996), I suggest accounting for this by performing the following additional step in calculating the test statistic

1' Regress $\hat{\varepsilon}_t$ on x_t and $x_t\widehat{G}_1(s_{1t})$ (and $\hat{\theta}_2' x_t \widehat{G}_{\gamma_1}(s_{1t})$ and $\hat{\theta}_2' x_t \widehat{G}_{c_1}(s_{1t})$ if these terms are not excluded), compute the residuals $\tilde{\varepsilon}_t$ from this regression, and the residual sum of squares $\text{SSR}_0 = \sum_{t=1}^{n} \tilde{\varepsilon}_t^2$.

The residuals $\tilde{\varepsilon}_t$ instead of $\hat{\varepsilon}_t$ then should be used in steps 2 and 3.

The LM-type test derived above is a generalization of the diagnostic test of Eitrheim and Teräsvirta (1996) against time-varying coefficients, in which s_{2t} is taken equal to time, $s_{2t} = t$, see Section 2.4. Obviously the test statistic (3.21) also can be interpreted and used as a diagnostic tool to evaluate estimated two-regime STAR models.

[4] This may also result from omitting the terms involving $\widehat{G}_{\gamma_1}(s_{1t})$ and $\widehat{G}_{c_1}(s_{1t})$.

3.2.2 Small sample properties

Before I turn to an empirical application of the MRSTAR models and the specification procedure discussed above, I evaluate the small sample properties of the LM-type test (3.21) by means of a Monte Carlo experiment.

To investigate the size of the $\mathrm{LM}_{\mathrm{EMR}}$ test, a two-regime LSTAR model (3.1)-(3.2) is used as data-generating process [DGP], with $p = 1$, $\phi_{1,0} = \phi_{2,0} = 0$, $s_t = y_{t-1}$, $\gamma = 2.5$, $c = 0$, and the errors ε_t standard normally distributed. The procedure that is followed in the simulation experiments mimics the setup of Eitrheim and Teräsvirta (1996). Each replication is subjected first to the LM_3 statistic based upon (2.24), which is used in the specification procedure for STAR models of Teräsvirta (1994), assuming that the true order of the model and the transition variable are known. The series is retained only if the null hypothesis is rejected at the 5% level of significance. The reason for doing this is to avoid estimating a STAR model on series in which very little or no evidence of nonlinearity is present. If the series is not discarded, a two-regime LSTAR model is estimated and, if the estimation algorithm converges, the $\mathrm{LM}_{\mathrm{EMR}}$ test statistic is computed as discussed above for $s_{2t} = y_{t-1}$, y_{t-2} and Δy_{t-1}. In computing the test statistic, the terms involving $\widehat{G}_{\gamma_1}(s_{1t})$ and $\widehat{G}_{c_1}(s_{1t})$ are always omitted, and the orthogonalization step 1' is always performed. I fix the total number of accepted replications at 1000 for all DGPs and consider series of $T = 200$ observations. The choice for this particular sample size is motivated by the length of the empirical time series on US GNP in Section 3.3. In all experiments, necessary starting values of the time series are set equal to zero. To eliminate possible dependencies of the results on this initialization, the first 100 observations of each series are discarded.

Table 3.1 shows the empirical size at 1, 5, and 10% nominal significance levels, using critical values from the appropriate F-distribution. It is seen that for all combinations of $\phi_{1,1}$ and $\phi_{2,1}$ that are considered, the empirical size of the $\mathrm{LM}_{\mathrm{EMR}}$ test statistic is below its nominal size. Especially if $s_{2t} = y_{t-1}$, which is the transition variable in the estimated LSTAR model, the test is very conservative. Unreported results for the $\mathrm{LM}_{\mathrm{EMR}}$ test statistic based on a first-order Taylor approximation of the transition function G_2 and the $\mathrm{LM}_{\mathrm{AMR}}$ test for no remaining (additive) nonlinearity of Eitrheim and Teräsvirta (1996) demonstrate that these tests suffer from the same problem.

The power properties of the $\mathrm{LM}_{\mathrm{EMR}}$ statistic are investigated in two different ways. First, I use a two-regime ESTAR model (3.1) with transition function

$$G(s_t;\gamma,c) = (1+\exp\{-\gamma(s_t-c_1)(s_t-c_2)\})^{-1}, \qquad c_1 \le c_2, \gamma > 0, \tag{3.22}$$

as DGP, with p, ϕ_1, ϕ_2 and s_{1t} as above, $\gamma = 10$, $c_1 = -1$, $c_2 = 1$, and ε_t again standard normally distributed. For replications for which linearity is rejected by the LM_3 statistic at the 5% significance level, I erroneously fit an LSTAR model to the series and, upon normal convergence of the estimation algorithm, apply the $\mathrm{LM}_{\mathrm{EMR}}$ test for the same choices of s_{2t} as above. Second, I use the sample MRSTAR model (3.9) as DGP, with $\gamma_1 = \gamma_2 = 2.5$ and $c_1 = c_2 = 0$ and several combinations of the AR parameters $\phi_1, \ldots, \phi_4$. Only series for which the LM_3 statistic rejects the

Table 3.1: Empirical size of LM_{EMR} test for MRSTAR nonlinearity

			Transition variable s_{2t}								
			y_{t-1}			y_{t-2}			Δy_{t-1}		
$\phi_{1,1}$	$\phi_{2,1}$	α	0.010	0.050	0.100	0.010	0.050	0.100	0.010	0.050	0.100
−0.5	−0.9		0.006	0.027	0.053	0.008	0.029	0.064	0.005	0.027	0.049
	0.0		0.006	0.032	0.059	0.004	0.044	0.101	0.004	0.019	0.053
	0.4		0.001	0.018	0.051	0.008	0.038	0.084	0.008	0.034	0.081
	0.9		0.007	0.024	0.037	0.012	0.044	0.082	0.007	0.038	0.086
0.5	−0.9		0.002	0.013	0.030	0.008	0.026	0.061	0.008	0.027	0.061
	−0.5		0.003	0.014	0.036	0.006	0.033	0.072	0.003	0.031	0.068
	0.0		0.002	0.024	0.058	0.007	0.038	0.095	0.006	0.038	0.083
	0.9		0.006	0.029	0.055	0.012	0.030	0.073	0.012	0.046	0.090

Empirical size of the LM_{EMR} test (3.21) of no remaining STAR-type nonlinearity at nominal significance levels $\alpha = 0.010$, 0.050, and 0.100, for series generated by the two-regime LSTAR model (3.1) with (3.2) with $\phi_{1,0} = \phi_{2,0} = 0$, $\gamma = 2.5$, $c = 0$, and $\varepsilon_t \sim$ i.i.d. $N(0,1)$. The table is based on 1000 replications for sample size $T = 200$.

null hypothesis at the 5% nominal significance level both when Δy_{t-1} and y_{t-2} are used as transition variable are retained. For these series, two different two-regime LSTAR models are estimated, with Δy_{t-1} and y_{t-2} as transition variables. The LM_{EMR} statistic and the LM_{AMR} test against the additive multiple-regime model are applied as diagnostic checks to test for remaining nonlinearity.

The results for the experiments with an ESTAR model as DGP are displayed in Table 3.2. It is seen that the power of the LM_{EMR} test against this alternative is reasonably good, provided that the nonlinearity is fairly strong, that is, if $\phi_{1,1}$ and $\phi_{2,1}$ are not too close. Also note that power is highest when y_{t-1} also is assumed to be the transition variable in the second transition function $G_2(s_{2t})$. As discussed before, strictly speaking the MRSTAR model is not appropriate in this case, as one of the four regimes in the model will not be realized. Still, the LM_{EMR} test can be applied as diagnostic test against this alternative, and the results in Table 3.2 suggest that it is quite useful in this respect.

The results for the experiments with the MRSTAR model (3.9) as DGP are shown in Table 3.3. The entries in this table show that the the LM_{EMR} test compares favorably with the LM_{AMR} test of Eitrheim and Teräsvirta (1996). This might have been expected of course as the LM_{EMR} was designed explicitly against the alternative of an MRSTAR model.

3.3 Multiple regimes in the business cycle?

Business cycle asymmetry has been investigated mainly by examining US output series, such as real GDP, GNP and industrial production, and US (un)employment series. I follow this practice here and explore whether multiple regimes in the dynamic behaviour of US real GNP over the business cycle can be identified and described by an MRSTAR model.

Previous studies applying statistical tests for asymmetry to US real GNP have

Table 3.2: Empirical power of LM_{EMR} test for MRSTAR nonlinearity

			Transition variable s_{2t}								
			y_{t-1}			y_{t-2}			Δy_{t-1}		
$\phi_{1,1}$	$\phi_{2,1}$	α	0.010	0.050	0.100	0.010	0.050	0.100	0.010	0.050	0.100
−0.5	−0.9		0.035	0.114	0.200	0.017	0.067	0.133	0.019	0.093	0.171
	0.0		0.133	0.341	0.467	0.012	0.044	0.087	0.038	0.151	0.241
	0.4		0.502	0.749	0.851	0.005	0.045	0.088	0.057	0.184	0.289
	0.9		0.731	0.867	0.919	0.103	0.263	0.385	0.020	0.077	0.149
0.5	−0.9		0.673	0.838	0.887	0.163	0.346	0.464	0.386	0.598	0.700
	−0.5		0.654	0.859	0.926	0.020	0.078	0.133	0.200	0.428	0.567
	0.0		0.133	0.307	0.430	0.009	0.038	0.097	0.034	0.116	0.202
	0.9		0.025	0.095	0.176	0.012	0.052	0.099	0.009	0.050	0.083

Empirical power of the LM_{EMR} test (3.21) of no remaining STAR-type nonlinearity at nominal significance levels $\alpha = 0.010$, 0.050, and 0.100, when series are generated according to the two-regime ESTAR model (3.1) with (3.22), with $\phi_{1,0} = \phi_{2,0} = 0$, $\gamma = 10$, $c_1 = -1$, $c_2 = 1$, and $\varepsilon_t \sim \text{i.i.d.}N(0,1)$, but an LSTAR model is erroneously fitted to the data. The table is based on 1000 replications for sample size $T = 200$.

provided mixed results. In particular, the evidence obtained from nonparametric procedures has not been very compelling. For example, Falk (1986) cannot reject symmetry when examining US real GNP for steepness, see also DeLong and Summers (1986) and Sichel (1993). Similarly, Brock and Sayers (1988) only marginally reject linearity, whereas Sichel (1993) finds only moderate evidence for deepness. An exception to the rule is Brunner (1992), who obtains fairly strong indications for asymmetry in GNP associated with an increase in variance during contractions. This is confirmed by Emery and Koenig (1992) who suggest that the variance of leading and coincident indices increases as contractions proceed.

The application of regime-switching models to GDP and GNP series has been more successful. Hamilton (1989) and Durland and McCurdy (1994), for example, find that a two-state Markov Switching model for the growth rate of post-war quarterly US real GNP offers a better description of the dynamic properties of GNP than linear models, see also Diebold and Rudebusch (1996). Boldin (1996) examines the stability of this model and demonstrates that the model is not robust to extension of the sample period. Tiao and Tsay (1994), Potter (1995b) and Clements and Krolzig (1998) all estimate a two-regime SETAR model consisting of AR(2) models (although Potter (1995b) adds an additional fifth lag). The growth rate two periods lagged is used as the transition variable, and the threshold is either fixed at zero (Potter (1995b)) or estimated to be equal to or close to zero (Tiao and Tsay (1994), Clements and Krolzig (1998)). Hence, a distinction is made between periods of positive and negative growth.

A common feature of most of the estimated models is that the dynamics in contractions are very different from those during expansions. In particular, the SETAR models of Tiao and Tsay (1994), Potter (1995b) and Clements and Krolzig (1998), which are estimated on data from 1948 until 1990, all contain a large negative coefficient on the second lag in the contraction regime, suggesting that US GNP moves quickly out of recessions. Notably, Clements and Krolzig (1998) find much

Table 3.3: Empirical power of LM_{EMR} and LM_{AMR} tests against MRSTAR nonlinearity

					Test with transition variable s_{2t}			
					LM_{EMR}		LM_{AMR}	
s_{1t}	ϕ_1	ϕ_2	ϕ_3	ϕ_4	y_{t-2}	Δy_{t-1}	y_{t-2}	Δy_{t-1}
Δy_{t-1}	−.7	.1	.1	.9	0.196	0.007	0.186	0.006
		−.4	.6		0.790	0.006	0.111	0.005
		.6	−.4		0.681	0.011	0.134	0.007
	−.3	.3	.3	.9	0.608	0.004	0.594	0.002
		.0	.6		0.773	0.007	0.630	0.006
		.6	.0		0.517	0.008	0.524	0.010
	.1	.5	.5	.9	0.966	0.000	0.975	0.001
		.3	.7		0.982	0.006	0.985	0.003
		.7	.3		0.947	0.002	0.964	0.000
y_{t-2}	−.7	.1	.1	.9	0.022	0.145	0.066	0.103
		−.4	.6		0.015	0.860	0.027	0.031
		.6	−.4		0.019	0.714	0.034	0.057
	−.3	.3	.3	.9	0.023	0.098	0.037	0.042
		.0	.6		0.022	0.324	0.038	0.027
		.6	.0		0.045	0.128	0.081	0.127
	.1	.5	.5	.9	0.022	0.031	0.043	0.030
		.3	.7		0.062	0.063	0.081	0.044
		.7	.3		0.013	0.020	0.014	0.035

Empirical power of the LM_{EMR} test (3.21) and the LM_{AMR} test based on (2.44) at 5% nominal significance level when series are generated according to the MRSTAR model (3.9) with G_1 and G_2 both equal to logistic functions (3.2) with $\gamma_1 = \gamma_2 = 2.5$, $c_1 = c_2 = 0$, and $\varepsilon_t \sim$ i.i.d. $N(0,1)$. A two-regime LSTAR model with transitions variable s_{1t} is fitted to the data, and the tests for no remaining nonlinearity are applied with transition variables s_{2t} in the additional transition function. The table is based on 1000 replications for sample size $T = 200$.

less evidence of this property when they re-estimate their model on a recent vintage of data ranging from 1960 until 1996.

Whereas most attention has been focused on the distinction between contractions and expansions, some indications for the existence of multiple regimes in output have been obtained as well. For example, Sichel (1994) demonstrates that growth in real GDP is larger immediately following a business cycle trough than during later parts of the expansion, suggesting that it might be worthwhile to decompose the expansion phase in a high-growth phase immediately following the trough of a cycle, and a subsequent moderate-growth phase. Wynne and Balke (1992) and Balke and Wynne (1996) document similar 'peak-reverting' behaviour in industrial production. Furthermore, they examine the relationship between growth during the first 12 months following a trough and the severity of the preceding contraction and show that deep recessions generally are followed by strong recoveries. Emery and Koenig (1992) also find that the mean growth rate in leading and coincident indexes is larger (in absolute value) in early (late) stages of the expansion (contraction). Finally, Cooper (1998) finds very strong evidence for the existence of multiple regimes in industrial production series using a regression tree approach.

The idea of a strong-recovery regime can be traced back to the 'plucking model' of business fluctuations of Friedman (1969, 1993), see also Goodwin and Sweeney (1993). Recently, several attempts have been made to capture the existence of such a regime by means of regime-switching models. For example, Boldin (1996) presents a three-regime MS-AR model in which the expansion regime is split into separate regimes for the post-trough rapid recovery period and the moderate growth period for the remainder of the expansion. In a similar vein, Koop *et al.* (1996) and Pesaran and Potter (1997) use SETAR-type models to construct a 'floor and ceiling' model that allows for three regimes, corresponding to low, normal, and high growth rates of output, respectively. Beaudry and Koop (1993) estimate a linear AR model in which the 'current depth of recession', which measures deviations from the historical maximum in the level of real GNP, is added as regressor. This variable is discussed in more detail below. Even though this CDR model is not a regime-switching model as such[5], it does capture the idea that the distance of the level of GNP to the historic maximum, which closely corresponds with the deepness of a recession, affects current and future growth. Finally, Tiao and Tsay (1994) develop a four-regime SETAR model for US real GNP in which the regimes are labeled worsening/improving recession/expansion.

Compared to the previous studies mentioned above, I use a relatively long span of quarterly observations on US real GNP, which ranges from 1947:1 to 1995:2. The data, which are at 1987 prices, are seasonally adjusted and are taken from Citibase. The corresponding growth rate, denoted y_t, is shown in the upper panel of Figure 3.3. The solid circles indicate NBER-dated peaks and troughs, which are marked with 'P' and 'T', respectively. The bar chart in the lower panel of Figure 3.3 shows the mean growth rates during contractions and different phases of expansions as identified by the NBER turning points, compare Sichel (1994, Figure 1). It is seen

[5] Although it is possible to interpret the CDR model as a SETAR model, see Pesaran and Potter (1997).

that in the first four quarters following a trough, growth is considerably higher than during the rest of the expansion, thus suggesting that a high-growth recovery phase indeed might be present.

Based on the values of AIC and SIC, defined in (2.12) and (2.13), respectively, and the Ljung-Box test for residual autocorrelation given in (2.14), an AR(2) model is found to offer a reasonable description of the linear properties of the time series. The estimated model over the period 1947:4-1995:2 is

$$y_t = \underset{(0.091)}{0.430} + \underset{(0.073)}{0.345}\, y_{t-1} + \underset{(0.073)}{0.095}\, y_{t-2} + \hat{\varepsilon}_t, \tag{3.23}$$

$\hat{\sigma}_\varepsilon = 0.917$, SK $= 0.01(0.48)$, EK $= 1.40(0.00)$, JB $= 15.58(0.00)$, ARCH(1) $= 3.03(0.08)$, ARCH(4) $= 9.27(0.06)$, LB(8) $= 5.05(0.41)$, LB(12) $= 14.00(0.12)$, AIC $= -0.142$, BIC $= -0.091$.

Normality of the residuals is rejected because of the considerable excess kurtosis. Closer inspection of the residuals reveals that this may be caused by large residuals in the first quarter of 1950 and the second quarter of 1980. These observations also may cause the ARCH tests to reject homoskedasticity. On the other hand, the LM test for ARCH is known to have power against alternatives other than ARCH as well, and, hence, it also may be that the significant values of this test statistic are caused by neglected nonlinearity.

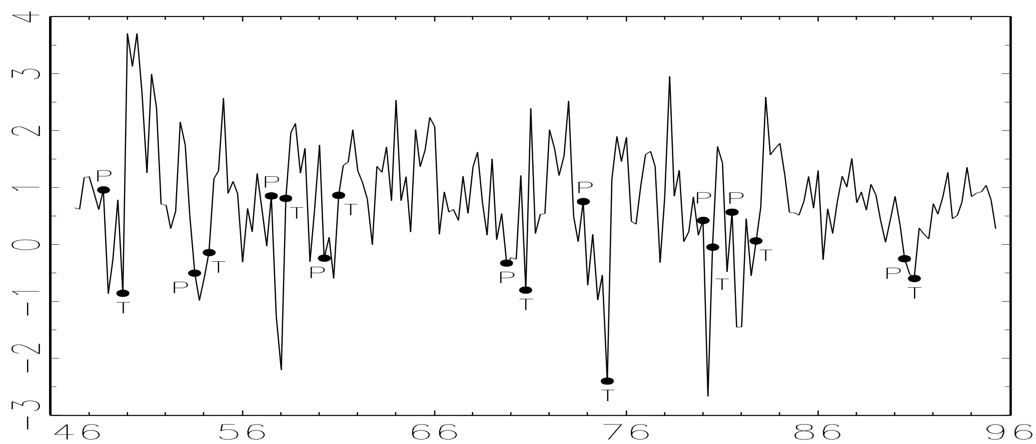
This final conjecture is investigated further by applying the LM-type linearity tests of Luukkonen *et al.* (1988), discussed in Section 2.2. I only report results for the LM_3 test, which is obtained by replacing the logistic transition function in (3.1) with a third-order Taylor approximation, as well as the parsimonious LM_3^e test. Apart from lagged growth rates and changes therein, I also consider a measure of the current depth of recession [CDR] as possible transition variable, following Beaudry and Koop (1993). I define CDR_t as

$$\text{CDR}_t = \max_{j \geq 1}\{x_{t-j}\} - x_t, \tag{3.24}$$

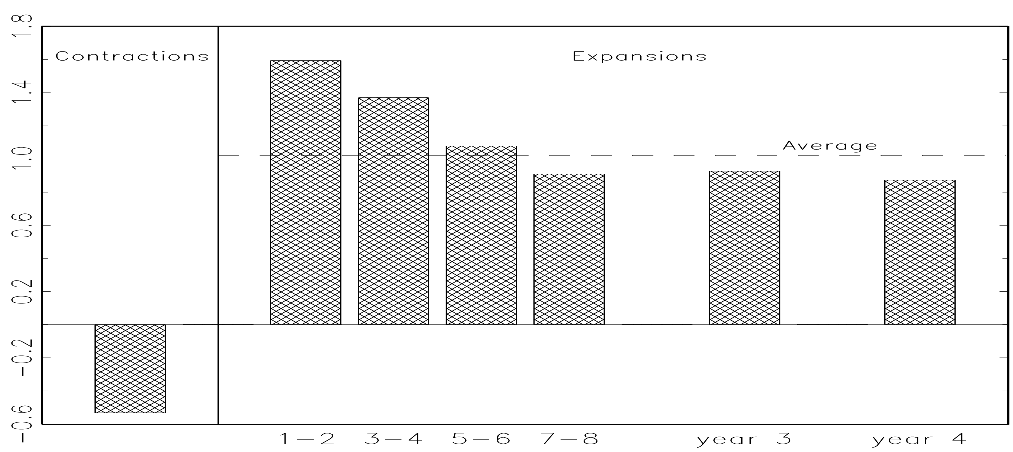
with x_t the log of US real GNP. As noted above, Beaudry and Koop (1993) include CDR_{t-1} as an additional regressor in an otherwise linear AR model for the GNP growth rate y_t. They claim that their CDR measure allows examination of the possibly different impact of positive and negative shocks. This is disputed by Elwood (1998), who argues that CDR_t only indicates (approximately) whether the economy is in contraction or expansion, but does not measure the impact of negative shocks *per se*[6]. Following this argument, I only consider the CDR measure as a possible transition variable in STAR models[7]. Note that the definition of the CDR variable

[6]See Hess and Iwata (1997a) for another critical assessment of the model of Beaudry and Koop (1993).

[7]Note that CDR_t resembles the growth rate y_t quite closely. Given that real GNP is upward trending, $\max_{j\geq 1} x_{t-j}$ will be equal to x_{t-1} most of the time. In that case, CDR_t equals $-y_t$. To be more precise, it is straightforward to show that $\text{CDR}_t = \max(\text{CDR}_{t-1}, 0) - y_t$. Hence, during expansions (that is, when $\text{CDR}_{t-1} < 0$), CDR_t and y_t coincide, whereas during contractions they might differ. The correlation between CDR_t and y_t equals -0.8, which confirms their similarity.



(a) Growth rate with peaks (P) and troughs (T)



(b) Mean growth rate during expansions and contractions

Figure 3.3: US real GNP, quarterly growth rate. Panel (a) shows quarterly growth rates of US real GNP, 1947:2-1995:2. Solid circles indicate NBER-dated peaks (P) and troughs (T). The lower panel displays average growth rates during contractions and several sub-periods of expansions. The dashed line labeled 'Average' is the average growth rate during complete expansions.

Table 3.4: LM-type tests for STAR nonlinearity in US GNP growth rates

Transition variable	Test	d = 1	2	3	4	5	6
y_{t-d}	LM_3	0.211	0.120	0.646	0.602	0.242	0.376
	LM_3^e	0.330	0.053	0.256	0.258	0.235	0.248
Δy_{t-d}	LM_3	0.089	0.065	0.982	0.819	0.291	0.220
	LM_3^e	0.074	0.248	0.971	0.840	0.287	0.460
CDR_{t-d}	LM_3	0.023	0.083	0.157	0.758	0.835	0.664
	LM_3^e	0.022	0.014	0.123	0.498	0.645	0.564
ΔCDR_{t-d}	LM_3	0.777	0.059	0.714	0.712	0.296	0.587
	LM_3^e	0.649	0.159	0.745	0.544	0.067	0.356

p-values for LM-type tests for smooth transition nonlinearity in quarterly growth rate of US real GNP. CDR_t measures the current depth of a recession, $CDR_t = \max_{j\geq 1}\{x_{t-j}\} - x_t$ with x_t the log of US GNP.

in (3.24) differs slightly from the original one of Beaudry and Koop (1993), which involves the maximum of past and *current* GNP. Hence, their CDR measure is equal to zero if real GNP is at an all time high, and greater than zero otherwise. Because using such a truncated variable as the transition variable in STAR models is not very convenient, I only consider the maximum up to time t.

General versions of the LM-type tests for STAR nonlinearity, in which the transition variable is assumed to be a lagged endogenous value y_{t-d} with the delay d left unspecified, reject the null hypothesis of linearity quite convincingly, with p-values of the LM_3 and LM_3^e tests equal to 0.029 and 0.057, respectively. However, if s_t is taken equal to y_{t-d} for a pre-specified value of d (to get an impression of the most appropriate transition variable(s)), the evidence for nonlinearity, in particular from the LM_3 test, disappears almost completely[8], as shown in Table 3.4. Only for $s_t = y_{t-2}$ can the null hypothesis of linearity be rejected at conventional significance levels. Application of the linearity tests with lagged changes in the growth rate or lagged CDR values are more succesful, in the sense that linearity can be rejected more often and more convincingly. In particular, the p-values of the tests shown in Table 3.4 suggest that Δy_{t-1}, Δy_{t-2}, CDR_{t-1}, and CDR_{t-2} might be considered as transition variables in a STAR model.

I decide to estimate an LSTAR model with CDR_{t-2} as the transition variable,

[8]Jansen and Oh (1996) also report that the tests for STAR-type nonlinearity do not reject the null hypothesis of linearity in case the transition variable is taken to be a lagged growth rate y_{t-d}, $d > 0$. Similarly, Hansen (1996) shows that tests for threshold-type nonlinearity do not provide very convincing evidence in favor of a threshold model. An alternative interpretation of these test results is that the regimes are determined by some function of the growth rates in the previous two quarters y_{t-1} and y_{t-2}. This also is suggested by the Bayesian analysis of a threshold model for US GDP by Koop (1996), who finds that the posterior for the delay parameter d, $p(d|\text{data})$ is such that $p(d = 1|\text{data}) \approx p(d = 2|\text{data}) \approx 0.5$.

because the p-value of the LM_3^e test is the smallest when this variable is used as transition variable. The parameters in this LSTAR model are estimated as

$$\begin{aligned} y_t = &\underset{(0.138)}{[0.160} + \underset{(0.090)}{0.346}\, y_{t-1} + \underset{(0.108)}{0.282}\, y_{t-2}] \times [1 - G(\mathrm{CDR}_{t-2})] \\ &+ \underset{(0.163)}{[0.665} + \underset{(0.121)}{0.308}\, y_{t-1} + \underset{(0.148)}{0.048}\, y_{t-2}] \times G(\mathrm{CDR}_{t-2}) + \varepsilon_t, \end{aligned} \tag{3.25}$$

$$G(\mathrm{CDR}_{t-2}) = (1 + \exp[-\underset{(-)}{200.0}\,(\mathrm{CDR}_{t-2} - \underset{(0.135)}{0.281})/\sigma_{\mathrm{CDR}_{t-2}}])^{-1}, \tag{3.26}$$

$\hat{\sigma}_\varepsilon = 0.899$, SK $= -0.17(0.16)$, EK $= 1.19(0.00)$, JB $= 12.21(0.00)$, ARCH(1) $=$ $2.74(0.09)$, ARCH(4) $= 7.09(0.13)$, $\mathrm{LM}_{\mathrm{SI}}(4) = 1.39(0.24)$, $\mathrm{LM}_{\mathrm{SI}}(8) = 1.48(0.17)$, $\mathrm{LM}_{\mathrm{C1}} = 1.12(0.35)$, $\mathrm{LM}_{\mathrm{C2}} = 1.01(0.44)$, $\mathrm{LM}_{\mathrm{C3}} = 0.87(0.62)$, AIC $= -0.129$, BIC $=$ 0.008,

where $\sigma_{\mathrm{CDR}_{t-2}}$ denotes the standard deviation of the transition variable CDR_{t-2}. $\mathrm{LM}_{\mathrm{SI}}(q)$ denotes the LM-type test for qth-order serial correlation in the residuals and $\mathrm{LM}_{\mathrm{C,i}}$, $i = 1, 2, 3$ denote the LM-type tests for parameter constancy, which were discussed in Section 2.4.

The exponent in the transition function is divided by the standard deviation of the transition variable in order to make γ scale-free. I do not report a standard error for $\hat{\gamma}$ for reasons discussed in Section 2.3. The sum of the autoregressive coefficients is considerably larger in the regime where $G(\mathrm{CDR}_{t-2})$ is equal to zero, which corresponds to expansions. This confirms the findings of Beaudry and Koop (1993) and Potter (1995b), among others, that contractions are less persistent than expansions. Also note the large constant in the upper regime, which might be taken as an additional indication of a quick recovery following contractions, see Sichel (1994) and Wynne and Balke (1992).

Apart from the diagnostic checks reported below the LSTAR model (3.25), I also apply the LM-type test against the MRSTAR alternative, developed in Section 3.2.1, as well as the LM-type tests of Eitrheim and Teräsvirta (1996) for remaining nonlinearity discussed in Section 2.4. Table 3.5 shows the p-values of the different tests for various choices of transition variables in the second transition function. The table also reports results of the same tests when the additional transition function is replaced by a first-order Taylor expansion, which, in theory at least, should be sufficient if only the logistic function is considered. The entries in Table 3.5 suggest that there is substantial evidence against the two-regime LSTAR model in favor of a MRSTAR model, especially if the change in the growth rate lagged one period is taken to be the transition variable in the second transition function.

Hence I proceed with estimating a four-regime MRSTAR model, with CDR_{t-2} and Δy_{t-1} as transition variables in the two logistic functions. The estimated model

Table 3.5: LM-type tests for multiple regimes in US GNP growth rates

Transition variable	Test			
	$LM_{AMR,1}$	$LM_{AMR,3}$	$LM_{EMR,1}$	$LM_{EMR,3}$
y_{t-1}	0.35	0.26	0.27	0.53
y_{t-2}	0.35	0.06	0.16	0.15
Δy_{t-1}	0.08	0.06	0.01	0.05
CDR_{t-1}	0.18	0.06	0.23	0.07
CDR_{t-2}	0.18	0.32	0.12	0.61
ΔCDR_{t-1}	0.56	0.56	0.22	0.41

The entries in columns $LM_{AMR,1}$ and $LM_{AMR,3}$ are p-values for the LM-type tests of Eitrheim and Teräsvirta (1996) for remaining nonlinearity, based on first- and third-order Taylor approximations of the second transition function, respectively. The entries in columns $LM_{EMR,1}$ and $LM_{EMR,3}$ are p-values for the tests of a basic LSTAR model against an MRSTAR alternative as developed in Section 3.2.1, also using first- and third-order Taylor approximations, respectively.

is given below.

$$\begin{aligned}
y_t = {} & [(\underset{(0.195)}{0.394} + \underset{(0.138)}{0.460}\, y_{t-1} + \underset{(0.156)}{0.092}\, y_{t-2}) \times (1 - G_1(\Delta y_{t-1})) \\
& + (\underset{(0.322)}{-0.121} + \underset{(0.284)}{0.442}\, y_{t-1} + \underset{(0.344)}{0.346}\, y_{t-2}) \times G_1(\Delta y_{t-1})] \times [1 - G_2(\mathrm{CDR}_{t-2})] \\
& + [(\underset{(0.283)}{0.360} - \underset{(0.362)}{0.530}\, y_{t-1} + \underset{(0.449)}{0.963}\, y_{t-2}) \times (1 - G_1(\Delta y_{t-1})) \\
& + (\underset{(0.283)}{-0.019} + \underset{(0.187)}{0.744}\, y_{t-1} - \underset{(0.215)}{0.235}\, y_{t-2}) \times G_1(\Delta y_{t-1})] \times G_2(\mathrm{CDR}_{t-2}) + \hat{\varepsilon}_t,
\end{aligned} \tag{3.27}$$

$$G_1(\Delta y_{t-1}) = (1 + \exp[-\underset{(-)}{500}\, (\Delta y_{t-1} - \underset{(0.032)}{0.250})/\sigma_{\Delta y_{t-1}}])^{-1}, \tag{3.28}$$

$$G_2(\mathrm{CDR}_{t-2}) = (1 + \exp[-\underset{(-)}{500}\, (\mathrm{CDR}_{t-2} - \underset{(0.259)}{0.064}/\sigma_{\mathrm{CDR}_{t-2}}])^{-1}. \tag{3.29}$$

$\hat{\sigma}_\varepsilon = 0.867$, SK $= -0.12(0.25)$, EK $= 0.55(0.06)$, JB $= 2.82(0.24)$, ARCH(1) $= 1.08(0.30)$, ARCH(4) $= 4.28(0.37)$, AIC $= -0.117$, BIC $= 0.155$.

The large estimates of γ_1 and γ_2 in (3.29) and (3.28) imply that for both $G_1(\Delta y_{t-1})$ and $G_2(\mathrm{CDR}_{t-2})$ the transition from zero to one is almost instantaneous at the estimated thresholds. The model is thus very similar to a NeTAR model. The model distinguishes between four different regimes, depending on whether the level of real GNP is above or below its historic high and whether growth is increasing or decreasing. This suggests the following interpretation of the four regimes.

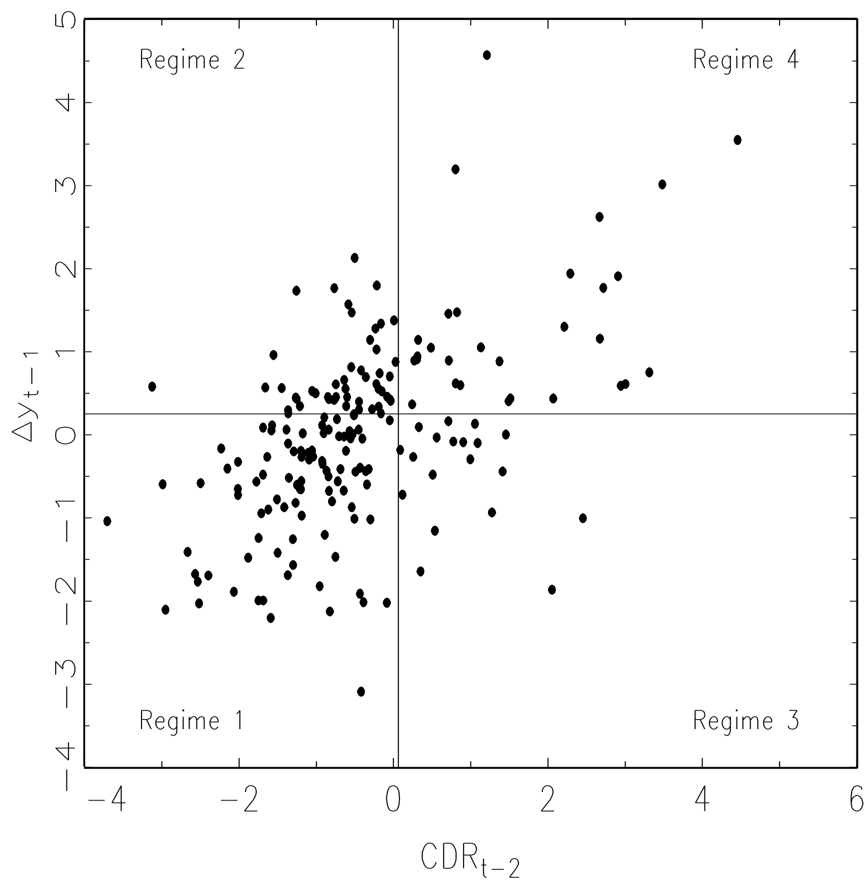


Figure 3.4: Distribution of observations on quarterly growth rates of US real GNP over the different regimes in the estimated MRSTAR model (3.27)-(3.29).

- $\Delta y_{t-1} < 0, \mathrm{CDR}_{t-2} < 0$. The economy is in expansion (recall that CDR_t as defined in (3.24) measures the distance in the level of real GNP relative to the previous all time high), but growth is declining.
- $\Delta y_{t-1} > 0, \mathrm{CDR}_{t-2} < 0$. The economy is in a strengthening expansion, as growth is accelerating.
- $\Delta y_{t-1} < 0, \mathrm{CDR}_{t-2} > 0$. The economy is in a worsening contraction.
- $\Delta y_{t-1} > 0, \mathrm{CDR}_{t-2} > 0$. The economy is in a contraction, but is improving given the positive change in growth.

The fourth regime more or less corresponds with the recovery phase identified by Sichel (1994), in which growth is strong immediately following a trough.

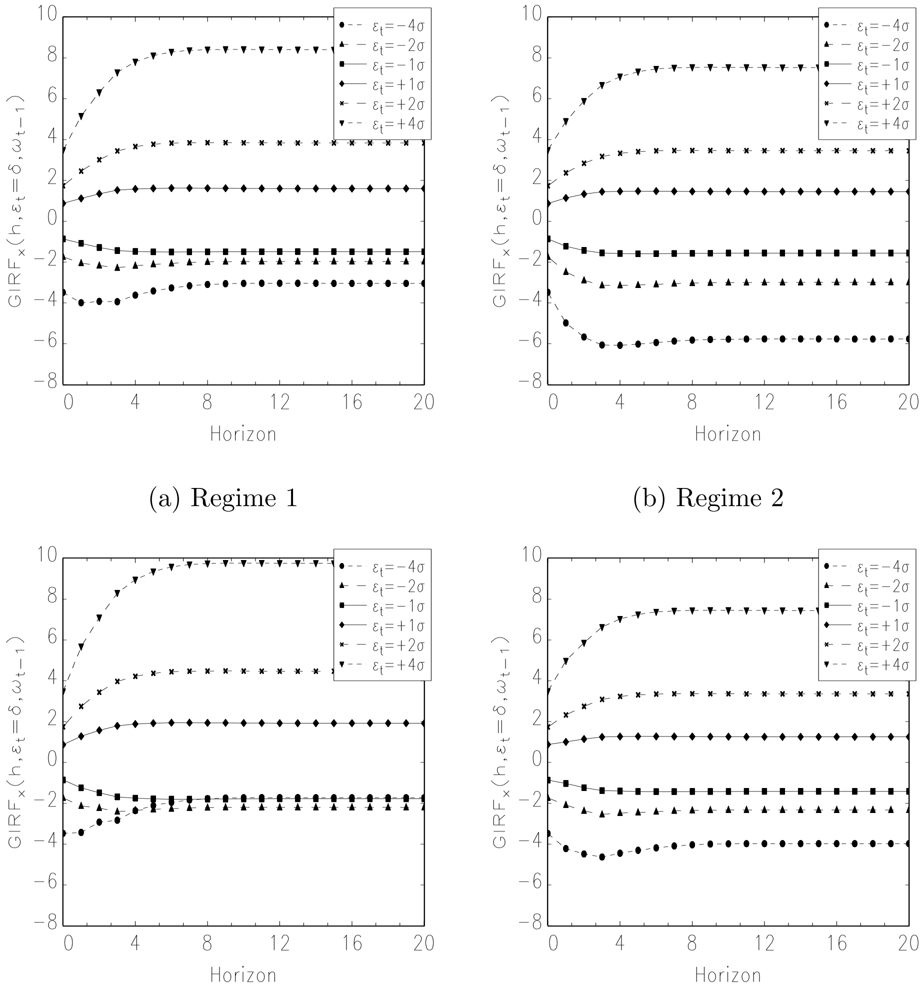
Figure 3.4 shows the distribution of the observations across the different regimes. When I take model (3.27), it is seen that the bulk of the observations is in regime 1, followed by regime 2. The worsening contraction regime (regime 3) contains only 19 observations, confirming that the US economy tends to recover quickly from recessions.

The various diagnostic tests for the MRSTAR model demonstrate that the residuals are much better behaved than the residuals from the AR and LSTAR models estimated before. For example, normality cannot be rejected anymore. On the other hand, comparing the residual standard deviations suggests that the additional regimes improve the fit of the model only slightly, whereas both information criteria clearly favor the parsimonious AR model. As an alternative way to evaluate the potential usefulness of the elaborate MRSTAR model, I examine the implied propagation of shocks occurring in different regimes. Toward this end I compute generalized impulse response functions [GIRFs] as developed in Koop *et al.* (1996) and discussed in Section 2.6. Recall that the GIRF can be used to examine the impact of different shocks under different circumstances by conditioning on particular subsets of shocks ε_t and histories ω_{t-1}, denoted A and B, respectively,

$$\mathrm{GIRF}_y(h, A, B) = E(y_{t+h}|\varepsilon_t \in A, \omega_{t-1} \in B) - E(y_{t+h}|\omega_{t-1} \in B), \tag{3.30}$$

where for the MRSTAR model the history ω_{t-1} can be summarized by the growth rates during the previous two quarters and the lagged current depth of recession, that is $\omega_{t-1} = \{y_{t-1}, y_{t-2}, \mathrm{CDR}_{t-2}\}$. I use a special case of (3.30) to obtain an impression of the dynamics in the different regimes of the estimated MRSTAR model and consider the GIRF for specific shocks, conditioning on all histories in a particular regime. That is, the set A is taken to consist of a single element δ, whereas the set B consists of all histories belonging to one of the four regimes in the MRSTAR model. For the shock ε_t, I consider values $\delta = \pm 1, \pm 2$, and ± 4 times the residual standard deviation. The GIRFs are estimated using the simulation procedure outlined by Koop *et al.* (1996). In particular, I use all observed histories in the estimation sample 1947:4-1995:2 and the corresponding residuals from the MRSTAR model to estimate the conditional expectations $E(y_{t+h}|\varepsilon_t = \delta, \omega_{t-1})$ and $E(y_{t+h}|\omega_{t-1})$ to obtain the shock- and history-specific GIRF as given in (2.87). The conditional GIRFs then are computed by averaging across histories in a particular regime. The GIRFs for the log level of US GNP (which are obtained by taking cumulative sums of the GIRFs for the growth rate) are shown in Figure 3.5.

Several conclusions can be drawn from this figure. First, negative shocks appear to be less persistent than positive shocks, in the sense that in three out of the four regimes the average long-run response to negative shocks is smaller than the long-run response to positive shocks of equal size. This corresponds with the conclusions of Beaudry and Koop (1993) and Potter (1995b), but contradicts the findings of Pesaran and Potter (1997). Second, whereas the response to positive shocks is quite similar in the different regimes, the response to negative shocks differs markedly. In the strengthening-expansion regime 2, negative shocks are magnified by a factor of 1.5 in the long run. In both the weakening-expansion and improving-contraction regimes, the long-run impact of negative shocks is approximately equal to the size of the shock. Finally, in the worsening-contraction regime 3, all negative shocks appear to generate approximately the same response, irrespective of their size. Inspection of the GIRFs for individual histories in this regime reveals that the long-run response to negative shocks can even be positive, while reversals also occur, that is, the largest (in absolute value) negative shock has the largest positive response.



(a) Regime 1

(b) Regime 2

(c) Regime 3

(d) Regime 4

Figure 3.5: Generalized impulse response functions for the log level of US real GNP for shocks ε_t equal to $\pm 1, \pm 2$, and ± 4 times the standard deviation based on the estimated MRSTAR model (3.27)-(3.29), conditional upon histories in the various regimes.

3.4 Concluding remarks

In this chapter I have explored possibilities of extending the basic STAR model to allow for more than two regimes. I have shown that a multiple-regime STAR model with independent behaviour in all regimes can be obtained by encapsulating different two-regime models. A (specific-to-general) specification procedure was proposed and a new LM-type test for nonlinearity was developed, which can be used to test for the presence of multiple regimes. Alternatively, this test might be used as a diagnostic tool to test the adequacy of a fitted two-regime STAR model, complementing the tests of Eitrheim and Teräsvirta (1996). The application of the multiple-regime STAR model to post-war US real GNP demonstrates that a multiple-regime characterization of the business cycle might indeed be useful.

This chapter offers several possibilities for further research. First, the effect of outliers on the detection of regimes seems to be of interest, as one does not want to fit spuriously a model that contains additional regimes only to capture some aberrant observations. It appears that a robust estimation method for STAR models needs to be developed to achieve proper protection against the influence of such anomalous observations. Alternative ways to compare different STAR models, possibly with a different number of regimes also might be explored. For example, it should be possible to use the techniques of Hess and Iwata (1997b) to examine explicitly whether the regime-switching models are capable of replicating basic stylized facts such as amplitude and duration of expansions and contractions. Finally, it might be worthwhile to extend the application to US real GNP to a multivariate model, following the ideas of Koop *et al.* (1996), or to model nonlinearity and time-varying parameters simultaneously. Some of these issues are addressed in the following chapters.

Chapter 4

Time-Varying Smooth Transition Models

Nonlinearity is only one of many different features which a time series can possess. Another important characteristic of macro-economic time series, especially when observed over long time spans, is structural instability. For example, Stock and Watson (1996) report an overwhelming amount of evidence for instability in both univariate and multivariate models for a large number of US postwar macroeconomic time series. Despite the evidence that both nonlinearity and instability are relevant features of time series, to date they mainly have been analyzed in isolation. It is illustrative that Stock and Watson (1996) only use linear models to examine the stability properties of their time series. Few attempts have been made to consider nonlinearity and structural change simultaneously. Diebold and Rudebusch (1992), Watson (1994) and Parker and Rothman (1996) apply nonparametric techniques to examine whether certain characteristics of the business cycle, such as the duration and amplitude of recessions and booms, have changed over time, while allowing for these properties to be different for the different business cycle phases. Cooper (1998) finds indications for both regime-switching behaviour and structural change following World War II in US industrial production in a regression tree analysis. The only attempt to capture both nonlinearity and structural instability in a parametric time series model that I am aware of is Kim and Nelson (1998), who allow for a structural change in the mean growth rate of US real GDP, while modeling different dynamic behaviour in recessions and expansions by means of a Markov Switching model.

The example used in Chapter 2 demonstrates that it is not all that difficult to set the parameters in a STAR model in such a way that the resultant time series resemble series that are subject to occasional level shifts, see Figure 2.4. Casual inspection of a graph of such series might suggest that a model with time-varying parameters is an appropriate characterization of its properties. This demonstrates that structural instability and regime-switching can be observationally equivalent. Garcia and Perron (1996) provide an illustrative empirical example of this phenomenon. The 3-regime Markov Switching model which they estimate for the US real interest rate exhibits only 2 regime shifts over the 40-year sample period. Statistical procedures also might have difficulty to distinguish nonlinearity from structural change. For example, Carrasco (1997) and Clements and Smith (1998c) find that tests for SETAR type nonlinearity reject the null hypothesis of linearity with high probability when the data in fact are generated by a structural change model, whereas the converse is also true.

Given the above, regime-switching behaviour and structural change can be regarded as competing alternative hypotheses to linearity. Of course, it is also possible that a time series displays both nonlinearity and structural instability, see Kim and Nelson (1998). The aim of this chapter is to consider a model based on the principle of smooth transition that simultaneously allows for nonlinear dynamics and time-varying parameters. This time varying smooth transition autoregressive [TV-STAR] model can be regarded as a special case of the multiple-regime STAR [MRSTAR] model discussed in the previous chapter.

This chapter is organized as follows. In Section 4.1, I briefly discuss representation of the TV-STAR model. In Section 4.2, I suggest two different specification procedures for TV-STAR models. Besides the specific-to-general procedure for MRSTAR models proposed in the previous chapter, the special character of the TV-STAR model makes a general-to-specific approach an attractive alternative specification method. The relative performance of these procedures is investigated by means of Monte Carlo simulation in Section 4.3. In Section 4.4, I apply the model to examine nonlinearity and stability of growth rates in industrial production in a number of OECD countries. Finally, Section 4.5 contains some further discussion and concluding remarks.

4.1 Representation

Consider again the basic smooth transition autoregressive [STAR] model

$$y_t = \phi_1' x_t (1 - G(s_t;\gamma,c)) + \phi_2' x_t G(s_t;\gamma,c) + \varepsilon_t, \tag{4.1}$$

where x_t is a vector consisting of lagged endogenous variables, $x_t = (1, \tilde{x}_t')'$ with $\tilde{x}_t = (y_{t-1}, \ldots, y_{t-p})'$. In this chapter, $G(s_t;\gamma,c)$ is taken to be the logistic function

$$G(s_t;\gamma,c) = \frac{1}{1+\exp\{-\gamma(s_t - c)\}}, \qquad \gamma > 0, \tag{4.2}$$

where s_t is the transition variable, and γ and c determine the smoothness and location, respectively, of the transition between the two regimes. It is straightforward to generalize the analysis presented here to models with other transition functions.

A special case of the STAR model that is of particular interest in this chapter results if the transition variable is taken to be time, $s_t = t$. In that case, the STAR model becomes a model with smoothly changing parameters, see Lin and Teräsvirta (1994). The limiting case if $\gamma \to \infty$ then gives rise to an AR model with a structural break in the parameters at $t = c$.

As discussed at length in Chapters 2 and 3, the STAR model (4.1) essentially allows for two regimes only, corresponding with $G(s_t;\gamma,c) = 0$ and $G(s_t;\gamma,c) = 1$.

The model can be extended to allow for more than two regimes by encapsulating different two-regime STAR models. For example, a multiple-regime STAR [MRSTAR] model with four regimes, as discussed in the previous chapter, is given by

$$\begin{aligned} y_t = [\phi_1' x_t(1 - G_1(s_{1t}; \gamma_1, c_1)) + \phi_2' x_t G_1(s_{1t}; \gamma_1, c_1)][1 - G_2(s_{2t}; \gamma_2, c_2)] + \\ [\phi_3' x_t(1 - G_1(s_{1t}; \gamma_1, c_1)) + \phi_4' x_t G_1(s_{1t}; \gamma_1, c_1)] G_2(s_{2t}; \gamma_2, c_2) + \varepsilon_t, \quad (4.3) \end{aligned}$$

where both $G_1(s_{1t}; \gamma_1, c_1)$ and $G_2(s_{2t}; \gamma_2, c_2)$ are of the form (4.2). The four regimes in the model (4.3) correspond with particular combinations of $G_1(s_{1t}; \gamma_1, c_1)$ and $G_2(s_{2t}; \gamma_2, c_2)$ being equal to 0 or 1. A time-varying STAR [TV-STAR] model now is obtained from (4.3) if the switching variables are taken to be a lagged endogenous variable and time, that is, $s_{1t} = y_{t-d}$ for certain $d > 0$ and $s_{2t} = t$. The model implies that y_t follows a STAR model at all times, with a smooth change in the autoregressive parameters in both regimes, from ϕ_1 to ϕ_3 for $G_1(y_{t-d}; \gamma_1, c_1) = 0$ and from ϕ_2 to ϕ_4 for $G_1(y_{t-d}; \gamma_1, c_1) = 1$. Obviously, by imposing appropriate restrictions on the autoregressive parameters in the different regimes, special cases can be obtained, such as a linear model which changes into a STAR model at some point during the sample, or a 2-regime STAR model with smoothly changing parameters in only one of the regimes. The various models that are nested within the general TV-STAR model will be considered in more detail in Section 4.3 below.

In the remainder of this chapter, I abbreviate $G_1(y_{t-d}; \gamma_1, c_1)$ and $G_2(t; \gamma_2, c_2)$ to $G(y_{t-d})$ and $G(t)$, respectively. Using this shorthand notation, the TV-STAR model is given by

$$\begin{aligned} y_t = [\phi_1' x_t(1 - G(y_{t-d})) + \phi_2' x_t G(y_{t-d})][1 - G(t)] + \\ [\phi_3' x_t(1 - G(y_{t-d})) + \phi_4' x_t G(y_{t-d})] G(t) + \varepsilon_t, \quad (4.4) \end{aligned}$$

In the next section I discuss two possible specification procedures for TV-STAR models.

4.2 Specification of TV-STAR models

As the TV-STAR model is a special case of the MRSTAR model, the specific-to-general procedure can readily be applied to specify TV-STAR models. Given the particular choice of transition variables in the TV-STAR model, a more direct approach to determining the appropriate model also is feasible. To be precise, linearity can be tested explicitly against the alternative of a TV-STAR model. If linearity is rejected, sub-hypotheses can be tested to examine whether either a 2-regime STAR model or a model with smoothly changing parameters is sufficient to characterize the time series at hand. This general-to-specific procedure is described in Section 4.2.3. The relative merits of the two possible specification procedures are investigated in Section 4.3 by means of Monte Carlo simulation. To keep this chapter self-contained, the various steps in the specific-to-general procedure are summarized in Section 4.2.2. This section starts with a brief description of the Lagrange Multiplier [LM] type test of linearity against the TV-STAR alternative.

As both the specific-to-general and general-to-specific procedures make heavy use of LM-type statistics to compare different models that are nested in the general TV-STAR model, this also serves as a reminder for the general principle underlying these tests.

4.2.1 An LM-type test against TV-STAR

Consider the problem of testing the null hypothesis of linearity against the alternative of a TV-STAR model (4.4). In order to obtain an appropriate test statistic, rewrite (4.4) as

$$y_t = \alpha' x_t + \beta' x_t G(y_{t-d}) + \pi' x_t G(t) + \theta' x_t G(y_{t-d}) G(t) + \varepsilon_t, \tag{4.5}$$

where $\alpha = \phi_1$, $\beta = \phi_2 - \phi_1$, $\pi = \phi_3 - \phi_1$, and $\theta = \phi_4 - \phi_3 - \phi_2 + \phi_1$. The null hypothesis of linearity is given by $H_0 : \gamma_1 = \gamma_2 = 0$. If the null hypothesis holds, the model is not identified and, therefore, this hypothesis can not be tested directly. To circumvent this identification problem the suggestion of Luukkonen *et al.* (1988) to approximate the transition functions by Taylor expansions can be followed again. To keep the notation clear, I adjust both $G(y_{t-d})$ and $G(t)$ by subtracting 0.5. This does not change the model, but has the advantage that under the null hypothesis the functions are equal to zero. The reparameterized model based on a first-order approximations for both transition functions is given by

$$y_t = \alpha^{*\prime} x_t + \beta^{*\prime} \tilde{x}_t y_{t-d} + \pi^{*\prime} x_t t + \theta^{*\prime} x_t y_{t-d} t + R(\gamma_1, \gamma_2) + \varepsilon_t, \tag{4.6}$$

where

$$\alpha^* = \alpha - \frac{1}{4} c_1 \beta \gamma_1 - \frac{1}{4} c_2 \pi \gamma_2 + \frac{1}{16} c_1 c_2 \theta \gamma_1 \gamma_2, \tag{4.7}$$

$$\beta^* = \frac{1}{4} \tilde{\beta} \gamma_1 - \frac{1}{16} c_2 \tilde{\theta} \gamma_1 \gamma_2, \tag{4.8}$$

$$\phi^* = \frac{1}{4} \pi \gamma_2 - \frac{1}{16} c_1 \theta \gamma_1 \gamma_2, \tag{4.9}$$

$$\theta^* = \frac{1}{16} \theta \gamma_1 \gamma_2, \tag{4.10}$$

with $\tilde{\beta}$ such that $\beta = (1, \tilde{\beta}')'$ and $\tilde{\theta}$ similarly defined. $R(\gamma_1, \gamma_2)$ is a remainder from the Taylor expansions. Under H_0, $R(\gamma_1, \gamma_2) \equiv 0$ so that this remainder does not affect the distribution theory. From (4.7)-(4.10) it follows that the null hypothesis of linearity in terms of the parameters of this reparameterized model is given by $H_0' : \beta^* = \pi^* = \theta^* = 0$, which can be tested by means of an LM-type test in a straightforward manner. Under certain regularity conditions, the test statistic has an asymptotic χ^2 distribution with $3(p+1) - 1$ degrees of freedom. In the Monte Carlo experiments reported below, I use the F-version of this and other LM-type tests, for the usual reason that it has better small sample properties than the corresponding χ^2-version.

The expressions for α^*, β^*, π^* and θ^* given in (4.7)-(4.10) also demonstrate under which restrictions the TV-STAR model reduces to a STAR or TV-parameter model. For example, if either $\gamma_2 = 0$ or, equivalently, $\phi_1 \neq \phi_2$, $\phi_1 = \phi_3$ and $\phi_2 = \phi_4$ in (4.4), the resultant model is a two-regime STAR model. From (4.7)-(4.10) and the relations between ϕ_i, $i = 1, \ldots, 4$, and α, β, π and θ given just below (4.5), it follows that these restrictions imply that $\beta^* \neq 0$, $\phi^* = \theta^* = 0$ in (4.6). Thus, linearity can be tested against the alternative of STAR-type nonlinearity by testing $H_0 : \beta^* = 0 | \pi^* = \theta^* = 0$. Similarly, linearity can be tested against the alternative of smoothly time-varying parameters by testing $H_0 : \pi^* = 0 | \beta^* = \theta^* = 0$ in (4.6).

4.2.2 Specific to general approach

The specific-to-general approach for specifying a TV-STAR model can be summarized as follows.

1. Specify an AR(p) model for the time series of interest y_t.
2. Test the null hypothesis of linearity against nonlinearity of STAR-type ($H_0 : \beta^* = 0 | \pi^* = \theta^* = 0$ in (4.6)) and against smoothly changing parameters ($H_0 : \pi^* = 0 | \beta^* = \theta^* = 0$).
3. Estimate the alternative for which the null hypothesis is rejected most convincingly and compute LM-type tests against additional nonlinear structure. For example, if the null hypothesis of linearity is rejected most convincingly against STAR-type nonlinearity in step 2, estimate a 2-regime STAR model and test the STAR model for parameter constancy using the test of Eitrheim and Teräsvirta (1996) discussed in Section 2.4.
4. Estimate the TV-STAR model under the alternative if the null of no remaining nonlinear structure is rejected in step 3 is rejected, and again compute evaluation tests for remaining additive structure.

If, at any of the above stages, the maintained model passes all the specification tests without rejection, tentatively accept that model. For more elaborate discussion of the various steps in this procedure, see Section 3.2.

4.2.3 General to specific approach

Although the specific-to-general procedure outlined above makes sense intuitively, it might have some drawbacks. First, the procedure can be quite time-consuming, as it involves repeated estimation of nonlinear models. Second, the model which is finally selected may depend crucially on decisions which are made at early stages in the specification search. For example, if the tests against STAR-type nonlinearity and time-varying parameters at step 2 both reject their respective null hypotheses, proceeding with a STAR model might ultimately lead to a very different model than proceeding with a model with smoothly changing parameters.

An alternative approach is to adopt a general-to-specific procedure, in which one starts with a TV-STAR model and then checks whether a STAR model or a model with smoothly changing parameters is sufficient to capture the essential features of the time series under investigation. In order to avoid estimating a possibly unidentified model, I suggest not to start immediately with estimating a TV-STAR model and then testing downward, but rather to combine several LM-type tests to obtain a rough idea of which type of model appears most adequate. A possible specification procedure which adopts this approach consists of the following steps.

1. Specify an AR(p) model for the time series of interest y_t.

2. Use the LM-type statistic developed in Section 4.2.1 to test linearity directly against the TV-STAR alternative ($H_0^{TVSTAR} : \beta^* = \phi^* = \theta^* = 0$).

3. If the null hypothesis of linearity is rejected, test sub-hypotheses which are nested in H_0^{TVSTAR} to assess whether a TV-STAR model is really necessary or whether either a STAR model or a model with smoothly changing parameters is sufficient to characterize the time series y_t. In particular, one can test

$$H_0^{STAR} : \beta^* = \theta^* = 0,$$
$$H_0^{TV} : \pi^* = \theta^* = 0,$$

   in the auxiliary model (4.6). The corresponding LM-statistics will be denoted as LM_{STAR} and LM_{TV}, which have asymptotic χ^2 distributions with $2p-1$ and $2p$ degrees of freedom, respectively. As noted above, it follows from the expressions of the parameters β^*, π^* and θ^* in terms of the parameters in the original TV-STAR model that under H_0^{STAR} the model reduces to a model with smoothly changing parameters, while under H_0^{TV} a two-regime STAR model results. These considerations lead to the following decision rule:

   - If both H_0^{STAR} and H_0^{TV} are rejected, select a TV-STAR model;
   - If H_0^{TV} is not rejected, and H_0^{STAR} is, select a STAR model;
   - If H_0^{STAR} is not rejected, and H_0^{TV} is, select a model with smoothly changing parameters;

   The only combination of test outcomes which does not lead to a clear-cut model choice is when both H_0^{STAR} and H_0^{TV} are not rejected, while the general null hypothesis H_0^{TVSTAR} is. In this case one may resort to the LM-type tests which test linearity against STAR-type nonlinearity and time-varying parameters as in step 2 of the specific-to-general approach, and see whether these tests indicate which model may be best suited to describe the time series at hand. It should be remarked however that this combination of test outcomes is very unlikely to occur in practice, since H_0^{STAR} and H_0^{TV} are sub-hypotheses of H_0^{TVSTAR}.

4. Estimate the model which is selected on the basis of the LM-type statistics. Since these tests only give a rough indication of which model is most appropriate, the estimated model should be thoroughly evaluated by means of the various available diagnostic tests and modified accordingly if necessary.

4.3 Performance of specification procedures

In this section I evaluate the (relative) performance of the two specification procedures for TV-STAR models by means of Monte Carlo simulation. In particular, I assess the size and power properties of the LM-type tests which are involved in the general-to-specific approach[1] as described in Section 4.2.3 and the success rate of the two procedures, that is, the frequency of selecting the correct model. In all cases I use F variants of the LM-type test statistics.

4.3.1 Monte Carlo design

I examine the properties of the specification procedures for 7 different data generating processes [DGPs], all of which are nested in the TV-STAR model (4.4) with $p = d = 1$, that is,

$$\begin{aligned} y_t = {} & [\phi_1 y_{t-1}(1 - G(y_{t-1}; \gamma_1, c_1)) + \phi_2 y_{t-1} G(y_{t-1}; \gamma_1, c_1)][1 - G(t; \gamma_2, c_2)] + \\ & [\phi_3 y_{t-1}(1 - G(y_{t-1}; \gamma_1, c_1)) + \phi_4 y_{t-1} G(y_{t-1}; \gamma_1, c_1)] G(t; \gamma_2, c_2) + \varepsilon_t, \end{aligned} \quad (4.11)$$

where both $G(y_{t-1}; \gamma_1, c_1)$ and $G(t; \gamma_2, c_2)$ are given by (4.2). In all experiments described below, I use 10000 replications and a sample size of 250 observations. Necessary starting values are always set equal to zero, while the first 100 observations in the artificial samples are discarded in order to eliminate any possible influence of this choice. The errors ε_t are taken to be i.i.d. standard normal distributed throughout. Finally, in all experiments, the autoregressive order and the transition variable(s) are assumed known.

The various DGPs can be characterized conveniently by the restrictions they impose on the autoregressive parameters in the different regimes in (4.11) as follows.

(i) $\phi_1 = \phi_2 = \phi_3 = \phi_4$. In this case the TV-STAR model reduces to a linear AR(1) model

$$y_t = \phi_1 y_{t-1} + \varepsilon_t. \quad (4.12)$$

This DGP is used only to evaluate the size properties of the LM-type tests in the general-to-specific approach. The AR parameter is varied among $\phi_1 = \{0.0, 0.1, 0.3, \ldots, 0.9\}$.

(ii) $\phi_1 = \phi_3$ and $\phi_2 = \phi_4$. This restriction implies that there is no time variation in the autoregressive parameters or, put differently, the DGP is a STAR model,

$$y_t = \phi_1 y_{t-1}(1 - G(y_{t-1}; \gamma_1, c_1)) + \phi_2 y_{t-1} G(y_{t-1}; \gamma_1, c_1) + \varepsilon_t. \quad (4.13)$$

[1] The size and power properties of the various LM-type tests which are used in the specific-to-general procedure have been investigated elsewhere. The properties of the tests against STAR nonlinearity and smoothly changing parameters are examined by Teräsvirta (1994) and Lin and Teräsvirta (1994), respectively. Eitrheim and Teräsvirta (1996) discuss the finite sample properties of the diagnostic tests for estimated STAR models against additional nonlinearity and smoothly changing parameters, see also Section 3.2.2.

The parameterizations for this DGP are taken from Luukkonen *et al.* (1988). I set ϕ_1 equal to -0.5 or 0.5, while ϕ_2 is varied among $\phi_2 \in \{-0.9, -0.7, -0.5, -0.3, 0, 0.3, 0.5, 0.7, 0.9\}$. The threshold c_1 in the logistic function $G(y_{t-1}; \gamma_1, c_1)$ is fixed at zero, while the smoothness parameter γ_1 is set equal to 5.

(iii) $\phi_1 = \phi_2$ and $\phi_3 = \phi_4$. In this case the resulting DGP is a model with smoothly changing parameters only and no STAR-type behaviour,

$$y_t = \phi_1 y_{t-1}(1 - G(t; \gamma_2, c_2)) + \phi_3 y_{t-1} G(t; \gamma_2, c_2) + \varepsilon_t. \tag{4.14}$$

In this case ϕ_1 is again fixed at -0.5 or 0.5, while ϕ_3 is varied as ϕ_2 in case (ii) discussed above. The threshold c_2 is varied among $c_2 = .25, .50, .75$, in order to examine the impact of the fraction of the sample at which the shift in parameters occurs. To save space, only the case $c_2 = 0.50$ is reported in full detail. Finally, γ_2 is set equal to 25, such that the parameter change takes about half of the sample to be completed.

(iv) $\phi_1 = \phi_2$. The DGP resulting from this restriction is a linear AR model which changes into a STAR model as $G(t; \gamma_2, c_2)$ changes from 0 to 1. The restriction $\phi_3 = \phi_4$ yields a mirror image, in the sense that in this case a STAR model changes smoothly into a linear model.

(v) $\phi_1 = \phi_3$. This DGP might be interpreted as a STAR model with a smoothly changing autoregressive parameter in the regime where $G(y_{t-1}; \gamma_1, c_1) = 1$, while the AR parameter in the regime where $G(y_{t-1}; \gamma_1, c_1) = 0$ remains constant. A mirror image of this DGP, that is, a STAR model with only a smoothly changing autoregressive parameter in the regime corresponding with $G(y_{t-1}; \gamma_1, c_1) = 0$, is obtained by imposing the restriction $\phi_2 = \phi_4$.

(vi) $\phi_4 - \phi_3 = \phi_2 - \phi_1$. In this case the autoregressive parameters in both regimes of the STAR model change over time, but in such a way that their difference (which in a sense can be interpreted as the degree of nonlinearity) remains the same.

(vii) $\phi_1 = \phi_4$. This final restriction which is considered here renders a model in which the dynamic behaviour in the regime $G(y_{t-1}; \gamma_1, c_1) = 0$ before the structural change is the same as the dynamic behaviour in the regime $G(y_{t-1}; \gamma_1, c_1) = 1$ after the change.

DGPs (iv), (v) and (vii) restrict two of the four autoregressive parameters ϕ_i, $i = 1, \ldots, 4$. In all cases, these restricted parameters are set equal to 0.5, while the two unrestricted parameters are varied (independently) among $\{-0.9, -0.7, -0.5, -0.3, 0, 0.3, 0.5, 0.7, 0.9\}$. In DGP (vi), I fix $\phi_1 = 0.5$ and vary ϕ_2 and ϕ_4 independently as in the other DGPs. For each combination of these unrestricted parameters, the value of ϕ_3 is obtained as $\phi_3 = \phi_1 - \phi_2 + \phi_4$. I only investigate combinations that yield realizations that are stationary, that is, configurations for which the implied ϕ_3 is larger than one in absolute value are not considered. The parameters in the transition functions $G(y_{t-1}; \gamma_1, c_1)$ and $G(t; \gamma_2, c_2)$ are set as discussed for cases (ii) and (iii), respectively.

Table 4.1: Empirical size of LM-type tests in general-to-specific procedure

		LM_{TVSTAR}			LM_{STAR}			LM_{TV}		
π_1	α	0.010	0.050	0.100	0.010	0.050	0.100	0.010	0.050	0.100
0.0		0.009	0.042	0.089	0.009	0.046	0.095	0.009	0.043	0.088
0.1		0.008	0.041	0.087	0.008	0.044	0.090	0.010	0.040	0.085
0.3		0.008	0.039	0.085	0.007	0.042	0.086	0.009	0.040	0.085
0.5		0.007	0.039	0.086	0.006	0.040	0.079	0.009	0.041	0.082
0.7		0.006	0.035	0.078	0.004	0.031	0.067	0.007	0.040	0.084
0.9		0.010	0.041	0.084	0.003	0.018	0.045	0.013	0.055	0.100

Empirical size of the LM-type test statistics which are involved in the general-to-specific procedure for specification of TV-STAR models, as described in Section 4.2.3. Series are generated according to the AR(1) model (4.12) with $\varepsilon_t \sim$ i.i.d. $N(0,1)$. The table is based on 10000 replications.

4.3.2 Rejection frequencies of LM-type statistics

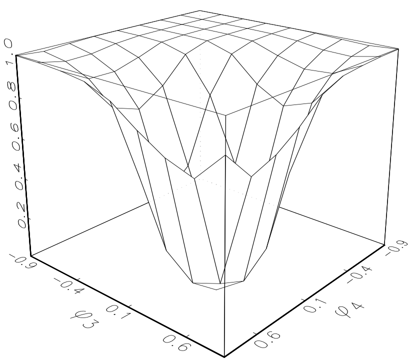
In this subsection I briefly discuss the properties of the three LM-type test statistics involved in the general-to-specific procedure under the different DGPs.

Rejection frequencies at nominal significance levels $\alpha = 0.01, 0.05$ and 0.10 when the DGP is the AR(1) model (4.12) are given in Table 4.1. It is seen that the empirical size of all three tests is reasonably close to the selected nominal significance levels.

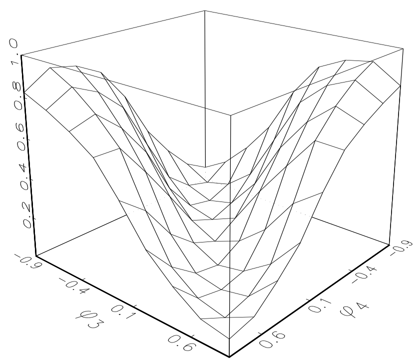
The rejection frequencies for the tests in case the DGP is the LSTAR model (4.13) [case (ii)] or the time-varying parameter model (4.14) [case (iii)] are conform prior expectations. For example, in case (ii), the rejection frequencies of the LM_{TVSTAR} and LM_{STAR} statistics are reasonably high and increase monotonically as the degree of nonlinearity, measured by the difference between the autoregressive parameters in the two regimes, $\phi_2 - \phi_1$, increases. The rejection frequency of the LM_{TV} statistic stays close to the nominal significance level (as it should). In case (iii), the power of the LM_{TVSTAR} and LM_{TV} statistics increases monotonically as the difference $\phi_4 - \phi_3$ increases, whereas the rejection frequency of LM_{STAR} stays close to the nominal significance level. The tests are most powerful if the change in parameters is centered around the middle of the sample.

For the other DGPs, the behaviour of the three statistics also corresponds with what one might expect intuitively. I illustrate this here by presenting some results for DGPs (iv) and (v). Figure 4.1 shows the rejection frequencies of the three statistics in case the DGP is a linear model which changes into a STAR model [case (iv)], with the change centered at the middle of the sample ($c_2 = 0.50$).

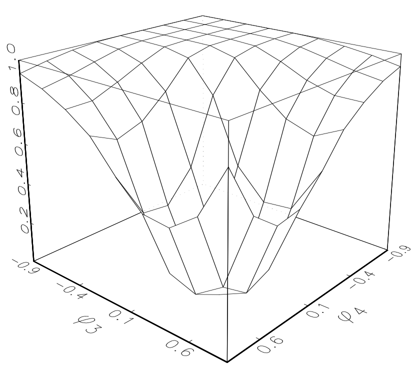
The power of the LM_{TVSTAR} and LM_{TV} tests is seen to increase as the differences between ϕ_3 and ϕ_4 and the restricted parameters ϕ_1 and ϕ_2 (which are fixed at 0.5) become larger. The power of the LM_{STAR} statistic on the other hand increases when the difference between ϕ_3 and ϕ_4 (which are the autoregressive parameters in the STAR model after the change) becomes larger. Comparing these results with the findings for DGPs with $c_2 = 0.25$ and $c_2 = 0.75$ shows that the power of the LM_{TVSTAR} and LM_{TV} statistics is highest when the change from the AR to the



(a) LM_{TVSTAR}

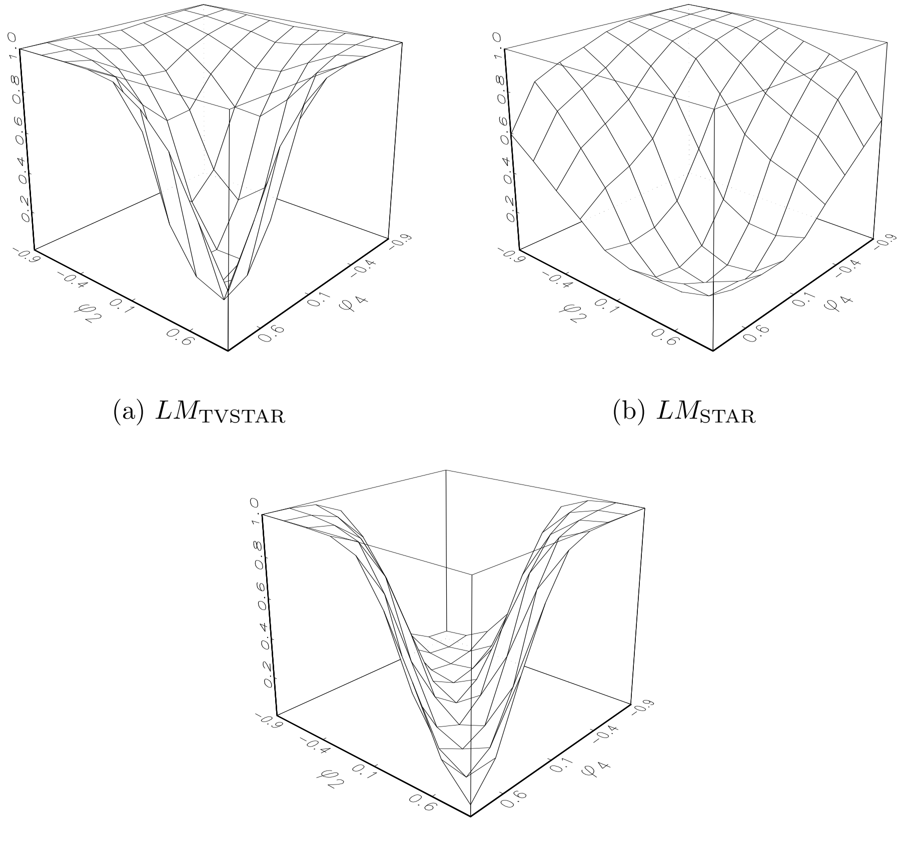


(b) LM_{STAR}



(c) LM_{TV}

Figure 4.1: Rejection frequencies of the appropriate null hypotheses by the various LM-type tests which are part of the general-to-specific procedure outlined in Section 4.2.3. Results are shown for F-variants of the tests, at 5% nominal significance level. Artificial series are generated according to DGP (iv) with $c_2 = 0.50$ as described in Section 4.3.1. The graphs are based on 10000 replications.



(a) LM_{TVSTAR} (b) LM_{STAR}

(c) LM_{TV}

Figure 4.2: Rejection frequencies of the appropriate null hypotheses by the various LM-type tests which are part of the general-to-specific procedure outlined in Section 4.2.3. Results are shown for F-variants of the tests, at 5% nominal significance level. Artificial series are generated according to DGP (v) with $c_2 = 0.50$ as described in Section 4.3.1. The graphs are based on 10000 replications.

STAR model is centered at the middle of the sample, while the power of the LM_{STAR} statistic is highest when the change occurs earlier in the sample.

The rejection frequencies for the various tests in case the DGP is a STAR model with a smoothly changing autoregressive parameter in the 'upper' regime corresponding to $G(y_{t-1}; \gamma_1, c_1) = 1$ [case (v)], with $c_2 = 0.50$, are shown in Figure 4.2.

The power of the $\text{LM}_{\text{TVSTAR}}$ statistic is seen to be an increasing function of the differences between the autoregressive parameters in the two STAR regimes before and after the change, that is, $\phi_2 - \phi_1$ and $\phi_4 - \phi_3$. Changing the parameter c_2 to 0.25 and 0.75 has intuitively plausible effects on the power of this statistic. For example, if the change occurs early in the sample, the difference $\phi_4 - \phi_3$ is much more important than the difference $\phi_2 - \phi_1$. This is even more so for the power of the LM_{STAR} statistic. In case $c_2 = 0.25$, the power of this test is determined almost entirely by the value of ϕ_4 (relative to ϕ_3), while the value of ϕ_2 hardly seems to matter. The power of the LM_{TV} statistic increases as the difference between ϕ_4 and

ϕ_2 increases, which is as expected as well.

4.3.3 Model selection frequencies

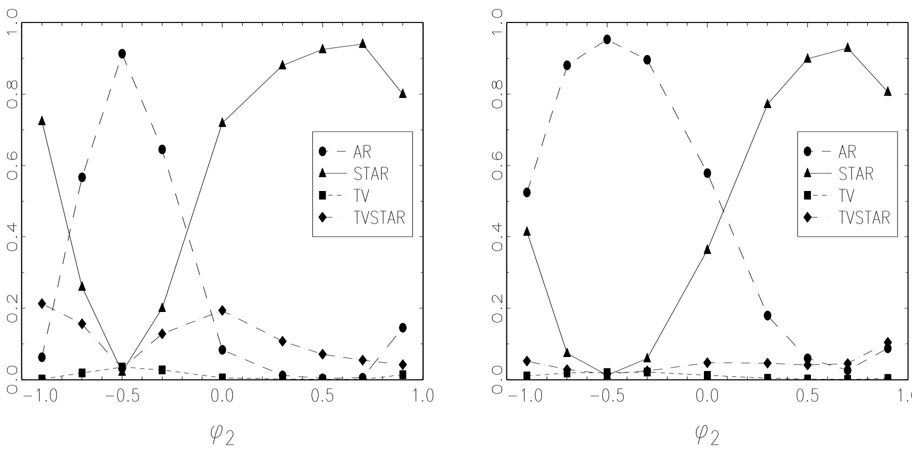
In this section I examine the selection frequencies of the various models when using the decision rules as discussed in section 4.2.2 and 4.2.3. To save space, for the DGPs which involve time-varying parameters, again I only report results for changes which are centered at the middle of the sample ($c_2 = 0.50$). Throughout I use a nominal significance level of 5% to assess the significance of the LM-type statistics that are use in the decision rules.

DGP (ii): LSTAR model. The frequency of selecting the various models using the decision rules in the specific-to-general and the general-to-specific procedures are displayed in Figure 4.3. The correct model is selected slightly more frequently when using the specific-to-general procedure compared to the general-to-specific procedure. The selection of the correct model becomes more frequent as the difference between the autoregressive parameters in the two regimes, $\phi_1 - \phi_2$, increases.

DGP (iii): TV-parameter model. The frequencies of selecting the different models for series generated according to (4.14) with $c_2 = 0.50$ are given in Figure 4.4. Again the true model is selected slightly more frequently for the specific-to-general procedure than for the general-to-specific procedure. The results show that the correct model [TV] is chosen more often as the AR models before and after the change differ to a larger extent. The unreported results for $c_2 = 0.25$ and $c_2 = 0.75$ show that both procedures select an AR model more often when the change occurs early or late in the sample.

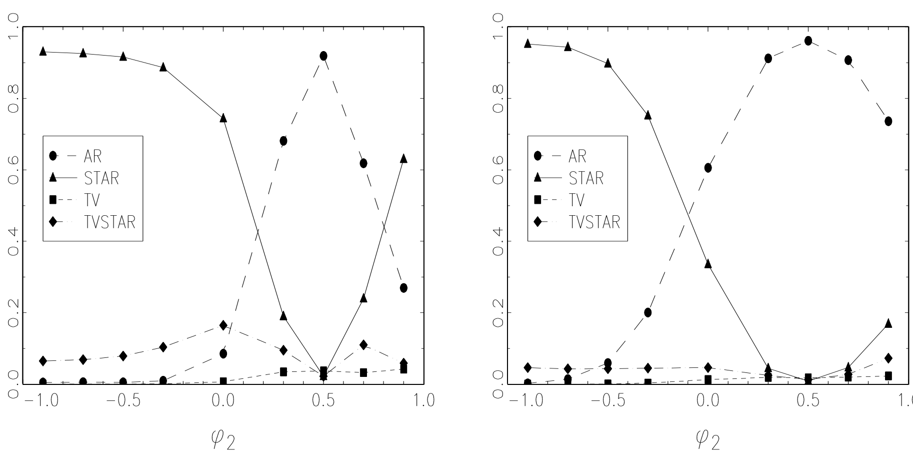
DGP (iv): AR model changing into a STAR model. Model selection frequencies for this DGP are displayed in Figure 4.5. For the general-to-specific procedure, the model selection frequencies correspond with the observations made in the previous subsection for the properties of the LM-type statistics which are involved. For ϕ_3 close to ϕ_4, both procedures tend to select a model with smoothly changing parameters only, especially if $\phi_3 - \phi_1$ and $\phi_4 - \phi_1$ are large (in absolute value). This makes sense intuitively, since the model reduces to a TV-parameter model if $\phi_3 = \phi_4$. As the difference between ϕ_3 and ϕ_4 grows the TV-STAR model is selected more often. The specific-to-general procedure selects the true model more frequently compared to the general-to-specific procedure.

DGP (v): STAR model with changing parameter in upper regime. Model selection frequencies for case (v) are depicted in Figure 4.6. Again the results the general-to-specific procedure conform with what one might expect, given the power surfaces discussed in the previous section. Along the diagonal $\phi_2 = \phi_4$ the model reduces to a STAR as in case (ii). For parameterizations which are close to this restriction, AR and STAR models seem to be selected more frequently. As the difference between ϕ_2 and ϕ_4 grows larger, the time-varying character of this parameter



(a) $\phi_1 = -0.5$, STG

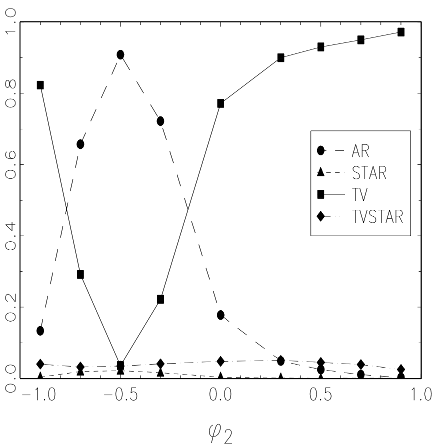
(b) $\phi_1 = -0.5$, GTS



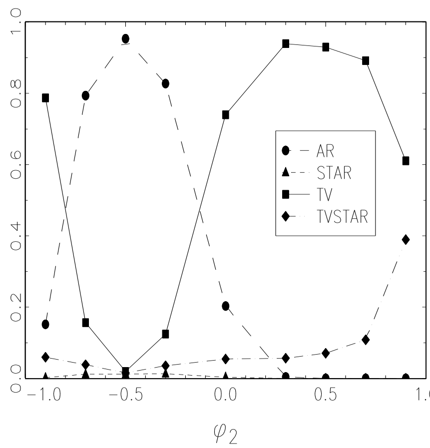
(c) $\phi_1 = 0.5$, STG

(d) $\phi_1 = 0.5$, GTS

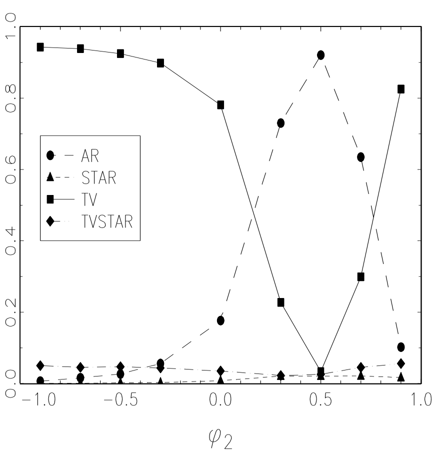
Figure 4.3: Frequencies of selecting the various models using the decision rules in the specific-to-general [STG] and general-to-specific [GTS] procedures as outlined in Sections 4.2.2 and 4.2.3, respectively. Artificial series are generated according to an LSTAR model (4.13) as described in Section 4.3.1. The graphs are based on 10000 replications.



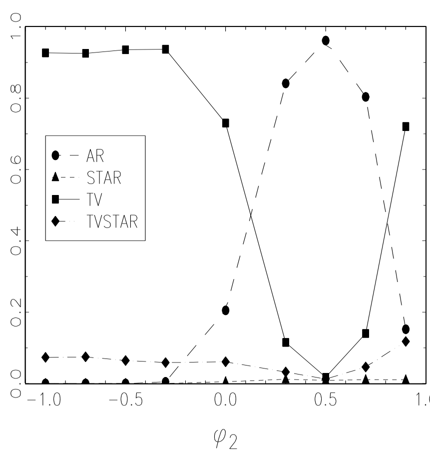
(a) $\phi_1 = -0.5$, STG



(b) $\phi_1 = -0.5$, GTS

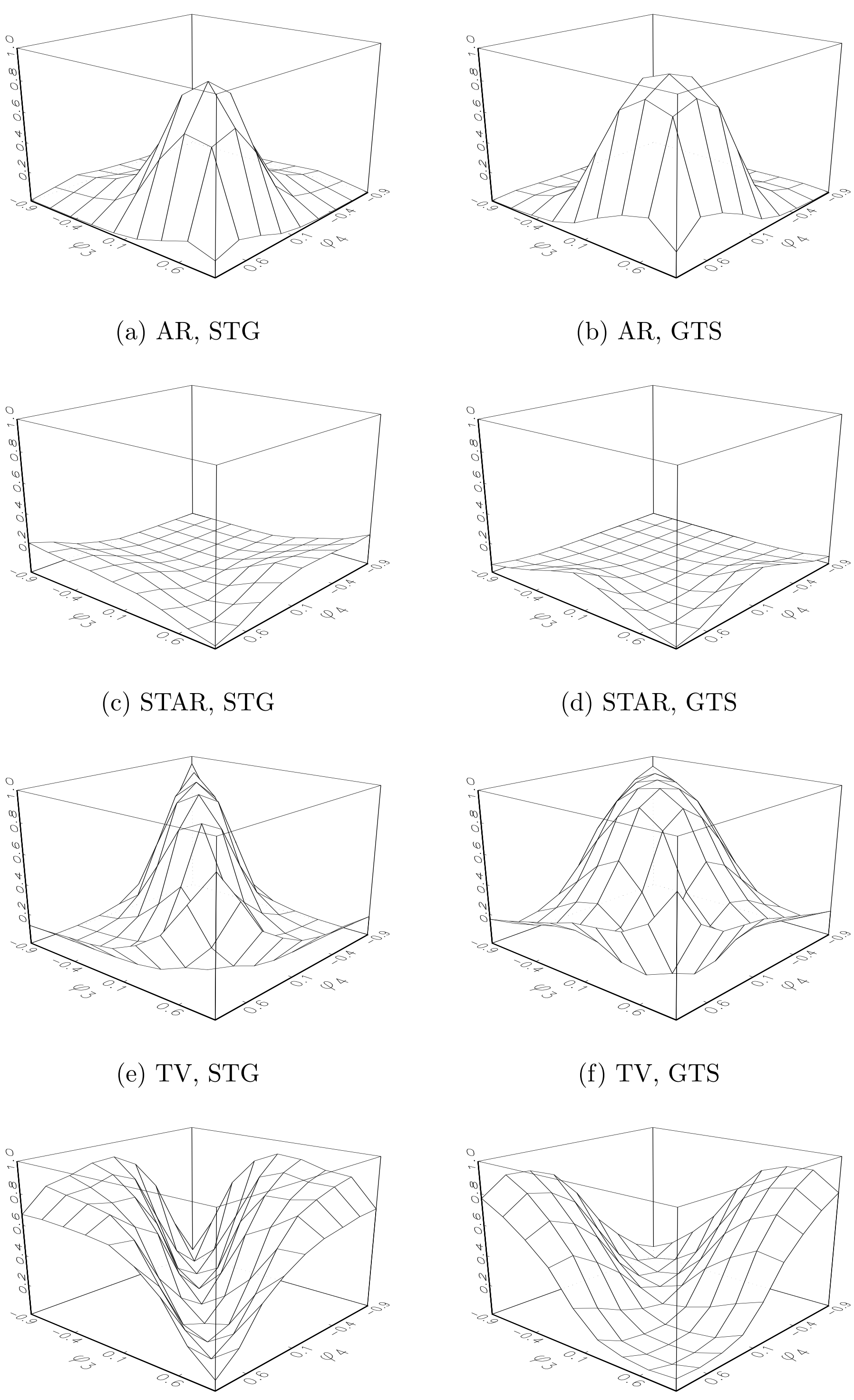


(c) $\phi_1 = 0.5$, STG



(d) $\phi_1 = 0.5$, GTS

Figure 4.4: Frequencies of selecting the various models using the decision rules in the specific-to-general [STG] and general-to-specific [GTS] procedures as outlined in Sections 4.2.2 and 4.2.3, respectively. Artificial series are generated according to a model with time-varying autoregressive parameter as described in Section 4.3.1. The graphs are based on 10000 replications.



(a) AR, STG

(b) AR, GTS

(c) STAR, STG

(d) STAR, GTS

(e) TV, STG

(f) TV, GTS

(g) TV-STAR, STG

(h) TV-STAR, GTS

Figure 4.5: Frequencies of selecting the various models using the decision rules in the specific-to-general [STG] and general-to-specific [GTS] procedures as outlined in Sections 4.2.2 and 4.2.3, respectively. Artificial series are generated according to DGP (iv) with $c_2 = 0.50$ as described in Section 4.3.1. The graphs are based on 10000 replications.

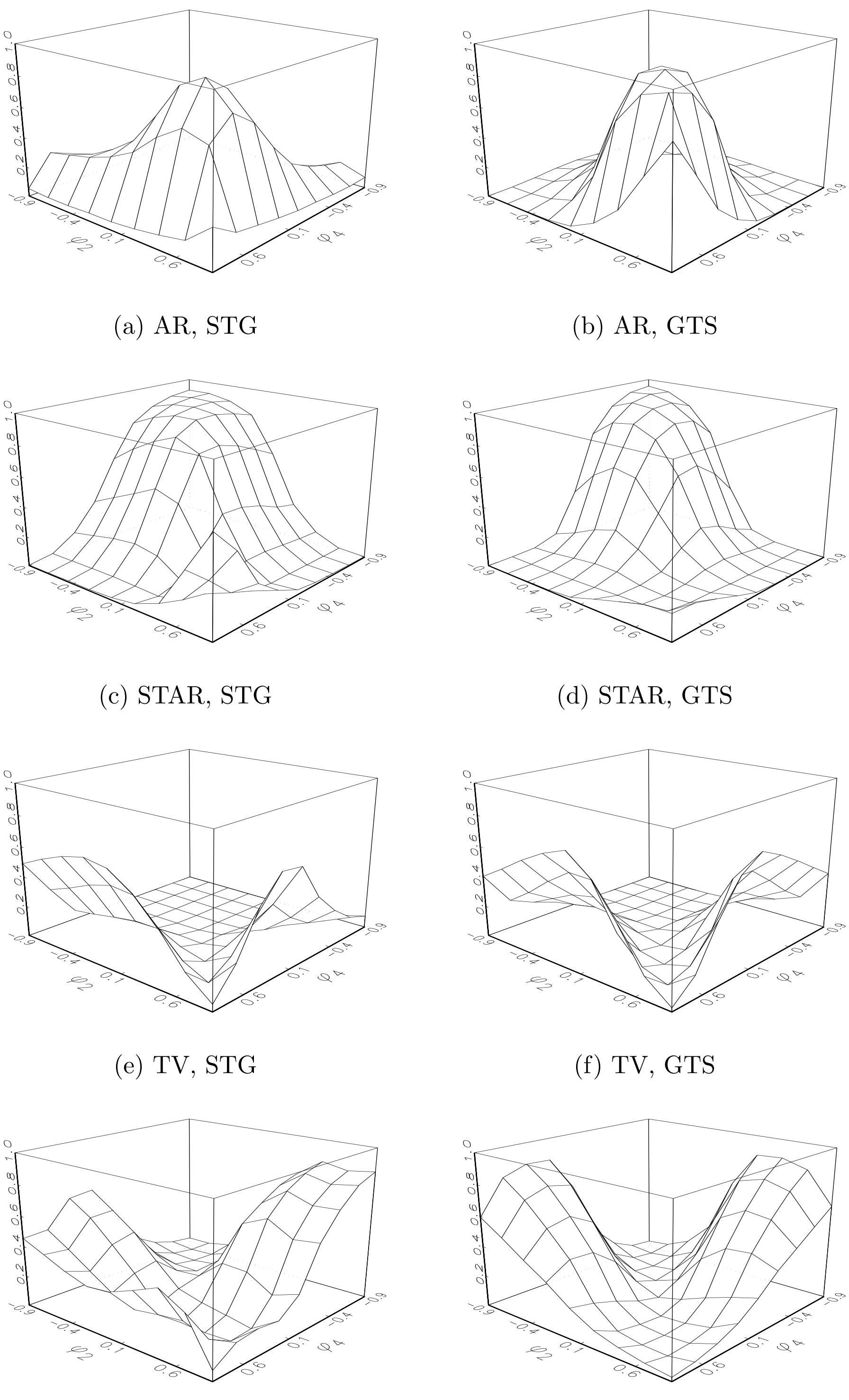
becomes more pronounced, and models which incorporate this are selected more often. When both ϕ_2 and ϕ_4 are reasonably close to ϕ_1 (in which case the nonlinearity is not very strong both before and after the parameter change) a TV-parameter model is selected most frequently. In other cases a TVSTAR model is the preferred specification. Also, the general-to-specific procedure clearly selects the true model more frequently compared to the specific-to-general procedure.

DGP (vi): TV-STAR model with constant degree of nonlinearity. Model selection frequencies for case (vi) are shown in Figure 4.7. The surfaces appear somewhat distorted, as I only investigate combinations of ϕ_2 and ϕ_4 that yield stationary realizations. At first glance, selection of the true model does not seem very frequent for both procedures. However, one has to keep in mind that on the diagonal $\phi_2 = \phi_4$ the model reduces to a STAR model [case (ii)], while the model reduces to a time-varying parameter model [case (iii)] if $\phi_2 = \phi_1 (= 0.5)$. For parameterizations which are close to these restrictions, STAR and TV-parameter models, respectively, tend to be selected most frequently. The specific-to-general procedure seems to select the correct model slightly more often compared to the general-to-specific procedure.

DGP (vii): TV-STAR model. Finally, Figure 4.8 shows model selection frequencies for DGP (vii). Both procedures tend to select the TVSTAR model as the difference between the AR parameters in the two regimes of the STAR models before and after the parameter change are sufficiently far apart. Also, the general-to-specific procedure selects the true model more frequently than the specific-to-general procedure.

The main conclusion from the Monte Carlo experiments is that both specification procedures appear to perform reasonably well. If both the time-variation in the parameters and the nonlinearity are sufficiently pronounced, both procedures are able to detect these features in the time series. If either one of the two characteristics is absent or only weak, the procedures tend to select a more parsimonious model. The relative performance of the two procedures is comparable. The specific-to-general procedure succeeds better in discriminating between TV and TV-STAR models than the general to specific procedure. For example, for almost all series generated from DGP (iv) for which the specific-to-general procedure selects a TV-STAR model while the general-to-specific procedure does not, a model with smoothly changing parameters is selected by the latter procedure, see Figure 4.5. On the other hand, when discriminating between STAR and TV-STAR the relative performance of the two procedures is the opposite, as can be seen in Figure 4.6.

The main advantage of the general-to-specific procedure is that it does not require estimation of a nonlinear or time-varying parameter model. The LM-type tests which are used can all be computed very quickly using simple auxiliary regressions. It therefore appears sensible to use this approach in case one wants to get a fast indication of whether STAR-type nonlinearity and/or time-varying parameters are relevant for a particular time series. Of course, the outcome of this procedure should not be interpreted as the final piece of evidence concerning the most appropriate



(a) AR, STG

(b) AR, GTS

(c) STAR, STG

(d) STAR, GTS

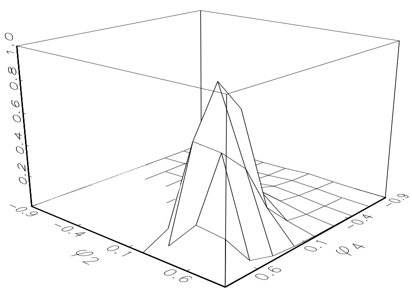
(e) TV, STG

(f) TV, GTS

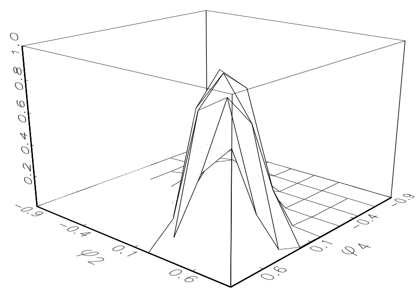
(g) TV-STAR, STG

(h) TV-STAR, GTS

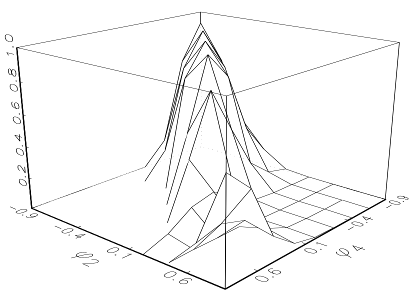
Figure 4.6: Frequencies of selecting the various models using the decision rules in the specific-to-general [STG] and general-to-specific [GTS] procedures as outlined in Sections 4.2.2 and 4.2.3, respectively. Artificial series are generated according to DGP (v) with $c_2 = 0.50$ as described in Section 4.3.1. The graphs are based on 10000 replications.



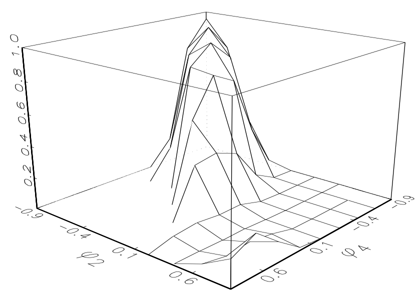
(a) AR, STG



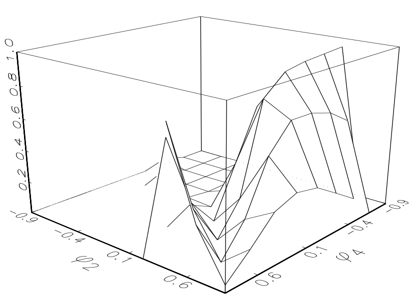
(b) AR, GTS



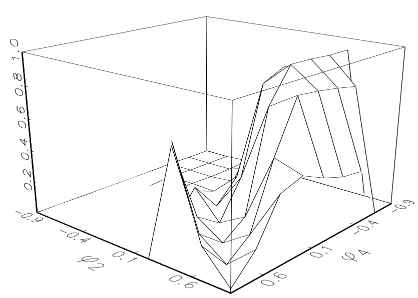
(c) STAR, STG



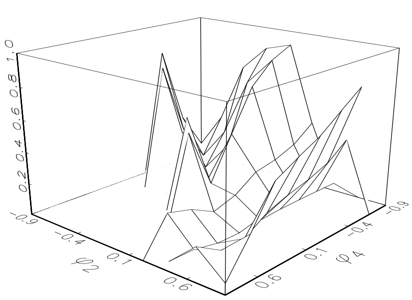
(d) STAR, GTS



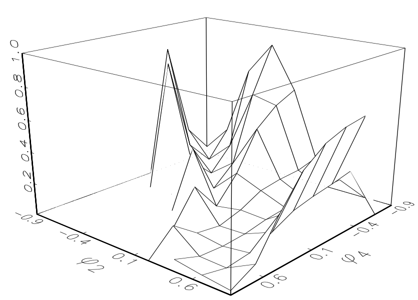
(e) TV, STG



(f) TV, GTS

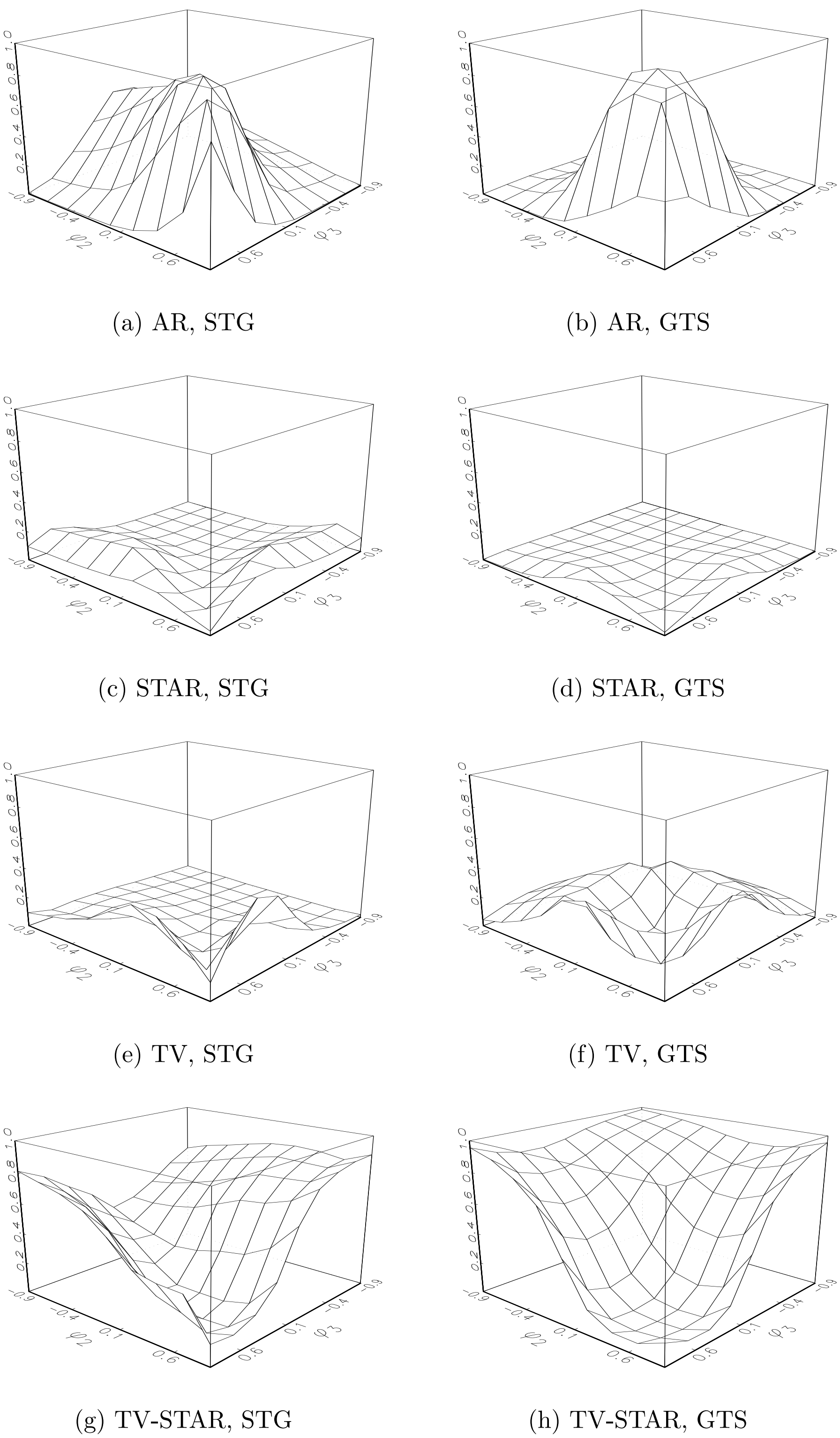


(g) TV-STAR, STG



(h) TV-STAR, GTS

Figure 4.7: Frequencies of selecting the various models using the decision rules in the specific-to-general [STG] and general-to-specific [GTS] procedures as outlined in Sections 4.2.2 and 4.2.3, respectively. Artificial series are generated according to DGP (vi) with $c_2 = 0.50$ as described in Section 4.3.1. The graphs are based on 10000 replications.



(a) AR, STG
(b) AR, GTS
(c) STAR, STG
(d) STAR, GTS
(e) TV, STG
(f) TV, GTS
(g) TV-STAR, STG
(h) TV-STAR, GTS

Figure 4.8: Frequencies of selecting the various models using the decision rules in the specific to general [STG] and general-to-specific [GTS] procedures as outlined in Sections 4.2.2 and 4.2.3, respectively. Artificial series are generated according to DGP (vii) with $c_2 = 0.50$ as described in Section 4.3.1. The graphs are based on 10000 replications.

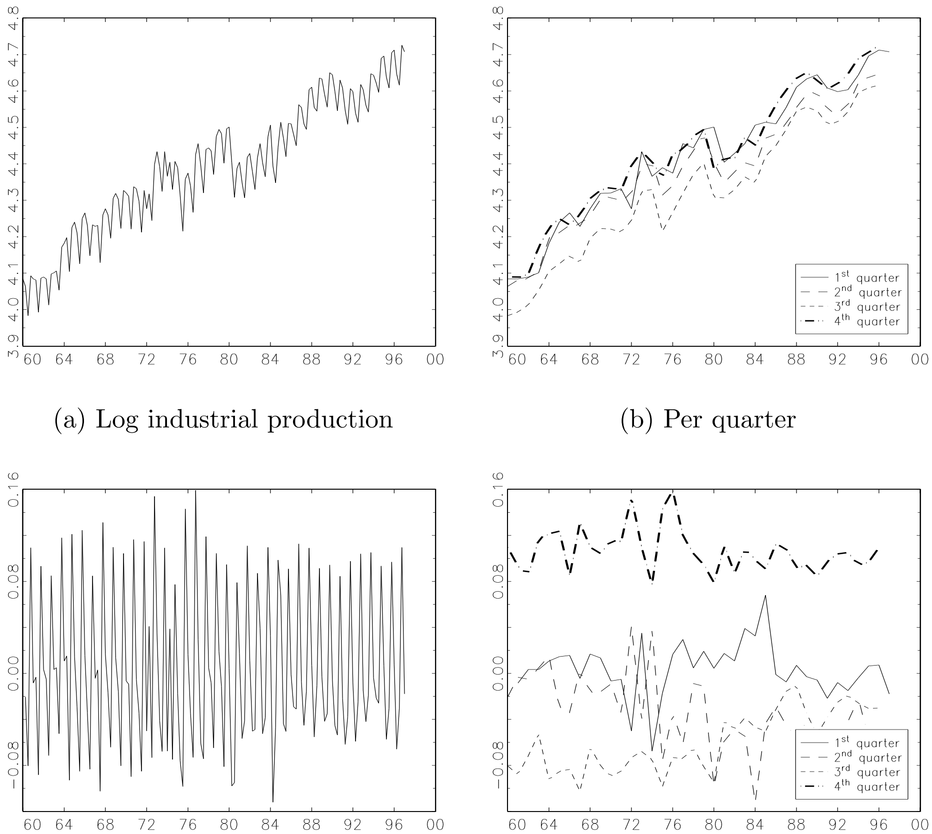
model, but should be followed by a careful specification. In that respect, the specific-to-general procedure might be relevant, as it allows for a more detailed comparison of different sub-models of the TVSTAR model.

4.4 Stability and nonlinearity in UK industrial production

Output series frequently have been examined for both nonlinearity and structural instability. As discussed in Section 3.3, regime-switching models have been used to capture the alleged asymmetry of output over the business cycle. At the same time, tests for stability and models with time varying parameters have been applied to investigate whether certain characteristics of output have changed over time, in particular since World War II or after the first oil crisis in 1973, see Perron (1989). It appears that these two lines of research have developed more or less independently. In applications of regime-switching models, it is common to assume parameter constancy, whereas in applications of stability tests or time-varying parameter models, linearity usually is not questioned. Some attempts have been made to compare regime-switching and linear time-varying parameter models, see Koop and Potter (1998). One of the possible explanations for the fact that instability and nonlinearity usually are considered in isolation is that the instability occurs in the trend and or seasonal characteristics of output series, whereas the nonlinearity is to be found in their cyclical properties. This implicitly assumes that trend, cyclical and seasonal components of these time series are independent.

In this section, I apply TV-STAR type models to quarterly observations on the seasonally unadjusted industrial production index for the United Kingdom. Similar series for a large number of OECD countries have been examined for regime-switching characteristics and time-varying properties before. For example, Luukkonen and Teräsvirta (1991) subject such series to the LM-type tests for STAR nonlinearity and find substantial evidence that regime-switching behaviour indeed might be an important characteristic of these series. Teräsvirta and Anderson (1992), Granger *et al.* (1993) and Teräsvirta *et al.* (1994) build upon these results and specify and estimate STAR models. Interestingly, Lin and Teräsvirta (1994) find evidence for structural instability in the industrial production index for the Netherlands and show that a model with time-varying parameters offers an alternative representation of this series.

The series I consider here are quarterly, seasonally unadjusted indexes of industrial production for the UK, obtained from the *OECD Main Economic Indicators*, covering the period from 1960:1 until 1997:1. Figure 4.9 shows various transformations of the industrial production index. From the graphs in panels a) and b), it is apparent that average growth decreased after 1975, say, while the seasonal patter changed markedly around the same time. The latter is confirmed by the graphs in panels c) and d), which depict the (log-)level and quarterly growth rates for each quarter separately. In fact, the seasonality in this series is quite dominant, and hides other features such as possible asymmetry in growth rates during expansions



(a) Log industrial production (b) Per quarter

(c) Quarterly growth rate (d) Per quarter

Figure 4.9: Quarterly, seasonally unadjusted index of industrial production for the United Kingdom, 1960:1-1997:1

and recessions. The latter property is brought out more clearly in the seasonally differenced series $\Delta_4 y_t = y_t - y_{t-4}$, with y_t the log of the index. This transformation is shown in Figure 4.10. This graph suggest that periods of positive growth (expansions) are longer on average than periods of negative growth (contractions). Also, the change from negative to positive growth (recovery) appears to occur faster than the opposite.

In the analysis below I focus on the properties of the quarterly growth rate $\Delta y_t = y_t - y_{t-1}$ and attempt to specify TV-STAR type models for this series. Panel (c) of Figure 4.9 shows that Δy_t is quite volatile and displays pronounced seasonal variation, suggesting that this variable is not a suitable indicator of the state of the economy. Hence, if the nonlinearity is thought to be associated with different phases of the business cycle, lagged quarterly growth rates are less useful as transition variable in the TV-STAR model. I therefore assume that the variable which characterizes the nonlinearity in the series is a lagged yearly growth rate, that is, $s_t = \Delta_4 y_{t-d}$. Furthermore, following Canova and Ghysels (1994), among others, I assume that the seasonal variation in Δy_t can be captured by means of seasonal dummies $D_{s,t}$, $s = 1, 2, 3, 4$, where $D_{s,t}$ takes a value of 1 in quarter s and a value

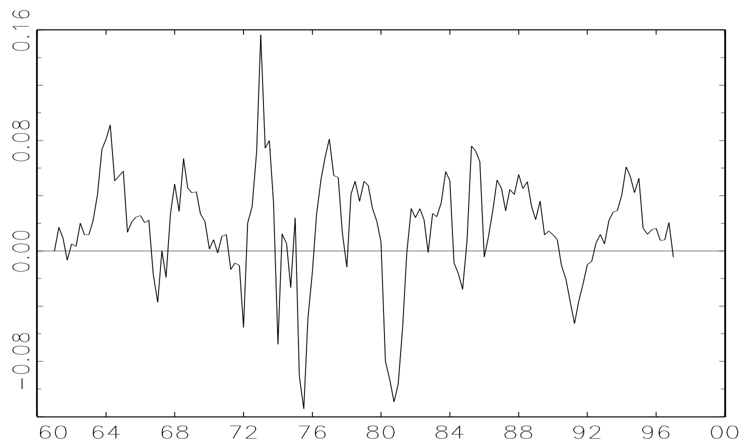


Figure 4.10: Seasonal differences of logarithmic transformed index of industrial production for the UK, 1960:1-1997:1

of 0 in other quarters.

As usual, the model building process begins with specifying a linear AR(p) model, which in this case includes seasonal intercepts. When AIC and SIC are applied to determine the appropriate order p, AIC selects $p = 8$, whereas SIC prefers $p = 1$. Given that the time series consists of only 145 observations, $p = 8$ is too large an order, as the TV-STAR model would contain too many parameters to be estimated. On the other hand, the AR(1) model suggested by SIC is inadequate, in the sense that the residuals exhibit very strong autocorrelation. It appears that an AR(4) model is sufficient to remove this residual autocorrelation, at least at the first 10 lags. This model therefore is used as the basis for the specification procedure.

To obtain a quick impression of whether the TV-STAR model might be appropriate for this time series, I first apply the LM-type test statistics that are used in the general-to-specific procedure discussed in Section 4.2.3. The tests are computed with $\Delta_4 y_{t-d}$ for $d = 1, \ldots, 4$ as transition variable in the nonlinear part. Columns 2-4 of Table 4.2 contain p-values of the LM-type statistics. It is seen that for all choices of d, linearity can be rejected in favor of the TV-STAR alternative at the 10% significance level. The p-values of the LM_{STAR} and LM_{TV} statistics suggest that a model with time-varying parameters might be sufficient, as the null hypothesis of LM_{STAR} can not be rejected.

In addition, Table 4.2 contains p-values for variants of the $\text{LM}_{\text{TVSTAR}}$ and LM_{STAR}, and LM_{TV} statistics, which test for linearity and constancy of the parameters corresponding with either the lagged dependent variables or the seasonal dummies only. The results in columns 5-7 suggest that the seasonality evolves over time and varies perhaps with the business cycle, as the null hypothesis of parameter constancy (LM_{TV})for these parameters is convincingly rejected, whereas the null hypothesis of linearity (LM_{STAR}) only attains a fairly small p-value when $s_{1t} = \Delta_4 y_{t-1}$. For the lagged dependent variables, the test results suggest that the corresponding parameters are constant over time and are not different in different phases of the business cycle, as none of the test statistics is significant.

Table 4.2: p-values of tests for TVSTAR-type nonlinearity in quarterly growth rates of UK industrial production

	All			$D_{s,t}$			Δy_{t-j}		
d	TVSTAR	STAR	TV	TVSTAR	STAR	TV	TVSTAR	STAR	TV
1	0.001	0.120	0.003	0.001	0.102	0.000	0.297	0.336	0.591
2	0.030	0.763	0.010	0.013	0.995	0.004	0.674	0.856	0.487
3	0.090	0.972	0.038	0.010	0.969	0.002	0.651	0.827	0.413
4	0.011	0.471	0.010	0.001	0.420	0.002	0.652	0.828	0.547

p-values of F variants of the LM-type statistics used in the general-to-specific procedure for TV-STAR models, based on first-order Taylor approximations of the logistic transition function. The test statistics are computed by considering $s_{2,t} = t$ and $s_{1,t} = \Delta_4 y_{t-d}$ with $d = 1, \ldots, 4$. The series are log first differences of quarterly observations on the industrial production index for the UK (1990=100), taken from the OECD *Main Economic Indicators*. The sample period covers 1960:1-1997:1.

Table 4.3: Tests for STAR-type nonlinearity in quarterly growth rates of UK industrial production

	All	$D_{s,t}$	Δy_{t-j}
$\Delta_4 y_{t-1}$	0.005	0.022	0.002
$\Delta_4 y_{t-2}$	0.500	0.702	0.526
$\Delta_4 y_{t-3}$	0.659	0.521	0.381
$\Delta_4 y_{t-4}$	0.047	0.077	0.155
t	0.002	0.000	0.004

p-values of F variants of the LM_3 statistic against the alternative of a STAR model with all variables (All), only the seasonal dummies ($D_{s,t}$), or only lagged dependent variables (Δy_{t-j}) entering nonlinearly. Test are computed with an AR(4) model under the null hypothesis.

Table 4.4: Diagnostic tests for no remaining nonlinearity in TV-parameter model for quarterly growth rates of UK industrial production

	All		$D_{s,t}$		Δy_{t-j}	
	$\text{LM}_{\text{AMR},3}$	$\text{LM}_{\text{EMR},3}$	$\text{LM}_{\text{AMR},3}$	$\text{LM}_{\text{EMR},3}$	$\text{LM}_{\text{AMR},3}$	$\text{LM}_{\text{EMR},3}$
$\Delta_4 y_{t-1}$	0.024	0.025	0.393	0.051	0.132	0.054
$\Delta_4 y_{t-2}$	0.631	0.354	0.856	0.855	0.861	0.654
$\Delta_4 y_{t-3}$	0.433	0.031	0.343	0.861	0.477	0.821
$\Delta_4 y_{t-4}$	0.061	0.160	0.030	0.342	0.148	0.398
t	0.187	0.129	0.117	0.129	0.225	0.206

p-values of F variants of the LM_{AMR} and LM_{EMR} diagnostic tests statistic for no remaining nonlinearity for all parameters (All), only for the seasonal dummies ($D_{s,t}$), or only for lagged dependent variables (Δy_{t-j}).

The LM-type tests in the general-to-specific approach provide at least some evidence in favor of nonlinearity and/or structural instability in the industrial production series. Therefore I attempt to determine an appropriate model by applying the specific-to-general approach. First, linearity is tested against the alternatives of STAR-type nonlinearity and smoothly changing parameters. Again the tests also are computed allowing only the parameters corresponding with the seasonal dummies or the lagged dependent variables to enter the nonlinear or time-varying part of the alternative model. Table 4.3 contains p-values of the LM_3 statistic. Linearity is rejected most convincingly against the alternative of smoothly changing parameters and the alternative of a STAR model with $s_t = \Delta_4 y_{t-1}$. Note that now linearity and parameter constancy also are rejected in case only the parameters corresponding with the lagged dependent variables are tested. This should not be taken as direct evidence against the null hypothesis of constancy or linearity of these parameters however, as the significant test statistics might be caused by the fact that the seasonal dummies are assumed to be constant and linear under the alternative as well.

Given the small p-value of the tests in case time is used as transition variable, I proceed with estimating a models with smoothly changing parameters, initially allowing all parameters to vary over time. Full details of the estimated model are not shown here. Instead, Table 4.4 contains p-values of the LM_{AMR} test for no remaining nonlinearity of Eitrheim and Teräsvirta (1996) and the LM_{EMR} test developed in Section 3.2.1, both based on a third-order Taylor approximation of the second transition function in the model under the alternative. Again three variants of the statistics are computed, testing for no remaining nonlinearity in all parameters, only the parameters corresponding with the seasonal dummies and only the lagged dependent variables. The entries in the table suggest that the null hypothesis can be rejected at reasonable significance levels, in case $\Delta_4 y_{t-1}$ is used as transition variable.

Based on these results, I estimate a TV-STAR model with $\Delta_4 y_{t-1}$ determining the regimes in the STAR part of the model. After sequentially deleting the least significant lagged dependent variables until all parameters have t-statistics larger than 1 in absolute value, the estimated model is given by

$$\begin{aligned}
\Delta y_t = [(&\underset{(0.024)}{0.075} - \underset{(0.228)}{0.479}\,\Delta y_{t-1} - \underset{(0.263)}{0.403}\,\Delta y_{t-2} \\
&- \underset{(0.046)}{0.076}\,D_{1,t} - \underset{(0.043)}{0.025}\,D_{2,t} - \underset{(0.019)}{0.200}\,D_{3,t}) \times (1 - G_1(\Delta_4 y_{t-1})) \\
&+ (\underset{(0.025)}{0.045} + \underset{(0.280)}{0.766}\,\Delta y_{t-1} - \underset{(0.223)}{0.318}\,\Delta y_{t-3} + \underset{(0.237)}{0.926}\,\Delta y_{t-4} \\
&- \underset{(0.039)}{0.122}\,D_{1,t} - \underset{(0.038)}{0.056}\,D_{2,t} + \underset{(0.046)}{0.052}\,D_{3,t})] \times G_1(\Delta_4 y_{t-1})] \times [1 - G_2(t)] \\
&+ [(\underset{(0.020)}{0.160} + \underset{(0.141)}{0.277}\,\Delta y_{t-1} + \underset{(0.121)}{0.209}\,\Delta y_{t-3} - \underset{(0.139)}{0.226}\,\Delta y_{t-4} \\
&- \underset{(0.032)}{0.185}\,D_{1,t} - \underset{(0.024)}{0.170}\,D_{2,t} - \underset{(0.031)}{0.279}\,D_{3,t}) \times (1 - G_1(\Delta_4 y_{t-1})) \\
&+ (\underset{(0.010)}{-0.078} - \underset{(0.148)}{0.222}\,\Delta y_{t-1} - \underset{(0.148)}{0.356}\,\Delta y_{t-2} \\
&- \underset{(0.021)}{0.064}\,D_{1,t} - \underset{(0.024)}{0.096}\,D_{2,t} - \underset{(0.011)}{0.119}\,D_{3,t})] \times G_1(\Delta_4 y_{t-1})] \times G_2(t) + \hat{\varepsilon}_t,
\end{aligned} \tag{4.15}$$

$$G_1(\Delta_4 y_{t-1}) = (1 + \exp\{-\underset{(-)}{28.53}\,(\Delta_4 y_{t-1} - \underset{(0.002)}{0.001})/\sigma_{\Delta_4 y_{t-1}}\})^{-1}, \tag{4.16}$$

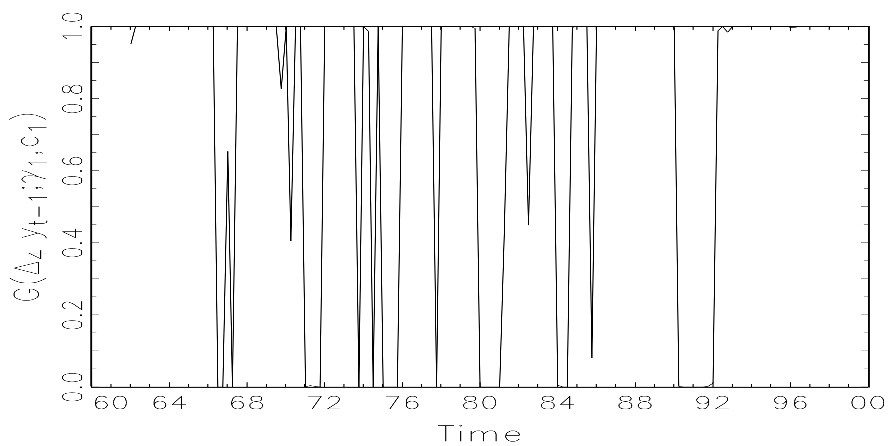
$$G_2(t) = (1 + \exp\{-\,\underset{(-)}{7.02}\,(t/T - \underset{(0.022)}{0.457})/\sigma_{t/T}\})^{-1}. \tag{4.17}$$

$\hat{\sigma}_\varepsilon = 0.017$, $\mathrm{S}_{\hat{\varepsilon}} = -0.16(0.21)$, $\mathrm{K}_{\hat{\varepsilon}} = 4.59(0.00)$, JB $= 15.20(0.00)$, ARCH(1) $= 0.22(0.64)$, ARCH(4) $= 5.85(0.21)$, AIC $= -7.715$, BIC $= -7.082$.

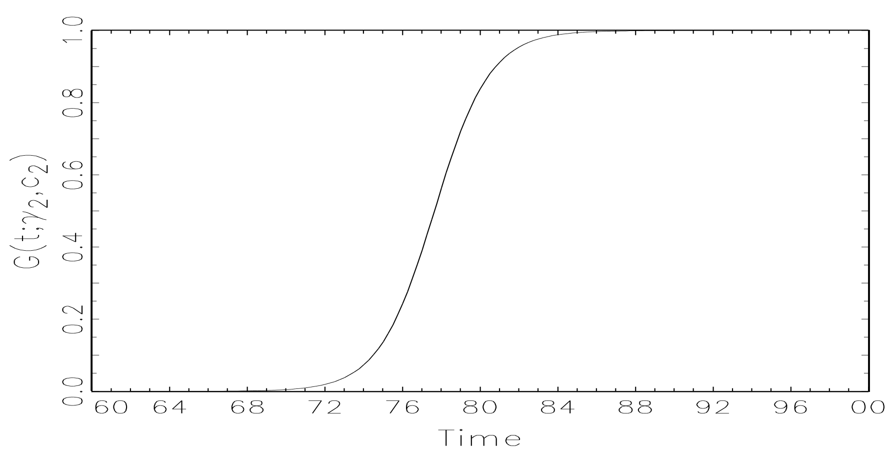
The estimate of c_1 is very close to zero, suggesting that the periods during which $G_1(\Delta_4 y_{t-1}; \gamma_1, c_1) = 0$ and 1 correspond quite closely with recessions and expansions, respectively. The estimate of γ_1 is fairly large, suggesting that transitions between the recession and expansion regimes occur rapidly. These observations are confirmed by the upper panel of Figure 4.11, which shows how estimated transition functions evolve over time. The lower panel of this figure shows that the transition of $G_2(t; \gamma_2, c_2)$ from 0 to 1 takes about 12 years in total, and occurs between 1972 and 1984.

The standard deviation of the residuals from the TV-STAR model is 29% and 15% less than the standard deviation of the residuals from the linear model and the model with smoothly changing parameters only, respectively. Especially the latter demonstrates that allowing for both time-varying and regime-switching behaviour is important for this series. Finally, AIC prefers the TV-STAR model, whereas BIC prefers the parsimonious AR(4) model[2].

[2] The values of the information criteria are AIC $= -7.349$, BIC $= -7.180$ for the linear model, and AIC $= -7.557$, BIC $= -7.177$ for the model with smoothly changing parameters.



(a) Transition function $G_1(\Delta_4 y_{t-1}; \gamma_1, c_1)$ versus time



(b) Transition function $G_2(t; \gamma_2, c_2)$ versus time

Figure 4.11: Transition functions in TV-STAR model for quarterly growth rates of UK industrial production.

4.5 Concluding remarks

In this chapter I have considered a model, based on the principle of smooth transition, which allows for regime-switching behaviour in conjunction with time-varying properties. This TV-STAR might be a useful tool to capture nonlinearity and structural instability simultaneously. The TV-STAR model model arises as a special case of the general MRSTAR model considered in the previous chapter. The specific-to-general procedure put forward there for the general MRSTAR can be readily applied to specify TV-STAR models. However, the special character of the TV-STAR model makes a general-to-specific procedure an attractive alternative. The Monte Carlo simulations demonstrated that the performance of the two specification procedures is comparable. Because the LM-type tests that are used in the general-to-specific procedure only require estimation of linear models, they can be used to obtain a quick impression of the importance of nonlinearity and/or structural instability for a particular time series. The application of the TV-STAR model to growth rates in UK industrial production demonstrated the importance of allowing for both nonlinear and time-varying characteristics. An interesting topic for further research is to examine whether the TV-STAR model is useful to describe industrial production series for other OECD countries as well.

Chapter 5

Multivariate Smooth Transition Models

In previous chapters attention has been restricted to univariate smooth transition models. Even though the general model as given in (2.1) allows for exogenous variables to enter the model, either as regressors or as the transition variable, the main purpose of the model is to describe and forecast a single variable. Sometimes it may be worthwhile to model several time series jointly, to exploit possible linkages that exist between them. In the context of empirical macro-econometrics, such models might be useful to examine whether the relationship between variables displays regime-switching characteristics and, for example, is different in different phases of the business cycle, see Diebold and Rudebusch (1996), Koop *et al.* (1996), Ravn and Sola (1995) and Weise (1999).

The interest in multivariate nonlinear modeling has started to develop only recently, and mainly has been application-oriented. The relevant statistical theory has not been fully developed yet and is a topic of much current research. The purpose of this chapter is two-fold. First, I consider representation and specification of a multivariate smooth transition autoregressive [STAR] model at a quite general level. In that sense, this chapter complements Krolzig (1997) and Tsay (1998), who treat similar issues for multivariate Markov-switching and threshold models, respectively. Second, I explore in somewhat more detail so-called smooth transition equilibrium correction models, which have proven useful in several applications. These models can describe situations where certain variables are linked by a linear (long-run) equilibrium relation, whereas adjustment toward this equilibrium is nonlinear.

The chapter is organized as follows. In Section 5.1, representation of the general multivariate STAR model is discussed. A bivariate example is used to derive a model with smooth transition equilibrium correction. In Section 5.2, a specification procedure for multivariate STAR models is outlined, which is based upon the specification procedure for univariate models discussed in Chapter 2. The main focus in this section is on the size and power properties of multivariate LM-type tests against STAR nonlinearity, which are investigated by means of Monte Carlo simulation. The concept of common nonlinearity also is touched upon. Section 5.3 contains an application of smooth transition equilibrium correction to spot and futures prices of the FTSE100 index.

5.1 Representation

Let $\mathbf{y}_t = (y_{1t}, \ldots, y_{kt})'$ be a $(k \times 1)$ vector time series. A k-dimensional analogue of the univariate 2-regime STAR model (2.2) then can be specified as

$$\mathbf{y}_t = (\mathbf{\Phi}_{1,0} + \mathbf{\Phi}_{1,1}\mathbf{y}_{t-1} + \cdots + \mathbf{\Phi}_{1,p}\mathbf{y}_{t-p})(1 - G(s_t; \gamma, c)) \\ + (\mathbf{\Phi}_{2,0} + \mathbf{\Phi}_{2,1}\mathbf{y}_{t-1} + \cdots + \mathbf{\Phi}_{2,p}\mathbf{y}_{t-p})G(s_t; \gamma, c) + \boldsymbol{\varepsilon}_t, \quad (5.1)$$

where $\mathbf{\Phi}_{i,0}$, $i = 1, 2$, are $(k \times 1)$ vectors, $\mathbf{\Phi}_{i,j}$, $i = 1, 2$, $j = 1, \ldots, p$, are $(k \times k)$ matrices, $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{kt})'$ is a k-dimensional vector white noise process with mean zero and $(k \times k)$ covariance matrix $\mathbf{\Sigma}$. As before, the transition function $G(s_t; \gamma, c)$ is assumed to be a continuous function bounded between zero and one, with γ and c parameters determining the smoothness and location of the change in the value of $G(s_t; \gamma, c)$. The transition variable s_t can be a lagged value of one the time series contained in $\mathbf{y}_t$, a linear combination of the k series, an exogenous variable z_t, or (a function of) a deterministic trend t. The latter choice renders a model with parameters that change smoothly over time.

Notice that in (5.1) it is assumed that the regimes are common to the k variables, in the sense that one and the same transition function determines the prevailing regime and the switches between regimes in all k equations of the model. It is straightforward to generalize the model to incorporate equation-specific transition functions $G_1(s_{1t}; \gamma_1, c_1), \ldots, G_k(s_{kt}; \gamma_k, c_k)$ and thus allow for equation-specific regime-switching behaviour. However, the case of common regime-switching appears most relevant in practice, and I therefore restrict attention to the model as specified in (5.1).

Judging from applications of multivariate regime-switching models that are available at present, it seems that a model of particular interest is one in which the components of $\mathbf{y}_t$ are linked by a linear long-run equilibrium relationship, whereas adjustment towards this equilibrium is nonlinear and can be characterized as regime-switching, with the regimes determined by the size and/or sign of the deviation from equilibrium. This class of so-called nonlinear equilibrium correction models is discussed next.

Smooth transition equilibrium correction

Many economic variables, while nonstationary individually, are linked by long-run equilibrium relationships, such that they tend to move together in the long-run. The concept of cointegration, introduced by Granger (1981) and Engle and Granger (1987), together with the corresponding equilibrium correction models [EqCMs], allows these characteristics to be modeled simultaneously[1]. In the standard EqCM,

[1]See Banerjee, Dolado, Galbraith and Hendry (1993), Johansen (1995), Hatanaka (1996) and Boswijk (1999) for in-depth treatments of cointegration and equilibrium correction models.

adjustment towards the long-run equilibrium is linear, in the sense that it is always present and of the same strength under all circumstances. There are however economic situations for which the validity of this assumption might be questioned and it might be worthwhile to explore generalizations of the linear framework, see Granger and Lee (1989) for an early attempt. Here I concentrate on incorporating the smooth transition mechanism in an EqCM to to allow for nonlinear or asymmetric adjustment, see Granger and Swanson (1996) for a more general discussion of nonlinear extensions of cointegration and EqCMs.

It appears that relevant forms of nonlinear equilibrium correction often concern some sort of asymmetry, that is, distinction is to be made between adjustment of positive and negative or between adjustment of large and small deviations from equilibrium[2]. Both types of asymmetry arise in a natural way when modeling prices of so-called equivalent assets in financial markets, see Yadav, Pope and Paudyal (1994) and Anderson (1997) for elaborate discussions[3]. Equivalent assets in a certain sense represent the same underlying value; examples of equivalent assets include stocks and futures, and bonds of different maturity. Since they are traded in the same market, or in markets which are linked by arbitrage-related forces, the prices of equivalent assets should be such that investors are indifferent between holding either one of them. If prices deviate from equilibrium, arbitrage opportunities are created which will result in the prices being driven back together again. However, market frictions can give rise to asymmetric adjustment of such deviations. Due to short-selling constraints, for example, the response to negative deviations from equilibrium might be different from the response to positive deviations. Alternatively, transaction costs prevent adjustment of equilibrium errors as long as the benefits from adjustment, which equal the price difference, are smaller than those costs. These market frictions suggest that the degree of equilibrium correction is a function of the sign and/or size of the deviation from equilibrium. Such asymmetric adjustment can be described by means of smooth transition models, as shown below.

Nonstationary variables that are linked by long-run equilibrium relations can be characterized in several different ways. As noted by Granger and Swanson (1996), in the traditional linear framework it is usual practice to start by defining the long-run equilibrium relationship that is presumed to hold between the variables involved, followed by deriving the corresponding equilibrium correction model and, finally, the representation for the permanent and transitory components of which the variables are composed. For nonlinear generalizations, it is more convenient to reverse this sequence. Consider therefore the series z_t and w_t, generated by a STAR model and

---

[2]In most applications of nonlinear equilibrium correction models, the equilibrium relationship(s) still is assumed to be linear. Even though this assumption often can be justified on grounds of economic theory, of course it can be relaxed, see Granger and Hallman (1991). Also, other forms of nonlinear equilibrium correction, which do not depend directly on the deviation from equilibrium itself, are possible, see Siklos and Granger (1997) for an example.

[3]See Escribano and Pfann (1998) for an alternative theoretical motivation for nonlinear equilibrium correction based on intertemporal choice models with asymmetric adjustment costs.

a random walk model, respectively, as

$$z_t = \rho_{1,1}(1 - G(z_{t-1};\gamma,c))z_{t-1} + \rho_{2,1}G(z_{t-1};\gamma,c)z_{t-1} + \nu_t, \tag{5.2}$$

$$w_t = w_{t-1} + \eta_t, \tag{5.3}$$

where

$$\begin{pmatrix}\nu_t \\ \eta_t\end{pmatrix} \sim \text{i.i.d. } (0, \mathbf{\Sigma}), \qquad \mathbf{\Sigma} = \begin{pmatrix}1 & \theta\sigma \\ \theta\sigma & \sigma^2\end{pmatrix}. \tag{5.4}$$

The variables z_t and w_t are not observed, but are assumed to be linked to the observed variables y_{1t} and y_{2t} through

$$z_t = y_{1t} - \beta y_{2t}, \tag{5.5}$$

$$w_t = y_{1t} - \alpha y_{2t}, \tag{5.6}$$

with $\alpha \neq \beta$, or

$$y_{1t} = \frac{1}{\alpha-\beta}(\alpha z_t - \beta w_t), \tag{5.7}$$

$$y_{2t} = \frac{1}{\alpha-\beta}(z_t - w_t), \tag{5.8}$$

The time series y_{1t} and y_{2t} thus consist of two components, a permanent random walk component[4] w_t and a nonlinear transitory component z_t.

The standard linear set-up, which is used by, Banerjee, Dolado, Hendry and Smith (1986) and Engle and Granger (1987), among others is obtained by taking $\rho_{1,1} = \rho_{2,1} \equiv \rho$ in (5.2) and imposing the restriction $|\rho| < 1$. In that case, the series y_{1t} and y_{2t} are cointegrated with cointegrating vector $(1, -\beta)'$. Put differently, the series y_{1t} and y_{2t} are linked by the (long-run) equilibrium relationship $y_{1t} = \beta y_{2t}$, and z_t represents the deviation from this equilibrium.

In the general set-up given above, z_t is assumed to follow a STAR model. To retain the interpretation of $y_{1t} = \beta y_{2t}$ as long-run equilibrium, z_t has to be stationary. This implies that, depending on the specific form of the function $G(z_{t-1};\gamma,c)$, certain restrictions have to be put on $\rho_{1,1}$ and $\rho_{2,1}$. For example, in case $s_t = z_{t-1}$, $|\rho_{1,1}| < 1$ and $|\rho_{2,1}| < 1$ are sufficient conditions for z_t to be stationary for all possible choices of $G(z_{t-1};\gamma,c)$ which are bounded between zero and one, see Section 2.1.

Using (5.2)-(5.5), (5.7) and (5.8) can be rewritten in equilibrium correction format as

$$\Delta y_{1t} = \frac{\alpha}{\alpha-\beta}\left[\rho_{1,1}(1 - G(z_{t-1})) + \rho_{2,1}G(z_{t-1}) - 1\right] z_{t-1} + \varepsilon_{1t}, \tag{5.9}$$

$$\Delta y_{2t} = \frac{1}{\alpha-\beta}\left[\rho_{1,1}(1 - G(z_{t-1})) + \rho_{2,1}G(z_{t-1}) - 1\right] z_{t-1} + \varepsilon_{2t}, \tag{5.10}$$

where $\varepsilon_{1t} = (\alpha\nu_t - \beta\eta_t)/(\alpha-\beta)$ and $\varepsilon_{2t} = (\nu_t - \eta_t)/(\alpha-\beta)$. From (5.9) and (5.10), the meaning of the term smooth transition equilibrium correction is obvious. For

[4]The assumption that w_t follows a random walk is made for convenience here. See Granger and Swanson (1996) for other possibilities.

example, in the equation for Δy_{1t}, the strength of equilibrium correction changes smoothly from $\alpha(\rho_{1,1}-1)/(\alpha-\beta)$ to $\alpha(\rho_{2,1}-1)/(\alpha-\beta)$ as $G(z_{t-1})$ changes from 0 to 1.

The function $G(z_{t-1};\gamma,c)$ can be used to obtain different kinds of nonlinear equilibrium correction behaviour. In empirical applications, in particular those involving financial variables, one might be especially interested in modeling asymmetric adjustment, as argued before. In this case, the regimes are defined directly in terms of the deviations from equilibrium[5]. Asymmetric effects of positive and negative deviations from equilibrium can be obtained by taking $G(z_{t-1};\gamma,c)$ to be the logistic function

$$G(z_{t-1};\gamma,c) = \frac{1}{1+\exp\{-\gamma(z_{t-1}-c)\}}, \qquad \gamma > 0. \tag{5.11}$$

In the resulting model the strength of reversion of z_t to its attractor changes monotonically from $\rho_{1,1}$ to $\rho_{2,1}$ for increasing values of z_{t-1}. The constant c in (5.11) can be set equal to zero to render the change symmetric around the equilibrium value of zero.

A second type of asymmetry that might be of interest is to distinguish between small and large equilibrium errors. This can be achieved by taking $G(z_{t-1})$ to be the exponential function

$$G(z_{t-1};\gamma,c) = 1-\exp\{-\gamma(z_{t-1}-c)^2\}, \qquad \gamma > 0, \tag{5.12}$$

where again c should be set equal to 0 to center the function at the equilibrium, or the quadratic logistic function

$$G(z_{t-1};\gamma,c) = \frac{1}{1+\exp\{\gamma(z_{t-1}+c)(z_{t-1}-c)\}}, \qquad \gamma > 0, \tag{5.13}$$

which result in gradually changing strength of adjustment for larger (both positive and negative) deviations from equilibrium. In the resulting model, the strength of attraction towards zero changes from $\rho_{2,1}$ to $\rho_{1,1}(1-G(0;\gamma,c))$ (recall that $G(0;\gamma,c)$ can be non-zero in case of (5.13)) and back again with increasing z_{t-1}, and this change is symmetric around 0.

The smooth transition equilibrium correction model with a quadratic logistic function resembles the threshold equilibrium correction model, introduced by Balke and Fomby (1997). The threshold equilibrium correction model is obtained by allowing γ in (5.13) to become very large and imposing the restrictions $\rho_{1,1}=1$ and $\rho_{2,1}<1$ in (5.2). Intuitively, z_t then is a random walk as long as $z_{t-1}\in(-c,c)$ and the time series y_{1t} and y_{2t} behave as unrelated nonstationary series in this middle regime, which follows from (5.9) and (5.10). For example, substituting $\rho_{1,1}=1$ and $G(z_{t-1};\gamma,c)=I[|z_{t-1}|>c]$ in (5.9) renders

$$\Delta y_{1t} = \frac{\alpha}{\alpha-\beta}(\rho_{2,1}-1)I[|z_{t-1}|>c]z_{t-1}+\varepsilon_{1t},$$

[5] Of course, other possibilities to define the regimes are possible, see Siklos and Granger (1997) for an example.

from which it is seen that $\Delta y_{1t} = \varepsilon_{1t}$ as long as the lagged deviation from equilibrium z_{t-1} is smaller than the threshold c in absolute value. When z_{t-1} becomes larger than c in absolute value, z_t becomes stationary and, consequently, y_{1t} and y_{2t} are cointegrated series in these outer regimes. Notice that in this specification, the attractor of z_t in the outer regimes is equal to 0, that is, if $|z_{t-1}| > c$, z_t will return to the middle of the non-stationary region $(-c, c)$ in the absence of a shock at time t. Other, perhaps more realistic, specifications, where z_t is attracted towards the closest edge of the region $(-c, c)$ for example, also are discussed in Balke and Fomby (1997).

Threshold equilibrium correction models are applied by Dwyer, Locke and Yu (1996), Martens, Kofman and Vorst (1998) and Tsay (1998) to describe the relationship between spot and futures prices of the S&P 500 index in the presence of transaction costs. These two prices are related to each other by means of a no-arbitrage relationship, and deviations from this relationship should exist for only a brief period of time. In the presence of transaction costs or other market imperfections however, small deviations may persist as they cannot be exploited for profitable arbitrage. The relationship between spot and futures prices is discussed in more detail in Section 5.3. Other applications of threshold equilibrium correction models include the term structure of interest rates (Anderson (1997), Balke and Fomby (1997), Kunst (1992,1995), and Enders and Granger (1998)), covered interest rate parity (Balke and Wohar (1998)), and (real) exchange rates (Obstfeld and Taylor (1997), O'Connell and Wei (1997), Prakash and Taylor (1997) and O'Connell (1998)).

5.2 Specification of multivariate STAR models

Tsay (1998) describes a specification procedure for multivariate threshold models, based upon the specification procedure for univariate models developed in Tsay (1989). In similar vein, the specific-to-general approach for specifying univariate STAR models, as discussed in Section 2.1.2, can be adapted to the multivariate case.

The procedure starts with specifying a vector autoregressive [VAR] model for $\mathbf{y}_t = (y_{1t}, \ldots, y_{kt})'$, that is,

$$\mathbf{y}_t = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \cdots + \mathbf{\Phi}_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \tag{5.14}$$

where the order p should be such that the residuals $\hat{\boldsymbol{\varepsilon}}_t$ have zero autocorrelations at all lags. VAR models and various ways to select the order p are described at length in Lütkepohl (1991) and Ooms (1994), among others.

The next step in the specification procedure consists of testing linearity against the alternative of a STAR model as given in (5.1). Testing linearity is hampered by the same problem that was encountered in the univariate case, in that the multivariate STAR model contains nuisance parameters that are not identified under the null hypothesis. This can be understood by noting that the null hypothesis of linearity can be expressed in multiple ways, either as $H_0 : \mathbf{\Phi}_{1,j} = \mathbf{\Phi}_{2,j}$ for $j = 0, 1, \ldots, p$, or as $H_0' : \gamma = 0$ in (5.1). Replacing the transition function $G(s_t; \gamma, c)$ with a suitable Taylor approximation again solves the identification problem. For example, in case the alternative is a multivariate STAR model with a logistic transition function, a third-order Taylor expansion yields the reparameterized model

$$\mathbf{y}_t = \boldsymbol{B}_{0,0} + \boldsymbol{B}_0(L)\mathbf{y}_{t-1} + \boldsymbol{B}_1(L)\mathbf{y}_{t-1}s_t + \boldsymbol{B}_2(L)\mathbf{y}_{t-1}s_t^2 + \boldsymbol{B}_3(L)\mathbf{y}_{t-1}s_t^3 + \mathbf{e}_t, \quad (5.15)$$

where $\boldsymbol{B}_i(L) = \boldsymbol{B}_{i,1} + \cdots + \boldsymbol{B}_{i,p}L^{p-1}$, $i = 0, 1, 2, 3$, and $\mathbf{e}_t$ consists of the original shocks $\boldsymbol{\varepsilon}_t$ and the error arising from the Taylor approximation. Alternatively, (5.15) can be written as

$$\mathbf{y}_t = \boldsymbol{B}_{0,0} + \boldsymbol{B}_0\tilde{\mathbf{x}}_t + \boldsymbol{B}_1\tilde{\mathbf{x}}_t s_t + \boldsymbol{B}_2\tilde{\mathbf{x}}_t s_t^2 + \boldsymbol{B}_3\tilde{\mathbf{x}}_t s_t^3 + \mathbf{e}_t, \quad (5.16)$$

where $\boldsymbol{B}_i = (\boldsymbol{B}_{i,1}, \ldots, \boldsymbol{B}_{i,p})$, $i = 0, 1, 2, 3$, and $\tilde{\mathbf{x}}_t = (\mathbf{y}_{t-1}', \ldots, \mathbf{y}_{t-p}')'$, which brings out the analogy with the univariate case perhaps more clearly, compare (2.24). The parameters in $\boldsymbol{B}_i$, $i = 0, 1, 2, 3$, are functions of the parameters in the STAR model such that the original null hypothesis of linearity is equivalent to the null hypothesis that all coefficients of the auxiliary regressors, $\tilde{\mathbf{x}}_t s_t^i$, $i = 1, 2, 3$, are equal to zero, that is, $H_0'' : \boldsymbol{B}_i = \mathbf{0}$, $i = 1, 2, 3$. This hypothesis can be tested by a standard variable addition test. The resulting Lagrange Multiplier [LM] statistic has an asymptotic χ^2 distribution with $3pk^2$ degrees of freedom under the null hypothesis. The statistic, which will be denoted as LM_3, can easily be computed from an auxiliary regression of the residuals from the VAR(p) model under the null hypothesis on $\tilde{\mathbf{x}}_t$ and $\tilde{\mathbf{x}}_t s_t^i$, $i = 1, 2, 3$, whereas an F version of the test can be used as well. The small sample properties of the LM_3 test are investigated below. Multivariate analogues of the other LM-type statistics discussed in Section 2.2 can be derived in a similar way.

To select an appropriate transition variable s_t, the LM_3 statistic can be computed for several candidates $s_{1t}, \ldots, s_{mt}$, say, and the one for which the p-value of the test statistic is smallest can be selected. As in the univariate case, the suitable form of the transition function $G(s_t; \gamma, c)$ can be determined by testing a short sequence of conditional hypotheses nested within the null hypothesis of LM_3, see Section 2.2 for details.

When linearity is rejected and s_t and $G(s_t; \gamma, c)$ have been selected, the parameters in the multivariate STAR model can be estimated by nonlinear least squares [NLS]. Under certain regularity conditions, the estimates are consistent and asymptotically normal distributed.

To evaluate an estimated multivariate STAR model, the residuals can be subjected to the usual diagnostic checks. Also, the tests for no residual autocorrelation, no remaining nonlinearity and parameter constancy as developed by Eitrheim and Teräsvirta (1996) for univariate models can be generalized to a multivariate setting, see Anderson and Vahid (1998, Appendix D). The generalized impulse response functions [GIRFs] of Koop *et al.* (1996) can be used to examine the propagation of shocks by the STAR model. Additional advantages of GIRFs in the multivariate case are that, unlike traditional impulse response functions, they do not require orthogonalization of the shock $\boldsymbol{\varepsilon}_t$ and they are invariant to the ordering of the variables $y_{1t}, \ldots, y_{kt}$, see Pesaran and Shin (1998).

Specification of smooth transition equilibrium correction models

Specification of STEqCMs proceeds along the same lines as the specification of general multivariate STAR models. An additional preliminary step concerns estimation of the long-run equilibrium relationship that is presumed to exist between $y_{1t}, \ldots, y_{kt}$, if its form is not suggested by economic theory. Escribano and Mira (1997) show that the parameters in a linear equilibrium relationship can still be estimated consistently by conventional techniques, such as a regression of y_{1t} on $y_{2t}, \ldots, y_{kt}$ suggested by Engle and Granger (1987), in the presence of nonlinear adjustment. Balke and Fomby (1997) and van Dijk and Franses (1998) provide simulation evidence that supports this.

Some care needs to be taken in case the long-run equilibrium contains deterministic components. Suppose for example that in the bivariate example considered before, the equilibrium relationship between y_{1t} and y_{2t} is given by

$$y_{1t} - \beta y_{2t} - \delta = z_t, \tag{5.17}$$

for some $\delta \neq 0$ and z_t given by (5.2). In this case, the familiar 'cointegrating regression' of y_{1t} on y_{2t} and a constant yields a (super-)consistent estimate of β as usual. By contrast, the estimate of δ is inconsistent. Intuitively this can be understood by noting that, in general, z_t will not have mean equal to zero (even though the attractor of z_t is equal to zero). As the residuals $\hat{z}_t$ from the regression y_{1t} on y_{2t} and a constant have mean equal to zero by construction, the non-zero mean of the true errors z_t will show up in the estimate of δ, rendering the estimate biased and inconsistent. A solution is to retain (5.5) as specifying the equilibrium relation between y_{1t} and y_{2t} and to incorporate δ in the specification of z_t as

$$z_t - \delta = (\rho_{1,1}(1 - G(z_{t-1}; \gamma, c) + \rho_{2,1} G(z_{t-1}; \gamma, c))(z_{t-1} - \delta) + \nu_t, \tag{5.18}$$

such that the attractor of z_t is equal to δ. If in addition the location parameter c in $G(z_{t-1}; \gamma, c)$ is set equal to δ, the change in the adjustment towards equilibrium is symmetric around δ. See Berben and van Dijk (1999) for more details.

The remaining steps of the specification procedure for STEqCMs are the same as described above for the general multivariate STAR model, see also Swanson (1999) for an explicit discussion of the properties of the (equation-by-equation) LM-type tests in this case.

Common nonlinearity

The nonlinearity in y_{1t} and y_{2t} as defined by (5.7) and (5.8) is common, in the sense that it is caused by the presence of z_t in both series. The possibility of common nonlinear components is, however, not restricted to series that are linked by long-run equilibrium relationships, but can be considered more generally. Following Anderson and Vahid (1998), the time series $\mathbf{y}_t = (y_{1t}, \ldots, y_{kt})'$ is said to contain a common nonlinear component if there exists a linear combination $\boldsymbol{\alpha}'\mathbf{y}_t$ whose conditional expectation is linear in the past of $\mathbf{y}_t$. For example, in the bivariate STEqCM, the conditional expectation of $y_{1t} - \alpha y_{2t}$ is linear in the past of y_{1t} and y_{2t}, which follows immediately from (5.3). To consider the the general case, rewrite the multivariate STAR model (5.1) as

$$\mathbf{y}_t = \mathbf{\Phi}_0 + \mathbf{\Phi}_1\mathbf{y}_{t-1} + \cdots + \mathbf{\Phi}_p\mathbf{y}_{t-p} \\ + (\mathbf{\Theta}_0 + \mathbf{\Theta}_1\mathbf{y}_{t-1} + \cdots + \mathbf{\Theta}_p\mathbf{y}_{t-p})G(s_t;\gamma,c) + \boldsymbol{\varepsilon}_t, \quad (5.19)$$

where $\mathbf{\Phi}_j = \mathbf{\Phi}_{1,j}$, $j = 0, 1, \ldots, p$, and $\mathbf{\Theta}_j = \mathbf{\Phi}_{2,j} - \mathbf{\Phi}_{2,j}$, $j = 0, 1, \ldots, p$. The existence of a common nonlinear component as defined above then means that there exists a $(k \times 1)$ vector $\boldsymbol{\alpha}$ such that

$$\boldsymbol{\alpha}'(\mathbf{\Theta}_0 + \mathbf{\Theta}_1\mathbf{y}_{t-1} + \cdots + \mathbf{\Theta}_p\mathbf{y}_{t-p})G(s_t;\gamma,c) = 0, \quad (5.20)$$

for all $\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-p}$ and s_t. Anderson and Vahid (1998) develop test statistics for the existence of common STAR-type nonlinearity based upon canonical correlations between $\boldsymbol{\alpha}'\mathbf{y}_t$ and $\tilde{\mathbf{x}}_t s_t^i$, $i = 1, 2, 3$.

5.2.1 Small sample properties of multivariate linearity tests

In this section I consider the small sample properties of the multivariate LM-type linearity tests discussed above. Only rejection frequencies for the LM_3 statistic based on (5.15) are reported in full detail, as the results for other statistics are very similar.

To examine the size properties of the LM-type test, I consider four VAR(p) systems for a k-dimensional time series $\mathbf{y}_t = (y_{1t}, \ldots, y_{kt})'$ as given in (5.14). Both p and k are varied to assess the effects of the dimensionality of the system on the behaviour of the tests.

The first DGP is based on Tsay (1998), and concerns the simplest possible case, that is, a VAR(1) model for a bivariate series $\mathbf{y}_t = (y_{1t}, y_{2t})'$, where

$$\textbf{DGP (i)} \qquad \mathbf{\Phi}_0 = \begin{pmatrix} 0.2 \\ 0.2 \end{pmatrix}, \quad \mathbf{\Phi}_1 = \begin{pmatrix} 0.7 & 0.2 \\ -0.2 & 0.7 \end{pmatrix}.$$

The second DGP is taken from Anderson and Vahid (1998), and concerns a VAR(2) model for a bivariate series, with

$$\textbf{DGP (ii)} \qquad \mathbf{\Phi}_0 = \begin{pmatrix} 0.2 \\ 0.1 \end{pmatrix}, \quad \mathbf{\Phi}_1 = \begin{pmatrix} -0.1 & -0.3 \\ 0.1 & -0.3 \end{pmatrix}, \quad \mathbf{\Phi}_2 = \begin{pmatrix} -0.36 & 0.53 \\ -0.54 & 0.18 \end{pmatrix}.$$

The remaining two DGPs are taken from Swanson, Ozyildirim and Pisu (1996) and concern trivariate systems, $\mathbf{y}_t = (y_{1t}, y_{2t}, y_{3t})'$, generated from either a VAR(1) model with

$$\textbf{DGP (iii)} \qquad \mathbf{\Phi}_0 = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \end{pmatrix}, \quad \mathbf{\Phi}_1 = \begin{pmatrix} 0.8 & 0.0 & 0.0 \\ 0.2 & 0.7 & 0.4 \\ -0.5 & 0.0 & 0.9 \end{pmatrix},$$

Table 5.1: Size of LM_3 statistic, DGP (i)

			$T = 200$			$T = 400$		
σ	d	α	0.010	0.050	0.100	0.010	0.050	0.100
0.0	1		0.009	0.047	0.096	0.010	0.048	0.094
	2		0.008	0.041	0.085	0.008	0.049	0.093
	3		0.011	0.050	0.096	0.009	0.048	0.093
0.3	1		0.011	0.047	0.090	0.009	0.047	0.095
	2		0.008	0.042	0.092	0.009	0.050	0.098
	3		0.011	0.056	0.106	0.009	0.049	0.097
0.7	1		0.010	0.045	0.085	0.009	0.046	0.096
	2		0.010	0.045	0.088	0.010	0.045	0.095
	3		0.009	0.053	0.105	0.008	0.048	0.096

The Table reports rejection frequencies of the null hypothesis of linearity by the F variant of the LM-type test based on (5.15) with $s_t = y_{1,t-d}$, at nominal significance levels $\alpha = 0.010$, 0.050 and 0.100. Artificial time series are generated according to DGP (i). The Table is based on 5000 replications.

or a VAR(3) model with

$$\textbf{DGP (iv)} \qquad \mathbf{\Phi}_0 = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \end{pmatrix}, \quad \mathbf{\Phi}_1 = \begin{pmatrix} 0.5 & 0.0 & 0.0 \\ 0.1 & 0.7 & 0.4 \\ -0.9 & 0.0 & 1.5 \end{pmatrix},$$

$$\mathbf{\Phi}_2 = \begin{pmatrix} 0.5 & 0.0 & 0.0 \\ 0.1 & -0.7 & 0.6 \\ 0.5 & 0.0 & -0.5 \end{pmatrix}, \quad \mathbf{\Phi}_3 = \begin{pmatrix} -0.2 & 0.0 & 0.0 \\ 0.0 & 0.7 & -0.6 \\ -0.1 & 0.0 & -0.1 \end{pmatrix}.$$

In all cases, $\boldsymbol{\varepsilon}_t$ is taken to be i.i.d. normally distributed, with mean zero and covariance matrix $\mathbf{\Sigma}$, where $\mathbf{\Sigma} = \{\sigma_{ij}\}$, with $\sigma_{ii} = 1$, $i = 1, \ldots, k$, and $\sigma_{ij} = \sigma$, for $i, j = 1, \ldots, k$, $i \neq j$. I vary the parameter σ among 0, 0.3 and 0.7, to examine whether the correlation among the elements of $\boldsymbol{\varepsilon}_t$ has a noticeable effect on the properties of the tests. The tests are computed with $y_{1,t-d}$, $d = 1, 2, 3$, as candidate transition variable for DGPs (i), (ii) and (iii), and $y_{2,t-d}$, $d = 1, 2, 3$, for DGP (iv). All experiments make use of 5000 replications, with the sample size set equal to $T = 200$ or 400.

Tables 5.1 through 5.4 show rejection frequencies of the F variant of the LM_3 statistic based on (5.15) at the 1, 5 and 10% nominal significance level for the four DGPs. The main conclusion that emerges from these tables is that the test is properly sized. The only exception appears to be DGP (iv) and the smaller sample size $T = 200$, in which case the rejection frequencies exceed the nominal significance level by quite a wide margin. The same pattern is observed for multivariate analogues of the LM_1, LM_2 and LM_4 statistics discussed in Section 2.2, but these are not reported here.

The power properties of the tests are investigated by means of the following three DGPs. First, I consider a STAR model with $p = 1$ for a bivariate series, given by

Table 5.2: Size of LM_3 statistic, DGP (ii)

			$T=200$			$T=400$		
σ	d	α	0.010	0.050	0.100	0.010	0.050	0.100
0.0	1		0.008	0.045	0.094	0.010	0.049	0.092
	2		0.008	0.042	0.092	0.008	0.048	0.098
	3		0.009	0.053	0.104	0.009	0.045	0.095
0.3	1		0.009	0.045	0.094	0.009	0.049	0.097
	2		0.008	0.049	0.099	0.009	0.051	0.099
	3		0.013	0.050	0.093	0.006	0.043	0.095
0.7	1		0.010	0.048	0.095	0.010	0.045	0.096
	2		0.009	0.049	0.102	0.009	0.049	0.094
	3		0.007	0.043	0.092	0.006	0.045	0.093

The Table reports rejection frequencies of the null hypothesis of linearity by the F variant of the LM-type test based on (5.15) with $s_t = y_{1,t-d}$, at nominal significance levels $\alpha = 0.010$, 0.050 and 0.100. Artificial time series are generated according to DGP (ii). The Table is based on 5000 replications.

Table 5.3: Size of LM_3 statistic, DGP (iii)

			$T=200$			$T=400$		
σ	d	α	0.010	0.050	0.100	0.010	0.050	0.100
0.0	1		0.010	0.057	0.106	0.010	0.052	0.105
	2		0.012	0.054	0.111	0.010	0.049	0.103
	3		0.012	0.060	0.116	0.013	0.054	0.107
0.3	1		0.010	0.051	0.103	0.010	0.046	0.092
	2		0.010	0.044	0.099	0.011	0.044	0.097
	3		0.012	0.057	0.102	0.011	0.054	0.100
0.7	1		0.010	0.046	0.094	0.006	0.043	0.091
	2		0.009	0.043	0.098	0.009	0.043	0.091
	3		0.010	0.052	0.103	0.009	0.047	0.101

The Table reports rejection frequencies of the null hypothesis of linearity by the F variant of the LM-type test based on (5.15) with $s_t = y_{1,t-d}$, at nominal significance levels $\alpha = 0.010$, 0.050 and 0.100. Artificial time series are generated according to DGP (iii). The Table is based on 5000 replications.

Table 5.4: Size of LM_3 statistic, DGP (iv)

			$T=200$			$T=400$		
σ	d	α	0.010	0.050	0.100	0.010	0.050	0.100
0.0	1		0.013	0.058	0.104	0.010	0.048	0.103
	2		0.010	0.057	0.111	0.009	0.051	0.108
	3		0.011	0.053	0.113	0.009	0.056	0.104
0.3	1		0.012	0.058	0.118	0.010	0.050	0.106
	2		0.012	0.055	0.104	0.011	0.057	0.108
	3		0.012	0.054	0.104	0.011	0.055	0.107
0.7	1		0.014	0.063	0.125	0.009	0.051	0.110
	2		0.014	0.064	0.123	0.012	0.059	0.112
	3		0.014	0.062	0.116	0.010	0.055	0.110

The Table reports rejection frequencies of the null hypothesis of linearity by the F variant of the LM-type test based on (5.15) with $s_t = y_{2,t-d}$, at nominal significance levels $\alpha = 0.010$, 0.050 and 0.100. Artificial time series are generated according to DGP (iv). The Table is based on 5000 replications.

(5.1), where

$$\textbf{DGP (v)} \qquad \mathbf{\Phi}_{1,0} = \mathbf{\Phi}_{2,0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \mathbf{\Phi}_{1,1} = \begin{pmatrix} 0.6 & 0.3 \\ 0.3 & 0.6 \end{pmatrix}, \quad \mathbf{\Phi}_{2,1} = \begin{pmatrix} -0.4 & 0 \\ 0 & -0.4 \end{pmatrix}.$$

The transition function is taken to be the logistic function $G(s_t;\gamma,c) = 1/(1+\exp\{-\gamma(s_t-c)\})$, with $\gamma=5$, $c=0$ and $s_t=y_{1,t-1}$. The second DGP is a bivariate STAR model with $p=2$, with $\mathbf{\Phi}_{1,0} = (0.2, 0.1)'$, $\mathbf{\Phi}_{2,0} = (0,0)'$,

$$\textbf{DGP (vi)} \qquad \mathbf{\Phi}_{1,1} = \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix}, \quad \mathbf{\Phi}_{1,2} = \begin{pmatrix} -0.4 & -0.2 \\ -0.2 & -0.4 \end{pmatrix},$$

$$\mathbf{\Phi}_{2,1} = \begin{pmatrix} 0.4 & -0.2 \\ -0.2 & 0.4 \end{pmatrix}, \quad \mathbf{\Phi}_{2,2} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

The transition function $G(s_t;\gamma,c)$ again is the logistic function, with $\gamma=3$, $c=0$, and $s_t=y_{2,t-1}$. In both DGP (v) and (vi), the covariance matrix $\mathbf{\Sigma}$ of the shocks $\boldsymbol{\varepsilon}_t$ is specified as before. The candidate transition variables that are considered are $s_t = y_{1,t-d}$ and $s_t = y_{2,t-d}$, $d=1,2,3$, for DGPs (v) and (vi), respectively. The final DGP (vii) is a bivariate STEqCM (5.9)-(5.10) with $\alpha=1$, $\beta=2$, $\rho_{1,1}=1$, $G(z_{t-1};\gamma,c)$ taken to be the exponential function (5.12) with $\gamma=5$ and $c=0$, $\sigma=1$ and $\theta=0$. The parameter $\rho_{2,1}$ is varied among $\rho_{2,1}\in\{0.8,0.6,0.4,0.2\}$. The linearity tests are applied in a linear EqCM representation of y_{1t} and y_{2t}, with $s_t = z_{t-d}$, $d=1,2,3$, as candidate transition variables.

Results for these experiments are shown in Tables 5.5 through 5.7. For DGP (v), the power of the test is excellent in case the transition variable is correctly specified, that is, in case $s_t=y_{1,t-1}$. Power is (much) smaller when $y_{1,t-2}$ or $y_{1,t-3}$ are considered as transition variables. Note however that as the sample size increases, the

Table 5.5: Power of LM_3 statistic, DGP (v)

			$T = 200$			$T = 400$		
σ	d	α	0.010	0.050	0.100	0.010	0.050	0.100
0.0	1		0.992	0.998	0.998	1.000	1.000	1.000
	2		0.261	0.473	0.600	0.738	0.889	0.934
	3		0.100	0.258	0.367	0.360	0.584	0.698
0.3	1		0.980	0.994	0.997	1.000	1.000	1.000
	2		0.251	0.482	0.612	0.743	0.886	0.937
	3		0.098	0.246	0.360	0.351	0.577	0.687
0.7	1		0.989	0.997	0.999	1.000	1.000	1.000
	2		0.359	0.594	0.714	0.849	0.942	0.970
	3		0.130	0.300	0.419	0.438	0.663	0.772

The Table reports rejection frequencies of the null hypothesis of linearity by the F variant of the LM-type test based on (5.15) with $s_t = y_{1,t-d}$, at nominal significance levels $\alpha = 0.010$, 0.050 and 0.100. Artificial time series are generated according to DGP (v). The Table is based on 5000 replications.

Table 5.6: Power of LM_3 statistic, DGP (vi)

			$T = 200$			$T = 400$		
σ	d	α	0.010	0.050	0.100	0.010	0.050	0.100
0.0	1		0.494	0.727	0.814	0.941	0.988	0.994
	2		0.027	0.105	0.187	0.081	0.219	0.327
	3		0.025	0.103	0.175	0.056	0.172	0.269
0.3	1		0.480	0.714	0.810	0.939	0.986	0.994
	2		0.025	0.092	0.168	0.069	0.195	0.300
	3		0.030	0.112	0.190	0.075	0.204	0.311
0.7	1		0.682	0.857	0.915	0.990	0.998	0.999
	2		0.024	0.101	0.178	0.073	0.211	0.316
	3		0.053	0.149	0.239	0.132	0.303	0.424

The Table reports rejection frequencies of the null hypothesis of linearity by the F-variant of the LM-type test based on (5.15) with $s_t = y_{2,t-d}$, at nominal significance levels $\alpha = 0.010$, 0.050 and 0.100. Artificial time series are generated according to DGP (vi). The Table is based on 5000 replications.

Table 5.7: Power of LM_3 statistic, DGP (vii)

			$T = 200$			$T = 400$		
σ	d	α	0.010	0.050	0.100	0.010	0.050	0.100
0.8	1		0.073	0.209	0.317	0.175	0.384	0.508
	2		0.047	0.158	0.260	0.114	0.296	0.433
	3		0.018	0.084	0.159	0.043	0.137	0.237
0.6	1		0.272	0.506	0.633	0.634	0.834	0.898
	2		0.176	0.405	0.541	0.518	0.755	0.851
	3		0.017	0.077	0.144	0.035	0.125	0.210
0.4	1		0.534	0.766	0.850	0.914	0.976	0.988
	2		0.374	0.640	0.759	0.820	0.939	0.969
	3		0.010	0.053	0.109	0.016	0.067	0.127
0.2	1		0.746	0.900	0.948	0.984	0.996	0.999
	2		0.528	0.766	0.860	0.928	0.983	0.993
	3		0.009	0.045	0.096	0.012	0.047	0.096

The Table reports rejection frequencies of the null hypothesis of linearity by the F-variant of the LM_3 statistic with $s_t = z_{t-d}$, at nominal significance levels $\alpha = 0.010$, 0.050 and 0.100. Artificial time series are generated according to DGP (vii), details of which are given in the text. The Table is based on 5000 replications.

rejection frequencies for incorrect transition variables increase considerably. Comparison of the p-values of the test for different choices of s_t shows that for most replications the null hypothesis still is rejected most convincingly in case the true transition variable is used. This suggests that selecting the transition variable based upon the minimum p-value rule should work reasonably well. For DGP (vi), power is somewhat lower for the smaller sample size $T = 200$. Note that the difference between the rejection frequencies for correct and incorrect transition variables is larger for this DGP. Finally, the results for DGP (vii) also show that power is highest when the correct transition variable z_{t-1} is used. Furthermore, the rejection frequencies increase as the strength of equilibrium correction for large (absolute) deviations from equilibrium, measured by $|\rho_{2,1} - 1|$, becomes larger.

5.3 SETS, arbitrage activity and stock price dynamics

In this section I employ the STEqCM framework to describe the behaviour of intraday spot and futures prices of the FTSE100 index. In particular, the model is used to study the effects of the introduction of a new electronic trading system on the London Stock Exchange in October 1997.

Exchanges throughout the world have introduced (for example, London and Frankfurt) or are about to introduce (for example, Sydney) electronic trading systems. There remains some uncertainty, however, concerning the benefits (or otherwise) of such systems versus traditional trading systems. The application in this section provides empirical evidence on the cost and efficiency improvements brought about by electronic trading systems. More specifically, the transaction costs and stock price dynamics associated with arbitrage activity in spot and futures markets in the UK before and after the introduction of an electronic trading system are measured.

On October 20, 1997, the London Stock Exchange introduced a new electronic trading system [SETS]. The system enables traders to place buy or sell orders for any FTSE100 shares in an electronic order book. These orders are then automatically matched with other orders placed. Before the introduction of this system orders were advertised on computer terminals but actual trades were carried out over the telephone. Under this old system market-makers would absorb the impact of large trades by putting their own capital at risk. Such generosity was compensated for by large bid-ask spreads. Gemmill (1998) reports a 39 basis point spread for large companies and a 79 basis point spread for small companies before the introduction of SETS. By contrast, the respective spreads after the introduction of SETS were 32 basis points and 53 basis points.

The reduction in average bid-ask spreads should have an effect on all arbitrage activity. The activity examined here concerns those trades that are conducted in order to lock into risk-less profits that arise because of perturbations in the contemporaneous relationship between FTSE100 spot and futures prices. Arbitrage activity involves simultaneous positions in both the spot and futures index. The length of time these positions are held depends upon whether or not it is profitable to unwind the position before the maturity of the contract. Brennan and Schwartz (1988,1990) thus consider such a position as both an arbitrage position and an option to unwind the position when positive profits can be obtained. As Neal (1992) and Sofianos (1993) find that most arbitrage positions are not held until maturity it follows that the option to unwind must have some positive value. This additional value presumably lowers the absolute value of the bounds outside which it is profitable to trade. Moreover, as the cost of exercising the option is the difference between the buy and sell prices of the security, then any decrease in bid-ask spread lowers the cost of unwinding the position and, thus, the arbitrage bounds.

The introduction of SETS offers an opportunity to study how arbitrage activity and stock price dynamics are affected by a change in transaction costs. I consider whether the introduction of SETS has changed the trading bounds outside which arbitrage activity takes place and whether markets have become more efficient.[6] The remainder of this section is organised as follows. First an economic model of arbitrager behaviour based on the cost-of-carry model and the econometric model

[6]SETS only pertains to spot positions in FTSE100 shares. It does not effect the mechanism by which futures contracts in the FTSE100 index are traded. The exchange within which these contracts are traded [LIFFE] currently operates open outcry trading. This difference in trading mechanisms should not detract from the fact that costs of arbitrage trading are likely to be reduced under SETS. This is because mispricing will lead to simultaneous trading in both spot and futures markets. As such, a decrease in the cost of trading the FTSE100 shares in the spot market will reduce overall transaction costs.

that is used are outlined. Section 5.3.2 provides a description of the data. Finally, Section 5.3.3 contains the empirical results.

5.3.1 The cost-of-carry model

The (contemporaneous) relationship between spot and forward prices can be described by the cost-of-carry model. This model is also capable of describing the relationship between spot and futures prices providing that the term structure of interest rates is flat and constant, see Brenner and Kroner (1995). Under the no-arbitrage condition with no transaction costs, the model has the following specification

$$F_{t,T} = S_t e^{(r_{t,T} - \delta_{t,T})(T-t)}, \tag{5.21}$$

where F_t is the futures price at time t for a contract with expiration date T, S_t is the spot price at time t, $r_{t,T}$ is the risk-free interest rate, $\delta_{t,T}$ is the expected dividend yield on the underlying asset, with both $r_{t,T}$ and $\delta_{t,T}$ measured over the time to maturity of the futures contract. If the contract is held to maturity then in the presence of proportional and symmetric[7] transaction costs c, arbitrage activity will take place when one of the following conditions holds,

$$\frac{F_t}{S_t} e^{(r_{t,T} - \delta_{t,T})(T-t)} < 1 - c, \tag{5.22}$$

$$\frac{F_t}{S_t} e^{(r_{t,T} - \delta_{t,T})(T-t)} > 1 + c. \tag{5.23}$$

As it takes time for arbitragers to take appropriate spot and futures positions, this arbitrage opportunity is necessarily lagged by d time periods. Therefore, providing c is small, the above inequalities can be expressed in the following (logarithmic) form

$$|z_{t-d}| > c, \tag{5.24}$$

where $z_t = \ln F_t - \ln S_t - (r_{t,T} - \delta_{t,T})(T-t)$ and is referred to as the basis or the pricing error, and d is the delay inherent in the arbitrage process. As arbitragers are expected to unwind the positions before the maturity of the contract, c represents approximately one half the total round-trip transaction costs incurred by arbitragers (Dwyer *et al.* (1996)).

Previous empirical studies have concluded that spot and futures stock indices are each non-stationary while the respective basis is stationary (Dwyer *et al.* (1996) and Martens *et al.* (1998)). This implies that spot and futures prices are cointegrated with a cointegrating vector equal to $(1, -1)$, and, consequently, they have an equilibrium correction representation. Such an equilibrium correction representation directly links changes in futures and spot prices to deviations from the arbitrage relation (5.21), that is, to pricing errors. Equation (5.24), however, states that arbitrage activity only occurs if it is profitable. Equivalently, arbitrage positions in spot and futures stock markets are taken only when the pricing error is outside a

[7]Symmetric here means that transaction costs are the same for positive and negative deviations from the cost-of-carry relation.

particular bound. Thus, spot and futures prices only adjust when the past pricing error is sufficiently large[8].

Earlier studies of the relationship between futures prices and spot prices in the presence of transaction costs have used the threshold equilibrium correction model [TEqCM] of Balke and Fomby (1997), see for example, Dwyer *et al.* (1996) and Martens *et al.* (1998). Essentially this model assumes that all arbitragers face identical transaction costs, that is, transaction costs are homogeneous[9]. Here I use a smooth transition equilibrium correction model [STEqCM], as discussed in Section 5.1 to model the behaviour of spot and futures prices. Ignoring lag dependence in differenced series, the model can be expressed as

$$\begin{pmatrix} \Delta f_t \\ \Delta s_t \end{pmatrix} = \begin{pmatrix} \alpha_f \\ \alpha_s \end{pmatrix} z_{t-d} G(z_{t-d}) + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}, \tag{5.25}$$

where Δf_t is the differenced logarithmic futures price series, Δs_t is the differenced logarithmic spot price series, ε_{1t} and ε_{2t} are (possibly heteroskedastic and cross-correlated) white noise series, $d \in \{1, 2, \ldots\}$, and $G(z_{t-d})$ is taken to be the exponential function

$$G(z_{t-d}; \gamma) = 1 - \exp\{-\gamma z_{t-d}^2\}, \tag{5.26}$$

where $\gamma > 0$. The strength of the relationship between Δf_t (Δs_t) and z_{t-d} will range from zero to α_f (α_s) as $G(z_{t-d})$ changes in a smooth fashion from 0 to 1. Hence, this model is capable of allowing for regime dependent arbitrage as given by equation (5.24). As argued by Anderson (1997), by allowing for smooth transition in the strength of adjustment of spot and futures prices, the STEqCM represents a more realistic representation of the heterogeneity of investors that each face different transaction costs.

The usual interpretation assigned to the parameter γ is that it measures the speed of transition from no adjustment ($G(z_{t-d}) = 0$) to full adjustment ($G(z_{t-d}) = 1$). In the present context, γ can also be regarded as measuring the degree of heterogeneity in transaction costs. Low γ values imply a wide range of transaction costs faced by investors. By contrast, high γ values imply a more uniform transaction cost structure.

The introduction of SETS should lower the transaction costs faced by all investors. Moreover, small (private) investors are expected to face similar transaction costs to those faced by large (institutional) investors. Such transaction cost homogeneity is conveniently measured by a large γ. It follows that γ should be larger after the introduction of SETS. Moreover, if γ is larger in the post-SETS period then transaction costs must be lower in this period. This is because the transition function equals zero when there is no pricing error ($z_{t-d} = 0$). As such, a large

[8] As argued by Martens *et al.* (1998), adjustment may also be observed for pricing errors inside the band $(-c, c)$ due to infrequent trading of the stocks underlying the index, see also Miller, Muthuswamy and Whaley (1994).

[9] An alternative motivation for the TEqCM model is that the most favorably positioned arbitrager, that is, the arbitrager with the lowest level of transaction costs, is able to fully exploit any arbitrage opportunities presented to him. This, however, might not be a realistic assumption in practice.

γ value means that the transition function is necessarily above the small γ value transition function. The hypothesis that is of particular interest is that γ takes the same value in the pre-SETS and post-SETS periods.[10]

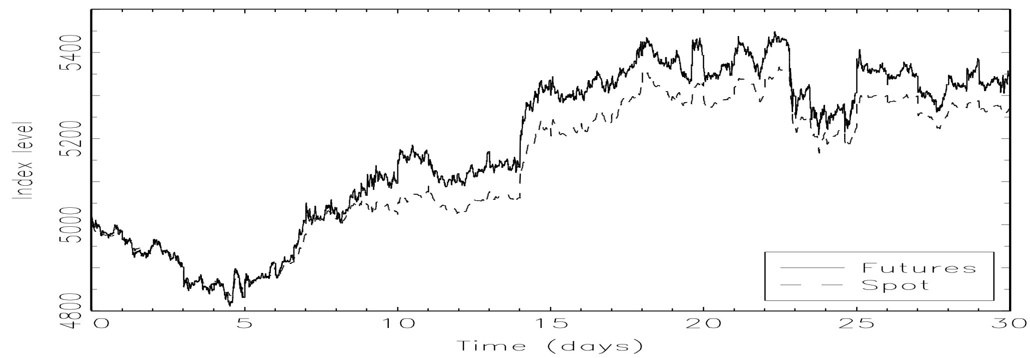
Note that the same transition function is used in the equations for futures and spot returns. This means that the same parameter γ enters both equations. Mathematically it is of course possible to have different parameters, say γ_f and γ_s, implying different transition functions. From an economic perspective, however, this is less plausible. The arbitrage mechanism is triggered by large values of the basis and requires taking a position in *both* the spot and futures market. Therefore, the parameter γ represents a measure of 'average' transaction cost heterogeneity over the two markets. It might be expected that SETS mainly reduces transaction cost heterogeneity in the spot market, thus lowering the 'average' heterogeneity. Given the present testing framework using arbitrage relations across markets, however, it is not possible to disentangle this average decrease into separate components for the spot and futures markets. This does not mean, however, that it is not possible to say anything on the relative contribution of SETS to spot and futures market efficiency improvements. In particular, the empirical results show that the efficiency of the spot market has increased relatively more than that of the futures market due to the introduction of SETS, as the strength of equilibrium correction for the spot market has increased more than that for the futures market.

5.3.2 Data

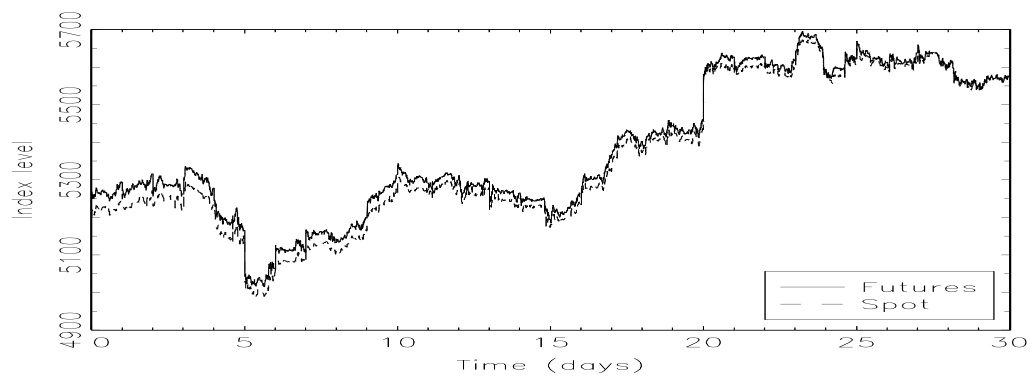
The futures price of the nearest FTSE100 contract is obtained for every transaction carried out. These data were obtained from LIFFE. The contract is changed when the volume of trading in the next nearest contract is greater than the volume of trading in the nearest contract.[11] To synchronise the futures and spot prices, the futures price series is converted to a price series with a frequency of one minute. As it is not known whether the price is a bid or ask price, the average of the last two prices is taken as the futures price. The (spot) level of the FTSE100 index was obtained from FTSE International. The trading hours of the futures market and the spot market are 8.30am to 5.30pm and 8.00am to 4.30pm, respectively. Thus overlapping futures and spot data are available for the period 8.30am to 4.30pm. However, since the introduction of SETS it has been noted that spreads are unusually

[10]The discussion in this paragraph that an increase in γ due to the introduction of SETS can be caused by either a reduction in the (average) level of transaction costs or a reduction in transaction cost heterogeneity (or both). It is not possible to disentangle these two possible causes with the exponential function (5.26). An alternative would be to use the quadratic logistic function (5.13). In that case, γ might be thought of as representing transaction cost heterogeneity and c as representing the (average) level of transaction costs. Comparing estimates of these parameters in the pre- and post-SETS period might suggest which is affected most by the introduction of the new trading system.

[11]The volume cross-over method of changing futures contracts results in one change in the pre-SETS period and no changes in the post-SETS period. The change involves a switch from the September 1997 contract to the December 1997 contract on September 19, 1997. On this date 1,422 September 1997 contracts are traded and 6,132 December 1997 contracts are traded. The post-SETS period futures prices make exclusive use of the March 1998 futures contract.



(a) Pre-SETS



(b) Post-SETS

Figure 5.1: FTSE100 futures and spot index levels

high during the first hour of trading. This is because few institutional orders are entered during this period. For this reason only prices observed between 9.00am and 4.30pm are used in the analysis. This results in 451 observations per day. The pre-SETS sample period covers the period September 8, 1997, to October 17, 1997. To allow traders to adapt to the new system, the post-SETS sample period will start on January 5, 1998, and end on February 13, 1998. These sample periods correspond to six weeks of data both before and after the introduction of SETS.

Time series plots of logarithmic futures and spot prices are presented in Figure 5.1. Sharp changes in these prices occur when the trading day changes. To avoid problems associated with these price discontinuities I remove overnight returns, following Dwyer *et al.* (1996) and Martens *et al.* (1998). This gives a total of 13,500 ($450 \times 5 \times 6$) one minute frequency returns in each of the sample periods. The analysis is also conducted using two and five minute frequency data over the same sample periods.

The pricing error is constructed using the daily demeaned futures and spot prices. This methodology follows Dwyer *et al.* (1996). Subtracting the daily mean from the futures prices ensures that any constant in the logarithmic price due to expected dividends or interest rates is removed. The pricing error is set equal to the difference

Table 5.8: Summary statistics

	Statistic				
Period	$\bar{s}_r$	$\hat{\mu}(f)$	$\hat{\mu}(s)$	$\hat{\sigma}(f)$	$\hat{\sigma}(s)$
Pre-SETS	0.62	0.23	0.19	1.12	0.94
Post-SETS	0.43	0.17	0.20	1.13	0.96

The mean daily spread in the FTSE100 stocks is denoted $\bar{s}_r$, the mean daily futures (f) and spot (s) returns and the standard deviation of daily returns are denoted $\hat{\mu}(.)$ and $\hat{\sigma}(.)$, respectively. All statistics are measured in percentage terms.

between the demeaned futures price and the demeaned spot price. Henceforth, the demeaned logarithmic futures and spot prices will be denoted by f_t and s_t, respectively, while the pricing error will be denoted by z_t.

5.3.3 Empirical results

One of the purposes of SETS is to reduce trading costs in the spot market. One way of measuring these costs is by calculating percentage bid-ask spreads. The mean spreads in the pre-SETS and post-SETS periods are given in Table 5.8. These spreads are calculated by taking the average of end-of-day spreads of all stocks in the FTSE100. The results indicate that there has been a large reduction in average bid-ask spreads since the introduction of SETS. These results confirm the results of more extensive studies, see for example, Gemmill (1998).

It could be argued that the subsequent analysis is sensitive to the particular sample periods used. For instance, one period may be more volatile than the other period. To examine this issue I calculate the standard deviation of daily futures and spot returns in the pre-SETS and post-SETS periods. The results are given in Table 5.8. Both spot and futures returns have approximately the same volatility in both periods. One can test the null hypothesis that the population standard deviations are the same in each period by comparing the ratio of sample standard deviations in the pre-SETS and post-SETS periods, with some upper percentile point on a $F(n, n)$-distribution, where n denotes the number of observations. The futures return ratio is 1.0069 and the spot return ratio is 1.0128. Comparing these values with various percentile points on an $F(30, 30)$-distribution leads to p-values of 0.4926 and 0.4862, respectively. Therefore, both futures and spot markets are equally volatile over the two periods.

Testing for non-stationarity

Augmented Dickey Fuller (ADF) tests are performed on various one minute and five minute frequency series. In each case a constant is included and the lag lengths are selected on the basis of the SIC. The one minute frequency results show that futures and spot prices are non-stationary. It should be remarked that these prices are not the same prices as those plotted in Figure 1. The non-stationarity tests are applied to intraday prices, which are are constructed as follows. First, logarithmic returns are calculated. Second, overnight returns are removed. Third, intraday prices are calculated by numerically integrating the intraday returns. Possible cointegration between these prices is investigated by testing for non-stationarity in the pricing error using the ADF test. The results (not shown here) indicate that the null hypothesis of non-stationarity can be rejected with a high level of confidence. Therefore, the cointegrating vector $(1, -1)$ provides a combination of non-stationary futures and spot prices that is stationary. As such, these prices have an equilibrium correction representation.

Testing for non-linearity

An EqCM with one lagged difference turns out to be an adequate (linear) representation of the spot and futures returns, that is,

$$\begin{pmatrix} \Delta f_t \\ \Delta s_t \end{pmatrix} = \boldsymbol{\Phi}_0 + \boldsymbol{\Phi}_1 \begin{pmatrix} \Delta f_{t-1} \\ \Delta s_{t-1} \end{pmatrix} + \begin{pmatrix} \alpha_f \\ \alpha_s \end{pmatrix} z_{t-d} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}, \tag{5.27}$$

where $\boldsymbol{\Phi}_0$ is a (2×1) vector and $\boldsymbol{\Phi}_1$ a (2×2) matrix. The number of lagged first differences is determined using SIC.

Linearity is tested against the alternative that returns follow a smooth transition equilibrium correction process

$$\begin{aligned} \begin{pmatrix} \Delta f_t \\ \Delta s_t \end{pmatrix} = & \left[\boldsymbol{\Phi}_{1,0} + \boldsymbol{\Phi}_{1,1} \begin{pmatrix} \Delta f_{t-1} \\ \Delta s_{t-1} \end{pmatrix}\right] [1 - G(z_{t-d})] \\ & + \left[\boldsymbol{\Phi}_{2,0} + \boldsymbol{\Phi}_{2,1} \begin{pmatrix} \Delta f_{t-1} \\ \Delta s_{t-1} \end{pmatrix} + \begin{pmatrix} \alpha_f \\ \alpha_s \end{pmatrix} z_{t-d}\right] G(z_{t-d}) + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}, \end{aligned} \tag{5.28}$$

using the LM_3 test discussed in Section 5.2. To avoid spurious indications of non-linearity due to the heteroskedasticity that is present in the high-frequency series, I use heteroskedasticity-consistent [HCC] variants of the test. Besides the multivariate test, equation-by-equation tests are computed as well. The results for the series sampled with one minute intervals are presented in Table 5.9 for $d = \{1, 2, 3, 4, 5\}$. The results for the multivariate test indicate that the null hypothesis can be rejected at conventional significance levels for all choices of d in the pre-SETS period and for $d = 1, 2, 3$ in the post-SETS period. The single equation tests are more informative concerning the proper choice for the delay parameter d, especially the test for nonlinearity in the futures equation. Here the null hypothesis is rejected at the 10% significance level only when $d = 1$ or 2 in the pre-SETS period, and $d = 2$ in the post-SETS period.

Estimating the smooth transition equilibrium correction models

Even though the results from the linearity tests give some indication concerning the appropriate value of the delay parameter, I decide to estimate STEqCms with

Table 5.9: Linearity tests

Period	Equation	d = 1	2	3	4	5
Pre-SETS	Both	37.59	39.71	32.71	29.56	30.82
		(0.00)	(0.00)	(0.02)	(0.04)	(0.03)
	Δf_t	16.26	20.12	13.33	12.79	13.50
		(0.06)	(0.02)	(0.15)	(0.17)	(0.14)
	Δs_t	19.16	24.36	21.60	20.57	23.55
		(0.02)	(0.00)	(0.01)	(0.01)	(0.01)
Post-SETS	Both	28.90	38.53	36.49	20.48	25.06
		(0.05)	(0.00)	(0.01)	(0.31)	(0.12)
	Δf_t	14.23	19.95	8.60	9.26	10.15
		(0.11)	(0.02)	(0.47)	(0.41)	(0.34)
	Δs_t	13.72	18.99	21.53	15.19	16.63
		(0.13)	(0.03)	(0.01)	(0.09)	(0.05)

HCC variants of multivariate and univariate LM_3 test statistics for linearity of the EqCM for spot and futures returns, with lagged pricing errors z_{t-d} as transition variable. p-values are given in parentheses.

exponential transition function with $d = \{1, 2, 3, 4, 5\}$ in both the pre-SETS and post-SETS periods, and to compare the different models using SIC. The exponential transition function is specified as $G(z_{t-d}; \gamma) = 1 - \exp\{-\gamma^* z_{t-d}^2 / \sigma_{z_{t-d}}^2\}$, where $\sigma_{z_{t-d}}^2$ is the variance of the pricing error. Dividing the exponent in the transition function by $\sigma_{z_{t-d}}^2$ follows the usual practice and facilitates estimation. Note however that here the interest is in the value of the transition function for 'absolute' values of z_{t-d} and not so much in the value of the transition function for z_{t-d} measured in multiples of its variance. Therefore, I report estimates of $\gamma = \gamma^*/\sigma_{z_{t-d}}^2$ instead of γ^*, also because the variance of the pricing error in the post-SETS period is considerably smaller than the variance in the pre-SETS period.[12] NLS estimates of the adjustment parameters $\hat{\alpha}_f$ and $\hat{\alpha}_s$ and of γ are presented in Panel A of Table 5.10, with heteroscedasticity-consistent standard errors in parentheses. Standard errors for $\hat{\gamma}$ are not reported, for reasons discussed in Section 2.3.

When using one minute frequency returns information criteria are minimised when the delay equals one minute in the pre-SETS period and two minutes in the post-SETS period. The adjustment coefficients have the expected signs, $\alpha_f < 0$ and $\alpha_s > 0$. Moreover, adjustment in the spot market is considerably larger, in absolute terms, than adjustment in the futures market during the post-SETS period.

The issue of stale prices

The analysis so far has made use of minute frequency index data. One problem with using such data is that it may be composed of stale prices. The inclusion of such

[12] In the pre-SETS period $\sigma_{z_{t-d}}^2 = 0.037$, whereas in the post-SETS period $\sigma_{z_{t-d}}^2 = 0.009$.

Table 5.10: Estimated STECM parameters

Period	d	$\hat{\alpha}_f$	$\hat{\alpha}_s$	$\hat{\gamma}$	AIC	SIC
			One minute frequency			
Pre-SETS	1	−0.0097	0.0135	8.7749	−285.4438	−285.4317
		(0.0048)	(0.0022)	–		
Post-SETS	1	−0.0038	0.1350	36.7303	−273.3028	−273.2908
		(0.0092)	(0.0198)	–		
Pre-SETS	2	−0.0145	0.0225	4.1029	−278.0397	−278.0277
		(0.0082)	(0.0063)	–		
Post-SETS	2	−0.0063	0.1199	77.8665	−274.0538	−274.0418
		(0.0051)	(0.0124)	–		
Pre-SETS	3	−0.0086	0.0108	18.7862	−284.3606	−248.3486
		(0.0035)	(0.0013)	–		
Post-SETS	3	−0.0036	0.0957	85.0978	−272.1318	−272.1198
		(0.0048)	(0.0118)	–		
Pre-SETS	4	−0.0065	0.0106	13.0187	−283.1522	−283.1403
		(0.0040)	(0.0016)	–		
Post-SETS	4	−0.0046	0.0793	59.7883	−270.3740	−270.3620
		(0.0052)	(0.0130)	–		
Pre-SETS	5	−0.0056	0.0083	22.3068	−283.0865	−283.0745
		(0.0032)	(0.0010)	–		
Post-SETS	5	−0.0021	0.0530	148.1662	−270.8448	−270.8328
		(0.0042)	(0.0064)	–		
			Two minute frequency			
Pre-SETS	1	−0.0285	0.0350	3.7280	−128.1952	−128.1843
		(0.0180)	(0.0111)	–		
Post-SETS	1	0.0093	0.2289	36.1349	−125.9378	−125.9269
		(0.0205)	(0.0369)	–		
Pre-SETS	2	−0.0257	0.0343	5.8313	−129.0821	−129.0712
		(0.0149)	(0.0099)	–		
Post-SETS	2	−0.0075	0.1976	107.4089	−125.2710	−125.2601
		(0.0102)	(0.0226)	–		
			Five minute frequency			
Pre-SETS	1	−0.0104	0.1033	2.8080	−444.0591	−443.9650
		(0.0638)	(0.0522)	–		
Post-SETS	1	0.0125	0.4433	48.0969	−443.4842	−443.3901
		(0.0581)	(0.0927)	–		

The numbers in parentheses are heteroscedasticity-consistent standard errors. The last two columns contain values of the Akaike Information Criterion (AIC) and the Schwarz Information Criterion (SIC).

prices occurs when prices are measured at regular intervals but are actually posted at irregular intervals. Such non-synchronous trading effects have been extensively studied, see Fisher (1966) and Scholes and Williams (1977) for early examples. Lo and MacKinlay (1990) show that in a portfolio consisting of homogeneously thinly traded securities the first-order autocorrelation in portfolio returns asymptotically equals the probability of observing a stale individual security price. Moreover, Lo and MacKinlay (1990) also show that this autocorrelation decreases rapidly, in a non-linear fashion, when the frequency of the data is decreased.

Non-synchronous trading effects are controlled for in two different ways. First, the STEqCM given in (5.28) includes a linear autoregressive component. Thus autocorrelation in returns is explicitly modeled. Second, the frequency of the data is decreased to show that the results are robust to changes in non-synchronous trading effects.

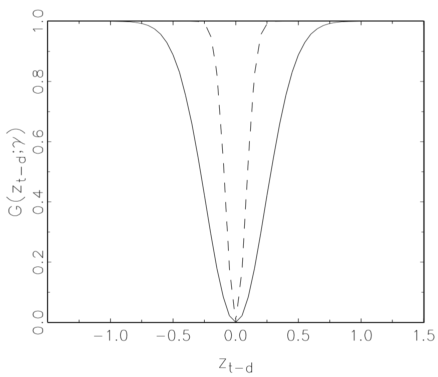
The estimated STEqCM parameters obtained using lower frequency data (two and five minute frequency data) are given in Panels B and C of Table 5.10. When two minute frequency data are used I set $d = \{1, 2\}$. These correspond to actual delays of two and four minutes, respectively. Using longer time delays is unrealistic given the likely speed of the arbitrage process. The results are very similar to those obtained using minute frequency data. The optimal time delays are $d = 2$ (four minutes) in the pre-SETS period and $d = 1$ (two minutes) in the post-SETS period. In most cases the adjustment coefficients take there expected signs and are significantly different from zero. Similar results are obtained when five minute frequency data are used with $d = 1$.

Comparing transaction cost profiles

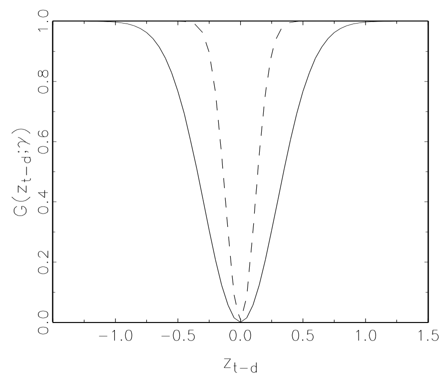
The results given in Table 5.10 indicate that when the exponential transition function is used the degree of transaction cost heterogeneity is greater in the pre-SETS period than in the post-SETS period. That is, $\hat{\gamma}$ is smaller in the former period. As the transition function must take a value of zero when there is no mispricing, this result implies that the transactions costs faced by arbitragers in the post-SETS period are smaller than those faced in the pre-SETS period.

The profiles presented in Figure 5.2 plot the estimated transition function against the pricing error using minute frequency data (panel a)), two minute frequency data (panel b)) and five minute frequency data (panel c)) for values of the delay parameter d selected by SIC. This figure shows that there is a sharper change from no adjustment ($G(z_{t-d}) = 0$) to full adjustment ($G(z_{t-d}) = 1$) in the post-SETS period than in the pre-SETS period. Put differently, there is full adjustment outside a narrow range of mispricing in the post-SETS period. By contrast, this range is considerably larger in the pre-SETS period.

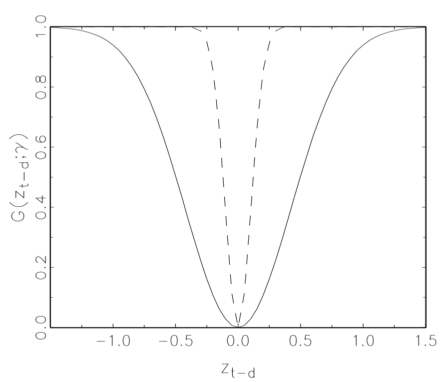
To formally test equality of the values of γ over the two sample periods a simple t-test based on heteroscedastic-consistent standard errors is performed for various delay values. In each case the same delay values are assumed in each period. In addition, the optimal delays, as given by the SIC, are used in each period and the t-statistic is calculated. The results pertaining to various sampling frequencies are



(a) One minute frequency



(b) Two minute frequency



(c) Five minute frequency

Figure 5.2: Estimated transition functions in pre-SETS (solid line) and post-SETS (dashed line) periods with d set equal to the value that is preferred by SIC.

Table 5.11: Transaction cost difference tests

	d					
Frequency	1	2	3	4	5	SIC
1M	0.10	4.00	0.35	0.34	1.33	2.64
	(0.46)	(0.00)	(0.36)	(0.37)	(0.09)	(0.00)
2M	2.81	3.44				1.26
	(0.00)	(0.00)				(0.10)
5M	3.35					
	(0.00)					

Frequencies of one minute (1M), two minutes (2M), and five minutes (5M) are considered. The null hypothesis that the γ coefficient in the pre-SETS period equals the γ coefficient in the post-SETS period is tested against the alternative that the pre-SETS γ is less than the post-SETS γ. The t-statistics associated with the difference between the pre-SETS γ and the post-SETS γ are reported. The standard error of this difference is calculated using the heteroscedasticity-consistent standard error. The numbers in parentheses are the p-values associated with this test.

presented in Table 5.11. When one minute frequency data are used transaction costs are significantly lower in the post-SETS period when delays of two and five minutes are assumed and when optimal delays are assumed. When two and five minute frequency data are used the results indicate that transaction costs are universally significantly lower in the post-SETS period.

Adjustment in response to pricing error

In using an STEqCM to model the arbitrage process one can obtain estimates of the adjustment in futures and spot markets for a given pricing error. Using (5.28), the respective adjustments due to previous pricing errors in futures and spot markets are

$$A\big(z_{t-d};\gamma;\alpha_f\big) = \alpha_f z_{t-d} G(z_{t-d};\gamma), \tag{5.29}$$

$$A\big(z_{t-d};\gamma;\alpha_s\big) = \alpha_s z_{t-d} G(z_{t-d};\gamma). \tag{5.30}$$

Using one minute frequency data and optimal delays, $A(.)$ is plotted against z_{t-d} and presented in Figure 5.3. Before the results are discussed consider a few presentation issues. First, as α_f and α_s take different signs the absolute values of $A(.)$ are used. Second, as $|A(.)|$ is symmetric about zero only positive pricing errors are considered. Third, the vertical axis in the figure is truncated so that a visual comparison of the various adjustments can be achieved.

Figure 5.3 shows that the introduction of SETS causes an increase in the level of adjustment in the spot market for all values of the past pricing error. In the futures market, the level of adjustment is greater (smaller) in the post-SETS period for small (large) pricing errors. The $A\big(z_{t-d};\gamma;\alpha_f\big)$ measures in the two periods are equal for $|z_{t-d}| \approx 0.3$. For the post-SETS period, the relevant range of z_{t-d} is in

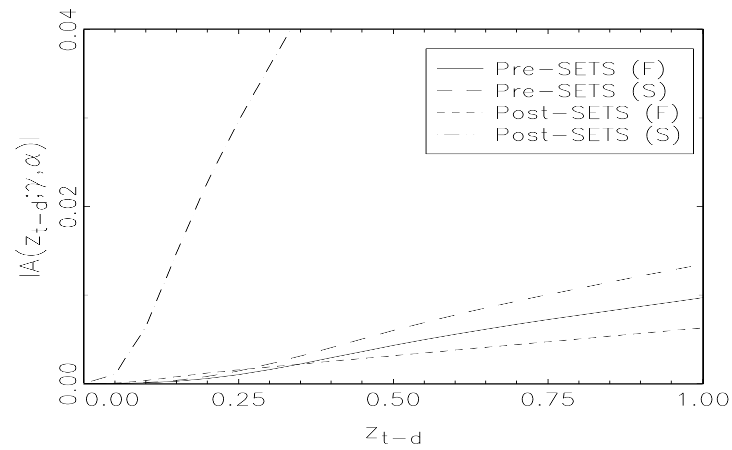


Figure 5.3: Adjustment in futures and spot markets for one minute frequency data, and the delay d set equal to the value selected by SIC.

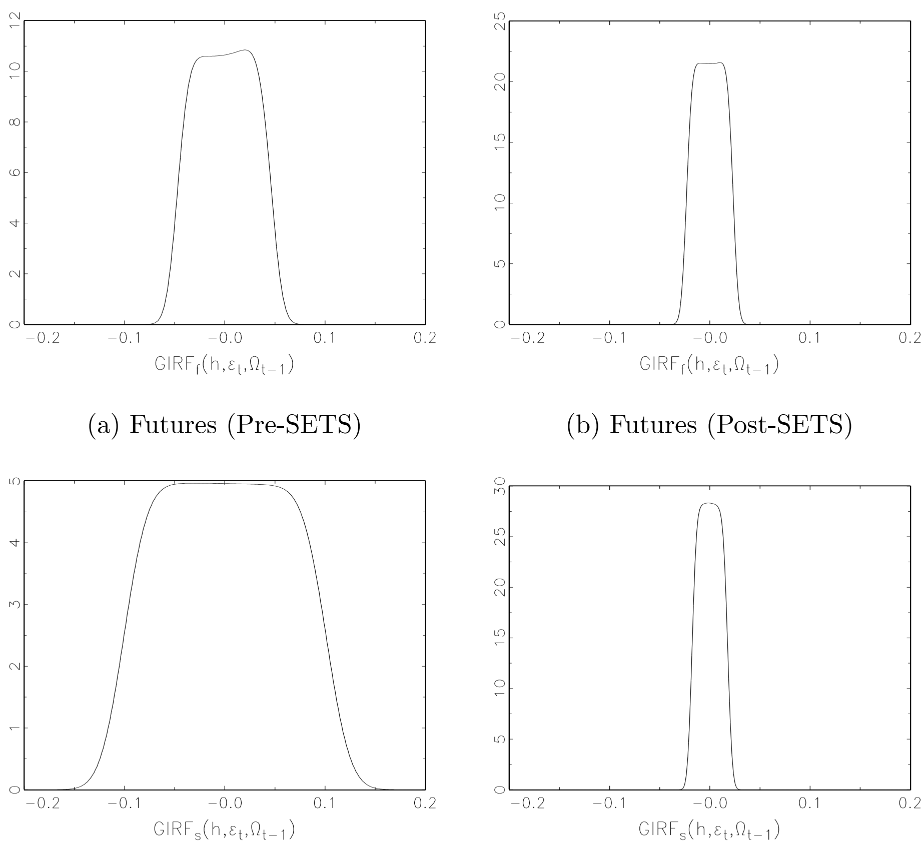
fact roughly equal to $(-0.3, 0.3)$, as values of z_{t-d} outside this range hardly ever occur. Thus prior to the introduction of SETS, prices were not adjusting as swiftly as possible due to prohibitive transaction costs. Since the introduction of SETS spot prices, in particular, are rapidly and fully adjusting to past mispricing because of lower transaction costs in the spot market. In this sense both spot and futures markets have become more efficient since the introduction of SETS.

Generalized impulse responses

Generalized impulse response functions are calculated using the smooth transition equilibrium correction models estimated in the pre-SETS and post-SETS periods. In both cases, one minute frequency data are used and the delay d is selected by the SIC. Shocks equal to -0.4, -0.35, -0.3, ..., 0.35, and 0.4 are assumed to affect both spot and futures markets. The effects that these shocks have on subsequent spot and futures returns are measured at various points within the pre-SETS and post-SETS sample periods.[13] The distribution of these innovations is estimated using a quartic kernel function at various time periods after the shock hits the system.[14] Due to the selected values of the shocks that hit the system, a uniform distribution taking values between -0.4 and 0.4 (inclusive) is observed when the shock occurs. Subsequent distributions are less uniform and have a smaller range as the effects of the shock gradually disappear. This rate of decay gives an indication of the speed of adjustment in the respective markets. The estimated distributions at two and five minutes after the initial shock occurs are presented in Figures 5.4 and 5.5, respectively.

[13] This estimation process is based on a sub-sample of the pre-SETS and post-SETS periods. The 'histories' used in the current context equal the 1st, 101st, 201st, ..., 13,301st, and 13,401st observations. These histories are selected from a sample consisting of 13,500 observations. Selection of these histories is used to reduce the computation time.

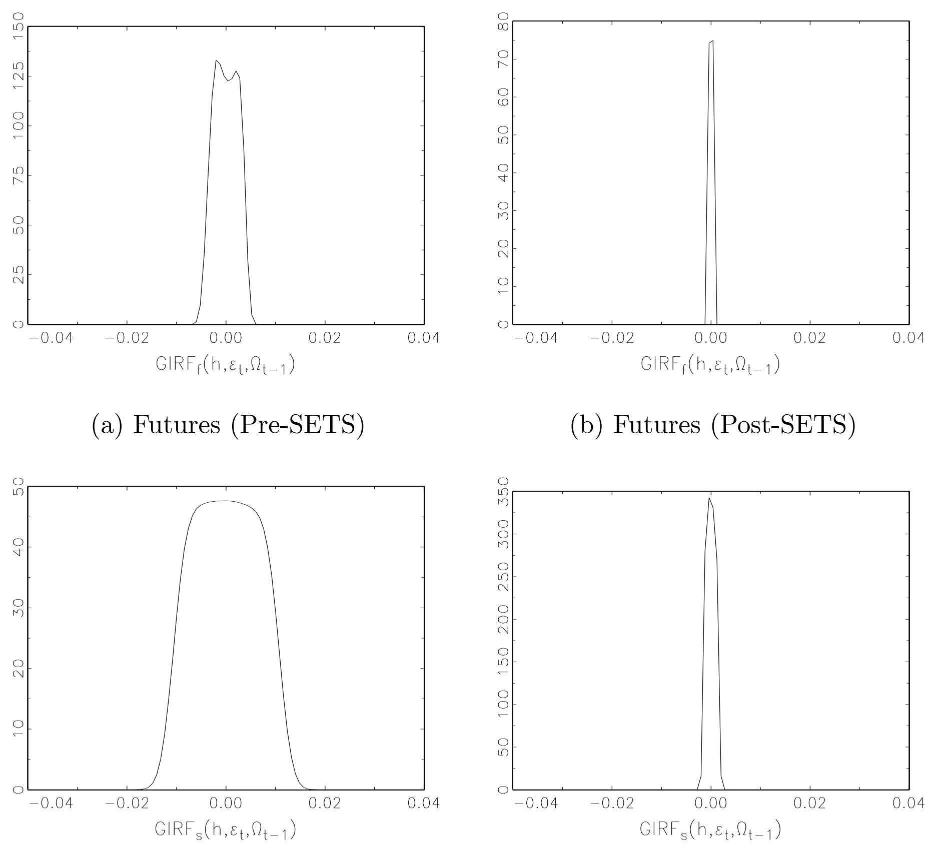
[14] For further details of kernel functions, see Wand and Jones (1995) and Fan and Gijbels (1996), among others. The optimal bandwidth is determined using equation (3.31) of Silverman (1986).



(a) Futures (Pre-SETS) (b) Futures (Post-SETS)

(c) Spot (Pre-SETS) (d) Spot (Post-SETS)

Figure 5.4: Generalized impulse response distributions (two minutes after shock)



(a) Futures (Pre-SETS)

(b) Futures (Post-SETS)

(c) Spot (Pre-SETS)

(d) Spot (Post-SETS)

Figure 5.5: Generalized impulse response distributions (five minutes after shock)

The introduction of SETS causes more rapid adjustment in both spot and futures markets. This can be observed by comparing Panels (a) and (c) with (b) and (d), respectively, in Figures 5.4 and 5.5. In each case the range of values taken by the innovations is smaller in the post-SETS period than in the pre-SETS period. Moreover, in the pre-SETS period adjustment in the futures market is faster than adjustment in the spot market. That is, after two and five minutes the range of the innovations in the futures equation is smaller than the range of innovations in the spot equation. By contrast, adjustment is similar in both markets in the post-SETS period. This increased speed of adjustment to exogenous shocks in the post-SETS period is clear evidence of improved efficiency under the new trading system.

Concluding, the transaction costs faced by arbitragers trading in the FTSE100 spot and futures markets have been significantly reduced since the introduction of the new electronic trading system SETS. As such, both markets have become more efficient. Analysis of generalized impulse response functions leads to two additional findings. First, shocks to the futures and spot markets have less effect in the post-SETS period. Indeed, the effects of such shocks almost disappear after five minutes. Second, the futures market is less affected by shocks than the spot market in the pre-SETS period. However, both markets appear to be equally affected in the post-SETS period. These two findings are consistent with the objectives of SETS, that is, to improve the efficiency of the FTSE100 market.

5.4 Concluding remarks

In this chapter I have described a generalization of the STAR model to a multivariate setting. The multivariate STAR model is potentially useful to describe relationships between time series that are subject to regime-switches, or to describe nonlinear adjustment towards (linear) long-run equilibrium relations. The latter is illustrated by the application to spot and futures prices in Section 5.3. The specification procedure for univariate STAR models, discussed in Chapter 2, can also be used to specify multivariate models. The Monte Carlo experiments examining the size and power properties of the LM-type tests for multivariate STAR nonlinearity suggest that they can yield reliable results. However, further empirical experience is needed to assess its usefulness in practice.

Chapter 6

Testing for Nonlinearity in the Presence of Outliers

Regime-switching models, like the smooth transition autoregressive [STAR] model, have been applied most frequently to study possible nonlinearity in the dynamic behaviour of macro-economic variables over the business cycle, see Teräsvirta and Anderson (1992), Beaudry and Koop (1993), Tiao and Tsay (1994), Potter (1995b), Pesaran and Potter (1997), and Chapter 3 of this thesis for extensive discussion and additional references. Most macro-economic variables are sampled only quarterly or annually. Therefore, only series of moderate length are available and, consequently, it may be that possible nonlinear properties are reflected in only a few observations. One may then be tempted to view these 'nonlinear data points' as aberrant observations and remove them using one of the familiar outlier removal techniques, see Balke and Fomby (1994), among others. This might even be justified by noting that nonlinear time series models typically involve many additional parameters and one may want to prevent estimating these parameters for only a few observations. However, removing outliers too drastically may accidentally destroy intrinsic nonlinearity that, for example, could have been exploited for forecasting. Conversely, in case of a linear time series that is contaminated with outliers, nonlinearity tests may point towards nonlinear structures, which in turn can lead to estimating unnecessarily complicated models. In sum, there seems to be a need for modeling strategies which are capable of distinguishing between nonlinearity and outliers. In the third part of this thesis, comprising Chapters 6 and 7, I focus on two important ingredients of such a modeling strategy: outlier robust testing for nonlinearity and outlier robust estimation of nonlinear time series models.

In the present chapter I propose modifications of the tests for smooth transition nonlinearity developed by Luukkonen *et al.* (1988) that are robust to the presence of outliers. By using outlier robust estimation techniques (see, e.g., Huber (1981), Martin (1981), Hampel *et al.* (1986) and Lucas (1999)), tests are obtained that have better size and power properties than standard non-robust tests in situations with outliers.

The outline of this chapter is as follows. In Section 6.1, I briefly discuss some outlier models that are considered relevant in the time series literature. I also review the effects of outliers on ordinary least squares [OLS] estimates of the parameters in linear time series models, as this proves to be of vital importance for the subject of this and the next chapter. Robust estimation methods for linear time series models are considered in Section 6.2. In Section 6.3 I summarize the Lagrange Multiplier [LM] type tests for STAR nonlinearity, discussed at length in Section 2.2. At the end of this section, I propose robustified versions of the test statistics. The effect of outliers on these nonlinearity tests is investigated analytically in Section 6.4, while the empirical size and power properties are evaluated by means of Monte Carlo experiments in Section 6.5. The robust testing procedure is found to work remarkably well. An empirical illustration is provided in Section 6.6, in which the tests are applied to various industrial production series. The general outcome is that I find similar evidence for nonlinear features based on standard and robust testing procedures, except for four series for which apparent nonlinearity appears to be due to few observations. Finally, Section 6.7 contains some concluding remarks.

6.1 A brief discussion on outliers

A useful starting point for a brief discussion on outliers in time series is the replacement model of Martin and Yohai (1986),

$$y_t = z_t(1-\delta_t) + \zeta_t\delta_t, \tag{6.1}$$

where δ_t is a binary random variable which equals 1 with probability π and is 0 otherwise. The observed time series y_t is a mixture of a *core* process z_t and a *contaminating* process ζ_t. These names for the two components of y_t stem from the fact that usually π is small (≤ 0.10), such that $y_t = z_t$ most of the time, and only occasionally $y_t = \zeta_t$.

Different specifications of the z_t, ζ_t and δ_t processes can generate a wide variety of outlier patterns. Intuitively it is clear that outliers are always defined with respect to the process that is assumed to describe regular data points: certain observations might be outliers in one specification for the core process z_t and at the same time be perfectly regular observations in another specification, see Davies and Gather (1993) for a more formal discussion. Most of the literature on outlier detection and estimation in the presence of outliers has concentrated on linear processes for z_t, and I follow this practice here. More specifically, in the following I assume that z_t is governed by an autoregressive [AR] process of order p

$$\phi(L)z_t = \eta_t, \tag{6.2}$$

where $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$ is a polynomial in the lag operator L, defined as $L^k z_t = z_{t-k}$ for all k, and where $\eta_t \sim$ i.i.d. $(0, \sigma_\eta^2)$. Furthermore, the process δ_t in (6.1) is assumed to be i.i.d., such that outliers occur in isolation[1].

[1] If δ_t is allowed to be serially correlated, blocks of outliers can occur, see Bruce and Martin (1989) and Wu, Hosking and Ravishankar (1993).

Following Fox (1972), two specific types of contamination usually are considered to be of special interest in the context of time series. First, an additive outlier [AO] model is obtained if $\zeta_t = z_t + \zeta$ for some constant ζ, such that (6.1) reduces to

$$y_t = z_t + \zeta\delta_t. \tag{6.3}$$

An innovation outlier [IO] model results if $\zeta_t = z_t + \zeta/\phi(L)$, which yields

$$y_t = z_t + \zeta\frac{\delta_t}{\phi(L)}, \tag{6.4}$$

The AO case gives a one time effect on the level of the time series because only the current observation y_t is affected. In the IO model, however, an outlier at time t also influences future observations $y_{t+1}, y_{t+2}, \ldots$, through the same dynamics as the core process z_t.[2]

To understand the effects of an isolated AO or IO occurring at time $t = \tau$, assume that the core process z_t is described by an AR(1) model

$$z_t = \phi_1 z_{t-1} + \eta_t, \qquad t = 1, \ldots, T, \tag{6.7}$$

with $|\phi_1| < 1$ and T denoting the sample size. In the AR(1) case, an AO corresponds with two outliers in the (y_{t-1}, y_t) plane. Using the classification of Rousseeuw and van Zomeren (1990), the point $(y_{\tau-1}, y_\tau)$ is a *vertical outlier* as (only) y_τ falls outside the range of the majority of the data. The next point $(y_\tau, y_{\tau+1})$ is a so-called *bad leverage point*, characterized by an abnormal value of the regressor. In case of an IO on the other hand, the vertical outlier at $t = \tau$ is followed by a number of *good leverage points*, characterized by large values for both y_{t-1} and y_t, that approximately satisfy the linear AR(1) model. An example is shown in Figure 6.1. A series z_t of length $T = 100$ is generated according to the AR(1) model (6.7) with $\phi_1 = 0.7$ and $\eta_t \sim$ i.i.d.$N(0, \sigma^2)$, $\sigma = 0.1$. A single outlier of magnitude 5σ occurs

[2] Two other types of outliers that are considered quite frequently (see Tsay (1988), Balke (1993) and Chen and Liu (1993), among others) are temporary changes and level shifts. A temporary change occurs when $\zeta_t = z_t + \zeta_t/(1 - \xi L)$ with $|\xi| < 1$, such that (6.1) becomes

$$y_t = z_t + \zeta\frac{\delta_t}{1 - \xi L}. \tag{6.5}$$

Comparing (6.5) with (6.4), a temporary change is seen to be similar to an IO, in the sense that a temporary change also affects future observations. The pattern of decline of the effect of the temporary change can be quite different than the pattern of regular shocks η_t however. A level shift occurs when $\xi = 1$ in (6.5). In that case, the unconditional mean of the observed time series y_t shifts permanently from 0 to $\zeta/(1 - \phi(1))$.

At this point it is useful to note that all outlier types discussed here can also be cast in terms of the general intervention model of Box and Tiao (1975),

$$y_t = \zeta\frac{\omega(L)}{\xi(L)}\delta_t + \varepsilon_t/\phi(L), \tag{6.6}$$

It is easily seen that suitable choices for the lag polynomials $\omega(L)$ and $\xi(L)$ render AO, IO, temporary change and level shift effects.

at $t = \tau = 50$. The AO gives a one-time spike in the series (see panel (a)) and the vertical outlier $(y_{\tau-1}, y_\tau)$ and bad leverage point $(y_\tau, y_{\tau+1})$ clearly stand out in the scatter of y_t against y_{t-1} (panel (c)). In case of an IO, the time series returns slowly to its normal level after the occurrence of the outlier (panel (b)). The scatter demonstrates that the vertical outlier now is followed by a sequence of good leverage points (panel (d)).

Denby and Martin (1979), Bustos and Yohai (1986) and Martin and Yohai (1986), among others, consider estimation of the parameters of AR models in the presence of outliers. In the presence of IOs, the ordinary least squares [OLS] estimates of the autoregressive parameters are consistent, although they are inefficient. AOs have a much more disastrous effect on the OLS estimates. If, for example, the core process z_t follows the stationary AR(1) model (6.7) and the observed time series y_t is contaminated with an isolated additive outlier of magnitude ζ at $t = \tau$, the OLS estimate of ϕ_1 calculated with the observed series is equal to

$$\hat{\phi}_1 = \frac{\sum_{t=1}^{T} y_{t-1} y_t}{\sum_{t=1}^{T} y_{t-1}^2} = \frac{\zeta(z_{\tau-1} + z_{\tau+1}) + \sum_{t=1}^{T} z_{t-1} z_t}{\zeta^2 + 2\zeta z_\tau + \sum_{t=1}^{T} z_{t-1}^2}. \tag{6.8}$$

This expression shows that for large AOs $\hat{\phi}_1$ will be biased towards zero, due to the bad leverage point $(y_\tau, y_{\tau+1})$, see Figure 6.1, panel (c). In general, if the probability of occurrence of an AO of size ζ is equal to π, the probability limit of the OLS estimator of ϕ_1 is given by

$$\operatorname*{plim}_{T\to\infty} \ \hat{\phi}_1 = \frac{\phi_1}{1 + \frac{\pi \zeta^2}{\sigma_z^2}}, \tag{6.9}$$

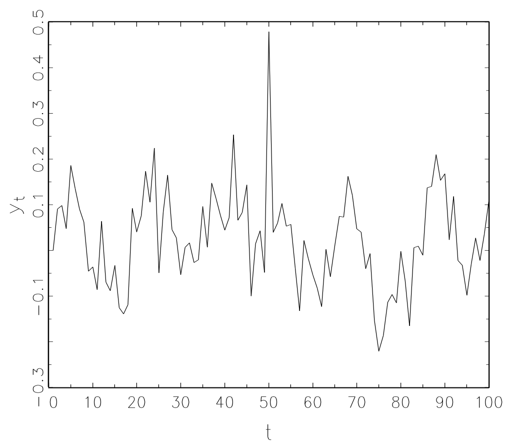
where σ_z^2 denotes the variance of z_t, $\sigma_z^2 = \sigma_\eta^2/(1-\phi_1^2)$, see Denby and Martin (1979).

Over the years, a number of outlier detection and correction procedures have been developed, see Tsay (1986a,1988), Chang, Tiao and Chen (1988) and Chen and Liu (1993), among others. All these procedures are characterized by an iterative 'estimation-detection-correction-estimation' scheme, which may make them subjective and time-consuming. An additional drawback of such methods is that the types of outliers which (may) occur have to be parametrized explicitly, as they are based on the intervention model given in (6.6). An alternative method to cope with outliers, which avoids both of these disadvantages, is to use robust estimation techniques. These latter techniques are the subject of the next section.

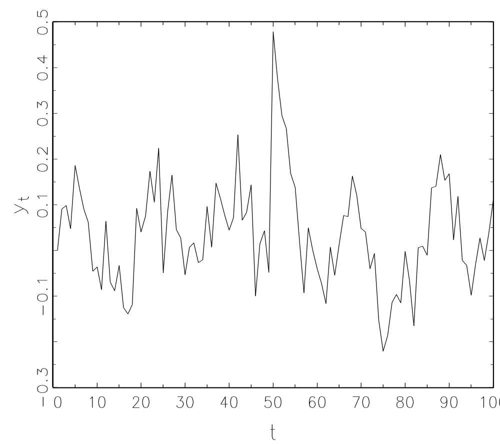
6.2 Robust estimation methods for linear time series models

In this section I highlight some issues in robust estimation of linear AR models. These estimation methods are used to modify the LM-type tests for STAR nonlinearity in the next section. For a more general discussion I refer to Denby and Martin (1979), Martin (1981), and Lucas (1999).

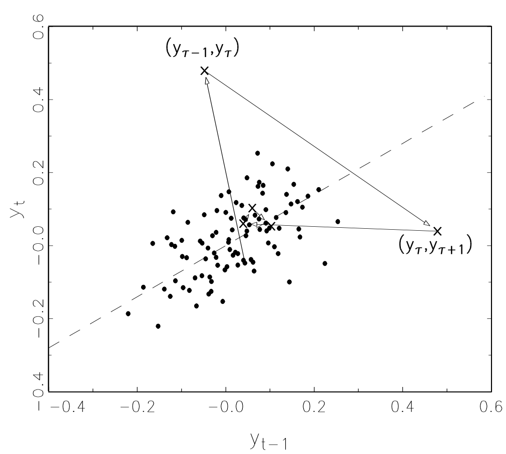
Consider again the simple AR(1) model (6.7) for the outlier-free process z_t with the errors η_t assumed to be independent white noise. Furthermore, assume that the



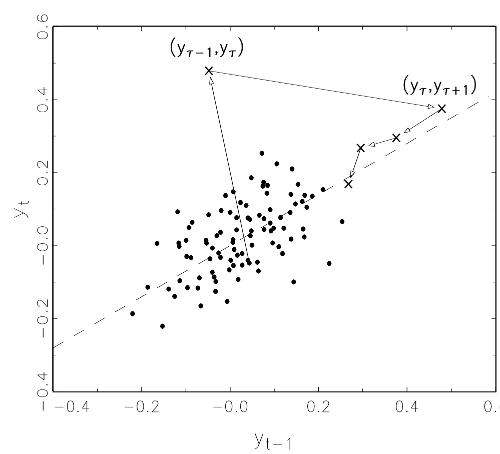
(a) Additive outlier



(b) Innovation outlier



(c) Additive outlier



(d) Innovation outlier

Figure 6.1: Example of the effects of a single AO (panels (a) and (c)) or IO (panels (b) and (d)). The core process z_t is generated according to the AR(1) model (6.7), with $\phi_1 = 0.7$ and $\eta_t \sim$ i.i.d. $N(0, \sigma^2)$ with $\sigma = 0.1$. A single outlier of size 5σ occurs at $t = \tau = 50$. The observations $(y_{\tau+j-1}, y_{\tau+j})$ for $j = 0, 1, \ldots, 4$ are indicated with crosses. The dashed line in panels (b) and (d) is the skeleton of the AR(1) model, $y_t = \phi_1 y_{t-1}$.

observed time series y_t follows the replacement model (6.1) for a general contamination process ζ_t. The OLS estimate of the autoregressive parameter ϕ_1 based on the observed series minimizes the sum of squared residuals, which can be characterized by the first order condition

$$\sum_{t=1}^{T} y_{t-1}(y_t - \phi_1 y_{t-1}) = 0. \tag{6.10}$$

To avoid the deficiencies of the OLS estimator in the presence of contamination as discussed previously, the autoregressive parameter can be estimated robustly using maximum likelihood type [M] or Generalized M [GM] estimators. The class of GM estimators is designed to obtain better estimates of ϕ_1 in the presence of contamination, by giving less weight to influential observations such as leverage points and vertical outliers. Here I consider the Schweppe type of GM estimators (Handschin *et al.* (1975)), which solves the alternative first order condition

$$\sum_{t=1}^{T} w_x(y_{t-1})y_{t-1} \cdot \psi\left(\frac{y_t - \phi_1 y_{t-1}}{\sigma_\varepsilon w_x(y_{t-1})}\right) = 0, \tag{6.11}$$

where σ_ε is a measure of scale of the residuals $\varepsilon_t \equiv y_t - \phi_1 y_{t-1}$, $\psi(r_t)$ is an odd and bounded function, with r_t denoting the t-th standardized residual, $r_t \equiv (y_t - \phi_1 y_{t-1})/(\sigma_\varepsilon w_x(y_{t-1}))$, and $w_x(\cdot)$ is a weight function that assigns weights between 0 and 1 to the regressor y_{t-1}. The function $\psi(\cdot)$ must satisfy certain additional regularity conditions in order for the GM estimator to be consistent and asymptotically normal, see Hampel *et al.* (1986).

If the regressors are not weighted, that is, if $w_x(y_{t-1}) = 1$ for all t, the GM estimator reduces to an M estimator, and the usual OLS estimator is obtained if, in addition, $\psi(r_t) = r_t$. Denby and Martin (1979) show that, in the presence of IOs, M estimators are efficient, whereas the asymptotic variance of GM estimators is larger due to the weighting of the regressors. Both M and GM estimators are asymptotically biased in an AO setting, although, if $w_x(\cdot)$ and $\psi(\cdot)$ are chosen properly, the bias of the GM estimator can be considerably smaller. I focus on the Schweppe form of the GM estimator, because this estimator only downweights vertical outliers and bad leverage points, while it fully exploits the correct signal in good leverage points, see Hampel *et al.* (1986).

Defining $w_r(r_t) = \psi(r_t)/r_t$ for $r_t \neq 0$ and $w_r(0) = 1$, the first order condition (6.11) can be rewritten as

$$\sum_{t=1}^{T} w_r\left(\frac{y_t - \phi_1 y_{t-1}}{\sigma_\varepsilon w_x(y_{t-1})}\right) y_{t-1}(y_t - \phi_1 y_{t-1}) = 0, \tag{6.12}$$

from which it can be inferred that the GM estimator is a type of weighted least squares estimator. The weight for the t-th observation is given by the value of $w_r(\cdot)$, which depends on the unknown parameter ϕ_1. The functions $w_x(\cdot)$ and $\psi(\cdot)$ now should be chosen such that the t-th observation receives a relatively small weight if either y_{t-1} or $(y_t - \phi_1 y_{t-1})/\sigma_\varepsilon$ becomes large (in absolute value).

Common choices for the $\psi(\cdot)$ function are the Huber and Tukey bisquare functions. The Huber ψ function is given by

$$\psi(r_t) = \begin{cases} -c & \text{if } r_t \leq -c, \\ r_t & \text{if } -c < r_t \leq c, \\ c & \text{if } r_t > c, \end{cases} \tag{6.13}$$

or $\psi(r) = \text{med}(-c, c, r)$, where med denotes the median and $c > 0$. The tuning constant c determines the robustness and efficiency of the resulting estimator. Because these properties are decreasing and increasing functions of c, respectively, the tuning constant should be chosen such that the two are balanced. Usually c is taken equal to 1.345 to produce an estimator that has an efficiency of 95% compared to the OLS estimator if ε_t is normally distributed. The weights $w_r(r_t)$ implied by the Huber function have the attractive property that $w_r(r_t) = 1$ if $-c \leq r_t < c$. Only observations for which the standardized residual is outside this region receive less weight. A disadvantage is that these weights decline to zero only very slowly. Thus, subjective judgement is required to decide whether a weight is small or not.

The Tukey bisquare function is given by

$$\psi(r_t) = \begin{cases} r_t(1 - (r_t/c)^2)^2 & \text{if } |r_t| \leq c, \\ 0 & \text{if } |r_t| > c. \end{cases} \tag{6.14}$$

The tuning constant c again determines the robustness and the efficiency of the resultant estimator. Usually c is set equal to 4.685, again to achieve 95% efficiency for normally distributed ε_t. The Tukey function might be considered as the mirror-image of the Huber function, in the sense that downweighting occurs for all nonzero values of r_t, but the resulting weights decline to 0 quite rapidly.

In this chapter I use the polynomial ψ function as proposed in Lucas *et al.* (1996), given by

$$\psi(r_t) = \begin{cases} r_t & \text{if } |r_t| \leq c_1, \\ \text{sgn}(r_t) g(|r_t|) & \text{if } c_1 < |r_t| \leq c_2, \\ 0 & \text{if } |r_t| > c_2, \end{cases} \tag{6.15}$$

or more compactly,

$$\psi(r_t) = r_t I[|r_t| \leq c_1] + I[|r_t| > c_1] I[|r_t| \leq c_2] \text{sgn}(r_t) g(|r_t|), \tag{6.16}$$

where c_1 and c_2 are tuning constants, $I[A]$ is the indicator function for the event A, defined as $I[A] = 1$ if A is true and $I[A] = 0$ otherwise, sgn is the signum function, and $g(|r_t|)$ is a fifth order polynomial such that $\psi(r_t)$ is twice continuously differentiable. This ψ function combines the attractive properties of the Huber and Tukey ψ-functions. Observations receive a weight $w_r(r_t) = \psi(r_t)/r_t$ equal to 1 if their standardized residuals are within $(-c_1, c_1)$ and a weight equal to zero if the residuals are larger than c_2 in absolute value. The polynomial $g(|r_t|)$ is such that

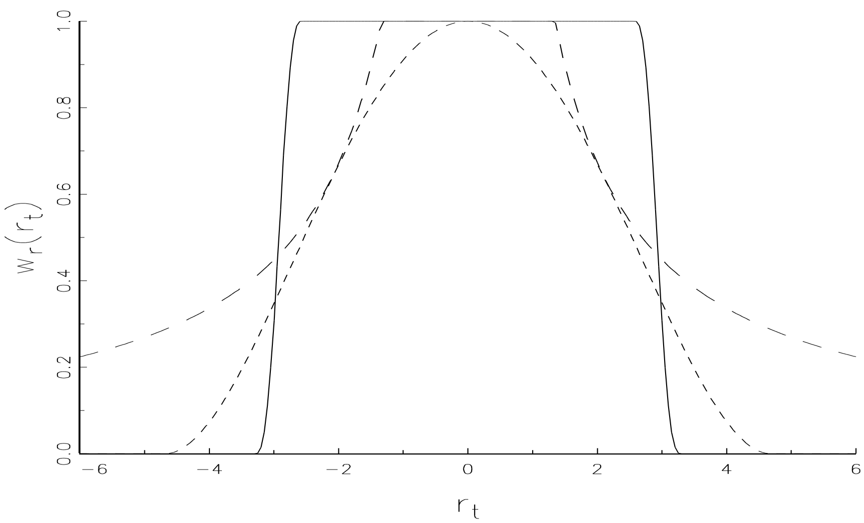


Figure 6.2: Weight functions $w_r(r_t)$ as implied by the polynomial ψ function given in (6.15), with $c_1 = 2.576$ and $c_2 = 3.291$ (solid line), the Huber function given in (6.13) with $c = 1.345$ (long dashed line), and the Tukey function given in (6.14) with $c = 4.685$ (short dashed line).

partial weighting occurs in-between. The tuning constants c_1 and c_2 are taken to be the square roots of the 0.99 and 0.999 quantiles of the $\chi^2(1)$ distribution, that is, $c_1 = 2.576$ and $c_2 = 3.291$.

The weights implied by the three ψ functions discussed above are shown in Figure 6.2, which clearly demonstrates the differences and similarities between the different functions.

The weight function $w_x(\cdot)$ for the regressor is specified as

$$w_x(y_{t-1}) = \psi(d(y_{t-1})^\alpha)/d(y_{t-1})^\alpha \ , \tag{6.17}$$

where $\psi(\cdot)$ again is given by (6.15), $d(y_{t-1})$ is the Mahalanobis distance of y_{t-1}, that is, $d(y_{t-1}) = |y_{t-1} - m_y|/\sigma_y$, with m_y and σ_y measures of location and scale of y_{t-1}, respectively. These measures are estimated robustly by the median $m_y = \text{med}(y_{t-1})$ and the median absolute deviation [MAD] $\sigma_y = 1.483 \cdot \text{med}|y_{t-1} - m_y|$, respectively. The constant 1.483 is used to make the MAD a consistent estimator of the standard deviation in case ε_t is normally distributed. Finally, following Simpson *et al.* (1992), the constant α in (6.17) is set equal to 2 to obtain robustness of standard errors.

The first order condition (6.12) is nonlinear in ϕ_1 and, therefore, estimation requires an iterative procedure. In fact, interpreting $w_r(\cdot)$ as a function of ϕ_1 and σ_ε, $w_r(\phi_1, \sigma_\varepsilon)$, and denoting the estimates of ϕ_1 and σ_ε at the kth iteration by $\hat{\phi}_1^{(k)}$ and $\hat{\sigma}_\varepsilon^{(k)}$, respectively, it follows from (6.12) that $\hat{\phi}_1^{(k+1)}$ might be computed as the weighted least squares estimate

$$\hat{\phi}_1^{(k+1)} = \frac{\sum_{t=1}^T w_r(\hat{\phi}_1^{(k)}, \hat{\sigma}_\varepsilon^{(k)}) y_{t-1} y_t}{\sum_{t=1}^T w_r(\hat{\phi}_1^{(k)}, \hat{\sigma}_\varepsilon^{(k)}) y_{t-1}^2}. \tag{6.18}$$

The estimate of σ_ε can be updated at each iteration using the MAD estimator given above.

To have maximum protection against outliers, the breakdown point of the estimator, that is, the maximum fraction of contaminated observations the estimator can cope with before producing nonsensical results, should be as high as possible. I follow Simpson *et al.* (1992) and Coakley and Hettmansperger (1993), who show that if a high breakdown point [HBP] estimator is used to construct starting values and if only one iteration according to (6.18) is performed, an efficient estimator is obtained which retains the high breakpoint of the initial estimator. I use the least median of squares [LMS] estimator of Rousseeuw (1984) to obtain a starting value for the autoregressive parameter, $\hat{\phi}_1^{(0)}$, and apply the MAD estimator to the corresponding residuals to obtain an initial scale estimate, $\hat{\sigma}^{(0)}$. In the context of regression models, the LMS estimator has a breakdown point of (approximately) 0.5, that is, the LMS estimator can give reliable parameter estimates even if (slightly less than) half of the observations in the sample is contaminated. As shown by Lucas (1997), in a time series context the breakdown point of the LMS estimator is much lower and depends on the true value of the autoregressive parameter(s). To the best of my knowledge, it has not been investigated whether alternative estimators which have a high breakdown point in a regression context retain this in a time series context.

Generalizing the robust HBP-(G)M estimator to allow the core process z_t to follow an AR(p) model with $p > 1$ as in (6.2) is fairly straightforward, except that now HBP estimators for multivariate (p-dimensional) location M_x and scatter Σ_x are required to compute the Mahalanobis distances for the regressors $\tilde{x}_t \equiv (y_{t-1}, \ldots, y_{t-p})'$

$$d(\tilde{x}_t) \equiv \sqrt{(\tilde{x}_t - M_x)' \, \Sigma_x^{-1} \, (\tilde{x}_t - M_x)}, \tag{6.19}$$

which are to be used in the weight function for the regressors, given in (6.17). For this purpose, I use the minimum volume ellipsoid [MVE] estimator proposed by Rousseeuw (1985). The projection algorithm of Rousseeuw and van Zomeren (1990) is used to approximate this estimator. In the next section I will use the HBP-GM estimator to modify tests for STAR nonlinearity.

6.3 Testing for smooth transition nonlinearity

The Lagrange Multiplier [LM] tests to identify STAR-type nonlinearity of Luukkonen, Saikkonen and Teräsvirta (1988) [LST hereafter] are discussed in detail in Section 2.2. Here I briefly recall the most important aspects of the standard tests, before introducing outlier robust variants.

Consider the basic STAR model of order p [STAR(p)] for a univariate time series y_t,

$$y_t = \phi_1' x_t (1 - G(s_t; \gamma, c)) + \phi_2' x_t G(s_t; \gamma, c) + \varepsilon_t, \tag{6.20}$$

where $x_t = (1, \tilde{x}_t)', \tilde{x}_t = (y_{t-1}, \ldots, y_{t-p})'$, $\phi_i = (\phi_{i,0}, \phi_{i,1}, \ldots, \phi_{i,p})'$, $i = 1, 2$, and the transition function $G(s_t; \gamma, c)$ is taken to be a continuous function that is bounded between 0 and 1. For a logistic STAR [LSTAR] model, the transition function $G(s_t; \gamma, c)$ is taken to be the logistic function

$$G(s_t; \gamma, c) = (1 + \exp\{-\gamma(s_t - c)\})^{-1}, \qquad \gamma > 0, \tag{6.21}$$

whereas for an exponential STAR [ESTAR] model, the transition function is defined as either the exponential function

$$G(s_t; \gamma, c) = 1 - \exp\{-\gamma(s_t - c)^2\}, \qquad \gamma > 0, \tag{6.22}$$

or the quadratic logistic function

$$G(s_t; \gamma, c) = (1 + \exp\{-\gamma(s_t - c_1)(s_t - c_2)\})^{-1}, \qquad c_1 \leq c_2,\ \gamma > 0, \tag{6.23}$$

where in the latter case $c \equiv (c_1, c_2)'$. In this chapter, I concentrate on the tests against STAR models $s_t = y_{t-d}$, for certain $0 < d \leq p$, where d is either known or unknown. Recall from Section 2.2 that the latter situation can be described as $s_t = \alpha'\tilde{x}_t$ with $\alpha = (0, \ldots, 0, 1, 0 \ldots, 0)'$, where the 1 is the d-th position of α. However, the analysis can be extended to allow for exogenous transition variables or time varying parameters (which would arise if $s_t = t$).

The null hypothesis of linearity can be taken as $H_0 : \phi_{1,j} = \phi_{2,j}$ for $j = 0, \ldots, p$ in (6.20). This hypothesis is tested against the alternative $H_1 : \phi_{1,j} \neq \phi_{2,j}$ for at least one j. As discussed extensively in Section 2.2, the testing problem is complicated by the fact that the model contains nuisance parameters (γ and c) that are not identified under the null hypothesis. As suggested by LST, this identification problem can be solved by replacing the transition function $G(s_t; \gamma, c)$ in (6.20) by a suitable Taylor approximation. In general, the reparameterized model that is used for linearity testing can be written as

$$y_t = \beta' x_t + \theta' q_t + e_t, \tag{6.24}$$

where q_t is an $m \times 1$ vector of auxiliary regressors containing higher-order and cross-product terms of $y_{t-1}, \ldots, y_{t-p}$. The exact contents of the vector q_t depends on the approximation of the transition function chosen, see Section 2.5 for details.

In all cases, the elements of the parameter vectors β and θ are defined in terms of the parameters $\phi_1, \phi_2, \gamma, \alpha$ and c in the original model, such that the null hypothesis of linearity, $H_0 : \phi_1 = \phi_2$, is equivalent to the hypothesis that the coefficients of the auxiliary regressors in (6.24) are 0, that is, $H_0' : \theta = 0$. This null hypothesis can be tested by a variable addition test in a straightforward manner. Specifically, the LM-type test is given by

$$\begin{aligned} \text{LM} &= \frac{\hat{\varepsilon}' X (X'X)^{-1} X' \hat{\varepsilon}}{\hat{\varepsilon}'\hat{\varepsilon}/T} \\ &= \frac{\hat{\varepsilon}'\hat{\varepsilon} - \hat{\varepsilon}'(I_T - X(X'X)^{-1}X')\hat{\varepsilon}}{\hat{\varepsilon}'\hat{\varepsilon}/T}, \end{aligned} \tag{6.25}$$

where $\hat{\varepsilon} = (\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_T)'$ contains residuals estimated under the null hypothesis of linearity, X is a $T \times (p + 1 + m)$ matrix with t-th row given by $(x_t', q_t')'$, and I_T is the identity matrix of size T. The LM statistic has an asymptotic $\chi^2(m)$ distribution under the null hypothesis. In small samples it is usually recommended to use an F version of the test. This version of the test can be computed as follows:

1. Estimate the model under the null hypothesis of linearity by regressing y_t on x_t. Compute the residuals $\hat{\varepsilon}_t$ and the sum of squared residuals $\text{SSR}_0 = \sum_{t=1}^{T} \hat{\varepsilon}_t^2$.

2. Perform the auxiliary regression of $\hat{\varepsilon}_t$ on x_t and q_t, and compute the sum of squared residuals from this regression, SSR_1.

3. The LM-type test statistic can now be computed as

$$\text{LM} = \frac{(\text{SSR}_0 - \text{SSR}_1)/m}{\text{SSR}_1/(T-m-p-1)}, \tag{6.26}$$

which is approximately F distributed under the null hypothesis with m and $T-m-p-1$ degrees of freedom, respectively.

6.3.1 Robust tests for smooth transition nonlinearity

Because of the properties of the OLS estimator in the presence of AOs as discussed in section 6.1, one expects that the LM-type tests can be severely affected by additive outliers. In the next section, I formally show that this is indeed the case. The robust estimators discussed in section 6.2 can be used to construct robust versions of these test statistics. In particular, a robust test can be obtained by using a robust estimator to estimate the model under the null hypothesis. Hampel *et al.* (1986), Peracchi (1991) and Markatou and He (1994) show that the robustness properties of estimators carry over to test statistics based on such estimators. Moreover, under conventional assumptions, the test statistics retain their standard limiting χ^2 distributions. Thus, it might be expected that if an M or GM estimator is used for constructing LM-type test statistics, the resulting statistics are protected against the influence of (additive) outliers.

By interpreting the first order condition for the AR model given in (6.11) as a pseudo-score, an LM test based on the GM estimator can easily be constructed. Let $\widehat{W}_x = \text{diag}(\hat{w}_x(x_1), \ldots, \hat{w}_x(x_T))'$ and $\widehat{\Psi} = (\hat{\psi}(r_1), \ldots, \hat{\psi}(r_T))'$, where $\widehat{W}_x$ and $\widehat{\Psi}$ are computed under the null hypothesis, and where r_t again denotes the t-th standardized residual, $r_t \equiv (y_t - \phi' x_t)/(\sigma_\varepsilon w_x(x_t))$. The robust version of the LM-test statistics to test $H_0 : \theta = 0$ in (6.24) can be computed as

$$\begin{aligned} \text{LM} &= \frac{\widehat{\Psi}'\widehat{W}_x X (X'\widehat{W}_x\widehat{W}_x X)^{-1} X'\widehat{W}_x \widehat{\Psi}}{\widehat{\Psi}'\widehat{\Psi}/T} \\ &= \frac{\widehat{\Psi}'\widehat{\Psi} - \widehat{\Psi}'(I - \widehat{W}_x X (X'\widehat{W}_x\widehat{W}_x X)^{-1} X'\widehat{W}_x)\widehat{\Psi}}{\widehat{\Psi}'\widehat{\Psi}/T}. \end{aligned} \tag{6.27}$$

Because $\psi(r_t) = w_r(r_t)r_t$, the term $\widehat{\Psi}'\widehat{\Psi} = \sum_{t=1}^{T}(\hat{w}_r(\hat{r}_t)\hat{\varepsilon}_t/\hat{\sigma}_\varepsilon \hat{w}_x(x_t))^2$ can be interpreted as a sum of squared weighted residuals under the null hypothesis, where the weights decline for large standardized residuals. As (6.27) shows, the F version of the test, corresponding to (6.26), can be computed by running an auxiliary OLS regression of the weighted residuals $\hat{\psi}(\hat{r}_t)$ on the weighted regressors $\hat{w}_x(x_t)x_t$ and

$\hat{w}_x(x_t)q_t$. By estimating the linear model under the null hypothesis with an M estimator and setting $w_x(\cdot) = 1$ in the above, one obviously obtains an LM test based on M estimators.

In the rest of this chapter, the LM-type tests based on OLS, M and GM estimators will be denoted as LM_i^{LS}, LM_i^{M}, and LM_i^{GM}, respectively, where $i = 1, 2, 3$. Recall from Section 2.2 that the subscripts 1 and 2 indicate the tests against LSTAR and ESTAR alternatives, respectively, based on a first–order Taylor approximation of the relevant transition function. The LM_3 statistic also tests against an LSTAR alternative, but is based on a third-order Taylor approximation of (6.21). The parsimonious version of the LM_3 statistic will be denoted as $\text{LM}_3^{e,j}$, with j =LS, M or GM.

6.4 Outliers and tests for STAR nonlinearity

In this section I formally investigate the effect of AOs on the LM-type test statistics. I restrict attention to the simple LM_1 statistic when used to test an AR(1) model against an LSTAR(1) alternative, applied under symmetric isolated AO contamination. Hence, the set-up is as follows. The observed time series y_t can be described as

$$z_t = \phi_1 z_{t-1} + \eta_t, \qquad t = 1, \ldots, T, \tag{6.28}$$

$$y_t = z_t + \zeta\delta_t, \tag{6.29}$$

where $|\phi_1| < 1$, η_t is a white noise process, and ζ is a non-zero constant. The process δ_t is assumed to be i.i.d., such that $P(\delta_t = 1) = P(\delta_t = -1) = \pi/2$ and $P(\delta_t = 0) = 1 - \pi$ for $0 < \pi < 1$, that is, positive and negative AOs occur with equal probability. The LM_1 statistic against the alternative of a STAR model with $s_t = y_{t-1}$ makes use of the auxiliary regression

$$\hat{\varepsilon}_t = \beta_1 y_{t-1} + \theta_1 y_{t-1}^2 + \nu_t, \tag{6.30}$$

where $\hat{\varepsilon}_t$ are residuals from an AR(1) model estimated for y_t.

The qualitative results derived below remain the same for higher order models and for the other LM-type tests discussed in Section 6.3. I show that the presence of AOs leads to higher rejection rates for both the robust and nonrobust tests. The distortion for the nonrobust test, however, is much larger. If the outliers become extremely large or if the fraction of contamination becomes relatively high, the level of the nonrobust test is recovered, but the power of the test drops to its size. The power of the GM-based test, in contrast, is significantly higher.

The remainder of this section is split in three parts. In Sections 6.4.1 through 6.4.3, I discuss the effect of AOs on the LM-type tests based on OLS, M, and GM estimators, respectively. In each of these sections, both a global and a local (non)robustness result for the tests are derived. The global result states that in the presence of AOs, the test statistics retain their asymptotic χ^2 distributions, only multiplied by a constant of proportionality. This constant of proportionality is a function of the autoregressive parameter (ϕ_1), the probability of occurrence of AOs (π), and the (absolute) magnitude of the outliers (ζ). The second result in each subsection is a local robustness result. It describes the behaviour of the LM_1 test for $\pi \downarrow 0$, that is, for (infinitesimally) small fractions of outliers. In this way, the local result can be compared to the derivation of an influence curve, see Hampel *et al.* (1986) for influence functions of estimators in the regression context, Martin and Yohai (1986) for influence functions of estimators in the time series context, and Peracchi (1991) and Markatou and He (1994) for influence functions of test statistics in the regression context. The local robustness results complement the results in the aforementioned articles by presenting the influence of infinitesimally small fractions of contamination on the distribution of a test statistic in the time series context. All proofs are gathered in the appendix to this chapter.

6.4.1 OLS-based tests

To start off the discussion on the robustness properties of the LM-type tests for STAR nonlinearity, I first consider the effect of outliers on the OLS-based testing procedure. The main result is summarized in the following theorem.

Theorem 1a (OLS) *Consider the AR(1) model without a constant under symmetric additive outlier contamination with standard Gaussian innovations η_t as given in* (6.28)-(6.29). *If the parameters of the model are estimated using OLS,*

$$\lim_{T\to\infty} LM_1^{LS} \xrightarrow{d} c_G \cdot \chi_1^2, \tag{6.31}$$

where χ_1^2 denotes a random variate with a chi-squared distribution with one degree of freedom, and where c_G denotes a fixed constant.

Furthermore, for sufficiently small fractions of contamination π,

$$\lim_{T\to\infty} LM_1^{LS} \xrightarrow{d} (1 + c_L \cdot \pi + O(\pi^2)) \cdot \chi_1^2 \ , \tag{6.32}$$

with

$$c_L = \frac{1}{3\sigma_z^4}\left(\phi_1^2\zeta^6 + (1+6\phi_1^2\sigma_z^2)\zeta^4 + (3\phi_1^2\sigma_z^4 + 6\sigma_z^2 + 3\sigma_z^4)\zeta^2\right) + \frac{-1}{3\sigma_z^4}\zeta^2(\zeta^2 + (7+\phi_1^2)\sigma_z^2) + \frac{2\phi_1\zeta^2}{3}(6\phi_1^2 - \zeta^2 - 3\phi_1^2\sigma_x^2), \tag{6.33}$$

and $\sigma_z^2 = (1-\phi_1^2)^{-1}$. *As a result,* $c_L = (1-\phi_1^2)^2\phi_1^2\zeta^6/3 + O(\zeta^4)$.

The value of the constant c_G in (6.31) is crucial for assessing the effect of additive outliers on the level of the LM_1^{LS} test. If $c_G = 1$, the test has the correct size. If $c_G > 1$, however, the test will be oversized, whereas the reverse holds if $c_G < 1$. The constant c_G in fact is a function of the parameters of the model, $c_G = c_G(\phi_1, \zeta, \pi)$. For the OLS estimator, an analytic expression for the function $c_G(\phi_1, \zeta, \pi)$ can be obtained, although it is rather lengthy and does not allow an immediate understanding of the way it depends on its arguments. For that reason, instead of presenting the expression in full detail, some plots of c_G are shown in Figure 6.3 for $\phi_1 \in [0, 0.9]$, $\zeta \in [0, 20]$, and $\pi = 0.01, 0.05, 0.10$ and 0.25.

From these graphs it can be observed that the function $c_G(\phi_1, \zeta, \pi)$ satisfies

(i) $c_G(\phi_1, 0, \pi) = 1$ for all combinations of ϕ_1 and π;

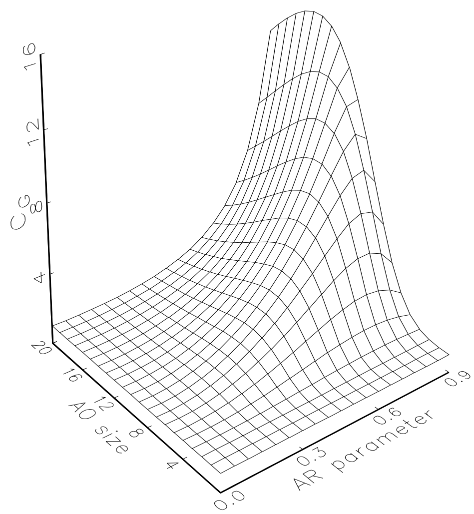
(ii) $\partial c_G/\partial\zeta > 0$ for $\zeta \in (0, q(\phi_1, \pi)]$, with q some function of ϕ_1 and π;

(iii) $\partial c_G/\partial\zeta < 0$ for $\zeta \in (q(\phi_1, \pi), \infty)$;

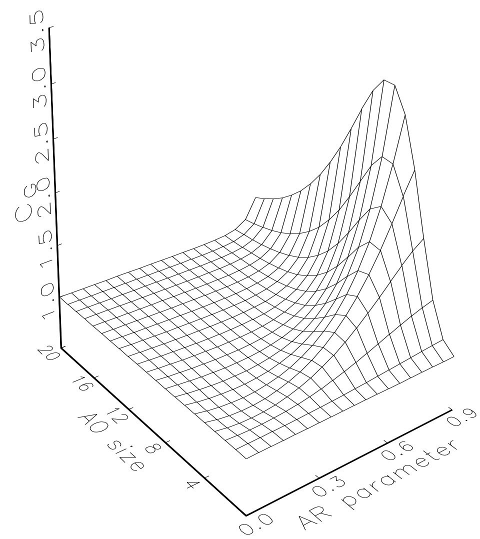
(iv) $c_G(\phi_1, \zeta, \pi) \to 1$ as $\zeta \to \infty$.

Property (i) is obvious: if there are no outliers, the LM_1^{LS} test has the correct size. Property (ii) states that additive outliers bias the OLS test results towards the detection of nonlinearity. The bias is increasing in the magnitude of the outliers ζ up to a certain threshold magnitude q, which depends on ϕ_1 and π. If ζ gets larger than $q(\phi_1, \pi)$, c_G starts to decrease again until it reaches 1 for infinitely large outliers, see properties (iii) and (iv). The fact that the size of the LM_1^{LS} test increases if AOs are present is intuitively clear. The dynamics of the time series surrounding the periods the outliers occur can be regarded as a different regime. Therefore, the LM_1^{LS} test is likely to mistake the outliers for regime switches. Properties (iii) and (iv) are less intuitive at first sight. They reveal that the size distortion of the LM_1^{LS} test becomes smaller and ultimately disappears if the outliers become extremely large. This follows from the nature of the additive outlier contamination, which, in fact, is a very specific form of white noise. As the model is estimated under the null, the OLS estimator cannot distinguish the original time series from white noise (that is, the contamination process) if the outliers are very large, see Hampel *et al.* (1986) or Hoek, Lucas and van Dijk (1995), among others. Moreover, due to the specific form of white noise contamination, the regressors y_{t-1} and y_{t-1}^2 are highly collinear for large outliers. As a result, it is unlikely that y_{t-1}^2 will be significant as an additional regressor to y_{t-1} in this case. This is exactly what is tested by the LM_1^{LS} test, see (6.30). These arguments explain why c_G tends to 1 for large values of ζ. Note, however, that a similar reasoning can be applied if the alternative hypothesis is in fact true. In that case the distribution of the LM_1^{LS} test also tends to the null distribution, that is, $c_G = 1$ for large outliers. This means that the small size distortion of the OLS procedure for large outliers are paid for in terms of a severe power loss of the test. This is supported by the simulations in the next section. Also note that similar results hold if ζ is fixed and the fraction of contamination π is increased, see Figure 6.3.

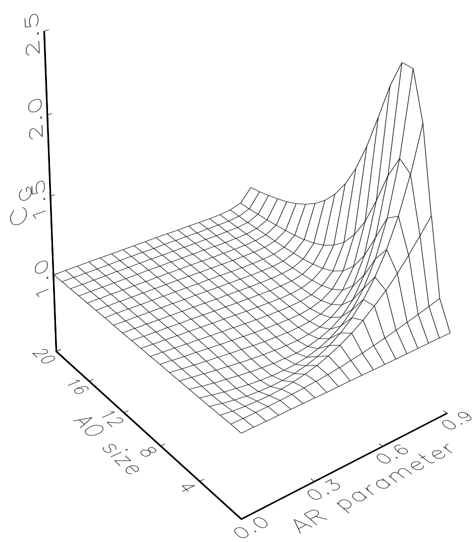
Before proceeding to the effect of AOs on nonlinearity tests based on M and GM estimators, two final remarks on the result in Theorem 1a should be made. First, in practice one usually includes a constant in the linear model under the null hypothesis. The first part of Theorem 1a remains valid in that case, albeit with a different proportionality constant c_G, which is more difficult to handle analytically. We therefore investigate it by means of simulation. For several combinations of



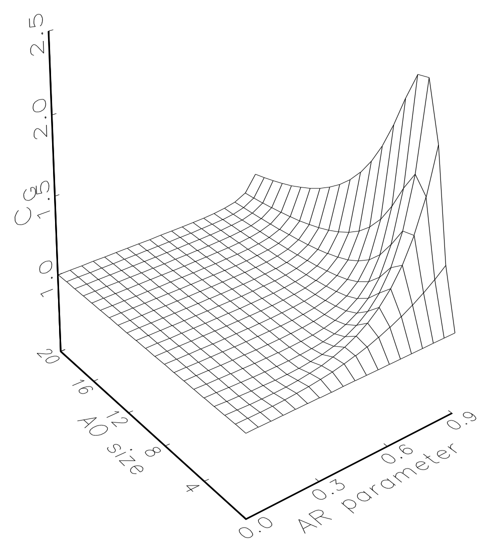
(a) $\pi = 0.01$



(b) $\pi = 0.05$



(c) $\pi = 0.10$



(d) $\pi = 0.25$

Figure 6.3: Values of the constant c_G in Theorem 1a for the OLS-based $\mathrm{LM}_1^{\mathrm{LS}}$ test statistic.

(ϕ_1, ζ, π), 1000 series of 100 observations are generated from a contaminated AR(1) model. For each series the LM_1^{LS} test statistic is computed including a constant in the regression model. Next, the test outcomes are regressed on the corresponding percentiles of the χ_1^2 distribution to obtain an estimate of c_G. These estimates are shown in the upper graphs of Figure 6.4, where a bivariate kernel is used to obtain a relatively smooth surface. Comparing these graphs with Figure 6.3, the main difference seems to be that, for fixed ϕ_1 and π, c_G attains its maximum for higher values of ζ, while the return to 1 after this point proceeds much slower. Moreover, for, e.g., $\pi = 0.05$, the maximum value of c_G in Figure 6.4 is much higher than the corresponding value of c_G in Figure 6.3.

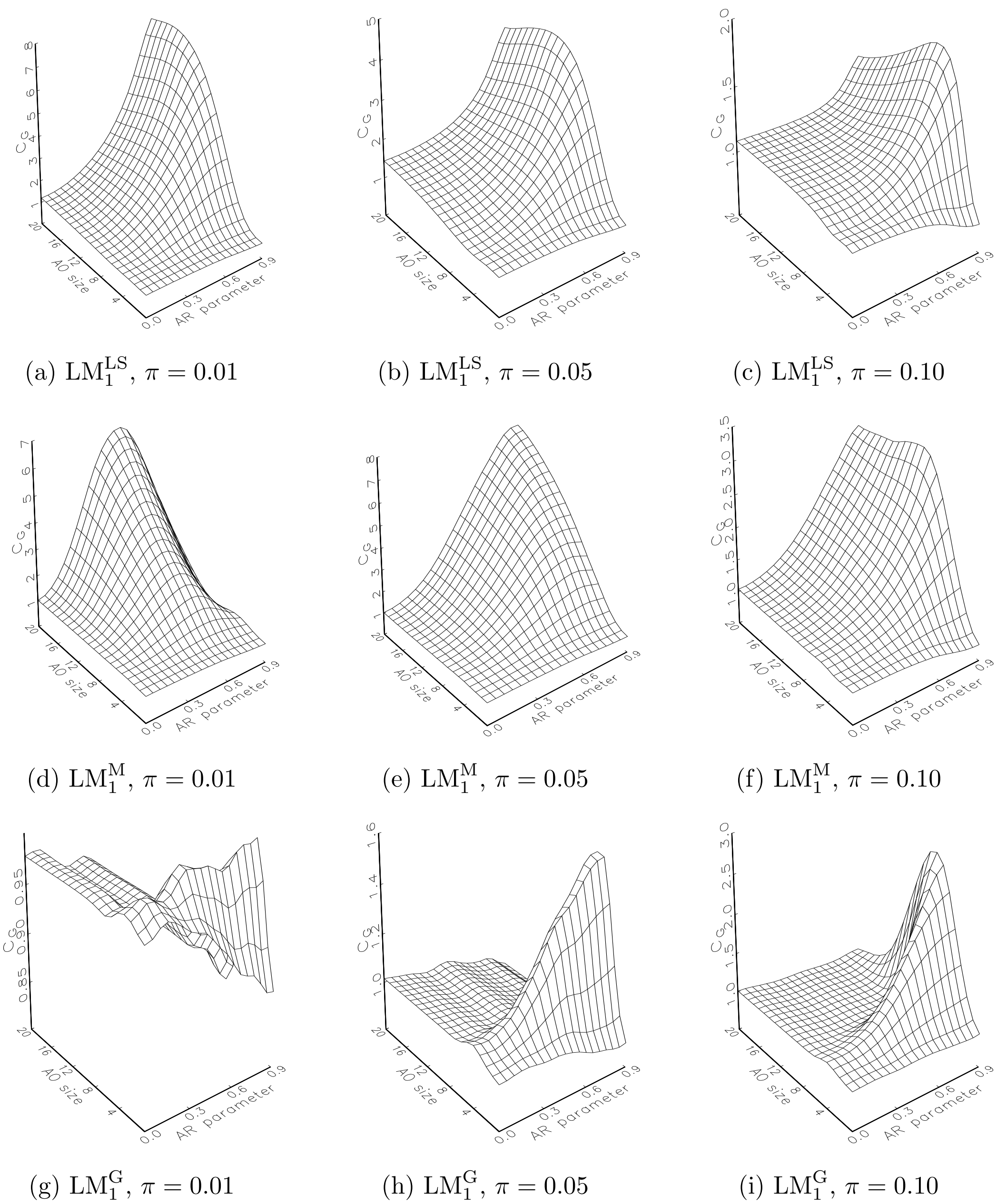
A second remark on Theorem 1a concerns the type of covariance matrix used in computing the LM_1^{LS} test. So far, I have used an ordinary covariance matrix estimator, $(\hat{\varepsilon}'\hat{\varepsilon}/T)(X'X)^{-1}$ to compute the test. In that case, the constant c_L in Theorem 1a is an unbounded function of ζ. The function increases to *plus* infinity for $\zeta \to \pm\infty$. Put differently, for local contamination the OLS-based test has a level above the nominal level. In contrast, one can construct a heteroskedasticity consistent (HCC) test as discussed at the end of Section 2.2. This effectively boils down to using a HCC covariance matrix estimator $(X'X)^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t^2 x_t x_t'(X'X)^{-1}$, see White (1980) and Hsieh (1983). In that case, the specification of k_2 in the proof of Theorem 1a becomes $k_2 = E(y_{t-1}^4\varepsilon_t^2)$. Consequently, (6A.1) in the appendix reduces to 1 and is independent of π. The first two lines in the expression for c_L in (6.33) then vanish and the dominant term in the expression for c_L for the OLS-based test becomes $-2\phi_1\zeta^4/3$. This is an unbounded function of ζ that tends to *minus* infinity for $\zeta \to \pm\infty$. In other words, the OLS-based test with a HCC covariance matrix estimator has a level below the nominal level under local contamination. Also, the order of c_L as a function of ζ is smaller if a HCC covariance matrix is used, which signifies that the use of such a covariance matrix estimator alleviates part of the nonrobustness of the OLS-based LM_1 test, compare Lucas (1995).

6.4.2 Tests based on M estimators

In the previous subsection I demonstrated the nonrobustness of the OLS-based LM_1^{LS} test using statistical robustness analysis. As explained in Section 6.2, the first way to tackle the nonrobustness of OLS is to use an M estimator. In the time series context, the class of M estimators is statistically robust to innovation outliers, but not necessarily to additive outliers, see Martin and Yohai (1986) and Hoek *et al.* (1995). The M estimator is defined by (6.11) with $w_x(\cdot) \equiv 1$. The analogue of Theorem 1a for M estimators is given below.

Theorem 1b (M estimators) *Given the setting of Theorem 1a with the OLS estimator replaced by an M estimator based on an anti-symmetric function $\psi(\cdot)$,* (6.31) *and* (6.32) *continue to hold for the LM_1^M test statistic, albeit with different values for the constants c_G and c_L.*

Again, an analogous result can be established if a constant is incorporated in the regressions. The middle panels of Figure 6.4 present estimates of the constant c_G



(a) $\mathrm{LM}_1^{\mathrm{LS}}$, $\pi = 0.01$ (b) $\mathrm{LM}_1^{\mathrm{LS}}$, $\pi = 0.05$ (c) $\mathrm{LM}_1^{\mathrm{LS}}$, $\pi = 0.10$

(d) $\mathrm{LM}_1^{\mathrm{M}}$, $\pi = 0.01$ (e) $\mathrm{LM}_1^{\mathrm{M}}$, $\pi = 0.05$ (f) $\mathrm{LM}_1^{\mathrm{M}}$, $\pi = 0.10$

(g) $\mathrm{LM}_1^{\mathrm{G}}$, $\pi = 0.01$ (h) $\mathrm{LM}_1^{\mathrm{G}}$, $\pi = 0.05$ (i) $\mathrm{LM}_1^{\mathrm{G}}$, $\pi = 0.10$

Figure 6.4: Values of the constant c_G in Theorem 1a for $\mathrm{LM}_1^{\mathrm{LS}}$ (panels (a)-(c)), $\mathrm{LM}_1^{\mathrm{M}}$ (panels (d)-(f)) and $\mathrm{LM}_1^{\mathrm{GM}}$ (panels (g)-(i)) test statistics if a constant is included in the estimation of the linear model under the null hypothesis. The figure is based on 1000 replications of a contaminated AR(1) model, $T = 100$.

for the M estimator based on simulations as in the previous subsection. I use the same weighting function as for the GM estimator described in the next subsection, with the weights for the regressors $w_x(\cdot)$ set to 1. It is seen in the figure that the behaviour of c_G crucially depends on the fraction of contamination π. For $\pi = 0.01$, c_G attains a maximum at $\phi_1 \approx 0.5$ and quickly reverts to 1 for higher values of the AR-parameter. For $\pi = 0.05$ and 0.10, c_G is increasing in ϕ_1. For $\pi = 0.01$ and 0.05 and fixed values of ϕ_1, c_G is strictly increasing in the absolute magnitude of the outliers over the range of ζ considered here. For $\pi = 0.10$, c_G levels off for $\zeta > 10$. Also note that for $\pi = 0.05$ and 0.10 the maximum of c_G is larger than the corresponding maximum for the LM_1^{LS} statistic.

6.4.3 Tests based on GM estimators

To conclude this section, I discuss the theoretical robustness properties of the LM-type tests based on GM estimators. The main result for GM estimators is stated in the following theorem.

Theorem 1c (GM estimators) *Given the setting of Theorem 1a with the OLS estimator replaced by a GM estimator as proposed in Section 6.3 with anti-symmetric function $\psi(\cdot)$ and symmetric function $w_x(\cdot)$,* (6.31) *and* (6.32) *continue to hold for the LM_1^{GM} test statistic, albeit with different values for the constants c_G and c_L.*

I use the same simulation methodology as in the previous subsections to obtain estimates of the constant c_G for the empirically relevant case in which the regression contains a constant term. The results for c_G, set out in the lower graphs of Figure 6.4, are remarkably different from the corresponding estimates for the OLS-based test. It is easily seen that, at least for $\pi = 0.05$, the maximum value of c_G is much lower for the robust test, while the decrease following this maximum is much faster as well. In fact, for large values of ζ and ϕ_1, the constant drops below 1, resulting in the test statistic being slightly biased toward the null. Notably, for the robust test c_G is larger for $\pi = 0.10$ than for $\pi = 0.05$. This contrasts to the findings for the OLS-based test. It will moreover appear from the simulations in the next section that the drop of c_G towards 1 for large ζ does not signal a loss in power of the test. This stands in sharp contrast to the result for the OLS-based test.

To conclude this section, some comments on the (dis-)advantages of M estimators versus GM estimators for constructing the LM_1 test are in order. As mentioned in Section 6.2, additive outliers in an autoregressive time series context are most adequately dealt with using GM estimators, see Martin (1981). The weights for the regressors employed by the GM estimator help to downweight bad leverage points caused by such outliers. Note that sometimes M estimators suffice to obtain robustness and GM estimators are not strictly necessary, see Hoek *et al.* (1995). Still, the impact of the outliers may in such cases be larger if M estimators are used instead of GM estimators. Therefore, if AOs are to be expected, it seems preferable to use GM estimators. For innovation outliers, however, the opposite holds. In that case the observations following the outlier give a clear signal of the dynamical pattern of the uncontaminated series. This signal is weakened if one downweights the leveraged observations in an autoregression, as one does with the GM estimator. Therefore, one can expect the GM estimator to be less efficient than the M estimator if only innovation outliers exist. This, however, is a well-known trade-off. If innovations are Gaussian and if there are no outliers, it is most efficient to use OLS instead of an M or GM estimator. Similarly, if there are only innovation outliers, it is most efficient to use the M estimator. This efficiency loss is the price one pays for being protected against the adverse effects of outliers, see also Section 6.2.

6.5 Monte Carlo experiments

In this section I evaluate the small sample performance of the standard and robust LM-type tests by means of Monte Carlo simulations. I focus on two effects that arise due to robustifying the test statistics. First, I consider the consequences of using a robust test in a setting in which robustness is not required - that is, in a setting without outliers. Second, I investigate the (relative) performance of the tests in the presence of additive and innovation outliers.

6.5.1 Monte Carlo design

The size of the LM-type tests is analyzed using series for which the core process z_t is generated from the AR(1) model (6.7) where $\eta_t \sim$ i.i.d. $N(0, \sigma_\eta^2)$, $\sigma_\eta = 1$. Contaminated series y_t are obtained by adding either AOs to z_t according to (6.3) or IOs according to (6.4). I consider symmetric contamination as described below (6.29), that is, the variable δ_t takes the values -1, 0 and 1 with probability $\pi/2$, $1-\pi$ and $\pi/2$, respectively, which ensures that positive and negative outliers occur with equal probability.

To consider the power properties of the tests, the AR(1) process for z_t is replaced by a first-order STAR model as in LST,

$$z_t = \phi_{1,1} z_{t-1}(1 - G(z_{t-1}; \gamma, c)) + \phi_{2,1} z_{t-1} G(z_{t-1}; \gamma, c) + \eta_t, \tag{6.34}$$

with transition function taken equal to

$$G(z_{t-1}; \gamma, c) = (1 + \exp\{-\gamma(z_{t-1} - c)\})^{-1}, \qquad \gamma > 0, \tag{6.35}$$

with $\gamma = 2.5$ and $c = 0$ or

$$G(z_{t-1}; \gamma, c) = (1 + \exp\{-\gamma(z_{t-1} - c_1)(z_{t-1} - c_2)\})^{-1}, \qquad \gamma > 0, \tag{6.36}$$

with $\gamma = 10$, $c_1 = -1$, and $c_2 = 1$. Model (6.34) with (6.35) represents an LSTAR model with a gradual transition between the two regimes, while model (6.34) with (6.36) renders an ESTAR model with a relatively swift transition between the three regimes at the threshold values c_1 and c_2.

In both the size and power experiments described below, 1000 replications of $T = 150$ observations are used. The choice for this particular sample size is motivated by the length of the empirical time series considered in Section 6.6. The necessary starting value z_0 is set equal to zero throughout. In order to eliminate possible dependencies of the results on this initial condition, series of $T + 100$ observations are generated and the first 100 observations are discarded. In all experiments, the probability of occurrence of outliers π is fixed at 0.05, while the absolute magnitude of the outliers is varied among $\zeta \in \{3, 5, 7\}$. The standard and robust LM-type tests are applied to both the clean and the contaminated series z_t and y_t, respectively, to get a precise measure of the influence of outliers on the performance of the tests. In the experiments, I consider the effects of varying the autoregressive parameters in the AR and STAR models and the magnitude of the outliers. The order of the AR(p) model under the null hypothesis is fixed at $p = 1$. In practice, one has to decide upon the order p by means of, for example, the AIC or SIC, or a procedure based on the (partial) autocorrelations of the observed time series, see Section 2.1.2. The reason for fixing the AR order at its true value here is that many of these order determination procedures also are affected by outliers, see Deutsch, Richards and Swain (1990) and Ronchetti (1997), among others. Application of such a procedure therefore would interfere with the effect of outliers on the LM-type test statistics.

Finally, in the size experiments and the power experiments involving the LSTAR model (6.34) with (6.35), I only consider the properties of the LM_1 statistic, whereas in the experiments involving the ESTAR alternative (6.34) with (6.36) the properties of the LM_2 test are examined. In all cases, F-variants of the tests are used, as given in (6.25) and (6.27). Results for other test statistics, as well as results for other contamination fractions, other sample sizes, and DGPs other than the ones described below are qualitatively similar.

6.5.2 Size

Table 6.1 shows rejection frequencies of the null hypothesis for the three variants of the LM_1 test statistic using 5% critical values, for series generated according to an AR(1) model with autoregressive parameter ϕ_1 as given in (6.7). Outliers of magnitude $\zeta = 3$, 5, and 7 are added to the model with probability $\pi = 0.05$. In addition, the columns headed $\zeta = 0$ show estimates of the size when the test is applied to the series without outliers. The upper and lower panels of the table present results for AO and IO contamination, respectively.

It is seen that for the clean series z_t and for series y_t which contain IOs the rejection frequencies of both the standard and the robust tests approximate the 5% significance level quite well, except for the LM_1^{LS} and LM_1^{M} tests for large values of ϕ_1. In the presence of AOs, however, marked differences appear. For the standard test, the rejection frequencies increase to rather high levels for large ζ and large values of ϕ_1. For the test based on the M estimator, the rejection frequencies become even higher for $\zeta = 5$ or 7. The distortions in the level of the GM-based test are much smaller, with rejection frequencies typically below 10%.

To demonstrate that these results are not specific for the 5% significance level,

Table 6.1: Size of LM_1 test for STAR nonlinearity

			LS				M				GM			
	ϕ_1	ζ	0	3	5	7	0	3	5	7	0	3	5	7
AO	0.0		4.7	5.9	5.5	6.3	5.4	5.6	5.6	5.6	5.0	5.1	5.9	6.0
	0.1		5.0	6.5	6.0	6.2	5.0	6.2	6.0	6.2	5.2	6.1	5.8	5.8
	0.3		4.6	6.8	7.7	7.4	4.8	6.6	8.9	9.7	5.0	6.6	8.0	6.8
	0.5		4.9	9.2	11.3	11.2	5.0	8.4	14.5	17.0	5.3	9.4	9.8	7.1
	0.7		4.0	12.6	18.1	16.8	4.2	12.1	22.5	25.8	4.0	11.0	12.4	4.5
	0.9		2.2	12.5	24.1	28.1	2.3	10.8	17.9	24.8	4.4	11.9	12.1	6.5
IO	0.0		4.7	4.1	4.8	5.3	5.4	4.9	5.0	5.5	5.0	6.3	5.8	5.9
	0.1		5.0	4.0	4.6	4.9	5.0	4.8	4.7	5.5	5.2	5.7	5.9	5.1
	0.3		4.6	3.6	3.3	4.2	4.8	4.4	4.4	4.5	5.0	5.8	5.5	5.6
	0.5		4.9	4.0	3.0	4.1	5.0	4.7	5.3	5.0	5.3	6.0	6.2	5.5
	0.7		4.0	2.8	2.7	3.4	4.2	4.2	5.0	5.1	4.0	5.2	5.3	6.0
	0.9		2.2	2.4	2.3	3.3	2.3	2.9	4.2	5.2	4.4	4.9	6.0	5.9

Rejection frequencies for F-versions of the $\mathrm{LM}_1^{\mathrm{LS}}$ test based on (6.25) and $\mathrm{LM}_1^{\mathrm{M}}$ and $\mathrm{LM}_1^{\mathrm{GM}}$ tests based on (6.27) at 5% significance level for series generated by the AO model (6.3) or the IO model (6.4), where the core process follows an AR(1) process (6.7) with $\sigma_\eta^2 = 1$. Outliers occur with probability $\pi = 0.05$. The table is based on 1000 replications for sample size $T = 150$.

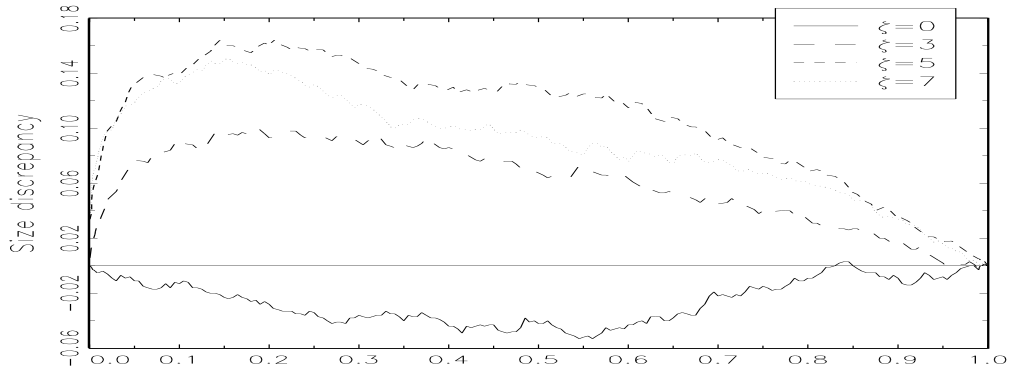
Figure 6.5 shows p-value discrepancy plots for the various LM_1 tests for $\phi_1 = 0.7$. These plots, advocated by Davidson and MacKinnon (1998), depict the difference between the actual and nominal size of the tests versus the nominal size[3].

It is seen that the actual size of the standard and robust tests is always larger than the nominal size in case $\zeta > 0$ (with the exception of the size of $\mathrm{LM}_1^{\mathrm{GM}}$ in case $\zeta = 7$). Comparing the scales on the vertical axis in the different panels of the figure shows that the size distortions are comparable for $\zeta = 3$, but for larger values of ζ the overrejection is most severe for the $\mathrm{LM}_1^{\mathrm{M}}$ test, followed by the $\mathrm{LM}_1^{\mathrm{LS}}$ and $\mathrm{LM}_1^{\mathrm{GM}}$ statistics. Results for other values of $\phi_1 (\geq 0.3)$ are qualitatively similar.

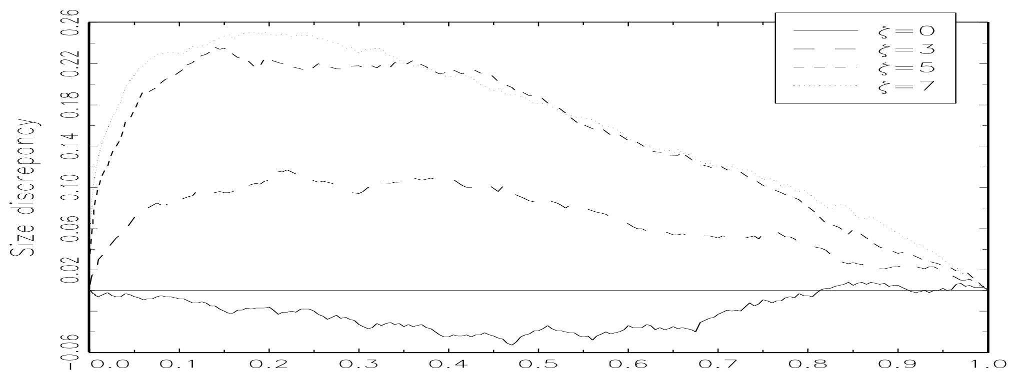
6.5.3 Power

Table 6.2 shows rejection frequencies for series generated by the LSTAR model (6.34) with (6.35) for various combinations of $\phi_{1,1}$ and $\phi_{2,1}$, which have been taken from LST. The probability and magnitudes of outliers are the same as above. Besides rejection frequencies using the 5% critical values from the F-distribution, Table 6.2 also contains size-corrected rejection frequencies, which have been obtained from

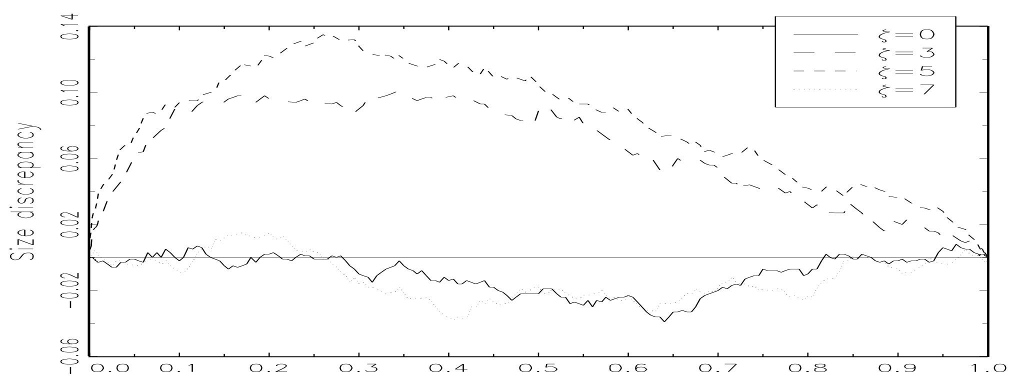
[3]To be more precise, p value discrepancy plots are constructed as follows. The N replications in the Monte Carlo experiments render p-values $p_1, \ldots, p_N$, where in our case $N = 1000$. The empirical distribution function of the p-values can be estimated by simply calculating $\hat{F}(x) = \frac{1}{N}\sum_{j=1}^{N} I[p_j \leq x]$, for any point x in the (0,1) interval. The function $\hat{F}(x)$ gives the actual rejection frequency of the test at nominal significance level x. If the distribution used to calculate the p-values p_j is correct, each of the p_j should be distributed as uniform (0,1), and $\hat{F}(x) \approx x$. By calculating $\hat{F}$ on a grid of points $x_1, \ldots, x_M$ on the (0,1) interval and plotting $\hat{F}(x_i) - x_i$ against x_i one can easily infer if the test statistic is under- or oversized at different nominal significance levels. Moreover, it allows easy comparison between different test statistics.



(a) $\mathrm{LM}_1^{\mathrm{LS}}$



(b) $\mathrm{LM}_1^{\mathrm{M}}$



(c) $\mathrm{LM}_1^{\mathrm{GM}}$

Figure 6.5: p-value discrepancy plots for the LM_1 test for STAR nonlinearity based on OLS, M- and GM-estimators. The plots are based on 1000 replications generated by (6.3) and (6.7) with $\phi_1 = 0.7$, $\sigma_\eta^2 = 1$. Additive outliers of magnitude plus or minus ζ are added with probability 0.05.

the following procedure. For each combination of $\phi_{1,1}$ and $\phi_{2,1}$, 1000 series of length $T = 150$ are generated according to model (6.34)-(6.35). For each of these series, an AR(1) model is estimated. The AR(1) model with autoregressive parameter equal to the mean of these estimates is regarded as pseudo-null model, in the sense that it is the best linear approximation (of the same lag order) to the LSTAR(1) model under consideration. This AR(1) model is then used to generate finite sample critical values for the LM_1 test.

The columns headed $\zeta = 0$ again show estimates of the power of the tests applied to the uncontaminated series. As expected, the power of the tests based on M and GM estimators is slightly lower than that of the nonrobust test, as the robust test downweights observations that are not outliers. The maximum difference, however, is below 5%, which is quite encouraging.

The results when the tests are applied to series with AO contamination as given in the upper part of Table 6.2 show that it pays off to use the test based on the GM estimator. The power of the OLS-based test decreases dramatically in the presence of AOs, and approaches its size for large values of ζ. The same appears to occur for the LM_1^{M} test. For small AOs, the GM-based test also loses some power, although the drop is by far smaller than for the standard and M-based tests. For large AOs, which obviously are easier to recognize, the power of the LM_1^{GM} test is hardly affected. The lower part of Table 6.2 contains rejection frequencies for the LM_1 test when applied to series with IO contamination. It is seen that IOs have quite different effects than AOs, in that the power of all tests increases as IOs are present. The increase becomes larger for larger values of ζ.

Table 6.3 shows rejection frequencies for the LM_2 test statistic when applied to series generated by the ESTAR model (6.34) with (6.36) for the same combinations of $\phi_{1,1}$ and $\phi_{2,1}$. I consider the LM_2 instead of the LM_1 statistic here because it was derived explicitly as the test against this alternative[4]. Again, size-corrected rejection frequencies are reported as well, which have been obtained according to the procedure described above.

First, notice that the entries in the column headed $\zeta = 0$ suggest that the LM_2 test is not particularly powerful against the alternatives considered here. Second, when AOs are present, it appears at first sight that for certain combinations of $\phi_{1,1}$ and $\phi_{2,1}$ the rejection frequencies of the tests *increase*. The size-corrected rejection frequencies reveal that this does not imply that the power of the tests increases. The power of the LM_2^{LS} and LM_2^{M} statistics in fact drops to the size of 5%. For the LM_2^{GM} test, size-corrected rejection frequencies also decrease for AOs of moderate magnitude. As ζ gets larger the AOs become more apparent and are easier detected, and the rejection frequencies of the GM-based test return to the same level as for clean series ($\zeta = 0$). For series with IO contamination, the power of the OLS- and M-based tests drops considerably. The power of the GM-based test statistic decreases as well, albeit to a much lesser extent.

[4] Accordingly, the LM_1 statistic has almost no power against ESTAR alternatives. Also note that for the ESTAR(1) case considered here, the LM_2 test is in fact identical to the LM_3 statistic as the auxiliary regressions for both tests contain y_{t-1}^2 and y_{t-1}^3 as additional regressors.

Table 6.2: Power of LM_1 test for STAR nonlinearity

AO			LS				M				GM			
$\phi_{1,1}$	$\phi_{2,1}$	ζ	0	3	5	7	0	3	5	7	0	3	5	7
−0.5	−0.9		53.3	25.1	19.3	16.3	51.7	27.5	25.5	29.1	49.0	25.1	38.5	45.1
			[55.3]	[11.5]	[7.2]	[5.1]	[54.4]	[15.2]	[7.7]	[5.9]	[48.4]	[18.2]	[27.8]	[42.0]
	0.0		43.7	14.4	8.5	8.3	42.5	15.0	10.5	8.5	42.9	26.2	40.4	43.2
			[44.6]	[10.0]	[6.4]	[5.3]	[43.0]	[11.4]	[6.2]	[6.0]	[42.7]	[23.1]	[37.5]	[39.2]
	0.4		86.9	32.8	8.6	7.0	85.3	33.7	12.0	6.3	84.2	54.8	80.0	82.3
			[87.9]	[27.4]	[6.9]	[4.9]	[86.6]	[27.4]	[9.9]	[5.2]	[83.9]	[51.3]	[77.5]	[79.8]
	0.9		91.0	66.6	34.8	22.6	85.4	63.1	43.4	35.5	87.2	73.5	76.9	84.1
			[93.0]	[45.5]	[10.6]	[4.7]	[88.7]	[46.5]	[18.0]	[8.4]	[87.9]	[56.4]	[61.4]	[83.8]
0.5	−0.9		99.8	64.6	17.5	7.3	99.0	64.2	25.2	10.3	99.2	81.5	96.9	98.4
			[99.8]	[61.2]	[13.5]	[5.4]	[99.0]	[60.2]	[21.6]	[7.0]	[99.2]	[78.8]	[95.6]	[97.6]
	−0.5		92.4	40.5	10.1	6.8	90.6	41.2	14.8	7.2	91.1	61.2	85.1	89.3
			[92.5]	[35.5]	[7.9]	[4.2]	[91.1]	[39.5]	[11.8]	[4.3]	[91.1]	[57.4]	[80.2]	[87.4]
	0.0		43.4	16.7	7.4	6.6	43.8	17.9	10.6	8.1	41.3	26.4	40.2	43.2
			[45.4]	[13.6]	[4.9]	[4.2]	[45.1]	[14.2]	[4.6]	[2.1]	[40.6]	[20.7]	[32.5]	[35.9]
	0.9		27.8	27.2	26.1	22.1	25.5	28.3	31.1	31.3	29.7	32.2	31.9	25.7
			[36.7]	[14.0]	[7.6]	[4.6]	[32.8]	[13.4]	[8.5]	[5.0]	[31.0]	[16.1]	[17.7]	[22.5]

IO			LS				M				GM			
$\phi_{1,1}$	$\phi_{2,1}$	ζ	0	3	5	7	0	3	5	7	0	3	5	7
−0.5	−0.9		53.3	52.4	57.8	63.4	51.7	53.5	79.1	92.6	49.0	50.7	65.7	73.6
			[55.3]	[50.7]	[56.5]	[64.3]	[54.4]	[49.0]	[79.5]	[92.3]	[48.4]	[49.6]	[62.1]	[71.3]
	0.0		43.7	47.3	54.0	62.3	42.5	49.5	75.5	87.7	42.9	44.0	47.4	45.6
			[44.6]	[46.1]	[52.5]	[55.5]	[43.0]	[46.4]	[76.8]	[87.5]	[42.7]	[39.0]	[45.9]	[44.1]
	0.4		86.9	89.6	93.5	94.8	85.3	87.9	92.2	75.9	84.2	82.8	83.2	87.2
			[87.9]	[90.4]	[94.4]	[95.0]	[86.6]	[89.3]	[92.4]	[75.4]	[83.9]	[80.2]	[82.4]	[85.8]
	0.9		91.0	90.4	90.1	91.0	85.4	75.9	55.9	30.7	87.2	79.1	70.9	60.9
			[93.0]	[93.0]	[91.7]	[92.4]	[88.7]	[80.7]	[56.3]	[32.2]	[87.9]	[78.9]	[71.7]	[58.1]
0.5	−0.9		99.8	99.8	99.8	99.5	99.0	98.0	89.3	64.7	99.2	98.0	98.2	98.6
			[99.8]	[99.8]	[99.8]	[99.6]	[99.0]	[98.1]	[89.9]	[65.2]	[99.2]	[97.5]	[98.1]	[98.5]
	−0.5		92.4	95.2	96.8	97.3	90.6	92.9	91.2	74.0	91.1	88.0	89.3	92.3
			[92.5]	[96.0]	[97.1]	[97.8]	[91.1]	[93.6]	[91.6]	[74.5]	[91.1]	[87.3]	[87.9]	[91.3]
	0.0		43.4	49.0	68.7	61.9	43.8	50.5	77.2	86.0	41.3	42.4	45.2	43.9
			[45.4]	[53.7]	[61.5]	[67.7]	[45.1]	[52.5]	[78.1]	[86.8]	[40.6]	[40.2]	[42.7]	[41.6]
	0.9		27.8	31.5	48.8	41.4	25.5	33.9	58.9	76.0	28.4	35.1	51.1	61.4
			[36.7]	[42.7]	[47.2]	[50.1]	[32.8]	[40.4]	[58.1]	[77.2]	[31.0]	[35.1]	[50.4]	[59.0]

Rejection frequencies for F-versions of the LM_1^{LS} test based on (6.26) and LM_1^{M} and LM_1^{GM} tests based on (6.27) at 5% significance level, for series generated by the AO model (6.3) or the IO model (6.4), where the core process follows the LSTAR process (6.34)-(6.35) with $\gamma = 2.5$ and $\sigma_\eta^2 = 1$. Outliers occur with probability $\pi = 0.05$. The table is based on 1000 replications for sample size $T = 150$. Entries in brackets denote size-corrected rejection frequencies using the procedure described in Section 6.5.3.

Table 6.3: Power of LM_2 test for STAR nonlinearity

AO			LS				M				GM		
$\phi_{1,1}$	$\phi_{2,1}$ ζ	0	3	5	7	0	3	5	7	0	3	5	7
−0.5	−0.9	4.2	22.3	67.4	92.2	4.4	18.8	43.9	40.7	6.7	16.0	19.2	8.7
		[4.4]	[4.1]	[3.5]	[4.2]	[4.5]	[2.8]	[3.2]	[2.8]	[6.7]	[2.8]	[3.0]	[6.5]
	0.0	16.9	8.4	7.7	7.3	17.1	10.0	10.4	10.4	21.3	17.5	22.1	22.9
		[18.0]	[7.4]	[5.3]	[5.8]	[17.1]	[8.4]	[6.6]	[5.8]	[19.1]	[16.6]	[18.8]	[20.2]
	0.4	39.1	7.7	12.1	14.1	39.3	8.0	18.9	33.0	46.6	17.3	30.3	43.3
		[41.2]	[3.5]	[3.8]	[5.1]	[41.9]	[3.6]	[3.3]	[4.4]	[44.9]	[12.0]	[24.8]	[39.7]
	0.9	9.8	23.2	71.1	91.5	9.8	19.6	57.7	59.4	17.8	21.6	31.1	17.2
		[11.3]	[2.8]	[2.6]	[3.9]	[11.1]	[2.3]	[0.6]	[1.1]	[17.8]	[3.3]	[3.0]	[13.8]
0.5	−0.9	10.7	13.1	57.0	87.7	11.1	12.3	44.0	51.7	20.6	15.9	22.3	20.0
		[11.6]	[1.3]	[1.1]	[2.8]	[11.1]	[1.1]	[0.5]	[1.2]	[18.2]	[2.9]	[2.1]	[16.6]
	−0.5	48.7	8.5	21.7	26.8	46.5	9.9	40.5	63.4	53.0	19.0	32.5	48.1
		[48.2]	[3.4]	[4.5]	[5.8]	[45.2]	[2.8]	[2.8]	[5.3]	[50.6]	[11.4]	[22.0]	[43.2]
	0.0	19.6	8.8	7.2	7.4	19.3	9.3	8.5	9.1	22.1	17.0	20.0	19.7
		[20.2]	[7.6]	[4.8]	[4.0]	[19.6]	[7.7]	[5.9]	[5.1]	[20.1]	[14.1]	[16.2]	[16.9]
	0.9	3.6	31.2	79.8	95.4	3.9	25.2	52.3	48.3	6.2	20.8	22.8	6.2
		[3.9]	[4.1]	[3.6]	[5.4]	[4.1]	[2.3]	[3.7]	[2.3]	[4.6]	[2.6]	[3.9]	[5.2]

IO			LS				M				GM		
$\phi_{1,1}$	$\phi_{2,1}$ ζ	0	3	5	7	0	3	5	7	0	3	5	7
−0.5	−0.9	4.2	3.6	4.7	5.4	4.4	3.3	3.6	4.7	6.7	4.7	4.4	5.2
		[4.4]	[3.8]	[4.5]	[4.2]	[4.5]	[3.0]	[5.7]	[5.4]	[6.7]	[4.1]	[6.3]	[6.3]
	0.0	16.9	8.4	6.6	7.0	17.1	10.9	10.7	11.2	21.3	16.0	21.2	23.0
		[18.0]	[9.4]	[5.4]	[4.1]	[17.1]	[10.7]	[10.3]	[12.7]	[19.1]	[12.9]	[19.0]	[20.8]
	0.4	39.1	15.3	7.0	5.7	39.3	17.8	17.9	17.4	46.6	30.6	36.2	37.1
		[41.2]	[19.4]	[7.9]	[5.3]	[41.9]	[19.4]	[19.9]	[19.9]	[44.9]	[28.7]	[32.5]	[33.5]
	0.9	9.8	5.3	4.6	6.3	9.7	6.2	7.6	5.9	17.8	12.2	11.5	10.0
		[11.3]	[5.4]	[4.5]	[5.0]	[11.1]	[7.1]	[9.4]	[6.1]	[17.8]	[10.9]	[10.1]	[8.6]
0.5	−0.9	10.7	4.9	4.1	4.6	11.1	6.3	6.0	5.8	20.6	12.8	10.6	10.6
		[11.6]	[4.5]	[4.0]	[4.2]	[11.1]	[6.6]	[7.2]	[6.2]	[18.2]	[12.4]	[10.9]	[8.9]
	−0.5	48.7	20.4	10.6	7.7	46.5	21.3	16.7	16.0	53.0	33.3	39.4	39.3
		[48.2]	[19.6]	[8.5]	[5.1]	[45.2]	[19.7]	[16.1]	[16.1]	[50.6]	[31.9]	[36.0]	[38.5]
	0.0	19.6	9.0	7.5	7.2	19.3	9.8	9.8	11.0	22.1	17.4	18.3	19.8
		[20.2]	[11.6]	[7.2]	[5.8]	[19.6]	[10.7]	[10.9]	[13.3]	[20.1]	[16.4]	[16.0]	[18.8]
	0.9	3.6	4.9	5.2	5.9	3.9	4.7	4.5	4.3	6.2	8.2	5.9	7.0
		[3.9]	[5.0]	[5.6]	[4.7]	[4.1]	[4.8]	[5.1]	[4.8]	[4.6]	[5.7]	[5.3]	[5.9]

Rejection frequencies for F-versions of the LM_2^{LS} test based on (6.26), and LM_2^{M} and LM_2^{GM} tests based on (6.27) at 5% significance level for series generated by the AO model (6.3) or the IO model (6.4), where the core process follows the ESTAR process (6.34)-(6.36) with $\gamma = 10$, $c_1 = -1$, $c_2 = 1$ and $\sigma_\eta^2 = 1$. Outliers occur with probability $\pi = 0.05$. The table is based on 1000 replications for sample size $T = 150$. Entries in brackets denote size-corrected rejection frequencies using the procedure described in Section 6.5.3.

6.5.4 Discussion

The Monte Carlo results discussed above suggest that the empirical performance of the tests based on the GM estimator is satisfactory. If applied to time series without outliers, the performance of these robust tests is similar to that of the more familiar OLS-based tests, with an expected slight power advantage for the OLS-based tests. Furthermore, in case of linear or nonlinear time series with not too many outliers, the GM-based tests point at the correct model more often than the standard test. Results for other DGPs, other sample sizes and other contamination fractions concur with these findings.

The relative performance of the standard and robust LM-type tests, as well as the sequences of LM-type tests that are used to decide between LSTAR and ESTAR models in the specification procedures of Teräsvirta (1994) and Escribano and Jordá (1999) (see Section 2.2) are investigated further in Escribano *et al.* (1998). Combined with the results presented here, the following guidelines can be given on how to proceed in practice, when one cannot be sure whether certain features of a particular time series are caused by genuine nonlinearity or by some outliers.

It seems worthwhile to apply both the standard and robust linearity tests and specification procedures and to combine the outcomes to reach a conclusion using the following decision rules. If both the standard and robust tests do not reject the null hypothesis of linearity, one can be reasonably confident that the DGP of the series is linear. When both standard and robust tests reject the null hypothesis, one might assume that the DGP of the series is genuinely nonlinear - although it is possible of course that it is linear with a high frequency of occurrence of large outliers, such that the sizes of both test procedures are heavily distorted. The case where the standard tests reject linearity and the robust tests do not, points towards the possibility that the nonlinearity which is detected by the standard test procedures is caused by only a few outliers. A further investigation of the series, especially the influential observations (that is, those observations that are downweighted by the robust estimation procedure) is strongly called for. The fact that the robust estimation procedure endogenously determines the weights for the different observations is advantageous here, as this allows one to determine which observations cause the standard tests to reject the null hypothesis. Alternatively, one might have encountered a case where the DGP is nonlinear but contaminated in such a way that the power of the standard test (when based on incorrect critical values from the distribution which holds under no contamination) increases while the power of the robust test does not. Also in this case it is advisable to further investigate the series for the presence of outliers before estimating a nonlinear model. Finally, if the standard test does not reject the null, while the robust test does, it is perhaps most likely that the DGP is nonlinear with some contamination such that the power of the standard test is decreased.

6.6 Nonlinearity in industrial production

As stated in the introduction to this chapter, regime-switching models have been applied in particular to study possible nonlinearity in business cycles. In this spirit, Luukkonen and Teräsvirta (1991), Teräsvirta and Anderson (1992) and Teräsvirta *et al.* (1994) consider modeling industrial production indexes for a number of OECD countries with STAR models. In this section, I apply the standard and robust tests for nonlinearity to similar series on an extended sample period.

I examine quarterly, seasonally unadjusted indexes of industrial production for several OECD countries, taken from the *OECD Main Economic Indicators.* The sample consists of 18 such series, covering the period 1960:1-1997:1. Following Teräsvirta and Anderson (1992), the data are made approximately stationary by transforming them to yearly growth rates - that is, by taking seasonal differences of logarithmic transformed data.

First of all, I apply the LM-type tests against STAR-type nonlinearity with the delay parameter d left unspecified, that is, with $s_t = \alpha' x_t$ for certain $\alpha = (0, \ldots, 0, 1, 0, \ldots, 0)'$, as discussed in Section 6.3. The orders of the linear AR models under the null are determined using the SIC. The p-values corresponding to the standard and robust variants of the LM_1, LM_3 and LM_3^e tests are shown in Table 6.4. It is seen that almost invariably the p-values for the robust tests are larger than for the OLS-based tests. For several countries the increase is quite considerable, see, for example, Austria and the US. In fact, quite often the conclusion concerning the possibility of nonlinearity in these series changes as one uses robust instead of standard tests. For example, based on the $\mathrm{LM}_3^{e,\mathrm{LS}}$ test, linearity can be rejected for 8 (11) countries at the 5% (10%) significance level, whereas using the $\mathrm{LM}_3^{e,\mathrm{M}}$ and $\mathrm{LM}_3^{e,\mathrm{GM}}$ tests one would do so only for 3 (6) and 5 (6) countries, respectively.

Table 6.5 shows the results from applying both the standard and robust methods to compute the various tests in the specification procedure of Teräsvirta (1994), as discussed in Section 2.2. The LM_3 test is computed against the alternative of a STAR model with $s_t = y_{t-d}$ for $1 \leq d \leq 6$. The value of d which renders the smallest p-value is selected as indicating the appropriate transition variable. The choice for an LSTAR or ESTAR model is based on the sequence of tests of sub-hypotheses of the null hypothesis that is tested by LM_3, see Section 2.2 for details. Columns 3-5 of Table 6.5 contain the minimum p-values of the LM_3 statistic based on OLS, M and GM estimates of the AR(p) model under the null. The value of d for which this minimum p-value occurs is shown in brackets in columns 6-8, following the selected model. A linear model is considered appropriate in case the minimum p-value of the LM_3 statistic is larger than 0.10.

Comparing the models that are selected by the OLS- and GM-based procedures, it is seen that for the majority of countries different conclusions are drawn. Only for France, Ireland, Luxemburg, The Netherlands, Portugal, and the US exactly the same results are obtained. Inspection of the weights for these countries resulting from the GM estimator (not shown here) reveals that outliers seem to be present around 1975, although the number and timing varies considerably across the different countries. Furthermore, the French industrial production series contain obvious AOs

Table 6.4: p-values of LM-type tests for quarterly industrial production series

		LS			M			GM		
Country	p^{a}	LM_1	LM_3	LM_3^e	LM_1	LM_3	LM_3^e	LM_1	LM_3	LM_3^e
Austria	5	0.040	0.110	0.026	0.803	0.494	0.785	0.803	0.494	0.785
Belgium	5	0.213	0.267	0.202	0.121	0.244	0.099	0.018	0.223	0.020
Canada	2	0.086	0.253	0.099	0.329	0.245	0.133	0.129	0.049	0.014
Finland	1	0.955	0.644	0.989	0.967	0.571	0.821	0.323	0.609	0.413
France	5	0.000	0.068	0.000	0.124	0.215	0.125	0.056	0.166	0.003
Greece	1	0.199	0.104	0.046	0.261	0.095	0.041	0.261	0.095	0.041
Ireland	5	0.452	0.705	0.723	0.425	0.536	0.674	0.425	0.536	0.674
Italy	5	0.057	0.011	0.024	0.451	0.126	0.085	0.451	0.126	0.085
Japan	5	0.087	0.021	0.080	0.155	0.144	0.159	0.155	0.144	0.159
Luxemburg	7	0.012	0.227	0.026	0.103	0.401	0.260	0.317	0.727	0.218
The Netherlands	5	0.396	0.033	0.134	0.389	0.281	0.352	0.400	0.252	0.201
Norway[b]	8	0.032	–	0.014	0.042	–	0.023	0.090	–	0.092
Portugal	1	0.026	0.054	0.048	0.017	0.042	0.038	0.002	0.023	0.010
Spain	5	0.038	0.397	0.051	0.405	0.809	0.274	0.405	0.809	0.274
Sweden	4	0.858	0.006	0.879	0.450	0.290	0.693	0.347	0.239	0.602
Switzerland	1	0.596	0.257	0.275	0.358	0.015	0.017	0.226	0.310	0.310
United Kingdom	5	0.013	0.003	0.017	0.065	0.001	0.073	0.400	0.004	0.421
United States	2	0.006	0.041	0.012	0.259	0.482	0.388	0.358	0.311	0.359

The series are seasonal differences of quarterly observations on the indices for industrial production (1990=100), taken from the OECD *Main Economic Indicators*. The sample period covers 1960:1-1997:1. For Canada and Spain the series start in 1961:1, for Greece in 1962:1. The sample for Austria ends in 1995:4, for Canada, Greece, Norway and Sweden in 1995:1, for Ireland in 1994:4, for Luxemburg and Switzerland in 1996:4. The tests are F-variants of the LM-type statistics as discussed in Section 6.3.

[a] AR orders have been determined by SIC.

[b] For Norway the LM_3 test could not be computed due to a shortage in degrees of freedom.

Table 6.5: Model selection for quarterly industrial production series

		Minimum p-value			Type of model[a]		
Country	p[b]	LS	M	GM	LS	M	GM
Austria	5	0.007	0.178	0.178	LSTAR[1]	Linear	Linear
Belgium	5	0.102	0.113	0.019	Linear	Linear	LSTAR[1]
Canada	2	0.193	0.193	0.010	Linear	Linear	LSTAR[2]
Finland	1	0.009	0.005	0.051	LSTAR[4]	LSTAR[4]	ESTAR[5]
France	5	0.000	0.011	0.002	LSTAR[2]	ESTAR[2]	LSTAR[2]
Greece	1	0.003	0.003	0.003	LSTAR[4)	LSTAR[4]	ESTAR[4]
Ireland	5	0.475	0.437	0.437	Linear	Linear	Linear
Italy	5	0.006	0.007	0.007	LSTAR[1]	LSTAR[3]	LSTAR[3]
Japan	5	0.001	0.112	0.112	ESTAR[1]	Linear	Linear
Luxemburg	7	0.000	0.009	0.003	LSTAR[1]	LSTAR[1]	LSTAR[1]
The Netherlands	5	0.005	0.008	0.007	ESTAR[1]	LSTAR[1]	ESTAR[1]
Norway	8	0.000	0.000	0.006	LSTAR[1]	LSTAR[2]	ESTAR[3]
Portugal	1	0.021	0.034	0.005	LSTAR[2]	LSTAR[4]	LSTAR[2]
Spain	5	0.094	0.546	0.546	ESTAR[1]	Linear	Linear
Sweden	4	0.085	0.147	0.101	LSTAR[4]	Linear	Linear
Switzerland	1	0.058	0.015	0.233	LSTAR[4]	ESTAR[1]	Linear
United Kingdom	5	0.031	0.018	0.020	LSTAR[5]	LSTAR[4]	ESTAR[4]
United States	2	0.000	0.001	0.000	LSTAR[4]	ESTAR[4]	LSTAR[4]

Columns 3-5 report the minimum p-value attained by the F variants of the standard and robust LM_3 statistic against the alternative of a STAR model with $s_t = y_{t-d}$. The minimum is computed by varying d over $1 \leq d \leq 6$.

[a] The value of the delay parameter d for which the LM_3 test statistic attains the minimum p-value is given in square brackets.

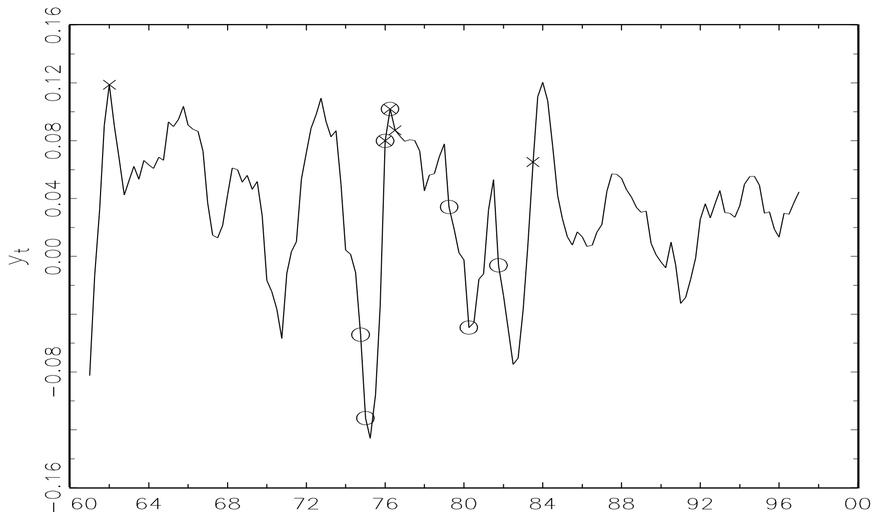
[b] AR orders have been determined by SIC.

due to nationwide strikes in 1963:1 and 1968:2. Apparently, the presence of these outlying observations does not affect the inference concerning potential nonlinear properties of these series. Comparing the p-values of the various tests in Tables 6.4 and 6.5 for these countries, it is seen that for France, Ireland, and Portugal the conclusions from the general nonlinearity tests (Table 6.4) and the tests with specific delay parameter (Table 6.5) more or less coincide. For Luxemburg, the Netherlands and the US on the other hand, using the general M- and GM-based tests, the null hypothesis of linearity can not be rejected at conventional significance levels, while the null can be rejected quite convincingly once specific transition variables are considered. For the US, for example, the p-values for the LM_3^{M} and LM_3^{GM} tests drop from 0.482 and 0.311 to 0.001 and 0.000, respectively. To investigate the possible cause of this marked difference, I estimate an LSTAR model with delay parameter $d = 4$ as selected by both the OLS- and GM-based procedures for this series. Figure 6.6 plots the time series for the seasonal differences of US industrial production in the upper panel, along with the weights assigned to the observations by the GM estimator which is used to estimate the AR(2) model under the null, and the values taken by the transition function $G(y_{t-d}; \gamma, c)$ in the fitted LSTAR model in the lower panel. Observations that receive a weight smaller than 1 in the GM estimation procedure and for which the value taken by the transition function is less than 0.5 are marked with circles and crosses, respectively, in the upper panel of this figure. These graphs reveal that the regime of the LSTAR model corresponding to $G(y_{t-d}; \gamma, c) = 0$ becomes active only for a few observations in 1976. The same observations are downweighted by the GM estimator. Hence, one might be tempted to consider these observations as outliers, and abstain from estimating STAR models for this series.

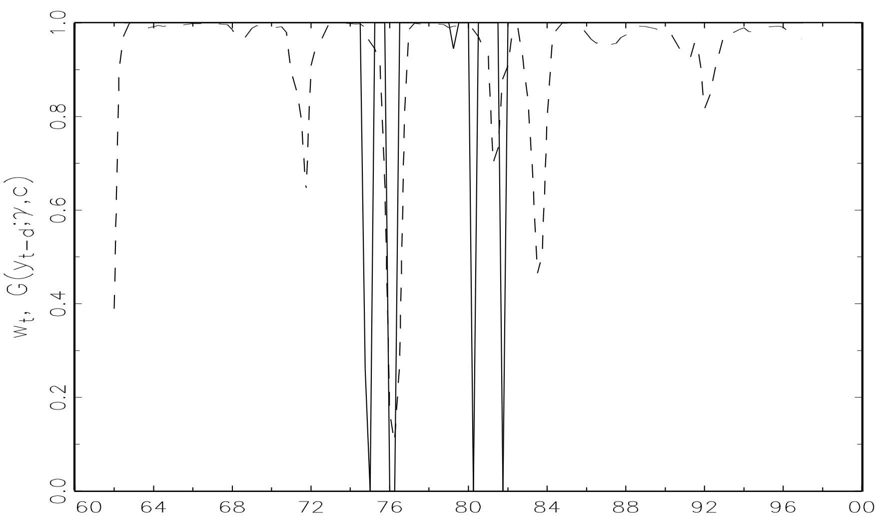
For Finland, Greece, Italy, Norway, and the UK the results coincide, albeit partially, in the sense that for these countries the standard and GM based procedures both indicate that a STAR model is appropriate. The selected delay parameter d and the transition function differ however. The Norwegian industrial production series contains an obvious (positive) AO due to the first oil crisis in the second quarter of 1975, whereas the series for Italy clearly is affected by widespread industrial action in 1969:4. Evidently, the presence of such AOs can influence the conclusion from the minimum p-value rule to determine the appropriate delay parameter d, and the sequence of tests to discriminate between LSTAR and ESTAR alternative.

For Austria, Japan, Spain, Sweden, and Switzerland the standard tests indicate that a STAR model might be appropriate, whereas the robust tests are unable to reject linearity at the 10% significance level. For all five countries, I estimate the STAR model that is selected by the OLS-based specification procedure and graph the time series, the weights from estimating the selected linear AR model using the GM estimator, and the values taken by the transition function in the fitted STAR model, as described previously in detail for the US. For illustration, Figures 6.7 and 6.8 show these graphs for Austria and Sweden, respectively.

Inspection of these figures shows that for these countries the apparent nonlinearity is due to only a few outlying observations. For Austria, the results are driven to a large extent by the exceptionally large observation in 1972:4. The estimates of

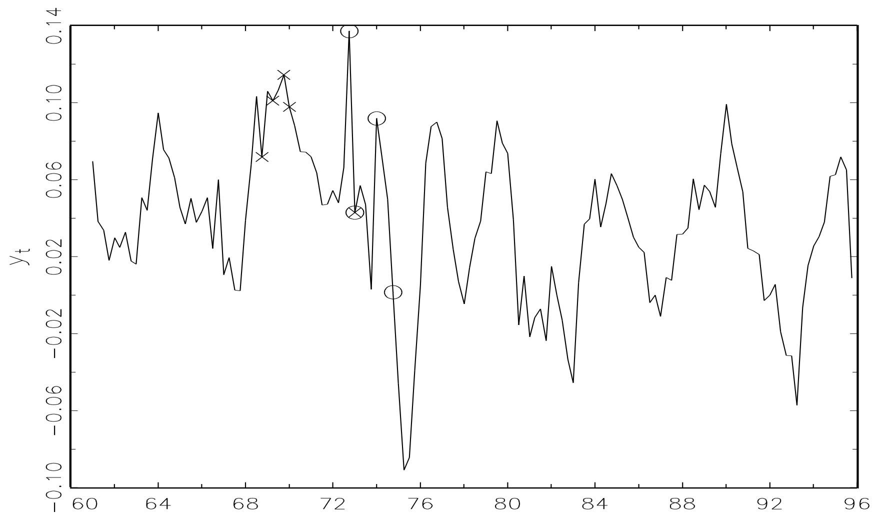


(a) Yearly growth rates

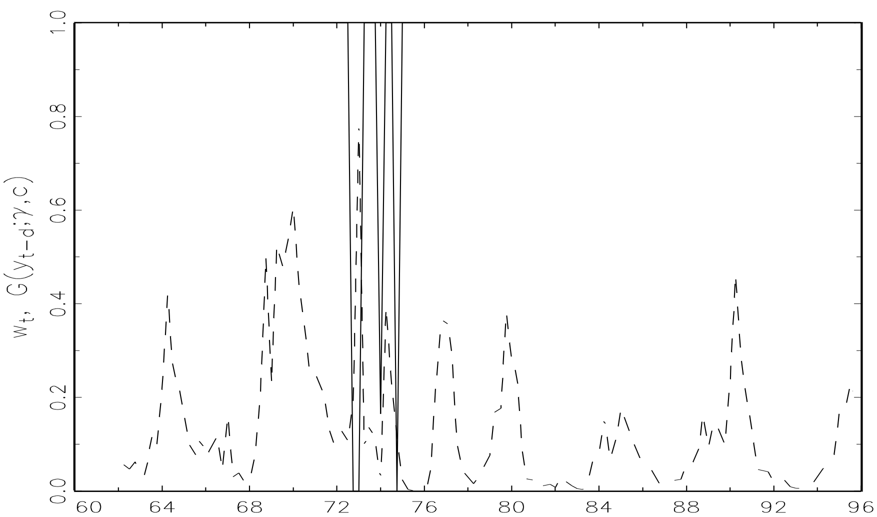


(b) Weights and transition function

Figure 6.6: Industrial production - United States. The upper graph contains observations on the seasonal difference of the logarithm of the index of US industrial production. Observations that receive a weight smaller than 1 in the HBP-GM estimation procedure, which is used to estimate an AR(2) model for this series, are marked with a circle. Observations for which the value of the transition function in the fitted LSTAR model, $G(y_{t-d};\gamma,c) = (1+\exp\{-\gamma(y_{t-d}-c)\})^{-1}$, $d=4, \gamma = 1.89 \times 21, c = -0.071$, is smaller than 0.5 are marked with a cross. The lower graph contains the actual weights w_t (solid line) and the values taken by the transition function (dashed line).

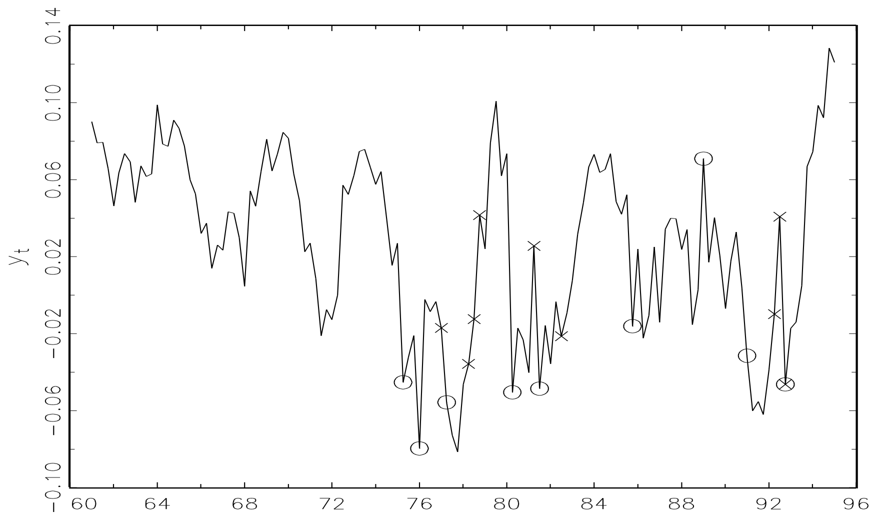


(a) Yearly growth rates

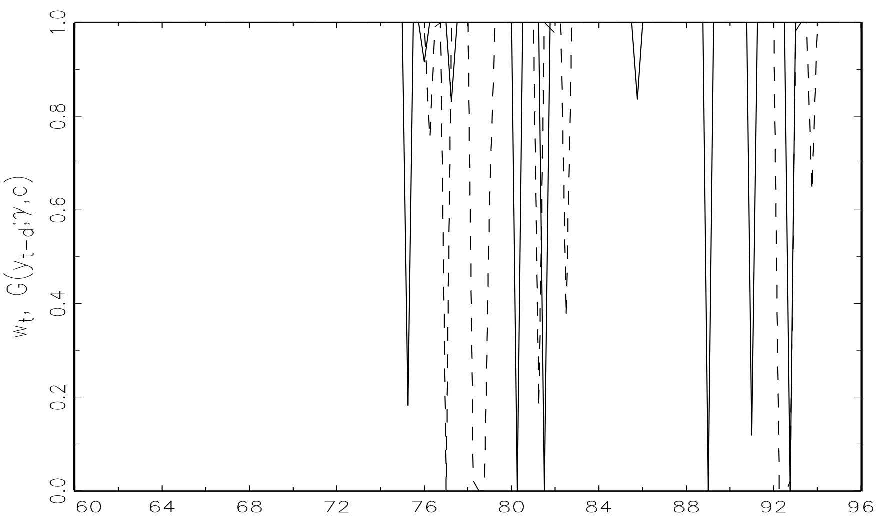


(b) Weights and transition function

Figure 6.7: Industrial production - Austria. The upper graph contains observations on the seasonal difference of the logarithm of the index of industrial production for Austria. Observations that receive a weight smaller than 1 in the HBP-GM estimation procedure, which is used to estimate an AR(5) model for this series, are marked with a circle. Observations for which the value of the transition function in the fitted LSTAR model, $G(y_{t-d};\gamma,c) = (1+\exp\{-\gamma(y_{t-d}-c)\})^{-1}$, $d = 1, \gamma = 1.54 \times 25.2, c = 0.103$, is larger than 0.5 are marked with a cross. The lower graph contains the actual weights w_t (solid line) and the values taken by the transition function (dashed line).



(a) Yearly growth rates



(b) Weights and transition function

Figure 6.8: Industrial production - Sweden. The upper graph contains observations on the seasonal difference of the logarithm of the index of Sweden industrial production. Observations that receive a weight smaller than 1 in the HBP-GM estimation procedure, which is used to estimate an AR(4) model for this series, are marked with a circle. Observations for which the value of the transition function in the fitted ESTAR model, $G(y_{t-d};\gamma,c) = (1+\exp\{-\gamma(y_{t-d}-c)\})^{-1}$, $d = 4, \gamma = 22.40\times 21.75, c = -0.047$, is smaller than 0.5 are marked with a cross. The lower graph contains the actual weights w_t (solid line) and the values taken by the transition function (dashed line).

γ and c are such that the logistic transition function $G(y_{t-1}; \gamma, c)$ is close to 1 only for the subsequent observation in 1973:1. Both observations evidently are marked as outlying, as they receive weights equal to zero in the robust estimation of the linear AR(5) model for this series. Note that a similar conclusion is reached by Teräsvirta *et al.* (1994), who estimate the same LSTAR model for this series on a shorter sample period. For Sweden, the regime corresponding to $G(y_{t-4}; \gamma, c) = 0$ appears necessary only to describe observations in the recovery phase following the recessions in 1977 and 1991. The weights from the GM-estimation procedure, as shown in Figure 6.8, suggest that this series contains a number of observations that are suspect and might be considered aberrant data points. In this case the apparent outliers do not clearly correspond with one of the two regimes in the estimated STAR model, and are not (all) located around the points where transitions from one regime to the other regime in the STAR model occur.

Finally, for Belgium and Canada the OLS-based tests do not reject linearity at the 10% significance level, whereas the GM-based tests do. This demonstrates that outliers also can mask nonlinear properties of a time series.

6.7 Concluding remarks

In this chapter I have proposed outlier robust LM-type tests for STAR nonlinearity. The tests, which are straightforward to compute, use a HBP-GM estimator to estimate the linear AR model under the null hypothesis. The Monte Carlo evidence suggests that the empirical performance of the tests is satisfactory. If applied to time series without outliers, they do not suffer from large size distortions or much loss of power. Furthermore, in case of linear or nonlinear time series with outliers, the robust tests hint toward the correct model more often than the standard tests. The application to several industrial production series indicates that one should carefully interpret evidence from standard tests, as the presence of only a few aberrant observations may cause spurious nonlinearity. It should be remarked however that I do not want to argue that all nonlinearity is caused entirely by outliers. In fact, as a guideline for practitioners I recommend applying both the standard and robust tests, and basing the conclusion concerning the presence of nonlinearity in empirical time series on the combined test results, as discussed in Section 6.5.4. In case the conclusions from the standard and robust test procedures coincide, one might be reasonably confident concerning the presence or absence of nonlinearity for the series under consideration. For example, for series with fairly obvious AOs such as the French, Norwegian and Italian industrial production series, the presence of these outliers appeared not to critically influence the conclusions concerning the possible nonlinearity in these series. Both the OLS- and GM-based LM-type tests convincingly rejected the null hypothesis. Hence, one may safely conclude that these series contain some intrinsic nonlinear properties.

Furthermore, whether certain observations are outliers or belong to a separate regime sometimes can be a more subjective than an objective decision. For example, for the US industrial production series, most of the observations which are downweighted in the robust estimation procedure belong to recessions or recoveries. Ultimately, it is a matter of subjective judgement whether the dynamic properties of the series are genuinely different in such periods and to incorporate this explicitly into a time series model, or whether such observations can safely be neglected for practical purposes. When deliberating over this decision, it appears sensible to compare the number of additional parameters which are to be estimated in a nonlinear time series model with the number of observations for which these parameters actually are necessary.

The results obtained in this chapter are not confined to the LM-type tests for STAR nonlinearity. The robust estimation techniques can be applied straightforwardly to construct robust tests for other types of misspecification. Motivated by the discussion on outliers and nonlinearity, an obvious possibility is to consider robust testing for autoregressive conditional heteroskedasticity [ARCH]. The clustering of large residuals typical in ARCH models, might well be mimicked by a sequence of AOs. The behaviour of tests for ARCH in the presence of outliers is considered in van Dijk, Franses and Lucas (1999b) and Franses, van Dijk and Lucas (1998). It is found that outliers have similar effects on the size and power properties of the ARCH test as encountered in this chapter. Outliers can suggest ARCH when it in fact is not present, whereas they also can mask true ARCH effects. Robust test statistics are shown to be useful to discriminate between outliers and genuine ARCH effects.

6.A Proof of the Theorem

Proof of Theorem 1a: The first part of the theorem follows directly from Theorem 1c with $\psi(\varepsilon_t) = \varepsilon_t$ and $w_x(\cdot) \equiv 1$, where $\varepsilon_t = y_t - \tilde{\phi}_1 y_{t-1}$, with $\tilde{\phi}_1$ the pseudo true parameter value under the null - that is, the value satisfying

$$E(y_{t-1} \cdot (y_t - \tilde{\phi}_1 y_{t-1})) = 0.$$

The second part of the theorem is proved if it is shown that $c_L = \partial(k_1/k_2)/\partial\pi$ evaluated in $\pi = 0$, with k_1 and k_2 from the proof of Theorem 1c. This is similar to deriving the influence function of a statistic, see, e.g., Hampel *et al.* (1986). In order to derive the local result, k_1/k_2 from the proof of Theorem 1c is split in two parts using the specific choice of $\psi(\cdot)$ and $w_x(\cdot)$, namely

$$\frac{E(y_{t-1}^4 \varepsilon_t^2)}{k_2} \tag{6A.1}$$

and

$$2\sum_{k=1}^{\infty} \frac{E(y_{t-1}^2 y_{t-k-1}^2 \varepsilon_t \varepsilon_{t-k})}{k_2}, \tag{6A.2}$$

with

$$k_2 = E(y_{t-1}^4)E(\varepsilon_t^2),$$

In order to put together the result, I first present the necessary individual derivatives with respect to π. Heavy use is made of the techniques for deriving influence functions in the time series context as presented in Martin and Yohai (1986). I obtain

$$\left.\frac{\partial E(\varepsilon_t^2)}{\partial \pi}\right|_{\pi=0} = E((\eta_t+\zeta)^2-(\eta_t)^2)+E((\eta_t-\phi_1\zeta)^2-(\eta_t)^2)$$
$$= (1-\phi_1^2)\cdot\zeta^2, \tag{6A.3}$$
$$\left.\frac{\partial E(y_{t-1}^4)}{\partial \pi}\right|_{\pi=0} = \zeta^4+6\zeta^2E(z_{t-1}^2), \tag{6A.4}$$
$$\left.\frac{\partial E(y_{t-1}^4\varepsilon_t^2)}{\partial \pi}\right|_{\pi=0} = E((z_{t-1}+\zeta)^4(\eta_t-\phi_1\zeta)^2-z_{t-1}^4\eta_t^2)+E(z_{t-1}^4((\eta_t+\zeta)^2-\eta_t^2))$$
$$= E((z_{t-1}^4+6\zeta^2z_{t-1}^2+\zeta^4)(\eta_t^2+\phi_1^2\zeta^2)-z_{t-1}^4\eta_t^2)+\zeta^2\cdot E(z_{t-1}^4), \tag{6A.5}$$
$$\left.\frac{\partial E(y_{t-1}^2y_{t-2}^2\varepsilon_t\varepsilon_{t-1})}{\partial \pi}\right|_{\pi=0} = -E(z_{t-1}^3z_{t-2}^2\eta_{t-1})IF_{\tilde{\phi}_1}$$
$$+E((z_{t-1}+\zeta)^2z_{t-2}^2(\eta_t-\phi_1\zeta)(\eta_{t-1}+\zeta)), \tag{6A.6}$$

where

$$IF_{\tilde{\phi}_1} = -\phi_1\zeta^2/E(z_{t-1}^2)$$

is the (time series) influence function of the estimator for the autoregressive parameter under AO contamination, and finally

$$\left.\frac{\partial E(y_{t-1}^2y_{t-k-1}^2\varepsilon_t\varepsilon_{t-k})}{\partial \pi}\right|_{\pi=0} = 0, \tag{6A.7}$$

for $k>1$. The result now follows by observing that

$$\left.\frac{\partial}{\partial\pi}\left(\frac{E(y_{t-1}^4\varepsilon_t^2)}{k_2}\right)\right|_{\pi=0} = \frac{\partial E(y_{t-1}^4\varepsilon_t^2)/\partial\pi - E(y_{t-1}^4)\partial E(\varepsilon_t^2)/\partial\pi}{k_2}$$
$$\left.\frac{-E(\varepsilon_t^2)\partial E(y_{t-1}^4)/\partial\pi}{k_2}\right|_{\pi=0} \tag{6A.8}$$

and

$$\left.\frac{\partial}{\partial\pi}\left(2\sum_{k=1}^{\infty}\frac{E(y_{t-1}^2y_{t-k-1}^2\varepsilon_t\varepsilon_{t-k})}{k_2}\right)\right|_{\pi=0} = 2\frac{\partial E(y_{t-1}^2y_{t-2}^2\varepsilon_t\varepsilon_{t-1})/\partial\pi}{k_2}. \tag{6A.9}$$

□

Proof of Theorem 1b: The first part of the theorem follows from Theorem 1c with $w_x(\cdot)\equiv 1$. The second part can be proved along similar lines as the second part of Theorem 1a. One only has to account for the additional fact that the function $\psi(\varepsilon_t/\tilde{\sigma})$ also depends on the estimated scale of the residuals under the null, $\tilde{\sigma}$. As a result, the calculations become somewhat more involved, and the time series influence function of the estimator for $\tilde{\sigma}$ will also enter the expressions. □

Proof of Theorem 1c: Recall the additive outlier model

$$z_t = \phi_1 z_{t-1}+\eta_t, \tag{6A.10}$$
$$y_t = z_t+\zeta\delta_t, \tag{6A.11}$$

where η_t is i.i.d. standard Gaussian distributed, $|\phi_1| < 1$, $z_0 = \eta_0/(1-\phi_1^2)^{1/2}$, and δ_t is i.i.d. with $P(\delta_t = 1) = P(\delta_t = -1) = \pi/2$, $P(\delta_t = 0) = 1-\pi$, and $0 \leq \pi \leq 1$. Define the regression residuals under the null as $\varepsilon_t = y_t - \tilde{\phi}_1 y_{t-1}$, where $\tilde{\phi}_1$ is the pseudo true value, satisfying

$$E(y_{t-1} \cdot w_{t-1} \cdot \psi_t) = 0,$$

with

$$\psi_t = \psi\left(\frac{y_t - \tilde{\phi}_1 y_{t-1}}{\tilde{\sigma} \cdot w_{t-1}}\right),$$
$$w_t = w_x\left(\frac{y_t - \tilde{m}}{\tilde{s}}\right),$$

and $\tilde{\sigma}$, $\tilde{s}$, and $\tilde{m}$ the pseudo true values for the scale of the residuals, the scale of the regressors, and the location of the regressors, respectively, which follow from taking the median absolute deviation of the median. The LM-type test statistic is given by

$$\mathrm{LM}_1^{\mathrm{GM}} = \frac{\left(T^{-1/2}\sum_{t=1}^{T} y_{t-1}^2 w_{t-1}\psi_t\right)^2}{\left(T^{-1}\sum_{t=1}^{T} y_{t-1}^4 w_{t-1}^2\right)\left(T^{-1}\sum_{t=1}^{T}\psi_t^2\right)}. \tag{6A.12}$$

I first prove that the terms in the summation in the numerator have expectation zero. To see this, note that

$$E(y_{t-1}^2 w_{t-1}\psi_t) = E((z_{t-1}+\zeta\delta_{t-1})^2 \cdot w_x\left(\frac{z_{t-1}+\zeta\delta_{t-1}-\tilde{m}}{\tilde{s}}\right) \cdot \psi\left(\frac{\eta_t + \zeta\delta_t - \tilde{\phi}_1\zeta\delta_{t-1} + (\phi_1 - \tilde{\phi}_1)z_{t-1}}{\tilde{\sigma}\cdot w_x\left(\frac{z_{t-1}+\zeta\delta_{t-1}-\tilde{m}}{\tilde{s}}\right)}\right) \tag{6A.13}$$

$$= E((-z_{t-1}-\zeta\delta_{t-1})^2 \cdot w_x\left(\frac{-z_{t-1}-\zeta\delta_{t-1}-\tilde{m}}{\tilde{s}}\right) \cdot \psi\left(\frac{-\eta_t - \zeta\delta_t + \tilde{\phi}_1\zeta\delta_{t-1} - (\phi_1 - \tilde{\phi}_1)z_{t-1}}{\tilde{\sigma}\cdot w_x\left(\frac{-z_{t-1}-\zeta\delta_{t-1}-\tilde{m}}{\tilde{s}}\right)}\right) \tag{6A.14}$$

$$= -E((z_{t-1}+\zeta\delta_{t-1})^2 \cdot w_x\left(\frac{z_{t-1}+\zeta\delta_{t-1}-\tilde{m}}{\tilde{s}}\right) \cdot \psi\left(\frac{\eta_t + \zeta\delta_t - \tilde{\phi}_1\zeta\delta_{t-1} + (\phi_1 - \tilde{\phi}_1)z_{t-1}}{\tilde{\sigma}\cdot w_x\left(\frac{z_{t-1}+\zeta\delta_{t-1}-\tilde{m}}{\tilde{s}}\right)}\right) \tag{6A.15}$$

$$= 0, \tag{6A.16}$$

where the equality between (6A.13) and (6A.14) follows from the assumption of a symmetric (i.i.d.) distribution for η_t and δ_t, the equality between (6A.14) and (6A.15) follows from the fact that the median $\tilde{m}$ is first order unbiased under symmetric contamination,

w_x is symmetric, and ψ is anti-symmetric. The last equality then follows from (6A.15) being equal to (6A.13). Therefore, the expression within the square in the numerator of (6A.12) satisfies a central limit theorem and converges in distribution to the square of a normal random variate with mean zero and variance k_1, with

$$k_1 = E(y_{t-1}^4 w_{t-1}^2 \psi_t^2) + 2\sum_{k=1}^{\infty} E(y_{t-1}^2 y_{t-k-1}^2 w_{t-1} w_{t-k-1} \psi_t \psi_{t-k}). \tag{6A.17}$$

Define

$$k_2 = E(y_{t-1}^4 w_{t-1}^2) E(\psi_t^2), \tag{6A.18}$$

then using the central limit theorem, $\mathrm{LM}_1^{\mathrm{GM}}$ tends in distribution to

$$\mathrm{LM}_1^{\mathrm{GM}} \xrightarrow{d} \frac{k_1}{k_2}\chi_1^2.$$

Defining $c_G \equiv k_1/k_2$, this proves the first part of the theorem. Note that $k_1 = k_2$ for $\pi = 0$.

The second part can be proved along similar lines as the second part of Theorem 1b. One only has to account for the additional fact that the weight function $w_x(y_{t-1})$ also depends on the estimated location ($\tilde{m}$) and scale ($\tilde{s}$) parameters of y_{t-1}. As a result, the calculations become even more involved than in the case of M estimators, and the time series influence function of the estimator for $\tilde{m}$ and $\tilde{s}$ also enter the expressions. □

Chapter 7

On Outlier Robust Estimation of Smooth Transition Models

The discussion in the previous chapter centers on the possibility that smooth transition type nonlinearity and outliers in an otherwise linear time series can be observationally equivalent and, hence, can easily be mistaken. The robust linearity tests that were developed are meant to avoid confusing these two alternatives. A third possibility is of course that a time series contains nonlinear features and is subject to contamination at the same time. For example, the Lagrange Multiplier [LM] type tests clearly rejected linearity for the French, Italian and Norwegian industrial production series in Section 6.6, whereas these series contain obvious aberrant observations, which are due to exogenous events. In this chapter I focus on this third possibility and consider outlier robust estimation methods for smooth transition autoregressive [STAR] models.

The effects of outliers on parameter estimates and on inference in linear models has been studied quite extensively. For nonlinear models, this is not the case. This is partly due to the fact that many of the concepts that are routinely used to assess the effects of outliers in linear models appear problematic, or even impossible, to apply in nonlinear models, see Stromberg and Ruppert (1992) for example. This chapter therefore takes a more pragmatic approach. First, I estimate STAR models for several of the industrial production series considered in the previous chapter using a variety of outlier robust estimation methods, which frequently are applied in the context of linear models. I concentrate on series for which the results from the standard and robust LM-type tests appeared to be in conflict. The similarities and differences between parameter estimates obtained with nonlinear least squares [NLS] and those obtained with the outlier-robust methods are used to infer which methods might be useful to obtain reasonable estimates of the parameters in a smooth transition model in the presence of aberrant observations. Next, Monte Carlo simulation is used to verify these conjectures.

The outline of this chapter is as follows. In Section 7.1, I extend the definition of additive and innovation outliers to STAR models. A simple example is used to demonstrate that the effects of such outliers on time series generated from a STAR model are similar to the effects on linear time series. Thus it might be expected that the effects on parameter estimates of STAR models also are more or less the same. In Section 7.2, outlier robust estimation methods for the parameters in a STAR model are derived, based on the principles of (G)M-estimators. Some pitfalls that are to be avoided and computational issues are discussed as well. In Section 7.3, I apply the robust methods to estimate STAR models for a number of industrial production series, for which the standard and robust linearity tests in Section 6.6 gave conflicting results, or for which least squares estimates of the specified STAR model appeared implausible. It is found that some of the robust estimation methods that are considered render more reasonable estimates, in the sense that the implied regimes in the model are more conform prior expectations. In Section 7.4, Monte Carlo experiments are performed to investigate the effects of outliers on the parameter estimates more rigorously. Finally, Section 7.5 concludes with some suggestions for further research.

7.1 Outliers in STAR models

It is straightforward to extend the definition of additive outliers [AOs] and innovation outliers [IOs], as discussed in Section 6.1 to the context of nonlinear time series models. Assume, for example, that the time series z_t follows a 2-regime STAR model

$$z_t = (\phi_{1,0} + \phi_{1,1}z_{t-1} + \cdots + \phi_{1,p}z_{t-p})\,(1 - G(s_t;\gamma,c)) \\ + (\phi_{2,0} + \phi_{2,1}z_{t-1} + \cdots + \phi_{2,p}z_{t-p})\,G(s_t;\gamma,c) + \eta_t, \quad (7.1)$$

where $\eta_t \sim$ i.i.d. $(0,\sigma_\eta^2)$. An additive outlier model for the observed time series y_t then can be defined as

$$y_t = z_t + \zeta\delta_t, \quad (7.2)$$

where the process δ_t is an i.i.d. random variable such that $P(\delta_t = 1) = P(\delta_t = -1) = \pi/2$ and $P(\delta_t = 0) = 1 - \pi$ for certain $0 < \pi < 1$, and ζ is a constant. Similarly, an innovation outlier model can be defined by assuming that y_t follows a STAR model, but is subject to occasionally large shocks, that is

$$y_t = \phi_1' x_t(1 - G(s_t;\gamma,c)) + \phi_2' x_t G(s_t;\gamma,c) + \eta_t + \zeta\delta_t, \quad (7.3)$$

where $\phi_i = (\phi_{i,0},\phi_{i,1},\ldots,\phi_{i,p})'$, $i = 1,2$, $x_t = (1, y_{t-1},\ldots,y_{t-p})'$, and δ_t and ζ as before.

The effects of AOs and IOs on, for example, nonlinear least squares [NLS] estimates of the parameters in the STAR model are difficult, if not impossible, to assess analytically. Some intuition of the likely effects can be obtained, however. Figure 7.1 depicts scatter-plots of time series y_t generated according to either the AO model (7.1)-(7.2) or the IO model (7.3). In both cases, the AR order p is set equal to 1, $\phi_{1,1} = -0.5$, $\phi_{2,1} = 0.5$, $G(s_t;\gamma,c)$ is taken to be a logistic function with $s_t = z_{t-1}$ (AO model) or y_{t-1} (IO model), $\gamma = 2.5$, and $c = 0.5$, and $\eta_t \sim$ i.i.d. $N(0,\sigma^2)$ with $\sigma = 1$. In panels (a) and (b), $\phi_{1,0} = -0.5$ and $\phi_{2,0} = 0.5$, whereas in panels (c) and (d), $\phi_{1,0} = 1.5$ and $\phi_{2,0} = -1.5$. A single outlier of magnitude $\zeta = 5\sigma$ occurs at

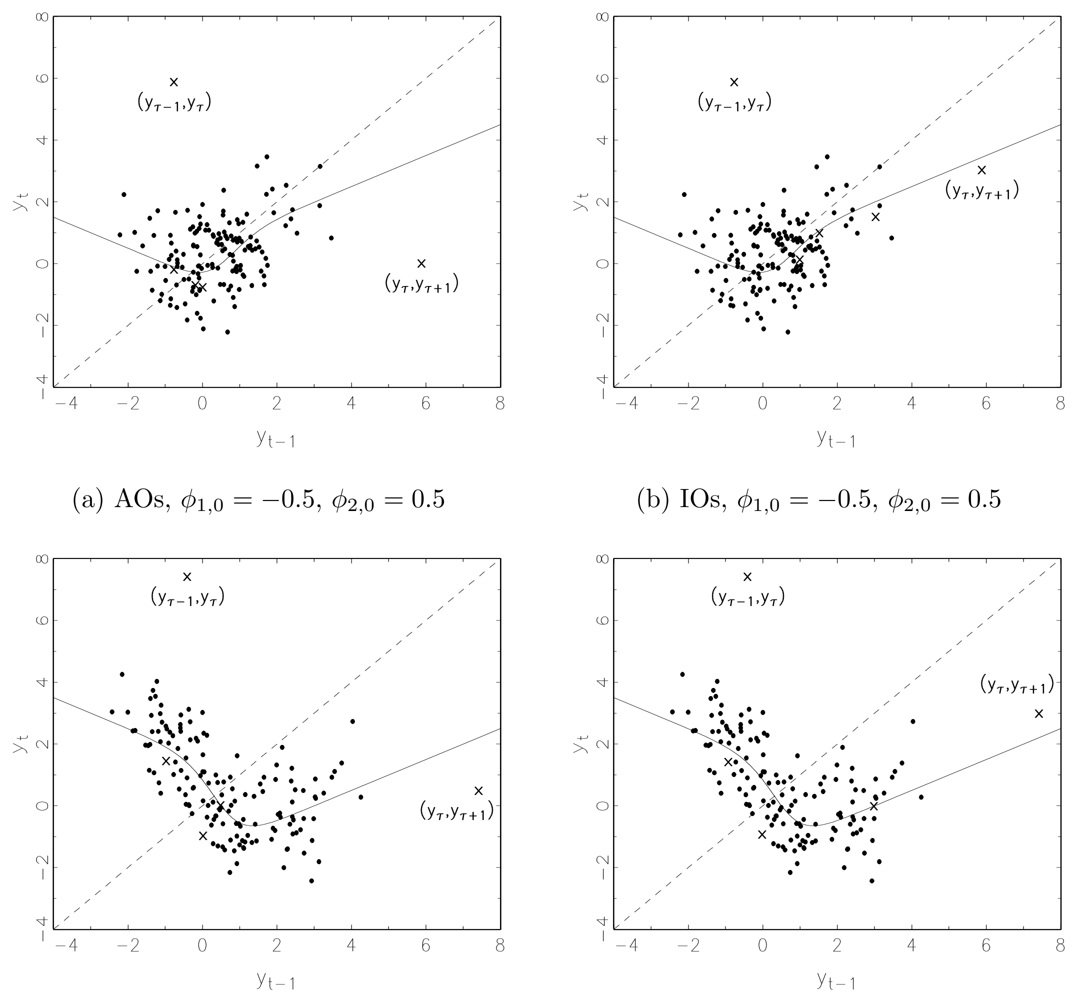
$t = \tau$ for which y_τ would have been larger than c if no outlier occurred. Thus, the time series would have been in the upper regime already at time $\tau + 1$, as the value of the transition function would have been larger than 0.5. By adding a positive outlier, the time series is 'pushed further into' the upper regime, in the sense that the value of $G(y_\tau; \gamma, c)$ is increased.

It is seen from Figure 7.1 that AOs and IOs show up in the same way as in linear models, compare Figure 6.1. At the time the outlier occurs, it gives rise to a vertical outlier, characterized by a large value of y_τ. In the AO case, this is followed by a bad leverage point at time $\tau + 1$, as the regressor y_τ is very different from the bulk of the data and the point $(y_{\tau+1}, y_\tau)$ is far removed from the skeleton of the STAR model. In the IO case, the vertical outlier is followed by a good leverage point. The value of the regressor y_τ falls outside the usual range in this case as well, but $(y_{\tau+1}, y_\tau)$ approximately satisfies the relationship implied by the STAR model. Due to the structure of the STAR model that is used here, the time series quickly returns to the main cloud of observations after $\tau + 1$, especially in panel (d). Obviously, it also is possible that an IO is followed by a longer sequence of good leverage points.

The similarity of the appearance of both outlier types in STAR and AR models suggests that they may have comparable effects on, for example, parameter estimates and test statistics in linear and nonlinear models. Intuitively, it might be expected that IOs will not harm NLS estimates of the parameters in the STAR model. The good leverage point(s) might even be beneficial. By contrast, AOs will have more serious effects on the NLS estimates. The exact influence is likely to depend on the sign and the magnitude of the outlier, and on the prevailing regime at the moment the outlier occurs. For example, in case a positive AO occurs when the time series is in the upper regime, as shown in panels (a) and (c) of Figure 7.1, it might be expected that the estimate of the AR parameter in the upper regime, $\phi_{2,1}$, will be driven towards zero. By analogy, negative outliers in the lower regime will affect the estimate of $\phi_{1,1}$. In case a negative (positive) outlier occurs in the upper (lower) regime, of such a magnitude that the implied regime is changed, it is more likely that the estimate of the AR parameter in the lower (upper) regime is affected.

In the above, it has been implicitly assumed that the outliers do not affect the estimate of the location parameter c, which determines the border between the two regimes in the STAR model. Another possible effect of AOs is, however, that the estimate of c changes such that the outliers constitute a separate regime. Whether or not this happens and whether the effect on the AR parameters or the effect on c dominates depends on the properties of the underlying STAR model. If the nonlinearity is not very pronounced as in panel (a) of Figure 7.1, the relationship between y_t and y_{t-1} for the main cloud of observations can be approximated reasonably well with a linear model. It might be expected that especially in such cases the change in the parameter c may be larger. By contrast, for the series in panel (c), it may be more difficult to approximate the nonlinear relationship with a linear model. In such cases, the effect on the AR parameter in the upper regime probably will be larger.

The Monte Carlo simulations to be presented in Section 7.4 confirm the intuitive discussion above, in that they demonstrate that the suggested effects of AOs on the



(a) AOs, $\phi_{1,0} = -0.5$, $\phi_{2,0} = 0.5$

(b) IOs, $\phi_{1,0} = -0.5$, $\phi_{2,0} = 0.5$

(c) AOs, $\phi_{1,0} = 1.5$, $\phi_{2,0} = -1.5$

(d) IOs, $\phi_{1,0} = 1.5$, $\phi_{2,0} = -1.5$

Figure 7.1: Examples of AOs (panels (a) and (c)) and IOs (panels (b) and (d)) occurring in STAR model. The time series y_t is generated either according to the AO model (7.1)-(7.2) or the IO model (7.3). In both cases, $p = 1$, $\phi_{1,1} = -0.5$, $\phi_{2,1} = 0.5$, $G(s_t; \gamma, c)$ is taken to be a logistic function with $s_t = z_{t-1}$ (AO model) or y_{t-1} (IO model), $\gamma = 2.5$, and $c = 0.5$, and $\eta_t \sim$ i.i.d. $N(0, \sigma^2)$ with $\sigma = 1$. A single outlier of magnitude 5σ occurs at $t = \tau$, for which y_τ would have been larger than c if no outlier occurred. The solid line is the skeleton of the STAR model.

NLS parameter estimates indeed can be observed in practice. In the next section, I derive an outlier robust estimation method for the parameters in STAR model, based on the principle of generalized maximum likelihood type estimation [GM]. See Chan and Cheung (1994) and Gabr (1998) for applications of GM estimators to SETAR and bilinear models, respectively.

7.2 Robust estimation of STAR models

The robust estimation methods discussed in Section 6.2 in the context of linear AR models can be extended to STAR models in a straightforward manner. To retain the analogy with Section 6.2, I will discuss the methods for a STAR model consisting of AR(1) models in both regimes and with transition variable $s_t = y_{t-1}$, that is,

$$y_t = (\phi_{1,0} + \phi_{1,1}y_{t-1})(1 - G(y_{t-1};\gamma,c)) + (\phi_{2,0} + \phi_{2,1}y_{t-1})G(y_{t-1};\gamma,c) + \varepsilon_t. \quad (7.4)$$

In the following, θ denotes the vector of parameters in the model, that is $\theta = (\phi_{1,0},\phi_{1,1},\phi_{2,0},\phi_{2,1},\gamma,c)'$. The NLS estimator, defined by (2.33), can alternatively be characterized by the first-order conditions

$$\sum_{t=1}^{T} \nabla F_t \cdot (y_t - F(y_{t-1};\theta)) = 0, \quad (7.5)$$

where $F(y_{t-1};\theta)$ the skeleton of the model, that is,

$$F(y_{t-1};\theta) = (\phi_{1,0} + \phi_{1,1}y_{t-1})(1 - G(y_{t-1};\gamma,c)) + (\phi_{2,0} + \phi_{2,1}y_{t-1})G(y_{t-1};\gamma,c),$$

and ∇F_t is the (6×1) vector of partial derivates $\nabla F_t = \partial F(y_{t-1};\theta)/\partial\theta$. A GM estimator for θ now can be defined as

$$\sum_{t=1}^{T} w\left(r_t;\nabla F_t\right)\nabla F_t \cdot (y_t - F(y_{t-1};\theta)) = 0, \quad (7.6)$$

where $w(\cdot,\cdot)$ is a weight function, $r_t = (y_t - F(y_{t-1};\theta))/\sigma_\varepsilon$ denotes the standardized residual, where σ_ε is a measure of scale of the residuals $\varepsilon_t = y_t - F(y_{t-1};\theta)$.

The Mallows version of the GM estimator is obtained by defining

$$w(r_t;\nabla F_t) = w_r(r_t)w_F(\nabla F_t), \quad (7.7)$$

with $w_r(r_t) = \psi(r_t)/r_t$ for $r_t \neq 0$, and $w_r(r_t) = 1$ for $r_t = 0$ (and thus $w(r_t;\nabla F_t) = w_F(\nabla F_t)$). The weight function $w_r(r_t)$ can be based on, for example, the Huber ψ function as given in (6.13) or Tukey's bisquare function (6.14). The weight function for $\nabla F_t(\theta)$ can be defined analogously to (6.17) as

$$w_F(\nabla F_t) = \psi(d(\nabla F_t)^\alpha)/d(\nabla F_t)^\alpha, \quad (7.8)$$

where $d(\nabla F_t)$ is the Mahalanobis distance

$$d(\nabla F_t) = \sqrt{(\nabla q_t - M_F)'\,\Sigma_F^{-1}\,(\nabla F_t - M_F)}, \quad (7.9)$$

with M_F and Σ_F measures of the location and scale of ∇F_t, respectively. To ensure robustness of the GM-estimator, robust estimates of M_F and Σ_F should be used. At first sight, it is tempting to use the minimum volume ellipsoid [MVE] estimator of Rousseeuw (1985) for this purpose, analogous to the linear model discussed in the previous chapter. Some reflection however reveals that the MVE estimator can not be employed in the present context of the STAR model. The MVE estimator looks for the ellipsoid with the smallest volume covering at least half of the observations ∇F_t, $t = 1, \ldots, T$. The center of the ellipsoid is taken as an estimate of location M_F, while an estimate of the scatter matrix Σ_F is based upon the metric matrix defining the ellipsoid. For the STAR model as given in (7.4), the vector ∇F_t takes the form

$$\nabla F_t = \begin{pmatrix} 1 - G(y_{t-1}) \\ (1 - G(y_{t-1}))y_{t-1} \\ G(y_{t-1}) \\ G(y_{t-1})y_{t-1} \\ (\phi_{2,1} - \phi_{1,1})y_{t-1}\frac{\partial G(y_{t-1})}{\partial \gamma} \\ (\phi_{2,1} - \phi_{1,1})y_{t-1}\frac{\partial G(y_{t-1})}{\partial c} \end{pmatrix}, \tag{7.10}$$

where $G(y_{t-1}) \equiv G(y_{t-1}; \gamma, c)$. In case the transition function is the logistic function $G(y_{t-1}) = (1 + \exp\{-\gamma(y_{t-1} - c)\})^{-1}$, it follows that

$$\frac{\partial G(y_{t-1})}{\partial \gamma} = (1 - G(y_{t-1}))G(y_{t-1})(y_{t-1} - c),$$
$$\frac{\partial G(y_{t-1})}{\partial \gamma} = -(1 - G(y_{t-1}))G(y_{t-1})\gamma.$$

Notice that if $G(y_{t-1}) = 0$ the score is equal to $\nabla F_t = (1, y_{t-1}, 0, 0, 0, 0)$, whereas $\nabla F_t = (0, 0, 1, y_{t-1}, 0, 0)$ if $G(y_{t-1}) = 1$. It should now be intuitively clear that the MVE estimator collapses if there are only few observations located in between the two extreme regimes, that is, observations for which $0 < G(y_{t-1}) < 1$, as the scale estimate for the final two elements of ∇F_t becomes equal to 0 in this case.

An alternative would be to use low breakdown estimators for multivariate location and scatter. For example, Maronna (1976) suggests to estimate M_F and Σ_F by solving

$$\frac{1}{T}\sum_{t=1}^{T} v_1(d_t)(\nabla F_t - M_F) = 0, \tag{7.11}$$

$$\frac{1}{T}\sum_{t=1}^{T} v_2(d_t)(\nabla F_t - M_F)(\nabla F_t - M_F)' = \Sigma_F, \tag{7.12}$$

where $d_t = d(\nabla F_t)$ is the Mahalanobis distance defined in (7.9) and v_1 and v_2 are weight functions, see also Hampel *et al.* (1986) and Lucas (1996) for discussion. In this chapter, I do not consider this possibility and restrict attention to M-estimators, which are obtained by setting $w_F(\nabla F_t) = 1$ in (7.7).

The first-order conditions in (7.6) show that the M-estimates can be obtained by an iterative weighted least squares algorithm. Denoting the estimates of θ and

σ_ε at the k-th iteration by $\hat{\theta}^{(k)}$ and $\hat{\sigma}_\varepsilon^{(k)}$, respectively, it follows that $\hat{\theta}^{(k+1)}$ can be computed as

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} + \left(\sum_{t=1}^{T} w_r(\hat{r}_t^{(k)}) \nabla \hat{F}_t \nabla \hat{F}_t' \right)^{-1} \sum_{t=1}^{T} w_r(\hat{r}_t^{(k)}) \nabla \hat{F}_t \cdot \hat{\varepsilon}_t^{(k)}, \tag{7.13}$$

for $k = 0, 1, 2, \ldots$, where $\hat{\varepsilon}_t^{(k)} = y_t - F(y_{t-1}; \hat{\theta}^{(k)})$, $\hat{r}_t^{(k)} = \hat{\varepsilon}_t^{(k)} / \hat{\sigma}_\varepsilon^{(k)}$, and $\nabla \hat{F}_t = \nabla F_t(\hat{\theta}^{(k)})$. To obtain an estimate of σ_ε, the scale of ε_t, I use the median absolute deviation [MAD], that is, $\hat{\sigma}_\varepsilon^{(k)} = 1.483 \cdot \text{med}|\hat{\varepsilon}_t^{(k)}) - \hat{m}_\varepsilon^{(k)}|$ with $\hat{m}_\varepsilon^{(k)} = \text{med}(\hat{\varepsilon}_t^{(k)})$.

Stromberg (1995) proves consistency of the least median of squares [LMS] estimator of Rousseeuw (1984) in nonlinear regression models, which suggests that LMS retains its high breakdown point [HBP] in nonlinear models. Consequently, a HBP-M estimator can be obtained by performing only a single iteration according to (7.13), using the LMS estimates as starting values $\hat{\theta}^{(0)}$.

Computing the LMS estimator can be very time-consuming already for linear models, see Rousseeuw and Leroy (1987). The situation is even worse for nonlinear models, as discussed by Stromberg (1993). Here I use the following two-step procedure to approximate the LMS estimator, making use of the fact that the model is linear in the autoregressive parameters for fixed values of the parameters in the transition function. First, estimates of $\phi_{i,j}$, $i = 1, 2$, $j = 0, 1$, are computed for a grid on γ and c, using an M-estimator. The grid-point (γ, c) and the corresponding estimates of the AR parameters for which the LMS function is minimized are used as starting values in the second step, in which the downhill simplex algorithm of Nelder and Mead (1965) is used to improve upon the LMS estimates. The simplex method does not require derivatives of the objective function that is to be minimized, which renders it a suitable algorithm for minimizing the LMS function. See Press *et al.* (1986, Section 10.4) for a description of the simplex method. In the grid search that is performed in the first step, γ is varied among $\gamma \in \{.5, 1, 2, 3, \ldots, 10, 25, 50\}$, whereas c is varied among the 5-th, 10-th,..., 95-th percentiles of the ordered transition variable. This grid certainly is not exhaustive, but it might be expected to provide reasonable starting values for the second step.

In the empirical analysis in the next section, I compare the NLS and HBP-M estimates of STAR models for a number of the industrial production series considered previously. In addition, I report results for the LMS estimator (computed as outlined above), the least absolute deviation [LAD] estimator, which minimizes the sum of absolute residuals, a pseudo maximum likelihood estimator based on a Student t distribution with 5 degrees of freedom, and an M-estimator which uses the LAD estimates as starting values. For the HBP-M and M estimators, the Huber ψ function as given in (6.13) is used, with the tuning constant c set equal to 1.345.

7.3 Nonlinearity in industrial production reconsidered

The application of the standard and robust linearity tests to the growth rates in industrial production in Section 6.6 rendered (partially) conflicting results for the majority of the series considered. In the most extreme case, the standard tests indicate that a STAR model might be appropriate whereas the robust tests are unable to reject linearity (Austria, Japan, Spain, Sweden, and Switzerland) or vice versa (Belgium and Canada). Even if the test results coincide completely, the NLS estimates of the preferred model can be inappropriate, as shown by, for example, the series for the US, see Section 6.6. In this section, I investigate whether these findings perhaps can be explained by the joint presence of STAR-type nonlinearity and aberrant observations.

First, I reconsider the series for the US. Both the standard and robust linearity tests indicate that an LSTAR model consisting of AR(2) models in the two regimes and delay parameter $d = 4$ might be an appropriate model for the yearly growth rate in industrial production, see Table 6.5. The NLS estimates of this model are not very plausible though, as the regime corresponding to $G(y_{t-d}; \gamma, c) = 0$ becomes active only for very few observations, see Figure 6.6, which shows that only for 5 observations the value of the transition function is smaller than 0.5. The NLS parameter estimates are shown in Table 7.1, together with estimates that are obtained with the various robust methods. The NLS estimates of the AR parameters in the lower regime, $\phi_{1,j}$, $j = 0, 1, 2$, are not reliable as they are based on a very limited sample size. It can be seen from the table that the LAD, Student $t(5)$ and Huber M estimates are roughly similar to the NLS estimates. By contrast, the LMS and HBP-M estimates are markedly different. In particular, the estimate of the location parameter c, which marks the border between the two regimes is much larger. Figure 7.2 depicts the estimated transition function from the HBP-M procedure, together with the weights that are assigned to the observations. It is seen that the lower regime is activated during the aftermath of all the recessions that occurred during the sample period. In fact, the transition function takes on small values

Table 7.1: Estimates of LSTAR model for yearly growth rates in US industrial production

Method	$\phi_{1,0}$	$\phi_{1,1}$	$\phi_{1,2}$	$\phi_{2,0}$	$\phi_{2,1}$	$\phi_{2,2}$	γ	c
NLS	0.10	0.08	−0.07	−0.00	1.32	−0.36	1.89	−0.071
LAD	0.10	0.20	−0.08	−0.00	1.34	−0.40	1.81	−0.073
$t(5)$	0.11	−0.06	0.02	−0.00	1.32	−0.37	1.74	−0.078
LMS	0.01	1.37	−0.64	−0.01	1.40	−0.30	10.03	0.025
Huber M	0.11	−0.02	−0.01	−0.00	1.32	−0.37	1.77	−0.077
HBP-M	0.01	1.34	−0.59	−0.00	1.30	−0.32	11.56	0.019

Estimates of the LSTAR model with AR order 2 and delay parameter $d = 4$ for yearly growth rates in US industrial production. The estimation methods are explained in the text.

Table 7.2: Estimates of LSTAR model for yearly growth rates in industrial production

	Austria		Belgium		Canada	
Method	γ	c	γ	c	γ	c
LS	1.54	0.103	2.65	−0.046	100.00	0.080
LAD	33.03	0.024	1.77	−0.073	23.43	0.054
$t(5)$	1.54	0.103	2.89	−0.050	100.00	0.080
LMS	2.03	0.071	3.38	−0.033	4.16	0.037
Huber M	1.54	0.103	2.88	−0.049	100.00	0.080
HBP-M	0.45	0.103	1.91	−0.050	9.42	0.053

Estimates of the LSTAR model for yearly growth rates in industrial production for Austria, Belgium and Canada. The estimation methods are explained in the text.

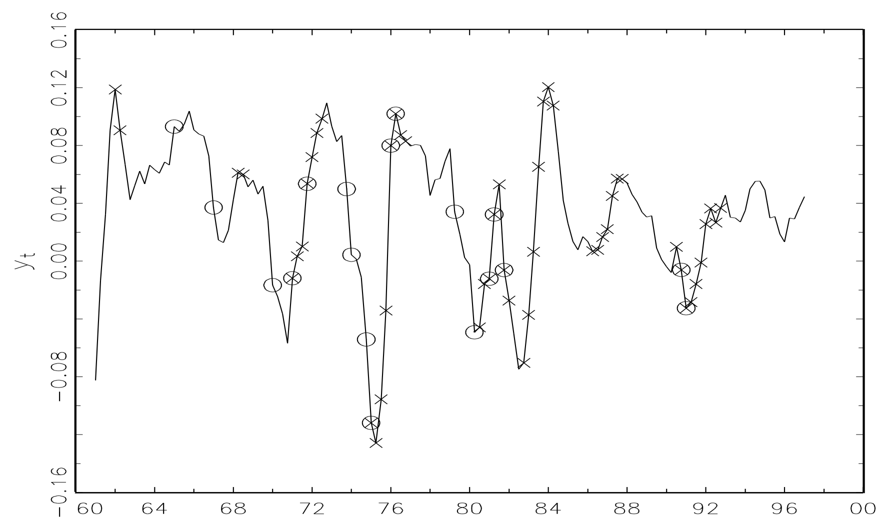
during the period 1986:2-1988:1 as well. Even though no recession occurred at this time, growth slowed down considerably and fell below the estimated value of c. The observations that receive weight (close to) zero are located around the points where transitions from one regime to the other regime in the STAR models occur. These regime shifts, which are seen to be very quick, are caused by the transition variable taking values opposite from the threshold c at consecutive points in time. Apparently, this coincides with y_{t-d} taking rather extreme values, which results in vertical outliers or bad leverage points, as suggested by the zero weights.

To illustrate that robust estimation methods can yield rather different estimates of the parameters in the transition function, consider the results in Table 7.2, which contains estimates of these parameters for Austria, Belgium and Canada. The AR orders and the delay parameter d are taken from Table 6.5.

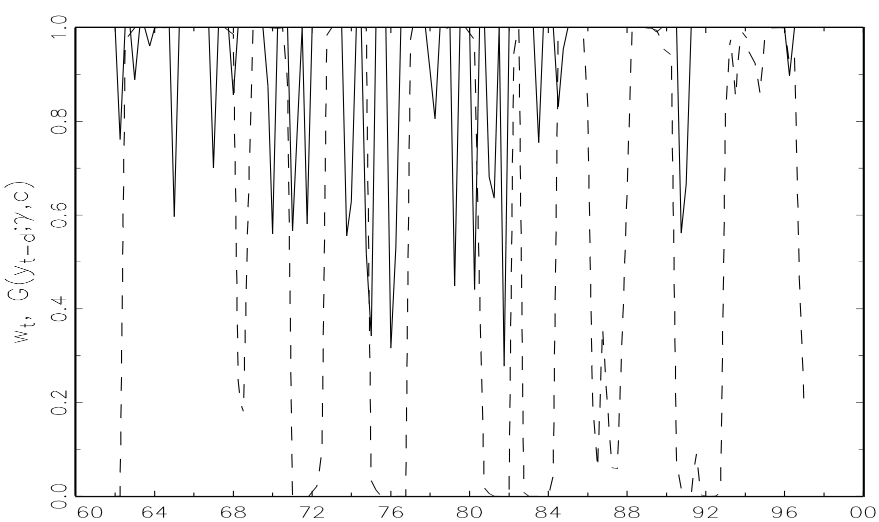
It was noted in Section 6.6 that for Austria the NLS estimation results are heavily influenced by the exceptionally large observation in 1972:4. In particular, the estimate of c becomes such that the upper regime essentially consists of this observation only. The results in Table 7.2 show that the estimates of c from the Student $t(5)$, Huber M and HBP-M methods are identical to the NLS estimates. The LMS and, especially, LAD estimates of c are considerably smaller, such that the two regimes in the LSTAR model correspond much closer with recessions and expansions.

For Canada a similar effect can be observed. The NLS, Student $t(5)$ and Huber M estimates of c are close to the maximum value that is attained by the transition variable y_{t-2}, such that the upper regime contains only few observations. The LAD, LMS and HBP-M estimates are smaller, such that the observations are more evenly distributed across the two regimes.

Finally, the results for Belgium are not so clear-cut, as the estimates of γ and c are comparable for all estimation methods. Anyhow, these empirical results suggest that matters may improve when using robust estimation methods, if there are indications for the presence of outliers.



(a) Yearly growth rates



(b) Weights and transition function

Figure 7.2: Industrial production - United States. The upper graph contains observations on the seasonal difference of the logarithm of the index of US industrial production. Observations that receive a weight smaller than 0.75 in the HBP-M estimation procedure of the LSTAR model are marked with a circle. Observations for which the value of the transition function in the fitted model, $G(y_{t-d};\gamma,c) = (1+\exp\{-\gamma(y_{t-d}-c)\})^{-1}$, $d = 4, \gamma = 11.56 \times 21, c = 0.019$, is smaller than 0.5 are marked with a cross. The lower graph contains the actual weights w_t (solid line) and the values taken by the transition function (dashed line).

7.4 Some simulation results

The empirical results in the previous section are quite suggestive about the properties of the robust estimation methods that are considered here. For example, the pseudo maximum likelihood estimator based on a Student $t(5)$ distribution and the M estimator with Huber ψ function yield estimates that are very similar to the NLS estimates in all cases. Hence, it might be expected that these estimation methods do not offer much protection against the influence of aberrant observations. The LAD, HBP-M estimators, and particularly the LMS estimator, might be more fit for this purpose. At least for the empirical examples, the parameter estimates from these methods are different from the NLS estimates, and in a sense more reasonable as they imply more plausible regimes in the STAR models. In this section, I examine whether the LMS estimator indeed is robust to the presence of contamination in observed time series that are generated from STAR models. In particular, I compare the performance of the NLS and LMS estimators in the presence of additive outliers.

The AO model given in (7.1)-(7.2) is used as data generating process [DGP]. The clean series z_t is generated according to a STAR model with AR order $p = 1$, with $\phi_{1,0} = -0.50$, $\phi_{1,1} = -0.50$, $\phi_{2,0} = 0.50$, and $\phi_{2,1} = 0.50$. The transition function $G(s_t;\gamma,c)$ is taken to be logistic function with $s_t = z_{t-1}$, $\gamma = 2.5$ and $c = 0.50$. Finally, the shocks ε_t are i.i.d. $N(0,\sigma^2)$ with $\sigma = 1$. The sample size is set equal to $T = 150$, which approximately corresponds with the length of the empirical time series considered in the previous section. To obtain the contaminated series y_t, AOs are not added randomly, as in (7.2), but in a more strategic fashion to obtain a better impression of their effects on the parameter estimates. To be precise, k outliers of absolute magnitude ζ occur, all being of the same sign, and all occurring at times τ for which y_τ would have been either larger or smaller than c if no outliers occurred. Thus, I consider four different contamination processes, to be summarized as 'negative/positive outliers in the lower/upper regime'. Notice that positive outlier in the lower regime and negative outliers in the upper regime might lead to regime-switches, in the sense that the clean series would have been in the one regime but the contaminated series is in the other. The number of outliers is set equal to $k = 1$, 3 and 5, whereas the absolute magnitude of the AOs is set equal to $\zeta = 3$, 5 and 7. Results based on 500 replications appear in Tables 7.3 to 7.6 for the different outlier configurations. For comparative purposes, Table 7.3 also contains results from estimating STAR models on the clean series z_t.

The simulation results in Table 7.3 suggest the following conclusions. For all values of k and ζ, the NLS method yields inaccurate estimates of γ, whereas the LMS method gives relatively precise estimates of this parameter. Furthermore, for all outlier scenarios, the LMS estimates of c are closer to the true value than those obtained using NLS. It must be stressed though that the absolute accuracy of LMS concerning this parameter decreases substantially as the number and/or the absolute magnitude of the outliers increases. Similar conclusions can be drawn for the AR parameters in the two regimes. Both NLS and LMS are heavily affected by outliers in the lower regime, where, as expected, the estimate of the AR parameter $\phi_{1,1}$ is

Table 7.3: Monte Carlo medians (and MADs in parentheses) for NLS and LMS estimators of the LSTAR model - negative outliers in the lower regime

k	ζ	$\phi_{1,0}$	$\phi_{1,1}$	$\phi_{2,0}$	$\phi_{2,1}$	γ	c
				NLS			
0	0	−0.30(0.46)	−0.36(0.39)	0.13(0.78)	0.59(0.36)	3.64(3.64)	0.54(1.06)
1	3	−0.17(0.46)	−0.18(0.33)	−0.10(0.48)	0.67(0.26)	5.84(5.84)	0.25(0.98)
	5	−0.01(0.54)	−0.07(0.30)	−0.18(0.43)	0.71(0.24)	7.19(7.19)	−0.07(1.16)
	7	0.03(0.48)	−0.02(0.20)	−0.19(0.42)	0.72(0.25)	10.20(10.20)	−0.34(1.24)
3	3	−0.14(0.58)	−0.08(0.33)	−0.17(0.44)	0.71(0.25)	5.21(5.21)	0.18(1.10)
	5	−0.03(0.55)	−0.03(0.23)	−0.24(0.42)	0.74(0.25)	4.95(4.95)	−0.33(1.11)
	7	0.01(0.51)	−0.01(0.15)	−0.32(0.38)	0.78(0.23)	10.17(10.17)	−0.41(1.04)
5	3	−0.20(0.39)	−0.09(0.24)	−0.17(0.51)	0.71(0.27)	6.93(6.93)	0.22(1.02)
	5	−0.20(0.43)	−0.03(0.17)	−0.25(0.51)	0.72(0.28)	10.42(10.42)	−0.10(1.10)
	7	−0.21(0.52)	−0.01(0.13)	−0.35(0.48)	0.77(0.28)	33.58(33.58)	−0.06(1.17)
				LMS			
0	0	−0.33(0.36)	−0.40(0.36)	0.89(1.10)	0.34(0.43)	3.95(1.44)	0.61(0.65)
1	3	−0.21(0.41)	−0.27(0.27)	0.51(0.98)	0.52(0.38)	2.77(1.53)	0.53(0.74)
	5	−0.04(0.45)	−0.07(0.23)	−0.16(0.71)	0.72(0.35)	3.00(1.25)	0.27(0.86)
	7	0.21(0.54)	0.00(0.16)	−0.34(0.68)	0.82(0.32)	2.89(1.20)	0.05(0.77)
3	3	−0.10(0.46)	−0.10(0.23)	−0.06(0.72)	0.67(0.33)	3.24(1.29)	0.27(0.90)
	5	0.06(0.55)	−0.01(0.17)	−0.39(0.59)	0.83(0.30)	2.52(1.19)	0.09(0.82)
	7	0.07(0.52)	0.01(0.14)	−0.47(0.66)	0.89(0.32)	2.59(1.04)	0.13(0.81)
5	3	−0.10(0.53)	−0.07(0.25)	−0.18(0.74)	0.73(0.37)	2.84(1.18)	0.23(0.89)
	5	0.02(0.57)	0.02(0.17)	−0.48(0.66)	0.88(0.33)	2.87(1.08)	0.02(0.79)
	7	0.04(0.62)	0.03(0.13)	−0.61(0.67)	0.95(0.35)	2.43(0.99)	0.04(0.79)

Median and median absolute deviation [MAD] (in parentheses) of NLS and LMS estimates of the parameters in the STAR model (7.4) for the observed series y_t, in the presence of AOs. Series z_t are generated from (7.1) with $p = 1$, $\phi_{1,0} = -0.50$, $\phi_{1,1} = -0.50$, $\phi_{2,0} = 0.50$, and $\phi_{2,1} = 0.50$. The transition function $G(s_t; \gamma, c)$ is taken to be logistic function with $s_t = z_{t-1}$, $\gamma = 2.5$ and $c = 0.50$. The shocks ε_t are i.i.d. $N(0,1)$. The series y_t is obtained by adding k AOs of magnitude $-\zeta$ to observations at $t = \tau$ such that y_τ would have been smaller than c if no outlier occurred. The Table is based on 500 replications.

Table 7.4: Monte Carlo medians (and MADs in parentheses) for NLS and LMS estimators of the LSTAR model - positive outliers in the lower regime

k	ζ	$\phi_{1,0}$	$\phi_{1,1}$	$\phi_{2,0}$	$\phi_{2,1}$	γ	c
				NLS			
1	3	−0.31(0.48)	−0.40(0.37)	0.41(0.96)	0.41(0.46)	3.07(3.07)	0.59(1.00)
	5	−0.62(0.66)	−0.60(0.43)	3.61(3.51)	−0.70(1.09)	1.82(1.45)	1.13(0.69)
	7	−0.62(0.56)	−0.62(0.41)	3.58(2.26)	−0.58(0.48)	1.73(1.02)	1.13(0.48)
3	3	−0.23(0.53)	−0.38(0.38)	0.64(1.25)	0.28(0.54)	3.12(3.04)	0.68(1.03)
	5	−0.52(0.49)	−0.55(0.38)	3.97(2.90)	−0.79(0.74)	1.83(1.03)	1.31(0.63)
	7	−0.34(0.54)	−0.50(0.39)	3.03(1.90)	−0.45(0.35)	1.96(1.25)	1.16(0.56)
5	3	−0.16(0.49)	−0.33(0.42)	0.96(1.15)	0.16(0.51)	3.78(3.77)	0.70(0.85)
	5	−0.24(0.58)	−0.43(0.43)	3.52(2.57)	−0.69(0.65)	2.03(1.25)	1.32(0.62)
	7	−0.08(0.61)	−0.36(0.52)	2.76(1.83)	−0.41(0.36)	2.14(1.98)	1.19(0.71)
				LMS			
1	3	−0.39(0.38)	−0.43(0.30)	1.15(1.31)	0.22(0.48)	2.56(1.72)	0.54(0.63)
	5	−0.53(0.49)	−0.60(0.42)	2.03(1.84)	−0.17(0.62)	2.19(1.43)	0.68(0.64)
	7	−0.80(0.78)	−0.78(0.54)	2.77(2.13)	−0.38(0.49)	1.97(1.36)	0.72(0.60)
3	3	−0.34(0.36)	−0.44(0.31)	1.09(1.17)	0.17(0.46)	2.41(1.56)	0.56(0.66)
	5	−0.59(0.59)	−0.66(0.46)	3.06(2.05)	−0.54(0.56)	2.01(1.27)	0.85(0.55)
	7	−0.73(0.79)	−0.77(0.55)	3.30(1.89)	−0.51(0.38)	1.75(1.10)	0.95(0.48)
5	3	−0.35(0.41)	−0.45(0.31)	1.46(1.28)	−0.07(0.50)	2.99(1.34)	0.59(0.63)
	5	−0.68(0.76)	−0.75(0.56)	3.55(2.11)	−0.66(0.51)	1.87(1.15)	0.92(0.52)
	7	−0.81(0.90)	−0.82(0.56)	3.35(1.81)	−0.49(0.32)	1.60(0.98)	0.85(0.55)

Median and median absolute deviation [MAD] (in parentheses) of NLS and LMS estimates of the parameters in the STAR model (7.4) for the observed series y_t, in the presence of AOs. Series z_t are generated from (7.1) with $p = 1$, $\phi_{1,0} = -0.50$, $\phi_{1,1} = -0.50$, $\phi_{2,0} = 0.50$, and $\phi_{2,1} = 0.50$. The transition function $G(s_t; \gamma, c)$ is taken to be logistic function with $s_t = z_{t-1}$, $\gamma = 2.5$ and $c = 0.50$. The shocks ε_t are i.i.d. $N(0, 1)$. The series y_t is obtained by adding k AOs of magnitude ζ to observations at $t = \tau$ such that y_τ would have been smaller than c if no outlier occurred. The Table is based on 500 replications.

Table 7.5: Monte Carlo medians (and MADs in parentheses) for NLS and LMS estimators of the LSTAR model - negative outliers in the upper regime

k	ζ	$\phi_{1,0}$	$\phi_{1,1}$	$\phi_{2,0}$	$\phi_{2,1}$	γ	c
				NLS			
1	3	−0.34(2.33)	−0.38(0.99)	0.25(1.85)	0.42(0.65)	2.09(2.01)	0.64(1.76)
	5	−0.25(1.59)	−0.26(0.70)	0.22(2.56)	0.43(0.80)	2.17(2.09)	0.59(1.73)
	7	−0.14(0.96)	−0.15(0.43)	0.34(2.77)	0.45(0.96)	2.16(2.08)	0.72(1.71)
3	3	−0.29(3.91)	−0.27(1.63)	0.20(3.07)	0.40(0.98)	2.24(2.16)	0.64(1.79)
	5	0.07(4.01)	0.01(1.09)	−0.01(4.58)	0.52(1.64)	2.04(1.96)	0.65(1.88)
	7	−0.02(2.44)	−0.03(0.51)	0.10(6.93)	0.40(1.95)	2.22(2.15)	0.64(1.91)
5	3	−0.24(3.64)	−0.28(1.45)	0.15(2.61)	0.41(0.76)	1.93(1.85)	0.44(1.92)
	5	0.23(7.01)	0.15(1.57)	−0.21(6.00)	0.59(2.15)	1.76(1.69)	0.22(2.22)
	7	−0.02(10.59)	0.02(1.58)	−0.05(8.89)	0.38(2.69)	1.69(1.63)	−0.81(3.33)
				LMS			
1	3	−0.25(0.27)	−0.33(0.25)	0.15(0.49)	0.55(0.25)	5.68(4.80)	0.51(0.64)
	5	−0.19(0.26)	−0.27(0.21)	0.04(0.40)	0.57(0.23)	6.30(4.62)	0.49(0.66)
	7	−0.09(0.26)	−0.18(0.18)	−0.14(0.37)	0.67(0.23)	6.01(4.67)	0.21(0.82)
3	3	−0.24(0.27)	−0.30(0.25)	0.01(0.41)	0.55(0.27)	5.77(4.79)	0.39(0.68)
	5	−0.20(0.24)	−0.22(0.18)	−0.10(0.45)	0.61(0.28)	5.49(4.20)	0.39(0.74)
	7	−0.12(0.27)	−0.14(0.13)	−0.22(0.43)	0.64(0.26)	5.02(4.20)	0.29(0.73)
5	3	−0.22(0.27)	−0.27(0.24)	−0.02(0.44)	0.57(0.26)	6.35(4.39)	0.36(0.69)
	5	−0.18(0.26)	−0.18(0.18)	−0.19(0.40)	0.62(0.27)	6.59(4.05)	0.32(0.68)
	7	−0.19(0.24)	−0.12(0.12)	−0.25(0.51)	0.62(0.33)	5.52(3.35)	0.19(0.77)

Median and median absolute deviation [MAD] (in parentheses) of NLS and LMS estimates of the parameters in the STAR model (7.4) for the observed series y_t, in the presence of AOs. Series z_t are generated from (7.1) with $p = 1$, $\phi_{1,0} = -0.50$, $\phi_{1,1} = -0.50$, $\phi_{2,0} = 0.50$, and $\phi_{2,1} = 0.50$. The transition function $G(s_t; \gamma, c)$ is taken to be logistic function with $s_t = z_{t-1}$, $\gamma = 2.5$ and $c = 0.50$. The shocks ε_t are i.i.d. $N(0, 1)$. The series y_t is obtained by adding k AOs of magnitude $-\zeta$ to observations at $t = \tau$ such that y_τ would have been larger than c if no outlier occurred. The Table is based on 500 replications.

Table 7.6: Monte Carlo medians (and MADs in parentheses) for NLS and LMS estimators of the LSTAR model - positive outliers in the upper regime

k	ζ	$\phi_{1,0}$	$\phi_{1,1}$	$\phi_{2,0}$	$\phi_{2,1}$	γ	c
				NLS			
1	3	−0.63(0.60)	−0.59(0.43)	2.45(2.31)	−0.27(0.69)	1.99(1.52)	0.93(0.65)
	5	−0.68(0.54)	−0.63(0.37)	2.72(1.79)	−0.27(0.43)	1.86(1.09)	1.05(0.43)
	7	−0.63(0.50)	−0.60(0.36)	2.54(1.27)	−0.20(0.25)	1.89(1.08)	0.97(0.36)
3	3	−0.51(0.50)	−0.51(0.39)	2.16(1.97)	−0.19(0.54)	2.04(1.38)	1.04(0.48)
	5	−0.46(0.60)	−0.54(0.45)	2.36(1.57)	−0.21(0.33)	2.07(1.34)	1.01(0.40)
	7	−0.38(0.54)	−0.52(0.42)	2.26(1.28)	−0.15(0.22)	2.35(1.88)	0.91(0.46)
5	3	−0.42(0.51)	−0.50(0.42)	1.77(1.50)	−0.06(0.40)	2.64(1.66)	0.90(0.44)
	5	−0.31(0.62)	−0.46(0.57)	1.82(1.65)	−0.06(0.39)	2.79(2.73)	0.94(0.56)
	7	−0.23(1.04)	−0.45(0.90)	1.86(1.71)	−0.06(0.32)	2.73(2.67)	0.93(1.04)
				LMS			
1	3	−0.30(0.33)	−0.42(0.31)	0.82(0.95)	0.29(0.37)	5.10(4.25)	0.63(0.61)
	5	−0.47(0.48)	−0.51(0.45)	1.51(1.43)	0.02(0.43)	4.05(3.34)	0.69(0.55)
	7	−0.47(0.51)	−0.56(0.47)	1.85(1.47)	−0.07(0.35)	3.68(2.98)	0.72(0.58)
3	3	−0.30(0.35)	−0.44(0.36)	1.33(1.27)	0.06(0.40)	4.88(4.09)	0.69(0.61)
	5	−0.46(0.51)	−0.56(0.48)	1.79(1.27)	−0.09(0.34)	3.68(3.02)	0.72(0.57)
	7	−0.49(0.56)	−0.60(0.51)	2.00(1.25)	−0.12(0.26)	3.27(2.68)	0.72(0.57)
5	3	−0.45(0.47)	−0.54(0.45)	1.47(1.32)	−0.01(0.41)	4.36(3.63)	0.60(0.59)
	5	−0.52(0.61)	−0.60(0.52)	2.14(1.36)	−0.16(0.32)	3.36(2.75)	0.71(0.59)
	7	−0.59(0.68)	−0.69(0.58)	2.03(1.17)	−0.13(0.23)	2.98(2.43)	0.66(0.59)

Median and median absolute deviation [MAD] (in parentheses) of NLS and LMS estimates of the parameters in the STAR model (7.4) for the observed series y_t, in the presence of AOs. Series z_t are generated from (7.1) with $p = 1$, $\phi_{1,0} = -0.50$, $\phi_{1,1} = -0.50$, $\phi_{2,0} = 0.50$, and $\phi_{2,1} = 0.50$. The transition function $G(s_t; \gamma, c)$ is taken to be logistic function with $s_t = z_{t-1}$, $\gamma = 2.5$ and $c = 0.50$. The shocks ε_t are i.i.d. $N(0, 1)$. The series y_t is obtained by adding k AOs of magnitude ζ to observations at $t = \tau$ such that y_τ would have been larger than c if no outlier occurred. The Table is based on 500 replications.

biased towards zero when the size and number of outliers increase. Only when $k = 1$ and $\zeta = 3$, the LMS method seems reliable.

Table 7.4 contains results for positive outliers occurring in the lower regime. It is seen that this type of contamination causes an upward bias in the estimate of c. Furthermore, the estimate of the autoregressive parameter in the upper regime is biased downward, whereas the estimates of the AR parameter in the lower regime is affected to a much lesser extent. This was to be expected, as the magnitude of the outliers ζ is such that the value of y_τ is likely to become larger than c and hence, the observation at $t = \tau + 1$ is shifted from the lower to the upper regime.

Results for negative outliers occurring in the upper regime are shown in Table 7.5. In this case, the estimate of c is biased downward, although for NLS this only is observed for $k = 5$. The estimate of $\phi_{2,1}$ is not affected, whereas the estimate of $\phi_{1,1}$ is driven towards zero as k and ζ increase. Again, these results make sense intuitively, as this type of contamination shifts the observations in the period following the occurrence of an outlier from the upper to the lower regime.

Finally, Table 7.6 summarizes the NLS and LMS estimates in the presence of positive outliers in the upper regime. This is the mirror-image of the case considered in Table 7.3, and comparing the results in these two tables shows that the parameter estimates are affected accordingly. Here, the estimate of c is biased upward, the AR parameter in the upper regime is biased towards zero, and the the AR parameter in the lower regime is not affected.

7.5 Concluding remarks

In this chapter I have made a modest attempt to examine whether several standard robust estimation methods, which often appear to work well for linear time series models, can be used to yield reliable estimates of STAR models in the presence of outliers. When considered for various industrial production series, I obtain evidence that some of these methods may turn out to be useful. Upon evaluating the subsequent Monte Carlo simulations, however, this apparent usefulness may emerge because of the possible fact that these series may suffer from occasional outliers, but that these outliers are not very large and do not appear very frequently. Indeed, the simulations indicated that robust estimation methods also break down when there are too many too large outliers.

In the end, this means that I conclude my thesis with a clear-cut outline of a potentially fruitful area for further research. It seems that the statistical theory for robust estimation of nonlinear time series models still has to be developed. Also, empirical experience with the methods to be developed should be obtained, by Monte Carlo simulation and by application to real-world data. In sum, it seems that these issues together would constitute another PhD thesis, for which the material in this chapter may be considered a starting-point.

Bibliography

Al-Qassam, M.S. and J.A. Lane, 1989, Forecasting exponential autoregressive models of order 1, *Journal of Time Series Analysis* **10**, 95–113.

Anderson, H.M., 1997, Transaction costs and nonlinear adjustment towards equilibrium in the US Treasury Bill market, *Oxford Bulletin of Economics and Statistics* **59**, 465–484.

Anderson, H.M. and F. Vahid, 1998, Testing multiple equation systems for common nonlinear components, *Journal of Econometrics* **84**, 1–36.

Anděl, J., 1989, Stationary distribution of some nonlinear AR(1) processes, *Kybernetika* **25**, 453–460.

Astatkie, T., D.G. Watts and W.E. Watt, 1997, Nested threshold autoregressive NeTAR models, *International Journal of Forecasting* **13**, 105–116.

Balke, N.S., 1993, Detecting level shifts in time series, *Journal of Business & Economic Statistics* **11**, 81–92.

Balke, N.S. and M.A. Wynne, 1996, Are deep recessions followed by strong recoveries? Results for the G-7 countries, *Applied Economics* **28**, 889–897.

Balke, N.S. and M.E. Wohar, 1998, Nonlinear dynamics and covered interest rate parity, *Empirical Economics* **23**, 535–559.

Balke, N.S. and T.B. Fomby, 1994, Large shocks, small shocks, and economic fluctuations: outliers in macroeconomic time series, *Journal of Applied Econometrics* **9**, 181–200.

Balke, N.S. and T.B. Fomby, 1997, Threshold cointegration, *International Economic Review* **38**, 627–646.

Banerjee, A., J.J. Dolado, D.F. Hendry and G.W. Smith, 1986, Exploring equilibrium relationships in econometrics through static models: some Monte Carlo evidence, *Oxford Bulletin of Economics and Statistics* **48**, 253–277.

Banerjee, A., J.J. Dolado, J.W. Galbraith and D.F. Hendry, 1993, *Co-integration, Error-correction, and the Econometric Analysis of Nonstationary Data*, Oxford: Oxford University Press.

Bates, D.M. and D.G. Watts, 1988, *Nonlinear regression and its applications*, New York: John Wiley.

Baum, C.F., M. Caglayan and J.T. Barkoulas, 1998, Nonlinear adjustment to purchasing power parity in the post-Bretton Woods era, Working paper No. 404, Department of Economics, Boston College.

Beaudry, P. and G. Koop, 1993, Do recessions permanently change output?, *Journal of Monetary Economics* **31**, 149–163.

Berben, R.-P. and D. van Dijk, 1999, Unit root tests and asymmetric adjustment - a reassessment, Econometric Institute Report 9902, Erasmus University Rotterdam.

Bianchi, M. and G. Zoega, 1998, Unemployment persistence: does the size of the shock matter?, *Journal of Applied Econometrics* **13**, 283–304.

Birchenhall, C., H. Jessen and D.R. Osborn, 1996, Predicting US business cycle regimes, unpublished manuscript, University of Manchester.

Blanchard, O.J. and L.H. Summers, 1987, Hysteresis in unemployment, *European Economic Review* **31**, 288–295.

Boldin, M.D., 1996, A check on the robustness of Hamilton's Markov Switching model approach to the economic analysis of the business cycle, *Studies in Nonlinear Dynamics and Econometrics* **1**, 35–46.

Boswijk, H.P., 1999, *Asymptotic Theory for Integrated Processes*, Oxford: Oxford University Press, to appear.

Box, G.E.P. and G.C. Tiao, 1975, Intervention analysis with application to economic and environmental problems, *Journal of the American Statistical Association* **70**, 70–79.

Box, G.E.P. and G.M. Jenkins, 1970, *Time Series Analysis; Forecasting and Control*, San Francisco: Holden-Day.

Brennan, M.J. and E. Schwartz, 1988, Optimal arbitrage strategies under basis variability, *Studies in Banking and Finance* **5**, 167–180.

Brennan, M.J. and E. Schwartz, 1990, Arbitrage in stock index futures, *Journal of Business* **63**, S7–S31.

Brenner, R.J. and K.F. Kroner, 1995, Arbitrage, cointegration, and testing the unbiasedness hypothesis in financial markets, *Journal of Financial and Quantitative Analysis* **30**, 23–42.

Breusch, T.S. and A.R. Pagan, 1979, A simple test for heteroscedasticity and random coefficient variation, *Econometrica* **47**, 1287–1294.

Brock, W.A. and C. Sayers, 1988, Is the business cycle characterized by deterministic chaos?, *Journal of Monetary Economics* **22**, 71–90.

Brown, B.Y. and R.S. Mariano, 1989, Predictors in dynamic nonlinear models: large sample behaviour, *Econometric Theory* **5**, 430–452.

Bruce, A.G. and R.D. Martin, 1989, Leave k-out diagnostics for time series, *Journal of the Royal Statistical Society* **B 51**, 363–424 (with discussion).

Brunner, A.D., 1992, Conditional asymmetries in real GNP: a seminonparametric approach, *Journal of Business & Economic Statistics* **10**, 65–72.

Burns, A.F. and W.C. Mitchell, 1946, *Measuring Business Cycles*, National Bureau of Economic Research, New York.

Bustos, O.H. and V.J. Yohai, 1986, Robust estimates for ARMA models, *Journal of the American Statistical Association* **81**, 155–168.

Caner, M. and B.E. Hansen, 1998, Threshold autoregression with a near unit root, unpublished manuscript, University of Wisconsin, Madison.

Canova, F., 1994, Detrending and turning points, *European Economic Review* **38**, 614–623.

Canova, F. and E. Ghysels, 1994, Changes in seasonal patterns: are they cyclical?, *Journal of Economic Dynamics and Control* **18**, 1143–1171.

Carrasco, M., 1997, Misspecified structural change, threshold and Markov switching models, unpublished manuscript, Ohio State University.

Carroll, S.M. and B.W. Dickinson, 1989, Construction of neural nets using the radon transform, *Proceedings of the IEEE Conference on Neural Networks (Washington DC)*, New York: IEEE Press, pp. 607–611.

Chan, K.S. and H. Tong, 1985, On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations, *Advances in Applied Probability* **17**, 666–678.

Chan, K.S., J.D. Petrucelli, H. Tong and S.W. Woolford, 1985, A multiple threshold AR(1) model, *Journal of Applied Probability* **22**, 267–279.

Chan, W.S. and S.H. Cheung, 1994, On robust estimation of threshold autoregressions, *Journal of Forecasting* **13**, 37–49.

Chang, I., G.C. Tiao and C. Chen, 1988, Estimation of time series parameters in the presence of outliers, *Technometrics* **30**, 193–204.

Chappell, D. and D.A. Peel, 1998, A note on some properties of the ESTAR model, *Economics Letters* **60**, 311–315.

Chen, C. and L-M. Liu, 1993, Joint estimation of model parameters and outlier effects in time series, *Journal of the American Statistical Association* **88**, 284–297.

Chen, R., 1998, Functional coefficient autoregressive models: estimation and tests of hypotheses, Working paper 98-10, Humboldt Universität Berlin.

Christoffersen, P.F., 1998, Evaluating interval forecasts, *International Economic Review* **39**, 841–862.

Clements, M.P. and D.F. Hendry, 1998, *Forecasting Economic Time Series*, Cambridge: Cambridge University Press.

Clements, M.P. and H.-M. Krolzig, 1998, A comparison of the forecast performance of Markov-switching and threshold autoregressive models of US GNP, *Econometrics Journal* **1**, C47–C75.

Clements, M.P. and J. Smith, 1997, The performance of alternative forecasting methods for SETAR models, *International Journal of Forecasting* **13**, 463–475.

Clements, M.P. and J. Smith, 1998a, Evaluating the forecast densities of linear and non-linear models: Applications to output growth and unemployment, unpublished manuscript, University of Warwick.

Clements, M.P. and J. Smith, 1998b, Nonlinearities in exchange rates, unpublished manuscript, University of Warwick.

Clements, M.P. and J. Smith, 1998c, Testing self-exciting threshold autoregressive models against structural change models, unpublished manuscript, University of Warwick.

Clements, M.P. and J. Smith, 1999, A Monte Carlo study of the forecasting performance of empirical SETAR models, *Journal of Applied Econometrics* **14**, 123–142.

Coakley, C.W. and T.P. Hettmansperger, 1993, A bounded influence, high breakdown, efficient regression estimator, *Journal of the American Statistical Association* **88**, 872–880.

Cooper, S.J., 1998, Multiple regimes in US output fluctuations, *Journal of Business & Economic Statistics* **16**, 92–100.

Cybenko, G., 1989, Approximation by superpositions of a sigmoid function, *Mathematics of Control Signals and Systems* **2**, 303–314.

Davidson, R. and J.G. MacKinnon, 1985, Heteroskedasticity-robust tests in regression directions, *Annales de l'INSEE* **59/60**, 183–218.

Davidson, R. and J.G. MacKinnon, 1998, Graphical methods for investigating the size and power of hypothesis tests, *The Manchester Scool* **66**, 1–26.

Davies, L. and U. Gather, 1993, The identification of multiple outliers - with comments and rejoinder, *Journal of the American Statistical Association* **88**, 782–801.

Davies, R.B., 1977, Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika* **64**, 247–254.

Davies, R.B., 1987, Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika* **74**, 33–43.

de Gooijer, J.G. and P. de Bruin, 1998, On forecasting SETAR processes, *Statistics and Probability Letters* **37**, 7–14.

de Gooijer, J.G. and K. Kumar, 1992, Some recent developments in non-linear time series modelling, testing and forecasting, *International Journal of Forecasting* **8**, 135–156.

DeLong, J.B. and L.H. Summers, 1986, Are business cycles asymmetrical?, in R.J. Gordon (editor), *The American Business Cycle - Continuity and Change*, University of Chicago Press, pp. 166–179.

Denby, L. and R.D. Martin, 1979, Robust estimation of the first-order autoregressive parameter, *Journal of the American Statistical Association* **74**, 140–146.

Deutsch, S.J., J.E. Richards and J.J. Swain, 1990, Effects of a single outlier on ARMA identification, *Communications in Statistics - Theory and Methods* **19**, 2207–2227.

Diebold, F.X. and R.S. Mariano, 1995, Comparing predictive accuracy, *Journal of Business & Economic Statistics* **13**, 253–263.

Diebold, F.X. and J.A. Nason, 1990, Nonparametric exchange rate prediction, *Journal of International Economics* **28**, 315–332.

Diebold, F.X. and G.D. Rudebusch, 1992, Have postwar economic fluctuations been stabilized?, *American Economic Review* **82**, 993–1004.

Diebold, F.X. and G.D. Rudebusch, 1996, Measuring business cycles: a modern perspective, *Review of Economics and Statistics* **78**, 67–77.

Diebold, F.X., T.A. Gunther and A.S. Tay, 1998, Evaluating density forecasts with applications to financial risk management, *International Economic Review* **39**, 863–883.

Dumas, B., 1992, Dynamic equilibrium and the real exchange rate in a spatially separated world, *Review of Financial Studies* **5**, 153–180.

Durland, J.M. and T.H. McCurdy, 1994, Duration dependent transitions in a Markkov model of US GNP growth, *Journal of Business & Economic Statistics* **12**, 279–288.

Dwyer, G.P., P. Locke and W. Yu, 1996, Index arbitrage and nonlinear dynamics between the S&P 500 futures and cash, *Review of Financial Studies* **9**, 301–332.

Eisinga, R., P.H. Franses and D. van Dijk, 1998, Timing of vote decision in first and second order Dutch elections, 1978-1995: Evidence from artificial neural networks, in W.R. Mebane (editor), *Political Analysis, Vol. 7*, Universtiy of Michigan Press, Ann Arbor, pp. 117–142.

Eitrheim, Ø. and T. Teräsvirta, 1996, Testing the adequacy of smooth transition autoregressive models, *Journal of Econometrics* **74**, 59–76.

Elwood, S.K., 1998, Is the persistence of shocks to output asymmetric?, *Journal of Monetary Economics* **41**, 411–426.

Emery, K.M. and E.F. Koenig, 1992, Forecasting turning points - Is a two-state characterization of the business cycle appropriate?, *Economics Letters* **39**, 431–435.

Enders, W. and C.W.J. Granger, 1998, Unit-root tests and asymetric adjustment with an example using the term structure of interest rates, *Journal of Business & Economic Statistics* **16**, 304–311.

Engle, R.F. and C.W.J. Granger, 1987, Co-integration and error-correction: representation, estimation and testing, *Econometrica* **55**, 251–276.

Escribano, A. and O. Jordá, 1999, Improved testing and specification of smooth transition regression models, in P. Rothman (editor), *Nonlinear Time Series Analysis of Economic and Financial Data*, Boston: Kluwer Academic Press, pp. 289–319.

Escribano, A. and S. Mira, 1997, Nonlinear error-correction models, Working Paper Series in Statistics and Econometrics 97-26, Universidad Carlos III de Madrid.

Escribano, A. and G.A. Pfann, 1998, Non-linear error correction, asymmetric adjustment and cointegration, *Economic Modelling* **15**, 197–216.

Escribano, A., P.H. Franses and D. van Dijk, 1998, Nonlinearities and outliers: robust specification of STAR models, Econometric Institute Report 9832, Erasmus University Rotterdam.

Falk, B., 1986, Further evidence on the asymmetric behavior of economic time series over the business cycle, *Journal of Political Economy* **94**, 1096–1109.

Fan, J. and I. Gijbels, 1996, *Local Polynomial Modeling and its Applications*, London: Chapman and Hall.

Fisher, L., 1966, Some new stock market indexes, *Journal of Business* **39**, 191–225.

Fox, A.J., 1972, Outliers in time series, *Journal of the Royal Statistical Society* **B 34**, 350–363.

Franses, P.H., 1998, *Time Series Models for Business and Economic Forecasting*, Cambridge: Cambridge University Press.

Franses, P.H. and R. Paap, 1999, Does seasonality influence the dating of business cycle turning points?, *Journal of Macroeconomics* **21**, 79–92.

Franses, P.H. and D. van Dijk, 1999, *Nonlinear Time Series Models in Empirical Finance*, Cambridge: Cambridge University Press, to appear.

Franses, P.H., D. van Dijk and A. Lucas, 1998, Short patches of outliers, ARCH, and volatility modeling, Tinbergen Institute discussion paper 98-057/4.

Friedman, M., 1969, *The Optimum Quantity of Money and Other Essays*, Chicago: Aldin.

Friedman, M., 1993, The 'plucking model' of business fluctuations revisited, *Economic Inquiry* **31**, 171–177.

Funabashi, K., 1989, On the approximate realization of continuous mappings by neural networks, *Neural Networks* **2**, 183–192.

Gabr, M.M., 1998, Robust estimation of bilinear time series models, *Communications in Statistics - Theory and Methods* **27**, 41–53.

Gallant, A.R., 1987, *Nonlinear Statistical Models*, New York: John Wiley.

Garcia, R. and P. Perron, 1996, An analysis of the real interest rate under regime switches, *Review of Economics and Statistics* **78**, 111–125.

Gemmill, G., 1998, Touched by Success, *Financial News Briefing Notes*, October 5.

Ghysels, E., C.W.J. Granger and P.L. Siklos, 1996, Is seasonal adjustment a linear or nonlinear data filtering process?, *Journal of Business & Economic Statistics* **14**, 374–386.

Goodwin, R.M., 1951, The nonlinear accelerator and the persistence of the business cycle, *Econometrica* **19**, 1–17.

Goodwin, T.H. and R.J. Sweeney, 1993, International evidence on Friedman's theory of the business cycle, *Economic Inquiry* **31**, 178–193.

Gordon, S., 1997, Stochastic trends, deterministic trends, and business cycle turning points, *Journal of Applied Econometrics* **12**, 411–434.

Granger, C.W.J., 1981, Some properties of time series data and their use in econometric model specification, *Journal of Econometrics* **16**, 121–130.

Granger, C.W.J., 1993, Strategies for modelling nonlinear time-series relationships, *The Economic Record* **69**, 233–238.

Granger, C.W.J. and J. Hallman, 1991, Long memory series with attractors, *Oxford Bulletin of Economics and Statistics* **53**, 11–26.

Granger, C.W.J. and T.H. Lee, 1989, Investigation of production, sales and inventory relationships using multicointegration and non-symmetric error correction models, *Journal of Applied Econometrics* **4**, S145–S159.

Granger, C.W.J. and N.R. Swanson, 1996, Future developments in the study of cointegrated variables, *Oxford Bulletin of Economics and Statistics* **58**, 537–553.

Granger, C.W.J. and T. Teräsvirta, 1993, *Modelling Nonlinear Economic Relationships*, Oxford: Oxford University Press.

Granger, C.W.J. and T. Teräsvirta, 1999, A simple nonlinear time series model with misleading linear properties, *Economics Letters* **62**, 161–165.

Granger, C.W.J., T. Teräsvirta and H. Anderson, 1993, Modeling non-linearity over the business cycle, in J.H. Stock and M.W. Watson (editors), *New research on business cycles, indicators and forecasting*, Chicago: Chicago University Press, pp. 311–325.

Hamilton, J.D., 1989, A new approach to the economic analysis of nonstationary time series subject to changes in regime, *Econometrica* **57**, 357–384.

Hampel, H.R., E.M. Ronchetti, P.J. Rousseeuw and W.A. Stahel, 1986, *Robust Statistics - The Approach based on Influence Functions*, New York: John Wiley.

Handschin, E., J. Kohlas, A. Fiechter and F. Schweppe, 1975, Bad data analysis for power system state estimation, *IEEE Transactions on Power Apparatus and Systems* **2**, 329–337.

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

Hansen, B.E., 1997, Inference in TAR models, *Studies in Nonlinear Dynamics and Econometrics* **2**, 1–14.

Harvey, D., S. Leybourne and P. Newbold, 1997, Testing the equality of prediction mean squared errors, *International Journal of Forecasting* **13**, 281–291.

Hatanaka, M., 1996, *Time-Series-Based Econometrics, Unit Roots and Cointegration*, Oxford: Oxford University Press.

Hendry, D.F., 1995, *Dynamic Econometrics*, Oxford: Oxford University Press.

Hess, G.D. and S. Iwata, 1997a, Asymmetric persistence in GDP? A deeper look at depth, *Journal of Monetary Economics* **40**, 535–554.

Hess, G.D. and S. Iwata, 1997b, Measuring and comparing business cycle features, *Journal of Business & Economic Statistics* **15**, 432–444.

Hoek, H., A. Lucas and H.K. van Dijk, 1995, Classical and Bayesian aspects of robust unit root inference, *Journal of Econometrics* **69**, 27–59.

Hornik, K., M. Stinchcombe and H. White, 1989, Multilayer feedforward networks are universal approximators, *Neural Networks* **2**, 359–366.

Hornik, K., M. Stinchcombe and H. White, 1990, Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks, *Neural Networks* **3**, 551–560.

Hsieh, D.A., 1983, A heteroscedasticity-consistent covariance matrix estimator for time series regressions, *Journal of Econometrics* **22**, 281–290.

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

Hyndman, R.J., 1995, Highest-density forecast regions for nonlinear and nonnormal time series, *Journal of Forecasting* **14**, 431–441.

Hyndman, R.J., 1996, Computing and graphing highest-density regions, *American Statistican* **50**, 120–126.

Jansen, D.W. and W. Oh, 1996, Modeling nonlinearities of business cycles: choosing between the CDR and STAR models, unpublished manuscript, Department of Economics, Texas A&M University.

Jansen, E.S. and T. Teräsvirta, 1996, Testing parameter constancy and super exogeneity in econometric equations, *Oxford Bulletin of Economics and Statistics* **58**, 735–768.

Johansen, S., 1995, *Likelihood-based inference in cointegrated vector autoregressive models*, Oxford: Oxford University Press.

Kaldor, N., 1940, A model of the trade cycle, *Economic Journal* **50**, 78–92.

Keynes, J.M., 1936, *The General Theory of Employment, Interest, and Money*, London: MacMillan.

Kim, C.-J. and C.R. Nelson, 1998, A test for structural change in Markov-Switching models: has the US economy become more stable?, unpublished manuscript, Department of Economics, University of Washington.

Koop, G., 1996, Parameter uncertainty and impulse response analysis, *Journal of Econometrics* **72**, 135–149.

Koop, G. and S. Potter, 1998, Nonlinearity, structural breaks or outliers in economic time series?, in W. A. Barnett, D. F. Hendry, S. Hylleberg, T. Teräsvirta, D. Tjøstheim and A. H. Würtz (editors), *Nonlinear dynamics*, Cambridge: Cambridge University Press, to appear.

Koop, G., M.H. Pesaran and S.M. Potter, 1996, Impulse response analysis in nonlinear multivariate models, *Journal of Econometrics* **74**, 119–147.

Krolzig, H.-M., 1997, *Markov-Switching Vector Autoregressions - Modelling, Statistical Inference and Applications to Business Cycle Analysis*, Lecture Notes in Economics and Mathematics **454**, Berlin: Springer-Verlag.

Kuan, C.-M. and H. White, 1994, Artificial neural networks: an econometric perspective, *Econometric Reviews* **13**, 1–143 (with discussion).

Kunst, R.M., 1992, Threshold cointegration in interest rates, Discussion paper 92-26, University of California, San Diego.

Kunst, R.M., 1995, Determining long-run equilibrium structures in bivariate threshold autoregressions: a multiple decision approach, Discussion paper, Institute for Advanced Studies, Vienna.

Leybourne, S., P. Newbold and D. Vougas, 1998, Unit roots and smooth transitions, *Journal of Time Series Analysis* **19**, 83–97.

Lin, C-F.J. and T. Teräsvirta, 1994, Testing the constancy of regression parameters against continuous structural change, *Journal of Econometrics* **62**, 211–228.

Lin, J-L. and C.W.J. Granger, 1994, Forecasting from nonlinear models in practice, *Journal of Forecasting* **13**, 1–9.

Lo, A. and C. MacKinlay, 1990, An econometric analysis of non-synchronous trading, *Journal of Econometrics* **45**, 181–212.

Lucas, A., 1995, Unit root tests based on M estimators, *Econometric Theory* **11**, 331–346.

Lucas, A., 1996, *Outlier Robust Unit Root Testing*, PhD thesis, Thesis/Tinbergen Institute.

Lucas, A., 1997, Asymptotic robustness of least median of squares for autoregressions with additive outliers, *Communications in Statistics - Theory and Methods* **26**, 2363–2380.

Lucas, A., 1999, *Outlier Robust Analysis of Trending Economic Time Series*, Oxford: Oxford University Press, to appear.

Lucas, A., R. van Dijk and T. Kloek, 1996, Outlier robust GMM estimation of leverage determinants in linear dynamic panel data models, Tinbergen Institute Discussion paper 94-132.

Lundbergh, S., T. Teräsvirta and D. van Dijk, 1999, Time-varying smooth transition autoregressive models, unpublished manuscript, Stockholm School of Economics.

Lütkepohl, H., 1991, *Introduction to Multiple Time Series Analysis*, Berlin: Springer-Verlag.

Luukkonen, R., P. Saikkonen and T. Teräsvirta, 1988, Testing linearity against smooth transition autoregressive models, *Biometrika* **75**, 491–499.

Luukkonen, R. and T. Teräsvirta, 1991, Testing linearity of economic time series against cyclical asymmetry, *Annales d'Économie et de Statistique* **21**, 125–142.

Markatou, M. and X. He, 1994, Bounded influence and high breakdown point testing procedures in linear models, *Journal of the American Statistical Association* **89**, 543–549.

Maronna, R.A., 1976, Robust M-estimators of multivariate location and scatter, *Annals of Statistics* **4**, 51–67.

Martens, M., P. Kofman and A.C.F. Vorst, 1998, A threshold error correction for intraday futures and index returns, *Journal of Applied Econometrics* **13**, 245–263.

Martin, R.D., 1981, Robust methods for time series, in D.F. Findley (editor), *Applied Time Series Analysis*, New York: Academic Press, pp. 6843–759.

Martin, R.D. and V.J. Yohai, 1986, Influence functionals for time series, *Annals of Statistics* **14**, 781–818.

McQueen, G. and S.R. Thorley, 1993, Asymmetric business cycle turning points, *Journal of Monetary Economics* **31**, 341–362.

Michael, P., A.R. Nobay and D.A. Peel, 1997, Transaction costs and nonlinear adjustment in real exchange rates: an empirical investigation, *Journal of Political Economy* **105**, 862–879.

Miller, M.H., J. Muthuswamy and R.E. Whaley, 1994, Mean reversion of S&P 500 index basis changes: arbitrage-induced or statistical illusion?, *Journal of Finance* **49**, 479–513.

Mittnik, S. and Z. Niu, 1994, Asymmetries in business cycles: econometric techniques and empirical evidence, in W. Semmler (editor), *Business Cycles: Theory and Empirical Methods*, Kluwer Academic Publishers, Boston, pp. 331–350.

Moeanaddin, R. and H. Tong, 1990, Numerical evaluation of distributions in nonlinear autoregression, *Journal of Time Series Analysis* **11**, 33–48.

Montgomery, A.L., V. Zarnowitz, R.S. Tsay and G.C. Tiao, 1998, Forecasting the US unemployment rate, *Journal of the American Statistical Association* **93**, 478–493.

Mullineux, A. and W. Peng, 1993, Nonlinear business cycle modelling, *Journal of Economic Surveys* **7**, 41–83.

Neal, R., 1992, Direct tests of index arbitrage models, unpublished manuscript, University of Washington.

Neftçi, S.N., 1984, Are economic time series asymmetric over the business cycle?, *Journal of Political Economy* **92**, 307–328.

Neftçi, S.N., 1993, Statistical analysis of shapes in macroeconomic time series: is there a business cycle?, *Journal of Business & Economic Statistics* **11**, 215–224.

Nelder, J.A. and R. Mead, 1965, A simplex method for function minimization, *The Computer Journal* **7**, 308.

Nicholls, D. and B.G. Quinn, 1982, *Random Coefficient Autoregressive Models: An Introduction*, Lecture notes in Statistics 11, Springer-Verlag, New York.

Obstfeld, M. and A.M. Taylor, 1997, Nonlinear aspects of goods-market arbitrage and adjustment: Heckscher's commodity points revisited, *Journal of the Japanese and International Economies* **11**, 441–479.

O'Connell, P.G.J., 1998, Market frictions and real exchange rates, *Journal of International Money and Finance* **17**, 71–95.

O'Connell, P.G.J. and S.-J. Wei, 1997, "The bigger they are, the harder they fall"; how price differences across U.S. cities are arbitraged, NBER working paper No. 6089.

Ooms, M., 1994, *Empirical Vector Autoregressive Modeling*, Berlin: Springer-Verlag.

Parker, R.E. and P. Rothman, 1996, Further evidence on the stabilization of postwar economic fluctucations, *Journal of Macroeconomics* **18**, 289–298.

Parker, R.E. and P. Rothman, 1997, The current depth of recession and unemployment rate forecasts, *Studies in Nonlinear Dynamics and Econometrics* **2**, 151–158.

Peel, D.A. and A.E.H. Speight, 1996, Is the US business cycle asymmetric? Some further evidence, *Applied Economics* **28**, 405–415.

Pemberton, J., 1987, Exact least squares multi-step prediction from nonlinear autoregressive models, *Journal of Time Series Analysis* **8**, 443–448.

Peracchi, F., 1991, Robust M-tests, *Econometric Theory* **7**, 69–84.

Perron, P., 1989, The great crash, the oil price shock, and the unit root hypothesis, *Econometrica* **57**, 1361–1401.

Pesaran, M.H. and S.M. Potter, 1997, A floor and ceiling model of US output, *Journal of Economic Dynamics and Control* **21**, 661–695.

Pesaran, M.H. and Y. Shin, 1998, Generalized impulse response analysis in linear multivariate models, *Economics Letters* **58**, 17–29.

Pesaran, M.H. and A. Timmermann, 1992, A simple nonparametric test of predictive performance, *Journal of Business & Economic Statistics* **10**, 461–465.

Pesaran, M.H. and A. Timmermann, 1994, A generalization of the non-parametric Henriksson-Merton test of market timing, *Economics Letters* **44**, 1–7.

Pötscher, B.M. and I.V. Prucha, 1997, *Dynamic Nonlinear Econometric Models - Asymptotic Theory*, Berlin: Springer-Verlag.

Potter, S.M., 1994, Asymmetric economic propagation mechanisms, in W. Semmler (editor), *Business Cycles: Theory and Empirical Methods*, Kluwer Academic Publishers, Boston, pp. 313–330.

Potter, S.M., 1995a, Nonlinear models of economic fluctuations, in K. Hoover (editor), *Macroeconometrics - Developments, Tensions and Prospects*, Kluwer, Boston, pp. 517–560.

Potter, S.M., 1995b, A nonlinear approach to US GNP, *Journal of Applied Econometrics* **10**, 109–125.

Prakash, G. and A.M. Taylor, 1997, Measuring market integration: a model of arbitrage with an application to the gold standard, 1879-1913, NBER Working paper No. 6073.

Press, W.H., B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, 1986, *Numerical Recipes - The Art of Scientific Computing*, Cambridge: Cambridge University Press.

Priestley, M.B., 1980, State-dependent models: a general approach to non-linear time series analysis, *Journal of Time Series Analysis* **1**, 47–71.

Priestley, M.B., 1988, *Nonlinear and Nonstationary Time Series Analysis*, London: Academic Press.

Quandt, R., 1983, Computational problems and methods, in Z. Griliches and M.D. Intriligator (editors), *Handbook of Econometrics I*, Amsterdam: Elsevier Science, pp. 699–746.

Ramsey, J.B. and P. Rothman, 1996, Time irreversibility and business cycle asymmetry, *Journal of Money, Credit and Banking* **28**, 1–21.

Ravn, M.O. and M. Sola, 1995, Stylized facts and regime changes: are prices procyclical?, *Journal of Monetary Economics* **36**, 497–526.

Ronchetti, E., 1997, Robustness aspects of model choice, *Statistica Sinica* **7**, 327–338.

Rothman, P., 1991, Further evidence on the asymmetric behavior of unemployment rates over the business cycle, *Journal of Macroeconomics* **13**, 291–298.

Rothman, P., 1998, Forecasting asymmetric unemployment rates, *Review of Economics and Statistics* **80**, 164–168.

Rousseeuw, P.J., 1984, Least median of squares regression, *Journal of the American Statistical Association* **79**, 871–880.

Rousseeuw, P.J., 1985, Multivariate estimation with high breakdown point, in I. Vincze W. Grossmann, G. Pflug and W. Wertz (editors), *Mathematical Statistics and Applications, Vol. B*, Dordrecht: Reidel Publishing, pp. 283–297.

Rousseeuw, P.J. and A.M. Leroy, 1987, *Robust Regression and Outlier Detection*, New York: John Wiley.

Rousseeuw, P.J. and B.C. van Zomeren, 1990, Unmasking multivariate outliers and leverage points - with comments and rejoinder, *Journal of the American Statistical Association* **85**, 633–651.

Saikkonen, P. and R. Luukkonen, 1988, Lagrange multiplier tests for testing non-linearities in time series models, *Scandinavian Journal of Statistics* **15**, 55–68.

Scholes, M. and J. Williams, 1977, Estimating betas from non-synchronous data, *Journal of Financial Economics* **5**, 309–328.

Sichel, D.E., 1989, Are business cycles asymmetric? A correction, *Journal of Political Economy* **97**, 1255–1260.

Sichel, D.E., 1993, Business cycles asymmetry: a deeper look, *Economic Inquiry* **31**, 224–236.

Sichel, D.E., 1994, Inventories and the three phases of the business cycle, *Journal of Business & Economic Statistics* **12**, 269–277.

Siklos, P.L. and C.W.J. Granger, 1997, Temporary cointegration with an application to interest rate parity, *Macroeconomic Dynamics* **1**, 640–657.

Silverman, B.W., 1986, *Density Estimation fpr Statistics and Data Analysis*, New York: Chapmann Hall.

Simpson, D.G., D. Ruppert and R.J. Carroll, 1992, On one-step GM estimates and stability of inferences in linear regression, *Journal of the American Statistical Association* **87**, 439–450.

Skalin, J. and T. Teräsvirta, 1998, Moving equilibria and asymmetries in unemployment rates, Working Paper Series in Economics and Finance No. 262, Stockholm School of Economics.

Sofianos, G., 1993, Index arbitrage profitability, *Journal of Derivatives* **1**, 6–20.

Staiger, D., J.H. Stock and M.W. Watson, 1997, The NAIRU, unemployment and monetary policy, *Journal of Economic Perspectives* **11**, 33–50.

Stock, J.H. and M.W. Watson, 1996, Evidence on structural instability in macroeconomic time series relations, *Journal of Business & Economic Statistics* **14**, 11–30.

Stromberg, A.J., 1993, Computation of high breakdown nonlinear regression parameters, *Journal of the American Statistical Association* **88**, 237–244.

Stromberg, A.J., 1995, Consistency of the least median of squares estimator in nonlinear regression, *Communications in Statistics - Theory and Methods* **24**, 1971–1984.

Stromberg, A.J. and D. Ruppert, 1992, Breakdown in nonlinear regression, *Journal of the American Statistical Association* **87**, 991–997.

Swanson, N.R., 1999, Finite sample properties of a simple LM test for neglected nonlinearity in error-correcting regression equations, *Statistica Neerlandica* **53**, 76–95.

Swanson, N.R., A. Ozyildirim and M. Pisu, 1996, A comparison of alternative causality and predictive accuracy tests in the presence of integrated and co-integrated economic variables, unpublished manuscript, Department of Economics, Pennsylvania State University.

Taylor, N., D. van Dijk, P.H. Franses and A. Lucas, 1999, SETS, arbitrage activity, and stock price dynamics, *Journal of Banking and Finance*, to appear.

Teräsvirta, T., 1994, Specification, estimation, and evaluation of smooth transition autoregressive models, *Journal of the American Statistical Association* **89**, 208–218.

Teräsvirta, T., 1995, Modelling nonlinearity in US gross national product 1889-1987, *Empirical Economics* **20**, 577–598.

Teräsvirta, T., 1998, Modelling economic relationships with smooth transition regressions, in A. Ullah and D.E.A. Giles (editors), *Handbook of Applied Economic Statistics*, New York: Marcel Dekker, pp. 507–552.

Teräsvirta, T. and H.M. Anderson, 1992, Characterizing nonlinearities in business cycles using smooth transition autoregressive models, *Journal of Applied Econometrics* **7**, S119–S136.

Teräsvirta, T. and C-F.J. Lin, 1993, Determining the number of hidden units in a single hidden-layer neural network model, unpublished manuscript, National Taiwan University, Taipei.

Teräsvirta, T., C-F.J. Lin and C.W.J. Granger, 1993, Power of the neural network linearity test, *Journal of Time Series Analysis* **14**, 209–220.

Teräsvirta, T., D. Tjøstheim and C.W.J. Granger, 1994, Aspects of modelling nonlinear time series, in R.F. Engle and D.L. McFadden (editors), *Handbook of econometrics, vol. IV*, Amsterdam: Elsevier Science.

Tiao, G.C. and R.S. Tsay, 1994, Some advances in non-linear and adaptive modelling in time-series (with discussion), *Journal of Forecasting* **13**, 109–140.

Tjøstheim, D., 1986, Some doubly stochastic time series models, *Journal of Time Series Analysis* **7**, 51–72.

Tong, H., 1990, *Non-Linear Time Series: a Dynamical Systems Approach*, Oxford: Oxford University Press.

Tong, H., 1995, A personal overview of non-linear time series analysis from a chaos perspective, *Scandinavian Journal of Statistics* **22**, 399–445.

Tsay, R.S., 1986a, Nonlinearity tests for time series, *Biometrika* **73**, 461–466.

Tsay, R.S., 1986b, Time series model specification in the presence of outliers, *Journal of the American Statistical Association* **81**, 132–141.

Tsay, R.S., 1988, Outliers, level shifts, and variance changes in time series, *Journal of Forecasting* **7**, 1–20.

Tsay, R.S., 1989, Testing and modeling threshold autoregressive processes, *Journal of the American Statistical Association* **84**, 231–240.

Tsay, R.S., 1998, Testing and modeling multivariate threshold models, *Journal of the American Statistical Association* **93**, 1188–1202.

van Dijk, D. and P.H. Franses, 1998, Nonlinear error-correction models for interest rates in the Netherlands, in W. A. Barnett, D. F. Hendry, S. Hylleberg, T. Teräsvirta, D. Tjøstheim and A. H. Würtz (editors), *Nonlinear Econometric Modeling*, Cambridge: Cambridge University Press, to appear.

van Dijk, D. and P.H. Franses, 1999, Modeling multiple regimes in the business cycle, *Macroeconomic Dynamics*, to appear.

van Dijk, D., P.H. Franses and A. Lucas, 1999a, Testing for smooth transition nonlinearity in the presence of additive outliers, *Journal of Business & Economic Statistics* **17**, 217–235.

van Dijk, D., P.H. Franses and A. Lucas, 1999b, Testing for ARCH in the presence of additive outliers, *Journal of Applied Econometrics*, to appear.

Verbrugge, R., 1997, Investigating cyclical asymmetries, *Studies in Nonlinear Dynamics and Econometrics* **2**, 15–22.

Wand, M.P. and M.C. Jones, 1995, *Kernel Smoothing*, London: Chapman and Hall.

Watson, M.W., 1994, Business cycle durations and postwar stabilization of the US economy, *American Economic Review* **84**, 24–46.

Weise, C.L., 1999, The asymmetric effects of monetary policy, *Journal of Money, Credit and Banking*, to appear.

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

White, H. and I. Domowitz, 1984, Nonlinear regression with dependent observations, *Econometrica* **52**, 143–161.

Wooldridge, J.M., 1990, A unified approach to robust, regression-based specification tests, *Econometric Theory* **6**, 17–43.

Wooldridge, J.M., 1991, On the application of robust, regression-based diagnostics to models of conditional means and conditional variances, *Journal of Econometrics* **47**, 5–46.

Wu, L.S.-Y., J.R.M. Hosking and N. Ravishankar, 1993, Reallocation outliers in time series, *Applied Statistics* **42**, 301–313.

Wynne, M.A. and N.S. Balke, 1992, Are deep recessions followed by strong recoveries?, *Economics Letters* **39**, 183–189.

Yadav, P.K., P.F. Pope and K. Paudyal, 1994, Threshold autoregressive modeling in finance: the price difference of equivalent assets, *Mathematical Finance* **4**, 205–221.

Nederlandse Samenvatting

Smooth Transition Models: Extensions and Outlier Robust Inference

Veel economische variabelen vertonen niet-lineaire patronen wanneer zij worden waargenomen op opeenvolgende tijdstippen. Een bekend voorbeeld betreft macro-economische variabelen, zoals industriële productie en werkloosheid, in verschillende fasen van de conjunctuurcyclus. In het algemeen wordt waargenomen dat de conjunctuurcyclus een aantal asymmetrische kenmerken heeft. Zo verschillen perioden van opgaande en neergaande conjunctuur (expansies en recessies) in lengte. Expansies duren gemiddeld langer dan recessies. Ook de omslagpunten in de conjunctuurcyclus verschillen van aard. De overgang van een recessie naar een expansie vindt vaak abrupt plaats, terwijl omgekeerd de overgang van een expansie naar een recessie meer geleidelijk verloopt. Deze asymmetrische kenmerken van de conjunctuurcyclus vertalen zich in niet-lineair gedrag van macro-economische variabelen met een cyclische component. Zo heeft het werkloosheidspercentage de neiging snel te stijgen gedurende recessies, en slechts langzaam te dalen gedurende expansies.

Dergelijke asymmetrische of niet-lineaire eigenschappen zijn niet adequaat te beschrijven met behulp van lineaire tijdreeksmodellen. Desondanks beperkte men zich in praktijk tot voor kort tot lineaire modellen. Een mogelijke verklaring voor dit gegeven is dat lineaire modellen vaak een redelijk goede benadering lijken te geven van de aanwezige niet-lineariteit. Een bijkomend probleem is dat het aantal mogelijke niet-lineaire modellen in principe onbegrensd is. Het is a priori vaak niet duidelijk welk niet-lineair model geschikt is voor het beschrijven van de intertemporele eigenschappen van een variabele. Daarnaast was het moeilijk, zo niet onmogelijk, betrouwbare schattingen te verkrijgen van parameters in niet-lineaire modellen. Mede dankzij de verbeterde computertechnologie, waardoor deze laatste verklaring min of meer teniet is gedaan, is de aandacht voor niet-lineaire tijdreeksmodellen recentelijk sterk toegenomen.

Een voor de hand liggende benadering voor het beschrijven van niet-lineariteit in macro-economische variabelen wordt ingegeven door de veronderstelling dat verschillende 'toestanden' of 'regimes' kunnen optreden, waarbij de eigenschappen van een variabele verschillend zijn in de diverse regimes. Het onderscheid tussen expansies en recessies in een conjunctuurcyclus is een voorbeeld van deze benadering. Recentelijk zijn een aantal niet-lineaire tijdreeksmodellen ontwikkeld, bestudeerd en toegepast, die dergelijke regime-veranderingen kunnen beschrijven. Dit proefschrift behandelt één van deze modellen, namelijk smooth transition autoregressieve [STAR] model. Het STAR model heeft een tweetal karakteristieke kenmerken die het onderscheiden van andere modellen. Ten eerste, de verschillende regimes kunnen worden geassocieerd met de waarden die bepaalde waarneembare variabelen aannemen. Zo kan bijvoorbeeld een recessie gedefineerd worden als een periode met negatieve groei in industriële productie. Ten tweede, de overgang in de eigenschappen van de variabele die met het STAR model worden beschreven tussen de verschillende regimes is geleidelijk. De verandering in de eigenschappen van bijvoorbeeld de groei in industriële productie vindt niet abrupt plaats wanneer een recessie omslaat in een expansie, maar deze treedt geleidelijk op.

Dit proefschrift valt uiteen in drie delen. Het eerste deel, omvattende de Hoofdstukken 1 en 2, beftreft een algemene inleiding en een beschrijving van het standaard STAR model zoals dit in het algemeen wordt toegepast. Uitgangspunt is een specificatieprocedure voor STAR modellen, aan de hand waarvan verschillende aspecten van het model worden behandeld. Een zelfde benadering is gebruikt in navolgende hoofdstukken. Het tweede deel, met de Hoofdstukken 3, 4 en 5, behandelt een aantal uitbreidingen van het standaard model, die toestaan dat meer dan twee regimes optreden, die toestaan dat de eigenschappen van een variabele ook over de tijd veranderen, en die meerdere tijdreeksen tegelijkertijd kunnen beschrijven. Het derde deel, met de Hoofdstukken 6 en 7, onderzoekt de invloed die uitschieters in de data hebben op de analyse van het STAR model. In het navolgende wordt de inhoud van de hoofdstukken in het tweede en derde deel nader toegelicht.

Uitbreidingen van het STAR model

Het STAR model zoals dit in het algemeen wordt toegepast, onderscheidt twee regimes. Hoewel dit voor veel toepassingen voldoende lijkt, kan het van belang zijn de mogelijkheid van meerdere regimes te onderzoeken. Als voorbeeld kan genoemd worden dat recent onderzoek naar de conjunctuurcyclus er op duidt dat het onderscheid tussen alleen expansies en recessies te restrictief is. De groei in industriële productie direct volgend op een dieptepunt van de cyclus blijkt in het algemeen veel sterker dan gedurende de rest van de expansie. Het zou derhalve nuttig kunnen zijn de expansie fase verder op te splitsen in een 'herstel' fase en een 'gematigde groei' fase.

In Hoofdstuk 3 wordt een uitbreiding van het STAR model onderzocht waarin meerdere regimes worden toegestaan. Dit meerdere regime STAR [MRSTAR] model kan op een simpele doch elegante wijze worden verkregen uitgaande van het standaard model. Een uitbreiding van de specificatieprocedure voor het standaard model kan worden gebruikt voor het construeren van een MRSTAR model. De toepassing in dit hoofdstuk op groei in het reëel bruto national product van de Verenigde Staten geeft aanwijzingen voor het bestaan van meerdere regimes in het gedrag van deze variabele over de conjunctuurcyclus.

Naast niet-lineariteit is instabiliteit een andere prominente eigenschap van veel economische variabelen. Met instabiliteit wordt hier bedoeld dat zekere kenmerken van een variabele veranderen over de tijd. Als voorbeeld kan genoemd worden het werkloosheidspercentage, dat in veel westerse geindustrialiseerde landen sinds de 80-er jaren op een permanent hoger niveau lijkt te liggen. Traditioneel worden mogelijke niet-lineariteit en instabiliteit in een variabele afzonderlijk van elkaar onderzocht. Een mogelijke verklaring voor dit gegeven is dat het moeilijk lijkt een onderscheid te maken tussen deze twee alternatieven. Immers, een variabele die gegenereerd wordt door een niet-lineair proces kan lijken op een variabele die onderhevig is aan enkele stucturele veranderingen.

In Hoofdstuk 4 wordt een model, gebaseerd op het principe van het STAR model, beschouwd dat tegelijkertijd niet-lineaire en tijds-variërende eigenschappen in een variabele kan beschrijven. Dit tijds-variërende STAR [TV-STAR] model kan worden verkregen als een bijzonder geval van het MRSTAR model in Hoofdstuk 3. De bijzondere structuur van het TV-STAR model maakt het mogelijk een alternatieve procedure voor specificatie te gebruiken. Een uitgebreide simulatie-studie vergelijkt de voor- en nadelen van deze aanpak ten opzichte van de specificatieprocedure zoals beschreven in Hoofdstuk 3. De resultaten duiden erop dat de statistische toetsen die gebruikt worden in beide procedures het mogelijk maken een onderscheid te maken tussen niet-lineariteit en instabiliteit. De toepassing in dit hoofdstuk op de groei in industriële productie in het Verenigd Koninkrijk toont aan dat deze variabele zowel niet-lineaire als tijds-variërende eigenschappen bezit, en dat het van belang kan zijn beide expliciet in een beschrijvend model op te nemen.

Het standaard STAR model is een univariaat model, in de zin dat het beoogt het gedrag van één enkele variabele te beschrijven. Het kan echter van belang zijn meerdere variabelen tegelijkertijd te modelleren, teneinde gebruik te maken van mogelijke verbanden die tussen de betreffende variabelen bestaan. Het gebruik van niet-lineaire modellen in deze context is zinvol wanneer het vermoeden bestaat dat dergelijke verbanden een niet-lineaire vorm hebben.

In hoofdstuk 5 wordt een uitbreiding van het STAR model behandeld welke het mogelijk maakt meerdere variabelen tegelijkertijd te beschrijven. Speciale aandacht wordt geschonken aan zogenaamde niet-lineaire evenwicht-correctie modellen. Deze modellen zijn geschikt om de situatie te beschrijven waarin twee of meer variabelen met elkaar verbonden zijn door middel van een lineaire evenwichtsrelatie, maar waarbij aanpassingen als gevolg van afwijkingen van dit evenwicht niet-lineair zijn. Plausibele vormen van niet-lineaire aanpassing zijn met name die waarbij de snelheid van aanpassing afhankelijk is van de grootte van de afwijking van het evenwicht. Een voorbeeld is de relatie tussen prijzen van een aandeel en een futures contract op ditzelfde aandeel. Deze prijzen zijn aan elkaar gerelateerd door middel van het cost-of-carry model, waarin de theoretische futures prijs bepaald wordt op basis van de aandeelprijs, de resterende looptijd van het futures contract, het dividend die gedurende de resterende looptijd worden uitgekeerd, en een rentestand. Wanneer de werkelijke futures prijs afwijkt van de theoretische maakt dit arbitrage mogelijk, waarbij posities in aandelen en futures kunnen worden ingenomen, zodanig dat op de afloopdatum van het futures contract een zekere winst wordt gerealiseerd. Aan het handelen in aandelen en futures zijn echter transactiekosten verbonden, welke ervoor zorgen dat winstgevende arbitrage niet mogelijk is bij kleine afwijkingen van de cost-of-carry relatie. Dit betekent dat aandeel en futures prijzen zich slechts dan zullen aanpassen naar deze evenwichtsrelatie wanneer het verschil tussen de werkelijke en theoretische futures prijs voldoende groot is. De toepassing in dit hoofdstuk toont aan dat dergelijk niet-lineair aanpassingsgedrag inderdaad gevonden kan worden in de spot- en futuresprijzen van de FTSE100 index. Tevens laat deze toepassing zien dat het handelssysteem dat gehanteerd wordt op een beurs van invloed is op de hoogte van de transactiekosten en daarmee op de vorm van de niet-lineaire aanpassing.

Uitschieter-robuuste analyse van STAR modellen

De parameters in het STAR model kunnen zodanig worden gekozen dat de resulterende tijdreeksen extreem asymmetrisch zijn, in die zin dat de overgrote meerderheid van de waarnemingen in een van beide regimes ligt. Dergelijke tijdreeksen lijken derhalve op tijdreeksen gegenereerd door een lineair model plus enkele uitzonderlijke waarnemingen, oftewel uitschieters. Omgekeerd kan een lineaire tijdreeks welke is besmet met enkele uitschieters niet-lineariteit suggereren. Dit suggereert dat het moeilijk is de twee mogelijkheden van elkaar te onderscheiden met behulp van standaard methoden. Standaard statistische toetsen voor niet-lineariteit bijvoorbeeld zijn bijzonder gevoelig voor de aanwezigheid van uitschieters, in de zin dat dergelijke waarnemingen ervoor kunnen zorgen dat de toetsen de aanwezigheid van niet-lineariteit suggereren. In het derde deel van dit proefschrift wordt onderzocht of een dergelijk onderscheid wel mogelijk is met behulp van alternatieve, uitschieter-robuuste methoden.

In hoofdstuk 6 worden uitschieter-robuuste schattingsmethoden voor lineaire tijdreeksmodellen gebruikt voor het construeren van statistische toetsen voor STAR-achtige niet-lineariteit. Aangetoond wordt dat deze toetsen resistent zijn voor uitschieters, en dat een lineaire tijdreeks die besmet is met enkele uitschieters niet per abuis wordt aangezien voor een niet-lineaire reeks. De toepassing in dit hoofdstuk op groeicijfers in industriële productie voor een aantal OECD landen suggereert dat voor een (beperkt) aantal landen de niet-lineariteit welke gevonden wordt met behulp van standaard toetsen in feite te wijten is aan enkele uitschieters.

Hoofdstuk 7 is gemotiveerd door de mogelijkheid dat niet-lineariteit en uitschieters ook tegelijkertijd kunnen voorkomen. De effecten van uitschieters op bijvoorbeeld schattingen van parameters in niet-lineaire modellen heeft tot op heden slechts weinig aandacht gekregen in de literatuur. Een mogelijke verklaring voor dit gegeven is dat een dergelijke analyse bijzonder gecompliceerd is en verder bemoeilijkt wordt door het feit dat aanpakken die gevolgd worden voor de analyse van dergelijke effecten in lineaire modellen niet toepasbaar zijn in de context van niet-lineaire modellen. In dit hoofdstuk is gekozen voor een pragmatische aanpak, en worden de standaard kleinste kwadraten methode en een aantal uitschieter-robuuste schattingsmethoden vergeleken op basis van een simulatie-experiment. De resultaten laten zien dat uitschieters vergelijkbare effecten op parameter-schattingen in een STAR model hebben als in een lineair model. Daarnaast lijken standaard uitschieter-robuuste methoden slechts onder bijzondere omstandigheden bruikbaar.

The Tinbergen Institute is the Netherlands Research Institute and Graduate School of Economics, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University in Rotterdam, the University of Amsterdam and the Free University in Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. If available, trade editions of the books which are published in the Tinbergen Institute Research Series can be ordered through Thela Thesis, Prinseneiland 305, 1013 LP Amsterdam, the Netherlands, phone: +3120 6255429; fax: +3120 6203395. The following books recently appeared in this series:

140. L.W.N. VAN DER TORRE, *Reasoning about obligations. Defeasibility in preference-based deontic logic.*
141. W. BOLT, *Theory of negotiation: a game-theoretic approach.*
142. M.A.J. SMITH, *On the availability of failure prone systems.*
143. R. SLOOF, *Game-theoretic models of the political influence of interest groups.*
144. J.N.M. WIT, *Dynamics and information aggregation in elections and markets.*
145. N.T. BUDINA, *Essays on consistency of fiscal and monetary policy in Eastern Europe.*
146. F. LIPPI, *Central Bank independence and credibility: essays on delegation arrangements for monetary policy.*
147. H. HOEK, *Variable trends: a Bayesian perspective.*
148. E.J.S. PLUG, *Leyden welfare and beyond.*
149. C.S.J. JIE-A-JOEN, *Strategic trade policy, multinational firms, and vertically related industries.*
150. R.G.J. DEN HERTOG, *Empirical essays on growth, uncertainty and prediction.*
151. M. GELDERMAN, *Success of management support systems: a review of the literature and an empirical investigation.*
152. M.A. VAN DER ENDE, *On the values of jobs and specific training.*
153. M. KLAPWIJK, *Rural industry clusters in central Java, Indonesia; an empirical assessment of their role in rural industrialization.*
154. R.J.G. VENNIKER, *Transition processes from constrained to Walrasian equilibrium.*
155. J.M. MORALEDA, *On the pricing of interest-rate options.*
156. J.F. STURM, *Primal-dual interior point approach to semidefinite programming.*
157. R. FRELING, *Models and techniques for integrating vehicle and crew scheduling.*
158. S. BORGHGRAEF, *Is more control better? The effects of organizational control on salespersons' job performance, job attitudes and sales organization effectiveness.*
159. C. VINKENBURG, *Managerial behavior and effectiveness. Determinants, measurement issues and gender differences.*

160. P.A. GAUTIER, *The flow approach to the labor market.*
161. A.A.R.J.M. VAN RIXTEL, *Informality, bureaucratic control and monetary policy: The case of Japan.*
162. Y. GAO, *Management support of information system project planning.*
163. J.G. PIEK, *Development and acquisition of technology in rural collective industries. The case of China's hinterland.*
164. F. LANGERAK, *The effect of market orientation on the performance of industrial firms: A study in the Dutch Metalelektro.*
165. A.K. SINHA, *Estimating failure probability when failure is rare: multi- dimensional case.*
166. J.P.M. SIJM, *Food security and policy interventions in sub-Saharan Africa. Lessons from the past two decades.*
167. W.A. LETTERIE, *Partisan politics, persuasion and information.*
168. R. VAN DIJK, *Corporate finance policy and equity investment: panel data analyses.*
169. M.A. CARREE, *Market dynamics, evolution and smallness.*
170. J.E. LIGTHART, *Environment, imperfect markets, and public finance.*
171. R. PAAP, *Markov trends in macroeconomic time series.*
172. J.H. ABBRING, *Essays in labour economics.*
173. J.P. COHEN, *Utility: a real thing. A study of utility's ontological status.*
174. G.A.R. VAN HOESEL, *Beyond export-led growth: the emergence of new multinational enterprises from Korea and Taiwan.*
175. H.H.J. ZEBREGS, *International capital movements and technology in economic development.*
176. M.F. CORNET, *Game-theoretic models of bargaining and externalities.*
177. M.N. BOUMAN, *Environmental costs and capital flight.*
178. L. PENG, *Second order condition and extreme value theory.*
179. P.W.C. KONING, *Structural empirical analysis of labor market frictions, job search and wage formation.*
180. C.W.M. NAASTEPAD, *The public sector budget and macroeconomic performance: A real-financial CGE analysis with portfolio choice with reference to India.*
181. S.T.M. STRAETMANS, *Extreme financial returns and their comovements.*
182. J.P. KOOIMAN, *Topics in the Economics of Environmental Regulation.*
183. P.P.A.A.H. KANDELAARS, *Material-product chains: Economic models and applications.*
184. N. VAN GIERSBERGEN, *Bootstrapping dynamic econometric models.*
185. K.M. SCHILSTRA, *Industrial relations and human resource management: A network approach.*
186. H. SUPRATIKNO, *Competitive strategy, production organization and interfirm collaboration. A case study of subcontracting arrangements in three manufacturing firms in Indonesia.*
187. J. GOUWENS, *Interne consistentie in personeelsmanagement: Naar een typering van HRM in grote bedrijven in Nederland.*

188. F. POT, *Continuity and change of human resource management: A comparative analysis of the impact of global change and cultural continuity on the management of labour between the Netherlands and the United States.*
189. M.J. KLEIJN, *Demand differentiation in inventory systems.*
190. J.F.M. SWEEGERS, *Coordination, cooperation and institutions.*
191. P.A. DE HEK, *Economic growth and fluctuations: A theoretical enquiry.*
192. J. MEIJAARD, *Decision making in research and development: A comparative study of multinational companies in the Netherlands and the United States.*
193. K.P.B. OLDENKAMP, *Derivatives in portfolio management.*
194. C.M.M.P. WETZELS, *Squeezing birth into working life. Household panel data analyses comparing Germany, Great Britain, Sweden and the Netherlands.*
195. J.M. VAN LEEUWEN, *Armoede en koopkrachtongelijkheid; theoretische beschouwingen over het effect van inkomensafhankelijke prijzen.*
196. P. FRIJTERS, *Explorations of welfare and well-being.*
197. J.A. CABRAL VIEIRA, *The evolution of wage structures in Portugal 1982-1992.*
198. M.J. SMITS, *Technology choice and the role of institutions. The case of wheat and the input market in Pakistan.*
199. M.T. KÖGEL, *Services and economic growth.*