Essays on nonlinearity in economic time series

Von der Wirtschaftswissenschaftlichen Fakultät der

Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des akademischen Grades

Doktor der Wirtschaftswissenschaften

– Doctor rerum politicarum –

genehmigte Dissertation

von

Diplom - Ökonom Florian Heinen

geboren am 29. Oktober 1981 in Bielefeld

2011

Referent: Prof. Dr. Philipp Sibbertsen, Leibniz Universität Hannover

Koreferent: Prof. Dr. Lukas Menkhoff, Leibniz Universität Hannover

Tag der Promotion: 16.08.2011

We have not succeeded in answering all our problems. The answers we have found only serve to raise a whole set of new questions. In some ways we feel we are as confused as ever, but we believe we are confused on a higher level and about more important things.

– Posted outside the mathematics reading room,
Tromsø University; cited from Øksendahl (2003).

Acknowledgements

Several people contributed to this thesis in different ways.
First and foremost I am greatly indebted to my advisor and co-author Prof. Dr. Philipp Sibbertsen. He gave me a lot of liberties and constant encouragement while working on this thesis. I benefited greatly from his insight and experience.
I am also thankful to Prof. Dr. Lukas Menkhoff for agreeing to be the second examiner of this thesis and to Prof. Dr. Olaf Hübler for chairing my examination board.

My other co-authors Dr. Stefanie Michael, Dipl.-Ök. Hendrik Kaufmann and Dipl.-Ök. Juliane Willert who worked with me on parts of this thesis deserve my thanks for sharing the workload with me.

During the last three years while working on this thesis my colleagues at the Institute of Statistics always provided a good working environment which made the research both fun and productive. Special thanks are due to Stefanie for helping me with several LaTeX problems and sharing various templates which made the work much easier and to Juliane for helping me with various computer and (especially) administrative issues.

Special thanks are also due to my parents for their constant support during the last years and for fostering my education. I hope that this thesis is a better answer to what I have done over the last years than all the "Good"s and "Lot's of work"s in the past.

Throughout my research I received financial support from the Deutsche Forschungsgemeinschaft (DFG) which I gratefully acknowledge.

Hannover, August 2011

Florian Heinen

Kurzfassung

Die Analyse nichtlinearen Verhaltens ökonomischer Zeitreihen, hat aufgrund ihrer Implikationen für die Anwendung und Analyse ökonomischer Theorien einer lange Tradition in der Statistik und Ökonometrie. Diese Arbeit beinhaltet fünf Kapitel, die sich mit unterschiedlichen Aspekten von nichtlinearem Verhalten in ökonomischen Zeitreihen beschäftigen.
Kapitel 2, verfasst mit Stefanie Michael und Philipp Sibbertsen, untersucht das exponential smooth transition autoregressive model (ESTAR). Dieses Modell besitzt die Eigenschaft, dass der Parameter der die Nichtlinearität steuert schwierig zu schätzen ist. Dieses Kapitel beinhaltet eine detailierte Studie der Eigenschaften dieses Parameters und es wird gezeigt, dass er unter bestimmten Bedingungen nicht identifizierbar ist. Es wird ein neues Modell, das TSTAR Modell, vorgeschlagen, dass diese Eigenschaft verbessert. Zusätzlich wird ein Linearitätstest sowie ein Einheitswurzeltest entwickelt. Kapitel 3 erweitert die Betrachtung des TSTAR Modells, indem verschiedene Fehlspezifikationstests für dieses Modell entwickelt werden.
Kapitel 4, verfasst mit Hendrik Kaufmann und Philipp Sibbertsen, untersucht die Theorie der Kaufkraftparität aus der Perspektive der Ökonometrie. Da die Modellierung realer Wechselkurse, als Maß für die Abweichung von der Kaufkraftparität, mit unterschiedlichen Modellen erfolgen kann, die wiederum unterschiedliche ökonomische Erklärungsansätze zulassen, ist es von Interesse diese Modelle unterscheiden zu können. In Kapitel 4 wird eine solche Methode auf Basis rechenintensiver Verfahren vorgeschlagen und auf verschiedene reale Wechselkurse angewand.
Das Kapitel 5 beschäftigt sich mit der Frage, ob Standardinferenztechniken der Statistik durch nichtlineare Datentransformation in ihren Eigenschaften beeinflusst werden. Da in der Ökonomie häufig eine logarithmische Transformation der Daten vorgenommen wird, ist es wichtig zu untersuchen, ob und inwiefern die Testresultate ihre Aussagekraft behalten.
Kapitel 6, verfasst mit Juliane Willert, nähert sich den Nichtlinearitäten aus einem anderen Blickwinkel. Da nichtlineares Verhalten häufig durch Regimewechselmodelle beschrieben wird, sind Strukturänderungen und Nichtlinearitäten eng verwand. Dieses Kapitel entwickelt eine Methode, die es erlaubt eine Strukturänderung der Persistenz einer Zeitreihe aufzudecken, sobald neue Daten eintreffen. Die zusätzliche Schwierigkeit ergibt sich dadurch, dass die Zeitreihe eine langfristige Korrelationstruktur aufweisen kann.

Schlagwörter: Nichtlinearitäten, Spezifikationstests, Identifikationsproblem, Kaufkraftparität

Short summary

The analysis of nonlinear behavior in economic time series has a long standing tradition in econometrics and statistics due to the implications for the application and analysis of economic theories. This collection of five essays deals with different aspects of nonlinearity in economic time series.

Chapter 2, co-authored with Stefanie Michael and Philipp Sibbertsen, analyzes the exponential smooth transition autoregressive model (ESTAR). This model possesses the property that it is hard to obtain a reliable estimate for the parameter that governs the nonlinearity. This chapter contains a detailed study of the properties of this parameter and we are able to show that it is unidentified under certain conditions. A new model, termed the TSTAR model, is proposed that improves on the ESTAR in terms of identifiability. Additionally a linearity test and an unit root test is proposed for this model. Chapter 3 extends the study of the TSTAR model by proposing various kinds of misspecification tests.

Chapter 4, co-authored with Hendrik Kaufmann and Philipp Sibbertsen, deals with the purchasing power parity from an econometric perspective. Because modeling real exchange rates, as a measure of deviation from purchasing power parity, can be accomplished using different models that might lead to different economic theories it is of interest to discriminate between these models. In chapter 4 such a method, based on computational intensive techniques, is proposed and applied to various real exchange rates. Chapter 5 deals with the question whether the properties of standard inference techniques are affected by nonlinear data transformation. Because in applied economics the logarithm is applied frequently to the data it is important to know whether the test results remain reliable. Chapter 6, co-authored with Juliane Willert, approaches nonlinearity from a different angle. As nonlinearities are frequently captured by regime switching models, structural change and nonlinearity are closely intertwined. This chapter develops a method that allows to detect changes in the persistence property of a time series whenever new data arrives. The additional difficulty is that the time series under study is allowed to display long range dependency.

Keywords: Nonlinearities, Specification testing, Identificationproblem, Purchasing Power Parity

Contents

1 Introduction **1**

2 Two competitive models and their identification problem: The ESTAR and TSTAR model **5**

2.1 Introduction . . . 6
2.2 Exponential Smooth Transition Autoregressive Models . . . 7
2.2.1 The Identification Problem of the ESTAR model . . . 10
2.3 The TSTAR Model . . . 12
2.3.1 Linearity testing . . . 14
2.3.2 Unit Root Testing . . . 16
2.3.3 Monte Carlo Simulations . . . 19
2.4 Empirical Illustration . . . 25
2.5 Conclusions . . . 29
2.6 Appendix: Proofs and Technical Lemmas . . . 30

3 Evaluating a class of nonlinear time series models **43**

3.1 Introduction . . . 44
3.2 The modeling cycle . . . 45
3.3 Choosing the transition function . . . 47
3.4 Residual based misspecification tests . . . 48
3.4.1 Test of serial independence . . . 48
3.4.2 Test of no remaining nonlinearity . . . 50
3.4.3 Test of parameter constancy . . . 51
3.5 Generalized impulse response function . . . 53
3.6 Finite sample properties . . . 54
3.7 Modeling real exchange rates . . . 58
3.8 Conclusion . . . 62
3.9 Appendix . . . 63
3.9.1 Regression coefficients in (3.23) . . . 63
3.9.2 Regression coefficients in (3.29) . . . 63

4 The dynamics of real exchange rates – A reconsideration **65**
4.1 Introduction . . . 66
4.2 Two Competing models for real exchange rates . . . 68
4.3 Testing non-nested hypotheses . . . 70
4.3.1 A bootstrap based likelihood ratio test . . . 71
4.3.2 Finite sample properties . . . 73
4.4 Modeling real exchange rates . . . 74
4.5 Conclusion . . . 80

5 A note on testing for purchasing power parity **81**
5.1 Introduction . . . 82
5.2 Unit root tests and their asymptotic behavior under misspecification . . . 83
5.3 Monte Carlo evidence . . . 85
5.4 Conclusion . . . 91
5.5 Appendix . . . 93
5.5.1 Proof of Theorem 5.2.1 . . . 93
5.5.2 Proof of Theorem 5.2.2 . . . 94

6 Monitoring a change in persistence of a long range dependent time series **96**
6.1 Introduction . . . 97
6.2 Monitoring a change in persistence . . . 98
6.3 Monte Carlo evidence . . . 101
6.4 Empirical Application . . . 106
6.5 Conclusion . . . 108
6.6 Appendix . . . 109
6.6.1 Proof of Theorem 6.2.1 . . . 109
6.6.2 Proof of Theorem 6.2.2 . . . 112

Bibliography **113**

Chapter 1

Introduction

Introduction

The study of nonlinearity in time series has a long standing tradition in the statistics and econometrics literature. Early contributions date back to Bacon and Watts (1971), Jones (1978), Ozaki (1980), Haggan and Ozaki (1981), White and Domowitz (1984), Tjøstheim (1986) or Tong (1990).
However, it was not until the contributions of Hamilton (1989) and Teräsvirta (1994) that economists have started to use nonlinear time series models more often in empirical applications. In the following two decades nonlinear time series models have been used to successfully describe various economic time series such as real exchange rates, nominal exchange rates, real interest rates, gross domestic product or US unemployment rate, to name but a few (see e.g Taylor et al. (2001), Franses and van Dijk (2000), Kapetanios et al. (2003), Potter (1995), Pesaran and Potter (1997) or van Dijk et al. (2002b)). For economic time series, models which provide a state- or regime-dependent dynamic have been most successful and enjoyed most attention both from theoretical studies and empirical applications (see e.g. the special issues of the *Journal of Econometrics (1996, 2002)* and *Studies in Nonlinear Dynamics & Econometrics (2008a,b)*). Teräsvirta et al. (2010) provide a recent up-to-date overview of nonlinear time series models and their application. This growing interest in nonlinear time series models can be explained by the fact that "linearity is at best a convenient artefact, and because the world is full of nonlinear phenomena such as limit cycles and jump resonance, we need to study nonlinearity to explore the nature" (Chen and Tsay (1993)). The additional increase in computer capacity has further reduced the burden of conducting nonlinear analysis and thus improved applicability.
Although substantial progress has been made over the last years to better understand the behavior of nonlinear time series and also various tools for statistical inference have been proposed many questions are still unanswered. This is mainly due to the myriad of possible forms of nonlinearity encountered.
Some open questions are for example:

- *How do parameters in nonlinear dynamic models interact?*
- *Do they affect each other?*
- *How can we discriminate between different, non-nested nonlinear models?*
- *Are standard inference techniques affected by nonlinear data transformations?*

The first two questions are vital as they refer to reliable parameter estimates. Although there are consistency results for the parameter estimates available in the literature (see Tjøstheim (1986)) there is at the same time a consensus that certain parameters are notoriously hard to estimate reliably (see e.g. Luukkonen et al. (1988) or Teräsvirta (2004)). This gap between theoretical results and empirical experience leaves room to explore the interaction between the parameters of nonlinear dynamic models in more depth as it directly affects the applicability of the models. The third question alludes to evaluate the adequacy of two competing non-nested models, a classical topic in statistical inference. Discriminating between two nonlinear models is

important because different models might lead to different economic intuitions and it is therefore of interest which theory best describes the empirical data. Further, if the models are nonlinear and non-nested, i.e. they are not special cases of each other, standard inference techniques fail. The last question is of a more technical nature. There are often good reasons to apply the logarithm to empirical data, especially with financial data. For example one often wants to stabilize the variance, wants to linearize exponential growth or wants to obtain approximately percentage growth if the logarithm is applied to first differences. In many applications however it is not a priori clear whether to apply the logarithm to the data or not. As many economic theories, especially equilibrium theories such as the law of one price or the Fisher hypotheses, rely on testing for a unit root it is of of interest for the applied economist whether the conclusions drawn from standard inference are reliable or biased in some direction if the logarithm is applied, possibly erroneously.
This thesis attempts to study these questions more thoroughly. It comprises of five self-contained chapters dealing with different aspects of nonlinear time series models. The common theme of these essays is statistical inference for univariate, nonlinear time series models.

Chapter 2, co-authored with Stefanie Michael and Philipp Sibbertsen, considers the exponential smooth transition autoregressive (ESTAR) model. Although popular in empirical applications, it is well known that it is very hard to obtain reliable estimates for the parameters of the ESTAR model. In this chapter we undertake a detailed study of the crucial parameter that governs the degree of nonlinearity. We are able show that if the variance of the innovation process becomes very small the parameter becomes unidentified in the sense that the standard deviation goes to infinity. This has important implications for the interpretability of such models. In order to improve on the ESTAR model in terms of identifiability we propose a new model formulation, termed the TSTAR model, that maintains the desirable properties of the ESTAR model. To ease the empirical application of this newly developed model we also propose a linearity test and a unit root test in this nonlinear framework.

Chapter 3 considers the TSTAR model in more detail. We focus mainly on misspecification testing for models of such type. Following the modeling cycle for nonlinear time series models of specification, estimation and evaluation we first treat how to choose an adequate transition function and then contribute to the evaluation stage by proposing tests against serial correlation, no remaining nonlinearity and parameter constancy.

Chapter 4, co-authored with Hendrik Kaufmann and Philipp Sibbertsen, considers a long standing problem in economics, namely the purchasing power parity (PPP). While it is widely agreed that PPP holds as a long-run concept the specific dynamic driving the process is largely build upon a priori economic belief rather than a thorough statistical modeling procedure. The two prevailing time series models, i.e. the ESTAR model and the Markov switching autoregressive (MSAR) model, are both able to support the PPP as a long-run concept. However, the dynamic behavior of real exchange rates implied by these two models is very different and leads to different economic interpretations. In this chapter we approach this problem by offering a bootstrap based testing procedure to discriminate between these two rival models. This helps us to base

the modeling of PPP on a thorough statistical ground.

Chapter 5 is a theoretical contribution that examines the behavior of unit root tests against nonlinear alternatives of the ESTAR type if the data is erroneously nonlinearly transformed. We show analytically and by a Monte Carlo study that the probability of rejecting the correct null of a random walk depends on the type of data transformation.

Nonlinear time series models such as the ESTAR or the MSAR model describe the data generating process as consisting of different regimes with different dynamic behavior at different points in time. Therefore nonlinear state-dependent dynamics and structural change are closely intertwined. Nonlinear time series model prove useful whenever there are two or more regimes assumed to be present in the data. A different perspective on this regime changing behavior is taken in the concluding chapter 6, co-authored with Juliane Willert. It deals with a single change in the long-run correlation structure of a time series at some unknown future point in time. We propose a monitoring technique that allows us to detect such a change in persistence whenever new data arrives. An additional difficulty arises because we cast this method into the framework of long range dependent time series.

Chapter 2

Two competitive models and their identification problem: The ESTAR and TSTAR model

Two competitive models and their identification problem: The ESTAR and TSTAR model

Co-authored with Stefanie Michael and Philipp Sibbertsen.
Published as Leibniz University of Hannover Discussion Paper No. 474.

2.1 Introduction

Nonlinear time series models have become more and more popular over the last decade. In particular, Exponential Smooth Transition Autoregressive (ESTAR) models have been found attractive for modeling real exchange rates. These models contain of two autoregressive regimes which are connected by a smooth transition function of an exponential type. Under certain regularity conditions they are globally stationary even if one regime is assumed to be a random walk. Moreover, the U-shape of the transition function is a desired property in the context of real exchange rates.

Contrary to linear models where parameter estimates are independent of the size of the error variable, we face an identification problem in ESTAR models due to the influence of the error term variance. Small or large error variances no longer allow to identify the ESTAR model in the sense that the variance of parameter estimates tends to infinity. This happens intuitively because the process stays only in one of the two regimes and does not switch between the regimes any more. This problem was first observed in Luukkonen et al. (1988) who mention it in a short remark. However, to the best of our knowledge it has not been further considered in the literature since.

The size of the error variance leads to an artificial linearization of the process and thus causes troubles in linearity and nonlinear unit root tests. Linearity tests against ESTAR have been developed by for example Teräsvirta (1994) and unit root tests against an ESTAR alternative can be found among others in Kapetanios et al. (2003). Both of these tests have the nonintuitive property of a low power when the error term variance is either very small or very large. For the linearity test this was stated in Luukkonen et al. (1988) and for the unit root test see Kruse et al. (2010). This effect is less surprising for a large error term variance as in this case the noise dominates the signal. However, it is rather surprising in the opposite case of a small error variance as in this case the signal dominates the noise. Therefore, an increase of the power would be expected. As real exchange rates have extremely small error variances (see for example Taylor et al., 2001) this problem is of a high practical relevance and can lead to false non-rejections of the null and therefore rejecting a nonlinear adjustment process for real exchange rates.

This effect is independent of the well known estimation problem of the transition parameter in ESTAR models. In order to circumvent this problem various ideas have been proposed (see e.g. Haggan and Ozaki (1981), Teräsvirta (2004)) to guarantee a better performance of the

estimators. Here a high though finite variance of the parameter estimate is obtained whereas in our situation the variance becomes unbounded. Therefore, no general estimation procedure will produce reasonable estimators, making some modification to the optimization procedure necessary. This has of course its limits as there is no theory saying that these methods work in general.

We introduce an alternative of the ESTAR model by using a different transition function, leading to the TSTAR model. This transition function possess the same desired properties as the exponential function and can therefore be applied to the same situations. The new transition function has however fatter tails which turns out to reduce the identification problem. We can improve the estimation procedure for extreme error term variances. In particular, standard optimization tools can be used. Moreover, we develop a linearity and a unit root test for this new model and study their performances in extensive simulations.

The rest of the paper is organized as follows. In the next section we define ESTAR models in more detail and analyze the identification problem, in particular with respect to small error term variances. In order to do so we derive results about the moment behavior of ESTAR models which might be of some interest also in another context. The new TSTAR model is examined in Section 2.3. After describing the model we derive the linearity as well as the unit root test in Sections 2.3.1 and 2.3.2, respectively. The simulation studies we performed are summarized in Section 2.3.3. A comparison of the ESTAR and TSTAR model is presented in Section 2.4, discussing real exchange data. Section 2.5 concludes whereas all proofs are collected in the Appendix, together with certain technical lemmas.

2.2 Exponential Smooth Transition Autoregressive Models

One speaks of a Smooth Transition Autoregressive (STAR) model, if two autoregressive regimes are connected by a transition function which satisfies certain smoothness conditions. In general, a univariate stochastic process $\{y_t\}_{t\in\mathbb{Z}}$ is called STAR(p), $p \geq 1$, if

$$y_t = [\Psi w_t] \cdot [1 - G(y_{t-d}; \gamma, c)] + [\Theta w_t] \cdot G(y_{t-d}; \gamma, c) + \varepsilon_t. \tag{2.1}$$

The parameter vectors Ψ and Θ as well as w_t are given by $\Psi = (\psi_0, \psi_1, \ldots, \psi_p)$, $\Theta = (\vartheta_0, \vartheta_1, \ldots, \vartheta_p)$, and $w_t = (1, y_{t-1}, \ldots, y_{t-p})'$. The transition function $G(\cdot; \gamma, c) : \mathbb{R} \to [0,1]$ governs the transition between the two autoregressive regimes Ψw_t and Θw_t in a smooth way. Alternatively, a STAR model can also be interpreted as a continuum of regimes which is passed through by the process. However, note that (2.1) can alternatively be written in somewhat less intuitive representations

$$y_t = [\Psi w_t] + [\Phi w_t] \cdot G(y_{t-d}; \gamma, c) + \varepsilon_t \tag{2.2}$$

$$= [\Xi w_t] \cdot [1 - G(y_{t-d}; \gamma, c)] + [\Theta w_t] + \varepsilon_t \tag{2.3}$$

with $\Phi = \Theta - \Psi$ and $\Xi = \Psi - \Theta$.

Different choices of the transition function G lead to different STAR models. Common choices

are the exponential function, leading to the Exponential STAR (ESTAR) model, or the logistic function, depending on the nature of the studied transition. The parameter γ is always the transition parameter that governs the speed of the regime changes. A recent overview of STAR models, estimation techniques and model building procedures can be found in Franses and van Dijk (2000). However, general STAR models have not yet been studied systematically, if possible at all.

In this paper we will consider functions G that are symmetrically U-shaped around the location parameter $c \in \mathbb{R}$ with

$$\lim_{\gamma \to +\infty} G(\cdot;\gamma,c) \equiv 1 - \mathbf{1}_c, \lim_{\gamma \to 0} G(\cdot;\gamma,c) \equiv 0 \text{ and } \lim_{z \to \pm\infty} G(z;\gamma,c) \equiv 1 \tag{2.4}$$

where $\mathbf{1}_c$ denotes the indicator function being one only at the value c. The ESTAR model, for instance, is defined by taking G as

$$G(z;\gamma,c) = 1 - \exp(-\gamma(z-c)^2), z \in \mathbb{R}. \tag{2.5}$$

As one often chooses $p = 1$ in practical applications, we restrict ourselves in this text to the case $p = d = 1$. Moreover, we assume according to (2.3) $\xi_0 = \vartheta_0 = 0$, $c = 0, \xi = \xi_1$ and $\vartheta = \vartheta_1$ and obtain the process

$$y_t = \left[\xi \exp\left(-\gamma y_{t-1}^2\right) + \vartheta\right] y_{t-1} + \varepsilon_t, t \in \mathbb{Z}. \tag{2.6}$$

Example 2.2.1.

Figures 2.1 and 2.2 show two realizations of the ESTAR(1) process (2.6) *of length* $T = 500$*, both generated with* $\vartheta = 1$ *and* $\xi = -0.45$*. The variances are chosen as* $\sigma^2 = 4$ *and* $\sigma^2 = 0.04$*, respectively.*

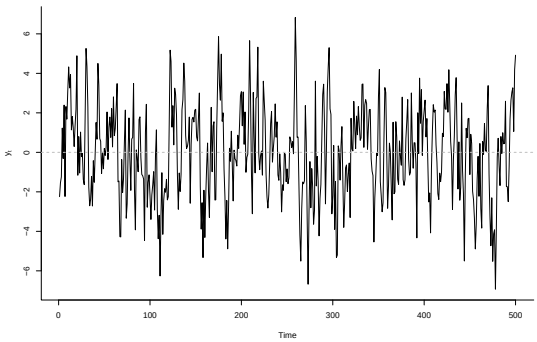

Figure 2.1: $\sigma = 2, \vartheta = 1$ and $\xi = -0.45$.

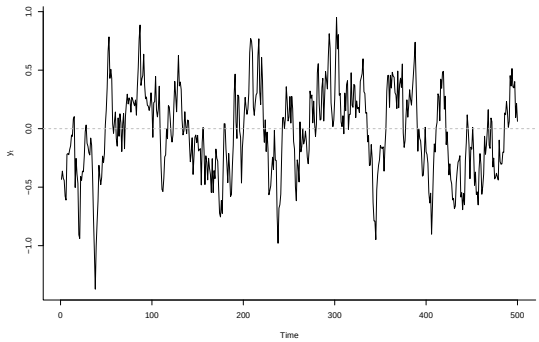

Figure 2.2: $\sigma = 0.2, \vartheta = 1$ and $\xi = -0.45$.

In order to prove statements about the ESTAR(1) model, the random error terms $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ are assumed to satisfy the following conditions.

Assumption 2.2.2.

(i) *The innovations* $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ *are assumed to have a symmetric density around zero with full support* $\mathbb{R}$ *and with* $\mathbb{E}[\varepsilon_t^{2k}] = c_{\varepsilon,k}\sigma^{2k}$ *for* $k \in \mathbb{N}_0$ *and constants* $c_{\varepsilon,k}$.

(ii) *For some* $p > \beta > 2$, $\{\varepsilon_t\}$ *is a strong mixing sequence with mixing coefficients* α_m *of size* $-p\beta/(p-\beta)$ *and* $\sup_{i\leq 1}\|\varepsilon_i\|_p = C < \infty$. *In addition,* $(1/T)E\left[(\sum_{i=1}^{T}\varepsilon_i)^2\right]\lambda^2 > 0$ *for* $T \to \infty$.

The first assumption is of technical nature to derive properties of y_t. It is in particular satisfied if $\varepsilon_t \sim \mathrm{N}(0,\sigma^2)$. Assumption (ii) makes sure that the error term has a flexible structure allowing for various forms of temporal dependence and heteroscedasticity.
In order to verify basic properties of the ESTAR(1) model, we interpret the process $\{y_t\}_{t\in\mathbb{Z}}$ as a functional coefficient autoregressive model (see Chen and Tsay (1993)) and hence as a homogeneous Markov chain with state space $\mathbb{R}$ equipped with the Borel $\sigma-$algebra. Then geometric ergodicity guarantees the existence and uniqueness of s stationary distribution F of y_t. Sufficient conditions for ergodicity (see also Tjøstheim (1990), Tweedie (1975)) and some properties of the moments of $\{y_t\}_{t\in\mathbb{Z}}$ that will play an important role in the following section are summarized in the following Lemma, proven in the Appendix.

Lemma 2.2.3 (Moment properties of y_t)**.**
Let $\{y_t\}_{t\in\mathbb{Z}}$ *be as in* (2.6) *with* $|\xi| + |\vartheta| < 1$ *and with innovations* $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ *that satisfy Assumption 2.2.2.*

(i) Then $\{y_t\}_{t\in\mathbb{Z}}$ *is geometrically ergodic, and in particular strictly stationary.*

(ii) The density of y_t *is symmetric for all* $t \in \mathbb{Z}$, *and in particular the stationary distribution has a symmetric density,*

(iii) For $n, k \in \mathbb{N}_0$ *and all* $t \in \mathbb{Z}$,

$$\mathbb{E}\left[\exp(-n\gamma y_t^2) y_t^{2k+1}\right] \;=\; 0. \tag{2.7}$$

(iv) For $n \in \mathbb{N}_0, k \in \mathbb{N}$ *and all* $t \in \mathbb{Z}$, *there exists a constant* c_k *such that*

$$\lim_{\sigma\downarrow 0}\left|\frac{\mathbb{E}\left[\exp(-n\gamma y_t^2) y_t^{2k}\right]}{\sigma^{2k}} - c_k\right| \;=\; 0 \tag{2.8}$$

and, for $J = \{j = 0,\ldots,2k : j \ \mathit{mod}\ 2 = 0\}$,

$$c_k \;=\; \frac{1}{1-(\vartheta+\xi)^{2k}} \sum_{j\in J\setminus\{0\}} \binom{2k}{j} c_{k-j/2}\, c_{\varepsilon,j}\, (\vartheta+\xi)^{2k-j},\ k \in \mathbb{N},$$

with $c_0 = 1$.

Remarks.

- Symmetry of the stationary distribution is in line with Jones (1978) who approximated the stationary distribution by Taylor approximations and derives a picture of the approximated stationary density for an ESTAR(1) process (see Figure 3c), p.93).
- Formula (2.7) implies that all odd moments of y_t vanish for all $t \geq 0$ by choosing $n = 0$. The even moments behave like (2.8) again for $n = 0$.

2.2.1 The Identification Problem of the ESTAR model

If transition functions G_1 and G_2, resulting from different parameter combinations in the ESTAR setting, cannot be distinguished, it is obviously nearly impossible to fit a 'good' model to given data. Whenever changes of the parameters do not result in significant changes of the transition function, we speak of the so-called identification problem. Due to (2.6) this happens in the ESTAR setting for extreme values (i.e. large values or values close to zero) of γy_{t-1}^2, caused either by γ or by y_{t-1}^2. The latter turns out to occur for very small or very large values of the error term variance σ^2. This observation is clearly in contrast to linear models and was mentioned by Luukkonen et al. (1988) but has not been studied further in the literature since.

We are aware that this problem is different from other settings of non-identification such as non-identified parameters under a null or alternative hypothesis as it occurs when testing STAR models. The problem we consider leads to the impossibility of estimating the transition parameter in the sense that its variance tends to infinity. Our problem is in line with findings by Nelson and Startz (2007) who show that parameter identification problems can occur in nonlinear models when a model parameter tends to a specific limit. However, our situation goes beyond the findings of Nelson and Startz (2007) as they consider only the case of explicitly used model parameters whereas in this paper we consider the effect of a scaling parameter of the error variance which has only implicit effects on the model.

Before giving more profound results, we want to describe the intuition behind the identification problem. With respect to γ, it is obvious that for different large values (say roughly $\gamma > 1$), the corresponding transition functions hardly change any more. As for y_t^2, one can already see from Figures 2.1 and 2.2, which show realizations of ESTAR processes with different values of σ^2, that -while only appearing implicitly in the definition (2.6)- the error term variance influences the behavior of the process. Large values for σ^2 allow the error term to dominate the process, resulting in large values for y_t and causing the identification problem, independent of the choice of γ. On the other hand, very small values for σ^2 result in small values of y_t.

As a consequence of the identification problem, the transition function G in the ESTAR model is either close to zero or close to one. This means that one of the two regimes is no longer present. The transition parameter γ as well as one of the autoregressive parameters are therefore unidentified and can not be estimated consistently. To illustrate this behavior, we estimate the parameter vector (ψ, φ, γ) by means of the conditional Maximum Likelihood method for the model $y_t = 0.3y_{t-1} + 0.65y_{t-1}G(\cdot) + \varepsilon_t$. The resulting highly biased estimators for γ using different choices of σ are summarized in Table 2.1. Note also, that ψ as well as φ are estimated quite well showing that it is indeed $G(\cdot)$ that cannot be determined in a good way.

γ \ σ		0.1		0.5	
		Mean	SD	Mean	SD
0.5	$\hat{\psi}$	0.315	0.117	0.311	0.358
	$\hat{\varphi}$	0.646	0.102	0.372	0.391
	$\hat{\gamma}$	2.850	75.09	78.736	1065.299
0.8	$\hat{\psi}$	0.312	0.121	0.294	0.354
	$\hat{\varphi}$	0.649	0.101	0.370	0.397
	$\hat{\gamma}$	1.694	12.288	64.777	717.136
1.5	$\hat{\psi}$	0.311	0.129	0.286	0.334
	$\hat{\varphi}$	0.646	0.109	0.357	0.404
	$\hat{\gamma}$	2.953	27.719	98.055	1498.332

Table 2.1: Estimation results for ESTAR: $y_t = 0.3y_{t-1} + 0.65y_{t-1}G(\cdot) + \varepsilon_t$

It is definitely worth studying this phenomena as it is in particular counter intuitive that tiny error term variances do not allow for good estimators as one would expect to observe (and estimate) the process well. Moreover, although not called identification problem, people are aware of the problems (e.g. Teräsvirta et al., 2010, p.381) and a lot of subjective 'tricks' have been proposed and used to circumvent them, allowing for a broader range for γ and σ without experiencing unidentified parameters. The common idea is to exclude γ from the estimation process and use an alternative way to fit the model. Haggan and Ozaki (1981) propose, for instance, to define a grid for γ and estimate only the remaining parameters, followed by a search for the best γ. By doing so, they do not estimate the transition variable γ. In order to reduce the influence of σ, Teräsvirta (2004, p.229) standardizes the exponent present in G by writing

$$G(y_t;\gamma,c) \;=\; 1-\exp(-\gamma y_t^2) \;=\; 1-\exp\left(-\gamma\hat{\sigma}^2\cdot\left(\frac{y_t^2}{\hat{\sigma}^2}\right)\right),$$

where $\hat{\sigma}$ is the standard deviation, in oder to obtain a scale free γ. However, this is not the case as the resulting Volterra series (see Priestley, 1988, p.25) is not bounded.

Although these modifications seem to help in certain situations they are not quite satisfying as it is hard to reproduce the parameter estimates and as they have not been studied well mathematically. However, it would indeed be desirable to have a mathematical unified approach for the estimation problem, in particular for very small σ^2, as one does find tiny estimated values $\hat{\sigma}$ in practical applications. See for example Gatti et al. (1998, p.56) or Öcal (2000, p.129), where small values for σ^2 together with huge estimates for γ are computed.

We close this section by proving that for small σ^2 one indeed will never find a good estimator for the unidentified γ. Tjøstheim (1986) derives in Theorem 3.2 asymptotic normality for a conditional Maximum Likelihood estimator $\hat{\beta}$ of $\beta = (\vartheta, \xi, \gamma)$ of a more general model than studied in this text. Specifying that result for the ESTAR model stated in (2.6) we obtain the following theorem, proved in the Appendix.

Theorem 2.2.4 (Asymptotic Variance of $\hat{\beta}$)**.**
Let y_t be as in (2.6) *where $\gamma > 0$ and $|\xi| + |\vartheta| < 1$. Let $\hat{\beta} = (\hat{\vartheta}, \hat{\xi}, \hat{\gamma})$ be the conditional ML estimator of $\beta = (\vartheta, \xi, \gamma)$. Assume that ε_t satisfies Assumption 2.2.2. Then*

$$\lim_{\sigma \downarrow 0} \mathrm{Var}(\hat{\gamma}) \quad \to \quad \infty. \tag{2.9}$$

Remarks.

- We are aware that the limiting situation in (2.9) never occurs in practical applications. However, the result should be read that the transition parameter γ can hardly be identified for very small sizes of the error variance, which results in biased estimators if no other correction is included in the optimization routine for deducing $\hat{\beta}$.

- We restrict the parameter vector in Theorem 2.2.4 to the three dimensional β not containing σ^2 only for technical reasons. In Tjøstheim (1986, Theorem 5.2) one can also find a general limiting result for $\tilde{\beta} = (\vartheta, \xi, \gamma, \sigma)$. That however neither yields any new information about the behavior of $\hat{\gamma}$, nor any substantial information about the remaining parameters and has therefore not been included in order to keep the proof of the above theorem somewhat readable.

- Theorem 2.2.4 only covers the case $\sigma \to 0$. As mentioned earlier, $\sigma \to \infty$ causes the identification problem, too. This is not just intuitive but has also been supported by simulation studies. Details are not included here as small values for σ^2 are the more interesting case in practical applications.

2.3 The TSTAR Model

In the ESTAR model, unidentified parameters occur for values (γ, σ) in a certain region, say $R^{\mathrm{E}}_{\gamma,\sigma}$. In particular estimating γ in the presence of a small $\sigma \in R^{\mathrm{E}}_{\gamma,\sigma}$ becomes impossible while on the other hand those values for σ^2 are used in the literature.
We now propose a new model within the STAR-framework, the so-called TSTAR model. The region $R^{\mathrm{T}}_{\kappa,\sigma}$ for which the identification of the parameters is not possible is smaller than $R^{\mathrm{E}}_{\gamma,\sigma}$, making the TSTAR model superior to the ESTAR-model.
The TSTAR(p) is defined by (2.1) using the transition function

$$G(z; \kappa, c) \quad = \quad \left[1 - \left(1 + (z-c)^2\right)^{-\kappa}\right], z \in \mathbb{R}, \tag{2.10}$$

with $\kappa > 0, 1 \le d \le p$ and $c \in \mathbb{R}$. The parameters κ and c can be interpreted as the transition and the location variable, respectively, just like in the ESTAR model. However, we denote the transition parameter differently as in the ESTAR setting to keep in mind that the mechanism with which κ affects the shape of the transition function is differently and thus the numerical values of γ for ESTAR and κ for TSTAR are not directly comparable. Nevertheless, properties like boundedness, the limit behavior for $z \to \pm\infty$ and $\kappa \to \pm\infty$ (see (2.4)) as well as the shape of G remain unchanged compared to the ESTAR transition function. The TSTAR model can therefore

been seen as an alternative model to the ESTAR model, applicable to the same situations.
We are aware that definition (2.10) is just one possible alternative to the ESTAR model. Formula (2.10) is motivated by the density of Student's t-distribution while the transition function (2.5) can be related to the normal density function. Moreover the chosen G has a series expansion which allows to prove linearity tests (see below) similar as to the well known tests in the ESTAR setting. However, more general forms of G are left for further research. The TSTAR model satisfies the analogous properties that are stated in Lemma 2.2.3 (i)-(iii), which follows directly from the proof of this Lemma which is stated in more general form for the ESTAR as well as the TSTAR setting.
As properties of the transition function G were mainly used to derive the results of Section 2.2 it is not surprising that we also encounter an identification problem in the TSTAR model. However, the identification problem causes less problems as different functions G are clearly distinct for a larger range of values for κ than in the ESTAR model. This is also visible in Figure 2.3 which illustrates for different values of κ the resulting transition functions in comparison to the ESTAR setting shown in Figure 2.4 (note the different scale on the x-axis).

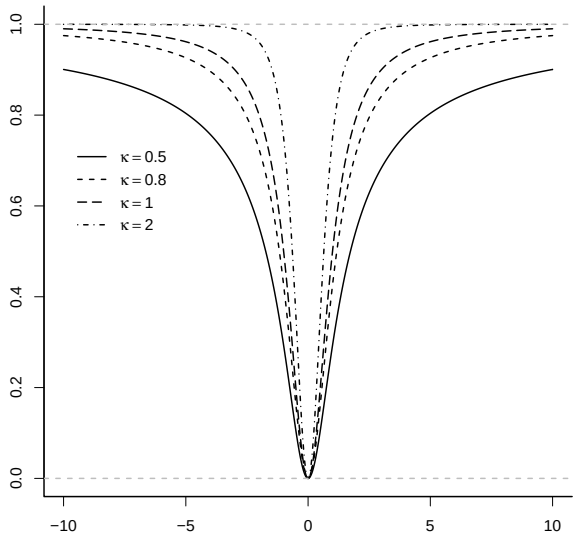

Figure 2.3: TSTAR: Transition function for different κ.

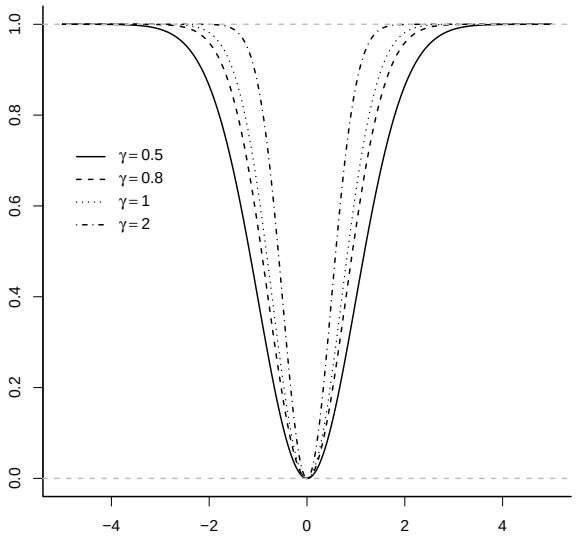

Figure 2.4: ESTAR: Transition function for different γ.

As κ and y_t^2 no longer appear as a product in the transition function (2.10), the interplay between these two parameters is reduced. For different values of κ the transition functions still differ even for small values of σ^2. This however implies that better estimates for $\beta = (\vartheta, \xi, \kappa, \sigma^2)$ are obtained even for parameter combinations of κ and σ^2 that cause problems in the ESTAR setting. This is also visible in Figure 2.5 which illustrated the quotient of the standard deviations of $\hat{\kappa}$ of the TSTAR model and the standard deviation of $\hat{\gamma}$ of the ESTAR model, each computed from 5000 repetitions. Note that all values stay below one pointing to the better performance of the TSTAR model.

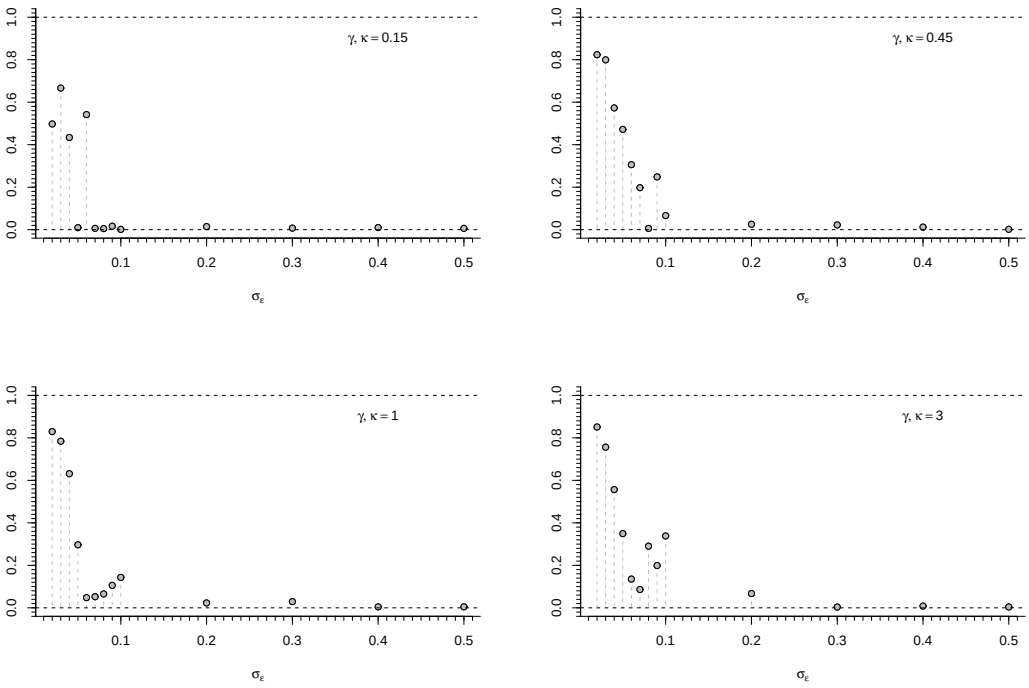

Figure 2.5: Comparison of the estimators in ESTAR and TSTAR settings.

2.3.1 Linearity testing

The procedure we derive in this section for testing linearity against non-linear TSTAR dynamics is related to the test against non-linearity proposed by Luukkonen et al. (1988). First, the transition function G is approximated by a suitable linear function; a common practice in non-linear time series analysis (see also Teräsvirta, 1994). Afterwards, a simple $F-$test is performed. For constructing the test it is convenient to use representation (2.2) of the TSTAR(p) model, i.e.

$$y_t = [\Psi w_t] + [\Phi w_t] \cdot G(y_{t-d}; \kappa, c) + \varepsilon_t. \tag{2.11}$$

Linearity then holds if the middle term on the right hand side vanishes, caused by either Φ or $G(\cdot)$ being equal to zero. In the first case the autoregressive parameters are then identical for both regimes (see also (2.1) with $\Theta = \Phi + \Psi = \Psi$). As we do not have to allow for switching between identical regimes a more parsimonious model is achieved by using a linear AR(p) model. In the latter case switching between different regimes it not performed as only one regime is considered. Hence, the pair of hypothesis we are interested in can be expressed either as

$$H_0 : \Phi = \mathbf{0}_{(1\times p)} \quad \text{vs.} \quad H_1 : \text{at least one } \varphi_i \neq 0;\ i = 1, \ldots, p$$

or

$$H_0 : \kappa = 0 \quad \text{vs.} \quad H_1 : \kappa > 0 \, .$$

In both cases the TSTAR model (2.11) reduces to a linear autoregressive model of order p. However, our test procedure employs the former pair of hypothesis.
Under H_0 the alternative is not identified, given that the vector Φ and c can take on any value without changing the value of the likelihood function when $\kappa = 0$ and vice versa. This can be circumvented by replacing G with a linear approximation. Based on the Binomial series, i.e.

$$(1+x)^{-m} = 1 + \sum_{n=1}^{\infty} (-1)^n \frac{m(m+1)(m+2)\ldots(m+n-1)}{n!} x^n, \, m > 0, \tag{2.12}$$

the transition function G in (2.11) can be approximated arbitrarily well by

$$G_k(\cdot) \quad = \quad \sum_{n=1}^{k} (-1)^n \frac{\kappa(\kappa+1)\ldots(\kappa+n-1)(y_{t-d}-c)^{2n}}{n!} \tag{2.13}$$

choosing $x = (y_{t-d} - c)^2$ and $m = \kappa$ in (2.12) as well as a sufficiently large k. We then obtain the auxiliary regression model for a fixed $d \leq p$ and k

$$y_t = \sum_{i=1}^{p} \psi_i y_{t-i} + \sum_{j=1}^{p} \delta_{j,0} y_{t-j} + \sum_{j=1}^{p} \delta_{j,1} y_{t-j} y_{t-d} + \ldots + \sum_{j=1}^{p} \delta_{j,2k} y_{t-j} y_{t-d}^{2k} + u_t \tag{2.14}$$

where the error terms are now denoted by u_t rather than ε_t as they are the sum of the original error terms and the approximation error caused by replacing G with G_k.
A test against non-linearity can then be carried out using a simple F-test for a subvector of parameters. Under the null the actual model is linear and hence the approximation error is zero leading to $u_t = \varepsilon_t$. Consequently the properties of the error term under the null and thus the asymptotic distribution of the F-test remain unaffected.

Example 2.3.1.
As an example consider the simple TSTAR(1) model,

$$y_t \quad = \quad \psi_1 y_{t-1} + \varphi_1 y_{t-1} \left[1 - \left(1 + (y_{t-d} - c)^2 \right)^{-\kappa} \right] + \varepsilon_t, \, t \geq 1 \, ,$$

with nonzero location parameter c. Approximating G by G_3 results in the regression model

$$\begin{aligned} y_t \quad &= \quad \psi_1 y_{t-1} \\ &\quad + \varphi_1 y_{t-1} \left[\kappa (y_{t-1} - c)^2 - \frac{1}{2} \kappa(\kappa+1)(y_{t-1}-c)^4 + \frac{1}{6} \kappa(\kappa+1)(\kappa+2)(y_{t-1}-c)^6 \right] + u_t \\ &= \quad \psi_1 y_{t-1} + \delta_{1,0} y_{t-1} + \delta_{1,1} y_{t-1}^2 + \delta_{1,2} y_{t-1}^3 + \delta_{1,3} y_{t-1}^4 + \delta_{1,4} y_{t-1}^5 + \delta_{1,5} y_{t-1}^6 + \delta_{1,6} y_{t-1}^7 + u_t \end{aligned}$$

where

$$\begin{aligned}
\delta_{1,0} &= \varphi_1\kappa c^2 + \frac{1}{6}\varphi\kappa(\kappa+1)(\kappa+2)c^6,\\
\delta_{1,1} &= -2c\varphi_1\kappa + 2\varphi_1\kappa(\kappa+1) - \varphi_1\kappa(\kappa+1)(\kappa+2)c^5,\\
\delta_{1,2} &= \varphi_1\kappa - 3\varphi_1\kappa(\kappa+1)c^2 + \frac{5}{2}\varphi_1\kappa(\kappa+1)(\kappa+2)c^4,\\
\delta_{1,3} &= 2\varphi_1\kappa(\kappa+1)c - \frac{1}{2}\varphi_1\kappa(\kappa+1)c^4 - \frac{10}{3}\varphi_1\kappa(\kappa+1)(\kappa+2)c^3,\\
\delta_{1,4} &= -\frac{1}{2}\varphi_1\kappa(\kappa+1) + \frac{5}{2}\varphi_1\kappa(\kappa+1)(\kappa+2)c^2,\\
\delta_{1,5} &= -\varphi_1\kappa(\kappa+1)(\kappa+2)c,\\
\delta_{1,6} &= \frac{1}{6}\varphi_1\kappa(\kappa+1)(\kappa+2)\,.
\end{aligned}$$

The hypothesis of linearity against TSTAR can now be tested via an F-test for the null

$$H_0 : \delta_{1,0} = \ldots = \delta_{1,6} = 0 \quad \textit{vs.} \quad H_1 : \textit{at least one } \delta_{1,i} \neq 0;\ i = 1,\ldots,6$$

for which extensive Monte Carlo simulations are summarized in Section 2.3.3.

2.3.2 Unit Root Testing

Kapetanios et al. (2003) develop a unit root test in the ESTAR framework and compute a Dickey-Fuller type t-test in this set-up based on a first order Taylor expansion. Our test is of the same type and thus we test the null of a linear unit root process against a globally stationary TSTAR process containing a partial unit root in one regime.
Based on parametrization (2.2) the TSTAR(1) model can be written in first differences as

$$\Delta y_t = \rho y_{t-1} + \varphi_1 y_{t-1}\left[1 - \left(1 + y_{t-1}^2\right)^{-\kappa}\right] + \varepsilon_t \tag{2.15}$$

where $\rho = \psi_1 - 1$. Setting the location parameter c equal to zero is motivated by simulation results in Kruse (2009) that show convincing power results even if the location parameter c is set to zero ex-ante. This is also consistent with Kapetanios et al. (2003). Hence, for the sake of simplicity we constrain ourselves to this case and further impose $d = 1$ which is in line with empirical applications of non-linear time series models (see e.g. Taylor et al. (2001) or Rapach and Wohar (2006)).
Setting $\rho = 0$ yields a unit root in the first regime and we have to distinguish between two cases:

(i) $\rho = 0$ and $\kappa > 0$: In this case we have a globally stationary TSTAR process that contains a partial unit root in the first regime, provided that $-2 < \varphi_1 < 0$ as we will assume henceforth.

(ii) $\rho = 0$ and $\kappa = 0$: In this case the model reduces to a linear random walk.

Thus we will test case (ii) against case (i) and formulate the pair of hypotheses as

$$H_0 : \kappa = 0 \quad \text{vs.} \quad H_1 : \kappa > 0\,. \tag{2.16}$$

We now proceed in the same way as in the previous section and approximate the nonlinearity with a Binomial expansion as in (2.13) setting the number of summands to $k = 3$. This yields the auxiliary regression (see also (2.14))

$$\Delta y_t = \delta_{1,2} y_{t-1}^3 + \delta_{1,4} y_{t-1}^5 + \delta_{1,6} y_{t-1}^7 + u_t \tag{2.17}$$

leading to the hypothesis

$$H_0 : \delta_{1,2} = \delta_{1,4} = \delta_{1,6} = 0 \quad \text{vs.} \quad H_1 : \text{at least one } \delta_{1,i} \neq 0;\ i = 2,4,6$$

which can also be expressed as

$$H_0 : \mathrm{I}_3 \beta \;=\; r \quad \text{vs.} \quad H_1 : \text{at least one } \delta_{1,i} \neq 0;\ i = 2,4,6$$

with $\beta = (\delta_{1,2}, \delta_{1,4}, \delta_{1,6})'$, $r = (0,0,0)'$ and

$$\mathrm{I}_3 \;=\; \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Due to the three parameter restrictions an F-statistic for the significance of the whole parameter vector β needs to be computed. Using $\hat{r} = R\hat{\beta}$, where $\hat{\beta}$ denotes the LS-estimator of β, we can write the F-statistic as

$$\begin{aligned} F^* \;&=\; \frac{1}{3}(\hat{r} - r)' \left[\hat{\sigma}^2 R(X'X)^{-1}R'\right]^{-1} (\hat{r} - r) \;=\; \frac{1}{3}(R\hat{\beta})' \left[\hat{\sigma}^2 R(X'X)^{-1}R'\right]^{-1} (R\hat{\beta}) \\ &=\; \frac{1}{3}\hat{\beta}' \hat{\sigma}^{-2} (X'X) \hat{\beta} \end{aligned} \tag{2.18}$$

where X is a $(T \times 3)$–design matrix with its t-th row given by $x_t = (y_{t-1}^3, y_{t-1}^5, y_{t-1}^7)$ and $\hat{\sigma}^2 = \frac{1}{T-4} \sum_{t=1}^{T} \left(\Delta y_t - \hat{\delta}_{1,2} y_{t-1}^3 - \hat{\delta}_{1,4} y_{t-1}^5 - \hat{\delta}_{1,6} y_{t-1}^7\right)^2$. The limit distribution of F^* under H_0 is computed in the next theorem and proven in the Appendix.

Theorem 2.3.2.
Consider the TSTAR(1) model (2.15) and let ε_t satisfy Assumption 2.2.2. Then the test statistic F^* as given in (2.18) *converges weakly under the null of a random walk as follows*

$$F^* \;\Rightarrow\; \frac{1}{3\sigma^2} \mathbf{v}' \mathbf{Q}^{-1} \mathbf{v},\ T \to \infty,$$

where the matrices $\mathbf{Q}$ *and* $\mathbf{v}$ *are given by*

$$\mathbf{Q} = \begin{bmatrix} \sigma^6 \int_0^1 B^6(r)\,dr & \sigma^8 \int_0^1 B^8(r)\,dr & \sigma^{10} \int_0^1 B^{10}(r)\,dr \\ \sigma^8 \int_0^1 B^8(r)\,dr & \sigma^{10} \int_0^1 B^{10}(r)\,dr & \sigma^{12} \int_0^1 B^{12}(r)\,dr \\ \sigma^{10} \int_0^1 B^{10}(r)\,dr & \sigma^{12} \int_0^1 B^{12}(r)\,dr & \sigma^{14} \int_0^1 B^{14}(r)\,dr \end{bmatrix}$$

and

$$\mathbf{v}' = \begin{bmatrix} \sigma^4 \left\{ \frac{1}{4} \int_0^1 B(1)^4 - \frac{3}{2} \int_0^1 B(r)\,dr \right\} \\ \sigma^6 \left\{ \frac{1}{6} \int_0^1 B(1)^6 - \frac{5}{2} \int_0^1 B(r)\,dr \right\} \\ \sigma^8 \left\{ \frac{1}{8} \int_0^1 B(1)^8 - \frac{7}{2} \int_0^1 B(r)\,dr \right\} \end{bmatrix}$$

with B *denoting the standard Brownian motion.*
Under the alternative the test is consistent.

In order to deal with deterministic components such as non-zero intercept terms or linear trends one can use a two-step approach and de-mean or de-trend the data prior to computing the test statistic F^*. In this case the true data generating process is given by

$$y_t = \omega' z_t + x_t$$

where $x_t = y_{t-1} + \varepsilon_t$ and ω' is a parameter vector of suitable dimensions and $z_t = 1$ for all t for the de-meaned case and $z_t = [1, t]$ for the de-trended case. The test can then be based on the OLS residuals $\hat{x}_t$, where the asymptotic distribution now depends on functionals of de-meaned and de-trended Brownian motion, respectively. These are given by

$$B(r) - \int_0^1 B(r)\,dr$$

for the de-meaned Brownian motion and by

$$B(r) + (6r - 4) \int_0^1 B(r)dr + (12r - 6) \int_0^1 rB(r)dr$$

for the de-trended Brownian motion. Considering the case of serially correlated errors and assuming that the dependence enters in a linear fashion we can generalize our results by augmenting the auxiliary regression with lagged differences as in Dickey and Fuller (1979) and Said

and Dickey (1984). The test regression then reads

$$\Delta y_t = \delta_{1,2} y_{t-1}^3 + \delta_{1,4} y_{t-1}^5 + \delta_{1,6} y_{t-1}^7 + \sum_{i=1}^{p} \pi_i \Delta y_{t-i} + u_t \,. \tag{2.19}$$

The pair of hypotheses as well as the test statistic in this more general set up do not change with respect to the auxiliary regression in (2.17).

Theorem 2.3.3.
Consider the test statistic F^* as in Theorem 2.3.2 but computed from (2.19). Under the null of a unit root the test statistic maintains the same asymptotic distribution as in Theorem 2.3.2. Under the alternative the test statistic is consistent.

Theorem 2.3.3 holds also true for the case of including deterministic terms as in auxiliary regression (2.19). The asymptotic distribution in this case is such as in Theorem 2.3.2 when deterministic terms are included, i.e. replacing the standard Brownian motion with the de-meaned or de-trended Brownian motion, respectively.
The large exponents in the auxiliary regressions necessarily lead to rather strong moment conditions. This could potentially be circumvented by using a different testing approach such as a sup-LM test. This is, however, beyond the scope of this paper as we only want to provide a simple test based on a well know procedure as a first approach.
Setting the approximation of the infinite sum from the Binomial series expansion to $k = 1$ it is readily seen that $\sqrt{F^*}$ has the same asymptotic distribution as the unit root test against ESTAR developed by Kapetanios et al. (2003) and thus the statistic F^* contains their test as a special case. It is also noteworthy that the F-test version of the ESTAR unit root test of Kapetanios et al. (2003) that would result if the location parameter c is not set equal to zero a priori is also a special case of our test. Setting the series expansion again to $k = 1$ and also letting the location parameter $c \neq 0$ the resulting limiting distribution of the test statistic from the related auxiliary regression is the same as for the respective ESTAR unit root test and thus we also contain this test version as a special case.
Containing these tests as special cases we expect a satisfying performance also against ESTAR processes but higher power against globally stationary alternatives than the Kapetanios et al. (2003) test as indicated by a faster rate of convergence in Theorem 2.3.2.

2.3.3 Monte Carlo Simulations

In this section we study the finite sample performance of the two tests developed above for the TSTAR(1) model given in (2.11). Two different data generating processes are investigated: $c = 0$ in the first scenario and $c = 1$ in the second one. If not stated differently, $M = 50000$ replications were performed combined with different sample sizes T. For all power simulations reported we consider a type 1 error of $\alpha = 0.05$ to save space.[1]

[1] The results for the cases $\alpha = 0.01$ and $\alpha = 0.1$ as well as all other unreported results are available from the authors upon request.

Linearity Testing:

In order to keep the experiment simple we conduct the simulations under the null of linearity using a simple AR(1) model and compute the auxiliary regression using the location parameter $c = 0$ for the first scenario and $c = 1$ for the second scenario. The empirical size results for the linearity test introduced in Section 2.3.1 are stated in Tables 2.2 and 2.3 for scenario one and scenario two, respectively. The test shows only minor deviations from its nominal level and it tends to under-reject somewhat. However, the test -although conservative- seems to be properly sized for reasonable sample sizes encountered in monthly or daily data.

	$\psi_1 = 0.3$			$\psi_1 = 0.5$			$\psi_1 = 0.8$		
T	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$
100	0.826	4.258	8.752	0.738	4.152	8.440	0.856	4.194	8.606
200	0.872	4.442	9.038	0.786	4.200	8.664	0.820	4.206	8.504
500	0.910	4.546	9.404	0.876	4.470	8.986	0.744	4.248	8.680
1000	0.858	4.594	9.554	0.840	4.360	9.008	0.878	4.152	8.578
5000	0.886	4.766	9.906	0.856	4.536	9.354	0.888	4.498	8.986

Table 2.2: Size results (linearity test) for scenario one.

	$\psi_1 = 0.3$			$\psi_1 = 0.5$			$\psi_1 = 0.8$		
T	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$
100	0.912	4.298	8.608	0.818	4.256	8.594	0.838	4.402	8.744
200	0.776	4.200	8.764	0.802	3.986	8.268	0.796	4.150	8.504
500	0.818	4.266	8.866	0.776	4.184	8.434	0.892	4.270	8.530
1000	0.882	4.626	9.276	0.868	4.318	8.708	0.848	4.218	8.528
5000	0.968	4.584	9.372	0.882	4.450	8.892	0.792	4.260	8.764

Table 2.3: Size results (linearity test) for scenario two.

In order to study the power of the test, TSTAR(1) models with several values for ψ_1, φ_1 and κ were used in the experiments. The results are shown in Table 2.4 for scenario one and in Table 2.5 for scenario two. The results suggest that the linearity test is a useful device to detect non-linearity in the data. As expected the rejection frequency becomes closer to 100% the more pronounced the difference between the regimes is and/or the larger the sample size is. Overall we obtain very similar power results against TSTAR compared to Luukkonen et al. (1988) for their linearity test against ESTAR. Unreported experiments confirmed that the proposed linearity test has also similar high power against the other non-linear alternatives ESTAR, LSTAR and Double LSTAR (see Jansen and Teräsvirta, 1996). Reasonable power results were also obtained against the Markov switching model proposed by Hamilton (1989).

		T = 200			T = 500			T = 1000		
ψ_1	φ_1 \ κ	0.5	0.8	1.0	0.5	0.8	1.0	0.5	0.8	1.0
0.3	0.5	5.452	5.468	5.650	7.502	8.628	8.708	11.796	14.488	15.014
0.3	0.6	6.638	7.414	7.686	12.534	16.060	16.154	24.166	32.020	31.766
0.3	0.7	9.434	11.456	11.570	22.000	29.264	29.014	44.362	57.616	58.062
0.3	0.8	13.642	17.872	18.238	36.348	49.194	48.964	69.112	83.718	83.360
0.3	0.9	21.276	29.262	28.098	56.726	74.030	71.142	89.640	97.252	96.482

Table 2.4: Power results (linearity test) for scenario one.

		T = 200			T = 500			T = 1000		
ψ_1	φ_1 \ κ	0.5	0.8	1.0	0.5	0.8	1.0	0.5	0.8	1.0
0.3	0.5	5.964	6.970	7.070	10.302	13.826	14.790	19.392	27.412	30.088
0.3	0.6	9.228	11.804	12.514	21.240	31.126	34.288	45.714	63.610	68.202
0.3	0.7	14.444	20.502	22.364	40.652	58.280	62.524	77.352	91.816	94.004
0.3	0.8	24.126	34.482	36.886	66.438	84.038	86.562	95.732	99.448	99.634
0.3	0.9	38.434	53.024	53.978	87.566	97.006	96.840	99.692	99.998	99.986

Table 2.5: Power results (linearity test) for scenario two.

Unit Root Testing:

We first report the asymptotic critical values for the unit root test in Table 2.6. Case 1 denotes raw data, i.e. no deterministic components, Case 2 denotes the case of de-meaned data and Case 3 denotes the case of de-trended data. Here, the sample size is set to $T = 10000$ and the number of replications to $M = 1000000$. The results from the size experiments are summarized in Table

α	Case 1	Case 2	Case 3
1%	4.730	5.477	6.595
2.5%	3.124	4.722	5.783
5%	3.458	4.137	5.136
7.5%	3.124	3.778	4.739
10%	2.884	3.515	4.450

Table 2.6: Asymptotic critical values for unit root testing.

2.7. For larger sample sizes ($T > 500$) the test is correctly sized and as the sample size increases it reaches its nominal level. For smaller sample sizes some minor size distortions are visible but the overall impression is that the test maintains good size properties also for smaller sample sizes. For the power experiment for the unit root test we exemplarily show our results for $T = 200$ and various values for κ, ψ and φ (see Table 2.8). For increasing sample size, the results improve which was seen from the same simulations for $T = 500$. The results indicate a good overall performance of the unit root test in all sample sizes considered. The ability to distinguish between a unit root process and a globally stationary TSTAR model increases if either the difference between

	Case 1			Case 2			Case 3		
T	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$
100	0.826	3.898	7.628	0.786	3.484	6.916	0.954	3.830	7.484
200	0.882	4.106	8.246	0.748	3.648	7.436	0.758	3.594	7.218
500	0.912	4.504	9.168	0.864	4.174	8.602	0.828	3.998	8.146
1000	0.932	4.920	9.528	0.904	4.370	9.076	0.862	4.324	8.868
5000	0.952	4.958	10.060	0.916	4.734	9.692	1.048	4.952	9.766
10000	0.984	5.036	10.006	1.004	4.982	9.990	0.966	4.918	9.886
50000	0.974	4.968	10.054	1.000	4.978	10.062	0.998	5.012	9.882

Table 2.7: Size for unit root testing using asymptotic critical values [in %].

the regimes becomes larger or even faster if the sample size increases.

$T = 200$		Case 1			Case 2			Case 3		
ψ_1	φ_1 \ κ	0.5	0.8	1.0	0.5	0.8	1.0	0.5	0.8	1.0
1.0	0.3	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
1.0	0.4	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
1.0	0.5	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
1.0	0.6	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
1.0	0.7	99.95	99.97	99.97	99.95	10.00	100.00	99.97	10.00	100.00
1.0	0.8	95.52	98.86	99.36	96.58	99.30	99.59	97.65	99.46	99.71
1.0	0.9	45.95	58.63	62.44	55.04	67.01	70.45	65.79	75.30	77.97
1.0	0.95	14.90	18.44	19.47	23.68	27.84	28.76	36.00	39.77	41.77

Table 2.8: Power results for unit root testing using asymptotic critical values [in %]

As empirical studies using smooth transition models such as ESTAR frequently find very small variances of the innovation term we examine the behavior of the newly developed unit root test against TSTAR in such a framework. Studying this behavior is critical since Kruse et al. (2010) show via Monte Carlo simulation that under small error term variances the power of unit root tests developed for non-linear models rapidly deteriorates. We report simulation results for small sample sizes of $T = 100$ (Table 2.9) and $T = 500$ (Table 2.10) and consider error term standard deviations of $\sigma_\varepsilon = 0.1$. The results show satisfying power results even for such small sample sizes. Low power results are only found for cases in which the difference between the regimes is only very small or the transition is so slow that only little observations are in the stationary regime. In these cases it is notoriously hard to distinguish between the two regimes and as a consequences the power decreases. However, the power is still high enough to deliver reliable test results and is in particular higher than found by Kruse et al. (2010) for extant test. In the case $T = 500$ no decline in power is visible and the test works under small error variances just as well as under white noise disturbances.

The newly developed test in particular shows better power properties as the test developed by Kapetanios et al. (2003) and therefore yields more reliable results in empirical applications as indicated by Kruse (2009).

$T = 100$		Case 1			Case 2			Case 3		
ψ_1	φ_1 \ κ	0.5	0.8	1.0	0.5	0.8	1.0	0.5	0.8	1.0
1.0	0.3	77.108	95.104	98.174	79.610	95.388	98.378	84.538	96.526	98.716
1.0	0.4	66.452	88.956	95.106	69.914	90.042	95.444	77.190	92.464	96.662
1.0	0.5	52.768	77.878	87.560	58.408	80.730	88.988	67.572	85.324	91.724
1.0	0.6	38.304	60.896	72.470	44.878	66.622	76.096	56.658	73.874	81.682
1.0	0.7	24.404	40.558	49.668	33.204	48.180	57.018	45.100	59.096	65.972
1.0	0.8	13.952	21.662	26.456	22.466	30.732	35.064	34.442	42.484	47.878
1.0	0.9	6.714	8.604	9.888	13.770	16.748	17.924	24.864	28.176	29.830
1.0	0.95	4.608	5.062	5.298	10.360	11.518	12.048	21.382	21.926	22.620

Table 2.9: Power results for unit root testing [in %] with $\sigma_\varepsilon = 0.1$

$T = 500$		Case 1			Case 2			Case 3		
ψ_1	φ_1 \ κ	0.5	0.8	1.0	0.5	0.8	1.0	0.5	0.8	1.0
1.0	0.3	100.00	100.00	100.00	100.00	100.00	100.00	100.000	100.00	100.00
1.0	0.4	100.00	100.00	100.00	100.00	100.00	100.00	100.000	100.00	100.00
1.0	0.5	100.00	100.00	100.00	100.00	100.00	100.00	100.000	100.00	100.0
1.0	0.6	100.00	100.00	100.00	100.00	100.00	100.00	100.000	100.00	100.00
1.0	0.7	100.00	100.00	100.00	100.00	100.00	100.00	100.000	100.00	100.00
1.0	0.8	100.00	100.00	100.00	100.00	100.00	100.00	99.99	100.00	100.00
1.0	0.9	93.25	99.09	99.70	94.01	99.19	99.74	95.57	99.38	99.77
1.0	0.95	44.70	63.54	70.72	53.48	70.29	76.52	64.24	77.55	82.39

Table 2.10: Power results for unit root testing [in %] with $\sigma_\varepsilon = 0.1$

With regards to the linearity test we find the power only slightly reduced under small error variances compared to the white noise assumption. These results however are unreported to save space.

Varying Error Term Variances:

Having described the behavior of the two tests extensively, we now consider two fixed parameter combinations and illustrate how the power of the two test changes as we let the error term variance σ_ε become smaller and smaller in accordance with the main statement of this text. Figures 2.6 and 2.7 are based on the unit root test for the parameters $\psi_1 = 0.3, \varphi_1 = 0.6, \kappa = 1$ and $\psi_1 = 0.7, \varphi_1 = 0.2, \kappa = 1$, respectively, while σ_ε varies from an almost vanishing value up to 1.5. One clearly sees that except for very small values of σ_ε we obtain a very high power, independent of the chosen α where $\alpha \in \{0.01, 0.05, 0.1\}$ is considered.

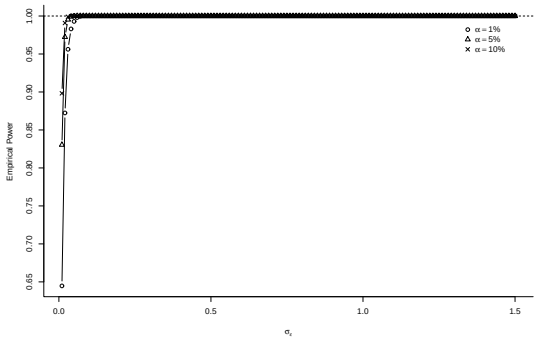

Figure 2.6: Power curve of the unit root test: $\psi_1 = 0.3, \varphi_1 = 0.6$ and $\kappa = 1$.

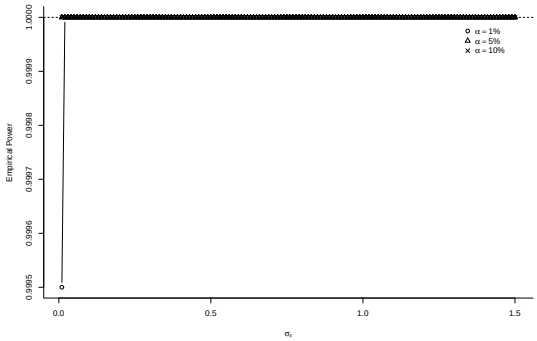

Figure 2.7: Power curve of the unit root test: $\psi_1 = 0.7, \varphi_1 = 0.2$ and $\kappa = 1$.

Using the same parameter combinations as mentioned above but now for the linearity test we obtain the curves for the power as shown in Figures 2.8 and 2.9. Although the range of σ_ε where the power lies below 100% is bigger than for the unit root test, the overall behavior is still identical and the power increases steeply.

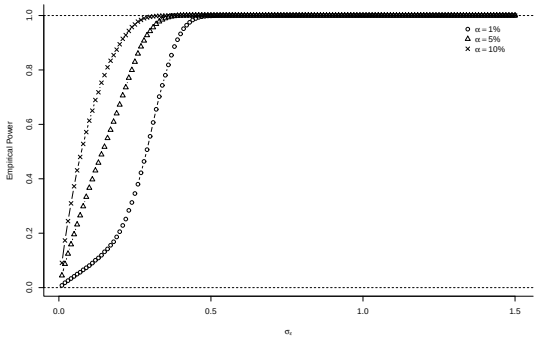

Figure 2.8: Power curve of the linearity test: $\psi_1 = 0.3, \varphi_1 = 0.6$ and $\kappa = 1$.

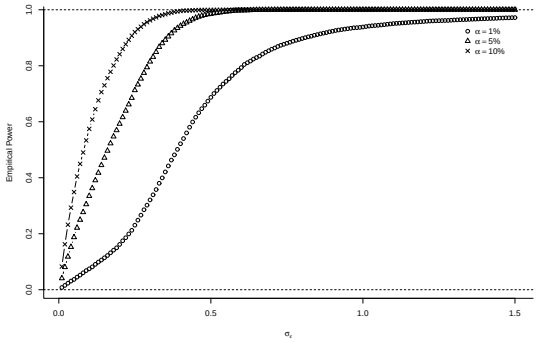

Figure 2.9: Power curve of the linearity test: $\psi_1 = 0.7, \varphi_1 = 0.2$ and $\kappa = 1$.

2.4 Empirical Illustration

To illustrate the application of the newly introduced TSTAR model with empirical data we investigate one of the most highly debated theories in international finance: the purchasing power parity (PPP). The initial finding of a unit root in real exchange rates by Meese and Rogoff (1988) subsequently shifted the interest in modeling real exchange rates to non-linear models (see e.g. Taylor et al., 2001). Technically spoken the real exchange rate should be non-linear but globally stationary (i.e. mean-reverting) and not behave like a unit root process to support PPP.

To ensure comparability we use the same data that has been analyzed by Taylor et al. (2001) and by Rapach and Wohar (2006). Namely, we analyze monthly real exchange data for Germany against the US from 1973:02 - 1996:12.[2] The series is depicted in Figure 2.10.

Figure 2.10: Monthly log real exchange rate for Germany

We choose the lag length to be used subsequently with the Bayesian information criterion (BIC) which yields a lag length of $p = 1$.

Applying the linearity test against TSTAR described in Section 2.3.1 we obtain a test statistic of 3.69 which is significant on the $\alpha = 5\%$ level of significance and thus we reject the null of linearity.

Validity of the PPP suggests that the real exchange rate should be a globally stationary process albeit non-linear. Applying the ESTAR unit root test developed by Kapetanios et al. (2003) as well as the unit root test against TSTAR developed in Section 2.3.2 yields support for the PPP. Both test are able to reject the null of a random walk on the $\alpha = 5\%$ level of significance. These test results support the theory that transaction costs in financial markets lead to a non-linear convergence to a long-run equilibrium and thus support the validity of the PPP as a long run

[2]The data set is available from David Rapach's website at: http://pages.slu.edu/faculty/rapachde/Nlfit.zip.

concept.

Since the data has already been under study by Taylor et al. (2001) we adopt the parameter estimates they found and which have also been confirmed by estimations undertaken by Rapach and Wohar (2006). It should be noted that Taylor et al. (2001) and Rapach and Wohar (2006) also estimated the location parameter c. However, as their estimate is very close to zero, namely $c = -0.007$, we restrict $c = 0$ in our estimation to keep it simple. Furthermore the authors fix the parameter of the second regime to be -1 which yields one unit root regime and one white noise regime. As this seems to be rather restrictive we only fix the unit root regime and estimate the autoregressive parameter of the second regime.

Table 2.11 shows the estimation results for the parameter γ for the model under the null for the ESTAR and the TSTAR model respectively.

ESTAR	TSTAR
$\hat{\varphi}$= -1	$\hat{\varphi}$= -0.023
$\hat{\gamma} = 0.264$	$\hat{\kappa} = 275.284$
$\hat{\sigma}_\varepsilon = 0.035$	$\hat{\sigma}_\varepsilon = 0.032$

Table 2.11: Estimation of the transition parameter under the null.

At a first glance the estimation result for ESTAR looks reasonable. But if we plot the estimated transition function against the transition variable y_{t-1} and against time (see Figure 2.11) we get to the conclusion that the ESTAR model basically reduces to a random walk model as the transition function is always close or equal to zero effectively switching off the stationary regime.

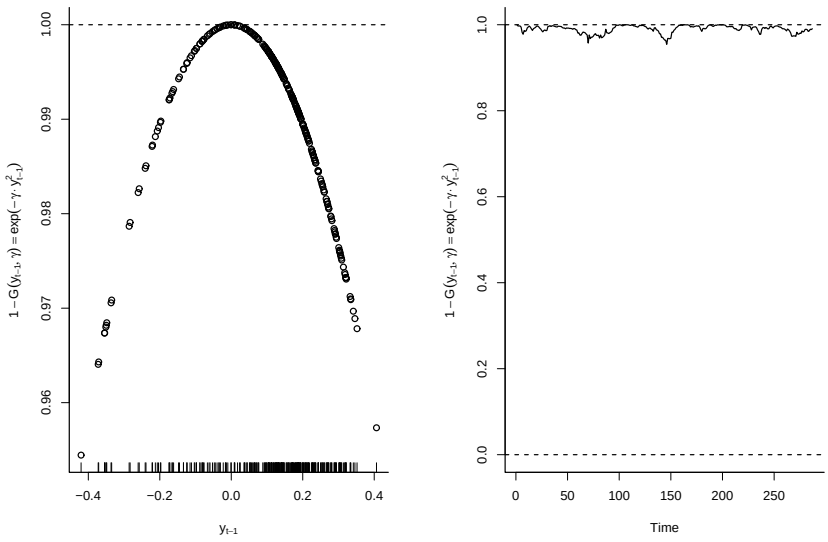

Figure 2.11: Left panel: One minus transition function against transition variable. Right panel: One minus transition function against time.

The figure supports the results that, albeit the parameter estimate for γ leads to a reasonable

looking transition function (note the range of the y-axis) plotting it against time, $\hat{\gamma} = 0.264$ actually produces a random walk model and by this contradicts PPP caused by a degenerated transition function. To further investigate whether an ESTAR specification seems appropriate we also estimate a STAR model using a double logistic form for the transition function

$$G(\cdot;\gamma,c_1,c_2) = \{1+\exp(-\gamma(y_{t-1}-c_1)(y_{t-1}-c_2))\}^{-1} \ .$$

The estimation led to $\hat{c}_1 = \hat{c}_2 = -0.0069$ and $\hat{\gamma} = 0.1719$. These parameters are very similar to the ESTAR results suggesting that the ESTAR model might be inappropriate.
Producing the same plots for TSTAR as for ESTAR in Figure 2.11 we obtain Figure 2.12.

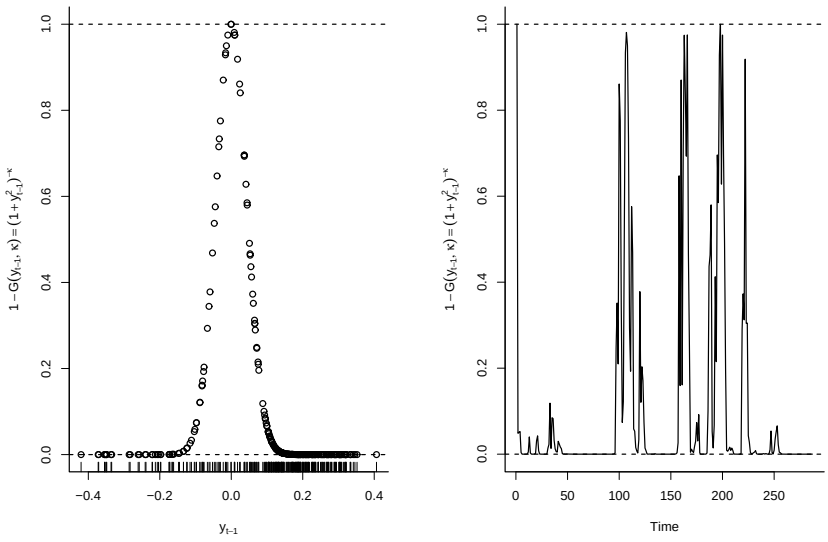

Figure 2.12: Left panel: One minus transition function against transition variable. Right panel: One minus transition function against time.

The estimation of a large $\hat{\kappa} = 275.284$ still produces a transition function that is by no means close to the limit for $\kappa \to \infty$ (see the properties in Section 2.2). In addition we see from the left panel that the estimated process is far more often in the stationary regime and becomes a random walk only on few occasions. This finding is in line with theoretical work on PPP. Deviations from the law of one price may stem from transaction costs between different markets (see e.g. Sercu et al. (1995)). This notion has subsequently been more refined by Coleman (1995) in whose model transaction costs create a band of no arbitrage for the real exchange rate. Once the real exchange rate, as a measure of deviation from PPP, hits the upper or lower threshold the process becomes mean-reverting to the equilibrium. Once within the transaction cost band, no trade takes place and the process diverges away from PPP. As a result the real exchange rate spends most of the time away from the equilibrium (see also the discussion in Taylor et al. (2001, p.1018) and Taylor (2003, p.444)).

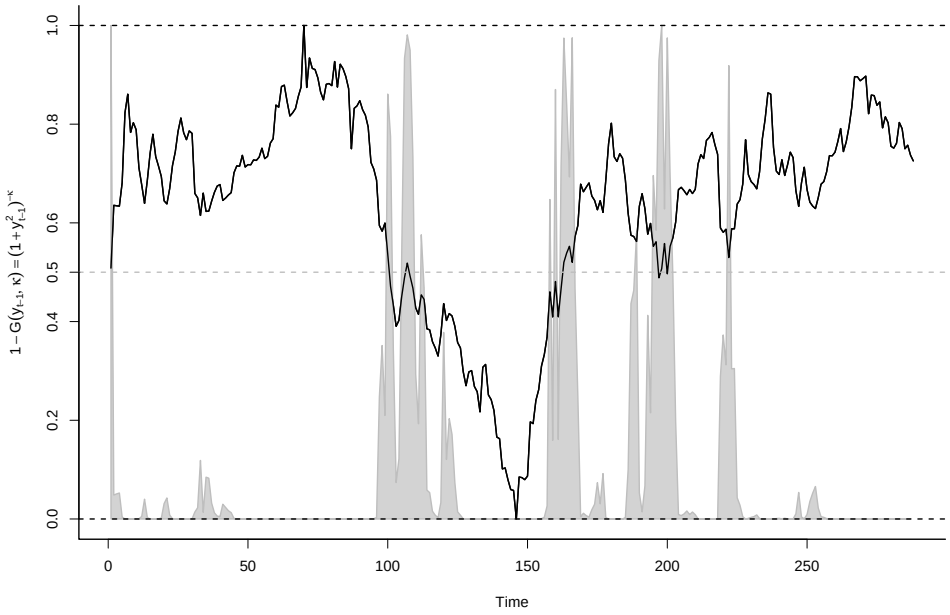

Figure 2.13: Rescaled real exchange rate with TSTAR transition function.

Figure 2.13 shows the real exchange rate from Figure 2.10 rescaled to be in $[0,1]$ and the transition function from the right panel of Figure 2.12. The gray shaded area shows the periods in which the process behaves like a random walk, i.e. PPP holds exactly. At the very beginning of the data, when the Bretton Woods system of fixed exchange rates was abandoned in favor of a free floating exchange rate regime, PPP holds exactly. It then starts to deviate from PPP until the upper threshold is reached and starts to revert back to the mean. Whenever the real exchange rate hits the equilibrium it quickly deviates away from exact PPP and we observe that the real exchange rate behaves like a nonlinear mean reverting process most of the time and like a random walk when deviation from PPP is near zero (note that the dotted line at 0.5 is the zero line of the unscaled series).

It is noteworthy that the plot in Figure 2.11 is not unique for this particular data set but a common finding in empirically estimated ESTAR models. Kruse et al. (2010) for example support these findings for other real exchange rates (see their Figure 1).

The panels in Figure 2.11 and Figure 2.12 also support the theoretical results that the estimation of the transition parameter heavily depends on the error term variance σ_ε derived in Section 2.2. Looking at the estimated values for γ and κ in Table 2.11 and the left panels in Figures 2.11 and 2.12 we obtain a reasonable form of the transition function from a mathematical point of view. However plotting the transition function against time we see that the estimated function does not support the hypothesis that the data comes from the assumed data generating process. This supports that the estimation of γ is heavily influenced by the small error standard deviation. The estimated κ for the TSTAR model might look awkward at first. However, looking at the plots in Figure 2.12 this yields a transition function that is not degenerated in the sense that the actual range exploits its whole domain and it does not behave as in the limiting case (see Section 2.2). This supports the assumed data generating process, i.e. a globally stationary

TSTAR model. The estimation of κ in the TSTAR case is by far not so heavily influenced from the small error standard deviation and thus we can extend the range of possible values for the transition parameter for which we obtain a non-degenerated transition function. This also supports the conclusion that we can largely reduce the influence of σ_ε on the transition parameter by reformulating the transition function G with respect to the ESTAR setting.

2.5 Conclusions

We have studied the ESTAR and TSTAR model, two competing models of the STAR family sharing the same characteristic properties of their transition functions. Due to their nonlinear structure, unidentified parameters occur for certain combinations of γ and σ^2, the transition parameter and the error term variance, respectively. This phenomena has not been studied systematically before although it is of importance in applications.
In the ESTAR setting, very small values of σ^2, among others, yield in particular an unidentified γ, making a consistent estimation of γ nearly impossible. In Theorem 2.2.4 we verified this by showing that the variance of the conditional Maximum Likelihood estimator $\hat{\gamma}$ tends to infinity as σ^2 vanishes. Hence, in order to estimate γ, somewhat unpleasant modifications need to be incorporate into the optimization routine.
In order to avoid this, we define the TSTAR model where the transition parameter becomes unidentified much later as $\sigma^2 \to 0$ compared to the ESTAR model. As a consequence, the parameter can be included in the parameter vector that is to be estimated. By deriving a linearity and a unit root test for the TSTAR model we support our opinion that this new model is indeed a worthy alternative, applicable to the same situations and should therefore be preferred to the ESTAR model.
This conclusion is illustrated by fitting both models to the same data set containing real exchange rates. The estimators one obtains in the ESTAR setting do not allow for a meaningful interpretation of the fitted model as one regime is basically switched off. One can clearly see that this is the result of the identification problem caused by a small error term variance. Contrary to that, the fitted TSTAR model allows for switching between the two regimes, leading to a better fit, although the estimators for γ and σ^2 look not very promising without interpreting them in the right context.
As this text deals with a topic that is not well studied yet, there are many possible open questions and possibilities how to go from here. The most interesting question at the moment is whether it is possible to quantify and compare the regions for which (γ, σ) cause the identification problem in both regimes. A more theoretical, in depth study of the TSTAR model would be needed for this. Also, a general theory about a whole class of transition functions G sharing some characteristic properties would be very helpful to fully understand STAR models.

2.6 Appendix: Proofs and Technical Lemmas

Lemma 2.2.3 is stated in terms of the ESTAR model. However, properties (i)-(iii) hold more generally. Therefore in the proof below we consider the models

$$y_t = \left[\vartheta + \xi \exp(-\gamma\eta(y_{t-1}^2))\right] y_{t-1} + \varepsilon_t,\ t \in \mathbb{Z},$$

for $\eta \in \{\eta_1 : [0,\infty) \to \mathbb{R}, \eta_1(z) = z;\ \eta_2 : [0,\infty) \to \mathbb{R}, \eta_2(z) = \log(1+z)\}$. Note that η_1 corresponds to the ESTAR model and η_2 to the TSTAR model. The proof of property (iv) depends heavily on η_1 and can therefore not be generalized although we believe that (iv) also holds for η_2.

Proof of Lemma 2.2.3.
Ad (i): Geometric ergodicity follows from general conditions for ergodicity of nonlinear time series which are satisfied due to $|\xi| + |\vartheta| < 1$ (see in particular Example 8.2 in Yao and Fan (2005)).

Ad (ii): In order to prove that the density of y_t is symmetric around zero for all t we use an inductive argument. Let $y_0 = \varepsilon_0$ which then has a symmetric density by choice. Now, let $t \geq 1$ and assume that y_{t-1} has a symmetric density and recall from (2.6) that

$$y_t = \left[\vartheta + \xi(1 - G(\eta_i(y_{t-1}^2)))\right] y_{t-1} + \varepsilon_t = \left[\vartheta + \xi \exp(-\gamma\eta_i(y_{t-1}^2))\right] y_{t-1} + \varepsilon_t,\ t \geq 1, i = 1,2.$$

We know from Lemma 2.6.1 that $\left[\vartheta + \xi \exp(-\gamma\eta_i(y_{t-1}^2))\right] y_{t-1}, i = 1,2$, has a symmetric density around zero. As the same holds for ε_t by assumption, it follows from Lemma 2.6.2 that also the denisty of y_t is symmetric around zero due to y_{t-1} and ε_t being independent.
The above argument requires the stationary distribution F_s of y_t to have a symmetric density.

Ad (iii): Let $n, k \in \mathbb{N}_0$ and $t \in \mathbb{Z}$. Then the density of $\exp(-n\gamma\eta_i(y_t^2))y_t^{2k+1}, i = 1,2$, is symmetric around zero by combining (ii) with an application of Lemma 2.6.1 choosing $a = 0, b = 1$ and $c = n\gamma > 0$. Hence, $\mathbb{E}\left[\exp(-n\gamma y_t^2)y_t^{2k+1}\right] = 0$.

Ad (iv): Let $n \in \mathbb{N}_0, k \in \mathbb{N}$. In order to show that $\mathbb{E}\left[\exp(-n\gamma y_t^2)y_t^{2k}\right]$ behaves asymptotically like σ^{2k} as σ goes to zero we use an inductive argument and determine two sequences $\{l_k(\sigma)\}_{\sigma>0}$ and $\{u_k(\sigma)\}_{\sigma>0}$ such that for all t

$$l_k(\sigma) \leq \frac{\mathbb{E}\left[\exp(-n\gamma y_t^2)y_t^{2k}\right]}{\sigma^{2k}} \leq u_k(\sigma)$$

with $\lim_{\sigma\downarrow 0} l_k(\sigma) = \lim_{\sigma\downarrow 0} u_k(\sigma) = c_k$ for some constants c_k.
The sequences $\{l_k(\sigma)\}_{\sigma>0}$ and $\{u_k(\sigma)\}_{\sigma>0}$ are determined by the even moments of the process y_t due to

$$\frac{\mathbb{E}\left[\exp(-n\gamma y_t^2)y_t^{2k}\right]}{\sigma^{2k}} \leq \frac{\mathbb{E}\left[y_t^{2k}\right]}{\sigma^{2k}} \tag{2.20}$$

and

$$\frac{\mathbb{E}\left[\exp(-n\gamma y_t^2)y_t^{2k}\right]}{\sigma^{2k}} \geq \frac{\mathbb{E}\left[(1-n\gamma y_t^2)y_t^{2k}\right]}{\sigma^{2k}} \geq \frac{\mathbb{E}\left[y_t^{2k}\right]}{\sigma^{2k}} - n\gamma\frac{\mathbb{E}\left[y_t^{2k+2}\right]}{\sigma^{2k+2}}\sigma^2 \tag{2.21}$$

using $\exp(x) \geq 1+x, x \in \mathbb{R}$. Note that we have for $k \in \mathbb{N}_0, t \in \mathbb{Z}$,

$$\begin{aligned}
\mathbb{E}\left[y_t^{2k}\right] &= \mathbb{E}\left[\left[\left[\vartheta + \xi\exp(-\gamma y_{t-1}^2)\right]y_{t-1} + \varepsilon_t\right]^{2k}\right] \\
&= \sum_{j\in J}\binom{2k}{j}\mathbb{E}\left[\left[\vartheta+\xi\exp\left(-\gamma y_{y-1}^2\right)\right]^{2k-j}y_{t-1}^{2k-j}\right]\mathbb{E}\left[\varepsilon_t^j\right] \\
&= \sum_{j\in J}\binom{2k}{j}\sum_{\nu=0}^{2k-j}\binom{2k-j}{\nu}\vartheta^{2k-j-\nu}\xi^\nu\,\mathbb{E}\left[\exp\left(-\nu\gamma y_{t-1}^2\right)y_{t-1}^{2k-j}\right]c_{\varepsilon,j}\sigma^j
\end{aligned} \tag{2.22}$$

where $J = \{j = 0,\ldots,2k : j \bmod 2 = 0\}$ as the expected value of any product of y_{t-1} and ε_t with odd powers vanishes and where $\mathbb{E}[\varepsilon_t^j] = c_{\varepsilon,j}\sigma^j$ with constants $c_{\varepsilon,j}$ due to Assumption 2.2.2. Let $t = 1$ and assume $y_0 \sim \mathrm{N}(0,\sigma^2)$. Then, for $m!!$ being the product of every odd number from 1 to m with $(-1)!! = 0!! = 1$,

$$\frac{\mathbb{E}\left[y_1^{2k}\right]}{\sigma^{2k}} \to \sum_{j\in J}\binom{2k}{j}(\vartheta+\xi)^{2k-j}c_{\varepsilon,j}(2k-j-1)!! \quad \text{as } \sigma \downarrow 0$$

due to the following argument. Formula (2.22) implies

$$\mathbb{E}\left[y_1^{2k}\right] = \sum_{j\in J}\binom{2k}{j}\sum_{\nu=0}^{2k-j}\binom{2k-j}{\nu}\vartheta^{2k-j-\nu}\xi^\nu c_{\varepsilon,j}\sigma^j\int_{\mathbb{R}} y^{2k-j}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\nu\gamma y^2 - \frac{1}{2}\frac{y^2}{\sigma^2}\right)dy.$$

As the exponent of the exponential term can be written as

$$-\nu\gamma y^2 - \frac{1}{2}\frac{y^2}{\sigma^2} = -\frac{1}{2}\left[\frac{y^2(1+2\sigma^2\nu\gamma)}{\sigma^2}\right] = y^2\left(\frac{\sigma}{\sqrt{1+2\sigma^2\nu\gamma}}\right)^{-2}$$

we obtain with f being the density of a $\mathrm{N}(0,\sigma^2/(1+2\sigma^2\nu\gamma))$ distribution

$$\begin{aligned}
\mathbb{E}\left[y_1^{2k}\right] &= \sum_{j\in J}\binom{2k}{j}\sum_{\nu=0}^{2k-j}\binom{2k-j}{\nu}\vartheta^{2k-j-\nu}\xi^\nu c_{\varepsilon,j}\sigma^j\frac{1}{\sqrt{1+2\sigma^2\nu\gamma}}\int_{\mathbb{R}} y^{2k-j}f(y)\,dy \\
&= \sum_{j\in J}\binom{2k}{j}\sum_{\nu=0}^{2k-j}\binom{2k-j}{\nu}\vartheta^{2k-j-\nu}\xi^\nu c_{\varepsilon,j}\sigma^j\frac{1}{\sqrt{1+2\sigma^2\nu\gamma}}\frac{\sigma^{2k-j}}{\left(\sqrt{1+2\sigma^2\nu\gamma}\right)^{2k-j}}(2k-j-1)!! \\
&= \sum_{j\in J}\binom{2k}{j}\sum_{\nu=0}^{2k-j}\binom{2k-j}{\nu}\vartheta^{2k-j-\nu}\xi^\nu c_{\varepsilon,j}\sigma^{2k}(2k-j-1)!!\frac{1}{(1+2\sigma^2\nu\gamma)^{(2k-j+1)/2}}.
\end{aligned}$$

As the last factor in the previous display converges to one as σ goes to zero, we obtain

$$\begin{aligned}\lim_{\sigma\downarrow 0}\frac{\mathbb{E}\left[y_1^{2k}\right]}{\sigma^{2k}} &= \sum_{j\in J}\binom{2k}{j}\sum_{\nu=0}^{2k-j}\binom{2k-j}{\nu}\vartheta^{2k-j-\nu}\xi^{\nu}c_{\varepsilon,j}(2k-j-1)!!\\ &= \sum_{j\in J}\binom{2k}{j}(\vartheta+\xi)^{2k-j}c_{\varepsilon,j}(2k-j-1)!!\,.\end{aligned}$$

This proves (2.8) for $t=1$ by combining the previous display with (2.20) and (2.21).
Now assume that (2.8) holds for $t-1, t\geq 2$. Then, due to stationarity $\mathbb{E}\left[y_t^{2k}\right]=m_{2k}$, independent of t. Formula (2.22) then implies

$$\begin{aligned}m_{2k} &= \vartheta^{2k}\mathbb{E}\left[y_{t-1}^{2k}\right]+\sum_{\nu=1}^{2k}\binom{2k}{\nu}\vartheta^{2k-\nu}\xi^{\nu}\mathbb{E}\left[\exp\left(-\nu\gamma y_{t-1}^2\right)y_{t-1}^{2k}\right]\\ &\quad+\sum_{j\in J\setminus\{0\}}\binom{2k}{j}\sum_{\nu=0}^{2k-j}\binom{2k-j}{\nu}\vartheta^{2k-j-\nu}\xi^{\nu}\mathbb{E}\left[\exp\left(-\nu\gamma y_{t-1}^2\right)y_{t-1}^{2k-j}\right]c_{\varepsilon,j}\sigma^j\end{aligned}$$

and eventually

$$\begin{aligned}\frac{m_{2k}}{\sigma^{2k}} &= \frac{1}{1-\vartheta^{2k}}\left[\sum_{\nu=1}^{2k}\binom{2k}{\nu}\vartheta^{2k-\nu}\xi^{\nu}\frac{\mathbb{E}\left[\exp\left(-\nu\gamma y_{t-1}^2\right)y_{t-1}^{2k}\right]}{\sigma^{2k}}\right.\\ &\quad\left.+\sum_{j\in J\setminus\{0\}}\binom{2k}{j}\sum_{\nu=0}^{2k-j}\binom{2k-j}{\nu}\vartheta^{2k-j-\nu}\xi^{\nu}\frac{\mathbb{E}\left[\exp\left(-\nu\gamma y_{t-1}^2\right)y_{t-1}^{2k-j}\right]}{\sigma^{2k-j}}c_{\varepsilon,j}\right]=:u_k(\sigma)\end{aligned}$$

with

$$l_k(\sigma) = \frac{m_{2k}}{\sigma^{2k}}-n\gamma\frac{m_{2k+2}}{\sigma^{2k+2}}\sigma^2 = u_k(\sigma)-n\gamma\frac{m_{2k+2}}{\sigma^{2k+2}}\sigma^2$$

and

$$\begin{aligned}\lim_{\sigma\downarrow 0}l_k(\sigma) &= \lim_{\sigma\downarrow 0}u_k(\sigma)\\ &= \frac{1}{1-\vartheta^{2k}}\left[\sum_{\nu=1}^{2k}\binom{2k}{\nu}\vartheta^{2k-\nu}\xi^{\nu}c_k+\sum_{j\in J\setminus\{0\}}\binom{2k}{j}\sum_{\nu=0}^{2k-j}\binom{2k-j}{\nu}\vartheta^{2k-j-\nu}\xi^{\nu}c_{k-j/2}\,c_{\varepsilon,j}\right]\end{aligned}$$

by assumption so that the constants $c_k, k\in\mathbb{N}$, can be determined recursively by

$$\begin{aligned}c_k &= \frac{1}{1-\frac{1}{1-\vartheta^2}\left[(\vartheta+\xi)^{2k}-\vartheta^{2k}\right]}\frac{1}{1-\vartheta^{2k}}\sum_{j\in J\setminus\{0\}}\binom{2k}{j}\sum_{\nu=0}^{2k-j}\binom{2k-j}{\nu}\vartheta^{2k-j-\nu}\xi^{\nu}c_{k-j/2}\,c_{\varepsilon,j}\\ &= \frac{1}{1-(\vartheta+\xi)^{2k}}\sum_{j\in J\setminus\{0\}}\binom{2k}{j}c_{k-j/2}\,c_{\varepsilon,j}(\vartheta+\xi)^{2k-j}\end{aligned}$$

with $c_0=1$.
According to statement (i) the effect of the initially chosen distribution for y_0 dies out as t

increases due to the ergodicity. Hence, condition (2.8) holds for the underlying stationary distribution F_s of y_t and hence for all $t \in \mathbb{Z}$ regardless the chosen distribution for y_0. □

Lemma 2.6.1.
Let X be a real valued random variable with symmetric density around zero. Then also the density of $[a + b\exp(-c\eta(X^2))]X^{2k+1}$ for some $k \in \mathbb{N}_0$ and $a, b, c \in \mathbb{R}$, $|a| + |b| > 0, c \geq 0$, with $\eta \in \{\eta_1 : [0,\infty) \to \mathbb{R}, \eta_1(z) = z;\ \eta_2 : [0,\infty) \to \mathbb{R}, \eta_2(z) = \log(1+z)\}$ is symmetric around zero.

Proof.
The result is obtained by applying theorems deriving the density of a transformed random variable (see e.g. Theorems 22.2 and 22.3 in Behnen and Neuhaus (1995)).
First note, that the condition $|a| + |b| > 0$ simply guarantees that a and b do not vanish at the same time making the statement of the lemma redundant. Wlog we restrict ourselves to the case $c > 0$ as $c = 0$ is incorporated in the computations for $b = 0$.
First let $b = 0$. For $a \in \mathbb{R} \setminus \{0\}$ and $c > 0$ define $g_k : \mathbb{R} \to \mathbb{R}$, $g_k(x) = ax^{2k+1}$ for some fixed $k \in \mathbb{N}_0$ as well as $Y_k = g_k(X)$. Since $g_k'(x) = (2k+1)ax^{2k}$, $x \in \mathbb{R}$, and $g_k^{-1}(x) = (x/a)^{1/(2k+1)}$, $x \in \mathbb{R}$, we obtain

$$f_{Y_k}(y) = \begin{cases} 0 & \text{for } y = 0, \\ \frac{f_X(g_k^{-1}(y))}{|g_k'(g_k^{-1}(y))|} = \frac{f_X\left(y^{\frac{1}{2k+1}} a^{-\frac{1}{2k+1}}\right)}{(2k+1)\left(y^{\frac{1}{2k+1}} a^{-\frac{1}{2k+1}}\right)^{2k}} & \text{for } y \neq 0, \end{cases}$$

where f_X and f_{Y_k} denote the densities of X and Y_k, respectively. Note that $(-y/a)^{1/(2k+1)} = (-1)^{1/(2k+1)}(y/a)^{1/(2k+1)} = -(y/a)^{1/(2k+1)}$ which implies, for $y \neq 0$,

$$f_{Y_k}(-y) = \frac{f_X\left(-y^{\frac{1}{2k+1}} a^{-\frac{1}{2k+1}}\right)}{(2k+1)\left(-y^{\frac{1}{2k+1}} a^{-\frac{1}{2k+1}}\right)^{2k}} = \frac{f_X\left(y^{\frac{1}{2k+1}} a^{-\frac{1}{2k+1}}\right)}{(2k+1)\left(y^{\frac{1}{2k+1}} a^{-\frac{1}{2k+1}}\right)^{2k}} = f_{Y_k}(y), \tag{2.23}$$

where the second last equality is due to the symmetry of f_X.
Now consider the case $b \neq 0$ which is incomparable more complex. Let $a \in \mathbb{R}$ and $c > 0$ with $g_{k,i} : \mathbb{R} \to \mathbb{R}$, $g_{k,i}(x) = (a + b\exp(-c\eta_i(x^2)))x^{2k+1}$ for $i = 1, 2$ and some fixed $k \in \mathbb{N}_0$ and $Y_{k,i} = g_{k,i}(X)$. Contrary to the case $b = 0$, $g_{k,i}$ can now change its monotonic behavior and might therefore be only piecewise invertible.
We can verify the following properties (see below) for all $k \in \mathbb{N}_0$:

(i) $g_{k,i}$ is continuous on $\mathbb{R}$ with $g_{k,i}(0) = 0$, $i = 1, 2$,

(ii) $g_{k,i}$ is point symmetric around zero, $i = 1, 2$,

(iii) the limit behavior for $x \to \infty$ can be described by

$$\lim_{x\to\infty} g_{k,i}(x) = \lim_{x\to\infty}\left(ax^{2k+1} + b\frac{x^{2k+1}}{e^{c\eta_i(x^2)}}\right) = \lim_{x\to\infty} ax^{2k+1} = \begin{cases} -\infty & \text{for } a < 0, \\ \kappa_i & \text{for } a = 0, \\ \infty & \text{for } a > 0 \end{cases} \tag{2.24}$$

with

$$\kappa_i = \begin{cases} 0 & \text{for } i=1, \\ 0 & \text{for } i=2, c > k+1/2, \\ 1 & \text{for } i=2, b>0, c=k+1/2, \\ -1 & \text{for } i=2, b<0, c=k+1/2, \\ \infty & \text{for } i=2, b>0, c<k+1/2, \\ -\infty & \text{for } i=2, b<0, c<k+1/2. \end{cases}$$

(iv) the monotonic behavior of $g_{k,1}$, $k \in \mathbb{N}_0$, (corresponding to the ESTAR model) can be summarized in the following table, where $\xi_{k,1} = \frac{2}{2k+1} e^{-\frac{2k+3}{2}}$:

$b>0$	$a > \xi_{k,1} b > 0$	$g_{k,1}$ is strictly monotone increasing
$b>0$	$a < -b < 0$	$g_{k,1}$ is strictly monotone decreasing
$b>0$	$a \in [-b, \xi_{k,1} b]$	
	$= [-b, 0]$	$g_{k,1}$ changes its monotone behavior twice, starting with being strictly decreasing
	$\cup (0, \xi_{k,1} b]$	$g_{k,1}$ changes its monotone behavior four times, starting with being strictly increasing
$b<0$	$a > -b > 0$	$g_{k,1}$ is strictly monotone increasing
$b<0$	$a < \xi_{k,1} b < 0$	$g_{k,1}$ is strictly monotone decreasing
$b<0$	$a \in [\xi_{k,1} b, -b]$	
	$= [\xi_{k,1} b, 0)$	$g_{k,1}$ changes its monotone behavior four times, starting with being strictly decreasing
	$\cup [0, -b]$	$g_{k,1}$ changes its monotone behavior twice, starting with being strictly increasing

If $g_{k,1}$ changes its monotonic behavior twice, it always happens at

$$w_{1,2} = \pm \sqrt{\frac{-W_0\left(-\frac{a}{b}\frac{2k+1}{2} e^{\frac{2k+1}{2}}\right) + \frac{2k+1}{2}}{c}}$$

no matter which parameter combination for a and b we consider. Here, W_0 denotes the principal branch of the Lambertsche W-function with domain $[-\exp(-1), \infty)$, i.e. the function that satisfies $x = W_0(x)\exp(W_0(x))$. If $g_{k,1}$ changes its behavior four times, it additionally

happens at

$$w_{3,4} = \pm\sqrt{\frac{-W_{-1}\left(-\frac{a}{b}\frac{2k+1}{2}e^{\frac{2k+1}{2}}\right)+\frac{2k+1}{2}}{c}}$$

where W_{-1} is the second real branch of the W-function defined on $[\exp(-1), 0)$.

Figures 2.14 and 2.15 illustrate the function $g_{0,1}$ for different parameters a and b with $c = 1$. While we choose $b = 1, a = -1.2 < -b$ in Figure 2.14, the latter corresponds to $b = 1, a = 0.2 \in (0, -\xi_{0,1}b] = (0, 0.4463]$.

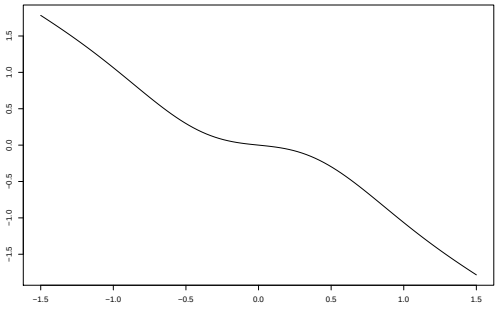

Figure 2.14: $g_{0,1}$ with $a = -1.2$ and $b = 1$.

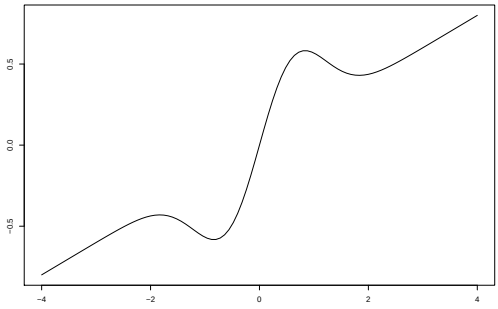

Figure 2.15: $g_{0,1}$ with $a = 0.2$ and $b = 1$.

(v) the monotonic behavior of $g_{k,2}, k \in \mathbb{N}_0$, (corresponding to the TSTAR model) could be summarized in a similar explicit table as in (iv) using $\xi_{k,c,2} = \frac{1(2c-2k-1)}{(2k+1)(2c+2)}$ depending on different ranges for the parameter c. Later on it is only important that $g_{k,2}$ is piecewise strictly monotone so that the details are skipped due to space. However the proof is exactly the same as for (iv).

Properties (i)-(iii) are easily verified. We prove property (iv) by writing, for $x \in \mathbb{R}$,

$$\begin{aligned} g'_{k,i}(x) &= (2k+1)bx^{2k}\left[\frac{a}{b} + \exp(-c\,\eta_i(x^2))\left(1 - \frac{2}{2k+1}\,cx^2\eta'_i(x^2)\right)\right] \\ &= (2k+1)bx^{2k}\left[\frac{a}{b} - h_{k,i}(x)\right], \end{aligned}$$

with

$$\begin{aligned} h_{k,i} : \mathbb{R} \to \mathbb{R},\ h_{k,i}(x) &= -\frac{2}{2k+1}\exp(-c\,\eta_i(x^2))\left(\frac{2k+1}{2} - cx^2\eta'_i(x^2)\right) \text{ and} \\ h'_{k,i} : \mathbb{R} \to \mathbb{R},\ h'_{k,i}(x) &= -\frac{2}{2k+1}\,cx\exp(-c\,\eta_i(x^2))\cdot \\ &\qquad \left(-2\eta'_i(x^2) - 2x^2\eta''_i(x^2) - \frac{2k+1}{2}2\eta'_i(x^2) + 2cx^2\eta'_i(x^2)^2\right) \end{aligned}$$

Consider $i = 1$. Let $k \geq 0$ and assume $b > 0$.

Then $(2k+1)bx^{2k} > 0$ for $x \in \mathbb{R} \setminus \{0\}$. Hence, $g'_{k,1}(x) > 0, x \in \mathbb{R} \setminus \{0\}$, (i.e. strictly monotone increasing) if $a/b > h_{k,1}(x)$ for all $x \in \mathbb{R} \setminus \{0\}$. As

$$\max_{x\in\mathbb{R}} h_{k,1}(x) \;=\; \max\left\{h_{k,1}(0), h_{k,1}\left(\pm\sqrt{\frac{2k+3}{2c}}\right)\right\} \;=\; \frac{2}{2k+1}\exp\left(-\frac{2k+3}{2}\right) =: \xi_{k,1},$$

we obtain a strictly increasing $g_{k,1}$ for $a/b > \xi_{k,1}$ and by a similar argument a strictly decreasing $g_{k,1}$ as long as $a/b < -1$ since

$$\min_{x\in\mathbb{R}} h_{k,1}(x) \;=\; h_{k,1}(0) \;=\; -1.$$

For $a/b \in [-1, \xi_{k,1}]$ the monotone behavior changes, driven by the sign of the parameter a (see (2.24)). Note that $g'_{k,1}(x) = 0$ whenever $x = 0$ $(k \geq 1)$ or

$$\frac{a}{b} + \left(\frac{2k+1}{2} - cx^2\right)\frac{2}{2k+1}\exp(-cx^2) \;=\; 0$$

which is equivalent to

$$\left(\frac{2k+1}{2} - cx^2\right)\exp\left(\frac{2k+1}{2} - cx^2\right) \;=\; -\frac{a}{b}\frac{2k+1}{2}\exp\left(\frac{2k+1}{2}\right). \tag{2.25}$$

For solving (2.25) we need to consider two different cases. Note that for $a \in (0, \xi_{k,1} b]$, the right hand side of (2.25) is contained in $[-\exp(-1), 0)$, hence in the range where W has two real-values branches, denoted by W_0 and W_{-1}. Therefore, from (2.25), for $j = 0, -1$,

$$\begin{aligned} \tfrac{2k+1}{2} - cx^2 \;&=\; W_j\left(-\tfrac{a}{b}\tfrac{2k+1}{2}\exp\left(\tfrac{2k+1}{2}\right)\right) \qquad (2.26)\\ \Leftrightarrow \quad w_{1,2,3,4} \;&=\; \pm\sqrt{\tfrac{-W_j\left(-\frac{a}{b}\frac{2k+1}{2}\exp\left(\frac{2k+1}{2}\right)\right)+\frac{2k+1}{2}}{c}} \end{aligned}$$

which are well defined as

$$-W_0\left(-\frac{a}{b}\frac{2k+1}{2}\exp\left(\frac{2k+1}{2}\right)\right) \;\geq\; -W_0\left(\frac{2k+1}{2}\exp\left(\frac{2k+1}{2}\right)\right) \geq -\frac{2k+1}{2}$$

and

$$-W_{-1}\left(-\frac{a}{b}\frac{2k+1}{2}\exp\left(\frac{2k+1}{2}\right)\right) \;\geq\; -W_{-1}(-\exp(-1)) \;=\; 1.$$

For $a \in [-1, 0]$ the argument of W_j in (2.26) is in the domain of only one real branch, namely W_0, so that (2.25) has only two solution leading to two monotone changes of $g_{k,1}$.
An analogue argument shows the behavior of g_k if $b < 0$.
For $i = 2$ one could work out the details in the same way as for $i = 1$.
For those combinations of a, b and c where $g_{k,i}, i = 1, 2$, is strictly monotone on $\mathbb{R}$, symmetry of the denisty of $f_{Y_{k,i}}$ can be derived in the same way as in (2.23). Note that $g'_{k,i}, i = 1, 2$, is symmetric around zero and that $g^{-1}_{k,i}, i = 1, 2$, is point symmetric around zero as an inverse function shares

this property with $g_{k,i}, i = 1,2$. Thus, for $y \in \mathbb{R}$, $i = 1,2$,

$$f_{Y_{k,i}}(-y) = \frac{f_X\left(g_{k,i}^{-1}(-y)\right)}{\left|g'_{k,i}\left(g_{k,i}^{-1}(-y)\right)\right|} = \frac{f_X\left(-g_{k,i}^{-1}(-y)\right)}{\left|g'_{k,i}\left(-g_{k,i}^{-1}(-y)\right)\right|} = \frac{f_X\left(g_{k,i}^{-1}(y)\right)}{\left|g'_{k,i}\left(g_{k,i}^{-1}(y)\right)\right|} = f_{Y_{k,i}}(y).$$

For the remaining parameter combinations of a, b and c we divide the real line in disjunct open intervals on which $g_{k,i}$ possesses a strictly monotone behavior. As an example we present the argument for the situation $i = 1$, $b > 0$ and $a \in (0, \xi_{k,1}\, b]$ where the open intervals can be determined explicitly. All other cases can be done in an analogous way and are omitted due to space. Define

$$G^{(1)} = (-\infty, -z_2) \cup (z_2, \infty),\ G^{(2)} = (-z_2, -z_1) \cup (z_1, z_2) \text{ and } G^{(3)} = (-z_1, z_1)$$

where $z_1 = w_1$ and $z_2 = w_3$. On the open sets $G^{(j)}$, $j = 1, \ldots, 3$, $g_{k,1}^{(j)} := g_{k,1} \mathbf{1}_{G^{(j)}}$, $j = 1, \ldots, 3$, is strictly monotone with derivatives and inverse functions $g_{k,1}^{(j)'}$ and $(g_{k,1}^{(j)})^{-1}$, respectively. Hence, from Theorem 22.3 of Behnen and Neuhaus (1995), $k \in \mathbb{N}_0$,

$$f_{Y_{k,1}}(y) = \sum_{j=1}^{3} \frac{f_X\left(\left(g_{k,1}^{(j)}\right)^{-1}(y)\right)}{\left|g_{k,1}^{(j)'}\left(\left(g_{k,1}^{(j)}\right)^{-1}(y)\right)\right|} \mathbf{1}_{y \in H_{k,i}},\ y \in \mathbb{R},$$

for $H_{k,i} = g_{k,1}(G^{(i)})$. As $H_{k,1} \cup H_{k,2} \cup H_{k,3} = \mathbb{R}$, as $g_{k,1}^{(j)'}$, $j = 1, \ldots, 3$, are symmetric around zero and as $(g_{k,1}^{(j)})^{-1}, i = 1, \ldots, 3$, are point symmetric around zero, we obtain $f_{Y_{k,1}}(-y) = f_{Y_{k,1}}(y)$ for all $y \in \mathbb{R}$. □

Lemma 2.6.2.
Let X and Y be independent real valued random variables with densities f_X and f_Y, respectively. Symmetries of f_X around $c \in \mathbb{R}$ ($f_X(x + c) = f_X(-x + c), x \in \mathbb{R}$) and f_Y around $d \in \mathbb{R}$ then imply that the convolution density f_Z of $Z = X + Y$ is symmetric around $c + d$.

Proof.
The density f_Z of Z is given by

$$f_Z(z) = \int_{\mathbb{R}} f_X(z - x) f_Y(x)\, dx,\ z \in \mathbb{R}.$$

Hence, for all $x \in \mathbb{R}$,

$$\begin{aligned} f_Z(-x + c + d) &= \int_{\mathbb{R}} f_X(-x + c + d - y) f_Y(y)\, dy = \int_{\mathbb{R}} f_X(-(x - d + y) + c) f_y(y)\, dy \\ &= \int_{\mathbb{R}} f_X(x - d + y + c) f_y(y)\, dy = \int_{\mathbb{R}} f_X(x + c + d + y - 2d) f_Y(y)\, dy \\ &= \int_{\mathbb{R}} f_X(x + c + d - z) f_Y(-z + 2d)\, dz = \int_{\mathbb{R}} f_X(x + c + d - z) f_Y(-(z - d) + d)\, dz \\ &= \int_{\mathbb{R}} f_X(x + c + d - z) f_Y(z - d + d)\, dz = f_Z(x + c + d). \quad \square \end{aligned}$$

Proof of Theorem 2.2.4.
Let $\hat{\beta}_n$ be the conditional Maximum Likelihood estimator of $\beta = (\vartheta, \xi, \gamma)$. Then by Theorem 3.2 of Tjøstheim (1986)

$$n^{1/2}(\hat{\beta}_n - \beta) \quad \xrightarrow{d} \quad \mathrm{N}(0, \sigma^2 U^{-1})$$

and where the matrix $U = \left(u_{ij}\right)_{i,j=1,\ldots,3}$ is given by

$$U \quad = \quad \begin{pmatrix} \mathbb{E}[y_{t-1}^2] & \mathbb{E}[y_{t-1}^2 \exp(-\gamma y_{t-1}^2)] & -\mathbb{E}[\xi y_{t-1}^4 \exp(-\gamma y_{t-1}^2)] \\ \mathbb{E}[y_{t-1}^2 \exp(-\gamma y_{t-1}^2)] & \mathbb{E}[y_{t-1}^2 \exp(-2\gamma y_{t-1}^2)] & -\mathbb{E}[\xi y_{t-1}^4 \exp(-2\gamma y_{t-1}^2)] \\ -\mathbb{E}[\xi y_{t-1}^4 \exp(-\gamma y_{t-1}^2)] & -\mathbb{E}[\xi y_{t-1}^4 \exp(-2\gamma y_{t-1}^2)] & \mathbb{E}[\xi^2 y_{t-1}^6 \exp(-2\gamma y_{t-1}^2)] \end{pmatrix}.$$

In order to obtain the limiting behavior of $\mathrm{Var}(\hat{\gamma})$ we therefore study

$$\sigma^2 \left(U^{-1}\right)_{33} \quad = \quad \frac{\sigma^2}{\det(U)} \det \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix}. \tag{2.27}$$

The last factor of (2.27) behaves like $O(\sigma^4)$ for $\sigma \to 0$ as

$$\begin{aligned} &\lim_{\sigma \downarrow 0} \sigma^{-4} \det \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix} \\ &\quad = \quad \lim_{\sigma \downarrow 0} \sigma^{-4} \left(\mathbb{E}[y_{t-1}^2] \, \mathbb{E}[y_{t-1}^2 \exp(-2\gamma y_{t-1}^2)] - \left[\mathbb{E}[y_{t-1}^2 \exp(-\gamma y_{t-1}^2)] \right]^2 \right) \; = \; c_4 \end{aligned}$$

due to Lemma 2.2.3 (iv) for some constant c_4. By a similar argument we obtain $\det(U) = O(\sigma^{10})$. Hence

$$\lim_{\sigma \downarrow 0} \frac{\sigma^2}{\det(U)} \det \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix} \quad = \quad \lim_{\sigma \downarrow 0} \frac{\sigma^6 \sigma^{-4} \det \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix}}{\sigma^{10} \sigma^{-10} \det(U)} \; = 0$$

which finishes the proof. □

Proof of Theorem 2.3.2.
In order to derive the asymptotic distribution of F^* we first study the asymptotic behavior of $\hat{\beta} = (\hat{\delta}_{1,2}, \hat{\delta}_{1,4}, \hat{\delta}_{1,6})$. Under the null $\Delta y_t = u_t$, $\hat{\beta}$ can be written as

$$\hat{\beta} \quad = \quad \left(\sum_{t=1}^{T} x_t' x_t \right)^{-1} \sum_{t=1}^{T} x_t' u_t \tag{2.28}$$

with

$$\sum_{t=1}^{T} x_t' x_t = \begin{bmatrix} \sum_{t=1}^{T} y_{t-1}^6 & \sum_{t=1}^{T} y_{t-1}^8 & \sum_{t=1}^{T} y_{t-1}^{10} \\ \sum_{t=1}^{T} y_{t-1}^8 & \sum_{t=1}^{T} y_{t-1}^{10} & \sum_{t=1}^{T} y_{t-1}^{12} \\ \sum_{t=1}^{T} y_{t-1}^{10} & \sum_{t=1}^{T} y_{t-1}^{12} & \sum_{t=1}^{T} y_{t-1}^{14} \end{bmatrix}$$

and

$$\sum_{t=1}^{T} x_t' u_t = \begin{bmatrix} \sum_{t=1}^{T} y_{t-1}^3 u_t & \sum_{t=1}^{T} y_{t-1}^5 u_t & \sum_{t=1}^{T} y_{t-1}^7 u_t \end{bmatrix}.$$

In order to determine the asymptotic behavior we have to scale the estimator $\hat{\beta}$ properly. We thus multiply $\hat{\beta}$ with the scaling matrix $\Gamma = diag(T^2, T^3, T^4)$ and obtain

$$\left[\Gamma^{-1} \sum_{t=1}^{T} x_t' x_t \Gamma^{-1} \right]^{-1} \left[\Gamma^{-1} \sum_{t=1}^{T} x_t' u_t \right] = \Gamma \hat{\beta}.$$

Now the asymptotic behavior of the first part of $\Gamma\hat{\beta}$ follows directly from Hamilton (1989) (p. 479 ff.) and the behavior of the second part follows from the convergence to stochastic integrals for products of I(1) variables (Theorem 4.2 from Hansen (1992), Sandberg (2009), the CMT and Itô's Lemma). This yields the following general result for $i \in \mathbb{N}_{>0}$ and $T \to \infty$

$$\frac{1}{T^{(i+1)/2}} \sum_{t=1}^{T} y_{t-1}^i u_t \quad \Rightarrow \quad \int_0^1 B^i(r)\, dB(r) = \sigma^{(i+1)} \left\{ \frac{1}{(i+1)} B(1)^{(i+1)} - \frac{i}{2} \int_0^1 B(r) dr \right\}.$$

Given these results the OLS estimator converges as $T \to \infty$ as follows

$$\Gamma\hat{\beta} \quad = \quad \left[\Gamma^{-1} \sum_{t=1}^{T} x_t' x_t \Gamma^{-1} \right]^{-1} \left[\Gamma^{-1} \sum_{t=1}^{T} x_t' u_t \right] \Rightarrow \mathbf{Q}^{-1}\mathbf{v}$$

where

$$\mathbf{Q} \quad = \quad \begin{bmatrix} \sigma^6 \int_0^1 B^6(r) dr & \sigma^8 \int_0^1 B^8(r) dr & \sigma^{10} \int_0^1 B^{10}(r) dr \\ \sigma^8 \int_0^1 B^8(r) dr & \sigma^{10} \int_0^1 B^{10}(r) dr & \sigma^{12} \int_0^1 B^{12}(r) dr \\ \sigma^{10} \int_0^1 B^{10}(r) dr & \sigma^{12} \int_0^1 B^{12}(r) dr & \sigma^{14} \int_0^1 B^{14}(r) dr \end{bmatrix}$$

and

$$\mathbf{v} \;=\; \begin{bmatrix} \sigma^4 \left\{ \frac{1}{4}\int_0^1 B(1)^4 - \frac{3}{2}\int_0^1 B(r)\,dr \right\} \\ \sigma^6 \left\{ \frac{1}{6}\int_0^1 B(1)^6 - \frac{5}{2}\int_0^1 B(r)\,dr \right\} \\ \sigma^8 \left\{ \frac{1}{8}\int_0^1 B(1)^8 - \frac{7}{2}\int_0^1 B(r)\,dr \right\} \end{bmatrix}.$$

The scaled F-statistic we are concerned with reads

$$F^* \;=\; \frac{1}{3}\Gamma\hat{\beta}'\left[\hat{\sigma}^2\Gamma(X'X)^{-1}\Gamma\right]^{-1}\Gamma\hat{\beta}\,.$$

By the law of large numbers it is easy to show that under the null as $T \to \infty$

$$\hat{\sigma}^2 \;=\; \frac{1}{T-4}\sum_{t=1}^{T}\left(\Delta y_t - \hat{\delta}_0 y_{t-1}^3 - \hat{\delta}_1 y_{t-1}^5 - \hat{\delta}_2 y_{t-1}^7\right)^2 \xrightarrow{P} \sigma^2$$

and hence

$$F^* \Rightarrow \frac{1}{3\sigma^2}(\mathbf{Q}^{-1}\mathbf{v})'(\mathbf{Q}^{-1})^{-1}(\mathbf{Q}^{-1}\mathbf{v}) = \frac{1}{3\sigma^2}\mathbf{v}'\mathbf{Q}^{-1}\mathbf{v}.$$

Under the alternative Δy_t and $y_{t-1}^i, \forall i \in \mathbb{N}_{>0}$ are $I(0)$ and thus it is readily seen that

$$\frac{1}{T}\sum_{t=1}^{T}\Delta y_t = O_P(1) \quad \text{and} \quad \frac{1}{T}\sum_{t=1}^{T} y_{t-1}^i = O_P(1)$$

are bounded in probability. Furthermore the innovation process u_t is by assumption $I(0)$ and thus

$$\frac{1}{T}\sum_{t=1}^{T} u_t \;=\; O_P(1)$$

as well. As

$$T\frac{1}{T}\sum_{t=1}^{T} x_t' x_t \;=\; T\,O_P(1) \;=\; O_P(T)\,.$$

and

$$\sum_{t=1}^{T} x_t' u_t \;=\; \sum_{t=1}^{T} x_t' \sum_{t=1}^{T} u_t \;=\; T\frac{1}{T}\left(\sum_{t=1}^{T} x_t' \sum_{t=1}^{T} u_t\right) \;=\; T\underbrace{\frac{1}{T}\sum_{t=1}^{T} x_t'}_{O_P(1)}\underbrace{\frac{1}{T}\sum_{t=1}^{T} u_t}_{O_P(1)} \;=\; T^2 O_P(1) \;=\; O_P(T^2)$$

we get for the OLS estimator $\hat{\beta}$, according to (2.28),

$$\hat{\beta} \;=\; (O_P(T))^{-1} O_P(T^2) \;=\; (TO_P(1))^{-1} T^2 O_P(1) \;=\; \frac{1}{T} T^2 O_P(1) \;=\; TO_P(1) \;=\; O_P(T)\,.$$

Analogously for the test statistic F^* it follows that

$$F^* \;=\; \frac{1}{3\hat{\sigma}^2}\hat{\beta}'(X'X)\hat{\beta} \;=\; \frac{1}{3\hat{\sigma}^2} O_P(T) O_P(T) O_P(T) \;=\; \frac{1}{3\hat{\sigma}^2} O_P(T^3)\,.$$

Hence as $T \to \infty$ the test statistic F^* diverges to infinity. □

Proof of Theorem 2.3.3.
We have to show that the inner product of the regressor matrix including the additional regressors is asymptotically block diagonal (see e.g. Hamilton (1994) or Hatanaka (1996)). The inner product $(X'X)$ of the regressor matrix from (2.19) reads

$$\left[\begin{array}{ccc:ccc}
\sum_{t=1}^{T} y_{t-1}^6 & \sum_{t=1}^{T} y_{t-1}^8 & \sum_{t=1}^{T} y_{t-1}^{10} & \sum_{t=1}^{T} y_{t-1}^3 \Delta y_{t-1} & \cdots & \sum_{t=1}^{T} y_{t-1}^3 \Delta y_{t-p} \\
\sum_{t=1}^{T} y_{t-1}^8 & \sum_{t=1}^{T} y_{t-1}^{10} & \sum_{t=1}^{T} y_{t-1}^{12} & \sum_{t=1}^{T} y_{t-1}^5 \Delta y_{t-1} & \cdots & \sum_{t=1}^{T} y_{t-1}^5 \Delta y_{t-p} \\
\sum_{t=1}^{T} y_{t-1}^{10} & \sum_{t=1}^{T} y_{t-1}^{12} & \sum_{t=1}^{T} y_{t-1}^{14} & \sum_{t=1}^{T} y_{t-1}^7 \Delta y_{t-1} & \cdots & \sum_{t=1}^{T} y_{t-1}^7 \Delta y_{t-p} \\
\hdashline
\sum_{t=1}^{T} \Delta y_{t-1} y_{t-1}^3 & \sum_{t=1}^{T} \Delta y_{t-1} y_{t-1}^5 & \sum_{t=1}^{T} \Delta y_{t-1} y_{t-1}^7 & \sum_{t=1}^{T} \Delta y_{t-1}^2 & \cdots & \sum_{t=1}^{T} \Delta y_{t-1} \Delta y_{t-p} \\
\vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\
\sum_{t=1}^{T} \Delta y_{t-p} y_{t-1}^3 & \sum_{t=1}^{T} \Delta y_{t-p} y_{t-1}^5 & \sum_{t=1}^{T} \Delta y_{t-p} y_{t-1}^7 & \sum_{t=1}^{T} \Delta y_{t-p} \Delta y_{t-1} & \cdots & \sum_{t=1}^{T} \Delta y_{t-p}^2
\end{array}\right].$$

Remember (see e.g Hamilton, 1994, p.517) that an AR(p) process

$$(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p) y_t \;=\; \varepsilon_t$$

can be written equivalently as

$$\{(1-\rho L) - (\zeta_1 L + \zeta_2 L^2 + \ldots + \zeta_{p-1} L^{p-1})(1-L)\} y_t \;=\; \varepsilon_t$$

where $\rho = \phi_1 + \phi_2 + \cdots + \phi_p$ and $\zeta_j = -[\phi_{j+1} + \phi_{j+2} + \ldots + \phi_p]$ for $j = 1, 2, \ldots, p-1$. Under the assumption of a unit root, i.e. $\rho = 1$, the process can be written as

$$(\zeta_1 L - \zeta_2 L^2 - \ldots - \zeta_{p-1} L^{p-1}) \Delta y_t \;=\; \varepsilon_t$$

or

$$\Delta y_t \;=\; e_t$$

where $e_t = (\zeta_1 L - \zeta_2 L^2 - \ldots - \zeta_{p-1} L^{p-1})^{-1}$. The behavior of the process y_t is such that it fulfills proposition 17.3 in Hamilton (1994, p.505).

First, letting $e_t = y_t - y_{t-1}$ we obtain

$$\left[\begin{array}{ccc:ccc} \sum_{t=1}^{T} y_{t-1}^6 & \sum_{t=1}^{T} y_{t-1}^8 & \sum_{t=1}^{T} y_{t-1}^{10} & \sum_{t=1}^{T} y_{t-1}^3 e_{t-1} & \cdots & \sum_{t=1}^{T} y_{t-1}^3 e_{t-p} \\ \sum_{t=1}^{T} y_{t-1}^8 & \sum_{t=1}^{T} y_{t-1}^{10} & \sum_{t=1}^{T} y_{t-1}^{12} & \sum_{t=1}^{T} y_{t-1}^5 e_{t-1} & \cdots & \sum_{t=1}^{T} y_{t-1}^5 e_{t-p} \\ \sum_{t=1}^{T} y_{t-1}^{10} & \sum_{t=1}^{T} y_{t-1}^{12} & \sum_{t=1}^{T} y_{t-1}^{14} & \sum_{t=1}^{T} y_{t-1}^7 e_{t-1} & \cdots & \sum_{t=1}^{T} y_{t-1}^7 e_{t-p} \\ \hdashline \sum_{t=1}^{T} e_{t-1} y_{t-1}^3 & \sum_{t=1}^{T} e_{t-1} y_{t-1}^5 & \sum_{t=1}^{T} e_{t-1} y_{t-1}^7 & \sum_{t=1}^{T} e_{t-1}^2 & \cdots & \sum_{t=1}^{T} e_{t-1} e_{t-p} \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ \sum_{t=1}^{T} e_{t-p} y_{t-1}^3 & \sum_{t=1}^{T} e_{t-p} y_{t-1}^5 & \sum_{t=1}^{T} e_{t-p} y_{t-1}^7 & \sum_{t=1}^{T} e_{t-p} e_{t-1} & \cdots & \sum_{t=1}^{T} e_{t-p}^2 \end{array}\right].$$

Using the results (c) and (e) stated in proposition 17.3 in Hamilton (1994, p.505) combined with the CMT and the results from theorem 2.3.2 we have

$$(X'X) \Rightarrow \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{W} \end{bmatrix}$$

where

$$\mathbf{W} = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-2} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-3} \\ \vdots & \vdots & \cdots & \vdots \\ \gamma_{p-2} & \gamma_{p-3} & \cdots & \gamma_0 \end{bmatrix}, \ \gamma_j = E[(\Delta y_t)(\Delta y_{t-j})]$$

and where $\mathbf{Q}$ is as given in Theorem 2.3.2 but with σ^k replaced by its long-run counterpart given by $\lambda = \sigma/(1-\zeta_1-\ldots-\zeta_{p-1})$. Thus the inner product of the regressor matrix is asymptotically block diagonal and therefore the distribution of the coefficients δ_1^2, δ_1^4 and δ_1^6 is independent of the distribution of the additional regressors.

Using similar arguments as in Theorem 2.3.2 it is straightforward to show that the test is consistent under (2.19). □

Chapter 3

Evaluating a class of nonlinear time series models

Evaluating a class of nonlinear time series models

Published as Leibniz University of Hannover Discussion Paper No. 445.

3.1 Introduction

Over the last decade and a half the interest in nonlinear time series models has been grown steadily. Especially the class of smooth transition autoregressive (STAR) models, initiated by the work of Bacon and Watts (1971) and popularized by Teräsvirta (1994), has enjoyed great success. A lot of work in this area has been devoted to estimation, specification, testing and applications such as forecasting. For a recent review of this field see e.g. Potter (1999) and van Dijk et al. (2002b).

Within the class of STAR models the two most notable are the logistic STAR (LSTAR) and the exponential STAR (ESTAR) model. The ESTAR model in particular has been very popular with empirical investigations of economic theories such as purchasing power parity (PPP) or the Fisher hypothesis (see e.g. Taylor et al. (2001) and Rose (1988)). It is by now well known that the estimation of the parameters in ESTAR models is notoriously hard as noticed early by Haggan and Ozaki (1981), Tong (1990) or Teräsvirta (1994). In fact, as shown by Heinen et al. (2011) some crucial parameters in ESTAR models are unidentified if the variance of the innovation term becomes very small. This is especially important within the context of real exchange rates as estimated innovation variances are usually extremely small (see e.g. Gatti et al. (1998), Öcal (2000), Taylor et al. (2001) or Rapach and Wohar (2006) among others). In order to remedy this problem Heinen et al. (2011) propose a new type of nonlinear model formulation called TSTAR that maintains the desirable properties of ESTAR but reduces the estimation problem for the most part. In addition the authors propose a linearity test and an unit root test for the new model.

A complete account of this new model however would require to run through all steps of the empirical modeling cycle devised by Teräsvirta (1994). This paper is mainly devoted to the evaluation stage of the modeling cycle and proposes a suite of Lagrange multiplier (LM) tests designed for this newly developed model. The development of specialized parametric tests for nonlinear models is important since standard misspecification tests such as the well known Ljung-Box test have been shown to be badly sized when the true data generating process is nonlinear (see Eitrheim and Teräsvirta (1996)). However, before considering residual based tests we treat the problem of how to choose between two different nonlinear models by proposing a direct test based on the encompassing principle.

The paper is organized as follows: In section 3.2 we review the class of models under study and the empirical modeling cycle for nonlinear models. A method to discriminate between competing nonlinear model formulations is described in section 3.3. In section 3.4 we propose different tests

against serial dependency of the residuals, no remaining nonlinearity of the residuals as well as parameter constancy of the estimated model. In addition we cover evaluation via impulse response analysis. In section 3.6 we study the finite sample performance of the afore mentioned tests. In section 3.7 we run through the whole modeling cycle to model real exchange rates before section 3.8 concludes. Some additional results are collected in the appendix 3.9.

3.2 The modeling cycle

A general STAR model is given by two autoregressive regimes connected by a smooth transition function. Smoothness means that the transition function changes continuously from zero to one and therefore governs the transition between the two regimes in a smooth way. Alternatively, a STAR model can also be interpreted as a continuum of regimes which is passed through by the process.

In general, univariate STAR(p) models, $p \geq 1$ and $d \leq p$, are given by

$$y_t = [\Psi w_t] \times [1 - \mathcal{G}(y_{t-d}; \gamma, c)] + [\Theta w_t] \times \mathcal{G}(y_{t-d}; \gamma, c) + \varepsilon_t \tag{3.1}$$

$$= [\Psi w_t] + [\Phi w_t] \times \mathcal{G}(y_{t-d}; \gamma, c) + \varepsilon_t, \quad t \geq 1, \tag{3.2}$$

with $\varepsilon_t \overset{iid}{\sim} (0, \sigma^2)$.

The parameter vectors Ψ and Θ as well as w_t are given by $\Psi = (\psi_0, \psi_1, \ldots, \psi_p)$, $\Theta = (\vartheta_0, \vartheta_1, \ldots, \vartheta_p)$, and $w_t = (1, y_{t-1}, \ldots, y_{t-p})'$. For the alternative parametrization in (3.2) we have $\Phi = (\varphi_0, \varphi_1, \ldots, \varphi_p) = (\psi_0 - \vartheta_0, \psi_1 - \vartheta_1, \ldots, \psi_p - \vartheta_p)$, i.e. the second regime realizes as sum of Ψ and Φ.

Different choices of the transition function $\mathcal{G}(\cdot; \gamma, c) : \mathbb{R} \to [0, 1]$ lead to different STAR models. A popular choice is the exponential form leading to the ESTAR model

$$\mathcal{G}(\cdot; \gamma, c) = 1 - \exp\left\{-\gamma (y_{t-d} - c)^2\right\}; \quad \gamma > 0\,. \tag{3.3}$$

This functional form for the transition function is popular for modeling real exchange rates or real interest rates (see e.g. Kapetanios et al. (2003)). However the estimation of the parameter γ that governs the functional form and thereby the transition speed is notoriously hard. In fact it is shown by Heinen et al. (2011) that

$$\lim_{\sigma_{\varepsilon_t} \downarrow 0} \mathrm{Var}(\hat{\gamma}) \to \infty\,. \tag{3.4}$$

As the parameter γ is absolutely crucial in STAR models (see Tong (1990)) this result makes it almost impossible to obtain reliable estimation results for reasonable small innovation variances. To remedy this problem Heinen et al. (2011) propose to reformulate the well known transition function in (3.3) as

$$\mathcal{G}(\cdot; \gamma, c) = 1 - \left\{1 + (y_{t-d} - c)^2\right\}^{-\gamma}; \quad \gamma > 0\,. \tag{3.5}$$

The resulting model is called TSTAR model and reduces the identification problem for the most part. Both transition functions in (3.3) and (3.5) share the same properties which make them

applicable in the same situations. For further details we refer to Heinen et al. (2011).
The modeling cycle for nonlinear models as proposed in Teräsvirta (1994) consists of three main steps:

Step 1: Specification

- Specifying a linear autoregressive model via an information criterion such as AIC (see Akaike (1974)) or BIC (see Schwarz (1978)).
- Testing linearity for different values of d and if it is rejected specify d by minimizing the p-value of the linearity test via a grid search over possible values of d (see Tsay (1986) or Teräsvirta (1994)).
- Choosing an adequate transition function by testing a series of nested hypotheses (see Teräsvirta and Anderson (1992), Teräsvirta (1994) and Escribano and Jordá (2001)).

Step 2: Estimation

- Estimate the specified model using either nonlinear least squares or conditional (quasi) maximum likelihood. Consistency for these techniques has been established by Klimko and Nelson (1978) and Tjøstheim (1986) respectively.

Step 3: Evaluation

- Perform residual based tests against serial dependence, no remaining nonlinearity and parameter constancy as proposed in Eitrheim and Teräsvirta (1996). Evaluate the dynamic behavior via generalized impulse response function developed by Koop et al. (1996). Modify the model if necessary.

After an adequate model is identified it can be used either for descriptive purposes or for computing forecasts. Forecasting techniques for nonlinear models are studied extensively by Clements and Smith (1997) (see also Granger and Teräsvirta (1993) and Clements and Hendry (1998)).
Specification testing on the first stage, i.e. linearity testing and selecting the transition function, is partially considered in Heinen et al. (2011) by developing a linearity test for the new model. This test will be only briefly reviewed here for the sake of completeness. Starting with the model formulation in (3.2) with transition function in (3.5) Heinen et al. (2011) proceed in the spirit of Luukkonen et al. (1988) and approximate the nonlinearity by expanding $\mathcal{G}(\cdot;\gamma,c)$ as a Binomial series which they truncate at a suitable length k. This yields the following general auxiliary regression for a fixed $d \leq p$ and k:

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{p} \delta_j^{(0)} y_{t-j} + \sum_{j=1}^{p} \delta_j^{(1)} y_{t-j} y_{t-d} + \sum_{j=1}^{p} \delta_j^{(2)} y_{t-j} y_{t-d}^2 + \ldots + \sum_{j=1}^{p} \delta_j^{(2k)} y_{t-j} y_{t-d}^{2k} + u_t \,. \tag{3.6}$$

The null in this auxiliary model reads:

$$H_0 : \delta_j^{(\ell)} = 0 \quad \text{vs.} \quad H_1 : \text{at least one } \delta_j^{(\ell)} \neq 0; \quad j = 1,\ldots,p \text{ and } \ell = 0,\ldots,2k \,.$$

This null can be tested by using a standard F-test for a subvector of parameters.

3.3 Choosing the transition function

Selecting an adequate transition function is crucial to capture the in-sample dynamic of the data generating process properly. A first approach has been made by Teräsvirta and Anderson (1992) and Teräsvirta (1994). These methods concentrate on discriminating between the two most popular STAR models, namely ESTAR and LSTAR. Their proposed technique exploits the properties of the Taylor series expansion of the two nonlinear alternatives and is based on testing a sequence of nested hypotheses. Using such a procedure involves some problems: First the size is not under control anymore because of sequential testing. A second problem is that the classical approach only considers a first order expansion. As Escribano and Jordá (2001) point out this approximation is only adequate if the location parameter c is restricted to zero a priori. To circumvent this and related problems they propose to always include cubic and fourth power terms in the auxiliary regression. This improves the discriminatory power of the test but as this procedure is still based on sequential testing the size is not under control in neither of these approaches.

A rather different approach is proposed by Chen (2003). Based on an encompassing principle for non-nested models from Chen and Kuan (2002, 2007) he directly test whether an ESTAR or LSTAR formulation is more adequate. In simulations he shows that his approach is more powerful in detecting the correct specification. It is however important to note that the test of Chen (2003) is erroneous due to a falsely estimated covariance matrix. Nevertheless, the principle approach could be applied by using the results in Chen and Kuan (2007).

The TSTAR model was proposed as an alternative to the ESTAR model and therefore we concentrate on distinguishing between TSTAR and LSTAR model formulations.

Our approach is related to Chen (2003) in the sense that we also aim to directly test whether TSTAR or LSTAR is more appropriate. The approach we take is a more straightforward encompassing approach to conditional mean testing for non-nested hypotheses as proposed by Wooldridge (1990a). The idea is that the residuals of the model under the null estimated by nonlinear least squares should be orthogonal to the gradient of the null model and the residuals should also be independent of the gradient of the model under the alternative. More formally, let

$$\{m_t(w_t, \alpha) : \alpha \in A, t = 1, 2, \ldots, T\}, \quad A \subset \mathbb{R}^P \tag{3.7}$$

be the model under the null. In our case this is the TSTAR model in (3.2) together with the transition function in (3.5). The model under the alternative reads

$$\{\mu_t(w_t, \beta) : \beta \in B, t = 1, 2, \ldots, T\}, \quad B \subset \mathbb{R}^Q . \tag{3.8}$$

In our case this is an LSTAR model defined as in (3.2) with the transition function

$$\mathcal{G}(\cdot; \gamma, c) = \{1 + \exp(-\gamma(y_{t-d} - c))\}^{-1} . \tag{3.9}$$

Define the residuals under the null as

$$\hat{e}_t = y_t - \hat{m}_t(w_t, \hat{\alpha}) .$$

The test considers to test for $\delta = 0$ in

$$y_t = m_t(w_t, \alpha) + \delta \nabla_\beta \mu(\hat{\beta}) + error_t, \tag{3.10}$$

where $\nabla_\beta \mu(\hat{\beta}) = \nabla_\beta \mu(w_t, \hat{\beta})$ is the $Q \times 1$ gradient of $\mu_t(w_t, \beta)$ evaluated at the nonlinear least squares estimate $\hat{\beta}$ of β. Denote by $\nabla_\alpha m(\hat{\alpha})$ the same $P \times 1$ quantity for the null model $m_t(w_t, \alpha)$. The test can be carried out by computing

$$LM = TR^2, \tag{3.11}$$

where T is the sample size and R^2 is the coefficient of determination from the regression of $\hat{e}_t$ on $\nabla_\beta \mu(\hat{\beta})$ and $\nabla_\alpha m(\hat{\alpha})$. Asymptotically this test statistic follows a $\chi^2(Q)$ distribution.
If no homoscedasticity is assumed the test can be robustified by using results from Wooldridge (1990b) and run through the steps described in *Procedure 3.1* in Wooldridge (1990a, p. 336).

3.4 Residual based misspecification tests

3.4.1 Test of serial independence

In deriving the evaluation tests for the TSTAR model we roughly follow the steps taken in Eitrheim and Teräsvirta (1996). However, there are some crucial differences between their tests and the ones developed during the course of this section.
Consider the following general TSTAR model of order p

$$y_t = [\Psi w_t] + [\Phi w_t] \times \left[1 - \left\{ 1 + (y_{t-d} - c)^2 \right\}^{-\gamma} \right] + u_t , \tag{3.12}$$

with $\Psi = (\psi_0, \psi_1, \ldots, \psi_p)$, $\Phi = (\varphi_0, \varphi_1, \ldots, \varphi_p)$ and $w_t = (1, y_{t-1}, \ldots, y_{t-p})'$. The innovation process u_t is autocorrelated and follows an invertible moving-average process of order q (MA(q))[1]

$$u_t = [1 + \alpha(L)]\varepsilon_t; \quad \varepsilon_t \overset{iid}{\sim} (0, \sigma^2) , \tag{3.13}$$

where L denotes the lag operator and $\alpha(L) = (\alpha_1 L + \alpha_2 L^2 + \ldots + \alpha_q L^q)$.
Under the null of no serial correlation the innovations u_t in (3.12) are *iid* and under the alternative they follow (3.13). The testable pair of hypotheses thus reads

$$H_0 : \alpha_i = 0 \quad \text{vs.} \quad H_1 : \text{at least one } \alpha_i \neq 0; \quad i = 1, \ldots, q .$$

[1] This is not restrictive as this is also the test against innovations following an AR(q) process since they are asymptotically local equivalent alternatives of each other (see Godfrey (1988, p. 114)).

Based on results given in Godfrey (1988), we can proceed to develop a Lagrange multiplier (LM) test against serially correlated innovations.
To keep the notation simple define the so-called skeleton (see Franses and van Dijk (2000))

$$f(w_t;\Gamma) := [\Psi w_t] + [\Phi w_t] \times \left[1 - \left\{1 + (y_{t-d} - c)^2\right\}^{-\gamma}\right], \tag{3.14}$$

with $\Gamma = (\Psi, \Phi, \gamma, c)$ and the model becomes

$$y_t = f(w_t;\Gamma) + u_t .$$

Further define

$$\hat{u}_t := y_t - f(w_t;\hat{\Gamma}),$$

where $\hat{\Gamma} = (\hat{\Psi}, \hat{\Phi}, \hat{\gamma}, \hat{c})$ is the minimizer of

$$Q = \sum_{t=1}^{T} [y_t - f(w_t;\Gamma)]^2 .$$

The test statistic for the null of no serial correlation can now be obtained as

$$LM_{(1)} = TR^2 , \tag{3.15}$$

where T is the sample size and R^2 is the coefficient of determination from the regression of $\hat{u}_t$ on lagged residuals $\hat{u}_{t-1}, \ldots, \hat{u}_{t-q}$ and $\hat{z}_t = \frac{\partial f(w_t;\Gamma)}{\partial \Gamma}\Big|_{\hat{\Gamma}}$. This '$TR^2$' variant of the LM test has been proposed for detecting misspecifications by Breusch (1978) and Godfrey (1978) (see also Godfrey (1988)).
Eitrheim and Teräsvirta (1996) provide the components needed for computing (3.15) for the ESTAR and LSTAR case. The respective components for the TSTAR case are as follows:

$$\frac{\partial f(w_t;\Gamma)}{\partial \Psi} = w_t \tag{3.16}$$

$$\frac{\partial f(w_t;\Gamma)}{\partial \Phi} = w_t \times \left[1 - \left\{1 + (y_{t-d} - c)^2\right\}^{-\gamma}\right] \tag{3.17}$$

$$\frac{\partial f(w_t;\Gamma)}{\partial \gamma} = [\Phi w_t] \times \left\{\left[1 + (y_{t-d} - c)^2\right]^{-\gamma} \ln\left(1 + (y_{t-d} - c)^2\right)\right\} \tag{3.18}$$

$$\frac{\partial f(w_t;\Gamma)}{\partial c} = [\Phi w_t] \times \left\{\frac{\left[1 + (y_{t-d} - c)^2\right]^{-\gamma} \gamma(-2\gamma + 2c)}{1 + (y_{t-d} - c)^2}\right\} . \tag{3.19}$$

Eitrheim and Teräsvirta (1996) point out that such a test can suffer from size distortions in finite samples because the estimation procedure in the first step may result in a solution in which the residuals are not perfectly orthogonal to the gradient $\hat{z}_t$. We adopt their proposed remedy for this situation and take an extra orthogonalization step after estimating the model in (3.12) and obtaining the residuals. The test can now be performed in three stages:

(i) Estimate the TSTAR model using either nonlinear least squares or conditional (quasi) maximum likelihood under the null and obtain the residuals $\hat{u}_t$.

(ii) Regress $\hat{u}_t$ on the gradient $\hat{z}_t$ and obtain the residuals $\breve{u}_t$.

(iii) Regress $\breve{u}_t$ on $\breve{u}_{t-1},\ldots,\breve{u}_{t-q}$ and the partial derivatives of $f(w_t;\Gamma)$ evaluated at $\hat{\Gamma}$ as detailed in (3.16) - (3.19) and compute the respective R^2.

Under the null of no serial correlation the test statistic follows

$$LM_{(1)} \sim \chi^2(q)\,.$$

3.4.2 Test of no remaining nonlinearity

Since there a numerous ways in which a nonlinear model can be misspecified we restrict ourselves to the case of additive nonlinearity. Therefore consider the model

$$y_t = [\Psi w_t] + [\Phi w_t] \times \mathcal{G}_1(y_{t-d};\gamma_1,c_1) + [\Xi w_t] \times \mathcal{G}_2(y_{t-e};\gamma_2,c_2) + \varepsilon_t\,, \tag{3.20}$$

where $\mathcal{G}_1(\cdot)$ and $\mathcal{G}_2(\cdot)$ are transition functions of the form in (3.5). As the null model we consider (3.20) but without the second nonlinear component. The respective null can be formulated as

$$H_0 : \gamma_2 = 0 \quad \text{vs.} \quad H_1 : \gamma_2 > 0\,. \tag{3.21}$$

This situation can also be interpreted as testing a two regime TSTAR model against a three regime TSTAR model. In this interpretation the test is readily extendable to more general models containing more than two or three regimes.[2]

If we want to test the pair of hypotheses in (3.21) we face a similar problem as Luukkonen et al. (1988) and Heinen et al. (2011) when constructing linearity tests, i.e. that under the null the model in (3.20) is not fully identified. We circumvent this problem similarly and approximate the second nonlinearity by using an adequate linear series expansion around $\gamma_2 = 0$. As a Taylorian expansion is impractical here we follow Heinen et al. (2011) and use a Binomial series expansion which we truncate after $k = 3$ summands.

The linear approximation to $\mathcal{G}_2(\cdot)$ reads

$$\mathcal{G}_2^{(3)} = \gamma_2(y_{t-e}-c_2)^2 - \frac{1}{2}\gamma_2(\gamma_2+1)(y_{t-e}-c_2)^4 + \frac{1}{6}\gamma_2(\gamma_2+1)(\gamma_2+2)(y_{t-e}-c_2)^6\,. \tag{3.22}$$

After substituting the transition function and combining terms we obtain the auxiliary model

$$\begin{aligned} y_t \;=\; & [\Psi w_t] + [\Phi w_t] \times \mathcal{G}_1(y_{t-d};\gamma_1,c_1) + \\ & \delta_0 w_t + \delta_1 w_t y_{t-e} + \delta_2 w_t y_{t-e}^2 + \delta_3 w_t y_{t-e}^3 + \delta_4 w_t y_{t-e}^4 + \delta_5 w_t y_{t-e}^5 + \delta_6 w_t y_{t-e}^6 + r_t\,, \end{aligned} \tag{3.23}$$

[2] To test against a very general form of remaining nonlinearity in the residuals, artificial neural network tests as studied by Lee et al. (1993) or nonparametric bootstrap tests as in Lee and Ullah (2001) could be used. See also Teräsvirta et al. (2008).

where δ_i, $i = 0,\ldots,6$, are functions of the parameters Ξ, γ_2 and c_2 given in the appendix 3.9.
This reformulation solves the identification problem as the parameters γ_2, c_2 and Ξ are now multiplicatively connected. The innovation term is now denoted by r_t as it not only contains ε_t but also the approximation error from truncating the infinite Binomial series. Notice that under H_0, $r_t = \varepsilon_t$.
The pair of hypotheses for the auxiliary model reads

$$H_0 : \delta_i = 0 \quad \text{vs.} \quad H_1 : \text{at least one } \delta_i \neq 0 \,; i = 0,\ldots,6\,.$$

This most general case simplifies if the location parameter c_2 is set to zero a priori which is frequently done in empirical applications. Then only the odd powers remain in the auxiliary model.
The test can be carried out using the test statistic in (3.15). The corresponding R^2 is obtained from regressing the residuals obtained under the null, i.e. model (3.20) without the second nonlinearity, on the partial derivatives of the regression function evaluated under the null, i.e. $\hat{z}_t$ given in (3.16) - (3.19) and the auxiliary regressors w_t and $w_t y_{t-e}^i$, $i = 1,\ldots,6$. After the estimation of the null model the additional orthogonalization step (ii) as for the test against serially correlated innovations can be performed to avoid numerical problems as described at the end of section 3.4.1.
The resulting test statistic follows

$$LM_{(2)} \sim \chi^2\,(7(p+1))\ .$$

Note that in the model formulation (3.20) the delay parameter of the second nonlinear component is assumed to be e with $e \neq d$ but $e \leq p$. Similar to determine the delay d described in step 1 of the modeling cycle the test can be carried out for various values of e and the test yielding the minimal p-value is chosen as the decisive test decision.

3.4.3 Test of parameter constancy

Testing for the constancy of estimated parameters a well established way of checking the adequacy of linear models (see e.g. Chow (1960), Quandt (1960) or Andrews (1993)). In the context of nonlinear time series this maintains its importance but the assumption of an abrupt break in the parameters is questionable. Therefore, we propose a parametric test of the null of parameter constancy against the alternative that the autoregressive parameters change smoothly over time. Assuming the parameters of the transition function fixed the model under the alternative reads

$$y_t = [\Psi(t)\, w_t] + [\Phi(t)\, w_t] \times \mathcal{G}(y_{t-d}; \gamma, c) + \varepsilon_t\,, \tag{3.24}$$

with $\varepsilon_t \overset{iid}{\sim} N(0,1)$. The parameter vectors $\Psi(t)$ and $\Phi(t)$ are now functions of time and can be represented as

$$\Psi(t) = \overline{\Psi} + \lambda_1 \mathcal{K}(t;\gamma_1,c_1) \tag{3.25}$$

$$\Phi(t) = \overline{\Phi} + \lambda_2 \mathcal{K}(t;\gamma_1,c_1)\,, \tag{3.26}$$

where λ_i, $i=1,2$, are vectors conformable to the dimension of Ψ and Φ and $\mathcal{K}(\cdot)$ has the functional form

$$\mathcal{K}(t;\gamma_1,c_1) = 1-\left\{1+(t-c_1)^2\right\}^{-\gamma_1};\ \gamma_1>0\,. \tag{3.27}$$

This function induces a nonmonotonic change which is symmetric around $t=c_1$. If $\mathcal{K}(\cdot)$ takes on the limiting case $\gamma_1\to\infty$ then $\mathcal{K}(\cdot)\to 1-\mathbb{1}_{c_1}$ which corresponds to a single abrupt break only at $t=c_1$. $\mathbb{1}_{c_1}$ denotes the indicator function at c_1. The null of parameter constancy against the alternative of smoothly changing parameters over time can now be expressed as

$$H_0:\gamma_1=0 \quad \text{vs.} \quad H_1:\gamma_1>0\,.$$

Again we face an identification problem under the null as γ_1 is not identified. We expand $\mathcal{K}(\cdot)$ as Binomial series and truncate after $k=3$ summands. This yields

$$\mathcal{K}^{(3)} = \gamma_1(t-c_1)^2 - \frac{1}{2}\gamma_1(\gamma_1+1)(t-c_1)^4 + \frac{1}{6}\gamma_1(\gamma_1+1)(\gamma_1+2)(t-c_1)^6\,. \tag{3.28}$$

Upon substitution of the approximation in (3.28) into the model in (3.24) we obtain after combining terms the auxiliary regression

$$\begin{aligned} y_t \;=\; & \left[\overline{\Psi}w_t+\delta_0 w_t+\delta_1 t w_t+\delta_2 t^2 w_t+\delta_3 t^3 w_t+\delta_4 t^4 w_t+\delta_5 t^5 w_t+\delta_6 t^6 w_t\right]+ \\ & \left[\overline{\Phi}w_t+\beta_0 w_t+\beta_1 t w_t+\beta_2 t^2 w_t+\beta_3 t^3 w_t+\beta_4 t^4 w_t+\beta_5 t^5 w_t+\beta_6 t^6 w_t\right]\times \mathcal{G}(y_{t-d};\gamma,c)+r_t\,, \end{aligned} \tag{3.29}$$

where under the null $r_t=\varepsilon_t$. The coefficients δ_i, $i=0,\ldots,6,$ and β_i, $i=0,\ldots,6,$ are functions of γ_1, c_1 and λ_i, $i=1,2$ and given in the appendix 3.9.
The pair of hypotheses for the auxiliary model reads

$$H_0:\delta_i=\beta_i=0 \quad \text{vs.} \quad H_1:\text{at least one } \delta_i \text{ or } \beta_i\neq 0\,;i=0,\ldots,6\,.$$

The test against smoothly changing parameters can now be computed using (3.15) where the R^2 is obtained from the regression of the residuals under the null on the gradient $\hat{z}_t$ and the auxiliary regressors. The additional orthogonalization step (ii) is again recommended. The additional regressors in (3.29) are trending but using Theorem 1 in Lin and Teräsvirta (1994) the OLS estimates are still normally distributed and the usual asymptotic holds.
The test statistic follows

$$LM_{(3)} \sim \chi^2\left(14(1+p)\right)\,.$$

As with the test against remaining nonlinearity the auxiliary regression simplifies when $c_1=0$ is assumed a priori. Then only the even powers of the trend remain.

3.5 Generalized impulse response function

Impulse response functions (IRF) are a well established way to analyze the effect of a shock on the behavior of a time series model. Traditional impulse response analysis therefore considers the question: 'What is the effect of a shock of size δ hitting the system at time t on the state of the system at time $t+n$, if no other shock hits the system in the meantime?'. Denote with δ the size of a shock hitting the system at time t and with ω_{t-1} a particular realization of the information set Ω_{t-1} then we can define the impulse response function more formally as

$$\begin{aligned} IRF(n,\delta,\omega_{t-1}) &= E[y_{t+n}|\varepsilon_t=\delta,\varepsilon_{t+1}=\ldots=\varepsilon_{t+n}=0,\omega_{t-1}]- \\ &\quad E[y_{t+n}|\varepsilon_t=\varepsilon_{t+1}=\ldots=\varepsilon_{t+n}=0,\omega_{t-1}]\,. \end{aligned} \tag{3.30}$$

The second conditional expectation is often called the 'baseline' which acts as a reference point. For linear models Koop et al. (1996) point out three properties of the impulse response functions: *Symmetry*, i.e. a shock of -1 has exactly the opposite effect of a shock of +1, *shock linearity*, i.e. a shock of size 2 has exactly twice the effect as a shock of 1, and the IRF is *history independent*, i.e. the past does not effect the response in any way. The authors also provide various examples to show that these properties do not carry over to the nonlinear case. To remedy this drawbacks Koop et al. (1996) propose a generalized impulse response function (GIRF) which is itself a random variable and is defined as

$$GIRF(n,\delta,\omega_{t-1}) = E[y_{t+n}|\varepsilon_t=\delta,\omega_{t-1}]-E[y_{t+n}|\omega_{t-1}]\,. \tag{3.31}$$

Here, the conditional expectation is conditioned only on the shock δ and the past ω_{t-1}. The shocks occurring in the meantime are handled by averaging them out. For computing the GIRF obviously we need the conditional expectation of a nonlinear model which is cumbersome as the dimension of the integral defining the conditional expectation grows with n (see Granger and Teräsvirta (1993)). To ease implementation Koop et al. (1996) propose a numerical technique to compute the conditional expectation by means of Monte Carlo integration (for further details see Koop et al. (1996, p. 135)).

The history on which we condition the GIRF can also be only a subset of the entire history such as $\omega_{t-1}\in\mathcal{A}$. Where $\mathcal{A}$ could be the subset containing only the observations coming from one regime. Such an approach is useful in determining whether the dynamic behavior is different in periods of recession compared with expansionary periods (see e.g. van Dijk et al. (2002a) and Kapetanios (2003)).

The GIRF can also be used to analyze whether the model under consideration produces asymmetric effects over time. This could be done as in Potter (1995) by defining

$$ASYM(n,\delta,\omega_{t-1}) = GIRF(n,\delta,\omega_{t-1})+GIRF(n,-\delta,\omega_{t-1})\,. \tag{3.32}$$

Another use of the GIRF is to examine the persistence of shocks (see Koop et al. (1996)). If a time series model is stationary, at least globally, then the effect of a shock should eventually fade

away to zero if the horizon n goes to infinity. As a consequence the density of the GIRF defined by (3.31) should collapse to a single spike at zero. Therefore, the dispersion of the densities of the GIRF at different horizons n can be used as a pragmatic measure of the persistence of shocks.

3.6 Finite sample properties

To study the behavior of the tests in finite samples we conduct a small scale simulation study. We report size results from simulating the following TSTAR process

$$y_t = 0.7y_{t-1} - 0.5y_{t-1}\left[1 - \left\{1 + y_{t-1}^2\right\}^{-1}\right] + \varepsilon_t\,, \tag{3.33}$$

where $\varepsilon_t \overset{iid}{\sim} N(0,1)$. The location parameter is set to $c = 0$ only to reduce computational burden in the estimation process and it does not effect the results reported here.
We study different sample sizes of $T = 300, 500, 1000$. For all time series generated we discard the first 500 observations in order to be independent of the initial values. The first step in each simulation is to compute the linearity test against TSTAR proposed by Heinen et al. (2011). If the null cannot be rejected at the 5% level of significance the series is discarded and a new one is simulated. If the null is rejected the size or power experiment is conducted. This is done until the number of replication $M = 50000$ is reached. Applying the linearity test in the first step is done to avoid the estimation of a series in which there is not much evidence of nonlinearity.
The results of the size experiment for the test against LSTAR models from section 3.3 are shown in Table 3.1.

α	$T = 300$	$T = 500$	$T = 1000$
1%	0.950	0.912	0.922
5%	4.822	4.844	4.750
10%	9.900	9.890	9.956

Table 3.1: Empirical size of the test against LSTAR [in %].

The size of the test procedure to choose between competing STAR formulations shows virtually no distortions in the considered sample sizes. This is especially notable as the encompassing test to discriminate between ESTAR and LSTAR proposed by Chen (2003) is generally undersized. The power of the test against LSTAR was simulated using the following LSTAR specification as alternative:

$$y_t = 0.7y_{t-1} - 0.5y_{t-1}\left[1 + \exp(-2y_{t-1})\right]^{-1} + \varepsilon_t\,.$$

The results are given in Table 3.2.

α	$T = 300$	$T = 500$	$T = 1000$
1%	21.050	39.078	67.472
5%	39.838	57.910	78.750
10%	50.675	66.920	83.512

Table 3.2: Empirical power of the test against LSTAR [in %].

The power of the testing procedure shows reasonable discriminatory power of the test. In particular it yields better results than the selection procedures of Teräsvirta and Anderson (1992) and Teräsvirta (1994), relying on the simulation results in Chen (2003). Compared to the test procedure for the ESTAR–LSTAR case of the latter author the power is comparable in most settings. In some cases the power of the test of Chen (2003) is clearly higher but given the serious size distortions[3] the power of his test is not readily interpretable.

When studying the empirical power of the test against serially correlated innovations described in section 3.4.1 we simulate from (3.33) but assume that the innovation process follows an AR(1) process $u_t = \rho u_{t-1} + \varepsilon_t$, with $\varepsilon_t \overset{iid}{\sim} N(0,1)$ and $\rho = 0.2, 0.4, 0.6$.

The results for the size and power experiment are summarized in Table 3.3 and Table 3.4 respectively.

	$T = 300$			$T = 500$			$T = 1000$		
α	$q = 1$	$q = 2$	$q = 5$	$q = 1$	$q = 2$	$q = 5$	$q = 1$	$q = 2$	$q = 5$
1%	1.066	1.044	0.990	0.982	1.072	0.910	1.074	0.982	1.014
5%	5.254	5.208	5.022	5.074	5.186	4.892	5.362	5.028	5.100
10%	10.302	10.506	10.104	10.234	10.438	10.032	10.164	10.240	10.100

Table 3.3: Empirical size of the test of no innovation correlation [in %].

The results in Table 3.3 show that the empirical size is always very close to its nominal level. Although some minor distortions are visible the overall result confirms a satisfactorily behavior of the test in finite samples.

[3] In some settings he obtains a size of only 0.5% at a nominal $\alpha = 5\%$ level.

$\rho = 0.2$	$T = 300$			$T = 500$			$T = 1000$		
α	$q = 1$	$q = 2$	$q = 5$	$q = 1$	$q = 2$	$q = 5$	$q = 1$	$q = 2$	$q = 5$
1%	4.030	7.210	4.112	18.004	12.424	7.038	38.910	29.094	17.606
5%	10.036	20.176	13.880	38.422	30.090	20.644	62.968	52.822	37.888
10%	14.276	30.718	23.064	50.936	42.136	31.470	74.304	65.226	50.902
$\rho = 0.4$	$T = 300$			$T = 500$			$T = 1000$		
α	$q = 1$	$q = 2$	$q = 5$	$q = 1$	$q = 2$	$q = 5$	$q = 1$	$q = 2$	$q = 5$
1%	43.138	32.416	19.384	68.482	57.274	40.060	95.498	91.406	81.562
5%	67.250	56.998	41.046	86.474	78.716	64.146	98.962	97.478	93.382
10%	77.660	68.994	54.452	92.132	86.608	75.366	99.568	98.828	96.610
$\rho = 0.6$	$T = 300$			$T = 500$			$T = 1000$		
α	$q = 1$	$q = 2$	$q = 5$	$q = 1$	$q = 2$	$q = 5$	$q = 1$	$q = 2$	$q = 5$
1%	75.702	65.084	46.986	94.442	89.720	77.784	99.950	99.814	99.100
5%	90.392	84.140	70.484	98.626	97.014	91.434	99.994	99.986	99.846
10%	94.778	90.658	80.636	99.398	98.576	95.508	99.996	99.992	99.946

Table 3.4: Empirical power of the test of no innovation correlation [in %].

The results for the empirical power displayed in Table 3.4 show a similar behavior to the test for the ESTAR case described in Eitrheim and Teräsvirta (1996). The power slightly decreases if the tested order of autocorrelation q increases. This might be expected as $q = 1$ is the true data generating process. Another factor that influences the power is the degree of autocorrelation ρ. In finite samples and a low degree of serial correlation the power is quite low but increases steeply if the sample size and/or the ρ becomes larger. Such a behavior is somewhat expected as the χ^2 distribution holds only asymptotically and if ρ increases the serial correlation becomes easier to detect. In general the test yields good results for most situations encountered in practice and helps to reveal severe misspecifications.

For the assessment of the test of no remaining nonlinearity from section 3.4.2 we simulate data from (3.33) to perform the size experiment. The results for the size and power experiments are displayed in Tables 3.5 and 3.6 respectively.

α	$T = 300$	$T = 500$	$T = 1000$
1%	0.806	0.834	0.852
5%	4.552	4.728	4.640
10%	9.540	9.600	9.566

Table 3.5: Empirical size of the test of no remaining nonlinearity [in %].

The results of the size experiment show that the empirical size is close to its nominal level. If anything, the test is slightly conservative.

For the power experiment we simulate data from

$$y_t = 0.3y_{t-1} - 0.1y_{t-1}\left[1 - \left\{1 + y_{t-1}^2\right\}^{-1}\right] + 0.75y_{t-1}\left[1 - \left\{1 + y_{t-1}^2\right\}^{-3.5}\right] + \varepsilon_t\,,$$

where $\varepsilon_t \overset{iid}{\sim} N(0,1)$. Additionally to this data generating process we simulate from an LSTAR(1) and ESTAR(1) process and induce remaining nonlinearity by fitting the wrong model, namely an TSTAR(1). The respective processes read

$$\begin{aligned} y_t &= 0.7y_{t-1} - 0.5y_{t-1}\left[1 + \exp(-2y_{t-1})\right]^{-1} + \varepsilon_t \\ y_t &= 0.7y_{t-1} - 0.5y_{t-1}\left[1 - \exp(-2y_{t-1}^2)\right] + \varepsilon_t\,, \end{aligned}$$

with $\varepsilon_t \overset{iid}{\sim} N(0,1)$. The results for these simulations are presented in Table 3.6.
Another variant of the power simulation inspired by Eitrheim and Teräsvirta (1996) is also explored: The data is generated by (3.33) but misspecification is assumed by fitting a STAR model of the wrong kind to the data, namely a LSTAR(1). These results are presented in Table 3.7.

H_1 : TSTAR			
α	$T = 300$	$T = 500$	$T = 1000$
1%	3.120	2.986	3.772
5%	12.224	11.424	12.184
10%	21.474	20.076	20.676
H_1 : LSTAR			
α	$T = 300$	$T = 500$	$T = 1000$
1%	4.430	5.522	8.274
5%	13.924	16.898	27.044
10%	24.070	28.614	43.822
H_1 : ESTAR			
α	$T = 300$	$T = 500$	$T = 1000$
1%	1.348	1.292	1.276
5%	6.848	6.480	6.210
10%	13.440	12.716	11.990

Table 3.6: Empirical power of the test of no remaining nonlinearity [in %].

For this test the empirical power results reveal a comparable performance for this test as for the test in the ESTAR case reported in Eitrheim and Teräsvirta (1996). The test appears to have reasonably good power against LSTAR models especially in larger samples where the power increases quite steeply.
Interestingly the test appears to have some nontrivial power also against ESTAR at least in small samples. This is surprising given that TSTAR has been designed to resemble the desirable properties of ESTAR. This power vanishes as T increases underlining that the TSTAR model can

very well serve as an alternative to ESTAR as they can hardly be distinguished for reasonable sample sizes.

α	$T = 300$	$T = 500$	$T = 1000$
1%	31.512	34.200	41.434
5%	49.670	52.924	60.868
10%	61.682	64.914	71.762

Remark: An LSTAR(1) model was fitted to data generated from (3.33).

Table 3.7: Empirical power of the test of no remaining nonlinearity [in %].

Analyzing the power experiment set up as in Eitrheim and Teräsvirta (1996) we obtain rather good results for the test against no remaining nonlinearity even in finite samples. In particular we obtain higher power as in the ESTAR case.
Turning to the results for the test of parameter constancy we report size and power results in Tables 3.8 and 3.9 respectively.

α	$T = 300$	$T = 500$	$T = 1000$
1%	0.778	0.850	0.940
5%	4.490	4.504	4.784
10%	9.278	9.354	9.522

Table 3.8: Empirical size of the test of parameter constancy [in %].

The test shows only minor size distortions in finite samples and approaches its nominal level as the sample size increases. Overall the test seems to be conservative, if anything.
For the power simulations we generate data from (3.24) of order one and set $\psi_0(t) = 2\mathcal{K}(\cdot)$, $\psi_1(t) = -0.2$, $\vartheta_0(t) = 0$ and $\vartheta_1(t) = (1.1 - 0.9\mathcal{K}(\cdot))$ where $\mathcal{K}(\cdot) = \mathcal{K}(t/T; 3, 0)$ as in (3.27).

α	$T = 300$	$T = 500$	$T = 1000$
1%	3.008	22.110	99.248
5%	11.458	44.252	99.746
10%	19.528	57.336	99.824

Table 3.9: Empirical power of the test of parameter constancy [in %].

The test shows reasonable power to detect parameter changes in finite samples. If the sample size increases the power of the test increases very steeply. Thus the test is a useful tool to detect parameter changes in most sample sizes.

3.7 Modeling real exchange rates

To demonstrate the application of the test developed in this paper we run through the whole modeling cycle described in section 3.2 to model real exchange rates.

We use the same data that has been analyzed by Taylor et al. (2001) and by Rapach and Wohar (2006). Namely, we analyze monthly real exchange data for Germany against the US from 1980:01 - 1994:12 ($T = 288$).[4] The series is depicted in Figure 3.1.

Figure 3.1: Monthly log real exchange rate for Germany.

Determining the lag length using the consistent BIC we obtain $p = 1$. The linearity test rejects the null of linearity on the $\alpha = 5\%$ level of significance. The test against LSTAR yields a test decision in favor of the null model, i.e. TSTAR.
As the data set has also been analyzed by Taylor et al. (2001) and by Rapach and Wohar (2006) we report also their estimates for an ESTAR model. The estimated model has been theoretically justified by the assumption that real exchange rates follow a nonlinear STAR model with one unit root regime and one stationary regime that pulls the real exchange rate back into its stable equilibrium once it wanders too far off. The model reads

$$y_t = y_{t-1} + \pi y_{t-1}\mathcal{G}(\cdot) + \varepsilon_t\,, \tag{3.34}$$

where $-2 < \pi < 0$ to ensure global stationarity of the model. The transition function $\mathcal{G}(\cdot)$ is either as in (3.3) for the ESTAR model or as in (3.5) for the TSTAR model. Additionally Taylor et al. (2001) and Rapach and Wohar (2006) set $\pi := -1$.
The estimation results are in table 3.10.

[4]The data set is available from David Rapach's website at: http://pages.slu.edu/faculty/rapachde/Nlfit.zip.

	ESTAR	TSTAR
$\hat{\pi}$	-1	-0.023
$\hat{\gamma}$	0.264	275.284
$\hat{\sigma}_\varepsilon$	0.035	0.032

Table 3.10: Estimation of STAR models.

Albeit the estimates for the TSTAR model might look puzzling at first Heinen et al. (2011) show that these estimates are much more reasonable than the corresponding ESTAR estimates as the ESTAR model actually degenerates to a random walk as opposed to the TSTAR model which maintains the regime switching behavior. Further support of the PPP can be seen in Figure 3.2.

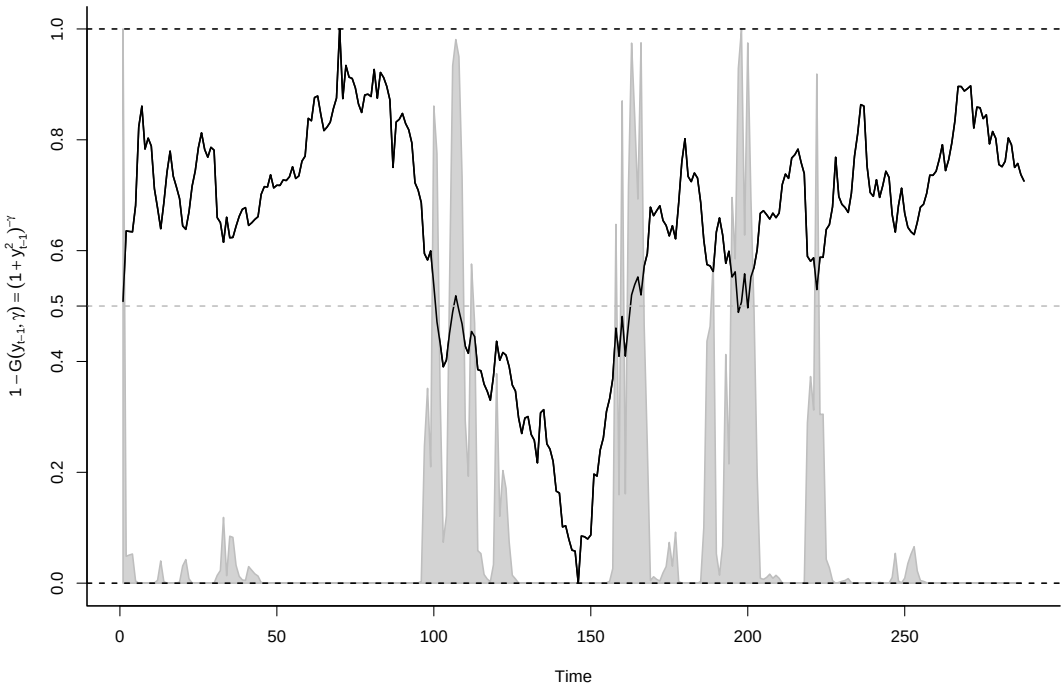

Figure 3.2: Monthly log real exchange rate for Germany and transition function.

The figure shows the time series, rescaled to be in the closed interval $[0, 1]$, and one minus the estimated transition function over time. The gray shaded area are the time periods in which the process behaves like a random walk. This is always the case when the process is close to its equilibrium near zero (note that the dotted line at 0.5 is the zero line of the unscaled series). Thus we have a stationary but nonlinear process most of the time which behaves like a random walk near the equilibrium as predicted by the PPP. Heinen et al. (2011) further show that the data is globally stationary although one unit root regime is present. Global stationarity is important for the asymptotic distributions of the misspecification tests to hold. Performing the misspecification tests yields the results in table 3.11.

	Test Statistic	Critical Value
Test against LSTAR	$LM = 5.779$	$\chi^2_{0.99;\nu=2} = 9.210$
Test of serial independence	$LM_{(1)} = 0.501$	$\chi^2_{0.99;\nu=1} = 6.635$
Test of no remaining nonlinearity	$LM_{(2)} = 8.935$	$\chi^2_{0.99;\nu=3} = 11.345$
Test of parameter constancy	$LM_{(3)} = 1.283$	$\chi^2_{0.99;\nu=6} = 16.812$

Table 3.11: Results of the misspecification tests.

The respective null hypotheses of the tests cannot be rejected at the $\alpha = 0.01$ level of significance hinting at a well specified model.

To gain further insights about the dynamic properties of the fitted model we estimate generalized impulse response functions. As the test against ARCH effects as described in Engle (1982) provides no evidence of conditional heteroscedasticity we randomly sample the innovations with replacement from the estimated model. The shocks we use are $\delta_t = \delta\hat{\sigma}_\varepsilon$, with $\delta = \pm 2, \pm 1$. We compute the GIRF for a horizon of $n = 150$ and estimate the conditional expectations in (3.31) as means over 5000 Monte Carlo repetitions. Figure 3.3 shows the estimated impulse response functions.

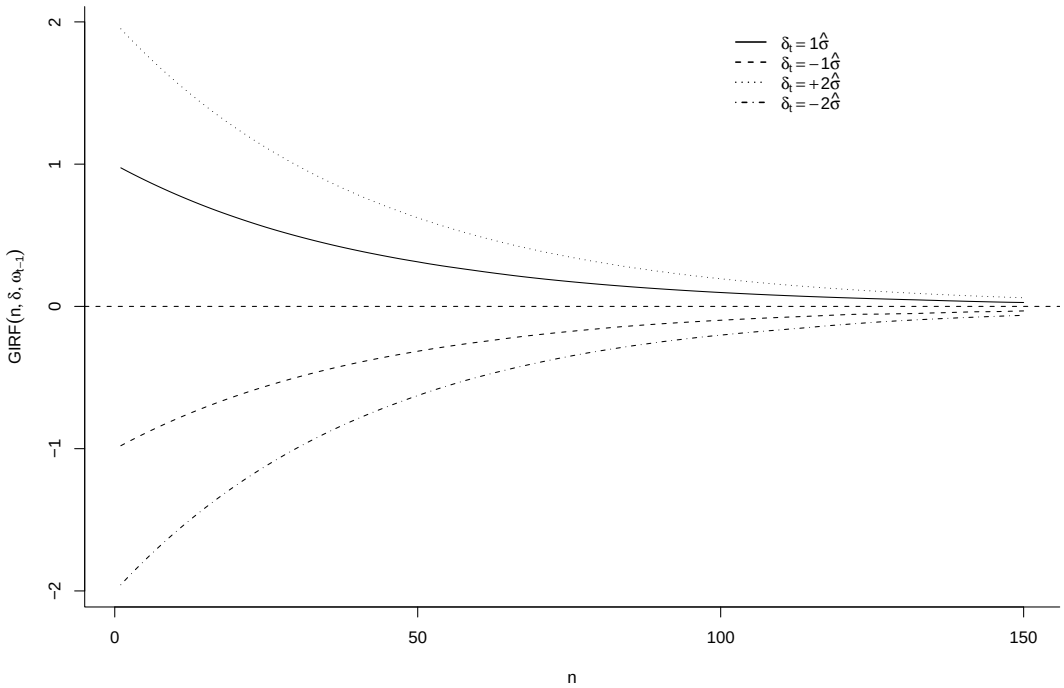

Figure 3.3: Generalized impulse response functions.

Obviously the shocks hitting the system are very persistent over time but eventually vanish. This supports the parameter estimates that show a highly persistent model. Additionally a high persistence of shocks in the model can also be induced by nonlinearities which in turn leads to such highly persistent impulse response functions (see e.g. van Dijk et al. (2002a) and Kuswanto and Sibbertsen (2008)). Another interesting aspect is whether the response to shocks

is asymmetric depending on the sign of the shock. A measure for asymmetry is defined in (3.32). Figure 3.4 shows the estimated quantities.

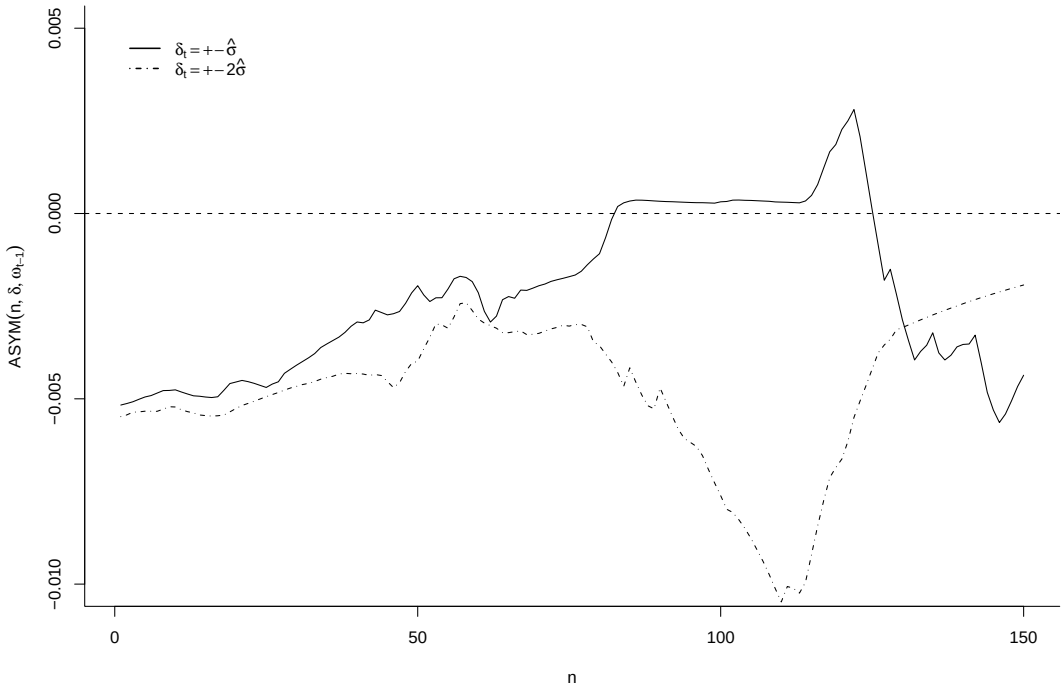

Figure 3.4: Measure of asymmetry.

For both shocks the behavior is asymmetric depending on the sign of the shock. If the shock is negative but relatively small the response to it is larger compared to a positive shock of the same size. This difference vanishes quite fast as the horizon increases. If however the shock is negative and relatively large ($\delta_t = |2\hat{\sigma}_\varepsilon|$) the response to a negative shock is heavier and decreases much slower. Indeed the asymmetry first increases before it decreases.

3.8 Conclusion

In this paper we extend the treatment of the newly developed nonlinear time series model named TSTAR developed by Heinen et al. (2011). We consider the modeling cycle for nonlinear time series models and contribute to the evaluation stage by proposing LM tests against serially correlated innovations, no remaining nonlinearity and parameter constancy. We also consider evaluation by generalized impulse response functions as proposed by Koop et al. (1996). In simulations we show that all the tests have reasonable power against their respective alternatives and are therefore an useful addition to the evaluation toolbox for nonlinear TSTAR models.
In an empirical application to real exchange data we put the evaluation techniques to the test and verify that a proposed TSTAR formulation adequately captures the nonlinear behavior of the data. Impulse response analysis is used to further evaluate the dynamic propagation behavior of the estimated model.

3.9 Appendix

3.9.1 Regression coefficients in (3.23)

Expanding the expressions containing c_2 as

$$\begin{aligned}
(y_{t-e}-c_2)^2 &= y_{t-e}^2 - 2y_{t-e}c_2 + c_2^2, \\
(y_{t-e}-c_2)^4 &= y_{t-e}^4 - 4y_{t-e}^3c_2 + 6y_{t-e}^2c_2^2 - 4y_{t-e}c_2^3 + c_2^4, \\
(y_{t-e}-c_2)^6 &= y_{t-e}^6 - 6y_{t-e}^5c_2 + 15y_{t-e}^4c_2^2 - 20y_{t-e}^3c_2^3 + 15y_{t-e}^2c_2^4 - 6y_{t-e}c_2^5 + c_2^6,
\end{aligned}$$

we obtain after some algebra

$$\begin{aligned}
\delta_0 &= \left[\left(\gamma_2 c_2 - \frac{1}{2}\gamma_2(\gamma_2+1) + \frac{1}{6}\gamma_2(\gamma_2+1)(\gamma_2+2)c_2^6\right)\Xi\right] \\
\delta_1 &= \left[\left(-2\gamma_2 c_2 + 2\gamma_2(\gamma_2+1)c_2^3 - \gamma_2(\gamma_2+1)(\gamma_2+2)c_2^5\right)\Xi\right] \\
\delta_2 &= \left[\left(-3\gamma_2(\gamma_2+1)c_2^2 + 2\frac{1}{2}\gamma_2(\gamma_2+1)(\gamma_2+2)c_2^4\right)\Xi\right] \\
\delta_3 &= \left[\left(2\gamma_2(\gamma_2+1)c_2 - 3\frac{1}{3}\gamma_2(\gamma_2+1)(\gamma_2+2)c_2^3\right)\Xi\right] \\
\delta_4 &= \left[\left(-\frac{1}{2}\gamma_2(\gamma_2+1) + 2\frac{1}{2}\gamma_2(\gamma_2+1)(\gamma_2+2)c_2^2\right)\Xi\right] \\
\delta_5 &= \left[(-\gamma_2(\gamma_2+1)(\gamma_2+2)c_2)\,\Xi\right] \\
\delta_6 &= \left[\left(\frac{1}{6}\gamma_2(\gamma_2+1)(\gamma_2+2)\right)\Xi\right].
\end{aligned}$$

3.9.2 Regression coefficients in (3.29)

Expanding the expressions containing c_1 as

$$\begin{aligned}
(t-c_1)^2 &= t^2 - 2tc_1 + c_1^2, \\
(t-c_1)^4 &= t^4 - 4t^3c_1 + 6t^2c_1^2 - 4tc_1^3 + c_1^4, \\
(t-c_1)^6 &= t^6 - 6t^5c_1 + 15t^4c_1^2 - 20t^3c_1^3 + 15t^2c_1^4 - 6tc_1^5 + c_1^6.
\end{aligned}$$

we obtain after some algebra for λ_1

$$\begin{aligned}
\delta_0 &= \left[\left(\gamma_1 c_1 - \frac{1}{2}\gamma_1(\gamma_1+1) + \frac{1}{6}\gamma_1(\gamma_1+1)(\gamma_1+2)c_1^6\right)\lambda_1\right] \\
\delta_1 &= \left[\left(-2\gamma_1 c_1 + 2\gamma_1(\gamma_1+1)c_1^3 - \gamma_1(\gamma_1+1)(\gamma_1+2)c_1^5\right)\lambda_1\right] \\
\delta_2 &= \left[\left(-3\gamma_1(\gamma_1+1)c_1^2 + 2\frac{1}{2}\gamma_1(\gamma_1+1)(\gamma_1+2)c_1^4\right)\lambda_1\right] \\
\delta_3 &= \left[\left(2\gamma_1(\gamma_1+1)c_1 - 3\frac{1}{3}\gamma_1(\gamma_1+1)(\gamma_1+2)c_1^3\right)\lambda_1\right] \\
\delta_4 &= \left[\left(-\frac{1}{2}\gamma_1(\gamma_1+1) + 2\frac{1}{2}\gamma_1(\gamma_1+1)(\gamma_1+2)c_1^2\right)\lambda_1\right] \\
\delta_5 &= \left[\left(-\gamma_1(\gamma_1+1)(\gamma_1+2)c_1\right)\lambda_1\right] \\
\delta_6 &= \left[\left(\frac{1}{6}\gamma_1(\gamma_1+1)(\gamma_1+2)\right)\lambda_1\right]
\end{aligned}$$

and for λ_2 respectively

$$\begin{aligned}
\beta_0 &= \left[\left(\gamma_1 c_1 - \frac{1}{2}\gamma_1(\gamma_1+1) + \frac{1}{6}\gamma_1(\gamma_1+1)(\gamma_1+2)c_1^6\right)\lambda_2\right] \\
\beta_1 &= \left[\left(-2\gamma_1 c_1 + 2\gamma_1(\gamma_1+1)c_1^3 - \gamma_1(\gamma_1+1)(\gamma_1+2)c_1^5\right)\lambda_2\right] \\
\beta_2 &= \left[\left(-3\gamma_1(\gamma_1+1)c_1^2 + 2\frac{1}{2}\gamma_1(\gamma_1+1)(\gamma_1+2)c_1^4\right)\lambda_2\right] \\
\beta_3 &= \left[\left(2\gamma_1(\gamma_1+1)c_1 - 3\frac{1}{3}\gamma_1(\gamma_1+1)(\gamma_1+2)c_1^3\right)\lambda_2\right] \\
\beta_4 &= \left[\left(-\frac{1}{2}\gamma_1(\gamma_1+1) + 2\frac{1}{2}\gamma_1(\gamma_1+1)(\gamma_1+2)c_1^2\right)\lambda_2\right] \\
\beta_5 &= \left[\left(-\gamma_1(\gamma_1+1)(\gamma_1+2)c_1\right)\lambda_2\right] \\
\beta_6 &= \left[\left(\frac{1}{6}\gamma_1(\gamma_1+1)(\gamma_1+2)\right)\lambda_2\right].
\end{aligned}$$

Chapter 4

The dynamics of real exchange rates – A reconsideration

The dynamics of real exchange rates – A reconsideration

Co-authored with Hendrik Kaufmann and Philipp Sibbertsen.
Published as Leibniz University of Hannover Discussion Paper No. 463.

4.1 Introduction

An ongoing debate about the behavior of real exchange rates suggests that Purchasing Power Parity (PPP) holds as a long-run concept (see e.g. Edison and Klovland (1987), MacDonald (1998) or Taylor et al. (2001)). Econometrically speaking PPP states that real exchange rates fluctuate finitely around an equilibrium, i.e. a time stable mean, and are thus weakly stationary. However, stationarity of real exchange rates does not say anything about the detailed dynamics driving them. Modeling real exchange rates by linear stationary models does not lead to convincing results as standard unit root tests can not reject the null of a random walk in the linear framework and thus can not confirm PPP (see e.g. Adler and Lehmann (1983), Meese and Rogoff (1983), Meese and Rogoff (1988) or Caporale et al. (2003)). In a theoretical framework Sercu et al. (1995) suggest among others that under transaction costs the adjustment process towards PPP is nonlinear (see also Sarno (2005) and the discussion in Taylor et al. (2001, p. 1018)). Empirical evidence that a nonlinear adjustment mechanism could solve the PPP puzzle is recently provided by Lo (2008) and Norman (2010). Therefore, recently nonlinear models came into the focus of economists.

The two prevailing approaches to model the dynamic of nonlinear adjustment towards PPP are exponential smooth transition autoregressive (ESTAR) models (see e.g. Michael et al. (1997), Taylor et al. (2001) and Kilian and Taylor (2003)) and Markov switching autoregressive (MSAR) models (see e.g. Bergman and Hansson (2005) and Kanas (2006)).[1] Both approaches imply under certain conditions that PPP may hold as a long-run concept. However, the nonlinear dynamics driving the respective processes are very distinct. The dynamic of an ESTAR process is driven by lagged values of the endogenous, and therefore observable, variable while an unobservable Markov process governs the dynamic in the MSAR case. This implies different economic intuition. The ESTAR approach is based on the idea that international trade only starts if the price differences between countries exceed a certain level which is determined by the costs of trading such as transportation costs, taxes and many others. As long as the price differences are smaller than this level no trading takes place and therefore real exchange rates fluctuate freely and behave like a random walk. As soon as the price differences exceed this level international trading starts and real exchange rates are pulled back to a long-run equilibrium. This economic

[1] An exception is Lahtinen (2006) who uses a symmetric, second order logistic STAR model. However, second order logistic STAR models are designed to closely resemble ESTAR models and the estimation results imply a dynamic very similar to a switching regression.

view of real exchange rates is expressed in the ESTAR model by a unit root regime close to the equilibrium and a stabilizing regime anywhere else. This approach has been applied to real exchange rates by Taylor et al. (2001), Kapetanios et al. (2003) and Rapach and Wohar (2006) amongst others.

In contrast, the MSAR view supports the overshooting or bubble theory by Dornbusch (1976) on real exchange rates. It emphasizes the point that real exchange rates are not just driven by a long-run behavior but that there are dominating short-run forces which are stabilized by some long-run adjustment. These short-run forces can be explained by short-run speculation or other exogenous shocks such as political regime changes, wars or economic crises. Taking the short-run speculation on real exchange rates as an example, these could be strong enough to create bubbles and therefore, in econometric terms, can create exploding regimes in MSAR models. A second stabilizing regime in the MSAR model determines the overall stationarity of the model and therefore the long-run convergence towards PPP. MSAR models are able to capture such occasional, sudden and large exchange rate changes caused by unobservable speculation (see e.g. van Norden (1996)). Regarding the MSAR approach there are two possibilities of modeling this economic behavior. One possibility is to implement Markov switching within the autoregressive parameter and thus having a model with two autoregressive regimes, one of them being explosive and the other stabilizing. This model was applied among others by Kanas (2006). Stationarity conditions for this model are given in Francq and Zakoïan (2001). A second option is to implement Markov switching in the mean rather than in the autoregressive parameter and thus to obtain a switching mean model. This approach was motivated by Engel and Hamilton (1990) and Hamilton (1993) who argue that switches in the mean are more appropriate for modeling the dynamics of financial data. Bergman and Hansson (2005) apply this approach to real exchange rates.

The application of either approach has so far been motivated by an a priori economic belief about the behavior of real exchange rates rather than a formal statistical model selection procedure.

In this paper we suggest for the first time a formal model selection procedure based on a parametric bootstrap to discriminate between ESTAR and MSAR models. We study the finite sample behavior of the proposed method and analyze several major real exchange rates to shed some light on the question which model and by that which theory describes the dynamic behavior of real exchange rates best.

The rest of this paper is organized as follows: In section 4.2 we describe the two competing models for real exchange rates. In section 4.3 we briefly discuss some specifics when testing non-nested models. In section 4.3.1 we describe the bootstrap procedure in order to distinguish the competing models in more detail and investigate the finite sample properties of this test in section 4.3.2. Section 4.4 applies the model selection procedure to empirical data before section 4.5 concludes.

4.2 Two Competing models for real exchange rates

The general ESTAR model is given by two autoregressive regimes connected by a smooth exponential transition function $\mathcal{G}(\cdot;\gamma,c):\mathbb{R}\to[0,1]$. This function governs the transition between the two regimes in a smooth way. Alternatively, an ESTAR model can also be interpreted as a continuum of regimes which is passed through by the process.
In general, univariate ESTAR(p) models, $p\geq 1$ and $d\leq p$, are given by

$$\begin{aligned} y_t &= [\Psi w_t]\times[1-\mathcal{G}(y_{t-d};\gamma,c)]+[\Theta w_t]\times\mathcal{G}(y_{t-d};\gamma,c)+\varepsilon_t \qquad (4.1)\\ &= [\Psi w_t]+[\Phi w_t]\times\mathcal{G}(y_{t-d};\gamma,c)+\varepsilon_t, \quad t\geq 1, \qquad (4.2)\end{aligned}$$

with $\varepsilon_t \overset{iid}{\sim}(0,\sigma^2)$.
The parameter vectors Ψ and Θ as well as w_t are given by $\Psi=(\psi_0,\psi_1,\ldots,\psi_p)$, $\Theta=(\vartheta_0,\vartheta_1,\ldots,\vartheta_p)$, and $w_t=(1,y_{t-1},\ldots,y_{t-p})'$. For the alternative parametrization in (4.2) we have $\Phi=(\varphi_0,\varphi_1,\ldots,\varphi_p)=(\psi_0-\vartheta_0,\psi_1-\vartheta_1,\ldots,\psi_p-\vartheta_p)$, i.e. the second regime realizes as sum of Ψ and Φ.
For an ESTAR model the transition function $\mathcal{G}(\cdot)$ is given by

$$\mathcal{G}(\cdot;\gamma,c)=1-\exp\left\{-\gamma(y_{t-d}-c)^2\right\};\quad \gamma>0\,. \qquad (4.3)$$

Due to the symmetry of $\mathcal{G}(\cdot)$ this functional form for the transition function is popular for modeling real exchange rates or real interest rates (see e.g. Taylor et al. (2001), Kapetanios et al. (2003) or Rapach and Wohar (2006)). Surveys of the broad field of nonlinear time series models in general and STAR models in particular are given by Potter (1999) and van Dijk et al. (2002b); see also Teräsvirta (1994). The most frequently used special case of the general ESTAR model in (4.2) is the ESTAR(1) model

$$y_t=\psi y_{t-1}+\phi y_{t-1}\left\{1-\exp\left(-\gamma(y_{t-1}-c)^2\right)\right\}+\varepsilon_t\,. \qquad (4.4)$$

Moreover, the additional restriction $c=0$ is also frequently imposed (e.g. Kapetanios et al. (2003)). To model the globally stationary but nonlinear behavior of real exchange rates one unit root regime is also commonly imposed (see e.g. Taylor et al. (2001) and Kapetanios et al. (2003)). This leads to the model

$$y_t=y_{t-1}+\phi y_{t-1}\left\{1-\exp\left(-\gamma(y_{t-1}-c)^2\right)\right\}+\varepsilon_t\,, \qquad (4.5)$$

with $-2<\phi<0$ to ensure global stationarity. Kapetanios et al. (2003) show that the model in (4.5) although locally non-stationary is globally stationary as long as $|\psi+\phi|<1$.
Estimation of these models either by nonlinear least squares or maximum likelihood techniques is treated by Klimko and Nelson (1978) and Tjøstheim (1986) respectively.
A standard MSAR model based on the work of Hamilton (1989) reads (see also Hamilton and Raj (2002) for an introduction)

$$y_t=\mu_{s_t}+\phi_{1,s_t}y_{t-1}+\ldots+\phi_{p,s_t}y_{t-p}+\varepsilon_t\,. \qquad (4.6)$$

The values of the autoregressive parameters $\phi_{1,s_t},\ldots,\phi_{p,s_t}$ and the mean μ_{s_t} and thus the regime switching is governed by an unobservable Markov chain

$$\mathbb{P}(s_t = j|s_{t-1} = i, s_{t-2} = k,\ldots,y_{t-1},y_{t-2},\ldots) = \mathbb{P}(s_t = j|s_{t-1} = i) = p_{ij}\,.$$

Extensions of this basic framework are possible, see e.g. Hamilton and Raj (2002) and the papers cited therein.

Although such a general model is possible to use in principle the models usually found in applied work are more restrictive and only have a few parameters that change depending on the regime. In the real exchange rate work in particular the most frequently used model is a first order, or at least low order, autoregression with a Markov switching mean (see e.g. Engel and Hamilton (1990), Bergman and Hansson (2005) or Kanas (2006)). This model is empirically justified by Hegwood and Papell (1998) and Montañés (1997) who find reversion to a mean which is subject to structural breaks.

Considering the special case of a first order MSAR model with a switching mean only we have two possible ways to describe such a process

$$(y_t - \mu_{s_t}) = \phi(y_{t-1} - \mu_{s_{t-1}}) + \varepsilon_t \tag{4.7}$$

$$(y_t - \mu_{s_t}) = \phi(y_{t-1} - \mu_{s_t}) + \varepsilon_t\,. \tag{4.8}$$

Although the difference between these two models seems negligible the dynamic patterns they produce are rather different (see Hamilton (1993)). Suppose that the process shifts from regime 1 to regime 2 at date t and stays there for j periods. Using the specification in (4.8) the value y_{t+j} will be $(\mu_2-\mu_1)(1+\phi+\phi^2+\ldots+\phi^j)$ units higher as a consequence of the regime switch. Thus, the consequences of a shift accumulate over time. If the specification (4.7) is used the value y_{t+j} will only be $(\mu_2-\mu_1)$ units higher. The formulation in (4.7) is the more promising one for economic and financial time series according to Hamilton (1993). Thus, we will consider such a model in our model selection approach.

Regarding stationarity conditions for MSAR models Francq and Zakoïan (2001) show that for the model in (4.7) to be globally stationary the following conditions must hold

$$(p_{11}+p_{22})\phi^2 + (1-p_{11}-p_{22})\phi^4 < 1 \tag{4.9}$$

$$(p_{11}+p_{22})\phi^2 < 2\,. \tag{4.10}$$

For the model in (4.7) to be locally stationary the usual condition $|\phi| < 1$ must hold.

Maximum likelihood estimation by direct, numerical maximization of the likelihood function of such models is treated in detail by Hamilton (1989). Alternatively an iterative EM algorithm is given by Hamilton (1990).

4.3 Testing non-nested hypotheses

Usually the comparison of different hypotheses, i.e. of competing models may be performed using standard approaches such as likelihood-ratio tests, Wald tests, Lagrange multiplier tests or the principle of Hausman (1978) (see Engle (1984) for a survey of these procedures). These principles assume that the null model can be obtained from the competing model by imposing some parameter restrictions. If this assumption is violated, i.e. none of the hypotheses is a special case of the other one, these approaches fail. For a survey of several aspects of testing non-nested hypotheses see Gourieroux and Monfort (1994).

Cox (1961, 1962) generalizes the likelihood-ratio principle to the case of non-nested models. He shows that the usual unadjusted likelihood-ratio statistic does not converge to zero if the two models are non-nested. Put more formally, let the two models for the conditional density be denoted by

$$H_f \quad : \quad \mathcal{F}_\theta = \{ f(y_t|\mathfrak{F}_{t-1};\theta),\, \theta \in \Theta \subseteq \mathbb{R}^r \},\, t = 1,\ldots,T \tag{4.11}$$

$$H_g \quad : \quad \mathcal{F}_\lambda = \{ g(y_t|\mathfrak{F}_{t-1};\lambda),\, \lambda \in \Lambda \subseteq \mathbb{R}^q \},\, t = 1,\ldots,T\,. \tag{4.12}$$

The process is denoted by y_t and $\mathfrak{F}_{t-1}$ is the sigma algebra generated by $(y_{t-1}, y_{t-2}, \ldots)$. For the sake of brevity we will suppress the dependence on y_t and $\mathfrak{F}_{t-1}$ and simply write $f(\theta)$ and $g(\lambda)$ whenever possible. The likelihood-ratio statistic of Cox (1961, 1962) reads

$$T_f(\hat{\theta},\hat{\lambda}) = \left[\log f(\hat{\theta}) - \log g(\hat{\lambda})\right] - E_f\left[\log f(\hat{\theta}) - \log g(\hat{\lambda})\right], \tag{4.13}$$

where $\hat{\theta}$ and $\hat{\lambda}$ denote the maximum likelihood estimates. $E_f[\cdot]$ is the expectation operator evaluated with respect to the 'true' density $f(\theta)$. This serves as a measure of closeness between the two densities and is defined by the Kullback-Leibler information criterion (KLIC) which is defined by

$$E_f[\cdot] = \int_{\mathbb{R}} \log \frac{f(\theta)}{g(\lambda)} f(y;\theta)\, dy\,. \tag{4.14}$$

The quantity in (4.14) is then minimized by choosing λ. This can be equivalently reformulated as

$$\max_\lambda E_f[\log g(\lambda)]\,. \tag{4.15}$$

The solution to this problem is called the pseudo-true value of λ given θ.

The main problem when using the Cox test statistic is that the measure of closeness can be analytically derived in a closed form only for very specific and simple cases such as the linear versus log-linear model (see Aneuryn-Evans and Deaton (1980)). It is however not possible in general. In cases where the exact expression of the KLIC is very complicated or even impossible to obtain, e.g. for nonlinear models, Pesaran and Pesaran (1993) and Lu et al. (2008) propose simulation based methods for approximating the quantity.

Regularity conditions for the applicability of Cox's test are given by White (1982b) for the *i.i.d.* case. It is basically required that consistent and asymptotically normally distributed estimators

for the parameters θ and λ can be obtained. This is especially important for the estimator of λ as we need a consistent and asymptotically normally distributed estimator which allows the likelihood function to fail to correspond to the true joint density of the observations. Such an estimator is the quasi-maximum-likelihood estimator (QMLE). Consistency and asymptotic normality results are provided by White (1982a) for the *i.i.d.* case, by White (1981) for the nonlinear case, by Bollerslev and Wooldridge (1992) for the dynamic linear case and by Gallant and White (1988) for the dynamic nonlinear case. These results ensure a wide applicability of Cox's test.

4.3.1 A bootstrap based likelihood ratio test

For the likelihood-ratio approach of Cox (1961, 1962) a simulation method for the test statistic is developed by Pesaran and Pesaran (1993) and compared to extant tests in the literature by Pesaran and Pesaran (1995). Lu et al. (2008) also consider simulation methods for the Cox test statistic and additionally propose a simulated Wald-type encompassing test. However, the latter test is computationally very costly in the context of nonlinear models as it involves a double loop design similar to a double bootstrap. Coulibaly and Brorsen (1999) report the results of a simulation study in which they compare different ways of computing the Cox test statistic. Their results suggest that simulation of the whole test statistic and the use of Monte Carlo p-values instead of simulating only parts of the test statistic and relying on asymptotic critical values is the more promising approach in finite samples. These results have also been confirmed more recently by Godfrey and Santos Silva (2005) and Kapetanios and Weeks (2003). The latter authors consider non-nested testing in a time series context to distinguish between several non-nested nonlinear time series models for the conditional mean. Different methods and test statistics based on the likelihood ratio principle are explored. Similar to Coulibaly and Brorsen (1999) Kapetanios and Weeks (2003) find that a studentized but not mean-adjusted test statistic with a simple variance estimator performs best over a variety of different settings; see also Lee and Brorsen (1997) for an application to nonlinear models for the conditional variance. Using the notation from section 4.3 we write the two rival non-nested models described in section 4.2 as

$$H_f \quad : \quad \mathcal{F}_\theta = \{ f(y_t|\mathfrak{F}_{t-1};\theta),\ \theta \in \Theta \subseteq \mathbb{R}^r \},\ t = 1,\dots,T \tag{4.16}$$

$$H_g \quad : \quad \mathcal{F}_\lambda = \{ g(y_t|\mathfrak{F}_{t-1};\lambda),\ \lambda \in \Lambda \subseteq \mathbb{R}^q \},\ t = 1,\dots,T\ . \tag{4.17}$$

For our test the model $\mathcal{F}_\theta$ in (4.16) is an ESTAR model as in (4.2) and the model $\mathcal{F}_\lambda$ in (4.17) is a MSAR model as in (4.6).

To keep the notation simple we write $f(\theta)$ and $g(\lambda)$ for (4.16) and (4.17) respectively whenever

possible. The log-likelihood functions for the models (4.16) and (4.17) can be written as

$$\begin{aligned} l_f &= \frac{1}{T}\sum_{t=1}^{T}\log f_t(\theta) \\ l_g &= \frac{1}{T}\sum_{t=1}^{T}\log g_t(\lambda)\,. \end{aligned}$$

Let $\hat{\theta}$ and $\hat{\lambda}$ denote the parameter values maximizing these functions. Then the log-likelihood ratio reads $l_f(\hat{\theta}) - l_g(\hat{\lambda})$. In order to studentize this likelihood ratio Coulibaly and Brorsen (1999) and Kapetanios and Weeks (2003) consider different estimators based on the outer-product of the scores of the models under consideration as in Berndt et al. (1974) and based on the information equality. A third alternative calculates

$$\hat{V}^2 = \frac{1}{T-1}\sum_{t=1}^{T}\left(d_t - \bar{d}\right)^2, \tag{4.18}$$

where $d_t = l_{f,t}(\hat{\theta}) - l_{g,t}(\hat{\lambda})$ is the likelihood ratio for the t-th observation of y_t and $\bar{d}$ is the respective arithmetic mean. These three methods are asymptotically equivalent but Pesaran and Pesaran (1993), Coulibaly and Brorsen (1999) and Kapetanios and Weeks (2003) report superior performance of the simple variance estimator in (4.18). Therefore, we adopt this approach in our test.

The test statistic we consider reads

$$S = \frac{\sqrt{T}\left\{l_f(\hat{\theta}) - l_g(\hat{\lambda})\right\}}{\sqrt{\hat{V}^2}}\,. \tag{4.19}$$

Note that the test statistic in (4.19) is not mean adjusted by an estimate of a measure of closeness of the two distributions in (4.16) and (4.17) such as the KLIC. This reduces the computational burden significantly compared to the methods of Pesaran and Pesaran (1993) and Lu et al. (2008) but renders the test statistic asymptotically non pivotal (see e.g. Pesaran and Pesaran (1993) and Godfrey (2007)).[2] However, as Hall and Titterington (1989) show, non pivotal statistics will have the same asymptotic accuracy regarding size and power as pivotal statistics. Thus we can reduce the computational burden by using the non mean adjusted statistic in (4.19) as we only need one bootstrap loop instead of two nested loops for computing an estimate of the KLIC (see Lee and Brorsen (1997) for a related approach).

We use the following parametric bootstrap to resample the likelihood ratio statistic in (4.19):

(i) Obtain the initial estimates $\hat{\theta}$ and $\hat{\lambda}$ from y_t and compute the test statistic S in (4.19).

(ii) Generate bootstrap samples by parametric resampling from the fitted model under the null $f(\hat{\theta})$. y_t^b denotes the t-th observation of the b-th bootstrap sample which is dependent on $\hat{\theta}$, i.e. $y_t^b\left(\hat{\theta}\right)$.

[2] A statistic η is called asymptotically pivotal if $a_n + b_n\eta$ has a proper nondegenerate limiting distribution not depending on unknowns for a sequence of known constants $\{a_n\}$ and $\{b_n\}$; see Hall (1992, p. 14).

(iii) For the b-th bootstrap sample let $\hat{\theta}^b$ and $\hat{\lambda}^b$ denote the parameter estimates obtained from maximizing $l_f^b = T^{-1}\sum_{t=1}^{T}\log f\left(y_t^b\left(\hat{\theta}\right)|\theta\right)$ and $l_g^b = T^{-1}\sum_{t=1}^{T}\log g\left(y_t^b\left(\hat{\theta}\right)|\lambda\right)$. Using $\hat{\theta}^b$ and $\hat{\lambda}^b$ compute the bootstrap analog to S in (4.19):

$$S^b = \frac{\sqrt{T}\left\{l_f^b\left(\hat{\theta}^b\right) - l_g^b\left(\hat{\lambda}^b\right)\right\}}{\sqrt{\hat{V}_b^2}}\,.$$

(iv) Repeat steps (ii) – (iii) B times and save the bootstrap test statistic S^b. This will give a small sample approximation of the distribution of S in (4.19).

(v) Compare S from (i) with critical values obtained from the distribution of S^b to decide which model captures the data best.

This bootstrap algorithm can be easily implemented using only a single loop design and thus the computational burden is not as heavy as the simulated Wald-type test of Lu et al. (2008). The applied hypotheses in (4.16) and (4.17) are beneficial compared to e.g. a Wald-type test because we have a closed form alternative. Common testing procedures claim that the model under the null is the correct model and test this model against a universe of alternative models. We, however, merely decide whether ESTAR is the better approximation to the data compared to an MSAR model. By this we are able to translate the competing economic theories about real exchange rates into econometric models and decide which theory is best supported by empirical data.

4.3.2 Finite sample properties

For the size experiment of the bootstrap likelihood ratio test we use the following data generating process

$$y_t = 0.95y_{t-1} - 0.55y_{t-1}\left\{1 - \exp\left(-(y_{t-1}-1)^2\right)\right\} + \varepsilon_t\,, \tag{4.20}$$

where $\varepsilon_t \overset{iid}{\sim} N(0,1)$. The first regime is moderately persistent while the second is highly persistent. This feature of our data generating process is consistent with the empirical literature on ESTAR models, see e.g. Rapach and Wohar (2006), Taylor et al. (2001), Öcal (2000) or Gatti et al. (1998). To keep the computational burden reasonable we consider $B = 200$ bootstrap replications and $M = 1000$ Monte Carlo repetitions.
Table 4.1 shows the size properties of the bootstrap likelihood ratio test for finite samples.

α	$T = 100$	$T = 200$	$T = 300$	$T = 400$
1%	2.10	1.10	1.10	1.00
5%	7.30	5.50	6.00	5.30
10%	13.60	11.70	11.60	11.20

Table 4.1: Empirical size of bootstrap likelihood ratio test [in %].

The test shows virtually no size distortions even in small samples which are frequently encoun-

tered when using monthly or quarterly macro data.
For the power results we consider a Markov switching model which is parameterized such as to resemble empirically estimated MSAR models as in Bergman and Hansson (2005). These mean switching AR(1) models contain a stationary but strongly persistent autoregressive parameter and a switching mean which is significantly different across both regimes. The model we use reads

$$y_t = 0.85(y_{t-1} + \mu_{s_{t-1}}) + \varepsilon_t \,. \tag{4.21}$$

The mean across regimes is specified as

$$\mu_{s_t} = \begin{cases} 1, & \text{if } s_t = 1 \\ 5, & \text{if } s_t = 2 \,. \end{cases}$$

The transition probabilities for the regimes are $p_{11} = p_{22} = 0.9$. This leads to two regimes with equal unconditional ergodic regime probabilities (see Hamilton (1994)) of

$$\mathbb{P}(s_t = 1) = \mathbb{P}(s_t = 2) = \frac{1 - p_{22}}{2 - p_{11} - p_{22}} = \frac{1 - p_{11}}{2 - p_{11} - p_{22}} = 0.5 \,.$$

The power results for this experiment are displayed in table 4.2.

α	$T = 100$	$T = 200$	$T = 300$	$T = 400$
1%	64.70	93.10	97.80	97.80
5%	86.50	98.30	98.60	98.30
10%	94.90	98.90	98.60	98.40

Table 4.2: Empirical power of bootstrap likelihood ratio test [in %].

The power results are very encouraging especially for the empirically most important cases of small sample sizes. The power increases steeply with the sample size and we also recommend to preferably use the 5% and 10% level of significance rather than the restrictive 1% level. An alternative is to increase the bootstrap replications to smooth the tails of the bootstrap distribution and obtain an even better approximation to the real distribution.[3]

4.4 Modeling real exchange rates

The real exchange rate data we use is the data set of Smallwood (2005) to ensure comparability with the existing literature.[4] However, the focus of Smallwood (2005) lies on the joint testing of long memory and nonlinearity in real exchange rates. The nonlinearity found by him is then modeled as the usual ESTAR transition function and no comparison to a nonlinear alternative is

[3] In another set of simulations we specified the data generating process such as to resemble empirical estimates of Bergman and Hansson (2005). These simulations yielded equally satisfactory results. However, they are unreported for the sake of brevity but can be obtained on request.
[4] The data is available at the article's website at: http://www.bepress.com/snde/vol9/iss2/art7/.

considered. Hence, our analysis adds to the literature by providing thorough statistical evidence of how the underlying nonlinear dynamic of real exchange rates is best described.

The data set comprises the exchange rates of 11 countries against the United States dollar as domestic currency. It includes data of Argentina, Brazil, France, Germany, Italy, Japan, Mexico, the Netherlands, Portugal, Spain, and the United Kingdom. The data starts in January 1973 and is sampled at a monthly frequency. The data set ends for all countries from the Euro area in December 1998 ($T = 312$), for Japan in February 2002 ($T = 350$) and for the rest in March 2002 ($T = 351$).

For real exchange rates Taylor et al. (2001) propose a formulation of the ESTAR model which reads

$$(y_t - \mu) = \psi(y_{t-1} - \mu) + \phi(y_{t-1} - \mu)\left\{1 - \exp\left(-\gamma(y_{t-1} - \mu)^2\right)\right\} + \varepsilon_t \,. \tag{4.22}$$

The real exchange rate fluctuates around the location parameter μ which models the long-run equilibrium. If we are in the extreme case of $y_{t-1} = \mu$ the process behaves like a random walk and is pulled back into the stationary region once the deviation becomes too large and $y_{t-1} \neq \mu$. The parameter γ governs the speed of the mean reversion. Thus, Taylor et al. (2001) assume that μ is the long-run mean around which the process fluctuates. We de-mean the data prior to testing. Thus, the long-run mean is $\mu = 0$ and we obtain a simplified model

$$y_t = \psi y_{t-1} + \phi y_{t-1}\left\{1 - \exp\left(-\gamma y_{t-1}^2\right)\right\} + \varepsilon_t \,. \tag{4.23}$$

Note that this only reduces the computational burden of estimating a nonlinear model and does not affect our results. Additionally, this is economically reasonable as the log real exchange rate is equal to zero if PPP holds.

We test this model against a MSAR(1) model with a switching mean as in (4.7) using $B = 500$ bootstrap repetitions.[5]

The results of the tests for the individual countries are displayed in table 4.3. The column 'Critical value' shows the critical value at the 5% level of significance from the bootstrap distribution. Note that the column 'Test statistic' displays negative values because we minimized the negative log-likelihood function.

[5] We confirmed by simulation that the procedure from section 4.3.1 maintains its good size and power properties in this special situation. The results are unreported to save space but can be obtained on request.

Country	Test statistic	Critical value	Suggested model
Argentina	-0.499	1.230	MSAR
Brazil	-0.727	-0.244	MSAR
France	-4.302	-17.758	ESTAR
Germany	-4.136	-17.324	ESTAR
Italy	-0.162	-3.815	ESTAR
Japan	-5.084	-0.981	MSAR
Mexico	-0.611	-0.215	MSAR
Netherlands	-4.480	-18.496	ESTAR
Portugal	-3.123	-16.744	ESTAR
Spain	-0.748	-5.806	ESTAR
United Kingdom	-0.210	-3.039	ESTAR

Table 4.3: Empirical test results for real exchange rates.

We can roughly classify the countries into three categories: Western European countries such as France and Germany, Latin-American countries such as Argentina and Brazil and Japan as another western-oriented major industrial nation.
The test decisions are obtained based on unrestricted models. This is in contrast to Taylor et al. (2001) who restrict the ESTAR model a priori to have a unit root in the first regime and a white noise process in the second regime. However, it turns out that the first regime is consistently estimated as a unit root and that the second regime is very close to white noise. Therefore, we perform a likelihood ratio test of the null that $\psi = 1$ and $\phi = -1$. This test supports evidence that the restricted model yields equally good results. As in Taylor et al. (2001) we can not reject the hypothesis that the restricted model is an equally reasonable description of the data. We subsequently estimate the restricted model. Estimation results and test results are displayed in table 4.4.

Country	$\hat{\gamma}_{restricted}$	LR test	p-value
France	0.1083	1.245	0.5367
Germany	0.0905	1.233	0.5400
Italy	0.2381	0.006	0.9968
Netherlands	0.1012	1.582	0.4533
Portugal	0.0771	0.399	0.8193
Spain	0.1034	0.645	0.7243
United Kingdom	0.3261	0.3811	0.8265

Table 4.4: Estimation and test results ESTAR.

The results are comparable with the results of Rapach and Wohar (2006) given that we use an extended data set. The conclusion is similar in that we find evidence of slight nonlinearity in the data but the speed of adjustment towards an equilibrium is slower, i.e. the interval of the fluctuation around PPP is wider.

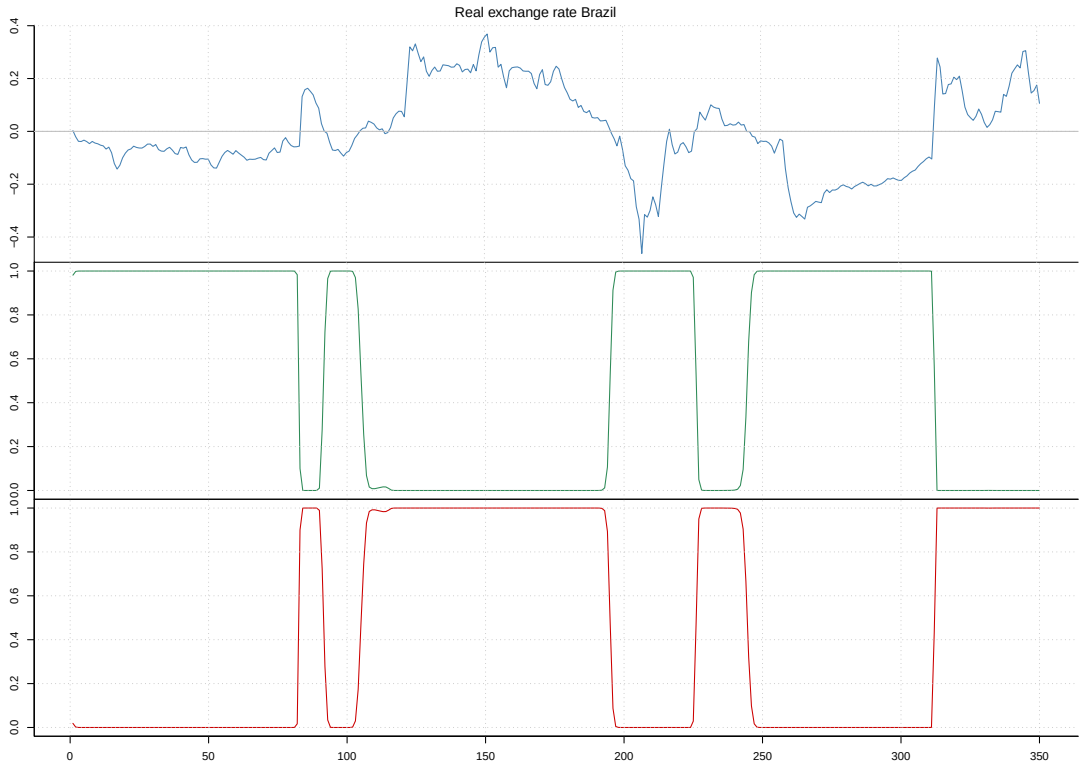

Figure 4.1: Real exchange rate Brazil with smoothed transition probabilities.

The estimation results for the MSAR models are displayed in table 4.5.

Country	$\hat{\mu}^{(1)}$	$\hat{\mu}^{(2)}$	$\hat{\phi}$	$\hat{p}_{11}$	$\hat{p}_{22}$	$\hat{\sigma}^2$
Argentina	0.090	0.733	0.582	0.983	0.976	0.018
Brazil	-0.126	0.150	-0.940	0.983	0.975	0.092
Japan	-0.392	0.005	0.588	0.990	0.989	0.109
Mexico	-0.191	0.123	-0.785	0.982	0.972	0.108

Table 4.5: Estimation results MSAR.

All estimated regime probabilities show a highly persistent behavior over time. All real exchange rates show a distinct mean switch between the regimes with a changing sign in each regime. An exception is Argentina. However, the difference between the mean coefficients is quite large given the range of the data. Note that our estimates are smaller than those of Bergman and Hansson (2005) because we did not multiply the data by 100. Looking at the stationarity conditions for our models in (4.9) and (4.10) we see that all estimated models fulfil these conditions. As an example of how the regime change dynamic behaves over time figures 4.1 and 4.2 show the real exchange rates of Brazil and Mexico along with their smoothed transition probabilities, respectively.

The investigated countries show some interesting and distinctive results. According to the theory reviewed in section 4.1, adequacy of the MSAR model supports the bubble theory for real exchange rates. Looking at the Latin-American countries a MSAR model seems to be more

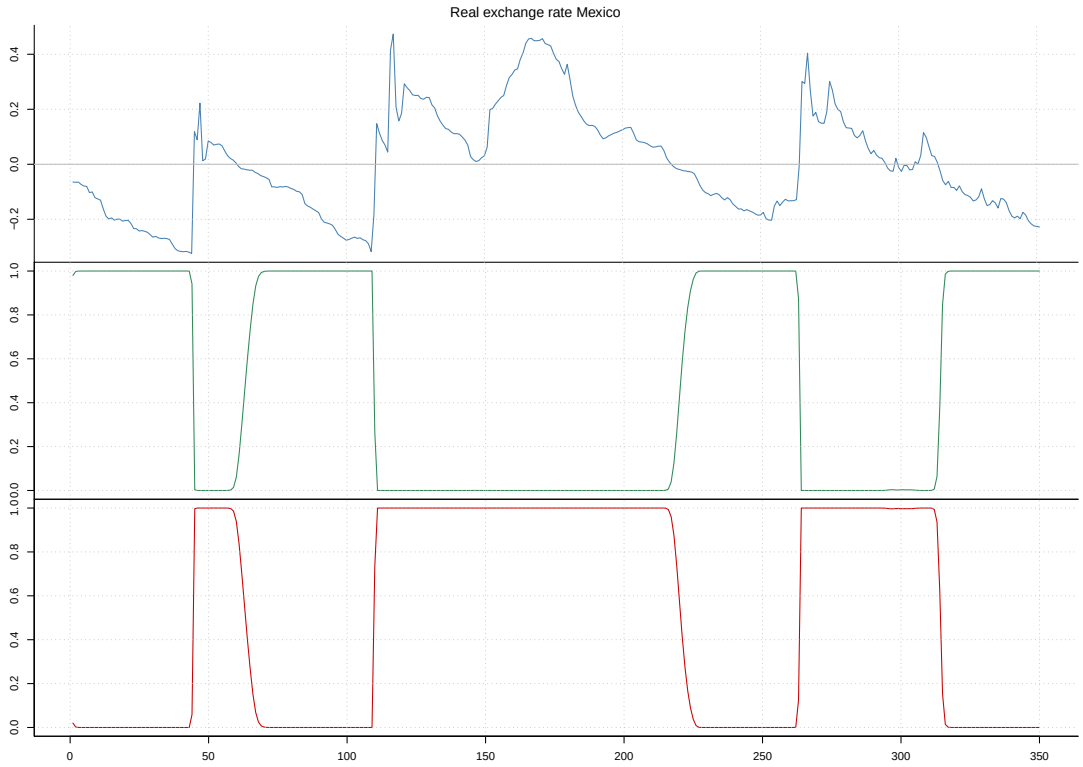

Figure 4.2: Real exchange rate Mexico with smoothed transition probabilities.

appropriate to model the dynamics of real exchange rates while the western European countries seem to be well-described by an ESTAR model. For Japan, again, the MSAR specification seems to be more appropriate. The question is: What is the difference between these groups that explains the different real exchange rate dynamic?

To answer this question for the Latin-American countries we focus our analysis on economic bubbles based on expectations about monetary policy and developments. In the 1960s and 1970s many of these countries borrowed huge amounts from international creditors to foster industrialization. With the beginning of the world recession and the first oil crisis during the first half of the 1970s the developing countries found themselves in a severe liquidity crunch. This situation raised interest rates in the United States and Europe and caused the deterioration of the exchange rates of the Latin-American countries. In the aftermath of the declaration of the national bankruptcy of Mexico in 1982 the incomes in the Latin-American countries dropped, economic growth stagnated and inflation rose to levels of hyperinflation.

In Argentina, the crisis was intensified by the military coup in 1976 which led to a monetarist financial liberalization. The excessive money supply increased the inflation rate to over 100% a year and rose the national foreign debt causing bank-runs and destroying business confidence. During the 1980s and early 1990s Argentina experienced periods of high inflation rates peaking in 1989 with 5000% a year which subsequently decreased to single digits by 1993 due to a more liberal economic policy. High government spending increased the national debt during the 1990s resulting in loss of confidence of foreign investors and causing another bank-run and a subsequent economic crisis in the late 1990s and early 2000s.

In Brazil, the oil crisis stopped economic growth and caused a recession which piled up foreign debt. Inflation remained a major problem during these periods due to the exchange rate devaluation of the austerity programm by the IMF and a growing public deficit. Inflation peaked in 1993 to 5000% per year. In the 1990s the first post-military government introduced a stabilization plan ('plano real') including a monetary reform which stabilized the inflation rate subsequently on a moderate level.
The history of Mexico can be classified into three phases. In the 1970s as a response to the oil shock the inflation rose and foreign debt was piled up. From 1983 to 1988 inflation rose to an average of 100% per year. In the 1990s Mexico ratified the North American Free Trade Agreement (NAFTA) which contributed to the economic recovery and helped controlling inflation.
All three countries are characterized by frequent and sudden economic regime changes in response to adverse economic situations which caused very volatile and high inflation rates. Despite of no formal cooperation on the political level, the economies in the countries behave very similar (see Singh (2006)). The changes could not have been anticipated and are thus better characterized by an abrupt change modeled by the MSAR model.
During the same period the western European countries started to cooperate politically as well as in economic policy in particular within the European Economic Community (EEC). This collaboration got even closer with the transition to the European Community (EC). This led to a rather similar behavior of the real exchange rates of these countries. In particular, central banks of the western European countries are independent from each other but nevertheless pursued a similar interest rate policy to control inflation on a rather low level. This led to the adoption of inflation targeting during the 1990's.
In Europe the aftermath of the oil crisis was not as severe as in the Latin-American countries because the foreign national deficit was not as high. Additionally the European countries were much more developed compared to the Latin-American countries and therefore able to deal with increasing inflation without risking hyperinflation. This is empirically investigated by Catão and Terrones (2005) who show a strong positive correlation between fiscal deficit and inflation for developing countries but not for industrialized countries.
In those stable regimes the information of past periods is always useful to describe real exchange rates. Therefore, the effect for the western European countries is described more appropriately by a smooth adjustment process such as postulated by the ESTAR model.
The finding that Japan, although a highly developed country, is better characterized by a MSAR model is due to the persistent deflation over the last two decades. After the oil crisis in 1973 inflation rose. During the late 1980 a lot of speculative money has been invested in Japan causing the Japanese asset price bubble which burst in 1991 followed by a decade of zero inflation rates and deflation in the 1990s. Since then, the Japanese central bank has kept the interest rate close to zero but could not stop deflation tendencies. These two regimes of high inflation during the 1970s and extremely low inflation during the late 1980s and 1990s can be captured well by a MSAR model. It is however important to note that in the case of Japan this result is much less clear because standard economic paradigms such as the relation regarding interest rate and inflation rate levels do not hold for Japan.

4.5 Conclusion

Different theories about the existence of the PPP support different economic views of the in-sample dynamic driving real exchange rates. These different views result in different models frequently used in the analysis of PPP; namely ESTAR and MSAR models. As both models are able to support PPP under certain conditions the question which model to use for the analysis is usually answered upon prior economic belief rather than statistical model selection procedures. However, as the dynamics of the competing models are rather different the question which model captures the data best is important as it results in different economic theories. In this paper we propose a bootstrap based likelihood ratio test that allows us to discriminate between both classes of nonlinear time series models. The bootstrap approximation of the asymptotic distribution of the test statistic allows us to obtain convincing power results for sample sizes frequently encountered in empirical studies. This is important as asymptotic tests have been shown to work unsatisfactorily in small samples (see Coulibaly and Brorsen (1999) and Kapetanios and Weeks (2003)).

In an empirical application we find that the real exchange rates of countries with high inflation rates such as Argentina or Japan are modeled best using MSAR models, thus supporting the theory of bubbles in real exchange rates. Countries not suffering from high inflation such as France or Germany are better described by an ESTAR model. Therefore, continuous adjustment towards a long-run PPP equilibrium can be concluded.

Chapter 5

A note on testing for purchasing power parity

A note on testing for purchasing power parity

Published as Leibniz University of Hannover Discussion Paper No. 471.

5.1 Introduction

Purchasing power parity (PPP) is arguably one of the most important parities in international economics because "most [economists] instinctively believe in some variant of purchasing power parity as an anchor for long-run real exchange rates" (Rogoff (1996)). At the same time there is an ongoing debate whether PPP holds or not. One popular way to test for PPP is to check whether real exchange rate data follows a mean reverting process and is thus (globally) stationary. The first studies used the classical Dickey-Fuller test (DF test) to check whether a random walk can be found in real exchange rates (see e.g. Adler and Lehmann (1983), Meese and Rogoff (1983), Meese and Rogoff (1988) or Caporale et al. (2003)). As these studies could not reject the null of a random walk and therefore found no evidence for PPP economists have turned to nonlinear models to capture a possibly nonlinear adjustment towards PPP (see e.g Taylor et al. (2001)). Especially models that allow for a symmetric adjustment towards PPP have been found useful. The most notable is the exponential smooth transition autoregressive (ESTAR) model (see Teräsvirta (1994)). In this framework Kapetanios et al. (2003) have developed a unit root test against an ESTAR alternative which subsequently has been successfully applied to confirm PPP (see e.g. Liew et al. (2004), Chortareas et al. (2002) and Chortareas and Kapetanios (2004)). Many papers construct real exchange rates from log data to obtain the log real exchange rate which is consistent with the theory. However, before the analysis is performed the logarithm is applied a second time to the constructed series although, based on theory, it is not a priori clear whether the suspected unit root behavior is in the logs or in the levels (see e.g. Chortareas et al. (2002), Caporale et al. (2003), Bergman and Hansson (2005), Kruse (2009)). It is therefore of interest whether the applied unit root tests maintain their properties if applied to erroneously transformed data. For the linear case this has been done by Krämer and Davies (2002) and Davies and Krämer (2003) (see also de Jong (2010)). Krämer and Davies (2002) are able to analytically confirm parts of the simulation results of Granger and Hallman (1991) that the DF test overrejects a correct null of a random walk. Another striking feature of real exchange rates are the unusual small estimated variances of the innovation term (see e.g. Taylor et al. (2001), Rapach and Wohar (2006) or Smallwood (2005)). An important question is whether or not the unit root tests maintain their properties if the innovation variance becomes very small or very large.

This paper extends the analysis of Krämer and Davies (2002) to the nonlinear unit root test of Kapetanios et al. (2003). We show analytically and by Monte Carlo evidence that the behavior of the test statistic depends heavily on the innovation variance of the underlying process if the

data is erroneously nonlinearly transformed.
The rest of the paper is structured as follows: Section 5.2 presents the considered unit root tests and derives their asymptotic behavior under nonlinear transformation. Section 5.3 underpins the analytical results with some Monte Carlo evidence before section 5.4 concludes. All proofs are collected in the appendix 5.5.

5.2 Unit root tests and their asymptotic behavior under misspecification

The first studies researching the PPP relied on the DF test for a unit root against a linear alternative by Dickey and Fuller (1979). The DF test regression reads

$$y_t = \rho y_{t-1} + \varepsilon_t \,. \tag{5.1}$$

The pair of hypotheses is given as

$$H_0 : \rho = 1 \quad \text{vs.} \quad H_1 : \rho < 1 \,. \tag{5.2}$$

The DF test checks if $T(\hat{\rho} - 1)$ is sufficiently close to zero. The quantiles of the resulting non-standard distribution of ρ are tabulated e.g. in Hamilton (1994, p. 762).
Kapetanios et al. (2003) extend the linear framework to the nonlinear ESTAR case (KSS test). The prototypical ESTAR process they consider reads

$$y_t = \phi y_{t-1} + \theta y_{t-1} \mathcal{G}(\gamma, c; \cdot) + \varepsilon_t, \text{ with } \varepsilon_t \overset{iid}{\sim} (0, \sigma^2) \,. \tag{5.3}$$

$\mathcal{G}(\gamma, c; \cdot) : \mathbb{R} \to [0, 1]$ is a symmetrically u-shaped transition function governing the regime switching behavior of the process in (5.3). It reads

$$\mathcal{G}(\gamma, c; y_{t-d}) = 1 - \exp\left(-\gamma (y_{t-d} - c)^2\right) \,. \tag{5.4}$$

For simplicity Kapetanios et al. (2003) set $c = 0$ and $d = 1$. If $\phi = 1$ the process is still globally mean reverting as long as $-2 < \theta < 0$ which is assumed henceforth.
To derive the unit root test against the model in (5.3) they approximate the transition function in (5.4) by a first order Taylor approximation resulting in the following auxiliary regression model

$$\Delta y_t = \delta y_{t-1}^3 + \varepsilon_t \,. \tag{5.5}$$

The pair of hypotheses in this framework reads

$$H_0 : \delta = 0 \quad \text{vs.} \quad H_1 : \delta < 0 \,, \tag{5.6}$$

which is also tested via a conventional t-test. Critical values are reported in Kapetanios et al. (2003, p. 364).

We first consider the case when the unit root test is applied to the levels of the time series but the random walk is in the logs.
In this case let $z_t := \ln(y_t)$ and assume

$$z_t = \mu + z_{t-1} + \sigma\varepsilon_t, \text{ with } \varepsilon_t \overset{iid}{\sim} (0,1)\,, \tag{5.7}$$

and $0 < \sigma < \infty$. The DF test is applied to the levels $y_t := \exp(z_t)$. The OLS estimator for ρ in the model

$$y_t = \rho y_{t-1} + \varepsilon_t \tag{5.8}$$

reads

$$\hat{\rho} = \frac{\sum_{t=1}^{T} y_{t-1}y_t}{\sum_{t=1}^{T} y_{t-1}^2}\,. \tag{5.9}$$

For the linear case the behavior of the DF test for $\sigma^2 \to 0$ and $\sigma^2 \to \infty$ (T fixed) and for $T \to \infty$ (σ^2 fixed) has been studied by Krämer and Davies (2002) and Davies and Krämer (2003). For the case of $T \to \infty$ and fixed σ^2 Davies and Krämer (2003) show that the rejection probability tends to unity.
Krämer and Davies (2002) show that if $\sigma \to \infty$ the probability for rejecting the null tends to zero if the random walk is in the logs but the DF test is erroneously applied to the levels. In the appendix we show that a similar result also holds for the nonlinear unit root test of Kapetanios et al. (2003).

Theorem 5.2.1.
If the data is generated by (5.7), and $\hat{\delta}$ is the OLS estimator in (5.5) where $y_t = \exp(z_t)$ we have:

(i) $\sigma^2 \to 0$ *implies* $\hat{\delta} \xrightarrow{D} 0$.
$\sigma^2 \to 0$ *implies* $s.e.\left(\hat{\delta}\right) \xrightarrow{D} 0$.

(ii) $\sigma^2 \to \infty$ *implies* $\hat{\delta} \xrightarrow{D} -1$.
$\sigma^2 \to \infty$ *implies* $s.e.\left(\hat{\delta}\right) \xrightarrow{D} 1$.

The theorem shows that analog to the linear case the probability of rejecting the unit root hypotheses tends to zero as $\sigma \to 0$ because the test statistic defined as

$$\hat{t}_{KSS} = \frac{\hat{\delta}}{s.e.\left(\hat{\delta}\right)}\,, \tag{5.10}$$

where $s.e.\left(\hat{\delta}\right)$ is the standard error of $\hat{\delta}$ tends to zero as well because $\hat{\delta}$ tends to zero more quickly than $s.e.\left(\hat{\delta}\right)$.
For the part (ii) of the theorem it is important to note that the test statistic formally tends to -1 but at such a slow rate that in finite samples with reasonable sample sizes and fairly large values of σ the test statistic will be large because $s.e.\left(\hat{\delta}\right)$ approaches 1 from below and much slower than $\hat{\delta}$ approaches -1.

Now we consider the situation when the unit root test is applied to the logs but the unit root process is in the levels. The true data generating process is now given by (5.7) but the unit root test is applied to $y_t := \ln(z_t)$. Of course this type of misspecification can only happen if $z_t > 0$ $(t = 0, 1, \ldots, T)$. Theoretically this only happens if z_0 is rather large and/or a substantial drift component is present in the data generating process. This is however of practical importance because given a finite sample of empirical data, there is a good chance to encounter only positive values although negative values might be possible from a theoretical argument. For the linear case Krämer and Davies (2002) show that $\hat{\rho}$ in (5.9) tends to one both as $\sigma \to 0$ and as $\sigma \to \infty$. The next theorem, which is proven in the appendix 5.5, gives the limiting behavior of the OLS estimate for δ in (5.5).

Theorem 5.2.2.
If the data is generated by (5.7), and $\hat{\delta}$ is the OLS estimator in (5.5) where $y_t = \log(z_t)$ we have:

(i) $\sigma^2 \to 0$ *implies* $\hat{\delta} \xrightarrow{D} 0$.
$\sigma^2 \to 0$ *implies* $s.e.\left(\hat{\delta}\right) \xrightarrow{D} 0$.

(ii) $\sigma^2 \to \infty$ *implies* $\hat{\delta} \xrightarrow{D} 0$.
$\sigma^2 \to \infty$ *implies* $s.e.\left(\hat{\delta}\right) \xrightarrow{D} 0$.

This theorem implies that the test statistic of the KSS test in (5.10) converges to 0 because $\hat{\delta}$ converges faster than its standard error and hence the null is never rejected given that the test is one-sided with negative critical values.

5.3 Monte Carlo evidence

This section illustrates the analytic results from the previous section by means of Monte Carlo experiments. We start with the case that the random walk is in the logs but the unit root test is applied to the levels.
To illustrate the asymptotic behavior from theorem 5.2.1 figure 5.1 shows plots of $\hat{\delta}$ and $s.e.(\hat{\delta})$ for various values of σ and $T = 100$. Each point is obtained as a mean over 5000 Monte Carlo repetitions.

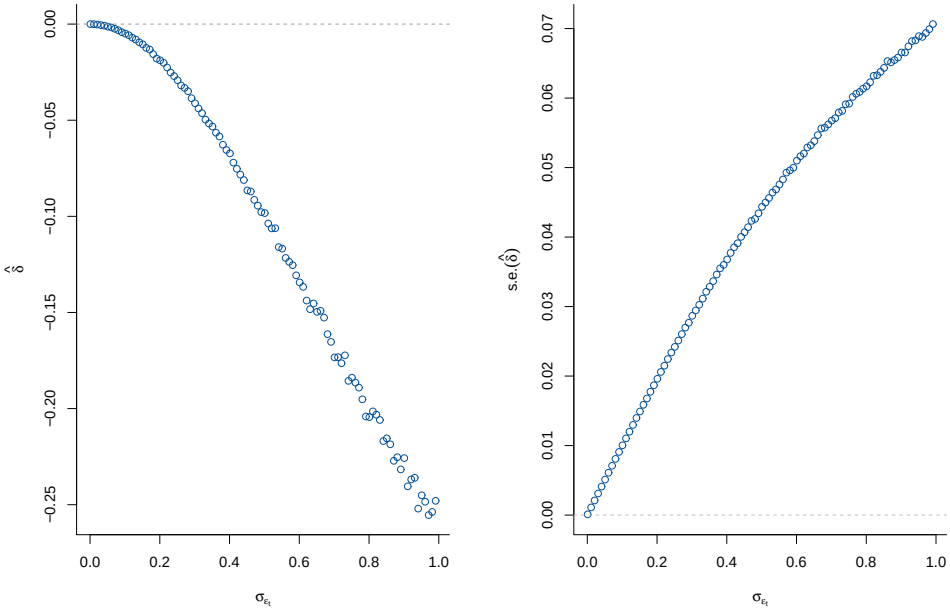

Figure 5.1: Illustration of theorem 5.2.1 part (i) I.

As predicted in the theorem both estimates tend to zero as $\sigma \to 0$. As a consequence also $\hat{t}_{KSS}$ tends to zero because $\hat{\delta}$ tends to zero more quickly. This is shown in figure 5.2. The right panel clearly shows a faster convergence of $\hat{\delta}$.

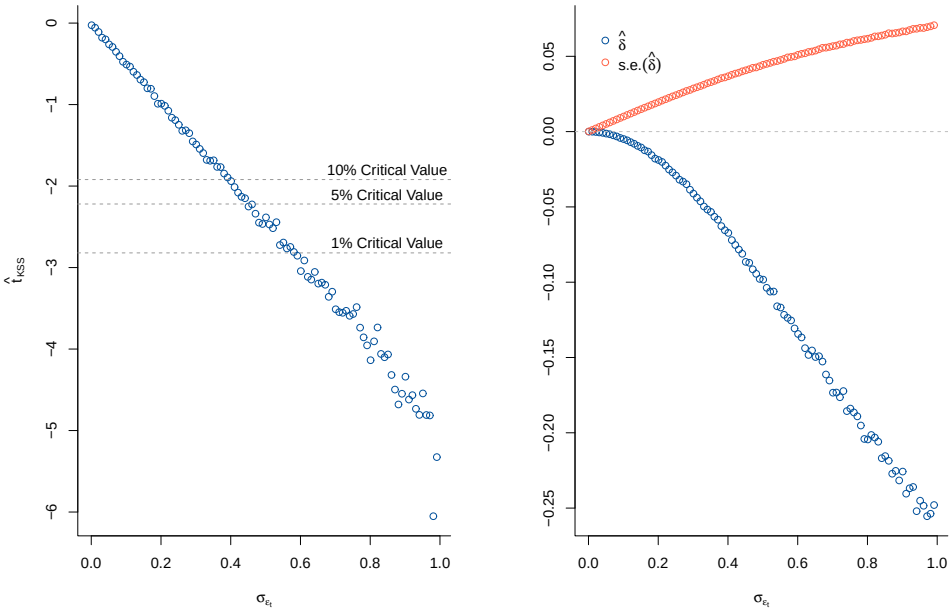

Figure 5.2: Illustration of theorem 5.2.1 part (i) II.

The associated power curves for the KSS test for this situation are depicted in figure 5.3.

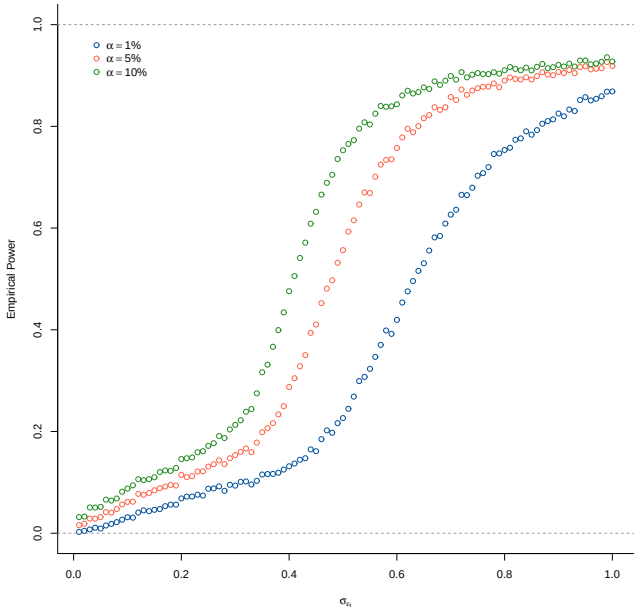

Figure 5.3: Power curves for exponential random walk I.

As expected the power curves are very low for small values of σ and eventually tend to one as σ increases.

Part (ii) of theorem 5.2.1 is illustrated in the following figures. Note that for these figures each point is obtained as the median over 5000 Monte Carlo repetitions rather than the mean. This is due to numerical instabilities when dealing with very large standard deviations. In such a case the noise dominates the signal and one obtains larger deviations from the mean estimate. However in order to get a notion about the central value of the distribution of the estimates we choose the more robust median as estimate.

Figure 5.4 shows the behavior of $\hat{\delta}$ and its standard deviation as $\sigma \to \infty$. As shown in the theorem $\hat{\delta}$ goes to -1 with increasing σ while the standard deviation goes very slowly to 1.

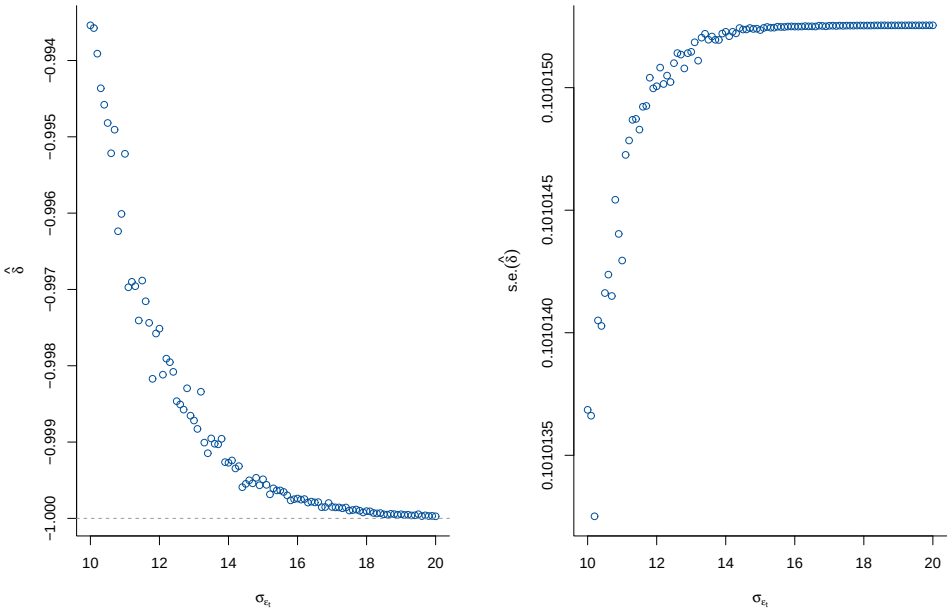

Figure 5.4: Illustration of theorem 5.2.1 part (ii) I.

Because of the slow convergence of $s.e.\left(\hat{\delta}\right)$ the associated test statistic $\hat{t}_{KSS}$ is rather large (see figure 5.5) leading to a power of 1 displayed in figure 5.6.

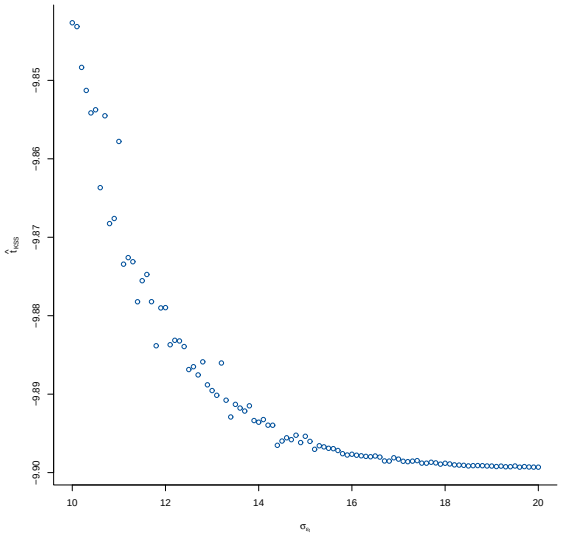

Figure 5.5: Behavior of $\hat{t}_{KSS}$ under theorem 5.2.1 part (ii).

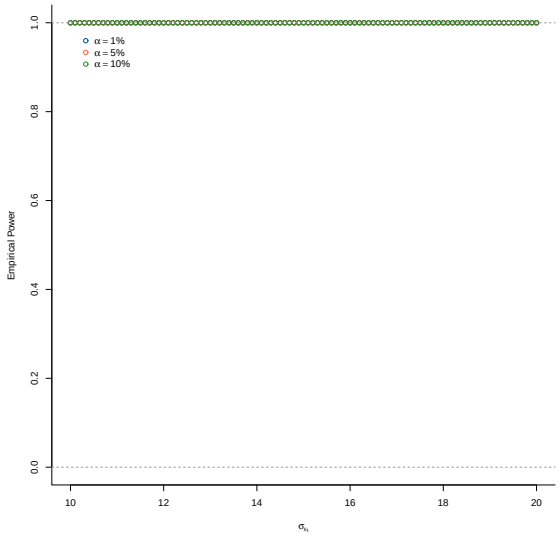

Figure 5.6: Power curves for exponential random walk II.

For the case that the random walk is in the levels but the unit root test is applied to the logs analog Monte Carlo experiments have been conducted. For part (i) of theorem 5.2.2 figure 5.7 shows the convergence of $\hat{\delta}$ and its associated standard error to zero as $\sigma \to 0$.

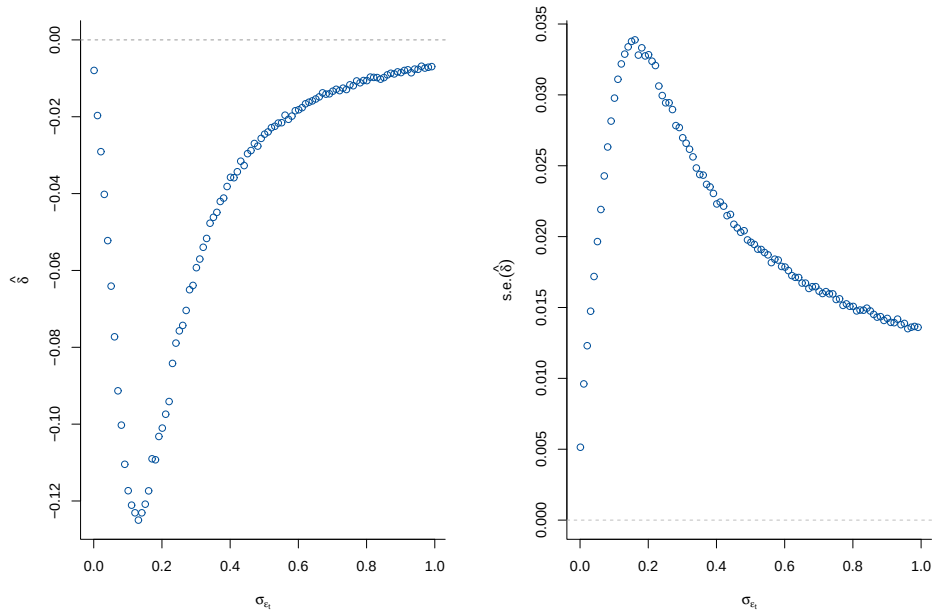

Figure 5.7: Illustration of theorem 5.2.2 part (i) I.

As predicted by theorem 5.2.2, $\hat{\delta}$ converges from negative values to zero and $s.e.\left(\hat{\delta}\right)$ converges from positive values to zero. Figure 5.8 shows the associated test statistic and illustrates the rates of convergence of $\hat{\delta}$ and $s.e.\left(\hat{\delta}\right)$.

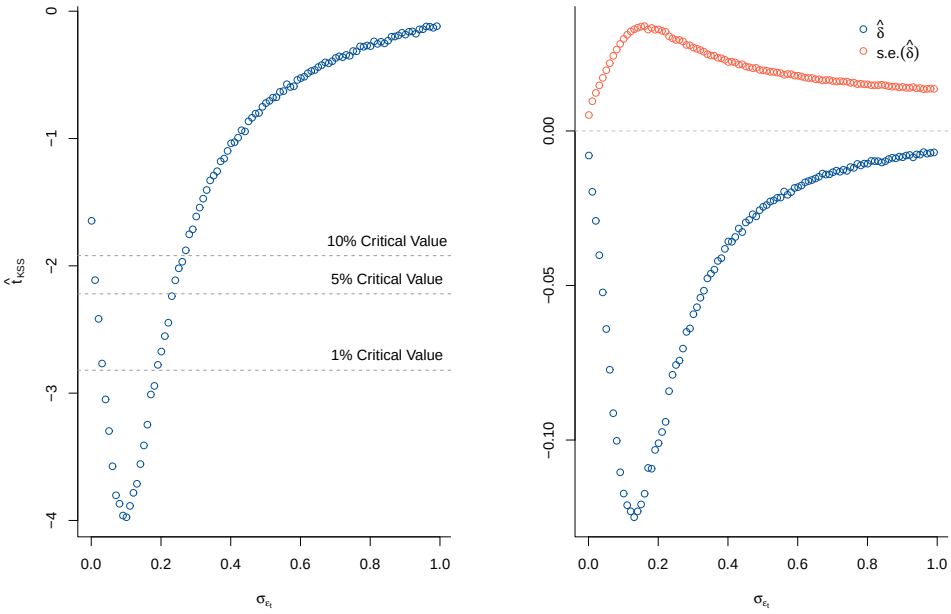

Figure 5.8: Illustration of theorem 5.2.2 part (i) II.

According to the theory the test statistic $\hat{t}_{KSS}$ converges to zero for small values of σ leading to non-rejection of the random walk hypotheses. The figures also clearly show an abnormal non-monotonic behavior of $\hat{\delta}$, $s.e.\left(\hat{\delta}\right)$ and $\hat{t}_{KSS}$. This induces even more uncertainty regarding the test decisions in this situation.

Figure 5.9 shows the associated power curves. As expected the power decreases with σ tending

to zero.

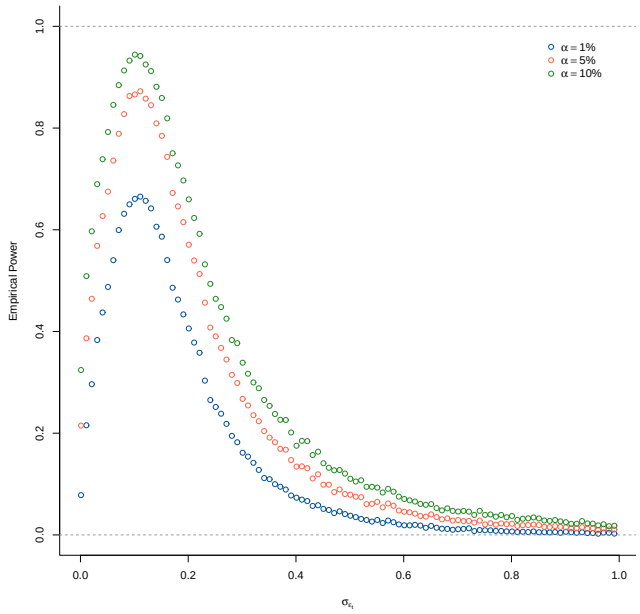

Figure 5.9: Power curves for logarithmic random walk.

The second part of theorem 5.2.2 analyzes the behavior of the KSS test for the case of $\sigma \to \infty$. Figure 5.10 shows the convergence of the OLS estimates $\hat{\delta}$ and $s.e.\left(\hat{\delta}\right)$ to zero as $\sigma \to \infty$.

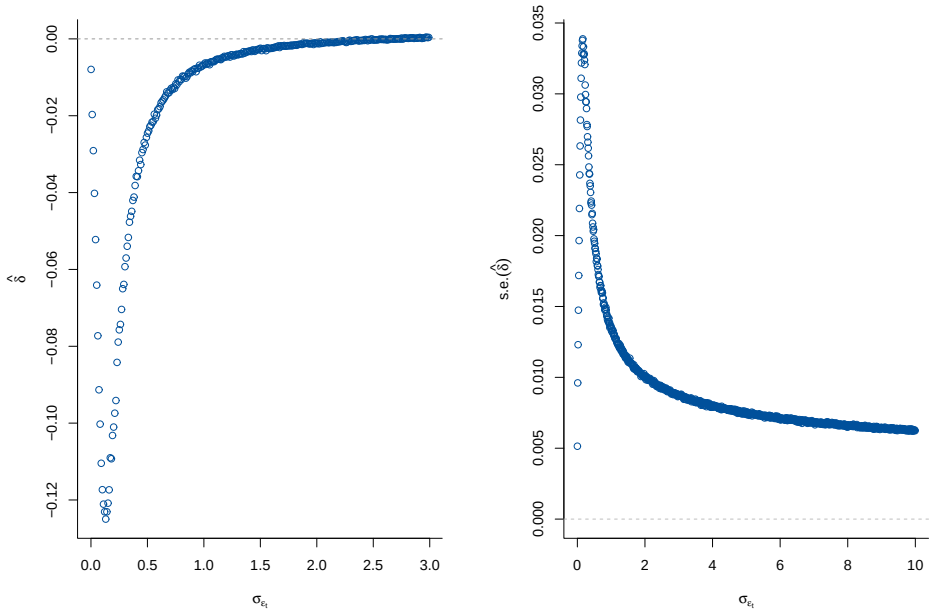

Figure 5.10: Illustration of theorem 5.2.2 part (ii).

The OLS estimates $\hat{\delta}$ is much faster than the convergence of $s.e.\left(\hat{\delta}\right)$ causing the test statistic $\hat{t}_{KSS}$ to rapidly converge to zero as well. This leads the KSS test to lose its power properties as the null cannot be rejected even for moderate values of σ. This is also supported by figure 5.11 that shows the power of the KSS test for large variances.

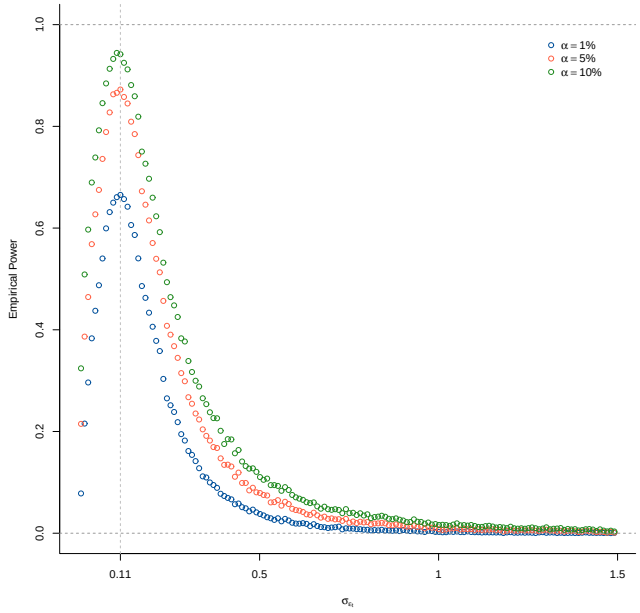

Figure 5.11: Power curves for logarithmic random walk and large variances.

The power decreases rapidly to zero. The vertical dashed line shows the maximal power value for the $\alpha = 5\%$ level. Except in the region around $\sigma \approx 0.11$ the power is virtually zero and thus the unit root hypothesis is never rejected in this case.

5.4 Conclusion

This paper studies the asymptotic behavior of the unit root test against a globally stationary ESTAR model proposed by Kapetanios et al. (2003). In particular we study the behavior of the test statistic if the innovation variance becomes very small or very large and the data has been erroneously nonlinearly transformed. I.e. the unit root is in the logs but the unit root test is applied to the levels or vice versa. We find that the behavior of the test depends heavily on the innovation variance if the data is transformed. If the unit root is in the logs but the unit root test is applied to the levels the test never rejects the random walk hypotheses if $\sigma \to 0$.
The situation when the unit root is in the levels but the unit root test is applied to the logs is even more severe. In this situation the test virtually never rejects the null. Only in a very small region of around $\sigma \approx 0.11$ the test rejects the null. This leads to a non-monotonic power of the KSS test.
The consequences for testing for PPP are that one should be very careful if nonlinear data transformation should be applied or not because in both considered cases the KSS test loses power and thus yields unreliable results.
Regarding data transformation in nonlinear time series future work should focus on discriminating between linear or logarithmic model formulation for nonlinear time series. A possible road is to extend the work of Kobayashi and McAleer (1999) or Spanos et al. (2008) to nonlinear time series models.
Possibly the asymptotic distributions of unit root tests for the case of erroneously transforming

the time series according to $\log(|z_t|)$ where z_t is $I(1)$ could be derived using the results in de Jong (2004) (see also de Jong and Schmidt (2002)).

5.5 Appendix

5.5.1 Proof of Theorem 5.2.1

Assume without loss of generality that $z_0 = 0$ and set the drift to $\mu = 0$. Then write

$$z_t = z_{t-1} + \sigma\varepsilon_t = \sigma \sum_{i=1}^{t} \varepsilon_i \,, \tag{5.11}$$

and notice that

$$\Delta y_t = y_t - y_{t-1} = \exp(z_t) - \exp(z_{t-1}) \,. \tag{5.12}$$

With this the OLS estimator of $\hat{\delta}$ in (5.5) reads

$$\begin{aligned}
\hat{\delta} &= \frac{\sum_{t=1}^{T} y_{t-1}^3 \Delta y_t}{\sum_{t=1}^{T} y_{t-1}^6} \\
&= \frac{\sum_{t=1}^{T} \left[\exp(3z_{t-1} + z_t) - \exp(4z_{t-1})\right]}{\sum_{t=1}^{T} \exp(6z_{t-1})} \\
&= \frac{\sum_{t=1}^{T} \left[\exp\left(3\sigma \sum_{i=1}^{t-1} \varepsilon_i + \sigma \sum_{j=1}^{t} \varepsilon_i\right) - \exp\left(4\sigma \sum_{i=1}^{t-1} \varepsilon_i\right)\right]}{\sum_{t=1}^{T} \exp\left(6\sigma \sum_{i=1}^{t-1} \varepsilon_i\right)} \,.
\end{aligned} \tag{5.13}$$

As the numerator tends to zero and the denominator tends to T as $\sigma \to 0$ the first part of (i) of the Theorem follows.
Now consider the OLS estimator of the standard deviation of $\hat{\delta}$

$$\begin{aligned}
s.e.\left(\hat{\delta}\right) &= \frac{\frac{1}{T-1} \sum_{t=1}^{T} \left(\Delta y_t - \hat{\delta} y_{t-1}^3\right)^2}{\sum_{t=1}^{T} y_{t-1}^6} \\
&= \frac{\sum_{t=1}^{T} \left[(\exp(z_t) - \exp(z_{t-1}))^2 - 2\hat{\delta}(\exp(z_t) - \exp(z_{t-1}))\exp(3z_{t-1}) + \hat{\delta}^2 \exp(6z_{t-1})\right]}{\sum_{t=1}^{T} \exp(6z_{t-1})} \,.
\end{aligned} \tag{5.14}$$

Given the representation from above we immediately see that the numerator tends to zero and the denominator tends to T as $\sigma \to 0$. Consequently $s.e.\left(\hat{\delta}\right)$ tends to zero as claimed in part the second part of (i) in the theorem. Note that $\hat{\delta}$ tends to zero more quickly so that $\hat{t}_{KSS}$ tends to zero as $\sigma \to 0$.
To prove the part (ii) note that for the case $\sigma \to \infty$ the slope of the exponential function becomes much steeper with increasing positive values as for negative values where the difference in the slopes eventually becomes negligible. We have from the representation in (5.13) that the

numerator tends to $-\infty$ because of the higher power of the second term. As the denominator tends to ∞ with $\sigma \to \infty$ the whole expression for $\hat{\delta}$ tends to -1 as claimed. In the representation (5.14) for the standard deviation the denominator clearly tends to ∞. In the numerator the crucial term is the last one. This expression also goes to ∞ but much slower than the denominator (because $\left|\hat{\delta}\right| < 1$). Thus the whole term will eventually go to one but very slowly. □

5.5.2 Proof of Theorem 5.2.2

Assume without loss of generality that $z_0 = 0$ and set the drift to $\mu = 0$. Then write

$$z_t = z_{t-1} + \sigma\varepsilon_t = \sigma \sum_{i=1}^{t} \varepsilon_i \,, \tag{5.15}$$

and notice that

$$\begin{aligned}
\Delta y_t = y_t - y_{t-1} &= \log(z_t) - \log(z_{t-1}) \\
&= \log\left(\sigma \sum_{i=1}^{t} \varepsilon_i\right) - \log\left(\sigma \sum_{i=1}^{t-1} \varepsilon_i\right) \\
&= \log(\sigma) + \log\left(\sum_{i=1}^{t} \varepsilon_i\right) - \log(\sigma) - \log\left(\sum_{i=1}^{t-1} \varepsilon_i\right) \\
&= \log\left(\frac{\sum_{i=1}^{t} \varepsilon_i}{\sum_{i=1}^{t-1} \varepsilon_i}\right) \\
&= \log\left(1 + \frac{\varepsilon_t}{\sum_{i=1}^{t-1} \varepsilon_i}\right) \\
&=: \log(\eta) \,.
\end{aligned} \tag{5.16}$$

With this the OLS estimator reads

$$\begin{aligned}
\hat{\delta} = \frac{\sum_{t=1}^{T} y_{t-1}^3 \Delta y_t}{\sum_{t=1}^{T} y_{t-1}^6} &= \frac{\sum_{t=1}^{T} \left[\log(z_{t-1})^3 \log(\eta)\right]}{\sum_{t=1}^{T} \left[\log(z_{t-1})^6\right]} \\
&= \frac{\sum_{t=1}^{T} \left[\left\{\log(\sigma) + \log\left(\sum_{i=1}^{t-1} \varepsilon_i\right)\right\}^3 \log(\eta)\right]}{\sum_{t=1}^{T} \left[\log(\sigma) + \log\left(\sum_{i=1}^{t-1} \varepsilon_i\right)\right]^6} \,.
\end{aligned}$$

Using the binomial theorem we can expand the powers of the binomials in the numerator and the denominator. Leading to the following expression for the numerator

$$\sum_{t=1}^{T}\left[\left\{\log(\sigma)^3+3\log(\sigma)^2\log\left(\sum_{i=1}^{t-1}\varepsilon_i\right)+3\log(\sigma)\log\left(\sum_{i=1}^{t-1}\varepsilon_i\right)^2+\log\left(\sum_{i=1}^{t-1}\varepsilon_i\right)^3\right\}\log(\eta)\right].$$

And for the denominator

$$\sum_{t=1}^{T}\left[\log(\sigma)^6+6\log(\sigma)^5\log\left(\sum_{i=1}^{t-1}\varepsilon_i\right)+15\log(\sigma)^4\log\left(\sum_{i=1}^{t-1}\varepsilon_i\right)^2+20\log(\sigma)^3\log\left(\sum_{i=1}^{t-1}\varepsilon_i\right)^3+\right.$$
$$\left.15\log(\sigma)^2\log\left(\sum_{i=1}^{t-1}\varepsilon_i\right)^4+6\log(\sigma)\log\left(\sum_{i=1}^{t-1}\varepsilon_i\right)^5+\log\left(\sum_{i=1}^{t-1}\varepsilon_i\right)^6\right].$$

The numerator tends to $-\infty$ as $\sigma\to 0$ and the denominator eventually tends to ∞. Thus $\hat{\delta}$ tends to 0 as claimed. Note that, because of the complex interplay of the powers of the logarithms involved and the associated changes of signs and the different slopes of the logarithmic functions in the interval $(0,1]$ the convergence is expected to be slow. However because the denominator converges at a higher rate than the numerator (because of the higher powers involved) the OLS estimate will eventually converge to zero.

Analyzing the standard deviation of $\hat{\delta}$ in (5.14) for $y_t=\log(z_t)$ we see for the same reasons as above that numerator slowly converges to zero (as $\hat{\delta}$ goes to zero) and the denominator diverges to ∞. Thus $s.e.\left(\hat{\delta}\right)\to 0$ as claimed.

For the first part of (ii) notice that the numerator as well as the denominator diverge to ∞. But again the denominator diverges much faster due to the high powers of $\log(\sigma)$ involved. Therefore $\hat{\delta}$ will eventually diverge to zero as claimed. The result for the standard error of $\hat{\delta}$ is obtained with the same reasoning. □

Chapter 6

Monitoring a change in persistence of a long range dependent time series

Monitoring a change in persistence of a long range dependent time series

Co-authored with Juliane Willert.

6.1 Introduction

The assumption of structural stability of an econometric model is a major issue in time series econometrics. If the parameter estimates stem from an unstable relationship they are not meaningful and additionally inference can be biased and forecasts yield inaccurate results (see e.g. Hansen (2001), Andrews and Fair (1988), Ghysel et al. (1997), Garcia and Perron (1996) or Clements and Hendry (1998)). In reaction to these findings a large amount of literature emerged that incorporated structural change in the inference techniques or analyzes forecasting subject to structural change more closely (see e.g. Perron (1989), Zivot and Andrews (1992) or Pesaran and Timmermann (2005)). Recently the possibility of a change in persistence, i.e. a change in the memory structure of the time series as a special case of structural instability, has become object of study (see e.g. Kim (2000), Kim et al. (2002), Busetti and Taylor (2004), Banerjee et al. (1992), Leybourne et al. (2003) or Leybourne et al. (2007)). This work has been placed within the $I(0)$ vs. $I(1)$, or vice versa, framework where the focus lies on short memory time series with an exponentially decaying autocorrelation structure.

However, since the seminal papers of Granger and Joyeux (1980) and Hosking (1981), long memory time series have become widely used in economics to model highly persistent time series as diverse as inflation rates or realized volatility (see e.g. Hassler and Wolters (1995) and Corsi et al. (2008)). Baillie (1996) provides an overview about various applications of long memory time series in economics.

Despite these facts little work has been done to test for a change in persistence in long range dependent time series. Notable exceptions are Beran and Terrin (1996), Ray and Tsay (2002), Sibbertsen and Kruse (2009) or Yamaguchi (2011). These test belong to the class of so-called "one-shot" tests (see Chu et al. (1996, p. 1045)), i.e. tests that are applied a posteriori to detect a structural break within a historical data set.

Because breaks can occur at any given time and also new data arrives steadily it is desirable for the applied econometrician to detect a change in persistence as soon as possible. This leads to a sequential testing problem (see Siegmund (1985) for an overview). As the usual "one-shot" tests work with constant critical values they cannot be applied sequentially given that the true null of no change would eventually be rejected with probability one (see Robbins (1970)). Starting with Bauer and Hackl (1978) a strand of literature has emerged that studies monitoring procedures that allow to detect structural change whenever new data arrives. Important contributions on this field are Chu et al. (1995), Kuan and Hornik (1995), Chu et al. (1996), Leisch et al. (2000), Altissimo and Corradi (2003), Zeileis et al. (2005), Andreou and Ghysels (2006) and Hsu (2007).

These papers contribute to the literature on monitoring structural stability on different levels ranging from theoretical contributions to detecting structural change in the conditional mean or the conditional variance or comparing different types of rejection regions for the null.
In this paper we use a monitoring approach based on moving sums of residuals and place it into a long memory framework. We develop a procedure to detect an increase in persistence for the case that the process becomes non-stationary. This is important because an increase in persistence implies a loss of controllability for important macroeconomic time series such as inflation rate or the European overnight rate (EONIA) (see Sibbertsen and Kruse (2009) and Hassler and Nautz (2008)). Further, a change in persistence also affects forecast accuracy in long memory time series (see Heinen et al. (2009)).
The rest of the paper is organized as follows: In section 6.2 we describe the test procedure we use and develop the asymptotic behavior. We further discuss and motivate different forms of boundary functions for the test. In section 6.3 we undertake a simulation study to asses the finite sample performance of the monitoring test. Section 6.4 contains an empirical application before section 6.5 concludes. All proofs are collected in the appendix 6.6.

6.2 Monitoring a change in persistence

We assume that the data generating process follows an ARFIMA(p,d,q) process as proposed by Granger and Joyeux (1980)

$$\Phi(L)(1-L)^d y_t = \Theta(L)\varepsilon_t, \text{ with } \varepsilon_t \overset{iid}{\sim} (0,\sigma^2) \text{ and } t = 1,2,\ldots,T\,. \tag{6.1}$$

The differencing parameter d can take fractional values but is assumed to be $|d| < \frac{1}{2}$. Thus the process y_t is in the stationary region (see e.g. Beran (1995)).
Bauer and Hackl (1978) propose the use of moving sums of cumulated residuals (MOSUM) to detect parameter changes in regression models. These tests are further investigated by Chu et al. (1995).
We are interested in detecting a change in persistence, i.e. a change in the fractional differencing parameter d, in the monitoring period $T+1$ up to $[T\tau]$, $\tau > 1$. Where $[\cdot]$ denotes the integer part of its argument.
In particular, we test the null of no change in persistence, i.e. $d = d_0$ within the monitoring period where $|d_0| < \frac{1}{2}$, against the alternative of an increase in persistence. More formally we test the null that

$$H_0 : d_\ell = d_0, \quad \ell = T+1,\ldots,[T\tau]\,, \tag{6.2}$$

against the alternative that at some point in the monitoring period the persistence increases and $\frac{1}{2} < d_\ell < \frac{3}{2}$. Thus we test whether the process stays in the stationary region throughout the whole monitoring period or changes into the non-stationary region with an infinite variance at some point in the monitoring period. For the period from $t = 1,\ldots,T$ we follow Chu et al. (1996)

and make the "noncontamination" assumption that

$$d_t = d_0, \quad t = 1, \ldots, T \ ,$$

with $|d_0| < \frac{1}{2}$. Consider for simplicity the case of an ARFIMA(0,d,0) process.
Let $\hat{e}_t$ be an ARFIMA(0,d,0) process as in (6.1) and $\hat{\sigma}^2 = T^{-1} \sum_{i=1}^{T} \hat{e}_i^2$ a consistent estimator of σ^2. Based on a moving sum of residuals obtained from a fixed window size $[Th]$, $0 < h \leq 1$, the prototypical MOSUM test reads

$$MS_{T,h,d} = \max_{T+1 \leq k \leq [T\tau]} \sigma^{-1} T^{-\frac{1}{2}-d} \left| \sum_{i=k-[Th]+1}^{k} \hat{e}_i - \frac{[Th]}{T} \sum_{i=1}^{T} \hat{e}_i \right| , \tag{6.3}$$

for each value k in the monitoring period $T+1$ through $[T\tau]$.
The next theorem gives the asymptotic behavior of the test statistic in (6.3) if y_t follows a long range dependent process as in (6.1) and (6.2).

Theorem 6.2.1.
Assume the process y_t follows an ARFIMA$(0,d,0)$ process as in (6.1) with $|d| < \frac{1}{2}$. Then, as $T \to \infty$, we have for $MS_{T,h,d}$ in (6.3) that

$$MS_{T,h,d} \Rightarrow \frac{1}{\sigma} \max_{t \in [1,\tau]} \left| BB^0(t,d) - BB^0(t-h,d) \right| ,$$

where $BB^0(t,d)$ denotes a fractional Brownian Bridge depending on fractional Brownian motion with parameter d. $\Rightarrow$ denotes weak convergence on a function space.
Under the alternative of a break in persistence the test is consistent.

The limiting distribution thus depends on the increments of a fractional Brownian bridge which in turn depends on the differencing parameter d of the data generating process. Therefore the asymptotic critical values of $MS_{T,h,d}$ are determined by the boundary crossing probabilities of the increments of a fractional Brownian bridge:

$$\mathbb{P}\{MS_{T,h,d} \leq b\} = \mathbb{P}\left\{ \left| BB^0(t,d) - BB^0(t-h,d) \right| \leq b \right\} . \tag{6.4}$$

The use of the test statistic in (6.3) is beneficial because the sequential application of usual CUSUM tests as in Sibbertsen and Kruse (2009) with constant critical values will eventually reject a correct null of no change in persistence with probability 1 (see Robbins (1970)).
Generally, every strictly increasing function $b(t) = zq(t)$ could serve as a boundary function where z is some suitable scaling factor and $q(t)$ is some monotonically increasing function in time. However if the boundary function grows too slowly the monitoring test will commit the type one error almost surely as it will detect a break in persistence with probability one. On the contrary if the boundary grows too quickly the test will loose power because a break in persistence cannot be detected anymore. For the short memory case a variety of different boundary functions have been proposed (see Andreou and Ghysels (2006, p. 92) for an overview). In particular Altissimo

and Corradi (2003) derive a boundary function based on the almost sure asymptotically uniform equicontinuity of the Brownian bridge obtaining an almost sure boundary function. This is convenient because it gives the rate of convergence with which the sequence of functions converges to a relatively compact set in the sense of an Arzelà-Ascoli theorem (see e.g. Davidson (1994, p. 335)). This provides useful information as we are interested in the behavior of the limiting distribution independently of the test statistic. We also derive almost sure results similar to the ones obtained by Altissimo and Corradi (2003) which are collected in the next theorem.

Theorem 6.2.2.
Let $BB^0(t,d) = B(t,d) - tB(1,d)$ *be a fractional Brownian bridge. Then,* $d_T^{-1}|BB^0(t,d)|$ *is almost surely asymptotically uniform equicontinuous in* $t \in [0,1]$. *With* $d_T := \sqrt{2T^{2H}\log\log(T)}$.

The use of this theorem is that it provides the rate with which the increment of the fractional brownian bridge becomes asymptotically uniform equicontinuous. In the proof this derived to be $\sqrt{2\log\log(T)}$. Hence, if we use this growth rate for the boundary function we will obtain a slowly growing function and therefore detect a change in persistence but at the same time the growth rate of this function is independent of the long memory parameter under the null d_0. Different forms of the boundary function are possible. For example one could use the boundary function

$$b_1(t) = z\sqrt{2t\log_2(t)}\,, \tag{6.5}$$

where $\log_2(t) := \log(\log(t))$. This boundary function is based on the law of iterated logarithm and is motivated by the fastest detection of change because it grows as slowly as possible. From theorem 6.2.2 we deduce the boundary function

$$b_2(t) = z\sqrt{2\log_2(t)}\,. \tag{6.6}$$

Because both boundary functions rely on the square root of a logarithm one needs to find a way to deal with values $\leq \log(1)$ to ensure real valued boundaries. One way of doing so is to define

$$\log_2'(t) := \begin{cases} 1 & \text{if } t \leq \exp(1) \\ \log\log(t) & \text{if } t > \exp(1)\,, \end{cases}$$

similar to Leisch et al. (2000). Another way which avoids the constant behavior of the boundary function at the beginning of the monitoring period is to define

$$\log_2''(t) := \begin{cases} t & \text{if } t \leq \exp(1) \\ \log\log(t) & \text{if } t > \exp(1)\,. \end{cases}$$

Formally this leads to four possible boundary functions

$$b_3(t) = z\sqrt{2t\log_2'(t)} \tag{6.7}$$

$$b_4(t) = z\sqrt{2t\log_2''(t)} \tag{6.8}$$

$$b_5(t) = z\sqrt{2\log_2'(t)} \tag{6.9}$$

$$b_6(t) = z\sqrt{2\log_2''(t)}\,. \tag{6.10}$$

One could think of different boundary functions such as functions that are dependent on the long memory parameter under the null to account for the gradually increasing variance of the process. However, unreported simulations showed that such a boundary function does not perform satisfactorily and we therefore restrict ourselves to the above boundary functions.

6.3 Monte Carlo evidence

We start by providing some Monte Carlo evidence on the small sample behavior of the usual MOSUM test as considered in Leisch et al. (2000) under long range dependence. Table 6.1 shows some of the simulation results.

	$\tau = 4$			$\tau = 6$			$\tau = 8$		
d	$h = 0.25$	$h = 0.5$	$h = 1$	$h = 0.25$	$h = 0.5$	$h = 1$	$h = 0.25$	$h = 0.5$	$h = 1$
0.1	66.22	57.26	51.08	70.04	62.80	53.72	72.52	67.12	57.02
0.2	98.06	96.38	91.90	99.28	97.94	94.90	99.76	98.98	96.86
0.3	99.96	99.90	99.64	100.00	100.00	99.96	100.00	100.00	99.96
0.4	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table 6.1: Empirical size of the fluctuation test by Leisch et al. (2000) [in %] for $T = 250$ and $\alpha = 5\%$.

As expected the generalized fluctuation test does not keep its size. Even if the long memory is only moderately present the test does not allow a secure conclusion whether a change in persistence is present or not because the boundary functions are too narrow.
In order to assess the finite sample performance of the monitoring procedure described in section 6.2 we consider different values for the long memory parameter $d = 0.1, 0.2, 0.3, 0.4$, the monitoring window $h = 0.25, 0.5, 0.75, 1$ and the out-of-sample monitoring period $\tau = 2, 4, 6, 8, 10$. We also consider different sample sizes of $T = 200, 250, 300$ and the different boundary functions $b_i(t)$, for $i = 3, \ldots, 6$, from (6.7) to (6.10) for the simulations. The number of Monte Carlo repetitions is set to $M = 10000$ and the levels of significance are set to $\alpha = 1\%, 5\%, 10\%$.[1]

[1] Some of the results here and in the sequel are unreported to save space but can be obtained from the authors on request.

Boundary function $b_3(t)$									
	$\tau = 4$			$\tau = 6$			$\tau = 8$		
d	$h = 0.5$	$h = 0.75$	$h = 1$	$h = 0.5$	$h = 0.75$	$h = 1$	$h = 0.5$	$h = 0.75$	$h = 1$
0.1	6.86	7.10	8.59	6.26	6.95	9.03	7.07	6.97	8.60
0.2	6.77	6.80	8.73	6.47	6.56	8.52	5.75	6.82	8.97
0.3	5.68	7.12	9.76	6.16	7.08	9.54	6.21	7.08	10.24
0.4	10.77	12.29	16.43	9.96	12.82	16.18	10.07	12.24	16.09

Table 6.2: Empirical size of the monitoring procedure [in %] for $T = 250$ and $\alpha = 5\%$.

Table 6.2 shows the size results for the boundary function motivated by the law of iterated logarithm. Using this boundary we obtain a procedure that is generally oversized. This overrejection of the correct null becomes more severe as the degree of persistence increases and/or the monitoring window h increases.

Table 6.3 displays the respective results based on the almost sure results from theorem 6.2.2. These results are more promising compared to the ones of boundary $b_3(t)$ as the nominal size level is better adhered to. Looking at the dependencies between the size, the long memory parameter d, the monitoring window h and the monitoring period τ we see that a moderate window size of $h = 0.5$ or $h = 0.75$ is generally preferable regardless of the monitoring period τ. If the persistence increases a reduced window size of $h = 0.5$ yields the most accurate size results. Reducing the window size even further to $h = 0.25$, however, leads to overrejection again as unreported results show.

As the boundary function $b_6(t)$ is only a slight modification of boundary function $b_5(t)$ the same argument as above applies to the results in table 6.4. The only difference is that the test overrejects somewhat when using boundary function $b_6(t)$.

Boundary function $b_5(t)$									
	$\tau = 4$			$\tau = 6$			$\tau = 8$		
d	$h = 0.5$	$h = 0.75$	$h = 1$	$h = 0.5$	$h = 0.75$	$h = 1$	$h = 0.5$	$h = 0.75$	$h = 1$
0.1	7.55	7.02	7.09	7.12	6.16	6.69	7.30	6.56	6.29
0.2	6.76	5.85	6.56	6.64	5.85	5.60	5.90	5.66	5.23
0.3	5.15	4.87	4.87	5.08	4.28	4.29	5.16	3.89	3.92
0.4	5.61	5.10	5.78	4.86	4.12	4.28	4.32	3.63	3.89

Table 6.3: Empirical size of the monitoring procedure [in %] for $T = 250$ and $\alpha = 5\%$.

Boundary function $b_6(t)$									
	$\tau = 4$			$\tau = 6$			$\tau = 8$		
d	$h = 0.5$	$h = 0.75$	$h = 1$	$h = 0.5$	$h = 0.75$	$h = 1$	$h = 0.5$	$h = 0.75$	$h = 1$
0.1	7.00	7.52	8.65	6.75	6.71	8.42	7.34	7.05	7.59
0.2	7.13	6.92	8.60	6.69	6.31	7.33	6.02	6.20	6.70
0.3	6.12	6.96	8.94	5.96	5.86	7.26	5.84	5.58	6.49
0.4	10.26	11.22	13.96	8.39	9.61	10.83	7.67	8.06	9.53

Table 6.4: Empirical size of the monitoring procedure [in %] for $T = 250$ and $\alpha = 5\%$.

The size results for the $\alpha = 10\%$ level are unreported but show the same general behavior of the previously discussed results. However, in this setting it becomes even more obvious that the boundary function $b_5(t)$ yields the best performance over all considered settings.
Generally the size distortions are minor and acceptable and also comparable to the short memory case as reported in Leisch et al. (2000).
In an empirical setting the long memory parameter d_0 is unknown and has to be estimated. We therefore conduct the size experiment again but this time using an estimated d_0. Generally every consistent estimation method is applicable but estimators that converge faster than the asymptotic distribution to the true value of d_0 are preferable. One such estimator is the approximate maximum likelihood estimator proposed by Beran (1995) which is $\sqrt{T}$ consistent. Another popular method to estimate d_0 is the log-periodogram regression (see Geweke and Porter-Hudak (1983)). The rate of convergence of this estimator is $\sqrt{m}$ where m is the number of frequencies used. The estimator is consistent as long as $(m\log(m))/n \to 0$ as $m, n \to \infty$, with n being the sample size (see Hurvich et al. (1998)). In our simulations we use this estimator with $T^{4/5}$ frequencies. The results are reported for the $\alpha = 5\%$ level in table 6.5.

Boundary function $b_5(t)$									
	$\tau = 4$			$\tau = 6$			$\tau = 8$		
d	$h = 0.5$	$h = 0.75$	$h = 1$	$h = 0.5$	$h = 0.75$	$h = 1$	$h = 0.5$	$h = 0.75$	$h = 1$
0.1	8.68	7.62	8.50	8.62	7.14	7.48	7.72	7.20	6.42
0.2	7.12	6.62	6.02	7.44	6.08	5.82	6.36	5.86	5.42
0.3	5.74	4.98	5.26	5.14	4.28	4.40	4.26	4.24	4.04
0.4	5.36	4.86	5.38	4.32	4.36	4.34	3.74	3.42	4.04

Table 6.5: Empirical size of the monitoring procedure with estimated d_0.

We observe small size distortions for smaller values of d_0 and larger monitoring periods but generally the size is well kept even if we estimate the long memory parameter.
When the persistence changes from stationary to non-stationary the MOSUM test will eventually detect this with probability one due to consistency (see theorem 6.2.1).[2] Therefore it is more interesting how fast a change in persistence can be detected.

[2]This has also been confirmed in unreported simulations.

To study the detection delay we consider breaks from the stationary region, namely $d_0 = 0.1, 0.2, 0.3, 0.4$, to the non-stationary region, $d_1 = 0.6, 0.7, 0.8, 0.9, 1$. The break occurs within the monitoring period at $t^* = [\rho\tau T]$, where $\rho = 0.3, 0.5, 0.7$ and $\tau = 2, 4, 6, 8, 10$ as above and $[\cdot]$ denotes the integer part of its argument. We use a sample size of $T = 250$ and the boundary functions $b_i(t)$, for $i = 3, \ldots, 6$, from (6.7) to (6.10). As an example the average detection delay for the $\alpha = 5\%$ level for the boundary function $b_5(t)$ for different breaks is displayed in tables 6.6, 6.7 and 6.8.

Boundary function $b_5(t)$									
$\tau = 2$		$h = 0.25$			$h = 0.5$			$h = 0.75$	
d_0	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$
0.1	53.33	43.40	37.23	77.61	63.15	49.58	91.34	72.67	56.20
0.2	79.23	63.41	54.45	112.58	95.82	81.91	126.86	112.28	95.50
0.3	112.44	92.12	76.31	144.79	137.08	122.53	158.95	156.00	148.96
0.4	133.96	118.80	99.73	150.59	158.25	151.10	157.33	172.92	172.75
$\tau = 4$		$h = 0.25$			$h = 0.5$			$h = 0.75$	
d_0	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$
0.1	57.89	46.73	38.50	91.25	71.51	57.13	116.31	90.09	69.96
0.2	94.31	72.12	59.49	149.92	118.26	98.78	186.03	152.03	127.12
0.3	160.65	109.40	86.93	222.47	183.72	152.57	265.43	240.25	210.58
0.4	225.64	167.07	121.51	272.32	258.26	214.56	293.90	309.68	282.91
$\tau = 6$		$h = 0.25$			$h = 0.5$			$h = 0.75$	
d_0	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$
0.1	53.52	41.29	32.36	96.13	71.78	56.47	124.22	94.77	72.26
0.2	98.55	69.55	55.80	167.16	124.39	104.44	217.05	169.95	138.28
0.3	190.16	111.91	85.96	282.65	204.74	165.30	339.06	276.86	231.03
0.4	315.36	192.13	126.20	380.04	315.26	239.72	419.28	396.05	320.28

Table 6.6: Average detection delay of the monitoring procedure for $T = 250$, $\alpha = 5\%$ and $\rho = 0.3$.

Table 6.6 shows the results for the case of an early break within the monitoring period. As one expects the detection is easier and therefore faster if the difference between d_0 and d_1 is large. Consequently the detection delay is rather small if the persistence changes from stationary, say $d_0 = 0.2$, long memory to non-stationary, say $d_1 = 0.8$, and even faster if the process becomes a unit root process after the break. In fact, the detection delay for larger breaks is comparable with the short memory case (see table 3 in Leisch et al. (2000)). This is encouraging given the well known slow rate of convergence in long memory time series. Another result is that it is easier and faster to detect a change in persistence if the width of the monitoring window $[Th]$ is rather small. Detection delays for values of $h = 0.25$ and $h = 0.5$ are generally smaller compared to larger values of h. This is also in line with the findings of Leisch et al. (2000) for the short memory case. It is well known also in related areas of the structural change literature (see e.g Pesaran and Timmermann (2005) for results regarding forecasts under structural breaks)

that smaller windows of data are usually better to detect and deal with structural change. The results for later breaks within the monitoring period are shown in tables 6.7 and 6.8. The general conclusions from above remain valid but the detection delay becomes even smaller if the breaks occurs later. This is also a similar behavior to the short memory case reported in Leisch et al. (2000).

Boundary function $b_5(t)$									
$\tau = 2$		$h = 0.25$			$h = 0.5$			$h = 0.75$	
d_0	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$
0.1	44.55	35.39	28.67	65.85	54.45	43.62	74.23	60.25	48.20
0.2	65.49	55.62	47.31	89.16	82.66	72.46	96.55	91.89	81.20
0.3	85.56	78.49	68.26	110.59	112.02	106.93	118.37	123.46	118.71
0.4	89.62	92.35	84.78	99.62	115.70	117.50	95.06	110.72	125.06
$\tau = 4$		$h = 0.25$			$h = 0.5$			$h = 0.75$	
d_0	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$
0.1	51.01	35.76	29.86	88.67	68.46	52.90	113.94	87.37	67.87
0.2	88.44	64.41	51.54	141.47	116.89	98.79	172.76	150.65	129.69
0.3	143.60	108.29	82.25	190.48	174.08	152.71	223.04	217.80	204.48
0.4	169.54	153.31	114.90	204.31	218.35	201.99	225.85	250.99	241.94
$\tau = 6$		$h = 0.25$			$h = 0.5$			$h = 0.75$	
d_0	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$
0.1	36.59	24.66	16.15	87.80	65.26	47.83	122.35	90.52	67.80
0.2	86.31	58.32	41.59	158.73	120.80	99.18	204.27	169.67	140.36
0.3	174.06	106.13	74.87	242.96	201.27	163.20	289.19	264.62	229.60
0.4	235.73	181.23	121.38	287.28	279.41	231.81	303.02	332.77	302.43

Table 6.7: Average detection delay of the monitoring procedure for $T = 250$, $\alpha = 5\%$ and $\rho = 0.5$.

Boundary function $b_5(t)$									
$\tau = 2$		$h = 0.25$			$h = 0.5$			$h = 0.75$	
d_0	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$
0.1	28.02	24.45	18.84	39.55	35.59	30.44	41.95	38.81	31.69
0.2	39.24	38.55	33.06	53.02	52.25	50.44	54.65	53.90	52.27
0.3	45.85	53.02	50.46	61.29	69.01	69.73	50.52	60.77	69.68
0.4	27.52	44.96	50.19	21.71	38.06	52.93	-77.72	-86.60	-61.95
$\tau = 4$		$h = 0.25$			$h = 0.5$			$h = 0.75$	
d_0	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$
0.1	38.81	24.41	16.72	69.65	58.49	44.95	86.56	79.32	63.26
0.2	66.51	51.73	40.02	101.49	101.71	92.10	121.38	117.65	112.30
0.3	90.24	82.78	69.57	124.50	133.20	131.38	146.83	156.16	159.54
0.4	77.71	100.90	92.53	96.81	138.45	144.58	93.48	140.91	166.83
$\tau = 6$		$h = 0.25$			$h = 0.5$			$h = 0.75$	
d_0	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$	$d_1 = 0.6$	$d_1 = 0.8$	$d_1 = 1$
0.1	12.08	4.92	-5.71	68.34	50.98	33.94	99.05	76.97	55.14
0.2	63.08	37.87	20.33	120.52	103.78	87.11	155.37	142.54	128.07
0.3	104.40	85.53	57.51	163.12	161.03	145.76	188.54	197.29	196.56
0.4	110.33	132.11	96.32	139.56	185.50	185.40	148.29	215.15	222.53

Table 6.8: Average detection delay of the monitoring procedure for $T = 250$, $\alpha = 5\%$ and $\rho = 0.7$.

6.4 Empirical Application

To illustrate the use of the monitoring approach we analyze monthly US price inflation series from Stock and Watson (2005).[3] In particular we consider the first difference of the logarithmic implied price deflator for durable goods. This series has also been under investigation from Cavaliere and Taylor (2008) who report a change in persistence from $I(0)$ to $I(1)$. However, they did not consider the possibility of fractional integration in the series although inflation related time series are likely to show long memory behavior (see e.g. Hassler and Wolters (1995)). The sample spans from 01/1959 to 12/2003. The series is depicted in figure 6.1.

[3]The data is available at Mark Watson's website at: http://www.princeton.edu/~mwatson/wp.html.

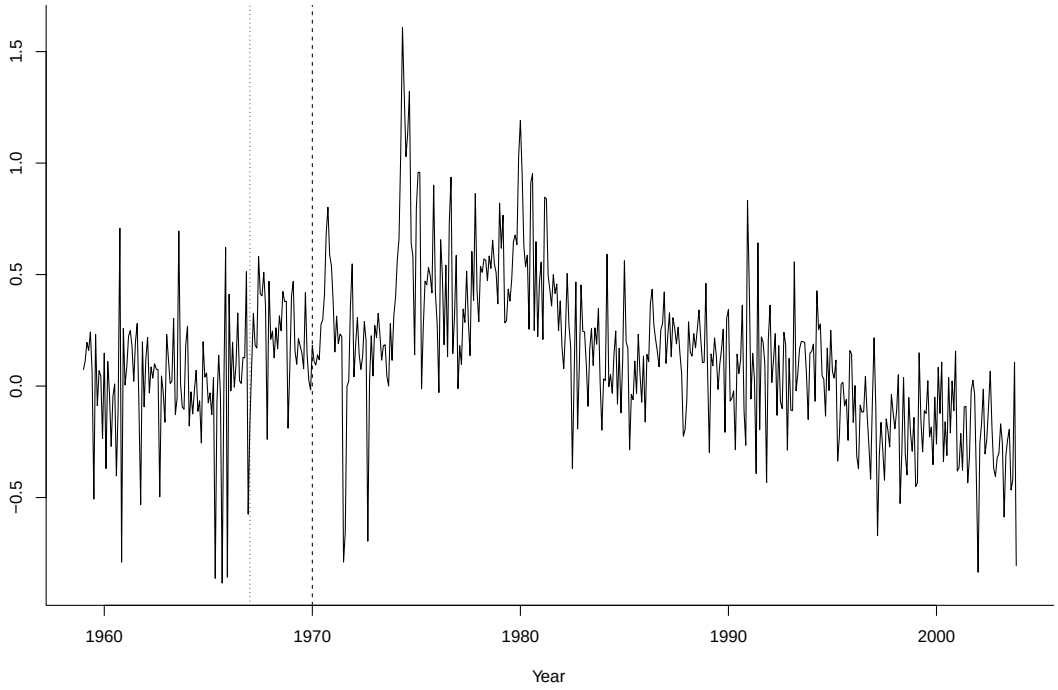

Figure 6.1: First difference of logarithmic price deflator for durable goods.

To determine the value of the long memory parameter we use log-periodogram regression as proposed by Geweke and Porter-Hudak (1983). The decision of how many frequencies should be used in the regression is a trade-off between reducing the bias and reducing the asymptotic variance. We use $T^{1/2}$ frequencies to deal with potential short memory components in the data (see e.g. Agiakloglou et al. (1993)). For the whole sample this yields an estimate of $\hat{d} = 0.61$. This value is highly significant as judged by its p-value which is $< 1e-03$.

To test whether a change in persistence can be detected in the data we apply the CUSUM of squares test for a change in persistence proposed by Sibbertsen and Kruse (2009) to the whole sample. This leads to a test statistic of $R = 0.0373$ which is significant at the $\alpha = 5\%$ level in favor of an increasing persistence. The estimated breakpoint is at $t^* = 107$ which is 11/1967 (the dotted line in figure 6.1).

To use the monitoring approach we split the sample in an in-sample part ranging from 01/1959 to 12/1965 and leave the rest as monitoring period. This yields a $\tau \approx 5$. The estimated d_0 within the in-sample period is $\hat{d}_0 = 0.23$.

For the application of the MOSUM test we use the boundary function $b_5(t)$ and set $h = 0.5$. The first time the sequence of test statistics exceeds the $\alpha = 1\%$ boundary function is at $t = 55$ in the monitoring period. This is equivalent to an estimated breakpoint at $t^* = 139$ which is 06/1970 (the dashed line in figure 6.1). The first time the sequence of test statistics exceeds the $\alpha = 5\%$ and $\alpha = 10\%$ boundary functions is only one period earlier.

The estimation of d_1 in the monitoring period yields $\hat{d}_1 = 0.68$. Thus we can confirm a change in persistence with high probability from stationary long memory to non-stationary long memory. Notably the detection delay is rather short and we obtain a fast indication of the change in persistence from using the monitoring procedure.

6.5 Conclusion

Detecting a change in persistence as soon as possible is of paramount interest because structural change affects the subsequent analysis of the data heavily. The usual approach is to use one-shot tests to detect a change in persistence a posteriori. However, these tests cannot be applied sequentially because a correct null of no change would eventually be rejected with probability one. We propose a monitoring procedure based on moving sums that allows to detect a change in the long memory parameter of a long range dependent time series whenever new data arrives. By means of a Monte Carlo experiment we show good size properties and also study the detection delay when a change in persistence occurs. Depending on the width of the monitoring window and the difference between the pre- and post-break long memory parameter the detection is rather fast. Smaller monitoring windows generally prove more useful to detect a change in persistence early and also larger differences between the long memory parameters are detected faster.

In an empirical illustration of the method we are able to confirm a change in persistence from stationary to non-stationary long memory in an inflation time series.

6.6 Appendix

6.6.1 Proof of Theorem 6.2.1

First, let $k = [Tt]$ for each value in the monitoring period then write the test statistic as

$$\begin{aligned} MS_{T,h,d} &= \max_{T+1\le k\le [T\tau]} \sigma^{-1} T^{-\frac{1}{2}-d} \left| \sum_{i=k-[Th]+1}^{k} \hat{e}_i - \frac{[Th]}{T} \sum_{i=1}^{T} \hat{e}_i \right| \\ &= \max_{T+1\le [Tt]\le [T\tau]} \sigma^{-1} T^{-\frac{1}{2}-d} \left| \sum_{i=1}^{[Tt]} \hat{e}_i - \frac{[Tt]}{T} \sum_{i=1}^{T} \hat{e}_i - \sum_{i=1}^{[Tt]-[Th]} \hat{e}_i + \frac{[Tt]-[Th]}{T} \sum_{i=1}^{T} \hat{e}_i \right| . \end{aligned}$$

Then using the FCLT for fractionally integrated processes (see Sowell (1990) and Davidson and de Jong (2000)) and the continuous mapping theorem (CMT) we have

$$\begin{aligned} MS_{T,h,d} &\Rightarrow \max_{T+1\le [Tt]\le [T\tau]} \sigma^{-1} \left| B(t,d) - tB(1,d) - B(t-h,d) + (t-h)B(1,d) \right| \\ &= \max_{T+1\le [Tt]\le [T\tau]} \sigma^{-1} \left| BB^0(t,d) - [B(t-h,d) - (t-h)B(1,d)] \right| \\ &= \max_{T+1\le [Tt]\le [T\tau]} \sigma^{-1} \left| BB^0(t,d) - BB^0(t-h,d) \right| , \end{aligned}$$

where $BB^0(t,d)$ denotes a fractional Brownian bridge.
To prove consistency we consider that at some point in the monitoring period, say k^*, the persistence changes from stationary long memory with $0 < d_0 < \frac{1}{2}$ to non-stationary long memory with $\frac{1}{2} < d_1 < \frac{3}{2}$ and then split the test statistic into its stationary and non-stationary parts. We write the test statistic as

$$\begin{aligned} MS_{T,h,d_0} &= \max_{T+1\le k\le [T\tau]} \sigma^{-1} T^{-\frac{1}{2}-d_0} \left| \sum_{i=k-[Th]+1}^{k} \hat{e}_i - \frac{[Th]}{T} \sum_{i=1}^{T} \hat{e}_i \right| \\ &= \max_{T+1\le [rT]\le [T\tau]} \sigma^{-1} T^{-\frac{1}{2}-d_0} \left| \underbrace{\sum_{i=1}^{[rT]} \hat{e}_i}_{I} - \underbrace{\sum_{i=1}^{[rT]-[hT]} \hat{e}_i}_{II} - \frac{[Th]}{T} \underbrace{\sum_{i=1}^{T} \hat{e}_i}_{III} \right| , \end{aligned}$$

where $k = [rT]$ for some $r > 1$. Part III only contains $I(d_0)$ variables due to the noncontamination assumption.
We have to distinguish two cases:

(i) $k^* \le [rT] - [Th] \Rightarrow$ in this case both I and II contain $I(d_1)$ variables

(ii) $[rT] - [Th] \le k^* \le [rT] \Rightarrow$ in this case only I contains $I(d_1)$ variables.

Ad (i):
The case (i) is depicted in figure 6.2 where $[rT]$ is denoted by k_1 and $[rT]-[Th]$ is denoted by k_0. The gray shaded area is the monitoring window.

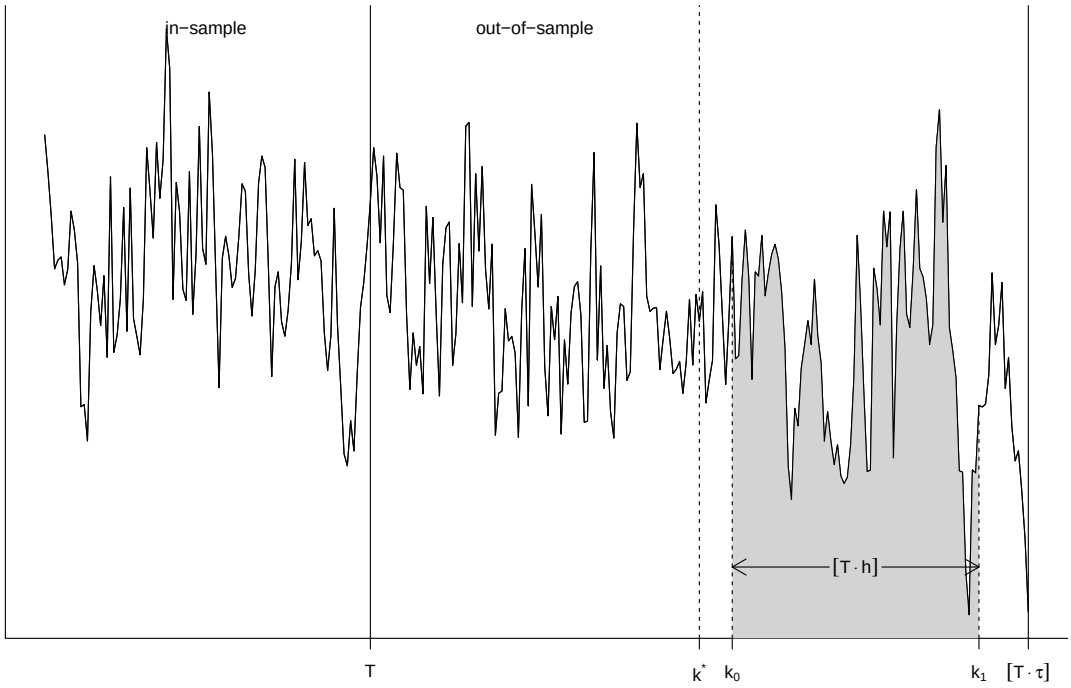

Figure 6.2: MOSUM case (i).

Write the test statistic as

$$\begin{aligned} MS_{T,h,d_0} &= \max_{T+1\le[rT]\le[T\tau]} \sigma^{-1}T^{-\frac{1}{2}-d_0}\left|\sum_{i=1}^{k^*}\hat{e}_i+\sum_{i=k^*+1}^{[rT]}\hat{e}_i-\sum_{i=1}^{k^*}\hat{e}_i-\sum_{i=k*+1}^{[rT]-[hT]}\hat{e}_i-\frac{[hT]}{T}\sum_{i=1}^{T}\hat{e}_i\right| \\ &= \max_{T+1\le[rT]\le[T\tau]} \sigma^{-1}T^{-\frac{1}{2}-d_0}\left|-\frac{[hT]}{T}\sum_{i=1}^{T}\hat{e}_i\right|+\max_{T+1\le[rT]\le[T\tau]} \sigma^{-1}T^{-\frac{1}{2}-d_0}\left|\sum_{i=[rT]-[hT]}^{[rT]}\hat{e}_i\right|. \end{aligned}$$

Now, the first part is $I(d_0)$ and is correctly standardized. Therefore, using the arguments from above it converges to $-hB(1,d_0)$ which is the standard deviation of the fractional Brownian motion. For the second part the standardization is obtained from d_0 but the variables are $I(d_1)$ and so the expression diverges and we obtain

$$MS_{T,h,d_0} = o_p(1)+O_p\left(T^{d_1-d_0}\right). \tag{6.11}$$

Ad (ii):
The situation (ii) is depicted in figure 6.3.

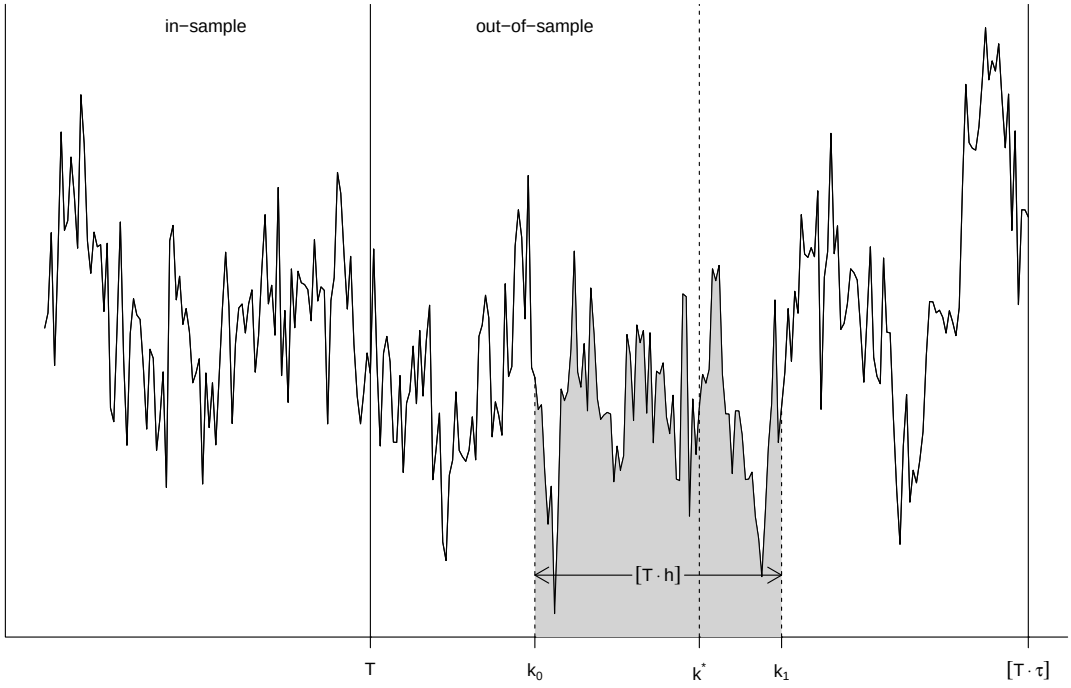

Figure 6.3: MOSUM case (ii).

Now only I contains $I(d_1)$ variables. Write the test statistic as

$$\begin{aligned}
MS_{T,h,d_0} &= \max_{T+1\le[rT]\le[T\tau]} \sigma^{-1}T^{-\frac{1}{2}-d_0}\left|\sum_{i=1}^{[rT]}\hat{e}_i - \sum_{i=1}^{[rT]-[hT]}\hat{e}_i - \frac{[hT]}{T}\sum_{i=1}^{T}\hat{e}_i\right| \\
&= \max_{T+1\le[rT]\le[T\tau]} \sigma^{-1}T^{-\frac{1}{2}-d_0}\left|\sum_{i=1}^{k^*}\hat{e}_i + \sum_{i=k^*+1}^{[rT]}\hat{e}_i - \sum_{i=1}^{[rT]-[hT]}\hat{e}_i - \frac{[hT]}{T}\sum_{i=1}^{T}\hat{e}_i\right| \\
&= \max_{T+1\le[rT]\le[T\tau]} \sigma^{-1}T^{-\frac{1}{2}-d_0}\left|\sum_{i=k^*+1}^{[rT]}\hat{e}_i\right| + \max_{T+1\le[rT]\le[T\tau]} \sigma^{-1}T^{-\frac{1}{2}-d_0}\left|\sum_{i=[rT]-[hT]}^{k^*}\hat{e}_i\right| \\
&\quad - \max_{T+1\le[rT]\le[T\tau]} \sigma^{-1}T^{-\frac{1}{2}-d_0}\left|\frac{[hT]}{T}\sum_{i=1}^{T}\hat{e}_i\right| .
\end{aligned}$$

With the arguments from case (i) we obtain

$$MS_{T,h,d_0} = O_p\left(T^{d_1-d_0}\right) + o_p(1) + o_p(1)\,, \tag{6.12}$$

where the second part of the above expression does not expand with T anymore and therefore vanishes as $T \to \infty$. □

6.6.2 Proof of Theorem 6.2.2

Denote by $d_T \coloneqq \sqrt{2T^{2d+1}\log\log(T)}$. By the reverse triangle inequality we have for some $r \in [0,1]$

$$d_T^{-1}\left|B(Tr,d) - rB(T,d) - (B(Tr',d) - r'B(T,d))\right| \leq d_T^{-1}\left|B(Tr,d) - B(Tr',d)\right| + d_T^{-1}\left|(r-r')B(T,d)\right| ,$$

for distinct values r and r'. Using the notation from Altissimo and Corradi (2003, p. 232) we write $S(r,\delta) = (r' : |r-r'| \leq \delta)$. Now, by the fact that (see Davidson (1994, p. 335 ff.))

$$\sup_{\theta\in\Theta}\sup_{\theta'\in S(\theta,\delta)}\left|f_n(\theta') - f_n(\theta)\right| \leq 2\sup_{\theta\in\Theta}|f_n(\theta)|$$

and the LIL for the fractional Brownian motion (see e.g. Taqqu (1977)) we have for the second part of the right side

$$\limsup_{T\to\infty}\sup_{r\in[0,1]}\sup_{r'\in S(r,\delta)} d_T^{-1}\left|(r-r')B(T,d)\right| \leq 2\delta\sigma ,$$

with σ the variance of the fractional Brownian Motion. As $\delta \to 0$ the whole part approaches zero which ensures the asymptotic uniform equicontinuity almost surely.
For the first part of the right hand side we have by self-similarity

$$\begin{aligned}\limsup_{T\to\infty}\sup_{r\in[0,1]}\sup_{r'\in S(r,\delta)} d_T^{-1}\left|B(Tr) - B(Tr')\right| &= \limsup_{T\to\infty}\sup_{r\in[0,1]}\sup_{r'\in S(r,\delta)} d_T^{-1}\left|T^{d+1/2}B(r) - T^{d+1/2}B(r')\right| \\ &= \limsup_{T\to\infty}\sup_{r\in[0,1]}\sup_{r'\in S(r,\delta)} T^{d+1/2}d_T^{-1}\left|B(r) - B(r')\right| .\end{aligned}$$

Now note that

$$d_T = \sqrt{2T^{2d+1}\log\log(T)} = \sqrt{T^{2d+1}}\sqrt{2\log\log(T)} = T^{d+1/2}\sqrt{2\log\log(T)} .$$

Therefore we obtain

$$\limsup_{T\to\infty}(2\log\log(T))^{-\frac{1}{2}}\sup_{r\in[0,1]}\sup_{r'\in S(r,\delta)}\left|B(r) - B(r')\right| .$$

Because $|B(r) - B(r')|$ is almost surely Hölder continuous of order strictly less than H (see Biagini et al. (2008, p. 11)) and $\limsup_{T\to\infty}(2\log\log(T))^{-\frac{1}{2}}$ tends to zero as $T \to \infty$ it follows that the above expression is almost surely asymptotically uniform equicontinuous. □

Bibliography

Adler, M. and Lehmann, B. (1983). Deviations from the purchasing power parity in the long run. *Journal of Finance*, 38:1471–1487.

Agiakloglou, C., Newbold, P., and Wohar, M. (1993). Bias of an estimator of the fractional difference parameter. *Journal of Time Series Analysis*, 14:235–246.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions in Automatic Control*, 19:716–723.

Altissimo, F. and Corradi, V. (2003). Strong rules for detecting the number of breaks in a time series. *Journal of Econometrics*, 117:207–244.

Andreou, E. and Ghysels, E. (2006). Monitoring disruptions in financial markets. *Journal of Econometrics*, 135:77–124.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61:821–856.

Andrews, D. W. K. and Fair, R. C. (1988). Inference in nonlinear econometric models wiht structural change. *The Review of Economic Studies*, 55:615–639.

Aneuryn-Evans, G. and Deaton, A. (1980). Testing linear versus logarithmic regression models. *The Review of Economic Studies*, 47:275–291.

Bacon, D. W. and Watts, D. G. (1971). Estimating the transition between two intersecting straight lines. *Biometrika*, 58:525–534.

Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, 73:5–59.

Banerjee, A., Lumsdaine, R., and Stock, J. (1992). Recursive and sequential tests of the unit root and trend break hypothesis: Theory and international evidence. *Journal of Business & Economic Statistics*, 10:271–288.

Bauer, P. and Hackl, P. (1978). The use of MOSUMs for quality control. *Technometrics*, 20:431–436.

Behnen, K. and Neuhaus, G. (1995). *Grundkurs Stochastik*. Teubner.

Beran, J. (1995). Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *Journal of the Royal Statistical Society. Series B*, 57:659–672.

Beran, J. and Terrin, N. (1996). Testing for a change of the long-memory parameter. *Biometrika*, 83:627–638.

Bergman, U. M. and Hansson, J. (2005). Real exchange rates and regime switching regimes. *Journal of International Money and Finance*, 24:121–138.

Berndt, E. K., Hall, B. H., Hall, R. E., and Hausman, J. A. (1974). Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement*, 3:653–665.

Biagini, F., Hu, Y., Øksendahl, B., and Zhang, T. (2008). *Stochastic calculus for fractional brownian motion and applications*. Springer, New York.

Bollerslev, T. and Wooldridge, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, 11:143–172.

Breusch, T. S. (1978). Testing for autocorrelation in dynamic linear models. *Australian Economic Papers*, 17:334–355.

Busetti, F. and Taylor, A. M. R. (2004). Tests of stationarity against a change in persistence. *Journal of Econometrics*, 123:33–66.

Caporale, G. M., Pittis, N., and Sakellis, P. (2003). Testing for PPP: The erratic behaviour of unit root tests. *Economics Letters*, 80:277–284.

Catão, L. A. V. and Terrones, M. E. (2005). Fiscal deficits and inflation. *Journal of Monetary Economcis*, 52:529–554.

Cavaliere, G. and Taylor, A. M. R. (2008). Testing for a change in persistence in the presence of non-stationary volatility. *Journal of Econometrics*, 147:84–98.

Chen, R. and Tsay, R. S. (1993). Functional-coefficient autoregressive models. *Journal of the American Statistical Association*, 88:298–308.

Chen, Y. T. (2003). Discriminating between competing STAR models. *Economics Letters*, 79:161–167.

Chen, Y. T. and Kuan, C. M. (2002). The peusdo-true score encompassing test for non-nested hypotheses. *Journal of Econometrics*, 106:271–295.

Chen, Y. T. and Kuan, C. M. (2007). Corrigendum to "The peusdo-true score encompassing test for non-nested hypotheses". *Journal of Econometrics*, 141:1412–1417.

Chortareas, G. and Kapetanios, G. (2004). The Yen real exchange rate may be stationary after all: Evidence from non-linear unit-root tests. *Oxford Bulletin of Economics and Statistics*, 66:113–131.

Chortareas, G., Kapetanios, G., and Shin, Y. (2002). Nonlinear mean reversion in real exchange rates. *Economics Letters*, 77:411–417.

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

Chu, C., Hornik, K., and Kuan, C. (1995). MOSUM tests for parameter constancy. *Biometrika*, 82:603–617.

Chu, C., Stinchcombe, M., and White, H. (1996). Monitoring structural change. *Econometrica*, 64:1045–1065.

Clements, M. P. and Hendry, D. F. (1998). *Forecasting economic time series*. Cambridge University Press.

Clements, M. P. and Smith, J. (1997). The performance of alternative forecasting methods for setar models. *International Journal of Forecasting*, 13:463–475.

Coleman, A. (1995). Arbitrage, storage and the law of one price: New theory for the time series analysis of an old problem. Working Paper Princeton University.

Corsi, F., Mittnik, S., Pigorsch, C., and Pigorsch, U. (2008). The volatility of realized volatility. *Econometric Reviews*, 27:46–78.

Coulibaly, N. and Brorsen, B. W. (1999). Monte carlo sampling approach to testing nonnested hypotheses: Monte carlo results. *Econometric Reviews*, 18:195–209.

Cox, D. R. (1961). Tests of seperate families of hypotheses. In *Proceedings of the Fourth Berkley Symposium on Mathematical Statistics and Probability*, volume 1, pages 105–123. University of California Press.

Cox, D. R. (1962). Further results on tests of seperate families of hypotheses. *Journal of the Royal Statistical Society. Series B*, 24:406–424.

Davidson, J. (1994). *Stochastic limit theory*. Oxford University Press.

Davidson, J. and de Jong, R. (2000). The functional central limit theorem and weak convergence to stochastic integrals II: Fractionally integrated processes. *Econometric Theory*, 16:643–666.

Davies, L. and Krämer, W. (2003). The Dickey-Fuller test for exponential random walks. *Econometric Theory*, 19:865–877.

de Jong, R. (2010). Exponentials of unit root processes. Working Paper Ohio State University.

de Jong, R. M. (2004). Addendum to "Asymptotics for nonlinear transformations of integrated time series". *Econometric Theory*, 20:627–635.

de Jong, R. M. and Schmidt, P. (2002). Spurious logarithms and the KPSS statistic. *Economics Letters*, 76:383–391.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74:427–431.

Dornbusch, R. (1976). Expectations and exchange rate dynamics. *The Journal of Political Economy*, 84:1161–1176.

Edison, H. J. and Klovland, J. T. (1987). A quantitative reassessment of the purchasing power parity hypotheses: Evidence from Norway and the United Kingdom. *Journal of Applied Econometrics*, 2:309–333.

Eitrheim, Ø. and Teräsvirta, T. (1996). Testing the adequacy of smooth transition autoregressive models. *Journal of Econometrics*, 74:59–75.

Engel, C. and Hamilton, J. D. (1990). Long swings in the Dollar: Are they in the data and do markets know it? *The American Economic Review*, 80:689–713.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50:987–1007.

Engle, R. F. (1984). Wald, Likelihood ratio and Lagrange multiplier tests in econometrics. In Griliches, Z. and Intriligator, M. D., editors, *Handbook of Econometrics*, volume 2. Elsevier Science Publisher B.V.

Escribano, A. and Jordá, O. (2001). Testing nonlinearity: Decision rules for selecting between logistic and exponential STAR models. *Spanish Economic Review*, 3:193–209.

Francq, C. and Zakoïan, J. M. (2001). Stationarity of mulitvariate Markov-switching ARMA models. *Journal of Econometrics*, 102:339–364.

Franses, P. H. and van Dijk, D. (2000). *Non-linear time series models in empirical finance*. Cambridge University Press.

Gallant, A. R. and White, H. (1988). *A unified theory of estimation and inference for nonlinear dynamic models*. Basil Blackwell.

Garcia, R. and Perron, P. (1996). An analysis of the real interest rate under regime shifts. *The Review of Economics and Statistics*, 78:111–125.

Gatti, D. D., Gallegati, M., and Mignacca, D. (1998). Nonlinear dynamics and european GNP data. *Studies in Nonlinear Dynamics and Econometrics*, 3:43–59.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long-memory time series models. *Journal of Time Series Analysis*, 4:221–238.

Ghysel, E., Guay, A., and Hall, A. (1997). Predictive tests for structural change with unknown breakpoint. *Journal of Econometrics*, 82:209–233.

Godfrey, L. G. (1978). Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variabeles. *Econometrica*, 46:1303–1310.

Godfrey, L. G. (1988). *Misspecification tests in econometrics - The lagrange multiplier principle and other approaches*. Cambridge University Press.

Godfrey, L. G. (2007). On the asymptotic validity of a bootstrap method for testing nonnested hypotheses. *Economics Letters*, 94:408–413.

Godfrey, L. G. and Santos Silva, J. M. C. (2005). Bootstrap tests of nonnested hypotheses: Some further results. *Econometric Reviews*, 23:325–340.

Gourieroux, C. and Monfort, A. (1994). Testing non-nested hypotheses. In Engle, R. F. and McFadden, D. L., editors, *Handbook of Econometrics*, volume 4. Elsevier Science Publisher B.V.

Granger, C. and Hallman, J. (1991). Nonlinear transformations of integrated time series. *Journal of Time Series Analysis*, 12:207–224.

Granger, C. W. J. and Joyeux, R. (1980). An introduction to long-range time series models and fractional integration. *Journal of Time Series Analysis*, 1:15–30.

Granger, C. W. J. and Teräsvirta, T. (1993). *Modelling nonlinear economic relationships*. Oxford University Press.

Haggan, V. and Ozaki, T. (1981). Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model. *Biometrika*, 68:189–196.

Hall, P. (1992). *The Bootstrap and Edgeworth expansion*. Springer, New York.

Hall, P. and Titterington, D. M. (1989). The effect of simulation order on level accuracy and power of monte carlo tests. *Journal of the Royal Statistical Society. Series B*, 51:459–467.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57:357–384.

Hamilton, J. D. (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics*, 45:39–70.

Hamilton, J. D. (1993). Estimation, inference and forecasting of time series subject to changes in regime. In Maddala, G., Rao, C., and Vinod, H., editors, *Handbook of Statistics*, volume 11. Elsevier Science Publisher B.V.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton, New Jersey.

Hamilton, J. D. and Raj, B. (2002). New directions in business cycle research and financial analysis. *Empirical Economics*, 27:149–162.

Hansen, B. E. (1992). Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory*, 8:489–500.

Hansen, B. E. (2001). The new econometrics of structural change: Dating breaks in U.S. labor productivity. *The Journal of Economic Perspectives*, 15:117–128.

Hassler, U. and Nautz, D. (2008). On the persistence of the Eonia spread. *Economics Letters*, 10:184–187.

Hassler, U. and Wolters, J. (1995). Long memory in inflation rates: International evidence. *Journal of Business & Economic Statistics*, 13:37–45.

Hatanaka, M. (1996). *Time Series based Econometrics - Unit roots and Co-Integration*. Oxford University Press, Oxford.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, 46:1251–1271.

Hegwood, N. D. and Papell, D. H. (1998). Quasi purchasing power parity. *International Journal of Finance and Economics*, 3:279–289.

Heinen, F., Michael, S., and Sibbertsen, P. (2011). Two competitive models and their identification problem: The ESTAR and TSTAR model. Discussion Paper No. 474 Leibniz University of Hannover.

Heinen, F., Sibbertsen, P., and Kruse, R. (2009). Forecasting long memory time series under a break in persistence. CREATES Research Paper 2009-53.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, 68:165–176.

Hsu, C. (2007). The MOSUM of squares test for monitoring variance changes. *Finance Research Letters*, 4:254–260.

Hurvich, C. M., Deo, R., and Brodsky, J. (1998). The mean-squared error of Geweke and Porter-Hudaks estimator of the memory parameter of a long-memory time series. *Journal of Time Series Analysis*, 19:19–46.

Jansen, E. S. and Teräsvirta, T. (1996). Testing parameter constancy and super exogeneity in econometric equations. *Oxford Bulletin of Economics and Statistics*, 58:735 – 768.

Jones, D. A. (1978). Nonlinear autoregressive processes. *Proceedings of the Royal Society of London. Series A*, 360:71–95.

Journal of Econometrics (1996). Special issue on *Asymmetries and nonlinearities in dynamic economic models . Journal of Econometrics*, 74.

Journal of Econometrics (2002). Special issue on *Long memory and nonlinear time series. Journal of Econometrics*, 110.

Kanas, A. (2006). Purchasing power parity and Markov regime switching. *Journal of Money, Credit and Banking*, 38:1669–1687.

Kapetanios, G. (2003). Threshold models for trended time series. *Empirical Economics*, 28:687–707.

Kapetanios, G., Shin, Y., and Snell, A. (2003). Testing for a unit root in the nonlinear STAR framework. *Journal of Econometrics*, 112:359–379.

Kapetanios, G. and Weeks, M. (2003). Non-nested models and the likelihood ratio statistic: A comparison of simulation and bootstrap based tests. Working Paper No. 490 Queen Mary, University of London.

Kilian, L. and Taylor, M. P. (2003). Why is it so difficult to beat the random walk for exchange rates? *Journal of International Economics*, 60:85–107.

Kim, J. (2000). Detection of a change in persistence of a linear time series. *Journal of Econometrics*, 95:97–116.

Kim, J., Belaire Franch, J., and Badilli Amador, R. (2002). Corrigendum to "Detection of a change in persistence of a liner times series". *Journal of Econometrics*, 109:389–392.

Klimko, L. A. and Nelson, P. I. (1978). On conditional least squares estimation for stochastic processes. *The Annals of Statistics*, 6:629–642.

Kobayashi, M. and McAleer, M. (1999). Tests of linear and logarithmic transformations for integrated processes. *Journal of the American Statistical Association*, 94:860–868.

Koop, G., Pesaran, M. H., and Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, 74:119–147.

Krämer, W. and Davies, L. (2002). Testing for unit roots in the context of misspecified logarithmic random walks. *Economics Letters*, 75:313–319.

Kruse, R. (2009). A new unit root test against ESTAR based on a class of modified statistics. *Statistical Papers*, 52:71–85.

Kruse, R., Frömmel, M., Menkhoff, L., and Sibbertsen, P. (2010). What do we know about real exchange nonlinearities? *Empirical Economics, forthcoming.*

Kuan, C. and Hornik, K. (1995). The generalized fluctuation test: A unifying view. *Econometric Reviews*, 14:135–161.

Kuswanto, H. and Sibbertsen, P. (2008). A study on "Spurious long memory in nonlinear time series models". Discussion Paper No. 410 Leibniz University of Hannover.

Lahtinen, M. (2006). The purchasing power parity puzzle: A sudden nonlinear perspective. *Applied Financial Economics*, 16:119–125.

Lee, J. H. and Brorsen, B. W. (1997). A non-nested test of GARCH vs. EGARCH models. *Applied Economics Letters*, 4:765–768.

Lee, T. H. and Ullah, A. (2001). Nonparametric bootstrap tests for neglected nonlinearity in time series regression models. *Journal of Nonparametric Statistics*, 13:425–451.

Lee, T. H., White, H., and Granger, C. W. J. (1993). Testing for neglected nonlinearity in time series models. *Journal of Econometrics*, 56:269–290.

Leisch, F., Hornik, K., and Kuan, C. (2000). Monitoring structural changes with the generalized fluctuation test. *Econometric Theory*, 16:835–854.

Leybourne, S., Kim, T., Smith, V., and Newbold, P. (2003). Tests for a change in persistence against the null of difference stationarity. *Econometrics Journal*, 6:291–311.

Leybourne, S., Taylor, R., and Kim, T. (2007). CUSUM of squares-based tests for a change in persistence. *Journal of Time Series Analysis*, 28:408–433.

Liew, V., Baharumshah, A., and Chong, T. (2004). Are Asian real exchange rates stationary? *Economics Letters*, 83:313–316.

Lin, C. F. J. and Teräsvirta, T. (1994). Testing the constancy of regression parameters against continuous structural change. *Journal of Econometrics*, 62:211–228.

Lo, M. C. (2008). Nonlinear PPP deviations: A monte carlo investigation of their unconditional half-life. *Studies in Nonlinear Dynamics and Econometrics*, 12:1–29.

Lu, M., Mizon, G. E., and Monfardini, C. (2008). Simulation encompassing: Testing non-nested hypotheses. *Oxford Bulletin of Economics and Statistics*, 70:781–806.

Luukkonen, R., Saikkonen, P., and Teräsvirta, T. (1988). Testing linearity against smooth transition autoregressive models. *Biometrika*, 75:491–499.

MacDonald, R. (1998). What determines real exchange rates? The long and the short of it. *Journal of International Financial Markets, Institutions and Money*, 8:117–153.

Meese, R. and Rogoff, K. (1983). Emprical exchange rate models of the seventies – Do they fit out of sample? *Journal of International Economics*, 14:3–24.

Meese, R. and Rogoff, K. (1988). Was it real? The exchange rate-interest differential relation over the modern floating-rate period. *Journal of Finance*, 43:933 – 948.

Michael, P., Nobay, A. R., and Peel, D. A. (1997). Transactions costs and nonlinear adjustments in real exchange rates: An empirical investigation. *The Journal of Political Economy*, 105:862–879.

Montañés, A. (1997). Unit roots, level shifts and purchasing power parity. Working Paper University of Zaragoza.

Nelson, C. R. and Startz, R. (2007). The zero-information-limit condition and spurious inference in weakly identified models. *Journal of Econometrics*, 138:47–62.

Norman, S. (2010). How well does nonlinear mean reversion solve the PPP puzzle? *Journal of International Money and Finance*, 29:919–937.

Öcal, N. (2000). Nonlinear models for U.K. macroeconomic time series. *Studies in Nonlinear Dynamics and Econometrics*, 4:123–135.

Øksendahl, B. (2003). *Stochastic differential equations - An introduction with applications*. Springer, New York.

Ozaki, T. (1980). Non-linear time series models for non-linear random vibrations. *Journal of Applied Probability*, 17:84–93.

Perron, P. (1989). The great crash, the oil price shock and the unit root hypothesis. *Econometrica*, 57:1361–1401.

Pesaran, B. and Pesaran, M. H. (1995). A non-nested test of level-differenced versus log-differenced stationary models. *Econometric Reviews*, 14:213–227.

Pesaran, M. H. and Pesaran, B. (1993). A simulation approach to the problem of computing Cox's statistic for testing nonnested models. *Journal of Econometrics*, 57:377–392.

Pesaran, M. H. and Potter, S. M. (1997). A floor and ceiling model of US output. *Journal of Economic Dynamics and Control*, 21:661–695.

Pesaran, M. H. and Timmermann, A. (2005). Small sample properties of forecasts from autoregressive models under structural breaks. *Journal of Econometrics*, 129:183–217.

Potter, S. M. (1995). A nonlinear approach to US GNP. *Journal of Applied Econometrics*, 10:109–125.

Potter, S. M. (1999). Nonlinear time series modelling: An introduction. *Journal of Economic Surveys*, 13:505–528.

Priestley, M. B. (1988). *Non-linear and non-stationary time series analysis*. Academic Press, London.

Quandt, R. (1960). Tests of the hypotheses that a linear regression system obeys two seperate regimes. *Journal of the American Statistical Association*, 55:324–330.

Rapach, D. E. and Wohar, M. E. (2006). The out-of-sample forecasting performance on nonlinear models of real exchange rate behavior. *International Journal of Forecasting*, 22:341–261.

Ray, B. K. and Tsay, R. S. (2002). Bayesian methods for change-point detection in long-range dependent processes. *Journal of Time Series Analysis*, 23:687–705.

Robbins, H. (1970). Statistical methods related to the law of iterated logarithms. *The Annals of Mathematical Statistics*, 41:1397–1409.

Rogoff, K. (1996). The purchasing power parity puzzle. *Journal of Economic Literature*, 34:647–668.

Rose, A. K. (1988). Is the real interest rate stable? *Journal of Finance*, 43:1095–1112.

Said, S. E. and Dickey, W. A. (1984). Testing for a unit root in autoregressive moving average models of unknown order. *Biometrika*, 71:599 – 607.

Sandberg, R. (2009). Convergence to stochastic power integrals for dependent heterogeneous processes. *Econometric Theory*, 25:739–747.

Sarno, L. (2005). Towards a solution to the puzzles in exchange rate economics: Where do we stand? *The Canadian Journal of Economics*, 38:673–708.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 4:461–464.

Sercu, P., Uppal, R., and van Hulle, C. (1995). The exchange rate in the presence of transaction costs: Implications for tests of purchasing power parity. *Journal of Finance*, 50:1309–1319.

Sibbertsen, P. and Kruse, R. (2009). Testing for a change in persistence under long-range dependencies. *Journal of Time Series Analysis*, 30:263–285.

Siegmund, D. (1985). *Sequential Analysis*. Springer, New York.

Singh, A. (2006). Macroeconomic volatility: The policy lessons from Latin America. IMF Working Paper 06/166.

Smallwood, A. D. (2005). Joint tests for non-linearity and long memory: The case of purchasing power parity. *Studies in Nonlinear Dynamics and Econometrics*, 9:1–28.

Sowell, F. (1990). The fractional unit root distribution. *Econometrica*, 58:495–505.

Spanos, A., Hendry, D., and Reade, J. (2008). Linear vs. log-linear unit-root specification: An application of misspecification encompassing. *Oxford Bulletin of Economics and Statistics*, 70:829–847.

Stock, J. H. and Watson, M. W. (2005). Implications of dynamic factor analysis for VAR models. NBER working paper No. 11467.

Studies in Nonlinear Dynamics & Econometrics (2008a). Special issue on *Nonlinear Dynamical Methods and Time Series Analysis. Studies in Nonlinear Dynamics and Econometrics*, 12(1).

Studies in Nonlinear Dynamics & Econometrics (2008b). Special issue on *Regime-Switching Models in Economics and Finance. Studies in Nonlinear Dynamics and Econometrics*, 12(3).

Taqqu, M. (1977). Law of the iterated logarithm for sums of non-linear functions of gaussian variables that exhibit a long range dependence. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 40:203–238.

Taylor, M. (2003). Purchasing power parity. *Review of International Economics*, 11:436–452.

Taylor, M. P., Peel, D. A., and Sarno, L. (2001). Nonlinear mean-reversion in real exchange rates: Toward a solution to the purchasing power parity puzzles. *International Economic Review*, 42:1015–1042.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, 89:208–218.

Teräsvirta, T. and Anderson, H. M. (1992). Characterizing nonlinearities in business cycles using smooth transition autoregressive models. *Journal of Applied Econometrics*, 7:119–136.

Teräsvirta, T., Lin, C. F., and Granger, C. W. J. (2008). Power of the neural network linearity test. *Journal of Time Series Analysis*, 14:209–220.

Teräsvirta, T. (2004). Smooth transition regression modeling. In Lütkepohl, H. and Krätzig, M., editors, *Applied Time Series Econometrics*. Cambridge University Press.

Teräsvirta, T., Tjøstheim, D., and Granger, C. W. J. (2010). *Modelling nonlinear economic time series*. Oxford University Press.

Tjøstheim, D. (1986). Estimation in nonlinear time series models. *Stochastic Processes and their Applications*, 21:251–273.

Tjøstheim, D. (1990). Non-linear time series and markov chains. *Advances in Applied Probability*, 22:587–611.

Tong, H. (1990). *Non-linear time series - A dynamical system approach*. Oxford University Press.

Tsay, R. (1986). Non-linearity tests for time series. *Biometrika*, 73:461–466.

Tweedie, R. L. (1975). Sufficient conditions for ergodicity and recurrence of markov chains on a general state space. *Stochastic Processes and their Applications*, 3:385–403.

van Dijk, D., Franses, P. H., and Paap, R. (2002a). A nonlinear long memory model, with an application to US unemployment. *Journal of Econometrics*, 110:135–165.

van Dijk, D., Teräsvirta, T., and Franses, P. H. (2002b). Smooth transition autoregressive models - a survey of recent developments. *Econometric Reviews*, 21:1–47.

van Norden, S. (1996). Regime switching as a test for exchange rate bubbles. *Journal of Applied Econometrics*, 11:219–251.

White, H. (1981). Consequences and detection of misspecified nonlinear regression models. *Journal of the American Statistical Association*, 76:419–433.

White, H. (1982a). Maximum likelihood estimation of misspecified models. *Econometrica*, 50:1–26.

White, H. (1982b). Regularity conditions for Cox's test of non-nested hypotheses. *Journal of Econometrics*, 19:301–318.

White, H. and Domowitz, I. (1984). Nonlinear regression with dependent observations. *Econometrica*, 52:143–161.

Wooldridge, J. M. (1990a). An encompassing approach to conditional mean tests with applications to testing nonnested hypotheses. *Journal of Econometrics*, 45:331–350.

Wooldridge, J. M. (1990b). A unified approach to robust, regression-based specification tests. *Econometric Theory*, 6:17–43.

Yamaguchi, K. (2011). Estimating a change point in the long memory parameter. *Journal of Time Series Analysis*, 32:304–314.

Yao, Q. and Fan, J. (2005). *Nonlinear time series: Nonparametric and parametic methods*. Springer, New York.

Zeileis, A., Leisch, F., Kleiber, C., and Hornik, K. (2005). Monitoring structural change in dynamic econometric models. *Journal of Applied Econometrics*, 20:99–121.

Zivot, E. and Andrews, D. W. K. (1992). Further evidence on the great crash, the oil-price shock and the unit-root hypothesis. *Journal of Business & Economic Statistics*, 10:251–270.