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

Florian Heinen, Stefanie Michael*and Philipp Sibbertsen

Leibniz Universität Hannover
Institute of Statistics
Königsworther Platz 1, 30167 Hannover, Germany

May 24, 2011

Abstract

Determining good parameter estimates in ESTAR models is known to be difficult. We show that the phenomena of getting strongly biased estimators is a consequence of the so-called identification problem, the problem of properly distinguishing the transition function in relation to extreme parameter combinations. This happens in particular for either very small or very large values of the error term variance. Furthermore, we introduce a new alternative model -the TSTAR model- which has similar properties as the ESTAR model but reduces the effects of the identification problem. We also derive a linearity and a unit root test for this model.

JEL-Numbers: C12, C22, C52

Keywords: Nonlinearities · Smooth transition · Linearity testing · Unit root testing · Real exchange rates

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

*Stefanie Michael, Königsworther Platz 1, 30167 Hannover, Germany, michael@statistik.uni-hannover.de, phone: +49-(0)511-762 3080, fax: +49-(0)511-762 3923

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. (2008). 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 3. After describing the model (see Section 3.1) we derive the linearity as well as the unit root test in Sections 3.2 and 3.3, respectively. The simulation studies we performed are summarized in Section 3.4. A comparison of the ESTAR and TSTAR model is presented in Section 4, discussing real exchange data. Section 5 concludes whereas all proofs are collected in the Appendix, together with certain technical lemmas.

2 Exponential Smooth Transition Autoregressive Models

In this section we introduce the Exponential Smooth Transition Autogressive (ESTAR) model and study its basic properties. Subsequently, the identification problem is described and analyzed in Section 2.2.

2.1 The ESTAR(p) Model

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

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

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 $\Psi\, w_t$ and $\Theta\, w_t$ 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 (1) can alternatively be written in somewhat less intuitive representations

$$y_t = [\Psi\, w_t] + [\Phi\, w_t] \cdot G(y_{t-d}; \gamma, c) + \varepsilon_t \tag{2}$$

$$= [\Xi\, w_t] \cdot [1 - G(y_{t-d}; \gamma, c)] + [\Theta\, w_t] + \varepsilon_t \tag{3}$$

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

$$\lim_{\gamma\to+\infty} G(\cdot; \gamma, c) \equiv 1 - \mathbb{1}_c, \quad \lim_{\gamma\to 0} G(\cdot; \gamma, c) \equiv 0 \quad \text{and} \quad \lim_{z\to\pm\infty} G(z; \gamma, c) \equiv 1 \tag{4}$$

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

$$G(z; \gamma, c) = 1 - \exp(-\gamma(z - c)^2),\ z \in \mathbb{R}. \tag{5}$$

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 (3) $\xi_0 = \vartheta_0 = 0$, $c = 0, \xi = \xi_1$ and $\vartheta = \vartheta_1$ and obtain the process

$$y_t = \left[\xi \exp\left(-\gamma y_{t-1}^2\right) + \vartheta\right] y_{t-1} + \varepsilon_t,\ t \in \mathbb{Z}. \tag{6}$$

Example 2.1.
Figures 1 and 2 show two realizations of the ESTAR(1) process (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.

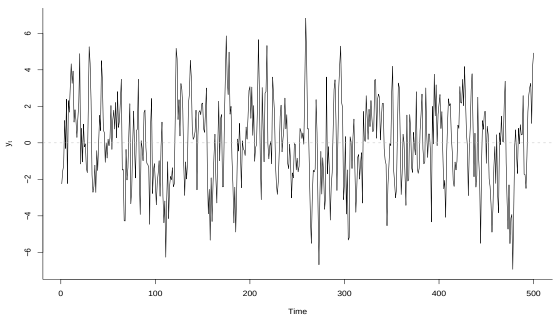

Figure 1: $\sigma = 2, \vartheta = 1$ and $\xi = -0.45$.

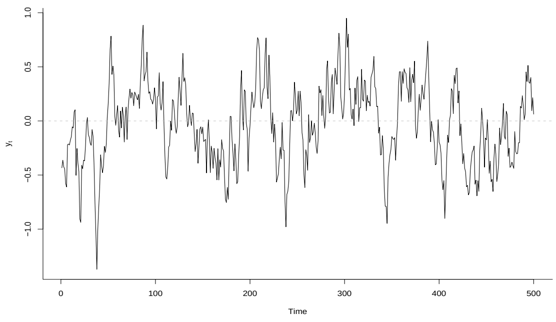

Figure 2: $\sigma = 0.2, \vartheta = 1$ and $\xi = -0.45$.

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.

(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 2.3 (Moment properties of y_t)**.**
Let $\{y_t\}_{t\in\mathbb{Z}}$ be as in (6) with $|\xi| + |\vartheta| < 1$ and with innovations $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ that satisfy Assumption 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 $t \in \mathbb{Z}$,

$$\mathbb{E}\left[\exp(-n\gamma y_t^2)y_t^{2k+1}\right] = 0. \tag{7}$$

(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

$$\lim_{\sigma\downarrow 0}\left|\frac{\mathbb{E}\left[\exp(-n\gamma y_t^2)\,y_t^{2k}\right]}{\sigma^{2k}} - c_k\right| = 0 \tag{8}$$

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

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

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 (7) implies that all odd moments of y_t vanish for all $t \geq 0$ by choosing $n = 0$. The even moments behave like (8) again for $n = 0$.

2.2 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 (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 Startza (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 Startza (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 1 and 2, which show realizations of ESTAR processes with different values of σ^2, that -while only appearing implicitly in the definition (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 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 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) standardize 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 (6) we obtain the following theorem, proved in the Appendix.

Theorem 2.4 (Asymptotic Variance of $\hat{\beta}$).
Let y_t be as in (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. Then

$$\lim_{\sigma \downarrow 0} \mathrm{Var}(\hat{\gamma}) \quad \to \quad \infty. \tag{9}$$

Remarks.

- We are aware that the limiting situation in (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 2.4 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 2.4 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.

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.
In Section 3.1 the TSTAR model is defined. A linearity and a unit root test are derived in Sections 3.2 and 3.3, respectively. Section 3.4 then gives an overview of the performed Monte Carlo Simulations.

3.1 The Model and Estimators

The TSTAR(p) is defined by (1) using the transition function

$$G(z; \kappa, c) \quad = \quad \left[1 - \left(1 + (z-c)^2\right)^{-\kappa}\right],\ z \in \mathbb{R}, \tag{10}$$

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 (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 (10) is just one possible alternative to the ESTAR model. Formula (10) is motivated by the density of Student's t-distribution while the transition function (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 2.3 (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 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 3 which illustrates for different values of κ the resulting transition functions in comparison to the ESTAR setting shown in Figure 4 (note the different scale on the x-axis).

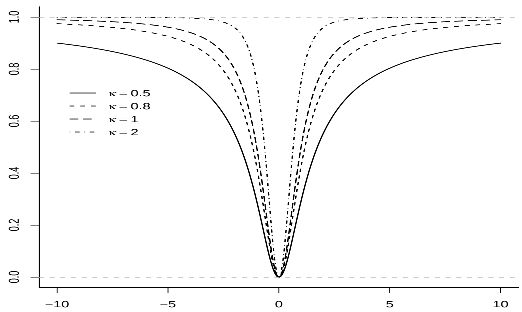

Figure 3: TSTAR: Transition function for different κ.

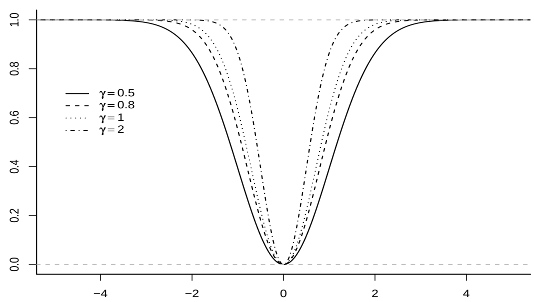

Figure 4: ESTAR: Transition function for different γ.

As κ and y_t^2 no longer appear as a product in the transition function (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 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.

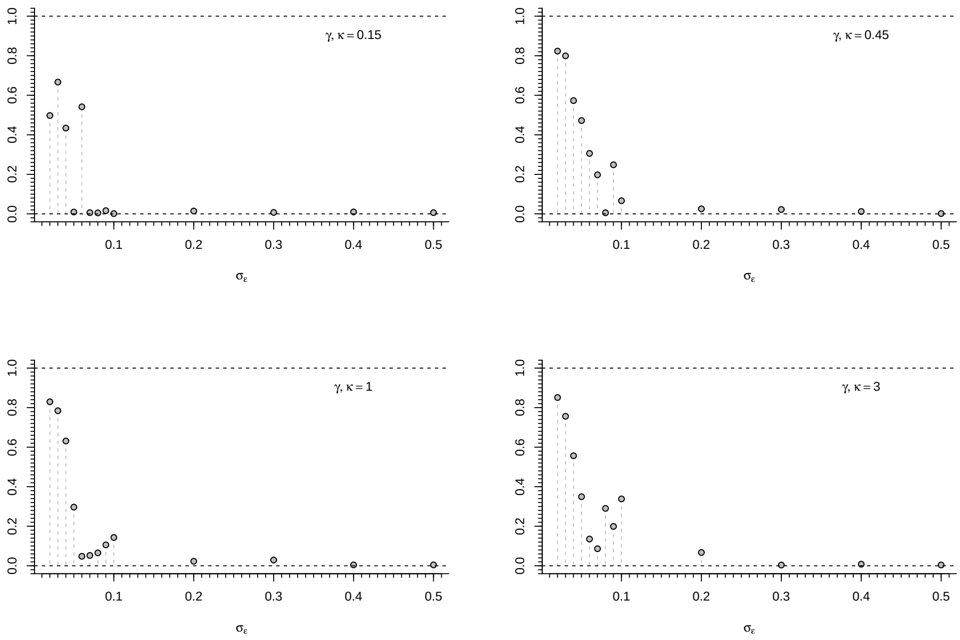

Figure 5: Comparison of the estimators in ESTAR and TSTAR settings.

3.2 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) of the TSTAR(p) model, i.e.

$$y_t \quad = \quad [\Psi w_t] + [\Phi w_t] \cdot G(y_{t-d}; \kappa, c) + \varepsilon_t. \tag{11}$$

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 (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 (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.

$$(1+x)^{-m} = 1 + \sum_{n=1}^{\infty}(-1)^n \frac{m(m+1)(m+2)\dots(m+n-1)}{n!}x^n,\ m > 0, \tag{12}$$

the transition function G in (11) can be approximated arbitrarily well by

$$G_k(\,\cdot\,) \;=\; \sum_{n=1}^{k}(-1)^n \frac{\kappa(\kappa+1)\dots(\kappa+n-1)(y_{t-d}-c)^{2n}}{n!} \tag{13}$$

choosing $x = (y_{t-d} - c)^2$ and $m = \kappa$ in (12) as well as a sufficiently large k. We then obtain the auxiliary regression model for a fixed $d \leq p$ and k

$$y_t = \sum_{i=1}^{p}\psi_i y_{t-i} + \sum_{j=1}^{p}\delta_{j,0}\, y_{t-j} + \sum_{j=1}^{p}\delta_{j,1}\, y_{t-j}y_{t-d} + \dots + \sum_{j=1}^{p}\delta_{j,2k}\, y_{t-j}y_{t-d}^{2k} + u_t \tag{14}$$

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 3.1.
As an example consider the simple TSTAR(1) model,

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

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

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

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

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

for which extensive Monte Carlo simulations are summarized in Section 3.4.

3.3 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) the TSTAR(1) model can be written in first differences as

$$\Delta y_t = \rho y_{t-1} + \varphi_1 y_{t-1} \left[1 - \left(1 + y_{t-1}^2\right)^{-\kappa}\right] + \varepsilon_t \tag{15}$$

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

$$H_0 : \kappa = 0 \quad \text{vs.} \quad H_1 : \kappa > 0\,. \tag{16}$$

We now proceed in the same way as in the previous section and approximate the nonlinearity with a Binomial expansion as in (13) setting the number of summands to $k = 3$. This yields the auxiliary regression (see also (14))

$$\Delta y_t = \delta_{1,2}\, y_{t-1}^3 + \delta_{1,4}\, y_{t-1}^5 + \delta_{1,6}\, y_{t-1}^7 + u_t \tag{17}$$

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

$$F^* = \frac{1}{3}(\hat{r}-r)' \left[\hat{\sigma}^2 R(X'X)^{-1}R'\right]^{-1}(\hat{r}-r) = \frac{1}{3}(R\hat{\beta})' \left[\hat{\sigma}^2 R(X'X)^{-1}R'\right]^{-1}(R\hat{\beta})$$
$$= \frac{1}{3}\hat{\beta}'\hat{\sigma}^2(X'X)\hat{\beta} \qquad (18)$$

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 3.2.
Consider the TSTAR(1) model (15) and let ε_t satisfy Assumption 2.2. Then the test statistic F^* as given in (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

$$\Delta y_t \quad = \quad \delta_{1,2} y_{t-1}^3 + \delta_{1,4} y_{t-1}^5 + \delta_{1,6} y_{t-1}^7 + \sum_{i=1}^{p} \pi_i \Delta y_{t-i} + u_t \,. \tag{19}$$

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 (17).

Theorem 3.3.
Consider the test statistic F^* as in Theorem 3.2 but computed from (19). Under the null of a unit root the test statistic maintains the same asymptotic distribution as in Theorem 3.2. Under the alternative the test statistic is consistent.

Theorem 3.3 holds also true for the case of including deterministic terms as in auxiliary regression (19). The asymptotic distribution in this case is such as in Theorem 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 3.2.

3.4 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 (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]

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 3.2 are stated in Tables 2 and 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: 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 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 4 for scenario one and in Table 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

[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.

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 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 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 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 $M = 1000000$.

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 6: Asymptotic critical values for unit root testing.

The results from the size experiments are summarized in Table 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.

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

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 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 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 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. (2008) 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 9) and $T = 500$ (Table 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. (2008) 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.

$T = 100$		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 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 10: Power results for unit root testing [in %] with $\sigma_\varepsilon = 0.1$

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).

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 test 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 6 and 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.

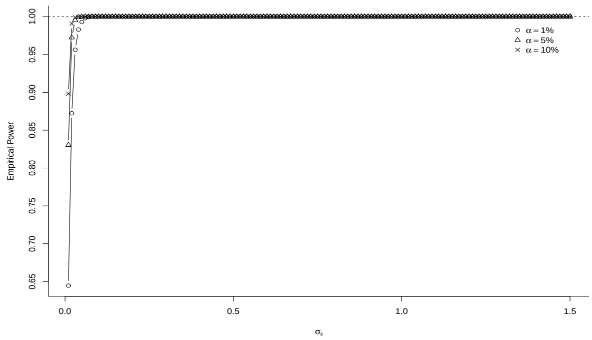

Figure 6: Power curve of the unit root test: $\psi_1 = 0.3, \varphi_1 = 0.6$ and $\kappa = 1$.

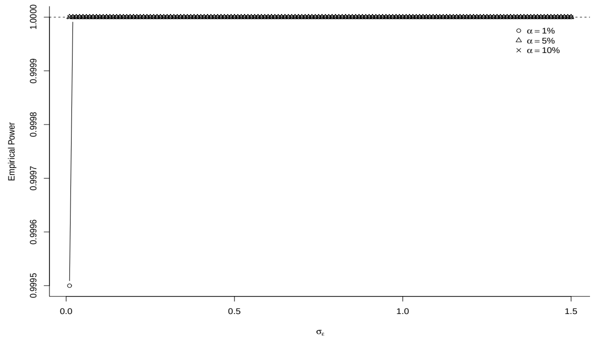

Figure 7: Power curve of the unit root test: $\psi_1 = 0.7, \varphi_1 = 0.2$ and $\kappa = 1$.

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 8 and 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.

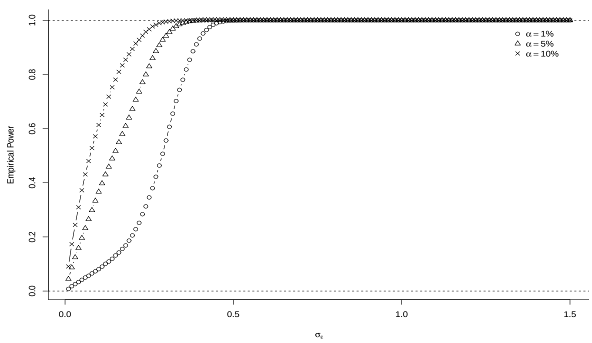

Figure 8: Power curve of the linearity test: $\psi_1 = 0.3, \varphi_1 = 0.6$ and $\kappa = 1$.

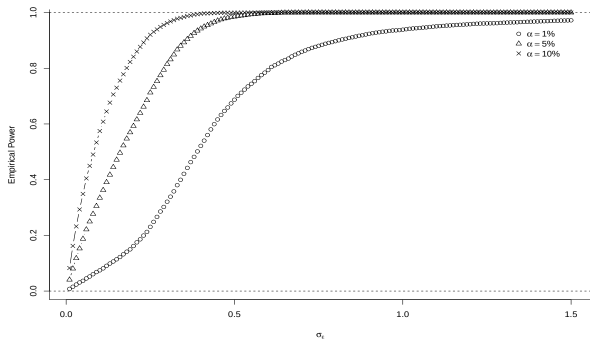

Figure 9: Power curve of the linearity test: $\psi_1 = 0.7, \varphi_1 = 0.2$ and $\kappa = 1$.

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 10.

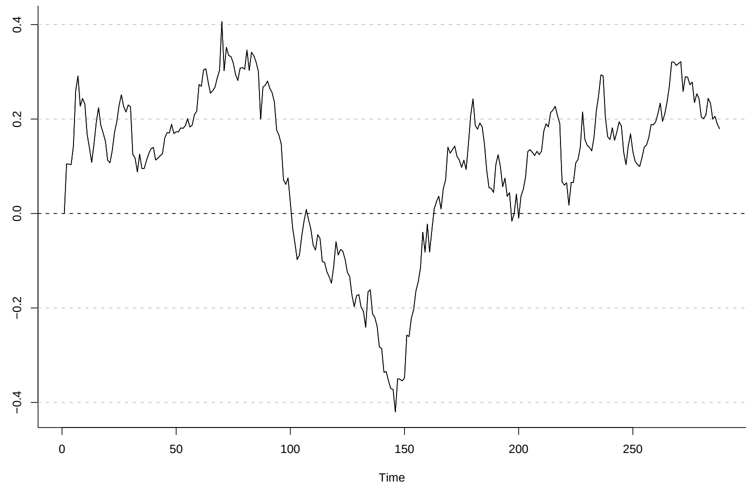

Figure 10: Monthly log real exchange rate for Germany

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 3.2 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 3.3 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 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 11 shows the estimation results for the parameter γ for the model under the null for the ESTAR and the TSTAR model respectively.

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

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 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 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.

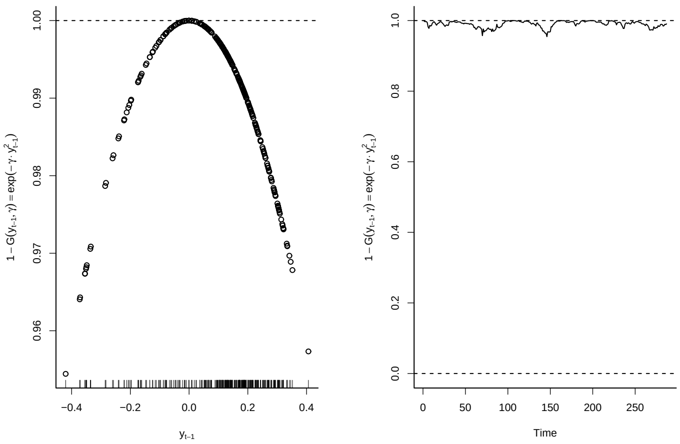

Figure 11: Left panel: One minus transition function against transition variable. Right panel: One minus transition function against time.

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 11 we obtain Figure 12.

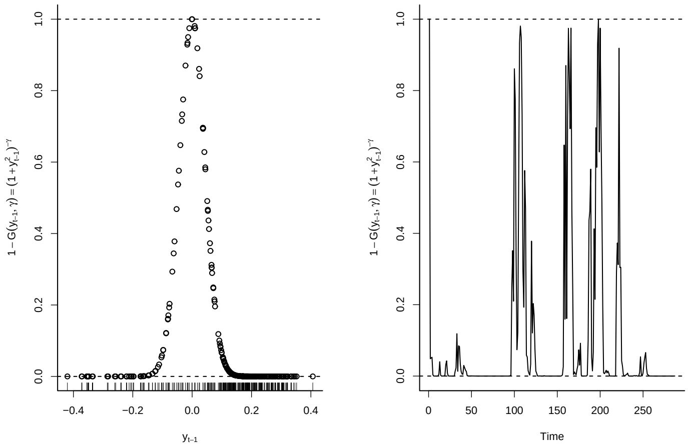

Figure 12: Left panel: One minus transition function against transition variable. Right panel: One minus transition function against time.

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). 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)).

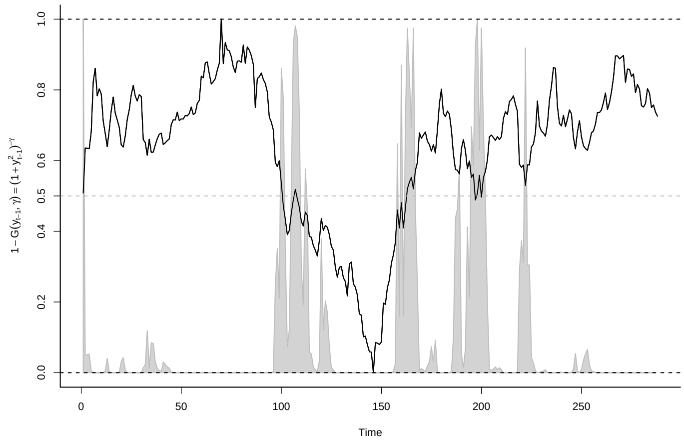

Figure 13: Rescaled real exchange rate with TSTAR transition function.

Figure 13 shows the real exchange rate from Figure 10 rescaled to be in [0,1] and the transition function from the right panel of Figure 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 11 is not unique for this particular data set but a common finding in empirically estimated ESTAR models. Kruse et al. (2008) for example support these findings for other real exchange rates (see their Figure 1).

The panels in Figure 11 and Figure 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. Looking at the estimated values for γ and κ in Table 11 and the left panels in Figures 11 and 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 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). 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.

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 2.4 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 models 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.

A Appendix: Proofs and Technical Lemmas

Lemma 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 2.3.

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 (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 A.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 A.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 A.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

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

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

$$\frac{\mathbb{E}\left[\exp(-n\gamma y_t^2)\, y_t^{2k}\right]}{\sigma^{2k}} \quad \leq \quad \frac{\mathbb{E}\left[y_t^{2k}\right]}{\sigma^{2k}} \tag{20}$$

and

$$\frac{\mathbb{E}\left[\exp(-n\gamma y_t^2)\, y_t^{2k}\right]}{\sigma^{2k}} \quad \geq \quad \frac{\mathbb{E}\left[(1-n\gamma y_t^2)\, y_t^{2k}\right]}{\sigma^{2k}} \quad \geq \quad \frac{\mathbb{E}\left[y_t^{2k}\right]}{\sigma^{2k}} - n\gamma\, \frac{\mathbb{E}\left[y_t^{2k+2}\right]}{\sigma^{2k+2}}\, \sigma^2 \tag{21}$$

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

$$\begin{aligned}
\mathbb{E}\left[y_t^{2k}\right] &= \mathbb{E}\left[\left[\left[\vartheta + \xi \exp(-\gamma y_{t-1}^2)\right] y_{t-1} + \varepsilon_t\right]^{2k}\right] \\
&= \sum_{j\in J} \binom{2k}{j} \mathbb{E}\left[\left[\vartheta + \xi \exp\left(-\gamma y_{y-1}^2\right)\right]^{2k-j} y_{t-1}^{2k-j}\right] \mathbb{E}\left[\varepsilon_t^j\right] \\
&= \sum_{j\in J} \binom{2k}{j} \sum_{\nu=0}^{2k-j} \binom{2k-j}{\nu} \vartheta^{2k-j-\nu} \xi^{\nu}\, \mathbb{E}\left[\exp\left(-\nu\gamma y_{t-1}^2\right) y_{t-1}^{2k-j}\right] c_{\varepsilon,j}\, \sigma^j
\end{aligned} \tag{22}$$

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.

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 $(-1)!! = 0!! = 1$,

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

due to the following argument. Formula (22) implies

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

As the exponent of the exponential term can be written as

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

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 (8) for $t=1$ by combining the previous display with (20) and (21).

Now assume that (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 (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 (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 A.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 $y \neq 0$,

$$f_{Y_k}(-y) = \frac{f_X\left(-y^{\frac{1}{2k+1}} a^{-\frac{1}{2k+1}}\right)}{(2k+1)\left(-y^{\frac{1}{2k+1}} a^{-\frac{1}{2k+1}}\right)^{2k}} = \frac{f_X\left(y^{\frac{1}{2k+1}} a^{-\frac{1}{2k+1}}\right)}{(2k+1)\left(y^{\frac{1}{2k+1}} a^{-\frac{1}{2k+1}}\right)^{2k}} = f_{Y_k}(y), \tag{23}$$

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

$$\lim_{x\to\infty} g_{k,i}(x) = \lim_{x\to\infty}\left(ax^{2k+1} + b\,\frac{x^{2k+1}}{e^{c\eta_i(x^2)}}\right) = \lim_{x\to\infty} ax^{2k+1} = \begin{cases} -\infty & \text{for } a<0,\\ \kappa_i & \text{for } a=0,\\ \infty & \text{for } a>0 \end{cases} \quad (24)$$

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

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

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

Figures 14 and 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 14, the latter corresponds to $b = 1$, $a = 0.2 \in (0, -\xi_{0,1}b] = (0, 0.4463]$.

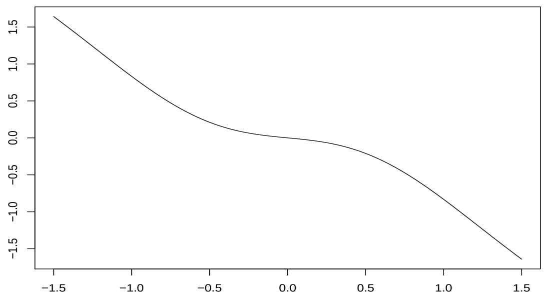

Figure 14: $g_{0,1}$ with $a = -1.2$ and $b = 1$.

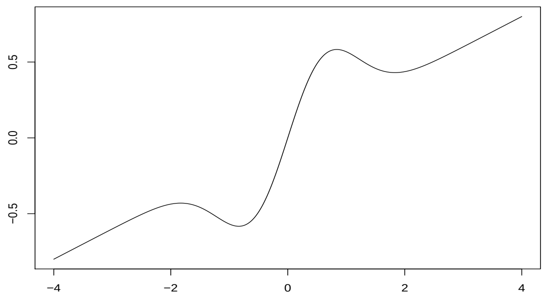

Figure 15: $g_{0,1}$ with $a = 0.2$ and $b = 1$.

(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

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

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 (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

$$\left(\frac{2k+1}{2} - cx^2\right)\exp\left(\frac{2k+1}{2} - cx^2\right) \;=\; -\frac{a}{b}\frac{2k+1}{2}\exp\left(\frac{2k+1}{2}\right). \tag{25}$$

For solving (25) we need to consider two different cases. Note that for $a \in (0, \xi_{k,1}\, b]$, the right hand side of (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 (25), for $j = 0, -1$,

$$\tfrac{2k+1}{2} - cx^2 \;=\; W_j\left(-\tfrac{a}{b}\tfrac{2k+1}{2}\exp\left(\tfrac{2k+1}{2}\right)\right) \tag{26}$$

$$\Leftrightarrow \quad w_{1,2,3,4} \;=\; \pm\sqrt{\frac{-W_j\left(-\frac{a}{b}\frac{2k+1}{2}\exp\left(\frac{2k+1}{2}\right)\right)+\frac{2k+1}{2}}{c}}$$

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

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

For $a \in [-1, 0]$ the argument of W_j in (26) is in the domain of only one real branch, namely W_0, so that (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 (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 $y \in \mathbb{R}$, $i = 1, 2$,

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

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 $G^{(j)}$, $j = 1, \ldots, 3$, $g_{k,1}^{(j)} := g_{k,1}\mathbb{1}_{G^{(j)}}$, $j = 1, \ldots, 3$, is strictly monotone with derivatives and inverse functions $g_{k,1}^{(j)\prime}$ and $(g_{k,1}^{(j)})^{-1}$, respectively. Hence, from Theorem 22.3 of Behnen and Neuhaus (1995), $k \in \mathbb{N}_0$,

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

for $H_{k,i} = g_{k,1}(G^{(i)})$. As $H_{k,1} \cup H_{k,2} \cup H_{k,3} = \mathbb{R}$, as $g_{k,1}^{(j)\prime}$, $j = 1, \ldots, 3$, 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 $y \in \mathbb{R}$. $\square$

Lemma A.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 2.4.
Let $\hat{\beta}_n$ be the conditional Maximum Likelihood estimator of $\beta = (\vartheta, \xi, \gamma)$. Then by Theorem 3.2 of Tjøstheim (1986)

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

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

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

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

$$\sigma^2 \left(U^{-1}\right)_{33} = \frac{\sigma^2}{\det(U)} \det \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix}. \tag{27}$$

The last factor of (27) behaves like $O(\sigma^4)$ for $\sigma \to 0$ as

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

due to Lemma 2.3 (iv) for some constant c_4. By a similar argument we obtain $\det(U) = O(\sigma^{10})$. Hence

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

which finishes the proof. □

Proof of Theorem 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

$$\hat{\beta} = \left(\sum_{t=1}^{T} x_t' x_t \right)^{-1} \sum_{t=1}^{T} x_t' u_t \tag{28}$$

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 $T \to \infty$

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

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

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

where

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

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 $T \to \infty$

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

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

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

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

$$\frac{1}{T}\sum_{t=1}^{T} u_t = \mathcal{O}_P(1)$$

as well. As

$$T\frac{1}{T}\sum_{t=1}^{T} x_t' x_t = T\,\mathcal{O}_P(1) = \mathcal{O}_P(T)\,.$$

and

$$\sum_{t=1}^{T} x_t' u_t = \sum_{t=1}^{T} x_t' \sum_{t=1}^{T} u_t = T\frac{1}{T}\left(\sum_{t=1}^{T} x_t' \sum_{t=1}^{T} u_t\right) = T\,\underbrace{\frac{1}{T}\sum_{t=1}^{T} x_t'}_{\mathcal{O}_P(1)}\,T\,\underbrace{\frac{1}{T}\sum_{t=1}^{T} u_t}_{\mathcal{O}_P(1)} = T^2\mathcal{O}_P(1) = \mathcal{O}_P(T^2)$$

we get for the OLS estimator $\hat{\beta}$, according to (28),

$$\hat{\beta} = (\mathcal{O}_P(T))^{-1}\,\mathcal{O}_P(T^2) = (T\mathcal{O}_P(1))^{-1}\,T^2\mathcal{O}_P(1) = \frac{1}{T}T^2\mathcal{O}_P(1) = T\mathcal{O}_P(1) = \mathcal{O}_P(T)\,.$$

Analogously for the test statistic F^* it follows that

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

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

Proof of Theorem 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 (19) reads

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

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

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

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 3.2 we have

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

where

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

and where $\mathbf{Q}$ is as given in Theorem 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 3.2 it is straightforward to show that the test is consistent under (19). □

Acknowledgements

The financial support of the DFG is gratefully acknowledged. The authors thank T. Teräsvirta and the participants of the research seminars of the Tinbergen Institute, Amsterdam, CREATES, Aarhus, University of Kiel and Humbold University Berlin for their inspiring comments.

References

Behnen and Neuhaus (1995). *Grundkurs Stochastik.* Teubner.

Chen, R. and Tsay, R. S. (1993). Functional-coefficient autoregressive models. *American Statistical Association*, 88:298–308.

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.

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.

Franses, P. H. and van Dijk, D. (2000). *Non-linear time series models in empirical finance.* Cambridge University Press.

Gatti, D. D., Gallegati, M., and Mignacca, D. (1998). Nonlinear dynamics and european GNP data. *Studies in Nonlinear Dynamics and Econometrics*, 3:43–59.

Haggan, V. and Ozaki, T. (1981). Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model. *Biometrika*, 68:198–196.

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. (1994). *Time Series Analysis.* Princeton University Press, Princeton, New Jersey.

Hansen, B. E. (1992). Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory*, 8:489–500.

Hatanaka, M. (1996). *Time Series based Econometrics - Unit roots and Co-Integration.* Oxford University Press, Oxford.

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 A*, 360:71–95.

Kapetanios, G., Shin, Y., and Snell, A. (2003). Testing for a unit root in the nonlinear STAR framework. *Journal of Econometrics*, 112:359–379.

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. (2008). What do we know about real exchange nonlinearities? forthcoming in Empirical Economics.

Luukkonen, R., Saikkonen, P., and Teräsvirta, T. (1988). Testing linearity against smooth transition autoregressive models. *Biometrika*, 75:491–499.

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.

Nelson, C. R. and Startza, R. (2007). The zero-information-limit condition and spurious inference in weakly identified models. *Journal of Econometrics*, 138:47–62.

Öcal, N. (2000). Nonlinear models for U.K. macroeconomic time series. *Studies in Nonlinear Dynamics and Econometrics*, 4:123–135.

Priestley, M. B. (1988). *Non-linear and non-stationary time series analysis.* Academic Press, London.

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.

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.

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.

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. (2004). Smooth transition regression modeling. In Lütkepohl, H. and Krätzig, M., editors, *Applied Time Series Econometrics*. Cambridge University Press, Cambridge.

Teräsvirta, T., Tjøstheim, D., and Granger, C. (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.

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., Teräsvirta, T., and Franses, P. H. (2002). Smooth transition autoregressive models - a survey of recent developments. *Econometric Reviews*, 21:1–47.

Yao, Q. and Fan, J. (2005). *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer, New York.