

A simple specification procedure for the transition function in persistent nonlinear time series models*

Hendrik Kaufmann[‡], Robinson Kruse[§], Philipp Sibbertsen[†]

Abstract

A simple procedure for the specification of the transition function describing the regime switch in nonlinear autoregressive models is proposed. This procedure is based on auxiliary regressions of unit root tests and is applicable to a variety of transition functions. In contrast to other procedures, complicated and computer-intense estimation of the candidate models is not necessary. Our approach entirely relies on OLS estimation of auxiliary regressions instead. We use standard information criteria for the selection of the unknown transition function. Our Monte Carlo simulations reveal that the approach works well in practice. Empirical applications to the S&P500 price-earnings ratio and the US interest spread highlight the merits of our suggested procedure.

JEL-Numbers: C15, C22, C52

Keywords: Nonlinearity · Smooth transition · Threshold model · Model selection · Unit root

*The authors are grateful to the participants of the 20th Symposium of the Society for Nonlinear Dynamics and Econometrics in Istanbul for their comments and suggestions. Financial support by the Deutsche Forschungsgemeinschaft (DFG) is gratefully acknowledged. Robinson Kruse gratefully acknowledges financial support from CREATES funded by the Danish National Research Foundation.

[‡]Leibniz University Hannover, School of Economics and Management, Institute of Statistics, Königsworther Platz 1, D-30167 Hannover, Germany. E-Mail: kaufmann@statistik.uni-hannover.de. Phone: +49-511-762-5636.

[§]Leibniz University Hannover, School of Economics and Management, Institute of Statistics, Königsworther Platz 1, D-30167 Hannover, Germany. E-mail address: kruse@statistik.uni-hannover.de and CREATES, Aarhus University, School of Economics and Management, Bartholins Alle 10, DK-8000 Aarhus C, Denmark. E-mail address: rkruse@creates.au.dk.

[†]Corresponding Author. Leibniz University Hannover, School of Economics and Management, Institute of Statistics, Königsworther Platz 1, D-30167 Hannover, Germany. E-Mail: sibbertsen@statistik.uni-hannover.de. Phone: +49-511-762-3783. Fax: +49-511-762-3923.

1 Introduction

During the last two decades it has become widely accepted that time series dynamics of a variety of macroeconomic and financial variables can be well described and modeled by stationary nonlinear processes. The most popular classes of regime-dependent time series models are smooth transition, threshold and Markov switching models. Nevertheless, model selection in a nonlinear framework is still a major challenge. This study contributes to the literature by considering the problem of model selection amongst smooth transition (see e.g. [Teräsvirta, 1994](#)) and threshold models (see e.g. [Tong, 1990](#)). To capture the well documented high degree of persistence in economic and financial time series, we focus on widely applied unit root tests against nonlinear models. It is remarkable that the theoretical and empirical literature on unit root tests against nonlinear smooth transition models is steadily growing. A survey of recent developments in unit root testing is given in [Haldrup, Kruse, Teräsvirta, and Varneskov \(2012\)](#).

The two different types of adjustment, i.e. threshold and smooth adjustment to an equilibrium, are economically and statistically appealing. For this reason, numerous empirical studies use these kind of transition functions to model the regime switch. Even though the exponential smooth transition autoregressive (ESTAR) model receives much attention, there are reasonable concerns regarding the exact shape of the transition function. Beside the fact that there are numerous different transition functions describing the adjustment to an equilibrium, evidence against an exponential shape has been documented. [Yoon \(2010\)](#) finds that ESTAR models cannot replicate the nonlinear serial dependence found in real exchange rates. [Kim and Moh \(2010\)](#) find evidence for nonlinearities in real exchange rates though not for an ESTAR model. Knowledge of the specific type and shape of the transition function is of importance for the analysis of generalized impulse-responses, forecasting and regime dating (see [Rapach and Wohar, 2006](#)). Another concern is a lack of asymmetry. [Norman and Phillips \(2009\)](#) suggest a skewed generalized transition function. [Sollis \(2009\)](#) concludes that an asymmetric ESTAR process is more appropriate than a symmetric one. Another criticism is the low flexibility of an exponential transition function in general. On the contrary, the threshold transition function implies an abrupt switch between regimes which can be seen as highly restrictive, too. The main goal of this study is to suggest and subsequently compare simple procedures for the selection of the most appropriate transition function, e.g. exponential, threshold or double logistic.

In a related article, [Psaradakis, Sola, Spagnolo, and Spagnolo \(2009\)](#) consider nonlinear model selection by using well-known and standard information criteria. The major drawback of such an approach stems from the fact that estimation of all nonlinear models under consideration is required. Beside time consuming and very intense computa-

tions, optimization routines may easily end up in local valleys of the likelihood function. Limited to the smooth transition framework, [Teräsvirta \(1994\)](#) suggests a procedure to discriminate between transition functions relying on a sequence of ordinary F -tests. Unfortunately, this procedure is problematic in the case of a linear unit root (see [Kiliç, 2004](#), [Harvey and Leybourne, 2007](#) and [Kruse and Sandberg, 2010](#)). To maintain generality and simplicity, we follow a similar philosophy as [Teräsvirta \(1994\)](#), but rely on auxiliary regressions which are built for unit root tests against nonlinear models. As pointed out by [Choi and Moh \(2007\)](#), a single unit root test is not able to distinguish between different nonlinear models. On the other hand, a battery of different unit root tests, each single one designed to detect a certain nonlinear model, are expected to be more informative. Therefore, this article aims at comparing simple statistics which are directly computable from a battery of auxiliary unit root regressions. These statistics shall be informative and indicative for selecting the transition function of a nonlinear transition model. This work focuses on the $\inf -t$ test by [Park and Shintani \(2009\)](#). The test does not rely on a Taylor approximation as the test suggested by [Kapetanios, Shin, and Snell \(2003\)](#) does. Linearization of the nonlinear model by applying a Taylor approximation can be costly in terms model selection precision. We therefore use the test proposed by [Park and Shintani \(2009\)](#) which uses a grid-search over the unidentified parameters under the null hypothesis instead.

As a possible selection rule, we consider the minimal sum of squared residuals which is in line with [Hansen \(1997\)](#). In a similar vein, we consider the performance of information criteria (similar to [Psaradakis et al., 2009](#)). The main difference is that information criteria are computed from auxiliary regressions and not from estimated nonlinear models themselves. The appeal of such a procedure lies in its simplicity and the fact that OLS estimation can be used throughout the analysis.

The remainder of the paper is organized as follows. Section 2 reviews the unit root $\inf -t$ test by [Park and Shintani \(2009\)](#) and the different transition functions. In Section 3 we discuss potential selection rules and computational details. Section 4 provides the results of our Monte Carlo simulation. In Section 5 we apply our procedure to the S&P500 price-earnings ratio and the US interest rate spread. Our results offer some new insights and conclusions on the nonlinear transition mechanism in these series. Section 6 concludes.

2 Unit root tests and nonlinear models

The $\inf -t$ unit root test of [Park and Shintani \(2009\)](#) covers many types of transition dynamics, including threshold, discrete and smooth transition functions in general. In this paper we consider nonlinear $AR(p)$ models with transition variable y_{t-1} which is

the most common choice in the related literature. The data generating process (DGP) is given by

$$\Delta y_t = \lambda y_{t-1} G(y_{t-1}, \theta) + \sum_{i=1}^{p-1} \rho_i \Delta y_{t-i} + \varepsilon_t, \quad (1)$$

where θ is an m -dimensional parameter vector, $G(\cdot)$ denotes a real-valued transition function and ε_t is a zero-mean white noise process. If $\lambda = 0$, the model becomes linear and there is only a single regime with a unit root implying no adjustment to any long-run equilibrium. Therefore, the null hypothesis of a unit root is given by $H_0 : \lambda = 0$. The alternative hypothesis of a globally stationary nonlinear model is given by $H_1 : \lambda < 0$. As θ is not identified under the null hypothesis, ordinary least squares estimates of λ for each value of $\theta \in \Theta$ are required to obtain¹

$$t(\theta) = \frac{\hat{\lambda}(\theta)}{s(\hat{\lambda}(\theta))}, \quad (2)$$

where $s(\hat{\lambda}(\theta))$ is the OLS standard error of the OLS estimate $\hat{\lambda}(\theta)$. The $\inf -t$ test is defined as

$$\inf -t \equiv \inf_{\theta \in \Theta} t(\theta). \quad (3)$$

The limiting distribution of the $\inf -t$ statistic is derived in [Park and Shintani \(2009\)](#). It depends on the transition function and the parameter space Θ . The lag length p is selected with the BIC criterion with a maximum of $p_{\max} = [4(n/100)^{1/4}]$ lags. It shall be noted that the limiting distribution does not change when additional lagged differences of the dependent variable are included on the right-hand side of the regression.

We consider widely applied models like the exponential smooth transition autoregressive (ESTAR), the double threshold autoregressive (D-TAR) and the double logistic smooth transition autoregressive (D-LSTAR) model. For simplicity, only symmetric transition functions are under consideration, but all procedures are applicable for asymmetric functions as well. Moreover, our approach is not limited to these particular types of transition functions in general.

The exponential transition function with location at zero is given by²

$$G(y_{t-1}, \theta) = 1 - \exp\{-\gamma^2 y_{t-1}^2\}, \quad \theta = \gamma. \quad (4)$$

¹Note that equation (1) is the linear OLS test regression with fixed θ .

²Please note that we use this uncommon notation of γ^2 to ensure comparability with [Park and Shintani \(2009\)](#).

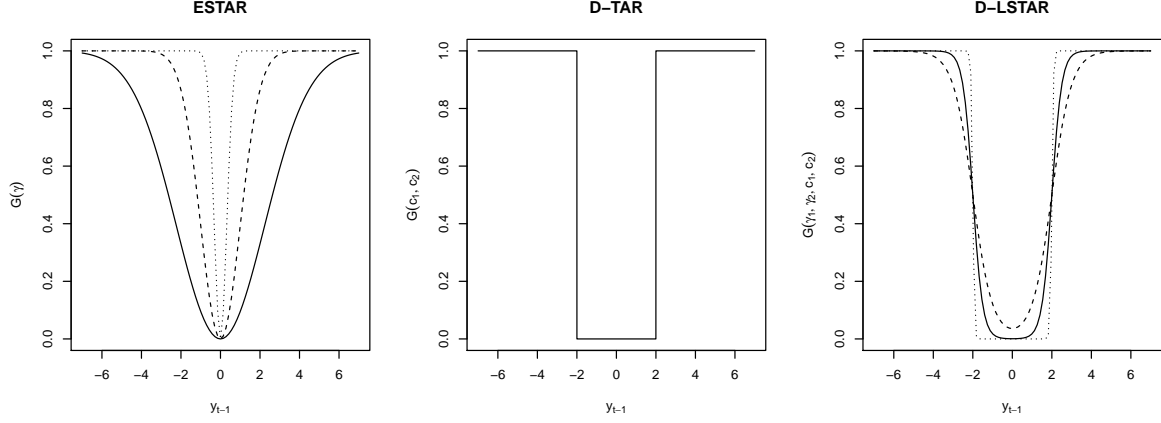


Figure 1: Transition functions for different parameterizations. For the ESTAR model, we set the scale parameter γ to $\sqrt{0.1}$ (solid), $\sqrt{0.5}$ (dashed) and $\sqrt{5}$ (dotted). For the D-TAR model, the location parameters are set to $-c_1 = c_2 = 2$. For the D-LSTAR model, the parameters are $-c_1 = c_2 = 2$ and $\gamma_1 = \gamma_2$ is 2 (solid), 1 (dashed) and 25 (dotted).

The parameter γ determines the smoothness of adjustment. The higher γ , the sharper is the transition function. For each limit, $\gamma \rightarrow \infty$ or $\gamma \rightarrow 0$, the ESTAR model becomes linear. In applied work, the inner regime is often modeled as unit root, while the outer regime is stationary, i.e. $\lambda < 0$, thereby ensuring global stationarity.

The D-TAR (or three regime TAR) model transition function is given by

$$G(y_{t-1}, \theta) = 1\{y_{t-1} < c_1\} + 1\{y_{t-1} > c_2\}, \quad \theta = (c_1, c_2)', \quad (5)$$

where $1\{\cdot\}$ is the indicator function. The D-TAR model reflects sudden changes between regimes and nests a linear process for $c_1 = c_2$. We consider the symmetric version with $-c_1 = c_2$. A non-stationary regime is active if y_t takes values between c_1 and c_2 . Again, the out regimes are stationary if $\lambda < 0$ which ensures global stationarity. The double logistic transition function is the smooth-adjustment version of the threshold function in 5 and given by

$$G(y_{t-1}, \theta) = [1 + \exp\{\gamma_1(y_{t-1} - c_1)\}]^{-1} + [1 + \exp\{\gamma_2(y_{t-1} - c_2)\}]^{-1}, \quad \theta = (\gamma_1, \gamma_2, c_1, c_2)'. \quad (6)$$

The D-LSTAR model nests the D-TAR model with $\gamma_1, \gamma_2 \rightarrow \infty$. A symmetric version can be obtained by setting the parameter restrictions $\gamma_1 = \gamma_2$ and $-c_1 = c_2$. It has also the ability to mimic the exponential transition function.

Graphs for different parameterizations of the three transition functions are shown in Figure 1.

3 Selection of nonlinear models

The goal of this paper is to discriminate between different transition functions in a nonlinear framework. To this end, we are interested in model selection criteria which are easy to implement and fast to compute. Therefore, we use only well-known criteria which are directly accessible from the auxiliary regression of the $\inf -t$ test.

The first selection rule is the residual sum of squares (RSS), which measures the goodness of fit. The RSS is given by

$$\text{RSS} \equiv \sum_{t=1}^n \hat{\varepsilon}_t^2(\tilde{\theta}),$$

where $\tilde{\theta}$ denotes the parameter vector θ which minimizes the $\inf -t$ statistic. Second, we compute well-known information criteria, Akaike's information criterion (AIC), the Bayesian information criterion (BIC) and the Hannan-Quinn information criterion (HQIC). These criteria reflect the goal of parsimony and penalize in different ways for the inclusion of an additional parameter. They share the following common structure:

$$IC \equiv \ln(\text{RSS}/n) + \varphi(n),$$

where $\varphi(n)$ denotes the penalty term which is specific to the particular criterion. Such information criteria are able to circumvent over-parametrization. Although all criteria are not designed for the selection of nonlinear models in the first place, reasonable results are presented by [Psaradakis et al. \(2009\)](#). The main difference in our approach is that we do not have to estimate a single model. The price for this simplification is that only a limited amount of parameter combinations are under consideration. However, our simulation results presented in the next section suggest that this issue is not problematic. The exact procedure for a collection of models (\mathcal{M}) under consideration is as follows:

1. Determine the appropriate lag length p as described in Section 2.
2. Perform an $\inf -t$ unit root test against model $m \in \mathcal{M}$. Find the minimal $t(\theta)$ -statistic. Save the auxiliary regression output and parameter vector $\tilde{\theta}$. Repeat this step for all models $m \in \mathcal{M}$.
3. Calculate a selection criterion for each model $m \in \mathcal{M}$ based on the saved regression outputs in the previous step.
4. Select the model m^* which minimizes the selection criterion, i.e.

$$m^* = \arg \min_{m \in \mathcal{M}} IC(m). \tag{7}$$

	1%	5%	10%	25%	50%	75%	90%	95%	99%
ESTAR	-3.64	-3.14	-2.85	-2.37	-1.86	-1.35	-0.71	-0.26	0.48
D-LSTAR	-4.07	-3.61	-3.33	-2.93	-2.49	-2.09	-1.71	-1.45	-0.78
D-TAR	-4.07	-3.54	-3.30	-2.90	-2.47	-2.06	-1.65	-1.34	-0.65

Note: Based on discrete approximation to the Brownian motion by partial sums of standard normal random variable with 1,000 steps and 10,000 replications for demeaned data.

Table 1: Asymptotic critical values of the $\inf -t$ test against different transition functions.

The following remarks to the procedure are in order. The critical values of the test of [Park and Shintani \(2009\)](#) depend on the transition function and the chosen grid for θ . Because of the grid-based approach the computer time grows exponentially with additional parameters. Even though only linear regressions are performed, we try to reduce the required time as much as possible. Therefore, we deal only with demeaned time series where the location parameters for symmetric models are often fixed or in a fixed relationship like in the presented transition functions in [Section 2](#).

Another important point is the exact specification of the grid which determines the parameter space Θ . For the location parameters we search over the parameter space $[\min(y), \max(y)]$, where $\min(y)$ and $\max(y)$ are respectively the minimum and maximum of (y_1, \dots, y_n) . This is a slight modification in comparison to [Caner and Hansen \(2001\)](#) and [Park and Shintani \(2009\)](#) who work with a grid containing values from the 15th to the 85th percent quantile. Although the smaller grid is reasonable if the true model is known, it turned out to be too restrictive if it has to approximate another data generating process (DGP). For the scale parameter γ a search over the interval $[10^{-1}P, 10^3P]$ is performed, where $P = \left(\sum_{t=1}^n y_t^2 / n\right)^{-1/2}$, which is the same grid as in [Park and Shintani \(2009\)](#). Due to the fact that changes in γ have a larger impact on the result when γ is small, a kinked grid is used. Two thirds of the available grid-points are in the interval $[10^{-1}P, S]$, $S = \frac{10^3P - 10^{-1}P}{100}$. Two thirds of the remaining grid-points are in the interval $[S, 3S]$ and the last grid-points are located in the interval $[3S, 10^3P]$.

Critical values for demeaned time series and the corresponding transition functions are given in [Table 1](#) for a grid of 50 data points per parameter.

4 Monte Carlo study

This section examines the empirical finite-sample performance of the RSS and information criteria to distinguish between different transition functions based on $\inf -t$ test auxiliary regressions. Additionally, we investigate whether a kinked grid with 50 points as described in [Section 3](#) is sufficient to approximate the true transition function well.

The DGPs of the experiments are based on the models in Section 2 which are given by

$$\Delta y_t = \lambda y_{t-1} G(y_{t-1}, \theta) + \varepsilon_t, \quad (8)$$

where $\varepsilon_t \sim N(0, \sigma^2)$. The parameter settings are $\lambda = \{-0.5, -0.9\}$ and $\sigma = 1$. For the ESTAR model we use the transition function in 4 and set $\gamma = \sqrt{0.1}$ to ensure a persistent time series with sufficient observations in the smooth transition region between the inner and outer regime. For the D-LSTAR model we use the transition function in 6 and set $\gamma_1 = \gamma_2 = 4$ and $-c_1 = c_2 = 2$ and for the D-TAR model the transition function is given in 5 with $-c_1 = c_2 = 2$. With these parametrizations all investigated transition functions are symmetric. Graphs of the transition functions for these parameter combinations are shown in Figure 1. All simulated time series are demeaned before we apply the unit root tests and the first 300 observations are discarded to reduce the effect of initial values. We use sample sizes of $n = \{200, 400, 800\}$ with $M = 2000$ replications. For all unit root tests we set $p = 1$, meaning that no further autoregressive components enter the auxiliary regressions. Even though we want to discriminate between nonlinear models, we also compare the performance with a linear AR model.

Table 2 shows the results for the ESTAR model. Without the AR model as competitor the BIC performs best in all combinations, but AIC and HQIC also perform quite well with a selection rate around 70% for $n = 800$. If $n = 200$ the D-TAR model is selected quite often but the selection error approaches zero with growing n . This result changes slightly if the AR model is added to the analysis. In this case the linear model is often selected by the BIC in small samples ($n = 200$) instead of the ESTAR or D-TAR model. But as sample size grows the difference between the results with and without the additional competitor vanish. The pairwise comparison shows that in the most draws the D-LSTAR model wins in terms of the RSS, but the gain from the additional parameter is not sufficient to win in terms of information criteria. In contrast the AR model has never the lowest RSS. It also turned out that the results of the direct comparisons of the ESTAR against the D-LSTAR model are quite stable for different sample sizes, whereas the selection rates for all other models decrease rapidly. While this analysis holds for both values of λ , the selection rates of the ESTAR model are usually higher the less persistent the time series is, especially in small samples. This result is not surprising since the importance of the connection between regimes increases with their difference. The results of the D-LSTAR model are shown in Table 3. With this transition function in the DGP the additional consideration of the AR model has almost no effect on the results, even for small samples the selection rates are close to zero. The best criteria to choose the true model are the AIC and HQIC. But even though a selection rate up to 100% is observed for $n = 800$, the sample size has to be large enough that the

ESTAR								
$\lambda = -0.5, \sigma = 1,$ $\gamma = \sqrt{0.1}$					$\lambda = -0.9, \sigma = 1,$ $\gamma = \sqrt{0.1}$			
	ESTAR	D-LSTAR	D-TAR	AR	ESTAR	D-LSTAR	D-TAR	AR
$n = 200$								
RSS	6	53	42		11	68	21	
	(6)	(53)	(42)	(0)	(11)	(68)	(21)	(0)
AIC	34	3	63		58	8	34	
	(32)	(2)	(53)	(13)	(57)	(8)	(32)	(4)
BIC	36	0	64		64	0	36	
	(23)	(0)	(37)	(40)	(57)	(0)	(29)	(15)
HQIC	33	5	63		55	12	33	
	(31)	(4)	(54)	(11)	(55)	(11)	(31)	(3)
RSS		7	36	100		11	64	100
AIC		78	36	80		77	64	94
BIC		98	36	50		97	64	83
HQIC		71	36	83		73	64	95
$n = 400$								
RSS	11	72	17		16	80	4	
	(11)	(72)	(17)	(0)	(16)	(80)	(4)	(0)
AIC	59	10	31		73	19	8	
	(59)	(9)	(29)	(3)	(73)	(19)	(8)	(0)
BIC	67	0	33		90	1	10	
	(57)	(0)	(27)	(16)	(88)	(1)	(9)	(3)
HQIC	58	12	30		70	23	8	
	(57)	(11)	(29)	(2)	(70)	(23)	(8)	(0)
RSS		12	67	100		16	90	100
AIC		77	67	96		77	90	99
BIC		98	67	82		98	90	97
HQIC		74	67	96		73	90	100
$n = 800$								
RSS	15	82	3		20	80	0	
	(15)	(82)	(3)	(0)	(20)	(80)	(0)	(0)
AIC	72	22	7		70	29	1	
	(72)	(22)	(7)	(0)	(70)	(29)	(1)	(0)
BIC	91	0	9		97	2	1	
	(90)	(0)	(8)	(2)	(97)	(2)	(1)	(0)
HQIC	70	24	7		68	32	1	
	(70)	(24)	(7)	(0)	(68)	(32)	(1)	(0)
RSS		15	91	100		20	99	100
AIC		75	91	100		70	99	100
BIC		99	91	98		98	99	100
HQIC		73	91	100		68	99	100

The top panel in each block shows the percentage of model selection on the basis of the RSS and different information criteria. The values in parentheses show the results if a linear model is an additional competitor. The bold numbers highlight the percentage of correct decisions. The bottom panel in each block shows the selection rates of the true models in percent in pairwise comparisons between the candidate models and the true model.

Table 2: Monte Carlo results for ESTAR DGPs.

D-LSTAR								
$\lambda = -0.5, \sigma = 1,$ $\gamma_1 = \gamma_2 = 4, -c_1 = c_2 = 2$					$\lambda = -0.9, \sigma = 1,$ $\gamma_1 = \gamma_2 = 4, -c_1 = c_2 = 2$			
	ESTAR	D-LSTAR	D-TAR	AR	ESTAR	D-LSTAR	D-TAR	AR
$n = 200$								
RSS	5	61	34		3	88	9	
	(5)	(61)	(34)	(0)	(3)	(88)	(9)	(0)
AIC	31	8	61		26	49	25	
	(31)	(8)	(60)	(2)	(26)	(49)	(25)	(0)
BIC	34	0	65		45	11	43	
	(30)	(0)	(61)	(8)	(45)	(11)	(43)	(0)
HQIC	30	12	58		23	55	22	
	(29)	(12)	(57)	(1)	(23)	(55)	(22)	(0)
RSS	95		66	100	97		91	100
AIC	50		27	96	72		74	100
BIC	19		5	70	41		42	99
HQIC	55		33	97	76		77	100
$n = 400$								
RSS	4	84	12		1	98	1	
	(4)	(84)	(12)	(0)	(1)	(98)	(1)	(0)
AIC	29	41	30		8	89	4	
	(29)	(41)	(30)	(0)	(8)	(89)	(4)	(0)
BIC	45	3	52		32	54	14	
	(44)	(3)	(52)	(1)	(32)	(54)	(14)	(0)
HQIC	27	44	28		7	90	3	
	(27)	(44)	(28)	(0)	(7)	(90)	(3)	(0)
RSS	96		88	100	99		99	100
AIC	68		67	100	92		96	100
BIC	32		23	96	67		84	100
HQIC	70		70	100	94		97	100
$n = 800$								
RSS	1	97	2		0	100	0	
	(1)	(97)	(2)	(0)	(0)	(100)	(0)	(0)
AIC	9	86	6		0	100	0	
	(9)	(86)	(6)	(0)	(0)	(100)	(0)	(0)
BIC	39	35	26		5	94	0	
	(39)	(35)	(26)	(0)	(5)	(94)	(0)	(0)
HQIC	8	87	5		0	100	0	
	(8)	(87)	(5)	(0)	(0)	(100)	(0)	(0)
RSS	99		98	100	100		100	100
AIC	91		94	100	100		100	100
BIC	57		69	100	95		100	100
HQIC	92		95	100	100		100	100

The top panel in each block shows the percentage of model selection on the basis of the RSS and different information criteria. The values in parentheses show the results if a linear model is an additional competitor. The bold numbers highlight the percentage of correct decisions. The bottom panel in each block shows the selection rates in percent in pairwise comparisons between the candidate models and the true model.

Table 3: Monte Carlo results for D-LSTAR DGPs.

D-TAR								
$\lambda = -0.5, \sigma = 1,$ $-c_1 = c_2 = 2$					$\lambda = -0.9, \sigma = 1,$ $-c_1 = c_2 = 2$			
	ESTAR	D-LSTAR	D-TAR	AR	ESTAR	D-LSTAR	D-TAR	AR
$n = 200$								
RSS	2	52	47		0	65	35	
	(2)	(52)	(47)	(0)	(0)	(65)	(35)	(0)
AIC	12	8	81		4	34	62	
	(12)	(8)	(80)	(1)	(4)	(34)	(62)	(0)
BIC	14	0	86		7	8	85	
	(12)	(0)	(85)	(3)	(7)	(8)	(85)	(0)
HQIC	11	10	79		3	39	58	
	(11)	(10)	(78)	(0)	(3)	(39)	(58)	(0)
RSS	86	47		100	92	35		100
AIC	86	85		99	92	63		100
BIC	86	98		96	92	89		100
HQIC	86	82		99	92	58		100
$n = 400$								
RSS	1	66	34		0	72	28	
	(1)	(66)	(34)	(0)	(0)	(72)	(28)	(0)
AIC	4	31	65		0	57	43	
	(4)	(31)	(65)	(0)	(0)	(57)	(43)	(0)
BIC	8	1	90		1	28	71	
	(8)	(1)	(90)	(0)	(1)	(28)	(71)	(0)
HQIC	4	35	62		0	59	41	
	(4)	(35)	(62)	(0)	(0)	(59)	(41)	(0)
RSS	92	34		100	98	28		100
AIC	92	66		100	98	43		100
BIC	92	95		100	98	71		100
HQIC	92	62		100	98	41		100
$n = 800$								
RSS	0	78	22		0	81	19	
	(0)	(78)	(22)	(0)	(0)	(81)	(19)	(0)
AIC	0	59	41		0	73	27	
	(0)	(59)	(41)	(0)	(0)	(73)	(27)	(0)
BIC	2	17	82		0	51	49	
	(2)	(17)	(82)	(0)	(0)	(51)	(49)	(0)
HQIC	0	60	40		0	74	26	
	(0)	(60)	(40)	(0)	(0)	(74)	(26)	(0)
RSS	97	22		100	100	19		100
AIC	97	41		100	100	27		100
BIC	97	82		100	100	49		100
HQIC	97	40		100	100	26		100

The top panel in each block shows the percentage of model selection on the basis of the RSS and different information criteria. The values in parentheses show the results if a linear model is an additional competitor. The bold numbers highlight the percentage of correct decisions. The bottom panel in each block shows the selection rates in percent in pairwise comparisons between the candidate models and the true model.

Table 4: Monte Carlo results for D-TAR DGPs.

n	ESTAR		D-LSTAR		D-TAR	
	$\lambda = -0.5$	$\lambda = -0.9$	$\lambda = -0.5$	$\lambda = -0.9$	$\lambda = -0.5$	$\lambda = -0.9$
200	0.15	0.12	0.12	0.07	0.06	0.03
400	0.10	0.08	0.08	0.05	0.03	0.02
800	0.07	0.04	0.05	0.03	0.02	0.01

Table 5: Average distance between the true transition function and the transition function of the correct model which minimizes the $\inf -t$ statistic.

gain from the additional parameter exceeds the penalty. Especially for $n = 200$ and the BIC the true model is hardly chosen at all. In those cases the wins are split between the other nonlinear models. In a direct comparison with its competitors the D-LSTAR model shows quickly increasing win ratios for AIC and HQIC whereas the BIC selection increases more slowly. As in the ESTAR case, a correct decision is more likely the more the regimes differ. For the DGP with $\lambda = -0.9$ and $n = 200$ the win rate is up to 55% for the HQIC instead of 12% for $\lambda = -0.5$. Therefore, the gain is higher for the D-LSTAR model in comparison to the former ESTAR results.

The results of the D-TAR model are presented in Table 4. As in the D-LSTAR case, the interpretation of the results is independent of the AR model which is hardly chosen at all. The BIC has the most correct decisions. The selection rate is best for small n and decreases in favor of the D-LSTAR model in larger samples. This result is also observable in the pairwise comparison. The selection rate of the ESTAR model is quite low and especially for $n \geq 400$ close to zero. Even in the rare cases where the ESTAR model wins in a pairwise comparison, the D-LSTAR model is better overall in most cases so that the ESTAR is not selected. This result shows that there is a substantial gain from the additional smoothness parameter of the D-LSTAR in large samples. Even though this behavior may not be desirable at first sight, it turned out that the selected value for the smoothness parameter γ is always close to the upper bound of the grid and thus close to the D-TAR model. In this case the transition functions are practically the same with only a few observations on the smooth part of the D-LSTAR transition function. Therefore, the economic interpretation of the final model is almost identically. It is also important to note that this behavior is not limited to our procedure, but also visible in the results of [Psaradakis et al. \(2009\)](#) where the models are estimated.

An overall comparison reveals that the AIC and HQIC exhibit the most balanced performance. The most likely error is to choose the D-LSTAR instead of the D-TAR in large samples. Fortunately the impact of this error on the interpretation of the final model should be quite small because of the practically identical transition functions.

The second question is if the specified grid in Section 3 with only 50 grid-points per

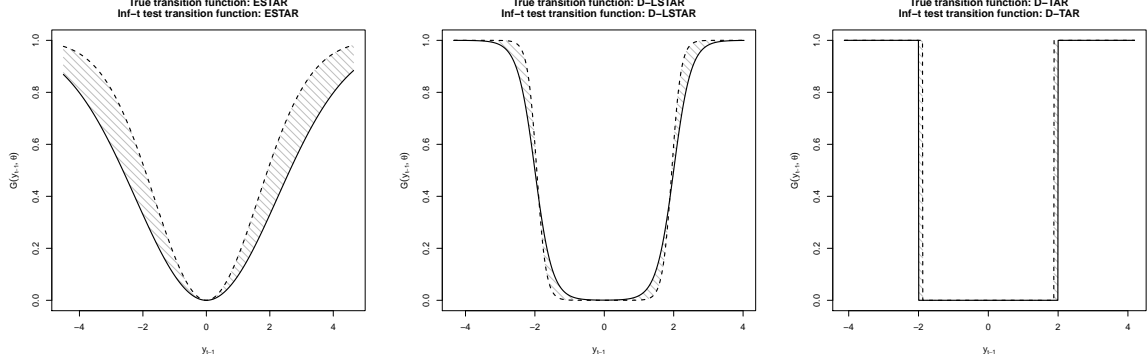


Figure 2: Representative draws of the DGP with $\lambda = -0.5$ and $n = 400$ (one per simulated transition function). The solid line shows the true transition function; the dashed line shows the transition function of the correct model evaluated at $\tilde{\theta}$. The grey shaded area is the computed distance between the two functions.

parameter is sufficient to approximate the true model. Therefore, we suggest to measure the distance between the true transition function and the transition function which minimizes the $\text{inf}-t$ statistic with

$$D = \frac{\int_{\min(y)}^{\max(y)} (G(y; \theta) - \tilde{G}(y; \tilde{\theta})) dy}{\max(y) - \min(y)}, \quad (9)$$

where $G(y; \theta)$ is the true and $\tilde{G}(y; \tilde{\theta})$ the transition function which corresponds to the $\text{inf}-t$ test. D measures the normalized area between the two functions. For all models the average distance $\bar{D} = M^{-1} \sum_{i=1}^M D_i$ between the true transition function and the transition function of the correct model evaluated at $\tilde{\theta}$ is computed, where D_i denotes the distance D of Monte Carlo replication i . These measures approach zero with an increasing sample size as shown in Table 5. Even for small samples the average distance is quite small. An example of D for the different transition functions is shown in Figure 2. Thus, we can conclude that even a small grid is sufficient for the proposed procedure. With this small grid computations are extremely fast and below one second for transition functions with one parameter and below ten seconds for transition functions with two parameters for $n = 200$. An increasing sample size raises the computer time only slightly.³ This makes the procedure useful for practitioners and ensures a wide applicability.

³All computations were made on a 3GHz Intel Core 2 (one core used) with R and the popular standard `lm()` function therein. A much better performance with other programs like GAUSS or MATLAB and parallel computing is expected.

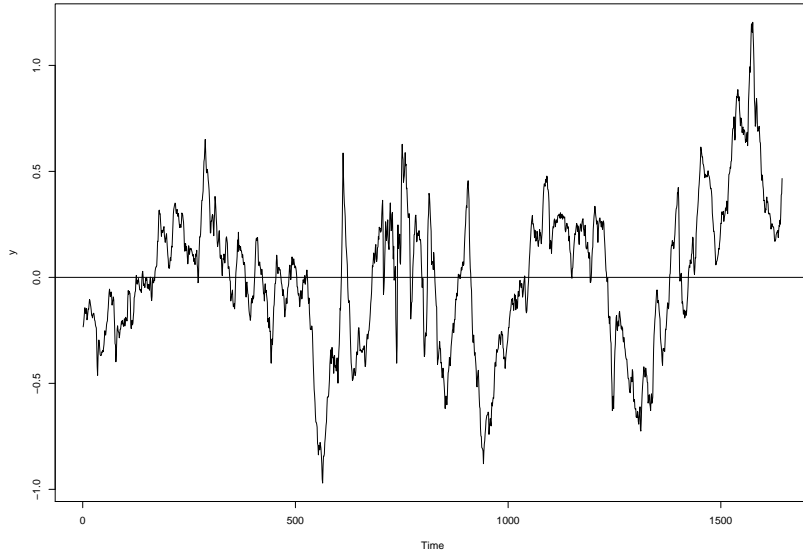


Figure 3: S&P500 price-earnings ratio (demeaned).

5 Empirical applications

We first report and discuss the empirical results of two important applications. The first one we consider is the S&P500 price-earnings (P/E) ratio. The general idea behind the P/E ratio is that market prices will not drift too far away from their normal levels relative to earnings (as long as no bubbles are present in the price index). If the P/E ratio is high above/below its historical mean, a price adjustment to the equilibrium is expected, i.e. a decrease/increase in the next periods. This idea was first applied in [Campbell and Shiller \(1998\)](#) in a linear setting. They conclude that the ratio has special significance to forecast stock prices, but also that the linear relation might not hold in the case of extreme events. [Rapach and Wohar \(2005\)](#) reexamine the predictability of stock prices based on the P/E ratio. They argue that the evidence of real stock price predictability at long, but not short horizons by the P/E ratio is difficult to explain in a linear framework. [Kilian \(1999\)](#) interprets this observation as indirect evidence of a nonlinear DGP. Therefore, [Rapach and Wohar \(2005\)](#) work with a parsimonious ESTAR model, similar to the one in [Section 2](#), to reconsider the mean-reverting behavior. [Kilian and Taylor \(2003\)](#) argue that such a model is able to capture price movements better in a world of noise trading and risky arbitrage. Nevertheless, [Rapach and Wohar \(2005\)](#) conclude that further analysis of nonlinear model specification for the P/E ratio is warranted. Another ESTAR approach to explain financial ratios is given in [McMillan \(2007\)](#).

We use data obtained from Robert Shiller’s website, consisting of 1644 monthly observations ranging from January 1871 to December 2007. We thus exclude observations

	$g = 50$				$g = 300$			
	ESTAR	D-LSTAR	D-TAR	AR	ESTAR	D-LSTAR	D-TAR	AR
RSS	2.8117	2.8098	2.8098	3.3158	2.8117	2.8098	2.8095	3.3158
AIC	-6.3668	-6.3663	-6.3675	-6.2043	-6.3668	-6.3663	-6.3676	-6.2043
BIC	-6.3570	-6.3531	-6.3577	-6.2011	-6.3570	-6.3531	-6.3578	-6.2011
HQIC	-6.3668	-6.3663	-6.3675	-6.2043	-6.3668	-6.3663	-6.3676	-6.2043
UR test	-3.9115***	-4.0529**	-4.0490**	-3.8469***	-3.9120***	-4.0755***	-4.0524**	-3.8469***
$\hat{\lambda}$	-0.0118	-0.0123	-0.0123	-0.0081	-0.0118	-0.0123	-0.0124	-0.0081

Table 6: Results obtained from $\inf -t$ auxiliary test regressions for the price-earnings ratio (columns ESTAR, D-LSTAR and D-TAR) and from an ADF test regression for the linear model (column AR).

belonging to the period of the recent financial crisis. These few observations can be judged as being highly influential on the results as they are very different from the remaining observations. We assume that including these observations would lead to structural breaks in the parameters of the nonlinear models we consider. This claim is supported by the strong empirical evidence for structural breaks found in [Lettau and Nieuwerburgh \(2008\)](#). Hence, time-varying STAR models allowing for both, nonlinearity and structural instability appear to be better suited in this case. This class of models is, however, beyond the scope of this paper. An important reference is [Lundbergh, Teräsvirta, and van Dijk \(2003\)](#).

In this application one additional lag is selected. The calculated criteria and t-values of the $\inf -t$ test can be found in Table 6. The procedure is performed with a grid of $g = 50$ and $g = 300$ points per parameter. Regardless which information criterion we consider (either AIC, BIC or HQIC), the D-TAR model is selected in all cases. The D-LSTAR model performs equivalently well in terms of RSS, which is not surprising because it is nesting the D-TAR. Nevertheless, minor deviations in favor of the D-TAR model can occur because of the upper bound of the grid for γ . Therefore, we judge the ESTAR and D-LSTAR model to be inferior. This finding contrasts previous results. When comparing the nonlinear models to the linear one, a clear improvement can be observed.

For the sake of a deeper investigation, we compare all transition functions in Figure 4. Each individual cell of the left side of Figure 4 plots a transition function against its argument y_{t-1} . The relevant parameters are obtained from the [Park and Shintani \(2009\)](#) unit root test regressions. For the ESTAR model, we have $\tilde{\gamma} = 7.45$ and $\tilde{c} = 0.2$ for the D-TAR model. The threshold parameter c has a clear interpretation: the three different regimes are $r_1 = (-\infty, -0.2)$, $r_2 = [-0.2, 0.2]$ and $r_3 = (0.2, \infty)$. For the D-LSTAR model we find exactly the same threshold parameter and a very high smoothness parameter close to, but lower than the maximal value of the grid.

A comparison to the suggested D-TAR model shows that the exponential transition

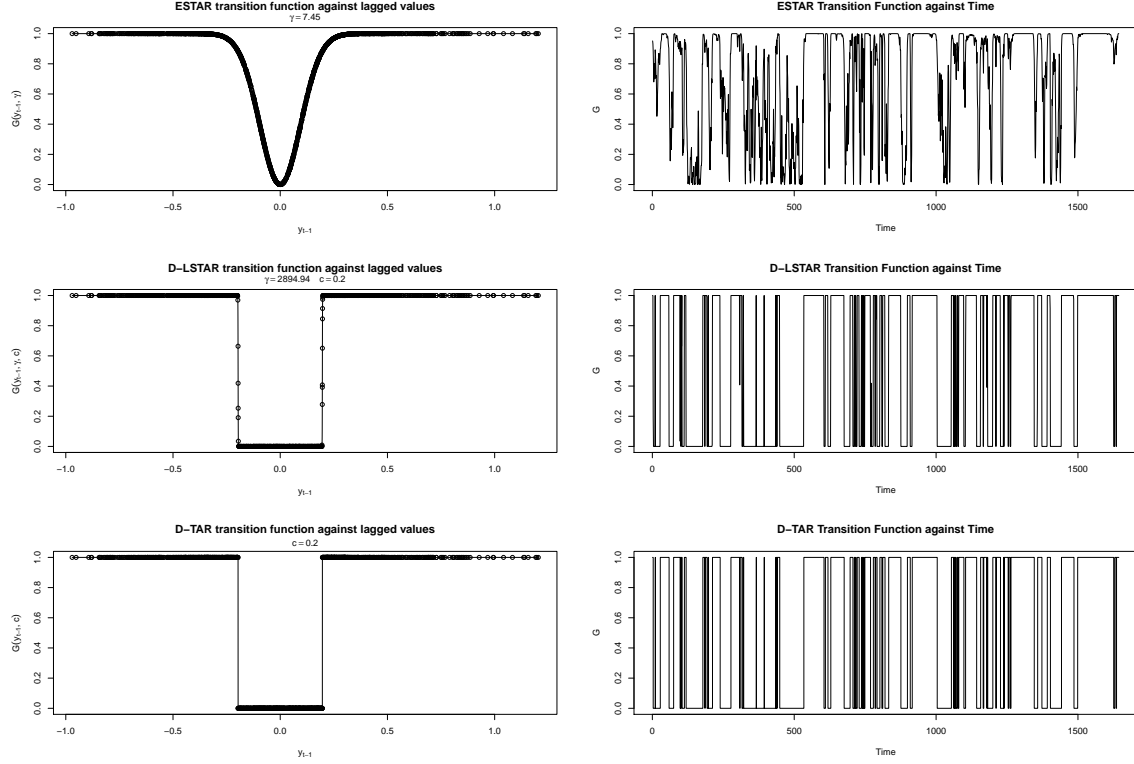


Figure 4: Transition functions obtained from Park-Shintani auxiliary test regressions for the S&P500 price-earnings ratio ($g = 50$).

function leads to a fairly smooth transition function which is in sharp contrast to the one for the D-TAR model. The D-LSTAR transition function looks similar to the D-TAR, but has an additional parameter. The transition functions against time on the right hand side of Figure 4 show similar features. Note that an observation belongs to the inner unit root regime (as defined in Section 2) when the transition function equals zero. A comparison suggests rather different behavior of the ESTAR and the D-TAR model although both mainly agree on which observations belong or are close to the unit root regime. Finally, we note that the minimal unit root t -statistics indicate a rejection at the nominal significance level of 5% for all nonlinear models. Similarly, the ADF test statistic for the linear model is also highly significant. The relatively small negative values for $\hat{\lambda}$ indicate a high persistence in the P/E ratio.

Our second application is the US interest rate spread, defined as $z_t = i_t^S - i_t^L$, where i_t^S and i_t^L are the short and long term interest rate respectively. [Stock and Watson \(1988\)](#) find that interest rates are random walks which share a common trend. From an economic angle this result confirms the expectation hypothesis of the term structure. Investors should be indifferent between the two investment opportunities, so that the expected returns are the same plus a risk premium for the long term investment (or for the short term investment, corresponding with investor preferences). The equilibrium relationship

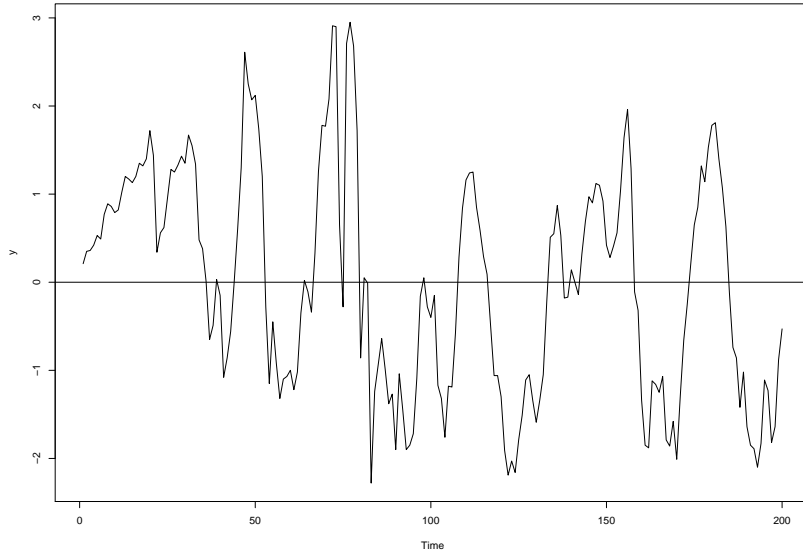


Figure 5: US interest rate spread (demeaned).

has the form $i_t^S - i_t^L = a$, where a is a constant. After demeaning, the equilibrium is at zero.

Balke and Fomby (1997) and Anderson (1997) argue that nonlinear error correction models provide an appropriate framework for studying how transaction costs affect yield movements in the primary US Treasury bill market. Balke and Fomby (1997) use a D-TAR model with a inner unit root to capture the argument that trading and arbitrage only takes place if the deviation from equilibrium is high enough. In this case the transaction costs are the same for all investors. Anderson (1997) uses an ESTAR model instead to allow for heterogeneous transaction costs. Both models outperform linear ones in terms of forecasting, but Anderson (1997) points out that her purpose was not to find the best nonlinear model. She also investigates asymmetry but concludes that the removal of this part changes the model and its properties only slightly.

	$g = 50$				$g = 300$			
	ESTAR	D-LSTAR	D-TAR	AR	ESTAR	D-LSTAR	D-TAR	AR
RSS	59.1668	59.1337	59.5234	64.8142	59.1554	59.1334	59.4263	64.8141
AIC	-1.1828	-1.1733	-1.1768	-1.1117	-1.1830	-1.1733	-1.1784	-1.1117
BIC	-1.1331	-1.1071	-1.1271	-1.0952	-1.1333	-1.1071	-1.1288	-1.0952
HQIC	-1.1878	-1.1800	-1.1818	-1.1134	-1.1880	-1.1800	-1.1834	-1.1134
UR test	-4.9891***	-4.8546***	-5.0015***	-4.0987***	-4.9933***	-4.8915***	-5.0016***	-4.0987***
$\hat{\lambda}$	-0.5031	-0.3504	-0.2130	-0.1086	-0.6441	-0.3442	-0.2072	-0.1086

Table 7: Results obtained from $\inf -t$ auxiliary test regressions for the price-earnings ratio (columns ESTAR, D-LSTAR and D-TAR) and from an ADF test regression for the linear model (column AR).

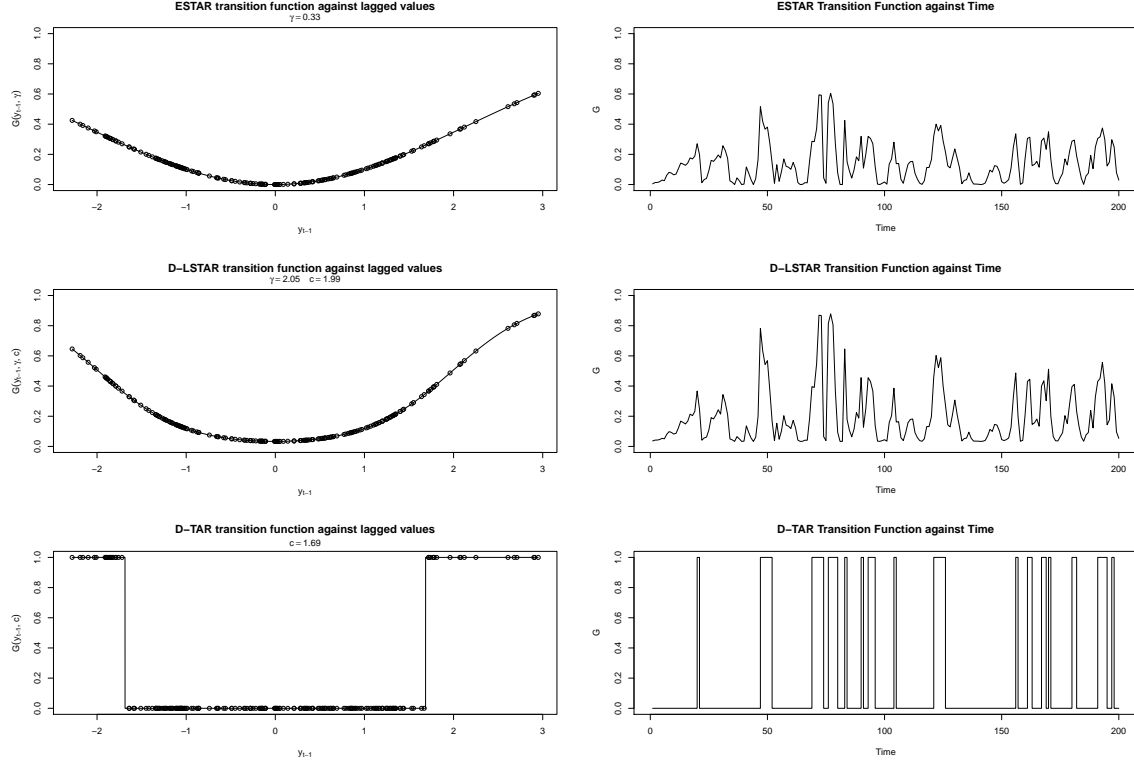


Figure 6: Transition functions obtained from Park-Shintani test regressions for the US interest rate spread ($g = 50$).

The data we use covers the span from 1962:1 to 2011:4 and is obtained from the FREDII database (Fed St. Louis). The number of quarterly observations equals $n = 200$. The short term interest rate has a maturity of three months and the long term interest rate is the ten year government bond yield. The lag length is equal to $p = 2$ using the same procedure as in the first application. A comparison of information criteria as shown in Table 7 clearly suggests the selection of an exponential transition function. All information criteria agree on such a selection. But as in the first example, the best fit is the D-LSTAR model in terms of the RSS. The selected parameters are $\tilde{\gamma} = 0.33$ for the ESTAR model which gives a very smooth transition function. For the D-LSTAR model, we find the grid-points $\tilde{\gamma} = 2.05$ and $\tilde{c} = 1.99$. As seen on the left side of Figure 6, the resulting transition function is similar (to a certain extent) to the one obtained for the ESTAR model. A completely different behavior is observed for the D-TAR model which appears to be least promising in this application. The threshold parameter is selected as $\tilde{c} = 1.69$ indicating a rather wide inner unit root regime. As the US interest rate spread seems to follow a stationary process, see Figure 5, the D-TAR model with a wide inner regime does not fit the data too well. This claim is supported by the unit root test results in Table 7 which strongly indicate evidence against the null hypothesis. The different values for $\hat{\lambda}$ indicate a different level in persistence at first sight. But a

combined consideration of $\hat{\lambda}$ and the maximum value of $G(y_{t-1}, \tilde{\theta})$ shows that this level is similar for all nonlinear models.

Another approach by [Pfann, Schotman, and Tschernig \(1996\)](#) suggests that interest rates in levels are stationary but nonlinear. They apply a two-regime TAR model to show that extremely high interest rates are corrected back. If we use our selection approach for this hypothesis, a D-TAR model is chosen with only two regimes (no data points are in the lower regime). The ESTAR model is not able to capture any dynamics and is very close to a random walk.

6 Conclusion

Model selection in nonlinear time series models is still a challenge in econometrics. While the literature about estimation and specification of a certain type of nonlinearity is quite extensive, procedures and tests to distinguish between different types of nonlinearity are rare. This work focuses on model selection between smooth transition and threshold time series models which are widely applied in practice. We propose a simple specification procedure based on standard regression output to select the appropriate transition function. The starting point of the procedure is the $\inf -t$ unit root test, where the problem of unidentified parameters under the null is solved by a grid-search. The auxiliary regression can be estimated by OLS and information criteria can be calculated easily in order to select the most appropriate transition function. Thereby the procedure itself does not require any complicated estimation of the competing nonlinear models as it uses linear auxiliary regressions instead. The use of standard tests and selection criteria as well as the low computational requirements make the procedure useful for practitioners. The Monte Carlo results suggest that the procedure works well in finite-samples in a variety of settings. The two empirical applications also underline the usefulness of the proposed method in practice.

References

- ANDERSON, H. M. (1997): "Transaction Costs and non-linear Adjustment towards Equilibrium in the US Treasury Bill Market," *Oxford Bulletin of Economics and Statistics*, 59, 465–484.
- BALKE, N. S. AND T. B. FOMBY (1997): "Threshold cointegration," *International Economic Review*, 38, 627–645.
- CAMPBELL, J. Y. AND R. J. SHILLER (1998): "Valuation Ratios and the Long-Run Stock Market Outlook," *The Journal of Portfolio Management*, 24, 11–26.
- CANER, M. AND B. E. HANSEN (2001): "Threshold Autoregression with a Unit Root," *Econometrica*, 69, 1555–1596.
- CHOI, C.-Y. AND Y.-K. MOH (2007): "How useful are tests for unit-root in distinguishing unit-root processes from stationary but non-linear processes?" *Econometrics Journal*, 10, 82–112.
- HALDRUP, N., R. KRUSE, T. TERÄSVIRTA, AND R. T. VARNESKOV (2012): "Unit roots, nonlinearities and structural breaks," CREATES Research Papers 2012-14, School of Economics and Management, University of Aarhus.
- HANSEN, B. E. (1997): "Inference in TAR Models," *Studies in Nonlinear Dynamics & Econometrics*, 2, 1–14.
- HARVEY, D. I. AND S. J. LEYBOURNE (2007): "Testing for time series linearity," *Econometrics Journal*, 10, 149–165.
- KAPETANIOS, G., Y. SHIN, AND A. SNELL (2003): "Testing for a unit root in the nonlinear STAR framework," *Journal of Econometrics*, 112, 359–379.
- KILIAN, L. (1999): "Exchange rates and monetary fundamentals: what do we learn from long-horizon regressions?" *Journal of Applied Econometrics*, 14, 491–510.
- KILIAN, L. AND M. P. TAYLOR (2003): "Why is it so difficult to beat the random walk for exchange rates?" *Journal of International Economics*, 60, 85–107.
- KILIÇ, R. (2004): "Linearity tests and stationarity," *Econometrics Journal*, 7, 55–62.
- KIM, H. AND Y.-K. MOH (2010): "A century of purchasing power parity confirmed: The role of nonlinearity," *Journal of International Money and Finance*, 29, 1398–1405.

- KRUSE, R. AND R. SANDBERG (2010): “Linearity Testing in Time-Varying Smooth Transition Autoregressive Models under Unknown Degree of Persistency,” CRE-ATES Research Papers 2010-36, School of Economics and Management, University of Aarhus.
- LETTAU, M. AND S. V. NIEUWERBURGH (2008): “Reconciling the return predictability evidence,” *Review of Financial Studies*, 21, 1607–1652.
- LUNDBERGH, S., T. TERÄSVIRTA, AND D. VAN DIJK (2003): “Time-Varying Smooth Transition Autoregressive Models,” *Journal of Business & Economic Statistics*, 21, 104–121.
- MCMILLAN, D. G. (2007): “Bubbles in the dividend-price ratio? Evidence from an asymmetric exponential smooth-transition model,” *Journal of Banking & Finance*, 31, 787–804.
- NORMAN, S. AND K. L. PHILLIPS (2009): “What is the Shape of Real Exchange Rate Nonlinearity?” *Working Paper*.
- PARK, J. Y. AND M. SHINTANI (2009): “Testing for a Unit Root against Transitional Autoregressive Models,” *Working Paper*.
- PFANN, G. A., P. C. SCHOTMAN, AND R. TSCHERNIG (1996): “Nonlinear interest rate dynamics and implications for the term structure,” *Journal of Econometrics*, 74, 149–176.
- PSARADAKIS, Z., M. SOLA, F. SPAGNOLO, AND N. SPAGNOLO (2009): “Selecting Nonlinear Time Series Models using Information Criteria,” *Journal of Time Series Analysis*, 30, 369–394.
- RAPACH, D. E. AND M. E. WOHR (2005): “Valuation Ratios and long-horizon Stock Price Predictability,” *Journal of Applied Econometrics*, 20, 327–344.
- (2006): “The out-of-sample forecasting performance on nonlinear models of real exchange rate behavior,” *International Journal of Forecasting*, 22, 341–261.
- SOLLIS, R. (2009): “A simple unit root test against asymmetric STAR nonlinearity with an application to real exchange rates in Nordic countries,” *Economic Modelling*, 26, 118–125.
- STOCK, J. H. AND M. W. WATSON (1988): “Testing for Common Trends,” *Journal of the American Statistical Association*, 83, 1097–1107.

- TERÄSVIRTA, T. (1994): “Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models,” *Journal of the American Statistical Association*, 89, 208–218.
- TERÄSVIRTA, T. (1994): “Estimation, and Evaluation of Smooth Transition Autoregressive Models,” *Journal of the American Statistical Association*, 89, 208–218.
- TONG, H. (1990): *Non-linear time series - A dynamical system approach*, Oxford University Press.
- YOON, G. (2010): “Do real exchange rates really follow threshold autoregressive or exponential smooth transition autoregressive models?” *Economic Modelling*, 27, 605–612.