Discussion Papers
Department of Economics
University of Copenhagen

No. 13-07

Parameter identification in the logistic STAR model*

Line Elvstrøm Ekner and Emil Nejstgaard

Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K., Denmark
Tel.: +45 35 32 30 01 – Fax: +45 35 32 30 00
http://www.econ.ku.dk

ISSN: 1601-2461 (E)

Parameter identification in the logistic STAR model*

Line Elvstrøm Ekner† and Emil Nejstgaard‡
Department of Economics, University of Copenhagen

September 19, 2013

ABSTRACT: We propose a new and simple parametrization of the so-called *speed of transition* parameter of the logistic smooth transition autoregressive (LSTAR) model. The new parametrization highlights that a consequence of the well-known identification problem of the speed of transition parameter is that the threshold autoregression (TAR) is a limiting case of the LSTAR process. We demonstrate how this fact impedes numerical optimization of the original parametrization, whereas this is not the case for the new parametrization. Next, we show that information criteria provide a tool to choose between an LSTAR model and a TAR model; a choice previously based solely on economic theory. Reestimation of two published applications illustrate the usefulness of our findings.

1. INTRODUCTION

Regime switching models have become increasingly popular in the time series literature over the last decades and applied to data from potential regime switching processes such as, e.g., the business cycle, the unemployment rate, exchange rates, prices, interest rates, etc. The majority of the models initiate from the threshold autoregressive (TAR) model first presented by Tong and Lim (1980). Nevertheless, the idea of smooth regime switching was first discussed by Bacon and Watts (1971), but not formalized in terms of a time series model until Chan and Tong (1986) proposed what they called a smoothed threshold autoregressive model as an extension to the TAR model of Tong and Lim (1980). Heavily cited contributions by Luukkonen *et al.* (1988) and Teräsvirta (1994) changed the label from "smoothed threshold" to "smooth transition" resulting in the label smooth transition

*We thank Anders Rahbek, Heino Bohn Nielsen, Andreas Noack Jensen, Rasmus Søndergaard Pedersen for useful comments and Søren Johansen for valuable discussions.
†Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K, Denmark, line.elvstrom.ekner@econ.ku.dk
‡Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K, Denmark, emil.nejstgaard@econ.ku.dk

autoregression (STAR) used today. For an overview of the TAR and STAR literature, see Tong (2011), Teräsvirta *et al.* (2010), and van Dijk *et al.* (2002).

The STAR model differentiates itself from the TAR model by smoothing out the regime switches which in the TAR model take place instantly at a given point in time. The primary economic motivation for the STAR model is that economic time series are often results of decisions made by a large number of economic agents. Even if agents are assumed to make only dichotomous decisions or change their behavior discretely, it is unlikely that they do so simultaneously. Hence, any regime switching in economic time series may be more accurately described as taking place smoothly over time. Moreover, the speed of the regime switching can be of separate interest to an economist, e.g., to analyze how fast the economy adapts to another regime or state of the economy.

The first contribution of this paper is a new and simple parameterization of the speed of transition parameter of the logistic STAR (LSTAR) model. This parameter is particular difficult to identify and, as a result, large and imprecise estimates are often reported in empirical applications. Using likelihood analysis, we study the consequences of this identification problem for estimation and inference in the LSTAR model. We show that the lack of identification, even with a relatively large sample, results in the TAR model (or a very close approximation to the TAR model) as the global maximum of the likelihood function. Due to poor parametrization, this fact impedes the numerical optimization of the likelihood function because the original speed of transition parameter in the TAR cases is infinity. Our new parametrization remedies this by mapping the speed of transition parameter into a much smaller interval. An additional advantage hereof is a simplification of data analysis that makes it easier to identify cases with insufficient support of an LSTAR model compared to a TAR model. Furthermore, we discuss numerical optimization of the LSTAR likelihood function. In particular, we consider the origin of multiple maxima on the likelihood function and grid search methods.

Having established that the TAR model can be the likelihood maximizing solution to the LSTAR model, the second contribution of this paper is a model selection procedure to select between these two models. In the literature of LSTAR models economic theory is used as the only motivation for modeling an LSTAR model instead of a TAR model, see, e.g., Granger and Teräsvirta (1993) and Teräsvirta (1998). However, our new parametrization facilitates a decision based upon the data, possibly in conjunction with economic theory. We show how information criteria provide a neat, but conservative, tool to select an LSTAR model over a TAR model that can be applied if the reseracher wishes to comment on the speed of transition. The related issue of selecting between an LSTAR model and an AR model is not treated in this paper. Although testing such hypothesis of linearity is non-standard, procedures are available and well-described in the literature of both the (L)STAR and TAR models, see Davies (1987), Luukkonen *et al.* (1988), Hansen (1996), and Kristensen and Rahbek (2013).

We show the advantages of the new parametrization by reestimating two published empirical applications. In the first application, the likelihood function for the reparametrized speed of transition parameter reveals that the published result is only a local maximum on the likelihood function, and that the global maximum is a TAR model. In the second example, data contains insufficient information about the speed of transition parameter which then becomes irrelevant, and, as a result, information criteria prefer a TAR model over the published LSTAR model.

The new parametrization can be applied to all kinds of regime switching models where the regime switching is governed by one or more logistic type transition functions. Identification of the speed of transition parameter in the related exponential STAR (ESTAR) model with an exponenetial transition function has recently been studied by Heinen *et al.* (2012). However, the problem is different in the ESTAR model since this model approaches an AR model when the speed of transition approaches infinity and not a TAR model. Hence, their results do no carry over to the LSTAR model. Nevertheless, the new paramtrization is also highly beneficial for estimation of the ESTAR model by facilitating numerical optimization as well as identification of the global maximum of the likelihood funcition.

2. THE MODEL AND THE IDENTIFICATION PROBLEM

We illustrate the identification problem and the benefits of our proposed reparametrization using a simple LSTAR model, cf., Teräsvirta (1994), given by the equations

$$y_t = \alpha y_{t-1} G\left(y_{t-1};\gamma,c\right) + \varepsilon_t,\ \varepsilon_t \sim i.i.d.(0,1) \tag{2.1}$$

and

$$G\left(y_{t-1};\gamma,c\right) = \left(1+\exp\left\{-\gamma(y_{t-1}-c)\right\}\right)^{-1} \tag{2.2}$$

where y_t is an observation at time t of some variable of interest and α is an autoregressive parameter. The transition function G_t (we use $G_t := G\left(y_{t-1};\gamma,c\right)$ as shorthand notation) facilitates smooth regime changes between the two extreme regimes of the process prevailing when G_t is close to its boundaries; $G_t = 0$ results in a pure white noise process and $G_t = 1$ results in a autoregressive process with adjustment α, where $|\alpha| < 1$. The two parameters of the transition function are $\gamma \in \bar{\mathbb{R}}_+$ and $c \in \mathbb{R}$, where $\bar{\mathbb{R}}_+ := \mathbb{R}_+ \cup \infty$. Note that we extend the original definition of the parameter space for γ to include infinity. This extension makes it possible to discuss both the LSTAR model and the TAR model within the same framework. The parameter γ is the speed of transition parameter and c can be interpreted as a threshold parameter. Figure 2.1 illustrates how the functional form of the transition function and the identification area changes with γ. The larger γ is, the smaller is the interval in which observations must lie to provide information about γ. Moreover,

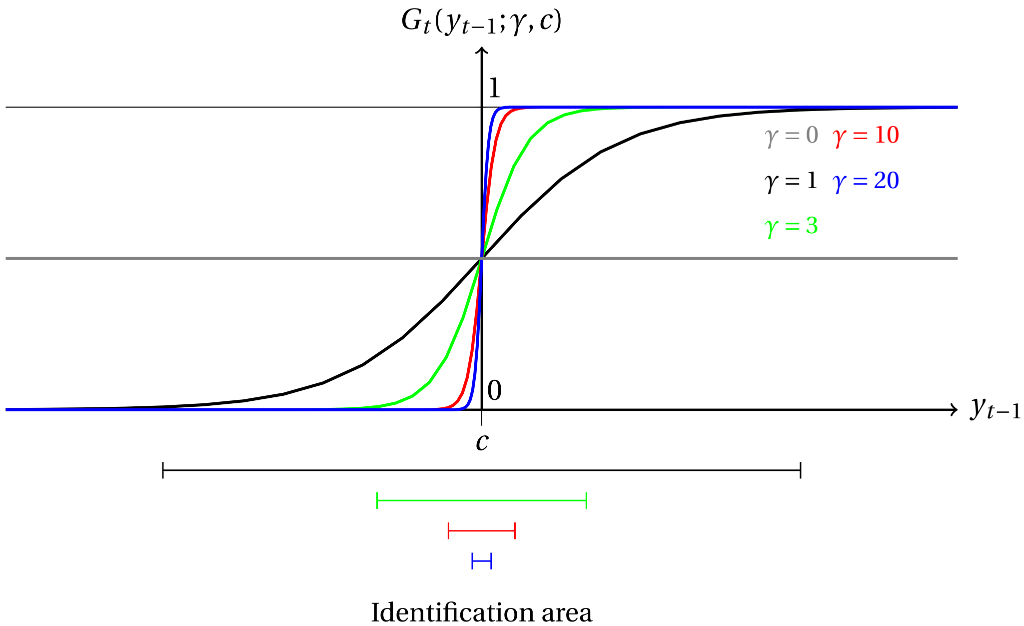


Figure 2.1: *Functional form and identification area of the logistic transition function* $G_t = \left\{1+\exp(-\gamma(y_{t-1}-c))\right\}^{-\gamma}$ *for different values of* γ.

observe from (2.2) and figure 2.1 that $G_t \to \mathbb{I}_{\{y_{t-1}-c>0\}}$ as $\gamma \to \infty$, and that, consequently, the TAR is a limiting case of the LSTAR model prevailing when $\gamma = \infty$. This feature of the model is the heart of the identification problem discussed in this paper. It has the consequence that the first and second order derivatives of the likelihood function tend to zero as $\gamma \to \infty$ (as will be shown), resulting in a log-likelihood function with large flat areas in the direction of γ that impede the numerical optimization.

The related ESTAR model is given by (2.1) and $G\left(y_{t-1};\gamma,c\right) = 1 - \exp\left\{-\gamma(y_{t-1}-c)^2\right\}$. When $\gamma \to \infty$, $G_t \to 0$ (with a single blip at $y_{t-1} = c$) and the ESTAR model approaches a white noise process or, in a more general case, an AR model. Hence, poor identification of the speed of transition parameter is, in contrast to the LSTAR model, often anticipated when testing against a linear model, which is standard in the STAR literature.

3. Likelihood analysis of the speed of transition parameter

LSTAR models are traditionally estimated by maximum likelihood (ML) or non-linear least squares (NLS). The two approaches are equivalent when the errors are assumed *i.i.d.* Gaussian, and thus the essential insights from the following ML analysis carry over to NLS. Before introducing the new parametrization, we illustrate the consequences of the poor original parametrization for estimation and inference in the LSTAR model. We are interested in analyzing only the properties of the ML estimator of γ, and, hence, we ignore estimation of α and c. The log-likelihood function (hereafter abbreviated "likelihood function") is given by

$$\ell_T(\gamma) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\varepsilon_t(\gamma)^2, \tag{3.1}$$

where

$$\varepsilon_t(\gamma) = y_t - \alpha y_{t-1} G(y_{t-1};\gamma, c)$$

and $G(\cdot)$ is the logistic transition function given by (2.2). The individual score in the direction of γ is given by

$$S_t(\gamma) = \frac{\partial \ell_T(\gamma)}{\partial \gamma} = \varepsilon_t(\gamma)\alpha y_{t-1}\left(\frac{\partial G_t}{\partial \gamma}\right) \tag{3.2}$$

where

$$\frac{\partial G_t}{\partial \gamma} = G_t(1-G_t)\left(y_{t-1}-c\right) \tag{3.3}$$

To analyze the information on γ, we consider the Hessian given by

$$H_t(\gamma) = \frac{\partial^2 \ell_T(\gamma)}{\partial \gamma \partial \gamma'} = -\alpha^2 y_{t-1}^2\left(\frac{\partial G_t}{\partial \gamma}\right)^2 + \alpha y_{t-1}\varepsilon_t(\gamma)\left(\frac{\partial^2 G_t}{\partial \gamma \partial \gamma}\right) \tag{3.4}$$

where

$$\frac{\partial^2 G_t}{\partial \gamma \partial \gamma} = G_t(1-G_t)(1-2G_t)\left(y_{t-1}-c\right)^2. \tag{3.5}$$

With the definition of G_t given in (2.2), it holds that $G_t \to \mathbb{I}_{\{y_{t-1}-c>0\}}$ when $\gamma \to \infty$ and, as a consequence, $G_t(1-G_t) \to 0$. Hence,

$$S_t(\gamma) \to 0 \quad \text{and} \quad H_t(\gamma) \to 0 \quad \text{as}\, \gamma \to \infty. \tag{3.6}$$

This result shows that the likelihood function becomes flat when the LSTAR model approximates the TAR model. Thus, the TAR model always represents at least a local maximum of the likelihood function. To illustrate the consequences of (3.6) for estimation, we simulate a dataset from an LSTAR model with $T = 150$, $\gamma = 2$, $c = 0$ and $\alpha = 0.5$. The series and transition function are graphed in figure 8.1 in appendix. The speed of transition is fairly slow and virtually all observations lie in the identification area. The corresponding likelihood function as a function of γ is depicted in figure 3.1(a) below. Observe that the likelihood function gets flatter as the value of γ grows and the maximum is found, roughly, somewhere in the interval $\gamma \in [35;\infty]$. Consequently, the exact value of $\widehat{\gamma}$ becomes arbitrary as it depends on the choice of stopping criteria for the numerical optimizer used to maximize the likelihood function.

In the literature on LSTAR models, $\widehat{\gamma}$ is often reported to have a positive sample bias, see, e.g., Chan and Tong (1986), Medeiros and Veiga (2005), Areosa *et al.* (2011), and Hillebrand *et al.* (2013). However, we suspect that this bias is influenced by estimating γ without recognizing the behavior of the numerical optimizer when the threshold alternative is the global maximum of a model with a logistic transition function. Table 3.1 shows results from a Monto Carlo study in which we have evaluated the estimation bias of $\hat{\gamma}$ while only changing the stopping criterion related to the score of the likelihood function. The positive bias in $\hat{\gamma}$ depends heavily on this criterion. This illustrates that one has to be careful when doing Monte Carlo analysis of LSTAR models.

In empirical applications, researchers tend to fix γ at some arbitrary large value when they are unable to estimate γ since it is infinity. As our parametrization below clarifies, this is approximative to estimating a TAR model, and a more satisfactory solution might be to switch to the TAR framework which by now has a well-developed theoretical framework, see Hansen (1997a), Hansen (2000), and references therein.

3.1. THE δ-PARAMETRIZATION

To avoid the drawbacks of the γ-parametrization in (2.2), we propose the following reparametrization. We define a new parameter $\delta \in (0;1]$, such that

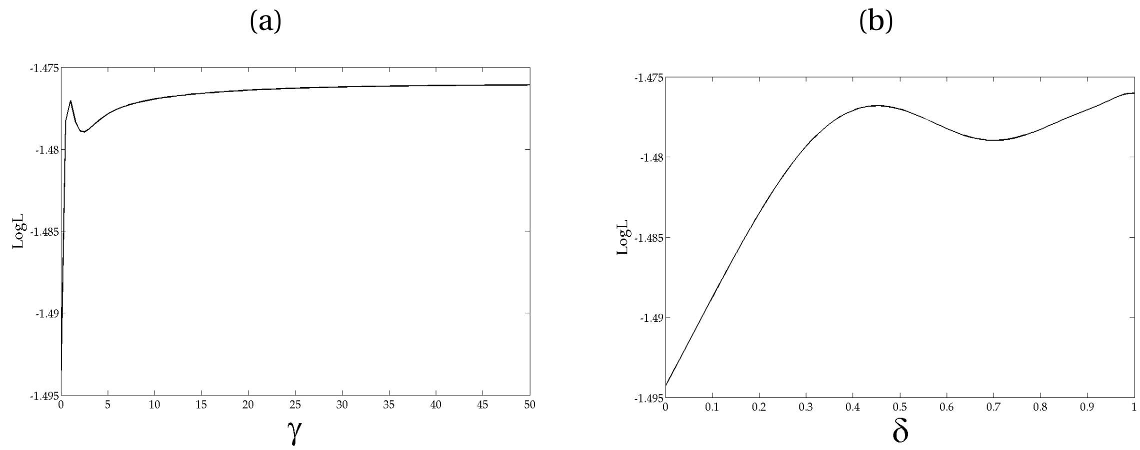
$$\delta = \frac{\gamma}{1+\gamma} \tag{3.7}$$

with $\delta \to 0$ as $\gamma \to 0$ and $\delta \to 1$ as $\gamma \to \infty$. Hence, the transition function in (2.2) is replaced by

$$G\left(y_{t-1};\delta,c\right) = \left(1+\exp\left\{-\frac{\delta}{1-\delta}(y_{t-1}-c)\right\}\right)^{-1} \tag{3.8}$$

The main advantage of this parametrization is that it emphasizes the part of the likelihood function that is of principal interest in an LSTAR model. Essentially, the reparametrization maps $\gamma \in \bar{\mathbb{R}}_{+}$ into $\delta \in (0;1]$, where $\delta \in (0;1)$ is an LSTAR model, $\delta = 1$ is a TAR model, and $\delta = 0$ is a AR model. Of particular importance is the mapping of $\gamma \in [9;\infty]$ into $\delta \in [0.9;1]$. This feature can facilitate numerical optimization of the likelihood function by compressing the large flat area seen in figure 3.1(a) into a much smaller area in figure 3.1(b).

Figure 3.1: *The profiled likelihood function as a function of γ (a) and δ (b). The data set is simulated for $T = 150$, $\gamma = 2/\delta = \frac{2}{3}$, $c = 0$, $\alpha = 0.5$.*



The reparametrization highlights two important aspects that were unclear with the original γ-parametrization. First, the likelihood function is bimodal with a well defined local maximum around $\delta = 0.45$, corresponding to an LSTAR model with $\gamma \approx 0.8$. Apparently, for this particular realization the local maximum undershoots the true value of the speed of transition. Second, the δ-parametrization stresses that the global maximum of the likelihood function is found close to or at the boundary of the parameter space, $\delta = 1$, i.e., corresponding to a TAR model. The fact that the likelihood function continues to increase until $\delta = 1$ is effectively masked in the γ-parametrization.

Next, as can be seen from table 3.1, the criterion dependent positive bias for $\hat{\gamma}$ dissappears when redoing the Monte Carlo experiments using the δ-parametrization. Note that the size of the bias is not comparable across parametrizations due to the different scaling of the parameters. Observe that since the ML estimator is consistent, the bias diminishes as T grows.

Table 3.1: *Estimated bias in $\hat{\gamma}$ and $\hat{\delta}$ as a function of the stopping criterion for the numerical optimizer.*

$S_T(\hat{x}) = \frac{\partial \ell_T(\hat{x})}{\partial \hat{x}}$	$\leq 10^{-2}$	$\leq 10^{-6}$	$\leq 10^{-16}$
$T = 150$			
$\widehat{BIAS}(\hat{\gamma}) = \sum_{m=1}^{M} (\hat{\gamma}_m - \gamma)$	0.9063	8.2439	49.509
$\widehat{BIAS}(\hat{\delta}) = \sum_{m=1}^{M} (\hat{\delta}_m - \delta)$	0.0533	0.0545	0.0545
$T = 300$			
$\widehat{BIAS}(\hat{\gamma}) = \sum_{m=1}^{M} (\hat{\gamma}_m - \gamma)$	0.6594	5.4652	21.781
$\widehat{BIAS}(\hat{\delta}) = \sum_{m=1}^{M} (\hat{\delta}_m - \delta)$	0.0146	0.0148	0.0148

Note: The DGP is $\gamma = 1/\delta = 0.5$, $c = 0$ and $\alpha = 0.5$. $T = 150$, $M = 10,000$ and c and α are fixed at the DGP values in estimation.

4. Selecting between LSTAR and TAR by means of information criteria

The bimodality of the likelihood function seen above is a common small sample property of LSTAR models. Typically, there exists one inner maximum corresponding to an LSTAR model and a maximum on the boundary of the parameter space ($\delta = 1$) corresponding to a TAR model. The simulation considered above in figure 3.1 is an extreme example of this, where the global optimum of the likelihood function is at $\delta = 1$. A more typical case is one with the inner maximum being the global maximum and a local, smaller maximum is found at the TAR solution, see for example figure 6.2(b). A relevant question is therefore whether the likelihood value of the inner maximum is large enough compared to the likelihood value of the TAR maximum to justify the estimation of a speed of transition parameter. One way to investigate this question would be to derive a test for the null-hypothesis of $\delta = 1$. However, such a test is highly non-standard and we save the analysis of this test to future research. Instead we consider model selection based on information criteria.

Information criteria combine a measure of goodness-of-fit with a penalty for model complexity. Comparing information criteria would therefore indicate whether the additional speed of transition parameter of an estimated LSTAR model leads to a notable improvement of fit compared to a corresponding TAR model. Psaradakis *et al.* (2009) pursue this idea and consider selecting between several non-linear autoregressive models by means of classical information criteria. In the following, we conduct similar simulation study for the choice between a TAR and an LSTAR using the proposed reparametrization.

We focus on the Bayesian Information Criterion (BIC), Schwarz (1978), and the Hannan-Quinn Information Criterion (HQIC), Hannan and Quinn (1979), defined as

$$\text{BIC} = -2\ell_T + k\log(T), \quad \text{and} \quad \text{HQIC} = -2\ell_T + 2k\log\big(\log(T)\big), \tag{4.1}$$

where ℓ_T is the log-likelihood value. The preferred model is the one that minimizes the information criteria. In appendix 8.2 we show formally that such a model selection procedure is consistent. We illustrate this selection method using four different models as data generating processes, an LSTAR model with $\delta = \{0.2, 0.5, 0.9\}$ and a TAR model. We simulate $M = 10^4$ data sets and estimate only the speed of transition parameter δ while keeping the remaning parameters fixed at the true values. For each replication, we calculate the percentage selected LSTAR models when applying the two information criteria. The experiment is done for a range of different sample lengths. The selection percentages are given in table 4.1.

Table 4.1: *Percentage selected LSTAR models using information criteria.* $c = 0$.

DGP	LSTAR, $\delta = 0.2$		LSTAR, $\delta = 0.5$		LSTAR, $\delta = 0.9$		TAR, $\delta = 1$	
T	BIC	HQIC	BIC	HQIC	BIC	HQIC	BIC	HQIC
100	48	64	13	25	1	4	0	0
250	82	92	26	45	0	3	0	0
500	98	99	47	69	1	3	0	0
1,000	100	100	76	90	1	4	0	0
10,000	100	100	100	100	1	6	0	0
50,000	100	100	100	100	2	21	0	0
100,000	100	100	100	100	6	42	0	0
1,000,000	100	100	100	100	100	100	0	0

Note: The information criteria BIC and HQIC are defined in (4.1). Only δ is estimated.

The slower the speed of transition, the better the performance of the information criteria. Nevertheless, even with a relatively slow transition speed of $\delta = 0.5$ and $T = 1{,}000$, BIC and HQIC still select a rather large number of incorrect TAR models, 24 and 10 percent, respectively. For the LSTAR model with $\delta = 0.9$ the information criteria are appearently punishing too severely for the additional parameter. The information criteria do not choose the LSTAR model until $T = 1{,}000{,}000$. Again, this shows that while the identification problem for δ is a small sample problem, the label "small sample" is misleading since T needs to be extremely large to get a clear distinction between an LSTAR and a TAR model. I.e., the number of observations in the identification area grows extremely slow with T, especially when the speed of transition is relatively fast. Observe that when the TAR model is the DGP, the information criteria perform surprisingly well.

The results of a repeated Monte Carlo experiment with $c = 1$ in table 4.2 show significantly improvements in the selection rates of the information criteria for small samples. It is peculiar that the power of the information criteria depends on the (fixed) value of the threshold parameter c.

Table 4.2: *Percentage selected LSTAR models using information criteria.* $c = 1$.

DGP	LSTAR, $\delta = 0.2$		LSTAR, $\delta = 0.5$		LSTAR, $\delta = 0.9$		TAR, $\delta = 1$	
T	BIC	HQIC	BIC	HQIC	BIC	HQIC	BIC	HQIC
100	52	67	24	42	4	11	0	0
250	84	92	51	71	6	17	0	0
500	98	99	77	89	9	27	0	0
1,000	100	100	96	98	18	45	0	0
10,000	100	100	100	100	98	100	0	0
50,000	100	100	100	100	100	100	0	0
100,000	100	100	100	100	100	100	0	0
1,000,000	100	100	100	100	100	100	0	0

Note: The information criteria BIC and HQIC are defined in (4.1). Only δ is estimated.

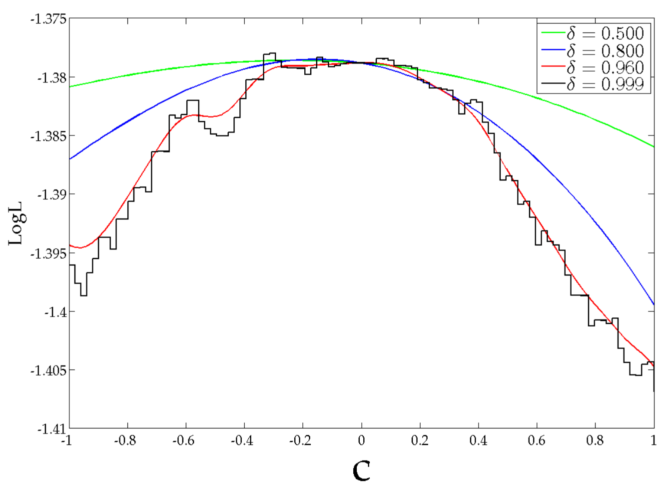
Overall, the results in table 4.1 and 4.2 show that if model selection based on information criteria prefer an LSTAR, it is a clear indication that the speed of transition is slow enough to make a difference compared to the TAR model. On the other hand, if the TAR is chosen, there is a risk that one has incorrectly fixed $\delta = 1$ and selected the TAR model. However, this only means that δ is irrelevant for the model. Hence, information criteria provide a conservative means of selecting LSTAR models over TAR models that can be used if the reseracher wishes to comment on the speed of transition.

5. Estimating LSTAR models

The properties of the likelihood function for LSTAR models discussed so far introduce a number of difficulties for numerical optimization. We observe two seperate problems that have to be taken into account. First, the likelihood function might have a few maxima in the direction of δ, as described in the previous sections. To handle this problem, it is useful to estimate δ with a derivative based optimizer and using different initial values from the parameter space $\delta \in (0;1]$. To ensure that the reached maximum is global, it is important to always calculate the additional likelihood value at the limit, $\delta = 1$.

The second difficulty is that the likelihood function approaches the step-wise likelihood function of a TAR model in the direction of c as $\delta \to 1$. Consequently, many local maxima exist in the direction of c and derivative based optimizers will not work well. For an illustration see figure 5.1, which shows the likelihood as a function of c for different values of δ.

Figure 5.1: *Simulated profiled likelihood functions in the direction of c for different values of* δ*. Data is simulated for* $T = 300$, $c = 0$ *and* $\alpha = 0.5$.



To circumvent the problem of a step-wise likelihood function, a grid search algorithm over c can be performed with an interval that covers observed values of y_{t-1} spanning from, e.g., the 10th to the 90th percentile of the distribution of y_{t-1}. This grid search technique for c is standard in the TAR litterature and ensures that all relevant points for threshold locations are examined. The rest of the parameters are estimated using least squares conditional on the transition function parameters.

When estimating simple models as the one analyzed in this paper, performing a two dimesional grid search over δ and c and drawing the profiled likelihood function is generally informative. This approach allows the researcher to take into account both problems. Note that this proposal is by no means new and is in fact standard practice in the litterature for finding candidates for initial values, see inter alia Bec *et al.* (2008) and Teräsvirta *et al.* (2010, ch. 12). Our contribution is that the δ-parametrization clarifies the reason for doing the grid search, and we emphasize that the main problem of multiple equilibria of the LSTAR model is related to the fact that the likelihood function approaches a step-wise likelihood function as $\delta \to 1$.

In more complex models with several transition function parameters, such a thorough approach might not be possible. For those applications, the researcher could consider performing grids over all the threshold-like parameters while letting the speed of transition parameters vary freely and using only a few initial values to capture the inner maxima corresponding to smooth transition models. Alternatively, heuristic optimization algorithms such as, e.g., simulated annealing or genetic algorithms might provide a means of estimating complex LSTAR models, see inter alia Maringer and Meyer (2008) for a discussion.

6. EMPIRICAL APPLICATIONS

This section reestimates two published LSTAR applications to demonstrate the advantages of the δ-parameterization over the γ-parametrization and model selection based on information criteria. The first application illustrates a situation where the δ-parametrization reveals that the reported maximum of the likelihood function is not the global maximum. In the second application, the δ-parametrization confirms that the global maximum is the reported one, but information criteria prefer the TAR model over the LSTAR model because the regime switching is so fast that estimating the additional speed of transition parameter is superfluous.

6.1. WOLF'S ANNUAL SUNSPOT NUMBERS

Teräsvirta *et al.* (2010, p. 390), illustrate a suggested STAR modeling procedure by analyzing Wolf's annual sunspot numbers dating from 1700 to 1979. The data is published at the Belgian webpage of Solar Influences Data Analysis Center.[1] Following Teräsvirta *et al.* (2010) the series is transformed as: $y_t = 2\left\{(1+z_t)^{1/2} - 1\right\}$ where z_t is the original series. The motivation for transformation is that the transformed series is easier to model than

[1] http://www.sidc.oma.be/sunspot-data/

the untransformed one. The original estimated LSTAR model is reproduced with both parametrizations and given by (standard errors in paranthesis)[2]

$$\begin{aligned} y_t &= \underset{(0.08)}{1.46} y_{t-1} - \underset{(0.13)}{0.76} y_{t-2} + \underset{(0.05)}{0.17} y_{t-7} + \underset{(0.04)}{0.11} y_{t-9} \\ &\quad + (\underset{(0.85)}{2.65} - \underset{(0.13)}{0.54} y_{t-1} + \underset{(0.18)}{0.75} y_{t-2} - \underset{(0.11)}{0.47} y_{t-3} \\ &\quad + \underset{(0.11)}{0.32} y_{t-4} - \underset{(0.07)}{0.26} y_{t-5} - \underset{(0.05)}{0.24} y_{t-8} + \underset{(0.06)}{0.17} y_{t-10}) \times \widehat{G}_t^x \end{aligned} \tag{6.1}$$

$$x = \gamma: \quad \widehat{G}^x = 1 + \exp\{-\underset{(1.11)}{5.46}(y_{t-2} - \underset{(0.36)}{7.88})/\widehat{\sigma}_{y_{t-2}}\}^{-1}$$

$$x = \delta: \quad \widehat{G}^x = 1 + \exp\left\{-\frac{\underset{(0.03)}{0.85}}{1 - \underset{(0.03)}{0.85}}(y_{t-2} - \underset{(0.36)}{7.88})/\widehat{\sigma}_{y_{t-2}}\right\}^{-1}$$

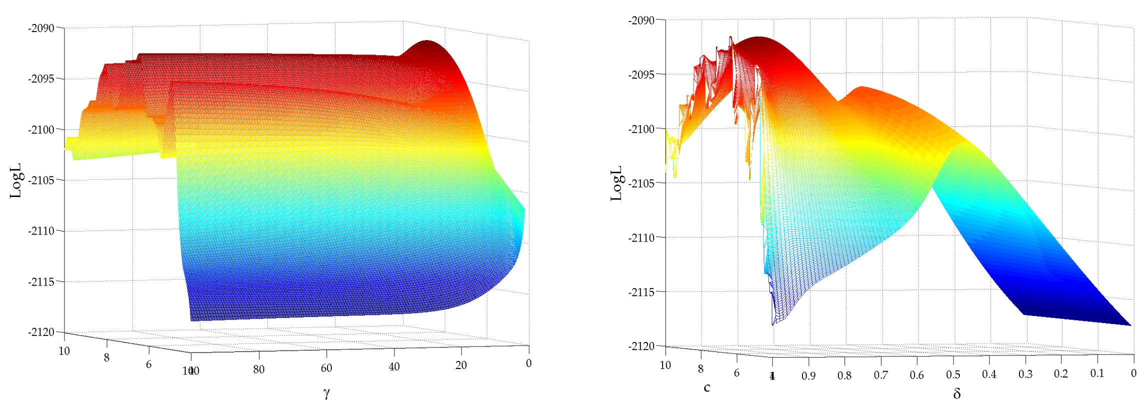
$$T = 270, \qquad RSS = 921.84, \qquad LogL = -2{,}091.2$$

$$BIC = 4{,}260.8, \qquad HQIC = 4{,}230.7$$

The profiled likelihood function in direction of c and γ is showed in figure 6.1(a). The

Figure 6.1: *Profiled likelihood functions of the LSTAR model for Wolf's sunspot numbers, 1710-1979. (a) is for the γ-parametrization and (b) is the for the δ-parametrization.*

(a) (b)



characteristically flatness in the direction of γ is pronounced, and the reported maximum

[2]The normalization by $\widehat{\sigma}_{y_{t-2}}$ in the transition function is standard in the litterature of applied STAR models because it facilitates the choice of grid or initial values for γ, see van Dijk *et al.* (2002).

for $(\widehat{c},\widehat{\gamma}) = (5.46, 7.88)$ appears relatively well-defined. However, figure 6.1(b) reveals that the global maximum is actually the TAR model at the boundary $\delta = 1$, whereas the LSTAR model is only a local maximum. The γ-parametrization has effectively blurred the shape of the likelihood function. At the boundary, the TAR likelihood function is characterized by discrete jumps over the range of c. This implies that performing a careful grid search over potential values of c is crucial for the estimation of c, as discussed in section 5 and, more importantly, that inference on c is non-standard, cf., Chan (1993) and Hansen (1997b). Estimating the TAR model yields[3]

$$\begin{aligned} y_t &= \underset{(0.08)}{1.43} y_{t-1} - \underset{(0.14)}{0.77} y_{t-2} + \underset{(0.05)}{0.17} y_{t-7} + \underset{(0.05)}{0.12} y_{t-9} \\ &\quad +(\underset{(0.70)}{2.69} - \underset{(0.11)}{0.45} y_{t-1} + \underset{(0.18)}{0.69} y_{t-2} - \underset{(0.11)}{0.48} y_{t-3} \\ &\quad +\underset{(0.11)}{0.36} y_{t-4} - \underset{(0.07)}{0.27} y_{t-5} - \underset{(0.05)}{0.21} y_{t-8} + \underset{(0.05)}{0.14} y_{t-10}) \times \mathbb{I}\left(y_{t-2} > 6.39\right). \end{aligned} \tag{6.2}$$

$$T = 270, \quad RSS = 920.66, \quad LogL = -2,090.9$$

$$BIC = 4,254.6, \quad HQIC = 4,226.6$$

While the autoregressive parameters are almost identical to those of the LSTAR model in (6.1), the threshold parameter differs between the models. This TAR maximum is preferred by the information criteria to the reported LSTAR model in (6.1) because the TAR model achieves a lower (higher) value of RSS (LogL) in addition to be one parameter short of the LSTAR model.[4] The TAR maximum (6.2) can easily be reproduced with the δ-parametrization by performing a two-dimensional grid search over c and $\delta \in (0;1]$. A similar exercise for the γ-paramterization produces, depending on the choice of grid for γ as well as the choice of stopping criterion, either the local LSTAR maximum of (6.1) or an invalid maximum with all observations in one regime. Hence, the model that truly maximizes the likelihood function is impossible to estimate with the γ-parametrization because γ is infinity.

Nevertheless, given that an LSTAR process has a TAR model as a small sample property, as found in section 4, and the relatively small sample size of 270, the LSTAR model cannot

[3] The grid search of c is performed over values of y_{t-2}, disregarding values in the lower 10% percentile and upper 90% percentile of the distribution of y_{t-2}. No standard error of $\widehat{c}$ is reported due to the non-standard inference on the threshold parameter in a TAR model.

[4] Teräsvirta *et al.* (2010) reach similiar conclusion when estimating a TAR model for the same data later in the book, though without specifying a measurement. Their TAR model is, however, specified differently and nonnested with (6.2) and (6.1) making direct comparisons infeasible.

be discarded as being the DGP of this sunspot data. In addition, the likelihood function in the region of the local LSTAR maximum in (6.1) and appearing in figure 6.1, seems closely approximated by a quadratic form, and is thus a well defined maximum. Based on these considerations, one could also argue that the LSTAR model may be the DGP of the process.

6.2. U.S. UNEMPLOYMENT RATE

The paper by van Dijk *et al.* (2002) illustrates a suggested STAR modeling cycle which includes, among others, impulse response and forecasting analysis. The date series is the monthly seasonally unadjusted unemployment rate for U.S. males aged 20 and over for the period 1968:6-1989:12.[5]

The LSTAR model is reproduced with both parametrizations and given by (standard errors in paranthesis)

$$\begin{aligned}
\Delta y_t \;=\; & \underset{(0.07)}{0.479} + \underset{(0.07)}{0.645} D_{1,t} - \underset{(0.10)}{0.342} D_{2,t} - \underset{(0.09)}{0.680} D_{3,t} - \underset{(0.11)}{0.725} D_{4,t} - \underset{(0.10)}{0.649} D_{5,t} \\
& - \underset{(0.09)}{0.317} D_{6,t} - \underset{(0.09)}{0.410} D_{6,t} - \underset{(0.09)}{0.501} D_{8,t} - \underset{(0.09)}{0.554} D_{9,t} - \underset{(0.07)}{0.306} D_{10,t} \\
& + [-\underset{(0.01)}{0.040} y_{t-1} - \underset{(0.08)}{0.1460} \Delta y_{t-1} - \underset{(0.06)}{0.101} \Delta y_{t-6} + \underset{(0.06)}{0.097} \Delta y_{t-8} - \underset{(0.06)}{0.123} \Delta y_{t-10} \\
& + \underset{(0.07)}{0.129} \Delta y_{t-13} - \underset{(0.06)}{0.103} \Delta y_{t-15}] \times [1 - \widehat{G}_t^x] \\
& + [-\underset{(0.01)}{0.011} y_{t-1} + \underset{(0.08)}{0.225} \Delta y_{t-1} + \underset{(0.08)}{0.307} \Delta y_{t-2} - \underset{(0.07)}{0.119} \Delta y_{t-7} - \underset{(0.09)}{0.155} \Delta y_{t-13} \\
& - \underset{(0.09)}{0.215} \Delta y_{t-14} - \underset{(0.09)}{0.235} \Delta y_{t-15}] \times \widehat{G}_t^x
\end{aligned} \tag{6.3}$$

$$\begin{aligned}
x = \gamma: &\quad \widehat{G}^x = 1 + \exp\{-\underset{(21.75)}{23.15} (\Delta_{12} y_{t-1} - \underset{(0.04)}{0.274}) / \widehat{\sigma}_{\Delta_{12} y_{t-1}}\}^{-1} \\
x = \delta: &\quad \widehat{G}^x = 1 + \exp\left\{ - \frac{\underset{(0.04)}{0.96}}{1 - \underset{(0.04)}{0.96}} (\Delta_{12} y_{t-1} - \underset{(0.04)}{0.274}) / \widehat{\sigma}_{\Delta_{12} y_{t-1}} \right\}^{-1}
\end{aligned}$$

$$T = 240, \quad RSS = 8.178, \quad LogL = -725.0$$

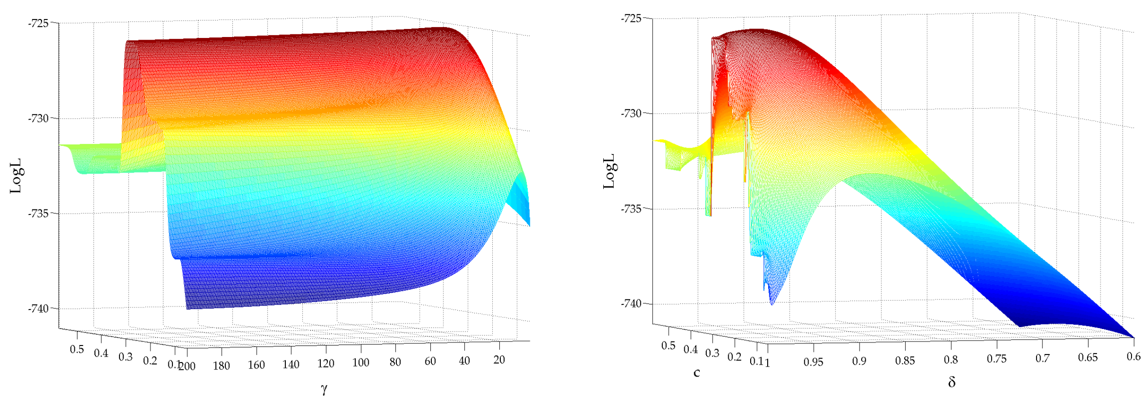
$$BIC = 1,597.9, \quad HQIC = 1,541.8$$

[5]The series is constructed from data on the unemployment level and labor force for the particular subpopulation. These two series are published together with Gauss programs used to estimate their model at http://swopec.hhs.se/hastef/abs/hastef0380.htm.

$D_{s,t}$ is monthly dummy variables where $D_{s,t} = 1$ if observation t corresponds to month s and $D_{s,t} = 0$ otherwise. van Dijk *et al.* (2002) have sequentially removed all variables with a t-statistic lower than 1 in absolute value. Observe that γ is rather large and imprecisely estimated indicating that data contains little information about the size of this parameter. The profiled likelihood functions for the two parametrizations are displayed in figure 6.2. Because $\widehat{\gamma}$ is so large, the maximum is almost blurred by the flatness of the γ-likelihood

Figure 6.2: *Profiled likelihood functions of the LSTAR model for U.S. male unemployment rate, 1968:6-1989:12. (a) is for the γ-parametrization and (b) is the for the δ-parametrization.*

(a) (b)



function in figure 6.2(a). In contrast, the δ-likelihood function in figure 6.2(b) confirms that the reported maximum is in fact the global maximum of the likelihood function. Interestingly, the δ-likelihood function shows that the local TAR maximum at the boundary leads to only a minor drop in likelihood value compared to the LSTAR model. To check whether this TAR model is preferred by information criteria, the TAR model is estimated and given by[6]

$$
\begin{aligned}
\Delta y_t \;=\; & \underset{(0.07)}{0.473} + \underset{(0.07)}{0.644} D_{1,t} - \underset{(0.10)}{0.343} D_{2,t} - \underset{(0.09)}{0.675} D_{3,t} - \underset{(0.11)}{0.721} D_{4,t} - \underset{(0.10)}{0.641} D_{5,t} \\
& - \underset{(0.09)}{0.308} D_{6,t} - \underset{(0.09)}{0.410} D_{6,t} - \underset{(0.08)}{0.505} D_{8,t} - \underset{(0.09)}{0.546} D_{9,t} - \underset{(0.07)}{0.295} D_{10,t} \\
& + [-\underset{(0.01)}{0.040} y_{t-1} - \underset{(0.08)}{0.140} \Delta y_{t-1} - \underset{(0.06)}{0.094} \Delta y_{t-6} + \underset{(0.06)}{0.092} \Delta y_{t-8} - \underset{(0.06)}{0.116} \Delta y_{t-10} \\
& + \underset{(0.07)}{0.136} \Delta y_{t-13} - \underset{(0.06)}{0.106} \Delta y_{t-15}] \times \mathbb{I}\left(\Delta_{12} y_{t-1} \leq 0.268\right) \\
& [-\underset{(0.01)}{0.012} y_{t-1} + \underset{(0.08)}{0.227} \Delta y_{t-1} + \underset{(0.08)}{0.307} \Delta y_{t-2} - \underset{(0.07)}{0.094} \Delta y_{t-7} - \underset{(0.09)}{0.146} \Delta y_{t-13}
\end{aligned}
$$

[6]Similar to the previous TAR estimation, the grid search of c is performed over values of $\Delta_{12} y_{t-1}$, disregarding values in the lower 10% percentile and upper 90% percentile of the distribution of $\Delta_{12} y_{t-1}$. No standard error of $\widehat{c}$ is reported due to the non-standard inference on the threshold parameter in a TAR model.

$$+ - \underset{(0.09)}{0.211}\Delta y_{t-14} - \underset{(0.09)}{0.216}\Delta y_{t-15}] \times \mathbb{I}\left(\Delta_{12} y_{t-1} > 0.268\right) \tag{6.4}$$

$$T = 240, \quad RSS = 8.191, \quad LogL = -725.3$$

$$BIC = 1,593.2, \quad HQIC = 1,539.1$$

The information criteria prefer this TAR model implying that the speed of transition is too poorly estimated to make a difference.

This application highlights one of the key points of the present paper, namely that a large and imprecise estimated γ implies that the LSTAR model is effectively a TAR model. Estimation of the LSTAR model is too much to ask of the data. The δ-parametrization clarifies this and, hence, such an analysis may continue by applying the TAR model.

7. Conclusion

Regime swicthing models characterized by smooth transitions only differ from discrete regime switching models by the speed of transition parameter. Thus, estimation and identification of this parameter is essential not only for economic interpretation but also for model selection. Nevertheless, the identification problem and its conseqences for estimation have received little attention in the STAR literature. We show that the original parameterization of the speed of transition parameter is impractical because the likelihood function is chacterized by large flat areas implying that the size of the estimate may depend on the arbitrary chosen stopping criteria of the numerical optimizer. To circumvent this problem, we propose a new and simple reparamterization of this parameter. The reparametrization maps the parameter space of the original parameter into a much smaller interval which facilitates identifying the global maximum of the likelihood function as well as numerical optimization. By means of this new speed of transition parameter we show that the TAR model can be the global maximum of a LSTAR likelihood function, while it, by construction, is always at least a local maximum. Hence, justifying the additional parameter of the LSTAR model becomes important. Instead of relying solely on economic theory when chosing between these two models, we suggest to use information criteria. We show that information criteria provide a conservative model selection tool that can be applied if the researcher wishes to comment of the speed of transition. Acknowledging that the LSTAR model considered in this paper is simple and the presented simulation results only apply to this particular framework, the new parametrization provides general insights on the shape of the likelihood function in directions of the two parameters of the transition function that can be generalized to a broad range of other models within the smooth switching litterature. For example, the double-logistic smooth transition (D-LSTAR), the Multi Regime Smooth Transition Autoregression (MR-STAR) and the logistic autoregressive conditional root (LACR) model, see, e.g. , **?** and Bec *et al.* (2008).

REFERENCES

AREOSA, W. D., M. MCALEER, AND M. C. MEDEIROS (2011): "Moment-based estimation of smooth transition regression models with endogenous variables." *Journal of Econometrics*, 165(1):100–111.

BACON, D. W. AND D. G. WATTS (1971): "Estimating the Transition Between Two Intersecting Straight Lines." *Biometrika*, 58:525–534.

BEC, F., A. RAHBEK, AND N. SHEPHARD (2008): "The ACR Model: A Multivariate Dynamic Mixture Autoregression*." *Oxford Bulletin of Economics and Statistics*, 70(5):583–618. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1468-0084.2008.00512.x/abstract.

CHAN, K. S. (1993): "Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model." *The Annals of Statistics*, 21(1):520–533.

CHAN, K. S. AND H. TONG (1986): "On estimating thresholds in autoregressive models." *Journal of Time Series Analysis*, 7(3):179–190.

DAVIES, R. B. (1987): "Hypothesis testing when a nuisance parameter is present only under the alternative." *Biometrika*, 74(1):33. URL http://biomet.oxfordjournals.org/content/74/1/33.abstract.

GRANGER, C. AND T. TERÄSVIRTA (1993): *Modelling nonlinear economic relationships.* Oxford University Press, USA.

HANNAN, E. J. AND B. G. QUINN (1979): "The determination of the order of an autoregression." *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 190–195.

HANSEN, B. (1997a): "Inference in TAR models." *Studies in nonlinear dynamics and econometrics*, 2(1):1–14.

HANSEN, B. E. (1996): “Inference when a nuisance parameter is not identified under the null hypothesis.” *Econometrica: Journal of the Econometric Society*, page 413–430. URL `http://www.jstor.org/stable/2171789`.

——— (1997b): “Inference in TAR Models.” *Studies in Nonlinear Dynamics & Econometrics*, 2(1):1.

——— (2000): “Sample splitting and threshold estimation.” *Econometrica*, 68(3):575–603.

HEINEN, F., S. MICHAEL, AND P. SIBBERTSEN (2012): “Weak identification in the ESTAR model and a new model.” *Journal of Time Series Analysis*.

HILLEBRAND, E., M. C. MEDEIROS, AND J. XU (2013): “Asymptotic Theory for Regressions with Smoothly Changing Parameters.” *Journal of Time Series Econometrics Economics and Business*.

JENSEN, S. T. AND A. RAHBEK (2007): “On the Law of Large Numbers for (geometrically) Ergodic Markov Chains.” *Econometric Theory*, 23(04):761–766.

KRISTENSEN, D. AND A. RAHBEK (2013): “Testing and Inference in Nonlinear Cointegrating Vector Error Correction Models.” *Econometric Theory, forthcoming*.

LUUKKONEN, R., P. SAIKKONEN, AND T. TERÄSVIRTA (1988): “Testing linearity against smooth transition autoregressive models.” *Biometrika*, 75:491–499.

MARINGER, D. AND M. MEYER (2008): “Smooth Transition Autoregressive Models–New Approaches to the Model Selection Problem.” *Studies in Nonlinear Dynamics & Econometrics*, 12(1):5.

MEDEIROS, M. C. AND Á. VEIGA (2005): “A flexible coefficient smooth transition time series model.” *Neural Networks, IEEE Transactions on*, 16(1):97–113.

PSARADAKIS, Z., M. SOLA, F. SPAGNOLO, AND N. SPAGNOLO (2009): “Selecting nonlinear time series models using information criteria.” *Journal of Time Series Analysis*, 30(4):369–394.

SCHWARZ, G. (1978): “Estimating the dimension of a model.” *The annals of statistics*, 6(2):461–464.

TERÄSVIRTA, T. (1994): “Specification, estimation, and evaluation of smooth transition autoregressive models.” *Journal of American Statistical Association*, 89:208–218.

——— (1998): “Modelling economic relationships with smooth transition regressions.” *Handbook of Applied Economic Statistics, Marcel Dekker: New York*, pages 507–552.

TERÄSVIRTA, T., D. TJØSTHEIM, AND C. GRANGER (2010): *Modelling Nonlinear Economic Time Series.* Advanced Texts in Econometrics Series. Oxford University Press.

TONG, H. (2011): “Threshold models in time series analysis–30 years on.” *Statistics and its Interface*, 4(2):107–118.

TONG, H. AND K. S. LIM (1980): “Threshold autoregression, limit cycles and cyclical data.” *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 245–292.

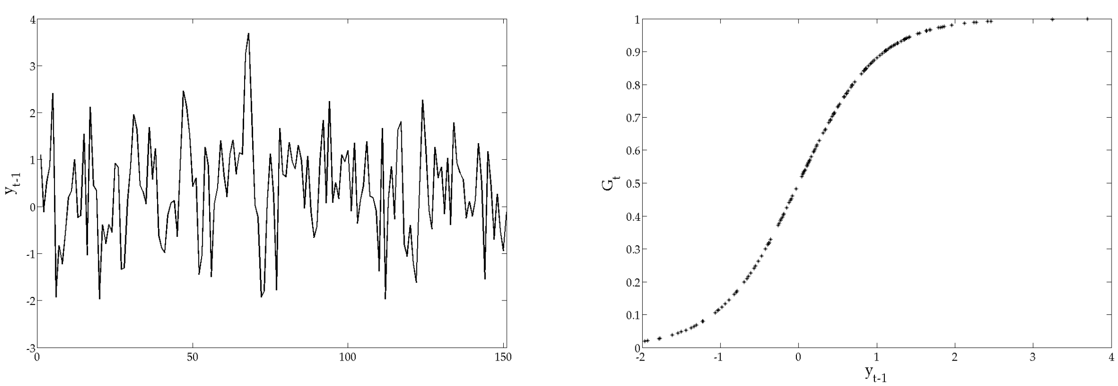
VAN DIJK, D., T. TERÄSVIRTA, AND P. H. FRANSES (2002): “Smooth transition autoregressive models - a survey of recent developments.” *Economic Reviews*, 21:1–47.

8. APPENDIX

8.1. SIMULATED LSTAR PROCESS AND LOGISTIC TRANSITION FUNCTION

Figure 8.1: *Simulated data series (a) and transition function (b) for the LSTAR model (2.1) with* $\gamma = 2$*,* $c = 0$*,* $\alpha = 0.5$ *and* $T = 150$*.*

(a) (b)



8.2. CONSISTENCY OF INFORMATION CRITERIA

We show in the following that the BIC and HQIC are consistent as model selection procedures when considering the choice between an LSTAR and TAR model. The log-likelihood functions of the LSTAR (ℓ_T^S) and TAR (ℓ_T^S) as functions of δ and y_t are (ignoring constants) given by

$$\ell_T^S = -\frac{1}{2}\sum_{t=1}^{T}\left(y_t - \alpha y_{t-1} G\left(\delta; y_{t-1}\right)\right)^2$$

with G_t defined in (3.8) and

$$\ell_T^T = -\frac{1}{2}\sum_{t=1}^{T}\left(y_t - \alpha y_{t-1} I\left(y_{t-1}\right)\right)^2$$

The information criteria have the form

$$IC = -2\ell_T + kc_T \tag{8.1}$$

where c_T is a penalty for model complexity that increases with T. The term c_T is given by $\log(T)$ and $2\log(\log(T))$ for BIC and HQIC, respectively. The number of estimated parameters is given by $k^S = 1$ for the LSTAR model and $k^T = 0$ for the TAR model. The true values of k and δ are denoted k_0 and δ_0, respectively. A selection procedure for k based on minimizing an information criterion is said to be consistent if $\hat{k} \xrightarrow{p} k_0$ as $T \to \infty$, where

$$\hat{k} = \underset{k=1,0}{\arg\min}\{IC(k)\} \tag{8.2}$$

and $IC(k)$ is of the form (8.1).

Proposition 1. *With $\hat{k}$, k_0 and c_T defined as in the previous section, it holds that $\hat{k} \xrightarrow{p} k_0$ if and only if*

$$\lim_{T\to\infty} c_T = \infty \tag{8.3}$$

and

$$\lim_{T\to\infty} (T^{-1} c_T) = 0. \tag{8.4}$$

Proof. First, consider the case where the TAR model is the DGP such that $k_0 = k^T = 0$. The probability of choosing the incorrect LSTAR model is given by $\Pr\left[IC^S < IC^T\right]$. Observe that as $T \to \infty$,

$$\Pr\left[IC^S < IC^T\right] = \Pr\left[-2\ell_T^S + k^S c_T < -2\ell_T^T + k^T c_T\right] = \Pr\left[\ell_T^S - \ell_T^T > \frac{1}{2}c_T\right] \to 0,$$

since by (8.3) $\lim_{T\to\infty}\left(\frac{1}{2}c_T\right) = \infty$ and $\lim_{T\to\infty}\left(\ell_T^S - \ell_T^T\right) = 0$.

Next, if an LSTAR model is the DGP, then the probability of choosing the incorrect TAR model is $\Pr\left[IC^T < IC^S\right]$. It holds that as $T \to \infty$,

$$\Pr\left[IC^T < IC^S\right] = \Pr\left[-T^{-1}2\left(\ell_T^T - \ell_T^S\right) < T^{-1}c_T\right] \to 0 \tag{8.5}$$

since by (8.4) $\lim_{T\to\infty}(T^{-1}c_T) = 0$ and $\lim_{T\to\infty}\left(-T^{-1}2\left(\ell_T^T - \ell_T^S\right)\right) = K$, where $K > 0$. The validity of the second statement is proven in the following.

Recall that the DGP is given by the LSTAR process,

$$y_t = \alpha y_{t-1} G\left(\delta_0; y_{t-1}\right) + \varepsilon_t \tag{8.6}$$

with $\varepsilon_t \sim i.i.d.(0,1)$. Consider now the statistic

$$-T^{-1}2\left(\ell_T^T - \ell_T^S\right) = T^{-1}\sum_{t=1}^{T}\left(y_t - \alpha y_{t-1} I\left(y_{t-1}\right)\right)^2 - T^{-1}\sum_{t=1}^{T}\left(y_t - \alpha y_{t-1} G\left(\hat{\delta}; y_{t-1}\right)\right)^2.$$

Replacing y_t by the DGP in (8.6) and using the definitions $\varphi_t := G\left(\delta_0; y_{t-1}\right) - I\left(y_{t-1}\right)$ and $\psi_t := G\left(\delta_0; y_{t-1}\right) - G\left(\hat{\delta}; y_{t-1}\right)$, one obtains

$$-T^{-1}2\left(\ell_T^T - \ell_T^S\right) = T^{-1}\sum_{t=1}^{T}\alpha^2 y_{t-1}^2\left\{\varphi_t^2 - \psi_t^2\right\} + T^{-1}\sum_{t=1}^{T}2\alpha y_{t-1}\varepsilon_t\left\{\varphi_t - \psi_t\right\}. \tag{8.7}$$

Note that

$$E\left[2\alpha y_{t-1}\varepsilon_t\left\{\varphi_t - \psi_t\right\} \mid y_{t-1}\right] = 2\alpha y_{t-1} E\left[\varepsilon_t | y_{t-1}\right]\left\{\varphi_t - \psi_t\right\} = 0,$$

where the second equality holds by the assumption of $E\left[\varepsilon_t\right] = 0$. Moreover, observe that by consistency of $\hat{\delta}$ and since $\delta_0 \in]0;1[$, we have

$$\Pr\left[\varphi_t^2 - \psi_t^2 \le 0\right] \to 0$$

as $T \to \infty$. Hence, by the law of large numbers for geometrically ergodic Markov chains (see e.g. Jensen and Rahbek (2007)) it holds that

$$T^{-1}\sum_{t=1}^{T}\alpha^2 y_{t-1}^2\left\{\varphi_t^2 - \psi_t^2\right\} \to E\left[T^{-1}\sum_{t=1}^{T}\alpha^2 y_{t-1}^2\left\{\varphi_t^2 - \psi_t^2\right\}\right] = K$$

where $K > 0$ provided that y_t has a non-zero variance. This completes the proof. $\square$