



Contemporaneous threshold autoregressive models: Estimation, testing and forecasting[☆]

Michael J. Dueker^{a,*}, Martin Sola^{b,c}, Fabio Spagnolo^d

^aResearch Division Federal Reserve Bank of St. Louis, Federal Reserve Bank of St. Louis, P.O. Box 442 St. Louis, MO 63166-0442, USA

^bUniversidad Torcuato di Tella, Argentina

^cBirkbeck College, University of London, UK

^dBrunel University, UK

Available online 6 December 2006

Abstract

This paper proposes a contemporaneous smooth transition threshold autoregressive model (C-STAR) as a modification of the smooth transition threshold autoregressive model surveyed in Teräsvirta [1998. Modelling economic relationships with smooth transition regressions. In: Ullah, A., Giles, D.E.A. (Eds.), Handbook of Applied Economic Statistics. Marcel Dekker, New York, pp. 507–552.], in which the regime weights depend on the ex ante probability that a latent regime-specific variable will exceed a threshold value. We argue that the contemporaneous model is well suited to rational expectations applications (and pricing exercises), in that it does not require the initial regimes to be predetermined. We investigate the properties of the model and evaluate its finite-sample maximum likelihood performance. We also propose a method to determine the number of regimes based on a modified Hansen [1992. The likelihood ratio test under nonstandard conditions: testing the Markov switching model of GNP. *Journal of Applied Econometrics* 7, S61–S82.] procedure. Furthermore, we construct multiple-step ahead forecasts and evaluate the forecasting performance of

[☆]Helpful comments by Arnold Zellner, an associate editor, and two anonymous referees are gratefully acknowledged. We especially would like to thank Zacharias Psaradakis for numerous conversations about the paper. We also would like to thank John Driffill, Walter Enders and Ron Smith for comments on a previous version of the paper. Finally we thank Maria Fernandez Vidal for excellent research assistance.

*Corresponding author. Tel.: +1 314 444 8594; fax: +1 314 444 8731.

E-mail address: mducker@stls.frb.org (M.J. Dueker).

the model. Finally, an empirical application of the short term interest rate yield is presented and discussed.

© 2006 Elsevier B.V. All rights reserved.

JEL classification: C22; E31; G12

Keywords: Smooth transition threshold autoregressive; Forecasting; Nonlinear models

1. Introduction

In recent years, a rich class of models has appeared in which economic time series are allowed to undergo regime shifts. One hallmark of these models is that the public cannot anticipate perfectly the regime shifts and, in many cases, the public can only infer regime shifts up to a probability. Such beliefs concerning the state of the business cycle or economic policies may, in turn, affect the stochastic properties of the economic time series under analysis. For example, [Hamilton \(1988\)](#) introduced a regime switching model of interest rates in which the unobserved states evolve according to a Markov chain process. In this type of model, the public is allowed to learn about the underlying state of the economy and to use this knowledge when pricing bonds. Other popular nonlinear autoregressive models that account for a similar phenomenon are the threshold models of [Tong \(1978, 1983\)](#) and [Tong and Lim \(1980\)](#) and the smooth transition threshold autoregressive (STAR) model of [Chan and Tong \(1986\)](#), [Luukkonen et al. \(1988\)](#), and [Teräsvirta \(1994\)](#). These models reflect the idea that variables such as interest rates might have different dynamics when rates are unusually high. In particular [Pfann et al. \(1996\)](#) used self-exciting threshold autoregressive (SETAR) models to characterize the evolution of the interest rates and found mean reversion only when the level of the interest rates was above a certain threshold. Moreover, [Aït-Sahalia \(1996\)](#) has shown that several anomalies of the term structure of interest rates can be accounted for only by using nonlinear modeling.¹

In this paper we propose a new class of contemporaneous smooth transition threshold autoregressive (C-STAR) model, in which the regime weights depend on the ex ante probability that a latent regime-specific variable will exceed a threshold value. Another key feature of the C-STAR is that its transition function depends on all the parameters of the model as well as on the data. These characteristics allow the model to generate a wide variety of empirical distributions. Therefore, we analyze in detail the response of the transition function to changes in all the parameters of the model, how different parameter configurations affect the empirical distribution of the data generated by the model and its stability properties.

Since the C-STAR model is continuous with respect to all the parameters of the model, we estimate them jointly by maximum likelihood (ML) and evaluate the quality of asymptotic approximations to the finite-sample distribution. We propose a procedure to

¹Applications of these models include: [Tiao and Tsay \(1994\)](#) and [Potter \(1995\)](#) to US GNP; [Rothman \(1998\)](#), [Caner and Hansen \(1998\)](#) and [Koop and Potter \(1999\)](#) to unemployment rates; [Obstfeld and Taylor \(1997\)](#) to real exchange rates; [Enders and Granger \(1998\)](#) to the term structure of interest rates; [Pesaran and Potter \(1997\)](#) to business cycle relationships. For excellent surveys of STAR models, see [Teräsvirta \(1998\)](#), [Potter \(1999\)](#) and [van Dijk et al. \(2002\)](#).

assess whether the model is a valid representation of the data based on testing a linear AR against the C-STAR alternative. We also propose a method for obtaining (analytically) multi-step out-of-sample forecasts for the C-STAR that computes the whole tree of the possible future values and evaluates the probability of the alternative future paths. A Monte Carlo study of its forecasting performance is presented. Finally we carry out an empirical application (using short-term U.S. interest rates) to assess the in- and out-of-sample performance of the C-STAR relative to alternative switching models.

The paper is organized as follows. Section 2 introduces the C-STAR model and discusses its properties. The finite-sample performance of the ML estimator and of the related statistics are evaluated by simulation in Section 3. Section 4 proposes a procedure to determine the number of regimes of the C-STAR model. Section 5 evaluates the forecasting performance of the C-STAR model. Section 6 presents the empirical application. Section 7 summarizes and concludes.

2. A contemporaneous threshold autoregressive model

The C-STAR model proposed in this paper is a special case of the STAR model. In a STAR model, the dependent variable, y_t , is a function of two (or more) autoregressive processes that are averaged according to a weighting function, $0 \leq G(z_{t-1}) \leq 1$, where the argument, z_{t-1} , is a predetermined variable

$$y_t = G(z_{t-1})y_{0t} + (1 - G(z_{t-1}))y_{1t},$$

where

$$y_{it} = \mu_i + \alpha_1^i y_{t-1} + \dots + \alpha_p^i y_{t-p} + \sigma_i \varepsilon_t, \quad i = 0, 1, \tag{1}$$

where $\{\varepsilon_t\}$ are independent, identically distributed (*i.i.d.*) random variables, independent of y_{t-1}, y_{t-2}, \dots , with $\mathbb{E}[\varepsilon_t] = 0$ and $\mathbb{E}[\varepsilon_t^2] = 1$; p is a positive integer; σ_0 and σ_1 are positive constants; μ_0, μ_1, α_j^0 and α_j^1 ($j = 1, \dots, p$) are real constants.

STAR models have been extensively used in the analysis of both economic and financial data. In this literature, the main feature that differentiates alternative STAR models is the choice of the transition function. For the specific model that we propose, let

$$z_t = (y_t, y_{t-1}, \dots, y_{t-p+1})', \quad \delta = (1, 0, \dots, 0) \tag{2}$$

and

$$A_i = \begin{bmatrix} \alpha_1^i & \alpha_2^i & \alpha_3^i & \dots & \alpha_{p-1}^i & \alpha_p^i \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix}, \quad i = 0, 1. \tag{3}$$

The weighting function we use is

$$G(z_{t-1}) = \frac{\Phi((y^* - \mu_0 - \delta A_0 z_{t-1})/\sigma_0)}{\Phi((y^* - \mu_0 - \delta A_0 z_{t-1})/\sigma_0) + [1 - \Phi((y^* - \mu_1 - \delta A_1 z_{t-1})/\sigma_1)]}, \tag{4}$$

where y^* is the threshold parameter and $\Phi(\cdot)$ is the standard normal distribution function. The key interpretation of this STAR model is that

$$\Phi((y^* - \mu_0 - \delta A_0 z_{t-1})/\sigma_0) = P(y_{0t} < y^* | z_{t-1}, \Theta_0) \tag{5}$$

and

$$[1 - \Phi((y^* - \mu_1 - \delta A_1 z_{t-1})/\sigma_1)] = P(y_{1t} \geq y^* | z_{t-1}, \Theta_1),$$

where Θ_i denotes the parameters corresponding with regime i , i.e., (μ_i, A_i, σ_i) .

Notice that we can rewrite Eq. (1) as

$$y_t = \frac{P(y_{0t} < y^* | z_{t-1}, \Theta_0)y_{0t} + P(y_{1t} \geq y^* | z_{t-1}, \Theta_1)y_{1t}}{P(y_{0t} < y^* | z_{t-1}, \Theta_0) + P(y_{1t} \geq y^* | z_{t-1}, \Theta_1)}. \tag{6}$$

Because the weighting function depends on the probability that the contemporaneous value of y_{it} will exceed the threshold level, we call this a contemporaneous threshold model.

The likelihood function of the C-STAR model is straightforward and easy to compute

$$l_t = \frac{P(y_{0t} < y^* | z_{t-1}, \Theta_0)f(y_{0t} | z_{t-1}, \Theta_0) + P(y_{1t} \geq y^* | z_{t-1}, \Theta_1)f(y_{1t} | z_{t-1}, \Theta_1)}{P(y_{0t} < y^* | z_{t-1}, \Theta_0) + P(y_{1t} \geq y^* | z_{t-1}, \Theta_1)}. \tag{7}$$

The likelihood function of the C-STAR model is continuous with respect to the threshold parameter so this parameter can be estimated jointly with the full parameter vector. In the following section we will further characterize the model using the simplest version of the model where $y_{it} \sim AR(1)$ for $i = 0, 1$.

2.1. Properties of the C-STAR model

In this section we use the following C-STAR(1) to illustrate the key properties of the model:

$$y_t = G(y_{t-1})y_{0t} + (1 - G(y_{t-1}))y_{1t}, \tag{8}$$

where

$$y_{it} = \mu_i + \alpha_i^j y_{i,t-1} + \sigma_i \varepsilon_{it}, \quad i = 0, 1$$

and

$$G(y_{t-1}) = \frac{\Phi((y^* - \mu_0 - \alpha_0^0 y_{t-1})/\sigma_0)}{\Phi((y^* - \mu_0 - \alpha_0^0 y_{t-1})/\sigma_0) + [1 - \Phi((y^* - \mu_1 - \alpha_1^1 y_{t-1})/\sigma_1)]}.$$

When generating the data we use the following identifying restriction: $\frac{\mu_0}{1-\alpha_0^0} < \frac{\mu_1}{1-\alpha_1^1}$. Notice that this restriction is sufficient to ensure that $\Phi((y^* - \mu_0 - \alpha_0^0 y_{t-1})/\sigma_0)$ and $[1 - \Phi((y^* - \mu_1 - \alpha_1^1 y_{t-1})/\sigma_1)]$ cannot both tend to zero at the same time.²

Since the model is capable of generating a wide variety of empirical distributions, in this section we analyze: (i) the response of the mixing function to changes in the parameters of the model; (ii) the empirical distribution of the data generated by the model; and (iii) the stability properties of the deterministic part of the model.

²The identifying restriction has been chosen to ensure that we do not introduce any inconsistencies when generating the data. Nevertheless, this restriction is sufficient but not necessary to ensure that the numerator and the denominator would not both tend to zero at the same time. As explained in footnote 6 using DGP 4 (one of the DGPs considered in the paper), this can happen only as the result of an inconsistency in designing the DGP.

Table 1
Properties of the mixing function

$\text{sign}\left(\frac{\partial G(y_{t-1})}{\partial \alpha_1^i}\right) = -\text{sign}(y_{t-1}), \text{ for } i = 0, 1$	$\frac{\partial G(y_{t-1})}{\partial y^*} > 0$
$\text{sign}\left(\frac{\partial G(y_{t-1})}{\partial \sigma_0}\right) = -\text{sign}(y^* - \mu_0 - \alpha_1^0 y_{t-1})$	$\text{sign}\left(\frac{\partial G(y_{t-1})}{\partial \sigma_1}\right) = \text{sign}(y^* - \mu_1 - \alpha_1^1 y_{t-1})$
$\frac{\partial G(y_{t-1})}{\partial \mu_0} < 0$	$\frac{\partial G(y_{t-1})}{\partial \mu_1} < 0$
$\frac{\partial G(y_{t-1})}{\partial y_{t-1}} < 0, \text{ if } \alpha_1^0, \alpha_1^1 > 0$	$\frac{\partial G(y_{t-1})}{\partial \varepsilon_{t-1}} < 0 \text{ if } \alpha_1^0, \alpha_1^1 > 0$

2.1.1. Properties of the mixing function

A key feature of the C-STAR model is that its mixing function, G , depends on all the parameters of the model as well as on y_{t-1} .³ In Table 1 we show the response of G to changes in the different parameters.

The first two rows of Table 1 show that the effect, on the mixing function, of a change in either the slope or the standard deviation is undetermined. The first row shows that $\text{sign}\left(\frac{\partial G(y_{t-1})}{\partial \alpha_1^i}\right) = -\text{sign}(y_{t-1})$. An increase in α_1^0 raises the conditional mean of y_{0t} , $\mu_0 + \alpha_1^0 y_{t-1}$, and reduces the probability $P(y_{0t} < y^* | y_{t-1})$ (and therefore $G(y_{t-1})$) for positive values of y_{t-1} (and thus increases the probability when $y_{t-1} < 0$). A similar argument applies for a change in α_1^1 where $\text{sign}\left(\frac{\partial G(y_{t-1})}{\partial \alpha_1^1}\right) = -\text{sign}(y_{t-1})$. The change in the slope raises the conditional mean of y_{1t} , $\mu_1 + \alpha_1^1 y_{t-1}$, increases the probability $P(y_{1t} > y^* | y_{t-1})$, thus reducing $G(y_{t-1})$ for positive values of y_{t-1} .

The second row shows that the sign of $\frac{\partial G(y_{t-1})}{\partial \sigma_i}$ depends on the sign of $y^* - \mu_i - \alpha_1^i y_{t-1}$, that is, the distance between the threshold and the conditional mean of y_{it} . In particular the sign of $\frac{\partial \Phi(\cdot)}{\partial \sigma_0}$ is inversely related to the sign of $y^* - \mu_0 - \alpha_1^0 y_{t-1}$. Notice that for $y^* - \mu_0 - \alpha_1^0 y_{t-1} > 0$, an increase in the volatility, σ_0 , reduces the value of $\Phi((y^* - \mu_0 - \alpha_1^0 y_{t-1})/\sigma_0)$ since, for a given conditional mean, a higher volatility reduces the area where the distribution of y_{0t} is smaller than the threshold. The opposite holds when $y^* - \mu_0 - \alpha_1^0 y_{t-1} < 0$. A similar argument applies for $\frac{\partial(1-\Phi(\cdot))}{\partial \sigma_1}$, which has the same sign as $y^* - \mu_1 - \alpha_1^1 y_{t-1}$.

The sign of $\frac{\partial G(y_{t-1})}{\partial y^*}$ is always positive since the higher is the threshold, the bigger is the area of the conditional density of y_{0t} (which is smaller than the threshold) and the smaller is the area of the conditional density of y_{1t} (which is greater than the threshold). In other words an increase in y^* increases $\Phi((y^* - \mu_0 - \alpha_1^0 y_{t-1})/\sigma_0)$ and reduces $[1 - \Phi((y^* - \mu_1 - \alpha_1^1 y_{t-1})/\sigma_1)]$. The sign of $\frac{\partial G(y_{t-1})}{\partial \mu_1}$, is always negative, since the larger is μ_1 , the larger is $[1 - \Phi((y^* - \mu_1 - \alpha_1^1 y_{t-1})/\sigma_1)]$. Analogously the sign of $\frac{\partial G(y_{t-1})}{\partial \mu_0}$ is always negative. Note also that the sign of $\frac{\partial G(y_{t-1})}{\partial y_{t-1}}$ is negative (provided that α_1^0 and α_1^1 are both positive).

³As we explain above, in C-STAR(p) models, G is a function of $z_{t-1} = (y_{t-1}, \dots, y_{t-p+1})$. Contrary to other STAR models, for the C-STAR there is no need to use any selection criteria to choose the appropriate threshold variable since, by construction, all the variables that enter in the information set also enter in the transition function.

Table 2
DGPs

	μ_0	α_1^0	σ_0	μ_1	α_1^1	σ_1	y^*
DGP 1	-0.5	0.9	3	0.5	0.9	2	1
DGP 2	-1	0.9	3	1	0.9	2	1
DGP 3	-1	0.9	3	1	0.9	3	0
DGP 4	-10	0.7	5	10	0.7	4	0

In Table 2 we present a selection of alternative DGPs used to illustrate the properties of the model.⁴ In Fig. 1 we present the conditional distributions of y_{it} , along with the threshold and the values taken by the mixing function $G(y_{t-1})$ once we condition on three arbitrary values, $y_{t-1} = \{-5, 0, 5\}$. The first row shows that, when most of the area of the two conditional distributions lies to the left of the threshold, $G(y_{t-1})$ tends to 1 and whenever they lie to the right, then $G(y_{t-1})$ tends to 0. The mixing function for DGP 1 takes values $\{G(y_{t-1} = -5) = 0.99, G(y_{t-1} = 0) = 0.63, G(y_{t-1} = 5) = 0.14\}$, which decrease with y_t since $\frac{\partial G(y_{t-1})}{\partial y_{t-1}} < 0$. DGP 2 has higher absolute values of μ_i and we find that $\{G(y_{t-1} = -5) = 0.98, G(y_{t-1} = 0) = 0.59, G(y_{t-1} = 5) = 0.17\}$. Comparing with the DGP 1 we find that these results come as the combination of a positive effect of a *negative* change in μ_0 and the negative effect of a *positive* change in μ_1 .⁵ In DGP 3 we increase σ_1 and reduce y^* (relative to DGP 2) and we find that $\{G(y_{t-1} = -5) = 0.89, G(y_{t-1} = 0) = 0.5, G(y_{t-1} = 5) = 0.11\}$. Even though a reduction in y^* always reduces $G(y_{t-1})$ and $\frac{\partial G(y_{t-1})}{\partial \sigma_1}$ depends on the sign of $y^* - \mu_1 - \alpha_1^1 y_{t-1}$, the total effect is to reduce the value of $G(y_{t-1})$ for the conditioning values under consideration. Finally we look at DGP 4 for which, for the chosen range, changes in the conditioning value do not affect substantially $G(y_{t-1})$. In the next section we analyze the empirical distribution of the data generated by the model and the behavior of the mixing function over time.

2.1.2. The empirical distribution of the data generated by the model

There is a large variety of empirical distributions and time series that can be generated using the C-STAR(1) model. In Fig. 2 we show, using the alternative DGPs presented in Table 2, the long-run state-dependent distributions alongside the threshold, the histogram of y_t generated by the C-STAR(1) model and the time series of y_t and $G(y_{t-1})$. We used the same 500,000 realizations of the shocks to draw the histograms, and the last 1000 realizations for the time series evolution of y_t and $G(y_{t-1})$.

DGP 1 and DGP 2 differ in the absolute value of μ_i , which is higher for DGP 2. This implies that for DGP 1 the long-run state-dependent distributions overlap for a larger part of their range than for DGP 2. This in turn implies, given that they share the same autoregressive parameters, that $G(y_{t-1})$ is more persistent for DGP 2 than for DGP 1.

⁴The DGPs have been arbitrarily chosen to highlight some relevant features of the model with respect to: the response of the mixing function to changes in the parameters of the model; the empirical distribution of the data generated by the model; and the stability properties of the deterministic part of the model.

⁵Even though it seems that this minor change from DGP 1 to DGP 2 does not greatly affect the model, we will see below that it has substantial effects on its stability.

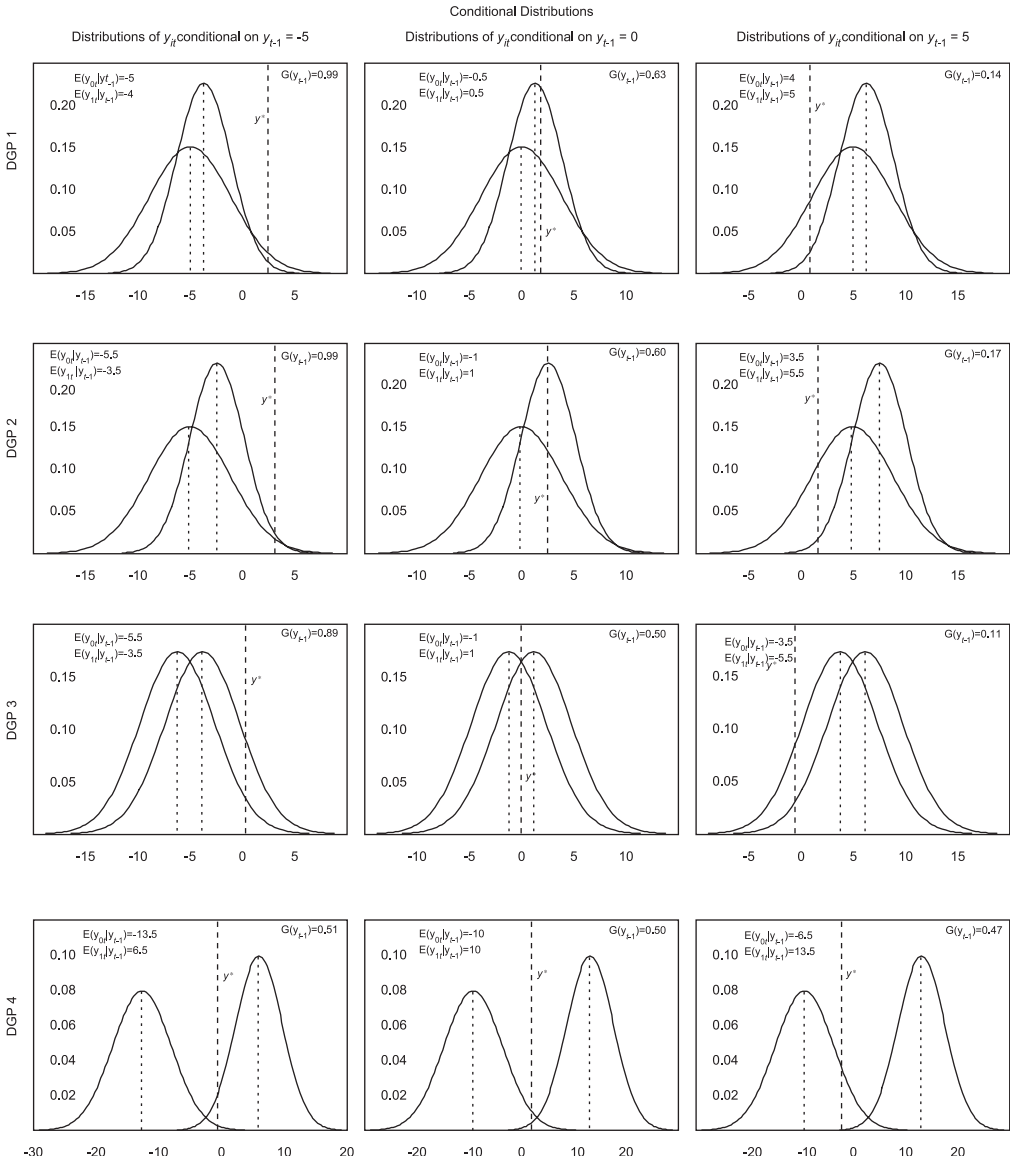


Fig. 1. Conditional distributions for the data generating processes presented in Table 2.

The plot shows, when considering DGP 2, that high values of y_{it} will most likely come from the distribution of $y_{i,t-1}$ since for those values $G(y_{i,t-1})$ is close to zero; the converse holds true for very low values of y_{it} . Turning our attention to the second column of Fig. 1, we can see that the histogram for DGP 1 is unimodal while that for DGP 2 is bimodal. This has implications for the stability properties of the model that are discussed in the next section.

DGP 3 is a mixture of the two distributions, which enter symmetrically in the sense that they have the same mean and standard deviation and that their means are equally apart

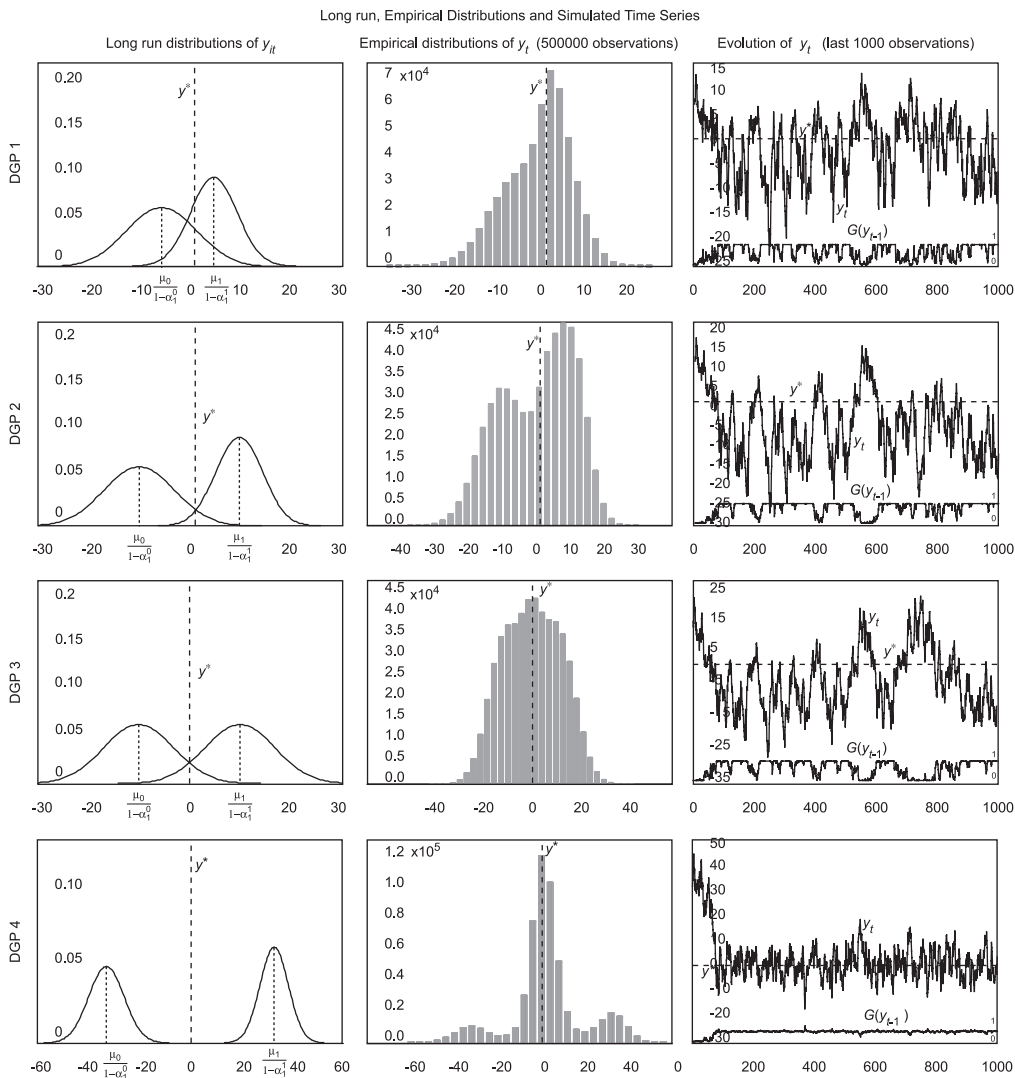


Fig. 2. Generated C-STAR data.

from the threshold. Then, even though $G(y_{t-1})$ takes values close to either 0 or 1 most of the time, the histogram of the generated data is unimodal and symmetric.

Finally using DGP 4, we find that the histogram has 3 modes and the model chooses most of the time a mixture of both distributions with probabilities, $G(y_{t-1})$, equal to one half.⁶ In the next section we derive the stability properties of the C-STAR based on the skeleton of the model, which will complement the characterization of the model.

⁶Consider DGP 4 (a very extreme case) and assume that you have inadvertently labeled the regimes incorrectly (swapped the regimes) in the data-generating process. Then, under this scenario, the long-run mean of y_{0t} would

2.2. Stability properties of the skeleton of the C-STAR

As Chan and Tong (1985) pointed out, we can analyze the properties of a nonlinear time series by considering the deterministic part of the model alone. This part is usually called the skeleton of the model and is defined as $y_t = F(y_{t-1}, \Theta)$, where

$$F(y_{t-1}, \Theta) = G(y_{t-1})(\mu_0 + \alpha_1^0 y_{t-1}) + (1 - G(y_{t-1}))(\mu_1 + \alpha_1^1 y_{t-1}), \tag{9}$$

and $\Theta = \{\Theta_0, \Theta_1, y^*\}$.⁷

Then a fixed point of the skeleton of the model is any value, y_L , that satisfies

$$y_L = F(y_L, \Theta) = G(y_L)(\mu_0 + \alpha_1^0 y_L) + (1 - G(y_L))(\mu_1 + \alpha_1^1 y_L). \tag{10}$$

Since the C-STAR(1) is a nonlinear model, there may be one, several or no equilibrium values that satisfy Eq. (10). Then, assessing which of the equilibria of the nonlinear first-order difference equation are stable is crucial for learning about the stability properties of the C-STAR model. It is relatively straightforward to assess the local stability of each of the equilibrium points, whenever the skeleton is only a function of the first lag.

We use each of the DGPs presented in Table 2 to assess: (i) the number of equilibria and (ii) the stability of the equilibria. We find the number of equilibria for the different DGPs presented in Table 1, using a grid of starting values to solve Eq. (10) numerically.⁸ For each equilibrium, we analyze whether it is locally stable by considering the following expansion around the fix point:

$$y_t - y_L = F(y_{t-1}, \Theta) - F(y_L, \Theta) \simeq \frac{\partial F(y_{t-1}, \Theta)}{\partial y_{t-1}} (y_{t-1} - y_L). \tag{11}$$

Whenever $|\frac{\partial F(y_{t-1}, \Theta)}{\partial y_{t-1}}| < 1$, the equilibrium is locally stable and $F(y_{t-1}, \Theta)$ is a contraction in the neighborhood of $y = y_L$, where

$$\frac{\partial F(y_L, \Theta)}{\partial y_{t-1}} = \alpha_1^1 + (\alpha_1^0 - \alpha_1^1)G(y_L) + [(\mu_0 - \mu_1) + (\alpha_1^0 - \alpha_1^1)y_L] \frac{\partial G(y_L)}{\partial y_{t-1}} \tag{12}$$

and

$$\frac{\partial G(y_L)}{\partial y_{t-1}} = \frac{-((\alpha^0/\sigma_0)\phi(w_0^L)\Phi(w_1^L) + (\alpha^1/\sigma_1)\phi(w_1^L)\Phi(w_0^L))}{(\Phi(w_0^L) + [1 - \Phi(w_1^L)])^2} \quad \text{where } \phi = \Phi',$$

$$w_0^L = (y^* - \mu_0 - \alpha_1^0 y_L)/\sigma_0 \quad \text{and} \quad w_1^L = (y^* - \mu_1 - \alpha_1^1 y_L)/\sigma_1. \tag{13}$$

(footnote continued)

be associated with the high-mean regime and the long-run mean of y_{1t} with the low-mean regime. When attempting to generate the data, both the numerator and the denominator of $G(y_{t-1})$ will tend to zero. Notice that this simply happens because of an inconsistency introduced when generating the data. The inconsistency is to label regime 0 as the low regime, which is identified by $P(y_{0t} < y^* | y_{t-1}, \Theta_0)$, when all the mass of the distribution of y_{0t} is higher than the threshold and, at the same time, to label regime 1 as the high regime, which is identified by $P(y_{1t} \geq y^* | y_{t-1}, \Theta_1)$, when all the mass of the distribution of y_{1t} is lower than the threshold.

⁷Instead of introducing new notation, in this section we denote y_t as the skeleton of the model. This is equivalent to setting the shock equal to zero.

⁸This is usually labeled “deterministic simulation”, see Teräsvirta and Anderson (1992) and Peel and Speight (1996).

Simulated Time Series and Skeletons

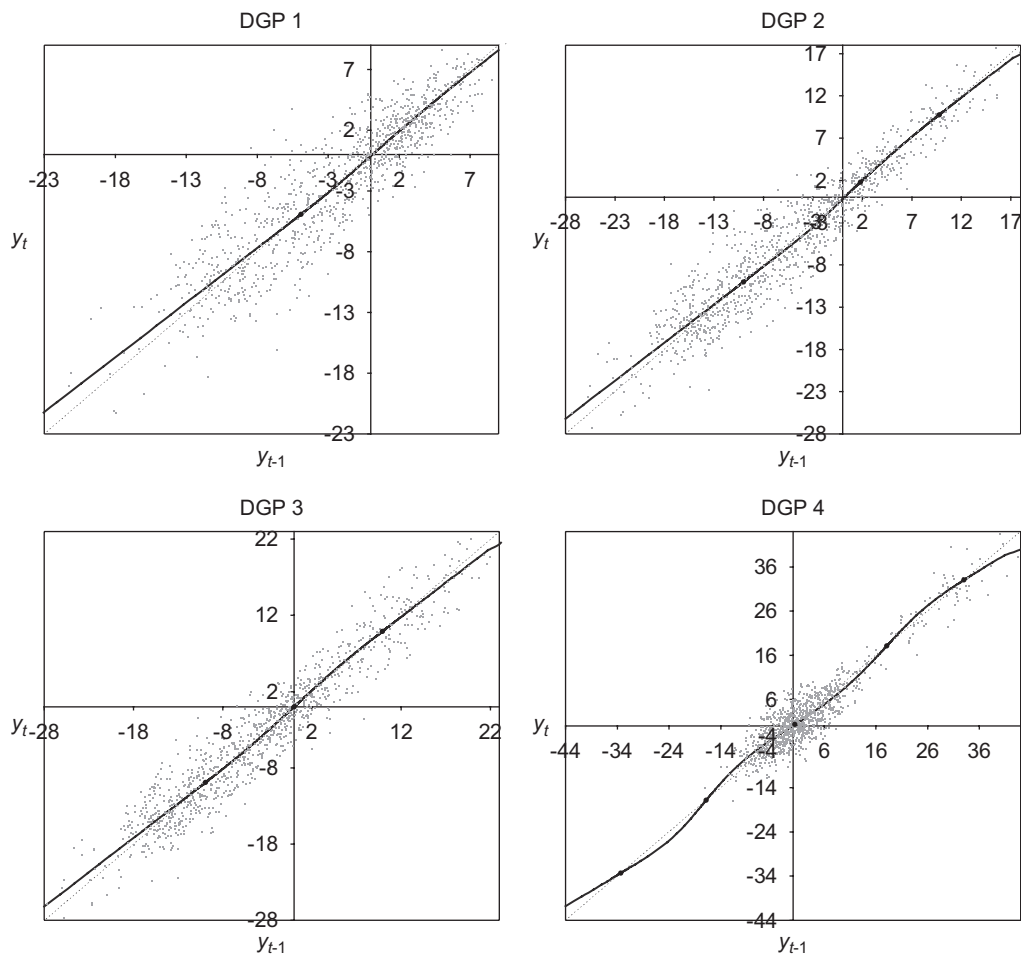


Fig. 3. Lung-run properties of the C-STAR model for the data generating processes presented in Table 2.

Fig. 3 shows the skeleton and scatter plots using the C-STAR model and the generated data (the last 1000 observations) from the DGPs presented in Table 1. The intersection of the skeleton and the 45° line in the plot between y_t and y_{t-1} shows the points where Eq. (10) is satisfied.

We found for DGP 1 a unique stable equilibrium for $y_L = \{-4.7\}$ with the associated $\frac{\partial F(y_L, \theta)}{\partial y_{t-1}} = \{0.93\}$. There is little that is known about stationarity conditions for STAR models in general and this is also the case with the C-STAR model. Nevertheless, in general, a way of checking the stationarity of nonlinear models is to determine whether the skeleton is stable using deterministic simulation. If the series generated from the skeleton explodes in this simulation exercise then the time series is not stationary.

For DGP 2 we increase (relative to DGP 1) the absolute value of the intercepts. The effect on the nonlinear model of increasing the absolute value of the intercepts is to

increase the number of fixed points to three.⁹ The fixed points for DGP 2 take the values $y_L = \{-9.99, 1.81, 9.77\}$, with the associated $\frac{\partial F(y_L, \Theta)}{\partial y_{t-1}} = \{0.9, 1.08, 0.91\}$. This implies that the first and the last fixed points are locally stable while the intermediate fixed point is locally unstable. We can see in Fig. 3 that whenever y_t lies near any of the stable equilibria, then it will take a large shock to cause a transition of the series from the one equilibrium to the other. Notice that in the absence of shocks, -9.99 and 9.77 are both attractors and 1.81 is the boundary between the domains of attraction, which implies that once we introduce shocks, we can expect the time series to switch occasionally between attractors, but we do not expect the time series to be explosive.

The figure for DGP 3 is qualitatively similar to that of DGP 2 since we also find three fixed points $y_L = \{-9.91, 0, 9.92\}$ with associated $\frac{\partial F(y_L, \Theta)}{\partial y_{t-1}} = \{0.91, 1.08, 0.90\}$. The points are evenly distributed since the distributions are symmetric and so is the domain of attraction.

Finally for DGP 4 we further increase the absolute value of the intercepts along with the variances and obtain the fixed points $y_L = \{-33.3, -16.8, 0.34, 18.12, 33.05\}$ with the associated $\frac{\partial F(y_L, \Theta)}{\partial y_{t-1}} = \{0.71, 1.41, 0.76, 1.28, 0.74\}$.¹⁰ Given that the smallest and highest equilibria are stable we expect y_t to revert to values which are in the range $\{-33.3, 33.05\}$.

We can also study the effect of changes in y^* on the fixed point y_L . The results are rather complex since they not only depend on the equilibrium under consideration but they also affect the number of fixed points. In general is easy to show that for very low values of y^* such that $G(y_L) \simeq 0$, $y_L \simeq \frac{\mu_1}{(1-\alpha_1^*)}$ and, for high values of y^* where $G(y_L) \simeq 1$, $y_L \simeq \frac{\mu_0}{(1-\alpha_1^*)}$.

In the following section, Monte Carlo methods are used to examine the quality of asymptotic approximations to the finite-sample distribution of the ML estimator and other related statistics using, among others, the DGPs analyzed above.

3. Finite-sample properties of C-STAR models

We begin by discussing the experimental design and Monte Carlo simulation of the statistics of interest. The numerical results follow.

3.1. Experimental design and simulation

The following C-STAR(1) model is used as the data-generating process (DGP) in the experiments carried out in this section

$$y_t = \frac{P(y_{0t} < y^* | z_{t-1}, \Theta_0)y_{0t} + P(y_{1t} \geq y^* | z_{t-1}, \Theta_1)y_{1t}}{P(y_{0t} < y^* | z_{t-1}, \Theta_0) + P(y_{1t} \geq y^* | z_{t-1}, \Theta_1)}, \tag{14}$$

$$y_{0t} = \mu_0 + \alpha_1^0 y_{t-1} + \sigma_0 \varepsilon_t,$$

$$y_{1t} = \mu_1 + \alpha_1^1 y_{t-1} + \sigma_1 \varepsilon_t.$$

⁹Franses and van Dijk (2000) find the same result using L-STAR models.

¹⁰We use this DGP to show the rather complex dynamic patterns that can be generated by the model even though we do not expect that financial or macro data will be generated by this configuration of parameters.

where ε_t are *i.i.d.* $N(0,1)$. The experiments are a full factorial design of

$$\begin{aligned}(\mu_0, \mu_1) &\in \{(-0.5, 0.5), (-1, 1), (-10, 10)\}, \\(\sigma_0, \sigma_1) &\in \{(3, 2), (3, 3), (3, 2)\}, \\(\alpha_1^0, \alpha_1^1) &\in \{(0.7, 0.7), (0.9, 0.9)\}, \\y^* &\in \{0, 1, 2, 3, 4\}, \\T &\in \{100, 200, 400, 800\}.\end{aligned}\tag{15}$$

The sample sizes selected are representative of the data sets that are typically used in empirical work (samples of 800 or more observations are not uncommon in studies using weekly or daily data). In all experiments, we generate $50 + T$ data points for y_t , starting with $y_0 = y^*$. However, in order to attenuate the effect of the initial values, only the last T of these observations are used in each Monte Carlo replication.

The ML estimates $\hat{\Theta} \equiv \{\hat{\mu}_0, \hat{\mu}_1, \hat{\alpha}_1^0, \hat{\alpha}_1^1, \hat{\sigma}_0, \hat{\sigma}_1, \hat{y}^*\}$, are obtained by means of a quasi-Newton algorithm that approximates the Hessian according to the Broyden–Fletcher–Goldfarb–Shanno (BFGS) update computed from numerical derivatives (see, e.g., Fletcher, 1987). In each case, a grid of seven initial values for each parameter (including the true parameter) are used as starting values for the BFGS iterations. The replication that achieves the higher likelihood value will then be selected.¹¹ Finally, since the computation of ML estimates for switching models is particularly time-consuming (given the large number of simulations and the grid for the initial values), the number of Monte Carlo replications per experiment is 2000.

In order to save space only a selection of simulation results are reported.¹² In particular, we consider several versions of DGP 2, where the threshold, y^* , is allowed to vary.

$$\begin{aligned}\text{DGP}(i) &= \{(\mu_0, \mu_1) = (-1, 1), (\alpha_1^0, \alpha_1^1) = (0.9, 0.9), (\sigma_0, \sigma_1) = (2, 3), y^* = i\} \\&\text{for } i = 0, 1, 2, 3.\end{aligned}\tag{16}$$

We analyze how this affects the estimation results since, for DGP(0) (where $y^* = 0$), the series spends on average 33% of the time below the threshold value. For DGP(1), on average it spends equal time above and below the threshold, while for DGP(2) and DGP(3) the time spent below the threshold is on average 66% and 80%, respectively.

3.1.1. Distribution of the ML estimator: biases, estimated standard errors and normality tests

We report some of the characteristics of the finite-sample distribution of the ML estimator of $\hat{\Theta}$. These include: (i) the deviation of the mean from the true parameter (bias), (ii) a measure of the accuracy of estimated large-sample standard errors as approximations to the correct sampling standard deviation of the ML estimator and (iii) tests for normality.

The top panel of Table 3 shows that for most of the design points the biases are only significantly different from zero when $T \leq 200$. The size of the bias in the ML estimator vary for the alternative DGPs under scrutiny. For example, while very large samples are

¹¹Notice that the estimation results appear to be robust to the choices of initial values.

¹²The full set of results is available on request.

Table 3

Characteristics of the empirical distribution of the MLE: mean bias and ratios of sampling SDs to estimated SEs

DGP		Maximum likelihood estimates						
y^*	T	$\hat{\mu}_0$	$\hat{\mu}_1$	$\hat{\alpha}_1^0$	$\hat{\alpha}_1^1$	$\hat{\sigma}_0$	$\hat{\sigma}_1$	\hat{y}^*
<i>Mean bias</i>								
0	100	-0.063	0.035	-0.047	-0.053	-0.031	-0.069	0.059
	200	0.040	-0.027	-0.015	-0.027	-0.034	-0.090	0.016
	400	0.016	-0.008	-0.011	0.006	-0.022	-0.015	-0.013
	800	0.009	0.002	0.001	0.000	-0.011	-0.009	-0.010
1	100	-0.068	0.008	-0.059	-0.014	-0.001	-0.014	-0.096
	200	-0.025	-0.006	-0.037	-0.022	-0.048	-0.020	-0.034
	400	0.017	-0.009	-0.011	0.005	-0.025	-0.019	-0.012
	800	0.001	-0.004	0.000	0.019	-0.006	-0.010	0.011
2	100	-0.075	0.095	-0.052	-0.074	-0.080	-0.093	-0.086
	200	-0.061	-0.040	-0.027	-0.045	-0.039	-0.085	-0.049
	400	-0.014	-0.054	-0.009	-0.016	-0.012	-0.018	-0.015
	800	0.002	-0.007	-0.002	0.004	-0.009	-0.015	-0.004
3	100	-0.096	-0.071	-0.073	-0.089	-0.065	-0.015	-0.070
	200	-0.091	-0.046	-0.026	-0.085	-0.042	-0.019	-0.058
	400	0.055	-0.033	-0.011	-0.032	-0.020	-0.007	-0.049
	800	-0.047	-0.020	-0.003	0.018	-0.012	-0.006	-0.021
<i>Ratios of sampling SDs to estimated SEs</i>								
0	100	1.120*	1.019	1.041	1.037	0.985	1.004	1.068*
	200	1.098*	1.012	0.971	1.025	0.991	1.002	1.029
	400	1.018	0.991	1.015	1.012	1.005	1.002	1.017
	800	1.014	1.006	1.010	1.010	0.994	1.000	1.016
1	100	1.061*	1.072*	1.020	1.011	0.971*	1.022	1.096*
	200	1.011	1.042*	1.025	1.021	0.981	1.016	1.080*
	400	1.011	1.009	1.019	1.006	0.997	1.006	1.046
	800	1.005	1.004	1.005	0.999	1.000	1.002	1.005
2	100	1.131*	1.047*	1.017	1.020	1.042*	0.951*	1.049*
	200	1.081*	1.014	1.010	1.023	1.026	0.987	1.008
	400	1.052	1.002	1.010	0.993	1.010	0.989	1.005
	800	1.041	1.002	1.000	1.002	0.998	0.994	1.002
3	100	1.113*	1.104*	1.244*	1.023	1.143*	0.964	1.100*
	200	1.082*	1.033	1.024	1.011	1.051	0.987	1.094*
	400	1.055	1.006	1.015	1.005	0.958	0.995	1.025
	800	1.012	1.000	1.009	0.999	0.983	0.999	1.011

Note: * indicates that the Kolmogorov–Smirnov statistic is significant at the 5% level.

needed to reduce the biases of $\hat{\mu}_0$, $\hat{\mu}_1$ and \hat{y}^* using DGP(3), for DGP(0) we find that the biases associated with the slopes $(\hat{\alpha}_1^0, \hat{\alpha}_1^1)$ approach zero for relatively small sample sizes. Overall the results show that the ML estimator is slightly biased only for the smallest sample under consideration. The bias clearly decreases as the sample increases and becomes negligible when $T = 400$.

Turning to the accuracy of the estimated standard errors, we show in the bottom panel of Table 3 the ratio of the exact standard deviation of the ML estimates to the estimated standard errors averaged across replications for each design point. The standard errors are calculated as $\sqrt{T^{-1}(\widehat{\Omega}_T^{-1})_{ii}}$, $ii \in (1, \dots, 7)$, where $(\widehat{\Omega}_T^{-1})_{ii}$ is the i th diagonal element of $\widehat{\Omega}_T^{-1}$, where $\widehat{\Omega}_T = -\frac{1}{T}(\frac{\partial^2 L(\theta)}{\partial \theta \partial \theta'} |_{\theta=\widehat{\theta}})$ and $L(\theta)$ is the logarithmic conditional likelihood function. For the vast majority of cases, the estimated asymptotic standard errors are downward biased. These biases however are not substantial (at least for samples larger than 200) and therefore should not affect inference significantly.

The Gaussianity of the finite-sample distributions of the ML estimates is also assessed by means of a Kolmogorov–Smirnov test that compares the empirical distribution function of the ML estimates (relocated and scaled so that the linearly transformed estimates have zero mean and unit variance) with the standard normal distribution function (see Lilliefors, 1967). The test statistic is calculated as $\max_{1 \leq j \leq r} |j/r - \Phi(z_{r;j})|$, where $z_{r;j}$ denotes the order statistic of rank j associated with the transformed estimates. The cases in which the Kolmogorov–Smirnov statistic is significant, at the 5% level, are indicated in the bottom panel of Table 3. We find, for the configurations under consideration, that the hypothesis of normality of all the estimators other than $\widehat{\mu}_0$, $\widehat{\mu}_1$, and \widehat{y}^* is never rejected for sample sizes larger than 100. Furthermore, the value of the Kolmogorov–Smirnov statistic decreases as T increases, suggesting that the quality of the normal approximation is likely to improve in larger samples. In fact, while for $T = 400$ the hypothesis of normality can be rejected only once, when $T = 800$ it can never be rejected.

3.1.2. Hypothesis tests

We now turn to hypothesis testing by examining the empirical distributions of conventional t -type statistics associated with the elements of $\widehat{\theta}$. These are calculated as $[(i)\widehat{\theta} - (i)\theta^0] / \sqrt{T^{-1}(\widehat{\Omega}_T^{-1})_{ii}}$, $i \in (1, \dots, 7)$, where $(i)\widehat{\theta}$ and $(i)\theta^0$ denote the i th element of $\widehat{\theta}$ and θ^0 , respectively. According to standard asymptotic theory, the t -statistics should be approximately distributed as $N(0, 1)$. The mean and standard deviation of the empirical distributions of the t -statistics are reported in Table 4. For some estimates, the mean and standard deviation differ substantially from the values associated with the approximating normal distribution. We find that the distributions of t -statistics based on $\widehat{\alpha}_1^0, \widehat{\alpha}_1^1, \widehat{\sigma}_0$ and $\widehat{\sigma}_1$ are mostly close to the true values, but the t -statistics based on $\widehat{\mu}_0, \widehat{\mu}_1$ and \widehat{y}^* have means significantly different from zero. However, the deviations from zero decrease (in absolute value) as T increases. More specifically, the distributions of the t -statistics based on the majority of the parameters in all DGPs are close to the theoretical values for sample sizes as small as 200. In addition, in some cases the standard deviation is larger than 1 (for the intercepts and the threshold) but approaches the theoretical value as the sample size increases (in particular the standard deviation of the slopes approaches one also for very small samples).

The outcome of Kolmogorov–Smirnov tests for the normality of the distribution of the t -statistics is indicated in the bottom panel of Table 4. The hypothesis of normality is rejected for only a few design points. The Gaussian approximation is generally adequate for t -statistics based on $\widehat{\alpha}_1^0, \widehat{\alpha}_1^1, \widehat{\sigma}_0$ and $\widehat{\sigma}_1$, and for statistics based on $\widehat{\mu}_1, \widehat{\mu}_2$ and \widehat{y}^* , provided that the sample size is not too small.

Table 4
Empirical moments of *t*-statistics

DGP		<i>t</i> -statistic						
<i>y</i> *	<i>T</i>	$\hat{\mu}_0$	$\hat{\mu}_1$	$\hat{\alpha}_1^0$	$\hat{\alpha}_1^1$	$\hat{\sigma}_0$	$\hat{\sigma}_1$	\hat{y}^*
<i>Mean</i>								
0	100	0.187	-0.185	0.083	0.044	-0.051	-0.074	-0.063
	200	0.093	-0.101	0.007	0.004	-0.045	-0.012	-0.042
	400	0.012	0.018	0.049	0.001	-0.025	0.008	0.026
	800	0.011	0.012	0.003	0.000	-0.008	0.004	0.012
1	100	0.126	-0.164	0.095	0.076	-0.038	-0.049	-0.097
	200	0.014	-0.156	0.061	0.022	-0.020	-0.004	-0.053
	400	-0.003	-0.098	0.038	0.006	-0.005	-0.003	-0.028
	800	-0.001	-0.012	-0.008	0.003	0.002	0.001	-0.002
2	100	0.159	-0.136	0.166	0.034	-0.064	-0.065	-0.043
	200	0.009	-0.123	0.097	0.021	-0.023	-0.011	-0.048
	400	0.012	-0.101	0.056	0.007	-0.009	-0.061	0.002
	800	0.001	-0.014	0.008	-0.002	-0.003	-0.002	0.001
3	100	-0.234	0.154	-0.095	0.054	-0.198	-0.073	-0.149
	200	-0.189	0.081	0.037	0.006	-0.032	-0.015	-0.123
	400	-0.131	-0.052	-0.023	-0.007	-0.018	0.004	-0.085
	800	-0.066	0.025	0.001	0.000	-0.005	0.002	0.021
<i>Standard deviation</i>								
0	100	1.271*	1.096*	1.071*	0.989	0.981	0.994	1.110*
	200	1.086*	1.021	1.044	0.993	1.052	1.005	1.088*
	400	1.035	1.006	1.029	1.004	1.031	1.003	1.072
	800	1.007	1.002	1.008	1.000	0.998	1.000	1.008
1	100	1.104*	1.077*	1.030	1.000	0.970	0.998	1.075*
	200	1.066	1.032	0.983	1.009	0.986	0.999	1.022
	400	1.023	1.008	1.014	0.997	1.007	0.997	0.999
	800	1.019	1.012	1.002	1.001	0.999	1.000	1.000
2	100	1.080*	1.113*	1.002	0.937	1.032	1.039	1.137*
	200	1.053	1.046	0.997	0.952	1.051	1.011	1.059
	400	1.022	1.040	1.000	0.969	1.009	1.005	1.071
	800	1.002	1.009	1.000	0.999	1.005	1.000	1.003
3	100	1.286*	1.075*	1.004	1.016	0.975	1.044	1.118*
	200	1.163*	1.044	1.002	1.003	0.964	1.006	1.052
	400	1.098*	1.032	0.998	1.000	0.987	1.003	1.028
	800	1.045	1.008	1.000	1.000	0.999	1.001	1.007

Note: * indicates that the Kolmogorov–Smirnov statistic is significant at the 5% level.

Finally, to examine the direct consequences of these results for hypothesis testing, Table 5 reports the empirical size of *t*-type tests of the null hypothesis $H_{0:(i)}\Theta_{=(i)}\Theta^0$ against the alternative $H_{0:(i)}\Theta \neq_{(i)}\Theta^0$. The entries are relative rejection frequencies based on standard normal critical regions of nominal size 0.05 and 0.10. It is clear that in some cases

Table 5
 Empirical size of two-tailed tests based on t -statistics

DGP		t -statistic						
y^*	T	$\hat{\mu}_0$	$\hat{\mu}_1$	$\hat{\alpha}_1^0$	$\hat{\alpha}_1^1$	$\hat{\sigma}_0$	$\hat{\sigma}_1$	\hat{y}^*
<i>Nominal size = 0.05</i>								
0	100	0.077	0.054	0.071	0.056	0.066	0.053	0.060
	200	0.065	0.052	0.066	0.052	0.062	0.054	0.061
	400	0.058	0.051	0.058	0.050	0.058	0.052	0.054
	800	0.053	0.050	0.049	0.050	0.054	0.051	0.050
1	100	0.067	0.079	0.066	0.076	0.063	0.071	0.081
	200	0.058	0.056	0.061	0.058	0.057	0.065	0.058
	400	0.045	0.053	0.058	0.052	0.055	0.052	0.053
	800	0.047	0.050	0.053	0.049	0.052	0.050	0.047
2	100	0.060	0.087	0.059	0.061	0.063	0.077	0.091
	200	0.058	0.062	0.054	0.057	0.061	0.068	0.067
	400	0.054	0.059	0.048	0.056	0.051	0.059	0.054
	800	0.051	0.056	0.050	0.052	0.050	0.054	0.052
3	100	0.130	0.054	0.128	0.053	0.111	0.052	0.071
	200	0.098	0.052	0.085	0.052	0.090	0.049	0.064
	400	0.078	0.050	0.072	0.050	0.084	0.051	0.060
	800	0.070	0.050	0.064	0.050	0.059	0.050	0.054
<i>Nominal size = 0.10</i>								
0	100	0.145	0.110	0.121	0.108	0.141	0.091	0.132
	200	0.132	0.108	0.117	0.103	0.114	0.107	0.117
	400	0.110	0.103	0.108	0.101	0.109	0.106	0.109
	800	0.106	0.100	0.103	0.100	0.103	0.105	0.102
1	100	0.121	0.109	0.116	0.129	0.136	0.120	0.133
	200	0.111	0.110	0.093	0.117	0.120	0.118	0.120
	400	0.102	0.103	0.106	0.109	0.117	0.106	0.108
	800	0.100	0.101	0.102	0.100	0.104	0.102	0.102
2	100	0.121	0.134	0.115	0.108	0.112	0.116	0.122
	200	0.115	0.126	0.091	0.108	0.110	0.108	0.110
	400	0.103	0.112	0.106	0.102	0.108	0.092	0.106
	800	0.100	0.104	0.102	0.100	0.103	0.100	0.105
3	100	0.120	0.112	0.131	0.102	0.128	0.110	0.144
	200	0.113	0.096	0.124	0.100	0.120	0.107	0.131
	400	0.096	0.092	0.108	0.102	0.118	0.104	0.129
	800	0.101	0.104	0.100	0.097	0.109	0.100	0.110

the tests suffer from size distortions (see in particular the empirical sizes associated to $\hat{\mu}_0$ and $\hat{\alpha}_1^0$ in DGP(3), with small sample size), having values well in excess of the nominal levels. However, these distortions do attenuate as the sample size increases; in fact, for $T = 200$ most of the tests have empirical rejection frequencies that are insignificantly different from the nominal values of 0.05 and 0.10. In summary, our results seem to suggest that samples of more than 200 observations are typically needed before asymptotic theory is a good guide for inference.

4. Forecasting with contemporaneous threshold autoregressive models

Parametric models that allow for nonlinear dynamics and changes in regime have attracted considerable interest in the literature. While such models have been shown to describe well the behavior of many time series, including economic and financial ones, the evidence concerning their ability to produce accurate out-of-sample forecasts is far from conclusive (see van Dijk et al., 2002 for a survey). Different authors have compared various methods for obtaining multi-step-ahead forecasts with threshold models and evaluated their empirical performance (e.g., Lin and Granger, 1994; Clements and Smith, 1999).¹³

In this section we discuss a method for obtaining multi-step, out-of-sample, forecasts for the C-STAR(p) model that computes the full tree of possible future values and evaluates the probability that the regimes would follow different paths in the future. We also undertake a systematic study to analyze the forecasting performance of the contemporaneous threshold autoregressive model.¹⁴ Even though our preferred forecasting procedure can be regarded as an approximation (as will be made clear below), our simulations results suggest that the cost of using this approximation is negligible.

4.1. One-step-ahead forecasts

Consider the C-STAR(p) model

$$y_t = \frac{P(y_{0t} < y^* | z_{t-1}, \Theta_0)y_{0t} + P(y_{1t} \geq y^* | z_{t-1}, \Theta_1)y_{1t}}{P(y_{0t} < y^* | z_{t-1}, \Theta_0) + P(y_{1t} \geq y^* | z_{t-1}, \Theta_1)},$$

$$y_{0t} = \mu_0 + \alpha_1^0 y_{t-1} + \dots + \alpha_p^0 y_{t-p} + \sigma_0 \varepsilon_t,$$

$$y_{1t} = \mu_1 + \alpha_1^1 y_{t-1} + \dots + \alpha_p^1 y_{t-p} + \sigma_1 \varepsilon_t. \tag{17}$$

The C-STAR model produces forecasts that involve a weighted average of the two linear relationships. The one-step-ahead forecast for the C-STAR model is straightforward to compute. Specifically, the minimum mean square error, one-step-ahead forecast (at the forecast origin t) $\hat{y}_t(1)$ is obtained as

$$\hat{y}_t(1) = E(y_{t+1} | \mathcal{F}_t) = E(y_{0t+1} | \Theta_0, \mathcal{F}_t) \frac{P(y_{0t+1} < y^* | \Theta_0, \mathcal{F}_t)}{P(y_{0t+1} < y^* | \Theta_0, \mathcal{F}_t) + P(y_{1t+1} \geq y^* | \Theta_1, \mathcal{F}_t)} + E(y_{1t+1} | \Theta_1, \mathcal{F}_t) \frac{P(y_{1t+1} \geq y^* | \Theta_1, \mathcal{F}_t)}{P(y_{0t+1} < y^* | \Theta_0, \mathcal{F}_t) + P(y_{1t+1} \geq y^* | \Theta_1, \mathcal{F}_t)}. \tag{18}$$

¹³For example, Tiao and Tsay (1994) and Clements and Smith (1997, 1999), using GDP data, find that threshold models can outperform their linear alternative when the economy is in a recession. By contrast, using US GNP data Clements and Krolzig (1998) suggest that a linear autoregressive model is a relatively robust forecasting device, even when such nonlinearities are a feature of the data. Also, Lundbergh and Teräsvirta (2002) use unemployment series and find that a threshold model outperforms the linear one over long horizons.

¹⁴Our approach is related to the approximation method for first-order exponential autoregressive threshold models first suggested by Al-Qassam and Lane (1989), and then extended by De Gooijer and Bruin (1998).

4.2. An approximate method for multi-step-ahead forecasts

Next we propose an approximate method to obtain multi-step-ahead forecasts for the C-STAR model. The approximation is that we do not evaluate all possible combinations of expected values of the future error terms, conditional on the paths the regimes might take through the tree of possible outcomes. The accuracy of our approach is then evaluated by comparing these values with those obtained by using a naive forecasting method, a Monte Carlo simulation approach, and a simple linear specification.

4.2.1. Approximate forecasts from the full tree of future states

To obtain the h -step-ahead forecast $\hat{y}_t(h)$, $h \geq 2$, we propose the following approximate formulae: let

$$H_{t+h} = \begin{cases} 0 & \text{if } \Theta_{t+h} = \Theta_0, \\ 1 & \text{otherwise,} \end{cases} \quad W_{t+h} = \begin{cases} 0 & \text{if } y_{t+h} < y^*, \\ 1 & \text{otherwise,} \end{cases} \tag{19}$$

and define $\mu_{t+h} = \mu_0 - (\mu_0 - \mu_1)H_{t+h}$, $\mathcal{H}_{t+h} = A_0 - (A_0 - A_1)H_{t+h}$, and $\sigma_{t+h} = (\sigma_0 - (\sigma_0 - \sigma_1))I H_{t+h}$. Then

$$E(y_{t+h} | \mathcal{F}_t) = \delta' \left(\begin{array}{l} \mu_{t+h} + \mathcal{H}_{t+h}\mu_{t+h-1} + \mathcal{H}_{t+h}\mathcal{H}_{t+h-1}\mu_{t+h-2} + \dots + \mathcal{H}_{t+h}\mathcal{H}_{t+h-1} \\ \times \dots \times \mathcal{H}_{t+2}\mu_{t+1} + \mathcal{H}_{t+h}\mathcal{H}_{t+h-1} \times \dots \times \mathcal{H}_{t+2}\mathcal{H}_{t+1}z_t \end{array} \right), \tag{20}$$

$$\times \mathbf{P}^*(W_{t+h}, W_{t+h-1}, W_{t+h-2}, \dots, W_{t+1} | \mathcal{F}_t),$$

where

$$\mathbf{P}^*(W_{t+h}, W_{t+h-1}, \dots, W_{t+1} | \mathcal{F}_t) = \frac{\mathbf{P}(W_{t+h}, W_{t+h-1}, W_{t+h-2}, \dots, W_{t+1} | \mathcal{F}_t)}{\sum_i^{2^h} \mathbf{P}(W_{t+h}, W_{t+h-1}, W_{t+h-2}, \dots, W_{t+1} | \mathcal{F}_t)}. \tag{21}$$

Much algebra goes into the calculation of these probabilities and it is available in Dueker et al. (2003). This forecast method evaluates the full tree of future state probabilities for the forecast horizon.

4.2.2. Naive forecasts

Finally, consider the naive method to obtain multi-step-ahead forecasts for the C-STAR model. The naive h -step-ahead forecast $\hat{y}_t(h)$, $h \geq 2$ is obtained as

$$\begin{aligned} \hat{y}_t(h) &= E(y_{t+h} | \hat{y}_t(h-1), \mathcal{F}_t) \\ &= E(y_{0t+h} | \hat{y}_t(h-1), \Theta_0, \mathcal{F}_t) \\ &\quad \times \frac{\mathbf{P}(y_{0t+h} < y^* | \hat{y}_t(h-1), \Theta_0, \mathcal{F}_t)}{\mathbf{P}(y_{0t+h} < y^* | \hat{y}_t(h-1), \Theta_0, \mathcal{F}_t) + \mathbf{P}(y_{1t+h} \geq y^* | \hat{y}_t(h-1), \Theta_1, \mathcal{F}_t)} \\ &\quad + E(y_{1t+h} | \hat{y}_t(h-1), \Theta_1, \mathcal{F}_t) \\ &\quad \times \frac{\mathbf{P}(y_{1t+h} \geq y^* | \hat{y}_t(h-1), \Theta_1, \mathcal{F}_t)}{\mathbf{P}(y_{0t+h} < y^* | \hat{y}_t(h-1), \Theta_0, \mathcal{F}_t) + \mathbf{P}(y_{1t+h} \geq y^* | \hat{y}_t(h-1), \Theta_1, \mathcal{F}_t)}. \end{aligned} \tag{22}$$

4.3. Forecast evaluation

We perform several Monte Carlo experiments to investigate the forecasting performance of the three different approaches described above (the approximation method, AC-STAR(1), the naive approach, N-STAR(1), and the Monte Carlo method, MC-STAR(1)). We also investigate the effect of using a linear autoregressive process when the true DGP is a C-STAR by analyzing their relative forecasting performance.¹⁵

Eqs. (14) and (16) are used as the DGP for the simulations. In all the experiments, the sample size is $T = 200$, the forecast horizon is $h \in \{1, 2, \dots, 7\}$ and $\{\varepsilon_t\}$ are *i.i.d.* Gaussian random variates such that $\mathbb{E}[\varepsilon_t] = 0$ and $\mathbb{E}[\varepsilon_t^2] = 1$. In each Monte Carlo replication, $50 + T + h$ data points for y_t are generated setting $y_0 = 0$. In order to attenuate the effect of the initial values, only the last $T + h$ data points are used for estimation and forecasting purposes. The forecasting comparisons are made in the following way: in each Monte Carlo replication, the first T observations are used to estimate the linear and the C-STAR model and then calculate the one- to seven-step-ahead forecasts.¹⁶ The procedure is repeated to generate 2500 forecast errors, for each forecasting horizon h . The forecast evaluation is based on the mean squared percent error, MSPE(h) defined on the forecast errors $e_{t+h} = y_{t+h} - \hat{y}_t(h)$, $h \geq 1$ (where $\hat{y}_t(h)$ denotes the h -step-ahead forecast at the forecast origin t).¹⁷

The simulation results are reported in Table 6. Overall, the results show that the cost of using the approximate method over the Monte Carlo approach, which should produce exact forecasts in the limit, is negligible. More specifically, the MSPE criterion shows an average gain of (i) 1% for the MC-STAR relative to the AC-STAR when using DGP(0) and DGP(3), and (ii) 2% for DGP(1) and DGP(2). The forecasting results obtained using the naive NC-STAR method are outperformed by those obtained using either the AC-STAR or MC-STAR approach. Finally, while we find that the MC-STAR and AC-STAR methods always outperform the linear specification, the results using the NC-STAR(1) are mixed.

5. Testing for the number of regimes

5.1. A modified Hansen test

Testing the hypothesis that the stochastic process under analysis can be characterized as an AR model against the C-STAR nonlinear alternative is subject to the usual difficulties that arise from the fact that the threshold parameter γ^* is not identified under the null hypothesis of linearity, thus violating conventional regularity conditions for likelihood-based inference (see for example Davies, 1977, 1987). In recent years, several methods for hypothesis testing under nonstandard conditions have been developed.¹⁸

¹⁵The order of the linear autoregressive process is chosen by means of a complexity-penalized likelihood criterion (e.g., the AIC).

¹⁶The MC-STAR multi-step forecasts are obtained by averaging over 2500 Monte Carlo replications.

¹⁷In addition we considered other measures of forecast evaluation such as forecast-encompassing tests, evaluation criteria based on correctly predicting the sign of the change of a variable and tests that evaluate the adequacy of density forecasts. However, since results are qualitatively similar, they are not reported and are available upon request.

¹⁸In the context of a threshold model, for example, Tsay (1989) suggests a graphical approach (based on the use of standardized t -ratios of an AR coefficient versus the threshold variable) to detect the number of regimes;

Table 6
Out-of-sample forecasts: MSPE(h)

y^*	h	AC-STAR	NC-STAR	MC-STAR	Linear AR
0	1	0.0021	0.0021	0.0021	0.0043
	2	0.0050	0.0064	0.0053	0.0075
	3	0.0115	0.0136	0.0113	0.0188
	4	0.0325	0.0369	0.0319	0.0405
	5	0.0522	0.0612	0.0519	0.0738
	6	0.0651	0.0698	0.0628	0.0932
	7	0.0730	0.0755	0.0722	0.1240
1	1	0.0124	0.0124	0.0124	0.0184
	2	0.0258	0.0267	0.0249	0.0302
	3	0.0297	0.0328	0.0298	0.0458
	4	0.0375	0.0389	0.0371	0.0533
	5	0.0510	0.0565	0.0496	0.0758
	6	0.0780	0.0812	0.0765	0.0976
	7	0.0983	0.1032	0.0956	0.1255
2	1	0.0179	0.0179	0.0179	0.0258
	2	0.0220	0.0263	0.0213	0.0369
	3	0.0277	0.0313	0.0255	0.0507
	4	0.0428	0.0540	0.0424	0.0651
	5	0.0512	0.0661	0.0508	0.0771
	6	0.0690	0.0811	0.0659	0.0860
	7	0.0782	0.0918	0.0758	0.0892
3	1	0.0216	0.0216	0.0216	0.0322
	2	0.0389	0.0412	0.0377	0.0398
	3	0.0501	0.0533	0.0492	0.0541
	4	0.0633	0.0681	0.0633	0.0712
	5	0.0718	0.0755	0.0716	0.0879
	6	0.0810	0.0853	0.0803	0.1002
	7	0.0828	0.0893	0.0812	0.1120

Note: MSPE(h) is the out-of-sample mean-squared percent error where h is the forecast horizon from the origin t .

Hansen (1992, 1996a) proposes a general theory for testing under such nonstandard conditions and applies it to the class of Markov switching models. He derives a bound for the asymptotic distribution of a suitably standardized likelihood ratio (LR) statistic by viewing the likelihood function as an empirical process of the unknown parameters. This asymptotic distribution is generally nonstandard, but an approximation may be obtained via simulation. In this paper we modify the Hansen procedure and apply it to the C-STAR model.

Let us consider the C-STAR(1) model (14) presented in Section 2. The C-STAR(1) reduces to a standard linear autoregressive process, AR(1), under the null hypotheses that

(footnote continued)

Hansen (1996b) proposes weighted average and supremum LM tests; Gonzalo and Pitarakis (2002) use a model selection criteria while van Dijk et al. (2003) consider using smooth transition probabilities for choosing between m and $m + 1$ thresholds.

Table 7
Modified Hansen test

y^*		$T = 100$	$T = 200$	$T = 400$	$T = 800$
0	$M = 0$	42.80	59.80	71.12	91.04
	$M = 1$	42.20	58.00	70.05	91.04
	$M = 2$	41.40	57.80	69.52	91.04
	$M = 3$	40.60	57.60	70.59	91.04
	$M = 4$	38.80	57.60	70.59	91.04
1	$M = 0$	44.50	71.17	91.00	99.50
	$M = 1$	42.50	72.39	91.00	99.50
	$M = 2$	41.50	71.78	90.50	99.00
	$M = 3$	40.50	71.78	90.00	99.00
	$M = 4$	41.00	71.17	89.50	99.00
2	$M = 0$	52.20	80.81	98.00	100.0
	$M = 1$	50.80	78.86	97.50	100.0
	$M = 2$	49.40	79.19	97.50	100.0
	$M = 3$	48.40	77.18	97.50	100.0
	$M = 4$	48.40	74.83	97.00	100.0
3	$M = 0$	59.60	74.36	92.00	100.0
	$M = 1$	59.40	74.31	91.50	100.0
	$M = 2$	57.25	73.54	91.50	99.50
	$M = 3$	57.43	72.69	91.00	99.50
	$M = 4$	57.20	72.21	89.50	99.50

Note: We report the empirical rejection of the test (calculated as the fraction of the 1000 Monte Carlo trials in which the test p -value was less than or equal to 0.05).

M is a bandwidth number. Following Hansen’s (1996a) suggestion, the test is carried using several choices of M .

$\mu_0 = \mu_1, \alpha_1^0 = \alpha_1^1$, and $\sigma_0 = \sigma_1$.¹⁹ However, even though C-STAR(1) and AR(1) are nested, conventional statistics used to test the null hypothesis (i.e., the LR statistic and the t -statistic) do not have standard null distributions. The reason for this nonstandard asymptotic behavior is that the threshold parameter y^* is unidentified under the single-state null hypothesis.

Using Hansen’s notation let $\beta = \{\mu_0 - \mu_1, \alpha_1^0 - \alpha_1^1, \sigma_0 - \sigma_1\}$, $\gamma = \{y^*\}$ and $\theta = \{\mu_0, \alpha_1^0, \sigma_0\}$. By viewing the likelihood as a function of the unknown parameters and eliminating the nuisance parameter vector θ by concentration, the likelihood function can be obtained as

$$\widehat{L}_n(\alpha) = L_n(\alpha, \widehat{\theta}(\alpha)) = \sum_{i=1}^n l_i(\alpha, \widehat{\theta}(\alpha)), \tag{23}$$

where $\alpha = (\beta, \gamma)$ and $\widehat{\theta}(\alpha) = \operatorname{argmax}_{\theta} L_n(\alpha, \theta)$. Accordingly, the LR function is defined as

$$\widehat{LR}_n(\alpha) = \widehat{L}_n(\alpha) - \widehat{L}_n(0, \gamma), \tag{24}$$

¹⁹Note that this test procedure can be easily extended to accommodate multiple regimes and lags.

while the standardized LR function is

$$\widehat{LR}_n^*(\alpha) = \frac{\widehat{LR}_n(\alpha)}{V_n(\alpha)^{1/2}}, \quad (25)$$

where $V_n(\alpha, \widehat{\theta}(\alpha)) = \sum_1^n q_i(\alpha, \widehat{\theta}(\alpha))^2$ and $q_i(\alpha, \widehat{\theta}(\alpha)) = l_i(\alpha, \widehat{\theta}(\alpha)) - l_i(0, \gamma, \widehat{\theta}(0, \gamma)) - (1/n) \widehat{LR}_n(\alpha)$. Then, the standardized LR statistic is given by

$$\widehat{LR}_n^* = \sup_{\alpha} \widehat{LR}_n^*(\alpha). \quad (26)$$

(See Appendix A for the derivation of the bound for the above standardized LR statistics.)

We conduct Monte Carlo experiments to assess the finite-sample properties of our proposed test procedure. The DGP for the simulations are those used in the previous two sections, defined in Eqs. (14) and (16). For each design point, we test the linear AR(1) model against the C-STAR(1) alternative using the modified Hansen standardized LR statistic. We use seven grid-points for both the state-dependent coefficients and the threshold parameter and the asymptotic p -values of the tests are calculated according to the method described above. Table 7 reports the empirical rejection probabilities of the tests (calculated as the fraction of 1000 Monte Carlo trials in which the test p -value was less than or equal to 0.05). The LR test seems to be powerful enough to detect C-STAR behavior, despite the fact that our test procedure uses asymptotic p -values, which are only an upper bound for the true p -values.

6. An empirical application: the short-term interest rate

Short-term interest rates have been widely modeled as processes subject to regime switching, using either Markov switching or threshold models. Both approaches attempt to capture the empirical regularity that the U.S. interest rates seem to display different dynamics across time and are used to prevent periods such as the Volcker era from affecting the estimation results. In this section we enquire whether the C-STAR model proposed in the previous sections can describe adequately the U.S. short-term interest rates. We compare the in-sample and out-of-sample performance of our C-STAR with the Markov switching and other threshold models.

We start the empirical analysis by inquiring whether the driving process under scrutiny (for the three-month U.S. T-bill on a quarterly basis from 1955:1 to 2005:2) can be characterized as a C-STAR model, using the method outlined in Section 4 to compare the linear model against the C-STAR alternative. The results of the modified Hansen test statistic for an AR(4) against a C-STAR(4) are reported in Table 8 and show that under the null hypothesis, the standardized LR statistic has, for all the choices of the band width parameter M , a p -value smaller than 0.05.²⁰ We interpret this result as strong evidence in favor of C-STAR since this test is conservative by construction.

²⁰As explained in Section 5, this procedure requires evaluation of the likelihood function across a grid of different values for the threshold parameter and for each set of regime-specific coefficients. For all cases, seven gridpoints are used. Furthermore, the p -values are calculated using 2500 random draws from the relevant limiting Gaussian processes and bandwidth parameter $M = 0, 1, \dots, 4$ (see Appendix A).

The lag order of the linear and nonlinear models is chosen using complexity-penalized likelihood criteria (e.g., the Akaike Information Criterion). In a related paper Dueker et al. (2006) propose the use of Bayesian procedures

Table 8
Modified Hansen test

$M = 0$	(0.039)
$M = 1$	(0.041)
$M = 2$	(0.047)
$M = 3$	(0.050)
$M = 4$	(0.046)
LR statistic	3.186

Notes: P -values are in brackets, M is a bandwidth number and LR is the standardized likelihood ratio statistic. Following Hansen's (1996a) suggestion, the test is carried using several choices of M .

Table 9, first column, reports the maximized log-likelihood values for the C-STAR(4) model. The estimated parameters show evidence of nonlinearity, with the estimated volatility being almost four times larger in regime 1 than in regime 0. Furthermore, the ML estimates suggest that the roots of the autoregressive regime-dependent processes are higher for regime 0 than for regime 1 (0.985 vs 0.962). The portmanteau Q statistics for the standardized residuals indicate that the fitted C-STAR model is well specified, having standardized residuals that exhibit no signs of either linear or nonlinear dependence.

We assess the stability of the model by numerical simulation. We calculate the skeleton of the model and find that there is only one stable fixed point, $y_L = 3.651$. In the top panel of Fig. 4 we plot the values taken by the skeleton and $G(y_{t-1}^{\text{skeleton}})$ along with the values of the estimated threshold, $y^* = 9.767$ and fixed point. The bottom panels show the three-month short-term interest rates and the evolution of the mixing function, $G(y_{t-1})$, which suggest that the regimes are highly persistent and that the separation is mostly associated with the Volcker period when the Federal Reserve operating instrument, between 1979 and 1982, was nonborrowed reserves.

We compare the C-STAR(4) with three other nonlinear models that have been proposed in the literature to characterize the short-term interest rate. Table 9 presents estimates of the logistic (L-STAR(4)), the exponential (E-STAR(4)) smooth transition model, along with estimates of the Markov switching (MS-AR(4)) model.²¹ Starting with the L-STAR(4), we find that the estimated threshold is close in magnitude to that obtained using the C-STAR(4) (respectively, 9.278 vs 9.767) and that the standardized residuals exhibit no signs of serial correlation. The estimated threshold value (11.995) for the E-STAR(4) is comparatively higher with residuals showing clear signs of nonlinear dependence. The plots of the estimated transition functions versus time, presented in Fig. 5 show that the separation of the regimes for the L-STAR(4) is very similar to the separation obtained using the C-STAR(4). On the other hand the transition function for the E-STAR(4) is comparatively less persistent. Turning to the MS-AR(4) model, the

(footnote continued)

to choose both the number of regimes and the number of lags for the C-STAR(p) model. Based on the same data used here, that procedure selects four lags and two regimes.

²¹For the L-STAR and E-STAR models, the appropriate lag order ($p = 4$) and the delay parameter ($d = 1$) in the threshold variable y_{t-d} are selected as suggested in Tong (1990). For the MS-AR model the lag order ($p = 4$) is selected on the basis of AIC and SBC criteria and by assessing whether the residuals of the selected model are uncorrelated. ML estimation is used on all the models.

Table 9
Maximum-likelihood estimates of STAR and Markov models

C-STAR		L-STAR		E-STAR		MS-AR	
μ_0	0.0526 (0.9559)	μ_0	2.3689 (3.5148)	μ_0	3.0550 (3.5146)	μ_0	5.2505 (0.9401)
μ_1	0.1191 (0.1155)	μ_1	0.1522 (0.1245)	μ_1	0.2110 (0.1127)	μ_1	8.1877 (0.4051)
α_1^0	1.1196 (0.0680)	α_1^0	1.0976 (0.0741)	α_1^0	1.0608 (0.0645)	α_1	1.1002 (0.0706)
α_2^0	-0.2062 (0.1010)	α_2^0	-0.1696 (0.1062)	α_2^0	-0.1476 (0.0944)	α_2	-0.2658 (0.1045)
α_3^0	0.2199 (0.0951)	α_3^0	0.2069 (0.1034)	α_3^0	0.2238 (0.0767)	α_3	0.4069 (0.0948)
α_4^0	-0.1477 (0.0607)	α_4^0	-0.1581 (0.0688)	α_4^0	-0.1745 (0.0548)	α_4	-0.2884 (0.0620)
α_1^1	0.5137 (0.2370)	α_1^1	0.3842 (0.3389)	α_1^1	0.2298 (0.3436)	–	–
α_2^1	0.0213 (0.2629)	α_2^1	-0.0438 (0.2415)	α_2^1	-0.0900 (0.2591)	–	–
α_3^1	0.6502 (0.2439)	α_3^1	0.6183 (0.2572)	α_3^1	1.1044 (0.3831)	–	–
α_4^1	-0.2227 (0.2730)	α_4^1	-0.1820 (0.2764)	α_4^1	-0.5591 (0.3934)	–	–
σ_0	0.5714 (0.0340)	σ_0	0.5807 (0.0379)	σ_0	0.5834 (0.0341)	σ_0	0.5823 (0.0309)
σ_1	2.0474 (0.2923)	σ_1	2.0513 (0.3770)	σ_1	1.9177 (0.3732)	σ_1	2.6048 (0.5488)
y^*	9.7666 (0.4427)	c	9.2784 (0.7528)	c	11.9947 (0.5283)	p	0.8417 (0.1015)
–	–	γ	1.4256 (0.5783)	γ	0.1249 (0.0474)	q	0.9895 (0.0074)
$Q(10)$	0.8701	$Q(10)$	0.8558	$Q(10)$	0.7917	$Q(10)$	0.9457
$Q(20)$	0.6353	$Q(20)$	0.6144	$Q(20)$	0.2364	$Q(20)$	0.6021
$Q_2(10)$	0.2520	$Q_2(10)$	0.8442	$Q_2(10)$	0.0000	$Q_2(10)$	0.7647
$Q_2(20)$	0.5163	$Q_2(20)$	0.9520	$Q_2(20)$	0.0001	$Q_2(20)$	0.2947
log L	-201.209	log L	-202.252	log L	-200.675	log L	-209.022
AIC	428.418	AIC	432.504	AIC	429.350	AIC	438.044
SBC	432.387	SBC	436.779	SBC	433.625	SBC	441.097

Note: The STAR models are defined as $y_t = G(\cdot)y_{0t} + (1 - G(\cdot))y_{1t}$, with $y_{jt} = \mu_j + \sum_{i=1}^4 \alpha_i^j y_{t-i} + \sigma_j \varepsilon_t$, $j = 0, 1$ with the different functions $G(\cdot)$ characterizing the three following models: C-STAR: $G(\cdot) = \frac{\Phi((y^* - \mu_0 - \sum_{i=1}^4 \alpha_i^0 y_{t-i})/\sigma_0)}{\Phi((y^* - \mu_0 - \sum_{i=1}^4 \alpha_i^0 y_{t-i})/\sigma_0) + (1 - \Phi((y^* - \mu_1 - \sum_{i=1}^4 \alpha_i^1 y_{t-i})/\sigma_1))}$; L-STAR: $G(\cdot) = (1 + \exp\{-\gamma(y_{t-1} - c)\})^{-1}$; E-STAR: $G(\cdot) = 1 - \exp\{-\gamma(y_{t-1} - c)^2\}$. The Markov switching model is defined as: MS-AR: $y_t - \mu_{s_t} = \sum_{i=1}^4 \alpha_i y_{t-i} - \mu_{s_{t-i}} + \sigma_{s_t} \varepsilon_t$, with $P(s_t = 0 | s_{t-1} = 0) = p$ and $P(s_t = 1 | s_{t-1} = 1) = q$. The figures in parentheses are autocorrelation- and heteroskedasticity-consistent standard errors. $Q(k)$ [$Q_2(k)$] is the p -value of the residual [squared-residual] Ljung–Box statistic at lag k .

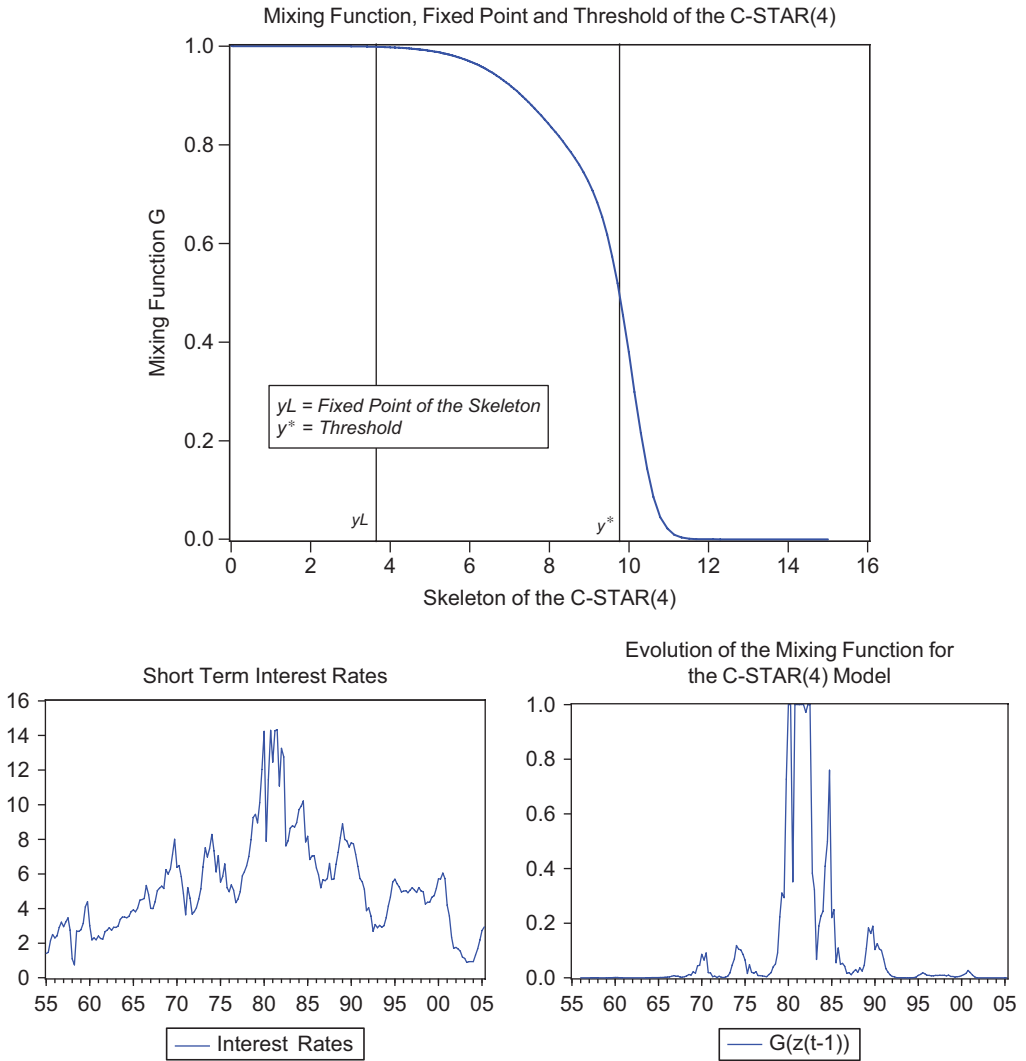


Fig. 4. Estimated short term interest rates using a C-STAR model.

estimated filtered probabilities single out the Volcker period as a different regime.²² We find, using the Akaike Information Criterion (AIC) and the Schwarz Bayesian Criterion (SBC), that the C-STAR(4) is the preferred model.²³

We finally compare the out-of-sample forecasting performance of the proposed C-STAR(4) model with that of alternative linear and nonlinear models. For the C-STAR(4)

²²We have chosen to use the Markov switching model of Hamilton (1988), which does not allow the autoregressive parameters to switch, as it is the most popular univariate Markov switching parameterization.

²³See Kapetanios (2001), Psaradakis and Spagnolo (2003, 2006) for the use of selection criteria for nonlinear models.

Separation of the Regimes Using E-STAR, L-STAR and MS Models

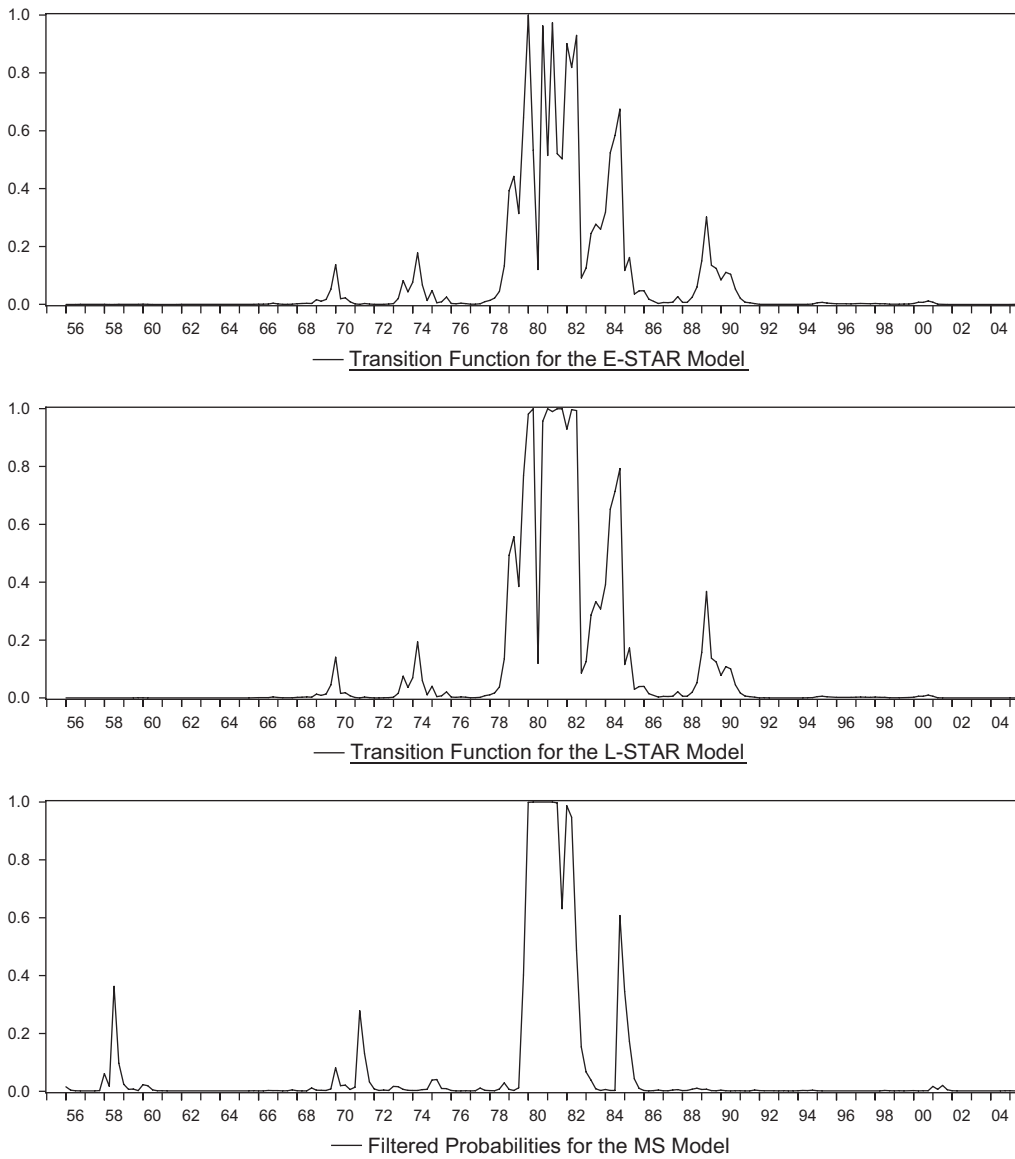


Fig. 5. Transition Functions and Filtered Probabilities of the alternative nonlinear models.

model, we based the forecasts on the approximation method, AC-STAR(4), described in Section 4 and on the Monte Carlo method, MC-STAR(4).²⁴ For the MS-AR(4) and linear AR(4) we calculated multi-step-ahead forecasts analytically, while for the E-STAR(4) and

²⁴Given the poor results obtained using N-STAR(4) approach in the simulation experiment carried out in Section 4, we exclude this method from the empirical investigation.

Table 10
Out-of-sample forecasts: MSPE(h)

h	AC-STAR	MC-STAR	L-STAR	E-STAR	MS-AR	Linear AR
1	0.0190	0.0190	0.0359	0.0426	0.0344	0.0315
2	0.0580	0.0602	0.1361	0.1780	0.0686	0.0991
3	0.1145	0.1162	0.2213	0.3810	0.1190	0.1809
4	0.2076	0.2119	0.4569	0.7708	0.1929	0.3455
5	0.3122	0.3169	0.7165	1.2297	0.2911	0.5238
6	0.4261	0.4306	1.0382	1.8349	0.5757	0.7087
7	0.5530	0.5625	1.4311	2.4816	1.9564	0.9494

Note: MSPE(h) is the out-of-sample mean squared percent error where h is the forecast horizon from the origin t .

the L-STAR(4) they are obtained via Monte Carlo simulation techniques.²⁵ The comparisons are based on series of recursive forecasts computed in the following way: for the interest rate time series $\{w_t\}_{t=1}^T$, the linear and nonlinear models are fitted to the sub-series $\{w_t\}_{t=1}^{T-\bar{h}-n}$, where \bar{h} ($= 7$) is the longest forecasting horizon under consideration, n ($= 80$) is the number of forecasts and T ($= 201$) is the sample size.²⁶ Using $t = T - \bar{h} - n$ as the forecast origin, a sequence of h -step-ahead forecasts are generated from the fitted models for $h \in \{1, \dots, \bar{h}\}$. The forecast origin is then rolled forward one period to $t = T - \bar{h} - n + 1$, the parameters of the forecast models are re-estimated and another sequence of one-step-ahead to \bar{h} -step-ahead forecasts is generated. The procedure is repeated until n forecasts are obtained for each $h \in \{1, \dots, \bar{h}\}$, which are then used to compute measures of forecast performance for each forecast horizon. All the results favor the C-STAR(4) in particular when using the AC-STAR(4) forecasting approach. A possible explanation of this result is that the mixing function of the C-STAR gives a probability forecast of the latent regime-specific variable at $t + h$ while the other STAR forecasting approaches would evaluate the mixing function at $t + h - 1$. This difference may be considerable in relatively nonpersistent regimes.

Table 10 shows that, based on the MSPE criterion, there is an average gain of almost 70% over the linear model when using the approximate AC-STAR(4). The gain decreases to almost 65% when the MC-STAR(4) is used. Turning to the other smooth transition models, the E-STAR(4) is always outperformed by the linear model, with the largest loss (over 250%) at the 7-step-ahead horizon.²⁷ The results are qualitatively similar for the L-STAR(4), with an average loss over the linear model of almost 20%. On the other hand, the results for the forecasts using the MS-AR(4) are better than those obtained using the linear model except for the 1- and 7-step-ahead forecasting horizons.

In summary, the results presented in this section are encouraging since they suggest that, for the data under scrutiny, the C-STAR model seems to not only successfully characterize the data but also, and perhaps more importantly, to have a good forecasting performance.

²⁵See Granger and Teräsvirta (1993).

²⁶Using quarterly data, the value of $\bar{h} = 7$ is associated with a two-year forecasting horizon and $n = 80$ is associated with a 20-year forecast evaluation period.

²⁷One possible reason for the poor empirical forecasting performance of the E-STAR model is (despite the fact that the fit of the model measured by the AIC and SBC criteria is similar to that of the C-STAR) that the estimated threshold for the E-STAR model is quite different from those obtained using the other STAR models. This might affect the regime-specific weights in the forecast.

The forecasting results are particularly noteworthy because one of the major weaknesses of many existing nonlinear models is their relatively poor out-of-sample performance.

7. Conclusions

In this paper we propose a new class of contemporaneous smooth transition threshold autoregressive (C-STAR) model, in which the probability that a latent variable (the current value of the regime-specific autoregressive process) exceeds a threshold value determines the regime weights. We argue that for this reason, the contemporaneous model seems to be well suited to rational expectations applications (and pricing exercises) in that it allows the regimes not to be predetermined. We discuss the properties of the model and evaluate its finite-sample ML performance. Furthermore, we propose a procedure to determine the number of regimes and evaluate the multiple-step-ahead forecasting performance of the model. Finally, an empirical application to the short-term interest rate shows that the proposed model is capable of outperforming some competing alternative nonlinear models, especially in terms of relative out-of-sample forecasting performance.

Appendix A. Modified Hansen test

For completeness we present the derivation of the bound for the standardized LR statistics. This relies on a few fairly weak regularity conditions:

Condition 1. $\sup_{\alpha} \sqrt{n} \|D(\alpha)\| = O_p(1)$, where $D(\alpha) = \widehat{\theta}(\alpha) - \theta(\alpha)$.

Condition 2. $\sup_{\alpha, \theta} \sqrt{n} \|M_n(\alpha, \theta)\| = O_p(n)$, where $M_n(\alpha, \theta) = \frac{\partial^2}{\partial \theta \partial \theta'} L_n(\alpha, \theta)$.

Condition 3. $Q_n^*(\alpha) \Rightarrow Q^*(\alpha)$, where $Q_n^*(\alpha) = \frac{Q_n^*(\alpha)}{V_n(\alpha)^{-1/2}}$ and $\widehat{Q}_n^*(\alpha) = \frac{\widehat{Q}_n^*(\alpha)}{V_n(\alpha)^{-1/2}}$ are centered stochastic process with $\widehat{Q}_n^*(\alpha) = V_n(\alpha)^{-1/2} \{\widehat{L}R_n(\alpha) - E[L_n(\alpha) - L_n(0, \gamma)]\}$, and $Q^*(\alpha) = \frac{Q(\alpha)}{V(\alpha)^{-1/2}}$ is a Gaussian process with covariance function $K^*(\alpha_1, \alpha_2) = \frac{K(\alpha_1, \alpha_2)}{V(\alpha_1)^{1/2} V(\alpha_2)^{1/2}}$, where

$$K(\alpha_1, \alpha_2) = \lim_{n \rightarrow \infty} \frac{1}{n} E[Q_n(\alpha_1) Q_n(\alpha_2)] \quad \text{and}$$

$$Q_n(\alpha) = [L_n(\alpha) - L_n(0, \gamma)] - E[L_n(\alpha) - L_n(0, \gamma)].$$

Condition 1 states that $\widehat{\theta}(\alpha)$ is consistent for $\theta(\alpha)$ at rate \sqrt{n} , uniformly in α ; condition 2 states that the matrix of second derivatives with respect to θ is well behaved, while condition 3 states that $Q_n^*(\alpha)$ satisfies an empirical process law.

Theorem. Under conditions 1–3

$$\Pr\{\widehat{L}R_n^* \geq x\} \leq \Pr\left\{\sup_{\alpha} \widehat{Q}_n^*(\alpha) \geq x\right\} \xrightarrow{n \rightarrow \infty} \Pr\{\sup Q^* \geq x\}, \tag{27}$$

Proof. See Hansen (1992).

This result provides a bound for the standardized LR statistics in terms of the distribution of the random variable $\sup Q^*$, which is generally nonstandard. The covariance function $K^*(\alpha_1, \alpha_2)$ (which completely characterizes $Q^*(\alpha)$) can be consistently

estimated by

$$K^*(\alpha_1, \alpha_2) = \frac{\widehat{K}_n(\alpha_1, \alpha_2)}{V(\alpha_1)^{1/2} V(\alpha_2)^{1/2}}, \tag{28}$$

where $\widehat{K}_n(\alpha_1, \alpha_2)$ is equal to

$$\sum_{i=1}^n q_i(\alpha_1, \widehat{\theta}(\alpha_1))q_i(\alpha_2, \widehat{\theta}(\alpha_2)) + \sum_{k=1}^M w_{kM} \left[\sum_{1 \leq i \leq n-k} q_i(\alpha_1, \widehat{\theta}(\alpha_1))q_{i+k}(\alpha_2, \widehat{\theta}(\alpha_2)) + \sum_{1+k \leq i \leq n} q_i(\alpha_1, \widehat{\theta}(\alpha_1))q_{i-k}(\alpha_2, \widehat{\theta}(\alpha_2)) \right], \tag{29}$$

$w_{kM} = 1 - |k|/(M + 1)$ is the Bartlett kernel, and M is a bandwidth number.²⁸ It follows that by repeated *i.i.d.* draws of Gaussian processes with covariance function $\widehat{K}_n^*(\alpha_1, \alpha_2)$, it is possible to obtain (approximately) the distribution $\sup_{\alpha} Q^*$ (and hence critical values and/or p -values for a test based on $\widetilde{LR}_n^*(\alpha)$). To obtain draws from the required family of Gaussian processes, Hansen (1992, 1996a) suggests to generate a random sample $\{u_i\}_{i=1}^{n+M}$ of $N(0, 1)$ variables and then construct

$$\widetilde{LR}^*(\alpha) = \frac{\sum_{k=0}^M \sum_{i=0}^n q_i(\alpha, \widehat{\theta}(\alpha))u_{i+k}}{\sqrt{1 + M V_n(\alpha)^{1/2}}}. \tag{30}$$

Then, conditional on the data, $\widetilde{LR}^*(\alpha)$ is a mean zero Gaussian process with exact covariance function $K_n^*(\alpha_1, \alpha_2)$, and the latter is an asymptotic approximation to $K^*(\alpha_1, \alpha_2)$.

Since we need to concentrate out the identified nuisance parameter θ , the constrained likelihood needs to be optimized for each value of $\alpha = (\beta, \gamma)$. A practical way to evaluate the maximal statistics is to form a grid search over a relatively small number of values of α . For every value of α at which the constrained likelihood is optimized, the sequence $\{q_i(\alpha, \widehat{\theta}(\alpha))\}$ is obtained, and from these numbers both the modified LR statistics and its asymptotic distribution are calculated.

References

Ait-Sahalia, Y., 1996. Testing continuous-time models of the spot interest rate. *The Review of Financial Studies* 9, 385–426.

Al-Qassam, M.S., Lane, J.A., 1989. Forecasting exponential autoregressive models of order 1. *Journal of Time Series Analysis* 10, 95–113.

Caner, M., Hansen, B., 1998. Threshold autoregressions with a unit root. *Econometrica* 69, 1555–1597.

Chan, K.S., Tong, H., 1985. On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations. *Advances in Applied Probability* 17, 666–678.

Chan, K.S., Tong, H., 1986. On estimating thresholds in autoregressive models. *Journal of Time Series Analysis* 7, 179–190.

Clements, M.P., Krolzig, H.-M., 1998. A comparison of the forecast performance of Markov-switching and threshold autoregressive models of US GNP. *Econometrics Journal* 1, C47–C75.

Clements, M.P., Smith, J., 1997. The performance of alternative forecasting methods for SETAR models. *International Journal of Forecasting* 13, 463–475.

Clements, M.P., Smith, J., 1999. A Monte Carlo study of the forecasting performance of empirical SETAR models. *Journal of Applied Econometrics* 14, 123–142.

²⁸Hansen (1996a) suggests carrying out the test for several choices of M .

- Davies, R.B., 1977. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 64, 247–254.
- Davies, R.B., 1987. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 74, 33–43.
- De Gooijer, J.G., De Bruin, P., 1998. On Forecasting SETAR Processes. *Statistics and Probability Letters* 37, 7–14.
- Dueker, M., Sola, M., Spagnolo, F., 2003. Contemporaneous threshold autoregressive models: estimation, forecasting and rational expectations applications. Federal Reserve Bank of St. Louis Working Paper 2003-025A.
- Dueker, M., Sola, M., Spagnolo, F., 2006. Bayesian inference in contemporaneous threshold autoregressive models. Mimeo, School of Economics, Mathematics and Statistics, Birkbeck College, University of London.
- Enders, W., Granger, C.W.J., 1998. Unit root tests and asymmetric adjustment with an example using the term structure of interest rates. *Journal of Business and Economic Statistics* 16, 304–311.
- Fletcher, R., 1987. *Practical Methods of Optimization*, second ed. Wiley, New York.
- Franses, P.H., van Dijk, D., 2000. *Nonlinear Time Series Models in Empirical Finance*. Cambridge University Press, Cambridge.
- Gonzalo, J., Pitarakis, J.Y., 2002. Estimation and model selection based inference in single and multiple threshold models. *Journal of Econometrics* 110, 319–352.
- Granger, C.W.J., Teräsvirta, T., 1993. *Modelling Nonlinear Economic Relationships*. Oxford University Press, Oxford.
- Hamilton, J.D., 1988. Rational expectations econometric analysis of changes in regime: an investigation of the term structure of interest rates. *Journal of Economic Dynamics and Control* 12, 385–423.
- Hansen, B.E., 1992. The likelihood ratio test under nonstandard conditions: testing the Markov switching model of GNP. *Journal of Applied Econometrics* 7, S61–S82.
- Hansen, B.E., 1996a. Erratum: the likelihood ratio test under nonstandard conditions: testing the Markov switching model of GNP. *Journal of Applied Econometrics* 11, 195–198.
- Hansen, B.E., 1996b. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 413–430.
- Kapetanios, G., 2001. Model selection in threshold models. *Journal of Time Series Analysis* 22, 733–754.
- Koop, G., Potter, S.M., 1999. Dynamic asymmetries in U.S. unemployment. *Journal of Business and Economic Statistics* 17, 298–312.
- Lilliefors, W.H., 1967. On the Kolmogorov–Smirnov Test for normality with mean and variance unknown. *Journal of the American Statistical Association* 62, 399–402.
- Lin, J.L., Granger, C.W.J., 1994. Forecasting from nonlinear models in practice. *Journal of Forecasting* 13, 1–9.
- Lundbergh, S., Teräsvirta, T., 2002. Forecasting with smooth transition autoregressive models. In: Clements, M.P., Hendry, D.F. (Eds.), *Companion to Economic Forecasting*. Blackwell, Oxford.
- Luukkonen, R., Saikkonen, P., Teräsvirta, T., 1988. Testing linearity against smooth transition autoregressive models. *Biometrika* 75, 491–499.
- Obstfeld, M., Taylor, A.M., 1997. Nonlinear aspects of goods-market arbitrage and adjustment: Heckscher's commodity points revisited. *Journal of the Japanese and International Economies* 11, 441–479.
- Peel, D.A., Speight, A.E.H., 1996. Is the US business cycle asymmetric? some further evidence. *Applied Economics* 28, 405–415.
- Pesaran, M.H., Potter, S.M., 1997. A floor and ceiling model of U.S. output. *Journal of Economic Dynamics and Control* 21, 661–695.
- Pfann, G.A., Schotman, P.C., Tschernig, R., 1996. Nonlinear interest rate dynamics and implications for the term structure. *Journal of Econometrics* 74, 149–176.
- Potter, S.M., 1995. A nonlinear approach to US GNP. *Journal of Applied Econometrics* 10, 109–125.
- Potter, S.M., 1999. Nonlinear time series modelling: an introduction. *Journal of Economic Surveys* 13, 505–528.
- Psaradakis, Z., Spagnolo, N., 2003. On the determination of the number of regimes in Markov-switching autoregressive models. *Journal of Time Series Analysis* 24, 237–252.
- Psaradakis, Z., Spagnolo, N., 2006. Joint determination of the state dimension and autoregressive order for models with Markov switching. *Journal of Time Series Analysis* 27, 753–756.
- Rothman, P., 1998. Forecasting asymmetric unemployment rates. *Review of Economics and Statistics* 80, 164–168.
- Teräsvirta, T., 1994. Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association* 89, 208–218.

- Teräsvirta, T., 1998. Modelling economic relationships with smooth transition regressions. In: Ullah, A., Giles, D.E.A. (Eds.), *Handbook of Applied Economic Statistics*. Marcel Dekker, New York, pp. 507–552.
- Teräsvirta, T., Anderson, H.M., 1992. Characterizing nonlinearities in business cycles using smooth transition autoregressive models. *Journal of Applied Econometrics* 7, S119–S136.
- Tiao, G.C., Tsay, R.S., 1994. Some advances in non-linear and adaptive modelling in time-series. *Journal of Forecasting* 13, 109–131.
- Tong, H., 1978. On a threshold model. In: Chen, C.H. (Ed.), *Pattern Recognition and Signal Processing*. Sijhoff and Noordoff, Amsterdam, pp. 100–141.
- Tong, H., 1983. *Threshold Models in Non-Linear Time Series Analysis*, lecture Notes in Statistics, vol. 21. Springer, Berlin.
- Tong, H., 1990. *Non-Linear Time Series: a Dynamical Systems Approach*. Oxford University Press, Oxford.
- Tong, H., Lim, K.S., 1980. Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society* 42, 245–292.
- Tsay, R.S., 1989. Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association* 84, 231–240.
- van Dijk, D., Teräsvirta, T., Franses, P.H., 2002. Smooth transition autoregressive models—a survey of recent developments. *Econometric Reviews* 21, 1–47.
- van Dijk, D., Strikholm, B., Teräsvirta, T., 2003. The effects of institutional and technological change and business cycle fluctuations on seasonal patterns in quarterly industrial production series. *Econometrics Journal* 6, 79–98.