

IHS Economics Series  
Working Paper 79  
March 2000

# Forecasting European GDP Using Self-Exciting Threshold Autoregressive Models: A Warning

Jesús Crespo-Cuaresma





INSTITUT FÜR HÖHERE STUDIEN  
INSTITUTE FOR ADVANCED STUDIES  
Vienna

## Impressum

---

**Author(s):**

Jesús Crespo-Cuaresma

**Title:**

Forecasting European GDP Using Self-Exciting Threshold Autoregressive Models: A Warning

**ISSN: Unspecified**

**2000 Institut für Höhere Studien - Institute for Advanced Studies (IHS)**

Josefstädter Straße 39, A-1080 Wien

[E-Mail: office@ihs.ac.at](mailto:office@ihs.ac.at)

Web: [www.ihs.ac.at](http://www.ihs.ac.at)

All IHS Working Papers are available online: [http://irihs.ihs.ac.at/view/ihs\\_series/](http://irihs.ihs.ac.at/view/ihs_series/)

This paper is available for download without charge at:

<https://irihs.ihs.ac.at/id/eprint/1254/>

**Forecasting European GDP Using Self-Exciting  
Threshold Autoregressive Models**  
A Warning

**Jesus Crespo-Cuaresma**

# **Forecasting European GDP Using Self-Exciting Threshold Autoregressive Models**

## **A Warning**

**Jesus Crespo-Cuaresma**

Reihe Ökonomie / Economics Series No. 79

**March 2000**

Institut für Höhere Studien  
Stumpergasse 56, A-1060 Wien  
Fax: +43/1/599 91-163

Jesus Crespo-Cuaresma  
Phone: +43/1/599 91-156  
E-mail: [crespo@ihs.ac.at](mailto:crespo@ihs.ac.at)

**Institut für Höhere Studien (IHS), Wien**  
**Institute for Advanced Studies, Vienna**

The Institute for Advanced Studies in Vienna is an independent center of postgraduate training and research in the social sciences. The **Economics Series** presents research carried out at the Department of Economics and Finance of the Institute. Department members, guests, visitors, and other researchers are invited to submit manuscripts for possible inclusion in the series. The submissions are subjected to an internal refereeing process.

**Editorial Board**

*Editor:*

Robert M. Kunst (Econometrics)

*Associate Editors:*

Walter Fisher (Macroeconomics)

Klaus Ritzberger (Microeconomics)

## **Abstract**

A two-regime self-exciting threshold autoregressive process is estimated for quarterly aggregate GDP of the fifteen countries that compose the European Union, and the forecasts from this nonlinear model are compared, by means of a Monte Carlo simulation, with those from a simple autoregressive model, whose lag length is chosen to minimize Akaike's AIC criterion. The results are very negative for the SETAR model when the Monte Carlo procedure is used to generate multi-step forecasts. When the "naive" procedure of generating forecasts is used, the results are surprisingly better for the SETAR model in long-term predictions. Due to the characteristics of the residuals, a bootstrapping method of forecasting was also used, yielding even poorer results for the nonlinear model.

## **Keywords**

Nonlinear Time Series Models, SETAR models, forecasting

## **JEL Classifications**

C53, C52, C22

**Comments**

The author acknowledges the helpful comments of Robert M. Kunst during the confection of the paper.

# Contents

- 1. Introduction 1**
- 2. The Linear Model 1**
- 3. The SETAR Model: Estimation 3**
  - 3.1. Hansen's Method 3
  - 3.2. A SETAR Model for European GDP 4
- 4. Forecasting with SETAR and Linear Models 5**
  - 4.1. The MC Method 6
  - 4.2. The SK Method 7
- 5. Forecasting European GDP: SETAR vs. AR Models 7**
- 6. Conclusions 13**
- 7. References 14**



# 1 Introduction

Threshold models have raised great interest among economists since Tong's seminal contributions [Tong (1983, 1990)]. Nevertheless, the evidence existing on the forecasting superiority of threshold autoregressive models (TAR) and, more precisely, self-exciting threshold autoregressive models (SETAR) when compared to linear models is relatively small. At first sight, the possibility of generating asymmetries that non-linear models possess could make them capable of better forecasting variables related to business cycle fluctuations such as employment or GDP, where asymmetric behaviour appears to be a stylized fact. Rothman (1998) brings out evidence on the superior forecasting capabilities of non-linear models (among them, self-exciting threshold autoregressive models) over univariate linear models when forecasting the unemployment rate in the USA. The same conclusion regarding the better forecasting properties of SETAR models is achieved by Pippenger and Goering (1998) and Chappell et al. (1996) on exchange rate data, where the scheme of exchange rate bands is also intuitively appealing for threshold modelling. This paper contributes to the comparison between forecasts from linear and nonlinear models by estimating both an autoregressive and a SETAR model for European GDP and comparing the forecasts induced by both of them out of the sample used for estimation. The paper is organized as follows: Section 2 deals with the estimation of the AR model for European GDP. Section 3 implements the estimation of the SETAR model; while the estimation method for the AR model is standard, the SETAR model is estimated through Hansen's (1997, 1999) procedure. Section 4 deals with the complications derived from forecasting with nonlinear models. Section 5 compares the forecasts from both models through Monte Carlo simulations, using the "naive" method of obtaining forecasts, and through a bootstrapping procedure. Section 6 concludes.

## 2 The Linear Model

Quarterly data on the aggregate GDP of the fifteen countries that currently compose the European Union is available from the OECD's *Main Economic Indicators* database from the first quarter of 1960 until the first quarter of 1999. First differences of the logarithmized series are used for the present study<sup>1</sup>. The series of interest is plotted in Figure 1 for the whole available range.

Observations up to 1994:4 are used for the estimation of both the linear and the SETAR model, and the interval 1995:1-1999:1 will be used for comparing the out-of-sample forecasting performance of both models. The optimal lag length of the autoregressive process to be estimated is chosen using Akaike's AIC criterion (Akaike, 1973): Autoregressive processes with lag length ranging from 1 to 8 are estimated and the AIC statistic is computed:

$$AIC = -2\hat{l}/n + 2k/n,$$

where  $\hat{l}$  is the estimated log-likelihood of the model,  $k$  is the number of estimated parameters and  $n$  is the number of observations. The number of lags to be included in the

---

<sup>1</sup>The results of the augmented Dickey-Fuller test suggested taking first differences from the logarithmized series of GDP in order to achieve stationarity. The null of a unit root was accepted at any sensible significance level (both a constant and a trend component were used in the testing procedure).

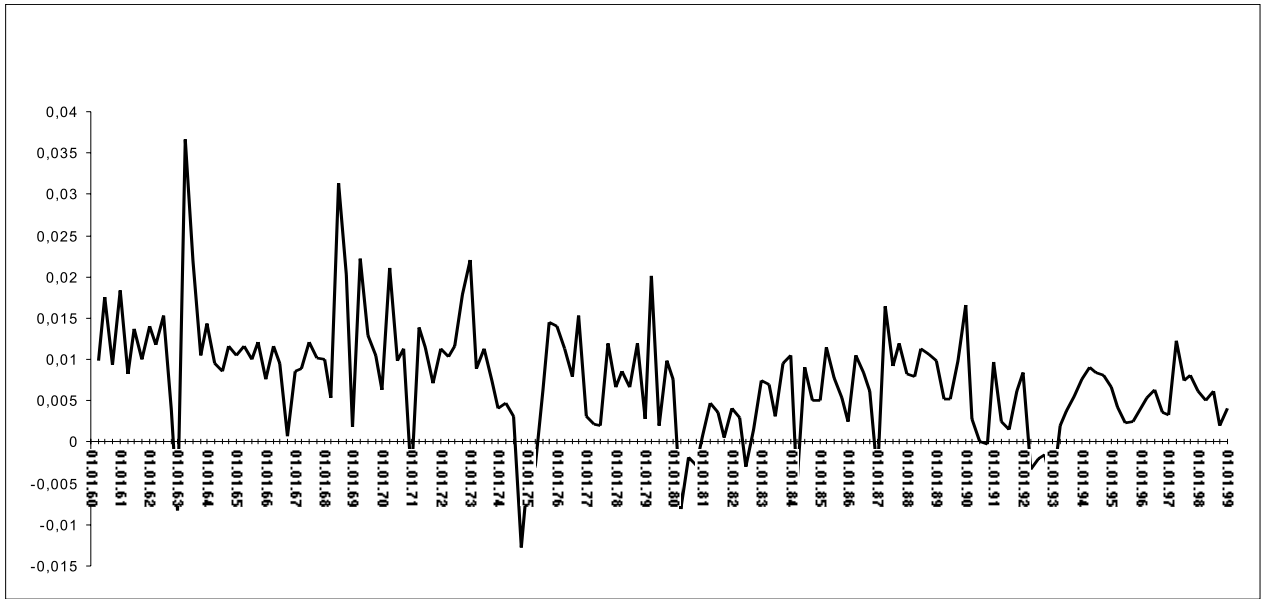


Figure 1: First difference of logarithmized GDP: Europe-15 (1960:1-1999:1)

autoregressive process is chosen from the model that minimizes the value of AIC. Table 1 reports the results for 1960:1-1994:4: the first column refers to the lag length of the estimated AR process, the second one to the AIC corresponding to that lag and the third column reports the sum of squared residuals of the estimated model.

Lag length (g)	AIC= $-2\hat{l}/n + 2k/n$	SSR
1	-7.09321	0.00652
2	-7.10281	0.00631
3	-7.11149	0.00612
4	-7.11072	0.00599
5	-7.09092	0.00597
6	-7.07177	0.00594
7	-7.04927	0.00594
8	-7.02864	0.00592

Table 1: AIC Statistics: AR(g) models

A lag length of three is optimal according to this procedure of model selection, and the estimated process is

$$y_t = 0.004016 + 0.20964y_{t-1} + 0.09376y_{t-2} + 0.17318y_{t-3} + \mu_t, \quad (1)$$

where the  $y$ s are first differences of logarithmized European GDP and  $\mu_t$  is assumed to

be a sequence of i.i.d. random variables distributed according to a Gaussian distribution with mean zero and constant variance  $\sigma_\mu^2$ . Except for the parameter of the second lag of  $y_t$ , whose  $t$ -statistic is 1.082, all estimated parameters are statistically significant at 5% significance level. Ljung and Box's  $Q$ -test [Box and Ljung (1978)], using up to 36 lags, was applied to the residuals of (1), and it showed no departure from iid-ness.

### 3 The SETAR Model: Estimation

Before reporting the results of the estimation of the 2-regime SETAR model for European GDP, a brief explanation of the method used is required. The next subsection is mainly based on Hansen (1997,1999).

#### 3.1 Hansen's Method

Let the two-regime SETAR model to be estimated be of the form

$$y_t = (\alpha_0 + \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p}) I(y_{t-d} \leq \gamma) + (\beta_0 + \beta_1 y_{t-1} + \dots + \beta_p y_{t-p}) I(y_{t-d} > \gamma) + \epsilon_t, \quad (2)$$

where  $I(\cdot)$  is the indicator function, which takes value 1 if its argument is true and zero if it is false, and  $\epsilon_t$  is assumed to be a sequence of independent, identically distributed random variables with expectation 0 and constant variance  $\sigma^2$ . The autoregressive order of the model is  $p \geq 1$ ,  $y_{t-d}$  is the threshold variable and  $\gamma$  is the threshold parameter.

Let us rewrite the model in equation (2), by defining

$$\begin{aligned} x_t &= (1 \quad y_{t-1} \quad \dots \quad y_{t-p})', \\ x_t(\gamma) &= (x_t' I(y_{t-d} \leq \gamma) \quad x_t' I(y_{t-d} > \gamma))', \\ \alpha &= (\alpha_0 \quad \alpha_1 \quad \dots \quad \alpha_p)', \\ \beta &= (\beta_0 \quad \beta_1 \quad \dots \quad \beta_p)' \end{aligned}$$

and

$$\psi = (\alpha' \quad \beta')$$

as

$$y_t = x_t(\gamma)' \psi + \epsilon_t. \quad (3)$$

The parameters to be estimated, apart from the delay parameter,  $d$ , are  $\gamma$  and  $\psi$ , and the estimation method to be used [Hansen(1997)] will be sequential conditional least squares. Using the least squares estimator of  $\psi$  for a given value of  $\gamma$ ,

$$\hat{\psi}(\gamma) = \left( \sum_{t=1}^n x_t(\gamma) x_t(\gamma)' \right)^{-1} \left( \sum_{t=1}^n x_t(\gamma) y_t \right),$$

the estimation problem is reduced to finding  $\hat{\gamma}$  that minimizes the sum of squared residuals of the model (that now depends exclusively on  $\gamma$ ), and recovering the estimates of  $\alpha$  and  $\beta$  through  $\hat{\psi}(\hat{\gamma})$ . The search for  $\hat{\gamma}$  is done by evaluating the sum of squared residuals of the model fixing the threshold variable to be equal to  $y_{t-d}$ ,  $t = d + 1, \dots, n$ , in the sample

space. Thus, we simplify the problem of finding  $\hat{\gamma}$  in a continuous interval  $\Gamma = [\gamma_1, \gamma_2]$  by replacing  $\Gamma$  by  $\tilde{\Gamma} = \{y_{t-d} \text{ for all } d+1 \leq t \leq n\}$ , which has a finite – and not too enormous – number of elements  $[n - (d+1)]$ . Estimating the delay parameter,  $d$ , amounts to extending the search just explained by allowing the variable  $d \in \mathcal{N}$ ,  $1 \leq d \leq D$ , to be also a choice variable in the discrete optimization problem. That implies that we need to evaluate the sum of squared residuals on  $[n - (d+1)]D$  models, and pick the pair  $(\hat{\gamma}, \hat{d})$  that minimizes it.

The procedure for testing for threshold models against linear ones allows us, as well, to construct confidence intervals for the threshold parameter: the LR-type test statistic to be used for such a testing procedure is

$$F_n = n \left( \frac{\tilde{\sigma}_n^2 - \hat{\sigma}_n^2}{\hat{\sigma}_n^2} \right),$$

where  $\tilde{\sigma}_n^2$  is the residual variance for the restricted ( $\alpha = \beta$ ) linear model and  $\hat{\sigma}_n^2$  is the residual variance of the alternative threshold model. The pointwise  $F$  statistic for the test  $H_0 : \alpha = \beta$  against  $H_1 : \alpha \neq \beta$  for a given  $\gamma$  is, thus,

$$F = \sup_{\gamma \in \Gamma} F_n(\gamma) = \sup_{\gamma \in \Gamma} n \left( \frac{\tilde{\sigma}_n^2 - \hat{\sigma}_n^2(\gamma)}{\hat{\sigma}_n^2(\gamma)} \right).$$

Such a test statistic has a non-standard distribution [see, e.g., Andrews and Ploberger (1994)], which has been approximated by Hansen (1999) using a bootstrapping procedure. Confidence intervals can be, therefore, calculated by inverting the  $F_n(\gamma)$  statistic. Let  $c(\phi)$  be the  $\phi$ -level critical value of the distribution of  $F_n(\gamma)$ . Then the 100  $\phi$  % confidence interval for  $\gamma$ ,  $\hat{\Gamma}$ , is

$$\hat{\Gamma} = \{\gamma : F_n(\gamma) \leq c(\phi)\}.$$

### 3.2 A SETAR Model for European GDP

In the framework of the procedure described hitherto, a simple two-regime SETAR model will be estimated for the data on European GDP, and its forecasting capabilities will be discussed and compared to the ones of the linear model described in the previous section. For comparison reasons, when estimating the model the number of lags of the autoregression to be included in each regime will be set to be equal to the number of lags in the competing autoregressive process, and the delay parameter will not be allowed to exceed this number of lags. Table 2 reports the sum of squared residuals for the models considered, and the p-value for the test of the null of linearity against the threshold model considered in each case calculated through Hansen's bootstrapping method. The method for computing this statistic is the one described above, and both the top and bottom 5% quantiles of the candidates to threshold variable have been trimmed. The SETAR(2;3,3) model estimated for the observations ranging from 1960:1 to 1994:4 is

$$y_t = (0.02161 + 2.8278y_{t-1} - 2.6206y_{t-2} + 3.03937y_{t-3})I(y_{t-1} \leq -0.004564) + \\ + (0.003469 + 0.30044y_{t-1} + 0.0945y_{t-2} + 0.1157y_{t-3})I(y_{t-1} > -0.004564) + \epsilon_t,$$

where  $\epsilon_t$  is a sequence of i.i.d. random variables, assumed to be normally distributed with zero mean and constant variance.

Threshold variable	$p$ -value	SSR
$y_{t-1}$	0.004	0.005065
$y_{t-2}$	0.479	0.005692
$y_{t-3}$	0.037	0.005343

Table 2: Choosing the appropriate SETAR model

Notice the non-stationary, explosive nature of the process for  $y_{t-1} \leq -0.004564$ , leading to “jumps” to the upper process. This model of GDP presents, thus, steep expansions followed by flatter contractions, a stylized fact constantly reported in the real business cycle literature. In order to illustrate such a feature, Figure 2 graphs a simulation of the process defined by (4), where the error process has been taken to have variance  $\hat{\sigma}_\epsilon$ , the estimated residual variance of equation (4).

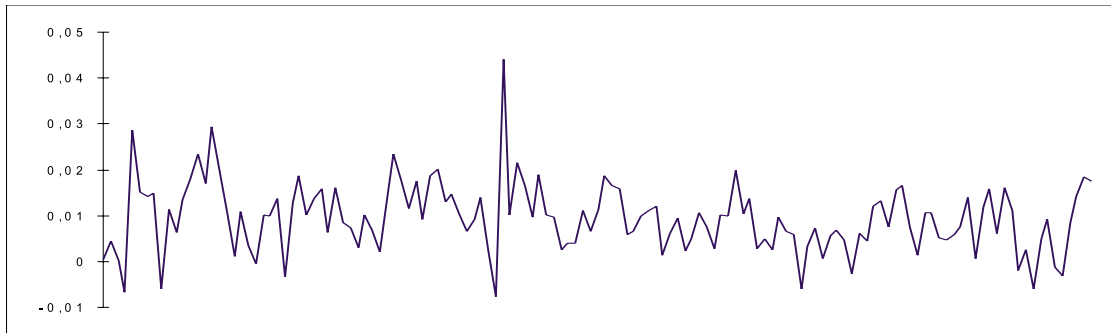


Figure 2: Simulation of the SETAR(2,3,3) process.

Figure (3) shows the in-sample fit of both the AR and the SETAR model for the period 1960:1-1994:4. As can be seen from the plot, the SETAR models accounts better for the deep recession of the mid-seventies and the explosive episode in the beginning of the sixties. Nevertheless, as Clements and Smith (1997) remark, “*neither in-sample fit nor the rejection of the null of linearity (...) guarantee that SETAR models (...) will forecast more accurately than linear AR models*”. The following sections justify such a comment quantitatively.

## 4 Forecasting with SETAR and Linear Models

The problems arising when calculating multi-step forecasts in nonlinear models are well known<sup>2</sup>. The complication that nonlinear models poses when trying to forecast  $h$ -steps ahead ( $h \geq 2$ ) is based on the fact that, if  $g(\cdot)$  is a nonlinear function,

$$\mathbb{E}[g(\cdot)] \neq g(\mathbb{E}[\cdot]),$$

<sup>2</sup>For a clear, brief review see Granger and Terasvirta (1993).

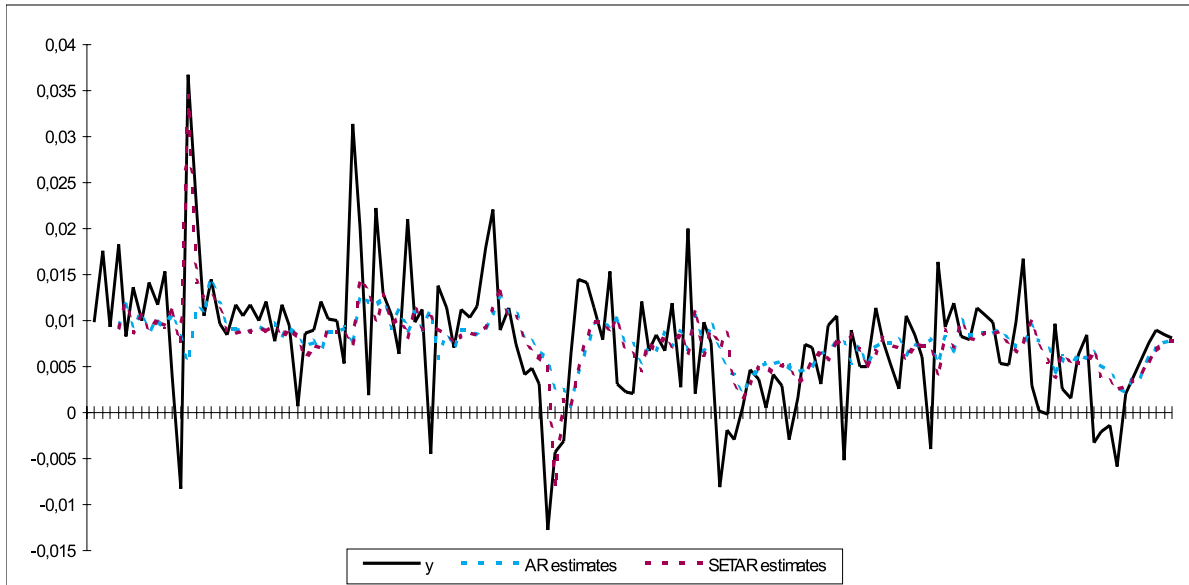


Figure 3: In-sample fit: AR and SETAR models

where  $\mathbb{E}[\cdot]$  is the expectation operator. Thus, closed-form analytic expressions for multi-step forecasts do not exist for the case of SETAR models. Clements and Smith (1997) compare several methods for obtaining  $h$ -steps ahead minimum mean square error forecasts for SETAR models, and conclude on the usefulness of a simple one: the Monte Carlo method. The next subsection briefly explains the procedure for obtaining forecasts of a SETAR model using Monte Carlo (MC) simulations. This method will be used in the comparison of the forecast power of the SETAR and AR model, together with the “naive” or “skeleton” (SK) method.

#### 4.1 The MC Method

Consider a simple SETAR model such as the one specified by equation (2). If we have data up to period  $T$  and want to forecast  $y_{T+1}$ , the regime of the SETAR process is known, and the one-step forecast for the MC procedure is just

$$\hat{y}_{T+1} = (\alpha_0 + \alpha_1 y_T + \dots + \alpha_p y_{T-p+1}) I(y_T \leq \gamma) + (\beta_0 + \beta_1 y_T + \dots + \beta_p y_{T-p+1}) I(y_T > \gamma).$$

However, when forecasting  $y_{T+2}$  we only have  $\hat{y}_{T+1}$  [which differs from  $y_{T+1}$  by an error term when considering (2) the DGP] in order to decide on the regime in which the process is. For a given realization of the error process, the forecasts for period  $T+2$  and  $T+3$  are

$$\begin{aligned} \hat{y}_{T+2}^j &= (\alpha_0 + \alpha_1 \hat{y}_{T+1} + \dots + \alpha_p y_{T-p+2}) I(\hat{y}_{T+1} \leq \gamma) + \\ &\quad + (\beta_0 + \beta_1 \hat{y}_{T+1} + \dots + \beta_p y_{T-p+2}) I(\hat{y}_{T+1} > \gamma) + \epsilon_{T+2}^j \\ \hat{y}_{T+3}^j &= (\alpha_0 + \alpha_1 \hat{y}_{T+2}^j + \dots + \alpha_p y_{T-p+3}) I(\hat{y}_{T+2}^j \leq \gamma) + \\ &\quad + (\beta_0 + \beta_1 \hat{y}_{T+2}^j + \dots + \beta_p y_{T-p+3}) I(\hat{y}_{T+2}^j > \gamma) + \epsilon_{T+3}^j \end{aligned}$$

Simulating  $J$  replications of the error process, the forecast for  $y_{T+h}$  would be

$$\hat{y}_{T+h} = 1/J \sum_{j=1}^J \hat{y}_{T+h}^j,$$

Following this reasoning, and using a high enough number of replications of the error process,  $h$ -steps ahead ( $h > 2$ ) forecasts can be computed.

## 4.2 The SK Method

The “naive” or SK method amounts to approximating  $\mathbf{E}[g(\cdot)]$  by  $g(\mathbf{E}[\cdot])$ , and can be interpreted as a special case of the MC method in which the errors are set to zero. Thus, through the SK method

$$\begin{aligned} \hat{y}_{T+2} = & (\alpha_0 + \alpha_1 \hat{y}_{T+1} + \dots + \alpha_p y_{T-p+2}) I(\hat{y}_{T+1} \leq \gamma) + \\ & + (\beta_0 + \beta_1 \hat{y}_{T+1} + \dots + \beta_p y_{T-p+2}) I(\hat{y}_{T+1} > \gamma), \end{aligned}$$

and

$$\begin{aligned} \hat{y}_{T+h} = & (\alpha_0 + \alpha_1 \hat{y}_{T+h-1} + \dots + \alpha_p \hat{y}_{T-p+h}) I(\hat{y}_{T+h-1} \leq \gamma) + \\ & + (\beta_0 + \beta_1 \hat{y}_{T+h-1} + \dots + \beta_p \hat{y}_{T-p+h}) I(\hat{y}_{T+h-1} > \gamma). \end{aligned}$$

## 5 Forecasting European GDP: SETAR vs. AR Models

The forecasting experiment is designed as follows. Starting in the last quarter of 1994, models (1) and (4) are using for computing  $s$ -steps ahead forecasts ( $s=1,2,\dots,8$ ) through both the MC and the SK method described above. When a new observation comes up, the AR model is reestimated, choosing its optimal lag length ( $d_1$ ) through AIC. Conditional on this lag length for both regimes, the SETAR(2;  $d_1, d_1$ ) is estimated by searching through the delay parameter and the threshold and choosing the best model measured by the  $p$ -value of the F-test described in section 2.1. The forecasts are then computed again, and the procedure is repeated for the new coming observation:  $d_2$ , the new delay parameter and the new threshold are chosen, and forecasts are computed again. 1000 replications of the error process -assumed iid normal with mean zero- were used in each step for the MC procedure, where the variance was chosen in each step to be equal to the residual variance of the model considered.

Once the forecasts were obtained, the following statistics were computed for each step, model and procedure:

- Root Mean Squared Error (RMSE):  $\{1/N \sum [A_t - F_t]^2\}^{1/2}$ ,
- Mean Absolute Deviation (MAD):  $1/N \sum |A_t - F_t|$ ,
- Theil’s U statistic (U):  $\text{RMSE}/[1/N \sum A^2]$ ,
- Confusion Rate (CR): Number of wrongly forecasted moves (up/down)/Number of observations to be forecasted.

AR model				
steps ahead	RMSE	MAD	U	CR
2	0.00326	0.00279	0.5656	0.7142
3	0.00326	0.00275	0.5579	0.6923
4	0.00334	0.00275	0.5540	0.75
5	0.00298	0.00245	0.4802	0.4545
6	0.00304	0.00250	0.4781	0.50
7	0.00314	0.00263	0.4866	0.4444
8	0.00316	0.00269	0.4883	0.25
SETAR model				
steps ahead	RMSE	MAD	U	CR
2	0.00359	0.00286	0.6229	0.6428
3	0.00357	0.00294	0.6112	0.6923
4	0.00348	0.00300	0.57725	0.5833
5	0.00336	0.00274	0.5408	0.4545
6	0.00334	0.00265	0.5254	0.60
7	0.00307	0.00253	0.4765	0.3333
8	0.00325	0.00281	0.5021	0.625

Table 3: Forecasting comparison: MC procedure



AR model				
steps ahead	RMSE	MAD	U	CR
1	0.00290	0.00231	0.4984	0.60
2	0.00326	0.00278	0.5655	0.7142
3	0.00331	0.00285	0.5665	0.7692
4	0.00326	0.00271	0.5411	0.75
5	0.00303	0.00253	0.4885	0.2727
6	0.00304	0.00251	0.4779	0.50
7	0.00307	0.00256	0.4762	0.4444
8	0.00318	0.00266	0.4914	0.625
SETAR model				
steps ahead	RMSE	MAD	U	CR
1	0.00314	0.00230	0.5404	0.60
2	0.00350	0.00276	0.6081	0.6428
3	0.00354	0.00293	0.6063	0.6153
4	0.00339	0.00286	0.5628	0.5833
5	0.00274	0.00223	0.4407	0.4545
6	0.00277	0.00222	0.4363	0.60
7	0.00245	0.00210	0.38028	0.4444
8	0.00260	0.00203	0.40307	0.625

Table 4: Forecasting comparison: SK procedure

$A$  refers to the actual values of the series and  $F$  to the forecasted ones. For a more extensive explanation of these concepts, see Holden et al. (1990) or Clements and Hendry (1998).

The results are shown in Tables 3 and 4. The optimal lag length for the AR process (and, therefore, the lag length of each one of the regimes of the SETAR process) was 3 for the interval 1994:4-1996:3. When the observation corresponding to 1996:4 enters the sample for estimation, the optimal lag length according to AIC is 4, and the optimal delay lag for the SETAR model, which was 1 until 1996:3, is 4 according to the procedure explained above. The threshold variable is, thus,  $y_{t-1}$  for the range 1994:4-1996:3, and  $y_{t-4}$  for 1996:4-1998:4. The threshold is -0.00456 in the first case and 0.00991 in the second one. Furthermore, both regimes of the SETAR processes for the range 1996:4-1998:4 have roots outside the unit circle.

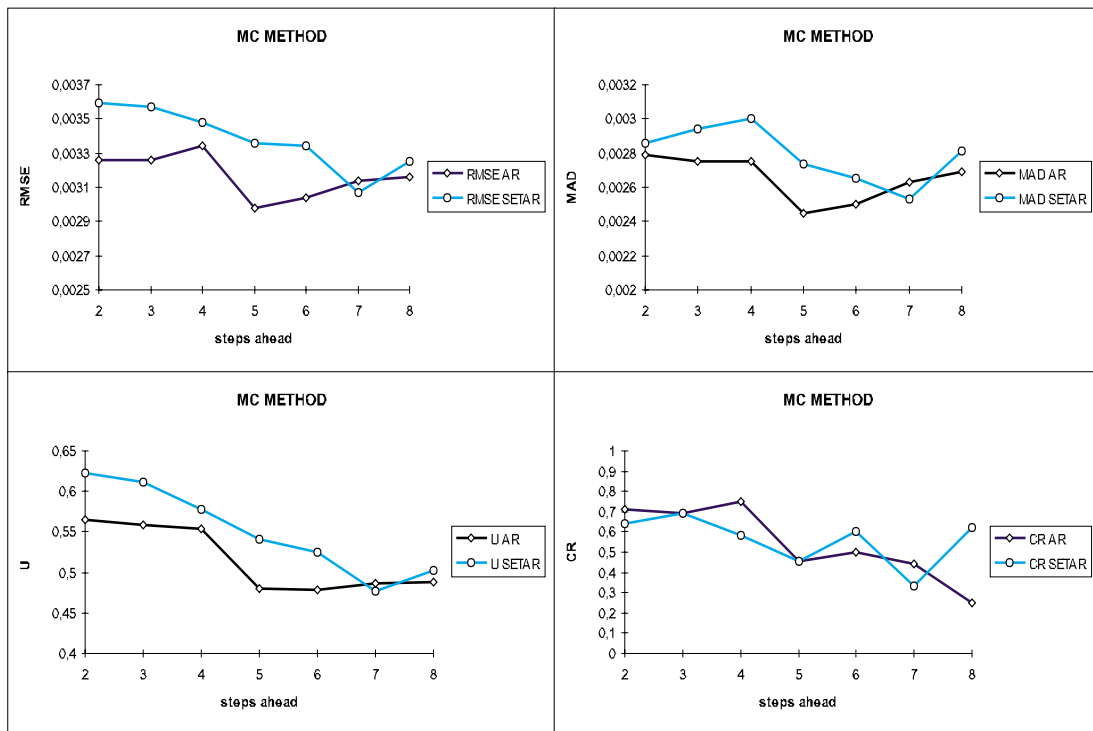


Figure 4: Forecasting comparison: MC procedure

Figures 4 and 5 present the comparison graphically. The results are rather surprising: for a start, the SETAR model performs much better when using the SK procedure than it does when using the MC, which is more adequate to its nonlinear nature. In the MC simulations, the AR model gives superior forecasts (measured in MSFE, MAD or U) for every forecasting horizon except for seven steps ahead. The results for the CR are not decisive for the case of the MC method of obtaining multi-step forecasts. The forecasting failure of the SETAR model can be explained by the non stationary nature of the contractive regime in the range 1994:4-1996:3, which could have “inflated” many of the replications of the Monte Carlo simulation, resulting in misleading forecasts. This hypothesis is sustained by

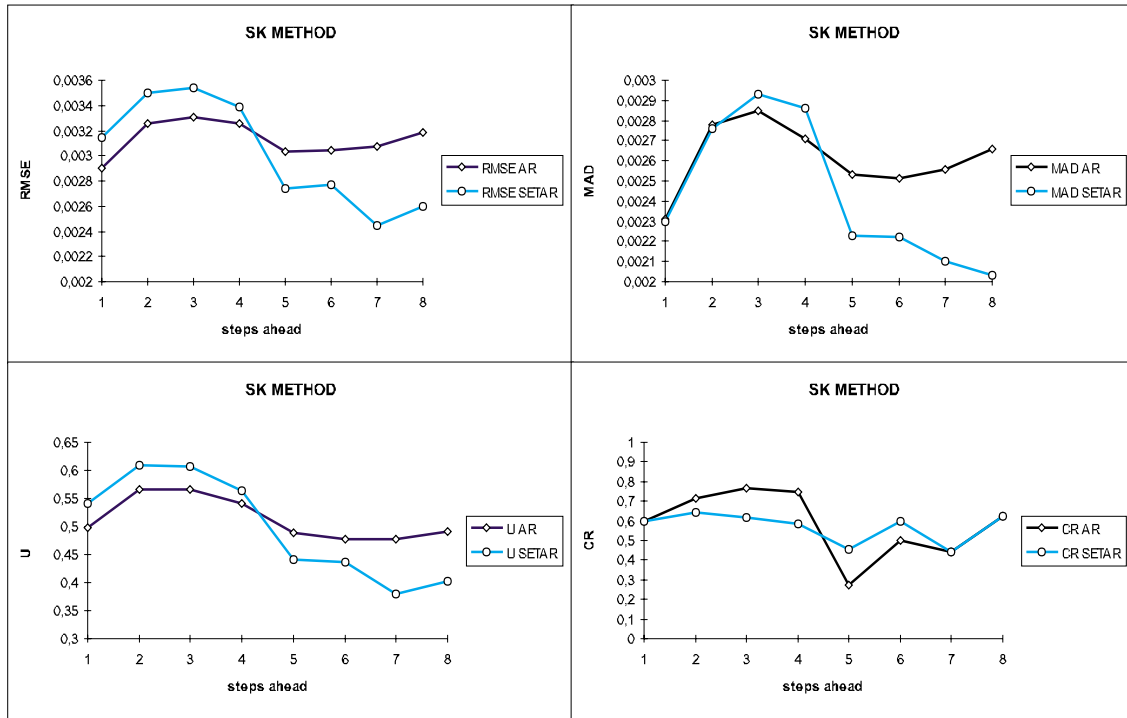


Figure 5: Forecasting comparison: SK procedure

the better performance in the SK method.

In any case, the results for the SETAR forecasts with the SK procedure are also rather disappointing: the AR model beats the SETAR model or performs equally well for all statistics when forecasting up to one year (four quarters) ahead. It is only for longer forecasting horizons that the SETAR model proves to give better forecasts than the simple linear model.

Another explanation for the poor performance of the SETAR model has to do with the features of the errors: in the MC simulation the replicated errors are normally distributed with mean zero and variance equal to the estimated variance of the model fitted to the available observations in each period. However, the residuals of both the AR and the SETAR model in (1) and (4) differ strongly from normality: the Jarque-Bera test [Jarque and Bera (1980)] applied to the residuals of (1) and (4) rejects the null of normality at 1% significance level. The same happens for the residuals of the out-of-sample reestimated models for every range of observations.

The sensible step to take is, thus, to repeat the forecasting experiment using a bootstrapping (BS) method for obtaining multi-step forecasts. The method is similar to MC, with the exception that the errors are drawn randomly from the residuals of the regression performed at each observational point; that is, once an AR(4) and a SETAR(2;4,4) have been fitted to, for example, the observations up to 1997:2, the residuals of this regression are used as source of the errors for the multi-step forecasts, in a similar fashion to the procedure explained in section 4.1. If the forecast failure of the SETAR model resides on the assumed normality of the error process, the forecasting exercise using bootstrapping should obtain better results for the nonlinear model when compared to the linear autore-

AR model				
steps ahead	RMSE	MAD	U	CR
2	0.00326	0.00276	0.5655	0.7857
3	0.00329	0.00276	0.5626	0.7692
4	0.00318	0.00265	0.5285	0.75
5	0.00306	0.00259	0.4926	0.3636
6	0.00291	0.00243	0.4585	0.60
7	0.00298	0.00248	0.4622	0.3333
8	0.00317	0.00265	0.4906	0.50
SETAR model				
steps ahead	RMSE	MAD	U	CR
2	0.00352	0.00283	0.6117	0.7142
3	0.00400	0.00328	0.6840	0.6153
4	0.00334	0.00273	0.5541	0.5833
5	0.00362	0.00248	0.5828	0.6363
6	0.00675	0.00458	1.0617	0.50
7	0.01492	0.01138	2.3151	0.3333
8	0.03524	0.02802	5.4533	0.75

Table 5: Forecasting comparison: BS procedure

gressive model. 1000 bootstrap replications of the error process have been generated for each step ahead forecast in each model, and the results are reproduced in Table 5. The forecasting superiority of the AR over the SETAR model when the forecasts are obtained through the BS procedure is even more devastating than when the other procedures of obtaining multi-step forecasts are used, and the SETAR forecasts are especially bad for longer horizon forecasts, which is the strength of the nonlinear model under the SK procedure. The fact that the empirical distribution of the residuals for the SETAR model is leptokurtic for all estimation ranges may explain the poor results, as errors would tend to drive the long horizon forecasts into the “contractive” regime, which is non-stationary for the SETAR models computed up to observation 1996:4, inflating the value of the forecasts. Is the non-stationary nature of the “expansive” regime the only responsible for the poor forecasting abilities of the SETAR(2;3,3) model? The following experiment is aimed to show if the potentially explosive expansive regime of the estimated SETAR model is causing the poor forecasting performance of the nonlinear model when compared to the simple linear AR process. We replaced our original SETAR(2;3,3) for a SETAR(2;3,1) for the estimation horizons where the expansive regime ( $y_{t-1} \leq -0.004564$ ) was nonstationary. The new expansive regime is

$$y_t = -0.8y_{t-1} + \epsilon_t,$$

Adjusted SETAR model				
steps ahead	RMSE	MAD	U	CR
2	0.00363	0.00290	0.6310	0.5714
3	0.00362	0.00299	0.6187	0.6153
4	0.00344	0.00298	0.5713	0.6666
5	0.00325	0.00264	0.5233	0.6363
6	0.00331	0.00265	0.5206	0.60
7	0.00301	0.00246	0.4866	0.4444
8	0.00299	0.00244	0.4631	0.50

Table 6: Forecasting statistics of the SETAR(2;3,1) model: MC procedure

which is expected to make observations in this regime be reflected into the “contractive” regime, and fulfills stationarity. Table 6 reports the forecasting statistics of this adjusted SETAR(2;3,1) model for the Monte Carlo procedure of computing  $h$ -steps ahead forecasts. Although 7 and 8-steps ahead forecasts improve with the new model, and the SETAR(2;3,1) is superior to the AR model in these forecasting horizons, still non-stationarity does not account for the failure in short and medium term forecasting, where the SETAR model with non-stationary expansive regime displays better forecasting features measured in mean square forecasting error.

## 6 Conclusions

We have compared the forecasts of a SETAR model on European GDP with those of a simple AR model by means of a Monte Carlo method, proved to perform well for self-exciting threshold models, and a simple, “naive” method. The nonlinear model performs rather poorly in terms of most of the statistics used to measure forecast accuracy, and is outperformed by the AR model in all except for one of the forecasts horizons used when we implement the MC method. Such a performance improves with the “naive” or SK method in the long run forecasts, but the SETAR model keeps on being beaten by the AR model for forecast horizons up to four periods. Due to the rejection of the null of normality when applying the Jarque-Bera test to the residuals of the estimated models, a bootstrapping method for obtaining multi-step forecasts was also implemented, and the results of the SETAR model were considerably worse than in the other cases. The study performed strengthens the results of, for example, Diebold and Nason (1990) or De Gooijer and Kumar (1992), concluding that the forecasting supremacy of nonlinear models when compared to linear ones is not evident. The sometimes excellent in-sample fit of SETAR models on real business variables does not mirror itself in excellence of out-of-sample forecasting, as has been demonstrated in the present paper for European GDP.

## References

- [1] Akaike, H. (1973): “Information Theory and an extension of the Maximum Likelihood principle,” in *Second International Symposium of Information Theory*, Petrov, B. N. and F. Csaki (eds.).
- [2] Andrews, D.W.K. and W. Ploberger (1994): Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* **62**, 1383–1414.
- [3] Box, G. E. P. and G. M. Ljung (1978): On a measure of lack of fit in time series models. *Biometrika* **66**, 262–270.
- [4] Chappell, D., Padmore, J., Mistry, P. and C. Ellis (1996): A threshold model for the French Franc/Deutschmark exchange rate. *Journal of Forecasting* **15**, 155–164.
- [5] Clements, M. P. and D. F. Hendry (1998): *Forecasting Economic Time Series*, Cambridge University Press, Cambridge.
- [6] Clements, M. P. and J. Smith (1997): The performance of alternative forecasting methods for SETAR models. *International Journal of Forecasting* **13**, 463–475.
- [7] De Gooijer, J. G. and K. Kumar (1992): Some recent developments in nonlinear time series modelling, testing and forecasting. *International Journal of Forecasting* **8**, 135–156.
- [8] Diebold, F. X. and J. A. Nason (1990): Nonparametric exchange rate prediction. *Journal of International Economics* **28**, 315–332.
- [9] Granger, C. W. J. and T. Terasvirta (1993): *Modelling Nonlinear Economic Relationships*. Oxford University Press, Oxford.
- [10] Hansen, B. E. (1997): Inference in TAR models. *Studies in Nonlinear Dynamics and Econometrics* **2**, 1–14.
- [11] Hansen, B. E. (1999): Sample splitting and threshold estimation. *Econometrica*, forthcoming.
- [12] Holden, K., Peel, D. A. and J. L. Thompson (1990): *Economic forecasting: an introduction*. Cambridge University Press, Cambridge.
- [13] Jarque, C. M. and A. K. Bera (1980): Efficient tests for normality, heteroscedasticity and serial independence of regression residuals. *Economic Letters* **6**, 255–259.
- [14] MacKinnon, J. G. (1991): Critical values for cointegration tests, in *Long run economic relationships: Readings in cointegration*; Engle, R. F. and C. W. J. Granger (eds.), Oxford University Press.
- [15] Pippenger, M. K. and G. E. Goering (1998): Exchange rate forecasting: Results from a threshold autoregressive model. *Open Economies Review*, **9**, 157–170.
- [16] Rothman, P. (1998): Forecasting asymmetric unemployment rates. *Review of Economic and Statistics*, **80**, 164–168.

- [17] Tong, H. (1983): *Threshold models in non-linear time series analysis. Lecture notes in Statistics, 21.* Berlin, Springer.
- [18] Tong, H. (1990): *Non-linear time series. A dynamical system approach.* Oxford University Press, Oxford.

**Institut für Höhere Studien**  
**Institute for Advanced Studies**

Stumpergasse 56

A-1060 Vienna

Austria

Phone: +43-1-599 91-145

Fax: +43-1-599 91-163