A COMPARISON OF ARMA AND SETAR FORECASTS

Paul T. De Bruin and Jan G. De Gooijer*

Department of Economic Statistics, University of Amsterdam,

Roetersstraat 11, 1018 WB Amsterdam, The Netherlands

Abstract: In this paper seven monthly Indian macroeconomic time series are investigated. The series are tested for threshold nonlinearity using various nonlinearity tests. Three types of models are fitted to the data, namely autoregressive moving average (ARMA), self-exciting threshold autoregressive (SETAR), and subset SETAR models. Next genuine multi-step ahead out-of-sample forecasts are computed for the series. The relative forecasting performances of the models are compared with two forecast criteria. It appears that the forecasting accuracy of the SETAR models is sensitive to the choice of various parameters in the model as well as to transformations of the series to stationarity.

Keywords: Forecasting, Nonlinear Time Series, Testing.

*Corresponding author.

1 Introduction

Interest in using nonlinear time series methods has recently become popular in the literature. Quite some papers have appeared on the identification, estimation, and testing of nonlinear models. However, relatively little attention has been given to their out-of-sample forecast performance. Indeed De Gooijer and Kumar (1992) argue that the evidence for a significant forecast performance of nonlinear models is patchy, when compared with linear models. The reason is that forecast comparisons usually concern one-step-ahead forecast evaluations of nonlinear models. One notable exception is the paper by Ray (1988). Using 10 Indian monthly macroeconomic time series he reports that bilinear models are the best followed by subsequently linear autoregressive moving average (ARMA) models and self-exciting threshold autoregressive (SETAR) models, in forecasting up to 12 periods ahead. Unfortunately, Ray's results are based upon within-sample comparisons of the fitted models. So his final conclusion is still open for questioning.

In this paper we report our experience with genuine multi-step ahead out-of-sample forecasting. To this end we re-analyze seven of the ten series considered by Ray (1988). Next, using 21 newly obtained observations for each series, we compare the ex post forecasting performance of ARMA models with SETAR models. These latter models have a wide variety of applications in economics. For instance, Kräger and Kugler (1993) and Chappell et al. (1996) fit SETAR models to exchange rate data. Other direct application include models of separating and multiple equilibria, and economic processes which follow an asymmetric business cycle. In contrast, application of bilinear models in the economic literature has been very limited.

The paper is organized as follows. Section 2 provides a brief discussion of the time series analyzed in this study. In Section 3, we introduce the basic SETAR process and three tests for threshold nonlinearity. We also apply these tests to the series. Section 4 gives an overview of the ARMA and SETAR models used in the forecasting competition. Section 5 discusses a simulation method to obtain multi-step ahead out-of-sample forecasts from a SETAR model. This section also introduces two criteria to evaluate the forecast performance of the various models. The relative out-of-sample forecast performance of SETAR and ARMA models will be assessed in Section 6. The final section concludes.

Table 1: Indian macroeconomic time series.

Series	Description of Series	Period	Effective Number of Observations (T)
1.	Currency notes in circulation	1/70-3/85	170
2.	Rupee securities	4/70-3/84	155
3.	Aggregate deposits of all scheduled commercial banks	4/70-3/84	155
4.	Bank credit of all scheduled commercial banks	4/70-3/84	155
5.	Money supply (M_3)	4/70-3/84	155
6.	Index numbers of wholesale prices	4/71-3/84	155
7.	Index numbers of consumer prices for industrial workers	4/70-3/84	155

2 Data

In this paper we investigate the forecasting performance of SETAR models fitted to the seven monthly seasonally unadjusted Indian macroeconomic time series listed in Table 1. The series are a subset of ten series analyzed by Ray (1988) who calibrated the forecasting performance of his models within the sample period. In contrast, using 21 additional observations for each series, we analyze the out-of-sample performance. However no extra data were available for the following series: gold and foreign exchange, index numbers of industrial production, and value of exports. Hence we omitted them from the forecasting comparison.

Following the conventional practice of using transformed and stationary data in time series analysis, Ray (1988) used the logarithms (base 10) for five series, with exceptions being the index number for wholesale prices and the index number of consumer prices for industrial workers. Furthermore, he made six variables stationary by using the first order differencing operator $\nabla = (1-B)$ in conjunction with seasonal differencing operator $\nabla_{12} = (1 - B^{12})$, where B denotes the delay operator. For the wholesale prices only the ∇ operator was used. The last column in Table 1 gives for each series the effective number of observations (T) examined by Ray (1988). Figure 1 contains the graphs of the transformed and differenced series, including the data in the out-of-sample period. Graphs of the original series, without the out-of-sample observations, are presented by Ray (1988).

Figure 1 about here

3 SETAR Model and Linearity Tests

A time series is said to follow a self-exciting autoregressive (SETAR) process of order $(2; p_1, p_2)$ if it satisfies the difference equation

$$Y_t = \begin{cases} \alpha_{10} + \sum_{i=1}^{p_1} \alpha_{1i} Y_{t-i} + \epsilon_{1t} & \text{if } Y_{t-d} \leq r, \\ \alpha_{20} + \sum_{i=1}^{p_2} \alpha_{2i} Y_{t-i} + \epsilon_{2t} & \text{if } Y_{t-d} > r. \end{cases} \tag{1}$$

Here $\{\epsilon_{jt}\}$ $(j = 1, 2)$ is a sequences of independent and identically distributed (*i.i.d.*) random variables with mean zero and variance σ_j^2 such that ϵ_{1t} and ϵ_{2t} are independent; d is a positive integer called the delay parameter (or threshold lag); and r is the threshold. The non-negative integers p_1, p_2 and d are assumed known and are such that $0 \leq p_2 \leq p_1$ and $1 \leq d \leq p_1$. For ease of deriving the first test it is also assumed that $\{\epsilon_{1t}\} = \{\epsilon_{2t}\} = \{\epsilon_t\}$. Then model (1) can be rewritten in the form

$$Y_t - \sum_{i=0}^{p_1} \alpha_{1i} Y_{t-i} - I(Y_{t-d} \leq r)(\sum_{j=0}^{p_2} \delta_j Y_{t-j}) = \epsilon_t, \tag{2}$$

where is $I(.)$ is the indicator function, and $\delta_j = (\alpha_{2j} - \alpha_{1j})$ $(j = 0, 1, \ldots, p_2)$. Clearly, if $p_1 = p_2 = p$ and δ_j=0 for every $j = 0, 1, \ldots, p_2$ (2) reduces to a linear autoregressive (AR) model of order p.

In this section the object is to test the null hypothesis of linearity in (2), i.e.

$$H_0 : \quad \delta_j = 0, \quad \text{for } j = 0, 1, \ldots, p_2 \quad \text{versus} \quad H_1 : \delta_j \neq 0, \quad \text{for some } 0 \leq j \leq p_2.$$

Under H_0 the nuisance parameter r is absent. Assume that $\{\epsilon_t\}$ is normally distributed, then the likelihood ratio (LR) test statistic is given by $\lambda = \{\hat{\sigma}^2(NL; r)/\hat{\sigma}^2(L)\}^{(T-p_2+1)/2}$ where T denotes the sample size, $\hat{\sigma}^2(L)$ and $\hat{\sigma}^2(NL; r)$ are the residual variances under H_0 and model (2) respectively after a least squares fit to the data; see, e.g., Chan and Tong (1990). Under H_0, $-2\ln(\lambda)$ converges asymptotically to a $\chi^2_{p_1+1}$. If the threshold r is unknown, the LR test statistic is

$$\lambda' = \{\hat{\sigma}^2(NL; \hat{r})/\hat{\sigma}^2(L)\}^{(T-p_2+1)/2} \tag{3}$$

where $\hat{r}$ is a least squares estimate of r. The asymptotic distribution of (3) is no longer χ^2. Indeed, Chan and Tong (1990) showed that the asymptotic null distribution of λ' is related to the maximum of a continuous parameter Gaussian process. Percentage points of the asymptotic null distribution of (3) have been tabulated by Chan (1991) for p_1 from 0 to 5 and from 6 to 18 in multiples of 3.

Three other tests have also been proposed in the literature for testing linearity against threshold nonlinearity: they are the CUSUM test of Petruccelli and Davis (1986), the TAR-F test of Tsay

(1989), and the New-F test of Tsay (1991). The CUSUM is based on the idea that if the estimation of a linear AR model is performed sequentially according to the increasing order of the value of the given threshold variable, one can effectively separate the data from regime to regime if the process is indeed a SETAR model. Also the predictive residuals can be computed recursively. If the process is linear, then the standardized predictive residuals are asymptotically normal random variables. The TAR-F test combines the idea of CUSUM and F-tests. The F-statistic is obtained by regressing the predictive residuals of an arranged autoregression (as in CUSUM) on the regressors $\{1, Y_{t-1}, \ldots, Y_{t-m}\}$, where m is a certain order. A large F-statistic implies that there are model changes in the series.

Tsay (1991) proposed a New-F test for testing linearity against threshold nonlinearity, exponential nonlinearity, and bilinearity. The procedure is as follows. First, choose a threshold variable Y_{t-d}, fit an arranged autoregression of order m to the process $\{Y_t\}$ and calculate the normalized predictive residuals $\hat{e}_t$ for $t = m+1, m+2, \ldots, T$. Then regress $\hat{e}_t$ on the regressor $\{1, Y_{t-1}, \ldots, Y_{t-m}\}$, $\{Y_{t-1}\hat{e}_{t-1}, \ldots, Y_{t-m}\hat{e}_{t-m}\}$, $\{\hat{e}_{t-1}\hat{e}_{t-2}, \ldots, \hat{e}_{t-m}\hat{e}_{t-m-1}\}$, and $\{Y_{t-1}\exp(-Y_{t-1}^2/\gamma), G(Z_{t-d}), \quad Y_{t-1}G(Z_{t-d})\}$, where $\gamma = \max|Y_{t-1}|$, $Z_{t-d} = (Y_{t-d} - \overline{Y}_d)/S_d$ with $\overline{Y}_d$ and S_d being the sample mean and standard deviation of Y_{t-d} respectively, and G(.) is the cumulative distribution function of the standard normal random variable. If $\{Y_t\}$ is a stationary AR(p) process of order $p \leq m$, the F-statistic of this regression follows asymptotically an F-distribution with degrees of freedom $3(m+1)$ and $T - m - 3(m+1)$.

For the LR test (3) we employed the STAR PC package of Tong (1990). The optimal AR order p was selected on the basis of Akaike's information criterion (AIC). Since the delay parameter is often unknown in applications, the set $\{1, 2, 3\}$ was used as the possible values for d. The values of λ' are given in the fourth column of Table 2. No values for the λ' are reported for $d > p$ since the STAR program cannot handle this case. For the TAR-F test and the New-F test columns five and six of Table 2 contain the p-values of respectively the TAR-F test and the New-F test. For each series, the last line gives test results for the maximum value of p=11 allowed in the STAR package and the value for d selected for the SETAR models fitted by Ray (1988).

Note that for Series 3 all tests suggest nonlinearity for $p = d$=1 and for $p = 4$, d=1 at the 5% level. The LR test seems to suggest some threshold nonlinearity for p=10 and d=2 in Series 5 and for p=11 and d=2 in Series 6 whereas the F tests suggest linearity for both series. The TAR-F test finds nonlinearity in Series 7 for p=4 and d=2,3. In contrast the New-F test suggests nonlinearity in this series for p=4 and d=1,3. Furthermore, the New-F test seems to suggest some nonlinearity

Table 2: Linearity test results.

Series	p	d	LR	TAR-F (p-value)	New-F (p-value)
1.	1	1	2.2	0.80	0.96
		2	-	0.43	0.90
		3	-	0.36	0.84
	11	2	3.5	0.58	0.49
2.	1	1	1.1	0.44	0.91
		2	-	0.67	0.71
		3	-	0.32	0.84
	11	2	2.1	0.62	0.26
3.	4	1	46.0*	0.00	0.00
		2	10.5	0.49	0.61
		3	9.7	0.83	0.28
	11	1	40.4*	0.01	0.16
4.	1	1	5.8	0.23	0.00
		2	-	0.54	0.10
		3	-	0.25	0.00
	11	5	6.0	0.79	0.63
5.	10	1	19.0	0.85	0.94
		2	34.5*	0.41	0.71
		3	18.9	0.50	0.85
	11	1	24.2	0.90	0.87
6.	1	1	2.7	0.93	0.87
		2	-	0.81	0.83
		3	-	0.78	0.94
	11	2	31.4*	0.52	0.64
7.	4	1	6.5	0.06	0.04
		2	9.4	0.02	0.35
		3	8.8	0.01	0.01
	11	2	25.4	0.20	0.53

Note: * denotes significant at the 5% level.

in Series 4 for p=1 and d=1,3 whereas the other tests suggest linearity. In addition we also tested for nonlinearity in the first differences of the logs (base 10) of the series of gold and foreign exchange prices, the $\nabla\nabla_{12}$ differenced series of the index of industrial production, and the first differences of the logs (base 10) of the value of exports. All the F-tests indicate nonlinearities for p=2 and d=1,2,3 in the gold and foreign exchange series. At $p = d$=2 this result is also confirmed by the LR test. The tests do not indicate nonlinearities for the other two series. Clearly our test results are different from those reported by Ray (1988). Using the bispectral test proposed by Subba Rao and Gabr (1980) he concludes that all ten time series are nonlinear. However, it has been reported in the literature that this test has very low power in detecting threshold nonlinearity.

4 Models

We now consider the ARMA models fitted by Ray (1988). For each transformed data series these models are reproduced in lines i) of Table 3. Using exact maximum likelihood we were able to reproduce reasonably well the parameter estimates for Series 1-5, and 7. For Series 6 our checking suggested that the sign of the MA parameter at lag 12 should be positive rather than negative as reported by Ray (1988). Unfortunately, it is difficult to interpret the parameters without their standard errors, which have not been given by this author. Indeed, our estimation results indicated that several non-seasonal parameters are not statistically different from zero at the 5% level. This suggests that the fitted ARMA models can be improved by a more systematic reparametrization and re-estimation of the parameters. Nevertheless, we accept Ray's models as a benchmark for the forecasting comparison without modifying them.

Ray (1988, Table 8) reports SETAR models fitted to the data selected by AIC. These models, with coefficients rounded to two decimal places, are reproduced in lines ii) of Table 3. There "0.00" denotes that the corresponding parameter value is less than 0.005. Note that the fit is rather peculiar. Specifically, the number of parameters in the SETAR models is extremely large: 17 for Series 1,2,5, and 6; 25 for Series 4; 32 for Series 7; and 36 parameters for Series 3. Now it is not easy to justify a model with 36 parameters for 155 observations. Moreover, the pooled residual variances, which we computed from the normalised AIC values given by Ray, are in almost all cases larger than the residual variances of the fitted ARMA models having only 2 to 5 parameters This casts some doubt on the validity of Ray's conclusion with respect to the forecasting performance of his parsimoniously fitted ARMA models versus his large SETAR models.

On checking Ray's SETAR models we noted that quite a few parameters were not significantly different from zero at the 5% level. Keeping the delay parameter d and the threshold value r fixed at the values selected by Ray we re-estimated the SETAR models dropping insignificant lags. The fitted models together with $\hat{\sigma}_j$ $(j = 1, 2)$ are given in lines iii) of Table 3. We shall refer to these latter models as the modified SETAR models. They are certainly a substantial improvement over the SETAR models in lines ii) in terms of interpretation, parsimony, and statistical goodness-of-fit. Note that we were unable to estimate a modified SETAR model for Series 7.

Table 3: Fitted models.

Series	Model	$\hat{\sigma}_\epsilon$ for i) $\hat{\sigma}_j$ for ii) and iii)
1.	i) $(1+0.10B^2)Y_t=(1-0.92B^{12})\epsilon_t$	0.58×10^{-2}
	ii) $Y_t=\begin{cases} 0.07\times10^{-2}-0.02Y_{t-1}+0.02Y_{t-2}+0.12Y_{t-3}-0.05Y_{t-4} \\ +0.01Y_{t-5}+0.04Y_{t-6}+0.12Y_{t-7}+0.11Y_{t-8}+0.19Y_{t-9} \\ +0.02Y_{t-10}+0.20Y_{t-11}-0.57Y_{t-12}-0.10Y_{t-13}+\epsilon_{1t} & \text{if } Y_{t-2}\le 0.0031 \\ -0.02\times10^{-2}+\epsilon_{2t} & \text{if } Y_{t-2}>0.0031 \end{cases}$	0.58×10^{-2} 0.86×10^{-2}
	iii) $Y_t=\begin{cases} 0.21Y_{t-9}+0.20Y_{t-11}-0.56Y_{t-12}+\epsilon_{1t} & \text{if } Y_{t-2}\le 0.0031 \\ -0.02\times10^{-2}+\epsilon_{2t} & \text{if } Y_{t-2}>0.0031 \end{cases}$	0.63×10^{-2} 0.87×10^{-2}
2.	i) $Y_t=(1+0.12B^3)(1-0.94B^{12})\epsilon_t$	0.76×10^{-2}
	ii) $Y_t=\begin{cases} 0.02\times10^{-2}+0.03Y_{t-1}-0.03Y_{t-2}+0.03Y_{t-3}-0.10Y_{t-4} \\ +0.00Y_{t-5}+0.04Y_{t-6}+0.11Y_{t-7}+0.12Y_{t-8}+0.15Y_{t-9} \\ +0.06Y_{t-10}+0.17Y_{t-11}-0.64Y_{t-12}-0.02Y_{t-13}+\epsilon_{1t} & \text{if } Y_{t-2}\le 0.0022 \\ 0.02\times10^{-2}+\epsilon_{2t} & \text{if } Y_{t-2}>0.0022 \end{cases}$	0.69×10^{-2} 0.11×10^{-1}
	iii) $Y_t=\begin{cases} -0.67Y_{t-12}+\epsilon_{1t} & \text{if } Y_{t-2}\le 0.0022 \\ -0.24Y_{t-1}+\epsilon_{2t} & \text{if } Y_{t-2}>0.0022 \end{cases}$	0.76×10^{-2} 0.11×10^{-1}
3.	i) $(1+0.23B-0.14B^3)(1-0.15B^{12})Y_t=(1-0.85B^{12})\epsilon_t$	0.36×10^{-2}
	ii) $Y_t=\begin{cases} 0.01\times10^{-2}+0.13Y_{t-1}-0.05Y_{t-2}+0.13Y_{t-3}-0.29Y_{t-4} \\ +0.04Y_{t-5}+0.02Y_{t-6}-0.19Y_{t-7}+0.09Y_{t-8}+0.12Y_{t-9} \\ +0.18Y_{t-10}-0.21Y_{t-11}-0.41Y_{t-12}-0.12Y_{t-13} \\ +0.08Y_{t-14}+\epsilon_{1t} & \text{if } Y_{t-1}\le -0.0010 \\ 0.19\times10^{-2}-0.66Y_{t-1}+0.27Y_{t-2}+0.12Y_{t-3} \\ +0.17Y_{t-4}+0.20Y_{t-5}+0.06Y_{t-6}+0.03Y_{t-7}-0.09Y_{t-8} \\ +0.10Y_{t-9}-0.03Y_{t-10}+0.07Y_{t-11}-0.38Y_{t-12} \\ -0.17Y_{t-13}-0.11Y_{t-14}+0.06Y_{t-15}-0.16Y_{t-16} \\ +0.35Y_{t-17}+0.09Y_{t-18}+\epsilon_{2t} & \text{if } Y_{t-1}>-0.0010 \end{cases}$	0.26×10^{-2} 0.36×10^{-2}
	iii) $Y_t=\begin{cases} -0.16Y_{t-2}-0.22Y_{t-4}-0.28Y_{t-12}+\epsilon_{1t} & \text{if } Y_{t-1}\le -0.0010 \\ 0.16\times10^{-2}-0.57Y_{t-1}+0.35Y_{t-2}+\epsilon_{2t} & \text{if } Y_{t-1}>-0.0010 \end{cases}$	0.29×10^{-2} 0.43×10^{-2}
4.	i) $(1+0.18B-0.12B^3)Y_t=(1-0.92B^{12})\epsilon_t$	0.78×10^{-2}
	ii) $Y_t=\begin{cases} 0.14\times10^{-2}-0.51Y_{t-1}-0.19Y_{t-2}+0.00Y_{t-3}-0.22Y_{t-4} \\ -0.16Y_{t-5}-0.11Y_{t-6}+0.07Y_{t-7}+0.56Y_{t-8}+0.18Y_{t-9} \\ +0.07Y_{t-10}-0.06Y_{t-11}-0.78Y_{t-12}-0.50Y_{t-13} \\ -0.07Y_{t-14}-0.24Y_{t-15}-0.54Y_{t-16}-0.39Y_{t-17} \\ -0.32Y_{t-18}+0.04Y_{t-19}+\epsilon_{1t} & \text{if } Y_{t-5}\le -0.0036 \\ -0.04\times10^{-2}-0.20Y_{t-1}+0.04Y_{t-2}+\epsilon_{2t} & \text{if } Y_{t-5}>-0.0036 \end{cases}$	0.51×10^{-2} 0.99×10^{-2}
	iii) $Y_t=\begin{cases} -0.42Y_{t-1}-0.24Y_{t-2}+0.41Y_{t-8}+0.27Y_{t-9}-0.73Y_{t-12} \\ -0.35Y_{t-13}-0.23Y_{t-15}-0.36Y_{t-16}-0.18Y_{t-17}+\epsilon_{1t} & \text{if } Y_{t-5}\le -0.0036 \\ -0.2227Y_{t-1}+\epsilon_{2t} & \text{if } Y_{t-5}>-0.0036 \end{cases}$	0.57×10^{-2} 0.10×10^{-1}
5.	i) $(1-0.17B^3)Y_t=(1-0.84B^{12})\epsilon_t$	0.31×10^{-2}
	ii) $Y_t=\begin{cases} -0.05\times10^{-2}-0.09Y_{t-1}+0.11Y_{t-2}+0.16Y_{t-3}+0.02Y_{t-4} \\ -0.16Y_{t-5}+0.14Y_{t-6}+0.10Y_{t-7}+0.06Y_{t-8}+0.07Y_{t-9} \\ -0.25Y_{t-10}+0.09Y_{t-11}-0.56Y_{t-12}+0.01Y_{t-13}+\epsilon_{1t} & \text{if } Y_{t-1}\le 0.0007 \\ -0.04\times10^{-2}+\epsilon_{2t} & \text{if } Y_{t-1}>0.0007 \end{cases}$	0.63×10^{-2} 0.87×10^{-2}
	iii) $Y_t=\begin{cases} 0.18Y_{t-3}-0.18Y_{t-10}-0.54Y_{t-12}+\epsilon_{1t} & \text{if } Y_{t-1}\le 0.0007 \\ 0.31Y_{t-7}+\epsilon_{2t} & \text{if } Y_{t-1}>0.0007 \end{cases}$	0.28×10^{-2} 0.38×10^{-2}
6.	i) $(1-0.44B)Y_t=(1+0.17B^{12})\epsilon_t$	2.44
	ii) $Y_t=\begin{cases} 0.68+0.43Y_{t-1}-0.01Y_{t-2}+\epsilon_{1t} & \text{if } Y_{t-2}\le 1.20 \\ 0.51+0.41Y_{t-1}+0.11Y_{t-2}+0.13Y_{t-3}-0.15Y_{t-4}-0.10Y_{t-5} \\ +0.09Y_{t-6}-0.02Y_{t-7}-0.39Y_{t-8}+0.25Y_{t-9}+0.18Y_{t-10} \\ +0.09Y_{t-11}+\epsilon_{2t} & \text{if } Y_{t-2}>1.20 \end{cases}$	2.22 2.32
	iii) $Y_t=\begin{cases} 0.68+0.42Y_{t-1}+\epsilon_{1t} & \text{if } Y_{t-2}\le 1.2 \\ 0.58Y_{t-1}-0.35Y_{t-8}+0.33Y_{t-9}+0.25Y_{t-10}+\epsilon_{2t} & \text{if } Y_{t-2}>1.2 \end{cases}$	2.27 2.46
7.	i) $(1-0.88B)(1+0.22B^{12})Y_t=(1-0.62B+0.13B^3)(1-0.85B^{12})\epsilon_t$	2.75
	ii) $Y_t=\begin{cases} -0.19+0.36Y_{t-1}+0.27Y_{t-2}+0.10Y_{t-3}+0.23Y_{t-4} \\ +0.07Y_{t-5}-0.03Y_{t-6}+0.07Y_{t-7}-0.02Y_{t-8}-0.01Y_{t-9} \\ -0.22Y_{t-10}-0.09Y_{t-11}-0.61Y_{t-12}+0.16Y_{t-13}+0.40Y_{t-14} \\ +0.15Y_{t-15}+\epsilon_{1t} & \text{if } Y_{t-2}\le 0 \\ 2.17+0.08Y_{t-1}-0.41Y_{t-2}+0.27Y_{t-3}+0.17Y_{t-4} \\ +0.26Y_{t-5}+0.07Y_{t-6}-0.07Y_{t-7}-0.01Y_{t-8}-0.16Y_{t-9} \\ +0.06Y_{t-10}+0.20Y_{t-11}-0.63Y_{t-12}+0.03Y_{t-13}+\epsilon_{2t} & \text{if } Y_{t-2}>0 \end{cases}$	2.88 2.76

5 Forecasting

Given the models in Table 3 forecasts for future values of series can be made. For the ARMA models we used the time series program PEST of Brockwell and Davis (1987). The program allows the options of transforming nonstationary data, taking logs, and differencing data. It calculates forecasts for future values with a user-prespecified ARMA model. Afterwards it transforms the forecasts back to their original scale. Unfortunately, as far as we know, there does not exist a software package which can compute multi-step ahead forecasts for SETAR models having large model orders. Tong's STAR package can handle orders up to lag 11. Moreover, it does not allow constraining redundant coefficients to zero. Hence, we developed our own software program which in broad lines works as follows. For a given SETAR$(2; p_1, p_2)$ model with $0 \le p_2 \le p_1$ one specifies the orders p_1 and p_2, the parameter values α_{1i} $(i = 0, 1, \ldots, p_1)$ and α_{2j} $(j = 0, 1, \ldots, p_2)$, the delay d, and the threshold value r. Now given the p_1 last transformed (stationary) observed time series values, the program simulates future values of the process for each step h using a large number, say N, of replications. In this paper we fixed N at 10,000. The innovations series $\{\epsilon_{1,T+h}\}$ and $\{\epsilon_{2,T+h}\}$ are replaced by numbers from a random number generator, with variances identical to the residual variances of the fitted SETAR model. Now a forecast $\hat{Y}_T(h)$ of Y_{T+h} is obtained by taking average of the N simulated values $Y_{T+h}(i)$ $(i = 1, 2, \ldots, N)$, i.e.

$$\hat{Y}_T(h) = \sum_{i=1}^{N} Y_{T+h}(i)/N \quad (h = 1, 2, \ldots). \tag{4}$$

It is assumed that computed value of (4) is a reasonable estimates of $E(Y_{T+h}|Y_t, t \le T, h = 1, 2, \ldots)$. In fact, Lai and Zhu (1991) proved that (4) is a strongly consistent estimate of $E(Y_{T+h}|Y_t, t \le T, h = 1, 2, \ldots)$ when the SETAR model (put in an arranged AR form) is estimated by least-squares. Next, in the final step, the forecasts are transformed back into the same scale as the original series. We denote these latter forecasts by $\hat{Z}_T(h)$.

Two criteria will be used to assess the relative out-of-sample forecasting performance of the fitted models, namely the mean square percentage error (MSPE) and the mean absolute percentage error (MAPE). They are defined as follows. Suppose that the models are fitted to data for time period $t = 1, 2, \ldots, T$. Hence, for each forecast horizon h $(h > 0)$ the forecast $\hat{Z}_T(h)$ can be computed. In addition the same model can be used to compute the forecasts $\hat{Z}_{T+k}(h)$, $(k = 0, 1, \ldots, K-1; h > 0)$, where K is a relatively small number as compared to T. This is reasonable under the assumption that the structure of the model doesn't change significantly when fitted to a longer period as opposed to fitting a model to a series having only T observations. In this way K different h-step ahead forecasts can be obtained. Now given these results the MSPE at forecast horizon h is defined as

$$MSPE(h) = \frac{1}{K}\sum_{k=0}^{K-1}\Big[\frac{Z_{T+h+k} - \hat{Z}_{T+k}(h)}{Z_{T+h+k}} \times 100\Big]^2, \quad (h = 1, 2, \ldots, H). \tag{5}$$

Similarly the MAPE at forecast horizon h is defined as

$$MAPE(h) = \frac{1}{K}\sum_{k=0}^{K-1}\Big|\frac{Z_{T+h+k} - \hat{Z}_{T+k}(h)}{Z_{T+h+k}} \times 100\Big|, \quad (h = 1, 2, \ldots, H). \tag{6}$$

The forecast results presented in the next section will be based on $H = 12$, i.e. the genuine out-of-sample period is twelve months, and on $K = 10$. Note that (5) was also used by Ray (1988) in his within-sample forecast experiment.

6 Forecast Comparison

Table 4 contains ratios of the MSPEs of the SETAR models relative to MSPEs of the ARMA models for forecast horizons $h = 1, 2, \ldots, 12$. Similarly, Table 5 provides ratios of the MAPEs of the SETAR models relative to the MAPEs of the ARMA models for $h = 1, 2, \ldots, 12$. For each series the first line reports results for the unmodified SETAR models (denoted by ii) in Table 3) and the ARMA models (denoted by i) in Table 3) while the second line shows results for the modified SETAR models (denoted by iii) in Table 3) and again the ARMA models. Note that for Series 7 only the MSPE and MAPE ratios are given for the unmodified SETAR model relative to the ARMA model. Hence in both tables the ARMA models are benchmarks for assessing the forecasting performance of the (un)modified SETAR models: If a reported value is larger than 1 then a respective SETAR model gives worse forecast results than an ARMA model.

It is evident from Tables 5 and 6 that overall there is hardly any gain in using a SETAR model relative to an ARMA model. However, looking at each series individually the following comments are in order:

- For Series 1 the ARMA model performs best for almost all forecast horizons as judged by the MSPE and MAPE ratios. This is not so surprising since no signs of nonlinearity were detected by the tests introduced in Section 3. Nevertheless it is interesting to see that the modified SETAR model gives better forecast results than the unmodified one. So removing insignificant parameters from the SETAR model seems to improve the forecast accuracy of the model.

Table 4: MSPE ratios of SETAR models relative to ARMA models.

h		Series						
		1	2	3	4	5	6	7
1	ii)/i	0.92	0.91	1.77	6.40	1.88	0.86	1.12
	iii)/i)	0.86	1.11	1.70	4.56	1.25	0.82	-
2	ii)/i)	1.21	0.81	1.38	4.64	1.78	0.87	0.77
	iii)/i)	1.11	0.99	1.34	2.34	1.23	0.88	-
3	ii)/i)	1.59	0.96	0.84	4.55	1.16	0.88	0.42
	iii)/i)	1.48	1.01	1.44	2.43	0.67	0.89	-
4	ii)/i)	1.55	0.90	1.30	6.46	1.07	0.87	0.49
	iii)/i)	1.50	0.90	1.61	3.19	0.72	0.87	-
5	ii)/i)	1.76	1.01	2.04	5.61	1.11	0.81	0.52
	iii)/i)	1.74	0.92	2.02	2.87	0.81	0.80	-
6	ii)/i)	1.88	1.11	3.49	6.81	1.35	0.71	0.46
	iii)/i)	1.82	0.94	3.42	3.93	0.87	0.68	-
7	ii)/i)	1.66	1.24	3.01	6.15	1.57	0.56	0.58
	iii)/i)	1.61	0.96	3.09	3.56	1.10	0.56	-
8	ii)/i)	1.45	1.47	2.21	5.89	1.74	0.38	0.81
	iii)/i)	1.40	1.04	2.66	3.67	1.19	0.41	-
9	ii)/i)	1.34	1.70	2.95	4.46	1.93	0.21	0.94
	iii)/i)	1.22	1.19	2.78	2.85	1.23	0.27	-
10	ii)/i)	1.34	1.81	2.41	4.20	2.23	0.12	1.25
	iii)/i)	1.16	1.29	2.55	2.86	1.43	0.23	-
11	ii)/i)	1.19	2.07	3.65	3.62	3.32	0.09	1.61
	iii)/i)	0.95	1.41	3.89	2.54	1.92	0.21	-
12	ii)/i)	0.98	2.26	6.48	3.07	7.13	0.09	1.61
	iii)/i)	0.74	1.57	8.00	2.30	3.66	0.20	-

Table 5: MAPE ratios of SETAR models relative to ARMA models.

h		Series						
		1	2	3	4	5	6	7
1	ii)/i)	0.98	0.94	1.46	2.82	1.51	0.97	1.13
	iii)/i)	0.94	0.99	1.30	2.53	1.23	0.96	-
2	ii)/i)	1.24	0.95	1.10	2.12	1.33	0.94	1.05
	iii)/i)	1.12	0.97	1.10	1.47	1.19	0.94	-
3	ii)/i)	1.20	0.94	0.87	2.05	0.93	1.08	0.64
	iii)/i)	1.15	1.00	1.31	1.52	0.81	1.02	-
4	ii)/i)	1.09	0.86	1.11	2.71	0.87	1.01	0.62
	iii)/i)	1.15	0.82	1.37	1.89	0.81	0.97	-
5	ii)/i)	1.26	0.88	1.36	2.47	0.81	0.98	0.64
	iii)/i)	1.31	0.86	1.55	1.70	0.83	0.94	-
6	ii)/i)	1.29	0.91	1.97	2.78	1.05	0.81	0.58
	iii)/i)	1.27	0.92	1.95	2.14	0.97	0.79	-
7	ii)/i)	1.23	1.00	1.73	2.52	1.24	0.71	0.67
	iii)/i)	1.20	0.95	1.80	1.91	1.12	0.72	-
8	ii)/i)	1.12	1.16	1.54	2.37	1.24	0.59	0.87
	iii)/i)	1.11	1.03	1.65	1.86	1.01	0.66	-
9	ii)/i)	1.03	1.32	2.18	2.31	1.40	0.51	0.97
	iii)/i)	0.97	1.12	2.00	1.86	1.06	0.58	-
10	ii)/i)	1.05	1.42	1.46	2.31	1.42	0.33	0.95
	iii)/i)	1.01	1.17	1.53	1.91	1.14	0.45	-
11	ii)/i)	0.97	1.53	1.96	2.05	1.92	0.27	1.11
	iii)/i)	0.86	1.24	1.99	1.76	1.39	0.41	-
12	ii)/i)	0.86	1.62	2.63	1.87	3.02	0.21	1.20
	iii)/i)	0.73	1.32	2.90	1.65	2.07	0.37	-

- For Series 2 the MSPE ratios at $h = 1, 2, \ldots, 5$ indicate that the SETAR model performs better than the ARMA model. Similar results can be seen for the MAPE ratios at $h = 1, 2, \ldots, 7$. This is curious since all linearity tests found no evidence of SETAR type nonlinearity. Also note that the unmodified SETAR model gives better forecast results than the modified SETAR model for short forecast horizons and both forecast criteria. For large values of h this conclusion is just the reverse. These findings seem to be in conflict with Series 1 where we concluded that a more parsimonious SETAR model is to be preferred.

- Clearly for Series 3 both SETAR models have a poor forecasting performance relative to the ARMA model as may be noted from the MSPE and MAPE ratios. Again this is quite surprising since all linearity tests in Section 3 suggested some type of nonlinearity for d=1.

- The New-F test suggest that Series 4 is nonlinear for delays d=1,2, and 3. No nonlinearity was detected by all tests for d=5 which is the value selected by Ray (1988). Clearly, as can be seen from the MSPE and MAPE ratios the forecasting results from both SETAR models with d=5 are terrible as compared to the ARMA model. We interpret this as an indication that the SETAR forecasting results are extremely sensitive to the optimal choice of d. Note that there is an improvement in forecasting accuracy for the modified SETAR model. So again removal of insignificant parameters is worthwhile.

- The forecast results for Series 5 show a somewhat mixed picture. In most cases the ARMA model performs better than both SETAR models. Only for h=3,4 and 5 the SETARs perform better according to the MAPE. The MSPE ratios indicate for these cases only a better forecasting performance for the modified SETAR model. Note that for this series nonlinearity was only detected by the LR-test for d=2 whereas the fitted SETAR models have a delay $d = 1$. This indicates that the result of the LR test depends very much on the threshold variable Y_{t-d}. Also note that again the modified SETAR model performs better than the unmodified one.

- For Series 6 the ARMA models are far worse than the SETAR models according to the MSPE and MAPE ratios. As h increases, the difference in forecast accuracy between ARMA and SETAR increases as well. These results are more decisive than those of the linearity tests. Also, for $h = 1, 2, \ldots, 8$ the difference between the two SETAR models reduces with the unmodified model forecasting slightly better than the modified one.

Table 6: Results of the BDS test statistic for the residuals of the time series models in Table 3; $\ell = 2, 3, \ldots, 5$ is the embedding dimension and e=(residual standard deviation)/range is a positive constant.

Series	Model		ℓ			
		e	2	3	4	5
1.	i)	0.13	0.64	1.53	1.26	0.98
	ii)	0.12	-2.25*	-0.82	-0.37	-0.34
	iii)	0.13	-1.25	-0.29	0.03	-0.15
2.	i)	0.15	-0.43	-0.12	-0.43	-0.38
	ii)	0.14	-1.24	-1.10	-0.71	-0.19
	iii)	0.14	0.58	0.26	0.39	0.66
3.	i)	0.16	1.49	2.20*	2.52*	2.10*
	ii)	0.17	1.02	0.78	1.09	1.06
	iii)	0.15	-0.38	-0.10	0.19	0.13
4.	i)	0.11	4.87*	4.09*	2.94*	2.55*
	ii)	0.10	0.32	0.30	0.31	-0.06
	iii)	0.11	0.29	0.12	0.06	-0.37
5.	i)	0.16	0.70	1.28	1.57	1.65
	ii)	0.17	0.93	0.69	0.76	0.85
	iii)	0.18	0.01	0.07	0.10	0.56
6.	i)	0.18	-1.15	-1.21	0.77	-0.28
	ii)	0.18	-0.64	-1.21	-1.30	-0.48
	iii)	0.16	1.00	0.23	0.16	0.60
7.	i)	0.18	3.22*	3.49*	3.88*	4.49*
	ii)	0.20	0.09	0.51	0.69	1.07

- Series 7 gives some mixed results. The unmodified SETAR model outperforms the ARMA model for forecast horizons $h = 2, 3, \ldots, 9$ if we look at the values of the MSPE ratios. For the MAPE ratios this is the case for horizons $h = 3, 4, \ldots, 10$. Some support for this result comes from the TAR-F and New-F test which both indicate some threshold-nonlinearity. On the other hand, the LR test suggest linearity for all values of d.

The residuals of the fitted models were tested for nonlinearity using the BDS test of Brock et al. (1987). Under the null hypothesis H_0 that a time series $\{X_t\}$ is i.i.d. the test will have an asymptotic normal distribution with mean zero and a variance that is a complicated function of ℓ, the so-called embedding dimension, and a distance parameter e. If H_0 is rejected, one can conclude that there is nonlinearity present, but its form is not determined. The results of the BDS test, standardized so that it has a standard normal asymptotic distribution under H_0, are reported in

Table 6. An asterisk indicates significance of the test at the 5% level. Significant nonlinearity is indicated for the ARMA residuals of Series 3, 4, and 7. These results confirm the test results for these series given in Table 2. Note that the BDS test fails to detect signs of nonlinearities in the SETAR residuals. This is also in agreement with the nonlinearity test results reported in Table 2.

7 Some Concluding Remarks

The two major findings of the paper are as follows. First, improved out-of-sample forecast performance cannot be achieved by using the SETAR models, as specified by Ray (1988), for five of the seven Indian macroeconomic time series considered in this study. Thus, there is support for Ray's (1988) claim that forecasts from his fitted ARMA models dominate forecasts from his fitted SETAR models. However, the relative forecasting performance of the SETAR models is sensitive to the correct choice of the delay d and the threshold value r. Moreover, it is sensitive to whether or not the series are transformed to stationarity by taking logs. Without such a transformation, the SETAR models fitted to Series 6 and 7 produce much better forecasts than the fitted linear ARMA models. Thus it is very important to appropriately transform and/or difference a series before specifying SETAR models. This is apparently because these transformations can swamp nonlinear features in the data.

The second main result of the paper is that the out-of-sample forecast ability of the SETAR models fitted by Ray (1988) can be significantly improved by deleting insignificant parameters. Thus it is worthwhile to fit subset SETAR models rather than fitting full models to the data. Also in view of the structural properties of the fitted seasonal ARMA models, subset SETAR models are much easier to interpret.

In closing this section we like to stress that we accepted the time series models fitted by Ray (1988) at face-value. No attempts were made to re-analyze the data starting from scratch. Hence, given the relative limited state-of-the art of nonlinear times series analysis in 1988 quite a few of uncertainties, apart from the ones already mentioned above, could have influenced the specification of the unmodified SETAR models. This is an area for further research, and we are happy to provide Ray's data to interesting parties for this purpose.

References

Brock, W.A., Dechert, W.D. and Scheinkman, J.-A. (1987), "A test for independence based on the correlation dimension", SSRI Working Paper No. 8702 (Dept. of Economics, University of Wisconsin).

Brockwell, P.J. and Davis, R.A. (1987), *Time Series: Theory and Methods*, New York: Springer-Verlag.

Chan, K.S. and Tong, H. (1990), "On likelihood ratio tests for threshold autoregression", *Journal of the Royal Statistical Society B*, 52, 469-476.

Chan, K.S. (1991), "Percentage points of likelihood ratio tests for threshold autoregression", *Journal of the Royal Statistical Society B*, 53, 691-696.

Chappell, D., Padmore, J., Mistry, P. and Ellis, C. (1996), "A threshold model for the French Franc/Deutschmark exchange rate", *Journal of Forecasting*, 15, 155-164.

De Gooijer, J.G. and Kumar, K. (1992), "Some recent developments in non-linear time series modelling, testing and forecasting", *International Journal of Forecasting*, 8, 135-156.

Kräger, H. and Kugler, P. (1993), "Nonlinearities in foreign exchange markets: a different perspective", *Journal of International Money and Finance*, 12, 195-208.

Lai, T.L. and Zhu, G. (1991), "Adaptive prediction in non-linear autoregressive models and control systems", *Statistica Sinica*, 1, 309-334.

Petruccelli, J.D. and Davies, N. (1986), "A portmanteau test for self-exciting threshold autoregressive-type nonlinearity in time series", *Biometrika*, 73, 687-694.

Ray, D. (1988), "Comparison of forecasts: an empirical investigation", *Sankhyā*, 50B, 258-277.

Subba Rao, T. and Gabr M.M. (1980), "A test for linearity of stationary time series", *Journal of Time Series Analysis*, 1, 145-57.

Tsay, R.S. (1989), "Testing and modeling threshold autoregressive processes", *Journal of the American Statistical Association*, 84, 231-240.

Tsay, R.S. (1991), 'Detecting and modeling nonlinearity in univariate time series analysis", *Statistica Sinica*, 1, 431-451.

Tong, H. (1990), *Non-Linear Time Series: A Dynamical System Approach*, Oxford: Oxford University Press.

Figure 1: Plots of Series 1 - 7 denoted by respectively a) - g).