A Monte Carlo study of the forecasting performance of empirical SETAR models

Michael P. Clements and Jeremy Smith*
Department of Economics,
University of Warwick,
Coventry CV4 7AL.
Email: M.P.Clements@Warwick.ac.uk or Jeremy.Smith@Warwick.ac.uk
Tel: 01203 523055
FAX: 01203 523032

December 6, 1997

Abstract

In this paper we investigate the multi-period forecast performance of a number of empirical self-exciting threshold autoregressive (SETAR) models that have been proposed in the literature for modelling exchange rates and GNP, amongst other variables. We take each of the empirical SETAR models in turn as the DGP to ensure that the 'non-linearity' characterises the future, and compare the forecast performance of SETAR and linear autoregressive models on a number of quantitative and qualitative criteria. Our results indicate that non-linear models have an edge in certain states of nature but not in others, and that this can be highlighted by evaluating forecasts conditional upon the regime.

*The first author acknowledges financial support under ESRC grant L116251015. The paper has benefitted from the comments of participants at the Forecasting Workshop, Nuffield College, June 1997, and the Econometric Society European Meeting, Toulouse, August 1997. We are especially grateful to Hashem Pesaran and three anonymous referees of this journal for helpful comments and suggestions.

1 Introduction

In recent years there has been considerable interest in testing for and modelling non-linearities in economic time series. Some of this activity has been based on allowing for non-linearities in traditional econometric equations variously described as ‘structural’ or ‘behavioural’, but much of it follows in the time-series tradition of Box and Jenkins (1970). The usefulness of linear time-series models is usually gauged by their predictive ability, and such models have sometimes been used as a benchmark for econometric models in forecast comparisons. However, in a recent review of non-linear time series models, De Gooijer and Kumar (1992) report that there is no clear evidence in favour of non-linear over linear models in terms of forecast performance, notwithstanding the ability of the former to capture asymmetries in important macro-aggregates over the business cycle (see, e.g., Hamilton, 1989, Tiao and Tsay, 1994 and Potter, 1995 for US GNP, Montgomery, Zarnowitz, Tsay and Tiao, 1997 for US unemployment, and Acemoglu and Scott, 1994 for UK labour market variables).

In this paper we investigate the forecast performance of a non-linear time series model that has been widely used in the literature to explain various empirical phenomena – the self-exciting threshold autoregressive (SETAR) model. We consider three studies which propose SETAR models of the foreign exchange market: Kräger and Kugler (1993) who model five currencies against the US dollar on weekly data over the last ten years, Peel and Speight (1994) who model three weekly sterling spot market rates over the inter-war period, and Chappell, Padmore, Mistry and Ellis (1996) who model the French franc to Deutschemark rate in the 1990’s on daily data. Rather than analysing all these models, a representative sample are selected. Kräger and Kugler (1993) do not evaluate the forecasting performance of these models, but, as is common in this literature, carry out in-sample residual-based tests, such as the Brock-Dechert-Scheinkman (BDS) test of non-linearity, tests of stability, and also comparisons against GARCH models. Peel and Speight (1994) report empirical mean-square forecast errors (MSFEs) but only for 1-step ahead forecasts. We also consider the models of US GNP of Tiao and Tsay (1994) and Potter (1995). These are qualitatively very similar. Potter (1995) presents non-linear impulse response functions but does not consider the forecast performance of the model. Tiao and Tsay (1994) calculate empirical MSFEs and show that the SETAR model performs favourably compared to a linear model particularly when the forecast origin happens to be in a recession, which is the regime with the minority of the data points. Other economic time series we include in the study are the UK savings ratio and GDP growth, analysed using threshold models by Peel and Speight (1995). Finally, for the purpose of comparison we briefly consider the Canadian lynx data and Wolf’s sunspot numbers (see, for example Tong, 1995a, chapter 7), which can be successfully forecast using non-linear time series models.

While some of the studies provide information on the forecast performance of the estimated models others neglect this aspect. The purpose of this paper is to fill out the relatively limited evidence that exists on the multi-step forecast performance of such models, and to see if we can go beyond the conclusion in De Gooijer and Kumar, 1992, p.151 that ‘no uniformity seems to exist in the evidence presented on the forecasting ability of non-linear models’ by isolating those features that may contribute to improved accuracy. What comes out strongly in our results is the importance of where the process is at the time the forecast is made, paralleling the importance of the history of the process in impulse response analysis of non-linear models: see Koop, Pesaran and Potter (1996).

The evidence on the forecast performance of non-linear models in the studies of economic and financial variables referred to above is based on empirical MSFEs. Without large data samples it is therefore difficult to calculate measures of forecast accuracy for multi-period forecasts with long horizons with much precision. We get around this by using a Monte Carlo study rather than an empirical study, with the data generating process (DGP) taken to be each of the estimated empirical models in turn. In many ways this casts the non-linear model in its best possible light. For example, the lack of forecast gain of non-linear models over linear models is often explained in terms of a failure of the 'non-linearity' to persist into the future (e.g., Granger and Teräsvirta, 1993, p.164) but by using the estimated SETAR model to generate data over the 'future', the future realizations of the process have the same non-linear imprint as the past as manifest in the estimated model. We are also in a position to better explore other aspects of forecasting with non-linear models, such as the dependence on the regime at the forecast origin. Finally, we can assess the impact on forecast performance of parameter estimation and model uncertainty.

The plan of the paper is as follows. In section 2 we briefly describe the SETAR model and the calculation of multi-period forecasts. Section 3 describes the Monte Carlo that we use to assess the relative forecast performance of the SETAR models and linear alternatives. Section 4 reviews some of the reasons why it may not be possible to exploit apparent non-linearities in the data to generate more accurate forecasts. In section 5 we propose as an indicator of the forecast performance of SETAR models a measure of 'regime persistence'. When there is no regime persistence, so that regimes are serially independently distributed, a linear model should forecast as well as the SETAR. Section 6 discusses 'qualitative measures' of forecast accuracy, and it is argued that these may provide additional information to the more traditional quantitative measures, such as MSFE, when the comparisons include non-linear models.

Section 7 discusses the results of the forecast comparisons of the non-linear to linear models and we assess the usefulness of the measure of regime persistence. Section 7.1 surveys some recent contributions to modelling and forecasting exchange rates using non-linear models, before discussing the Monte Carlo forecast comparison of the SETAR models of Kräger and Kugler (1993), Peel and Speight (1994) and Chappell *et al.* (1996) to linear models. The results for US GNP are reported in section 7.2, for the UK savings ratio and GDP in section 7.3, and for the Lynx data and Sunspot numbers in section 7.4. Section 7.5 collects together the results concerning the outcomes of the tests of serially independently distributed regimes and the relative forecast performance of the SETAR and linear models. Section 8 concludes.

2 SETAR models and multi-period forecasts

The threshold autoregressive (TAR) model first proposed by Tong (1978), Tong and Lim (1980) and Tong (1983) (see also Tong, 1995a) assumes that a variable y_t is a linear autoregression within a regime but may move between regimes depending on the value taken by the threshold variable. When the threshold variable is a lag of y_t, say, y_{t-d}, so that d is the length of the delay, then the model is 'self-exciting', giving rise to the acronym SETAR. When there are two regimes, then the process is in regime $i = 1$ at period t when $y_{t-d} \leq r$, and otherwise ($y_{t-d} > r$) in regime $i = 2$:

$$y_t = \phi_0^{\{i\}} + \phi_1^{\{i\}} y_{t-1} + \ldots + \phi_p^{\{i\}} y_{t-p} + \epsilon_t^{\{i\}}, \quad \epsilon_t^{\{i\}} \sim \mathsf{iid}\left(0, \sigma^{2\{i\}}\right), \quad i = 1, 2 \tag{1}$$

where the parameters super-scripted by $\{i\}$ may vary across regime. The orders of the autoregressions may differ across regimes (so that p is the maximum lag order and some of the $\phi_p^{\{i\}}$ may be zero for some i). Stationarity and ergodicity conditions are discussed in, e.g., Tong (1995a).

The SETAR model is a special case of the 'endogenous selection' Markov Switching (MS) model of e.g., Durland and McCurdy (1994). In the general model, the thresholds depend on the regime. Potter (1995) shows that the MS model of Hamilton (1989) and the mixture of distributions model are also special cases: the former arises when the probability of switching regimes does not depend on the realized values of the process, and the latter when the regimes are serially independently distributed.

One of the reasons for the equivocal conclusion in De Gooijer and Kumar (1992) concerning the usefulness of non-linear models for forecasting is that obtaining multi-period forecasts is more difficult than for linear models, because exact analytical solutions are generally not available. For example, suppose $y_t = g(y_{t-1}) + \epsilon_t$, where $g(\cdot)$ is a non-linear function. For a 2-regime SETAR model, for example, we might have:

$$g(y_{t-1}) = \left[\phi^{\{1\}} + 1(y_{t-1} > r)\left(\phi^{\{2\}} - \phi^{\{1\}}\right)\right] y_{t-1}$$

where $1(\cdot)$ is the indicator function – equal to unity if the argument is true and zero if false. ϵ_t is iid with mean zero and distribution function D_ε. The exact 1-step ahead forecast defined by $\hat{y}_{t,1} \equiv \mathsf{E}[y_{t+1} \mid \mathcal{I}_t]$, where $\mathcal{I}_t = y_t, y_{t-1}, \ldots \equiv Y^t_{-\infty}$, is given by:

$$\hat{y}_{t,1} = \mathsf{E}\left[\left(g\left(y_t\right) + \epsilon_{t+1}\right) \mid \mathcal{I}_t\right] = g\left(y_t\right).$$

However, for 2-steps ahead:

$$\hat{y}_{t,2} \equiv \mathsf{E}[y_{t+2} \mid \mathcal{I}_t] = \mathsf{E}\left[\left(g\left(y_{t+1}\right) + \epsilon_{t+2}\right) \mid \mathcal{I}_t\right] = \mathsf{E}\left[g\left(y_{t+1}\right) \mid \mathcal{I}_t\right] \tag{2}$$

But, $\mathsf{E}[g(y_{t+1})] \neq g(\mathsf{E}[y_{t+1}]) = g(\hat{y}_{t+1})$ when $g(\cdot)$ is non-linear.

Exact numerical solutions require computer-intensive sequences of numerical integrations (see, e.g., Tong, 1995a sections 4.2.4 and 6.2) based on the Chapman-Kolmogorov relation. As an alternative, one might use a Monte Carlo method (such as that implemented in the STAR3 program of Tong: see Tong, 1995a, and used by, e.g., Tiao and Tsay, 1994 and Clements and Smith, 1997)[1] particularly for high-order autoregressions. Another possibility is the Normal Forecast Error (NFE) method proposed by Al-Qassam and Lane (1989) for the exponential-autoregressive model, and adapted by De Gooijer and De Bruin (1997) to forecasting SETAR models. Clements and Smith (1997) compare a number of methods of obtaining multi-period forecasts from SETAR models and conclude that the Monte Carlo method performs reasonably well, and is the method we use in this paper.

3 Design of the Monte Carlo

For each of the empirical models mentioned in section 1 we explore a number of aspects concerning the forecast performance of a SETAR model versus a linear alternative via Monte Carlo. In the first instance, we assume that the SETAR model is the SETAR DGP, and compare the forecasts from the SETAR model calculated by the Monte Carlo method to those of a linear model. We refer to this as the 'Known Model' case. The comparison is based on quantitative and qualitative measures of forecast accuracy: see section 6. For each of $N_j = 1000$ replications a single realization $\{y\}_1^{T+H}$ is generated from the SETAR DGP by replacing the disturbances by normal random variates. The linear model is estimated on $\{y\}_1^T$ and used to forecast the observations $\{y\}_{T+1}^{T+H}$, and the resulting linear model forecast errors are then stored. The SETAR model forecasts are obtained by averaging over an additional $N_f = 500$ realizations of $\{y\}_{T+1}^{T+H}$ for each of the N_j replications. These realizations are generated from drawings of the errors from the normal distribution with appropriate regime-specific error variances.

We present results which condition on the regime, whereby for conditioning upon the regime r_i we discard drawings of $\{y\}$ for which the condition $r_{i-1} \leq y_{T+1-d} < r_i$ fails. Thus for a delay parameter

[1] The dynamic simulation method used by Pesaran and Potter (1997) to calculate 2-step forecasts from their non-linear 'floor and ceiling' model is a Monte Carlo method. Gallant, Rossi and Tauchen (1993) and Koop *et al.* (1996) provide analyses of the construction of conditional densities for non-linear time series models, in the context of impulse response analysis, and emphasise the importance of allowing for, and integrating out, non-zero future realizations of the disturbances (ϵ_{t+1} in equation (2)

of $d = 1$ we require that $r_{i-1} \leq y_T < r_i$ so that the 1-step ahead forecast is generated from regime r_i. We also present unconditional results, where the number of times y_{T+1-d} falls in regime i will be approximately equal to the unconditional probability of the process being in regime r_i. This parallels the practice of reporting empirical MSFEs for specific regimes versus for all regimes together, as in Tiao and Tsay (1994), for example. Tong (1995b) p. 409 – 410 argues strongly that for non-linear models 'how well we can predict depends on where we are' and that there are 'windows of opportunity for substantial reduction in prediction errors' (p.409). An important aspect of our evaluation of the forecasts from the non-linear SETAR models relative to the linear AR models is to make the comparison in a way which highlights the favourable performance of the former for certain states.

So far we have described a situation in which the SETAR model is assumed to be the DGP: the number of regimes and threshold values, the delay lag, the orders of the process in each regime and the model coefficients, are all assumed known. This abstracts from sources of forecast uncertainty emanating from parameter estimation and model selection. We continue to condition upon the number of regimes (N_r) being known. If the threshold values and the delay, d, are known, the sample can simply be split into 2 (for $N_r = 2$) and an OLS regression run on the observations belonging to each regime separately, or indicator functions can be used in a single regression, constraining the residual error variance to be constant across regimes (see, for example, Potter, 1995, p.113.) However, we wish to allow for the situation in which r and d are unknown, so the model is estimated by searching over all admissable values of these parameters, where r is allowed to take on each of the sample period values of y_{t-d} in turn[2], and d typically takes on the values $0, 1, 2, \ldots$ up to the maximum lag length allowed. For a known lag order, the selected model is that for which the pair (r, d) minimize the overall residual sum of squares.

We also allow the lag orders in the regimes to be unknown. This requires a search over all lag orders (not constrained to be the same across regimes) less than some maximum, based on minimizing AIC (see, Akaike, 1973). In section 7 the term 'Unknown Model' is used as a shorthand for the case when the SETAR model parameters, including r and d, and the lag orders, are unknown and are estimated as described above.

4 Factors inhibiting non-linear model first moment prediction

A number of suggestions have been made as to why apparent non-linearities can not be exploited in forecasting. In this section we review a number of the arguments, including those put forward in the context of forecasting exchange rates.

In reviewing Smooth Transition Autoregressive Regression (STAR)[3] models of US industrial production, Granger and Teräsvirta (1993), chapter 9 (see also Teräsvirta and Anderson, 1992) argue that the superior in-sample performance of such models will only be matched out-of-sample if that period contains 'non-linear features'. As discussed in section 1 the approach that we adopt does not allow the non-linear models to perform poorly for this reason: in the Monte Carlo evaluation of the forecast performance of the SETAR models the future is simulated to mimic the past and reproduces the 'non-linear features' that characterised the past. Thus, even if the non-linear structure captured in the empirical model was primarily due to 'outliers' and therefore of a variety not conducive to improved empirical forecasts, by taking the model to be the DGP we ensure it is a feature of the period to be forecast.

---

[2] In practice the range of values of y_{t-d} is restricted to those between the 15^{th} and 85^{th} percentile of the empirical distribution, following Andrews (1993) and Hansen (1996).

[3] The STAR model is a general class of 'smooth' regime-switching models. The TAR model is a special case in which the movement between regimes is discontinuous. STAR models were first formulated by Chan and Tong (1986) as smooth 'threshold' autoregressive models

Secondly, it has been shown that forecast performance may depend on the regime at the forecast origin. As explained in section 3 we assess the connection between forecast performance and the regime the forecast origin falls in.

With respect to exchange rate prediction, Diebold and Nason (1990) give four reasons why non-linear models may fail to forecast better than the simplest linear model even when linearity is routinely rejected statistically. Their suggestions may be relevant for variables other than exchange rates. The first is that there is linear dependence in exchange rates of the form that large (small) changes tend to be followed by large (small) changes of either sign. Thus there are non-linearities in even-ordered conditional moments, which explains the success of ARCH (Engle, 1982) and GARCH (Bollerslev, 1986) type models of exchange rates, but can not be used for improved point (as opposed to interval) prediction. Secondly, the apparent non-linearities detected by tests for linearity are due to outliers or structural breaks, but these offer no gain in improved out-of-sample performance. Third, conditional-mean non-linearities are a feature of the DGP, but are not large enough to yield much of an improvement to forecasting, and finally, they are present and important but the wrong types of non-linear models have been used to try and capture them. Our approach of simulating the estimated empirical models suggests that the first, second and fourth explanations could not account for any inability of the non-linear models to outperform linear models. This leaves the third reason.

In the next section we discuss one approach to assessing when a SETAR model might be expected to yield useful forecasts.

5 A measure of regime persistence.

The hallmark of the SETAR model is that the movement between regimes is internally generated (subject to the realization of a random disturbance term) and has a cyclical structure. Tong (1983) p. 109 gives an example of a SETAR process for which the cyclical movement between regimes is absolutely regular, in that the process alternates between the two regimes. He then goes on to show how this characteristic can be exploited to simplify multi-step ahead forecasting. In Tong's example the movement between regimes is perfectly predictable. The polar case is that the regimes are serially independently distributed. This might occur if the empirical SETAR model is essentially capturing non-linearities of a non-SETAR type. In that case, we would not expect a markedly better forecast performance relative to a linear model (see section 4).

In this section we discuss the calculation of a measure of regime persistence that can be applied to the empirical SETAR models in a Monte Carlo setting, and in section 7 we consider how useful this measure is as a predictor of the relative forecast gains of the SETAR model over a linear model.

The idea is to see whether or not the degree of regime-persistence in the data generated by the SETAR model is consistent with what would result from using a linear AR model. The approach we adopt is the following. For each replication j of the Monte Carlo, we record the proportion of times the process remained in a particular regime i for R consecutive periods, $p^j_{i,R}$. Thus, if $\{y^j\}^T_1$ denotes the artificial data vector of length T simulated from the empirical SETAR model on the j^{th} replication, then:

$$p^j_{i,R} = \frac{1}{T-R+1}\sum_{s=R}^{T}\left[1\left(y^j_s \in i\right) \times \cdots \times 1\left(y^j_{s-R+1} \in i\right)\right].$$

for $i = 1, \ldots, N_r$, $R = 1, 2, 3, \ldots$, and where e.g., $1(y^j_s \in i) = 1$ when y^j_s is in regime i. When $R = 1$ we simply have the proportion of observations in regime i. We compare these proportions with what would result from using a linear model, where the linear model is obtained by weighting the linear autoregressions of the SETAR model by the relevant $p^j_{i,1}$. Thus, for illustrative purposes consider the simple SETAR$(2;1,1)$:

$$\begin{aligned} y_t &= \phi^{\{1\}} y_{t-1} + \epsilon_t, \quad \text{when } y_{t-1} \leq r \\ y_t &= \phi^{\{2\}} y_{t-1} + \epsilon_t, \quad \text{when } y_{t-1} > r \end{aligned} \tag{3}$$

then on iteration j we consider an AR model with slope parameter given by $p^j_{1,1} \times \phi^{\{1\}} + (1 - p^j_{2,1}) \times \phi^{\{2\}}$. This AR model is simulated N_l times. For each of the N_l simulated data vectors, we calculate the proportion of times R consecutive observations fall in regime i, and denote this by $\hat{p}^{j,l}_{i,R}$. We then rank the $l = 1, \ldots, N_l$ values $\hat{p}^{j,l}_{i,R}$ by size in the vector $\hat{\mathbf{p}}^j_{i,R}$,[4] and for iteration j we conclude that the degree of serial dependence in the $\{y^j\}^T_1$ is not compatible with the data being generated by a linear model if $p^j_{i,R}$ is greater than the $(0.025 \times N_l)^{th}$ largest element in $\hat{\mathbf{p}}^j_{i,R}$, or is smaller than the $(0.025 \times N_l)^{th}$ smallest element. This procedure is then repeated for $j = 1, \ldots, N_j$, and we calculate the proportion of times the serial dependence is not consistent with a linear model for each i and R.

By construction, the 'test' is correctly-sized, and will reject 5% of the time subject to sampling variability. In table 1 we record for illustrative purposes the rejection frequencies for the simple SETAR model given by (3) for a number of parameter values and two sample sizes, $T = 100, 200$. The test is applied exactly as described above in a Monte Carlo setting. In an empirical setting where the SETAR model is unknown the test is not directly applicable, but a suitably modified version may be useful, but we do not pursue that here. Table 1 indicates that for the DGP given by (3) the tests based on the individual regimes are much more powerful than the test based on how often the process remains in either regime for two consecutive periods. As expected, when $\phi^{\{1\}} = \phi^{\{2\}}$, so that the DGP is linear, the rejection frequency is (barring sampling uncertainty) 5%.

Table 1 Rejection frequencies of whether the regime dependence can be explained by a linear model.

$\phi^{\{1\}}$	$\phi^{\{2\}}$	Regime ≤ 0	Regime > 0	Regime Either	Regime ≤ 0	Regime > 0	Regime Either
			$T = 100$			$T = 200$	
0	0	0.052	0.047	0.043	0.064	0.051	0.056
0	0.3	0.191	0.130	0.053	0.266	0.218	0.051
0	0.6	0.542	0.441	0.051	0.822	0.756	0.069
0	0.9	0.925	0.881	0.087	0.998	0.999	0.155
0.3	0.3	0.058	0.048	0.046	0.048	0.049	0.051
0.3	0.6	0.229	0.157	0.050	0.331	0.281	0.042
0.3	0.9	0.667	0.584	0.065	0.954	0.938	0.101
0.6	0.6	0.047	0.057	0.045	0.039	0.041	0.057
0.6	0.9	0.263	0.216	0.053	0.498	0.470	0.071
0.9	0.9	0.050	0.049	0.045	0.060	0.060	0.047

Notes on table. The data is generated by (3). The calculations reported are for $R = 2$ based on $N_j = 1000$ and $N_l = 500$.

[4] Rather than calculate $\hat{\mathbf{p}}^j_{i,R}$ for each j, we in fact calculate this vector once only, where the weights used to form the AR model are the average proportion of times the simulated data is in each regime, where the averaging is over the N_j replications. The loss in precision is small, and we save a great deal of computer time.

6 Forecast accuracy comparison of linear versus non-linear models

Most forecasts of macroeconomic variables are quantitative in nature, and quantitative measures of forecast accuracy based on the distance between the forecast and realization (i.e., the magnitude of the forecast error) have dominated the forecast evaluation literature. However, regime-switching models may be better suited to predicting movements between regimes rather than small movements within a regime. An evaluation criterion based on how often the direction of change of a variable is correctly predicted is one way of capturing this idea.

Direction-of-change tests were originally developed in the context of predicting rates of return on market investments by Henriksson and Merton (1981). Schnader and Stekler (1990) and Stekler (1994) applied the approach to macroeconomic prediction, and Pesaran and Timmermann (1992) suggested a number of refinements and extensions. Such tests are closely related to the standard χ^2 test of independence between actual and predicted directions of change based on the 2×2 contingency table. Where applicable, we report the p-value of the Pesaran and Timmermann (1992) test, the null of which is that the actual and predicted directions of change (in our case, the regimes) are independent. Relatedly, we also report a number of measures (rather than tests) of how well the model forecasts the direction of change/regime. One is the conditional probability of correctly predicting the regime, defined for regime i and step-ahead h as the proportion of times the model predicts (h-steps ahead) the process is in regime i conditional on the process being in regime i. Thus, we record the proportion of the N_j replications of the Monte Carlo for which the model correctly predicted the regime and divide this by the proportion of replications for which the process was in that regime. These are referred to as CRPs – conditional regime predictions. A value of unity indicates that the regime is always correctly predicted (but note some caution is required – such an outcome would arise if the model always predicted that regime regardless of which regime the process is actually in). We report CRPs for the AR and SETAR models when the process is not conditioned upon being in a particular regime. For forecast evaluation conditional upon the regime, CRPs would appear to be less informative – for example, consider calculating the CRP for being in regime i 1-step ahead conditional on the process being in regime i at the time the forecast is made. A single summary statistic (for each horizon) is given by the 'Non-Confusion' Rate (NCR), which is the number of times the regimes are correctly predicted divided by the total number of predictions. For the two regime case the NCR is bounded between plus and minus one.

An obvious limitation to the use of such criteria is that a forecast of very small increase, when a small decline occurred, will be counted one-for-one with a forecast of a large increase when a large decline occurred. Moreover, as noted by Schnader and Stekler (1990) and Stekler (1994), we may wish to evaluate forecasts relative to some baseline other than zero growth, e.g., in terms of how well the model predicts 'high' growth (say, growth above 2%) relative to low growth and declines ($\leq 2\%$). In both the 'Known Model' and 'Unknown Model' cases, we report CRPs where the regimes are as given by the empirical SETAR model DGPs.

We also calculate the more traditional squared error loss measures, such as the mean squared forecast error (MSFE), for each horizon, using the variability over the replications of the Monte Carlo to obtain our estimates. We use a test proposed by Diebold and Mariano (1995) to see if the differences in MSFEs between models are statistically significant. The test is implemented using a uniform lag window to estimate the variance of the sample mean of the loss differential series and assumes that the h-step forecasts exhibit $h-1$ dependence (see Diebold and Mariano, 1995 for details). Harvey, Leybourne and Newbold (1997) propose some modifications to the test statistic that correct for the tendency of the statistic to be over-sized, but we do not use them here given the large number of forecasts we have.

7 Monte Carlo studies of SETAR versus linear model prediction

7.1 SETAR forecasts of exchange rates

The literature on conditional mean exchange rate prediction over the post-war period suggests it is difficult to better a random walk. Diebold and Nason (1990) estimate non-parametrically the conditional expectation (or regression function) for non-parametric prediction to guard against the failure to benefit from non-linearities due to the incorrect choice of functional form. Using a nearest-neighbor (NN) technique of locally-weighted regression (LWR) (see, e.g., Cleveland, Devlin and Grosse, 1988) they find no improvement over a simple random walk for predicting 10 major dollar exchange rates over the post 1973 period. Meese and Rose (1991) allow for non-linear extensions to a number of structural exchange rate models using parametric and non-parametric models but with no significant improvement in forecast accuracy.

We apply the NN technique as in Diebold and Nason (1990), using a Euclidean distance measure and a tricube weighting function. The number of 'nearest neighbours' is set equal to the sample size, and the regression surface is re-estimated for each step ahead we forecast. We do not consider alternative values for these parameter or ways of implementing the method, since the NN technique is used only as a rough check of whether the empirical SETAR models fashion the data with sufficiently marked non-linear features that non-parametric forecasts are superior to linear ones. Both for the financial, and the economic, time series we find the NN forecasts are not significantly better than the simple linear AR model forecasts.

7.1.1 Kräger and Kugler (1993)

The exchange rate models estimated by Kräger and Kugler (1993) for the French franc, the Italian Lira, the Japanese Yen and the Swiss franc against the US dollar all follow a similar pattern. There are 3 regimes ($N_r = 3$), the delay is one period ($d = 1$), the middle regime is a third-order AR ($p = 3$) in the difference of the log of the exchange rate, and the first and third set the growth rate equal to a constant ($p = 0$). The estimated standard deviations of the first and third regimes exceed that of the middle regime, which is explained by central bank interventions in response to large appreciations (regime 1) and depreciations (regime 3). The model for the German mark differs in that the first regime is an AR with $p = 2$. The reader is referred to Table 2 of Kräger and Kugler (1993) for the details. The theoretical rationale is that threshold models of this sort approximate the solution to a rational expectations monetary model with stochastic intervention rules (see Hsieh, 1989) that may characterise the managed floating of the 1980's. The suggestion is that the authorities react to large appreciations and depreciations (rates of change) whereas for the target zone approach to managed floating the level of the exchange rate (or rather, its proximity to ceilings or floors) is relevant for signalling interventions.

Table 2 summarises the results of our Monte Carlo evaluation of the multi-step forecast performance of the SETAR exchange rate models of Kräger and Kugler (1993) of the Italian Lira and Japanese Yen. The linear competitor is an AR(0) model for the differences (of the logs) – i.e., a random walk (including a constant term). Higher-order linear models were generally dominated by the random walk (RW). The tables all follow a similar format. Panel [A] reports, for the 'Known SETAR Model' case, unconditional (on the regime) MSFEs for the RW model divided by those for SETAR for 1, 2 and 5-steps ahead. For longer horizons the ratio is approximately unity. Panel [B] reports the ratio of the MSFEs, again for the 'Known SETAR Model' case, but this time conditioning the forecast origin on each of the three regimes: Lower, Middle and Upper. The p-values are of the Diebold and Mariano (1995) test of equal forecast accuracy (as measured by MSFE), discussed in section 6, and are the probabilities under the null (of equal accuracy) of obtaining lower test statistics than we record. The CRPs in panel [A] record the proportion of times the regimes were correctly predicted when we do not condition on the regime: see section 6. The NCRs in panel [A] are the non-confusion rates. Finally, panels [C] and [D] repeat the information in panels [A] and [B] for the case when the SETAR model is not known. The CRPs and NCRs for the AR model are the same by construction in panels [A] and [C] and hence are reported only once.

Italian Lira. Relative to a RW there is a gain of just over 3% at 1-step ahead unconditionally when the SETAR model is known. Conditional on being in the Middle regime, there is a 1-step gain of 16%, and at 2-steps ahead the SETAR forecast are also significantly more accurate statistically. The 1-step gain remains, conditional on being in the Middle regime, when the SETAR model has to be estimated, and is of the order of 10%. At 2-steps ahead there is a much smaller, though still significant gain. The CRPs reflect this finding – the AR model never correctly predicts the Middle or Upper regimes, but the SETAR model correctly predicts the Middle regime 12% of the time 1-step ahead (model known, and 10% of the time when the model is estimated).

Japanese Yen. There is a gain of over 40%, conditional on being in the Middle regime, when the SETAR model is known, and of 15% when it is estimated. However, even conditionally there is nothing to choose between the models at further steps ahead, on the basis of MSFE. The CRPs indicate that the SETAR is better at predicting the Middle regime even at 5 steps ahead, getting it right nearly 40% of the time. The NCRs indicate that the SETAR is less confused at 1-step ahead. For the Japanese Yen the qualitative and quantitative measures tell a similar story.

7.1.2 Peel and Speight (1994)

The inter-war exchange rate models of Peel and Speight (1994) are for three weekly sterling spot rates. They estimate the US dollar and French franc models on 217 observations and that for the reichsmark on 139 observations. The estimated models are given in Peel and Speight (1994), Table 3, p.407. In each case the data are transformed by taking differences of the logs of the original series. We report results for the US dollar rate (table 3), which is modelled as a SETAR($3; 0, 2, 0$), with $d = 3$. The thresholds are such that a depreciation of the dollar three weeks ago exceeding 0.40% results in an expected appreciation of 0.15%, and at the other extreme an appreciation of over 0.41% gives an appreciation of 0.086%. We consider both a RW and an AR(3) model, as the latter appears to have some predictive ability. We find significant gains up to 3-steps ahead unconditionally, when the SETAR model is known, mainly due to the performance in the Middle Regime, but not when the SETAR is estimated. In tune with the MSFE gains in the ‘Known Model’ case, the SETAR model fares better at predicting the Lower and Upper regimes.

7.1.3 Chappell *et al.* (1996)

Chappell *et al.* (1996) fitted one and two threshold models to various ERM cross rates for daily data (1^{st} May 1990 to 20^{th} January 1992 for fitting, 21^{st} January 1992 to 30^{th} March 1992 for forecasting) but only found a superior forecast performance for the French franc/Deutschmark rate. They expect the operation of the ERM to set ceilings and floors on exchange rates, so that while the random walk model is appropriate within the prescribed bands it may not be as the exchange rate approaches either extreme. We focus on their preferred two-regime model for the FFr/DM.

By way of contrast to Kräger and Kugler (1993) and Peel and Speight (1994), they fit the levels of the data and compare the ability of the linear and non-linear models to predict levels, whereas the other two studies fitted models to the first differences and we have compared forecast performance in terms of predicting differences. For multi-period forecasts using an MSFE measure of forecast accuracy the choice of data transformation on which to assess forecast accuracy is not neutral between rival models: see Clements and Hendry (1993). Evaluation in differences is likely to play down any gains relative to evaluation in levels. The choice of the transformation on which to evaluate accuracy (e.g., levels or differences) is distinct from the issue of whether to estimate models in levels or differences.

In table 4 we give the results for two linear models: RW and an AR(3), both with constant terms, against a SETAR$(2;1,3)$, with a threshold of 5.831 and a delay of 1. Unconditionally upon the regime, there is a 1-period gain of nearly 10% relative to the RW, which at first declines in h and then rises to 20% at $h = 20$, due to a markedly superior Upper regime performance. Forecasts from the estimated SETAR model are not statistically superior to those from the AR models, unconditionally, and conditional on being in the Upper regime, there is only an apparent improvement at 1-step ahead. Neither the CRPs or the NCRs signal the gains apparent in terms of MSFE. By comparison, Chappell *et al.* (1996) find gains of 2% and 18% at $h = 1, 2$ relative to a random walk, but of nearly 50% and 350% at $h = 5$ and 10!

7.2 SETAR model forecasts of US GNP

Tiao and Tsay (1994) compare the forecast performance of an AR(2) and a two-regime SETAR model for real US quarterly GNP growth. They find that the maximum gain to the SETAR is no more than 6%, and this occurs at 3-steps ahead. However, dividing up the forecast errors into two groups depending upon the regime at the forecast origin, and then assessing forecast accuracy for each regime separately, the SETAR records gains of up to 15% in the first regime. The rationale for this effect is that over the sample period a clear majority of the data points (approximately 78%) fall in the second regime, so that the linear AR(2) model, which will largely be determined by these points, will be close to the TAR model in the second regime. Thus the forecast performance of the two models is broadly similar for data points in the second regime. However, data points in the first regime are characterised by a different process, captured by the first regime of the TAR model, so it is here that the TAR model can gain relative to the linear model.

We analyse a SETAR model similar to that estimated by Potter (1995)[5], who estimates a SETAR$(2;5,5)$ but with the third and fourth lags restricted to zero under both regimes. The delay lag $d = 2$, and the model is in the expansionary regime when $y_{t-2} > 0$ (where y_t is the difference of the log of quarterly US GNP) and otherwise in the contractionary phase. The model we use is the same except that the zero values of the coefficients on the third and fourth lags are not imposed (so that the model corresponds to Potter, 1995, Table 2, p.113). A summary of the results is given in table 5.

Comparing the SETAR model to an AR(2), we find a gain of around 16% at 1-step rising to 21% at 2-steps with nothing to choose between the two thereafter. Conditional on being in the Lower (recessionary) regime, which occurs only 24% of the time (and again for the Known Model case), the gains at 1 and 2 steps are of the order of 35%. This mirrors the empirical finding of improved forecast accuracy (relative to the linear model) when the economy happens to be in the lower regime. Conditional on the upper regime, the gain is only around 8% at $h = 1$ relative to an AR(2), rising to 16% at $h = 2$.

The CRPs are in tune with the outcome on the MSFE measure of accuracy. For example, at 2-steps ahead the SETAR correctly predicts the Lower regime around 20% of the time compared to 3% for the AR(2).

Estimating the SETAR model reverses the situation – the AR(2) is now clearly preferable on MSFE. Further investigation suggested that choosing the SETAR lag orders endogenously was particularly harm-

[5] Building on Beaudry and Koop (1993), Pesaran and Potter (1997) propose a threshold model of US GNP with an endogenously changing floor and ceiling.

ful to the SETAR. Interestingly, the CRPs indicate that even in these circumstances the SETAR correctly predicts the Lower regime 2-periods ahead nearly 25% of the time, whereas the linear model does so less than 3% of the time. Moreover, the Pesaran and Timmermann (1992) test is clearly rejected for the estimated SETAR model at 2-steps ahead, but not for the AR, so that the qualitative and quantitative indicators are apparently at odds.

7.3 SETAR model forecasts of the UK savings ratio and GDP

Peel and Speight (1995) estimate SETAR variables for a number of variables. We analyse the multi-period forecast performance of their models of the difference of the UK quarterly personal sector savings ratio, ΔSY (1955:1 - 1994:1, with the 17 latest observations held back for out-of-sample forecasting) and the rate of growth of annual GDP (1855 - 1993, last 8 years retained for forecasting).

The model for ΔSR is a SETAR$(2;1,3)$, $d=4$, and the model for GDP is SETAR$(2;0,1)$, $d=1$. For ΔSR there are statistically significant gains up to 4-steps ahead (10% at $h=1$), unconditionally when the SETAR model is known. The larger gains are in the Lower regime, and the CRPs indicate the SETAR is much better at predicting this regime 2 and 3-steps ahead. However, the MSFE-ranking of the models is reversed when the SETAR model has to be estimated from the data, and there is little to favour the SETAR on the quantitative measures. The SETAR and the AR both reject on the Pesaran and Timmermann (1992) test at 1-step.

For GDP growth, relative to an AR(1) there is only a significant gain at 1-step ahead, and then only when the SETAR is known. When the SETAR model has to be estimated the results follow the case of ΔSR – using a linear model yields more accurate forecasts.

7.4 Non-economic data

7.4.1 Lynx data.

These are the number of lynx trapped in a certain district of Canada each year from 1821 to 1934. If these are proportional to the population, then a threshold model which captures the underlying population dynamics (a rising number of births below a critical population size, declining above that ‘threshold’, with a ‘delay’ as the young mature to reproductive age) may be appropriate.

The model we investigate is taken from Tong (1995a), p.387, equation 7.7, and is a SETAR$(2;7,2)$, with a delay lag $d=2$ and a threshold value of 3.116. Since the unconditional results establish the superiority of the SETAR model, to save space we do not report the conditional results. When the SETAR model is known (see table 7) the ratio of the MSFE of an AR(4) model (Tong, 1995a gives a SETAR$(2;2,2)$ as a first approximation to the Lynx data) to the SETAR model MSFE is around 1.8 for 1-step ahead (indicating an 80% gain), rising to over 2 for 3-steps ahead, and then declines but is still around 1.1 at 20-steps ahead. Estimating the SETAR roughly halves the gains at 1, 2 and 5-steps ahead. The CRPs indicate that the SETAR is better at predicting the Upper regime at 5 and 10 -steps ahead, and the NCRs indicate that the SETAR model is less confused. Both models have predictive ability judged by the Pesaran and Timmermann (1992) test.

7.4.2 Wolf’s annual mean sunspot numbers.

The physical process that constitutes the solar cycle is not well understood, in contrast to the biological process of the lynx population dynamics, but both have become classic data sets in non-linear modelling.

Here we consider the SETAR$(2;10,2)$, with $d=8$ and a threshold of 11.93, taken from Tong (1995a), p.421, equation 7.15. Following Ghaddar and Tong (1981), the SETAR model is fit to the series

(y_t) after applying an instantaneous square-root transformation to the original numbers (x_t). Hence we generate realizations of the series $\{y\}_1^{T+H}$ using the SETAR model but transform these to the original units of observation (using the inverse of the square-root transformation) for fitting the linear models and evaluating the forecasts from both models.

Tong (1995a), p.425-427 reports the results of an empirical forecast comparison between a SETAR$(2;3,11)$ model and an AR(9) model fitted to the original sunspot numbers over the period 1700-1920, and used to forecast 1956-1979. Non-linear prediction is found to do better over the troughs but worse over the peaks, and overall linear prediction is superior on MSFE at all but the shortest horizons. The 1956 observation is somewhat anomalous, indicating an unusually steep rise, and arguably affects the SETAR more than the linear model.

Generating data by simulating the SETAR$(2;10,2)$ of the transformed series, we find gains of the known SETAR model relative to an AR(9) of 40%, 34% and 12% at horizons of $1, 2$ and 10, respectively, and these are reduced to 30%, 24% and 10% when the SETAR model is estimated. The CRPs for the Lower regime at short and medium horizons are higher for the SETAR, matching the empirical finding reported above. The SETAR model is less confused at short and moderate horizons, and both models reject on the Pesaran and Timmermann (1992) test.

7.5 The indicator of the serial dependence of regimes

Table 8 records the proportion of the replications of the Monte Carlo for which the number of times the simulated data remained in regime i (Lower, Middle or Upper) for 2 periods was not consistent with the data being generated by a linear model. Comparing these rejection frequencies with the relative MSFE performances discussed above, it is apparent that there is a clear positive association: when linearity is rejected a large number of times, there is usually a clear gain in relative forecast accuracy to the SETAR: witness the results for the Lynx and Sunspot data sets. Conversely, for both the dollar exchange rates the null of linearity is rejected less than 5% of the time for one of the three regimes, which matches the finding that any forecast gains that there might be at 1-step ahead when the SETAR model is known, disappear when it has to be estimated. However, for the Lira the linear null is rejected 20% of the time for the middle regime, while conditional on being in the middle regime there are significant gains to the estimated SETAR model over a random walk.

For the sterling-dollar inter-war exchange rate the linear model is seldom rejected, while the same finding for the French franc is more surprising, given the MSFE gains.

For US GNP the rejection of around 20% in the lower regime matches the earlier indications (mainly qualitative) that the SETAR model has some ability to predict this regime. The findings for ΔSR are reasonably in tune with those from the qualitative and quantitative measures reported earlier, while the high rates of rejection for UK GDP are anomalous.

8 Conclusions

In this paper we have undertaken a comprehensive analysis of the multi-period forecast performance of a number of empirical self-exciting threshold autoregressive (SETAR) models that have been proposed in the literature, using both quantitative and qualitative measures of forecast accuracy. The Monte Carlo approach we adopt favours the SETAR model relative to an empirical comparison by ruling out the possibility that the SETAR forecast model is capturing ‘non-linearities’ (outliers, non-SETAR type non-linearities) which cannot be exploited for forecasting or which do not persist in to the future.

Our findings indicate that the ability to exploit non-linearities for forecasting may turn on whether forecasts are evaluated conditional upon the regime, reflecting the ability of non-linear models to forecast well in certain states of nature, but not always sufficiently well to score better than linear models on average (across all states of nature).

We found that non-linear models appear to be favoured by qualitative measures of forecast performance, and we considered a number of these.

We also proposed and implemented an indicator of when SETAR models are likely to forecast well relative to linear rivals, based on the serial dependence of regimes. For the most part, the SETAR model forecast gains relative to the linear model are more marked when the indicator of regime dependence clearly rejects the linear model. Nevertheless, from our results it is apparent that this is not the whole story, and factors other than the serial dependence of regimes have a role to play in determining when the SETAR model will yield an improved forecast performance relative to a linear model.

In summary, our study has shown that it may not always be possible to exploit non-linearities of a SETAR type to yield markedly 'better' forecasts than a linear model on average, or unconditionally, even when such non-linearities are a feature of the data (by construction). Model uncertainty and parameter estimation uncertainty adversely affect the SETAR model forecasts, particularly when the threshold values and the delay, together with the lag orders and the autoregressive coefficients, all have to be determined form the data. Nevertheless, conditional on the regime in force at the time the forecast is made, substantial improvements over linear models are possible.

References

Acemoglu, D., and Scott, A. (1994). Asymmetries in the cyclical behaviour of UK labour markets. *Economic Journal*, **104**, 1303–1323.

Akaike, A. (1973). Information theory and an extension of the maximum likelihood principle. In Petrov, B. N., and Saki, F. L. (eds.), *Second International Symposium of Information Theory*. Budapest.

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

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, **61**, 821–856.

Beaudry, P., and Koop, G. (1993). Do recessions permanently affect output. *Journal of Monetary Economics*, **31**, 149–163.

Bollerslev, T. (1986). Generalised autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **51**, 307–327.

Box, G. E. P., and Jenkins, G. M. (1970). *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day.

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

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.

Clements, M. P., and Hendry, D. F. (1993). On the limitations of comparing mean squared forecast errors. *Journal of Forecasting*, **12**, 617–637. With discussion.

Clements, M. P., and Smith, J. (1997). The performance of alternative forecasting methods for SETAR models. *International Journal of Forecasting*. Forthcoming.

Cleveland, W. S., Devlin, S. J., and Grosse, E. (1988). Regression by local fitting: Methods, properties and computational algorithms. *Journal of Econometrics*, **37**, 87–114.

De Gooijer, J. G., and De Bruin, P. (1997). On SETAR forecasting. *Statistics and Probability Letters*. Forthcoming.

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.

Diebold, F. X., and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, **13**, 253–263.

Diebold, F. X., and Nason, J. A. (1990). Nonparametric exchange rate prediction. *Journal of International Economics*, **28**, 315–332.

Durland, J. M., and McCurdy, T. H. (1994). Duration dependent transitions in a Markov model of U.S. GNP growth. *Journal of Business and Economic Statistics*, **12**, 279–288.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**, 987–1007.

Gallant, A. R., Rossi, P. E., and Tauchen, G. (1993). Nonlinear dynamic structures. *Econometrica*, **61**, 871–907.

Ghaddar, D. K., and Tong, H. (1981). Data transformation and self-exciting threshold autoregression. *Applied Statistics*, **30**, 238–248.

Granger, C. W. J., and Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships*. Oxford: Oxford University Press.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, **57**, 357–384.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, **64**, 413–430.

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*. Forthcoming.

Henriksson, R. D., and Merton, R. C. (1981). On market timing and investment performance. II statistical procedures for evaluating forecast skills. *Journal of Business*, **54**, 513–533.

Hsieh, D. A. (1989). A non-linear stochastic rational expectations model of exchange rates. Unpublished manuscript, Fuqua School of Business, Duke University.

Koop, G., Pesaran, M. H., and Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, **74**, 119–147.

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

Meese, R. A., and Rose, A. K. (1991). An empirical assessment of non-linearities in models of exchange rate determination. *Review of Economic Studies*, **58**, 603–619.

Montgomery, A. L., Zarnowitz, V., Tsay, R. S., and Tiao, G. C. (1997). Forecasting the U.S. unemployment rate. Mimeo, Wharton School, University of Pennsylvania, Philadelphia.

Peel, D. A., and Speight, A. E. H. (1994). Testing for non-linear dependence in inter-war exchange rates. *Weltwirtschaftliches Archiv*, **130**, 391–417.

Peel, D. A., and Speight, A. E. H. (1995). Modelling macroeconomic time series: A comparative analysis of parametric and nonparametric methods. Discussion paper, University of Wales, Aberystwyth.

Pesaran, M. H., and Potter, S. M. (1997). A floor and ceiling model of US Output. *Journal of Economic Dynamics and Control*, **21**, 661–695.

Pesaran, M., and Timmermann, A. (1992). A simple nonparametric test of predictive performance. *Journal of Business and Econometic Statistics*, **10**, 461–465.

Potter, S. (1995). A nonlinear approach to U.S. GNP. *Journal of Applied Econometrics*, **10**, 109–125.

Schnader, M. H., and Stekler, H. O. (1990). Evaluating predictions of change. *Journal of Business*, **63**, 99–107.

Stekler, H. O. (1994). Are economic forecasts valuable?. *Journal of Forecasting*, **13**, 495–505.

Teräsvirta, T., and Anderson, H. M. (1992). Characterizing nonlinearities in business cycles using smooth transition autoregressive models. *Journal of Applied Econometrics*, **7**, 119–139.

Tiao, G. C., and 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*, pp. 101–141. Amsterdam: Sijhoff and Noordoff.

Tong, H. (1983). *Threshold Models in Non-Linear Time Series Analysis*: Springer-Verlag, New York.

Tong, H. (1995a). *Non-linear Time Series. A Dynamical System Approach*. Oxford: Clarendon Press. First published 1990.

Tong, H. (1995b). A personal overview of non-linear time series analysis from a chaos perspective. *Scandinavian Journal of Statistics*, **22**, 399–445.

Tong, H., and Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data. *Journal of The Royal Statistical Society*, **B 42**, 245–292.

Table 2 Kräger and Kugler (1993), Dollar Exchange Rates.

Italian Lira

[A] Known SETAR Model – Unconditional

Horizon	MSFE : AR/SETAR		CRPs : AR			CRPs : SETAR			NCRs	
	RW	p-value	Lower	Middle	Upper	Lower	Middle	Upper	AR	SETAR
1	1.033	0.000	1.000	0.000	0.000	0.954	0.126	0.000	.593	.596
2	1.004	0.260	1.000	0.000	0.000	0.995	0.023	0.000	.589	.592
5	0.996	0.864	1.000	0.000	0.000	1.000	0.000	0.000	.571	.571

[B] Known SETAR Model – Conditional. MSFE : AR/SETAR

	Lower		Middle		Upper	
	RW	p-value	RW	p-value	RW	p-value
1	1.003	0.272	1.160	0.000	0.999	0.560
2	1.008	0.086	1.015	0.016	0.999	0.574
5	0.997	0.705	1.003	0.265	1.002	0.345

[C] Unknown SETAR Model – Unconditional

Horizon	MSFE : AR/SETAR		CRPs : SETAR			NCR
	RW	p-value	Lower	Middle	Upper	SETAR
1	1.007	0.261	0.911	0.100	0.000	.564
2	0.990	0.972	0.964	0.051	0.000	.581
5	0.995	0.949	0.996	0.000	0.000	.569

[D] Unknown SETAR Model – Conditional. MSFE : AR/SETAR

	Lower		Middle		Upper	
	RW	p-value	RW	p-value	RW	p-value
1	0.993	0.847	1.097	0.000	0.987	0.893
2	1.002	0.383	1.012	0.037	0.996	0.689
5	0.995	0.910	1.000	0.557	0.998	0.725

Japanese Yen

[A] Known SETAR Model – Unconditional

Horizon	MSFE : AR/SETAR		CRPs : AR			CRPs : SETAR			NCRs	
	RW	p-value	Lower	Middle	Upper	Lower	Middle	Upper	SETAR	AR
1	1.046	0.000	0.634	0.344	0.000	0.643	0.521	0.069	.367	.415
2	1.003	0.340	0.647	0.333	0.000	0.611	0.455	0.000	.372	.370
5	0.997	0.774	0.643	0.318	0.000	0.663	0.374	0.000	.347	.363

[B] Known SETAR Model – Conditional. MSFE : AR/SETAR

	Lower		Middle		Upper	
	RW	p-value	RW	p-value	RW	p-value
1	1.028	0.000	1.411	0.000	1.024	0.010
2	1.006	0.141	1.005	0.210	1.001	0.438
5	1.011	0.021	1.004	0.198	0.999	0.605

[C] Unknown SETAR Model – Unconditional

Horizon	MSFE : AR/SETAR		CRPs : SETAR			NCR
	RW	p-value	Lower	Middle	Upper	SETAR
1	1.011	0.192	0.603	0.469	0.072	.390
2	0.993	0.877	0.602	0.402	0.011	.362
5	0.995	0.958	0.620	0.393	0.000	.344

[D] Unknown SETAR Model – Conditional. MSFE : AR/SETAR

	Lower		Middle		Upper	
	RW	p-value	RW	p-value	RW	p-value
1	1.001	0.466	1.149	0.000	0.994	0.673
2	1.000	0.478	1.005	0.204	1.003	0.271
5	1.004	0.085	0.999	0.594	1.000	0.557

Table 3 Peel and Speight (1994), Sterling – US Dollar rate.

US Dollar

[A] Known SETAR Model – Unconditional

	MSFE		MSFE		CRPs : AR(3)			CRPs : SETAR			NCRs	
	AR/SETAR		AR/SETAR		AR	SETAR						
	RW	*p*-value	AR(3)	*p*-value	Lower	Middle	Upper	Lower	Middle	Upper		
1	1.128	0.000	1.103	0.000	0.025	0.951	0.081	0.122	0.849	0.170	.382	.404
2	1.100	0.000	1.073	0.000	0.015	0.957	0.061	0.112	0.919	0.128	.381	.417
3	1.030	0.002	1.042	0.000	0.000	0.992	0.009	0.000	0.981	0.009	.372	.368
4	1.008	0.162	1.013	0.086	0.000	1.000	0.000	0.000	1.000	0.000	.371	.371
5	1.003	0.338	1.005	0.241	0.000	0.997	0.000	0.000	1.000	0.000	.363	.364

[B] Known SETAR Model – Conditional. MSFE : AR/SETAR

	Lower				Middle				Upper			
	RW	*p*-value	AR(3)	*p*-value	RW	*p*-value	AR(3)	*p*-value	RW	*p*-value	AR(3)	*p*-value
1	1.004	0.293	1.049	0.000	1.690	0.000	1.327	0.000	1.009	0.089	1.056	0.000
2	1.073	0.000	1.076	0.000	1.137	0.000	1.115	0.000	1.065	0.000	1.075	0.000
3	1.014	0.053	1.014	0.123	1.043	0.006	1.039	0.010	0.999	0.545	1.008	0.185
4	0.998	0.621	1.005	0.229	1.015	0.051	1.018	0.060	1.010	0.114	1.014	0.044
5	1.012	0.060	1.011	0.082	0.999	0.530	1.001	0.445	1.013	0.042	1.012	0.050

[C] Unknown SETAR Model – Unconditional

	MSFE		MSFE		CRPs : SETAR			NCR
	AR/SETAR		AR/SETAR		SETAR			
	RW	*p*-value	AR(3)	*p*-value	Lower	Middle	Upper	
1	0.997	0.565	0.974	0.897	0.061	0.929	0.067	.379
2	0.987	0.818	0.963	0.990	0.022	0.946	0.028	.367
3	0.983	0.979	0.994	0.776	0.003	0.968	0.009	.364
4	0.997	0.700	1.002	0.369	0.000	0.995	0.000	.369
5	0.997	0.686	1.000	0.553	0.000	0.997	0.000	.363

[B] Unknown SETAR Model – Conditional. MSFE : AR/SETAR

	Lower				Middle				Upper			
	RW	*p*-value	AR(3)	*p*-value	RW	*p*-value	AR(3)	*p*-value	RW	*p*-value	AR(3)	*p*-value
1	0.946	1.000	0.989	0.775	1.140	0.000	0.896	1.000	0.952	0.999	0.997	0.604
2	0.962	0.987	0.964	0.991	1.016	0.179	0.996	0.605	0.989	0.776	0.997	0.587
3	0.980	0.982	0.980	0.978	0.998	0.587	0.994	0.733	0.989	0.833	0.998	0.609
4	0.991	0.805	0.999	0.561	0.987	0.990	0.991	0.902	0.999	0.548	1.003	0.342
5	0.994	0.834	0.992	0.905	0.996	0.742	0.998	0.641	0.998	0.624	0.997	0.734

Table 4 Chappell et al. (1995), Fr. Franc / Deutschmark.

[A] Known SETAR Model – Unconditional

Horizon	MSFE : AR/SETAR		MSFE : AR/SETAR		CRPs : AR		CRPs : SETAR		NCRs	
	RW	p-value	AR(3)	p-value	Lower	Upper	Lower	Upper	AR	SETAR
1	1.097	0.000	1.095	0.000	0.947	0.691	0.955	0.701	.897	.906
2	1.065	0.004	1.058	0.003	0.945	0.647	0.958	0.636	.889	.898
5	1.078	0.012	1.069	0.023	0.959	0.485	0.974	0.423	.867	.867
10	1.116	0.026	1.084	0.061	0.965	0.280	0.980	0.213	.820	.818
20	1.192	0.052	1.110	0.113	0.976	0.081	0.999	0.028	.787	.794

[B] Known SETAR Model – Conditional. MSFE : AR/SETAR

	Lower				Upper			
	RW	p-value	AR(3)	p-value	RW	p-value	AR(3)	p-value
1	1.005	0.201	1.034	0.001	1.476	0.000	1.215	0.000
2	1.001	0.479	1.037	0.005	1.410	0.000	1.128	0.001
5	1.023	0.160	1.060	0.013	1.396	0.000	1.059	0.083
10	1.049	0.157	1.099	0.028	1.610	0.000	1.080	0.059
20	1.089	0.173	1.137	0.067	1.786	0.000	1.081	0.118

[C] Unknown SETAR Model – Unconditional

Horizon	MSFE : AR/SETAR		MSFE : AR/SETAR		CRPs : SETAR		NCR
	RW	p-value	AR(3)	p-value	Lower	Upper	SETAR
1	1.031	0.085	1.029	0.064	0.944	0.680	.893
2	1.021	0.228	1.015	0.246	0.954	0.572	.883
5	0.999	0.509	0.990	0.639	0.971	0.412	.863
10	1.011	0.439	0.983	0.651	0.971	0.218	.812
20	1.041	0.371	0.970	0.635	0.976	0.057	.782

[D] Unknown SETAR Model – Conditional. MSFE : AR/SETAR

	Lower				Upper			
	RW	p-value	AR(3)	p-value	RW	p-value	AR(3)	p-value
1	0.993	0.820	1.022	0.014	1.325	0.000	1.090	0.001
2	0.986	0.833	1.022	0.045	1.273	0.000	1.018	0.289
5	0.974	0.788	1.009	0.332	1.244	0.004	0.943	0.894
10	0.961	0.737	1.007	0.419	1.446	0.000	0.970	0.707
20	0.940	0.699	0.981	0.595	1.499	0.007	0.907	0.811

Table 5 Potter (1995), US GNP.

[A] Known SETAR Model – Unconditional								
Horizon	MSFE : AR/SETAR		CRPs : AR		CRPs : SETAR		NCRs	
	AR(2)	*p*-value	Lower	Upper	Lower	Upper	AR	SETAR
1	1.167	0.000	0.167	0.959	0.290	0.954	.784	.807
2	1.219	0.000	0.024	0.968	0.220	0.950	.728	.765
3	1.009	0.399	0.008	0.997	0.081	0.966	.742	.738
5	1.033	0.103	0.000	1.000	0.000	1.000	.761	.761
10	1.012	0.154	0.000	1.000	0.000	1.000	.762	.762

[B] Known SETAR Model – Conditional. MSFE : AR/SETAR				
	Lower		Upper	
	AR(2)	*p*-value	AR(2)	*p*-value
1	1.351	0.000	1.084	0.000
2	1.346	0.000	1.167	0.000
3	1.048	0.181	1.018	0.204
5	1.046	0.121	1.018	0.157
10	1.012	0.124	1.016	0.060

[C] Unknown SETAR Model – Unconditional					
Horizon	MSFE : AR/SETAR		CRPs : SETAR		NCR
	AR(2)	*p*-value	Lower	Upper	SETAR
1	0.864	1.000	0.258	0.908	.764
2	0.985	0.633	0.244	0.908	.739
3	0.915	0.951	0.058	0.926	.702
5	0.834	0.970	0.042	0.966	.745
10	0.825	0.849	0.038	0.972	.750

[D] Unknown SETAR Model – Conditional. MSFE : AR/SETAR				
	Lower		Upper	
	AR(2)	*p*-value	AR(2)	*p*-value
1	1.001	0.485	0.769	1.000
2	0.944	0.901	0.846	0.999
3	0.951	0.798	0.833	1.000
5	0.839	0.969	0.880	0.978
10	0.851	0.838	0.771	0.824

Table 6 Peel and Speight (1995).

UK ΔSR

[A] Known SETAR Model – Unconditional

Horizon	MSFE – AR/SETAR		CRPs : AR		CRPs : SETAR		NCR	
	AR(3)	*p*-value	Lower	Upper	Lower	Upper	AR	SETAR
1	1.104	0.000	0.415	0.814	0.526	0.841	.666	.724
2	1.073	0.000	0.040	0.963	0.230	0.892	.613	.641
5	1.005	0.066	0.000	1.000	0.046	0.982	.654	.658

[B] Known SETAR Model – Conditional. MSFE : AR/SETAR

	Lower		Upper	
	AR(3)	*p*-value	AR(3)	*p*-value
1	1.245	0.000	1.043	0.000
2	1.147	0.000	1.075	0.000
5	1.016	0.091	1.021	0.033

[C] Unknown Model – Unconditional

Horizon	MSFE – AR/SETAR		CRPs : SETAR		NCR
	AR(3)	*p*-value	Lower	Upper	SETAR
1	0.814	1.000	0.383	0.806	.649
2	0.908	0.978	0.161	0.857	.593
5	0.947	0.977	0.084	0.913	.626

[D] Unknown SETAR Model – Conditional. MSFE : AR/SETAR

	Lower		Upper	
	AR(3)	*p*-value	AR(3)	*p*-value
1	0.937	0.999	0.807	1.000
2	0.901	0.839	0.847	0.999
5	0.961	0.913	0.938	0.998

UK GDP growth

[A] Known SETAR Model – Unconditional

Horizon	MSFE – AR/SETAR		CRPs : AR		CRPs : SETAR		NCRs	
	AR(1)	*p*-value	Lower	Upper	Lower	Upper	AR	SETAR
1	1.069	0.000	0.000	1.000	0.143	0.928	.629	.637
2	1.006	0.555	0.000	1.000	0.000	1.000	.613	.613
5	1.010	0.130	0.000	1.000	0.000	1.000	.597	.597

[B] Known SETAR Model – Conditional. MSFE : AR/SETAR

	Lower		Upper	
	AR(1)	*p*-value	AR(1)	*p*-value
1	1.032	0.089	1.168	0.000
2	1.017	0.006	1.010	0.199
5	1.005	0.754	1.016	0.069

[C] Unknown SETAR Model – Unconditional

Horizon	MSFE – AR/SETAR		CRPs : SETAR		NCR
	AR(1)	*p*-value	Lower	Upper	SETAR
1	0.903	1.000	0.132	0.844	.580
2	0.900	1.000	0.158	0.832	.571
5	0.889	0.998	0.136	0.849	.567

[D] Unknown SETAR Model – Conditional. MSFE : AR/SETAR

	Lower		Upper	
	AR(1)	*p*-value	AR(1)	*p*-value
1	0.947	1.000	0.879	1.000
2	0.908	1.000	0.916	1.000
5	0.915	1.000	0.903	0.996

Table 7 Tong (1995), Non-economics data.

Tong (1995), Lynx data								
[A] Known SETAR Model – Unconditional								
Horizon	MSFE – AR/SETAR		CRPs : AR		CRPs : SETAR		NCRs	
	AR(4)	p-value	Lower	Upper	Lower	Upper		
1	1.777	0.000	0.922	0.836	0.931	0.870	.886	.905
2	1.897	0.000	0.814	0.629	0.843	0.789	.742	.822
5	1.484	0.000	0.885	0.394	0.833	0.683	.693	.744
10	1.120	0.009	0.942	0.273	0.873	0.527	.668	.731
20	1.083	0.060	0.990	0.021	0.900	0.259	.612	.650
[C] Unknown SETAR Model – Unconditional								
Horizon	MSFE – AR/SETAR		CRPs : SETAR		NCR			
	AR(4)	p-value	Lower	Upper	SETAR			
1	1.397	0.000	0.919	0.867	.897			
2	1.511	0.000	0.838	0.714	.790			
5	1.218	0.003	0.819	0.624	.743			
10	0.996	0.543	0.873	0.446	.698			
20	0.992	0.578	0.933	0.151	.628			

Tong (1995), Sunspot numbers								
[A] Known SETAR Model – Unconditional								
Horizon	MSFE – AR/SETAR		CRPs : AR		CRPs : SETAR		NCRs	
	AR(9)	p-value	Lower	Upper	Lower	Upper	AR	SETAR
1	1.397	0.000	0.923	0.841	0.936	0.873	.892	.913
2	1.338	0.000	0.850	0.770	0.897	0.738	.820	.837
5	1.343	0.002	0.777	0.674	0.848	0.668	.743	.789
10	1.122	0.153	0.812	0.650	0.790	0.686	.753	.725
20	1.106	0.196	0.738	0.501	0.778	0.501	.654	.680
[C] Unknown SETAR Model – Unconditional								
Horizon	MSFE – AR/SETAR		CRPs : SETAR		NCR			
	AR(9)	p-value	Lower	Upper	SETAR			
1	1.309	0.000	0.939	0.873	.914			
2	1.243	0.003	0.892	0.757	.841			
5	1.263	0.014	0.850	0.649	.784			
10	1.099	0.183	0.811	0.661	.756			
20	1.085	0.227	0.806	0.456	.682			

Table 8 Rejection frequencies of the linear null based on the indicator of the serial dependence of regimes. $R = 2$.

		Lower	Middle	Upper
Kräger and Kugler (1993)	Italian Lira	0.097	0.217	0.029
	Japanese Yen	0.113	0.030	0.119
Peel and Speight (1994)	US Dollar	0.058	0.137	0.022
Chappell *et al.* (1996)	French Franc	0.008	-	0.003
Potter (1995)	US GNP	0.206	-	0.161
Tong (1995a)	Lynx	0.694	-	0.611
	Sunspot	0.726	-	0.324
Peel and Speight (1995)	Δ SR	0.042	-	0.101
	UK GDP	0.531	-	0.256