An Automatic Procedure for Identification, Estimation and Forecasting Univariate Self Exiting Threshold Autoregressive Models

Author(s): N. Davies, J. Pemberton and J. D. Petruccelli
Source: *Journal of the Royal Statistical Society. Series D (The Statistician)*, Vol. 37, No. 2, Special Issue: Statistical Forecasting and Decision-Making (1988), pp. 199-204
Published by: Wiley for the Royal Statistical Society
Stable URL: https://www.jstor.org/stable/2348693
Accessed: 31-07-2019 20:28 UTC

*The Statistician* (1988) 37, pp. 199–204

# An automatic procedure for identification, estimation and forecasting univariate self exiting threshold autoregressive models

N. DAVIES[1], J. PEMBERTON[2], J. D. PETRUCCELLI[3]

[1]*Trent Polytechnic, Nottingham,* [2]*Salford University, Salford, United Kingdom and* [3]*Worcester Polytechnic Institute, Worcester, MA, U.S.A.*

**Abstract.** We outline the procedure for identification and estimation of self exiting threshold autoregressive—SETAR—models, based on cusum tests. Forecasting for a general nonlinear autoregressive—NLAR—model is then discussed and a recurrence relation for quantities related to the forecast distribution is given. Some preliminary results from fitting and forecasting SETAR models are then summarised and discussed.

## 1 Identification and estimation of SETAR models

We are concerned with the fitting of self exiting threshold autoregressive (SETAR) models of the form

$$X_t = a_0^{(j)} + \sum_{i=1}^{k} a_i^{(j)} X_{t-1} + \varepsilon_t^{(j)} \text{ if } X_{t-d} \,\varepsilon\, (\mathrm{r}_{j-1}, r_j]$$

where $j=1, 2, \ldots, l$, $-\infty = r_0 < r_1 < \ldots < r_l = \infty$ are the threshold values, $k$=largest order of the model $>d$=delay lag, and $\varepsilon_t^{(j)}$ are i.i.d. sequences of random variables with zero mean and variance $\sigma_j^2$. These sequences are also independent. The acronym SETAR ($l$; $k_1$, $k_2$, $\ldots$, $k_l$) is used for such a model, signifying that there are $l$ submodels, the $j$th of which has order $k_j$, the last non-zero coefficient being at this lag. Further information about this class of models can be found in, e.g. Tong (1983).

Given $n$ observed values of a time series, $x_1$, $x_2$, $\ldots$, $x_n$, we have used two methods of identifying and fitting a SETAR model. The first is based on the cusum test for detecting nonlinearity developed by Petruccelli & Davies (1986) and it proceeds in four steps:

(i) For $1 \leqslant d \leqslant k \leqslant$ KMAX, conduct a cusum test to select the ($k$, $d$) values for which the series is suggested as nonlinear. If none are selected, then the series is taken to be linear.

(ii) If (i) throws out some values, use a *vmask* and *runs test* to find initial threshold estimates (up to a maximum number to be selected by the investigator).

(iii) Fit these models and compute their BIC.

(iv) Perturb the threshold values slightly, go to (iii), unless a local minimum of BIC has been attained.

The second procedure is based on a likelihood ratio test (LRT). It is more time consuming and as a result we have only used one threshold. The three stages of this procedure are as follows:

(i) For each $1 \leqslant d \leqslant k \leqslant$ KMAX, obtain the least squares estimate of the threshold value. Fit the model and denote its MSE by MSE($k$, $d$).

(ii) Choose the parameter estimates $\mathrm{k}_0$, $\mathrm{d}_0$, for which MSE($k_0$, $d_0$) is the minimum of MSE($k$, $d$) over all ($k$, $d$) allowed in (i). Find the likelihood ratio statistic of this model versus the null model, i.e. a linear AR($k_0$).

(iii) To assess the significance of the LRT statistic, simulate n observations from the AR($k_0$), and repeat from (i).

## 2 Forecasting with general nonlinear AR models

The problem of obtaining the minimum mean squared error (MMSE) $m$-steps ahead forecast of a nonlinear autoregressive (NLAR) model was first tackled by Jones (1976, 1978). Recently, a different approach was proposed by Pemberton (1987) and this is the one adopted here. We now give a brief outline of the derivation of this method.

Let $X_t$ be given by the NLAR of order $k$, i.e. a NLAR($k$).

$$X_t=\lambda(\mathbf{X}_{t-1})+\varepsilon_t \tag{2.1}$$

where $\lambda$: $\mathrm{IR}^k \to \mathrm{IR}$, $\mathbf{X}_{t-1}=(X_{t-1}, X_{t-2}, \ldots, X_{t-k})^T$ and $\{\varepsilon_t\}$ is a sequence of zero mean, independent, identically distributed random variables with variance $\sigma^2$. Let $g(\cdot)$ be the p.d.f. of $\varepsilon_t$ and $f_m(x|\mathbf{x}_t)$ denote the conditional p.d.f. of $X_{t+m}$ given $\mathbf{X}_t=\mathbf{x}_t$ (i.e. the $m$-step predictive p.d.f.). Then

$$f_m(x|\mathbf{x}_t)=\int_{-\infty}^{\infty} f_{m-1}(x|\mathbf{x}_{t+1})\, g\{x_{t+1}-\lambda(\mathbf{x}_t)\}dx_{t+1} \tag{2.2}$$

where

$$\mathbf{x}_s=(x_s, x_{s-1}, \ldots, x_{s-k+1})^T,\ s=t,\ t+1.$$

This equation can be obtained by considering the joint p.d.f. of $X_{t+m}, X_{t+m-1}, \ldots, X_{t+1}$ conditional on $\mathbf{X}_t=\mathbf{x}_t$ and integrating out the unwanted variables as in Pemberton (1987). Alternatively, we can follow Tong & Moeanaddin (1987) and use the Chapman-Kolmogorov relation for the conditional p.d.f. of $\mathbf{X}_{t+m}$ given $\mathbf{X}_t=\mathbf{x}_t$, which can be written with an abuse of notation as

$$\begin{aligned} f_m(\mathbf{x}_{t+1}|\mathbf{x}_t) &= \int_{\mathrm{IR}^k} f_{m-1}(\mathbf{x}_{t+m}|\mathbf{x}_{t+1})\, f(\mathbf{x}_{t+1}|\mathbf{x}_t)\, d\mathbf{x}_{t+1} \\ &= \int_{-\infty}^{\infty} f_{m-1}(\mathbf{x}_{t+m}|\mathbf{x}_{t+1}) f(x_{t+1}|\mathbf{x}_t)\, dx_{t+1}. \end{aligned}$$

The last expression follows from the fact that for the model (2.1), all but the first element of $\mathbf{X}_{t+1}$ is fixed by knowledge that $\mathbf{X}_t=\mathbf{x}_t$. Hence the conditional p.d.f. of the former given the latter contains a product of delta functions and so we obtain the last expression. It remains to integrate out the unwanted variables as before.

Although equation (2.2) contains all the information about the future observation $X_{t+m}$ given knowledge of the past and present, $\mathbf{X}_t$, we are interested in the performance of the MMSE forecast i.e. the mean of the $m$-step predictive p.d.f. given by the solution of the recurrence

$$P_m(\mathbf{x}_t)=\int_{-\infty}^{\infty} P_{m-1}(\mathbf{x}_{t+1})\, g\{x_{t+1}-\lambda(\mathbf{x}_t)\}\, dx_{t+1} \tag{2.3}$$

with $P_1(\mathbf{x}_t)\equiv\lambda(\mathbf{x}_t)$. In fact, we can use (2.2) to obtain a similar recurrence for $E[H(X_{t+m}|\mathbf{X}_t=\mathbf{x}_t]$ for any suitably well behaved function $H(\cdot)$, by taking $P_1(\mathbf{x}_t)\equiv E[H(X_{t+1})|\mathbf{X}_t=\mathbf{x}_t]$.

We have developed a numerical procedure for solving equation (2.3), based on Guass quadrature rules for the threshold autoregression (1.1), when $g(\cdot)$ is the density of the $N(0, \sigma^2)$ distribution in the cases $H(x)\equiv x$ and $H(x)\equiv x^2$. We also allow for the noise variances to be different in the different regions of the model. This simply requires $g(z)$ to be replaced by $\phi(z/\sigma_j)/\sigma_j$ if $r_{j-1}<x_{t-d+1}\leqslant r_j$, where $z=x_{t+1}-\lambda(\mathbf{x}_t)$ and $\phi(\cdot)$ is the $N(0, 1)$ density.

## 3 Performance of MMSE forecasts of SETAR models

Several experiments to assess the forecasting performance of SETAR models driven by Gaussian noise were carried out. In all of these the forecast errors for 1, 2 and 3 steps ahead are compared with those for a linear AR model. The relative accuracy of the numerical calculation of SETAR forecasts was four significant figures. This is an adjustable parameter in our computer program and in practice would probably be set to two or three significant figures. We only present here the results of three experiments, these being typical of what we have observed and sufficient for our discussion in section four.

In experiment I we carried out the identification and estimation of SETAR models with KMAX=3 and up to 4 thresholds for 155 real time series. Of these, 63 were identified as being nonlinear. Experiment II was similar, except that we used simulated series of length 100 from each of 50 randomly selected stationary SETAR(2; 1, 1) models of the form

$$X_t=\begin{cases} aX_{t-1}+\varepsilon_t \text{ if } X_{t-1}\leqslant 0 \\ bX_{t-1}+\varepsilon_t \text{ if } X_{t-1}>0 \end{cases} \tag{2.4}$$

with the noise taken to be independent $N(0, 1)$ in both regions. This time, 45 were identified as being nonlinear.

For all of these series, the last three observations were not used in the identification stage, but were used to compute the forecast errors for forecasts from the origin at the fourth observation from the end of the series. For any series identified as nonlinear, a linear AR model was identified and fitted using the BIC selection criterion. The maximum order was taken as the number of estimated parameters in the fitted SETAR model i.e. the number of thresholds plus the number of coefficients plus 1 for the delay lag, and this came to at most 25.

Experiment III was intended to be the ultimate 'benchmark'. The models used in experiment II were once again employed for simulation with $N(0, 1)$ noise, but this time 10003 observations were obtained. A linear AR model was fitted to the first 10000 and the remaining 3 were used to compare forecasts with the origin at the 10000th observation. The linear AR order was selected in one of two ways, (a) fixed to be 1 and (b) selected by BIC with maximum 20. These were thought of as being the theoretical MMSE AR approximations to the SETAR models.

We summarise our results in Table 1 by reporting the percentage of times that the absolute forecast error is smaller for the nonlinear than for the linear model.

**Table 1.** Percentage of times absolute forecast error is smaller for nonlinear than for linear model in MMSE forecasts

| | No. of steps ahead | | |
|---|---|---|---|
| Experiment | 1 | 2 | 3 |
| I | 48 | 56 | 52 |
| II | 69 | 51 | 58 |
| IIIa) | 66 | 48 | 58 |
| IIIb) | 62 | 48 | 54 |

The differences in magnitude of the forecast errors between the linear and the nonlinear models is represented for experiment IIIb for one, two and three steps ahead by dotplots of the performance measure

$$\mathrm{PM} = (L^2 - NL^2)/(L^2 + NL^2).$$

Here, L and NL are used to denote the forecast errors for the linear and nonlinear models respectively. Clearly $-1 \leqslant \mathrm{PM} \leqslant 1$, positive values indicating that the nonlinear error is the smaller. The plots are given in Fig. 1, and are typical of those from the other experiments.

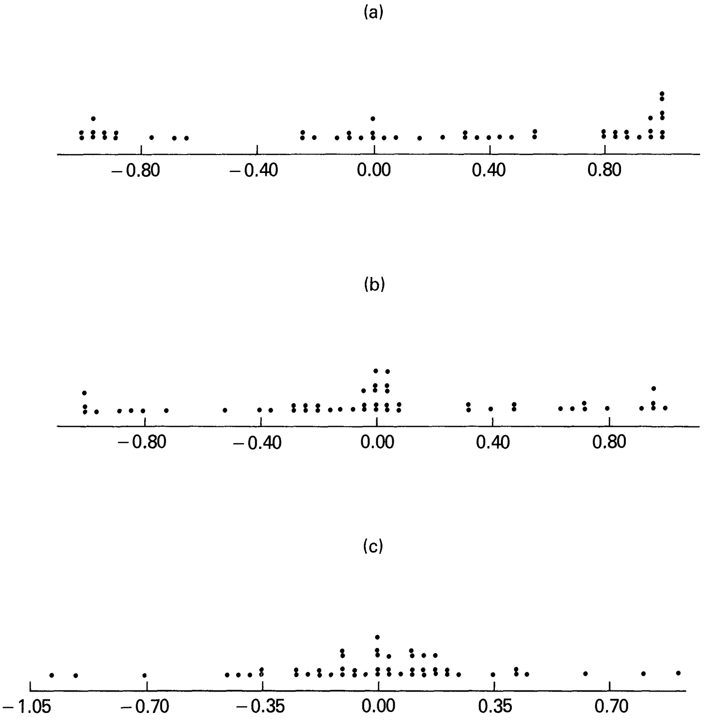


**Fig. 1.** Dotplots of the performance measure, PM, for experiment IIIb. (a) one step ahead, (b) two steps ahead, (c) three steps ahead.

As can be seen, the results do not suggest that there is a substantial improvement to be made when using a SETAR model for forecasting, even when we use the exact SETAR model that has generated the data.

There is, however, a simple explanation of what might at first seem to be an anomaly. It is simply that we do indeed have the MMSE forecasts, but our criterion for comparing these with the linear AR forecasts is based on *absolute error*. The latter is not necessarily minimised by the same forecast, although the two forecasts will be equal when the predictive distribution has equal mean and median. When we have a SETAR model with Gaussian noise, this will always be the case for one step ahead, but will not in general be true for more than one step.

If we use the conditional median instead of the conditional mean i.e. the minimum mean absolute deviation (MMAD) forecast instead of minimum mean square forecast, then it is easily seen that any other forecast based on the same information cannot have smaller absolute error in more than 50% of the ensembles passing through the same state at time $t$. In other words, if $X_t(m)$ is the median of the distribution of $X_{t+m}$ given $\mathbf{X}_t = \mathbf{x}_t$ and $FCT(\mathbf{x}_t)$ is any other forecast, then

$$P(|X_{t+m} - X_t(m)| > |X_{t+m} - FCT(\mathbf{x}_t)|\,|\mathbf{X}_t = \mathbf{x}_t) < \tfrac{1}{2}$$

The conditional median has other nice properties, such as (i) the conditional distribution need not have any moments and (ii) it is invariant under monotonic transformations of the data i.e. If we wish to forecast $Y_t = T(X_t)$, then the conditional median of $Y_t$ is given by $T(X_t(m))$ when $T(\cdot)$ is a monotonic transformation.

We do have some preliminary results from a simulation study of the performance of the MMAD forecast for the 50 series simulated from the SETAR models of the form (3.1). These are along the same lines as for the MMSE results reported in section 3 and are contained in Table 2. These results do not show a significant improvement over those obtained previously using the MMSE criterion.

**Table 2.** Percentage of times absolute forecast error is smaller for nonlinear than for linear model in MMAD forecasts

| | No. of steps ahead | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| Method of fit | | | |
| Least squares[1] | 66 | 58 | 64 |
| LRT[2] | 71 | 57 | 60 |
| cusum test[3] | 60 | 47 | 62 |

[1] SETAR(2; 1, 1) fitted by least squares to all 50 series.

[2] LRT first, then SETAR(2; 1, 1) fitted by least squares: 42 series selected as being nonlinear by this method.

[3] Full cusum procedure, with KMAX=3 and up to 4 thresholds, which yielded 45 out of the 50 as nonlinear. Linear AR fitted using BIC with maximum order equal to the number of fitted SETAR parameters.

## 4 Conclusions

We have developed a package for the identification, estimation and forecasting of SETAR models. Some preliminary results on the performance of the MMSE forecast in comparison with that of a linear AR model have been obtained, and it has been seen that on the basis of absolute error, the performance is not good. However, since in general the MMSE forecast and the MMAD forecast will not be the same, this is only to be expected. We do not universally advocate the use of the MMAD criterion although it does seem to have some advantage over MMSE including the fact that for nonlinear models the usual computational advantage of the latter disappears. On the other hand, when we used the MMAD forecast, the improvement was seen to be marginal. Indeed, from some preliminary results of a further study, it appears that both criteria produce similar forecasts for a wide selection of first order SETAR models, and thus at present it looks as though the linear model is doing much better than expected. We need to do a much more detailed investigation before we can reach any concrete conclusions and the results of this will be published elsewhere.

## Acknowledgement

We are most grateful to Dr Mike Wadsworth of the Mathematics department at Salford University for developing and coding the algorithm for solving the forecasting equation (2.3).

## References

JONES, D.A. (1976) *Ph.D. Thesis* [Unpublished] University of London.

JONES, D.A. (1978) Nonlinear autoregressive processes. *Proceedings of the Royal Society of London A360*, pp. 71–95.

PEMBERTON, J. (1987) Exact least squares multi-step prediction from nonlinear autoregressive models. *Journal of Time series Analysis*, 8, pp. 443–448.

PETRUCCELLI, J.D. & DAVIES, N. (1986) Detecting nonlinearity in time series, *Biometrika*, 73, pp. 687–694.

TONG, H. (1983) Threshold models in nonlinear time series analysis, *Lecture Notes in Statistics 21* (Berlin, Springer Verlag).

TONG, H. & MOEANADDIN, R. (1987) On multi-step non-linear least squares prediction, Paper presented at the IoS conference, Cambridge, July 1987.