Jour. Ind. Soc. Ag. Statistics
53(2), 2000 : 151-160

Threshold Autoregressive Model for a Time Series Data

C. Kesavan Nampoothiri and N. Balakrishna
Cochin University of Science and Technology, Cochin-682022
(Received : May, 1999)

SUMMARY

In this paper we try to fit a threshold autoregressive (TAR) model to time series data of monthly coconut oil prices at Cochin market. The procedure proposed by Tsay [7] for fitting the TAR model is briefly presented. The fitted model is compared with a simple autoregressive (AR) model. The results are in favour of TAR process. Thus the monthly coconut oil prices exhibit a type of non-linearity which can be accounted for by a threshold model.

Key Words : Autocorrelation functions, Coconut oil, Threshold autoregressive model.

1. Introduction

Coconut is one of the most important crops cultivated by a large number of farmers in Kerala, India. Coconut oil (a major product of coconut) is a vegetable oil used in every household for culinary and toiletry purposes. Since the production of coconuts involves large investments, stability in price is needed for the development of this crop. But the coconut oil prices undergo wide and violent fluctuations. These fluctuations can generally be attributed to various factors. Since the observed prices arise in a time sequence, it is possible that the consecutive observations are dependent. In this study, a time series model based approach has been tried to explain the variations other than trend and seasonal variations.

Box and Jenkins [1] provided a method for constructing a time series model in practice. Using this method the current value of the series can be linearly expressed as a function of its previous values. But in practice, non-linearity can often be detected in the underlying structure of the time-series.

There are several non-linear models introduced by Tong [4]. Here we consider only the piece-wise models, that is, the models for which discontinuities arise resulting in the change of regimes. Because of the complexities of the method introduced by Tong, it is not widely used in practice. Tsay [7] proposed a simultaneous method for testing the non-linearity and the identification of the delay parameter. In this paper, a threshold autoregressive (TAR) model is tried for the monthly coconut oil prices using the method proposed by Tsay and compared it with that of a simple autoregressive model. In the following sections the TAR modelling technique is briefly discussed followed by the results and discussions.

2. *Methodology*

A class of threshold autoregressive models can be defined by

$$Z_t = a_0^{(j)} + \Sigma a_i^{(j)} Z_{t-1} + \xi_t^{(j)} \text{ if } r_{j-1} \le Z_{t-d} \le r_j \tag{2.1}$$

where $j = 1, 2, \ldots k$ is the number of regimes being separated by $(k-1)$ threshold values $(-\infty \le r_0 < r_1 < \ldots < r_k = +\infty)$, d is the delay parameter $(d \le p)$, $(a_0^{(j)}, a_i^{(j)})$ $i = 1, 2, \ldots p$ and $j = 1, 2, \ldots, k$ are model parameters of the regime j and $(\xi^{(j)}, j = 1, 2, \ldots, k)$ is a sequence of independent normal variables with mean zero and variance $\sigma^2_{\xi t}$. Such a process partitions the one-dimensional Eucleadian space into k regimes. The model (2.1) is non-linear when there are at least two regimes with different linear models. For these models, the autoregressive coefficients change over time, changes which are determined by comparing the previous values (back shifted by a delay parameter d) to fixed threshold values. The order of AR models may also change over regimes.

The first step in the Tsay's procedure is to test for non- linearity. Consider a TAR model with only two regimes. Let the delay parameter be 'd' and threshold value be r_1, then the model can be written as

$$\begin{aligned} Z_t &= a_0^{(1)} + \Sigma a_i^{(1)} Z_{t-i} + \xi_t^{(1)} \text{ if } Z_{t-d} \le r_1 \\ &= a_0^{(2)} + \Sigma a_1^{(2)} Z_{t-i} + \xi^{(2)} \text{ if } Z_{t-d} > r_1 \end{aligned} \tag{2.2}$$

where $t \in (p+1, \ldots, n)$, n is the number of observations. Now arrange the observations in ascending order. Let π_i be the time index of the i^{th} smallest observation, then the model can be rewritten as

$$\begin{aligned} Z_{\pi i+d} &= a_0^{(1)} + \Sigma a_i^{(1)} Z_{\pi i+d-i} + \xi_{\pi i+d}^{(1)} \quad \text{if } i \le s \\ &= a_0^{(2)} + \Sigma a_i^{(2)} Z_{\pi i+d-i} + \xi_{\pi i+d}^{(2)} \quad \text{if } i > s \end{aligned} \tag{2.3}$$

where $i \in (p+1-d, \ldots n-d)$ and $Z_{\pi s} < Z_{\pi s+1}$. The arranged autoregression grouped the data points into two, so that all the observations in one group follow the same AR process.

The non-linearity can be tested using the classical F-statistics (Tsay [7]) corresponding to the predictive residuals $\xi_{\pi i+d}$ (recursively estimated) of the arranged autoregresion on the regressors $(1, Z_{\pi i+d-1}, \dots, Z_{\pi i+p-d})$ $(i = p+1-d, \dots, n-d)$. The recursive estimate of the parameters were computed by the algorithm developed by Ertel and Fowlkes [2]. The recursion starts with $m = \left(\frac{n}{10}\right) + p$ observations so that there are $(n-p-m)$ predictive residuals available. Consider the least squares regression

$$\xi_{\pi i+d} = w_0 + \Sigma w_v Z_{\pi i+d-v} + \eta_{\pi i+d} \tag{2.4}$$

for $i = m+1, \dots, n-p$ and compute the statistic

$$F(p,d) = \frac{\left[\dfrac{\Sigma\xi_t^2 - \eta_t^2}{(p+1)}\right]}{\left[\dfrac{\Sigma\eta_t^2}{(n-1-2p-m)}\right]} \tag{2.5}$$

follows approximately an F-distribution with $(p+1)$ and $(n-1-2p-m)$ degrees of freedom. The significance of the F-distribution detects the presence of non-linearity. Now for fixed 'p' and 'd', the value of 'd_p' is chosen from the values $(1, 2, \dots p)$ as

$$d_p = \max_{1 \le \delta \le p} (F(p,\delta))$$

where $F(p,\delta)$ is the statistic as defined by (2.5). After obtaining the delay parameter the next step is to obtain the values of the thresholds. To obtain these values, use the scatter plots of the t ratios of the recursive estimates of the AR coefficients versus Z_{t-dp}. The t ratio starts to turn at the threshold values, otherwise it would smoothly converge to a fixed point as the number of observations in the recursive estimation increases. After identifying AR order 'p' and the delay parameter 'd_p' we can refine the order for each regime and the value of the threshold based on the Akaike's Information Criterion (AIC). The AIC is calculated for different values 'p' and for each potential threshold values. This criterion is defined as

$$AIC(p) = N\log(RSS/N) + (p+1)\log(N)$$

RSS is the residual sum of squares, N is the effective number of observations and p is the number of independent parameters of the model.

3. Analysis of the Monthly Coconut Oil Prices

The monthly coconut oil prices show an upward trend during the period 1978-96 (Fig-1). Apart from the sharp increase in prices, fluctuations in the prices of coconut oil within the year have also been seen. Generally speaking, the highest price in any year was received during the period November to January whereas the price tended to decline during the peak monsoon months (June- July) or the months immediately preceding it. These seasonal fluctuations in prices may be due to the seasonality in the production of the coconuts. About 60 percent of the nuts were harvested during the first six months of the year. Because of the abundant availability of oil in the market, the prices are generally low during the summer months and during the early periods of monsoon. During the second half of the year, the production of the coconuts and the availability of copra are low and these cause the rise in the market prices of coconut oil.

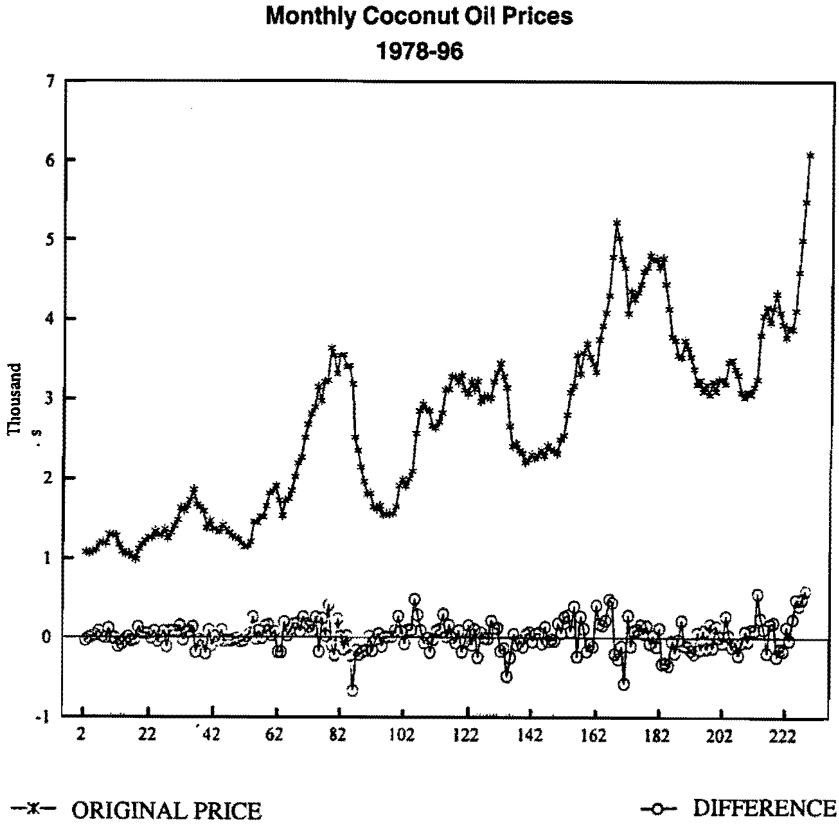

Fig. 1 Monthly coconut oil prices (1978-96)

In this paper we are interested in the univariate time series analysis of the monthly coconut oil prices in the Cochin market. The data consists of 228 monthly observations from January 1978 to December 1996. Since a fairly good estimate of the parameters of the series in obtained only if the series is stationary, we take the first order difference of the prices (i.e., if Z_t is the price then the first order difference is $\nabla Z_t = Z_t - Z_{t-1}$) for further analysis (Fig-1).

Firstly we try to model the monthly coconut oil prices using Univariate Box-Jenkins (UBJ) method and then using the threshold autoregressive (TAR) method. In the UBJ technique a model can be fitted to the data by studying the characteristics of the auto correlation functions (ACF) and partial autocorrelation functions (PACF). After identifying the order and nature of the relationship, the parameters can be estimated. These models are used for short-term forecasting, because most of the autoregressive integrated moving average models place emphasis on recent past rather than the distant past. The ACF and PACF of the differenced series converge to zero reasonably quickly. The cutoff of the PACF after the lag two (Fig. 2) recommends an autoregressive process of order two for the series. Also an examination of the AIC and SSR

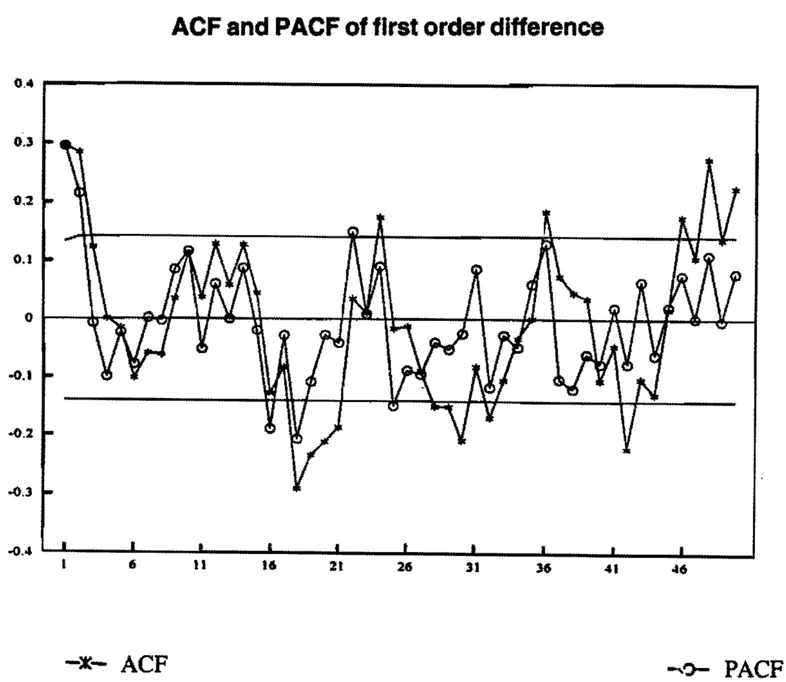

Fig. 2. ACF and PACF of first order difference

for different orders of 'p' and 'q' (Table-1) suggests an AR(2) is more appropriate for the series. After fitting a UBJ model a non-linear model was also tried to get a better representation.

Table 1. Estimates of the parameters of ARMA (p, q)

	p = 1 q = 0	p = 2 q = 0	p = 1 q = 1	p = 2 q = 1	p = 2 q = 2
Constant	22.77 (16.37)	24.89 (21.16)	25.10 (21.31)	24.90 (21.23)	22.36 (16.92)
AR (1)	–0.308 (0.065)	0.243 (0.066)	0.688 (0.136)	0.247 (0.286)	1.446 (0.217)
AR (2)	–	0.236 (0.067)	–	0.235 (0.108)	–0.616 (0.163)
MA (1)	–	–	0.401 (0.166)	0.004 (0.293)	1.193 (0.217)
MA (2)	–	–	–	–	–0.575 (0.129)
AIC	2981	2306	2974	2973	2971
RSS	6592777	6262937	6362139	6243104	6169479
Error Variance	40833	27844	28381	27699	27753

AIC – Akaike Information Criteria
RSS – Residual Sum of Square
AR (p) – Autoregressive process of order p
MA (q) – Moving average process of order q

After selecting the order 'p' of the AR process the next step is to test the non-linearity using the Tsay's statistic (2.5). Here the AR order is p = 2 therefore the delay parameter can take the value either 1 or 2. The recursion starts with 25 observations so that there are 200 predictive residuals. The values of the F-statistics for the delay parameter d = 1 and d = 2 are given in Table-2. Since the F value is maximum for d = 1, we choose the delay parameter as $d_p = 1$. The second step is to locate the threshold value using the scatter plot of the t-ratios of the recursive estimates of an AR coefficient versus ∇Z_{t-dp}. The scatter diagram tells the location of the delay parameter directly. The t ratios of the estimates behave exactly as those of a linear time series before the recursion reaches the threshold value r_i. Once

r_i is reached t ratio begins to deviate. The scatter plot (Fig-3) of the t ratios indicate the possible threshold values are around -100, 40 and 100. Since it needs a minimum of 50 observations for an accurate parameter estimation, we choose the values as -100 and 40, that is, a threshold model with three regimes. There are 51, 74, and 100 observations in the first, second and third regimes respectively. In order to refine the AR order in each regimes and threshold values, we used the Akaike Information Criterion (AIC). Since AIC is minimum for $p = 2$ and for $r_1 = -100$ and $r_2 = 40$ we choose the order of AR as two and the threshold values as -100 and 40. The parameters are estimated for all the models (Table-2). The identified TAR (2, 2, 1) is as follows

$$\begin{aligned} \nabla Z_t &= -45.11 + 0.1088\,\nabla Z_{t-1} - 0.00418\nabla Z_{t-2} + \xi_t^{(1)} \text{ if } \nabla Z_{t-1} \le -100 \\ &= 10.88 + 0.0297\,\nabla Z_{t-1} - 0.12370\,\nabla Z_{t-2} + \xi_t^{(2)} \text{ if } -100 < \nabla Z_{t-1} \le 40 \\ &= 67.69 + 0.3419\nabla Z_{t-1} - 0.3250\,\nabla Z_{t-2} + \xi_t^{(3)} \text{ other wise} \end{aligned} \tag{3.1}$$

Table 2. Estimates of the autoregressive parameters for TAR (2, 3, 1) and AR (2)

	TAR (2, 3, 1) model			AR (2) model
Constant	–45.11	10.88	67.67	24.89
1	0.1088 (0.145)	0.0297 (0.118)	0.3419 (0.092)	0.243 (0.066)
2	–0.00418 (0.146)	–0.123 (0.121)	0.0169 (0.125)	0.236 (0.067)
AIC	528	702	1022	2306
RSS	2168589	1044825	3191737	6262937
Residual Variance	44257	14313	32569	27844

Table 2a . Values of the F-statistic

d	F (δ, d)
1	1.81
2	1.20

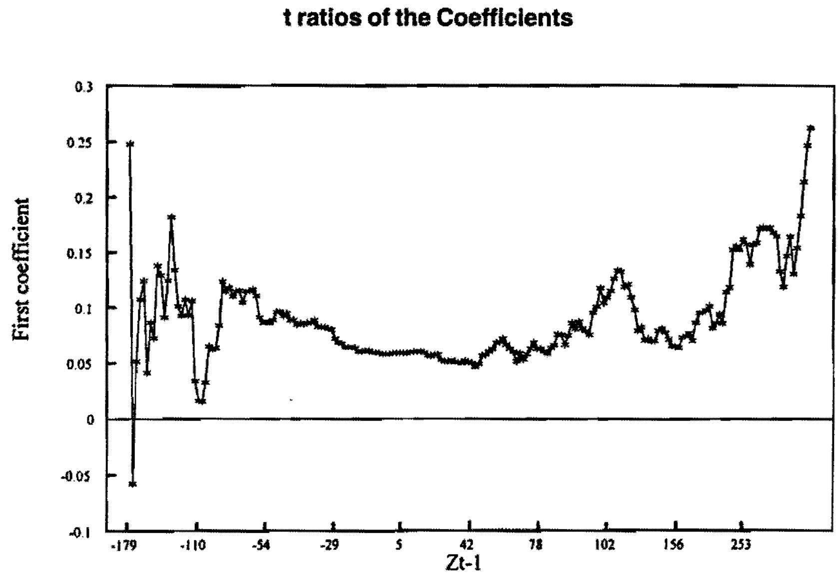

Fig. 3a. t ratios of the Coefficients

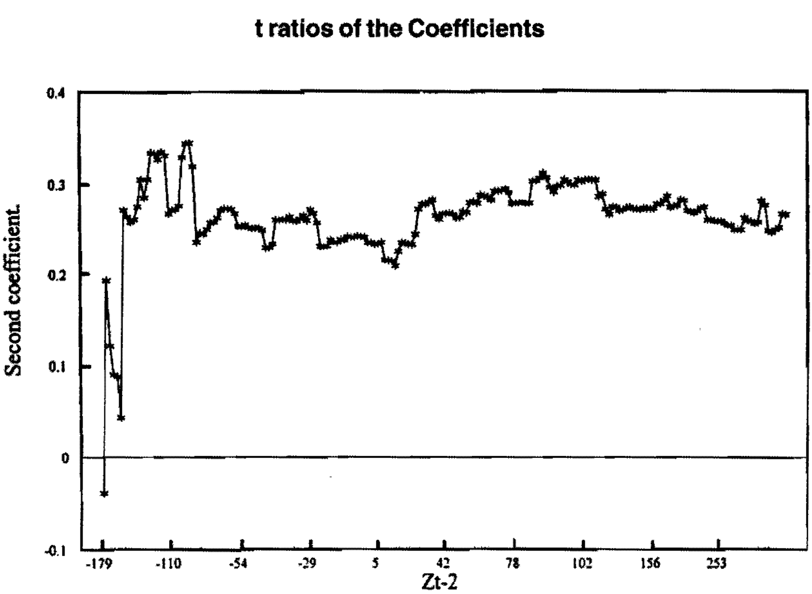

Fig. 3b. t ratios of the Coefficients

In the diagnostic stage, we compute the ACF of the residuals for each models. Most of the ACF are within the 2σ limits shows the independence of the residuals. The sum of squared residuals and the AIC are less for TAR model than that for ordinary AR model. The forecast percent error (Fig-4) is also minimum for TAR model. These observations are in favour of modeling the series by a TAR process.

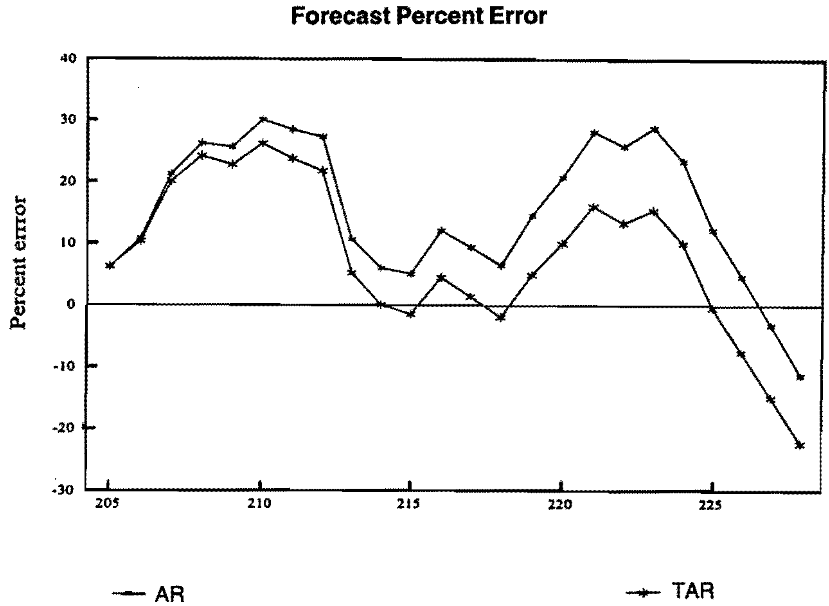

Fig. 4. Forecast Percent Error

4. *Concluding Remarks*

The values of the F-statistic do not show any non-linearity in the series. But other factors like the scatter plot of the t ratios, the values of RSS and AIC are in favour of the TAR process. The percent forecast error for TAR process is lower than that for an AR process. These factors are in favour of modeling the coconut oil prices using TAR process.

Tsay used the scatter plot to identify the threshold values. The pattern of the gradual convergence of the t ratios starts to turn when the recursion reaches the threshold values. Here the scatter diagram starts to turn when the value of difference between the previous two observations reaches around - 100 and 40. Thus the TAR process gives a better fit for time series of coconut oil prices. For an observation Z_t, the model change is identified using the difference of previous two observations.

REFERENCES

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

[2] Ertel, J.E. and Fowlkes, E.B. (1976). Some algorithms for linear spline and piecewise multiple regressions. *J. Am. Statist. Assoc.,* **71,** 640-648.

[3] Jacob, Mathew (1978). Trend and fluctuations in the prices of coconut and coconut oil. M. Phil dissertation, Jawaharlal Nehru University, New Delhi.

[4] Tong, H. (1990). Non-linear Time Series : *A Dynamic System Approach.* Oxford University Press, New York .

[5] Tong, H. and Lim, K.S. (1980). Threshold autoregression, limit cycles and cyclical data (with discussion). *J. Roy. Statist. Soc.,* **B 142,** 245-292.

[6] Tsay, R.S. (1986). Non-linearity tests for time series. *Biometrica,* **73,** 461-466.

[7] Tsay, R.S. (1989). Testing and modelling threshold autoregressive processes. *J. Am. Statist. Assoc.,* **84,** 231-240.

[8] Watier, L. and Richardson, S. (1995). Modelling of an epidemiological time series by a threshold autoregressive model. *The Statistician,* **44,** 353-364.