Application of Threshold Autoregressive Model: Modeling and Forecasting Using U.S. Exports of Crude Oil Data

2 authors, including:

Yakoub Boularouk
Centre universitaire de Mila
12 PUBLICATIONS   **25** CITATIONS

Some of the authors of this publication are also working on these related projects:

Project   Chaos in fractional order systems View project

Project   Stochastical modeling View project

Application of Threshold Autoregressive Model: Modeling and Forecasting Using U.S. Export Crude Oil Data

K . Djeddour [1] , Y . Boularouk [2]

1. Laboratory MSTD, Faculty of Mathematics, USTHB Algiers, Algeria
2. Department of Mathematics and Informatics, University of Mila, Algeria

Corresponding Author Email: kdjaballah@usthb.dz

Abstract:

This work focuses on the specification of the threshold autoregressive model and forecasting. We consider U.S. Export is the amount of oil exported from January 1991 to December 2004. We present threshold models that are special cases of the procedure for non-linear models on average above TAR (threshold autoregressive). This means that we start with a simple model and we use a more complicated model if the diagnostic tests indicate that the model obtained is not satisfactory. We will use this procedure to compare an approach of ARMA models and approach the nonlinear threshold for the series. Between the two methods, the prediction threshold autoregressive model is better in the mean square error.

Keyword: Threshold Autoregressive Model, Nonlinearity, test, TAR models, forecasting

1. Introduction

Over millions of years, organic life has mingled with the earth. Thus was born the oil. In less than two centuries, man has learned to explore and exploit the fossil resource. Our daily is now largely dependent: it is used in chemical production, fuel production, drug production and power generation, etc...

Because of the importance of its production and its oil and gas exports, Algeria has emerged on the world energy markets. Among its market potential include the U.S. oil market. In the light of preliminary studies the U.S. oil market is promising for Algerian oil.

The work was entrusted to us is part of the prospecting of U.S markets, which in our case would be the U.S. exportation of crude oil.

The United States consumed 19.1 million barrels per day of petroleum products during 2010, making us the world's largest petroleum consumer. The United States was third in crude oil production. In 2010 the United States imported 11.8 million barrels per day of crude oil and refined petroleum products. We also exported 2.3 MMbd of crude oil and petroleum products during 2010.

American Journal of Oil and Chemical Technologies;
ISSN (online): 2326-6589; ISSN (print): 2326-6570
Volume 1, Issue 9, November 2013

Because the United States is the world's largest oil importer, it may seem surprising that it also exports about 2 million barrels a day of oil, almost all of it in the form of refined petroleum products. Due to various logistical, regulatory, and quality considerations, it turns out that exporting some barrels and replacing them with additional imports is the most economic way to meet the market's needs. For example, refiners in the U.S. Gulf Coast region frequently find that it makes economic sense to export some of their gasoline to Mexico rather than shipping the product to the U.S. East Coast because lower-cost gasoline imports are available from Europe.

The threshold autoregressive model describes complex dynamic data as an extension to autoregressive models. It is popular in application to nonlinear time series data. The TAR model is first introduced by Tong [7]. And then, Tong and Lim [8] first complete the TAR model's exposition, and give effective technology for the practical issues in application [3]. Then, there are several scholars researched and developed test methods for the TAR model. For instance, Tsay [10] proposed the F test which combined three studies of nonlinearity tests of Keenan [4], Tsay [9], and Petruccelli and Davies [5]. In recently years, many scholars apply the TAR model to analyze real exchange rates, interest rates, and stock return etc.

Such models can be useful for the description of the data and the forecasts. The thresholds models are particularly interesting since they allow to take into account the asymmetry phenomena and outs of high amplitude. The most commonly used models are the SETAR models. These model classes assume the presence of two or more regimes, within which the data require different (linear) models for description and forecasting.

2. Modeling

The time series has long been dominated by linear modeling, and especially by the ARMA models:

$$X_t = \phi_0 + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=0}^{q} \theta_j \varepsilon_{t-j} , \quad \theta_0 = 1 \qquad (1)$$

This class of model has had a quasi-systematic use. Indeed, this type of specification has many advantages: linear difference equations, statistical inference for well-controlled linear Gaussian process, minimum variance estimates, practical tool for prediction and control.

However, this representation is limited insofar as the linearity restricts the possible dynamics.

If we abandon the assumption of linearity and that we accept nonlinear functions of past and present values, then the past may contain useful information for the future.

Following areas that are of interest, the class of linear ARMA models may not work. The series are indeed characterized by nonlinear dynamics.

2.1. The Threshold Autoregressive Model

We consider the case when there are two or more regimes, and the process is a pth order autoregression in each. This is the threshold models whose aim is to model phenomena of asymmetry in the cycles. The underlying idea is that a relationship may be nonlinear over the

whole period considered, but linear sub-periods. TAR model is formed by pieces of linear relations. It is a model with k equations of the form:

$$
X_t = \begin{cases} \phi_0^{(1)} + \sum_{i=1}^{p_1} \phi_i^1 X_{t-i} + \varepsilon_t^{(1)} & si\ X_{t-d} \prec c_1 \\ \phi_0^{(2)} + \sum_{i=1}^{p_2} \phi_i^{(2)} X_{t-i} + \varepsilon_t^2 & si\ c_1 \leq X_{t-d} \prec c_2 \\ \qquad \cdots\cdots\cdots \\ \phi_0^k + \sum_{i=1}^{p_K} \phi_i^k X_{t-i} + \varepsilon_t^k & si\ X_{t-d} \geq c_{k-1} \end{cases} \quad .(2)
$$

In this model, k is the number of regimes. We see that each equation represents a linear AR model of order p_i, where i = 1,…, k. The parameter d is a positive integer called delay and the coefficients c_j, j = 1,…, k-1, are the threshold parameters for which the system switches from one regime to another.

The model requires estimation of several parameters: the number of regimes (that is to say the number of thresholds), the delay parameter d, the threshold values for j = 1,…,k-1, and the autoregressive coefficients on each regime.

Nonlinear time series models open a new door for estimation and forecasting this kind of economic time series data. There is a typical nonlinear time series model (the threshold autoregressive (TAR) model) which is easy to do modeling with regime-switching data.

2.3. Linearity test

Assume that a linear series is tantamount to considering that it is characterized by a single dynamic. To test the linearity of the series we test the null hypothesis of linearity against the alternative hypothesis of the existence of a threshold model. To simplify the presentation of the test, consider a process TAR (p) with two regimes. We then test:

$$
H_0: \quad X_t = \phi_0 + \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t\ , \quad \theta_0 = 1
$$

Versus

$$
H_1: \quad X_t = \begin{cases} \phi_0^{(1)} + \sum_{i=1}^{p} \phi_i^{(1)} X_{t-i} + \varepsilon_t^{(1)} & si\ X_{t-d} \prec c \\ \phi_0^{(2)} + \sum_{i=1}^{p} \phi_i^{(2)} X_{t-i} + \varepsilon_t^2 & si\ X_{t-d} \geq c \end{cases} \quad (3)
$$

With $X_t = (X_{t-1},…,X_{t-p})$ and $\phi^{(j)} = (\phi_1^{(j)},…,\phi_p^{(j)})$ for j = 1,2.

2.2. Tsay's Approach

Tsay [10] transforms the problem of test for breakpoints using the concept of arranged autoregression and uses the predictions to build residual test statistics that do not involve the parameters undefined.

Tsay [10] proposed a very interesting graphical approach for detecting the number and location of the thresholds. He transform the testing problem into detecting change points using the concept of arranged auto regression and employ predictive residuals to construct test statistics that do not involve undefined parameters.

Observations need to be sorted according to the threshold values from the smallest observation to the largest observation.

Let $(Y_t, 1, Y_{t-1}, \ldots, Y_{t-p})$, a set of observations where $t = p + 1, \ldots, n$. For the threshold variable Y_{t-d}, there exist two situations. When $d \leq p + 1$, the threshold variables are $(Y_{p+1-d}, \ldots, Y_{n-d})$. On the other hand, when $d > p + 1$, the threshold variables are $(Y_1, \ldots, Y_{n-d})$. Therefore, we combine two situations together: threshold variables $(Y_h, \ldots, Y_{n-d})$, where $h = \max(1, p + 1 - d)$. We sort them by a new time index π_i which express new order from the ith smallest observation in the set $(Y_1, \ldots, Y_{n-d})$. Therefore, $i = 1, 2, \ldots, n-d-h+1$ and $n - d - h + 1$ is the effective sample size. Here, we use $Y_{\pi i}$ instead of Y_{t-d} to show the threshold variable. For example, if the tenth observation in $(Y_h, \ldots, Y_{n-d})$ is the smallest, then $\pi_i = 10-d$.

And then, the model (3) can be arranged as follow:

$$X_{\pi_i+d} = \begin{cases} \phi_0^{(1)} + \sum_{v=1}^{p} \phi_v^{(1)} X_{\pi_i+d-i} + \varepsilon_{\pi_i+d}^{(1)} & si \ X_{\pi_i} \prec c_1 \\ \phi_0^{(2)} + \sum_{v=1}^{p} \phi_v^{(2)} X_{\pi_i+d-i} + \varepsilon_{\pi_i+d}^{(2)} & si \ X_{\pi_i} \geq c_2 \end{cases} \qquad (4)$$

So the threshold variables X_{π_i} which are smaller than c_1 will fit to the first equation, while the threshold variables X_{π_i} which are larger than c_2 will fit to the second equation. And then, Tsay uses recursive least squares (RLS) estimates of ϕ_v in model (4) to calculate the F-statistic for testing the threshold nonlinearity. The RLS estimates are calculated as follows:

$$\hat{\beta}_{m+1} = \hat{\beta}_m + K_{m+1} \left[Y_{m+1} - x'_{m+1} \hat{\beta}_m \right],$$

$$D_{m+1} = 1 + x'_{m+1} P_m x_{m+1},$$

$$K_{m+1} = P_m x_{m+1} / D_{m+1},$$

and

$$P_{m+1} = (I - P_m \frac{x_{m+1} x'_{m+1}}{D_{m+1}}) P_m$$

American Journal of Oil and Chemical Technologies;
ISSN (online): 2326-6589; ISSN (print): 2326-6570
Volume 1, Issue 9, November 2013

With $\hat{\beta}_m$, estimated by ordinary least squares for the m first data and Pm the inverse matrix of XX^t associated to x_{m+1} and Y_{m+1} represent the vector of regressors and the dependent variable of the next observation, respectively.

$$\hat{a}_{m+1} = Y_{m+1} - x'_{m+1}\hat{\beta}_m$$
$$\hat{e}_{m+1} = \hat{a}_{m+1}\big/\sqrt{D_{m+1}}.$$

With $\hat{a}_{m+1}$ the recursive residuals and $\hat{e}_{m+1}$ recursive standardized residuals. One way to test the nonlinearity of regression residuals is normalized according to the observations (4) on the regressors Y_{m+1}:

$$\hat{e}_{m+1} = w_0 + \sum_{v}^{p} w_v Y_{m+1} + \mu_{m+1}$$

Based on the model (4), the null hypothesis tested is: $\phi_i^{(1)} = \phi_i^{(2)}$ (1) for i = 0,…,p. And then, Tsay computes F-statistic for testing threshold nonlinearity as follows:

$$Q\,(p,\,d)= \frac{\left(\ \sum_t \hat{e}_t^2 - \sum_t \hat{u}_t^2\ \right)}{\sum_t \hat{u}_t^2} * \frac{T-k-2p-1}{p+1} \qquad (5)$$

Under the null hypothesis of linearity, the Q statistic follows a Fisher with (p+1, T-k-2p-1) degrees of freedom. If the null hypothesis is rejected, the procedure continues and is retained as the threshold variable (i.e as the value of d) that maximizes the test statistic of Tsay.
He proposes to determine the threshold value by a graphical analysis. This method is very approximate because the graphical observation is tricky. The threshold model is linear until it reaches the threshold ensures that the regime change. It then passes to a second regime. Therefore, one need only look at the graphs (coefficients, residuals) for any value of the variable transition takes place this sudden change. The estimated TAR model is obtained by standard regression methods, knowing the threshold and the delay parameter.

3.   Application of the threshold models to behaviors of U.S Export oil

In this section, we will use U.S. Exports data to do modeling and forecasting.

3.1. Data
We apply the estimation method for nonlinear models in the series: "Export": quantities of oil exported from January 1991 to December 2004. The unit is 1,000 barrels / day. It is denoted by EXP.
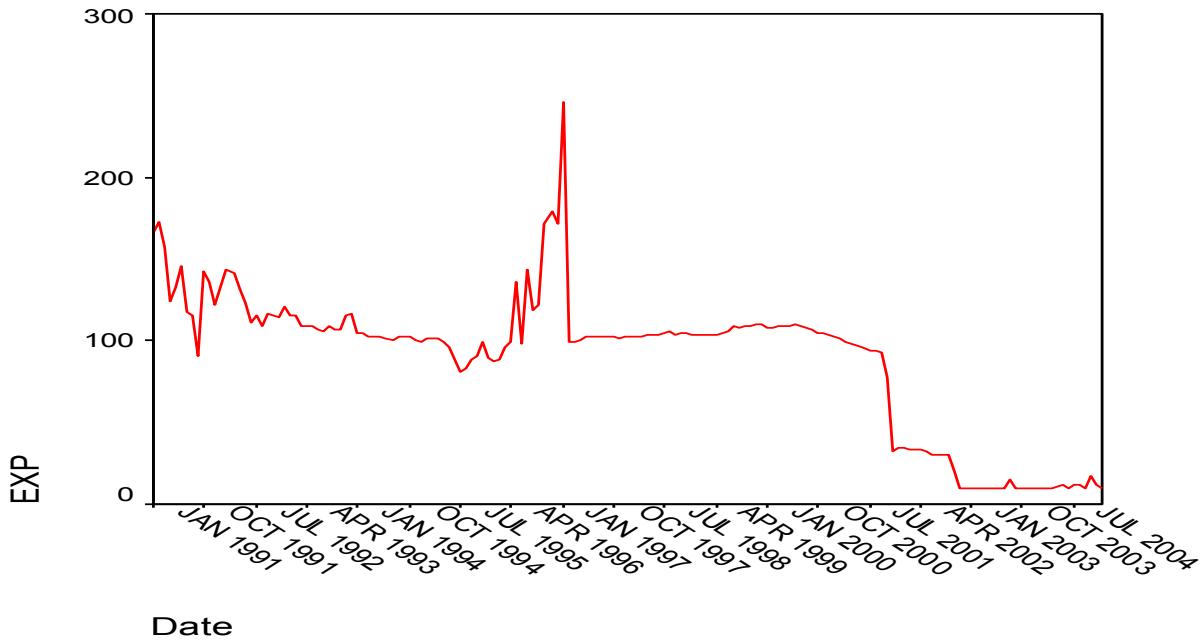
Figure 1. EXP Series

The application of the estimation method for nonlinear models best known requires the stationary of the series (EXP). Given the large variability of the series, we begin with a logarithmic transformation of the series and denote the resulting series EXPL. The use of the log transform is widespread, primarily to reduce the asymmetry of the series.
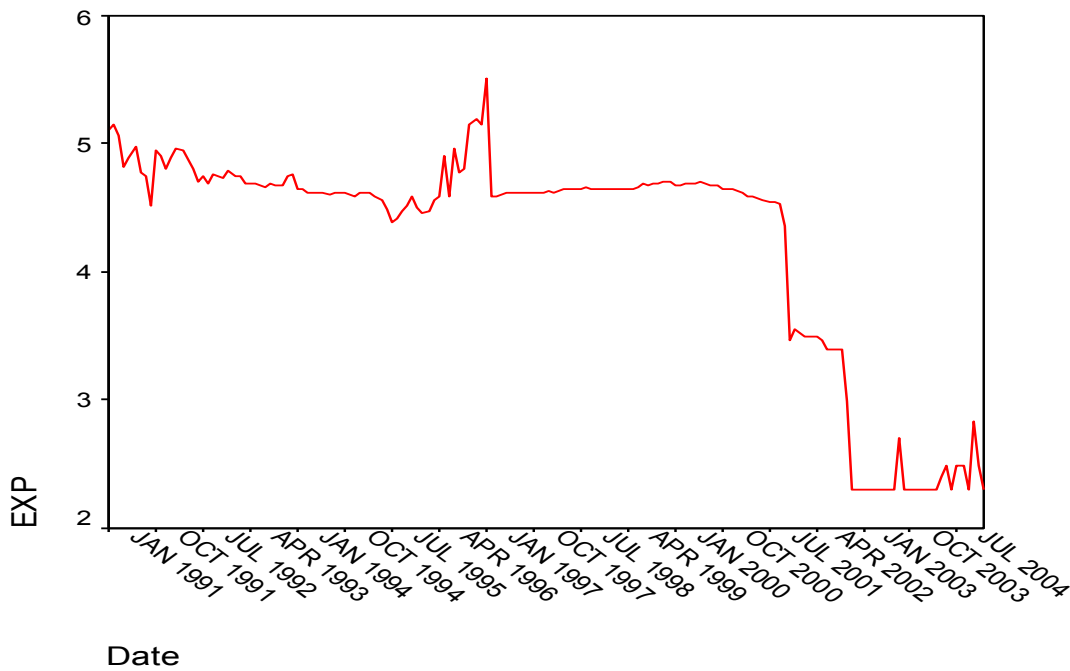

Figure 2. EXPL Series

The unit root tests conducted showed that our variable is integrated of order 1. Differentiation can make it stationary. This ensures that the variables studied are stationary.
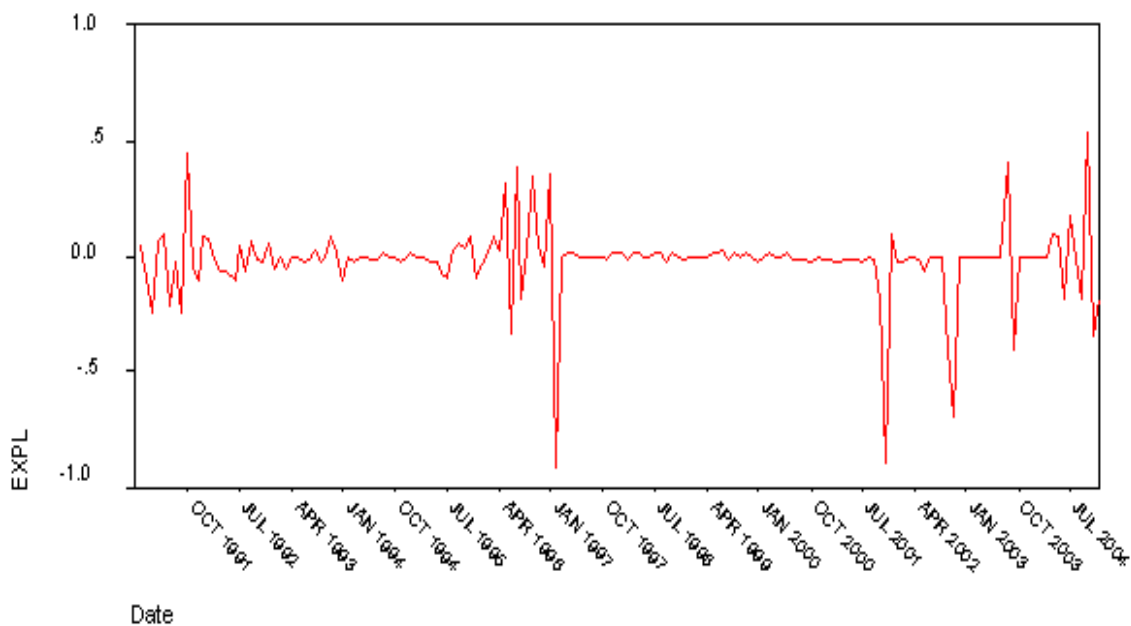

Figure 3. DEXPL Differentiated series

The first observation one can make, is that the series is stationary, but is also observed disruptions of high amplitude. These breaks suggest a type of nonlinearity (threshold) of the series. And we detect three regimes in the curve of exports, corresponding to December 2001 and January 1997.

3.2. Results:

3.2.1. Choose Order p:

For modeling the TAR model, we firstly select the order p via autocorrelation function (ACF) and partial autocorrelation function (PACF). According to order p, we usually make the range of d, which is d< p. For all possible lags d, the number of possible (p,d) is p. And then, we can calculate the test statistic Q, p times. If we reject the null hypothesis of linearity, it is possible to choose lag d when the maximum F statistic is obtained. That means we choose the lag d when the P-value of Q is minimum. Generally, we assume that d is no more than p in model. For a given AR order p. First, ACF & PACF help us to determine order p:

Table 1: Correlogram of DEXPL

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.196	-0.196	6.5566	0.010
		2	0.005	-0.035	6.5611	0.038
		3	0.058	0.054	7.1402	0.068
		4	-0.111	-0.092	9.2709	0.055
		5	0.030	-0.009	9.4249	0.093
		6	0.077	0.081	10.472	0.106
		7	-0.139	-0.104	13.873	0.053
		8	0.118	0.066	16.334	0.038
		9	-0.099	-0.075	18.081	0.034
		10	-0.046	-0.057	18.463	0.048
		11	0.120	0.076	21.057	0.033
		12	0.157	0.230	25.562	0.012
		13	-0.040	0.037	25.849	0.018
		14	0.112	0.085	28.167	0.014
		15	-0.048	0.028	28.593	0.018
		16	-0.021	-0.016	28.679	0.026
		17	-0.051	-0.094	29.163	0.033
		18	-0.042	-0.064	29.503	0.043
		19	0.033	0.026	29.707	0.056
		20	-0.004	-0.020	29.710	0.075
		21	-0.101	-0.055	31.672	0.063
		22	0.100	0.065	33.618	0.054
		23	-0.068	-0.062	34.514	0.058

After p = 1, the PACF function is falling down rapidly. We check that the residuals are white noise. Since high order AR model can fit nonlinear dynamics well, we try lower order in nonlinear model (p = 1). We also consider an AR (2) model, we retain the AR (1).

3.2.2. Nonlinearity Test and Selecting the Delay Parameter *d*
The first step in the specification of these models is to check that the model is characterized by nonlinear dynamics. We will therefore apply the Tsay's test to verify the linearity of the series.

$\sum e_t^2 = 0.2902$ $\qquad$ $\sum u_t^2 = 0.2254$ $\quad$ with $\quad$ T=167 $\quad$ k=18.7 $\quad$ ((T-k-2p-1)/ (p+1)) =47.76.

Then **Q = 13.7403 > $F_{0,05}$(2; 144) = 3**

We reject the null hypothesis of linearity. Then we consider there is threshold nonlinearity.

3.2.3. Locating Threshold Value

We identify the threshold value using scatter plot of t-statistic of the recursive least squares. For do this, we have drawn two graphs:

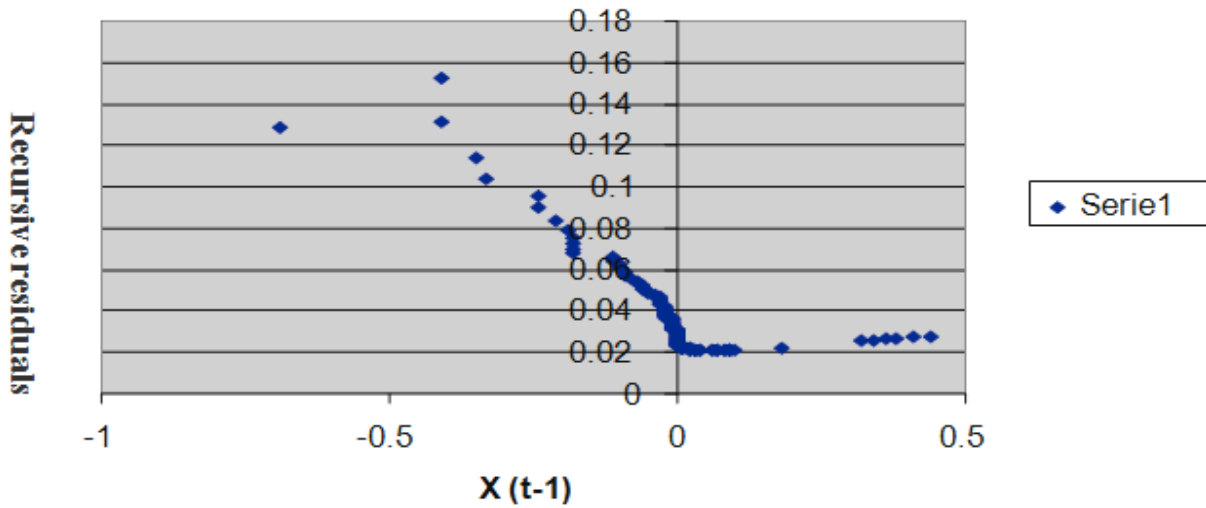- The first is the x-axis variable ordered, and on the ordinate axis recursive residuals.

American Journal of Oil and Chemical Technologies;
ISSN (online): 2326-6589; ISSN (print): 2326-6570
Volume 1, Issue 9, November 2013

Figure 4. Graph of Recursive Residuals

- The second is related to the x-axis variable ordered, and on the ordinate axis the recursive autoregressive coefficients estimated.
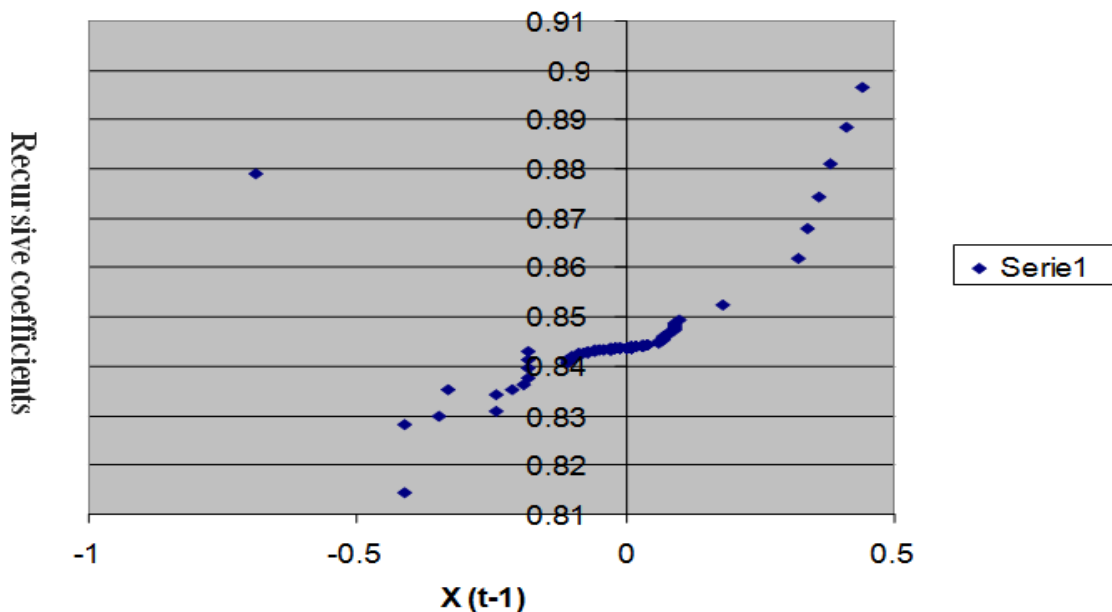

Figure 5. Graph of recursive coefficients

From the second graph, we locate two thresholds $c_1 = -0.09$ and $c_2 = 0.06$. We identify three regimes in the curve of exports, corresponding to December 2001 and January 1997.
We identify threshold value $c_1 = -0.09$ clearly, because there is an obvious jump around December 2001. The plot also shows there is a jump around January 1997.

American Journal of Oil and Chemical Technologies;
ISSN (online): 2326-6589; ISSN (print): 2326-6570
Volume 1, Issue 9, November 2013

3.3. Estimation:

We obtain the following estimated models by using the results above. We get model with three-regime and the two threshold values are $c_1 = -0.09$ and $c_2 = 0.06$. The estimation gives:

$$X_t = \begin{cases} \underset{(24,27760)}{0{,}841983}\, X_{t-1} + \varepsilon_t^{(1)} & si\ X_{t-1} \prec -0{,}09 \\[1em] \underset{(85,05554)}{0{,}944523}\, X_{t-1} + \varepsilon_t^{(2)} & si\ -0{,}09 \leq X_{t-1} \prec 0{,}06 \\[1em] \underset{(14,867545)}{0{,}645453}\, X_{t-1} + \varepsilon_t^{(3)} & si\ \ X_{t-1} \geq 0{,}06 \end{cases}$$

Furthermore, a study conducted on these same data using the methodology of Box and Jenkins found that EXPL had behavior ARIMA (0, 1, 1) with the following model:

(1-B) EXPL$_t$ = (1-0,18B) ε_t

The root of mean squared error is RMSE; as shown in table below.

Table 2. RMSE

	Model with three regimes	ARIMA(0,1,1)
RMSE	0.09441198	0.20456716

In terms of RMSE, nonlinear modeling is better than linear modeling.

3.4. Forecasting:

Table 3. Forecasts of the EXP series in 12 months of the year

Jan	Feb	Mar	Apr	May	Jun
14.18861	10.73681	10.95995	11.71913	11.32981	10.89427

Jul	Aug	Sep	Oct	Nov	Dec
11.67006	11.24347	11.10500	11.49689	11.27125	11.18646

The following graph represents the realization and the fitted values:
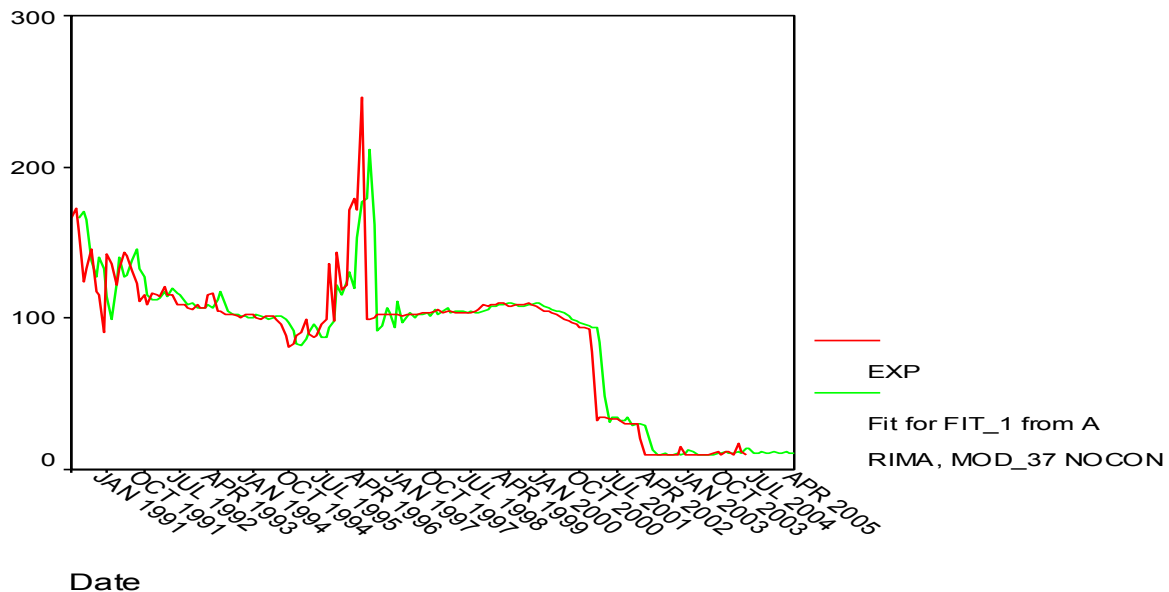
Figure 6: Fitted values

4. References:

[1] G.E.P. Box & G.M. Jenkins (1976). *Time series analysis, forecasting and control*. Holden-Day, San Francisco, (revised edition).

[2] Ertel, J. E., and Fowlkes, E. B. (1976). *Some Algorithms for Linear Spline and Piecewise Multiple Linear Regression*. Journal of the American Statistical Association, 71, 640-648.

[3] Geweke, J., (2007). *The SETAR Model of Tong and Lim and Advances in Computation*. In: Chan, K.S., ed. (2009). Exploration of a Nonlinear World: An Appreciation of Howell Tong.s Contributions to statistics. World Scienti.c Publishing, pp.85-94

[4] Keenan D.M A Tukey *non additivity type test for time series non linearity*. Biometrika. 72 (1) 39-44. (1985)

[5] Petruccelli, J. and Davies, N. (1986), *A Portmanteau test for self-exciting threshold Autoregressive-type nonlinearity in time series*. Biometrika 73, 687-694.

[6] Tiao, G.C. and G.E.P. Box (1981). *Modeling multiple time series with applications*. Journal of the American Statistical Association 76, 802-816. 18

[7] Tong, H. (1978). *On a Threshold Model in Pattern Recognition and Signal Processing*, ed. C. H. Chen, Amsterdam: Sijho¤&Noordhoff.

[8] Tong, H., and Lim, K. S. (1980). *Threshold Autoregression, Limit Cycles and Cyclical Data* (with discussion). Journal of the Royal Statistical Society, Ser. B, 42, 245-292.

[9] Tsay, R. S. (1986). *Nonlinearity Tests For Time Series*. Biometrika, 73, 461-466.

[10] Tsay, R.S., (1989). *Testing and modeling threshold autoregressive processes*. Journal of the American Statistical Association, 84(405), 231-240.