Journal of Engineering Technology and Applied Sciences, 2022 *e-ISSN:* 2548-0391
Received 8 April 2022, Accepted 2 June 2022 *Vol. 7, No. 2, 115-124*
Published online 29 August 2022, Research Article *doi: 10.30931/jetas.1099407*

Citation: Kahraman, U. M., Arsel, I., Genc, A., Oturanc, G. "Estimation of Wind Speed Data with Setar Model". Journal of Engineering Technology and Applied Sciences 7 (2) 2022 : 115-124.

ESTIMATION OF WIND SPEED DATA WITH SETAR MODEL

Ümran Münire Kahraman[a], İsmail Arsel[b*], Aşır Genç[c], Galip Oturanç[b]

[a] *Department of Political Science, Necmettin Erbakan University, Konya, Turkey*
uteksen@gmail.com
[b*] *Department of Physics, Karamanoğlu Mehmetbey University, Karaman, Turkey*
*ismailarsel@kmu.edu.tr (*corresponding author), ismail.arsel@gmail.com*
[b] *Department of Mathematics, Karamanoğlu Mehmetbey University, Karaman, Turkey*
goturanc@gmail.com
[c] *Department of Statistics, Necmettin Erbakan University, Konya, Turkey*
agenc@erbakan.edu.tr

Abstract

The threshold model allows expression with different Autoregressive Moving Average (ARMA) models sorted according to the threshold value of the observations. In this study, nineteen years of observed wind speed data have been modeled with the Self Exciting Threshold Autoregressive (SETAR) model. Two different Autoregressive (AR(3)) models have been obtained for the situation where the wind speed was below and above 2.5 m / s of the previous observation in the time series. In addition, in the SETAR (1,3,3) model, it has been determined that the residual terms have the effect of GARCH (1,1) and a range has been estimated for model predictions.

Keywords: Threshold autoregressive Model, GARCH, wind speed, SETAR prediction, weibull distribution

1. Introduction

There are many researches on wind energy studies in the literature. It is possible to reach many studies on the technologies [1,2,3] and local potentials [4,5,14] of alternative energy resources. Karaman is 328 km away an important international oil transportation line (Bakü (Azerbaijan) -Tbilisi (Georgia) - Ceyhan (Turkey)) in Turkey. In this study, a different statistical analysis of wind energy data of Karaman has been made. Karaman is located at 37.18 north latitude and 33.23 east longitude and is 1039 m above sea level. Although the region is an agricultural area, it has been developing in industry especially in recent years with state support. Especially food industry investments have been increasing in recent years. In parallel with the industrial investments in the region, investments made in alternative energy resources with the support of the state have increased remarkably in recent years. Currently, there are many privately owned solar power plants in the region. In this study, nineteen years of data obtained from General Directorate of Meteorology (MGM) for Karaman were evaluated with the SETAR model.

Wind speed data moves by following per under in accordance with the Weibull probability distribution [14]. In this study, it has been studied to model as a time series instead of estimating Weibull parameters for daily compiled data.

The SETAR model of wind speed is useful in modeling time series in many areas according to the range where the data of the observations allow for different models. Mostly, the SETAR model was created in stock market data and financial data [6,7, 9, 10, 13, 17, 18,19, 21, 25, 29], and it was also applied in biological and medical data ([8, 16, 28, 30]), climate and meteorology data ([20]), traffic data ([32]).

2. SETAR model

SETAR models are among the most widely used nonlinear time series models [27]. The threshold model was first discussed by Tong [23] and later Tong and Lim [24] developed the model.

A SETAR $(d; p_1, p_2, ..., p_k)$ model with k regime is established $Y_t = \phi_0^{(k)} + \sum_{j=1}^{p_k} \phi_j^{(k)} Y_{t-j} + a_t^{(k)}$ [33]. Here, k (k>1) is the number of regimes in the model. $r_{k-1} < Y_{t-d} < r_k$, d > 0 delay parameter and p_i are the order of the autoregressive process in the i th regime of the model. Threshold parameters satisfy limits $(-\infty = r_0 < r_1 < r_2 < ... < r_{k-1} < r_k = \infty)$. Normal independent and identical distribution (iid) random variables $a_t^{(i)}$ in each i th regime have zero mean and $\sigma_i^2 < \infty$ $(i = 1, 2, ..., k)$ constant variance. The superscripts in the model show the regimes. The dynamic behavior of the time series variable in each regime is a linear autoregressive process.

If $p_1 = p_2 = ... = p_k$ and $\phi_i^{(1)} = \phi_i^{(2)} = ... = \phi_i^{(k)}$, $i = 0,1,..., p$ is and then the SETAR model transforms into a linear autoregressive process with p in-line. To test the existence of a nonlinear threshold state, Tsay [22] has developed a test based on the sequential autoregression process. Sequential autoregressions are generally separated by the variable, which Y_{t-d} is the regime indicator of the SETAR model. Here, for each j, one step forward prediction error $\hat{e}_{(j+1)+d}$ can be calculated. Under the assumption that the model has a linear AR (p) process, standardized predictive errors are not only independent and uniformly distributed, but it is also orthogonal to $\{Y_{(j+1)+d-1}, ..., Y_{(j+1)+d-p}\}$ regressors. But if the real model is a nonlinear SETAR process, the orthogonality is broken. Using this feature, the $\underline{e} = \underline{Y}\underline{\beta} + \underline{\eta}$ regression model is created. Here $\underline{e}$ is the $(\hat{e}_{(m+1)+d}, ..., \hat{e}_{(n-p)+d})'$ vector, $\underline{Y}$ is the matrix of regressors $\{Y_{(j+1)+d-1}, ..., Y_{(j+1)+d-p}\}$ with $j = m, m+1, ..., n-p-1$, $\underline{\beta}$ is the p-dimensional parameter vector, $\underline{\eta}$ is the error vector. The F statistic, which is used to test $H_0 : \underline{\beta} = \underline{0}$ the hypothesis in the regression model, can also be used to test orthogonality and SETAR type nonlinearity. The F statistic is calculated with the help of the variance analysis table that tests the significance of the linear regression model. To apply the F test, p and d values must be given. However, in practice, these are generally unknown.

For the selection of p, it is suggested to use the partial autocorrelation function (PACF) of the Y_t series. After determining the p-value, d is chosen which gives the highest F statistic value. The scatter plot of the Y_t series against a Y_{t-d} variable can be examined to determine the location of the regime number k and threshold parameters [27]. In the graph, the regression line obtained by the locally weighted regression (loess) method, where Y_t is the dependent variable, gives information about the number of regimes and the threshold value. The loess method is a non-parametric method created by converting the linearity condition in classical regression methods to flexible [11].

The least-squares estimates $\underline{\phi}^{(j)}$ can be obtained by the classical least squares method for $j = 1,2,...,k$. Hence, it is in the shape.

$$\hat{\underline{\phi}}^{(j)} = (\underline{A}_j^{'} \underline{A}_j)^{-1} (\underline{A}_j^{'} \underline{Y}_j) \qquad (1)$$

After the model is estimated, the ACF plot and Ljung-Box statistic for residuals provide information on whether residuals contain models.

3. Autoregressive (ARCH) models with conditionally variable variance

The model that allows varying variance in time series is a conditionally varying autoregressive (ARCH) model. In the first-order ARCH model,

$$h_t = \omega + \alpha_1 a_{t-1}^2 \qquad (2)$$

in the form. Clearly, the h_t (conditional) variance should not be negative. To guarantee this, the parameters of the ARCH (1) model must satisfy the states $\omega > 0$ and $\alpha_1 \geq 0$. By adding h_{t-1} to the ARCH (1) model, (1,1) ordered generalized ARCH (GARCH) model is obtained.

$$h_t = \omega + \alpha_1 a_{t-1}^2 + \beta_1 h_{t-1} \qquad (3)$$

In this model, for $h_t \geq 0$ to be, the states $\omega > 0$, $\alpha_1 > 0$, and $\beta_1 \geq 0$ must be satisfied. The general form of the GARCH (p, q) model is given as $h_t = \omega + \sum_{i=1}^{q} \alpha_i a_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}$.

ARCH-LM method is based on an autoregressive model for residuals squares. ARCH-LM model is $\hat{\varepsilon}_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \hat{a}_{t-i}^2 + v_t$. Here, q shows the lag length of the autoregressive model. The appropriate delay length can be determined by pattern determination criteria such as AIC or SIC. The null hypothesis is defined as H_0: $\alpha_1 = \alpha_2 = \cdots = \alpha_q = 0$. When the H_0 hypothesis is true, the value nR^2 has asymptotically a chi-square distribution with q degrees of freedom ($nR^2 \sim \chi_q^2$)[12]. Here, n is the number of observations, q is the constraint number and it is the number of lags in which the ARCH effect is investigated. If the obtained chi-square value is greater than the table value, the null hypothesis is rejected.

4. Wind speed time series

Wind speed data has n = 6940 observations and was shown as series $\{Y_t\}$. Here $t = 1, \ldots, n$. The series are daily observations from 1.1.2000-30.12.2018. In order to demonstrate the seasonality structure in the ARIMA model, the series was made weekly. Weekly data were obtained as an average of 7 days. In this case, the number of observations is n = 992 and the model created was found to be SARIMA (0,1,3) (2,0,0) [52] as in Table 1. series related graphics are given in Figure 1.

Table 1. SARIMA (0,1,3) (2,0,0) [52] model information

	MA(1)	MA(2)	MA(3)	SAR(1)	SAR(2)
	-0.7018	-0.0717	-0.0448	0.0963	0.0064
s.e.	0.0334	0.0437	0.037	0.0349	0.0362

sigma^2 estimated as 0.3967: log likelihood=-943.35

AIC=1898.7 AICc=1898.79 BIC=1928.08

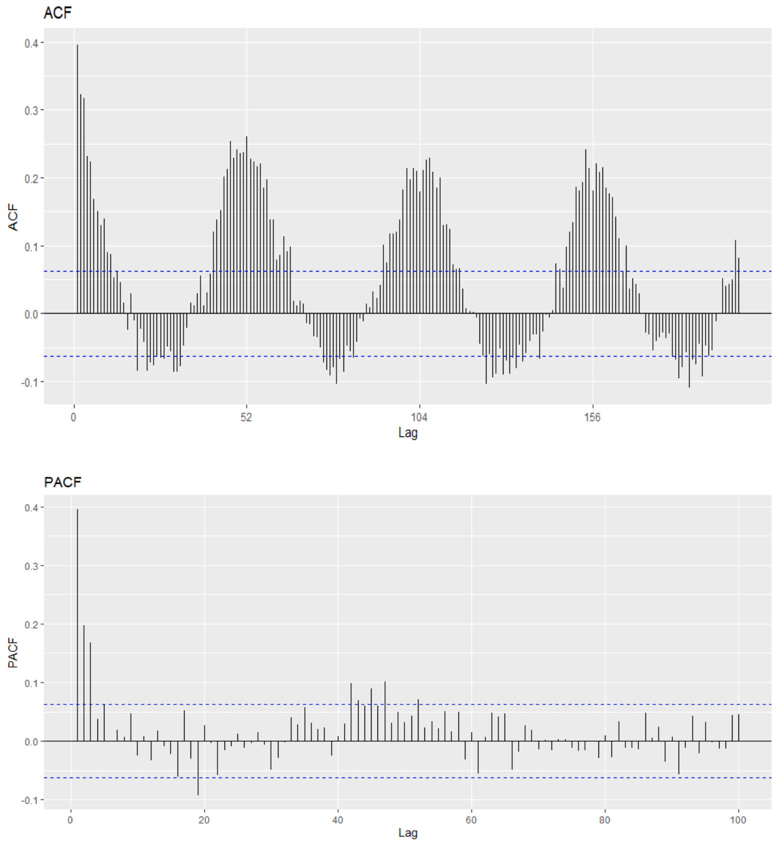

Figure 1. Weekly wind speed graphs

A serial correlation analysis was made of the residuals obtained from models for weekly data (Figure 1). According to the Ljung-Box for weekly data; χ^2 = 13.624, df = 10, p-value = 0.1908 it appears that the residuals are not serially dependent (p > 0.05)

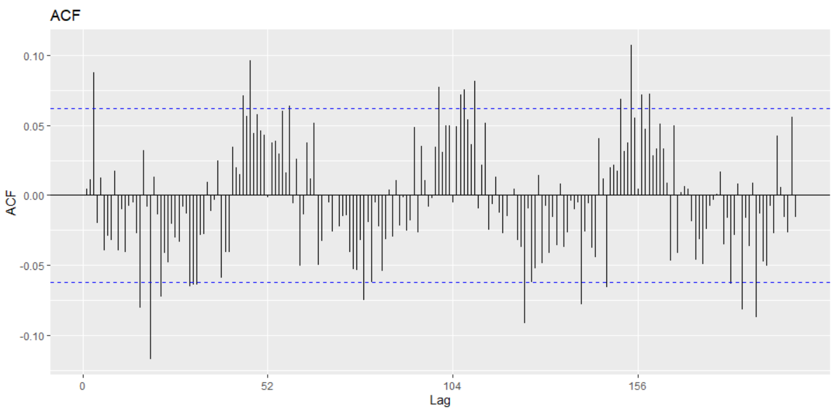

Figure 2. ACF graphs for residuals of weekly data

5. SETAR model estimation

The F statistic, which tests whether there is nonlinearity in the weekly series and gives suggestions for the AR order p, was calculated (F = 2.635, p = 0.0006). According to the F statistics, a nonlinear structure is observed in the data and the recommended number of p lags is 5. This value is also observed from the PACF chart of the series. Later, Tsay test was used to reveal the threshold structure in the data [22].

Table 2. Tsay's F test

p	d	F
5	1	3.8073*
5	2	1.1161
5	3	2.5724*
5	4	1.8575
5	5	1.1233

*: Significant at 5% level

According to the test of Tsay applied to reveal the threshold structure, Y_{t-1} with the value of d = 1, which has the highest F statistic value in weekly data, is taken as the threshold variable (Table 2).
When the scatter plot of Y_t series against Y_{t-1} variable (Figure 3) is examined, the nonlinear representation the data is observed. It was decided to split the data into two regimes and estimates were obtained.

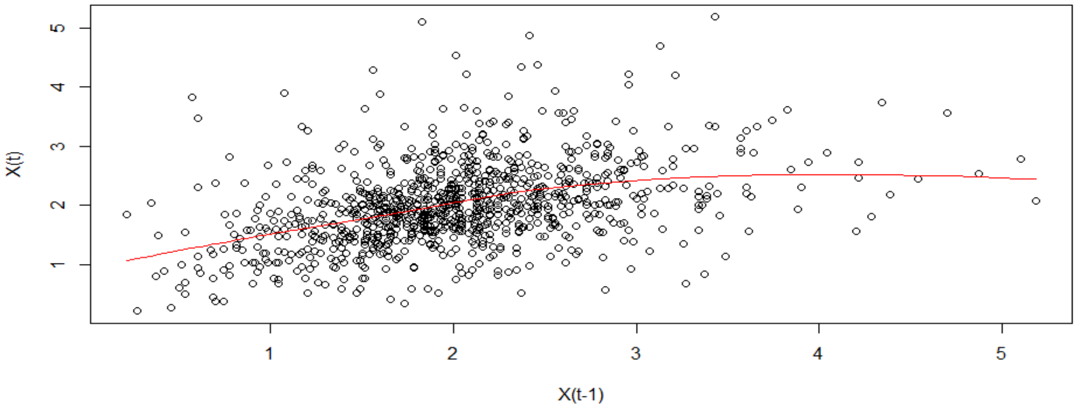

Figure 3. Graph of threshold value determination

The threshold value for the two-regime model is estimated as 2.46. The significant model coefficients with the lowest AIC value for the two regimes of the SETAR model are given in the table. In the model, it is calculated as AIC = 1822 (Table 3).

Table 3. SETAR (1,3,3) model

$Y_{t-1} \leq 2.46$		$Y_{t-1} > 2.46$	
Constant	0.75786	**Constant**	1.48053
AR(1)	0.38528	AR(1)	-0.0301
AR(2)	0.15180	AR(2)	0.11051
AR(3)	0.09361	AR(3)	0.33012

The residuals of the SETAR model obtained for weekly data do not contain autocorrelation (Ljung-Box test; $\chi^2 = 7.8283$, df = 10, p-value = 0.6456).

6. GARCH model estimation for residuals

Since the information criterion AIC is at a lower level in the estimates obtained with weekly data, the residuals estimated from weekly data were examined. The ACF and PACF graphs (Figure 4) generated for the residuals' squares indicate that there may be autocorrelation in the residuals.

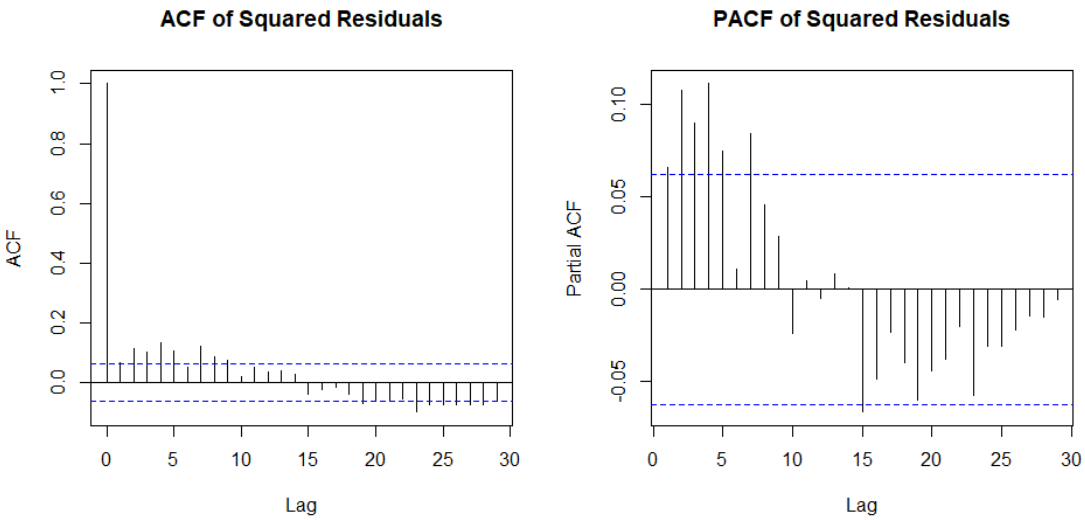

Figure 4. ACF and PACF values for SETAR (1,3,3) model residual squares

The ARCH (LM) test statistic applied to see if there is any ARCH effect on residuals was calculated as 84.242 (p = 0.000). There is varying variance in the series.

Estimates of the GARCH (1,1) model with the lowest AIC value are given in Table 4. The AIC value of the model is 1,752. According to the Ljung-Box and ARCH (LM) tests, autocorrelation and ARCH effects were eliminated (Table 5).

Table 4. GARCH(1,1) model coefficient estimates

OMEGA	ALPHA(1)	BETA(1)
0.02660**	0.16811**	0.77055**

*: Significant at 1% level

Table 5. Remaining tests

			Statistic	p-Value
Ljung-Box Test	R	Q(10)	10.86855	0.367851
Ljung-Box Test	R	Q(15)	12.96366	0.605105
Ljung-Box Test	R	Q(20)	28.33866	0.101622
Ljung-Box Test	R^2	Q(10)	14.16865	0.165437
Ljung-Box Test	R^2	Q(15)	16.8809	0.326034
Ljung-Box Test	R^2	Q(20)	22.3538	0.321625
LM Arch Test	R	TR^2	14.77919	0.253735

7. SETAR Model Prediction

SETAR model estimated for weekly data can be given as in Equation (4). Contrary to the linear time series models, in nonlinear models, when the prediction length is greater than 1, there is usually no formula to predict from the obtained model [26]. For this reason, the parametric bootstrap method will be used to make predictions in the SETAR model. The averages of the 60-unit predictive value obtained from the SETAR model and the prediction interval are given in the graph (Figure 5).

$$Y_t = \begin{cases} 0.75786 + 0.38528 * Y_{t-1} + 0.1518 * Y_{t-2} + 0.09361 * Y_{t-3}, & Y_{t-1} \leq 2.46 \\ 1.48053 - 0.0301 * Y_{t-1} + 0.11051 * Y_{t-2} + 0.33012 * Y_{t-3}, & Y_{t-1} > 2.46 \end{cases}$$

$$h_t = 0.0266 + 0.16811 * \varepsilon_{t-1}^2 + 0.77055 h_{t-1} \quad (4)$$

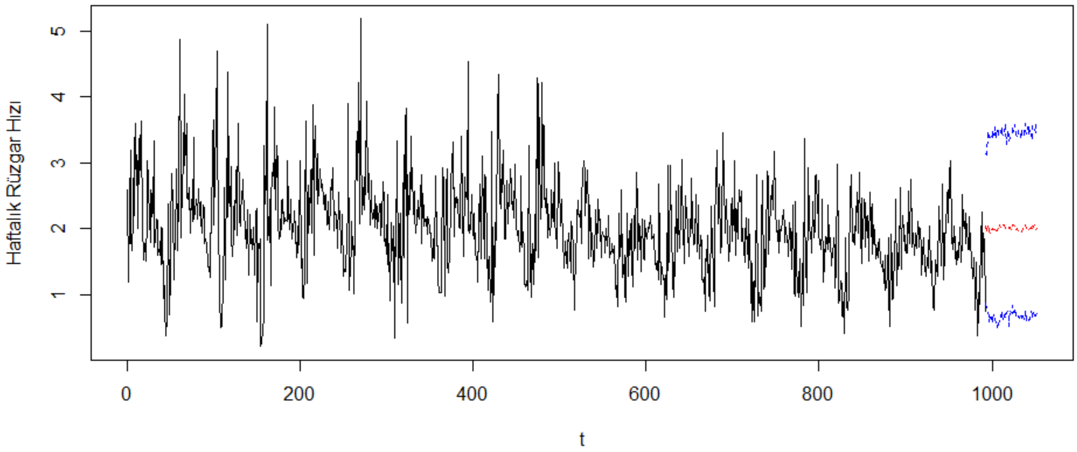

Figure 5. SETAR (1,3,3) -GARCH (1,1) model prediction interval

8. Conclusion

In the study, the predictive values were obtained by predicting the wind speed data with the SETAR model. The SETAR model was obtained for the average estimation of the time series by converting the wind speed into weekly data. However, by revealing the changing variance structure in residual terms with a GARCH model, a range has been stipulated for the predicted values. The resulting SETAR (d = 1, $p_1 = 3$, $p_2 = 3$) - GARCH (1,1) model has a lower AIC value than the wind speed time series from the SARIMA (0,1,3)(2,0,0) [52] model and residues of the model do not contain autocorrelation and variance problems.

Threshold autoregressive models among nonlinear time series models are preferred because they allow data to be modeled in different values. In this study, it has been shown that the SETAR model is useful in predicting wind speed.

References

[1] Narayana M., Sunderland K. M., Putrus G., Conlon M. F., "Adaptive linear prediction for optimal control of wind turbines", Renewable Energy 113 (2017) : 895-906.

[2] Kamjoo A., Maheri A., and Putrus G., "Wind speed and solar irradiance variation simulation using ARMA models in design of hybrid wind-PVBattery system", Journal of Clean Energy Technologies 1(1) (2013).
[3] Zhang M., Wei Yu w., Xu J., "Aerodynamic physics of smart load control for wind turbine due to extreme wind shear", Renewable Energy 70 (2014) : 204-210.
[4] Kose R., Ozgur M.A., Erbas O., Tugcu A., "The analysis of wind data and wind energy potential in Kutahya, Turkey", Renewable and Sustainable Energy Reviews 8(3) (2004) : 277-288.
[5] Mathew S., Pandey K.P., Kumar A.V., "Analysis of wind regimes for energy estimation", Renewable Energy 25(3) (2002) : 381-399.
[6] Ahmed, N.B., "A comparative analysis of forecast performance between SARIMA and SETAR models using macroeconomic variables in Ghana", Master Thesis, Unıversıty of Ghana, (2018).
[7] Albuquerquemello, V.P., Medeiros R.K., Besarria, C.N., Maia, S.F., "Forecasting crude oil price: Does exist an optimal econometric model? ", Energy 155 (2018) : 578-591.
[8] Campenhout B.V., "Modelling trends in food market integration: Method and an application to Tanzanian maize markets", Food Policy 32(1) (2006).
[9] Chen, J., "Crisis, capital controls and covered interest parity: evidence from China in transformation", Paris-Jourdan Sciences Economiques, CNRS : UMR8545, (2012).
[10] Clements, M., Smith, J., "Evaluating forecasts from SETAR models of exchange rates", Journal of International Money and Finance 20 (2001) : 133-148.
[11] Cleveland, W.S., "Robust locally weighted regression and smoothing scatterplots", Journal of the American Statistical Association 74(368) (1979) : 829-836.
[12] Engle, R.F., "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation", Econometrica 50(4) (1982) : 987-1008.
[13] Feng, H., Liu, J., "A SETAR model for Canadian GDP: non-linearities and forecast comparisons", Applied Economics 35(18) (2003).
[14] Genc, A., Erisoglu, M., Pekgor, A., Oturanc, G., Hepbasli, A., Ulgen, K., "Estimation of Wind Power Potential Using Weibull Distribution", Energy Sources 27(9), (2005) : 809-822.
[15] Hasan, H., Nordin, M. A. B. C., & Salleh, N. H. M., "Modeling daily maximum temperature for thermal comfort in northern Malaysia". Advances in Environmental Biology, 9(26 SI) (2015) : 12+.
[16] Hsu, B., Sherina, V., McCall, M.N., "Auto-regressive modeling and diagnostics for qPCR amplification", bioRxiv (2019).
[17] Huang, B.N., Hwang, M.J., Peng, H.P., "The asymmetry of the impact of oil price shocks on economic activities: An application of the multivariate threshold model", Energy Economics 27(3) (2005).
[18] Hutchison, M., Kendall, J., Pasricha, G., Singh, N., "Indian capital control liberalization: evidence from NDF markets", Munich Personal RePEc Archive (2010).
[19] Khadaroo, A.J., "A Threshold in inflation dynamics: Evidence from emerging countries", Applied Economics, 37(6) (2005).

[20] Pinson, P., Christensen, L.E.A., Madsen, H., Sørensen, P.E., Donovan, M.H., Jensen, L.E., "Regime-switching modeling of the fluctuations of offshore wind generation", Journal of Wind Engineering and Industrial Aerodynamics 96(12) (2008)
[21] Singh, T., "Testing nonlinearities in economic growth in the OECD countries: An evidence from SETAR and STAR models", Applied Economics 44 (2012) : 3887–3908 .
[22] Tsay, R., "Testing and modelling threshold autogressive processes", Journal of the American Statistical Association 84 (1989) : 231-240.
[23] Tong, H., "On a threshold model", In Pattern Recognition and Signal Processing (C. H. Chen, ed.), 101-141. Sijthoff and Noordhoff, Amsterdam, (1978).
[24] Tong, H. and Lim, K.S., "Threshold autoregression, Limit cycles and cyclial data", Journal of the Royal Statistical Society Ser. B 42 (1980) : 245-292.
[25] Tong, H., Yeung, I., "On tests for self-exciting threshold autoregressive-type non-linearity in partially observed time series", Applied Statistics 40 (1991) : 43-62.
[26] Tsay, R.S., "Analysis of financial time series", Johns Wiley & Sons, Inc., Publication, Third Edition, Canada, (2010).
[27] Tsay, R.S., Chen, R., "Nonlinear time series analysis", Wiley Series in Probability and Statistics (2019).
[28] Watier, L., Richardson, S., "Modelling of an epidemiological time series by a threshold autoregressive model", The Statistician 44(3) (1999) : 353-364.
[29] Yadav, P.K., Pope, P.F., Paudyal, K., "Threshold autoregressive modelling in finance: the price difference of equivalent assets", Mathematical Finance 4 (1994) : 205-221.
[30] Yang, X.H., Li, Y.Q., "DNA optimization threshold autoregressive prediction model and its application in ice condition time series", Hindawi Publishing Corporation Mathematical Problems in Engineering 2012 (2012), Article ID 191902, 10 pages.
[31] Zhao, Z., Wang, X., Qiao, Y., Sun, H., "Wind speed prediction based on improved self excitation threshold auto regressive model", 37th Chinese Control Conference (CCC), Wuhan 2018 (2018) : 1498-1503.
[32] Zhang, J. J., Shao, C. F., Wang, F., "Research on short-term traffic flow prediction model based on threshold autoregression", Special Issue 3 (2018) : 79-84.
[33] Chan, W., Wong, A., Tong H., "some nonlinear threshold autoregressive time series models for actuarial use", North American Actuarial Journal 8(4) (2004) : 37-61.