Journal of Applied and Natural Science 8 (1): 50 - 54 (2016)

JANS

An agricultural price forecasting model under nonstationarity using functional coefficient autoregression

B. S. Yashavanth*, K. N. Singh and Amrit Kumar Paul

Indian Agricultural Statistics Research Institute, Pusa, New Delhi – 110012, INDIA
*Corresponding author. E-mail: yashavanthbs@gmail.com

Received: July 9, 2015; Revised received: August 30, 2015; Accepted: January 20, 2016

Abstract: In this globalized world, management of food security in the developing countries like India where agriculture is dominated needs efficient and reliable price forecasting models more than ever. Forecasts of agricultural prices are handy to the policymakers, agribusiness industries and farmers. In the present study, Functional Coefficient Autoregression (FCAR) has been applied for modeling and forecasting the monthly wholesale price of clean coffee seeds in Hyderabad coffee consuming center using the data from Jan, 2001 to Sep, 2014. FCAR (2,2) model was found suitable based on the minimum Average Prediction Error (APE) criterion. The FCAR model thus obtained was compared with the Autoregressive Integrated Moving Average (ARIMA) model. Since the original series was found to be nonstationary from Augmented Dickey-Fuller test (ADF statistic=-2.84, p=0.22), the differenced series (ADF statistic=-4.20, p<0.01) was used and ARIMA (12,1,0) was found suitable. The FCAR model obtained was compared with the ARIMA model with respect to forecast accuracy measures viz., Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). The RMSE and MAPE for the FCAR (2,2) were found to be 17.16 and 4.41%, respectively, whereas for the ARIMA (12,1,0) models, 62.64 and 26.15%, respectively. The results indicated that the FCAR model was efficient than the ARIMA model in forecasting the future prices.

Keywords: ARIMA, FCAR, Forecasting, Stationarity.

INTRODUCTION

Forecasting the market price is an essential part of commodity trading and price analysis. Agricultural production and price are highly varying as they are largely influenced by several eventualities. Natural calamities like droughts, floods and attacks by pests and diseases make these unpredictable leading to a considerable risk and uncertainty in the process of price modeling and forecasting. Forecasts of agricultural production and prices are intended to be useful to the farmers, governments and agribusiness industries. Policy makers need internal forecasts to execute policies that provide technical and market support for the agricultural sector. Before liberalization and globalization, prices were controlled by the government, rendering price forecasting a low value-added activity. Presently, the prices are determined by international and domestic market forces. This leads to increased price variability making it imperative to study the trends in prices of different commodities by employing efficient statistical modeling techniques which in turn, will help the planners in formulating suitable policies to face the challenges ahead. The agricultural price forecasts are also important to farmers as it helps them to strategize their production and marketing on the expected prices that may have financial repercussions many months later (Jha and Sinha, 2013).

In time series modeling, the past observations of the same variable are collected and analyzed to develop a model describing the underlying relationship. During the past few decades, a lot of effort has been directed towards developing and improving time series forecasting models. One of the most important and widely used time series models is the Auto Regressive Integrated Moving Average (ARIMA) model. The popularity of ARIMA model is due to its statistical properties as well as use of well-known Box-Jenkins methodology in the model building process (Box *et al*., 2007). The ARIMA methodology has been used by several authors for agriculture related forecasting such as cultivated areas (Prabakaran *et al*., 2013), price (Assis *et al*., 2010 and Paul, 2010), productions (Paul *et al*., 2013a, Paul and Das, 2013 and Paul *et al*., 2014) and productivity (Padhan, 2012) of different crops. If the seasonality is observed in the data, Seasonal ARIMA (Paul *et al*., 2013b) can be made use of.

However ARIMA requires the series to be stationary. Data points are often non-stationary or have means, variances and covariances that change over time. Non-stationary data, as a rule, are unpredictable and cannot be modeled or forecasted. The results obtained by using non-stationary time series may be spurious in that they may indicate a relationship between two variables where one does not exist. In order to receive consistent, reliable results, the non-

ISSN : 0974-9411 (Print), 2231-5209 (Online) All Rights Reserved © Applied and Natural Science Foundation www.ansfoundation.org

stationary data needs to be transformed into stationary data. In contrast to the non-stationary process that has a variable variance and a mean that does not remain near, or returns to a long-run mean over time, the stationary process reverts around a constant long-term mean and has a constant variance independent of time. Widely used technique to achieve stationarity is to difference the series until it becomes stationary and then proceed for modelling. Alternatively, one can go for non-parametric time series models, like FCAR, which do not make any assumption about the behavior of the series (Cai *et al*., 2009).

The present study is directed at application of FCAR time series model to forecast the agricultural prices under nonstationary conditions. As an illustration, the monthly wholesale prices of clean coffee seeds in Hyderabad coffee consuming center is modeled and forecasted using FCAR model. An attempt is also made to compare the results obtained with the ARIMA models.

MATERIALS AND METHODS

FCAR model: Functional Coefficient Autoregressive (FCAR) nonparametric time-series model, introduced by Chen and Tsay (1993), admits the form

$$X_t = f_1(X_{t-d})X_{t-1} + \ldots + f_p(X_{t-p}) + \varepsilon_t \quad (1)$$

where ε_t is white noise with finite variance σ^2 and is independent of X_{t-1}, X_{t-2}, ... The coefficient functions $f_1(.)$, $f_2(.)$, ... , $f_p(.)$ are unknown and change gradually. It is a direct extension of the linear AR model, but allows the coefficients to vary according to a threshold variable X_{t-d} which is one among the lagged variables. The functions in the model can be determined using a non-parametric procedure called local linear regression technique hence allowing 'data to speak for themselves'.

FCAR model can be regarded as a Stochastic regression model by introducing dependent variable Y_t as current observation X_t , the ith independent variable X_i as lag i variable X_{t-i} (i=1,2,…,p), and U as lag d variable X_{t-d} ($d \leq p$). With induced variables, FCAR model (Cai *et al*., 2000) can be written as

$$Y_t = f_1(U)X_{t-1} + f_2(U)X_{t-2} + \ldots + f_p(U)X_{t-p} + \varepsilon_t \quad (2)$$

In order to apply FCAR model to data, the coefficient functions may be estimated by using a local linear regression technique. These coefficient functions are expanded by Taylor's series expansion in which unknown coefficients are estimated by the method of weighted least squares, weights being the kernel density function (Fan and Yao, 2003). For any given u_0 and u in a neighborhood of u_0 and using Taylor's series expansion:

$$f_j(u) \approx f_j(u_0) + f_j'(u_0)(u-u_0) \equiv a_j + b_j(u-u_0) \quad (3)$$

where a_j and b_j are local intercept and slope, respectively. Using the data with U_i around u_0 and local model (3), the following expression is minimized:

$$\sum_{i=1}^{n}\left[Y_t - \sum_{j=1}^{p}\{a_j + b_j(U_i - u_0)\}X_{ij}\right]^2 K_h(U_i - u_0) \quad (4)$$

where $K_h(.) = h^{-1}K(./h)$ where $K(.)$ is a kernel function and h is the bandwidth. Then, the local linear regression estimator is simply $\hat{f}_j(u_0) = \hat{a}_j$.

The local linear regression estimator a_j and b_j can be easily obtained. Let $\mathbf{e}_{j,2p}$ be the $2p \times 1$ unit vector with 1 at the jth position, $\mathbf{X}$ denote an $n \times 2p$ matrix with $(\mathbf{X}_i', \mathbf{X}_i(U_i - u_0))$ as its ith row, and $\mathbf{Y} = [Y_1 :, \ldots, Y_n]^T$, where superscript T indicates transpose. Set $W = diag\{K_h(U_1 - u_0), \ldots, K_h(U_n - u_0)\}$. Then the local regression problem reduces to minimizing $(\mathbf{Y} - \tilde{\mathbf{X}}\boldsymbol{\beta})^T \mathbf{W}(\mathbf{Y} - \tilde{\mathbf{X}}\boldsymbol{\beta})$, where $\boldsymbol{\beta} = (a_1, a_2, \ldots, a_p, b_1, b_1, \ldots, b_p)^T$. The local least square estimator is $\hat{\boldsymbol{\beta}} = (\tilde{\mathbf{X}}'\mathbf{W}^{-1}\tilde{\mathbf{X}})'\tilde{\mathbf{X}}'\mathbf{W}^{-1}\mathbf{Y}$.

The optimal bandwidth h in the local linear regression methodology is selected by "Modified multifold cross-validation" criterion of Cai *et al*. (2000). Taking m and Q as two positive integers such that $n > mQ$, the parameters are estimated using various bandwidth values h_j and the Q-fitted models are used for carrying out one -step forecasting error of the next section of the time-series of length m based on the estimated models.

Let $\{\hat{a}_{j,q}(.)\}$ be the estimated coefficients using q^{th}, q=1, 2,…, Q subseries $\{(U_i, X_i, Y_i), 1 \leq i \leq n\text{-}qm\}$. The

Table 1. Forecasting performance of ARIMA and FCAR models.

Model	Measure of accuracy	
	RMSE	MAPE (%)
ARIMA (12,1,0)	62.644	17.163
FCAR (2,2)	26.154	4.410

Table 2. The actual and forecasted price values.

Observation number	Month and year	Actual price (Rs/kg)	Forecasted price (Rs/kg)	
			ARIMA	FCAR
160	Apr,2014	280	268.70	277.00
161	May,2014	345	276.63	285.78
162	Jun,2014	342	274.71	366.08
163	Jul,2014	345	273.92	345.04
164	Aug,2014	347	278.42	349.66
165	Sep,2014	350	283.22	351.42

average prediction error using q^{th} subseries is given by

$$APE_q(h) = \frac{1}{m}\sum_{i=n-qm+1}^{n-qm+m}\left[Y_i - \sum_{j=1}^{p}\hat{a}_{j,q}(U_i)X_{i,j}\right]^2$$

The overall average prediction error is given by

$$APE(h) = Q^{-1}\sum_{q=1}^{Q} APE_q(h) \quad (5)$$

The proposed data driven bandwidth is the one that minimizes APE(h). In practice, generally m = [0.1n] and Q = 4 are considered. The selected bandwidth does not depend critically on the choice of m and Q so long as mQ is reasonably large; thereby ensuring that evaluation of prediction errors is stable.

Choosing an appropriate model dependent variable U is also very important. Knowledge of physical background of data may be very helpful. Without any prior information, it is pertinent to choose U in terms of some data–driven methods, such as Akaike information criterion (AIC), cross-validation and other criteria. Let APE(h,d) be the average prediction error defined by (5) using lagged variable $U = X_{t-d}$. A simple and practical approach is to minimize APE(h, d) simultaneously for h in a certain range and d over the set {1, 2, …, p}. The order p can also be chosen to minimize the APE.

Forecast evaluation methods: The forecasting ability of different models is assessed with respect to two common performance measures, viz. the root mean squared error (RMSE) and the mean absolute percentage error (MAPE). The RMSE measures the overall performance of a model and is given by Equation (6)

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\hat{y}_t - y_t\right)^2} \quad (6)$$

where, y_t is the actual value for time t, $\hat{y}_t$ is the predicted value for time t, and n is the number of predictions. The second criterion, the mean absolute percentage error is a measure of average error for each point forecast and is given by Equation (7)

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\hat{y}_t - y_t}{y_t}\right| \times 100 \quad (7)$$

where the symbols have the same meaning as above. The model with least RMSE and MAPE values is considered as the best model for the data.

Collection of data: The monthly wholesale prices (per kilogram of clean coffee seeds) of Arabica coffee seeds at Hyderabad coffee consuming center are used for the study. The data covered a period of 165 months (January, 2001 to September, 2014). The first 159 data points are used for model fitting and the last 6 data points are used for model validation. The data were obtained from various issues of Coffee Data, published by Coffee Board, Government of India available at the website *www.indiacoffee.org*.

RESULTS AND DISCUSSION

Any time series analysis begins with plotting the data against the time (Tsay, 2010). Fig. 1 shows the time series plot of monthly wholesale prices of clean seeds of Arabica coffee for the period January 2001 to

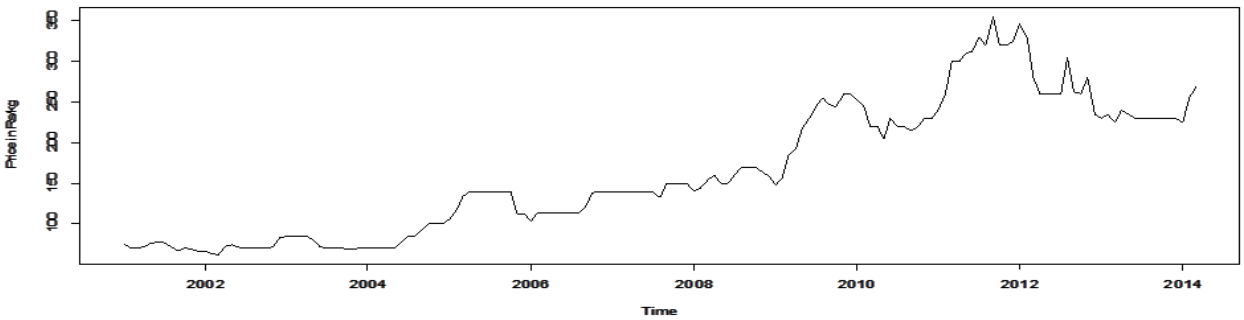

Fig. 1. *Monthly wholesale price of clean coffee seeds at Hyderabad center.*

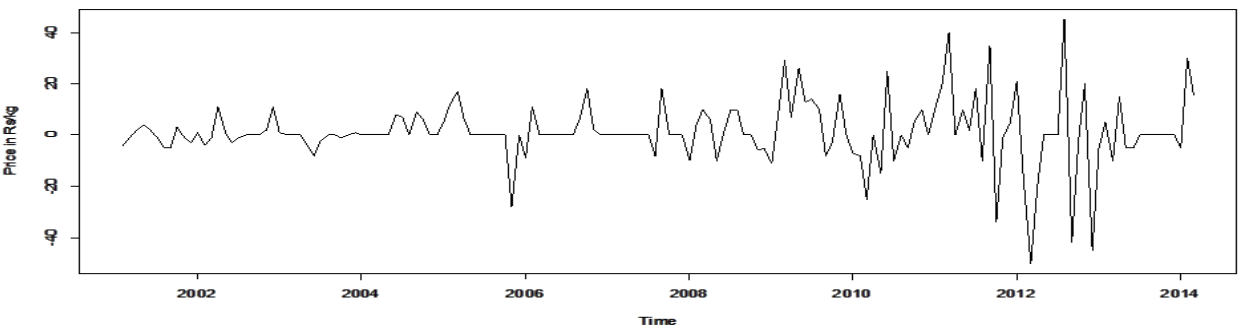

Fig. 2. *Monthly wholesale price of clean coffee seeds in Hyderabad center after first differencing.*

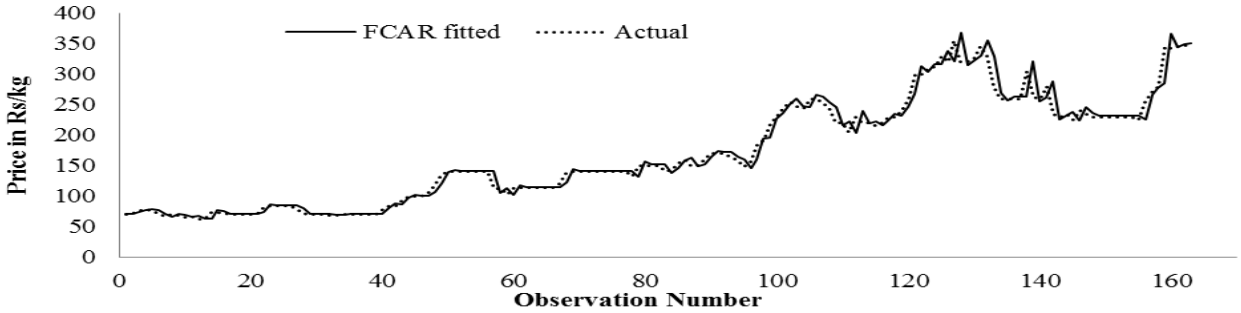

Fig. 3. *Actual and forecasted monthly wholesale price of clean coffee seeds in Hyderabad center.*

March 2014. A perusal of Fig. 1 reveals a positive trend over time which indicates the nonstationary nature of the time series. To confirm the presence of nonstationarity in the original data, a unit root test called Augmented Dickey-Fuller (ADF) test is performed. This test has been used by several authors (Gupta and Basu, 2007, Su and Deng, 2014 and Toumache *et al*., 2014) to check stationarity of a time series. The original series was found to be nonstationary (ADF statistic = -2.84, p = 0.22) whereas, after first differencing, the series became stationary (ADF statistic = -4.20, $p < 0.01$). Fig. 2 shows the plot of monthly wholesale prices of clean seeds of Arabica coffee at Hyderabad center after first differencing.

The candidate ARIMA models for the differenced series are found out based on the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). Among the candidate models, the model with least AIC is chosen as the best model (Makridakis *et al*., 2003). ARIMA (12,1,0) was found to be the best based AIC and only this model had a significant parameter estimate. The model obtained is given in equations (8). The values in the parenthesis are the standard errors of the parameters.

$$\Delta y_t = 1.310 + 0.326\, \Delta y_{t-12} \quad (0.503) \quad (0.083) \qquad (8)$$

Where, Δy_t is the first differenced price values at time t.

Optimum FCAR model is selected by the aforesaid methodology. Estimates of parameters of this model are obtained by using computer program in SAS-IML. The optimum values for p and d are found as p = 2 and d = 2. The function APE against the bandwidth over a grid of point h_j =0.1 (j=1,2,…,10) is computed. The selected bandwidth along with autoregressive order p and delay parameter d which minimizes the APE defined in (5) is found as h = 0.1. The fitted FCAR (2,2) model is as follows:

$$Y_t = \{1.2345 - 0.000464(Y_{t-2} - 164.239)\} Y_{t-1} + \{-0.2196 + 0.000507\,(Y_{t-2} - 164.239)\} Y_{t-2}$$

Here, Y_{t-2} is the threshold value and 164.239 is the mean of the data.

The comparative results for the ARIMA and FCAR models with respect to measures of forecast accuracy viz., RMSE and MAPE are given in Table 1. The model with least values of RMSE and MAPE is considered to be the best for forecasting (Jha and Sinha, 2013). From the table, it is evident that both MAPE and RMSE values of FCAR model are less than those of ARIMA model. One-step ahead forecasts of wholesale prices during the months April, 2014 to September, for ARIMA (12,1,0), and FCAR(2,2) models are carried out and reported in Table 2. The price values forecasted by using FCAR(2,2) model were found to be par with the actual values than those forecasted using ARIMA(12,1,0). Figure 3 shows the plot of actual (dotted lines) and FCAR forecasted (solid lines) monthly wholesale prices of clean seeds of arabica coffee in Hyderabad center. These results indicated that the FCAR model outperformed ARIMA model in forecasting the prices. A similar result was obtained when Indian Lac export data was forecasted using FCAR model and compared with ARIMA and SETARMA models (Ghosh *et al*., 2010). Also, by going for FCAR model instead of ARIMA model, one can skip the ADF test or any other such test for stationarity and can directly proceed to fit the data. This also avoids the need for data transformations like differencing (Makridakis *et al*., 2003), detrending (Chuang *et al*., 2012), Box-Cox transformation and Centering (Tan *et al*., 2014) and Logarithmic transformations (Ayekple *et al*., 2015) to achieve stationarity which is needed for ARIMA modeling. Hence a better performing model can be obtained in a fewer steps than the ARIMA technique. Bharadwaj *et al*. (2014) used differencing to achieve stationarity for forecasting daily prices of Gram. Hassan *et al*. (2013) applied logarithmic transformation for achieving stationarity for forecasting wholesale price of coarse rice in Bangladesh. In such instances, data transformation can be avoided by making use of aforesaid FCAR technique of forecasting.

Conclusion

Due to globalization and market integration, there is an increased need for agricultural price information at all levels of decision making. This marks the need for statistical techniques to provide accurate and timely price forecast by taking into account the local information to the farmers, traders and policymakers so that they may make production, marketing and policy decisions well in advance. Time series models are vastly

used for this purpose of model building and forecasting. In this paper a nonparametric time series modeling technique called Functional Coefficient autoregression (FCAR) has been used to model and forecast the monthly wholesale prices of arabica coffee in Hyderabad coffee consuming center. The advantage of the FCAR model is that it does not make any assumption on the stationarity behavior of the series, unlike ARIMA model, making it suitable for modeling both stationary and nonstationary series. The FCAR model arrived is also compared with the ARIMA model and it is found that FCAR model fits better than ARIMA model with respect to forecast accuracy measures and forecasted values.

REFERENCES

Assis, K., Amran, A. and Remali, Y. (2010). Forecasting cocoa bean prices using univariate time series models. *Journal of Arts Science & Commerce*. 1(1): 71-80.

Ayekple, Y.E., Harris, E., Frempong, N.K. and Amevialor, J. (2015). Time Series Analysis of the Exchange Rate of the Ghanaian Cedi to the American Dollar. *Journal of Mathematics Research*. 7(3): 46-53.

Bharadwaj, S.P., Paul, R.K., Singh, D.R. and Singh, K.N. (2014). An Empirical Investigation of Arima and Garch Models in Agricultural Price Forecasting. *Economic Affairs*. 59 (3): 415-428.

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (2007). Time-Series Analysis: Forecasting and Control. Pearson Education. India.

Cai, Z., Fan, J. and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time-series. *Journal of American Statistical Association*. 95: 941–956.

Cai, Z., Li, Q. and Park, J.Y. (2009). Functional-coefficient models for nonstationary time series data. *Journal of Econometrics*. 148: 101-113.

Chen, R. and Tsay, R. S. (1993). Functional coefficient autoregressive models. *Journal of American Statistical Association.* 88: 298-308.

Chuang, W.I., Liu, H.H. and Susmel, R. (2012). The bivariate GARCH approach to investigating the relation between stock returns, trading volume, and return volatility. *Global Finance Journal*. 23 (1): 1-15.

Fan, J. and Yao, Q. (2003). Nonlinear Time-Series: Nonparametric and Parametric Methods. Springer, U.S.A.

Ghosh, H., Paul, R. K. and Prajneshu. (2010). Functional coefficient autoregressive model for forecasting Indian lac export data. *Model Assisted Statistics and Applications*. 5 (2): 101-108.

Gupta, R and Basu, P. K. (2007). Weak form efficiency in Indian stock markets. *International Business and Economics Research Journal*. 6 (3): 57-64.

Hassan, M.F., Islam, M.A., Imam, M.F. and Sayem, S.M. (2013). Forecasting wholesale price of coarse rice in Bangladesh: A seasonal autoregressive integrated moving average approach. *Journal of the Bangladesh Agricultural University*. 11(2): 271-276.

Jha, G.K. and Sinha, K. (2013). Agricultural price forecasting using neural network model: An innovative information delivery system. *Agricultural Economics Research Review.* 26 (2): 229-239.

Makridakis, S., Wheelright, S.C. and Hyndman, R.J. (2003). Forecasting: Methods and Applications. Wiley-India. New Delhi.

Padhan, P. C. (2012). Application of ARIMA model for forecasting agricultural productivity in India*, Journal of Agriculture & Social Sciences.* 8: 50-56.

Paul, R. K. (2010). Stochastic Modeling of Wholesale Prices of Rohu in West Bengal, India. *Interestat*.

Paul, R. K. and Das, M. K. (2013). Forecasting of average annual fish landing in Ganga Basin. *Fishing chimes*. 33 (3): 51-54.

Paul, R. K., Alam, W. and Paul, A. K. (2014). Prospects of livestock and dairy production in India under time series framework. *Indian Journal of Animal Sciences*. 84 (4): 130-134.

Paul, R.K., Panwar, S., Sarkar, S.K., Kumar, A., Singh, K.N., Farooqi, S. and Choudhary, V.K. (2013b). Modelling and forecasting of meat exports from India. *Agricultural Economics Research Review.* 26 (2): 249-255.

Paul, R.K., Prajneshu and Ghosh, H. (2013a). Statistical modelling for forecasting of wheat yield based on weather variables. *Indian Journal of Agricultural Sciences*. 83 (2): 180-183.

Prabakaran, K., Sivapragasam, C., Jeevapriya, C. and Narmatha, A. (2013). Forecasting cultivated areas and production of wheat in India using ARIMA model, *Golden Research Thoughts*. 3 (3).

Su, J. and Deng, G. (2014). The Chinese urban and rural per capita income and trend analysis. *Applied Mathematics*, 5: 106-109.

Tan, S., Bhaduri, M. and Ho, C.H. (2014). A Statistical Model for Long-Term Forecasts of Strong Sand Dust Storms. *Journal of Geoscience and Environment Protection*. 2:16-26.

Toumache, R., Rouaski, K. and Talbi, B. (2014). The impact of fluctuating oil prices on inflation in Algeria. *Journal of Business and Retail Management Research*, 9 (1): 64 -72.

Tsay, R. S. (2010). Analysis of Financial Time series. John Wiley and Sons, New Jersey.