RESEARCH PAPER | Statistics | Volume : 4 | Issue : 7 | July 2014 | ISSN - 2249-555X

F-ARIMA Model for Forecasting Natural Rubber Production in India

KEYWORDS: Time series, F-ARIMA, Box-Jenkins Methodology, AIC, R-Package.

P. Arumugam
Reader, Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli-627012, India

V. Anithakumari
Research Scholar, Department of Statistics Manonmaniam Sundaranar University Tirunelveli-627012, India

ABSTRACT *Fundamental ARIMA model furnish an attractive approach to time series data analysis. This article proposes the forecasting models that can detect trend, seasonality, auto regression, moving average and error term for F-ARIMA. This model is specific to forecast seasonal demand from a time series where the demand of a period is correlated with the demand of the other periods in a business cycle.*

INTRODUCTION

Time series modeling holds a great promise as a tool for studying network traffic. However, traditional models can only capture short-range dependence; for examples, Poisson process, Markov processes, AR, MA, ARMA and ARIMA processes [1].Therefore, models are required to describe both long range and short-range dependence simultaneously. We consider F-ARIMA (*p,d,q*) (fractional autoregressive integrated moving average) model as one of good models with this capability. This paper studied the F-ARIMA models in its implementation detail. We provide a procedure to fit a F-ARIMA model to the actual traffic trace, as well as a method to generate a F-ARIMA process with given parameters. We use the techniques of backward prediction, fractional differencing and a combination of rough estimation and accurate estimation to provide guidelines for simplifying the F-ARIMA model fitting procedure, in order for us to reduce the time of traffic modeling.

METHODOLOGY

A time series is a set of values represented by a linear combination of independent random variable $y_t(t = 1,2,3, \ldots n)$, where index t indicates the intervals of time. For seasonal time series data, the direct scale of time is not always necessary to develop the model. Any mean difference of the series or logarithmic transformation of data can be used to develop the model. The development of the model involves two basic tasks: (a) identifying the nature of the demand represented by the sequence of observation, and (b) predicting future demands of the time series. To achieve the first goal, the following are the steps considered: First, identify the pattern of observed time t series data, and then the parameters of the model are estimated. The second goal is achieved by extrapolating the model to predict the future demand. In a given time series the following can be recognized as: A long-term component of variability termed as trend represents the pattern of the series, A short-term component, whose shape occurs periodically at intervals of s lags of the index variable, is known as seasonality.

The theory of time series analysis has been developed as a set of linear operators. According to Box and Jenkins (1970), a highly useful operator in time series theory is the lag or backshift linear operator (B) to eliminate the linear or seasonal trend.

Difference Operator to Remove the Increasing Trend

In time-series analysis, the lag or backshift linear operator (B) is used to eliminate the linear trend. If the operator B makes $BZ_t = Z_{t-1}$, which shifts backward in time by one period, B is called the 1st order delay operator. For example, $BZ25= Z24$. The double application of lag operator is indicated by $B2$. Applying the lag operator twice to a series, the result is given by

$$B(BZ_t) = BZ_{t-1} = Z_{t-1}$$

Definition: The k-th order delay operator is defined as

$$B^k Z_t = Z_{t-k}$$

Therefore, any integer k is written as $B^k Z_t = Z_{t-k}$. Using the back operator from Definition 1, the Equation (2) can be rewritten as

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} - \ . \ . \ . - \phi_p Z_{t-p} = \varepsilon_t = \phi(B) Z_t$$

The autoregressive and moving average components can be combined in an autoregressive Moving average (ARMA) (p, q) model as

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + . . . + \phi_p Z_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - . . . - \theta_q \varepsilon_{t-q}$$

The lag operator used in the above equation is

$$(1 - \theta_1 B - \theta_2 B^2 - \ . \ . \ . - \theta_q B^q) Z_t = \left(1 - \theta_1 B - \theta_2 B^2 - \ . \ . \ . - \theta_q B^q\right) \varepsilon_t$$

$$\phi(B) Z_t = \theta(B) \varepsilon_t$$

Once linear trends of time series are removed, the periodic trends are eliminated as following. The analysis of a series begins by evaluating the long and short-term periodic components, which are essential to define the regular structure of the series. The trend components are evaluated by fitting a regular function. According to Box and Jenkins (1970), the seasonal component is estimated by a seasonal decomposition procedure, which calculates a seasonal index based on the ratio of the observed values to the moving average. In the final stage of series modeling, both the trend and the seasonal component are integrated in the ARMA (p, q) process. For the trend, such integration is obtained by using the difference linear operator $\nabla = 1 - B$, $\nabla Y_t = (1 - B) Y_t$. A single application of the ∇ operator transforms the data to a linearly increasing trend, and repeated use of the ∇ operator for d times $\left(\nabla^d\right)$ transforms the trend to stationary which can be fitted by a d-order polynomial. Stationary series Zt obtained after the d^{th} difference $\left(\nabla^d\right)$ of Yt, which is given by

$$Z_t = \nabla^d Y_t = (1 - B)^d Y_t$$

The combination of ∇ operator in Equation (6) and the ARMA (p, q) process results in an ARIMA (p, d, q) model. Again, ARIMA can be used for the seasonal component of s lag period, by using both correlations between Zt and $Zt\text{-}s$ values and those between the corresponding residuals εt and $\varepsilon t\text{-}s$. A seasonal ARIMA model is an ARIMA (p, d, q) model Whose residuals εt are further modeled by an ARIMA (P, D, Q)s. The operators of a seasonal ARIMA model is defined as (p, d, q)x(P,D,Q)s.

F-ARIMA MODEL IDENTIFICATION

ARIMA model is estimated only after transforming the variable for forecasting into a stationary series. The stationary series is the one whose values vary over time only around a constant mean and constant variance. The stationary of data series after various differencing is shown in Figure 1.

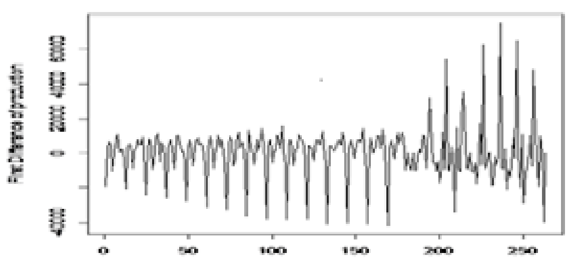

Figure 1: Actual Rubber Production Data Series

Figure 1indicates a strong periodic seasonal pattern and increasing trend. This appears to be non-stationary. After differencing the series, the newly constructed variable is Zt, which is $Z_t = \nabla^d \nabla^D = (1 - B)^d (1 - B)^D$. Thus, Zt is determined after differencing the data, which is

$$Z_t = (y_t - y_{t-1}) - (y_{t-12} - t_{t-13})$$

Volume : 4 | Issue : 7 | July 2014 | ISSN - 2249-555X

The next part of this step is to identify the values of *p* and *q*, which are the AR(*p*) and MA(*q*) components for both seasonal and non-seasonal series. The ACF and Partial ACF show that the order of *p* and *q* can at the most be one. Two goodness of fit statistics are most commonly used for model selection AIC and SBC. The AIC and BIC is determined based on a likelihood function. Several (seven) tentative ARIMA models are tested for the data series and the corresponding AIC values for the models are shown in Table 1.

TABLE -1
AIC VALUE FOR VARIOUS F-ARIMA MODELS

Model	ARIMA (p, d, q) *(P, D, Q)$_{12}$	AIC
1	(1, 1, 1) (0, 1, 0)	5500.64
2	(1, 1, 1) (0, 1, 1)	5476.98
3	(1, 1, 1) (2, 1, 1)	5443.99
4	(1, 1, 1) (1, 1, 1)	5442
5	**(2, 1, 2) (1, 1, 1)**	**5425.62**
6	(2, 1, 2) (2, 1, 1)	5427.6
7	(1, 1, 1) (1, 1, 0)	5473.08

The models that have the lowest AIC value are *F-ARIMA* (2,1,2) (1,1,1)$_{12}$ and (2,1,2) (2,1,1)$_{12}$. Since two models are identified, the most suitable model is selected by checking the residuals of both models and selected the one with the most significant residuals. The AIC values residual test and the estimation of model parameters are performed by the R package. The results indicates that *F-ARIMA* (2,1,2) (1,1,1)$_{12}$ is a significantly better model.

Parameters Estimation of F-ARIMA Model

Once a suitable F-ARIMA (p, d, q) × (P, D, Q)$_{12}$ structure is identified, the second step is the parameter estimation or fitting stage. The parameters are estimated by the maximum likelihood method. It is also important to check that the parameters contained in the model are significant. Both the moving average and the autoregressive parameters have significant *t* values. The subsequent step after the parameter estimation is the Diagnostic Checking or model verification. The Box and Jenkins (1970) estimation process for seasonal *F-ARIMA* model is the forms of ACF and PACF with differencing of data series and after first and seasonal differencing of the data series are shown in Figure 2 and Figure 3.

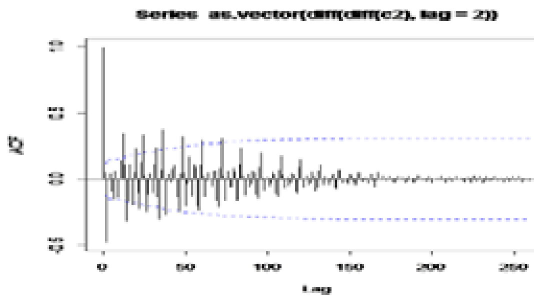

Figure 2: ACF of Seasonal differenced production data

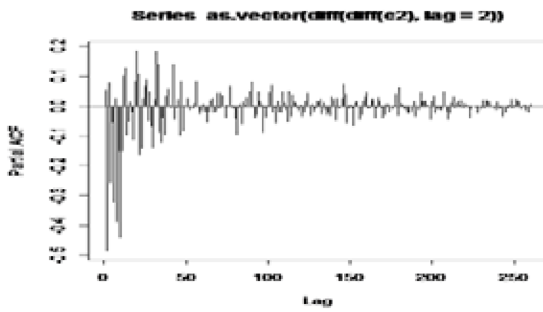

Figure 3: PACF of Seasonal differenced production data

DIAGNOSTIC CHECKING AND MODEL VALIDATION

The model verification is concerned with checking the residuals of the model to determine if the model contains any systematic pattern which can be removed to improve on the selected ARIMA model. Although the selected model may appear to be the best among a number of models considered, it is also necessary to do diagnostic checking to verify that the model is adequate. Verification of an ARIMA model is tested (i) by verifying the ACF of residuals using the chi squared test, (ii) by verifying the normal probability plot of the residuals. The χ^2 tests indicated that the hypothesis cannot be rejected and residuals are uncorrelated. The ARIMA model (2,1,2) (1,1,1)$_{12}$ is selected to forecast the demand variable, where autoregressive term $p = 2$, $P = 1$ differencing term $d = 1$, $D = 1$ (seasonal difference) [that is, $(1-B)(1-B^{12})$] and moving average term $q = 2$, $Q = 1$. The fitted model is given by

Volume : 4 | Issue : 7 | July 2014 | ISSN - 2249-555X

$$\left(1-\phi_1 B-\phi_2 B^2\right)\left(1-\beta_1 B^{12}\right)\left(1-B\right)\left(1-B^{12}\right)y_t = C+\left(1-\psi_1 B-\psi_2 B^2\right)\left(1-\theta_1 B^{2s}\right)e_t$$

The model has (1 and 2) period differencing, the autoregressive factors is the moving average factors is

$$(1-\phi_1 B-\phi_2 B^2) = (1+0.1313B+0.3443B^2)$$
$$(1-\psi_1 B-\psi_2 B) = (1-0.0796B-0.8641B)$$

and the estimated mean, C = 11906.3414 Transforming autoregressive terms and coefficient, the form of equation is given by

$$Y_t = \frac{C+\left(1-\psi_1 B-\psi_2 B^2\right)\left(1-\theta_2 B^{24}\right)e_t}{\left(1-\phi_1 B-\phi_2 B^2\right)\left(1-\beta_1 B^{12}\right)\left(1-B\right)\left(1-B^{12}\right)}$$

$$Y_t = \frac{C+\left(1-\theta_2 B^{24}+\psi_1\theta_2 B^{25}+\psi_2\theta_2 B^{23}\right)e_t}{\left(1-B^2-\theta_2 B^2+\beta_1 B^{13}-\beta_1\phi_1 B^{13}+\beta_1\phi_2 B^{14}\right)}$$

$$Y_t = Y_t\left[B^2\left(1+\theta_2+\beta_1 B^{11}+\beta_1\phi_1 B^{11}+\beta_1\phi_2 B^{12}\right)\right]+\theta_2 B^{23}\left(1-\theta_2 B+\psi_1\theta_2 B^2+\psi_2\right)e_t+C$$

Transforming the back operator is given by

$$Y_t = Y_{t-1}+(1+\phi)Y_{t-2}+(1-\beta_1)Y_{t-11}+(1-\phi_1)Y_{t-11}(1-\beta_1)Y_{t-11}+(1-\beta_1)Y_{t-12}(1-\phi_2)Y_{t-12}$$
$$+(1-\theta)Y_{t-23}(1-\psi_2)Y_{t-1}+(1-\psi_1)Y_{t-2}(1-\theta_2)Y_{t-2}+(1-\psi_2)Y_{t-1}+e_t+\mu$$

FORECAST RESULTS BY *F-ARIMA* MODEL

In order to forecast one period ahead, that is, Y_{t+1} is Increased by one unit, throughout, as given by

$$Y_{t+1} = Y_t+(1-\phi)Y_{t-1}+(1-\beta_1)Y_{t-10}+(1-\phi_2)Y_{t-22}(1-\beta_1)Y_{t-22}+(1-\theta_1)+(1-\psi_2)Y_{t-26}+e_t+\mu$$

The term e_{t+1} is not known because the expected value of future random errors has to be taken as zero. For the forecast of the second period onward, the term e_t is also taken as zero as the actual value is not known so the forecast errors cannot be found. There are 264 data points from January 1991 to December 2012 used to build the F-ARIMA model. Using $\phi_1 = -0.1313,\ \phi_2 = 0.3443, \psi_1 = -0.0796, \psi_2 = -0.8641,\ \beta = 0.3297, \theta = -0.9941$ *and* μ = 119063. In order to forecast for the period 265 is given by

$$Y_{265} = Y_{264}-0.1313Y_{262}+0.3297Y_{251}-0.04328Y_{251}+113515Y_{252}+\hat{e}_{265}-0.13\hat{e}_{241}$$
$$+0.9145\hat{e}_{262}-0.8641\hat{e}_{263}+119063$$

The value of e_{265} is not known, so $\hat{e}_{265}$ is replaced by zero. The value for $\hat{e}_{264}$ is the difference actual production and the forecasted value for the period 264, which is 104816.

CONCLUSION

In this paper we have studied how to fit F-ARIMA models is used to forecast the demand of a natural rubber production. It is possible to explore a number of interrelated models where the demand process is correlated across time. Based on the demand pattern, the F-ARIMA (2, 1, 2) (1, 1, 1)$_{12}$ model was found to be the best model for the dataset. F-ARIMA processes are much more flexible and capable of simultaneously modeling both the long-range and the short-range dependent behavior of a time series. Compared with other short-range dependent processes such as ARMA models, less parameters are required by the F-ARIMA models for a non stationary stochastic time series such as winter apparel, the forecasting model of tern becomes complicated. In ARIMA model, forecast errors are incorporated to refine the predicted value, so the model gradually improves toward the end of the time series and provides satisfactory forecasting accuracy.

ACKNOWLEDGMENT

"The authors would like to thank the reference for given that very helpful comments and suggestions, their insight and comments led to a better presentation of the idea expressed in this paper. This research work was supported by the university Grant Commission, New Delhi."

REFERENCE [1]. Li Song and Pascal Bondon (2011) , "Piecewise F-ARIMA Models for Long-Memory Time Series" CNRS UMR 8506, University Paris-Sud, France.,pp 553-564. | [2]. Oliver W.W. Yang, Fei Xue, Jiakun Liu, Yantai Shu, Lianfang Zhang.," Traffic Modeling Based on F-ARIMA Models", School of Information Technology and Engineering, University of Ottawa, | [3]. L.K. Ibrahim, B.K. Asare, M. Yakubu and 1. Dauda (2011),." Modeling of Sokoto Daily Average Temperature: A Fractional Integration Approach", Nigerian Journal of Basic and Applied Science (2011), 19(1):21- 30, ISSN 0794-5698 | [4]. J.R.M. Hosking (1981):" Fractional Differencing", Biometrika, Vol 68, No:1, Pp,165-176. | [5]. Fallaw Sowell91992):"Modeling long run behavior with the fractional ARIMA Model". Journal of Monetary Economics 29, Pp, 277-302, north Holland, Elsevier Science Publisher. | [6]. Oliver W.W.Yang, Yantaishu:"Traffic Prediction Using F-ARIMA Models". School of Information Technology and Engineering, University of Ottawa. | [7]. Philippe Barbe and Michel Broniatowski (1998):"Note on functional large Deviation principle for fractional ARIMA process", Statistical Inference for stochastic process, Vol 1: Pp,17-27. kilwer Academic publisher, printed in the Netherlands. |