Advances in Information Mining
ISSN: 0975-3265 & E-ISSN: 0975-9093, Volume 4, Issue 1, 2012, pp.-52-56.
Available online at http://www.bioinfo.in/contents.php?id=32

SHORT-TERM TRANSACTIONS FORECASTING USING TIME SERIES ANALYSIS: A CASE STUDY FOR INDIA

SHARMA S.A.[1] AND BHATIA M.P.S.[2]
[1]Department of Information Technology, Krishna Engg. College, Ghaziabad-201 007, UP, India.
[2]Department of Computer Engineering, Netaji Subhas Institute of Technology, New Delhi-110 075, Delhi, India.
*Corresponding Author: Email- surbhi263@yahoo.com

Received: June 18, 2012; Accepted: June 23, 2012

Abstract- This paper presents time series analysis for short-term Indian bank transactions (NEFT) forecasting. Two time series models are proposed, namely, the multiplicative decomposition model and the seasonal ARIMA Model. Forecasting errors of both models are computed and compared. The proposed models are implemented to predict one year transactions data. The accuracy of the two models are calculated and compared. The paper utilizes the mean absolute percentage error (MAPE) as a measure of forecast accuracy. Results show that both time series models can accurately predict the short-term Transactions load demand and that the Multiplicative decomposition model slightly outperforms the seasonal ARIMA model.
Key words- Indian Bank Data, Short-Term Transactions (NEFT) Forecasting, Time Series Analysis.

Citation: Sharma S.A. and Bhatia M.P.S. (2012) Short-Term Transactions Forecasting Using Time Series Analysis: A Case Study for India. Advances in Information Mining, ISSN: 0975-3265 & E-ISSN: 0975-9093, Volume 4, Issue 1, pp.-52-56.

Copyright: Copyright©2012 Sharma S.A. and Bhatia M.P.S. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Introduction

Transactions forecasting is a vital and fundamental factor for a successful operation of banking system. In order to operate the banking system effectively and efficiently, the system transactions load should be correctly predicted. Transaction demand forecasting has significant implications on the costs and speed of the Banking transactions. Accurate forecasting models are needed for secure and reliable banking system operations. If the system load forecast is overstated, the system may over-commit the generation and leads to costly operation. On the other hand, if the system load forecast is understated, the reliability and security of the system may be compromised which may result in system failures.

The approach using time series analysis is among the main areas with rich research effort [1-7], with specially formulated methods for data in various contexts, for example, the Box and Jenkins' ARIMA models as applied in [1-5], neural-network-based algorithms in [8], Principle Component Analysis for European data in [7] and transfer function models in [9].

This paper presents the application of Box-Jenkins' ARIMA time series model with the application to National Electronic Fund Transfer (NEFT) data for banks in India. Two time series models, namely, the multiplicative decomposition model and the ARIMA model are employed The multiplicative decomposition technique has been profoundly used for forecasting tasks in the business sector, such as sales projection or financial forecasting; however, it has not been commonly employed for electronic transactions forecasting. This may due to the fact that transactions usually vary to a large extent and it is challenging to fit a trend line.

The proposed models are implemented to predict one year transaction data. The accuracy of the two models are calculated and compared. The paper utilizes the mean absolute percentage error (MAPE) as a measure of forecast accuracy.

The paper is organized as follows. Section II introduces the two above mentioned models and explains how they would be used in analyzing and forecasting the load consumption with a brief note on forecasting error measurement. Section III is the case study whereby Indian bank electronic transactions (NEFT) data sets are first described for Government banks, Private banks, and Foreign banks. Then models would be applied with forecasts generated and compared with the transactions actually occured.

Time Series Models

A time series is defined as a set of data generated sequentially in time. The time series models assume that in the absence of major disruptions to critical factors of a recurring event, the data of this event in the future will be related to that of the past events and can be expressed via models developed from the past events. In this analysis, two time series models, the Multiplicative Decomposition Model and the Seasonal ARIMA Model, are employed and presented in the followings.

Multiplicative Decomposition Model

Multiplicative Decomposition Model assumes that a time series can be described as (1).

$$x(t) = T(t)*S(t)*C(t)*R(t),\ t = \ldots -1,0,1,2 \qquad (1)$$

where x(t) is the time series, T(t) is the trend component, S(t) is the seasonal component, C(t) is the cyclic component and R(t) represents the random component.

Cyclic component is usually in the duration of many years and is not applicable in short term transactions forecasting. Thus, we propose to simplify the above combination to only three terms as shown in (2):

$$x(t) = T(t)*S(t)*R(t),\ t = \ldots -1,0,1,2 \qquad (2)$$

In order to apply this model, it requires that the trend of the time series be found and extended into the future. The trend component of a series made up of these three components can be found if the other two could be taken off the series. A typical transaction series contains monthly seasonal indexes, which when used to divide a typical month's data, would remove the seasonal component from the series. To find the indexes, data of the most recent month is divided by the average of a few recent months and using the average of these weeks tends to minimize the random effect. With these two components removed or minimized, the series now contains mainly the trend component. The equation of this trend line is extrapolated to estimate the trend in the future. Seasonal effects can then be incorporated into these future trend forecasts to account for the variation and obtain reasonably comprehensive forecasts.

Steps

1. Identify the seasonal period (Quarter, months e.t.c)
2. Develop a Moving Average (MA) forecast (Regular MA can be used for forecasting)
3. Find the ratio of each observation to the MA forecast: $r_t = Y_t/S_t$, where Y_t is the absolute value and S_t is the observation which we developed.
4. Find the average of the ratios for each month, season or periodic unit. If there is k periods then we have k average ratios. These are unadjusted seasonal indexes.
5. Adjust the ratios: Divide each of the k ratios by the average of the k ratios, these are the adjusted seasonal indexes.
6. Adjust the series: For each observation, divide the observation by its adjusted seasonal index. It is the deseasonalized series which gives the seasonally adjusted series.

Deseasonalized Monthly Trend

A deseasonalized monthly performance is obtained by dividing the latest month's data (or the previous month's, whichever is used) by the monthly seasonal indexes correspondingly. This is the deseasonalized monthly trend which is expected to appear as a straight trend line with small variations due to random effects that we cannot remove completely.

Forecasting

Equation to the weekly trend can be easily obtained by mathematical derivation or software tools, such as MATLAB. This will then be used to project into the future for future trend in a process called trend forecasting. The actual forecast is done by multiplying this forecasted trend with the monthly seasonal indexes.

Seasonal ARIMA Model

Two of the most basic models in time series are the autoregressive model (AR) and the moving average model (MA). In autoregressive models, the next value in the time series is represented as a linear combination of p previous values and a random shock.

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + \omega_t \qquad (3)$$

where

x_t = An observation at time t of a time series

ϕ_i = Autoregressive component parameter of lag i observation

ω_t = Random shock component of a time series

The backshift operator $B\ x_t = x_{t-1}$ or $B^m\ x_t = x_{t-m}$ and the autoregressive operator $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$ are introduced so that the expression (3) simplifies to (4):

$$\phi(B)\ x_t = \omega_t \qquad (4)$$

where B^i is a backshift operator of lag i.

MA models assume that the next observation is made up of q previous random shocks.

$$x_t = \omega_t + \Theta_1 \omega_{t-1} + \Theta_2 \omega_{t-2} + \ldots + \Theta_q \omega_{t-q} \qquad (5)$$

where Θ_i is a moving average component parameter of lag i observation.

Similarly, moving average operator is defined as:

$$\Theta(B) = 1 + \Theta_1 B + \Theta_2 B^2 + \ldots + \Theta_p B^p$$

And (5) can be converted to the form (6)

$$x_t = \Theta(B)\ \omega_t \qquad (6)$$

When a process involves characteristics of both AR and MA models, an autoregressive moving average model, or ARMA can be used.

$$x_t = \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + \omega_t + \Theta_1 \omega_{t-1} + \ldots + \Theta_q \omega_{t-q}$$

Equivalently we have (7)

$$\phi(B)\ x_t = \Theta(B)\ \omega_t \qquad (7)$$

However, if the series is non-stationary, i.e. when its statistical properties such as mean and variance change over time, differencing is required to transform the series into a stationary one. Differencing is essentially finding the difference between values in the series separated at certain lags k, denoted by $\nabla^d{}_k$ where d is the order of differencing, i.e. $\nabla^d = (1-B)^d$ and k is the number of lags. Differencing results in autoregressive integrated moving average model, or ARIMA, represented as:

$$(B)\ \nabla^d{}_k\ x_t = \Theta(B)\ \omega_t \qquad (8)$$

Seasonal variations can be observed in the transactions data in that a transaction at any point of time might be similar to that of the previous month and that of the previous year, it is hence advantageous to use a Seasonal ARIMA model. Incorporating seasonal effects of ARIMA order (P,D,Q), the Seasonal ARIMA Model can be written as (9).

$$\phi_P(B^S)\phi_p(B)\ \nabla^D{}_S\ \nabla^d{}_k\ x_t = \Theta_Q(B^S)\ \Theta_q(B)\ \omega_t \qquad (9)$$

Identification of models usually relies on analysis of the autocorrelation function (ACF) and partial autocorrelation function (PACF). The autocorrelation function measures the correlation between values in a time series separated by k, which represents the number of lags between these observations. The partial autocorrelation function provides indication in determining the number of lags in the AR models. Table- 1 summarizes rough guideline for using these parameters in initial model identification.

Table 1- Model Identification using ACF and PACF

	AR(p)	MA(q)	ARMA(p,q)
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

For both models, error measurement is carried out using the mean absolute percentage error, or MAPE, calculated as follows.

$$\mathrm{M} = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|,$$

where A_t is the actual data and F_t represents the forecast and n denotes the number of forecasts made.

Steps

1. Plot the data : Time series Plot
2. Identify the orders (p,d,q) of the model: Inspection of the TS plot may help to identify the differencing order d, while inspection of the sample ACF and PACF of the differenced data may help to identify the AR order p and MA order q.
3. Estimation of the Model Parameters ϕ and Θ
4. Prediction: forecast future values of the time series and also generate confidence intervals for these forecasts from the ARIMA model.

Case Studies With Indian Bank NEFT Data

Monthly data is collected from rbi.org for NEFT (National electronic fund transfer) in India for the years 2009, 2010, 2011 from which two variables are taken: time period and transactions. Preprocessing is done and Missing values are replaced and the transactions data is classified into three categories (Government Banks, Private bank and Foreign banks). Data is given for each month for NEFT for about 90 Banks operating in India. Out of it clusters of 5 Government Banks, 5 Private Banks and 5 Foreign Banks are taken and total transactions (inward + outward) is calculated for each month. The banks which are chosen are the top five banks in each category which do the maximum business.

Multiplicative Decomposition Model

In this case, the multiplicative decomposition technique is used to predict the transactions for the months of 2012, data from the months of three years(36 months) preceding 2012, namely 2011 (Jan to Dec), 2010(Jan to Dec) and 2009(Jan to Dec) are used.
First, the 12-months MA is calculated and dividing the observed transactions with the MA gives the ratio-to-moving-average. After that we find the average of the ratios for each month. In our case we have 12 periods so we get 12 average ratios. These are unadjusted seasonal indexes.

Next, we adjust each of the ratios by the average of the 12 ratios, these are the adjusted seasonal indexes. After that for each observation, we divide the observation by its adjusted seasonal index. It is the deasonalized monthly series which gives the seasonally adjusted series.

The trend line equation is found from simple regression to be y = 2.503x – 2.500 for Government banks data, y = 2.979x + 10.14 for private banks data and y = 1.114x + 14.61 for Foreign banks data. The x is then set to be from 37 to 48 and applied to the trend equation to project into future 12 months. This way, trend forecasts are obtained. Multiplying them with the weekly seasonal indexes will give the final forecasts. The overall forecasts and actual transactions are put into a plot for comparison as shown in Fig. 1, Fig. 2, and Fig. 3 for Government, Private, and Foreign banks respectively. The MAPE calculated from different years are given in Table- 2.

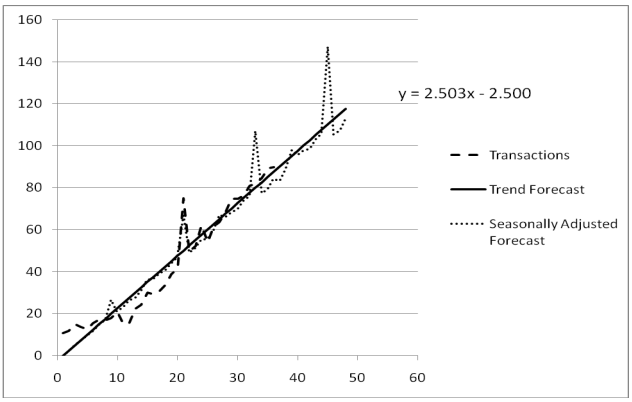

Fig. 1- Shows forecasts and actual load comparisons for Government Banks with the trend line equation, x-axis shows time period in months while y-axis shows transactions in millions

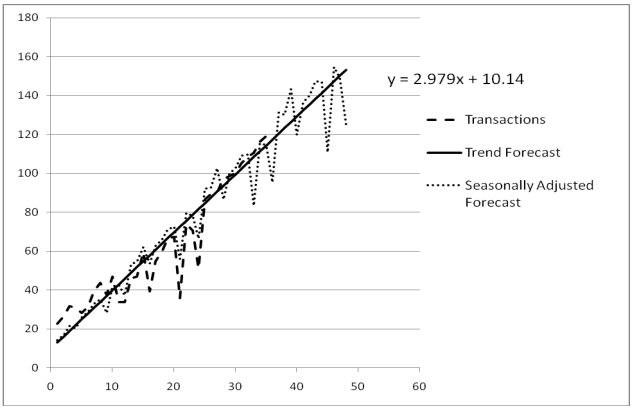

Fig. 2- Shows forecasts and actual load comparisons for Private Banks with the trend line equation, x-axis shows time period in months while y-axis shows transactions in millions

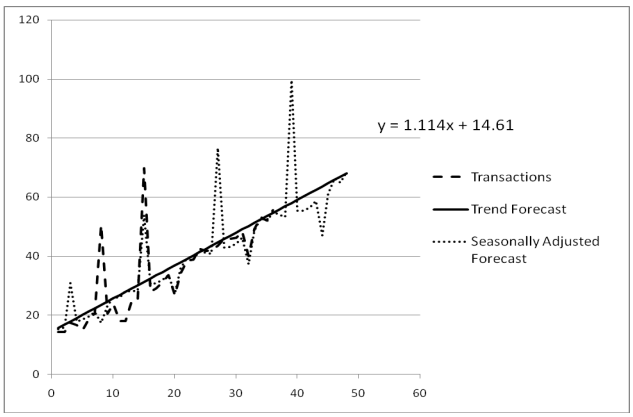

Fig. 3- Shows forecasts and actual load comparisons for Foreign Banks with the trend line equation, x-axis shows time period in months while y-axis shows transactions in millions

Table 2- Mape Using Multiplicative Decomposition Model

Year	MAPE (Government)	MAPE (Private)	MAPE (Foreign)
1	3.70%	1.80%	2.20%
2	1.40%	1.50%	0.50%
3	0.60%	0.60%	0.80%
Overall	1.90%	1.30%	1.20%

As observed, the MAPE generally decreases as the forecasts are done for the years further into the future.

Seasonal ARIMA Model

In this case, the Seasonal ARIMA technique is used to predict the transactions for the months of 2012, data from the months of three years(36 months) preceding 2012, namely 2011(Jan to Dec), 2010 (Jan to Dec) and 2009(Jan to Dec) are used.

The first step in identifying a suitable model is to examine the ACF plot. The ACF plots indicate non-stationary series shown in Fig. 4, Fig. 5, and Fig. 6 for Government, Private and Foreign banks respectively. Differencing is needed to obtain a stationary one. A few differencing schemes have been tried and the resulting ACF plots are shown in Fig. 7, Fig. 8, and Fig. 9 for Government, Private and Foreign banks respectively.

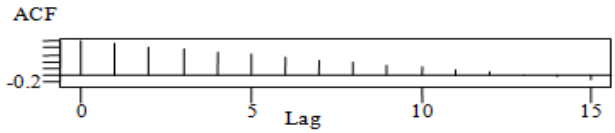

Fig. 4- ACF Plot of years 2009, 2010, 2011 Data for Government Banks

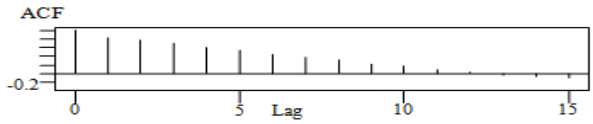

Fig. 5- ACF Plot of years 2009, 2010, 2011 Data for Private Banks

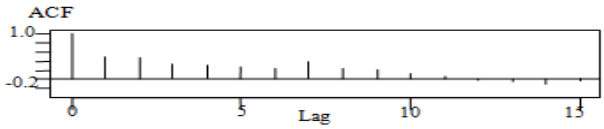

Fig. 6- ACF Plot of years 2009, 2010, 2011 Data for Foreign Banks

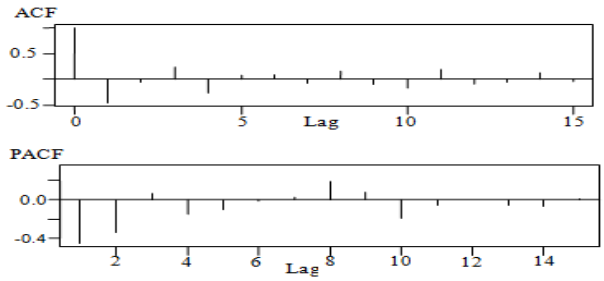

Fig. 7- ACF & PACF Plot after Differencing for Government Banks – ARIMA(0,1,0)

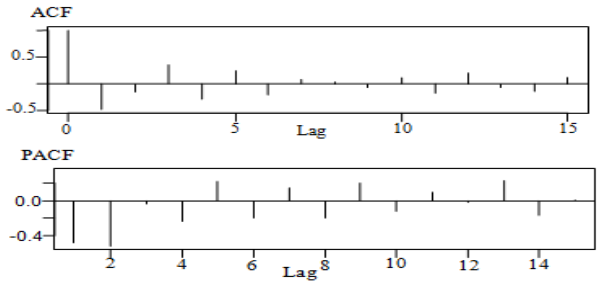

Fig. 8- ACF & PACF Plot after Differencing for Private Banks – ARIMA (0,1,0)

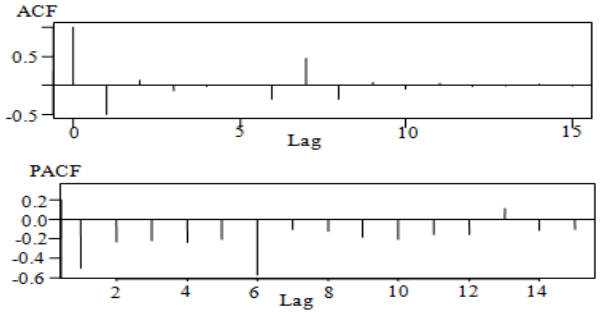

Fig. 9- ACF & PACF Plot after Differencing for Foreign Banks – ARIMA(0,1,0)

For Government banks, examination of the ACF and PACF plot suggests AR(1) and MA(0) components with 1-differencing scheme. Using statistics software R, the following seasonal ARIMA model of the form ARIMA(1,1,0)*(1,1,0)$_{12}$ is obtained.

$$(1-(-0.9063)B^{12})\,(1-(-0.3092)B)\,\nabla_1\,\nabla_{12}\,x_t = \omega_t$$

For Private banks, examination of the ACF and PACF plot suggests AR(0) and MA(0) components with 1-differencing scheme. Using statistics software R, the following seasonal ARIMA model of the form ARIMA(1,1,0)*(0,1,0)$_{12}$ is obtained.

$$(1-(-0.4628)B)\,\nabla_1\,\nabla_{12}\,x_t = \omega_t$$

For Foreign banks, examination of the ACF and PACF plot suggests AR(1) and MA(0) components with 1-differencing scheme. Using statistics software R, the following seasonal ARIMA model of the form ARIMA(0,1,0)*(1,1,0)$_{12}$ is obtained.

$$(1-(-0.6821)B^{12})\,\nabla_1\,\nabla_{12}x_t = \omega_t$$

These models are then used to forecast the data for the next year, i.e. the months in 2012. Fig. 10, Fig. 11, and Fig. 12 is showing the future predictions for Government, Private and Foreign banks respectively, prediction bounds are also included in the figures because it expresses how uncertain it is where the value may fall in. The prediction errors are calculated and shown in Table 3.

Table 3- Mape Using Seasonal Arima Model

	MAPE (Government)	MAPE (Private)	MAPE (Foreign)
Overall	2.20%	3.30%	2.10%

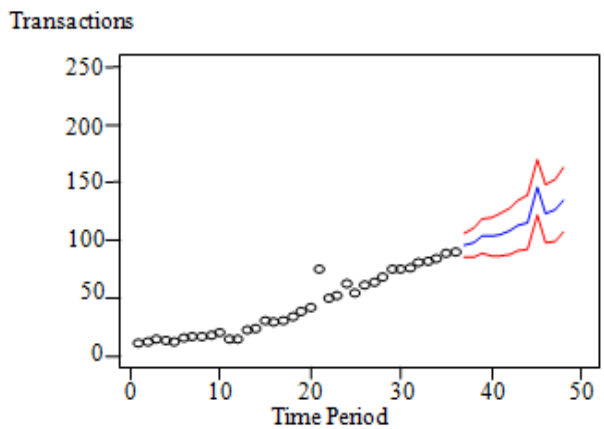

Fig. 10- showing future transactions predictions for next year(2012) for Government Banks

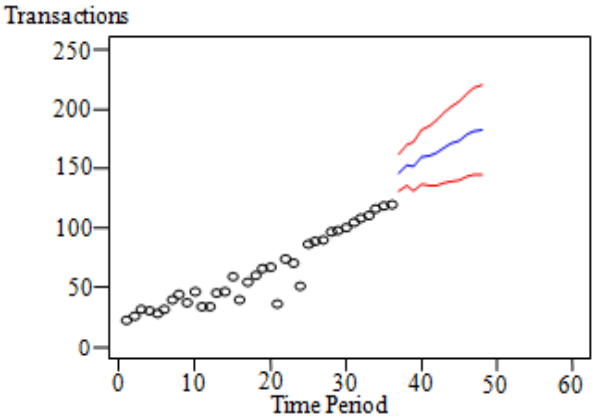

Fig. 11- showing future transactions predictions for next year(2012) for Private Banks

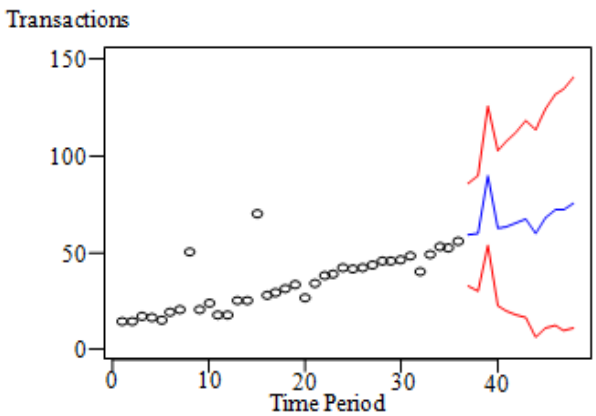

Fig. 12- showing future transactions predictions for next year(2012) for Foreign Banks

Model Performance Comparison

In this section, forecasts from the two models are compared together with the actual usage. It is observed that forecasts from the Multiplicative Decomposition Model are closer to the actual transactions than the forecasts from ARIMA Model. The decomposition technique tends to match the forecasts with the transaction behavior of the previous year which is very similar in this case and hence better captures the trend.

Table IV compares the MAPE of the forecasts generated by each of the models. It can be seen that the MAPE from Multiplicative decomposition model is considerably less than that from the ARIMA Model.

Table 4- Mape Using (1) Multiplicative Decomposition Model (2) Seasonal Arima Model

BANKS	MAPE(1)	MAPE(2)
Government	1.90%	2.20%
Private	1.20%	3.30%
Foreign	1.30%	2.10%

Conclusion

Transactions load forecasting has significant implications on the costs and speed of the Banking transactions. Accurate forecasting models are needed for secure and reliable banking system operations. This paper introduces and applies two time series methodologies to short-term transactions forecasting, with case study using India's banking transactions load demand. These two models have shown favorable forecasting accuracy with the Multiplicative Decomposition Model outperforms the Seasonal ARIMA Model.

References

[1] Hagan M.T. and Behr S.M. (1987) *IEEE Trans. Power Syst.*, PWRS-2(3).
[2] Conejo A.J., Plazas M.A., Espinola R. and Molina A.B. (2005) *IEEE Trans. Power Syst.*, 20(2).
[3] Yang H.T., Huang C.M. and Huang C.L. (1996) *IEEE Trans. Power Syst.*, 11(1).
[4] Liu K., Subbarayan S., Shoults R.R., Manry M.T., Kwan C., Lewis F.L. and Naccarino J. (1996) *IEEE Trans. Power Syst.*, 11(2).
[5] Espinoza M., Joye C., Belmans R. and Moor B.D. (2005) *IEEE Trans. Power Syst.*, 20(3).
[6] Papalexopoulos A.D. and Hesterberg T.C. (1990) *IEEE Trans. Power Syst.*, 5(4).
[7] Taylor J.W. and McSharry P.E. (2007) *IEEE Trans. Power Syst.*, 22(4).
[8] Abu-El-Magd M.A. and Findlay R.D. (2003) *IEEE CCECE Canada Conf.*, 3, 1723-1726.
[9] Cho M.Y., Hwang J.C. and Chen C.S. (1995) *EMPD International Conf.*, 1, 317-322.
[10] Povinelli R.J. and Feng X. (1999) *Artificial Neural Networks in Engineering*, St. Louis, Missouri, 511-516.
[11] Povinelli R.J. and Feng X. (1998) *Artificial Neural Networks in Engineering*, St. Louis, Missouri, 691-696.
[12] Pandit S.M. and Wu S.M. (1983) *Time Series and System Analysis with Applications*, New York: Wiley.
[13] Box G.E.P. and Jenkins G.M. (1994) *Time Series Analysis: Forecasting and Control*, 3rd ed. Englewood Cliffs, N.J.: Prentice Hall.
[14] Bowerman L. and O'Connell R.T. (1993) *Forecasting and Time Series: an Applied Approach, 3rd ed*. Belmont.
[15] Povinelli R.J. (1999) *Electrical and Computer Engineering Department*. Milwaukee, Wisconsin: Marquette University.