Improving efficacy of library Services: ARIMA modelling for predicting book borrowing for optimizing resource utilization

K Kumar
kumarkutty@gmail.com

M M. Alpha Raj

Follow this and additional works at: http://digitalcommons.unl.edu/libphilprac

Part of the Library and Information Science Commons

http://digitalcommons.unl.edu/libphilprac/1395
Improving efficacy of library Services: ARIMA modelling for predicting book borrowing for optimizing resource utilization

K. Kumar\textsuperscript{a} and M.Alpha Raj\textsuperscript{b}

\textsuperscript{a}Assistant Professor, Library and Information Science, College of Veterinary Science, Sri Venkateswara Veterinary University, Proddatur – 516 360, Email: kumarkkutty@gmail.com

\textsuperscript{b}Assistant Professor, Dept. of Pharmacology and Toxicology, College of Veterinary Science, Sri Venkateswara Veterinary University, Proddatur – 516 360, Email: alpharajm@gmail.com

Abstract

Book borrowing is a key service in libraries. Library users frequently visit the library for borrowing books compared to other library services. To predict book-borrowing service in a college library, Auto Regressive Integrated Moving Average (ARIMA) model has been developed from the data pertaining to book borrowing during the year 1998 to 2013. The study found that the number of books borrowed one month and twelve months earlier could estimate the number of books borrowed in a month. The study used a fitted model for predicting book borrowing for the year 2014 by two alternative approaches: 12-steps ahead versus 1-step ahead. The calculations show that there was no significant difference (P=0.928; Wilcoxon signed rank test) between 1-step and 12-steps ahead approach for predicting book borrowing. However, the Root Mean Squared Error (RMSE) in 1-step ahead approach (109.57) was lower than 12-steps approach (131.33). The study findings indicate that ARIMA models are useful for monitoring book borrowing in institutional libraries. Furthermore, these models can predict library usage trends.

Keywords: ARIMA Modelling; Book borrowing; Library; Prediction

Introduction

Academic libraries are an essential component of colleges and universities throughout the world and play a central role in academic life\textsuperscript{1}. Regular library usage by students improves critical thinking\textsuperscript{2},\textsuperscript{3}, student learning and engagement, collaborative learning, student-faculty interactions, and academic challenges\textsuperscript{4},\textsuperscript{5}. Even the academic performance of students improves with library usage\textsuperscript{6},\textsuperscript{7}. Academic librarians face immense pressure to meet the needs of ever-increasing users and commitments for modernization and improvements\textsuperscript{8}. Further, with the advent of information-communication revolution, academic libraries are exploring service developments to support a series of new scenarios such as publication and scholarly communications; extensive use of digital resources; increasing heterogeneous student
population; ICT-based learning and distance learning. Hence, libraries with limited resources, particularly in terms of staffing face enormous challenges in management. Librarians need to predict library usage over time, as this helps in planning the resources required to meet the usage. The future usage might not be stationary and there could be periods of huge demand, interleaved with long periods. A model for predicting the library usage should take into account this variability to be accurate in its prediction. Hence, seasonal prediction models are likely to improve the accuracy of prediction. The study proposes a prediction model that forecasts library usage for a period of 12 months.

Time series analysis allows examining temporal dependency of data and generating predictions. Time series consists of two components: trend and seasonality. Trend depicts a linear or non-linear component that is non-repetitive within the time range, whereas seasonality repeats at systematic intervals over time. Time series analysis using Autoregressive Integrated Moving Average (ARIMA) models are very effective in modelling the time-dependent data as they structure time dependence between the observations. The predictions made using ARIMA models are more accurate than those obtained by other statistical methods. Earlier, ARIMA models are successfully forecasted household electric consumption; software evolution; incidence of several infectious diseases and use of health facilities. However, such predictions are not available in the context of library usage. Hence, in this study, the authors used Box-Jenkins ARIMA modelling to monitor and predict book borrowing in Sri Venkateswara College of Engineering and Technology, Chittoor. The data pertaining to book borrowing from 1998 to 2013 fit into an ARIMA model using Box-Jenkins approach. The fitted model has been used to predict book borrowing for the year 2014 with 12-steps ahead and 1-step ahead approach. Root mean squared error (RMSE) was used to compare the predictive power of the two approaches. Further, both approaches have been tested using Wilcoxon signed rank test.

Objectives of the study

- To detect any seasonal pattern in the book borrowing services and identify the underlying factors associated with it;
- To identify trends in the distribution of book borrowing over time; and
- To validate the usage of ARIMA model with seasonal information to accurate forecasting of book borrowing.
Methodology

The study was conducted at the college library of Sri Venkateswara College of Engineering and Technology, Chittoor established in 1998. For the analyses, monthly data on book borrowing for the period 1998–2013 was used to fit an ARIMA model, which later used to predict out-of-sample book borrowing for the year 2014. Since estimation procedures are available only for stationary series, Augmented Dickey-Fuller (ADF) unit root test and Box-Jenkins approach on ARIMA modelling of time series, used to check the stationary series of the data, which was ascertained by observing the plot of book borrowing that involves a four-step process.\textsuperscript{11,15,24,28-33}

Based on the correlogram of Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF), the order of non-seasonal (p, d, q) and seasonal (P, D, Q) component; Auto Regressive (AR) parameters (p and P); Moving Average (MA) parameters (q and Q); the need for non-seasonal and seasonal differencing (d and D)\textsuperscript{34,35} determined. The goodness of fit of the model validated using Bayesian Information Criterion (BIC) and Ljung-Box statistics. In the identified ARIMA model, two approaches employed to estimate the values for the year 2014, in which out of sample predicted values with 95% prediction intervals are calculated.

- The first approach was to predict the subsequent 12-steps ahead for the entire year 2014 and second to predict 1-step ahead, that is, January 2014 only.
Then, as the observed value for January 2014 obtained, the authors included the same in the model and re-estimated the ARIMA model, and predicted next 1-step, that is February 2014, this process continued until the end of the year 2014.

The predictive power of both approaches by RMSE accessed, which measures the difference between fitted and observed values. Lower RMSE indicates increased predictive power. Further, the analysis applied Wilcoxon signed-rank test to assess the statistical difference in the errors between both approaches\(^{15, 29}\). ADF unit root test and Wilcoxon signed rank test carried out in Microsoft Excel 2013 and ARIMA modelling in the Statistical Package for Social Sciences (IBM SPSS 19.0 V). Expert modeller of SPSS used for 1-step ahead predictions.

**Analysis**

The box plot of book borrowing showed no trend and the data remained stationary (Figure 1). In addition, ADF test (statistic = -4.75; p=0.01; lag order = 5) showed that the data was stationary without significant trend. Correlogram of sample plots of ACF and PACF indicated that non-seasonal and seasonal parameters are required in the model (Figure 2 and 3). The sinusoidal exponential decay in ACF associated with an oscillating PACF suggested that the model should contain a non-seasonal AR and MA components. In the time lag (lag=12) correlogram, the ACF showed exponential decay, whereas PACF showed spikes at lags one and two, and later no significant spikes suggesting a seasonal AR (P =2) (Figure 4 and 5). After analysis, the best model was found to be a multiplicative ARIMA \((1,0,1) \times (2,0)_{12}\) (Normalized BIC=8.898, Table 1) on the non-differenced book borrowing data (multiplicative ARMA\([2,1]\times[2,0]_{12}\), as no differencing required). After developing the model, the plots of residuals of ACF and PACF showed no remaining temporal correlation (Figure 6) indicating that, the model is a good fit. Higher seasonal ARIMA models like \((2,0,1) \times (2,0,0), (2,0,2) \times (2,0,0)\) and \((1,0,2) \times (2,0,0)\) also fitted the data, but they were rejected based on Occam’s razor-among competing models that predict equally well, the one with the least complexity should be selected. Seasonal ARIMA \((1,0,1) \times (2,0,0)_{12}\) model equation (Table 2)

\[
\hat{Y}_t = Y_t - \phi_1 Y_{t-1} + \Phi_1 Y_{t-12} + \Phi_1 \phi_1 Y_{t-13} + \Phi_2 Y_{t-24} + \Phi_2 \phi_1 Y_{t-25} + \theta_1 Y_{t-1} - \theta_1 \theta_1 Y_{t-2} - \Phi_1 \theta_1 Y_{t-13} + \Phi_1 \phi_1 \theta_1 Y_{t-14}
\]

Where, \(\hat{Y}_t\) = Predicted value; \(Y_t\) = Constant
\(\phi_1\) = Non-seasonal Auto Regression coefficient
\(\theta_1\) = Non-seasonal Moving Average coefficient
$\Phi_1, \Phi_2 =$ Seasonal Auto Régression coefficients

Fig. 1 -- Box plot of book borrowing cases from the year 1998 to 2013 presenting no particular trend

Fig. 2 -- ACF shows a sigmoid curve with exponential decay
Fig. 3-- PACF showing multiple peaks until lag25 and then cutting off

Fig. 4-- ACF at periodic lag of 12 indicating exponential decaying
Fig. 5-- PACF at a periodic lag of 12 indicating spikes at lags 1 and 2 and cutting off later

In addition, Ljung–Box test statistic was calculated, which supported the independence of the time series residuals (Statistic=12.81;df=15; P=0.617). The model fitted (1998–2013) and predicted values (the year 2014) followed the observed book borrowing in the same year (Figure 7). Graphically, the predicted values estimated using the 12-step ahead approach showed less dispersion from observed values (RMSE=139.87) than 12-step ahead approach (RMSE=140.30) (Figure 8). The Wilcoxon signed rank test statistic "W" is simply the smallest of the rank totals. The smaller it is then the less likely it is to have occurred by chance. Wilcoxon signed rank test showed that the predictions of 12-step ahead approach were significantly different than 1-step ahead approach (W= 9; Critical value = 10).
Fig. 6--Residual ACF and PACF plots after modelling illustrate no remaining temporal correlation

Fig. 7--Plot of the observed and forecasted book borrowing
Fig. 8--Plot of actual book borrowing Vs 12-step ahead and 1-step ahead approach predictions

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Stationary R-squared</td>
<td>0.629</td>
</tr>
<tr>
<td>2</td>
<td>R-squared</td>
<td>0.629</td>
</tr>
<tr>
<td>3</td>
<td>Root Mean Squared Error (RMSE)</td>
<td>80.978</td>
</tr>
<tr>
<td>4</td>
<td>Mean Absolute Percentage Error (MAPE)</td>
<td>20.725</td>
</tr>
<tr>
<td>5</td>
<td>Mean Absolute Error (MAE)</td>
<td>57.568</td>
</tr>
<tr>
<td>6</td>
<td>Maximum Absolute Percentage Error</td>
<td>357.159</td>
</tr>
<tr>
<td>7</td>
<td>Maximum Absolute Error</td>
<td>291.353</td>
</tr>
<tr>
<td>8</td>
<td>Normalized Bayesian Information Criterion</td>
<td>8.898</td>
</tr>
</tbody>
</table>
Discussion and Conclusion

Book borrowing is an important aspect of libraries and attracts many people to the library. Academic libraries provide a unique opportunity for students to borrow books free of cost. This often benefits students who generally have limited financial resources and cannot afford to buy high priced books. Hence, they prefer to borrow books rather than to buy them, at the same time, book lending is a challenging aspect of librarians requiring more resources to meet the demand. Therefore, prior information on the number of book borrowings in a month will help the librarian to plan and prepare for the task effectively. Various techniques proposed to forecast a given time series; studies show that models from the ARIMA successfully used, as well as regression approaches based on e.g. linear, non-linear regression, neural networks, and Support Vector Regression. The difference lies in forecasting methodology. The application to plenty of time series underlines its usefulness.

In this study, an ARIMA model developed, that closely fits book borrowing in the college library of SVCET, Chittoor. The Auto Regressive and Moving Average parameters of the model entail the number of book borrowing cases in a month can be estimated by previous cases occurring one or twelve months prior; found from the analysis that 12-step ahead approach better agrees with the observed book borrowing and was significantly different (P<0.05) from 1-step ahead prediction. The model indicated that peak book borrowing occurs in August (539.99 [310.43-769.55] and least in September (364.72 [131.62-597.81]). However, an increasing trend in book borrowing observed starting from June (505.69 [286.21-725.17] through August (539.99) [310.43-769.55]. This increase in trend is a result of semester examinations, scheduled during the months of June and July every year. Similarly, the actual book borrowing data showed peak borrowing during the month of June (517).

ARIMA models of time series applied to a variety of library functions including

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Lag No.</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>T-value</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td></td>
<td>356.727</td>
<td>74.41</td>
<td>4.92</td>
<td>0.000</td>
</tr>
<tr>
<td>AR</td>
<td>Lag 1</td>
<td>0.900</td>
<td>0.037</td>
<td>24.64</td>
<td>0.000</td>
</tr>
<tr>
<td>MA</td>
<td>Lag 1</td>
<td>0.346</td>
<td>0.084</td>
<td>4.120</td>
<td>0.000</td>
</tr>
<tr>
<td>AR, Seasonal</td>
<td>Lag 2</td>
<td>0.590</td>
<td>0.061</td>
<td>9.724</td>
<td>0.000</td>
</tr>
</tbody>
</table>
If the numerical data can be obtained. This procedure also used to evaluate predictive scenarios for search engine transactional logs, resource allocation, and usage for server monthly circulation of library books, and more. The study concludes that ARIMA modelling successfully predicted book borrowing in 12-step ahead approaches. However, forecasting requires the presence of ‘cause and effect’ relationship successfully predict into the future. In this study, book borrowing in academic library appears directly related to the examination schedule. Hence, these predictions used successfully deploy working staff in libraries. Moreover, such predictions are also helpful in making a strong case for increasing budgetary allocations and workforce for libraries.

References
11. Helfenstein U, The use of transfer function models, intervention analysis and related time


(Accessed on 05 May 2015).


