Journal of Engineering and Applied Sciences 12 (12): 3102-3107, 2017
ISSN: 1816-949X

Analysis of Time-Series Method for Demand Forecasting

Aayushi Verma, Aishwarya Karan, AarohMathur and S. Chethan
Department of Information and Communication Technology,
Manipal Institute of Technology, Manipal, India

Abstract: This study focuses on analyzing the demand for raw material supply for a university dining mess using time series method which falls under quantitative approach of demand forecasting and gives a detailed step by step process of analysis. This method accurately forecasts the daily demand within a span of four months, thereby eliminating the concern of mess authorities on their budgeting and supply of raw material for university dining mess on a daily basis.

Key words: Time-series, forecasting, quantitative method, data analysis, ARIMA Model

INTRODUCTION

Demand forecasting has always been a critical business parameter which helps in resource optimization and improvement in productivity. It essentially consists of determining the expected level of demand during a certain time period under consideration (Meimanda *et al*., 2012). It depends on prediction of consumer trends, competitor and supplier actions, seasonal variations, government and regulatory framework. It makes use of both, data sources which provide statistical framework and judgmental sources which are based on analogy and domain knowledge of expert. In order to estimate future demand it is important to plan the production level and make arrangements for consuming resources.

Figure 1 throws light on the two approaches of demand forecasting which are qualitative and quantitative. Quantitative approach include the following methods which are time series, moving average, regression analysis, econometric model, etc. qualitative methods can be categorized into-expert opinion method and consumer survey method (Box and Jenkins, 1970). Quantitative methods require the existence of past data and thus help in forecasting future demand of mainly the existing data and can forecast over a long period of time whereas qualitative approaches don't require any historic data since they rely on pooled expert opinions and are suited for a short term demand forecasting. In other words qualitative methods require awareness of experts with the current developments in their field and thus provide reasonably good forecast. Qualitative approaches are used where the past data are inappropriate for processing.

This information helps us in understanding how the above two categorization of demand forecasting are based on the nature and the scope of the product whether it is an existing one or to be released in the market. Therefore, helping us in identifying where which method can be applied and analysis can be done accurately. Through our case study on university dining mess we focus on one of the quantitative methods for demand forecasting-time series.

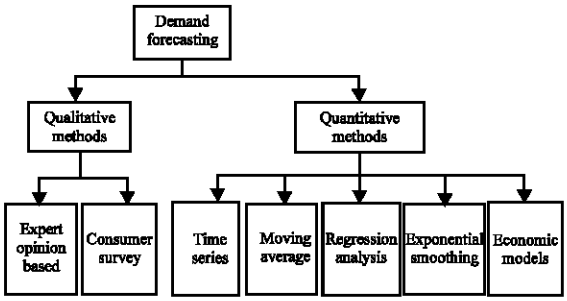

Fig. 1: Classification of forecasting techniques

MATERIALS AND METHODS

Review stage: Our database consists of university dining mess records for a particular semester which spans a period of four months (122 days). It contains the details of all the active students and the total number of meals they have consumed in a month.

Traditional qualitative approaches such as delphi method, Judgment-aided model requires experts to agree on a common forecast. Identifying of experts in the suitable field is also one big challenge in this approach.

We found that quantitative approach is more appropriate for our case study as it consists of methods

Corresponding Author: Aayushi Verma, Department of Information and Communication Technology, Manipal Institute of Technology, Manipal, India

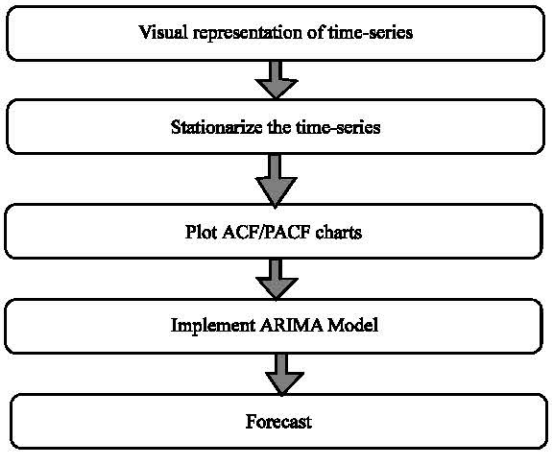

Fig. 2: The schema for the application of ARIMA time-series modeling

such as time series which is a banal tool in forecasting demands. We can use time series method to analyze the trends and also the seasonal and cyclic factors that can influence the demand for a certain product. It consists of working on discrete data which are time-based to produce hidden acumen for revised decision making. Time-series modeling is very useful when we have a correlated temporal data (De Gooijer and Hyndman, 2006). The commonly used model in time series modeling is known as the ARMA Model in which AR means autoregression and MA means moving average. This model is applicable only on a stationary series. AR and MA Models differ on the basis of the analogue between time series objects at various different points. Instead of using regression model, MA Model uses past prediction errors in a regression model. Regression over a variable itself is implied by autoregression. In autoregressive models, using a linear consolidation of the previous values of the variable, the variable of interest is forecasted (Fig. 2).

RESULTS AND DISCUSSION

Following we present the steps which are followed to do a time series analysis on the university dining mess dataset.

Step 1 (visual representation of time-series): Brockwell and Davis (1996) prior to building a time series model, it is of prime importance to analyze the trends. Since, our dataset corresponds to a non-seasonal time series, it consists of two components. One being the trend component and other being an irregular component. To estimate these two components the time series is decomposed.

To analyse the trend component of a non-seasonal time series which can be visualized using an additive model equation, smoothing methods are used for measuring simple moving average of a time series (Fig. 3). Smoothed time series data when plotted can help in estimating the trend component. It removes or smoothens the random fluctuations found in the dataset (Fig. 4).

Inferences:

- Over the span of 4 months (Aug'16-Nov'16-122 days), a decreasing trend in observed
- There is a sudden increase in the number of students in the initial days as day by day, more students join thedining mess as soon as they arrive at the university
- The deep dip found during late September is because of the one week holidays given to freshers
- Many fluctuations are found in the graph which are mainly because of the holidays occurring in that period and thus students tend to eat outside the campus instead of at the mess

This exploration of data through plotting graphs is very important part in determining the stationary or non-stationary nature of the dataset.

Step 2 (stationarize the time series): After having a clear understanding of the pattern, trend and cycle, now we can check if the time series is stationary or not. There are three common techniques which can be used to transform a non-stationary time series to a stationary time series.

Detrending: Removing the trend component from the series.

Differencing: Differences of the terms are modeled instead of the actual term. Differencing contributes to the Integration part in AR(I)MA Model. The parameters associated with this can be found using Auto Correlation Function (ACF) and Partial ACF plots.

Seasonality: Incorporated into the ARIMA Model directly (Fig. 5). Now to address the issue of trend component of our dataset being stationary or not, we perform differencing which might help in stabilizing the mean of a time series by removing changes in the level of a time series which in turn eliminates the trend and seasonality (Dickey and Fuller, 1979).

We perform the Augmented Dickey-Fuller (ADF) test for the null hypothesis of a unit root of the univariate (non-stationary) time series using the function adf.test. The augmented Dickey-Fuller test includes three

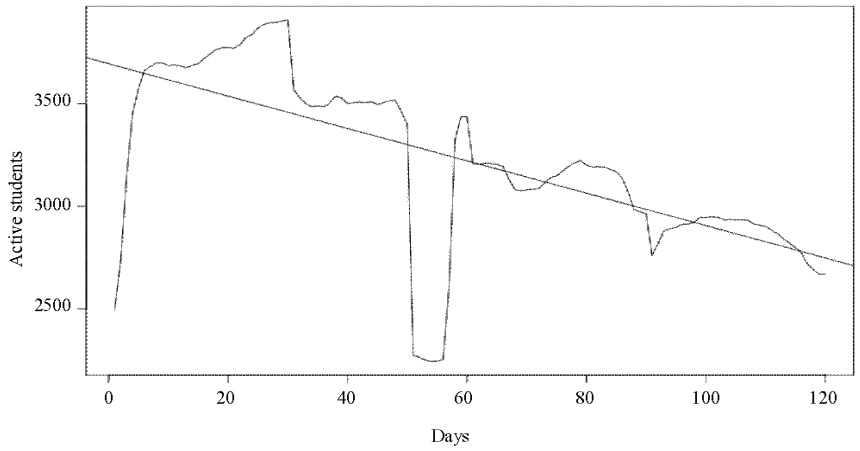

Fig. 3: Plot of No. of active students vs. each day in four month span with linear regression applied

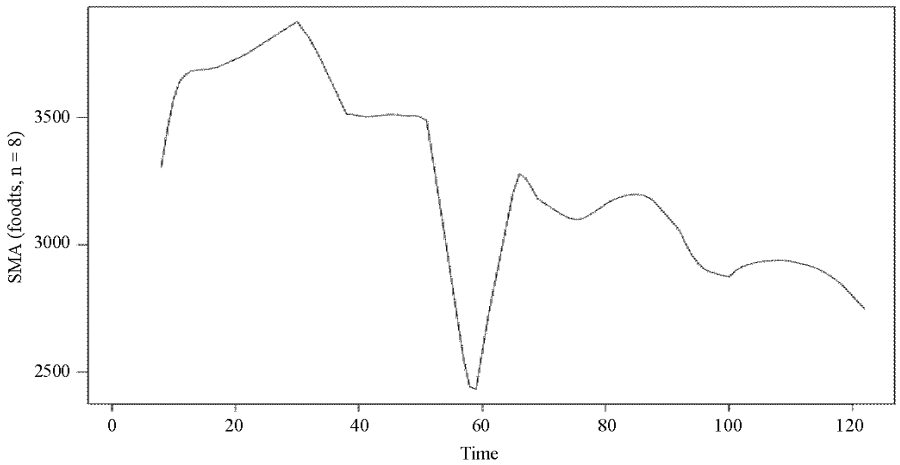

Fig. 4: Smoothed time series

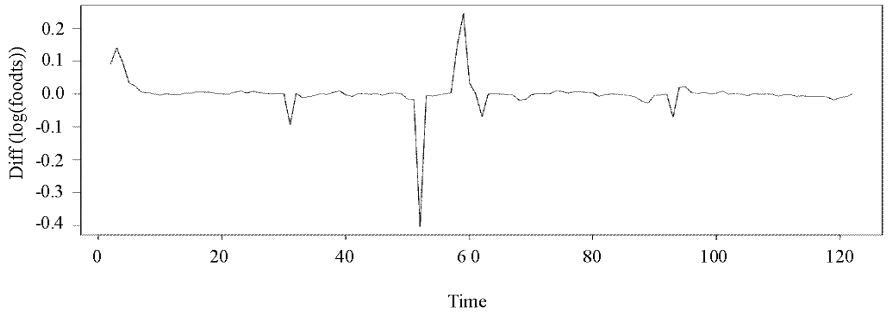

Fig. 5: Differencing performed on the time series

kinds of regression models. The statistic used here in the test is a negative number. The rejection of the null hypothesis of a unit root becomes more stronger as the statistic becomes more negative.

Algorithm 1 (ADF test summary):

Augumented Dickey-Fuller Test

data: diff (fc_ts)
Dickey-Fuller =-8.2788, Lag order = 0,p-value = 0.01
Alternative hypothesis: stationary

We observe that the series is stationary and is ready for any kind of time series modeling.

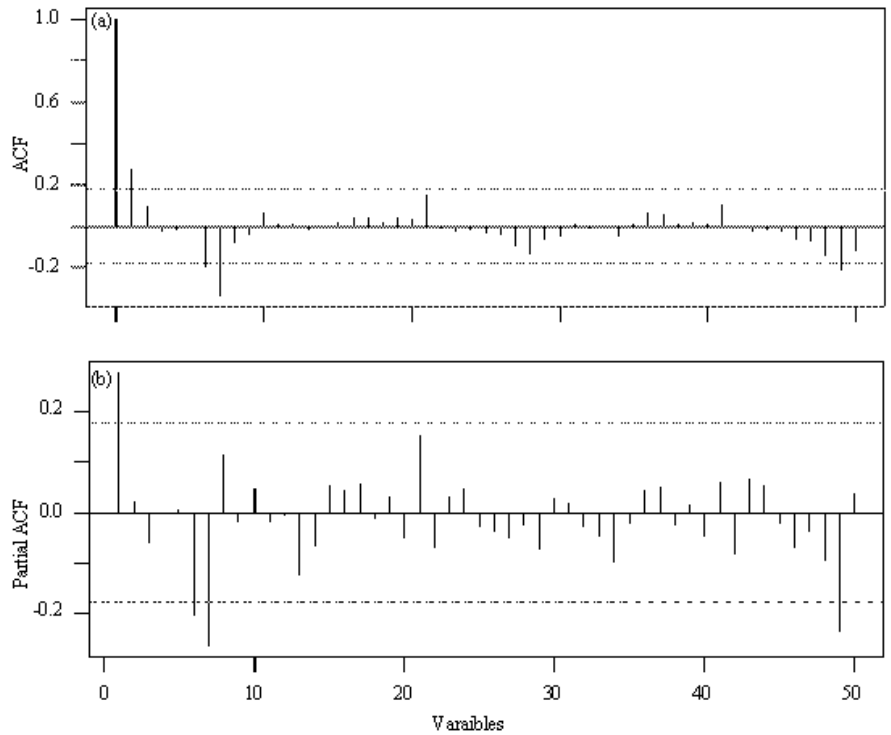

Fig. 6: ACF and Partial ACF Plot on difference of log

Step 3 (plot ACF/PACF charts): The ACF (Auto Correlation Function) plot provides an insight about the covariance of the variables in the time series and their underlying structure. The PACF (Partial Auto Correlation Function) charts are plotted to find a measure of correlation between two variables in a time series excluding the effect of the variables lying in between them. The ACF and PACF charts are plotted and then the data in these plots are analyzed (Brockwell and Davis, 1996) (Fig. 6).

We propose to regress on the difference of log rather than log directly as differencing removes the non-stationary component in the time series. We first address the seasonal component and then the non-seasonal component of an ARIMA Model.

It is observed that the ACF plot cuts off at the first lag only and the first lag has a value closest to 1.0. We can say that it is an AR(1) Model. The PACF plot cuts is observed to find the absolute value is greatest at the fifth lag and then at lag 49. Thus, we can conclude that AR(1) is to be considered which would make the value of p as 1, d as 1(as differencing has been done once) and q as 0. The parameters required in the ARIMA Model are identified by observing the above plots. (1, 1, 0) is the best values for (p, d, q) for the ARIMA Model.

Step 4 (implement the ARIMA Model): Chu (2009) with the parameters (p, d, q) obtained from the ACF and PACF plots, the ARIMA Model can be implemented and can used as a model for predicting future values of our time series data set. For that we call the arima function in which we pass the time series and the parameter as order argument.

Algorithm 2 (ARIMA Model summary):

```
Series: foodts
ARIMA (1, 1, 0)
Coefficients:
        ar1
    0.2825
s.e. 0.0877
Sigma^2 estimated as 19524: log likelihood =-768.94
AIC = 1541.87   AICc = 1549.97   BIC = 1547.46
```

ARIMA(1, 1, 0)Model on our time series depicts that ARMA(1,0) model is being fitted to the time series of first differences. An ARIMA(1, 1, 0) Model can be represented as $X(T) = Z(T)-(\theta*Z(T-1))$ where X(T) is the value to be found at time t. While passing the time series data to the arima function we estimate the value of theta which is 0.2825.

Step 5 (forecast and accuracy): Once the ARIMA Model is ready, it can be used to forecast the future values of the

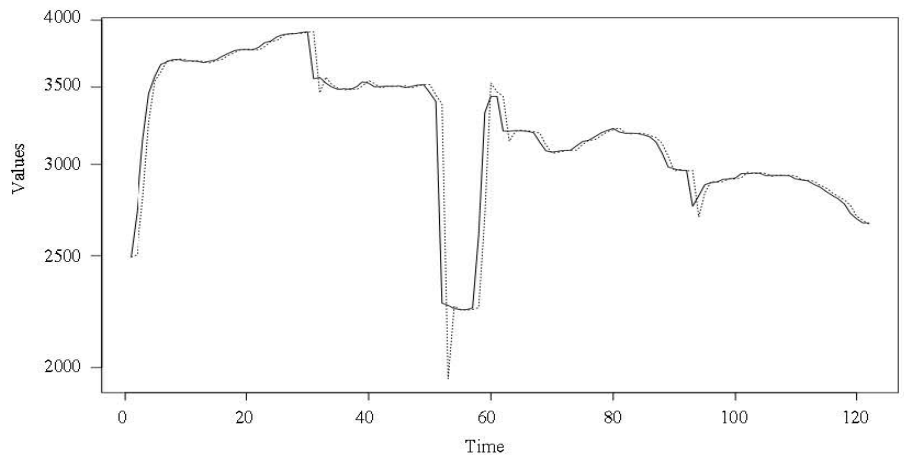

Fig. 7: Dotted line represents the predicted active students for the next 4 months

time series using forecast. Arima function in the "forecast" R package. We can set the confidence level of our prediction intervals by using the "level" argument. The default confidence levels are 80 and 95%. Also we mention the range upto which the model will calculate the forecasted values. The fitted values of the forecast using the ARIMA (1,1,0) Model are then plotted (Archer, 1987) (Fig. 7).

Now, we move on to investigate for forecast errors in the ARIMA Model. This is done by checking whether the model is normally distributed with mean value equal to zero and shows constant variance. For this we can plot the forecast error values or residuals to check for constant variance and histogram (with an overlaid normal curve) for checking the distribution to be normal (Fig. 8).

The time plot shows that the variance of the error values appear to be approximately persistent over time and the histogram depicts that the errors are about normally distributed with mean value closer to zero (Hyndman and Koehler, 2006; Chambers and Hastie, 1992; Syntetos and Boylan, 2005). Accuracy of the fitted values of the forecast using the ARIMA (1,1,0) Model can be estimated by the accuracy function. Training set: ME = 0.964595; RMSE = 138.5794; MAE = 47.8; MPE = -0.04019555; MAPE = 1.661669; MASE = 0.9668672; ACF1 = -0.01347402.

The data above is the range of summary measures returned by forecast accuracy functionin R. If a data series t overlapping with the time of ob(any forecast object like arima or lm object) is given, it calculates the test setforecast accuracy based on t-ob. If t is not given, it calculates the training set accuracy of the forecasts according to the fitted values.

Algorithm 3 (common error parameters):
The folowing are the measures calculated:

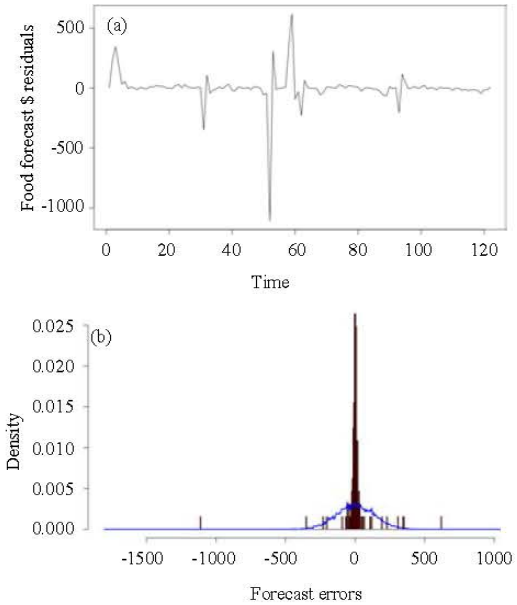

Fig. 8: Time plot and histogramof forecast residuals/errors

ME: Mean Error
RMSE: Root Mean Squared Errpr
MAE: Mean Absolute Error
MPE: Mean Percentage Error
MAPE: Mean Absolute Scaled Error
MASE: Mean Absolute Scaled Error
ACFI: Aotocorrelation of errors at lag 1

MPE computes the percentage error by which the forecast result of a givenmodel differs from the actualdata. MAPE is scale independent thus it is used to compare forecast performance of data series that are on different scale whereas RMSE and MAE are scale dependent errors, thus cannot be used for series that are different scale. Another metric that is mentioned above is MAPE

which compares the forecasts using naive method. As it never gives undefined or infinite result it is very useful for data series occurring at irregular intervals.

CONCLUSION

In this study, ARIMA forecasting model was used to predict the number of active students for the next semester of university. This model is based on time series analysis. The number of active students predicted are accurate enough to be used by the mess authorities to estimate the amount of raw material for meals on a daily basis. The forecasted values keeps the trend and the non-seasonality of the data into account. Quantitative approach was the first choice for this case as the database has discrete time based data.

We can thus apply this method to other university dining messes to illustrate the raw material supply required on a daily, monthly and even per student vise basis as per the demand of the dining mess. Predicting the pattern helps the mess authorities to estimate days with high and low number of students coming to mess and thus can also estimate the manpower required tomanage the crowd and be prepared accordingly. Such a model can help the authorities to plan their mess budget accordingly and also reduces food wastage which is of prime importance in areas where scarcity of food is common.

ACKNOWLEDGEMENTS

We wish to thank the mess authorities and the chief warden of the university for providing us the data used in the study. We would like to show our gratitude to AakilMathur, an undergrad student in computer science MPSTME, Mumbai for sharing his inputs and thoughts with us during the course of this study. We also wish to thank the reviewers for their helpful suggestions and comments on the earlier version of the manuscript.

REFERENCES

Archer, B., 1987. Demand Forecasting and Estimation. In: Travel, Tourism and Hospitality Research: A Handbook for Managers and Researchers, Ritchie, J.R.B. and C.R. Goeldner (Eds.). John Wiley & Sons, Hoboken, New York, USA., pp: 77-85.

Box, G.E.P. and G.M. Jenkins, 1970. Time Series Analysis: Forecasting and Control. 2nd Edn., Holden-Day, San Francisco, California, Pages: 553.

Brockwell, P.J. and R.A. Davis, 1996. Introduction to Time Series and Forecasting. Springer, Berlin.

Chambers, J.M. and T.J. Hastie, 1992. Statistical Models. Brooks/Cole Publishing Company, Pacific Grove, California,.

Chu, F.L., 2009. Forecasting tourism demand with ARMA-based methods. Tourism Manage., 30: 740-751.

De Gooijer, J.G. and R.J. Hyndman, 2006. 25 years of time series forecasting. Int. J. Forecasting, 22: 443-473.

Dickey, D.A. and W.A. Fuller, 1979. Distribution of the estimators for autoregressive time series with a unit root. J. Am. Stat. Assoc., 74: 427-431.

Hyndman, R.J. and A.B. Koehler, 2006. Another look at measures of forecast accuracy. Int. J. Forecast., 22: 679-688.

Meimanda, S.E., G.H. Hadi and N.I. Hairul, 2012. Demand forecasting: A review on qualitative methods and surveyed done by Delphi technique. Faculty of Built Environment, University Technology Malaysia, Johor, Malaysia. http: //staff. estem-uc.edu. au/taipham/files/2013/01/Demand_Forecasting_Beijing-libre.pdf

Syntetos, A.A. and J.E. Boylan, 2005. The accuracy of intermittent demand estimates. Intl. J. Forecasting, 21: 303-314.