Forecasting Call Arrivals to Call Center of a Telecommunication Call Centre in Sri Lanka

W.D.A.S.Weerawardana* and D.J.C. Suriyaarachchi
Department of Statistics & Computer Science
Faculty of Science, University of Kelaniya
Kelaniya, Sri Lanka
anushkaweerawardana@gmail.com and suri@kln.ac.lk

***Abstract*— Call centers (which is also known as Customer care units) have become one of the increasingly popular topics in today's society since it helps to develop a good interaction between customers and the company. One of the most important decisions making the task for the call centers is determining the appropriate number of staff officers to meet the demand for ensuring customer satisfaction while minimizing the service cost. For that, it is important to identify the accurate predictions on call arrivals. This study is focused on developing a model for the call arrival data of a major call center of a telecommunication company in Sri Lanka with the purpose of getting precise predictions of hourly call arrival count. The mixed model approach is used in this study for model the call arrival data by incorporating both the fixed and random effects which affect to the call arrival count. For confirming the accuracy of the mixed model, here it included the comparison of the mixed model with the model with only fixed effects. And the mixed model which contains the day of week and hour of the day as fixed effects and daily volume deviation and hourly random deviation as random effects is identified as the precise model for predicting the call arrival count of the call center.**

Keywords; Call Center; Call Arrivals; Fixed Effects Model; Mixed Effects Model

I. Introduction

Call centers have become one of the most increasingly popular topics in today's society, because with increasing the competitiveness of the industries most of the companies focused on providing better quality service to their customers. In the point of satisfying the customers, call centers have become very important. Call centers perform an important role in developing a good interaction among the customers and the company.

For increasing the customer satisfaction about the call center the system of call center must be well organized. When looking at the call center the quality of the service is normally measured by looking at the waiting time of the customers before connecting with the agent. So improve the quality it is a must to reduce the customer delays (waiting time of customers). It is a responsibility of call center managers to determine the precise number of staff officers to meet their demand for ensures customer satisfaction. To reduce the customer delays there must be enough number of call center agents to handle the calls receiving. When allocating staff, managers should also focus on the cost associated with the call center operations. Because operating a call center is also a cost to the company, so it is important to reduce the cost of call center while providing better service to the customers in order to achieve the organization's goals.

So for determining the most efficient scheduling of call center agents it is important to know the accurate forecasts for the number of call arrivals of the call center. These call arrival forecasts may be also needed over periods of the day and also for several days in advance.

II. Literature Review

Since the accurate call arrival forecasts are very important in determining the number of staff officers appropriately many researchers have been placed efforts to analyze and forecast call arrival rate of call centers using different approaches.

The process of a call center can be well identified by the Queuing models; the most simple and widely used queuing model for the call center is M/M/N model. In [1] M/M/M+N (Erlang- A) model was used for the predictions.

Earlier call arrivals forecasting studies focused on applying time series methods such as BOX and Jenkins Auto Regressive Moving Average (ARMA) models, Auto Regressive Integrated Moving Average (ARIMA). In [2] several univariate time series models were applied to forecast intraday arrivals of two call centers in the UK and the results of this study indicated that the using seasonal ARIMA model and an extension of Holt-Winters exponential smoothing is potential in predicting call arrivals. In [3] Andrews and Cunningham also used the time series methodology to forecast arrivals to L.L Bean's call center. They used ARIMA/transfer function to model the data of call center.

Some of the researchers have analysed call center arrival data using Singular Value Decomposition (SVD). This is a non-Bayesian approach. As an example, Sheng and Huang [4] analyse the data of a U.S. financial organization call center using this method. They visualize data and also extracted the features from noisy data, reduce the dimensions of data and made a model to intraday forecasting using the singular value decomposition. Then they expand that study to account both interday and intraday patterns of call arrivals [5].

In near year's researchers uses Bayesian techniques to call forecasting. This helps to forecast arrival counts as well as their distributions (i.e. these methods providing more information than just point estimates). In [6] Soyer and Tarimcilar analysed the effectiveness of different advertising strategies on call arrivals by using Bayesian analysis based on Poisson process model to analyse call arrival data from a call center. They model arrival rate function using a mixed model approach. Here they extend the Poisson process model to mixed model approach by adding random effects in order to include heterogeneity in different advertisement strategies. And in [7] also Bayesian techniques was used to model the data of call center of North American commercial bank. They provide a Gaussian model to predict the call arrival rate of an inhomogeneous Poisson process.

In this study call arrivals are modelled using linear mixed model approach by incorporating both fixed and random effects.

In [8] researchers introduce an arrival count model which is based on a mixed Poisson process approach using the Israeli Telecom company call center data includes arrival count of a call center per half hour in six months period. Their mixed model includes fixed effects, such as day of week and period of day effects and also daily volume deviations and period by period random deviation as random effects. When modelling the call arrivals they assumed number of call arrivals follows a Poisson distribution and then used square root transformation (Assumes call arrivals ($X_{i,j}$) follows Poisson distribution and then transformed arrivals ($Y_{i,j} = \sqrt{X_{i,j} + 1/4}$) is normally distributed) to stabilize the variance. And also mixed Poisson process approach was used in [9] to model the call arrivals of a Canadian company call center using the data of 275 days and each day contains 22 half-hour periods.

Reference [10] is a comparative study on forecasting call center arrivals which used both time series and mixed model approaches. First, they present a comparative study of several time series methods and secondly they show the importance of modelling different types of correlations in the data (mixed model).

III. METHODOLOGY

A. Mixed effects models

Mixed effects model is also like many other types of statistical models. It also describes a relationship between a response (dependent) variable and some other covariates that have been observed along with the response variable. The parameters associated with the particular levels of a covariate are called as effects. There are two types of effects as fixed effects and random effects. Fixed effects are ones in which the possible values of the variable are fixed. Random effects refer to variables in which the set of potential values can change.

Mixed Effects Model can be used to model both linear and nonlinear relationships. Linear mixed effects models simply model the fixed and random effects as having a linear form. Similar to the General Linear Model, an outcome variable is contributed to by additive fixed and random effects (as well as an error term). Using the familiar notation, the linear mixed effect model takes the form:

$$y_{ij} = \beta_1 x_{1ij} + \beta_2 x_{2ij} \ldots \beta_n x_{nij} + b_{i1} z_{1ij} + b_{i2} z_{2ij} \ldots b_{in} z_{nij} + \varepsilon_{ij}$$

Where,

yij - Value of the outcome variable for a particular ij case

$\beta_1, \beta_{2\ldots}$ βn - Fixed effect coefficients

$x_{1ij}, x_{2ij\ldots}\ x_{nij}$ - Fixed effect variables (predictors) for observation j in group i (usually the first is reserved for the intercept/constant; $x_{1ij} = 1$)

$b_{i1}, b_{i2,\ldots,}\ b_{in}$ - Random effect coefficients which are assumed to be multivariate normally distributed

$z_{1ij}, z_{2ij\ldots}\ z_{nij}$ - Random effect variables (predictors)

ε_{ij} - Error for case j in group i where each group's error is assumed to be multivariate normally distributed

IV. RESULTS AND DISCUSSION

A. Preliminary Data Analysis

The data set includes the call arrival counts per hour from 01-Mar-2014 to 28-Feb-2015. This call center operates all 24 hours every day (including holidays). Here they count the hour starting from 12.00 a.m. to 1.00 a.m. as the first hour (H1) and respectively.

Since Saturday and Sunday have fewer call volumes than other days, it has decided to remove the arrival counts of Saturdays and Sundays from the data set for more precise model formulation.

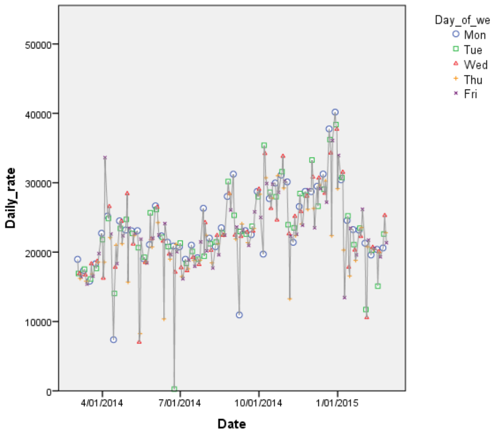

Figure 1. Daily arrival counts between 01-Mar-2014 and 28-Feb-2015

As in the weekly pattern shown in the Figure 1, one can observe two important characteristics. One is there are some outlier points in the dataset.

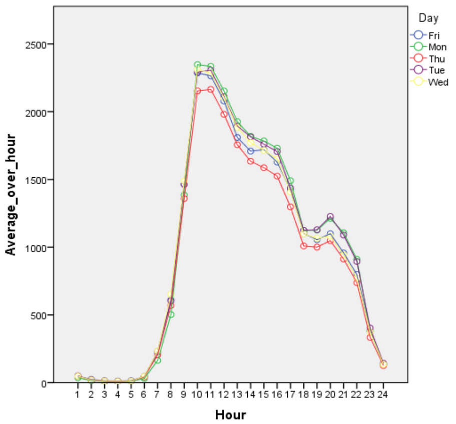

Figure 2. Interday arrival patters for each weekday between 01-Mar 2014 and 28-Feb-2015 after removing the Saturdays and Sundays

Most of those outlier points are contain unusual call volumes than other days. And also there are some unique patterns in call arrival count in each weekday.

According to the Figure 2, it is clear that there are interday patterns between arrival counts over the day. And also there are intraday patterns between arrival counts over periods/hours of the same day.

TABLE I. ESTIMATES OF THE CORRELATIONS BETWEEN ARRIVAL COUNTS (CORRECTED TO SEASONAL TRENDS) OVER WEEKDAYS

	Mon	Tue	Wed	Thu	Fri
Mon	1.0000	0.6887	0.6522	0.4279	0.6057
Tue		1.0000	0.8087	0.5383	0.6322
Wed			1.0000	0.7136	0.5642
Thu				1.0000	0.5702
Fri					1.0000

As we see in above Table 1 correlations are higher in near days and correlations get smaller when the distance between days gets larger. So it is clear that there are correlations between the weekdays that mean there are interday correlations.

Not only interday correlations there are also heavy intraday correlations in the dataset.

According to the above Figure 3, we can clearly see there are intraday strong correlations.

According to the above analysis, we can conclude that there is a positive intraday correlation between the call arrival counts over hours in same weekday and also there is a strong positive interday correlation between call arrival counts over weekdays. These two correlations are

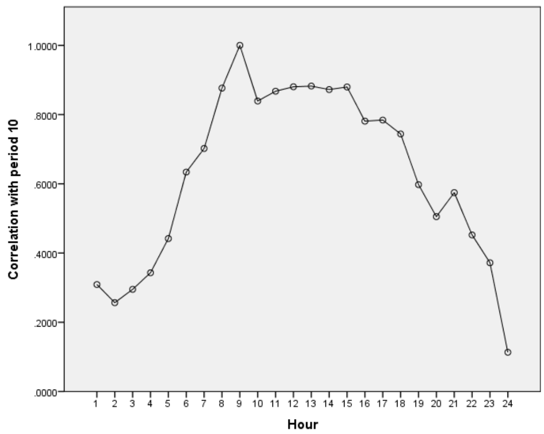

Figure 3. Correlation between hour 10 and remaining hours of Monday for the call arrivals

incorporate in formulating a suitable model for the call arrival counts.

B. Replacing Missing Values and Data Transformation

Since there were some missing values and outliers then those were replaced by using the method of linear interpolation. To stabilize the variance of data here used the following data transformation, If N is a random variable and N ~ Poisson (λ), then $y = \sqrt{(N + 1/4)}$ is approximately normally distributed with mean $\sqrt{\lambda}$ and variance ¼.

According to this transformation method as $\lambda \to \infty$, y approximately normally distributed. Since in the data λ (mean call arrival count per hour) has value around 875 per hour it is reasonable to apply this approximation.

C. Model Formulation

1. Fixed Effects Model

In this study, the day effect and the period effect were considered as the fixed effects in this model, and also included another effect as cross term between the day of week and period of the day to identify whether there is an interaction between those two effects.
Then the Fixed Effects model is as follows,

$$Y_{ij} = \sum_{k=1}^{5} \alpha_k I_{di}^k + \sum_{l=1}^{24} \beta_l I_j^l + \sum_{k=1}^{5} \sum_{l=1}^{24} \theta_{k,l} I_{di}^k I_j^l + \varepsilon_{i,j} \qquad \text{(Model 1)}$$

Where,

Y_{ij} - Transformed call volume in period j of day i
I_{di}^k , I_j^l - Indicators for day d_i and period j respectively (i.e. if di=k, $I_{di}^k = 1$ and if j=l, $I_j^l = 1$ and otherwise both are 0)
α_k , β_l , $\theta_{k,l}$ - Coefficients that need to be estimated from real data.

When considering the cross terms between day of the week effect and period of the day effect all those terms are not significant. So by removing all the cross terms from the model fit a model only with day of week effect and period of day effects as follows,

$$Y_{ij} = \sum_{k=1}^{5} \alpha_k I_{di}^k + \sum_{l=1}^{24} \beta_l I_j^l + \varepsilon_{i,j} \qquad \text{(Model 2)}$$

2. Linear Mixed Effects Model

Then a mixed effect model is formulated by incorporating the both fixed and also the random effects. Daily volume deviation and the period by period random deviation are used as the random terms of the mixed model. And the same fixed effects as in the above fixed effects model has

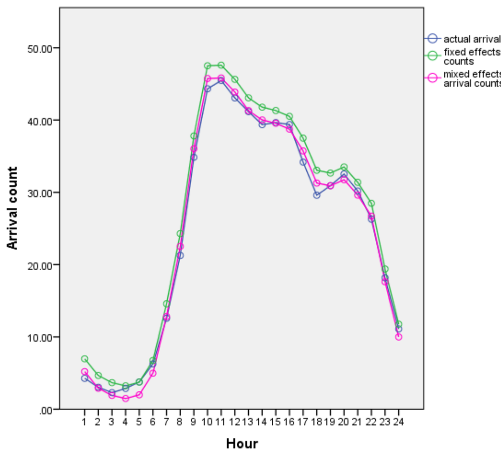

Figure 4. Comparison of actual arrival counts and predicted arrival counts for 10th Feb 2015

used for the mixed model also. Following is the mixed model formulated,

$$Y = X_D\beta_D + X_P\beta_P + Z\Upsilon + \varepsilon \qquad \text{(Model 3)}$$

Where,
Y – (y1,1,…,y1,k,y2,1,…,yd,1,…,yd,k)T
X_D – Design matrix for the fixed day-level effects
β_D – Vector of fixed effects of day of week coefficients
X_P – Design matrix for the fixed period-level effects
β_P – Vector of fixed effects of period coefficients
Υ – Vector of random day level effects
ε – Vector of random residual effects

First order autoregressive covariance structure /AR (1) (As in reference [8]) is used as the pre specified covariance structure for the random effects here.

According to the results of the models, it can be seen that the mixed model leads to more accurate forecasts since the root mean squared error (RMSE) and mean absolute error (MAE) gives the smallest value in mixed model predictions. And AIC and AICC values also confirms the linear mixed model is more appropriate than the fixed effects model by giving smaller values to both AIC and AICC. So according to the results got it is clear that developing a model with both fixed and random effects is most suitable than the model with only fixed effects to predict the call arrival count of the call center. Following figure 4 also confirms that the mixed effect model gives more accurate forecast than the fixed effects model.

TABLE III. COMPARISON OF FIT STATISTICS OF THE ALTERNATIVE MODELS

	Model 1	Model 2	Model 3
-2 Log Likelihood	34796.7	34887	31649
AIC	35036.7	34943	31711
AICC	35041.7	34943.3	31711.3
BICC	35839.2	35130.3	31819.8

TABLE IV. RMSE AND MAE VALUES FOR TWO MODELS

	Fixed effects model	Mixed effects model
RMSE	4.5888	3.2018
MAE	2.9998	2.1356

V. CONCLUSION

In this study, three models were developed to describe the hourly call arrival counts of a major call center in Sri Lanka.

Here the model is developed by using the intraday and interday structures which are a common property of many of the call center arrival counts. Then as the main and important part of the study, a linear mixed effects model which contains both fixed and random effects was developed. Here intraday and interday effects were added as the fixed effects and the daily volume deviation and the hourly random deviation was incorporates as the random effects which have AR(1) covariance structure.

As further study directions, this study was focused on developing a univariate model to the call arrival data, that means here the model was developed without considering the type of call and taking all the calls as whole since there was no sufficient data. But this model can be developed further by extending this model into a bivariate model or a multivariate model by considering the type of calls as two types or multiple types of calls.

REFERENCES

[1] L. Brown, N. Gans, A. Mandelbaum, A. Sakov, H. Shen, S. Zeltyn, L. Zhao, “Statistical analysis of a telephone call center: A queueing-science perspective,” Journal of the American Statistical Association,March 2005, Vol. 100,pp. 36-50.

[2] J. W. Taylor, “A Comparison of Univariate Time Series Methods for Forecasting Intraday Arrivals at a Call Center,” Management Science, 2008, Vol. 54, pp. 253-265.

[3] B. Andrews & S. Cunningham, “L.L.bean improves call-canter forecasting”,1995.

[4] H. Shen & J. Z. Huang, “Analysis of call center arrivals data using singular value decomposition,” Appl. Stochastic Models Bus. Ind., 2005, pp. 251–263.

[5] H. Shen & J. Z. Huang,” Forecasting time series of inhomogeneous poisson processes with application to call center workforce management,” The Annals of Applied Statistics 2008, Vol. 2, No. 2,pp. 601–623.

[6] R. Soyer & M.Tarimcilar,” Modeling and Analysis of Call Center Arrival Data:A Bayesian Approach,” 2008.

[7] J. Weinberg, L. Brown & J. Stroud, “ Bayesian Forecasting of an inhomogeneous Poisson process with applications to call center data,” 2006.

[8] S.A. Noiman, P.D. Feigin, & A. Mandelbaum, “ Workload forecasting for a call center: Methodology and a case study,” Annals of Applied Statistics, 2009, Vol. 3, No. 4,pp. 1403-1447.

[9] R. Ibrahim, P. L. 'Ecuyer, N. Regnard & H. Shen, “On the modeling and forecasting of call canter arrivals,” Proceedings of the 2012 Winter Simulation Conference, pp. 256-267.

[10] R. Ibrahim, P. L. 'Ecuyer ,”Forecasting call center arrivals:A comparative study”.