

TIME SERIES MODELING OF DAILY ABANDONED CALLS IN A CALL CENTRE

F.I. UGWUOWO

(Received 29, March 2010; Revision Accepted 3, May 2010)

ABSTRACT

Models for evaluating and predicting the short periodic time series in daily abandoned calls in a call center are developed. Abandonment of calls due to impatient is an identified problem among most call centers. The two competing models were derived using Fourier series and the Box and Jenkins modeling approaches. The selected models were shown to be both parsimonious and adequate using the P-P plots, Q-Q plots and residual analysis. In validating and comparing the models some efficiency measures were used with a view to determining the model which best represent the population under investigation. Fourier series model was found to be more efficient than Box and Jenkins model. The data for application were got from a GSM telephone provider.

KEY WORDS: Call center; Tele-queue; abandoned calls; Fourier time series; Periodogram; Moving average, Autoregressive process

1. INTRODUCTION

Call centers provides tele-services to their customers with the aid of an agent. As they speak with customers over the phone, they interact with a computer terminal, inputting and retrieving information related to customers and their requests. Customers, who are virtually present, are either being served or are waiting in what we call a tele-queue, a phantom queue which they share, invisible to each other and to the agents who serve them. Customers wait in this queue until one or two things happen: an agent is allocated to serve them (through supporting software), or they become impatient and abandon the tele-queue.

In recent years, several documented studies of the incoming call arrival process have been conducted and tested, thanks to technology advances in the call center industry. Earlier studies focused on classical Box and Jenkins, Auto-Regressive-Moving-Average (ARMA) models such as the Fedex company study. A well-known study, also employing Auto-Regressive-Integrated-Moving-Average (ARIMA) model techniques, was carried out by Andrews and Cuningham (1995) to produce L.L Beans Call Center daily forecasts. The study focuses on modeling two different arrival queues each with its own characteristics. Their models incorporated exogenous variables along-side the moving average (MA) and Auto-regressive (AR) variables, using transfer functions to help predict outliers such as holidays and special sales promotion periods.

A more recent study was carried out by Taylor (2006) in this study several different time series models were investigated on two different sources of data. Among these models were seasonal ARMA models, exponential smoothing for double seasonality methods, and dynamic harmonic regression. His result indicated that for short term forecasting horizons, the exponential smoothing for double seasonality method performs quite well but for a practical horizons (longer than one day) a very basic averaging model out-performs all of the suggested alternatives.

More so, Aldor-Noiman (2006) in his research, count models which are based on a mixed Poisson process approach. The first model uses the normal-Poisson stabilization in order to employ linear mixed mode techniques and the second model employs the Bayesian approach, implementing Gibbs sampling techniques, and using "open bugs" software, to produce the predictive distributions for the future arrival counts.

SenGupta & Ugwuowo (2006) proposed asymmetric angular- linear multivariate regression models, which were motivated by the need to predict some environmental characteristics based on some circular and linear predictors. In the work, they adopted a simple cosine function for the wind energy and solar energy data. A measure of fit was provided through the residual analysis.

The purpose of this paper is to model the short periodic time series in daily abandoned calls in a call center. The Fourier series model and the Box and Jenkins models were developed with a view to determining the most efficient model. In section 2 we introduce the data and the background of this study. Section 3 discusses the two modeling approaches while the model fitting were given in section 4. A measure of the efficiency of the fitted model is given in section 5. Discussions on the overall fit of the model and other conclusive notes were shown in section 6.

2. Data and background

2.1 Description of Call Center Operation

When a customer calls one of the several telephone numbers associated with the call center, except for rare busy signals, the customer is then connected to VRU and identifies herself. While using the VRU, the customer receives recorded information, general and customized. It is also possible for customers to perform some self service

The Models

2.1 Fourier Series Model

The practical value of this approach to time series analysis rests most obviously on the empirical observation that many time series data exhibit cyclic fluctuations in value, but at frequencies which are not always predictable before the data are examined. The periodogram is a summary description based on a representation of an observed time series as a super position of sinusoidal waves of various frequencies. To motivate the periodogram, we shall examine a very simple model for a time series (y_t) exhibiting cyclic fluctuations with a known period, p say. We assume that

$$Y_t = \alpha \cos (wt) + \beta \sin (wt) + Z_t \quad t = 1,2,\dots,n \quad \dots (1)$$

Where $\{Z_t\}$ is a white noise sequence, $w = 2\pi/p$, $p = k/n$ is the known frequency of the cyclic fluctuations, and α, β are the parameters. Where $2\pi = 6.28571429$ in radius or 360 degrees. Now consider a slightly less simplistic model in which we contemplate not one, but several sinusoidal components. This gives a model of the form

$$Y_t = \sum_{k=1}^m \left\{ \alpha_k \cos(w_k t) + \beta_k \sin(w_k t) \right\} + \mu_t \quad \dots \dots \dots (2)$$

$$t = 1, 2, 3 \dots n; k = 1, 2 \dots m$$

Where $\{\mu_t\}$ is again a white noise sequence and each w_k is one of the Fourier frequencies. The estimate of the parameters can be written explicitly as:

$$\alpha_k = \frac{2}{n} \left\{ \sum_{t=1}^n z_t \cos(w_k t) \right\}, \quad \beta_k = \frac{2}{n} \left\{ \sum_{t=1}^n z_t \sin(w_k t) \right\}, \quad \dots (3)$$

Figure 1 shows the data under consideration which is made up of Sine and Cosine waves with different frequencies; Z_t is the observed value for period t , 'n' is the data size. The periodogram was originally used to detect and estimate the amplitude of a Sine component, of known frequency, buried in noise. Since the set of data under consideration is odd, hence the periodogram then consist of the $m = (N-1)/2$ values i.e. $k = 1, 2 \dots m$ values.

2.2 Seasonal model

We consider the family of seasonal models since the series shows a marked seasonal pattern (See Figure 1 below).The nature of abandoned calls has been found to exhibit periodic behavior with period $S = 1$ week. The general multiplicative model as given in Box et al (2008) is,

$$\phi_p(B) \Phi_P(B^S) \nabla^d \nabla_s^D z_t = \theta_q(B) \Theta_Q(B^S) a_t \dots (4)$$

Where for this particular model $S = 7$ days and the order of various operators are severally represented by p, P, q and Q . Hence the model is a seasonal autoregressive integrated moving average model of order $(p,d,q) \times (P, D, Q)_s$

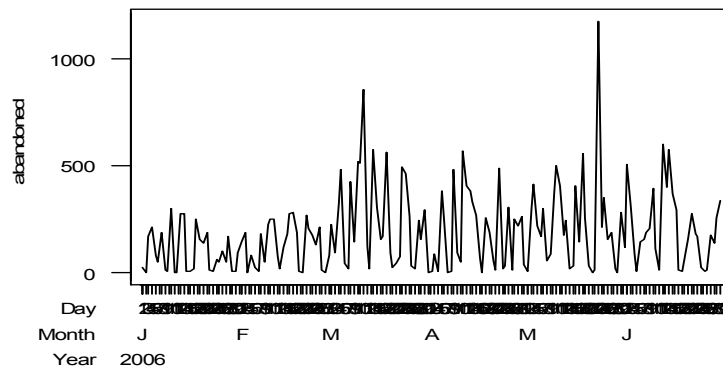


Figure 1 Plot of the original data

3. Model Fitting

3.1 The Fourier Model

Detecting the best parsimonious model is an important step in the analysis of data sets. After estimating the parameters of the model, several diagnostic tools were used to assess the goodness-of-fit of the model. Going by some set of assumptions, the standardized residuals ought to resemble a sample from a standard normal distribution. Hence, in Figure 2, we consider the normal P-P plot and the Q-Q plot of the standardized residuals to assess the goodness-of-fit of the model. The plot shows an amazingly good imitation of a straight line suggesting that the fitted model is adequate. Summary of the analysis for the parameters of the model as obtained using SPSS is given in Tables 1, 2 and 3. Figure 3 shows the plot of the standardized residual with the fitted values. Once more we observe

that there are no definite patterns that will cause us to doubt the fitted model. The plot of the histogram of the standardized residual can be seen in figure 4. The plot is somewhat symmetric and tails off at both the high and low ends as a normal distribution does.

Identification of the model was made after comparing the strength of the relationship between several models and the dependent variable.

Table 1: Model Summary

Model	R	R Square	Adjusted Square	R	Std. Error of the Estimate
1	.579(a)	.335	.316		149.880

Predictors: (Constant), SINw2t, SINw1t, SINw26t, COSw1t, COSw52t

Table 2: Anova Table

Model		Sum Squares	Df	Mean Square	F	Sig.
1	Regression	1981097.343	5	396219.469	17.638	.000(a)
	Residual	3931181.099	175	22463.892		
	Total	5912278.442	180			

Predictors: (Constant), SINw2t, SINw1t, SINw26t, COSw1t, COSw52t

Dependent Variable: h_ung

Table 3: Parameters Estimates

Coefficients (a)

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
1	(Constant)	185.330	11.140		16.636	.000
	COSw1t	-37.639	15.758	-.147	-2.389	.018
	SINw26t	-98.101	15.753	-.384	-6.228	.000
	COSw52t	81.422	15.754	.319	5.168	.000
	SINw1t	-39.730	15.752	-.155	-2.522	.013
	SINw2t	-51.289	15.752	-.201	-3.256	.001

Dependent Variable: h_ung

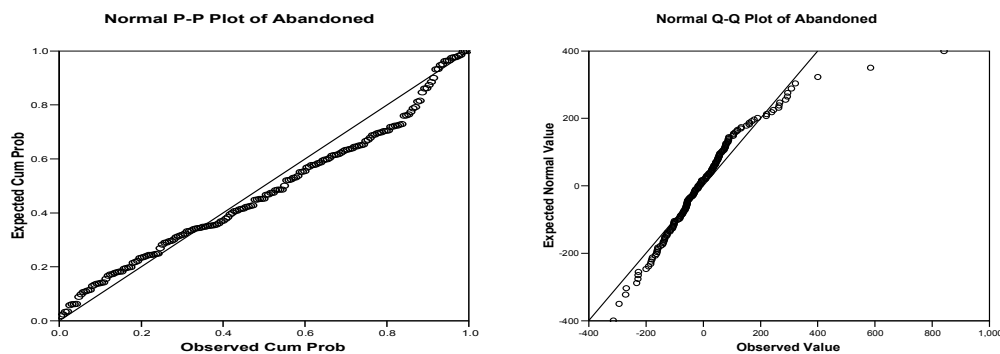


Figure 2: Normal P-P plot and Q-Q plot

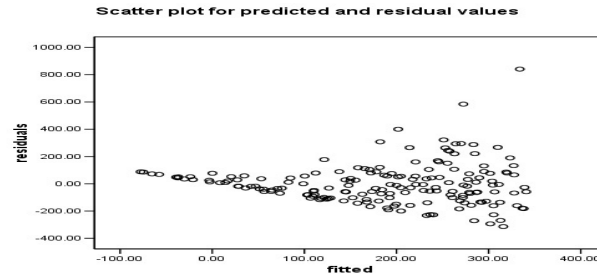


Figure 3: Scatter Plot for predicted and residual values

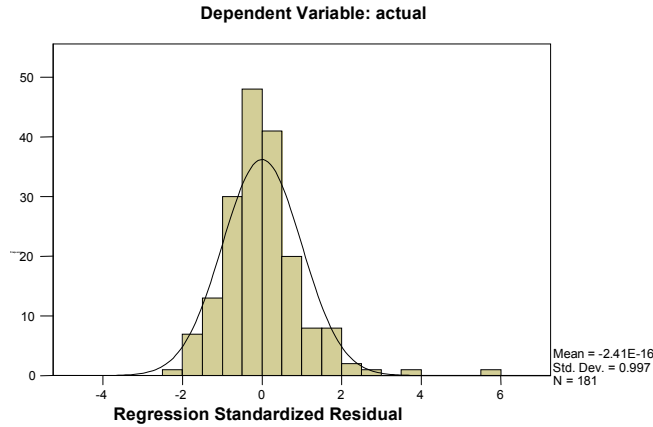


Figure 4: Histogram of the residual

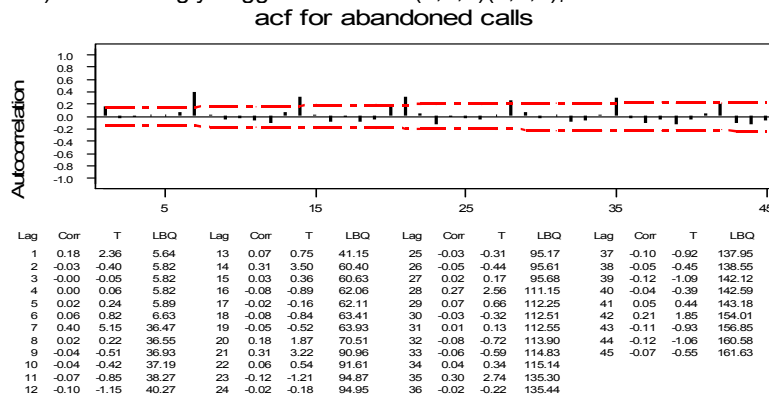
Using the parameter estimates in Table 3, the fitted Fourier series model is thus given in Eq. (5) below.

$$z_t(l) = 185.330 - 37.639 \cos\left(\frac{2\pi l}{181}\right) - 39.730 \sin\left(\frac{2\pi l}{181}\right) - 51.289 \sin\left(\frac{4\pi l}{181}\right) + 81.422 \cos\left(\frac{104\pi l}{181}\right) - 98.101 \sin\left(\frac{52\pi l}{181}\right) \dots \quad (5)$$

for $l=1,2,3,\dots,n$

3.2 The Seasonal Autoregressive Integrated Moving Average (SARIMA) Model

After examining the graph of the original data, we discovered that it was stationary. Inspecting the ACF and PACF of the series, we discovered presence of seasonal variation at every lag 7, hence, the need for seasonal differencing of order 1 ($D=1$). This strongly suggest SARIMA (1,0,1)(1,1,1)₇ model.



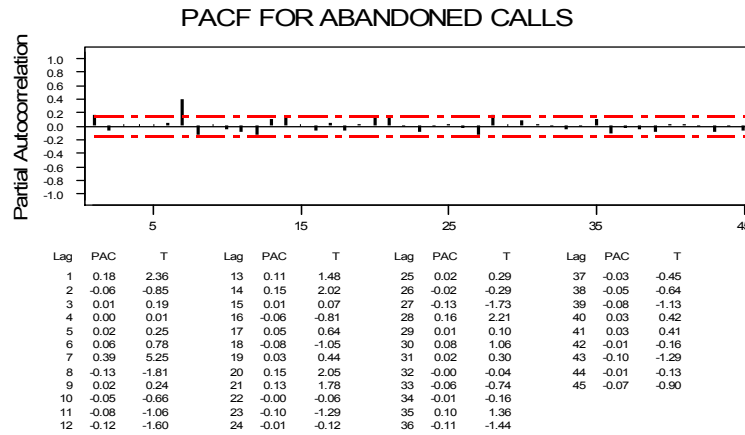


Figure 5: A.C.F. and P.A.C.F. of the fitted model

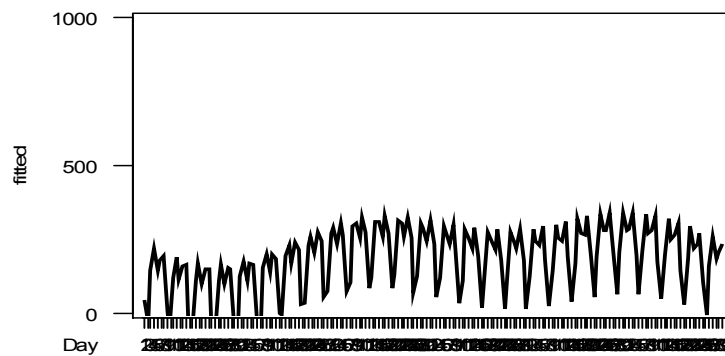


Figure 6: Shows the graph of the fitted data with some forecast.

After an appropriate time series model has been tentatively identified, the next thing is to estimate the parameters of the model. Table 4 gives the final estimates of the model while Figure 6 shows the graph of the fitted model.

Table 4: Final Estimates of Parameters

Type	Coef	StDev	T
AR 1	0.1051	0.4161	0.25
SAR 7	0.1407	0.0868	1.62
MA 1	-0.0814	0.4171	-0.20
SMA 7	0.9356	0.0440	21.27
Constant	3.288	1.243	2.65

From the above table 4, the SARIMA model can be stated fully in equation (6) below as,

$$z_t = 1.246z_{t-1} - 0.155z_{t-2} + 0.015z_{t-3} + a_t - 0.9356a_{t-7} + 0.0814a_{t-1} - 0.076a_{t-8} \dots (6)$$

The chi-square test for adequacy of the model gave the calculated value to be 27.30 as against the tabulated value which gave 55.6. Hence we can conclude that the fitted model is very adequate since the calculated value is less than the tabulated value. Further test of adequacy of the model was done by examining the residuals from the actual and fitted model and this was found to be a white noise process. Fig. 7 shows the plot of the residual versus the fitted values and we can observe that the points are clustered around zero which is an indication of a good fit. Fig. 8 also shows that the histogram of the residual is normally distributed as expected. Figures 9 & 10 shows the P-P plot and Q-Q plot respectively. The correlogram of the standardized residual is shown in figure 11. We observe that all values are within the horizontal dashed line which indicates that the stochastic component of the series is white noise. The model was also found to be both stationary and invertible.

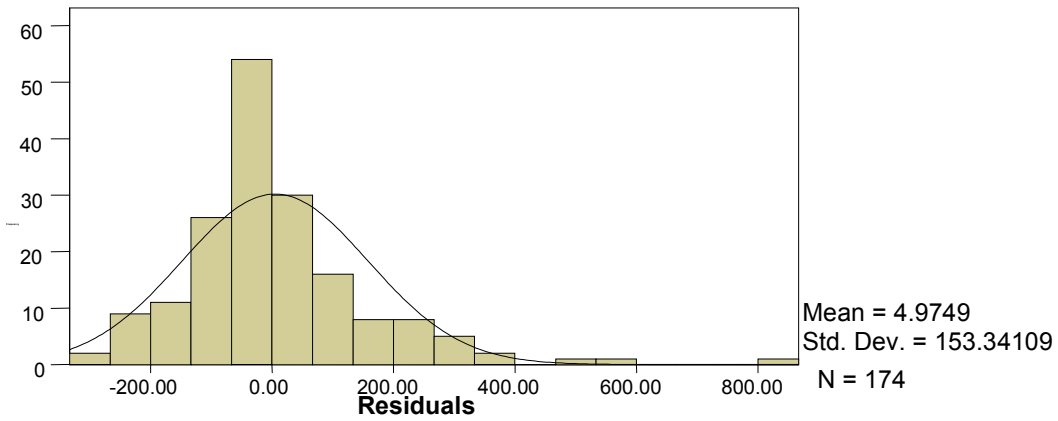
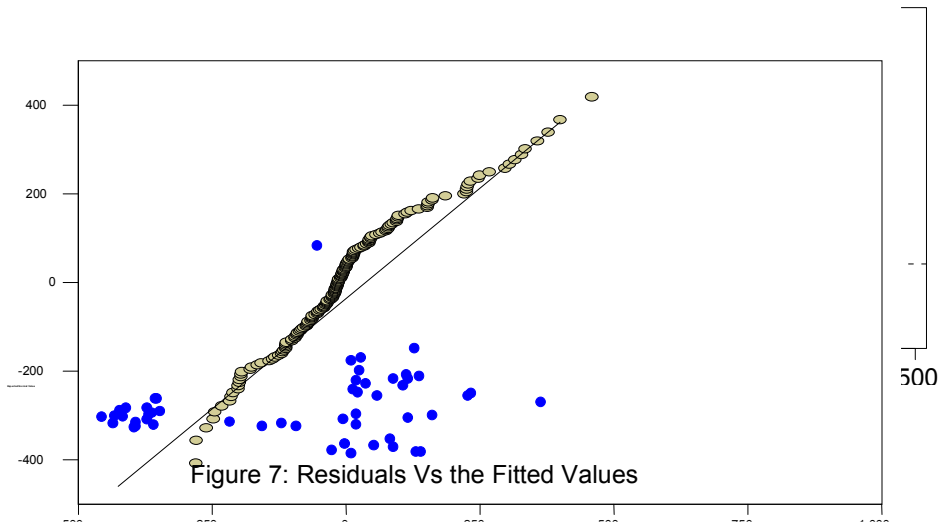
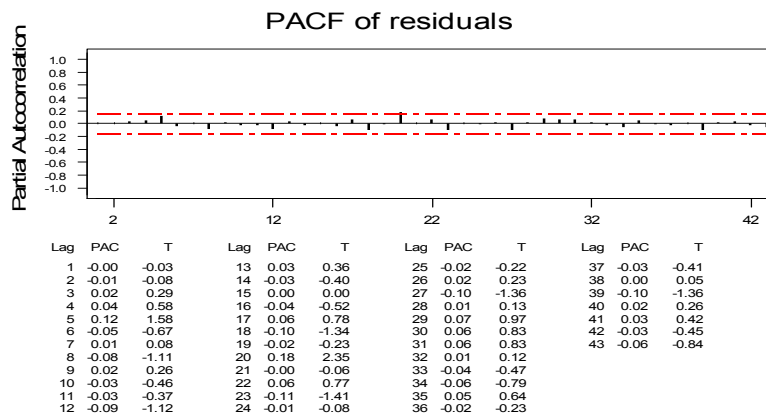


Figure 8: Residual analysis of abandoned calls

Figure 10: Normal Q-Q plot of residuals



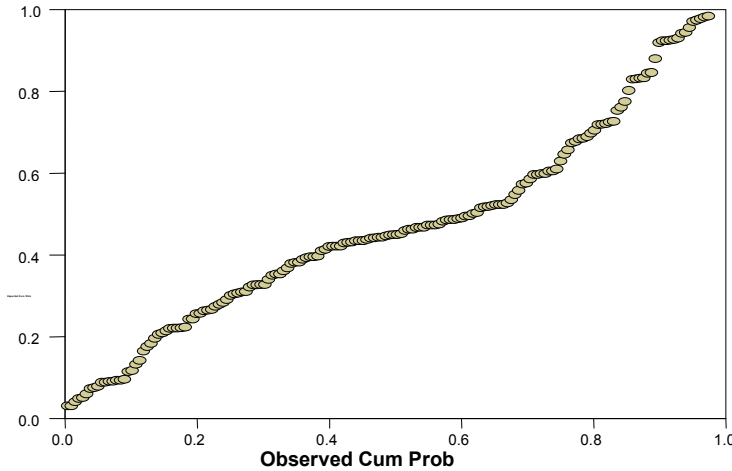


Figure 9: Normal P-P plot of Residuals

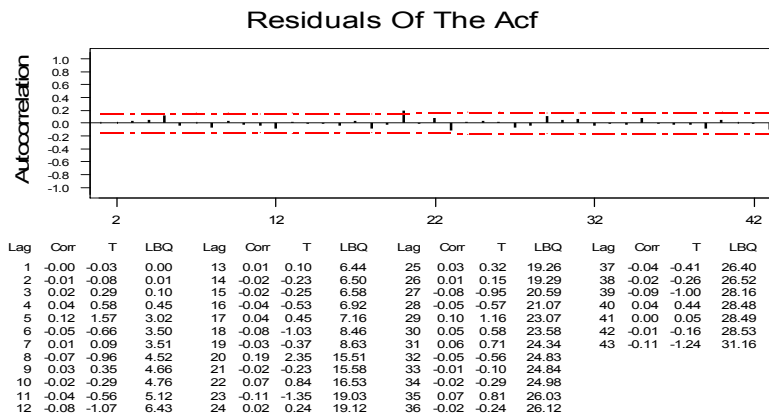


Figure 11: The a.c.f and p.a.c.f of the residuals

4. Efficiency of the fitted models

The foregoing analysis of the two modeling approaches has shown that the fitted models are both adequate and parsimonious. In this circumstance, the identified models using the two approaches could be considered as competing models for a parsimonious representation of the short periodic time series. However, if the Fourier series model is an adequate model, how much do we lose if we use the less parsimonious SARIMA (p,d,q)(P,D,Q)₇ model? To approach this problem, we must first consider how to compare the two models. The parameters themselves are not directly comparable, but we can compare the estimates of the model at comparable time points. Consider the estimates of the two models at $t = 1$. With the Fourier model which has the stochastic component to be white noise then,

$$Var(Z_t) = \frac{5\gamma_0}{n} \quad \text{where } \gamma_0 \text{ is the autocovariance function at lag zero}$$

For the SARIMA model with the stochastic component also being white noise, then the variance is simply given by,

$$Var(Z_t) = \frac{2.465\gamma_0}{N} \quad \text{where } \gamma_0 \text{ is the autocovariance function at lag zero}$$

Thus the ratio of the standard deviation in the Fourier series model to that in the SARIMA model is,

$$\sqrt{\frac{5\gamma_0/n}{2.465\gamma_0/N}} = \sqrt{\frac{5N}{2.465n}}$$

In particular, for the daily abandoned calls, we have $n = 174$ after differencing and $N = 26$ weeks. Thus the ratio is

$$\sqrt{\frac{5(26)}{2.465 \times 174}} = 0.5505$$

Thus in the Fourier series model, we estimate the no of abandoned calls on the first day with a standard deviation that is about 55% as large as it would be if we estimated with the SARIMA model. This however shows that the Fourier series model is more effective than the SARIMA model.

6. CONCLUSION

In this paper we have presented two approaches to modeling a periodic time series. The main emphasis is on Fourier series modeling of daily abandoned calls in a GSM call center and its comparison with the Box and Jenkins model. The analysis involved development of several models using both approaches and selecting the best out of those models. Assessment of the accuracy of the estimates under the approaches was considered. The reliability and efficiency of the approaches were also investigated. The residual standard deviation gives the absolute measure of the goodness of fit of the estimated model and since the value is very low, it indicates a good fit. The t-value or t-ratio shows that in each case the null hypothesis will be rejected. A plot of the standardized residuals versus the corresponding fitted values shows that there is no definite pattern. The frequency histogram of the standardized residuals for the two models shows they are somewhat symmetric and has a tail off at both the high and low ends as a normal distribution does. Further check for normality can be seen from the plot of normal scores or quantile-quantile (QQ) plot. The plot shows the quantiles of the data versus the theoretical quantiles of a normal distribution. With normally distributed data, the QQ plot looks approximately like a straight line. The straight-line pattern of the QQ normal scores plot for the standardized residuals supports the assumption of a normally distributed stochastic component in this model. Finally, a test for possible dependence in the stochastic component was done by plotting the sample autocorrelation of the standardized residuals. It is observed that all values are within the horizontal dashed lines which are placed at zero plus and minus two approximate standard errors of the sample autocorrelation. This shows that it is reasonable to infer that the stochastic component of the series is white noise.

REFERENCES

- Aldor-Noiman S., 2006. Forecasting demand for telephone call center: analysis of desire versus attainable precision, M.Sc thesis, Technion-Israel institute of technology.
- Andrews, B., S.M. Cunningham, 1995. L.L. Bean improves call centre forecasting, *interfaces* 25(6) pg. 1-13.
- Box G.E.P, Jenkins G.M and Reisel G.C., 2008. *Time Sseries Analysis: Forecasting and Control*, Fourth Edition J.Wiley and sons.
- Gans et al, 2003. Telephone call centers: Tutorials, Review, and Research prospects, *MSOM*, vol. 5; no 2, spring 2003 pg.79-141.
- SenGupta A and Ugwuowo, 2006. Asymmetric circular-linear multivariate regression models with applications to environmental data, *Environ Ecol stat* 13: 299-309.
- Taylor, J.W., 2006. A comparison of univariate time series methods for forecasting intraday arrivals at a call center, under review for a focused issue of management science.
- Whitt, W., 2004. Fluid models for many server queues with abandonments. *Operations Research*. To appear.