International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

Queuing Models - A Call Center Case

Ditila Ekmekçiu

[1]Department of Finance, Teleperformance Albania, AMS shpk, Tirana, Albania

Abstract: ***We look at the Erlang C model, as a queuing model generally used to examine call center performance. Erlang C is a simple model, which does not take into consideration caller abandonment and is the model most usually used by experts and researchers. We compare the theoretical performance forecasts of the Erlang C model to a call center simulation model in which a lot of the Erlang C assumptions are liberated. Our discoveries demonstrate that the Erlang C model is subject to important error in forecasting system performance, but that these errors are massively biased and most probably to be pessimistic; the system has the tendency to perform better than forecasted. It can be the case that the model's tendency to give pessimistic estimates helps clarify its continued popularity. Forecast error is powerfully correlated with the abandonment rate thus the model works best in call centers with large numbers of agents and almost low utilization rates.***

Keywords: Call Center, Erlang C, Simulation Model

1. Introduction

A call center is composed by a set of resources (generally staff, computers and telecommunication supplies) that guaranties the delivery of services through the phone. The working environment of a typical large call center could be imagined as very big room with numerous open-space workstations, in which people with earphones sit in front of computer terminals, providing services/products to "unseen" customers (Ekmekçiu, 2015). During the last years the growth of the call center's industry has been very high. In particular in Albania, the number of outsourcing call centers is more than 300 with 20.000 employees (Mapo online, Call-center: Biznesi që vijon lulëzimin). Large range call centers are technologically and organizationally sophisticated operations and have been the object of important academic research. The literature concentrated on call centers is really large, with accurate and comprehensive reviews given in (Aksin, Armony et al. 2007) and (Gans, Koole et al. 2003). Experimental analysis of call center data is provided in (Gans, Brown et al. 2005).

Call centers are illustration of queuing systems; calls come, wait in a virtual line, and then are serviced by an available agent. Call centers are frequently modeled as M/M/N queuing systems, or saying it differently- the Erlang C model. The Erlang C model builds different assumptions that are doubtful in the context of the environment of a call center. In specific the Erlang C model takes for granted that calls arrive at a known average rate, and that they are serviced by a determined number of statistically equal agents with service times that pursues an exponential distribution. Most considerably, Erlang C takes for granted all callers wait as long as it takes for service without hanging up. The model is utilized broadly by both experts and researchers.

Admitting the deficiencies of the Erlang C model, different papers have defended utilizing alternative queuing models and staffing heuristics that take into consideration conditions ignored in the Erlang C model.

The most common alternative is the Erlang A model, an easy extension of the Erlang C model which permits caller abandonment. For example, in a broadly quoted review of the call center literature (Koole, Gans et al. 2003), the authors declare "For this reason, we advise the utilization of Erlang A as the standard to replace the widespread Erlang C model." One more broadly cited paper analyzes experimental data collected from a call center (Gans, Brown et al. 2005) and these authors make a much alike declaration; "utilizing Erlang- A for capacity-planning scopes should make better operational performance. Actually, the model is by now beyond general current practice (that is Erlang-C monopolized), and one goal of this paper is to help change this general situation."

The scope of this study is to regularly examine the fit of the Erlang C model in real call center situations. We look for understanding the nature and magnitude of the error related with the model, and developing a better understanding of what determinants affect prediction error.

2. Queuing Models and the Associated Literature

Queuing models are utilized to forecast system performance of call centers so that the suitable staffing level can be established to obtain a desired performance metric like the Average Speed to Answer, or like the Abandonment rate. The most usual queuing model utilized for inbound call centers is the Erlang C model (Brown, Gans et al. 2005; Gans, Koole et al. 2003). A Google research on "Erlang C Calculator" generates about 8,180,000 items including a large number applications that can be downloaded to calculate staffing requirements based on the Erlang C model.

The Erlang C model (or the M/M/N queue) is a really simple multi-server queuing system. Calls arrive in accordance with a Poisson process at an average rate of λ. Based on the nature of the Poisson process inter-arrival times are independent and equally distributed exponential random variables with mean 1/λ. Calls enter a queue with infinite length and are served on a First In – First Out (FIFO) basis. All calls which enter the queue are served by a group of n homogeneous (or statistically identical) agents at an average rate of nμ. Service times pursue an exponential distribution with a mean service time of 1/μ.

The steady state attitude of the Erlang C queuing model is simply defined, see for example (Koole, Gans et al. 2003). The offered amount, a unit-less quantity frequently cited as the number of Erlangs, is determined as R □ λ/μ. The traffic intensity (or occupancy) is determined as:

$$\rho \,\square\, \lambda/(N\mu) = R/N.$$

Provided the assumption that all calls are served, the traffic intensity should be rigidly less than one or the system becomes not stable; the queue increases without bound. This system may be examined by resolving a set of balance equations and the steady state probability that results that all N agents are busy is

$$\text{P\{Wait > 0\}} = 1\text{-} \left(\sum_{m=0}^{N-1} \frac{R^m}{m!}\right) \Big/ \left(\sum_{m=0}^{N-1} \frac{R^m}{m!} + \left(\frac{R^m}{N!}\right)\left(\frac{1}{1-R/N}\right)\right) \quad (1)$$

Equation (1) computes the proportion of callers which should wait before being served, an important measure of system performance. The Average Speed to Answer is one more important performance measure for call centers managers.

$$\text{ASA} \,\square\, \text{E[Wait]} = \text{P\{Wait > 0\}}\,^{\cdot}\, \text{E[Wait | Wait > 0]}$$
$$= \text{P\{Wait > 0\}} \cdot \left(\frac{1}{N}\right) \cdot \left(\frac{1}{\mu_i}\right) \cdot \left(\frac{1}{1-\rho_i}\right) \quad (2)$$

A third relevant performance metric for call center managers is the Telephone Service Factor, differently called the service level. The TSF is the portion of calls entered that are sooner or later served and for which the delay is lower than a specified level. For example, a call center can report the TSF as the percentage of callers that hold less than 20 seconds. The TSF metric then can be formulated as

$$\text{TSF} \,\square\, \text{P\{Wait} \le \text{T\}} = 1 - \text{P\{Wait > 0\}} \cdot \text{P\{Wait > T | Wait >}$$
$$\}$$
$$= 1 - \text{C}(N, R_i) \cdot e^{-N\mu_i(1-\rho_i)T} \quad (3)$$

A fourth performance metric checked by call center managers is the Abandonment Rate; the quantity of all calls that leave the queue (differently saying, hang up) before being served. Abandonment rates cannot be forecasted immediately utilizing the Erlang C model because the model does not consider abandonment.

An important amount of research examines the behavior of Erlang C model, plenty of it looks for establishing simple staffing heuristics based on asymptotic frameworks implemented to large call centers. (Whitt and Halfin 1981) develop a formal variant of the square root staffing foundation for M/M/N queues in what has become noted as the Quality and Efficiency Driven regime. (Mandelbaum, Borst et al. 2004) develop a foundation for asymptotic optimization of a big call center with no abandonment.

Like in the cases with any analytical model, the Erlang C model does many assumptions, different of which are not totally accurate. In the case of the Erlang C model various assumptions are doubtful, but obviously the most problematic is the lack of abandonment assumption, because even low levels of abandonment can greatly impact system performance (Koole, Gans et al. 2003). A lot of call center research articles however examine call center characteristics under the hypothesis of no abandonment (Green, Kolesar et al. 2001; Jennings and Mandelbaum 1996; Green, Kolesar et al. 2003; Gans and Zhou 2007, Wallace and Whitt 2005; Borst, Mandelbaum et al. 2004).

The Erlang C model takes for granted also that calls arrive based on a Poisson process. The inter-arrival time is a random variable peaked from an exponential distribution with a known arrival rate. Various authors affirm that the hypothesis of a known arrival rate is problematic. The two bigger call center reviews (Aksin, Armony et al. 2007, Gans, Koole et al. 2003) have sections dedicated to arrival rate uncertainty. (Gans, Brown et al. 2005) complete an accurate experimental analysis of call center data. As they find that a time-inhomogeneous Poisson practice suits their data, they also find that arrival rate is difficult to forecast and suggest that the arrival rate has to be modeled as a stochastic process. Different authors discuss that call center arrivals pursue a double stochastic process, a Poisson process where the arrival rate is also a random variable (Whitt 2006; Aksin, Armony et al. 2007; Chen and Henderson 2001). Arrival rate uncertainty can exist for numerous reasons. Arrivals can show randomness higher than that forecasted by the Poisson process because of unobserved variables like the weather or advertising. Call center managers try to account for these factors when they develop predictions, yet predictions can be subject to important error. (Steckley, Henderson et al. 2009) compare estimated and actual volumes for 9 weeks of data captured from 4 call centers. They demonstrate that the forecasting errors are big and modeling arrivals as a Poisson practice with the predicted call volume as the arrival rate can present important error. (Medeiros, Robbins et al. 2006) utilize simulation analysis to evaluate the impact of prediction error on performance measures showing the important impact prediction error can have on system performance.

Some articles address staffing requirements when arrival rates are uncertain. (Harrison, Bassamboo et al. 2005) develop a model that tries to minimize the cost of staffing plus a supposed cost for customer abandonment for a call center with numerous customer and server kinds when arrival rates are uncertain and variable (Harrison and Zeevi 2005) utilize a fluid approximation to resolve the sizing problem for call centers with numerous call types, numerous agent types, and with uncertain arrivals. (Whitt 2006) permits arrival rate uncertainty and uncertain staffing, like absenteeism, too, when calculating staffing requirements. (Henderson, Steckley et al. 2004) analyze the kind of performance measures to utilize when staffing under arrival rate uncertainty. (Harrison and Robbins 2010) develop a scheduling algorithm utilizing a stochastic programming model that is based on uncertain arrival rate predictions.

The Erlang C model as well assumes that the service time follows an exponential distribution. The lack of memory property of the exponential distribution considerably facilitates the calculations necessary to characterize the system's performance, and makes possible the nearly simple

equations (1)-(3). If the hypothesis of exponentially distributed talk time is relaxed, the queuing model that result is the M / G / N queue that is analytically difficult (Koole, Gans et al. 2003) and approximations are requested. Nevertheless, empirical analysis advises that the exponential distribution is an almost poor fit for service times. Most detailed analysis of service time distributions find that the lognormal distribution is a better fit (Gans et al. 2005; Gans, Koole et al. 2003; Brown; Mandelbaum, Sakov et al. 2001).
In the end, the Erlang C model presumes that agents are homogeneous. More accurately, it is presumed that the service times follow the identical statistical distribution independent of the particular agent handling the call. Empirical evidence supports the concept that some agents are more efficient than others and the distribution of call time depends on the agent where the call is routed. Particularly more experienced agents generally handle calls faster than new hire agents (Ward and Armony 2008).

3. Call Center Simulation

3.1 The Altered Model

In this section we present a corrected model of a call center, relaxing key hypothesis discussed earlier. In our model calls arrive at the call center according to a Poisson process. Calls are predicted to arrive at an average rate of λˆ. The realized arrival rate is λ, where λ is a normally distributed random variable with mean λˆ, standard deviation σλ and coefficient of variation cλ= σλ/ λˆ. The choice of the normal distribution provides us a symmetric distribution centered on the predicted value. A disadvantage of the normal distribution is the probability of generating negative values. Nevertheless, in our experiments the mean value is adequately positive, a minimum of 5 standard deviations, which is not a problem. The time requested to process a call by an average agent is a lognormally distributed random variable with mean 1/μ and standard deviation σμ. Arriving calls are routed to the agent who has been ineffective for the longest time if one is available. If all agents are busy the call is place in a FIFO queue. Once put in a queue a proportion of callers will balk because they directly hang up. Callers that join the queue have a patience time that follows a Weibull distribution. If wait time surpasses their patience time the caller will abandon.

Calls are served by agents who have variable relative productivity ri. Agent productivity is hypothesized to be a normally distributed random variable with a mean of 1 and a standard deviation of σr. An agent with a relative productivity level of 1, for instance, serves calls at the average rate. An agent with a relative productivity level of 1.75 serves calls at 1.75 times the average rate, an agent with a productivity level of .5 serves calls at .5 times the average rate. Provided the mean productivity level of 1, calls are served at the rate l, on average.

3.2 Experimental Design

In order to judge the performance of the Erlang C versus the simulation model we organize a series of designed experiments. Based on the hypothesis for our call center discussed earlier, we determine the following set of nine empirical factors.

Table 1: Experimental Factors

	Factor	Low	High
1	Number of Agents	10	100
2	Offered Utilization ($\hat{\rho}$)	65%	95%
3	Talk Time (mins)	2	20
4	Patience β	60	600
5	Forecast Error CV (c_λ)	0	.2
6	Patience α	.75	1.25
7	Talk time CV	.75	1.25
8	Probability of Balking	0	.25
9	Agent Productivity Standard Deviation	0	.15

The predicted arrival rate in the simulation is a quantity that derives from other experimental factors by

$$\hat{\lambda} = \hat{\rho} N \mu \quad (4)$$

Provided the relatively large number of empirical factors, a well-designed experimental approach is required to evaluate without difficulty the experimental region. A standard approach to designing computer simulation experiments is to apply a full or fractional factorial design (Law 2007). Nevertheless, the factorial model evaluates only corner points of the experimental region and inevitably hypothesized that responses are linear in the design space. Provided the anticipated non-linear relationship of errors we preferred to implement a Space Filling Design based on Latin Hypercube Sampling as examined in (Williams, Santner et al. 2003). Provided a wanted sample of n points, the experimental region is divided into nd cells. A sample of n cells is chosen in a way that the centers of these cells are uniformly spread when projected into each axis of the design space. While the LHS design is not absolutely orthogonal like a factorial design, the design does give for a low correlation between input factors highly reducing the risk of multicollinearity. We picked our design point as the center of any selected cell.

3.3 Simulation Model

The model is judged utilizing a simple discrete event simulation model. The scope of the model is to forecast the long term, steady state behavior of the queuing system. The model produces random numbers utilizing a combined multiple recursive generator (CMRG) based on the Mrg32k3a generator explained in (L'Ecuyer 1999). Ordinary random numbers are utilized across design points to reduce output variance. To diminish any start up bias we utilize a warm up period of 5,000 calls, after which all statistics are reset. Then the model is run for an evaluation period of 25,000 calls and summary statistics are put together.

For each design point we repeat this process for 500 replications and report the average value across replications.

The particular process for each replication is in this manner. The input factors are picked based on the empirical design. The average arrival rate is calculated based on the stated number of agents, talk time, and offered utilization rate in accordance with equation (4). A random number is drawn and the accomplished arrival rate is set based on the probability distribution of the prediction error. That arrival rate is after that utilized to generate Poisson arrivals for the reproduction. Agent productivities are generated utilizing a normal distribution with mean 1 and standard deviation σρ. Every new call generated contains a lognormally distributed average talk time, an exponentially distributed inter-arrival time, a Weibull distributed time before abandonment, and a Bernoulli distributed balking indicator. When the call comes it is assigned to the longest ineffective agent, or put in the queue if all agents are already busy. If it is sent to the queue the simulation model examines the balking indicator. If the call has been recognized as a balker it is directly abandoned, in case it is not, an abandonment event is scheduled based on the accomplished time to abandon. When the call has been assigned to an agent, the accomplished talk time is calculated by multiplying the average talk time with the agent's productivity. The agent is dedicated for the accomplished talk time. Once the call terminates the agent processes the next call from the queue, or in case of no calls are queued becomes ineffective. If a call is processed before its time to abandon, the abandonment event is cancelled. If not, the call is abandoned and cancelled from the queue when the patience time concludes.

After all replications of the design point have been performed the results are compared to the theoretical forecasts of the Erlang C model. We measure the error as the difference between the theoretical value and the simulated value. We make a relative error calculation in order that the sign of the error shows the bias in the calculation. In our experiment we evaluated an LHS sample of 1000 points.

4. Experimental Analysis

4.1. Summary Observations

Based on our examination we can make the summary observations as below:

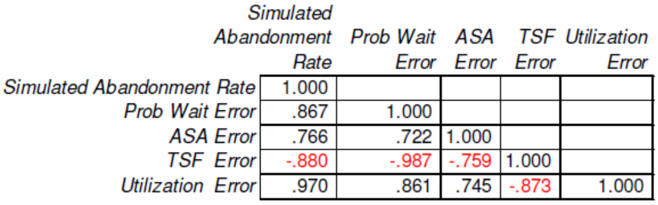
- The Erlang C model is subject to a fairly large error over this range of parameter values, on average,
- Measurement errors are greatly positively correlated across performance measures.
- The Erlang C model is pessimistically biased, on average (the real system performs better than forecasted), but can become optimistically biased when usage is high and arrival rates are uncertain.
- Measurement error is high when the real system demonstrates higher levels of abandonment. The error is powerfully positively correlated with accomplished abandonment rate and forecasted ASA.
- The Erlang C model is most correct when the number of agents is large and usage is low.
- Errors diminish as caller patience increases.

Now we will review our empirical results in more detail.

4.2. Correlation and Magnitude of Errors

The magnitude of errors produced by utilizing the Erlang C model across our test space is on average high and very high in a few cases. The errors across the key metrics are greatly correlated with each other, and highly correlated with the accomplished abandonment rate. In Table 2 is demonstrated a correlation matrix of the errors produced by the Erlang C model.

Table 2: Error Correlation Matrix



	Simulated Abandonment Rate	Prob Wait Error	ASA Error	TSF Error	Utilization Error
Simulated Abandonment Rate	1.000				
Prob Wait Error	.867	1.000			
ASA Error	.766	.722	1.000		
TSF Error	-.880	-.987	-.759	1.000	
Utilization Error	.970	.861	.745	-.873	1.000

Correlations between measure errors are powerful. The measured errors move all, on average, in an optimistic or pessimistic direction together. Prob -Wait and ASA are positively correlated; it is wanted for the two of these measures to be low. Prob -Wait is correlated with TSF negatively; a measure for which a high value is wanted. Measurement error is also greatly correlated with abandonment rate. Provided the high correlation between measures we will use Prob - Wait as a proxy for the global error of the Erlang C model.

Average error rates are fairly high under the Erlang C model, with errors being pessimistically altered. Figure 1 gives a histogram of the Prob - Wait error.

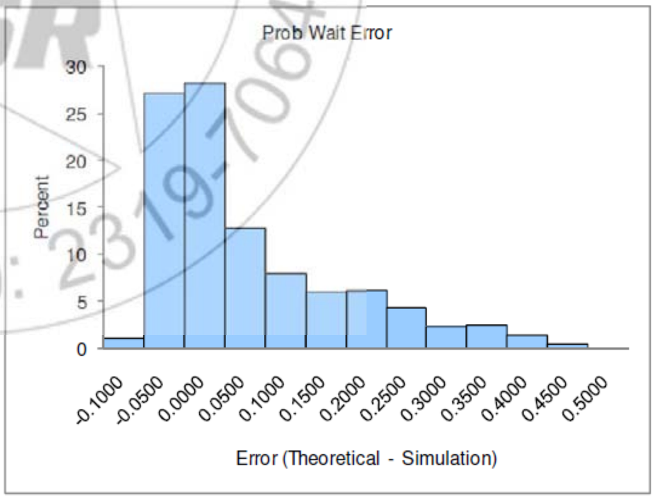


Figure 1: Histogram of Erlang C Prob Wait Errors

The average error is 7.96%, and the data has a strong positive alter; 72% of the errors being positive.
The largest error is 49.4%, the smallest is -8.0 %.

4.3. Drivers of Erlang C Error

Having settled that error rates are high under the Erlang C model, now we show attention to characterizing the drivers of that error. As argued in the earlier section, Erlang C errors are gretaly correlated with the accomplished abandonment

DOI: 10.21275/ART2017812

rate. The notion that abandonment is a dominant driver of errors in the Erlang C model is additionally shown in Figure 2. This graph demonstrates the error in the Prob - Wait measure on the vertical axis and the abandonment rate from the simulation examination on the horizontal axis

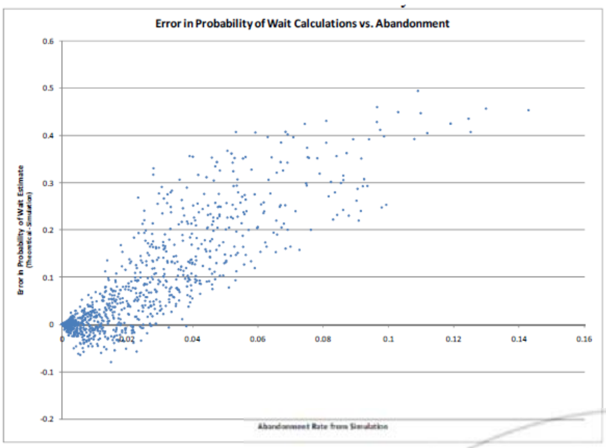


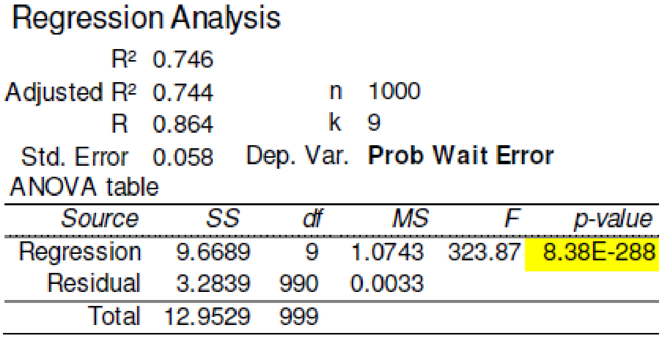
Figure 2: Scatter Plot of Erlang C Errors and Abandonment Rate

The graph shows without any doubt that as abandonment gets higher, the error in the Prob - Wait measure increases as well. The graph also shows that optimistic errors, that are errors in which the system performed worse than forecasted, happen only with relatively low abandonment rates. The average abandonment rate for optimistic forecasts was .74%. The graph shows also that significant error can be related with even low to moderate abandonment rates. As an illustration, for all test points with abandonment rates of less than 5%, the average error for Prob - Wait is 4.8%. For test points where abandonment covered the range between 2% and 5% the average Prob - Wait error is 12.2%.

To determine how each of the nine empirical factors impacts the error, we perform a regression analysis.

The dependent variable is the Prob - Wait error. For the independent variable we utilize the nine experimental factors normalized to a [-1, 1] scale. This normalization permits us to better determine the relative impact of every factor. The LHS sampling method gives an experimental design where the correlation between experimental factors is low, highly reducing risks of multicollinearity. The results of the regression analysis are demonstrated in Table 3.

Table 3: Regression Analysis of Prob - Wait Error

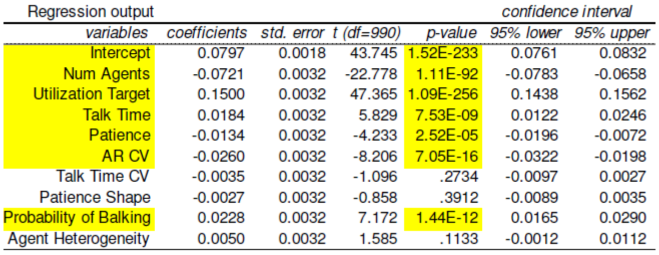


Regression Analysis

R²	0.746		
Adjusted R²	0.744	n	1000
R	0.864	k	9
Std. Error	0.058	Dep. Var.	Prob Wait Error

ANOVA table

Source	SS	df	MS	F	p-value
Regression	9.6689	9	1.0743	323.87	8.38E-288
Residual	3.2839	990	0.0033		
Total	12.9529	999			



Regression output

variables	coefficients	std. error	t (df=990)	p-value	confidence interval 95% lower	95% upper
Intercept	0.0797	0.0018	43.745	1.52E-233	0.0761	0.0832
Num Agents	-0.0721	0.0032	-22.778	1.11E-92	-0.0783	-0.0658
Utilization Target	0.1500	0.0032	47.365	1.09E-256	0.1438	0.1562
Talk Time	0.0184	0.0032	5.829	7.53E-09	0.0122	0.0246
Patience	-0.0134	0.0032	-4.233	2.52E-05	-0.0196	-0.0072
AR CV	-0.0260	0.0032	-8.206	7.05E-16	-0.0322	-0.0198
Talk Time CV	-0.0035	0.0032	-1.096	.2734	-0.0097	0.0027
Patience Shape	-0.0027	0.0032	-0.858	.3912	-0.0089	0.0035
Probability of Balking	0.0228	0.0032	7.172	1.44E-12	0.0165	0.0290
Agent Heterogeneity	0.0050	0.0032	1.585	.1133	-0.0012	0.0112

Provided the normalization of the experimental factors, the magnitude of the regression coefficients gives a direct evaluation of the impact that a factor has on the measurement error. The factor which most powerfully influences the error is the offered usage, the magnitude of its coefficient being more than twice the value of the next measure and more than five times the magnitude of all other factors. The size of the call center, calculated as the number of agents, has a bigger influence on errors. Factors connected to disposition to wait, that are Patience, Probability of Balking, and Patience Shape, all have low to temperate impacts, but with exception of Patience Shape are statistically important. Also talk time is a statistically important factor with a balanced impact. The volatility of talk time and agent heterogeneity both have low impacts that are not statistically important.

The most significant drivers of Erlang C errors are the size and usage of the call center. This is additionally shown in Figure 3. This graph illustrates the results of an experiment where the number of agents and usage factors are altered in a controlled fashion. All other experimental factors are kept at their mid-point.

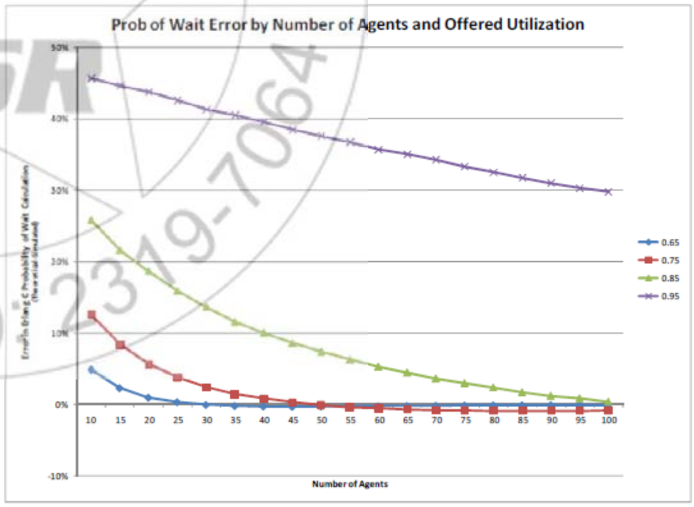


Figure 3: Erlang C ProbWait Errors by Call Center Size and Utilization

This graph shows that the Erlang C model has the tendency to provide almost poor forecasts for small call centers. This error has the tendency to diminish as the size of the call center increases. Nevertheless, the graph shows also that for busy centers the error stays high. Running at 95% offered usage, for a very busy call center, the error rate stays at 30%, even with a group of 100 agents. The errors have the tendency to track with abandonment; abandonment rates increase with usage and diminish with the agent group.

DOI: 10.21275/ART2017812

The conclusion that abandonment behavior drives the Erlang C error is then shown in Figure 4.

In this experiment we regularly change two Willingness to Wait parameters. In specific, we change the balking probability and the b factor of patience distribution.

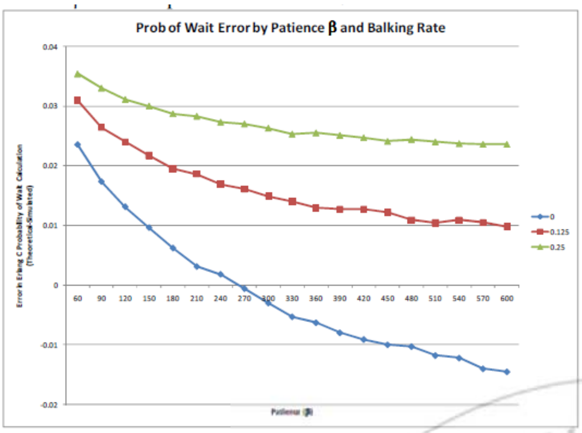


Figure 4: Erlang C ProbWait Errors by Willingness to Wait

This examination validates that the higher the probability that callers balk, the higher the error rate. The analysis demonstrates, too, that the more patient the callers are, the higher are the error rates. The higher the probability that callers abandon, either directly or just after being queued, the higher the abandonment rate and the less accurate become the Erlang C measures.

Another factor of interest is the uncertainty related with the arrival rate. While its global effect is not large (about 1.8%), it has effects that are not alike to other empirical factors as demonstrated in Figure 5. This graph illustrates the results of an experiment that changes the coefficient of variation of the arrival rate error and the number of agents while keeping all other indicators at their mid-points.

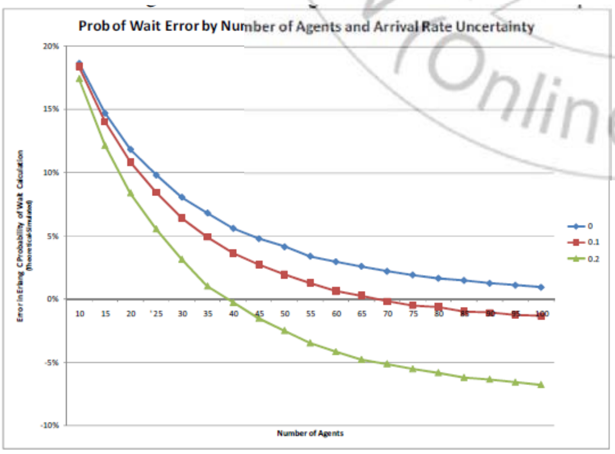


Figure 5: Erlang C Pro - bWait Errors by Call Center Size and Forecast Error

This experiment demonstrates that for small call centers arrival rate uncertainty has a small impact, but that effect becomes more noticeable for larger call centers. Also it is worth noticing that arrival rate uncertainty has a confident effect, and for high levels of uncertainty the model displays an optimistic bias. Arrival rate uncertainty is a dominant factor leading to an optimistic estimate from the Erlang C model; from the 21.9% of test points with a confident bias the average arrival rate uncertainty cv was 14.0%. As the arrival rate uncertainty has the tendency to bias the forecast in the opposite direction of most other factors, it has also the effect of reducing error in a lot of situations. As in this case, high usage tends to bias the estimate pessimistically, a bias diminished when arrival rate uncertainty exists.

5. Summary and Conclusions

The Erlang C model is usually implemented to forecast queuing system behavior in call center applications. Our analysis demonstrates that when we test the Erlang C model over a range of acceptable conditions forecasted performance measures are dependent to large errors. The Erlang C model works fairly well for large call centers with low to moderate usage rates, but factors that have the tendency to generate caller abandonment; like high usage, impatient callers, and small agent pools cause the model error to become quite large. While the model has the tendency to give a pessimistic estimate, arrival rate uncertainty will either diminish that bias or lead to an optimistic bias. It can be the case that the model's tendency to give pessimistic (that is conservative) forecasts helps explain its continued popularity. It is obvious that high care must be done before utilizing the Erlang C model to do any calculations which require a high level of precision.

Our future research is concentrated on examining the increasingly popular Erlang A model and comparing its performance to the Erlang C model to test the growing agreement that Erlang A is a greater model for call center analysis.

References

[1] Aksin, Z., M. Armony and V. Mehrotra. 2007. The Modern Call-Center: A Multi-Disciplinary Perspective on Operations Management Research. Production and Operations Management 16: 665-668.

[2] Armony, M. and A. R. Ward 2008. Fair Dynamic Routing in Large-Scale Heterogeneous-Server Systems, Stern School of Business, NYU.

[3] Bassamboo, A., J. M. Harrison and A. Zeevi. 2005. Design and Control of a Large Call Center: Asymptotic Analysis of an LP-based Method. Operations Research 54: 419-435.

[4] Borst, S., A. Mandelbaum and M. I. Reiman. 2004. Dimensioning Large Call Centers. Operations Research 52: 17-35.

[5] Brown, L., N. Gans, A. Mandelbaum, et al. 2005. Statistical Analysis of a Telephone Call Center: A Queueing-Science Perspective. Journal of the American Statistical Association 100: 36-50.

[6] Chen, B. P. K. and S. G. Henderson. 2001. Two Issues in Setting Call Centre Staffing Levels. Annals of Operations Research 108: 175-192.

[7] Ekmekçiu, 2015. Optimizing a call center performance using queueing models – an Albanian Case. 5th International Conference - "Compliance of the Standards

in South-Eastern European Countries with the Harmonized Standards of European Union", 15-16 June, 2015 Peja, Republic Of Kosovo.

[8] Gans, N., G. Koole and A. Mandelbaum. 2003. Telephone call centers: Tutorial, review, and research prospects. Manufacturing & Service Operations Management 5: 79-141.

[9] Gans, N. and Y.-P. Zhou. 2007. Call-Routing Schemes for Call-Center Outsourcing. Manufacturing & Service Operations Management 9: 33-51.

[10] Green, L. V., P. Kolesar and J. Soares. 2003. An Improved Heuristic for Staffing Telephone Call Centers with Limited Operating Hours. Production and Operations Management 12: 46-61.

[11] Green, L. V., P. J. Kolesar and J. Soares. 2001. Improving the SIPP Approach for Staffing Service Systems That Have Cyclic Demands. Operations Research 49: 549-564.

[12] Halfin, S. and W. Whitt. 1981. Heavy-Traffic Limits for Queues with Many Exponential Servers. Operations Research 29: 567-588.

[13] Harrison, J. M. and A. Zeevi. 2005. A Method for Staffing Large Call Centers Based on Stochastic Fluid Models. Manufacturing & Service Operations Management 7: 20-36.

[14] Jennings, O. B. and A. Mandelbaum. 1996. Server staffing to meet time-varying demand. Management Science 42: 1383.

[15] L'Ecuyer, P. 1999. Good Parameters and Implementations for Combined Multiple Recursive Random Number Generators. Operations Research 47: 159-164.

[16] Law, A. M. 2007. Simulation modeling and analysis. Boston, McGraw-Hill.

[17] Mandelbaum, A., A. Sakov and S. Zeltyn 2001. Empirical Analysis of a Call Center, Technion - Israel Institute of Technology.

[18] Robbins, T. R. 2007. Managing Service Capacity Under Uncertainty - Unpublished PhD Dissertation Pennsylvania State University. University Park, PA. Avaialble via (http://personal.ecu.edu/robbinst/)

[19] Robbins, T. R. and T. P. Harrison. 2010. Call Center Scheduling with Uncertain Arrivals and Global Service Level Agreements. European Journal of Operational Research Forthcoming.

[20] Robbins, T. R., D. J. Medeiros and P. Dum 2006. Evaluating Arrival Rate Uncertainty in Call Centers. 2006 Winter Simulation Conference, Monterey, CA.

[21] Santner, T. J., B. J. Williams and W. Notz 2003. The design and analysis of computer experiments. New York, Springer.

[22] Steckley, S. G., S. G. Henderson and V. Mehrotra. 2009. Forecast Errors in Service Systems. Probability in the Engineering and Informational Sciences: 305-332.

[23] Steckley, S. G., W. B. Henderson and V. Mehrotra 2004. Service System Planning in the Presence of a Random Arrival Rate, Cornell University.

[24] Wallace, R. B. and W. Whitt. 2005. A Staffing Algorithm for Call Centers with Skill-Based Routing. Manufacturing & Service Operations Management 7: 276-294.

[25] Whitt, W. 2006. Staffing a Call Center with Uncertain Arrival Rate and Absenteeism. Production and Operations Management 15: 88-102.

.