$\mathcal{I}^2\mathcal{SDS}$

The Institute for Integrating Statistics in Decision Sciences

Technical Report TR-2008-18

October 6, 2008

Call Center Arrival Modeling: A Bayesian State Space Approach

Tevfik Aktekin
Department of Decision Sciences
The George Washington University, USA

Refik Soyer
Department of Decision Sciences
The George Washington University, USA

CALL CENTER ARRIVAL MODELING: A BAYESIAN STATE SPACE APPROACH

Tevfik Aktekin
Department of Decision Sciences
The George Washington University

Refik Soyer
Department of Decision Sciences
The George Washington University

10/06/2008

Abstract

In this paper we introduce a discrete time Bayesian state space model with Poisson measurements for intra-day call arrivals. We present the properties of our model and develop Bayesian inference. In so doing, we provide analytically tractable expressions for sequential updating for parameters, for smoothing and prediction of call arrivals and discuss how the model can be used for inter-weekly forecasts. We illustrate the implementation of the model by using actual intra-day arrival data from a US commercial bank's call center.

1 Introduction

Call centers provide vital contact points between firms and their potential and existing customers. In general, customers who are looking for service are served either by an interactive voice response unit (IVR) or by agents of the call center. The focus of this paper will be on calls handled by agents. In recent years, academic research in call center modeling has shown a drastic increase in volume. In a survey paper by Gans et al. (2003), 164 call center related papers are referenced. A detailed overview of recent call center related literature can be found in Aksin et al. (2007), where different research areas such as forecasting, queueing, scheduling and behavioural issues are considered. Call center management consists of technology and labour intensive operations, where staffing comprise 60-80% of the overall operating budget (Aksin et al. (2007)). Therefore modeling uncertainty in call center arrivals is an important issue for call center management. In this paper, we address the issue of modeling daily call volumes in an inbound call center with particular emphasis on intra-day forecasting and illustrate how inter-week forecasts can be developed from the proposed models.

Due to the increased availability of call center databases, forecasting related studies have been an active area for academic research. Some of the recent work in this area can be summarized as follows. Jongbloed and Koole (2001) discuss that the arrival process at a Dutch insurance company call center follow a non-homogeneous Poisson process where the arrival rate is not constant over time. Avramidis et al. (2004) develop a model based on the correlation structure of the intra-day arrivals at a Bell Canada call center using doubly stochastic Poisson processes. Recent work by Weinberg et al. (2007) and Soyer and Tarimcilar (2008) are the first two known attempts to model call volumes using a Bayesian point of view. Soyer and Tarimcilar (2008) model the arrival process at a consumer electronics producer call center via a modulated Poisson process whose rate is a function of advertising specific characteristics as well as time. Weinberg et al. (2007) consider the intra-day arrival behaviour of a US commercial bank's call center and introduce a non-homogeneous Poisson model whose arrival rates are changing both intra-day and inter-day. A sequential Monte Carlo algorithm is developed in order to estimate both the latent states and the model parameters. In order to develop the MCMC algorithm, the arrival count data is transformed into a Gaussian time series using a square root transformation. As noted by Aksin et al. (2007)the approach proposed by Weinberg et al. (2007) is computationally intensive especially for intra-day forecast updating.

Intra-day forecasting is an important issue in call center staffing decisions. Whitt (1999) mentions the importance of near term staffing needs of a call center in order to carry out dynamic staffing with the aim of answering all calls. Near term staffing is usually referred to as the staffing requirements in the next minute, five minutes or ten minutes. Conventional methods used in the industry for call center staffing problems require that mean of call volume at any given time interval is forecasted so that the required staffing level is $c\sqrt(m)$, where c is a constant defined over $1 \leq c \leq 10$ and m is the predicted mean for a given time interval. Another approach proposed by Whitt (1999) for dynamic staffing is as follows. Let $s(t)$ be the required number of servers at time t, then

$$s(t) = [E(D(t)) + z_\alpha \sqrt(Var(D(t))) + 0.5] \tag{1.1}$$

where $P(N(0,1) > z_\alpha) = \alpha$, $[x]$ is the least integer greater than x, $D(t)$ and $V(t)$ are the respective mean and the variance of the predicted demand for time t. Therefore, availability of intra-day updating call enables call center managers adjust their staffing dynamically on a given day. Whitt (1999) also introduces the idea of flexible staffing where the call center agents have alternative work on a given day. In other words if the call volume is predicted to be low in the morning they can be assigned to other tasks, as soon as intra-day forecasts indicate an increase in the call arrivals they can be reassigned to answering inbound calls. Thus, it is important to able to update forecasts rather easily at any time period in a given day.

In this paper, we propose an alternative to Weinberg et al. (2007) via modeling the original count data rather than using the square root transformation and consider a discrete time Bayesian state space model with Poisson measurements. The attractive feature of our model is that, under certain conditions, it provides analytically tractable expressions for sequential updating for the model parameters, for smoothing and predictions of call arrivals. Thus, it can easily be used with decision models for updating intra-day staffing schedules. Furthermore, the proposed approach is quite valuable in inferring inter-day as well as intra-day differences in arrival patterns and it can be extended to obtain for inter-week forecasts.

A synopsis of our paper is as follows: In Section 2, we introduce a Bayesian state space model for the daily within day call arrivals, show how the sequential updating of the parameters is implemented and discuss the underlying properties of the proposed model. In section 3, we present an extension of the model to consider interweekly forecasts. In section 4, a numerical example is illustrated for the model. Finally, in section 5 we conclude with comments and suggestions for future extensions of our work.

2 A Discrete Time Poisson Model for Intra-day Call Arrivals

Queuing theory has been commonly used in the call center modelling literature. However, appropriateness of the standard assumptions of queuing models have been questioned by many researchers. For example, Jongbloed and Koole (2001) and Weinberg et al. (2007) note that the assumption of Poisson arrivals with a constant rate over time does not apply to call center arrivals. Therefore they propose a stochastic arrival rate, θ_t, which evolves over time. Similarly, the independent increments property of Poisson processes has been criticized by Avramidis et al. (2004) who propose a Poisson processes with stochastic arrival intensities. In this paper, we consider a discrete time Poisson process with a stochastic arrival rate that evolves over time according to a discrete time Markov process. Our model and its properties are discussed in the sequel.

Let N_t be the number of call arrivals during the time period $t-1$ to t in a given day and θ_t be the corresponding arrival rate during the same time period. Note that the length of the time interval may change depending on the call arrival data at hand. For example, N_t may represent the call arrivals during the tth minute, tth hour or tth day depending on the nature of the data, but the proposed model can still be applicable.

Given the arrival rate θ_t, we assume that the number of call arrivals during period t is described by a discrete time nonhomogeneous Poisson process with probability distribution

$$p(N_t|\theta_t) = \frac{\theta_t^{N_t} e^{-\theta_t}}{N_t!}. \tag{2.1}$$

It is assumed that given θ_t's, N_t's are independent. In other words, the model implies that the independent increments property holds only conditional on θ_t and N_t's are correlated. In the above, (2.1) acts as a measurement equation and is defined over discrete space in time.

For time evolution of θ_t's, we assume a Markovian structure where the relationship between the arrival rates is described by

$$\theta_t = \frac{\theta_{t-1}}{\gamma}\epsilon_t \tag{2.2}$$

where $(\epsilon_t|N^{(t-1)}) \sim Beta(\gamma\alpha_{t-1}, (1-\gamma)\alpha_{t-1})$ with $\alpha_{t-1} > 0$, $0 < \gamma < 1$, and $N^{(t-1)} = (N_1, \cdots, N_{t-1})$. In (2.2), γ acts like a discounting term. It follows from the above that $\theta_t < \frac{\theta_{t-1}}{\gamma}$. We can show that

$\theta_t|\theta_{t-1}, N^{(t-1)} \sim Beta(\gamma\alpha_{t-1}, (1-\gamma)\alpha_{t-1}, 0, \frac{\theta_{t-1}}{\gamma})$, that is, a truncated Beta density given by

$$p(\theta_t|\theta_{t-1}, N^{(t-1)}) = \frac{\Gamma(\alpha_{t-1})}{\Gamma(\gamma\alpha_{t-1})\Gamma((1-\gamma)\alpha_{t-1})}(\frac{\gamma}{\theta_{t-1}})^{\alpha_{t-1}-1}\theta_t^{\gamma\alpha_{t-1}-1}(\frac{\theta_{t-1}}{\gamma} - \theta_t)^{(1-\gamma)\alpha_{t-1}-1}. \quad (2.3)$$

The state equation (2.2) also implies that $E(\theta_t|\theta_{t-1}, N^{(t-1)}) = \theta_{t-1}$, that is, a random walk type of evolution for the arrival rates. We note that a modified version of the evolution model (2.2) was first introduced in Smith and Miller (1986) where the measurement equation was exponential and later was used by Morali and Soyer (2003) in the context of software reliability.

A Poisson measurement model has not been considered in the earlier work, but it is possible to develop an analytically tractable Bayesian analysis for the model if we assume that at time 0, $(\theta_0|N^{(0)})$ is a gamma distribution. In other words, before any of the call arrivals have been observed, we define our prior uncertainty regarding the call arrival rate as

$$(\theta_0|N^{(0)}) \sim Gamma(\alpha_0, \beta_o). \quad (2.4)$$

Following Smith and Miller (1986) we assume that

$$(\theta_{t-1}|N^{(t-1)}) \sim Gamma(\alpha_{t-1}, \beta_{t-1}) \quad (2.5)$$

which we can show by induction.

Using (2.3) and (2.5), we can obtain the prior of θ_t given $N^{(t-1)}$ via

$$p(\theta_t|N^{(t-1)}) = \int_{\gamma\theta_t}^{\infty} p(\theta_t|\theta_{t-1}, N^{(t-1)})p(\theta_{t-1}|N^{(t-1)})d\theta_{t-1}, \quad (2.6)$$

which reduces to a gamma density as

$$(\theta_t|N^{(t-1)}) \sim Gamma(\gamma\alpha_{t-1}, \gamma\beta_{t-1}). \quad (2.7)$$

It follows from the above that $E(\theta_t|N^{(t-1)}) = E(\theta_{t-1}|N^{(t-1)})$, whereas $V(\theta_t|N^{(t-1)}) = \frac{V(\theta_{t-1}|N^{(t-1)})}{\gamma}$. In other words, the model implies that as we move forward in time, expected call arrival rate stays the same but our uncertainty about arrival rate increases as a function of the discount factor γ.

Given the prior (2.5) and the Poisson observation model (2.1) we obtain the posterior distribution $\theta_t|N^{(t)}$ using the Bayes' Rule,

$$p(\theta_t|N^{(t)}) \propto p(N_t|\theta_t)p(\theta_t|N^{(t-1)}). \tag{2.8}$$

The above implies that

$$p(\theta_t|N^{(t)}) \propto \theta_t^{\gamma\alpha_{t-1}+N_t-1} e^{-(\gamma\beta_{t-1}+1)\theta_t},$$

that is, the posterior distribution of the call arrival rate at time t is a gamma density

$$\theta_t|N^{(t)} \sim Gamma(\alpha_t, \beta_t) \tag{2.9}$$

where $\alpha_t = \gamma\alpha_{t-1} + N_t$ and $\beta_t = \gamma\beta_{t-1} + 1$.

The one-step ahead predictive distribution of call arrivals at time t given $N^{(t-1)}$ can be obtained via

$$p(N_t|N^{(t-1)}) = \int_0^\infty p(N_t|\theta_t)p(\theta_t|N^{(t-1)})d\theta_t \tag{2.10}$$

where $N_t|\theta_t \sim Poisson(\theta_t)$ and $\theta_t|N^{(t-1)} \sim Gamma(\gamma\alpha_{t-1}, \gamma\beta_{t-1})$. Therefore,

$$p(N_t|N^{(t-1)}) = \binom{\gamma\alpha_{t-1} + N_t - 1}{N_t}(1 - \frac{1}{\gamma\beta_{t-1}+1})^{\gamma\alpha_{t-1}}(\frac{1}{\gamma\beta_{t-1}+1})^{N_t}. \tag{2.11}$$

which is a negative binomial model denoted as

$$N_t|N^{(t-1)} \sim Negbin(\gamma\alpha_{t-1}, \gamma\beta_{t-1}). \tag{2.12}$$

Availability of one-step ahead predictive density in closed form is an attractive feature of the model from a practical point of view. Given (2.12), one can carry out one step ahead predictions and credibility interval calculations in a straightforward manner. Further results regarding one step ahead predictive distributions will be discussed in our numerical example section.

Although the k-step ahead predictive density is not analytically available, the k-step ahead predictive means can be easily obtained. Using a standard conditional expectation argument one can obtain $E(N_{t+k}|N^{(t)})$ as follows

$$E(N_{t+k}|N^{(t)}) = E_{\theta_{t+k}}(E(N_{t+k}|\theta_{t+k}, N^{(t)})) = E(\theta_{t+k}|N^{(t)}). \tag{2.13}$$

Furthermore, using the state equation we have

$$E(\theta_{t+k}|N^{(t)}) = E(\theta_t|N^{(t)}) \prod_{n=t+1}^{t+k} \frac{E(\epsilon_n|N^{(t)})}{\gamma} = E(\theta_t|N^{(t)}) = \frac{\alpha_t}{\beta_t} \tag{2.14}$$

where $E(\epsilon_n|N^{(t)}) = \gamma$ for any n. Therefore, combining (2.15) and (2.14), we can write

$$E(N_{t+k}|N^{(t)}) = E(\theta_{t+k}|N^{(t)}) = \frac{\alpha_t}{\beta_t}. \tag{2.15}$$

Due to the random walk type of structure introduced in (2.3), the above result simply indicates that k-step ahead forecasts given that we have observed the call arrivals up to time t are equal to α_t/β_t.

2.1 Long run behaviour of the call arrival rate

Long run behaviour of the call arrival rate is of interest to call center practitioners. Under the proposed model we can analyze the long run behaviour of the call arrival rate θ_t via a study of the behaviours of α_t and β_t. In other words, if the call center system has been in operation for a long time, a steady state type behaviour might have been reached for the arrival rate. We note that the long run behaviour of α_t and β_t would determine how the mean and the variance of the arrival rate θ_t would change as functions of γ.

We define the steady-state value of β_t as $\beta = \lim_{t\to\infty} \beta_t$. By taking the limit on both sides of $\beta_t = \gamma\beta_{t-1} + 1$, we obtain

$$\beta = \frac{1}{1-\gamma} \tag{2.16}$$

For the long run behavior of α_t, using $\alpha_t = \gamma\alpha_{t-1} + N_t$ we can show that for a large k

$$\alpha_k = N_k + \gamma N_{k-1} + \cdots + \gamma^{k-1} N_1 + \gamma^k \alpha_0 \tag{2.17}$$

Therefore given (2.16) and (2.17), in the long run $E(\theta_t|N^{(t)}) = \frac{\alpha_t}{\beta_t} = (1-\gamma)\alpha_k$. That is

$$E(\theta_k|N^{(k)}) = (1-\gamma)N_k + (1-\gamma)\gamma N_{k-1} + \cdots + (1-\gamma)\gamma^{k-1}N_1 + (1-\gamma)\gamma^k\alpha_0. \tag{2.18}$$

We note from (2.18) that the posterior mean of θ_t is an exponentially weighted average of the past observed call arrivals up to time t. Similarly, from (2.12) $E(N_t|N^{t-1}) = \frac{\alpha_{t-1}}{\beta_{t-1}} = (1-\gamma)\alpha_{k-1}$.

$$E(N_k|N^{k-1}) = (1-\gamma)N_{k-1} + (1-\gamma)\gamma N_{k-2} + \cdots + (1-\gamma)\gamma^{k-1}N_1 + (1-\gamma)\gamma^k\alpha_0 \tag{2.19}$$

Therefore, we can also conclude that one step ahead predictive mean of the call center arrivals at time $k-1$ is a weighted average of the past arrivals.

It follows from the above that for $\gamma < 0.5$ both for $E(\theta_k|N^{(k)})$ and $E(N_k|N^{k-1})$, the model assigns a relatively larger weight on the most recent call arrival, whereas for $\gamma > 0.5$ the weight is relatively smaller. Following a similar logic, we can also obtain the variance of $\theta_t|N^{(t)}$ and $N_t|N^{(t-1)}$ in the long run as follows

$$V(\theta_k|N^{(k)}) = (1-\gamma)^2\alpha_k \tag{2.20}$$

and

$$V(N_k|N^{(k-1)}) = \frac{(1-\gamma)}{\gamma}\alpha_{k-1} \tag{2.21}$$

where α_k is defined as in (2.17) for a large k. A quick note about the initial values of α_0 and β_0 is that if the call center system has been in operation for a long time their affect on both the mean and the variance in the long run becomes negligible.

2.2 Bayesian learning about discount parameter γ

In previous sections, the discount factor γ has been assumed to be known in order for us to obtain a tractable Bayesian analysis. However, we can also treat γ as an unknown quantity and describe our uncertainty about it via a prior distribution, say $p(\gamma)$. Given, $N^{(t)}$, that is, call arrivals up to time t, the likelihood function of γ is given by

$$L(\gamma; N^{(t)}) = \prod_{k=1}^{t} p(N_k|N^{(k-1)}, \gamma), \tag{2.22}$$

where $p(N_k|N^{(k-1)}, \gamma)$ is given (2.12). The posterior distribution of γ can then be obtained via the Bayes' rule as

$$p(\gamma|N^{(t)}) \propto \prod_{k=1}^{t} p(N_k|N^{(k-1)}, \gamma)p(\gamma). \tag{2.23}$$

For any choice of prior $p(\gamma)$ in (2.23) the posterior distribution can not be obtained analytically. However, we can always sample from the posterior distribution of γ using a Markov chain Monte Carlo (MCMC) method such as the Metropolis-Hastings algorithm. Alternatively, a discrete prior can be used for γ over $(0, 1)$. For example, a discrete uniform prior between 0.01 and 0.99 can be a reasonable choice and this will be considered in our examples.

2.3 Filtering distribution and retrospective analysis

In the previous sections our focus has been on prediction of call arrivals in the future. For inference on mean arrival rates at different points in time one is more interested in a retrospective type of analysis. In other words, given that we have observed the data $N^{(t)}$ at time t, we will be interested in the distribution of $\theta_{t-k}|N^{(t)}$ for all $k \geq 1$.

We can write

$$p(\theta_{t-k}|N^{(t)}) = \int p(\theta_{t-k}|\theta_{t-k+1}, N^{(t)})p(\theta_{t-k+1}|N^{(t)})d\theta_{t-k+1} \tag{2.24}$$

where $p(\theta_{t-k}|\theta_{t-k+1}, N^{(t)})$ is obtained via the Bayes' rule as

$$\begin{aligned} p(\theta_{t-k}|\theta_{t-k+1}, N^{(t)}) &= \frac{p(\theta_{t-k}|\theta_{t-k+1}, N^{(t-k)})p(N^*|\theta_{t-k}, \theta_{t-k+1}, N^{(t-k)})}{p(N^*|\theta_{t-k+1}, N^{(t-k)})} \\ &= p(\theta_{t-k}|\theta_{t-k+1}, N^{(t-k)}) \end{aligned}$$

with $N^* = (N_{t-k+1}, \cdots, N_t)$. Here, given θ_{t-k+1}, N^* is independent of θ_{t-k}, i.e. $p(N^*|\theta_{t-k}, \theta_{t-k+1}, N^{(t-k)}) = p(N^*|\theta_{t-k+1}, N^{(t-k)})$. Thus, (2.24) reduces to

$$p(\theta_{t-k}|N^{(t)}) = \int p(\theta_{t-k}|\theta_{t-k+1}, N^{(t-k)})p(\theta_{t-k+1}|N^{(t)})d\theta_{t-k+1}. \tag{2.25}$$

We can not obtain (2.25) analytically, but it is possible to evaluate it using Monte Carlo methods. More specifically, we can draw samples from $p(\theta_{t-k}|N^{(t)})$. This requires us to develop an efficient algorithm which would lead us to sample from the joint density, i.e $p(\theta_1, \cdots, \theta_t|\gamma, N^{(t)})$, and then collect the samples corresponding to $p(\theta_{t-k}|\gamma, N^{(t)})$ for all $k \geq 1$.

Due to the Markovian nature of the state parameters, we can rewrite $p(\theta_1, \cdots, \theta_t|\gamma, N^{(t)})$ as

$$p(\theta_t|\gamma, N^{(t)})p(\theta_{t-1}|\theta_t, \gamma, N^{(t-1)}) \cdots p(\theta_1|\theta_2, \gamma, N^{(1)}). \tag{2.26}$$

We note that $p(\theta_t|\gamma, N^{(t)})$ is available from (2.9) and $p(\theta_{n-1}|\theta_n, \gamma, N^{(n-1)})$ for any n can be obtained as follows

$$p(\theta_{n-1}|\theta_n, \gamma, N^{(n-1)}) \propto p(\theta_n|\theta_{n-1}, \gamma, N^{(n-1)})p(\theta_{n-1}|\gamma, N^{(n-1)}) \tag{2.27}$$

where the first term is available from (2.3) and the second term from (2.7). It would be straightforward to show that $\theta_{n-1}|\theta_n, \gamma, N^{(n-1)} \sim Gamma((1-\gamma)\alpha_{n-1}, \beta_{n-1}+1)$ where $\gamma\theta_n < \theta_{n-1} < \infty$.

Therefore, given (2.26) and the posterior distribution of γ from 2.23, we can sample from $p(\theta_1, \cdots, \theta_t|\gamma, N^{(t)})$ sequentially simulating the individual state parameters as follows:

1. Generate $\gamma^{(i)}$ from $p(\gamma|N^{(t)})$.
2. Using the generated $\gamma^{(i)}$, sample $\theta_t^{(i)}$ from $\theta_t|\gamma, N^{(t)}$.
3. Using the generated $\gamma^{(i)}$, for each $n = t-1, \cdots, 1$ to generate $\theta_n^{(i)}$ from $\theta_n|\theta_{n+1}^{(i)}, \gamma, N^{(n)}$ where $\theta_{n+1}^{(i)}$ is the value generated in the previous step.

If we repeat the above large number of times, then we obtain samples from $p(\theta_1, \cdots, \theta_t|\gamma, N^{(t)})$ which allows us to obtain a density estimate for $p(\theta_{t-k}|\gamma, N^{(t)})$ for all $k \geq 1$. Further details about the filtering distributions are presented in our numerical example section.

3 A Model for Inter-week Call Arrivals

It is possible to extend the model discussed in the previous section to describe arrivals for each day of the week. In so doing, we consider each day of the week separately and assume that the behaviour of a given day is the same from one week to another. In other words we assume that call arrival process for Monday on any week will exhibit similar behaviour to any other Monday in another week. We consider such a behaviour by assuming the weekly call arrivals for any day are exchangeable over time, that is, from one week to another.

Such an extension is helpful in providing one week ahead forecasts for staffing decisions. These forecasts are of interest to call center managers who would like to be able to determine staff schedules in advance for different time intervals in a given day.

Let $N_{t,j,k}$ denote the number of call arrivals during the time interval $(t-1,t]$ on the k^{th} day of the j^{th} week. We will refer to the interval $(t-1,t]$ as the t^{th} within day interval. Also, let $\theta_{t,k}$ be the corresponding arrival rate. We assume that for each different day of the week the arrival rates follow the Markovian evolution

$$\theta_{t,k} = \frac{\theta_{t-1,k}}{\gamma_k}\epsilon_{t,k} \tag{3.1}$$

where $\epsilon_{t,k}|D^{(j)}_{t-1,k} \sim Beta(\gamma_k\alpha_{t-1,k},(1-\gamma_k)\alpha_{t-1,k})$ and $D^{(j)}_{t-1,k} = \{\mathbf{N_1^{(t-1,k)}},\mathbf{N_2^{(t-1,k)}},\cdots,\mathbf{N_j^{(t-1,k)}}\}$. In other words, $D^{(t,k)}_j$ denotes all call arrivals observed up to time period t for the last j weeks for day k and $\mathbf{N_j^{(t,k)}}$ represents the call arrivals observed up to time period t for the j^{th} week of day k. In the above, γ_k acts like a common discounting term for day k. Thus, there is a different Markov evolution for each day of the week with a different discounting term.

In what follows, we will suppress the index k for notational convenience and focus on a particular day of the week. Thus, we write (3.1) as $\theta_t = \frac{\theta_{t-1}}{\gamma}\epsilon_t$. Given the arrival rate θ_t the distribution for the number of call arrivals for the j^{th} week during the t^{th} interval for a given day is described via the discrete time Poisson model

$$p(N_{t,j}|\theta_t) = \frac{\theta_t^{N_{t,j}}e^{-\theta_t}}{N_{t,j}!} \tag{3.2}$$

where $t=1,\cdots,T$ and $j=1,\cdots,J$. Here T is the number if time intervals available in a given day and J is the number of weeks available in the whole data. We assume that given $\Theta^{(T)} = (\theta_1,\cdots,\theta_T)$, N_j^T's are independent of each other, with $N_j^T = (N_{1,j},\cdots,N_{T,j})$. Thus, $N_j^T; j=1,\cdots,J$ form an exchangeable sequence of random variables. Our objective is to obtain Bayesian updating for call arrival rates after observing J weeks each with T time intervals and to use this information to predict the call arrivals in each interval of $(J+1)th$ week.

To develop Bayesian inference for the model, we will process data in a sequential manner. Similar to our development in Section 2, prior to observing any data we assume that

$$(\theta_0|N^{(0)}) \sim Gamma(\alpha_0,\beta_0). \tag{3.3}$$

Given $D^{(j)}_{t-1}$, that is, given arrival counts up to time $t-1$ for the last j weeks, we can show by induction that

$$(\theta_{t-1}|D^{(j)}_{t-1}) \sim Gamma(\alpha_{t-1},\beta_{t-1}). \tag{3.4}$$

and by combining (3.1) and (3.4), we can obtain

$$(\theta_t|D_{t-1}^{(j)}) \sim Gamma(\gamma\alpha_{t-1}, \gamma\beta_{t-1}). \tag{3.5}$$

Furthermore, the posterior distribution $\theta^k|D_t^{(j)}$ can be obtained using the Bayes' Rule, that is, by combining the likelihood function (3.2) with the prior (3.5), as

$$(\theta_t|D_t^{(j)}) \sim Gamma(\alpha_t, \beta_t) \tag{3.6}$$

where $\alpha_t = \gamma\alpha_{t-1} + \sum_{i=1}^{j} N_{t,i}$ and $\beta_t = \gamma\beta_{t-1} + j$. In other words, the posterior distribution of the call arrival rate at time t given that we have observed the call arrivals for the last j weeks up to time period t is again a gamma density.

An attractive feature of the extended model is that we can now obtain the probability distribution of the call arrivals at time t for the j^{th} week given observed data from previous $j-1$ weeks. The posterior predictive distribution of $N_{t,j}|D_t^{(j-1)}$ will be

$$N_{t,j}|D_t^{(j-1)} \sim Negbin(\gamma\alpha_t, \gamma\beta_t). \tag{3.7}$$

The predictive distribution (3.7)provides a mechanism to make forecasts for the same day of the future weeks. For example, given call arrivals data for Monday from the previous $j-1$ weeks, we can make predictions for all the time periods for the next Monday. Similarly if we have the call arrivals data for all other days, we can provide predictions for the whole next week. This is a helpful feature for a week-ahead staffing decisions of call center managers. Also, the within day posterior predictive distribution of $N_{t,j}|D_{t-1}^{(j)}$ can be obtained as

$$N^{t,j}|D_{t-1}^{(j)} \sim Negbin(\gamma\alpha_{t-1}, \gamma\beta_{t-1}). \tag{3.8}$$

Similar to our original model proposed, we can assume that the discount factor γ is unknown and learn about it based on arrival counts data on different weeks of the given day. We can define our uncertainty about γ by a prior distribution $p(\gamma)$ over $(0, 1)$ and obtain the posterior distribution

$p(\gamma|D_{(T)}^{(J)})$ via

$$p(\gamma|D_{(T)}^{(J)}) \propto \prod_{j=1}^{J}\prod_{t=1}^{T} p(N_{t,j}|D_{(t-1)}^{(j)}, \gamma)p(\gamma) \tag{3.9}$$

where $p(N_{t,j}|D_{(t-1)}^{(j)}, \gamma)$ for any t is given by (3.8). As before we can either use MCMC methods for drawing posterior samples from the above distribution or alternatively can specify a discrete prior distribution for γ and evaluate posterior accordingly. Once the posterior distribution is available, using (3.7) we can obtain one week ahead predictions.

4 Numerical Examples

In this section we use actual data to show the implementation of the proposed models. The available data is on call center arrivals during different intervals of a day from a US commercial bank. Each day consists of 170 time intervals each of which is 5 minutes duration. In a given day, the call center is operational between 7 AM and 9:05 PM. A detailed summary of the data is available in Weinberg et al. (2007).

In our analysis, we initially focus on the first five days of the first week in the data set. In so doing, we assume a discrete uniform prior over $(0.01, 0.99)$ for the discounting term γ and study its behaviour via the posterior distribution. Using the discrete prior, the posterior distribution of γ given $N^{(t)}$ can be obtained via

$$p(\gamma = i|N^{(t)}) = \frac{p(N_1, \cdots, N_t|\gamma = i)p(\gamma = i)}{\sum_{j=0.01}^{0.99} p(N_1, \cdots, N_t|\gamma = j)p(\gamma = j)} \tag{4.1}$$

for $i = 0.01, \cdots, 0.99$.

We investigate the effect of the gamma prior parameters, α_0 and β_0 on the posterior distribution of γ. As shown in Figure 1, posterior distribution of γ is not very sensitive to the initial prior values as long as the prior mean, (α_0/β_0) is within the same order of magnitude as the actual call arrival data.

In Figure 2 we present the posterior distributions of γ for different days of the week. We note that the discount term γ behaves more or less the same for different days of the week. Table 1 shows the posterior means and the standard deviations of γ's for different days of the week. Again, we note that there are no large differences between the means and the standard deviations for

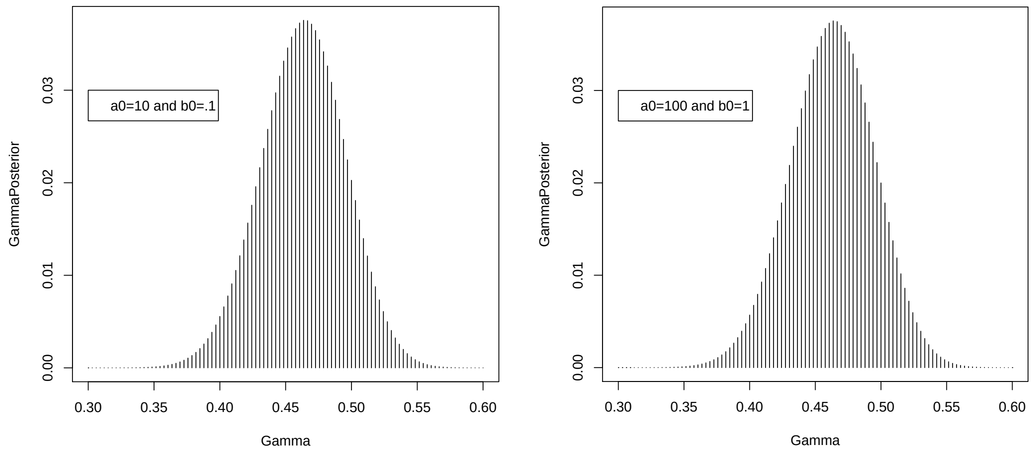

Figure 1: for $\alpha_0 = 10$, $\beta_0 = .1$ (left) and $\alpha_0 = 100$, $\beta_0 = 1$ (right)

different days.

Days	Means	Std
Monday	.4628	0.03200
Tuesday	.4730	0.03425
Wednesday	.5147	0.03477
Thursday	.4817	0.03478
Friday	.5026	0.03441

Table 1: Posterior means and standard deviations for different days of the week

One step ahead forecasts for call arrivals can be obtained using the conditional posterior predictive distribution given by (2.12) with the posterior distribution of γ given by (4.1). More specifically, to obtain $p(N_t|N^{(t-1)})$ for $t \geq 1$, we evaluate

$$p(N_t|N^{(t-1)}) = \sum_{i=0.01}^{0.99} p(N_t|N^{(t-1)}, \gamma = i)p(\gamma = i|N^{(t-1)}) \tag{4.2}$$

where $p(N_t|N^{(t-1)}, \gamma = i)$ is available from (2.12) and $p(\gamma = i|N^{(t-1)})$ is from (4.1). One step ahead posterior predictive distributions $p(N_{141}|N^{(140)})$ and $p(N_{142}|N^{(141)})$, using the first 140 and 141 observations, for Monday are given in Figure 3. Using (4.2) we have obtained one step ahead forecasts for different days of the week. In doing so, we have used the first 140 time periods of each day as historical data and sequentially obtained one step ahead forecasts. Based on the one step

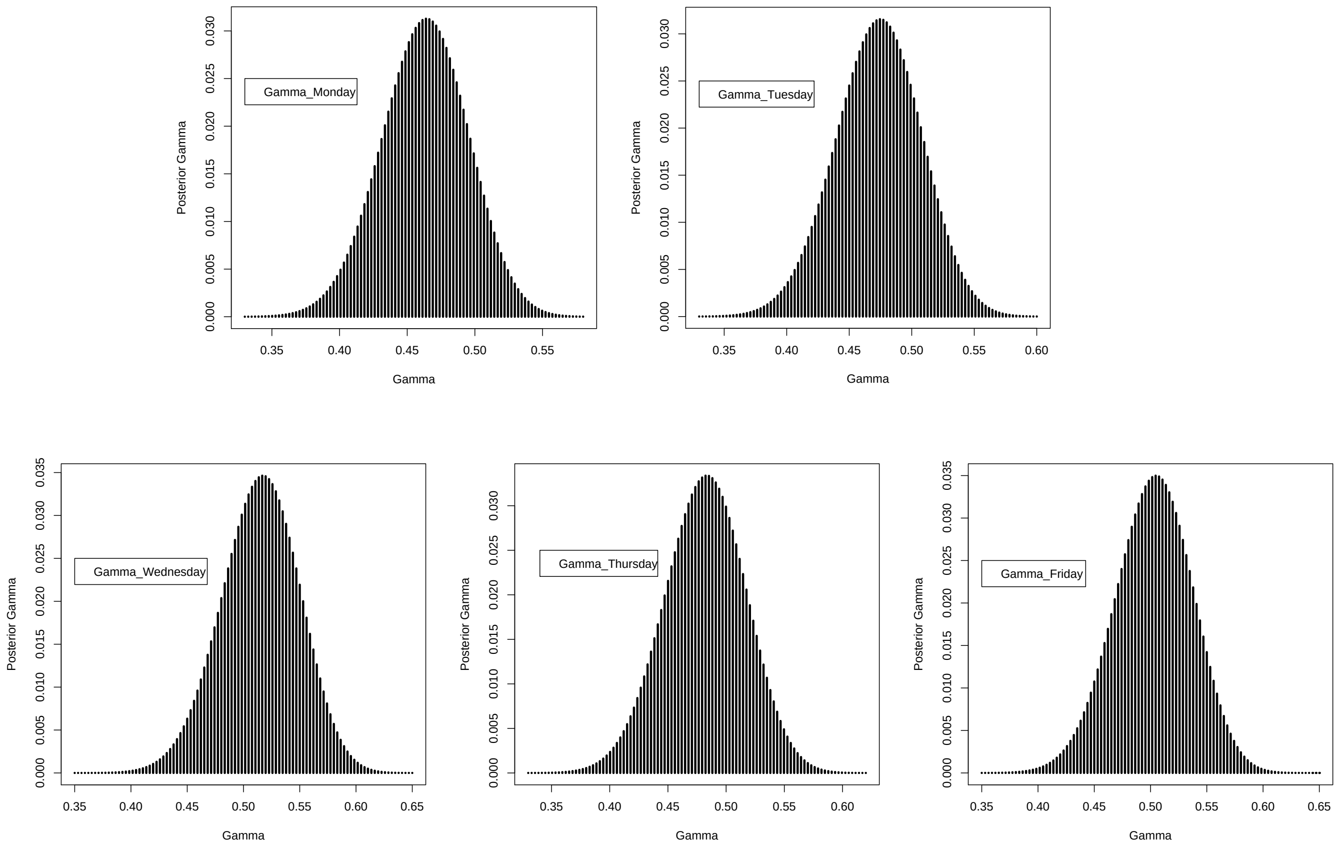

Figure 2: Posterior γ for different days of the week

ahead forecasts we have computed mean absolute percentage error for day d as

$$MAPE_d = \sum_{i=141}^{170} \frac{|N_i - \hat{N}|}{30N_i}, \tag{4.3}$$

where d =Monday, $\cdots$, Friday. We have also obtained mean absolute percentage errors for each day using the respective posterior means in 1 as the fixed values of γ. The results are identical as can be seen from the summary in Table 2.

One way to infer how good the proposed model fits to the actual call arrival data on a given day is by a comparison of the posterior means, $E(\theta_t|N^{(t)})$ with the actual values. Figure 4 shows the posterior means and the call arrivals on the same plot for the whole day on Monday and suggests that the model provides a reasonably good fit to the data.

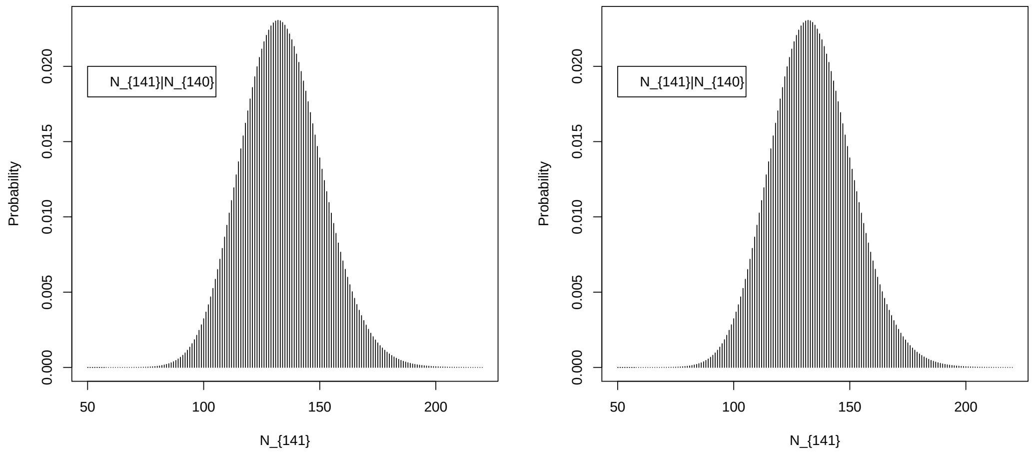

Figure 3: $p(N_{141}|N^{(140)})$ (left) and $p(N_{142}|N^{(141)})$ (right)

Days	MAPE (γ unknown)	MAPE (γ fixed)
Monday	.081	.081
Tuesday	.093	.094
Wednesday	.137	.137
Thursday	.117	.117
Friday	.098	.099

Table 2: One step ahead MAPE for different days

A comparison of mean arrival arrival rates during different periods of each day can be made retrospectively by looking at the filtered state means $E(\theta_t|N^{(170)})$'s based on all observed data in a given day. As discussed in section 2.3, the forward filtering backward sampling algorithm was implemented with 1,000 iterations for each day. Figure 5 shows the behaviour of the filtered state means for different days of the week. We note that, retrospectively, arrival rates on Mondays are mostly higher than those on the other days. It may be of interest to look at the arrival patterns during different time periods in the day. As shown in Figure 6, between 7:00 AM and 12:00 noon, call arrival rates are higher on Monday, but we also note that Friday arrivals are almost as high as Monday in the morning whereas the lowest rate is observed on Thursday. Similar insights can be obtained from Figure 7 which illustrates the arrival rate patterns between 2:30 and 9:05 PM.

Our analysis so far has focused on comparison of mean arrival rates during different days and intra-day call arrival predictions. As noted before, short term predictions are important for dynamic staffing. An equally important issue is the ability to predict the arrival pattern in future weeks for

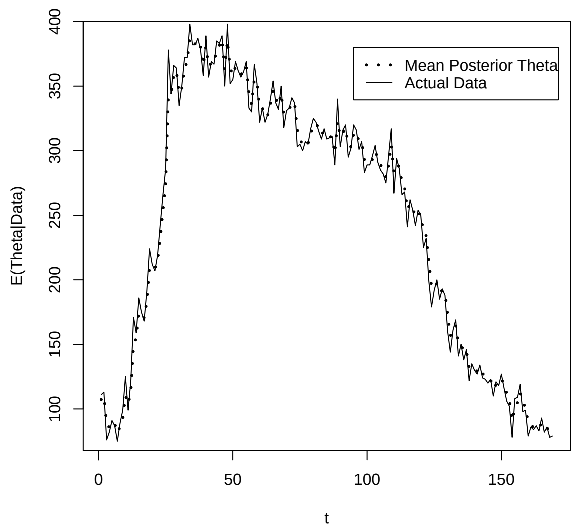

Figure 4: Posterior means and the actual arrivals at different time intervals during Monday.

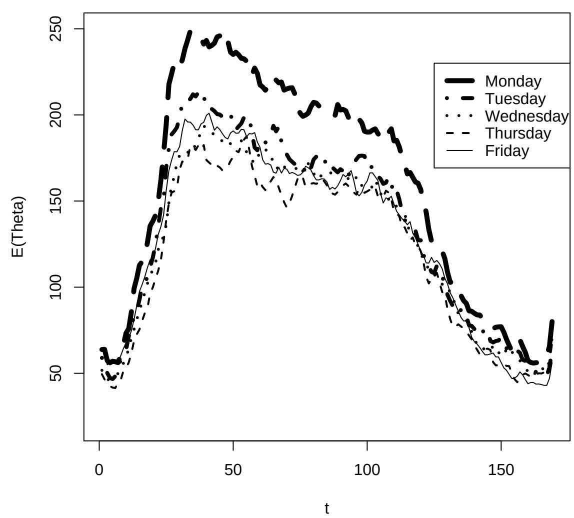

Figure 5: Posterior means $E(\theta_t|N^{(170)})$ versus time t for different days.

long term staffing decisions. As discussed in Section 3, our model enables us to obtain inter-weekly forecasts as well as the intra-day ones. Using our development in Section 3 and more specifically using (3.7) we can obtain one week ahead predictions for each day of the week. For illustrative purposes we have used the first three weeks of call arrivals in order to predict the arrivals during

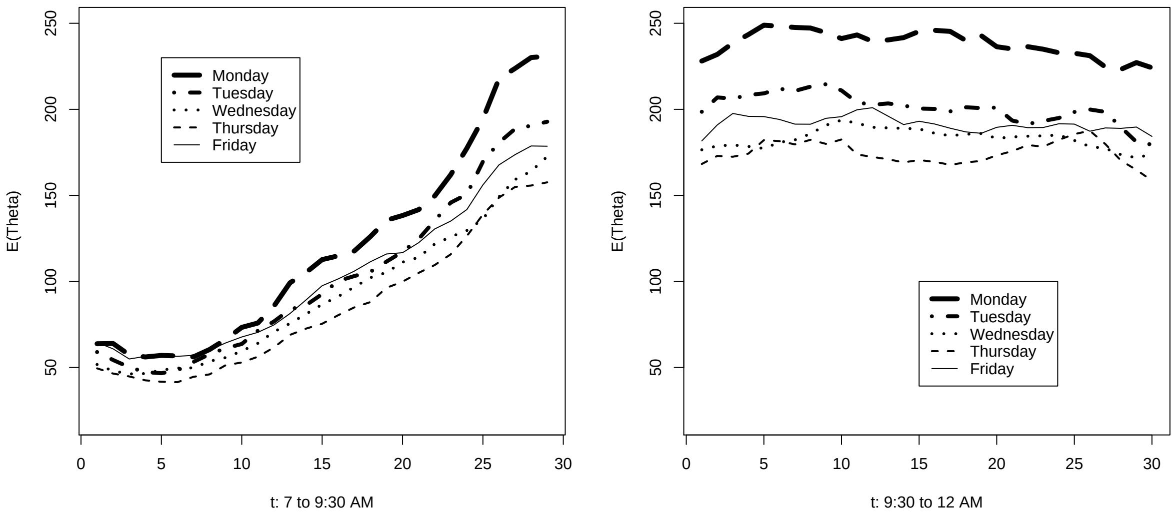

Figure 6: Mean arrival rates during 7:00 to 9:30 AM(left) and 9:30 AM to 12: noon (right)

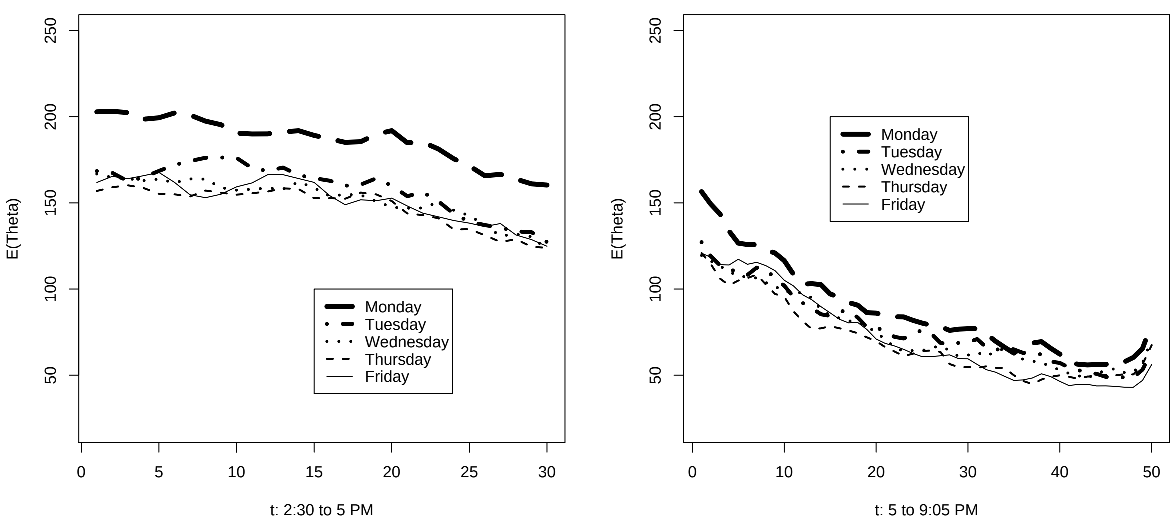

Figure 7: Mean arrival rates during 2:30 to 5:00 PM(left) and 5:00 to 9:05 PM (right)

all time periods of each day of the following week. In Figure 8 we present a comparison of one week ahead predictions for Monday and Tuesday arrivals with the actual call arrivals during the whole 170 time periods. In table 3 we present the mean absolute percentile errors (MAPEs) for each day of the week based on 170 time periods. We note that the daily error rate changes between 7.5 and

8.5 per cent which is comparable our intra-day predictions given in Table 2.

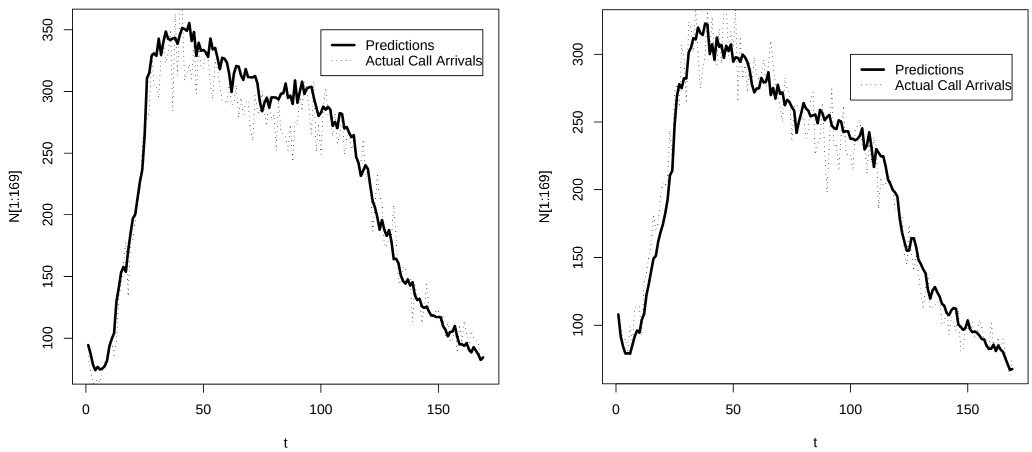

Figure 8: One week ahead forecasts for Monday (left) and Tuesday (right) versus actual call arrivals

Days	MAPE
Monday	0.0756
Tuesday	0.0759
Wednesday	0.074
Thursday	0.080
Friday	0.0855

Table 3: One week ahead MAPEs

5 Concluding Remarks

In this article we have developed a Bayesian state space model whose measurement equation follows a discrete time Poisson process and whose rates follow a discrete time gamma process with beta distributed error terms. We provided closed form updating procedures for the model parameters under certain conditions, developed one step ahead predictions and introduced a forward filtering backward sampling algorithm in order to estimate filtered arrival rates for different days of the week for inference purposes. Our model also allows us to obtain analytically tractable inter-week forecasts based on the previously observed call arrivals. We applied the proposed model to data from an anonymous U.S. commercial bank. Based on our comparison of the expected posterior means of the arrival rates and the actual arrival counts for each period of a given day we have shown that the proposed model provided a good fit to the data at hand. Furthermore, the model provided us with accurate one-week ahead forecasts.

We believe that modeling the call center arrivals using a Bayasian perspective provides additional insights for call center practitioners. For instance, instead of obtaining just one step ahead point estimates, now one can talk about the distributions of the estimates, the state parameters and their respective distributional properties such as the variance, mode and credibility intervals. Furthermore, the proposed model incorporates managerial insights from call center practitioners via the prior distribution parameters. From a practical point of view, the fact that state parameter updating and one step ahead call forecasting is analytically tractable makes the proposed model attractive. From a staffing point of view, parameter updating efficiency and adaptability of the proposed model are other features that might be of interest to call center managers.

Our findings led us to believe that further extensions to the proposed model is possible if the discount term γ were to be treated as a common parameter for different days of the week which would introduce a dependence structure between the call arrivals of different days. Therefore, if one can develop a mechanism which can efficiently update the prior parameters of each day based on past data, the proposed model can be further extended. However we expect that this type of approach would lead us to lose the analytical tractability of our model and is currently being investigated as an extension to the problem at hand.

References

Aksin, Z., Armony, M., and Mehrotra, V. (2007). The modern call center: A multi-disciplinary perspective on operations management research. *Production and Operations Management*, 16(6):665–688.

Avramidis, A. N., Deslauriers, A., and L'Ecuyer, P. (2004). Modeling daily arrivals to a telephone call center. *Management Science*, 50(7):896–908.

Gans, N., Koole, G., and Mandelbaum, A. (2003). Telephone call centers: Tutorial, review and research prospects. *Manufacturing and Service Operations Management*, 5(2):79–141.

Jongbloed, G. and Koole, G. (2001). Managing uncertainty in call centers using poisson mixtures. *Applied Stochastic Models in Business and Industry*, 17:307–318.

Morali, N. and Soyer, R. (2003). Optimal stopping in software reliability. *Naval Research Logistics*, 50.

Smith, R. and Miller, J. (1986). A non-gaussian state space model and application to prediction of records. *Journal of the Royal Statistical Society. Series B*, 48(1):79–88.

Soyer, R. and Tarimcilar, M. (2008). Modeling and analysis of call center arrival data: A bayesian approach. *Management Science*, 54(2):226–278.

Weinberg, J., D.Brown, L., and Stroud, J. R. (2007). Bayesian forecasting of an inhomogeneous poisson process with applications to call center data. *Journal of the American Statistical Association*, 102(102).

Whitt, W. (1999). Dynamic staffing in a telephone call center aiming to immediately answer all calls. *Operations Research Letters*, 24:205–212.