FORECASTING CALL CENTER ARRIVALS: FIXED-EFFECTS, MIXED-EFFECTS, AND BIVARIATE MODELS

Rouba Ibrahim
Desautels Faculty of Management
McGill University

Pierre L'Ecuyer
Department of Operations Research
University of Montreal

Abstract

We consider different statistical models for the call arrival process in telephone call centers. We evaluate the forecasting accuracy of those models by describing results from an empirical study analyzing real-life call center data. We test forecasting accuracy using different lead times, ranging from weeks to hours in advance, to mimic real-life challenges faced by call center managers. The models considered are: (i) a benchmark fixed-effects model which does not exploit any dependence structures in the data; (i) a mixed-effects model which takes into account both interday (day-to-day) and intraday (within day) correlations; (ii) two new bivariate mixed-effects models, for the joint distribution of the arrival counts to two separate queues, which exploit correlations between different call types. Our study shows the importance of accounting for different correlation structures in the data.

Keywords: forecasting; arrival process; dynamic updating; time series; correlation; call centers.

1. Introduction

Unlike tangible products, services are experienced and not consumed. To increase customer satisfaction, service systems compete in improving the quality of service provided, while maintaining high levels of operational efficiency. As a result, service system managers often need to weigh contradictory objectives. In the context of call centers, quality of service is typically measured by customer delay in the system (i.e., the amount of time that callers spend waiting on hold before being handled by an agent), whereas operational efficiency is measured by the proportion of time that agents are busy handling calls. The quality of service in a call center is usually regulated by a service-level agreement (SLA) which need be respected. The SLA specifies target performance levels, such as the wait-time level or the proportion of abandoning customers. For background on call centers, see Gans et al. (2003) and Aksin et al. (2007).

In order to achieve the right balance between quality of service and operational efficiency, call center managers are faced with multiple challenges. First, there is the problem of determining appropriate staffing levels, weeks or even months in advance, based on long-term forecasts of future incoming demand which is typically both time-varying and stochastic. In the words of Aksin et al. (2007), that is a problem of "resource acquisition". Second, there is the problem of scheduling (and re-scheduling) the available pool of agents based on updated forecasts, typically made several days or weeks in advance. That is a problem of "resource deployment"; see Avramidis et al. (2010). Finally, there are short-term decisions that need be made, such as routing incoming calls in real-time to available agents, or mobilizing agents on short notice due to unforseen fluctuations in incoming demand. Those decisions are based on short-term forecasts, updated one day or a few hours in advance. As an initial step, pending the analysis of effective scheduling and routing designs, it is crucial to develop accurate forecasts of future call volumes, and to study ways of updating those forecasts at different points in time.

1.1. Main Contributions

In this paper, we consider different statistical models for the call arrival process. Specifically, we consider Gaussian linear mixed-effects models for the square-root transformed call arrival counts. For background on linear mixed-effects models, see Muller and Stewart (2006). We conduct an empirical study using real-life call center data, and generate point and confidence interval forecasts of future arrival counts. We test the accuracy of our forecasts using lead times ranging from weeks to hours in advance: We do so to mimic real-life challenges faced by call center managers. Our study shows the importance of accounting for different correlation structures in the data. For example, our mixed-effects models take into account both interday (day-to-day) and intraday (within-day) correlations in the time series of arrival counts. This paper was motivated by an industry research project with a major telecommunications company in Canada; see §3.

The main novelty of this work lies in jointly modeling the arrival counts for different call types handled at the call center. In particular, we use a bivariate linear mixed-effects model for the joint distribution of arrival counts to two separate queues. We exploit inter-(call)type correlations and show that bivariate mixed-effects models can lead to more accurate forecasts than univariate mixed-effects models. Bivariate mixed-effects models are traditionally used in the field of biostatistics, e.g., when analyzing longitudinal data of two associated markers; see Barry and Bowman (2007), and Thiébaut et al. (2007). To the best of our knowledge, ours is the first work that proposes using those models in the context of call center applications.

1.2. Fixed-Effects, Mixed-Effects, and Bivariate Models

We now briefly describe the statistical models considered in this paper; for details, see §4. We first consider a simple fixed-effects (FE) model with day-of-week and period-of-day covariates, and independent residuals. This model also includes cross terms to capture the interaction between the day-of-week and period-of-day effects. The FE model is equivalent to a historical average approach since it essentially uses past averages as forecasts of future call volumes; it was used as a benchmark model in both Weinberg et al. (2007) and Shen and Huang (2008b). The FE model serves as a useful reference point because it does not incorporate any dependence structures in the data. Since there is strong evidence for correlations in the time-series of arrival counts (e.g., see §3.2), we anticipate that forecasts based on the FE model will not be very accurate. In §6 and §7, we show that the FE model is useful with relatively long forecasting lead times, but not otherwise.

In §4.2, we extend the FE model and consider a mixed-effects (ME) model incorporating both fixed and random effects. Random effects, which are Gaussian deviates with a specific covariance structure, are used to model interday correlations. Intraday correlations are modeled by imposing a specific covariance structure on the residuals of the model. The resulting ME model also includes the FE model's fixed effects. Consistent with Aldor-Noiman et al. (2009), we show that the ME model generally leads to accurate point and interval forecasts of future call volumes; see §6 and §7. In particular, it is superior to the FE model with relatively short forecasting lead times.

In real-life call centers, there is typically evidence for correlations between the arrival counts of different call types. Intertype correlations may arise in multi-lingual call centers where certain service requests are handled in different languages. More generally, intertype correlations may arise when arrivals to different queues are driven by the same underlying causes, e.g., specific marketing campaigns or special events. In §5, we extend the ME model to bivariate mixed-effects (BM) models which exploit inter-type correlations. A BM model jointly models arrivals to two separate queues; it consists of two dependent ME models, each modeling arrivals to one of the two queues.

We propose two ways of modeling correlations across call types. In our first bivariate model, BM1, we specify a correlation structure between the random effects of the two underlying ME models; see §5.1. In our second bivariate model, BM2, we specify a correlation structure between the residuals of the two underlying ME models; see §5.2. The resulting BM models exploit interday, intraday, and inter-type correlations. In §6 and §7, we show that BM models generally yield more accurate point and and interval forecasts than both FE and univariate ME models.

1.3. Organization

The remainder of this paper is organized as follows. In §2, we review some of the relevant literature. In §3, we describe the data set that motivated this research. In §4, we describe the fixed-effects and univariate mixed-effects models. In §5, we describe the bivariate models. In §6, we compare the forecasting accuracy of our different models using the data set described in §3. In §7, we provide further empirical evidence by considering two additional data sets. In §8, we draw conclusions.

2. Literature Review

We now review some of the existing literature on forecasting call center arrivals. Much of the earlier work focuses on applying standard time series methods, such as Autoregressive Integrated Moving Average (ARIMA) models. For example, Andrews and Cunningham (1995) used the ARIMA/transfer function methodology to forecast arrivals to L. L. Bean's call center, and emphasized the impact of holidays and marketing campaigns on the arrival process. Bianchi et al. (1998) also used ARIMA models and found that they outperform simple Holt-Winters smoothing.

More recent work includes Weinberg et al. (2007) who used a Bayesian approach to forecast incoming calls at a United States bank's call center. They used the same square-root data transformation that we use in this paper, and exploited the resulting normality of data in their model. Taylor (2008) compared the forecasting accuracy of alternative time series models, including a version of Holt-Winters smoothing which accommodates multiple seasonal patterns. He showed that, with long forecasting lead times, simple forecasting techniques such as taking historical averages are difficult to beat. We reach a similar conclusion in this work as well. Shen and Huang (2008a, b) used a Singular Value Decomposition (SVD) approach to to create a prediction model which allows for interday forecasting and intraday updating of arrival rates. Aldor-Noiman et al. (2009) proposed an arrival count model which is based on a Gaussian linear mixed-effects model day-of-week, periodic, and exogenous effects. We use a similar mixed-effects model in this paper as well.

Other empirical studies have shown several important features of the call arrival process. Avramidis et al. (2004) proposed several stochastic models including a doubly stochastic Poisson arrival process with a random arrival rate. Their models reproduce essential characteristics of call center arrivals, such as: (i) a variance considerably higher than with Poisson arrivals, as observed by Jongbloed and Koole (2001), and (ii) strong correlations between the arrivals in successive periods of the same day, as in Tanir and Booth (2001). Then, they tested the goodness of fit of their models via an empirical study of real-life data. We also model intraday correlations in this paper. Additionally, we account for interday correlations. Interday correlations were shown to be significant in the seminal paper by Brown et al. (2005). One last feature of the call arrival process, which we also take into account here, is the time-variability of arrival rates. Indeed, there is strong empirical evidence suggesting that arrival rates in call centers are usually not stationary; e.g., see Gans et al. (2003), Brown et al. (2005) and Aksin et al. (2007).

3. Preliminary Data Analysis

The present data were gathered at the call center of a major telecommunications company in Canada. They were collected over 329 days (excluding days when the call center is closed, such as holidays and Sundays) ranging from October 19, 2009 to November 11, 2010. The data consist of arrival counts for two call types, Type A and Type B, whose incoming calls originate in the Canadian provinces of Ontario and Quebec, respectively. In §7, we briefly describe two additional data sets, each consisting of arrival counts for two call types as well.

The call center operates from 8:00 AM to 7:00 PM on weekdays (Monday to Friday), and from 8:00 AM to 6:00 PM on Saturdays. Because the call arrival pattern is very different between weekdays and Saturdays, we focus solely on weekdays in this paper. We thus remove a total of 47 Saturdays from the data set. There are "special" days in the data, such as days with missing values or irregular days (i.e., days on which arrival volumes are unusually high or low). In particular, there is a total of 15 special days (including 9 outlier days) which we remove from the set. This leaves us with $D = 329 - 47 - 15 = 267$ remaining days. Arrival counts for each day are aggregated in consecutive time periods of length thirty minutes each. There are $P = 22$ consecutive thirty-minute periods on a weekday, and a total of $D \times P = 267 \times 22 = 5874$ observations in our data set.

3.1. Overview

In Figures 1 and 2, we plot the average number of arrivals per half-hour period (for each weekday) for Type A and Type B, respectively. Figures 1 and 2 show that Type A has higher average arrival counts than Type B. Moreover, the daily profiles for the two call types are different. Figure 1 shows that half-hourly averages for Type A do not fluctuate substantially between the hours of 11:00 AM and 5:00 PM, for each weekday. In contrast, Figure 2 shows that there are two major daily peaks for Type B arrivals. The first peak occurs in the morning, shortly before 11:00 AM, and the second peak occurs in the early afternoon, around 1:30 PM. (There is also a third "peak", smaller in magnitude, which occurs shortly before 4:00 PM on Mondays, Tuesdays, and Wednesdays.) Such intraday arrival patterns are commonly observed in call centers; e.g., see Gans et al. (2003).

3.2. Correlation Structures

Exploratory analysis of our data shows evidence of: (i) strong positive interday correlations between the arrival counts over successive days; (ii) strong positive intraday correlations between the arrival counts over successive half-hour periods of the same day; and (iii) strong positive intertype

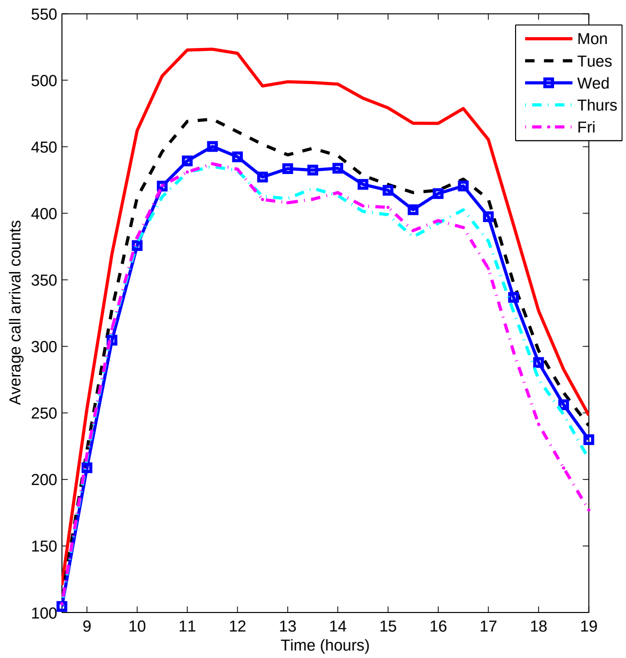

Figure 1: Average arrival counts per half-hour period for Type A.

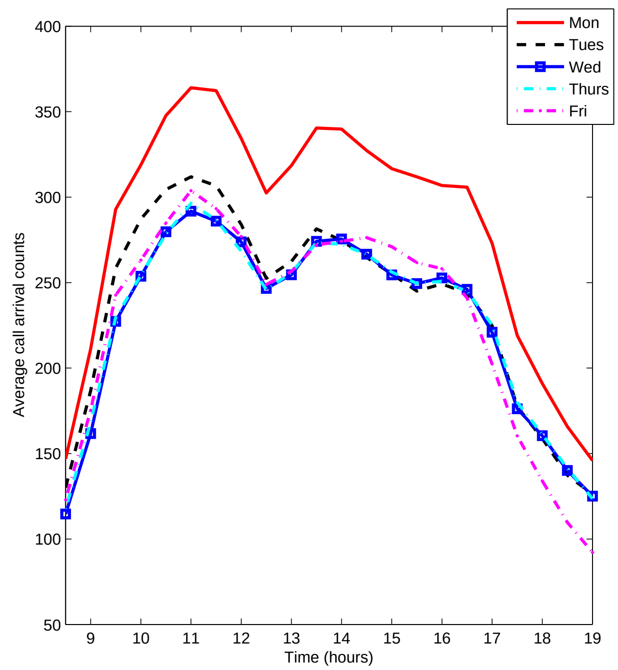

Figure 2: Average arrival counts per half-hour period for Type B.

correlations between the arrival counts of different call types. In Tables 1-3, we present point estimates for the values of those correlations in our data set. The p-values for testing the hypothesis of no correlation at the 0.95 confidence level are all substantially smaller than 0.05, and are therefore not included in the corresponding tables.

3.2.1. Interday Correlations

In Table 1, we present estimates of correlations between daily arrival counts over successive weekdays for Type B. (We first subtract from each daily total the average arrival volume for the corresponding weekday.) Table 1 shows that there are significant positive correlations between days of the same week. In particular, correlations are strong between successive weekdays, and are slightly smaller with longer lags; e.g., the correlation between (the total call volume on) Tuesday and (the total call volume on) Wednesday is 0.68, whereas the correlation between Tuesday and Friday is 0.62. Additionally, Table 1 shows that Mondays are less correlated with the remaining weekdays; e.g., the correlation between Monday and Tuesday is 0.48. Results for Type A calls are largely similar, and are therefore not reported separately.

3.2.2. Intraday Correlations

There are strong intraday correlations in the data set. In Table 2, we present estimates of correlations between half-hourly arrival counts for Type A on a given day. In particular, we show correlation estimates for the five consecutive half-hourly periods between 10:00 AM and 12:30 PM on Wednesdays. (We first subtract from each half-hourly count the average arrival count for the corresponding period.) Table 2 shows that correlations between counts on successive half-hour periods are uniformly strong and positive. Moreover, correlations are slightly smaller with longer lags, as expected.

3.2.3. Intertype Correlations

In Table 3, we present estimates of correlations between half-hourly arrival counts for Type A and Type B. We focus on the same consecutive half-hour periods as in Table 2. (As before, we first

Weekday	Mon	Tues.	Wed.	Thurs.	Fri.
Mon.	1.0	0.48	0.35	0.35	0.34
Tues.		1.0	0.68	0.62	0.62
Wed.			1.0	0.72	0.67
Thurs.				1.0	0.80
Fri.					1.0

Table 1: Correlations between arrival counts on successive weekdays for Type B.

Half-hour periods	(10, 10:30)	(10:30, 11)	(11, 11:30)	(11:30, 12)	(12, 12:30)
(10, 10:30)	1.0	0.87	0.80	0.73	0.66
(10:30, 11)		1.0	0.82	0.74	0.71
(11, 11:30)			1.0	0.83	0.80
(11:30, 12)				1.0	0.81
(12, 12:30)					1.0

Table 2: Correlations between arrivals in consecutive half-hour periods on Wednesday morning for Type A.

subtract from each half-hourly count the average arrival count for the corresponding period.) Table 3 shows that intertype correlations are uniformly strong and positive. Consistent with intuition, intertype correlations are slightly smaller for longer lags. Intertype correlations are relatively easy to interpret in this data set. Indeed, call arrivals to the Type A queue originate in the province of Ontario, and are mainly handled in English, whereas arrivals to the Type B queue originate in the province of Quebec, and are mainly handled in French. Otherwise, arrivals to both queues have similar service requests. Thus, we anticipate that there be correlations between their respective arrival processes. In §5, we propose two bivariate mixed-effects models which exploit such intertype correlations.

Type B \ Type A	(10, 10:30)	(10:30, 11)	(11, 11:30)	(11:30, 12)	(12, 12:30)
(10, 10:30)	0.75	0.72	0.67	0.60	0.59
(10:30, 11)	0.76	0.73	0.72	0.64	0.62
(11, 11:30)	0.66	0.65	0.67	0.67	0.63
(11:30, 12)	0.60	0.56	0.63	0.63	0.63
(12, 12:30)	0.58	0.54	0.58	0.65	0.62

Table 3: Correlations between Type A and Type B arrivals in consecutive half-hour periods on Wednesday.

3.3. Data Transformation

Let $N_{i,j}$ be the number of arrivals in the j^{th} period of day i, where $i = 1, 2, ..., D$ and $j = 1, 2, ..., P$. As in Whitt (1999) and Avramidis et al. (2004), we model the arrival process as a doubly stochastic Poisson process with a random arrival rate $\Lambda_{i,j}$. In particular, conditional on $\Lambda_{i,j} = \lambda_{i,j}$ where $\lambda_{i,j} > 0$ is a deterministic value, we assume that $N_{i,j}$ follows a Poisson distribution with arrival rate $\lambda_{i,j}$. As in Jongbloed and Koole (2001), our data possesses overdispersion relative to the Poisson distribution, e.g., the variance of the arrival counts is roughly equal to ten times the mean. To stabilize the variance, we use the "root-unroot" method which is commonly used in the literature; e.g, see Brown et al. (2005). In particular, letting $y_{i,j} = \sqrt{N_{i,j} + 1/4}$, it was shown in Brown et al. (2001) that for large values of $\lambda_{i,j}$, $y_{i,j}$ is approximately normally distributed, conditional on $\lambda_{i,j}$, with a mean value of $\sqrt{\lambda_{i,j}}$ and a variance equal to 1/4. Since there are hundreds of calls per period on average in a given weekday, it is reasonable to assume that our square-root transformed counts are roughly normally distributed with the above mean and variance. In §4.2, we exploit normality to fit Gaussian linear mixed-effects models to the transformed data. In Figures 3 and 4, we present the qqplots for the residuals of the ME and BM1 models, respectively, for Type B calls. (We also include in the plots the corresponding envelope at the 95% confidence level.) We use a forecasting lead time of half a day, and make predictions from from August 19, 2010 to November 11, 2010. Figure 4 shows that the normal distribution is a slightly better fit (at both the upper

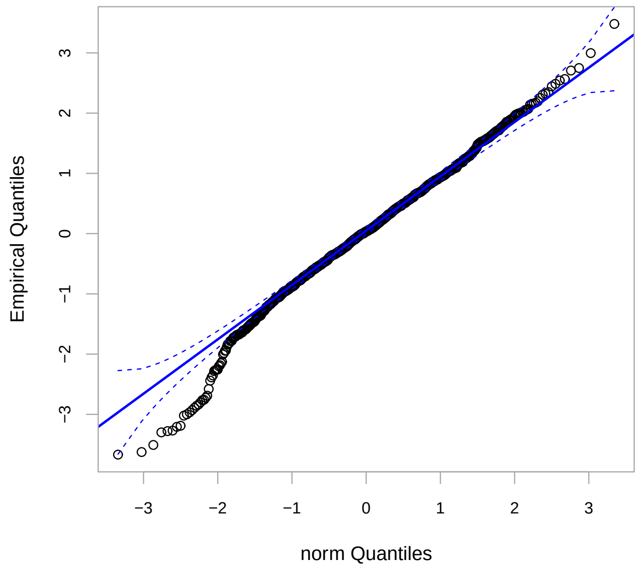

Figure 3: QQ plot for Type B residuals of the ME model with a forecasting lead time of half a day.

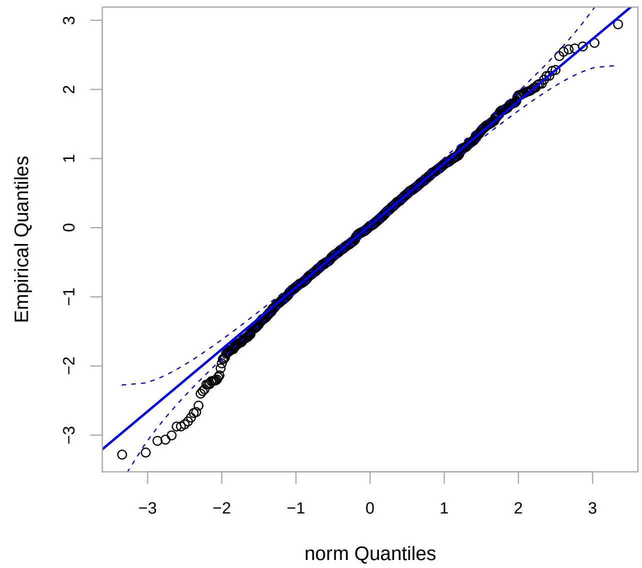

Figure 4: QQ plot for Type B residuals of the BM1 model with a forecasting lead time of half a day.

and lower tails) for BM1 residuals than for ME residuals.

4. Fixed-Effects and Mixed-Effects Models

In this section, we describe the fixed-effects and mixed-effects models for the call arrival counts. In §6 and §7, we compare the alternative models based on forecasting performance.

4.1. Fixed-Effects (FE) Model with Independent Residuals

The preliminary data analysis of §3 shows that the five weekdays have different expected daily total call volumes. Moreover, the expected number of calls per period for each weekday varies depending on the period; see Figures 1 and 2. We capture those two properties in our first model which is a simple linear additive model incorporating both day-of-week and period-of-day covariates. This model also includes cross terms to capture the interaction between the day-of-week and period-of-day effects. The additional cross terms allow for a different intraday profile for each weekday.

We consider the FE model because similar models are often used for forecasting future demand in real-life call centers. As pointed out in §1.2, the FE model is equivalent to a historical average approach since it essentially uses past averages as forecasts of future call volumes. It is a useful reference point because it does not incorporate any correlation structures in the data.

Let d_i be the day-of-week of day i, where $i = 1, 2, ..., D$. That is, $d_i \in \{1, 2, 3, 4, 5\}$ where $d_i = 1$ denotes a Monday, $d_i = 2$ denotes a Tuesday, ..., and $d_i = 5$ denotes a Friday. Let j denote the half-hour period index in day i, where $j = 1, 2, ..., P$. We model $y_{i,j}$, the square-root transformed call volume in period j of day i, as:

$$y_{i,j} = \sum_{k=1}^{5} \alpha_k I_{d_i}^k + \sum_{l=1}^{22} \beta_l I_j^l + \sum_{k=1}^{5} \sum_{l=1}^{22} \theta_{k,l} I_{d_i}^k I_j^l + \mu_{i,j} \ , \tag{4.1}$$

where $I_{d_i}^k$ and I_j^l are the indicators for day d_i and period j, respectively. That is, $I_{d_i}^k$ (I_j^l) equals 1 if $d_i = k$ ($j = l$), and 0 otherwise. The products $I_{d_i}^k I_j^l$ are indicators for the cross terms between the day-of-week and period-of-day effects. The coefficients α_k, β_l, and $\theta_{k,l}$ are real-valued constants that need be estimated from data, and $\mu_{i,j}$ are independent and identically distributed (i.i.d.) normal random variables with mean 0. The normality assumption enables us to obtain prediction intervals for future observations; see §6 and §7. Equation (4.2) simplifies to

$$y_{i,j} = \alpha_{d_i} + \beta_j + \theta_{d_i,j} + \mu_{i,j} \ . \tag{4.2}$$

We estimate model parameters using the method of least squares. Least squares estimates are equivalent to maximum likelihood estimates with normal i.i.d. residuals, as in (4.2).

4.2. Gaussian Linear Mixed-Effects (ME) Model

As discussed in §3.2, there is evidence of strong correlations in the data at both the interday and intraday levels. In this subsection, we extend the FE model of §4.1 and consider an ME model incorporating both fixed and random effects. We consider the same fixed effects as in (4.1). Random effects, which are Gaussian deviates with a pre-specified covariance structure, are used to model the interday correlations. Intraday correlations are modeled by imposing a specific covariance structure on the residuals of the model. We fit the ME model to data by computing maximum likelihood estimates of all model parameters. Mixed-effects models have been previously considered to model call center arrivals. In particular, the model that we describe in this subsection has been proposed by Aldor-Noiman et al. (2009).

4.2.1. Random Effects

Let γ_i denote the daily volume deviation from the fixed weekday effect on day i, where $i = 1, 2, ..., D$. Then, γ_i is the random effect on day i. Let G denote the $D \times D$ covariance matrix for the sequence of random effects. The random effects, γ_i, are identically normally distributed with expected value $E[\gamma] = 0$ and variance $\text{Var}[\gamma] = \sigma_G^2$. (We omit the subscript from a random variable when the specific index is not important.) We assume that G has a first-order autoregressive covariance structure, AR(1). That is,

$$\gamma_i = \rho_G \gamma_{i-1} + \psi_i \ , \tag{4.3}$$

where ψ_i are i.i.d. normally distributed random variables with $E[\psi] = 0$ and $\text{Var}[\psi] = \sigma_G^2(1 - \rho_G^2)$. The covariance between γ_i and γ_j is given by

$$\text{cov}(\gamma_i, \gamma_j) = g_{i,j} = \sigma_G^2 \rho_G^{|i-j|} \quad \text{for} \quad 1 \leq i, j \leq D \ , \tag{4.4}$$

where ρ_G is the autocorrelation parameter.

Considering an AR(1) covariance structure for G is both useful and computationally effective, because it requires the estimation of only two parameters, σ_G and ρ_G. We preserve the true numerical distance between days by fitting the power transformation covariance structure to G, using the actual duration between days; e.g., the lag between Monday and Tuesday of the same week is equal to 1, whereas the lag between Friday and the following Monday is equal to 3.

4.2.2. Model Residuals

Let $\epsilon_{i,j}$ denote the residual effect on period j of day i, where $i = 1, 2, .., D$ and $j = 1, 2, ..., P$. That is, $\epsilon_{i,j}$ is the normally distributed periodic deviation from the sum of fixed and random effects. Let

R denote the within-day $P \times P$ covariance matrix of the residual effects. We assume that R has an AR(1) covariance structure with variance parameter σ_R^2 and autocorrelation parameter ρ_R. Thus, paralleling (4.3), we have that

$$\epsilon_{i,j} = \rho_R \epsilon_{i,j-1} + \tau_{i,j} \ , \tag{4.5}$$

where $\tau_{i,j}$ are i.i.d. normally distributed random variables with $E[\tau] = 0$ and $\mathrm{Var}[\tau] = \sigma_R^2(1 - \rho_R^2)$. We also assume that residual effects are independent across different days.

4.2.3. Mixed Model Formulation

We assume that γ and ϵ are independent. In our ME model, we also include the fixed effects of the FE model in (4.1). The resulting model for $y_{i,j}$ is

$$y_{i,j} = \alpha_{d_i} + \beta_j + \theta_{d_i,j} + \gamma_i + \epsilon_{i,j} \ , \tag{4.6}$$

where γ_i and $\epsilon_{i,j}$ satisfy equations (4.3) and (4.5), respectively, and α_{d_i}, β_j, and $\theta_{d_i,j}$ are the fixed effects of (4.2). In Table 4, we present maximum likelihood estimates of the covariance parameters σ_G^2, ρ_G, σ_R^2, and ρ_R, for both Type A and Type B arrivals. We use a learning set consisting of 42 days, as in §6 and §7. Positive interday and intraday correlations are indicated by the values of ρ_G and ρ_R. For details on the estimation of linear mixed models, see Henderson (1975).

4.3. Distributional forecasts

Call center managers need both point and distributional forecasts of future arrival counts. Distributional forecasts may consist of confidence interval estimates and densities. In this paper, we assume that the square-root transformed counts, $y_{i,j}$, are normally distributed for all $i = 1, 2, ..., D$ and $j = 1, 2, ..., P$. Therefore, we rely on conditional multivariate Gaussian theory to obtain interval forecasts for future counts; see Henderson (1975). The conditioning set for a single prediction is the corresponding learning set: With additional information about recent arrivals, the learning set is updated and, consequently, so are the predictions. In Tables 5-7, we assess the forecasting accuracy of our models by computing both point and interval predictions for future arrival counts.

	FE	ME	BM1	BM2
$\sigma^2_{G,A}$	–	0.33	0.20	0.16
$\rho_{G,A}$	–	0.83	0.83	0.81
$\sigma^2_{G,B}$	–	0.58	0.75	0.60
$\rho_{G,B}$	–	0.73	0.81	0.64
$\sigma^2_{R,A}$	0.67	0.32	0.64	0.67
$\rho_{R,A}$	–	0.15	0.10	0.13
$\sigma^2_{R,B}$	0.84	0.34	0.56	0.66
$\rho_{R,B}$	–	0.24	0.25	0.13
$\rho_{R,AB}$	–	–	–	0.46
$\rho_{G,AB}$	–	–	0.25	–

Table 4: Estimates of covariance parameters for Type A and Type B queues with a learning period of 42 days.

5. Bivariate Mixed-Effects Models

In this section, we describe bivariate mixed-effects (BM) models which extend the univariate ME model of §4.2 by exploiting the dependence structure between arrivals to separate queues.

Correlations between the arrival counts of different call types are commonly observed in practice. For example, they may arise in multi-lingual call centers where certain service requests are handled in different languages; e.g., see Table 3. More generally, they may arise when arrivals to different queues are driven by the same underlying causes, e.g., specific marketing campaigns or special events. Thus, it is important to propose and study alternative statistical models which exploit intertype correlations. Here, we describe two such models, BM1 and BM2. The BM1 and BM2 models exploit intertype correlations at the daily and half-hourly levels, respectively. In §6 and §7, we study the forecasting accuracy of the BM1 and BM2 models, and show that they uniformly lead to accurate point and interval predictions.

5.1. The BM1 Model

Let $y^A_{i,j}$ ($y^B_{i,j}$) denote the square-root transformed arrival count for Type A (Type B) in period j of day i, where $i = 1, 2, ..., D$ and $j = 1, 2, ..., P$. As in §4.1 and §4.2, we assume that both $y^A_{i,j}$ and $y^B_{i,j}$ are normally distributed random variables. In this subsection and the next, we propose different ways of modeling correlations between $y^A_{i,j}$ and $y^B_{i,j}$.

In our first bivariate model, BM1, we propose modeling intertype correlations at the daily level. We begin by modeling the marginal distribution of the arrival counts for each call type. In particular, we assume that $y^A_{i,j}$ and $y^B_{i,j}$ are each (separately) modeled by a univariate ME model, as in §4.2. That is, paralleling (4.6), we assume that,

$$y^A_{i,j} = \alpha^A_{d_i} + \beta^A_j + \theta^A_{d_i,j} + \gamma^A_i + \epsilon^A_{i,j} \text{ , and,} \tag{5.1}$$

$$y^B_{i,j} = \alpha^B_{d_i} + \beta^B_j + \theta^B_{d_i,j} + \gamma^B_i + \epsilon^B_{i,j} \text{ ,} \tag{5.2}$$

for $i = 1, 2, ..., D$ and $j = 1, 2, ..., P$. We assume that the random effects, γ^A_i, follow an AR(1) structure with parameters $\sigma_{G,A}$ and $\rho_{G,A}$, as in (4.3). We assume that the residuals $\epsilon^A_{i,j}$ follow an AR(1) structure with parameters $\sigma_{R,A}$ and $\rho_{R,A}$, as in (4.5). We make the same assumptions for Type B. That is, we assume that γ^B_i follow an AR(1) structure with parameters $\sigma_{G,B}$ and $\rho_{G,B}$, and $\epsilon^B_{i,j}$ follow an AR(1) structure with parameters $\sigma_{R,B}$ and $\rho_{R,B}$. In (5.1) and (5.2), we also use the same fixed effects, α, β, and θ, as in (4.2).

In order to model the dependence between $y^A_{i,j}$ and $y^B_{i,j}$, we assume that the random effects γ^A_i and γ^B_i are correlated for every $i = 1, 2, ..., D$. In particular, we assume that they satisfy the following set of equations

$$\gamma^A_i = \rho_{G,A}\gamma^A_{i-1} + \nu^A_i \text{ , and,} \tag{5.3}$$

$$\gamma^B_i = \rho_{G,B}\gamma^B_{i-1} + \nu^B_i \text{ ,} \tag{5.4}$$

where ν^A_i are i.i.d. normally distributed random variables with $E[\nu^A] = 0$ and $\text{Var}[\nu^A] = \sigma^2_{G,A}(1 - \rho^2_{G,A})$, and ν^B_i are i.i.d. normally distributed random variables with $E[\nu^B] = 0$ and $\text{Var}[\nu^B] = \sigma^2_{G,B}(1 - \rho^2_{G,B})$. We assume that ν^A_i and ν^B_i are correlated for every i, and denote the correlation between them by $\text{cor}(\nu^A_i, \nu^B_i) = \rho_{G,AB}$. As a result, $y^A_{i,j}$ and $y^B_{i,j}$ are correlated for every i and j, i.e., they are correlated across different days, and across the periods of the same day. In Table 4, we present point estimates for the covariance parameters of the BM1 model. We use a learning set consisting of 42 days. The value of $\rho_{G,AB}$ (which is roughly equal to 0.25) indicates that the two call types are positively correlated at the interday level.

5.2. The BM2 model

For the BM2 model, we also assume that $y_{i,j}^A$ and $y_{i,j}^B$ are each separately modeled by a univariate ME model. In particular, we assume that (5.1) and (5.2) continue to hold. As in §5.1, we assume that γ_i^A and γ_i^B each follow an AR(1) structure with covariance parameters ($\sigma_{G,A}$, $\rho_{G,A}$) and ($\sigma_{G,B}$, $\rho_{G,B}$), respectively, for $i = 1, 2, ..., D$. However, in contrast with the BM1 model, we now assume that ν_i^A and ν_i^B in (5.3) and (5.4) are independent across call types, for every $i = 1, 2, ..., D$.

To model intertype correlations, we assume that the residuals of the two underlying ME models, i.e., $\epsilon_{i,j}^A$ and $\epsilon_{i,j}^B$ in (5.1) and (5.2), are correlated for each given i and $j = 1, 2, ..., P$. As in the BM1 model, we let the residuals $\epsilon_{i,j}^A$ follow an AR(1) structure with parameters $\sigma_{R,A}$ and $\rho_{R,A}$, for each given i and $j = 1, 2, ..., P$. Similarly, we let the residuals $\epsilon_{i,j}^B$ follow an AR(1) structure with parameters $\sigma_{R,B}$ and $\rho_{R,B}$. To model correlations between $\epsilon_{i,j}^A$ and $\epsilon_{i,j}^B$, we assume that

$$\epsilon_{i,j}^A = \rho_{R,A}\epsilon_{i,j-1}^A + \kappa_{i,j}^A \text{ , and,} \tag{5.5}$$

$$\epsilon_{i,j}^B = \rho_{R,B}\epsilon_{i,j-1}^B + \kappa_{i,j}^B \text{ ,} \tag{5.6}$$

for $i = 1, 2, ..., D$ and $j = 1, 2, ..., P$. In (5.5), $\kappa_{i,j}^A$ are i.i.d. normally distributed random variables with $E[\kappa^A] = 0$ and $\text{Var}[\kappa^A] = \sigma_{R,A}^2(1-\rho_{R,A}^2)$. Similarly, in (5.6), $\kappa_{i,j}^B$ are i.i.d. normally distributed random variables with $E[\kappa^B] = 0$ and $\text{Var}[\kappa^B] = \sigma_{R,B}^2(1-\rho_{R,B}^2)$. We assume that $\kappa_{i,j}^A$ and $\kappa_{i,j}^B$ are correlated for every fixed i, and $j = 1, 2, ..., P$. We denote the correlation between them by $\text{cor}(\kappa_{i,j}^A, \kappa_{i,j}^B) = \rho_{R,AB}$. We assume that κ^A and κ^B are independent across different days. As a result, $y_{i,j}^A$ and $y_{i,j}^B$ are correlated within a given day i, for all $j = 1, 2, ..., P$. However, they are independent across different days. In Table 4, we present point estimates for the covariance parameters of the BM2 model. The estimated value of $\rho_{R,AB}$ is roughly equal to 0.46, and indicates that the two call types are positively correlated at the intraday level.

5.3. Distributional forecasts

It is relatively easy to generate distributional forecasts for bivariate models. Indeed, the joint distribution of the arrival counts for both call types, Type A and Type B, is assumed to be multivariate normal. Therefore, as with the ME model, we can derive distributional forecasts by relying on conditional multivariate Gaussian theory, where the conditioning set is the learning set for a given prediction. In Tables 5-7, we assess the forecasting accuracy of our models by computing both point and interval predictions for future arrival counts.

6. Model Comparison

In this section, we compare the statistical models of §4 and §5 based on their forecasting performance. We conduct an empirical study using the data set described in §3. In particular, we make out-of-sample forecasts for alternative forecasting lead times, and quantify the accuracy of the forecasts generated by the candidate models. In Table 5, we present point and interval predictions for both Type A and Type B arrivals. The best values are highlighted in bold. In §7, we consider additional data sets to further substantiate our results.

6.1. Lead Times and Learning Period

We generate out-of-sample forecasts for the forecasting horizon ranging from August 19, 2010 to November 11, 2010. That is, we make forecasts for a total of 55 days (excluding weekends and removed outlier days) and generate $55 \times 22 = 1210$ predicted values for each call type. We consider four different forecasting lead times to mimic real-life challenges faced by call center managers; see §1. In particular, we consider lead times of 2 weeks, 1 week, 1 day, and half a day (which corresponds to 11 half-hour periods in our context). We let the learning period consist of 42 days, i.e., 6 weeks. When we generate a forecast for all periods of a given day, we roll the learning period forward so as to preserve the length of the forecasting lead time. We re-estimate all model parameters after each forecast. We do the estimation using our own code, written in MATLAB.

6.2. Performance Measures

We quantify the accuracy of a point prediction by computing the *average squared error* (ASE) per half-hour period, defined by:

$$ASE \equiv \frac{1}{K} \sum_{i,j} (N_{i,j} - \hat{N}_{i,j})^2 \ , \tag{6.1}$$

where $N_{i,j}$ is the number of arrivals in the j^{th} period of a given day i, $\hat{N}_{i,j}$ is the predicted value of $N_{i,j}$, and K is the total number of predictions made. We also compute the *average percent error* (APE), for a relative measure of accuracy, defined by:

$$\text{APE} \equiv 100 \cdot \frac{1}{K} \sum_{i,j} \frac{|N_{i,j} - \hat{N}_{i,j}|}{N_{i,j}} \ . \tag{6.2}$$

To evaluate the distributional forecasts generated by the candidate models, we use performance measures to describe the prediction intervals for $N_{i,j}$; see §4.3 and §5.3. In particular, we define the *average cover* (Cover) of a prediction interval for a given model as:

$$\text{Cover} = \frac{1}{K} \sum_{i,j} I(N_{i,j} \in (\hat{L}_{i,j}, \hat{U}_{i,j})) \ , \tag{6.3}$$

where $I(\cdot)$ denotes the indicator random variable, and $\hat{L}_{i,j}$ and $\hat{U}_{i,j}$ are the lower and upper bounds of the prediction interval, respectively. In this paper, we compute prediction intervals with a confidence level of 95%. If the chosen model adequately captures the correlation structure in the data, then we expect that the average cover be close to 95%. The average cover corresponds to the unconditional coverage of Christoffersen (1998). Finally, we also compute the *relative half width* (Width) of a prediction interval, defined as:

$$\text{Width} = \frac{1}{K} \sum_{i,j} \frac{0.5|\hat{U}_{i,j} - \hat{L}_{i,j}|}{\hat{N}_{i,j}} \ . \tag{6.4}$$

6.3. Forecasting Performance

6.3.1. Two Weeks Ahead Forecasts

Predictions for Type A calls. With a forecasting lead time of 2 weeks, Table 5 shows that the FE model generates the most accurate point forecasts, among all models considered. Consistent with Taylor (2008), this shows that a simple historical average can be difficult to beat with relatively long forecasting lead times. Indeed, Table 5 shows that both ASE(ME) and ASE(BM1) are roughly 8% larger than ASE(FE). Moreover, ASE(BM2) is roughly 12% larger than ASE(FE). Table 5 also shows that APE(FE) is roughly 0.6% smaller than APE(ME), and roughly 0.4% smaller than APE(BM1). However, the predictive power of bivariate models becomes evident when comparing the prediction intervals generated by the alternative models. Indeed, the average cover for both BM1 and BM2 is close to 0.90, whereas Cover(ME) is roughly equal to 0.86 and Cover(FE) lags behind at 0.47. The widths of the prediction intervals for BM1 and BM2 are slightly larger than for ME, and significantly larger than for FE. Indeed, the FE model considerably underestimates uncertainty in the data by not capturing any correlation structure between the arrival counts.

Predictions for Type B calls. With a forecasting lead time of 2 weeks, Table 5 shows that all models perform nearly the same from the ASE and APE perspectives. Indeed, ASE(BM1) is the smallest among all models considered, and ASE(FE)/ASE(BM1) is roughly equal to 1.02. Additionally, Table 5 shows that the APE's for all models are roughly the same. With Type B arrivals, the average cover of prediction intervals for the ME and BM models is the same (approximately equal to 0.93), whereas the cover for the FE model is much lower, and is roughly equal to 0.5.

6.3.2. One Week Ahead Forecasts

Predictions for Type A calls. With a forecasting lead time of 1 week, Table 5 shows that the FE model continues to generate the most accurate point forecasts. Indeed, ASE(ME)/ASE(FE) is roughly equal to 1.15, whereas ASE(BM1)/ASE(FE) is roughly equal to 1.10. Both the BM1 and BM2 models generate more accurate point forecasts than the ME model. For example, the difference between APE(ME) and APE(BM1) is roughly equal to 0.5%. Moreover, the cover for the

BM models is better than for the ME model; e.g., Cover(BM1) ≈ 0.92 whereas Cover(ME) ≈ 0.85. In short, both BM models generate more accurate point and interval forecasts than the ME model.

Predictions for Type B calls. Table 5 shows that the BM1 model yields the most accurate point predictions. For example, ASE(ME)is roughly 5% larger than ASE(BM1), and APE(ME) is roughly 0.3% larger than APE(BM1). Moreover, the cover of prediction intervals for the BM models is slightly larger than for the ME model. The FE model continues to yield accurate point forecasts, e.g., ASE(FE)/ASE(BM1) ≈ 1.03. Consistent with previous results, the cover of the prediction intervals generated by the FE model is poor. Indeed, Cover(FE) ≈ 0.47.

6.3.3. One Day Ahead Forecasts

Predictions for Type A calls. With a forecasting lead time of 1 day, Table 5 shows that the BM models yield more accurate point and interval forecasts than both the ME and the FE models. For example, ASE(ME) is roughly 8% larger than ASE(BM2), and APE(ME) is roughly 0.4% larger than APE(BM1). Additionally, Cover(BM1) and Cover(BM2) are both roughly equal to 0.95, whereas Cover(ME) is roughly equal to 0.88. The widths of the prediction intervals for the BM models is slightly larger than for the ME model, since they account for more uncertainty in the data. The FE model yields considerably less accurate point and interval forecasts than the BM and ME models; e.g., ASE(FE)/ASE(BM2) is roughly equal to 1.31.

Predictions for Type B calls. The ME, BM1, and BM2 models perform roughly the same in this case. Indeed, ASE(BM1) is only slightly smaller than ASE(ME), and APE(ME) is roughly equal to APE(BM1). The average cover of prediction intervals for the BM1 model (0.95) is slightly larger than for the ME model (0.92).

6.3.4. Within-day Forecasts

Predictions for Type A calls. Table 5 shows that the ME and BM models perform roughly the same in this case. Indeed, ASE(ME) is only marginally smaller than the ASE(BM2). Moreover, the cover and widths for the BM and ME models are similar, and show that both models fit the

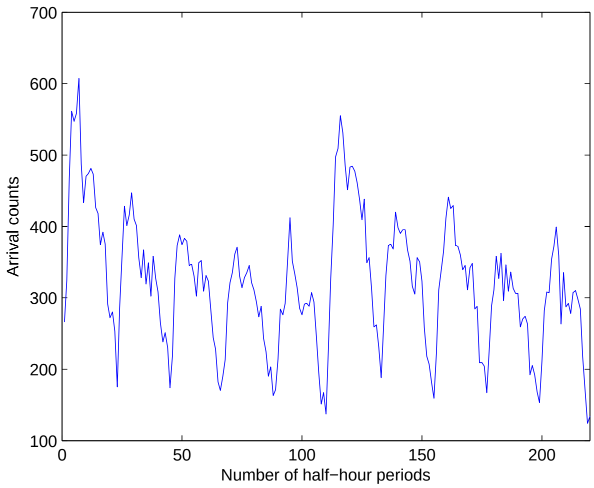

Figure 5: Type C arrivals for two weeks starting June 7, 2010.

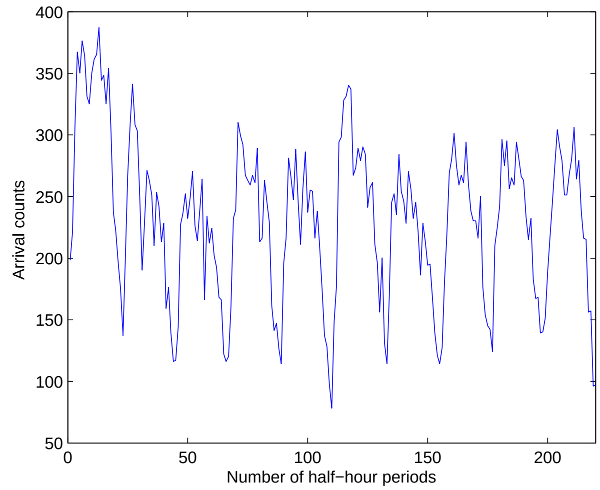

Figure 6: Type E arrivals for two weeks starting June 7, 2010.

data well. The FE model performs considerably worse than the ME and BM models, in terms of both point and interval predictions. For example, ASE(FE)/ASE(BM1) is roughly equal to 2.

Predictions for Type B calls. With a forecasting lead time of half a day, the BM2 model yields more accurate point and interval forecasts than the ME model. For example, ASE(ME) is roughly 8% larger than ASE(BM2), and APE(ME) is roughly 0.4% larger than APE(BM2). The BM1 model performs slightly worse than the BM2 model, but better than the ME model. Finally, the cover of the prediction intervals for the BM and ME models are all roughly equal to 0.9.

The results of this section show that the BM models usually lead to more accurate point and interval forecasts than both the ME and FE models. In §7, we provide further empirical evidence supporting our conclusions, by considering two additional data sets.

7. Additional Data Sets

In this section, we consider two additional data sets, taken from the same call center as in §3, and study the forecasting accuracy of our candidate models using each set. Preliminary data analysis of those two sets is largely similar to our previous analysis in §3. Therefore, we omit detailed data descriptions in this section. In §7.1 and §7.2, we present estimates for the forecasting performance of the FE, ME, and BM models. For our predictions, we continue to use a learning period of 42 days, and forecasting lead times of 2 weeks, 1 week, 1 day, and half a day; see §6.

7.1. First Additional Data Set: Call Type C

The data from the first set were collected over 305 days (excluding weekends) ranging from June 1, 2010 to July 25, 2011. The data consist of arrival counts for two call types, Type C and Type D, whose incoming calls originate in the Canadian provinces of Ontario and Quebec, respectively. These two types correspond, otherwise, to similar service requests. There are 8 outlier days which were removed with the data. Thus, we are left with a total of $D = 297$ remaining days. In this section, we report forecasting performance estimates for Type C counts since those have higher volumes than Type D counts. In Figure 5, we present a plot of the time series of Type C arrival counts for the two weeks ranging from June 7, 2010 to June 21, 2010. For the BM models in Table 6, we jointly model Type C and Type D arrivals.

7.1.1. Two Weeks Ahead Forecasts.

Table 6 shows that, with a forecasting lead time of 2 weeks, the BM2 model generates the most accurate point predictions. The BM1 model performs similarly to BM2. For example, ASE(ME) is roughly 16% larger than ASE(BM2), and is roughly 15% larger than ASE(BM1). From the APE perspective, both APE(BM2) and APE(BM1) are roughly 0.6% smaller than APE(ME). The FE model yields more accurate point predictions than the ME model. Indeed, ASE(FE) is roughly 11% larger than ASE(BM2). The ME and BM models all yield similar prediction intervals of future counts. Indeed, Cover(BM2) is roughly equal to 0.92, whereas Cover(ME) is roughly equal to 0.94. Additionally, the average widths of corresponding prediction intervals are roughly the same. Consistent with §6, the FE model yields prediction intervals which have a relatively small cover. Indeed, Cover(FE) ≈ 0.53 in this case.

7.1.2. One Week Ahead Forecasts.

Table 6 shows that, with a forecasting lead time of 1 week, the BM1 model yields the most accurate point forecasts. Indeed, ASE(ME) is roughly 11% larger than ASE(BM1). The BM2 model continues to outperform the ME model as well: ASE(ME) is roughly 6% larger than ASE(BM2). The FE model yields similar point estimates as the ME model. The widths of the prediction intervals for the ME and BM models are nearly the same. The model which best fits the data, yielding the best average cover for prediction intervals, is the BM2 model; indeed, Cover(BM2) ≈ 0.96.

7.1.3. One Day Ahead Forecasts.

In this case, the BM1 and BM2 models perform nearly the same, and continue to yield more accurate point and interval forecasts than the ME model. Indeed, ASE(ME) is about 6% larger than ASE(BM2). The FE model performs considerably worse: e.g., ASE(FE) is 35% larger than ASE(BM2). The ME and BM models have good average covers, both close to 95%. Additionally, all prediction intervals have similar relative widths.

7.1.4. Within-Day Forecasts.

As before, the BM1 and BM2 models perform nearly the same, and both outperform the ME model. (The FE model is no longer competitive in this case.) Indeed, ASE(ME) is roughly 10% larger than both ASE(BM1) and ASE(BM2). The cover and prediction intervals for the ME and BM models are roughly the same; e.g., Cover(ME) ≈ 0.96.

7.2. Second Additional Data Set: Call Type E

As in §7.1, the data were also collected over 305 days (excluding weekends) ranging from June 1, 2010 to July 25, 2011. The data consist of arrival counts for two call types, Type E and Type F. Both call types originate in the Quebec province, but correspond to different service requests. Call center managers at the company informed us that, based on their experience, arrivals to these two call types are dependent in practice. In this subsection, we investigate whether exploiting this dependence leads to more accurate forecasts. There are 9 outlier days which were removed from the data set. Thus, we are left with a total of $D = 296$ remaining days. In Figure 6, we present a plot of the time series of the arrivals from June 7, 2010 to June 21, 2010. In this section, we report forecasting estimates for Type E calls only, since they have larger volumes than Type F calls. For the BM models in Table 7, we jointly model Type E and Type F arrivals.

7.2.1. Two Weeks Ahead Forecasts.

Table 7 shows that the FE model yields the most accurate point forecasts, among all models considered. The ME model performs nearly the same as the FE model. Indeed, ASE(ME) is roughly equal to ASE(FE). The BM2 model yields the least accurate forecasts in this case, with ASE(BM2) roughly 10% larger than ASE(FE). From the APE perspective, APE(ME) and APE(BM1) are nearly the same. The BM2 model falls behind, and APE(BM2) is roughly 0.7% larger than APE(FE). Consistent with prior results, the cover for the FE model is poor (roughly equal to 0.4). The cover of prediction intervals for the BM models (roughly equal to 0.85) are slightly smaller than for the ME model (roughly equal to 0.87).

7.2.2. One Week Ahead Forecasts.

Table 7 shows that the most accurate model, from the ASE and APE perspectives, is the BM1 model. For example, ASE(ME) is roughly 3% larger than ASE(BM1), and APE(ME) is nearly 0.4% larger than APE(BM1). The BM2 model performs roughly the same as the ME model. Additionally, the average covers and widths of the prediction intervals for the ME and BM models are all roughly the same (approximately 0.86).

7.2.3. One Day Ahead Forecasts.

Table 7 shows that the BM1 and BM2 models perform nearly the same in this case. The model which yields the most accurate point forecasts is the BM2 model, but ASE(BM2) is only 2% smaller than ASE(ME). In terms of APE, APE(ME) is about 0.2% larger than APE(BM1). The cover and widths for the ME and BM models are nearly the same. The FE model performs very poorly with ASE(FE)/ASE(BM1) roughly equal to 1.36.

7.2.4. Within-Day Forecast.

Consistent with previous results, the model which yields the most accurate point predictions is the BM1 model. Indeed, ASE(ME) is roughly 5% larger than ASE(BM1). Moreover, APE(ME) is roughly 0.4% larger than APE(BM1). The BM1 model is only slightly less accurate than the BM2 model, with ASE(BM2)/ASE(BM1) roughly equal to 1.01. The cover of the prediction intervals for the ME model are better than for the BM models. For example, Cover(ME) ≈ 0.88 whereas Cover(BM1) ≈ 0.82.

8. Conclusions

8.1. Summary of Main Contributions

In this paper, we evaluated alternative statistical models for the half-hourly arrival counts to a call center. Our different models incorporate different dependence structures in the data. We compared the forecasting accuracy of those models based on three real-life call center data sets from a major telecommunications company in Canada. We compared the forecasting accuracy of all models based on forecasting lead times ranging from weeks to hours in advance, to mimic the challenges faced by call center managers in real life. Our study shows the importance of accounting for different correlation structures in the data when forecasting future arrival counts.

We modeled interday and intraday dependence structures, commonly observed in call center arrivals, via a Gaussian linear mixed-effects model, ME (§4.2). The mixed model incorporates fixed effects such as day-of-week, period-of-day, and cross-terms between the two. As a useful reference model, we also considered a fixed-effects model, FE (§4.1) which does not incorporate any dependence structure in the data. Tables 5, 6, and 7 showed that the FE model can be competitive with a relatively long forecasting lead time, but not otherwise.

Correlations between the arrival counts of different call types are commonly observed in practice; e.g., see Table 3. In §5, we extended the mixed model into two bivariate mixed models, thus innovatively exploiting correlations between the arrival counts of different call types. Our two bivariate models, BM1 and BM2, exploit intertype correlations at the interday and intraday levels, respectively. In §6 and 7, we showed that bivariate models generally lead to more accurate point and interval forecasts than both ME and FE models; see Tables 5, 6, and 7. As a result, we quantified the impact of modeling correlations between alternative call types when making forecasts of future call volumes. We reached consistent conclusions with all three data sets considered.

8.2. Future Research Directions

One possible direction for future research is to extend bivariate models into multivariate models which exploit the dependence structure between multiple related call types. Indeed, experience indicates that the arrival processes of several queues in call centers are often correlated, and exploiting this correlation promises to improve predictions of future call volumes, as we saw in §6 and §7. One potential difficulty in that computational difficulties are bound to arise when estimating multivariate models involving many different call types.

As indicated in §1, forecasting call center arrivals is an essential first step towards the better management of call centers. There remains to study the more complicated problem of developing efficient algorithms for scheduling agents, and updating the resulting schedules, based on distributional forecasts of future call volumes. In this paper, we used a square-root transformation of the data (§3.3) and exploited the resulting normality of the transformed counts. There remains to characterize the marginal and joint densities of the untransformed arrival counts, as in Avramidis et al. (2004). Simulation-based methods may be used in complicated systems where there are multiple customer classes and multiple service pools with some form of skill-based routing. Distributional forecasts of future arrivals could potentially be implemented in those simulation models to produce approximate solutions to the agent scheduling problem, see Avramidis et al. (2010).

9. References

Aksin, O.Z., Armony, M. and V. Mehrotra 2007. The Modern Call-Center: A Multi-Disciplinary Perspective on Operations Management Research, *Production and Operations Management*, 16: 665 – 688.

Aldor-Noiman, S. 2006. Forecasting demand for a telephone call center: Analysis of desired versus attainable precision. Unpublished masters thesis, Technion-Israel Institute of Technology, Haifa, Israel.

Aldor-Noiman, S., Feigin, P. and A. Mandelbaum. 2009. Workload forecasting for a call center: Methodology and a case study. *Annals of Applied Statistics*, 3: 1403–1447

Andrews, B. H. and S. M. Cunningham. 1995. L.L. Bean improves call-center forecasting. *Interfaces*, 25: 1–13.

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

Avramidis, A.N., Chan, W., Gendreau, M., L'Ecuyer, P. and O. Pisacane. 2010. Optimizing daily agent scheduling in a multiskill call center. *European Journal of Operational Research*, 200: 822–832.

Barry, S.J.E. and A. Bowman. 2007. Linear mixed models for longitudinal shape data with applications to facial modeling. *Journal of Biostatistics*, 9: 555–565.

Bianchi, L., Jarrett, J. and R.C. Hanumara. 1998. Improving forecasting for telemarketing centers by ARIMA modeling with intervention. *International Journal of Forecasting*, 14: 497–504.

Brown, L. D., Zhang, R. and L. Zhao. 2001. Root un-root methodology for non parametric density estimation. Technical report, The Wharton, Univ. Pennsylvania.

Brown, L., Gans, N., Mandelbaum, A., Sakov, A., Shen, H., Zeltyn, S. and L. Zhao. 2005. Statistical analysis of a telephone call center: a queueing-science perspective. *Journal of American Statistical Association*, 100: 36–50.

CIA World Factbook. 2010. Available online at: https://www.cia.gov.

Channouf, N. L'Ecuyer, P., Ingolfsson, A. and A. Avramidis. 2007. The application of forecasting techniques to modeling emergency medical system calls in Calgary, Alberta. *Health Care Management Science*, 10: 25-45.

Chatfield . C. and M. Yar. Holt-Winters forecasting: Some practical issues. 1988. *The Statistician*, 37: 129–140.

Christoffersen, P. 1998. Evaluating Interval Forecasts, International Economic Review, 39: 841-862.

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

Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 31: 423-447.

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

Mandelbaum, A. 2002. Call Centers (Centres): Research Bibliography with Abstracts, Version 3. Downloadable from ie.technion.ac.il/ serveng/References/ccbib.pdf.

Muller, K. and P. Stewart. 2006. Linear Model Theory: Univariate, Multivariate, and Mixed Models. Wiley, New York.

SAS. 2009. SAS OnlineDoc® 9.1 SAS Institute Inc., Cary, NC.

Soyer, R. and Tarimcilar, M. M. 2008. Modeling and Analysis of call center arrival data: A Bayesian approach. *Management Science*, 54: 266–278.

Shen H. and J. Z. Huang 2008a. Forecasting time series of inhomogeneous poisson process with application to call center management software. *Annals of Applied Statistics*, 2: 601–623.

Shen H. and J. Z. Huang 2008b. Intraday forecasting and interday updating of call center arrivals. *Manufacturing and Service Operations Management*, 10: 391–410.

Steckley, S.G., Henderson, S.G. and V. Mehrotra. 2005. Performance measures for service systems with a random arrival rate. *Proceedings of the 2005 Winter Simulation Conference.* M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, eds, 566–575.

Tanir, O. and R. J. Booth. 1999. Call center simulation in Bell Canada. *Proc 1999 Winter Simulation Conference*. P. A. Farrington, H. B. Nemhard, D.T. Sturrock, G.W. Evans, eds , 1640–1647.

Taylor, J. W. 2008. A comparison of univariate time series methods for forecasting intraday arrivals at a call center. *Management Science*, 54: 253–265.

Thiébaut, R., H. Jacqmin-Gadda, G. Chne, Leport, C. and D. Commenges. 2007. Bivariate linear mixed models using SAS proc MIXED. *Computer Methods and Programs in Biomedicine*, 69: 249-56

Weinberg J., Brown, L. D. and J. R. Stroud. 2007. Bayesian forecasting of an inhomogeneous Poisson process with applications to call center data. *Journal of American Statistical Association*, 102: 1185–1199.

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

Winters PR. 1960. Forecasting sales by exponentially weighted moving averages. *Management Science*, 6 : 324-342.

Predictions for a forecast lead time of 14 days

	Type A				Type B			
	ASE	APE	Cover	Width	ASE	APE	Cover	Width
FE	**3059**	**13.9**	0.47	0.10	1292	**13.3**	0.50	0.11
ME	3301	14.5	0.86	0.29	1290	13.3	0.92	0.30
BM1	3293	14.3	0.90	0.31	**1264**	13.4	**0.93**	0.29
BM2	3414	14.7	**0.90**	0.34	1269	13.4	0.92	0.31

Predictions for a forecast lead time of 7 days

	Type A				Type B			
	ASE	APE	Cover	Width	ASE	APE	Cover	Width
FE	**3186**	**14.1**	0.45	0.10	1330	13.6	0.47	0.10
ME	3663	15.0	0.85	0.26	1351	13.8	0.92	0.32
BM1	3506	14.5	0.93	0.42	**1287**	**13.5**	**0.95**	0.28
BM2	3366	14.5	**0.93**	0.34	1288	13.5	0.95	0.32

Predictions for a forecast lead time of 1 day

	Type A				Type B			
	ASE	APE	Cover	Width	ASE	APE	Cover	Width
FE	3193	14.1	0.50	0.11	1273	13.4	0.50	0.11
ME	2636	12.6	0.88	0.24	1127	12.5	0.92	0.30
BM1	2463	**12.2**	**0.95**	0.32	**1113**	**12.4**	**0.95**	0.25
BM2	**2440**	12.3	0.94	0.34	1126	12.5	0.94	0.30

Predictions for a forecast lead time of 0.5 days

	Type A				Type B			
	ASE	APE	Cover	Width	ASE	APE	Cover	Width
FE	3194	14.1	0.50	0.12	1273	13.4	0.50	0.11
ME	**1612**	**9.80**	0.91	0.21	871	10.6	**0.91**	0.22
BM1	1639	9.80	**0.93**	0.24	832	10.4	0.89	0.22
BM2	1616	9.80	0.92	0.25	**806**	**10.2**	0.90	0.23

Table 5: Accuracy of point and interval predictions for Type A and Type B calls for alternative forecasting lead times and a learning period of 42 days.

Predictions for Type C Calls

	14 Days ASE	14 Days APE	14 Days Cover	14 Days Width	7 Days ASE	7 Days APE	7 Days Cover	7 Days Width
FE	585	12.8	0.53	0.12	546	12.4	0.52	0.11
ME	617	13.1	**0.94**	0.34	548	12.4	0.93	0.31
BM1	538	12.6	0.91	0.32	**493**	**12.1**	**0.96**	0.39
BM2	**531**	**12.5**	0.92	0.32	514	12.2	0.93	0.36

	1 Day ASE	1 Day APE	1 Day Cover	1 Day Width	0.5 Days ASE	0.5 Days APE	0.5 Days Cover	0.5 Days Width
FE	523	12.0	0.56	0.12	548	12.3	0.52	0.11
ME	410	10.6	0.94	0.27	365	9.9	0.95	0.28
BM1	**387**	**10.6**	**0.96**	0.28	336	9.9	0.96	0.27
BM2	389	10.7	0.94	0.27	**335**	**9.8**	**0.96**	0.27

Table 6: Predictions for Type C calls for alternative forecasting lead times and a learning period of 42 days.

Predictions for Type E Calls

	14 Days				7 Days			
	ASE	APE	Cover	Width	ASE	APE	Cover	Width
FE	1781	**16.2**	0.40	0.10	2034	17.6	0.44	0.10
ME	**1779**	16.2	**0.87**	0.29	2056	17.8	0.86	0.28
BM1	1860	16.3	0.84	0.29	**2008**	**17.4**	**0.86**	0.29
BM2	1979	16.9	0.85	0.34	2048	17.6	0.84	0.30

	1 Day				0.5 Days			
	ASE	APE	Cover	Width	ASE	APE	Cover	Width
FE	2010	17.4	0.48	0.11	2032	17.4	0.47	0.11
ME	1514	15.0	0.88	0.26	1068	12.0	**0.88**	0.19
BM1	**1483**	**14.8**	**0.88**	0.25	1026	11.7	0.82	0.22
BM2	1484	14.8	0.86	0.27	**1015**	**11.6**	0.82	0.20

Table 7: Predictions for Type E calls for alternative forecasting lead times and a learning period of 42 days.