Forecasting Demand for a Telephone Call Center: Analysis of Desired versus Attainable Precision

Reasearch Thesis

Submitted in Partial Fullfillment of the
Requirements for the Degree of
Master of Science in Statistics

Sivan Aldor-Noiman

Submitted to the Senate of the
Technion - Israel Institute of Technology
TAMUZ, 5766 – HAIFA – JULY, 2006

Acknowledgments

The Research Thesis was written under the supervision of Professor Paul Feigin in the Faculty of Industrial Engineering and Management. I would like to express my deep gratitude for the countless hours we have worked together. I feel privileged to have learned from this brilliant statistician and more importantly this dedicated advisor. His desire to teach and to continue learning made this research both challenging and at the same time a fascinating experience. This research could not have been completed without his open-door policy and is everlasting patience; and for that I am truly grateful.

The Research Thesis was also written under the co-supervision of Professor Avishai Mandelbaum in the Faculty of Industrial Engineering and Management. I would like to thank him for giving me a different prespective on things. The knowledge I gained during our work together is priceless. I thank him for the time and effort he so willingly contributed to this research.

Also I would like to thank my family for supporting and encouraging me to attain my goals.

Finally, the generous financial help of the Technion Graduate School is gratefully acknowledged.

Contents

List of Tables

List of Figures

Abstract

Today's call centers managers face multiple operational decision making tasks. One of their most common chores is determining the weekly staffing levels to ensure customer satisfaction and needs while minimizing service costs. An initial step for producing the weekly schedule is forecasting the future system loads comprising both the predicted arrival counts and the average service times.

After obtaining the forecasted system load, in large call centers, a manager can implement the QED (Quality-Efficiency Driven) regime "square-root staffing" rule to allow balancing between the offered load per server and quality of service. Implementing this staffing rule requires that the forecasted values maintain certain levels of precision. One of the aims of this thesis is to determine whether or not these levels can be achieved by practical algorithms.

In this thesis we introduce two arrival count models which are based on a *mixed* Poisson process approach. The first model uses the Normal-Poisson stabilization transformation in order to employ linear mixed model techniques. The model is implemented and analyzed on two different data sets. In one of the call centers the data include billing cycles information and we also demonstrate how to incorporate it as exogenous variables in this model. We develop different goodness-of-fit criteria that help determine the models performance under the QED regime. These show that during most hours of the day the model can reach the desired precision levels. Actually, whenever the QED regime and square root staffing formula are appropriate, the model performs well. We also demonstrate the effect the forecasting lead time (that is, the time between the last learning data and the first forecasted time) has on this model precision.

We also demonstrate how our mixed model can achieve very similar levels of precision when compared to other models, such as the Bayesian model developed by Weinberg *et al.* in [22]. This similarity holds even though our model's predictions are based on smaller amounts of learning data.

Our second model employs the Bayesian approach, implementing Gibbs sampling techniques, and using ‘OpenBugs’ software, to produce the predictive distributions for the future arrival counts. Due to computational limitations we only show a ‘proof of concept’ for this model by applying it to predicting a single day’s arrivals and comparing it to the mixed model results.

We also develop a fairly simple quadratic regression model to predict the average service times needed for producing the future system loads.

List of Symbols

The following table summaries all the symbols appearing in this thesis. Some of the symbols have different meanings in different sections of the thesis. Next to these symbols, the section numbers and a brief definition. Symbols that do not have a section number indicated next to them are general symbols used throughout the thesis. Symbols which have several section numbers indicated have different meanings in those sections.

Symbol	Section	Definition
$\sim$	-	distributed as (for example, $X \sim$ Poisson(λ) means that X is a random variable that is Poisson distributed with parameter λ)
$\approx$	-	$a_n \approx b_n$ if $a_n/b_n \to 1$, as $n \to \infty$
$\hat{[]}$	-	an estimate (for example, $\hat{\theta}$ is the estimated value for θ)
ϵ_{dk}	-	inherent error term related to the k^{th} period of day d
λ_{dk}		the arrival rate during the k^{th} period of day d
λ_d		is the arrival rate during day d
μ_{dk}	-	the service rate during the k^{th} period of day d
ρ	-	correlation parameter
cov	-	covariance
$\mathbb{E}$	-	Expectation
n	-	number of observations
N_{dk}	-	the arrival count during the k^{th} period of day d
N_d	-	the number of arrival counts during day d
S	-	the number of agents
Var or σ^2	-	Variance
y_{dk}	-	$= \sqrt{N_{dk} + 0.25}$
q_d	-	the day of the week corresponding to day d
α_q	5.1	the day-of-week fixed effect for the q^{th} weekday

Symbol	Section	Definition
	5.3	is the vector of prior parameters for the (discrete) Dirichlet distribution
	5.2	the intercept for q^{th} weekday
β	5.1	the fixed-effects coefficient vector
	5.2	the auto-regressive coefficient
	7	the QED staffing policies coefficient
β_d	5.1	the billing cycles effects during day d
η_{dk}	5.1	the within-day (period) error term during the k^{th} period of day d
θ_{dk}	5.1	the expected value of y_{dk} during the k^{th} period of day d
μ_q	5.3	the q^{th} weekday mean value
υ_q	5.3	is the averaged q^{th} weekday mean value according to the learning data
π_q	5.3	the mean proportion vector for the discrete Dirichlet distribution for the q^{th} weekday.
B_i^j	6.2.1	indicates whether cycle j's billing period falls on the i$^{\text{th}}$ day
D_i^j	6.2.1	indicates whether cycle j's delivery period falls on the i^{th} day
g_{q_d}	5.2	$= \sqrt{R_{q_d}}$
G_d	5.3	day d deviation from its corresponding weekday average
M_q	5.3	a variable which governs the variability of the Dirichlet distribution about its mean α_q
$p_{q,k}$	5.1	the fixed (interaction) effect for period k of the q^{th} weekday
p_q	5.3	the vector of daily volume proportions on the q^{th} weekday for the K periods
$R_{q_d}(t_k)$	5.2	the proportion of daily volume on the q^{th} weekday for the k^{th} period
υ_d	5.2	represents the daily volume during day d
V_d	5.1	the random daily volume effect during day d
W_q	6.2.1	the q^{th} weekday effect
x_d	5.2	$= \sqrt{v_d}$
z_{q_d}	5.2	$=\{g_{q_d}(t_k), dg_{q_d}(t_k)/dt_k\}$

List of Acronyms

ANOVA	ANalysis Of VAriance
APE	Averaged Percentage Error
AR	AutoRegressive
ARIMA	AutoRegressive Integerated Moving Average
ARMA	AutoRegressive Moving Average
BUGS	Bayesian inference Using Gibbs Sampling
ChiSq	Chi-Square distribution
Cover	Prediction coverage probability
Data-MOCCA	Data MOdel for Call Center Analysis
Den DF	Denominator Degrees of Freedom
DF	Degrees of Freedom
ILCellular	Israeli Cellular phone company
LR	Likelihood Ratio
MA	Moving Average
MCMC	Markov Chain Monte Carlo
Num DF	Numerator Degrees of Freedom
ODA	One Day Ahead
QED	Quality-Efficiency Driven
RMSE	Root Mean Squared Error
US	United States (of America)
USBank	US Bank
VRU	Voice Response Unit
Width	Prediction interval width

Chapter 1

Introduction

Many companies today invest large amounts of resources in order to provide full customer service, with much or all of the customer interaction based on telephone or internet access. Contact and Call centers provide a direct contact point between companies and their customers, thus making them especially important in the battle for market share. These contact points accumulate large amounts of data that can later be analyzed and utilized for short term operational decisions, medium term managerial decisions or for long term tactical, strategic decisions.

The manager of a call center faces the classical scheduling problem of determining the number of service agents that offers the best trade-off between maintaining high service levels and low operating costs. Forecasting higher call loads than will actually be realized will cause overstaffing, leading to unnecessary costs. On the other hand, forecasting a lower number of incoming calls may result in long periods of waiting and high abandonment rates, which eventually could lead to a loss of customers and revenues. Therefore, modelling the load arrival process is the first and basic step of the scheduling problem. Forecasting the system load requires the knowledge of two components: the arrival process and the service time distribution.

With the help of recent advances, we now understand how to choose appropriate staffing levels, given sufficiently accurate predictions of the load, in order to balance service quality and efficiency. This leads one to the so-called QED (Quality-Efficiency Driven) regime. The basic concepts of this regime have been developed in both the M/M/N queue, usually referred to as Erlang-C, and in the M/M/N+M queue, usually referred to as Erlang-A. (The latter notation denotes an M/M/N queue with customers whose patience has an exponential distribution.) The approach is relevant when there is a high customer arrival

rate per unit time (λ), the service rate of customers served per unit time per agent (μ) is fixed and the number of agents (S) is a function of the offered load ($R = \lambda/\mu$). For maintaining high service levels while preserving high offered load per server ($\rho = R/S$), the QED regime prescribes that, for some constant β, the number of scheduled agents should equal $S = R + \beta \cdot \sqrt{R}$ (β is positive for Erlang-C and can take either sign, as well as being zero, for Erlang-A). We call this prescription "square-root staffing".

Following square-root staffing, the requirements for prediction accuracy in the QED regime is that the estimated $\hat{\lambda}$ (arrival rate) must not exceed a square-root deviation from the real arrival rate. Otherwise the QED rule will not be effective since, with a larger error, it will either grossly over- or under-estimate the needed number of agents.

One of the questions that we consider in this thesis is whether in real systems, one can achieve the prediction accuracy required in order to operate in the QED regime. For an extensive review on call centers and the QED regime one can refer to [8]. A detailed bibliography detailing further literature on this subject and other call centers related papers is described by Mandelbaum in [12].

In recent years, several different modelling techniques have been suggested as alternatives for forecasting the arrival process. These alternatives range from classical ARMA (see [23]) to complex Bayesian models (see for example [22]).

In this thesis we introduce two arrival count models which are based on a *mixed* Poisson process approach. This approach stipulates that the arrival counts follow a Poisson distribution and that the arrival rate, λ, is itself a stochastic process. The additional variability created by the arrival rate creates a mechanism that can account for the well known *over-dipersion* phenomena often encountered in call center arrival counts data. Moreover, this approach also allows for introducing correlations between different time intervals.

In the first model we use Gaussian linear mixed model formulations to describe a suitably transformed version of the arrival process. Mixed model techniques allow us the much needed flexibility to describe different seasonality effects using correlation structures. Motivated by an Israeli cellular phone company forecast procedure, we evaluate our model's results using six weeks of past data as the learning data and producing a ten-day-ahead weekly forecast for each week.

As previously mentioned, the scheduling problem requires the system load prediction which means that forecasts for the service rate should also be provided (in addition to the arrival process predictions). We introduce a simple average service time forecasting model based on the weekday and different daily period effects. Based on the results

of both the mixed arrival process model and the average service time predictions, we introduce a new measure to evaluate these load predictions. This measure directly reflects the extent to which the system's operational quality and efficiency goals are achieved when using the forecasts.

The mixed model results are also compared to results for a similar Bayesian model presented in [22] and applied to data from a US Bank.

Our second model employs Bayesian techniques to produce the arrival count predictions. By employing Gibbs sampling techniques, we produce the forecast distributions for the arrival counts. Due to computational limitations we only show a 'proof of concept' for this model by applying it to predicting a single day's arrivals. Further work on this approach is required in order to make it a computationally practical alternative.

The mixed model developed in this paper is relatively simple to implement using standard softwares such as SAS®. From a practical prospective, the model is very flexible and can be adapted to different period lengths (or *resolutions*) as well as various lead times (the time between the last learning data and the first forecasted time). This model provides good precision when compared to similar models. From a managerial prospective, we also conclude that the lead time has a significant effect on prediction precision: generally, shorter lead times are better. Hence, we advise a two stage weekly forecasting process where, except for the first day of the week, a one-day-ahead forecast is used to update subsequent forecasts.

As for the question of obtaining the desired level of the QED regime precision for load predictions, we conclude that during most of the day these levels can be maintained. During early morning hours, the algorithm precision is insufficient. However, this fact makes sense since the QED regime staffing rule is also inadequate during these hours.

The outline of this thesis is as follows. In Chapter 2 we review past and recent studies that have been conducted on call center arrival processes. In Chapter 3 we describe two different sources of call center data that are later used to illustrate our methodologies. In Chapter 4 we describe performance measures for comparing different forecast methods. Chapter 5 describes three different arrival counts prediction models. We describe both our mixed model and our Bayesian model in sections 5.1 and 5.3, respectively. The third model, briefly described in Section 5.2, was developed by Weinberg, Brown and Stroud and is more fully described in [22]. An additional model, considered in Section 5.4, is the average service time forecasting model.

In Chapter 6, for the arrival process mixed models, we explain the comprehensive process

of determining the fixed and random effects for the two different data sets. Comparisons between the different models and analyses of their results are discussed in Chapter 7. Conclusions are presented in Chapter 8.

Chapter 2

Literature Review

In recent years, several documented studies of the incoming call arrival process have been conducted and tested thanks to technology advances in the call center industry. Earlier studies focused on classical Box and Jenkins, Auto-Regressive-Moving-Average (ARMA) models such as the Fedex company study [23]. A well-known study, also employing Auto-Regressive-Integrated-Moving-Average (ARIMA) models techniques, was carried out by B.Andrews and S.Cunningham (see [1]) to produce L.L.Beans call center daily forecasts. The study focuses on modelling two different arrival queues each with its own characteristics. Their models incorporated exogenous variables along-side the MA (Moving Average) and AR (Auto-Regressive) variables, using transfer functions to help predict outliers such as holidays and special sales promotion periods.

A slightly different approach was described in an article published by Antipov and Meade [2]. In this article the authors tackle the problem of including advertising response and special calender effects by adding these variables in a multiplicative manner.

A more recent study was carried out by Taylor [18]. In this study several different time series models were investigated on two different sources of data. Among these models were: seasonal ARMA models; exponential smoothing for double seasonality methods; and dynamic harmonic regression. His results indicated that for short term forecasting horizons, the exponential smoothing for double seasonality method performs quite well but for practical horizons (longer than one day) a very basic averaging model outperforms all of the suggested alternatives. One of the author's conclusions is that a more useful picture could be obtained if both the prediction interval, defined in a previous article (see [6]) by Chatfield, and the predictions density would be incorporated together with the point estimates. In this study we follow the author's suggestions as we are particulary

interested in the precision of our estimates with respect to implementing the "square-root staffing" rule.
Recent empirical work has revealed several important characteristics that underly the arrival process:

1. the arrival rate changes over the course of a day;

2. the arrival counts exhibit a common phenomena called *over-dispersion*. Over-dispersion means that the call volume data sometimes show a variance that substantially dominates the mean value, contradicting the assumption that the data is generated by a simple Poisson distribution. A mechanism that accounts for this phenomena was suggested by Jongbloed and Koole in [9]. They proposed the Poisson mixture model which incorporates a stochastic arrival rate process to generate the additional variability;

3. there is a significant dependency between arrival counts on successive days. In [5] Brown *et al.* suggest an arrival forecasting model which incorporates a random daily variable that has an autoregressive structure to explain the intra-day correlations;

4. successive periods within the same day exhibit strong correlations. This correlation was empirically analyzed in an article by Avramidis *et al.* (see [3]).

In the latter article three models were suggested that take account of these correlations. The first two are different versions of the mixed Poisson model: (a) the first assumes that the arrival rate has the following form $\Lambda(t) = W \cdot f(t)$ where $W \sim \text{Gamma}(\gamma, 1)$. Here $f(t)$ characterizes the time variation of the arrival rate over a day (yielding a negative multinomial distribution of the arrival count vector); (b) the second assumes that the arrival counts vector has a compound negative multinomial distribution, which generalizes the negative multinomial distribution by allowing the parameters to be randomly distributed according to a Dirichlet distribution. The third model assumes a more general structure in that the daily volume, Y, is randomly distributed according to a general distribution G. It incorporates a vector, Q, of the proportions of daily demand allocated to the different K periods. It assumes that Q is independent of Y and is distributed according to a Dirichlet distribution. The vector of observed arrival counts, X, is obtained by rounding up each element of the product of Y and Q. However, in this article the authors do not tackle the intra-daily correlations presented in 3.

In recent years technology has allowed researchers to employ advanced Bayesian techniques to this type of forecasting problem. These techniques include Markov Chain Monte Carlo sampling mechanisms such as the Gibbs sampling algorithm. These algorithms produce the forecasted arrival rate and the arrival counts distributions and so give more information then just the point estimates.

An example of such a study was conducted by Soyer and Tarimcilar [16]. In the article the authors analyzed the effect of marketing strategies on call arrivals. Their Bayesian analysis is based on the Poisson distribution of arrivals over (possibly varying) time periods measured in days, with cumulative rate function of the form:

$$\Lambda_i(t) = \Lambda_0(t) \exp(\beta' \mathbf{Z}_i) \tag{2.1}$$

where i denotes an advertising campaign, with its covariate vector $\mathbf{Z}_i$, and

$$\Lambda_0(t) = \gamma t^{\alpha} \quad . \tag{2.2}$$

The parameters (α, γ, β) are given a prior distribution, and the posterior distribution of these parameters is then discussed. In a random effect or mixed model approach, they allow γ to have a random component by modelling:

$$\log(\gamma) = \theta + \phi_i \tag{2.3}$$

where the ϕ_i are iid N$(0, 1/\tau)$, and the precision τ has a Gamma prior. By considering the DIC statistic, they conclude that the random effects model fits much better than the fixed effects model and conclude that the data cannot be adequately described by assuming a model that explains arrivals solely using the $\mathbf{Z}_i$ information without some additional random variability. The mixture model also provides the within advertisement correlations over different time periods.

Another interesting paper, modelling incoming call arrivals to the US Bank call center used here, also employing Bayesian techniques, was written by Weinberg *et al.* [22]. In this paper the authors use the Normal-Poisson stabilization transformation to transform the Poisson arrival counts to normal variables. The normally transformed observations allowed them the necessary flexibility to incorporate conjugate multivariate normal priors with a wide variety of covariance structures. The authors provide a detailed description of both the one-day-ahead forecast and within-day learning algorithms. Both algorithms use Gibbs sampling techniques and Metropolis-Hastings steps to sample from the forecast distributions. The model and its results will be further discussed in sections 5.2 and 7.2.1 of this thesis.

Chapter 3

The Data

Our datasets originated from an ongoing basic research project called Data-MOCCA (Data MOdel for Call Center Analysis), conducted by the Technion's Statistics Laboratory.(For more information on the Data-MOCCA project see [21].) Data-MOCCA's databases contain detailed call-by-call histories obtained from several different call centers. This study will focus on data from two different call centers belonging to: an Israeli cellular phone company; and a North American commercial bank. The next two sections describe these databases.

3.1 The Israeli Cellular Phone Company Data

The Technion's collaboration with the cellular company, that began at the end of 2003, is providing a monthly updated database which preserves call histories dating from January 2004.

The call center handles calls from several main queues: Private clients; Business clients; Technical Support problems; Foreign languages queues; and a few minor queues. In general, queues are operated by different service provider groups. Almost 30% of incoming calls enter the Private customers queue which is operated by a dedicated team of 150 telephone agents. The load generated from each of the remaining queues is much smaller (for example, the second largest queue is the Business queue and it generates 18% of the overall incoming load). Hence, we shall limit our discussion to modelling the Private queue (and bear in mind that our model techniques can be applied to the other queues as well).

The Private queue's call center operates six days a week, closing only on Saturdays and Jewish holidays. On regular weekdays, operating hours are between 7AM and 11PM and

on Fridays it closes earlier, at around 4PM.

We divide each day into half-hour intervals. There are two alternative justifications for choosing a half-hour analysis resolution: (a) currently shifts scheduling is carried out at this resolution; and (b) from a computational complexity point of view taking shorter intervals significantly increases the computing time for many models and may make their implementation completely impractical. Another justification comes from analyzing our mixed model (detailed later on) behavior under different resolutions, where the half-hour resolution exhibits "good enough" behavior. This analysis is fully detailed in Appendix A. Consequently, we consider for each day 33 half-hourly arrival intervals between 7AM and 11:30PM.

Note that if the arrival rate was *very* inhomogeneous during a particular half-hour interval, then using the average arrival rate could lead to under-staffing. Specifically, the staff level assigned to meet the average load would not be able to cope with the peak load in that particular half-hour interval. We do basically assume in the sequel that the within interval inhomogeneity is mild.

The learning stage of the model is based on the arrivals between mid-February, 2004 and December 31, 2004.

Figure 3.1 demonstrates the weekly pattern which occurs between April, 2004 and September, 2004. By examining the above graph one can reach several conclusions: Sundays and Mondays have the highest arrival counts; the number of arrivals gradually decreases over the week until it reaches its lowest point on Fridays; there are quite a few outliers which occur in April.

Examination of outlying observations singles out twenty-two days with strange arrival counts. Among these days were the holidays listed in Table 3.1, which exhibit different daily patterns and unusual daily volumes when compared to similar regular weekdays. The additional five days that were assigned to the set of outliers are detailed in Table 3.2. As mentioned earlier, April 2004 has an unusual weekly pattern. Out of the list of twenty-two outliers, nine occur in April which explains the peculiar pattern that we saw in Figure 3.1. Among these nine outliers is the Passover holiday, Memorial day and Independence day. During these holidays there was a lower arrival counts than on similar non-holiday weekdays. Another event that took place on April is the country-wide change in the first three cellular digits. During that day and the previous day there was a significant increase in the arrival counts to the cellular company's call center.

In conclusion, the outlying days were excluded from the learning stage of our model but

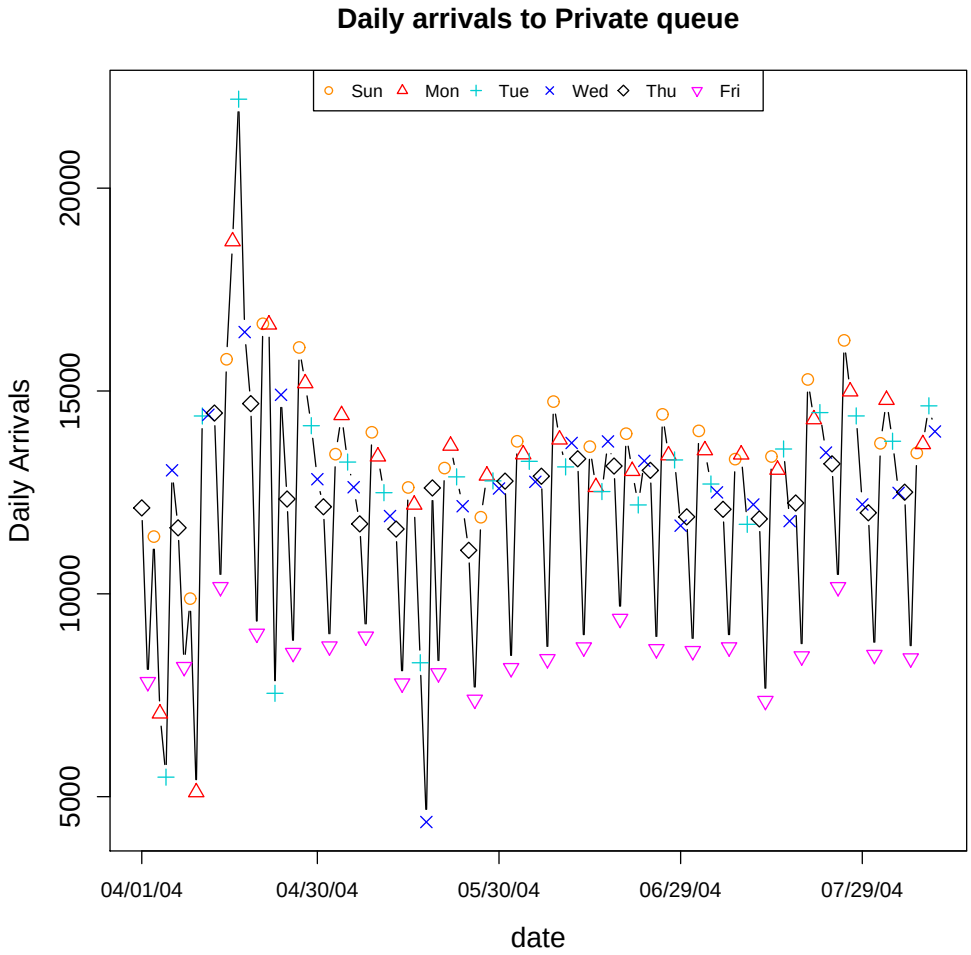

Figure 3.1: Daily arrivals to the Private queue between April 1st, 2004 and September 1st, 2004 including holidays.

kept for later evaluation purposes.

Study of intra-day arrival patterns for the regular days reveals some interesting characteristics.

- The weekdays, Monday through Thursday have a similar pattern. Figure 3.3 illustrates the last fact by depicting the *normalized* weekday patterns — each half-hour is divided by the mean half-hour arrival rate for that day, and the normalized values for corresponding weekdays are averaged. There are two major peaks during the day: one at around 2PM and the higher one at around 7PM. The higher peak occurs probably due to the fact that people finish working at around this hour and so are free to phone the call center. From 7PM there is a gradual decrease (except for a small increase at around 9PM).

Holiday Name	Date	Day of week
First Passover Eve	04/05/2004	Monday
First day of Passover	04/06/2004	Tuesday
Last Eve of Passover	04/11/2004	Sunday
Last day of Passover	04/12/2004	Monday
Memorial (day's) eve	04/25/2004	Sunday
Memorial day	04/26/2004	Monday
Independence day	04/27/2004	Tuesday
Eve of Feast of Weeks	05/25/2004	Tuesday
Feast of Weeks	05/26/2004	Wednesday
New Year's eve	09/15/2004	Wednesday
New Year's day	09/16/2004	Thursday
The Second day of New Year	09/17/2004	Friday
Yom Kippur's eve	09/24/2004	Friday
Eve of Feast of Tabernacles	09/29/2004	Wednesday
Feast of Tabernacles	09/30/2004	Thursday
Sixth day of Tabernacles	10/06/2004	Wednesday
Simchat Torah	10/07/2004	Thursday

Table 3.1: Holidays — this list contains all the holidays

Irregular Date	Day of week	Event Description
03/01/04	Wednesday	An unexplained irregular day.
04/19/04	Monday	A day before a country-wide change in the first three cellular digits.
04/20/04	Tuesday	The day of a country-wide change in the first three cellular digits.
08/22/04	Sunday	An unexplained irregular day.
10/03/04	Sunday	The third day of Tabernacles which comes after a long weekend.

Table 3.2: Irregular days, 2004

- Fridays have a completely different pattern from the rest of the weekdays. This can be seen in Figure 3.2. For each day of the week, the 33 arrivals were smoothed using the default smoothing method in R statistical analysis software [19]. Because Friday is a half work day for most people in Israel it is very reasonable for its daily pattern to differ from the rest of the weekdays.

- Sunday's pattern also differs from the rest of the weekdays. Figure 3.3 exhibits how Sunday has an earlier increase than the other weekdays (Monday through Thursday), possibly as a result of customers who were not able able to contact the call center on the weekend (Saturday).

The cellular company's major complaint regarding their current forecasting algorithm is that it does not incorporate billing cycles effects. Their own experience leads them to believe that on billing days the number of incoming calls is higher than on non-billing days. There are six billing cycles each month. Customers are assigned to one of the cycles when they purchase a service contract. Table 3.3 summarizes the distribution of the billing cycles among the Private queue customers. From these results it is clear that cycles 10 and 17 are negligible. These two billing cycles main customers are the company's employees which can account for these results. Hence, we focus our attention on the remaining four cycles.

Each billing cycle is defined according to two periods: the delivery period — prior to the bank billing day, each customer receives a letter detailing his cellular expenses; the billing period — the day on which the customer's bank account will be debited. The delivery period extends over two working days (depending solely on the Israeli postal services). The billing period usually covers only one day. There is usually a full week between the delivery period and the billing period but this can vary due to weekends and

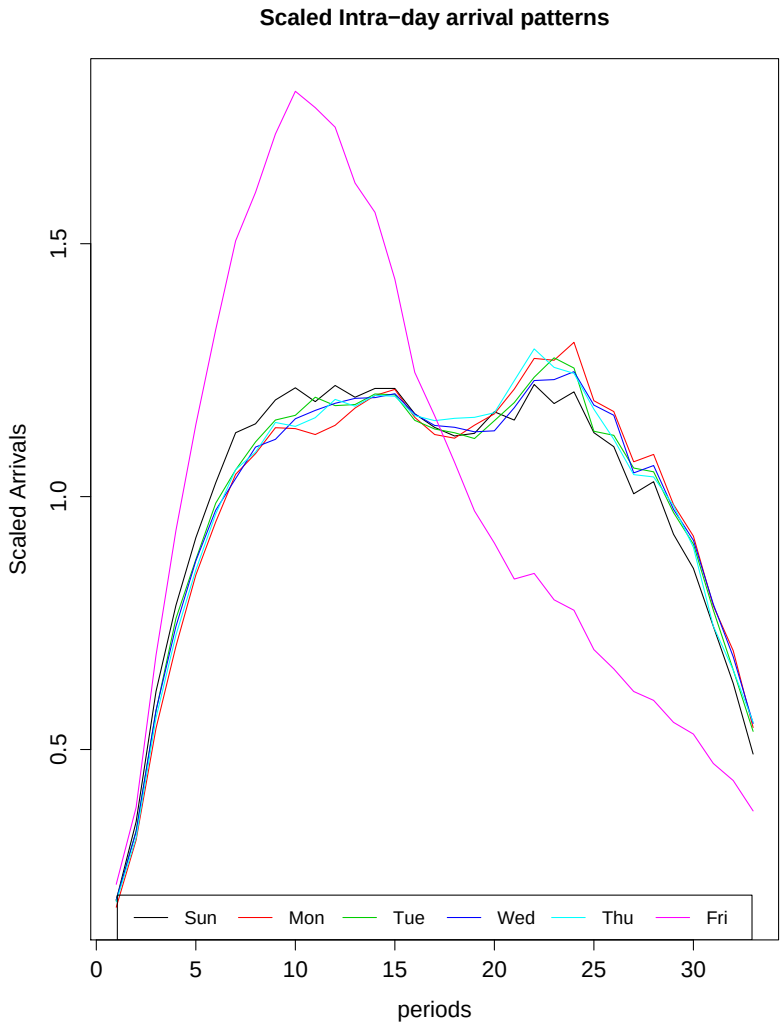

Figure 3.2: Normalized intra-day arrival patterns

holidays. We decided therefore to describe *each* cycle using two indicators: a delivery indicator - marking the two working days of the delivery period; and a billing indicator - marking the first and second day of the billing period and zero otherwise. By describing each cycle using two indicators we actually differentiate between the influence of the actual billing date and that of the delivery of the bill. According to the cellular company's past experience, the different queues are affected by different periods. For example, the Private queue is strongly affected by the delivery periods and not so much by the billing periods. On the other hand, the Finance queue is strongly affected by the billing periods and less by the delivery periods. Section 6.2.1 demonstrates how we examined which indicators are significant for the Private queue's arrival process.

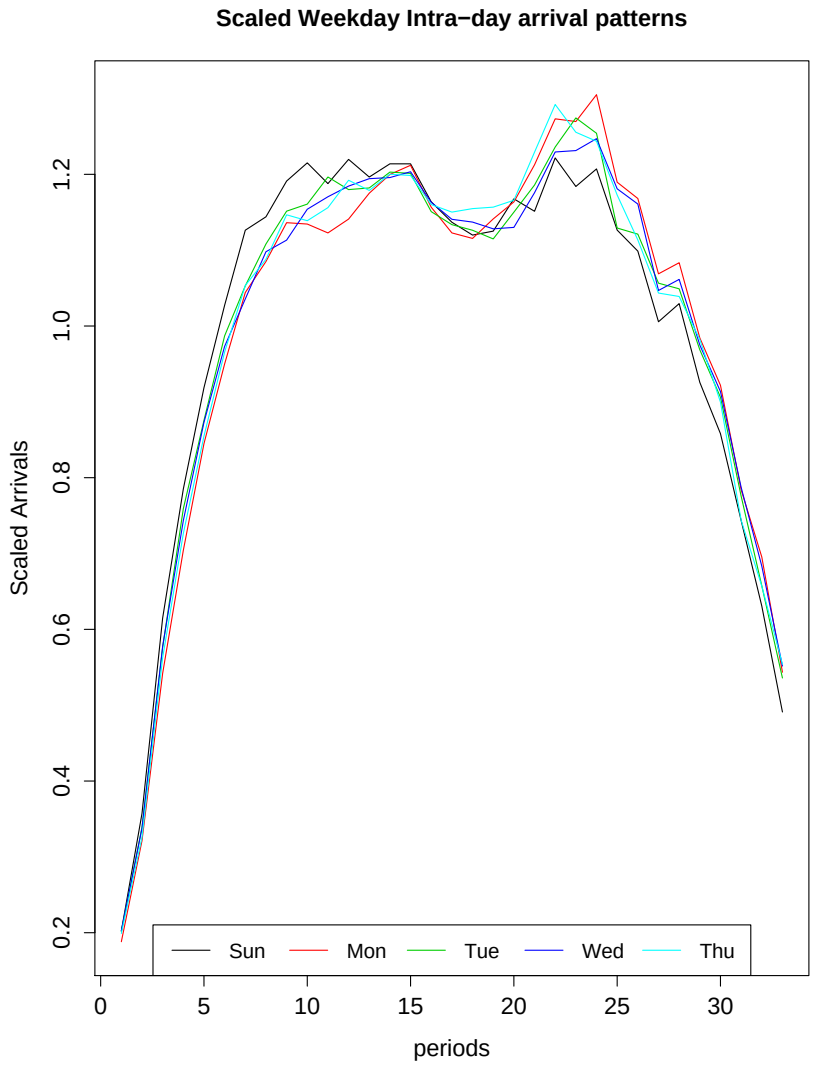

Figure 3.3: Scaled weekdays intra-day arrival patterns

Billing Cycle name	Proportion of Customers
1	0.31
7	0.27
10	0.00
14	0.26
17	0.00
21	0.16

Table 3.3: The Private queue customer distribution over billing cycles.

3.2 The US Bank data

DataMOCCA preserves all call-by-call data originating from the US Bank during 2002-2003. This call center is very large and handles approximately 300,000 calls each day. It has a voice response unit (VRU) which greets each call upon entrance. Only 20% of entering calls advance to be handled by a human service provider. Since this call center also provides various types of services, we focus on the largest queue which is the Retail service. The calls entering this queue account for approximately 68% of all the incoming calls which require human-agent services. Our research database contains calls arriving between March 3 and October 24, 2003.

Since one of our goals is to compare our model with the results of the model described in [22] we generated the same database. This means that we concentrate our attention on calls generated during weekdays between 7AM and 9:05PM (the most active periods of the call center). Each day is divided into 169 five-minutes intervals. We assume that during each interval the arrival rate remains relatively constant. We point out here that we do not think that this high resolution is required for practical scheduling purposes. Currently, five minute intervals are both impractical from a managerial point of view and also the computations are time consuming.

Between March 3 and October 24 there are only four holidays: 1. May 26 Memorial Day; 2. Jul 4 Independence Day; 3. Sep 1 Labor Day; 4. Oct 13 Columbus Day. We removed these days from our database since they exhibit irregular patterns and daily volumes, compared to similar weekdays. In the previously mentioned article the authors indicate that the day after Labor day depicts an unusual pattern. This day is a Tuesday but because it has a similar pattern to a Monday and a peculiar high volume they have decided to model it as a Monday. Following this same reasoning, we identified another abnormal Tuesday which takes place a day after Columbus day. We modelled both these days as Mondays.

Studying the daily weekday patterns two interesting characteristics appear.

- The weekdays, Tuesday through Thursday have a similar pattern. Figure 3.4 illustrates the last fact by depicting the *normalized* weekday patterns — each five-minute interval is divided by the mean five-minute arrival rate for that day, and the normalized values for corresponding weekdays are averaged. There is a major peak during the morning hours followed by a slow decrease until 5PM. Afterwards we see a sharper decrease in the patterns.

- Fridays and Mondays have patterns that differ from the rest of the weekdays. Monday has a lower starting point compared to the rest of the weekdays. One explanation can be that Monday is the first working day of the week and so customers begin their day later. Friday has a lower tail, as also observed in Figure 3.4. This fact may not be surprising because people want to finish their business before the weekend (Saturdays and Sundays).

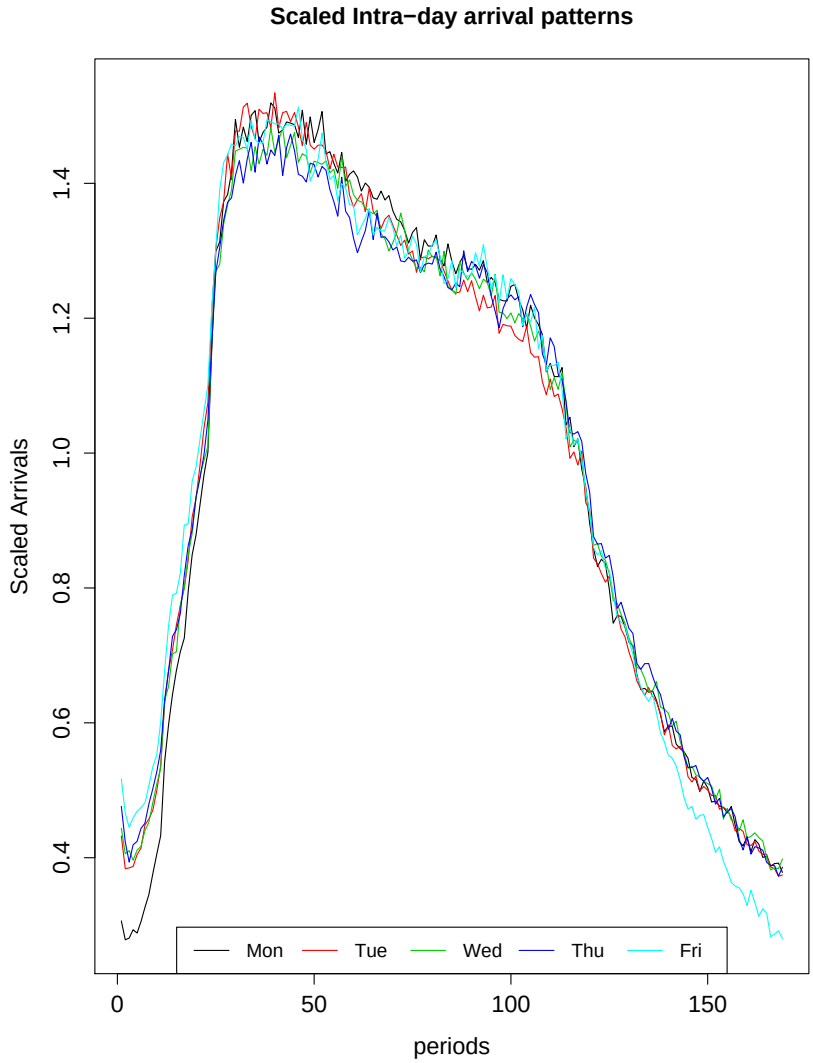

Figure 3.4: The US bank normalized intra-day arrival patterns

Chapter 4

Evaluation of Models

The Israeli cellular company utilizes the arrivals forecast for determining its call centers weekly staffing schedules. Each Thursday, using the past six weeks data as the learning data, it predicts the week starting ten-days ahead. We will refer to this forecasting strategy as the ten-day-ahead weekly predictions. Accordingly, we define three periods: the learning period; the prediction lead time which is the duration between the last learning day and the first predicted day; and the forecast period.

In the Israeli Cellular data we implement the same strategy as the company, i.e. we predict the arrivals to the Private queue for each week between April 11, 2004 and December 25, 2004 (37 weeks which are $D = 37 \cdot 6 = 222$ days). The forecasting procedure is carried out 37 times since there are 37 weeks. For each of the 6 weekdays, we predict the arrivals for the $K = 33$ half-hour intervals between 7AM and 11:30PM using six weeks of learning data and a lead time of ten days. All together we have a total of $n = 37 \cdot 6 \cdot 33 = 7326$ predicted values. Excluding the 19 irregular days (which occur during the mentioned period) we evaluate the results using a total of 203 days or $203 \cdot 33 = 6699$ observations.

In the US Bank data we take a slightly different approach. Emulating the same procedure carried out by the authors of [22] we generate one-day-ahead predictions. Using our own notation: the learning period is five weeks because of computational limitations (since each day has $K = 169$ intervals); the prediction lead time period is set to zero; the forecast period consists of one day. We predict the arrivals to the Retail queue for the $D = 64$ days between July 25 and October 24, 2003. As previously mentioned, each day consists of 169 five-minute intervals between 7AM and 9:05PM. In conclusion, we have a total of $n = 169 \cdot 64 = 10816$ predicted values.

4.1 Prediction Accuracy

Here are a few basic definitions:

- The subscript $d = 1, \ldots, D$ denotes the d^{th} day in the predicted data set.
- The subscript $k = 1, \ldots, K$ denotes the k^{th} period during a day.
- n, the number of predicted values, equals the product of D and K, i.e. $n = D \cdot K$.

Let $\hat{N}_{dk}$ denote the predicted value of N_{dk}, which is the number of arrivals in the k^{th} period for day d. We define two measures to compare between the observed and the predicted values:

- The Squared Error: $\text{SE}_{dk} = (\hat{N}_{dk} - N_{dk})^2$.
- The Relative Error: $\text{RE}_{dk} = 100 \cdot \frac{|\hat{N}_{dk} - N_{dk}|}{N_{dk}}$

The following two measures are used to evaluate confidence statements concerning N_{dk}:

- Cover_{dk} = $\mathbf{I}(N_{dk} \in (\text{Lower}_{dk}, \text{Upper}_{dk}))$
- Width_{dk} = Upper_{dk} - Lower_{dk}

In the above, Lower_{dk} and Upper_{dk} denote the lower and upper (nominally 95%) confidence limits.
The comparison between different forecasting models is performed over the entire set of n observations. We first average the measures for each day over the K periods:

- $\text{RMSE}_d = \sqrt{\frac{\sum_{k=1}^{K} \text{SE}_{dk}}{K}}$
- $\text{APE}_d = \frac{\sum_{k=1}^{K} \text{RE}_{dk}}{K}$
- $\text{Cover}_d = \frac{\sum_{k=1}^{K} \text{Cover}_{dk}}{K}$
- $\text{Width}_d = \frac{\sum_{k=1}^{K} \text{Width}_{dk}}{K}$

Alternatively, the basic performance statistics can be averaged over days for each of the K periods, in order to consider accuracy and precision by period of the day.
The summary statistics which are reported for each measure include the lower quartile, the median, the mean and the upper quartile values of these daily summary statistics.

4.2 Goodness of Fit

Our model, which will be introduced later on, stipulates through (5.1) that during each interval of each day the square-root of the arrival counts is a function of two elements: θ_{dk} which we can model and the inherent random error term ϵ_{dk}. Following the Normal-Poisson stabilizing transformation (which will also be described in the sequel), we assume that ϵ_{dk} follows a Normal distribution with zero expected value and a variance equal to 0.25.

We will explore both the normality assumption and the values of residual variances in order to evaluate the goodness of fit of our model.

Chapter 5

Prediction Models

5.1 Gaussian Mixed Model for Arrival Counts

In this subsection we present a forecast model which is based on the mixed linear modelling (MLM) theory. Figures 3.2 and 3.4 clearly show that the assumption of a homogeneous Poisson arrival process does not hold for our data. However, as assumed above and explained in Appendix A, the arrival process during relatively short intervals can be treated as being homogeneous. The MLM approach allows us the needed flexibility to both model different arrival rates for different intervals, as well as to incorporate a random component in the variation of those interval-specific arrival rates. This extra randomness will help account for the observed *over-dispersion* when looking at the variation in arrivals for a given period over similar weekdays. For more on mixed models, the reader is referred to [7].

5.1.1 Definition

Let N_{dk} denote the number of arrivals to the queue on day $d = 1, ..., D$ and during the time interval $[t_{k-1}, t_k)$ where $k = 1, ..., K$ is the k^{th} period of the day. Our basic model assumption is that N_{dk} follows a Poisson distribution, with expected value (λ_{dk}). We follow the work of Brown *et al.* [4] and take advantage of the variance stabilizing transformation for Poisson data in the following manner. If $N_{dk} \sim$ Poisson(λ_{dk}), then $y_{dk} = \sqrt{N_{dk} + \frac{1}{4}}$ has approximately a mean value of $\sqrt{\lambda_{dk}}$ and variance $\frac{1}{4}$.
For $\lambda \to \infty$, y_{dk} is approximately normally distributed. In our data sets, for most parts of the day, λ_{dk} has high values: either around 300 per five minutes for the US Bank's data; or

around 500 per half-hour for the Israeli cellular phone company's data. Hence, it seems reasonable to use this approximation for our modelling.

The transformed observations y_{dk} allow us to exploit the benefits of the linear mixed modelling approach. Our aim is to model the expected value of these observations (since the variance is known). The expected value is a function of the relevant interval rate itself. In the mixed model the square-root of the arrival rate ($\sqrt{\lambda_{dk}}$) is regarded as a linear function of both fixed and random effects.

The fixed effects include the weekday effects and the interaction between them and the period effects. These two effects express the weekday differences in the daily levels and the intra-day profiles (over the different periods). In the Israeli cellular company we also add exogenous variables to these fixed effects (i.e., the billing cycles explanatory variables).

The random effects are normal deviates with a pre-specified covariance structure. One random effect is the daily volume deviation from the fixed weekday effect. In concert with other modelling attempts, a first-order autoregressive covariance structure (over successive days) has been considered for this daily deviation. It involves the estimation of one variance parameter and one autocorrelation parameter. The other random effects are also called the noise or residual effects, and refer to the period-by-period random deviations from the values after accounting for the fixed weekday and period effects. We considered a few different covariance structures that can describe a reasonable relationship between the periods, such as an AR(1) structure.

The general formulation of our linear mixed model can be written as:

$$
\begin{aligned}
y_{dk} = \quad & \theta_{dk} + \epsilon_{dk} && (5.1)\\
\theta_{dk} = \quad & V_d + \alpha_{q_d} + p_{q_d,k} + \beta_d + \eta_{dk} && (5.2)\\
\text{with} \quad & \epsilon_{dk} \sim N(0, \frac{1}{4}) \quad \text{i.i.d.},\\
& (V_1, \ldots, V_D)^T \sim N_D(\vec{0}, G) \quad \text{and} \quad \eta_d = (\eta_{d1}, \ldots, \eta_{dK})^T \sim N_K(\vec{0}, R)
\end{aligned}
$$

where

- V_d is the random daily volume effect that has the first-order autoregressive structure. The vector $V = (V_1, \ldots, V_D)^T$ represents the vector of daily volume random effects and is assumed to follow a D-variate normal distribution with zero expected value and covariance matrix G.
- q_d denotes the weekday (Sun, Mon, ...) corresponding to day d.

- α_q is the qth weekday fixed effect ($q = 1, \ldots, Q;\ Q = 5$ or 6 or 7).
- $p_{q,k}$ is the fixed (interaction) effect for period k of the q^{th} weekday;
- Define π_d as the set of all relevant billing cycle effects that occur on the d^{th} day. Then $\beta_d = \sum_{j \in \pi_d} \xi_j$ where ξ_j indicates the j^{th} billing cycle effect. (The set of possible ξ_j is determined by the appropriate model setup);
- $\eta_d = (\eta_{d1}, \ldots, \eta_{dK})^T$ is the within day vector of errors. $\eta_1, \ldots, \eta_D$ is an i.i.d sequence of K-variate normal with zero mean and covariance matrix R.

5.1.2 Estimation Method

The additional random effects complicate matters and so we cannot simply use linear regression model techniques. Instead, we exploit the normality assumptions of these random effects to obtain maximum likelihood estimators for the random effects covariance matrices.

Define T as the covariance matrix of Y. We can write the elements of T explicitly using only elements from the R and G matrices in the following manner (and assuming the AR(1) structure for G):

$$\begin{aligned}
\text{cov}(y_{dk}, y_{dk}) &= G_{d,d} + R_{k,k} + \frac{1}{4} = \sigma_V^2 + \sigma_\eta^2 + \frac{1}{4};\ k = 1, \ldots, K;\ d = 1, \ldots, D \\
\text{cov}(y_{di}, y_{dj}) &= G_{d,d} + R_{i,j} = \sigma_V^2 + \text{cov}(\eta_{di}, \eta_{dj});\ i \neq j;\ d = 1, \ldots, D \\
\text{cov}(y_{mi}, y_{tj}) &= G_{m,t} = \sigma_V^2 \cdot \rho_V^{d(t,m)};\ i, j = 1, \ldots, K;\ m \neq t
\end{aligned}$$

where $d(t, m)$ is the number of days between the t^{th} and m^{th} days.

Recognizing this last fact, one can rewrite the log-likelihood function for the transformed observations Y in the following manner:

$$l = -\frac{1}{2} \log |T| - \frac{1}{2} r' T^{-1} r - \frac{n}{2} \log 2\pi \tag{5.3}$$

where $r = Y - X(X'T^{-1}X)^{-1}X'T^{-1}Y$ and the matrix X (the fixed effect matrix) is of rank p. Of course, in order to use these equations we assume that all the necessary matrices are nonsingular (otherwise we need to use generalized inverse matrices). For models such as these, we can estimate the fixed effects and estimate R and G by minimizing twice the negative of the log-likelihood using methods such as the SAS® *Mixed*

procedure (the code for implementing a model of this type is presented in Appendix B.1 — see Section 5.1.6 for a discussion of the choice of implementable models). This SAS® procedure uses a ridge-stabilized Newton-Raphson algorithm to search for the maximum likelihood estimators.

After obtaining the estimators for the covariance matrices of the random effects, SAS produces the vector of the fixed-effects estimators using the following formula:

$$\hat{\beta} = (X'\hat{T}^{-1}X)^{-1}X'\hat{T}^{-1}Y \tag{5.4}$$

5.1.3 Prediction Method

Based on the normality assumption for the random effects, SAS® uses multivariate normal conditional expectations to obtain the empirical best linear unbiased predictors (BLUPs). Using the fixed effect explanatory matrix, X_m, for the data to be predicted together with the past data matrices, Y and X, the prediction vector, $\hat{m}$, can be obtained using the following formula:

$$\hat{m} = X_m\hat{\beta} + \hat{C}_m\hat{T}^{-1}(Y - X\hat{\beta}) \tag{5.5}$$

where $\hat{T}$ is the maximum likelihood estimate of the covariance matrix; $\hat{\beta}$ is the maximum likelihood estimate of the fixed-effects coefficients defined in (5.4); and $\hat{C_m}$ is the model-based estimated covariance matrix between the observed Y and m.

The estimated prediction variance can be obtained as follows:

$$\begin{aligned} \hat{\text{Var}}(\hat{m} - m) &= \hat{T}_m - \hat{C}_m\hat{T}^{-1}\hat{C}'_m \\ &+ [X_m - \hat{C}_m\hat{T}^{-1}X](X'\hat{T}^{-1}X)^{-1}[X_m - \hat{C}_m\hat{T}^{-1}X]' \end{aligned} \tag{5.6}$$

where $\hat{T}_m$ is the estimated model-based covariance matrix for the predicted observations.

5.1.4 Goodness of Fit

Following (5.1) and (5.2) we may check to see whether our prediction residuals are normally distributed. We use a QQ-plot to examine this assumption.

Furthermore, we may examine the estimated variance of ϵ_{dk}. According to the model presented in (5.1) and (5.2) its value should be approximately 0.25.

Assuming that both random effects covariance matrices, R and G have an AR(1) structure we can formulate them in the following manner:

- The daily random effects covariance matrix G:

$$G = \sigma_V^2 \cdot \begin{pmatrix} 1 & \rho_V^{d_{12}} & \dots & \rho_V^{d_{1D}} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_V^{d_{D1}} & \rho_V^{d_{D2}} & \dots & 1 \end{pmatrix}$$

- The within-day periods random effects covariance matrix R for a specific day:

$$R = \sigma_\eta^2 \cdot \begin{pmatrix} 1 & \rho_\eta^{1} & \dots & \rho_\eta^{K-1} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_\eta^{K-1} & \rho_\eta^{K-2} & \dots & 1 \end{pmatrix}$$

Using this notation it is easy to see that the variance of each observation is: Var$(y_{dk}) = \sigma_V^2 + \sigma_\eta^2 + \frac{1}{4}$. Where 0.25 corresponds to the value of ϵ's variance. We shall examine an alternative model which has the following representation:

$$\begin{aligned} y_{dk} &= \theta_{dk} + \epsilon_{dk} && (5.7) \\ \theta_{dk} &= V_d + \alpha_{q_d} + p_{q_d,k} + \beta_d + \eta_{dk} && (5.8) \\ \text{with} \quad & \epsilon_{dk} \sim N(0, \sigma_\epsilon^{\mathbf{2}}) \quad \text{i.i.d.}, \\ & (V_1, \dots, V_D)^T \sim N_D(\vec{0}, G) \quad \text{and} \quad \eta_d = (\eta_{d1}, \dots, \eta_{dK})^T \sim N_K(\vec{0}, R) \end{aligned}$$

This model differs from the model presented in (5.1) and (5.2) since we are not constraining the variance of ϵ to 0.25. Our goal is to examine the estimated variance of ϵ from this model. If in fact its value is close to 0.25 then this can help justify our theoretical model. We will examine results from two modelling alternatives, both allowing us to incorporate the three variance components. The first modelling technique is carried out by using the ARMA(1,1) structure extra parameter γ to construct both the within-periods AR(1) structure and the needed extra ϵ error term variance. The G matrix is modelled using a standard spatial power structure, while the $R^* = R + \sigma_\epsilon^2 \cdot I$ matrix is modelled using an ARMA(1,1) formulation. Here is the manner in which we use the standard ARMA(1,1) covariance matrix to achieve this goal:

$$R^* = \sigma_{R^*}^2 \cdot \begin{pmatrix} 1 & \gamma & \gamma \cdot \rho_{R^*} & \dots & \gamma\rho_{R^*}^{K-2} \\ \gamma & 1 & \gamma & \dots & \gamma\rho_{R^*}^{K-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma \cdot \rho_{R^*}^{K-2} & \gamma \cdot \rho_{R^*}^{K-3} & \dots & \gamma & 1 \end{pmatrix}$$

Where:

$$\sigma^2_{R^*} = \sigma^2_\eta + \sigma^2_\epsilon \tag{5.9}$$

$$\gamma = \rho_\eta \cdot \frac{\sigma^2_\eta}{\sigma^2_\eta + \sigma^2_\epsilon} \tag{5.10}$$

$$\rho_{R^*} = \rho_\eta \tag{5.11}$$

$$\Rightarrow \sigma^2_{R^*} \cdot (1 - \frac{\gamma}{\rho_\eta}) = \sigma^2_\epsilon \tag{5.12}$$

By plugging-in the relevant estimators in (5.12) we can estimate the required variance of ϵ_{dk} and compare its value to the theoretical value of 0.25.
The second technique is quite straight forward. We model equations (5.7) and (5.8) using the *mixed* procedure in SAS®. The G matrix is modelled using a standard spatial power structure and the R matrix is modelled according to a first order auto-regressive structure. We add the 'local' option to incorporate the diagonal covariance matrix of ϵ. As part of the SAS® default output one obtains the estimated value of σ^2_ϵ.

5.1.5 Benchmark Models

An elementary prediction model would simply average past data in order to produce a forecast. This model is referred to as the industry model. Specifically, let q_i denote the weekday corresponding to the i^{th} day. Denote $W_{is} = \{i^{'} : i^{'} \leq i \text{ and } q_{i'} = s\}$ and let $\mid W_{is} \mid$ denote the cardinality of W_{is}. Then the forecast arrival count, $N_{D+h,j}$, based on the information up to day D can be expressed as

$$\hat{N}_{D+h,j} = \frac{\sum_{i \in W_{D,D+h}} N_{ij}}{\mid W_{D,D+h} \mid} \tag{5.13}$$

Based on this intuitive approach, we develop two similar baseline models. These models will serve as benchmarks for our more complicated models.
The first basic model only considers the weekday fixed effects and their interactions with the periods. Basically, this model states that each day of the week has its own baseline level and its own intra-day pattern and that consecutive days and periods are uncorrelated (as opposed to our initial correlated mixed model defined in (5.2)). The formulation of

this model can be written as follows:

$$\begin{aligned} y_{dk} &= \theta_{dk} + \epsilon_{dk} \\ \theta_{dk} &= \alpha_{q_d} + p_{q_d,k} + \eta_{dk} \\ \text{with} \quad & \epsilon_{dk} \sim \mathrm{N}(0, \frac{1}{4}) \quad \text{i.i.d.} \quad \text{and} \\ & \eta_d = (\eta_{d1}, \ldots, \eta_{dK})^T \sim \mathrm{N}_K(\vec{0}, \sigma^2 I) \end{aligned} \tag{5.14}$$

where the sequence $\eta_1, \ldots, \eta_D$ is an i.i.d sequence of K-variate normal vectors with zero mean and covariance matrix $\sigma^2 I$.
This model basically corresponds to the industry model (defined in (5.13)). It assume the days are independent of each other. Alternatively, one can think of a different benchmark model which is similar to the above model but also includes exogenous variables. Hence, in the Israeli Cellular company our second benchmark model also incorporates the billing cycles variables. Using the above notation we can define the second benchmark model in the following manner:

$$\begin{aligned} y_{dk} &= \theta_{dk} + \epsilon_{dk} \\ \theta_{dk} &= V_d + \alpha_{q_d} + p_{q_d,k} + \beta_d + \eta_{dk} \\ \text{with} \quad & \epsilon_{dk} \sim \mathrm{N}(0, \frac{1}{4}) \quad \text{i.i.d.} \quad \text{and} \\ & \eta_d = (\eta_{d1}, \ldots, \eta_{dK})^T \sim \mathrm{N}_K(\vec{0}, \sigma^2 I) \end{aligned} \tag{5.15}$$

where the sequence $\eta_1, \ldots, \eta_D$ is an i.i.d sequence of K-variate normal vectors with a zero mean and covariance matrix $\sigma^2 I$.
This second benchmark model will have the same fixed-effects settings as our final mixed model but includes an underlying assumption that all of its observations are uncorrelated. Hence it represents a baseline to our more elaborate, correlated model.
Both of the benchmark models are, in fact, linear regression models and are quite fast and efficient in providing the necessary predictions using standard programs.

5.1.6 Theoretical versus Practical models

In Section 5.1.1 we introduced our theoretical model. This model which constrains the variance of ϵ to a value of 0.25 was implemented using the *mixed* procedure in SAS®

adding both the 'local' and 'hold' options in order to incorporate the extra variance. However, the results we obtained made little sense since the estimated value of daily variance (i.e., σ_V^2) was zero. This result contradicts previous research that was done on similar call center data, and it probably is the consequence of an algorithmic failure due to the large dimension of the model.

Consequently we implemented two alternative models. The first model is the general model presented in (5.7) and (5.8). This model allows ϵ's variance to have any non-negative value. This model will be referred to as the general variance model.

The second model is a special case of the first model where the variance of ϵ is given the value of zero. This means that we actually do not model the ϵ extra variance parameter. To some extent the extra variance is actually incorporated in the remaining two variance components (i.e., σ_V^2 and σ_η^2) instead. This model will be referred to as the zero variance model.

We compare the prediction results of the two models in detail in Section 6.1.

5.2 Gaussian Bayesian Model for Arrival Counts

The next section is based on the forecast model specified in a paper by Weinberg *et al.* ([22]). We use the results presented in that paper to compare our mixed model results in the sequel.

5.2.1 Definition

Using the same terminology as in the mixed model, define N_{dk} as the number of arrivals to the queue on day $d = 1, ..., D$, and during the time interval $[t_{k-1}, t_k)$, where $k = 1, ..., K$ is the k^{th} period of the day. The model assumes:

$$N_{dk} \sim \text{Poisson}(\lambda_{dk}), \quad \lambda_{dk} = R_{q_d}(t_k)v_d + \epsilon_{dk} \tag{5.16}$$

where λ_{dk} is the arrival rate for day d during period k, $R_q(t_k)$ is the proportion of daily volume on the q^{th} weekday during the time interval $[t_{k-1}, t_k]$, v_d represents the daily volume during day d and ϵ_{dk} is the random error. The models assumes that each day has its own within-day pattern and as a result, the following restriction is enforced:

$$\sum_{k=1}^{K} R_{q_d}(t_k) = 1 \quad \text{for} \quad q_d = 1, \ldots, 5. \tag{5.17}$$

Following the same approximation technique used in the previous mixed model, define $y_{dk} = \sqrt{N_{dk} + \frac{1}{4}}$. This normal approximation enables one to redefine the model in the following manner:

$$y_{dk} = g_{q_d}(t_k)x_d + \epsilon_{dk}, \quad \epsilon_{dk} \sim N(0, \sigma^2) \quad \text{iid} \tag{5.18}$$

where $g_{q_d} = \sqrt{R_{q_d}}$ and $x_d = \sqrt{v_d}$. The daily correlations are modelled using an AR(1) structure in the following manner:

$$x_d - \alpha_{q_d} = \beta(x_{d-1} - \alpha_{q_{d-1}}) + \eta_d, \quad \eta_d \sim N(0, \psi^2) \quad \text{iid} \tag{5.19}$$

where α_{q_d} denotes the intercept for day q_d. To incorporate the different weekday patterns and ensure some smoothness in them, two extra constraints are added:

$$\frac{d^2 g_q(t_k)}{dt_k^2} = \tau_q \frac{dW_{q_d}(t_k)}{dt_k} \tag{5.20}$$

where

$$\sum_{k=1}^{K} g_q(t_k)^2 = 1, \quad \text{for} \quad q = 1, \ldots, 5. \tag{5.21}$$

Here $W_q(t)$ are 5 independent Wiener processes with $W_q(0) = 0$ and Var$\{W_q(t)\} = t$ (and $\frac{dW_{q_d}(t_k)}{dt_k}$ is a notation for a white noise).

For computational reasons the authors reformulate their model by introducing a new variable $z_{q_d} = \{g_{q_d}(t_k), dg_{q_d}(t_k)/dt_k\}$ and rewriting the model in the following manner:

$$y_{dk} = h' z_{q_d}(t_k)x_d + \epsilon_{dk}, \quad \epsilon_{dk} \sim N(0, \sigma^2) \tag{5.22}$$

$$x_d - \alpha_{q_d} = \beta(x_{d-1} - \alpha_{q_{d-1}}) + \eta_d, \quad \eta_d \sim N(0, \psi^2) \tag{5.23}$$

$$z_{q_d}(t_k) = F(\delta) z_{q_d}(t_{k-1}) + u_k, \quad u_k \sim N(0, \tau_{q_d}^2 U(\delta)) \tag{5.24}$$

where ϵ_{dk}, η_d and u_k are mutually independent, $\delta = t_k - t_{k-1}$ and vector $h' = [1, 0]$. The matrices $F(\delta)$ and $U(\delta)$ are defined as

$$F(\delta) = \begin{pmatrix} 1 & \delta \\ 0 & 1 \end{pmatrix}, \quad U(\delta) = \begin{pmatrix} \delta^3/3 & \delta^2/2 \\ \delta^2/2 & \delta \end{pmatrix}$$

Finally the authors also incorporate the weekday constraints defined in (5.21) and use diffuse distributions for the initial states x_1 and $z_{q_d}(t_1)$, for $q_d = 1, \ldots, 5$. It is now apparent that the model presented here is a multiplicative model with two latent states each evolving on its own time scale. Conditional on each latent state variable, the model can be cast into a linear state space form. This property enables the use of state-space model techniques to help overcome computational problems.

5.2.2 Estimation and Prediction

The authors of [22] give an extensive description of the hybrid Markov Chain Monte Carlo (MCMC) algorithm which they developed to sample from the relevant parameters' posterior distributions. The algorithm utilizes Gibbs sampling techniques and Metropolis-Hastings steps to sample from the required conditional distributions.

The first step of the one-day-ahead forecasting algorithm is to run MCMC simulations based on past data. Assuming that the goal is to forecast day d arrival counts, the purpose of this step is to obtain samples from the posterior distribution $p(\Omega|Y_{1:d-1})$ (where $\Omega = (\alpha, \beta, \sigma^2, \psi^2, \tau^2, x_{d-1}, z)$ is the vector of required parameters and $Y_{1:d-1}$ is the vector of transformed past observations). These samples enable the authors to generate the empirical distribution of the untransformed Poisson arrival counts (i.e., N_{dk}) for each period k. The one-day-ahead forecasting algorithm for predicting day d is summarized in the next few lines:

1. Start by generating an MCMC sample, $\Omega^{(1)}, \ldots, \Omega^{(M)}$, drawn from $p(\Omega|Y_{1:d-1})$. M is approximately 4899 samples which are obtained using the Gibbs sampler and Metropolis algorithm. The authors state that they used 49,000 iterations after a burn-in period of 1000. They sample the parameters every 10th iteration.
2. Draw $x_d^{(i)} \sim N\left(\alpha_{q_d}^{(i)} + \beta^{(i)}\left(x_{d-1}^{(i)} - \alpha_{q_{d-1}}^{(i)}\right), (\psi^2)^{(i)}\right)$, for each $i = 1, \ldots, M$.
3. For each period $k = 1, \ldots, K$ and each $i = 1, \ldots, M$.
   - Set $\lambda_{dk}^{(i)} = \left(x_d^{(i)} g_{q_d}(t_k)^{(i)}\right)^2$.
   - Draw $y_{dk}^{(i)} \sim N\left(\sqrt{\lambda_{dk}^{(i)}}, (\sigma^2)^{(i)}\right)$.
   - Set $N_{dk}^{(i)} = \left(y_{dk}^{(i)}\right)^2 - 0.25$.

5.3 Poisson Bayesian Model for Arrival Counts

The model described in this section, similar to that in the last section, also employs Bayesian modelling techniques to predict the arrival counts. However, it directly models the untransformed Poisson arrival counts without using the Normal-Poisson stabilization approximation. The model was implemented using both the *BRugs* package [20] in the R statistical software and 'OpenBugs' software [17]. The actual code used for implementing this model is provided in Appendix B.2.

5.3.1 Definition

The formulation of a hierarchical Bayesian model for the untransformed counts can be presented as follows:

$$
\begin{aligned}
N_{dk}|\lambda_{dk} &\sim \text{Poisson}(\lambda_{dk}) && (5.25)\\
\lambda_{dk}|V_d, p_{q_d} = (p_{q_d,1}, \ldots, p_{q_d,K}) &= V_d \cdot p_{q_d,k} && (5.26)\\
V_d &= \mu_{q_d} \cdot G_d && (5.27)\\
G_d &= F(G_{d-1}, U_d) && (5.28)\\
p_q|\alpha_q &\sim \text{Dirichlet}(\alpha_{q,1}, \ldots, \alpha_{q,K}) && (5.29)\\
\alpha_q|M_q &= \exp(M_q) \cdot (\pi_{q,1}, \ldots, \pi_{q,K}) && (5.30)\\
M_q &\sim N(5,5) && (5.31)\\
\mu_q &\sim N(\upsilon_q, 0.01 \cdot \upsilon_q^2), \quad \text{for } q = 1, \ldots, 7 && (5.32)
\end{aligned}
$$

The arrival rate for each period of each day, λ_{dk}, is comprised of two sets of parameters: the daily volume parameters and the daily pattern parameters. We shall begin by explaining the latter:

- $p_{q_d} = (p_{q_d,1}, \ldots, p_{q_d,K})$ is the vector of daily proportions assigned to each of the K periods. The subscript q denotes the day of the week ($q = 1, \ldots, 7$). Each weekday has its own pattern. These vectors are distributed according to the discrete Dirichlet distribution.

- $\alpha_q = (\alpha_{q,1}, \ldots, \alpha_{q,K})$ is the vector of parameters for the (discrete) Dirichlet distribution. The vector α_q can be written as $\exp(M_q) \cdot \pi_q$.

- $\exp(M_q)$ governs the variability of the Dirichlet distribution about its mean $\pi_q = (\pi_{q,1}, \ldots, \pi_{q,K})$, which is a probability vector itself. The M_q factors have a diffuse prior and π_q is the estimated pattern according to the learning data. M_q has been given a normal prior distribution with expectation and variance equal to 5. This prior choice allows M_q to have both negative and positive values. Large positive values of M_q imply small variability of the Dirichlet distribution while very negative values of M_q imply large variability of the Dirichlet distribution. The normal distribution is very easily simulated from and so is a natural candidate for this prior.

The daily volume has two components: μ_q, the weekday mean value for day q; and G_d, which describes the day d deviation from its corresponding weekday mean. The different

weekday means $(\mu_1, \ldots, \mu_7)$ are given a diffuse normal prior with means corresponding to the actual average daily total values and variance which is 1% of those daily totals. In fact, υ_q is the averaged q^{th} weekday mean value according to the learning data. The weekday means prior variance were adjusted to allow coefficients of variation of 10%. This implies that the weekdays means are not the main source of variation and the variation comes from a different source, mainly G_d.

Equation (5.28) describes the dynamic dependence of daily volumes between successive days. It is expressed in a general form here, where G_d is assumed to have an expectation of 1. Different alternatives can be chosen in order to define the G_d process. These alternatives need to have two main characteristics: the process should always have positive values since G_d represents a *multiplicative* deviation from the weekday average; G_d should have a mean value of 1 basically implying that on average the daily volume V_d equals the appropriate weekday mean. We have selected the Beta-Gamma Auto-Regressive process which accommodates both characteristics. This process is defined by:

$$G_d = G_{d-1} \cdot B_d + U_d \quad \text{for } d = 2, \ldots, D \tag{5.33}$$

$$B_d \sim \text{Beta}(\gamma\rho, \gamma(1-\rho)) \tag{5.34}$$

$$U_d \sim \text{Gamma}(\gamma(1-\rho), \gamma) \tag{5.35}$$

$$G_1 \sim \text{Gamma}(\gamma, \gamma) \tag{5.36}$$

which has a stationary marginal Gamma(γ, γ) distribution (with mean 1). In order to complete the Bayesian formulation the non-informative Gamma(0.01,0.01) prior was used for the γ parameter. Also a normal prior distribution was used for the $-\log\rho$ parameter. It is quite straightforward to show that ρ is the autocorrelation between G_d and G_{d-1} and that the expected value of ρ, using Normal(2,5) as the prior distribution, is approximately 0.69.

5.3.2 Estimation and Prediction

As mentioned earlier, we use the 'OpenBugs' environment to implement the Poisson Bayesian model. 'OpenBugs' is an open-source software for Bayesian analysis of complex statistical models using Markov Chain Monte Carlo (MCMC) methods. The estimation procedure utilizes Gibbs sampling techniques and Metropolis steps to sample from the relevant parameters' posterior distributions. Based on these samples, one can create the empirical posterior distributions of each of the parameters.

The forecasting procedure, using 'OpenBugs', simply regards the predictions as additional parameters. Hence, one can create for each day d and period k an estimated distribution for the predicted N_{dk}.

5.4 Regression Model for Service Times

Aside from predicting the arrival rate, forecasting a queuing system load also requires predicting the average service patterns (or alternatively the service rate pattern for each day). Since our arrivals model focuses on a specific resolution, we need to predict the average service time during those same intervals (i.e., periods).

Our model involves two explanatory variables that may affect service rates: the weekday and the period. We compare two alternatives models where one is a generalization of the other. The first model describes the average service time using a quadratic regression in the periods, with interactions with the weekday effect, where the period is included as a numeric variable (rather than as a categorical variable). Intuitively speaking, this model states that the daily service time curves differ among the different weekdays but they are confined to be of a quadratic form. The formulation of this model can be written as follows:

$$\text{Model 1} \quad \varphi_{dk} = \alpha_{q_d} + \varsigma k^2 + \vartheta k + \chi_{q_d} k^2 + \phi_{q_d} k + \phi d + \epsilon_{dk};\ \epsilon_{dk} \sim N(0, \sigma^2) \quad (5.37)$$

where α_q is the constant term related to the q^{th} weekday; ϑ and ς are, respectively, the quadratic and linear coefficients; χ_q and ϕ_q are the weekday-specific quadratic and linear period effects and make up the weekday-period interaction effects. The last effect is a postulated linear daily trend coefficient denoted ϕ. We naturally, added a random error term denoted by ϵ_{dk}.

The second model is a generalization of the first model and it assumes that the periods variable is a categorial variable. It basically assumes that each weekday has its own average service times pattern with no other restriction on its shape (hence it is a generalization of the last quadratic-shaped model). We add both the linear daily trend effect to this model and the error term as well. This model can be formulated in the following manner:

$$\text{Model 2} \quad \varphi_{dk} = \beta_{q_d,k} + \phi d + \epsilon_{dk};\ \epsilon_{dk} \sim N(0, \sigma^2). \quad (5.38)$$

where $\beta_{q,k}$ is the interaction between the weekday and the effect of the k^{th} period.

Chapter 6

Mixed Model — Determining Fixed Effects and Covariance Structure

This next section gives details of the selection process for both the fixed effects (i.e., the weekday patterns, or the Israeli Cellular company's billing cycles), and the random effects covariance structures for the mixed model.

6.1 Analysis of Practical Models

As previously explained in Section 5.1.6, we shall explore the predictions results of two alternative models. This analysis will be carried out on the Israeli Cellular company's data. We encountered convergence problems when we implemented the first general model and so we base the following comparison only on 135 days out of the 203 between April 11, 2004 and December 25, 2004. For this analysis, we shall set the between-periods (within-day) correlation according to a first-order autoregressive structure. The fixed effects set-up is carried out according to the model presented in the end of Section . Tables 6.1,6.2, 6.3 and 6.4 present the results of the two models. The model where ϵ's variance is constrained to the value of zero (zero variance model) shows much better results then the general variance model. A possible explanation for this is that the general model is over-fitting the learning data and hence the predictions turn out to be quite poor. Following this analysis we decided to continue investigating only the second model where ϵ's variance is restricted to the value of zero. This model can be formulated in the following manner:

$$y_{dk} = V_d + \alpha_{q_d} + p_{q_d,k} + \beta_d + \eta_{dk} \tag{6.1}$$

with

$$(V_1, \ldots, V_D)^T \sim N_D(\vec{0}, G) \quad \text{and} \quad \eta_d = (\eta_{d1}, \ldots, \eta_{dK})^T \sim N_K(\vec{0}, R)$$

From here on we shall refer to this model as the mixed model.

	RMSE	
N=135	general variance model	zero variance model
1^{st} Quartile	42.45	31.34
Median	72.50	37.07
Mean	103.92	42.06
3^{rd} Quartile	170.73	49.93

Table 6.1: Analysis of practical models – RMSE results .

	APE	
N=135	general variance model	zero variance model
1^{st} Quartile	10.06	7.72
Median	18.57	9.68
Mean	34.47	11.10
3^{rd} Quartile	39.29	13.51

Table 6.2: Analysis of practical models – APE results.

	Coverage Probability	
N=135	general variance model	zero variance model
1^{st} Quartile	0.36	0.88
Median	0.70	0.97
Mean	0.64	0.92
3^{rd} Quartile	0.97	1

Table 6.3: Analysis of practical models – Coverage probabilities results.

	Width	
N=135	general variance model	zero variance model
1^{st} Quartile	36.21	136.81
Median	152.22	155.76
Mean	159.45	159.57
3^{rd} Quartile	181.85	182.76

Table 6.4: Analysis of practical models – Confidence interval widths.

6.2 Israeli Cellular Company

Guided by the principle of parsimony, we conducted a preliminary investigation of the billing cycles indicators described in Section **??**. We sequentially examined the different effects using out-of-sample performance measures. In effect we tested each effect (random or fixed) alone while the rest of the model terms were kept fixed. When we established the best setting for a specific effect we continued to the next one.

6.2.1 Preliminary Analysis of Billing Cycles

As mentioned in the data description, we have eight indicators which represent the four major billing cycles (i.e. four delivery period indicators and four billing period indicators). Based on the company's information we were made aware that some of these indicators might not have a significant influence on the Private queue's arrival process. The purpose of this preliminary examination is to highlight those indicators which are significant so they may later be incorporated in the final forecasting model. For this coarse investigation, we use the aggregated *daily* arrivals between February 14^{th}, 2004 and December 31^{st}, 2004 (excluding all twenty-two outliers).

Let N_i denote the number of arrivals to the Private queue on day i=1,....M. In our study, we have M=254 days. In addition, let q_i denote the weekday corresponding to day i. The daily arrivals are modelled using a Poisson log-linear model (for further details on this approach the reader is referred to [13]). Our initial model is of the following form:

$$\begin{aligned} N_i &\sim \text{Poisson}(\lambda_i) \\ \log(\lambda_i) &= \sum_{l=1}^{6} \text{W}_{q_i l} \cdot X_{q_i l} + \sum_{j \in \text{cycles}} \text{B}^j \cdot \text{M}_i^j + \sum_{j \in \text{cycles}} \text{D}^j \cdot \text{U}_i^j \end{aligned} \tag{6.2}$$

where

q_i is the weekday corresponding to day i

W_q is the qth weekday coefficient

$\vec{X_{q_i}}$ is a vector which has six elements.It's q^{th} element has a value of one indicating the weekday corresponding to the i^{th} day.

B^j is the j^{th} billing period coefficient

$$M_i^j = \begin{cases} 1 & \text{if cycle j's billing period falls on the i}^{th} \text{ day} \\ 0 & \text{otherwise} \end{cases}$$

D^j is the j^{th} delivery period coefficient

$$U_i^j = \begin{cases} 1 & \text{if cycle j's delivery period falls on the } i^{th} \text{ day} \\ 0 & \text{otherwise} \end{cases}$$

The model is implemented using the *GENMOD* procedure in SAS® based on the 254 observations in the current learning set. Some of the results are summarized in tables 6.5 and 6.6. The results indicate the following: the six weekdays have significantly different effects, each having a different baseline mean (no intercept was included in the model); the delivery period indicators are significant and have a positive effect on the mean value of the number of incoming calls; on the other hand, most of the billing period indicators seem less significant, which confirms the cellular company's beliefs. The Cycle 14 billing period seems to have an exceptional effect. First, it is statistically significant as opposed to the billing indicators of the rest of the cycles. Furthermore, its estimator is the only negative value among those of all of the effects. This strange result would suggest that the number of incoming calls is reduced during the Cycle 14 billing period. We could not attribute this phenomenon to any outlying data problems. One possibility is that this negative value can be compensating for other oversized billing cycle effects, which could arise due to the overlap of delivery and billing days among the 4 cycles.

These results led us to believe that some of the explanatory variables are statistically redundant. We proceed by comparing different models with this initial model (defined in (6.2)) using the 'contrast' statement in the *GENMOD* procedure (which computes likelihood-ratio statistics). The different models are variations of the initial model. They exclude different covariates in order to establish the importance of the omitted variables. The retained explanatory variables for each examined model are listed below:

1. the weekday effect and the delivery period indicators;

Parameter	Category	Estimate	Chi-Square	Pr > ChiSq
W_q	Sunday	9.5358	718177	<.0001
W_q	Monday	9.4977	541484	<.0001
W_q	Tuesday	9.4880	572102	<.0001
W_q	Wednesday	9.4719	649509	<.0001
W_q	Thursday	9.4326	677894	<.0001
W_q	Friday	9.0385	453474	<.0001
D^1	.	0.0586	11.59	0.0007
D^7	.	0.0327	3.71	0.0540
D^{14}	.	0.0449	5.39	0.0202
D^{21}	.	0.0935	17.98	<.0001
B^1	.	0.0242	2.10	0.1473
B^7	.	0.0279	2.17	0.1403
B^{14}	.	-0.0592	6.83	0.0090
B^{21}	.	0.0276	2.54	0.1109

Table 6.5: Analysis of parameter estimates for the Poisson Log-Linear model.

2. the weekday effect, one global billing indicator (which takes the value one when at least one of the cycles is during its billing period) and the four delivery period indicators

3. the weekday effect, Cycle 14 billing period indicator and the four delivery period indicators

4. the weekday effect, Cycle 14 billing period indicator and the one global delivery period indicator (which takes the value one when at least one of the cycles is during its delivery period)

5. the weekday effect, one global billing period indicator and one global delivery period indicator

Source	DF	Chi-Square	Pr > ChiSq
W_q	6	518663	<.0001
D^1	1	11.46	0.0007
D^7	1	3.69	0.0548
D^{14}	1	5.35	0.0207
D^{21}	1	17.79	<.0001
B^1	1	2.09	0.1485
B^7	1	2.16	0.1415
B^{14}	1	6.87	0.0088
B^{21}	1	2.53	0.1119

Table 6.6: The Poisson Log-Linear model. Likelihood-ratio comparison for Type 3 analysis.

The analyses of the contrasts are shown in Table 6.7. By using the 'contrast' statement we are actually examining which variables are statistically (in)significant. The results indicated that billing periods 1, 7 and 21 are redundant. It is therefore clear that there are three main factors contributing to the daily volumes: the weekday, the delivery periods and billing period of Cycle 14. Since both models 3 and 4 seemed to be reasonable we decide to pursue them both.

We conclude this section by defining four different alternative settings for the billing cycle indicators (ξ):

- Setup 1: $(\xi_1, \ldots, \xi_5) =$ (Cycle 14 billing period and four delivery period indicators);
- Setup 2: $(\xi_1, \ldots, \xi_4) =$ (four delivery period indicators);
- Setup 3: $(\xi_1, \xi_2) =$ (Cycle 14 billing period and one global delivery period indicator);
- Setup 4: $(\xi_1) =$ (one global delivery period indicator).

We shall analyze the ”best” settings using out-of-sample performance measures in the following subsection.

6.2.2 Fixed Effects Selection

Our process of model selection does not rely on classical inference methods or measures, such as Akaike's Information Criteria [15]. Instead, we explore the influence of the models' elements on the prediction performance based on the 2004 validation-set. The evaluation method is detailed in Chapter 4.

As mentioned in the above section we begin with four alternatives for the fixed-effects, all including the different weekday effects and their interaction with periods. However, they do differ in their billing cycles indicators. Our first step is to determine the best candidate model out of these four. For now, we shall set the between-periods (within-day) correlation according to a first-order autoregressive structure. We choose this specific structure for its simplicity. In the next section we will also consider other correlation structures.

In addition we compare the four models with the performance of the two benchmark models, mentioned in Section 5.1.5. The first model only includes the weekday and weekday patterns. The alternative benchmark model has two more additional billing cycles indicators: one global delivery indicator and the billing period indicator associated with cycle 14. The reason why these specific billing cycles settings were chosen will be explained later in this section.

These models are evaluated using the same out-of-sample prediction procedure, i.e. for a specific week we calculate the appropriate linear regression estimates based on six weeks of past data and then we predict the week starting 10-days-ahead. This procedure is carried out 37 times since there are 37 weeks between April 11, 2004 and December 25, 2004.

Model No.	The Alternative model explanatory variables	Num DF	Den DF	F Value	Pr>F	ChiSq	Pr>ChiSq	Type
1	• Weekday • four Delivery periods indicators	4	240	3.42	0.0096	13.69	0.0084	LR
2	• Weekday • Global Billing indicator • four Delivery periods indicators	3	240	4.11	0.0072	12.32	0.0064	LR
3	• Weekday • Billing 14 period indicator • Global Delivery period indicator	6	240	1.89	0.0834	11.33	0.0786	LR
4	• Weekday • Billing 14 period indicator • four Delivery period indicators	3	240	2.00	0.1152	5.99	0.1121	LR
5	• Weekday • Global Billing period indicator • Global Delivery period indicator	6	240	2.38	0.0296	14.30	0.0264	LR

Table 6.7: Log-Linear models contrasts analyses. Each row depicts a different model which is compared to the initial model. Comparing models 3 and 4 to the initial model shows that the extra variables are not significant at a significance level of 5%.

	RMSE					
N=203	Setup 1	Setup 2	Setup 3	Setup 4	BM1	BM2
1^{st} Quartile	34.26	34.14	33.33	33.37	33.23	33.12
Median	40.62	41.12	40.46	40.77	41.39	40.71
Mean	45.51	45.87	44.57	44.8	46.05	45.6
3^{rd} Quartile	53.21	52.95	51.72	51.88	54.36	55.8

Table 6.8: RMSE results for the four fixed effects models and the benchmark models.

	APE					
N=203	Setup 1	Setup 2	Setup 3	Setup 4	BM1	BM2
1^{st} Quartile	8.35	8.33	8.15	8.17	8.30	8.25
Median	10.13	10.13	9.95	9.98	10.46	10.48
Mean	11.25	11.29	11.06	11.07	11.24	11.3
3^{rd} Quartile	13.81	13.45	13.37	13.01	13.49	13.44

Table 6.9: APE results for the four fixed effects models and the benchmark models.

	Coverage Probability					
N=203	Setup 1	Setup 2	Setup 3	Setup 4	BM1	BM2
1^{st} Quartile	0.85	0.85	0.88	0.88	0.39	0.39
Median	0.94	0.94	0.94	0.94	0.51	0.48
Mean	0.91	0.91	0.92	0.92	0.51	0.5
3^{rd} Quartile	1	1	1	1	0.61	0.61

Table 6.10: Coverage probabilities for the four fixed effects models and the benchmark models.

	Width					
N=203	Setup 1	Setup 2	Setup 3	Setup 4	BM1	BM2
1^{st} Quartile	134.87	135.06	138.15	137.59	51.57	51.29
Median	156.87	155.54	157.22	157.31	59.56	58.7
Mean	161.61	161.74	151.71	162.15	63.2	62.04
3^{rd} Quartile	184.79	185.33	184.48	184.56	70.07	68.78

Table 6.11: Confidence interval widths for the four fixed effects models and the benchmark models.

We use the SAS® *Mixed* procedure in order to implement and evaluate the candidate models. Tables 6.9, 6.8, 6.10 and 6.11 present the results of the four different fixed model setups and the two benchmark models. Out of the four different models, the 3^{rd} alternative seems to exhibit the best results. Its results are also better than the first benchmark model. One can argue that the shorter confidence intervals imply that the benchmark outperforms the third model. However, since its coverage probability is very far from the nominal 95%, we conclude that these narrow intervals are unreliable and probably result from an under-estimated error variance.

The second benchmark model is in fact the "fixed" version of this third model; that is, without the random effects. It is interesting to see that introducing the intra- and inter-day correlations improves the forecasting results. Intuitively speaking, the additional correlation parameters result in wider confidence intervals to compensate for the extra uncertainty.

Based on the above results, we continue analyzing models which include only two billing cycle indicators (one for any delivery period and one for the Cycle 14 billing period). At this stage we will refer to this chosen model as the *Multi-Pattern* model because it incorporates a different intra-day pattern for each weekday.

In the preceding models, each day has its own pattern of arrivals over periods. Keeping the parsimony concept in mind, Figure 3.3 suggests that some weekday patterns resemble others (at least during most periods of the day). To examine if indeed this is the case we first normalize each period's arrivals, N_{dk}, dividing by the day d mean value, $\bar{N}_d$. Then we averaged each normalized period value separately over each weekday type (for Sunday through Friday). By carefully examining the above mentioned Figure 3.3, we can conclude that Sundays and Fridays have patterns that differ from those of the rest of the weekdays. Based on this observation, we set aside Sundays and Fridays and test the hypothesis, separately for each period k, that there is no significant difference between the remaining 4 weekdays based on the normalized mean values. In effect, we fit the following simple model for each period k:

$$\frac{N_{dk}}{\bar{N}_d} = \upsilon_{q_d};\ \text{for} \quad d = 1, \ldots, 254;\ k = 1, \ldots, 33 \tag{6.3}$$

where υ_q is the effect of the q^{th} weekday.

We thus formally test our observation using a simple ANOVA model. The ANOVA underline assumptions of normality and homoscedasticity were examined using different plots and were found to hold for the data.

Rejecting the null hypothesis for a specific period k indicates that there is a difference between the weekdays for that period. According to the results shown in Table 6.12, for most periods of the day the 4 weekdays patterns are similar.

Period	Start Time	F-value	Pr>F	Period	Start Time	F-value	Pr>F
1	7:00	0.96	0.4131	18	15:30	1.54	0.2070
2	7:30	0.93	0.4268	19	16:00	2.34	0.0757
3	8:00	1.70	0.1692	20	16:30	1.63	0.1841
4	8:30	2.43	0.0673	**21**	**17:00**	**3.08**	**0.0290**
5	9:00	1.28	0.2831	**22**	**17:30**	**4.80**	**0.0031**
6	9:30	1.54	0.1069	23	18:00	1.32	0.2683
7	10:00	0.49	0.6873	**24**	**18:30**	**3.57**	**0.0153**
8	10:30	0.53	0.6598	**25**	**19:00**	**3.23**	**0.0238**
9	11:00	1.53	0.2097	**26**	**19:30**	**3.22**	**0.0242**
10	11:30	0.70	0.5512	27	20:00	0.71	0.5467
11	**12:00**	**5.02**	**0.0024**	28	20:30	2.53	0.0593
12	**12:30**	**3.00**	**0.0321**	29	21:00	0.25	0.8607
13	13:00	0.54	0.6564	30	21:30	0.46	0.7076
14	13:30	0.05	0.9852	31	22:00	1.63	0.1836
15	14:00	0.18	0.9097	32	22:30	1.71	0.1663
16	14:30	0.15	0.9276	33	23:00	0.35	0.7865
17	15:00	0.82	0.4851	-	-	-	-

Table 6.12: ANOVA Results of Monday through Thursday effects for each period. According to the p-values, most of the periods do not differ between the 4 weekdays. The bolded rows indicate those 5% significant periods.

As a result of the above analysis we shall evaluate two additional models. The Three-Pattern model has three different patterns: one for Sunday, one for Friday and one for the remaining weekdays. In addition to the Multi-Pattern model, we also try and fit a Two-Pattern model which includes two patterns: one for Friday and one for the rest of the weekdays. Both models still have the weekday effect, a general delivery periods effect and the Cycle 14 billing period effect, together with one of the patterns. The Two-Pattern model assumes that all weekdays, except Friday, have the same relative intra-day behavior but may show different absolute levels. The results of the three model evaluations are

presented in tables 6.13, 6.14, 6.15 and 6.16. Based on the RMSE and the APE measures the Three-Pattern model provides the best performance. However, looking at the coverage and confidence widths we see that the Multi-Pattern model achieves the best outcomes. Based on parsimony considerations, we decided to choose the Three-Pattern model. This concludes our discussion of how the fixed effects were selected.

	RMSE		
N=203	Two-Pattern	Three-Pattern	Multi-Pattern
1^{st} Quartile	32.51	32.65	33.33
Median	38.81	38.25	40.46
Mean	43.32	43.3	44.57
3^{rd} Quartile	50.45	50.94	51.72

Table 6.13: Comparing models with different numbers of weekday patterns. Results for RMSE.

	APE		
N=203	Two-Pattern	Three-Pattern	Multi-Pattern
1^{st} Quartile	7.83	7.83	8.15
Median	9.71	9.68	9.95
Mean	10.8	10.8	11.06
3^{rd} Quartile	13.27	12.86	13.37

Table 6.14: Comparing models with different numbers of weekday patterns. Results for APE.

	Coverage Probability		
N=203	Two-Pattern	Three-Pattern	Multi-Pattern
1^{st} Quartile	0.91	0.88	0.88
Median	0.97	0.97	0.94
Mean	0.93	0.93	0.92
3^{rd} Quartile	1	1	1

Table 6.15: Comparing models with different numbers of weekday patterns.Results for coverage probability.

	Width		
N=203	Two-Pattern	Three-Pattern	Multi-Pattern
1^{st} Quartile	141.67	141.53	138.15
Median	160.15	160.76	157.22
Mean	164.33	162.84	161.71
3^{rd} Quartile	186.79	185.4	184.48

Table 6.16: Comparing models with different numbers of weekday patterns. Results for width.

6.2.3 Determining the Covariance Structure — Random Effects

Having chosen the fixed effects that will be incorporated in our model we now discuss the modelling of the random effects. There are two sources of variation in our model: one is from the daily volume effect V_d and the other is the within-day error vector η_d.

We begin by examining different structures for the matrix R which is the within-day covariance matrix. Because of a certain indeterminacy in solving for V_d and the K residual error variances one cannot allow the variances (diagonal elements) in R to be unconstrained. Either a lower bound must be set on these variances — 1/4 would be a logical lower bound since this is the approximate variance of the square-root-Poisson variable — or one may choose a structure for R that has the same variance for each component. Amongst the latter we shall try the AR(1), ARMA(1,1) and Toeplitz forms for R. Other covariance structures theoretically may also be incorporated here but since most of them are more complex (i.e., include more parameters) we did not consider them because of computational limitations. Another reason is that other forms of covariance matrices in SAS® are not directly related to time series structures.

We evaluate and compare the models using the same technique we developed for selecting the fixed effects. We use the same learning data as before.

The results are shown in tables 6.17, 6.18, 6.19 and 6.20. The Toeplitz form did not converge for the learning data. The results for the ARMA(1,1) show only slight improvements compared to the AR(1) structure. However, one more factor that should be taken into consideration is that the CPU time was markedly higher for the ARMA(1,1) model (by several hours). In conclusion, the model chosen for R in this approach is the AR(1) model for the residual error vector.

The last source of variability is the daily volume effect, V_d. We assume its covariance

N=203	RMSE	
R covariance structure	AR(1)	ARMA(1,1)
1^{st} Quartile	32.65	32.39
Median	38.35	38.47
Mean	43.40	43.23
3^{rd} Quartile	50.94	50.74

Table 6.17: Different within-day errors covariance structure. RMSE results.

N=203	APE	
R covariance structure	AR(1)	ARMA(1,1)
1^{st} Quartile	7.83	7.83
Median	9.68	9.54
Mean	10.8	10.73
3^{rd} Quartile	12.86	12.83

Table 6.18: Different within-day errors covariance structure. APE results.

N=203	Coverage Probability	
R covariance structure	AR(1)	ARMA(1,1)
1^{st} Quartile	0.88	0.88
Median	0.97	0.97
Mean	0.93	0.93
3^{rd} Quartile	1	1

Table 6.19: Different within-day errors covariance structure comparison. Coverage results.

N=203	Width	
R covariance structure	AR(1)	ARMA(1,1)
1^{st} Quartile	141.53	138.98
Median	160.76	158.15
Mean	162.84	161.48
3^{rd} Quartile	185.40	161.70

Table 6.20: Different within-day errors covariance structure. Width results.

structure also has a first-order autoregressive form. This basic assumption means that if on a certain day the call center experienced a rise in the amount of incoming calls (compared to the fixed effects prediction) then we would also expect to see a similar increase during the following days. As the days become farther apart from that day we expect its influence to decline.

We investigated the influence of the V_d correlations by comparing our Three-Pattern model with an alternative model which does not include this random effect. For the company's current (10-day-ahead) strategy for prediction, one would hardly expect to see any difference between the two models. Since 10 days is such a long lead time we anticipated that the daily random effect would have a small influence on the results, if any.

The results are summarized in tables 6.21, 6.22, 6.23 and 6.24. In contradiction to our expectations, it seems that the daily random effect is an important one. One possible explanation for such an outcome is that by modelling the between day correlations we also influence other parameter estimates in the model (making them more *smooth*) which in turn improves the overall forecasting model.

	RMSE	
N=203	Three-Pattern without daily random effect	Three-Pattern (with daily random effect)
1^{st} Quartile	32.03	32.65
Median	38.63	38.35
Mean	44.53	43.40
3^{rd} Quartile	54.12	50.94

Table 6.21: Testing the influence of the daily random effect. RMSE results.

	APE	
N=203	Three-Pattern without daily random effect	Three-Pattern (with daily random effect)
1^{st} Quartile	7.86	7.83
Median	10.09	9.68
Mean	11.03	10.80
3^{rd} Quartile	13.28	12.86

Table 6.22: Testing the influence of the daily random effect. APE results.

	Coverage Probability	
N=203	Three-Pattern without daily random effect	Three-Pattern (with daily random effect)
1^{st} Quartile	0.85	0.88
Median	0.94	0.97
Mean	0.91	0.93
3^{rd} Quartile	1	1

Table 6.23: Testing the influence of the daily random effect. Coverage results.

	Width	
N=203	Three-Pattern without daily random effect	Three-Pattern (with daily random effect)
1^{st} Quartile	136.12	141.53
Median	151.61	160.76
Mean	154.63	162.84
3^{rd} Quartile	174.95	185.40

Table 6.24: Testing the influence of the daily random effect. Width results.

6.3 US Bank

We do not repeat the whole selection process for the US bank data for several reasons: (1) The bank's data does not include billing cycles and as such only requires the daily pattern analysis which was carried out; (2) Since each day is divided into 169 intervals the computational efforts are significantly increased and as a result the ARMA(1,1) covariance structure leads to a model which does not converge, and therefore we are left only with the AR(1) choice; (3) In the bank data we only carry out one-day-ahead predictions so we do not examine the importance of the inter-day autoregressive structure.

6.3.1 Determining Fixed Effects and Covariance Structure

The random effects are modelled using the same settings as our original model; i.e. both follow an AR(1) structure. However, we fitted a slightly different version of the daily-pattern fixed effects since the USA working days and weekday patterns differ from the Israeli cellular call center ones. Recall that the normalized version of the data was plotted in Figure 3.4. As already noted, three interesting facts are depicted in this figure: (a) Mondays have an early start compared to the rest of the weekdays; (b) Fridays have a slower decrease at the end of the day; (c) the remaining weekdays appear to be similar. Based on these observations, we set aside Mondays and Fridays and tested the hypothesis that during each period k all the remaining weekdays are not significantly different. We tested the last statement using a simple ANOVA model similar to the one described in (6.3). The ANOVA underline assumptions of normality and homoscedasticity were examined using different plots and were found to hold for the data.

Figure 6.1 is a QQ-plot of 169 P-values corresponding to the ANOVA results for each period. The results show that for most parts of the day the 3 weekdays are quite similar. The results show that the percentage of significant p-values are far greater than the 5% of significant periods that one would expect under the intersection null hypothesis. Despite these results, due to parsimony considerations, we restricted the analysis to 3 different weekday patterns: one for Monday, one for Friday and one for the rest of the weekdays.

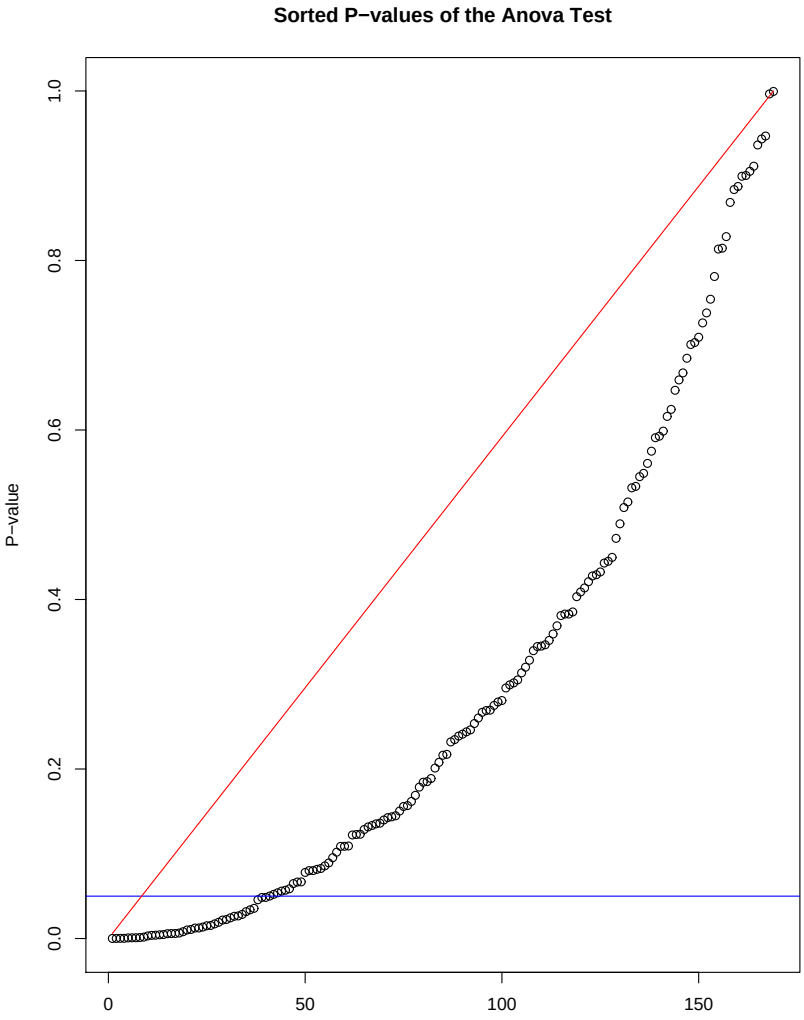

Figure 6.1: P-values QQ-plot for the ANOVA by periods of the US Bank data. The blue line corresponds to the 5% value. The red line represents the uniform distribution.

Chapter 7

Results of Prediction

7.1 Israeli Cellular Phone Company

In this section we will define and analyze a few goodness-of-fit criteria based on the mixed model predictions. Some of these criteria evaluate how well our mixed model can perform if implemented with the aim of achieving a particular QED regime. We will also compare the Poisson Bayesian model to the mixed model results.

7.1.1 Mixed Model Analysis

Goodness of Fit Figure 7.1 presents the QQ-plot for the residuals of the Three-Pattern mixed model. Consequently, it is apparent that the residuals normality assumption holds for most of observations.

Tables 7.1 and 7.2 shows the different values of the estimated error variance (σ_ϵ^2) using the two alternatives, once using an ARMA(1,1) structure for the between-periods covariance structure and once with an AR(1) structure (both techniques are detailed in Section 5.1.4). Both analyses exhibit very similar values. The two averages of the estimated variance are 0.309 and 0.31 which are quite close to the expected theoretical value of 0.25.

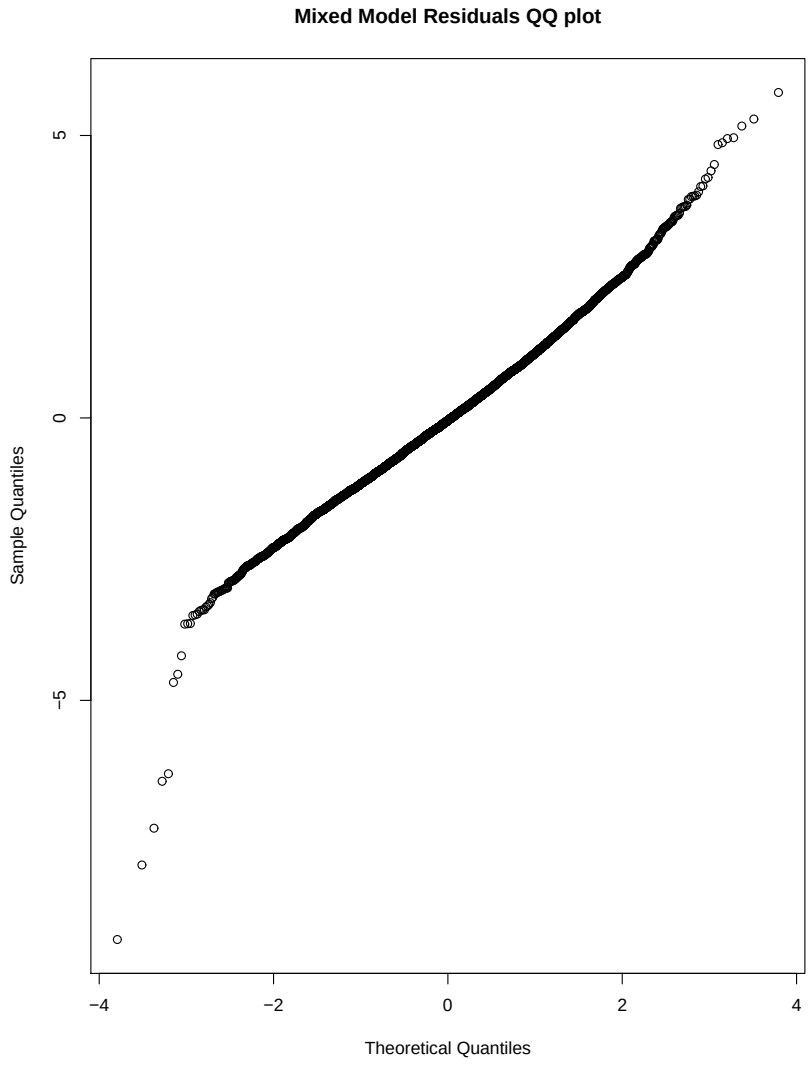

Figure 7.1: The Three-Pattern mixed model residuals QQ plot

Learning Set No.	Estimated Variance	Learning Set No.	Estimated Variance
1	0.199	20	0.320
2	0.185	21	0.352
3	0.183	22	0.390
4	0.136	23	0.376
5	0.092	24	0.382
6	0.230	25	0.387
7	0.288	26	0.416
8	0.309	27	0.380
9	0.293	28	0.369
10	0.309	29	0.333
11	0.319	30	0.324
12	0.319	31	0.338
13	0.340	32	0.334
14	0.324	33	0.340
15	0.314	34	0.320
16	0.309	35	0.333
17	0.326	36	0.320
18	0.338	37	0.301
19	0.311	-	-

Table 7.1: The estimated variance of ϵ using the AMRA(1,1) structure. The average value is 0.309 and the standard deviation is 0.07.

Learning Set No.	Estimated Variance	Learning Set No.	Estimated Variance
1	0.260	20	0.320
2	0.227	21	0.352
3	0.193	22	0.390
4	0.192	23	0.376
5	0.086	24	0.381
6	0.131	25	0.387
7	0.288	26	0.416
8	0.310	27	0.379
9	0.293	28	0.369
10	0.301	29	0.333
11	0.319	30	0.325
12	0.319	31	0.338
13	0.340	32	0.334
14	0.324	33	0.340
15	0.314	34	0.319
16	0.310	35	0.333
17	0.326	36	0.300
18	0.338	37	0.291
19	0.311	-	-

Table 7.2: The estimated variance of ϵ using an AR(1) as the between-period covariance structure. The average value is 0.31 and the standard deviation is 0.07.

Lead Time Effect Our prediction process has three user defined elements: the learning time; the prediction lead time; the forecasting horizon. During our model's training stage we did not change these parameters.

Some academic studies conducted in the past concentrate on producing one-day-ahead predictions or sometimes online updating forecasting algorithms. These methods, however, do not tackle the industry problem of attaining good predictions in order to produce the weekly schedule sufficiently ahead of time. Trying to cope with this problem, the cellular company actually uses a two stage process. It first produces a somewhat inaccurate forecast ten days before the desired week and then it generates another one, five days before. The second forecast, it says, slightly differs from the first one and so it is essential in order to adequately schedule agents. One interesting question which arises from the cellular company's method is what is the extent of the prediction lead time effect.

In order to test the prediction lead time effect, we ran our forecasting procedure using ten different lead times, ranging from one-day-ahead to ten-days-ahead. The learning period and the forecasting horizon stay the same; i.e., six weeks and one week, respectively.

Figure 7.2 illustrates both the average RMSE and APE behaviors as the prediction lead times grows. By looking at the average results we can confirm that the company is right: using more recent learning data does improve prediction results. However, from the figure it seems that the six-days-ahead predictions are more accurate, on average, than the five-days-ahead ones. We also examined ten boxplots for each of these measures to investigate their dispersion across the ten different lead times. The results show that the dispersion is quite the same over different lead times. We also note that there is an advantage in producing one-day-ahead predictions (i.e., zero lead time) since they seem to be considerably more accurate. Knowledge such as this may be useful for managing the workforce even if it is not possible to change the scheduling itself from one day to the next. Since we are referring to a weekly prediction, the manager might use this recent forecast to update the schedule for days later in the week.

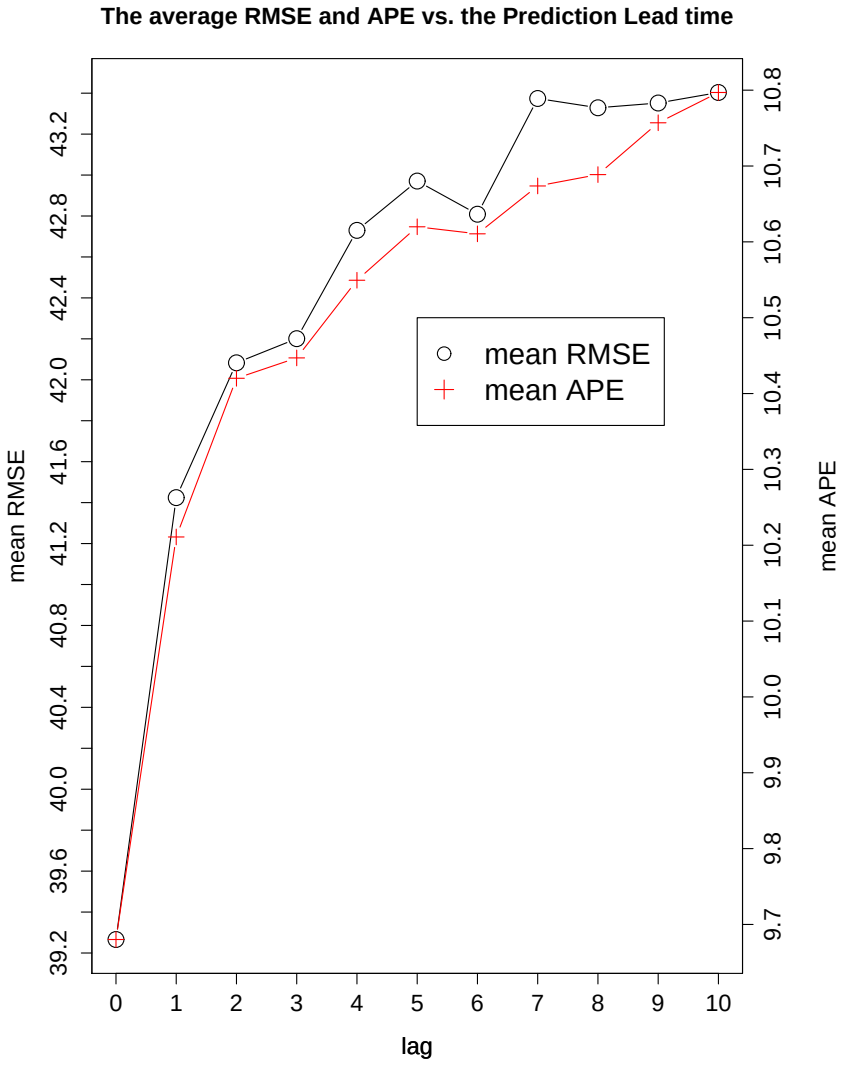

Figure 7.2: The average RMSE and APE versus the prediction lead time

Service Times Analysis We estimate parameters of the two average service times models using the SAS® *GLM* procedure. The learning data include dates between mid-February, 2004 and the end of December, 2004. Examining the first model results shows that the interaction between the quadratic period term and the weekday is not significant. Consequently, we also examined the first model excluding the insignificant term. This last model is referred to as Model 3. Since our data has a large number of observations (8382 which correspond to 254 regular days), we use the asymptotic log-likelihood ratio chi-square test to compare the models. Table 7.3 summarizes the results of the different models. We compare Model 3 to Model 2 to check if the generalized model is significantly better than the reduced quadratic model. The relevant chi-square statistic equals 58.104 and the appropriate p-value is approximately one. Hence it seems that the generalized model (i.e., Model 2) is not significantly better in modelling the average service times. Hence, we choose Model 3 for our forecasting model. Figure 7.3 illustrates a typical average service time prediction curves for each weekday.

By comparing the predictions to the true service means in the same manner as we did with the arrival process analysis, we calculate the mean APE. Its value is 8.54%. The predictions and this last result will later be used to estimate different measures of system loads.

Model No.	No. of parameters	Error SS
1	19	947.388
2	198	890.217
3	14	948.321

Table 7.3: Average service time models. Model 1 assumes a different quadratic curve for each weekday. Model 3 is the same as the Model 1 excluding the interaction between the quadratic period term and the weekday. Model 2 is the generalized model which assumes a different pattern for each weekday.

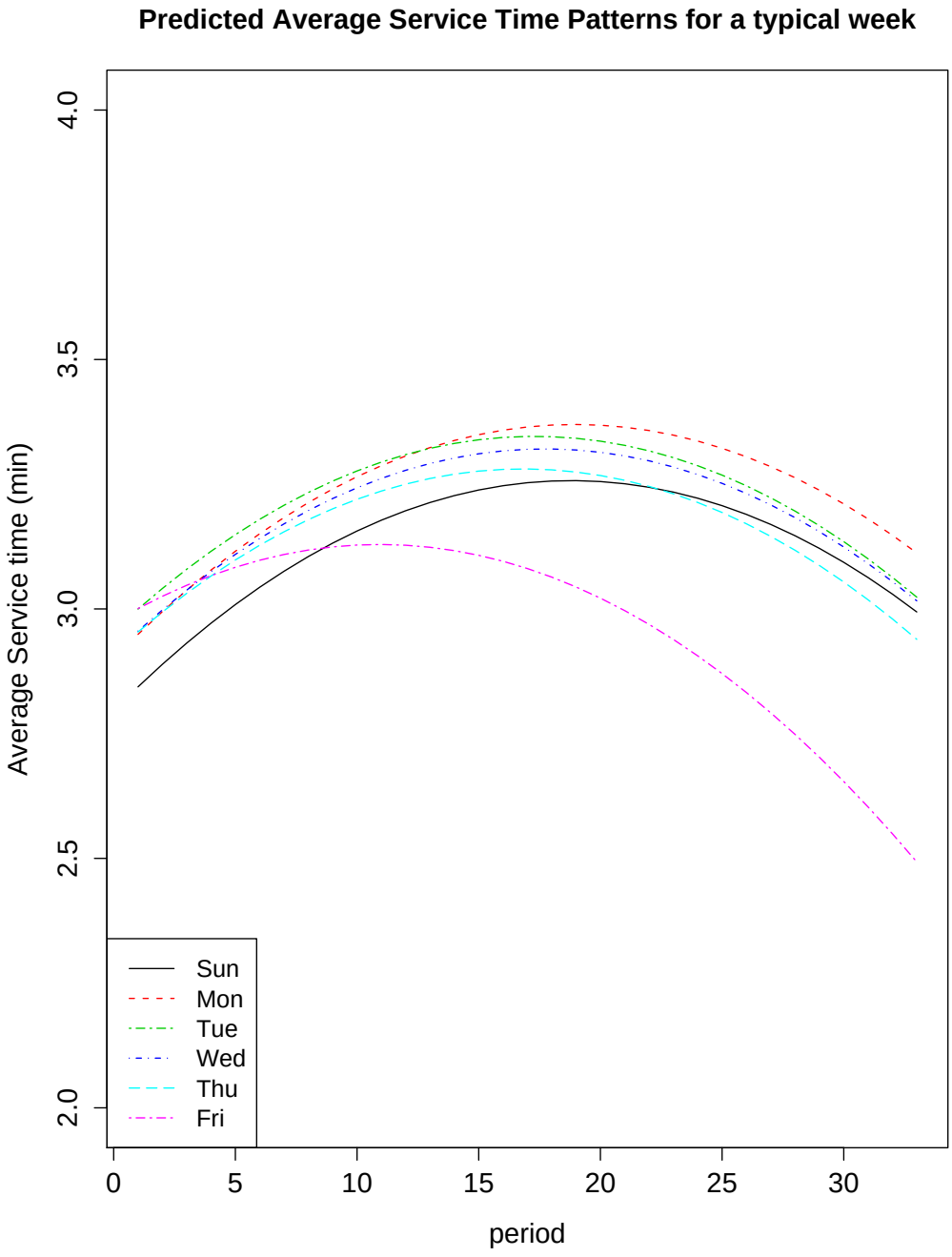

Figure 7.3: The average service pattern for typical weekdays as a function of period.

QED regime Analysis According to the QED regime, the number of service providers, S, can be determined by the following relation:

$$S = \lceil \frac{\lambda}{\mu} + \beta \cdot \sqrt{\frac{\lambda}{\mu}} \rceil \tag{7.1}$$

where λ is the arrival rate and μ is the service rate. By determining β, the call center manager actually sets the company's costs and staffing policies. In a queueing system with abandonments, such as ours, β can have both negative and positive values. The quantity β is a function of the ratio between the cost of customers' delays and for abandonment and the cost of staffing an agent. Practical values of β are between -1 and 2. Figures demonstrating the β function in an Erlang-C system, where there are no abandonments can be found in [14].

In this paragraph, we introduce two measures that evaluate forecast performance with respect to the QED "square-root staffing" rule. The first one estimates the deviation between the pre-determined β (β_u) and the true value of β (β_a) for the load actually observed assuming that the average service time can be perfectly forecast. The second measure estimates the deviation between the scheduled number of agents ($\hat{S}$) and the actual required number of agents (S) using units of the square-root of the offered load. In addition, this second measure incorporates the average percentage error between the predicted and the actual average service time. The two measures provide similar managerial information, however, the second measure has a more natural interpretation.

We begin by defining β_u as the user (i.e. the call center manager) chosen β. Now we know that the user will use the predicted value of the arrival and service rates in order to set the number of required agents, $\hat{S}$. Hence we know that:

$$\hat{S} = \frac{\hat{\lambda}}{\hat{\mu}} + \beta_u \cdot \sqrt{\frac{\hat{\lambda}}{\hat{\mu}}} \tag{7.2}$$

In practice, the assigned agents need to deal with the true value (λ) of the arrival rate and the actual service rate (μ). Realizing the above, one can say that with the real values of λ and μ, the call center is in effect operating under a different value of β. This adjusted value of β will be referred to as β_a. Formally we can write:

$$\hat{S} = \frac{\lambda}{\mu} + \beta_a \cdot \sqrt{\frac{\lambda}{\mu}} \tag{7.3}$$

By equating the above equations we come to the following results:

$$\begin{aligned} \frac{\lambda}{\mu} + \beta_a \cdot \sqrt{\frac{\lambda}{\mu}} &= \frac{\hat{\lambda}}{\hat{\mu}} + \beta_u \cdot \sqrt{\frac{\hat{\lambda}}{\hat{\mu}}} \\ &\Rightarrow \\ \beta_a - \beta_u \cdot \sqrt{\frac{\hat{\lambda}}{\lambda} \cdot \frac{\mu}{\hat{\mu}}} &= \frac{\frac{\hat{\lambda}}{\hat{\mu}} - \frac{\lambda}{\mu}}{\sqrt{\frac{\lambda}{\mu}}} \end{aligned} \tag{7.4}$$

Let us assume for now that we can predict the rate μ perfectly which means that $\mu = \hat{\mu}$ (which is a fairly reasonable assumption in practice). From earlier results we know that the average arrival rate APE is about 0.1 and hence the square root term is approximately $1.049 \approx 1$. Under the above assumptions we conclude that:

$$\Delta\beta \triangleq \beta_a - \beta_u \approx \frac{\hat{\lambda} - \lambda}{\sqrt{\lambda \cdot \mu}} \tag{7.5}$$

This difference will be referred to as the $\Delta\beta$ measure. Examining it can help determine how well our forecasting method behaves. It can answer questions like: does this forecasting algorithm usually over-estimate or under-estimate the number of arrivals and by how many agents. Note that the desired value of $\Delta\beta$ is zero, indicating a perfect point-prediction of the arrival counts.

We evaluate $\Delta\beta$ using the estimated average service times (i.e., average service time= 1/service rate). In Figure 7.4, we examine the averaged $\Delta\beta$ values across the 33 periods of the day using our final mixed predictions. To obtain these values of the estimated $\Delta\beta$, we first estimate $\Delta\beta$ for each day in our learning data during each period. Afterwards, we average for each period separately over all the days, excluding holidays and irregular days.

The small values of $\Delta\beta$ indicate that our model does quite well in predicting the value of the arrival rate. It also indicates that the user's β_us are very close to the real β_as. The estimated average values are close to zero but are usually greater than zero which means that for most parts of the day the predictions would lead to some over-staffing. During the early morning hours and at the end of the day, the values which we observe make sense since the arrival rates during those hours are usually low which corresponds to large values of $\Delta\beta$.

After examining the values of average $\Delta\beta$ one can also examine the dispersion of $\Delta\beta$ throughout the day by looking at boxplots for each period as shown in Figure 7.5. From these results we can say that most of the $\Delta\beta$ have absolute value less than 2. There are

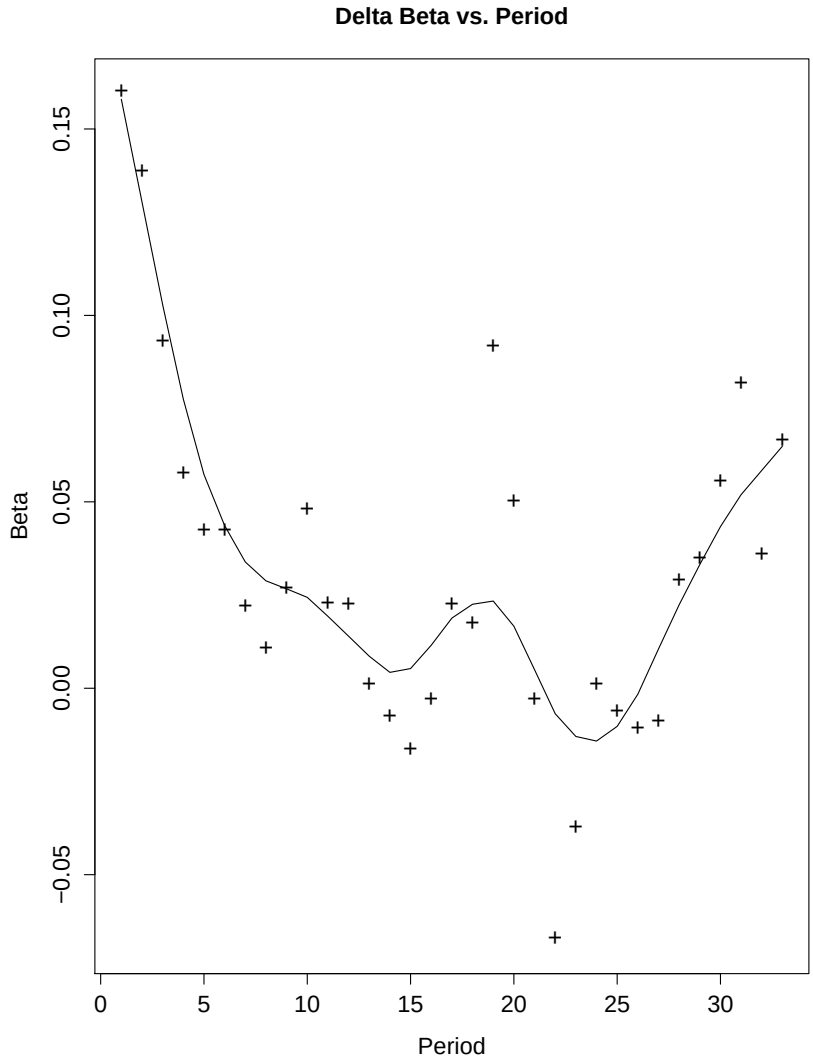

Figure 7.4: The average estimated $\Delta\beta$ as a function of period.

five relatively large $\Delta\beta$ values. The three which occur during periods 31,32 and 33 all come from the same day, July 27th, which has an unusual drop in the number of incoming calls during those periods. The two outliers located during the 5th and 6th periods are also related to the same date, September 19th. During these periods there is also a peculiar drop in the number of incoming calls which is unexplained.

In conclusion, the boxplots due indicate that 50% of $\Delta\beta$ values are between -0.5 and 0.5 and on average are zero. Given that practical values of β are between -1 and 2, on a large proportion of the periods the mixed model will not make a gross error in the staffing.

Now let us remove the assumption made before about the service rate. This leads to the

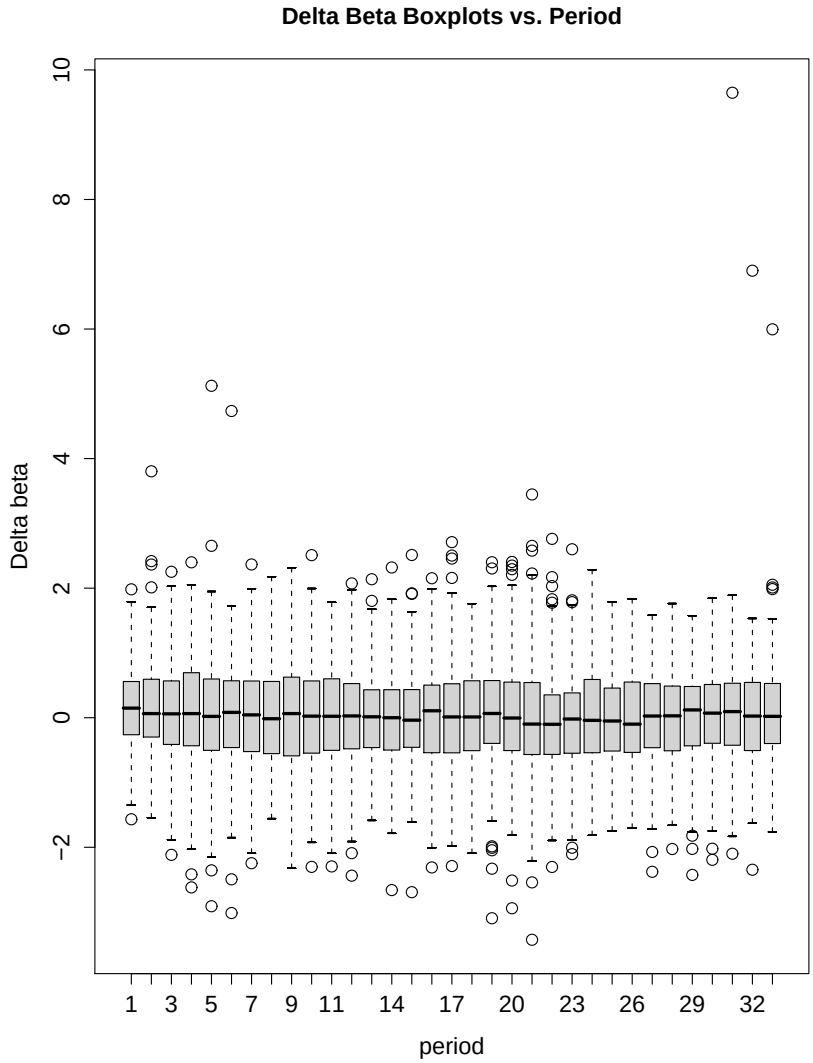

Figure 7.5: Boxplots of $\Delta\beta$ for the different periods.

problem of predicting the mean service times in addition to the arrival rate during each half-hour. Using the same methodology as before, one can say that the user has to predetermine a value for β, and based on the predicted values of both λ and μ set the number of required agents $\hat{S}$ in the following manner:

$$\hat{S} = \frac{\hat{\lambda}}{\hat{\mu}} + \beta \cdot \sqrt{\frac{\hat{\lambda}}{\hat{\mu}}} \tag{7.6}$$

Knowing the true values of λ and μ the user would still use the same β but would now get the desired number of agents, i.e. S. This number is the correct number of agents needed to handle the actual system load at the desired quality and efficiency trade-off. Following

this notation we can write the following equation:

$$S = \frac{\lambda}{\mu} + \beta \cdot \sqrt{\frac{\lambda}{\mu}} \tag{7.7}$$

From the above two formulas one can deduce the following:

$$\begin{aligned} \frac{\hat{S} - S}{\sqrt{\frac{\lambda}{\mu}}} &= \frac{\frac{\hat{\lambda}}{\hat{\mu}} - \frac{\lambda}{\mu}}{\sqrt{\frac{\lambda}{\mu}}} + \beta \left(\frac{\sqrt{\frac{\hat{\lambda}}{\hat{\mu}}} - \sqrt{\frac{\lambda}{\mu}}}{\sqrt{\frac{\lambda}{\mu}}} \right) \\ \frac{\hat{S} - S}{\sqrt{\frac{\lambda}{\mu}}} &= \frac{\frac{\hat{\lambda}}{\hat{\mu}} - \frac{\lambda}{\mu}}{\sqrt{\frac{\lambda}{\mu}}} + \beta \left(\sqrt{\frac{\hat{\lambda}\mu}{\lambda\hat{\mu}}} - 1 \right) \end{aligned} \tag{7.8}$$

Using the mean APE for the service rate and arrival rate the second term in (7.8) is small. Hence, we come to the conclusion that:

$$\frac{\hat{S} - S}{\sqrt{\frac{\lambda}{\mu}}} \approx \frac{\frac{\hat{\lambda}}{\hat{\mu}} - \frac{\lambda}{\mu}}{\sqrt{\frac{\lambda}{\mu}}} \tag{7.9}$$

We shall refer to the above measure as ΔQED . It enables one to evaluate the difference between the actual and the desired number of agents, normalized by the square root of the actual offered load. Normalizing with the square-root of the offered load is natural using the following reasoning: in the QED regime, we add (or deduct) a factor (β) of square-root of the offered load to ensure adequate staffing levels. Using ΔQED we can evaluate by how many units of the square-root of the offered load our staffing level $\hat{S}$ deviated from the required level S. This is similar to evaluating the number of unit deviations between the user defined β_u, and the adjusted β_a (defined above).

We proceed by calculating ΔQED using the same data as before but this time we are also incorporating the mean service times predictions produced by the model described in the previous paragraph. Figure 7.6 demonstrates the results. Because the first two values of ΔQED are very large they distort the figure and its very hard to make sense of things. For this reason, we add Figure 7.7 which gives a zoomed-in view for the periods between 8:30AM and 11:30PM.

From the results we can see that during the afternoon hours we are predicting the offered load quite well. During the second peak period (which occurs at around 19:30) we are under-estimating the offered load and hence under-staffing the call center by 1 – 2 agents. In the early morning hours, this deviation might cause problems since only a few agents

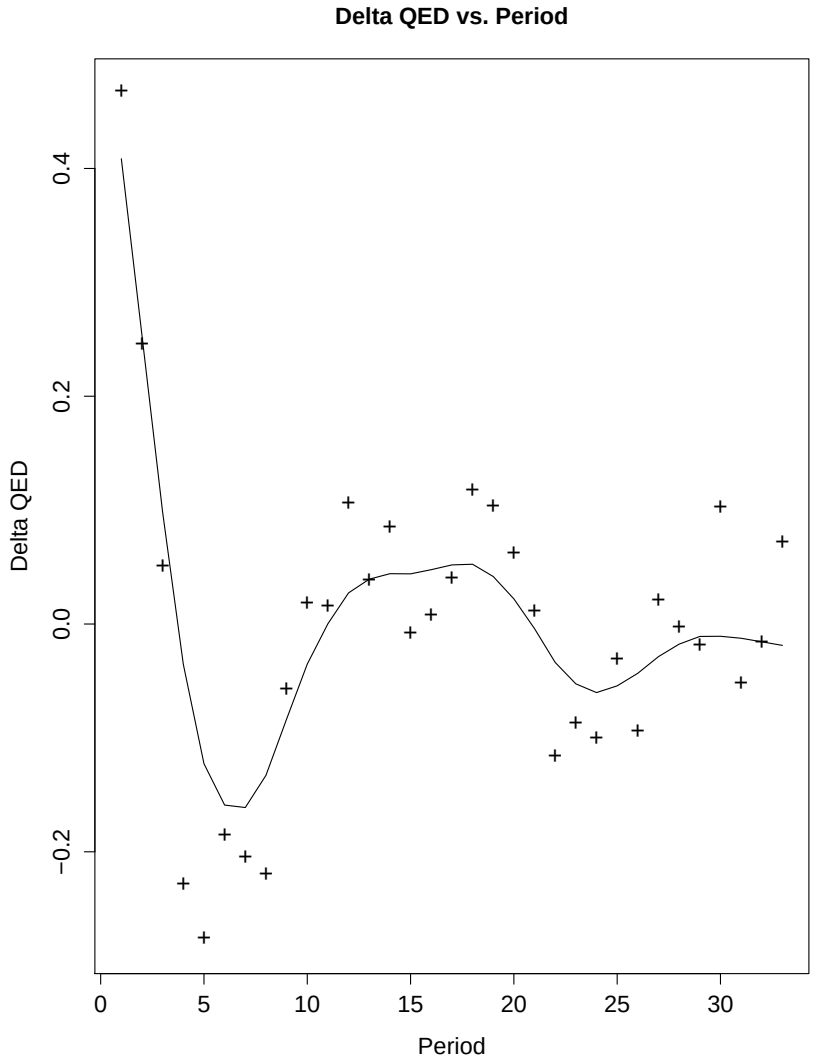

Figure 7.6: The average estimated Δ QED as a function of period.

are available. However, it is very important to also note that during these hours the QED regime is likely to be inappropriate and hence this measure is irrelevant during those periods.

The next natural step of this analysis is to study the effect of the predictions deviations on service level measures. In the process of scheduling, a manager will ultimately decide what is the required number of agents according to some pre-determined service level goals. These goals are translated into measures such as: the customer's probability to wait, the average waiting time or the probability of a customer to abandon. It would be interesting to study the sensitivity of these measures to errors of the predicted service times and arrival counts, compared to their actual values.

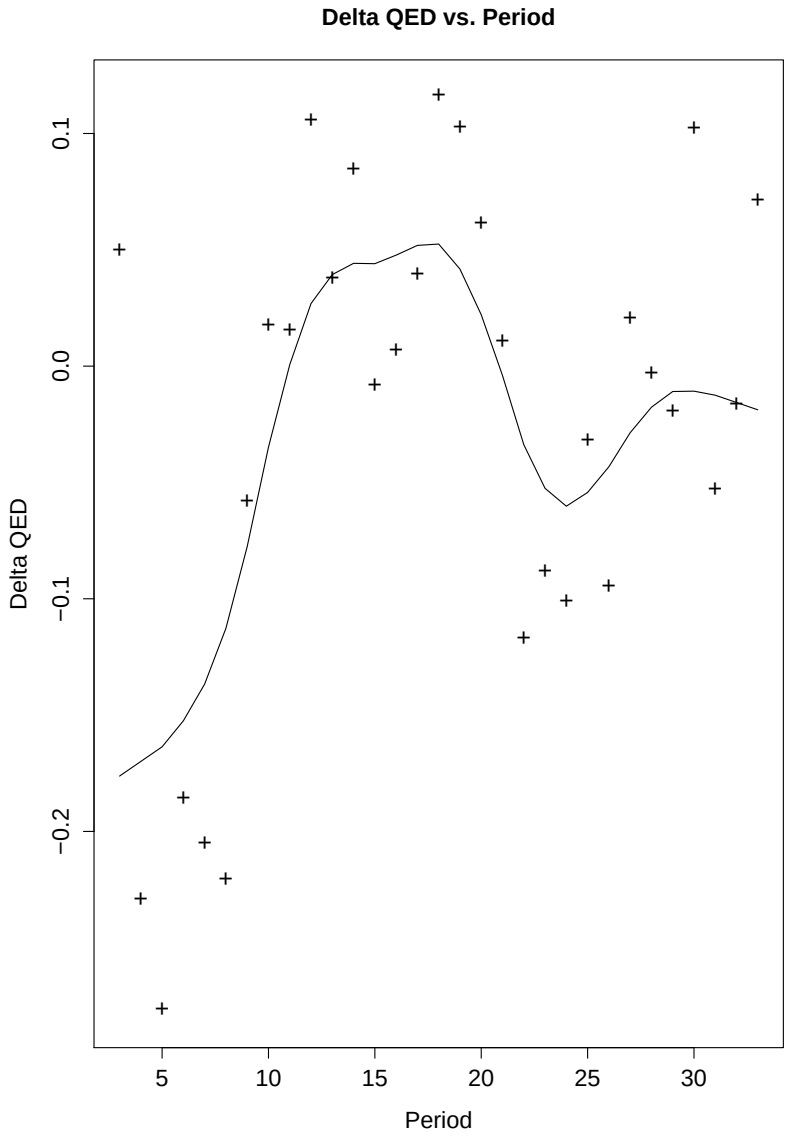

Figure 7.7: The average estimated Δ QED versus periods between 8:30 and 23:30.

Variability Measure Analysis One of our research goals is to quantify the uncertainty in the arrival process; for example, trying to decide if there are specific periods during the day which are harder to predict than others.

As we already saw, the daily patterns differ among weekdays. Taking a slightly different approach than before we begin by investigating the natural cluster hierarchy hidden in the daily patterns. We explore this by implementing an agglomerative clustering algorithm on the normalized daily patterns. We used the 'agnes' command located in the 'cluster' package (see [11]) available in the R statistical software. The results show that there are two large clusters: one contains only weekends and the other has all the rest of the week-

days. We continued exploring the weekdays cluster which contains most of the dataset (including irregular days). We studied this cluster by trying to divide it into different smaller groups. We did this by running the k-medoids algorithm (detailed in the second chapter of [10]) with different k-values ranging from 2 to 10. Based on the silhouette coefficient we arrived at the conclusion that the weekdays cluster can be further divided into two clusters which can be identified as holidays and regular weekdays. The holidays identified using this procedure correspond to the list of holidays provided in Table 3.1. Taking out the five outlier days that we initially listed in Table 3.2 we are left with only the regular weekdays in the other cluster.

After this primary procedure we now have a cluster containing days which are similar to each other based on the Euclidian distance measure between daily arrival profiles.

Let us assume for a moment that we would have to predict each period arrival count based on the naive estimator, i.e. the daily average during that period. Since all the days in the cluster are similar we would use all of them to calculate each period mean value. Note that this procedure only uses periods information to produce the prediction as opposed to our first benchmark approach (detailed in (5.14)) which also incorporated the weekday data.

We shall now investigate how much variability is present in each period when the forecast is restricted to this basis (i.e.,period) by comparing between each period mean value and its RMSE. Figure 7.8 shows the results of the naive model comparison.

The next evident step that suggests itself from the above figure, is to calculate the linear regression curve estimators since the points seem to fall on an almost straight line. In a perfect Poisson distributed environment with an arrival rate of λ, we know that the variance, σ^2, equals the arrival rate, λ. Taking log on both sides of this last equation one would expect the following relationship between the log standard deviation and the log expected value:

$$\log \sigma = \frac{1}{2} \cdot \log \lambda \tag{7.10}$$

So for a collection of Poisson variates one would expect to see a straight line when plotting their log-RMSE's against their log-mean values. The estimated value of the slope should be close to 0.5 and the intercept estimator would be close to zero. Our naive linear regression results are presented in Table 7.4, and show that the intercept is not significantly different from zero. Looking at the slope estimator we can observe the well known *over-dispersion* phenomena mentioned in several articles such as [3] and [5].

Using the same method as implemented on the naive model predictions, we can compare

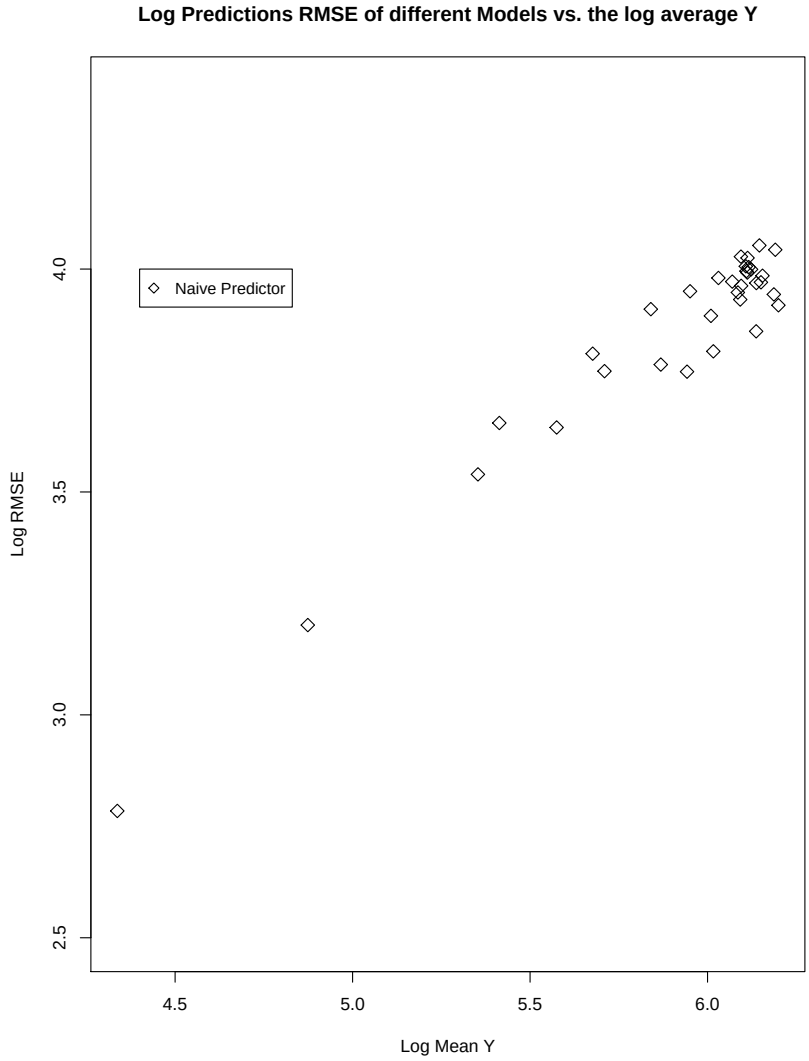

Figure 7.8: Plot of the log RMSE versus the log mean arrival value based on the naive predictor. Each point is for a different period.

Models Name	Intercept	Intercept P-value	Slope	Slope P-value	R-Squared
Naive Model	0.202	0.231	0.617	$1.55 \cdot 10^{-20}$	0.9403

Table 7.4: The naive model linear regression estimators.

between different models. Specifically, plotting the RMSE against the mean arrival value for each period. The analysis is carried out on two additional models: the first benchmark (industry) model, defined in (5.14) and the final three-pattern mixed-model, defined in (5.2).

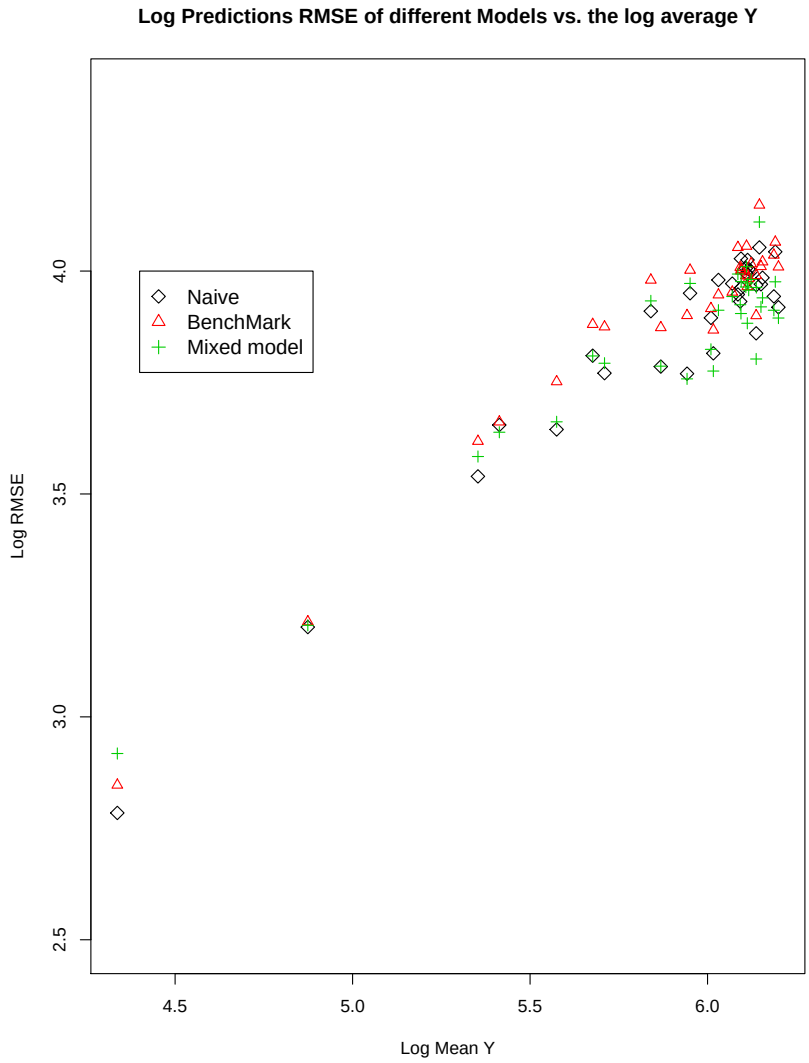

Figure 7.9: For each period, a plot of the log RMSE versus the log mean arrival value based on the naive predictor, the first benchmark model and the mixed-model.

Figure 7.9 shows the comparison between the three models: the naive model and the two additional ones. There is a large cluster on the right side of the figure. Figure 7.10 focuses on periods with higher arrival mean values which corresponds to this cluster. By looking at both graphs, one can see that the mixed model has lower RMSE during most of the periods. The naive model has lower RMSE than the benchmark model. The observations located on the lower left side of the graph correspond to the morning hours. During these morning hours, it seems that the naive model is doing just as well and even better than the other two models. However, one must consider that this naive model's 'predictions' (i.e.,

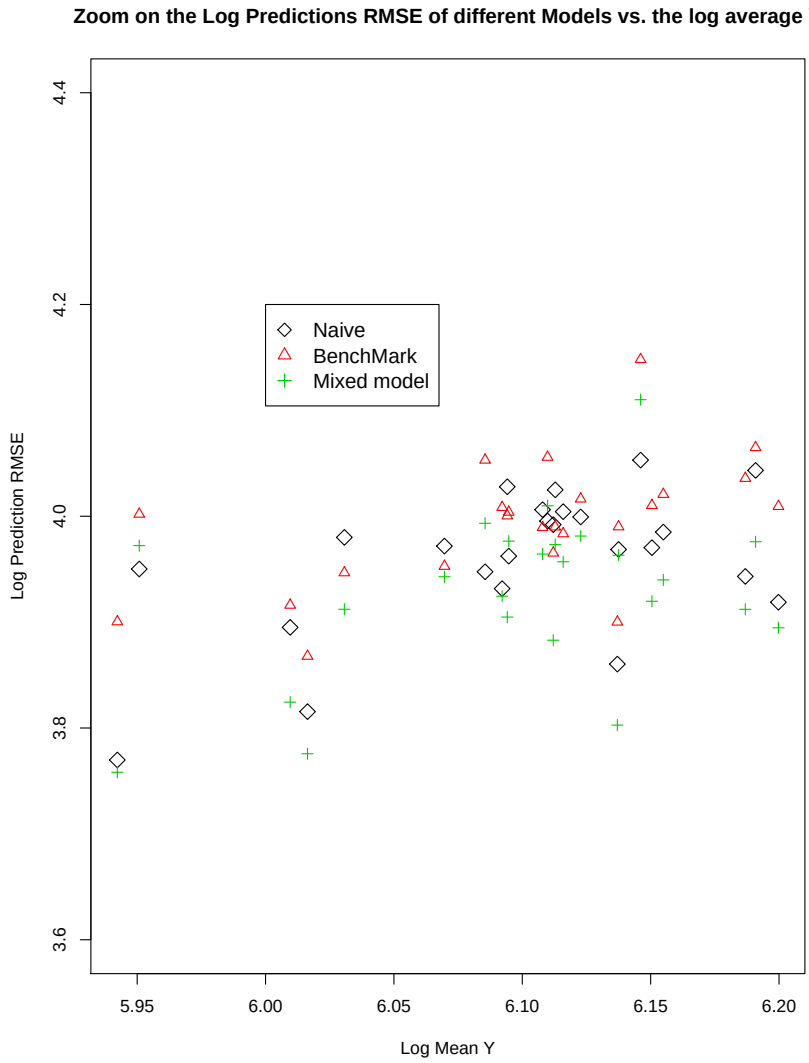

Figure 7.10: For each period, a plot of the log RMSE versus the log mean arrival value based on the naive predictor, the first benchmark model and the mixed-model. This is a zoom-in graph that focuses on periods with higher mean arrival counts.

mean values) are based on the entire 203 days for each of the periods whereas the other predictions are based on only 6 weeks.

Looking at Table 7.5, we can compare the linear regression estimators of the three models. As can be seen, the mixed model slope is the closest to the “natural” 0.5 value but it also has a significant intercept. The positive intercept on the logarithmic scale means that even if the dependence of standard deviation on the mean has a square-root form, there remains nevertheless some unexplained over-dispersion.

Models Name	Intercept	Intercept P-value	Slope	Slope P-value	R-Squared
Naive Model	0.202	0.231	0.617	$1.55 \cdot 10^{-20}$	0.9403
Benchmark Model	0.344651	0.0492	0.601	$5.699 \cdot 10^{-20}$	0.9315
Three Pattern Mixed-Model	0.622	0.003	0.543	$3.158 \cdot 10^{-20}$	0.901

Table 7.5: The naive, benchmark and mixed model linear regression estimators.

The last thing we shall examine is the one-day-ahead predictions versus the ten-day-ahead ones. Adding these two additional models brings us to a total of five models to compare: the naive model, the benchmark model ten-day-ahead, the benchmark model one-day-ahead (ODA), the three-pattern mixed model ten-day-ahead and the three-pattern mixed model one-day-ahead (ODA).

Figure 7.11 displays the comparison between the five models. Figure 7.12 zooms in on periods with higher arrival mean values. It is evident that there is a reduction in the RMSE when we use a shorter lead time. As mentioned before, this comment should be considered by call center managers since even when predicting the full week, one might nevertheless consider making changes as the week unfolds.

The one-day-ahead results of the two models linear regression estimators are presented in Table 7.6. From these results one can see that the one-day-ahead estimators of both the benchmark and the mixed models have slopes closer to the theoretical Poisson slope. Comparing both the one-day-ahead and the ten-day-ahead behaviors, one can see that the difference between the slopes of the two models is about 0.05.

Models Name	Intercept	Intercept P-value	Slope	Slope P-value	R-Squared
Benchmark Model ODA	0.3091	0.0709	0.602	$3.146 \cdot 10^{-20}$	0.9375
Three Pattern Mixed-Model ODA	0.4636	0.0356	0.5522	$3.781 \cdot 10^{-16}$	0.8856

Table 7.6: One-day-ahead benchmark and mixed models — linear regression estimators.

In summary, from the above results it appears that the mixed model is successful in capturing more of the predictable variability than do the naive model or the benchmark model; in particular during the busier periods of the day. As a result, for the mixed model, the residual variability is closer to that corresponding to the inherent randomness of the Poisson distribution. Also the short, one-day-ahead, lead time plays an important role in reducing this residual variability.

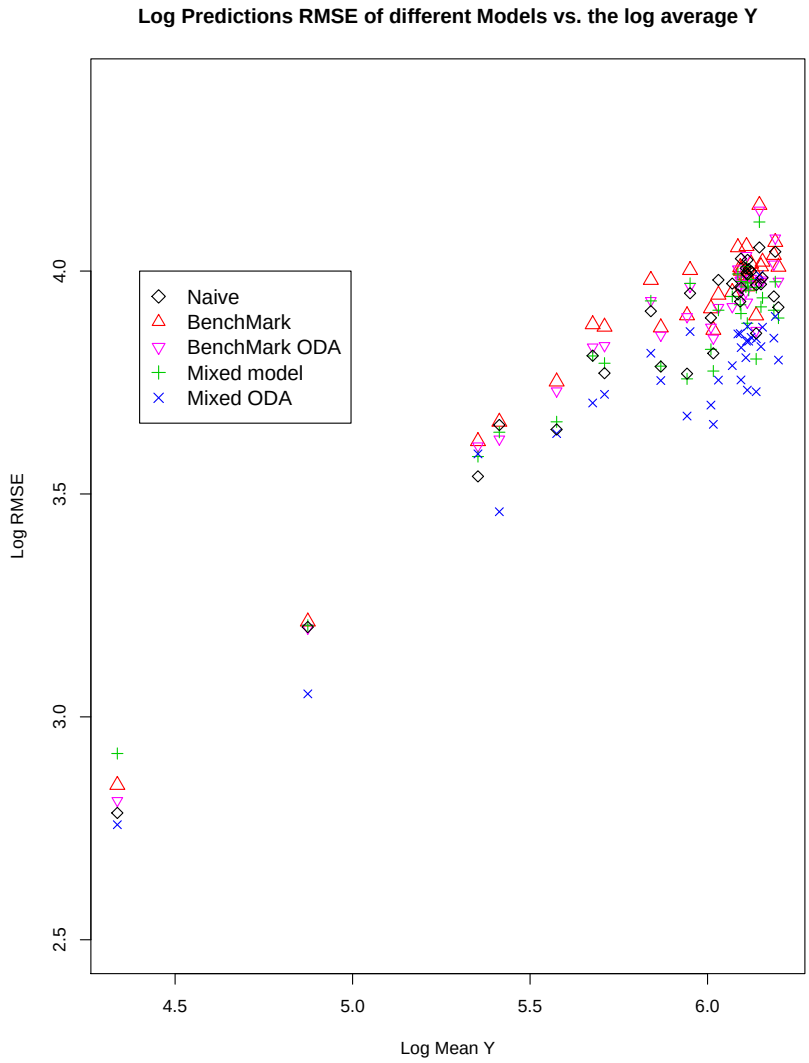

Figure 7.11: For each period, a plot of the log RMSE versus the log mean arrival value based on the naive model, the benchmark model ten-day-ahead, the benchmark model one-day-ahead, the three-pattern mixed model ten-day-ahead and the three-pattern mixed model one-day-ahead.

7.1.2 Comparison between the Poisson Bayesian Model and the Mixed Model

When applying the Poisson Bayesian model to the current version of ‘OpenBugs’, our programs, unfortunately, could not handle our original prediction problem of forecasting the full week on a ten-day-ahead basis. Moreover, because of computational difficulties,

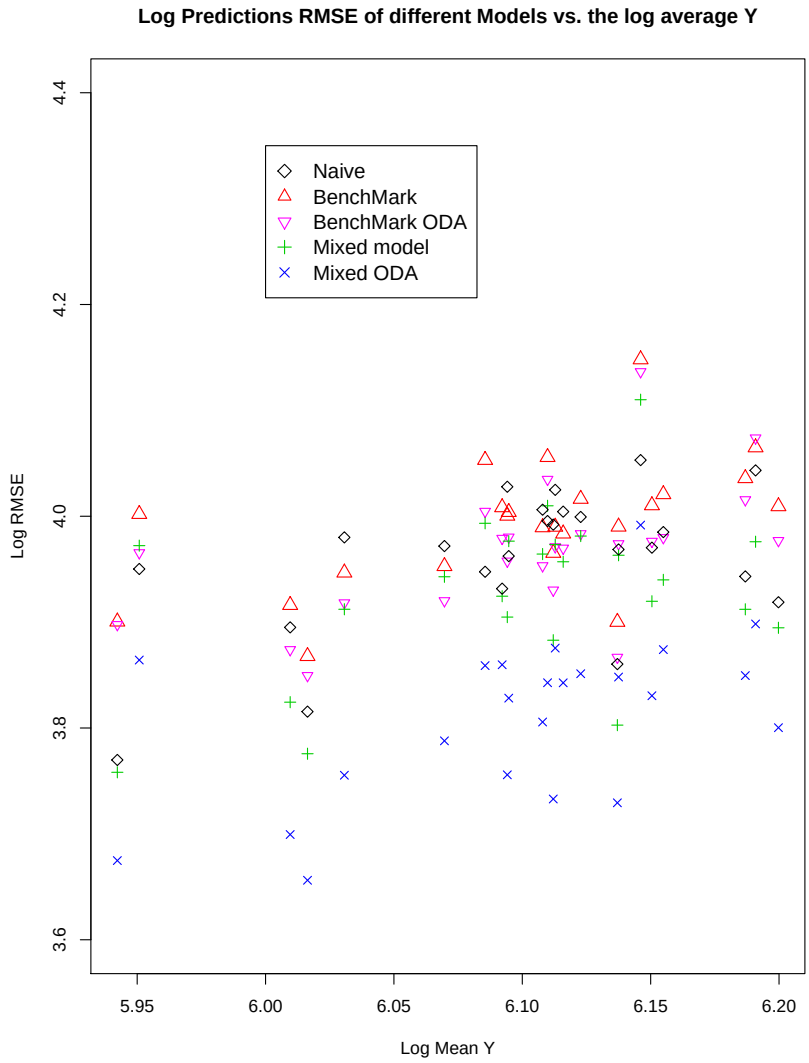

Figure 7.12: For each period, a plot of the log RMSE versus the log mean arrival value based on the naive model, the benchmark model ten-day-ahead, the benchmark model one-day-ahead (ODA), the three-pattern mixed model ten-day-ahead and the three-pattern mixed model one-day-ahead (ODA). This graph shows only periods with high arrival mean values.

the highest resolution we were able to apply to the Poisson Bayesian model was two and half hours over six periods of the day. As a result, this section will only present a 'proof of concept' for this model. We will investigate the one-day-ahead forecast results for July 1, 2004 during the six periods between 7AM-9:30PM. We used 6 weeks of past data as the learning data.

The mixed-model was also adjusted using the same settings as the Poisson Bayesian model to enable a fair comparison. For example, the model was expanded to include all seven days of the week. The billing cycles indicators were not included in the model because the computational complexity was excessively hard for the 'OpenBugs' software to handle.

We ran the Poisson Bayesian model using two Markov chains to reduce the correlations between the samples. Each chain was run for 1750 iterations after a burn-in period of 1000. The inference is carried out using the combined samples. The effective number of iterations, which is used as a crude measure of effective sample size, was approximately 1900 (for the predictive distributions of the 6 periods).

One of the main advantages of implementing the model using 'OpenBugs' is that the forecasted periods are considered to be parameters and as such one can generate their *posterior* distributions. The periods histograms and density plots are presented in Figure 7.13.

Figure 7.14 presents the predicted values attained from both the mixed-model and the Poisson Bayesian model. We used the mean value of the forecast distribution as the predicted value for each period. For most periods of *this* day the mixed-model outperforms the Poisson Bayesian model.

We also compare the prediction intervals of both methods. For the Poisson Bayesian model we used the forecast distribution 0.025 and 0.975 quantiles to determine the 95% prediction interval. From Figure 7.15 it is apparent that the Poisson Bayesian model has wider prediction intervals than the mixed-model.

Although, based on these results one might be discouraged from pursuing the Poisson Bayesian model, we believe that further investigation of this model is appropriate since one day is hardly enough to determine model adequacy.

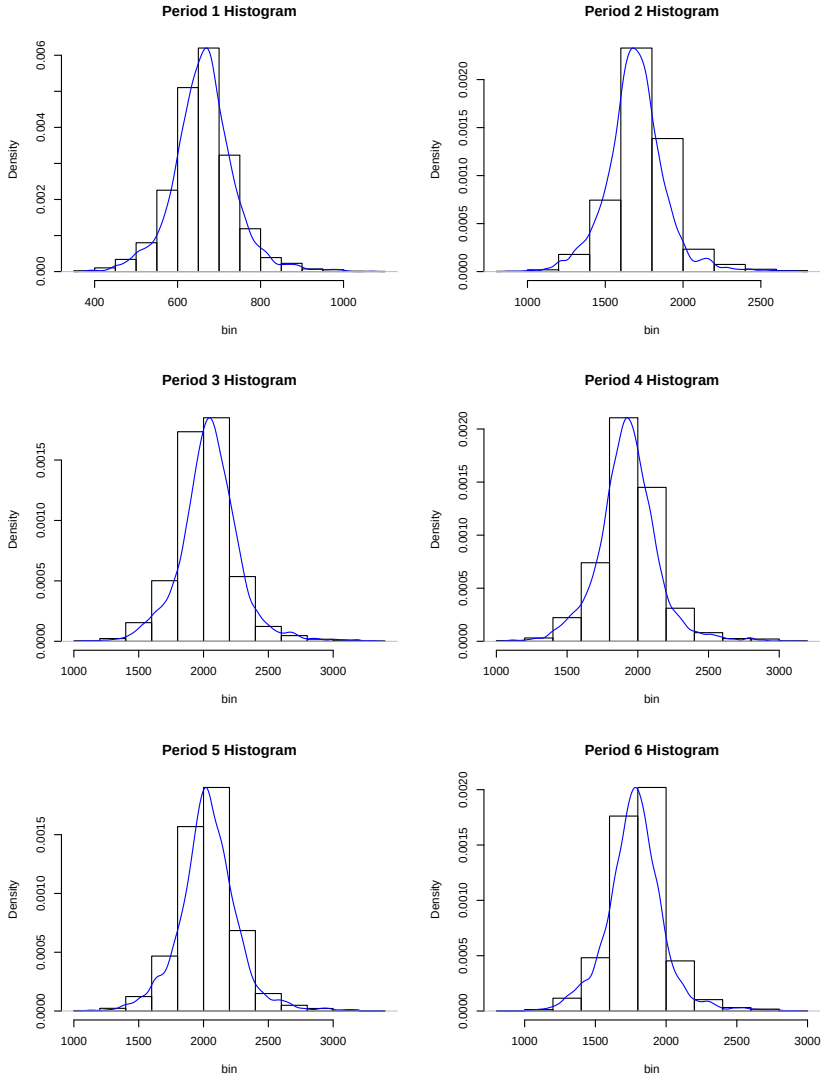

Figure 7.13: The Poisson Bayesian model predicted periods. For each of the six periods between 7AM-9:30PM the forecast distribution is plotted.

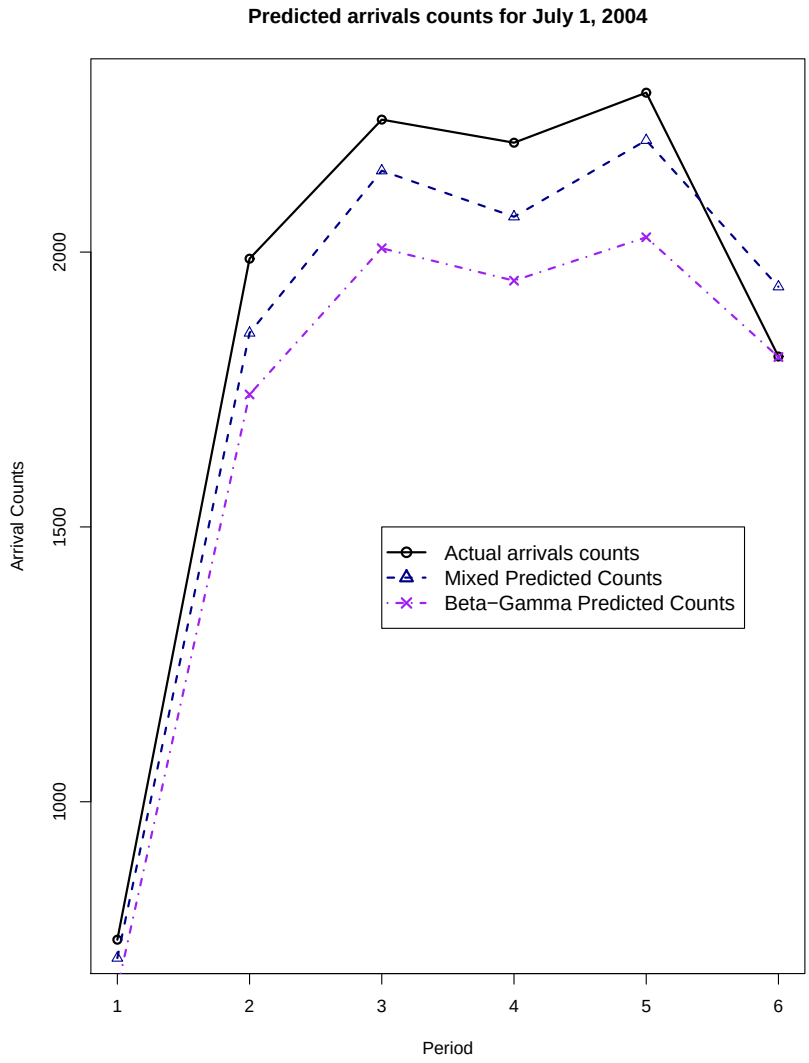

Figure 7.14: Comparison between the Poisson Bayesian model and the mixed-model predictions results.

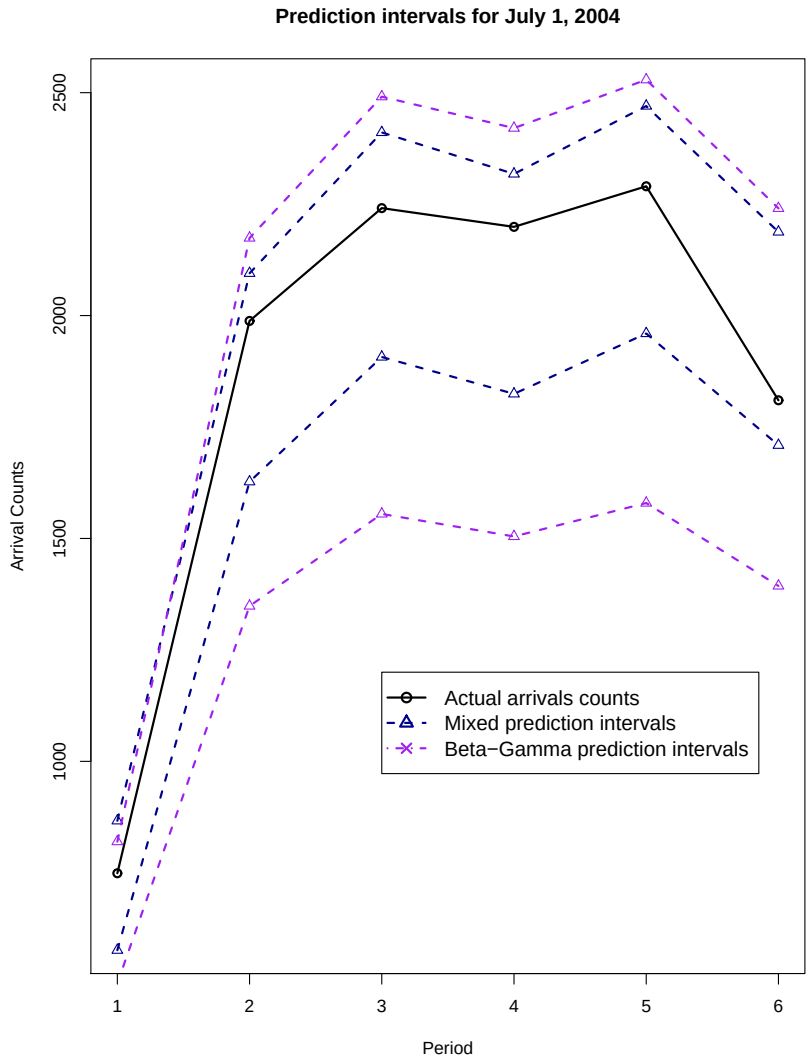

Figure 7.15: Comparison between the Poisson Bayesian model and the mixed-model prediction intervals.

7.2 US Bank

In this section we examine the normality assumption of the mixed model and compare between the Gaussian Bayesian model and the mixed model prediction results. We also discuss the practical applications of these models in a business environment.

Goodness of Fit Figure 7.16 presents the QQ-plot for the residuals of the mixed model on the US Bank data. Near the ends of the QQ-plot there are deviations from the normal distribution. However, these deviations correspond to only a few days and hence for most of the observations the normality assumption holds.

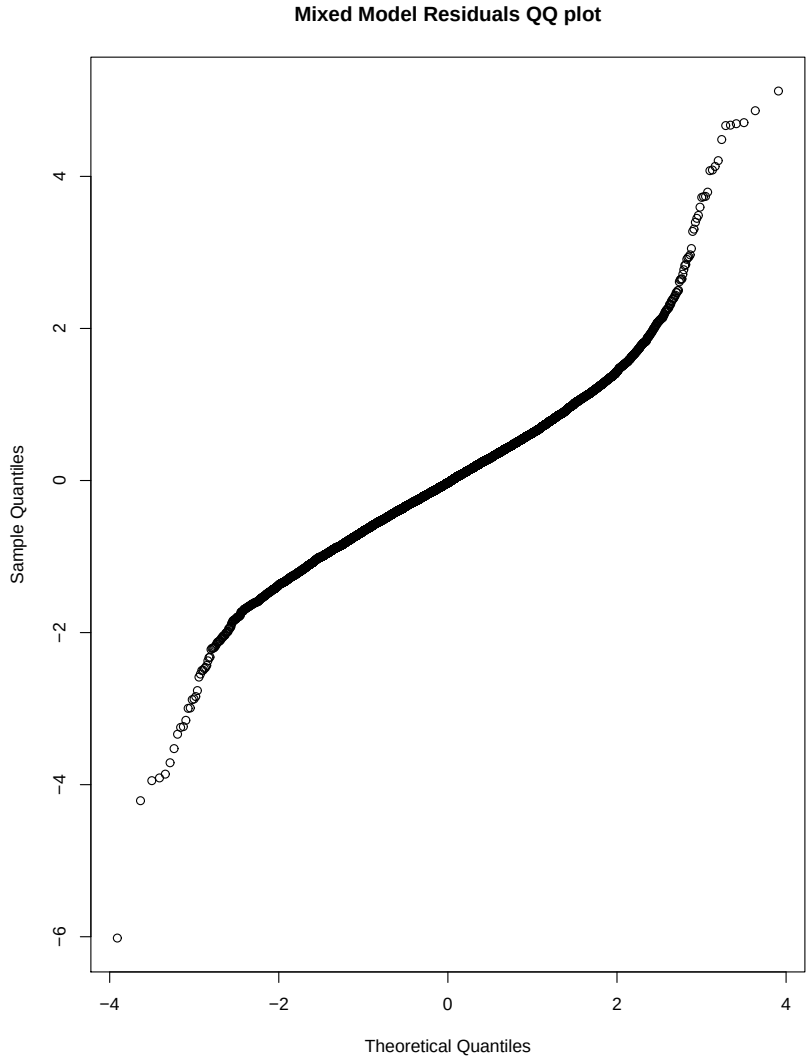

Figure 7.16: The mixed model residuals QQ plot.

7.2.1 Comparison between the Gaussian Bayesian Model and the Mixed Model

In this section we compare between the results of two main models: the mixed model and the Gaussian Bayesian model. We compare both models' forecasting performances over a period of 64 regular weekdays. This period includes weekdays between July 25 and October 24, 2003.

For the Gaussian Bayesian model, we use the performance measures reported in [22]. We did not implement this model. As described in the above mentioned article, for each day the 100 previous days were used as the learning data.

For the mixed model, we ran a one-day-ahead prediction process on the data. Because of computational problems we could not preform our original forecasting process, i.e. using six weeks of historical data as the learning data. Instead we ran the forecasting procedure twice: once using five weeks (referred to as Model 5) of past data and once using four weeks (referred to as Model 4) as the learning period.

In the previous mixed-model (i.e. for the Israeli cellular data), the prediction intervals used degrees of freedom that were calculated according to a general Satterthwaite approximation as recommended when dealing with random effects. For further details the reader is referred to the SAS® help manual located in the software itself or to the proc mixed website.

However, using this approximation on the current US bank data yielded some confusing results.

N=64	RMSE	APE	Cover	Width
Min	12.31	5.85	0.63	60.60
1^{st} Quartile	15.18	7.29	0.92	63.63
Median	17.07	7.59	0.97	69.70
Mean	19.10	8.76	0.94	139.88
3^{rd} Quartile	21.12	9.09	0.99	81.16
Max	41.01	28.02	1	501.28

Table 7.7: The mixed-model results with 4 weeks of learning data.

Table 7.7 presents the prediction results for the mixed-model using four weeks as the learning period. Looking at the width statistics, specifically the mean and maximum val-

ues, it is apparent that there are a number of days which have very wide average prediction interval. Closer examination of the results revealed that out of the 64 predicted days, 13 exhibited this abnormal behavior. These days do not have an exceptionally large prediction standard deviations. However they do have an unusual number of denominator degrees of freedom which equals one (during all periods of the day). To investigate this phenomenon, we compared the above results with those of the mixed-model procedure using the default degrees of freedom calculation method (all the rest of the settings remain the same). This default method is referred to, by the SAS® help manual, as the containment method. Using the default method, the 13 'abnormal' days exhibited significantly smaller widths. The remaining question is: what effect does the default method have on the width of the remaining 51 days. Table 7.8 summarizes the 51 days width and coverage results of both the Satterthwaite and the containment (default) method. We do not explore the RMSE and APE since they are not affected by the change in the degrees of freedom.

	Cover		Width	
N=64	Satterthwaite	Containment (Default)	Satterthwaite	Containment (Default)
Min	0.63	0.63	60.60	58.37
1^{st} Quartile	0.91	0.89	63.17	61.96
Median	0.96	0.95	67.13	66.37
Mean	0.93	0.93	68.62	67.25
3^{rd} Quartile	0.98	0.98	72.32	71.85
Max	0.99	0.99	84.06	81.07

Table 7.8: Comparing the mixed-model coverage probability and width using Satterthwaite approximation versus the containment method to calculate the degrees of freedom.

From this comparison one can see that the coverage probability statistics are very similar (some even have exactly the same values). However, the prediction interval widths are consistently shorter using the default method. This may indicate that using the Satterthwaite approximation leads to unnecessarily large prediction intervals. The reason as to why such results occur specifically when using Satterthwaite approximation on the Bank's data and not on the Cellular's data is unknown to us and might need to be further explored.

Following these results, the two versions of the mixed-model (once using four weeks of past data and once using five weeks of past data as the learning periods) are implemented

using the default method for calculating the degree of freedoms and the associated prediction intervals. Tables 7.9 and 7.10 present the results of the Bayesian model and both versions of the mixed model using all 64 predicted days data. Generally speaking, the Bayesian model outperforms the mixed model although the results are very close. The RMSE and APE maximum values are smaller in the mixed model. The mixed-models width values are smaller than for the Bayesian model but the coverage probabilities are smaller as well.

In spite of the above results, it is important to emphasize that the mixed-models daily predictions only require 20-25 learning days in comparison to the 100 days that the Bayesian model required. Moreover, the daily forecast took about twenty-thirty minutes for each day. Article [22] authors do not specify the learning time or the prediction time but it seems that using the Bayesian approach is much more time consuming and hence may be impractical industry-wise. As it was implied by the authors, their algorithm was unable to accurately predict call volumes at horizons greater than a week for five minute intervals.

	RMSE			APE		
N=64	Bayesian	Mixed 4	Mixed 5	Bayesian	Mixed 4	Mixed 5
Min	11.14	12.31	12.27	5.6	5.85	5.89
1^{st} Quartile	14.25	15.18	14.93	7.0	7.29	7.17
Median	15.83	17.07	17.11	7.4	7.59	7.71
Mean	18.28	19.10	19.05	8.4	8.76	8.74
3^{rd} Quartile	19.83	21.12	20.31	8.5	8.9	9.09
Max	43.42	41.01	41.30	28.6	28.02	27.84

Table 7.9: Comparing different models using RMSE and APE performance measures on the bank's data. The table presents the following approaches: the Bayesian and the mixed. Mixed 4 and Mixed 5 represent the results of the mixed model using four weeks of past data and five weeks of past data as the learning periods, respectively.

	Cover			Width		
N=64	Bayesian	Mixed 4	Mixed 5	Bayesian	Mixed 4	Mixed 5
Min	0.686	0.63	0.63	64.54	58.37	58.32
1^{st} Quartile	0.935	0.90	0.91	68.13	62.44	63.53
Median	0.97	0.96	0.96	69.23	66.68	66.39
Mean	0.947	0.93	0.93	70.10	67.80	67.93
3^{rd} Quartile	0.988	0.98	0.98	72.41	73.15	71.60
Max	1	0.99	0.99	79.3	82.60	92.19

Table 7.10: Comparing different models Coverage and Width performances on the bank's data. The table presents the following approaches: the Bayesian and the mixed. Mixed 4 and Mixed 5 represent the results of the mixed model using four weeks of past data and five weeks of past data as the learning periods, respectively.

Chapter 8

Conclusions and Future Research

In this thesis we have developed two variations of Poisson process models for describing count data of call center arrivals. These models utilize different techniques to tackle the modelling problem. One approach uses mixed models techniques while the other uses modern Bayesian techniques to analyze the data.

Our mixed model was customized to the specific requirements of an Israeli Cellular phone company. The company requires that the weekly forecast be available to the decision makers at least ten days in advance and should be based on six weeks of past data. Recent research, on the other hand, has focused on producing one-day-ahead forecasts or within-day learning algorithms. These issues are very important and may be very useful for call centers that can mobilize their agents on short term notice. As we show, our mixed model does contain the much needed practical flexibility to also support long lead times and short learning periods. It is also relatively easy to implement with standard software such as SAS®.

The mixed model incorporates fixed effects, such as day-of-week and its interaction with the daily periods; but it also models the day-to-day and the period-to-period correlations. We have detailed how to determine the significance of such effects. This process was illustrated on two different data sets: from an Israeli cellular phone company and from a US bank. The Israeli cellular phone company data allowed us to examine another feature of the mixed model. The data contained billing cycle dates which were used as exogenous variables after a preliminary examination was carried out to help reduce the size of the problem.

Unfortunately, we were unable to compare our mixed model to the current forecasting algorithm of the Israeli cellular phone company (since the company does not regularly

maintain its past predictions). An interesting future study would compare our mixed model to other industry-used models to see whether it may be useful in such surroundings. We have examined the mixed model results using several different measures such as the variability measure. Also its behavior under different lead times was examined. From a practical prospective, a manager might wish to consider a two stage prediction forecast: First, producing an early weekly forecast for the scheduling process and next re-producing another forecast one day before the week begins. This later prediction provides a much more reliable forecast. Using this one-day-ahead forecast, the manager of a call center might be able to incorporate immediate changes to the week schedule.

The model was also compared to the Gaussian Bayesian model detailed in [22]. The results were very similar and quite good considering that the mixed model used only a quarter of the learning data that the Gaussian Bayesian model used. The results are comparable even without considering other improvements that might be made to the mixed model (such as incorporating a different daily pattern for *each* day of the week). Based on these results and earlier analyses, we conclude that our mixed model is a very flexible, easy to implement and time efficient model. We would recommend it to the Israeli cellular phone company as an alternative to their current "black-box" algorithm. Also other exogenous variables, such as marketing effects, can be easily incorporated into the model using either the same basic procedure we have described for the billing cycles effects or a similar procedure.

An alternative Bayesian model was also proposed which models directly the Poisson arrival counts. This model was implemented using the 'OpenBugs' software. Unfortunately, the model was too complex and the data were too large to allow a thorough examination of this algorithm. For future research, this model needs to be implemented using other software such as C or C++, to further evaluate its capabilities. Nevertheless, the partial results are promising and the approach is worthy of further consideration.

Our original forecast problem was to predict the system load which involves the average service times as well as the arrival rates. We suggested and fitted a fairly easy quadratic regression model which incorporates weekdays and period effects.

Having both these predictions and the arrival counts, we constructed a QED regime performance measure. This measure helps determine how well our model would perform when planning a particular QED regime. Our results show that during busy periods, when the QED regime "square-root staffing" rule is relevant, the system will be able to perform at a desired level commensurate with the actual load. By comparison with benchmark

and naive models, the mixed model time series approach has improved the precision of prediction for the busier periods. This result gives some evidence that one can ensure a pre-specified QED regime using load forecasts that are sufficiently precise.

Appendix A

Analysis of Interval Resolution

An interesting debate might be held between practitioners and theoreticians as to what is the appropriate interval resolution to analyze. Theoreticians might say that in order to fully maintain the homogeneity assumption the intervals should be as small as possible. Alternatively, from a practitioner point of view, the resolution should be determined as a function of the possible shift starting times. If it is possible to change the number of available agents every 5 minutes then this should be the appropriate interval resolution. This may occur, for example, in large call centers where there are also agents who are occupied doing different off-line tasks and who can become immediately available. However, it is fairly common that call centers plan their daily schedule according to either half-hour or 15-minute resolutions.

As previously shown, our Gaussian mixed model can be easily modified to deal with different interval resolutions. An interesting question is by how much predictions based on lower-level resolutions are worse than 15 minute predictions as evaluated at the 15 minute period level. For example, if one predicted accurately the total arrivals over a half-hour period, but in that period the first 15 minutes had 0.5 times the average arrival rate, and the second 15 minutes had 1.5 times the average arrival rate, then using the half-hour prediction would lead one to seriously overstaff in the first 15 minutes and understaff in the second 15 minutes. This problem would not happen if one had good predictions at the 15 minute resolution. Hence, we are interested in analyzing the effect of the interval resolution on the forecast accuracy at the finest practical resolution.

In order to examine this last subject we used the Israeli Cellular data and predicted the arrival counts between 7AM and 11PM during the 203 regular weekdays between April 11 and December 24, 2004. The forecast procedure is the same as implemented by the

company; meaning that we used 6 weeks of past data as learning data to predict the week which begins ten days ahead.

Our baseline data resolution is 15-minute intervals. We compared 15-minute intervals with three additional interval resolutions: half-hour, one-hour and four-hours. In order to fairly assess the behavior of the different interval widths we compared their 15-minute predictions. The lower resolution forecasts were simply uniformly distributed between the 15-minutes intervals. For example, we took the predicted arrival count for a specific hour (on a certain day) and equally divided it into the four quarter hours. Tables A.1 and A.2 describe the results for both the RMSE and APE, respectively. The results show that the most precise predictions are obtained using the highest resolution. However, it is also noticeable that the differences between the half-hour resolution and the 15-minute one are quite small. The four-hour resolution results are quite bad in comparison with the other interval resolutions. These results can be used to justify the use of half-hour intervals in our study — only a minor practical improvement to the precision can be achieved by using a higher resolution.

	RMSE			
	15-minutes	half-hour	One-Hour	Four-Hour
Min	12.28	12.07	14.13	32.18
1^{st} Quartile	19.67	19.81	20.86	38.18
Median	22.48	22.59	23.56	41.37
Mean	24.97	25.06	25.87	42.80
3^{rd} Quartile	28.45	28.29	29.18	46.14
Max	60.00	60.01	60.21	73.12

Table A.1: Prediction accuracy as a function of interval resolution. We compare the RMSE result of the mixed model for four different resolutions: 15-minute, half-hour, one-hour and four-hour

	APE			
	15-minutes	half-hour	One-Hour	Four-Hour
Min	6.28	6.71	7.48	20.29
1^{st} Quartile	9.28	11.05	12.38	14.26
Median	11.05	11.43	12.64	28.42
Mean	12.38	12.70	14.01	30.97
3^{rd} Quartile	14.26	14.39	15.14	32.58
Max	53.19	60.32	80.88	217.00

Table A.2: Prediction accuracy comparison as a function of interval resolution. We compare the APE result of the mixed model with four different resolutions: 15-minute, half-hour, one-hour and four-hour

Appendix B

Computer Code for Two Models

This section presents the codes used to implement the mixed model and the Poisson Bayesian model.

B.1 Mixed Model – SAS Code

```
ods listing close ;
ods output CovParms = CovP ;
    proc mixed data= forpred method= ml ;
    class weekday kperiod date ;
    model y = weekday Sun*kperiod kperiod Fri*kperiod /noint ddfm= satterth outp= predict solution ;
    random date / type=SP(POW)(numDate) ;
    repeated kperiod / subject= numDate type= AR(1) ;
    run ;
ods output close ;
ods listing ;
```

B.2 Poisson Bayesian Model - OpenBugs Code

The source code for implementing the Beta-Gamma Bayesian model is presented below. Remarks were colored in green for readers convenience.

```
model {
```

```
The following loop defines the daily Gamma-Beta Process
        for (i in 2:D) {
        G[i] ← G[i-1]*B[i] + W[i]
            B[i] ∼ dbeta(arg1,arg2)
            W[i] ∼ dgamma(arg2,gam)
}
The following three lines define the Gamma-Beta Process priors
G[1] ∼ dgamma(gam,gam)
arg1 ← rho*gam
arg2 ← gam*(1-rho)
The following loop defines the daily counts distributions for the learning data
for (i in 1:(D-1) ) {
    V[i] ← G[i] * mu[QD[i]]
        for (j in 1:K) {
            lambda[i,j] ← V[i]*p[QD[i],j]
            NDK[i , j] ∼ dpois(lambda[i,j])
        }
}
The following loop defines the daily counts distributions for the predicted values
V[D] ← G[D] * mu[QD[D]]
for (j in 1:K) {
    lambda[D,j] ← V[D]*p[QD[D],j]
    NDD[j] ∼ dpois(lambda[D,j])
}
The following loop defines the vectors of proportion of daily volume on each of the weekdays
for( q in 1 : 7 ) {
    p[q,1:K] ∼ ddirch(alpha[q,])
}
The following loop defines the vectors of proportion of daily volume on each of the weekdays priors
for( q in 1 : 7 ) {
    mu[q] ∼ dnorm(mn[q], precis[q])
    precis[q] ← 100*pow(mn[q],-2)
    M[q] ∼ dnorm(5,0.2)
        for (j in 1:K) {
```

```
            alpha[q,j] ← exp(M[q])*pi[q,j]
        }
}
The following loop defines the vectors of proportion of daily volume on
each of the weekdays parameters priors
gam ∼ dgamma(0.01,0.01)
rho ← exp(-prerho)
prerho ∼ dgamma(2,5)
}
```

Bibliography

[1] Bruce H. Andrews and Shwan M. Cunningham. L. L. bean improves call-center forecasting. *INTERFACES*, 25(6):1–13, 1995.

[2] A. Antipov and N. Meade. Forecasting call frequency at a financial services call centre. *Journal of Operational Research Society*, 53(9):953–960, 2002.

[3] Athanassios N. Avramidis, Alexandre Deslauriers, and Pierre L'Ecuyer. Modeling daily arrivals to a telephone call center. *Management Science*, 50(7):896–908, July 2004.

[4] Lawerence D. Brown, R. Zhang, and Linda Zhao. Root un-root methodology for non parameteric density estimation. Technical report, The Wharton School, University of Pennsylvenia, 2001.

[5] Lawrence D. Brown, Noah Gans, Avishai Mandelbaum, Anat Sakov, Haipeng Shen, Sergey Zeltyn, and Linda Zhao. Statistical analysis of a telephone call center: A queueing-science prespective. *JASA*, 100(469):36–55, March 2005.

[6] Chris Chatfield. Prediction interval. Department of Mathematical Sciences, University of Bath, 1998.

[7] Eugene Demidenko. *Mixed Models: Theory and Applications*. Wiley-Interscience, 2004.

[8] Noah Gans, Ger Koole, and Avishai Mandelbaum. Telephone call centers: Tutorial, review, and research prospects. *Manufacturing and Service Operations Management*, 5:79–141, 2003.

[9] Geurt Jongbloed and Ger Koole. Managing uncertainty in call centers using Poisson mixtures. *Applied Stochastic Models in Bussinesss and Industry*, 17:307–318, 2001.

[10] Leonard Kaufman and Peter J. Rousseeuw. Finding groups in data : An introduction to cluster analysis. 1990.

[11] M. Maechler, P. Rousseeuw, A. Struyf, and M. Hubert. Cluster analysis basics and extensions. Rousseeuw et al provided the S original which has been ported to R by Kurt Hornik and has since been enhanced by Martin Maechler: speed improvements, silhouette functionality, bug fixes, etc. See the 'Changelog' file (in the package source), 2005.

[12] Avishai Mandelbuam. Call centers (centres) research bibiliography with abstracts. Chapter two details Statistics and Forecasting related papers. The serveng website, 2004.

[13] Peter McCullagh and John A. Nedler. *Generalized Linear Models*. Chapman and Hall, second edition, 1989.

[14] Sem Borst Avishai Mandelbaum Martin Reiman. Dimensioning large call centers. *Operations Research*, 52.

[15] Ishiguro M. Sakamoto Y. and Kitagawa G. *Akaike Information Criterion Statistics*. Springer, 1999.

[16] Refik Soyer and M. Murat Tarimcilar. Modeling and analysis of call center arrival data: A bayesian approach. Preprint, Department of Management Science, The George Washington University, Washington, DC, 2005.

[17] David Spiegelhalter, Andrew Thomas, Nicky Best, and Dave Lunn. *WinBUGS User Manual*, April 2005. WinBugs website.

[18] James W. Taylor. A comparison of univariate time series methods for forecasting intraday arrivals at a call center. 2006. under review for a Focused Issue of Managment Science.

[19] R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2005. ISBN 3-900051-07-0. R website.

[20] Andrew Thomas. *BRugs User Manual (the R interface to BUGS)*. Dept of Mathematics and Statistics, University of Helsinki, Finland, 2004. Version 1.0.

[21] Valery Trofimov, Paul Feigin, Avishai Mandelbaum, and Eva Ishay. DataMOCCA – Data MOdel for Call Center Analysis. Technical report, Technion, Israel, 2003. DataMOCCA.

[22] Jonathan Weinberg, Lawrence D. Brown, and Jonathan R. Stroud. Bayesian forecasting of an inhomogeneous Poisson process with applications to call center data. May 2005.

[23] Weidong Xu. Long range planning for call centers at Fedex. *The Journal of Bussiness Forecasting Methods and Systems*, 18(4):7–11, 2000.