Forecasting Arrivals to a Telephone Call Center

Haipeng Shen • Jianhua Z. Huang

Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

Department of Statistics, Texas A&M University, College Station, TX 77843, USA

haipeng@email.unc.edu • jianhua@stat.tamu.edu

Accurate forecasting of call arrivals is critical for staffing and scheduling of a telephone call center. We develop methods for day-to-day and dynamic within-day forecasting of incoming call volumes. Our approach is to treat the intra-day call volume profiles as a high-dimensional vector time series. We propose to first reduce the dimensionality by singular value decomposition of the matrix of historical intra-day profiles and then apply time series and regression techniques. Both day-to-day dynamics and within-day patterns of call arrivals are taken into account by our approach. The proposed methods are applied to a real data set and show substantial improvements over existing industry standards in an out-of-sample forecast comparison. Our methods are computationally fast and therefore it is feasible to use them for real-time dynamic forecasting.

1. Introduction

Call centers have become a primary contact point between service providers and their customers in modern business. For example, it is estimated that in 2002 more than 70% of all customer-business interactions are handled in call centers, and the call center industry in US employs more than 3.5 million people, or 2.6% of its workforce (Brown et al. 2005). With their growing presence and importance in various organizations, managing call center operations more efficiently is of significant economic interest.

For most call centers, 60-70% of operating expenses are capacity costs, and in particular human resource costs (Gans et al. 2003). Effective management of a call center requires call center managers to match up call center resources with workload. To forecast workload accurately, the first and critical step is to provide an accurate forecast of future call volumes. Usually two kinds of forecasts are needed by a call center manager for staffing and scheduling purpose: 1. Forecast the call volumes several days or weeks ahead; 2. On a particular day, dynamically update the forecast using newly available information as additional calls arrive throughout the day. The dynamic updating is useful because it adds flexibility to a call center manager's allocation of available resources, which then leads to higher efficiency and productivity. By reevaluating her forecast for the remainder of the day, the manager can then schedule meetings or training sessions for agents free from work at short-notice, or call in back-up agents.

Quantitative approaches to call arrival forecasting make use of historical call arrival data, which record the times at which calls arrive to the center. Typically call arrival data are aggregated into total numbers of arrivals during short time periods such as 15-minute or 30-minute intervals and subsequently the target of forecasting is future call volumes over such periods. We refer to the collection of such aggregated call volumes for a given day as the *intra-day call volume profile* of that day (Section 3.1). See Jongbloed and Koole (2001) and Brown et al. (2005) for examples of this data aggregation scheme. There exist some regularities within call-arrival data, such as inter-day dependence and intra-day dependence. By inter-day dependence, we mean correlation between daily volumes and possible weekly, monthly, seasonal and even yearly cycles in call volumes. Intra-day dependence refers to correlation between arrivals during different time periods (morning/afternoon/night) within the same day. Such regularities and dependences serve as the basis for developing good forecasts.

Early work on forecasting call arrivals usually ignores intra-day dependence, partly due to lack of relevant data. Andrews and Cunningham (1995) forecast daily volumes at a retailer's call center using an ARIMA model, which incorporates advertising effects via transform functions. Bianchi et al. (1993, 1998) forecast daily arrivals at a telemarketing call center using ARIMA modelling as well, and they use intervention analysis to control for outliers. Recently, with the availability of call-by-call data and driven by the increasing need for more effective management of telephone call centers, stochastic models of call arrivals that take into account the intra-day dependence have been introduced. For example, Avramidis et al. (2004) consider a doubly stochastic Poisson model and Brown et al. (2005) propose a multiplicative two-way mixed-effects model. Bayesian methods have also been applied to call volume forecasting. Soyer and Tarimcilar (2004) analyze a modulated Poisson model using a Bayesian approach, with a focus on assessing the effectiveness of advertising campaigns. Weinberg et al. (2005) extend the work of Avramidis et al. (2004) and Brown et al. (2005) to model both inter-day and intra-day dependences, and propose a Bayesian MCMC estimation algorithm. They consider day-of-the-week effect as well as intra-day forecast updating.

Despite recent progress in analyzing call center data, Gans et al. (2003) urge more serious research effort for developing time series forecasting methods to improve arrival forecasting. In this paper, we introduce a system of methods that can generate both day-to-day and *especially* dynamic within-day forecasts. The current study extends Shen and Huang (2005), where singular value decomposition (SVD) is proposed as a tool for preliminary data analysis of call arrival data before serious modelling, to detect anomalous days with unusual arrival patterns, and to extract significant inter-day features such as day-of-the-week effect and working hour switching, as well as intra-day features like average daily arrival patterns. Shen and Huang (2005) point out the potential of using the reduced data from SVD to develop call volume forecasting models. Such a potential is substantially developed in the current paper to offer methods for generating one- or multi-day-ahead forecasts and within-day dynamic updates of call volume profiles.

We consider the intra-day call volume profiles as a vector time series across days. Each intra-day profile is treated as one observation of the vector time series. Thus, the dimensionality of the time series is high. We propose to first reduce the dimensionality by applying SVD to the matrix of the intra-day profiles. As a result, the high dimensional intra-day profile series is summarized by a few pairs of inter-day and intra-day features, (see (1) in Section 3.1). We then keep the intra-day features fixed, and apply univariate time series forecasting techniques to forecast the inter-day feature series separately. The resulting forecasts are subsequently combined with the intra-day features to produce day-to-day forecasts of future intra-day profiles. Our case study in Section 5 suggests that only two or three pairs of inter-day and intra-day features are sufficient to achieve very good forecasting performance. As it turns out, the first intra-day feature summarizes the average call arrival pattern during a regular weekday while the first inter-day feature is highly correlated with total daily call volume. As a second-order effect, the second intra-day feature captures the following specific phenomenon: Fridays usually have above average call volumes in early mornings and below average volumes in late afternoons and evenings, while the opposite is true for Mondays.

Our procedure for within-day dynamic updating is based on a technique known as *penalized least squares*. For a particular day, suppose we have obtained some day-to-day time series forecasts for the inter-day features. As calls arrive when the day progresses, one direct approach to update the time series forecasts is via a least squares regression using the newly available call volume profile as the response and the corresponding part of the intra-day feature vectors as the independent variables, (see (5) in Section 3.3.1). The corresponding regression coefficients then give the updated inter-day feature forecasts. However, this approach relies solely on the new arrival information of the present day, and makes no use of the time series forecasts. Our penalized least squares method adjusts the direct approach by pushing the least squares updates towards the time series forecasts and finding a balance point between them, (see (7) in Section 3.3.2). This method effectively combines the newly available information with the information up to the end of the previous day. We show in Section 5 that the proposed updating method can reduce the forecasting error dramatically.

This paper is structured as follows. Section 2 introduces the call center arrival data that are used to illustrate our methodology. Our forecasting approach is described in detail in Section 3. Section 3.1 introduces dimensionality reduction through SVD, and followed by Section 3.2 on the method for one- or multi-day-ahead intra-day profile forecasting, and Section 3.3 on dynamic within-day updating. Section 4 describes a few performance measures for comparing forecasts from different forecasting methods. A comparative study of several competing methods based on their out-of-sample forecasting performance is presented in Section 5. We conclude in Section 6.

2. The Data

The data that will be used to illustrate our forecasting methods were gathered at an inbound call center of a major northeastern US financial firm between January 1 and October 26, 2003. The original database has detailed information about every call that got connected to this call center during this period (except three days when the data collecting equipment went out of order). The center opens normally between 7:00AM and midnight. For the current study, we are interested in understanding the arrival pattern of calls to the service queue and generating out-of-sample forecasts of future call volumes. As a result, the portion of the data of interest to us involves the information about the time every call arrives to the service queue. The call arrival pattern is very different between weekdays and weekends. For this reason, we restrict our attention to weekdays. In particular, we focus on the 42 whole weeks between January 6 and October 24.

For a particular day, we divide the 17-hour operating period into sixty-eight 15-minute intervals, and record the number of calls arriving to the service queue during each interval. The aggregated data form a 210×68 count matrix, with each row corresponding to a particular day in the 42 weeks considered, and each column corresponding to one specific

15-minute interval between 7:00AM and midnight.

We found by plotting the data that call volumes almost always occur in predictable, repeating patterns. For example, it is typical that there is a peak around 11:00AM followed by a second, lower peak around 2:00PM. As an illustration, Figure 1 plots the intra-day arrival patterns for the five days in the final week of the data set. Such regularity of arrival patterns has important implications for our forecasting task.

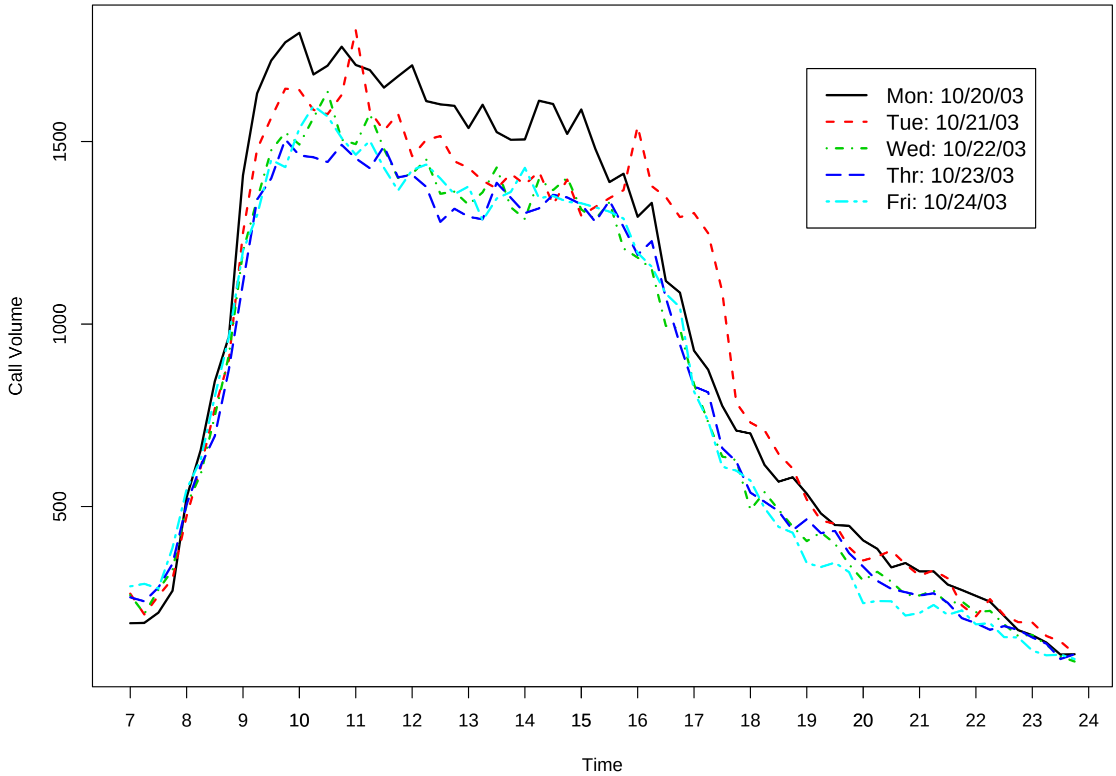

Figure 1: Intra-day arrival patterns during the final week of our data. Both predictive regularity and abnormality are present in call center arrival data.

Figure 1 also exhibits some abnormal arrival pattern for the Tuesday, during which there are too many calls arriving after 3:30PM. There are some other abnormal days in the data such as holidays (for example, May 26, 2003, Memorial Day), a missing day (April 4, 2003, no data), and days with other unusual arrival patterns (like the Tuesday shown in Figure 1). The call arrival patterns for these days are so different from the rest of the days that forecasting methods are generally unable to produce reasonable forecasts. Inclusion of these days in our study is likely to be unhelpful in the assessment of different methods. Prior to applying the forecasting methods, we cleaned the data as follows. We applied the technique in Shen and Huang (2005) to identify the anomalous days. The call arrival profiles of these days were then replaced by the average of the corresponding periods (i.e. same weekday and time-of-day) in the two adjacent weeks.

Our data appear to possess heteroscedasticity (i.e. nonconstant variance), with the variance being approximately proportional to the call volume. In fact, for our data set, it is a good approximation to assume that the call volume over a short time interval follows a Poisson distribution. To stablize the variance, we employed the root-unroot method (Brown et al., 2005) as follows. Let N denote the call volume for a certain time interval. Set $X = \sqrt{N + 1/4}$. Denote the forecast of X by $\widehat{X}$. Then the forecast of N is given by $\widehat{N} = \widehat{X}^2 - 1/4$. The rational of this method is that, if $N \sim \text{Poisson}(\lambda)$, then X has approximately a mean of $\sqrt{\lambda}$ and a variance of $1/4$.

3. Forecasting Methods

3.1 Dimension Reduction via Singular Value Decomposition

Let $\mathbf{X} = (x_{ij})$ be an $n \times m$ matrix that records the call volumes (or a transformation of the volumes as described in Section 2) for n days with each day having m time periods. The rows and columns of $\mathbf{X}$ correspond respectively to days and time periods within a day. The ith row of $\mathbf{X}$, denoted as $\mathbf{x}_i^T = (x_{i1}, \ldots, x_{im})$, is referred to as the *intra-day call volume profile* of the ith day. The intra-day profiles, $\mathbf{x}_1, \mathbf{x}_2, \ldots$, form a vector-valued time series taking values in $\mathbb{R}^m$. We want to build a time series model for this series and use it for forecasting. However, commonly used multivariate time series models such as VAR (Vector AutoRegressive models) and more general VARMA (Vector AutoRegressive and Moving Average models) (Reinsel 2003) are not directly applicable due to the large dimensionality of the time series we consider. For example, the dimensionality m is 68 for our data.

Our approach starts from dimension reduction. We first seek a few basis vectors, denoted as $\mathbf{f}_k$, $k = 1, \ldots, K$, such that all elements in the time series $\{\mathbf{x}_i\}$ can be represented (or approximated well) by these basis vectors. The number of the basis vectors K should be much smaller than the dimensionality m of the time series. Specifically, we consider the following decomposition,

$$\mathbf{x}_i = \beta_{i1}\mathbf{f}_1 + \cdots + \beta_{iK}\mathbf{f}_K + \boldsymbol{\epsilon}_i, \qquad i = 1, \ldots, n, \tag{1}$$

where $\mathbf{f}_1, \ldots, \mathbf{f}_K \in \mathbb{R}^m$ are the basis vectors and $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n \in \mathbb{R}^m$ are the error terms. We expect that the main features of $\mathbf{x}_i$ can be summarized by a linear combination of the basis vectors so that the error terms in (1) would be small in magnitude. This can be achieved by solving the following minimization problem,

$$\min_{\substack{\beta_{i1},\ldots,\beta_{iK}\\ \mathbf{f}_1,\ldots,\mathbf{f}_K}} \sum_{i=1}^{n} \|\boldsymbol{\epsilon}_i\|^2 = \min_{\substack{\beta_{i1},\ldots,\beta_{iK}\\ \mathbf{f}_1,\ldots,\mathbf{f}_K}} \sum_{i=1}^{n} \left\| \mathbf{x}_i - (\beta_{i1}\mathbf{f}_1 + \cdots + \beta_{iK}\mathbf{f}_K) \right\|^2. \tag{2}$$

For identifiability, we require in (2) that $\mathbf{f}_i^T\mathbf{f}_j = \delta_{ij}$, where δ_{ij} is the Kronecker delta which equals 1 for $i = j$ and 0 otherwise. The solution to this problem is actually given by the singular value decomposition (SVD) of the matrix $\mathbf{X}$ as shown below.

The SVD of the matrix $\mathbf{X}$ can be expressed as

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T, \tag{3}$$

where $\mathbf{U}$ is an $n \times m$ matrix with orthonormal columns, $\mathbf{S}$ is an $m \times m$ diagonal matrix, and $\mathbf{V}$ is an $m \times m$ orthogonal matrix. The columns of $\mathbf{U}$, $\{\mathbf{u}_k = (u_{1k}, \ldots, u_{nk})^T\}$, namely the left singular vectors, satisfy $\mathbf{u}_i^T\mathbf{u}_j = \delta_{ij}$. The columns of $\mathbf{V}$, $\{\mathbf{v}_k = (v_{1k}, \ldots, v_{mk})^T\}$, or the right singular vectors, satisfy $\mathbf{v}_i^T\mathbf{v}_j = \delta_{ij}$. The diagonal elements of $\mathbf{S}$ are the singular values, which are usually ordered decreasingly. Let $\mathbf{S} = \mathrm{diag}(s_1, \ldots, s_m)$ and $r = \mathrm{rank}(\mathbf{X})$. Then $s_1 \geq s_2 \geq \cdots \geq s_r > 0$, and $s_k = 0$ for $r + 1 \leq k \leq m$. It follows from (3) that

$$\mathbf{x}_i = s_1 u_{i1}\mathbf{v}_1 + \cdots + s_r u_{ir}\mathbf{v}_r.$$

Keeping only the terms associated with the largest K singular values, we have the following approximation,

$$\mathbf{x}_i \simeq (s_1 u_{i1})\mathbf{v}_1 + \cdots + (s_K u_{iK})\mathbf{v}_K.$$

This K-term approximation has the following optimal property that it is exactly the solution for the minimization problem (2). More precisely, $\beta_{ik} = s_k u_{ik}$ and $\mathbf{f}_k = \mathbf{v}_k$, $i = 1, \ldots, n$, $k = 1, \ldots, K$, solve (2), and the solution is unique up to a sign change to $\mathbf{f}_k$ (Harville 1997). Thus, we have formally obtained the decomposion (1) using the SVD of $\mathbf{X}$.

We call $\mathbf{f}_1, \ldots, \mathbf{f}_K$ the *intra-day feature vectors* and $\{\beta_{i1}\}, \ldots, \{\beta_{iK}\}$ the *inter-day feature series*. From the decomposition (1), the high-dimensional intra-day profiles $\mathbf{x}_i$ are concisely summarized by a small number of inter-day feature series, using the same number of intra-day feature vectors as the bases. By using a small K, we achieve a big dimension reduction — from m to K. In our real data example, $K = 2$ or 3 already gives good forecasting results while m is 68. One attractive feature of the decomposition (1) is that, it effectively separates out the intra-day and inter-day variations, both of which are present in the intra-day profile time series.

3.2 Day-to-day Forecasting

Consider forecasting the intra-day call volume profile $\mathbf{x}_{n+h}$ ($h > 0$) using the historical data $\mathbf{x}_1, \ldots, \mathbf{x}_n$. By making use of the model (1), forecasting a m-dimensional time series $\{\mathbf{x}_i\}$ reduces to forecasting K one-dimensional inter-day feature series $\{\beta_{i1}\}, \ldots, \{\beta_{iK}\}$. As the basis of our forecasting method, we assume that the decomposition (1) is valid for future intra-day profiles, that is, it holds for $i = n+1, \ldots, n+h$. If we can obtain a forecast of the coefficients $\beta_{n+h,1}, \ldots, \beta_{n+h,K}$, then a forecast of $\mathbf{x}_{n+h}$ is given by

$$\hat{\mathbf{x}}_{n+h} = \hat{\beta}_{n+h,1}\mathbf{f}_1 + \cdots + \hat{\beta}_{n+h,K}\mathbf{f}_K,$$

where $\hat{\beta}_{n+h,k}$ is a forecast of $\beta_{n+h,k}$, $k = 1, \ldots, K$. Denote $\mathbf{f}_k = (f_{1k}, \ldots, f_{mk})^T$. The forecast of the call volume of the j-th time period of day $n+h$ is given by

$$\hat{x}_{n+h,j} = \hat{\beta}_{n+h,1} f_{j1} + \cdots + \hat{\beta}_{n+h,K} f_{jK}, \qquad j = 1, \ldots, m.$$

Because of the way how the SVD is constructed, $(\beta_{1k}, \ldots, \beta_{nk})$ is orthogonal to $(\beta_{1l}, \ldots, \beta_{nl})$ for $k \neq l$. This lack of contemporaneous correlation suggests that the cross-correlations at non-zero lags are likely to be small. Therefore it is reasonable to believe that it is adequate to forecast each series $\{\beta_{ik}\}$ separately using univariate time series methods, for $k = 1, \ldots, K$. To forecast each series, we propose to use univariate time series models such as exponential smoothing, ARIMA models (Box et al. 1994), state space models (Harvey 1990) and other suitable models. Which model to use would depend on the actual situation, and can be decided on by analyzing historical data.

In our case study of the call center data described in Section 2, we apply a varying coefficient AR(1) model that takes into account the day-of-the-week effect. Specifically, we model the series $\{\beta_{ik}\}$ by

$$\beta_{ik} = a_k(d_{i-1}) + b_k \beta_{i-1,k} + \epsilon_i,$$

where d_{i-1} denotes the day-of-the-week of day $i-1$. The same model has been obtained in Shen and Huang (2005) when analyzing the arrival data from the same call center for a different year. This varying intercept AR(1) model works well in our out-of-sample forecasting exercise in Section 5.

3.3 Dynamic Within-day Updating

Using the method in the previous section, a call center manager is able to forecast the call volume profile for day $n+h$ at the end of day n. In particular, $h = 1$ corresponds to the next day. As calls arrive during day $n+1$, the manager may want to dynamically update her forecast for the remainder of the day using the information from the early part of that day. There are empirical justifications for such updating, as both Avramidis et al. (2004) and Steckley et al. (2004) report evidence of positive correlation among time periods within a given day. This updating is operationally beneficial and feasible as well. If done appropriately, one can reduce the forecast error for the rest of the day. In turn, operational benefits would follow from the ability to change agent schedules according to the updated forecast. For example, to adjust for a revised forecast, call center managers may send people home early or have them take training sessions or perform alternative work, or they may arrange agents to work overtime or call in part-time or work-from-home agents.

Suppose we have available the call volumes for the early part, say, m_0 time periods of day $n+1$. Denote them collectively as $\mathbf{x}^e_{n+1} = (x_{n+1,1}, \dots, x_{n+1,m_0})^T$, a vector contains the first m_0 elements of $\mathbf{x}_{n+1}$. Denote $\mathbf{x}^l_{n+1} = (x_{n+1,m_0+1}, \dots, x_{n+1,m})^T$ to be the call volumes for the later part of day $n+1$. Here, for notational simplicity, we suppress the dependence of $\mathbf{x}^e_{n+1}$ and $\mathbf{x}^l_{n+1}$ on m_0. Let $\hat{\beta}^{\mathrm{TS}}_{n+1,k}$ be a time series forecast based on $\{\beta_{i,k}\}$, for $k = 1, \dots, K$, using information up to the end of day n. The time series forecast of $\mathbf{x}^l_{n+1}$ is then given by

$$\hat{x}^{\mathrm{TS}}_{n+1,j} = \hat{\beta}^{\mathrm{TS}}_{n+1,1} f_{j1} + \cdots + \hat{\beta}^{\mathrm{TS}}_{n+1,K} f_{jK}, \quad j = m_0+1, \dots, m. \tag{4}$$

These forecasts do not utilize any new information of day $n+1$. Below, we discuss two ways to incorporate the new information in $\mathbf{x}^e_{n+1}$ to get an updated forecast of $\mathbf{x}^l_{n+1}$.

3.3.1 Direct Least Squares Updating

When being applied to the intra-day profile of day $n+1$, the decomposition (1) can be written as

$$x_{n+1,j} = \beta_{n+1,1} f_{j1} + \cdots + \beta_{n+1,K} f_{jK} + \epsilon_{n+1,j}, \qquad j = 1, \dots, m. \tag{5}$$

Let $\mathbf{F}^e$ be a $m_0 \times K$ matrix whose (j,k)-th entry is f_{jk}, $1 \le j \le m_0$, $1 \le k \le K$, $\boldsymbol{\beta}_{n+1} = (\beta_{n+1,1}, \dots, \beta_{n+1,K})^T$, and $\boldsymbol{\epsilon}^e_{n+1} = (\epsilon_{n+1,1}, \dots, \epsilon_{n+1,m_0})^T$. Then, with the availability of $\mathbf{x}^e_{n+1} = (x_{n+1,1}, \dots, x_{n+1,m_0})^T$, we have the following linear regression model,

$$\mathbf{x}^e_{n+1} = \mathbf{F}^e \boldsymbol{\beta}_{n+1} + \boldsymbol{\epsilon}^e_{n+1}.$$

This suggests that we can forecast $\boldsymbol{\beta}_{n+1}$ by the method of least squares. Solving

$$\min_{\boldsymbol{\beta}_{n+1}} \|\mathbf{x}^e_{n+1} - \mathbf{F}^e \boldsymbol{\beta}_{n+1}\|^2,$$

we obtain that

$$\hat{\boldsymbol{\beta}}_{n+1}^{\text{LS}} = (\mathbf{F}^{eT}\mathbf{F}^{e})^{-1}\mathbf{F}^{eT}\mathbf{x}_{n+1}^{e}.$$

To make the LS forecast of $\boldsymbol{\beta}_{n+1}$ uniquely defined, we require that $m_0 \geq K$. This makes sense intuitively in that one needs more observations than parameters, in order to estimate the parameters well. The forecast of $\mathbf{x}_{n+1}^{l}$ is then given by

$$\hat{x}_{n+1,j}^{\text{LS}} = \hat{\beta}_{n+1,1}^{\text{LS}} f_{j1} + \cdots + \hat{\beta}_{n+1,K}^{\text{LS}} f_{jK}, \quad j = m_0 + 1, \ldots, m. \tag{6}$$

As an example, if $K = 1$, the LS forecast of $\boldsymbol{\beta}_{n+1}$ is simply

$$\hat{\beta}_{n+1,1}^{\text{LS}} = \frac{\sum_{j=1}^{m_0} f_{j1} x_{n+1,1}}{\sum_{j=1}^{m_0} f_{j1}^2},$$

and the corresponding forecast of $\mathbf{x}_{n+1}^{l}$ is

$$\hat{x}_{n+1,j}^{\text{LS}} = \frac{\sum_{j=1}^{m_0} f_{j1} x_{n+1,1}}{\sum_{j=1}^{m_0} f_{j1}^2} f_{j1}, \qquad j = m_0 + 1, \ldots, m.$$

3.3.2 Penalized Least Squares Updating

The direct least squares updating makes use of the additional information available at the early part of day $n+1$, but it needs a sufficient amount of data (i.e. a large enough m_0) in order for $\hat{\boldsymbol{\beta}}_{n+1}^{\text{LS}}$ to be reliable. This might create a problem if the manager wants to update her forecast early in the morning, for example, at 8:00AM with $m_0 = 4$ or 10:00AM with $m_0 = 12$ for the data in Section 2. Another disadvantage of the direct least squares updating is that it does not make full use of the historical information other than the estimated intra-day feature vectors. In particular, it totally disregards the day-to-day dependence among the inter-day profiles.

We propose to combine the least squares forecast with the time series forecast of $\boldsymbol{\beta}_{n+1}$ by using the idea of penalization. Specifically, we minimize with respect to $\beta_{n+1,1}, \ldots, \beta_{n+1,K}$ the following *penalized least squares* criterion,

$$\sum_{j=1}^{m_0} \left| x_{n+1,j} - (\beta_{n+1,1} f_{j1} + \cdots + \beta_{n+1,K} f_{jK}) \right|^2 + \lambda \sum_{k=1}^{K} |\beta_{n+1,k} - \hat{\beta}_{n+1,k}^{\text{TS}}|^2, \tag{7}$$

where $\hat{\beta}_{n+1,k}^{\text{TS}}$ is a time series forecast based on the information up to the end of day n, and $\lambda > 0$ is a penalty parameter. The penalized least squares criterion involves two terms: the first term measures how well the model prediction matches the observed call volumes in the early part of the day, while the second term penalizes a large departure from the time series forecast. The $\boldsymbol{\beta}_{n+1}$ obtained as the solution to the minimization problem (7) will be a compromise between the two terms based on the size of λ, the penalty parameter. In practice, λ can be selected using a rolling hold-out sample; see Section 5.3 for a detailed description of one selection procedure through an example.

For each given λ, the penalized least squares problem (7) can be rewritten in a constrained optimization form:

$$\text{minimize } \sum_{k=1}^{K} |\beta_{n+1,k} - \hat{\beta}^{\text{TS}}_{n+1,k}|^2 \text{ subject to}$$
$$\sum_{j=1}^{m_0} \Big| x_{n+1,j} - (\beta_{n+1,1} f_{j1} + \cdots + \beta_{n+1,K} f_{jK}) \Big|^2 \le \xi,$$

where $\xi = \xi(\lambda)$ is chosen appropriately so that the solution to this problem and the previous one are the same, at the given value of λ. The problem can be interpreted as follows, among all the $\boldsymbol{\beta}_{n+1}$'s that yield forecasts which adequately match the observed call volumes at the early part of the day, find the one that is closest to the time series forecast.

The penalized least squares criterion (7) can be expressed in the following matrix form,

$$(\mathbf{x}^e_{n+1} - \mathbf{F}^e \boldsymbol{\beta}_{n+1})^T (\mathbf{x}^e_{n+1} - \mathbf{F}^e \boldsymbol{\beta}_{n+1}) + \lambda(\boldsymbol{\beta}_{n+1} - \hat{\boldsymbol{\beta}}^{\text{TS}}_{n+1})(\boldsymbol{\beta}_{n+1} - \hat{\boldsymbol{\beta}}^{\text{TS}}_{n+1}).$$

Minimizing this criterion gives us the penalized least squares forecast of $\boldsymbol{\beta}_{n+1}$,

$$\hat{\boldsymbol{\beta}}^{\text{PLS}}_{n+1} = (\mathbf{F}^{eT} \mathbf{F}^e + \lambda \mathbf{I})^{-1} (\mathbf{F}^{eT} \mathbf{x}^e_{n+1} + \lambda \hat{\boldsymbol{\beta}}^{\text{TS}}_{n+1}). \tag{8}$$

It is a weighted average of the least squares forecast and the time series forecast with weighs being controlled by the penalty parameter λ. This can be easily seen as follows,

$$\begin{aligned}\hat{\boldsymbol{\beta}}^{\text{PLS}}_{n+1} =& (\mathbf{F}^{eT} \mathbf{F}^e + \lambda \mathbf{I})^{-1} \mathbf{F}^{eT} \mathbf{F}^e (\mathbf{F}^{eT} \mathbf{F}^e)^{-1} \mathbf{F}^{eT} \mathbf{x}^e_{n+1} + \lambda (\mathbf{F}^{eT} \mathbf{F}^e + \lambda \mathbf{I})^{-1} \hat{\boldsymbol{\beta}}^{\text{TS}}_{n+1} \\ =& (\mathbf{F}^{eT} \mathbf{F}^e + \lambda \mathbf{I})^{-1} \mathbf{F}^{eT} \mathbf{F}^e \hat{\boldsymbol{\beta}}^{\text{LS}}_{n+1} + \lambda (\mathbf{F}^{eT} \mathbf{F}^e + \lambda \mathbf{I})^{-1} \hat{\boldsymbol{\beta}}^{\text{TS}}_{n+1}.\end{aligned}$$

Note that, if $\lambda = 0$, $\hat{\boldsymbol{\beta}}^{\text{PLS}}_{n+1}$ is simply $\hat{\boldsymbol{\beta}}^{\text{LS}}_{n+1}$; if $\lambda \to \infty$, $\hat{\boldsymbol{\beta}}^{\text{PLS}}_{n+1}$ then reduces to $\hat{\boldsymbol{\beta}}^{\text{TS}}_{n+1}$. The forecast of $\mathbf{x}^l_{n+1} = (x_{n+1,m_0+1}, \ldots, x_{n+1,m})^T$ based on $\hat{\boldsymbol{\beta}}^{\text{PLS}}_{n+1}$ is given by,

$$\hat{x}^{\text{PLS}}_{n+1,j} = \hat{\beta}^{\text{PLS}}_{n+1,1} f_{j1} + \cdots + \hat{\beta}^{\text{PLS}}_{n+1,K} f_{jK}, \qquad j = m_0 + 1, \ldots, m. \tag{9}$$

We end this section by pointing out a connection between our penalized least squares approach and the widely used ridge regression (Draper and Smith 1998) in regression analysis.

Define $\widetilde{\mathbf{x}}^e_{n+1} = \mathbf{x}^e_{n+1} - \mathbf{F}^e\hat{\boldsymbol{\beta}}^{\mathrm{TS}}_{n+1}$ and $\widetilde{\boldsymbol{\beta}}_{n+1} = \boldsymbol{\beta}_{n+1} - \hat{\boldsymbol{\beta}}^{\mathrm{TS}}_{n+1}$. Then the penalized least squares criterion becomes

$$(\widetilde{\mathbf{x}}^e_{n+1} - \mathbf{F}^e\widetilde{\boldsymbol{\beta}}_{n+1})^T(\widetilde{\mathbf{x}}^e_{n+1} - \mathbf{F}^e\widetilde{\boldsymbol{\beta}}_{n+1}) + \lambda\widetilde{\boldsymbol{\beta}}^T_{n+1}\widetilde{\boldsymbol{\beta}}_{n+1},$$

which is the same minimizing criterion for a ridge regression with $\widetilde{\mathbf{x}}^e_{n+1}$ as the response, $\mathbf{F}^e$ as the design matrix, and $\widetilde{\boldsymbol{\beta}}_{n+1}$ as the regression coefficient vector. The ridge estimator is given by

$$\widetilde{\boldsymbol{\beta}}_{n+1} = (\mathbf{F}^{eT}\mathbf{F}^e + \lambda\mathbf{I})^{-1}\mathbf{F}^{eT}\widetilde{\mathbf{x}}^e_{n+1},$$

which is a rewriting of the expression (8), since

$$(\mathbf{F}^{eT}\mathbf{F}^e + \lambda\mathbf{I})^{-1}\mathbf{F}^{eT}\mathbf{F}^e\hat{\boldsymbol{\beta}}^{\mathrm{TS}}_{n+1} = \hat{\boldsymbol{\beta}}^{\mathrm{TS}}_{n+1} - \lambda(\mathbf{F}^{eT}\mathbf{F}^e + \lambda\mathbf{I})^{-1}\hat{\boldsymbol{\beta}}^{\mathrm{TS}}_{n+1}.$$

4. Performance Measures

We shall conduct an out-of-sample rolling forecast exercise to evaluate the proposed forecasting methods, using the data described in Section 2. In this study, we only evaluate one-day-ahead forecasting and dynamic within-day updating. There are in total 210 intra-day call volume profiles in the data set. We use the final 60 days of the data, from August 4 to October 24, as the forecasting set. For each day in the forecasting set, we use its 150 preceding days as the historical data to generate a forecast for that day.

Two measures are used for assessing and comparing forecast performances. Let N_{ij} denote the call volume on day i during period j. Suppose $\widehat{N}_{ij}$ is a forecast of N_{ij}, then the absolute error (AE) and relative error (RE) are defined respectively as follows,

$$\mathrm{AE}_{ij} = |\widehat{N}_{ij} - N_{ij}| \quad \text{and} \quad \mathrm{RE}_{ij} = 100|\widehat{N}_{ij} - N_{ij}|/N_{ij}.$$

These two measures are defined for each day and each time period. For each given time period within a day, summary statistics (over days in the forecasting set) such as mean, median, quartiles of AE and RE can be reported to show the performance of different forecasting methods.

We now define several summary measures to gauge the overall performance over many time periods. Suppose we are interested in forecasting the whole intra-day profile for day i, as for example, in one-day-ahead forecasting. We define the mean absolute error (MAE) and mean relative error (MRE) by averaging AE and RE over all time periods in day i as,

$$\mathrm{MAE}_i = \frac{1}{m}\sum_{j=1}^{m}\mathrm{AE}_{ij},$$

and

$$\mathrm{MRE}_i = \frac{1}{m}\sum_{j=1}^{m}\mathrm{RE}_{ij}.$$

We also define the root mean squared error as

$$\mathrm{RMSE}_i = \sqrt{\frac{1}{m}\sum_{j=1}^{m}\mathrm{AE}_{ij}^2}.$$

Suppose, as in dynamic updating, in addition to the historical data, we have information available up to time period j for the current day i, we are interested in forecasting the remaining periods of day i. We define the mean absolute error (MAE) and mean relative error (MRE) by averaging AE and RE over time periods after period j in day i as,

$$\mathrm{MAE}_{ij} = \frac{1}{m-j}\sum_{j'=j+1}^{m}\mathrm{AE}_{ij'},$$

and

$$\mathrm{MRE}_{ij} = \frac{1}{m-j}\sum_{j'=j+1}^{m}\mathrm{RE}_{ij'}.$$

The root mean squared error is correspondingly defined as

$$\mathrm{RMSE}_{ij} = \sqrt{\frac{1}{m-j}\sum_{j'=j+1}^{m}\mathrm{AE}_{ij'}^2}.$$

5. Forecasting Performance Comparison

In this section, out-of-sample performances of different forecasting methods are compared using the evaluation methodology discussed in the previous section. The root-unroot method described in Section 2 is applied to all competing methods to deal with heteroscedasticity. We observe that the forecasting methods without the square root transformation do not perform as well (results not shown). This is not surprising, since SVD and least squares regression are both based on squared errors and therefore won't work well for data from a highly skewed distribution. A varying-coefficient AR(1) model as described in Section 3.2 is used to model the inter-day feature series when applying our methods.

5.1 The Benchmark

We consider a simple approach that is currently used in the call center industry as the benchmark. This approach, referred as the historical average (HA) approach, uses historical averages of the same day-of-the-week as a forecast for the current day's call volumes. Specifically, let d_i denote the day-of-the-week of day i. Denote $D_{iq} = \{i' : i' \leq i \text{ and } d_{i'} = d_q\}$ and let $|D_{iq}|$ denote the cardinality of D_{iq}. Then the HA forecast of $N_{n+h,j}$ based on the information up to day n can be expressed as

$$\widehat{N}_{n+h,j} = \frac{\sum_{i \in D_{n,n+h}} N_{ij}}{|D_{n,n+h}|}.$$

Below we describe a variant of the HA approach based on the root-unroot method aforementioned in Section 2. It will be used as our benchmark in the forecast comparison in Section 5. Let $X_{ij} = \sqrt{N_{ij} + 1/4}$. The HA forecast of $X_{n+h,j}$ is

$$\widehat{X}_{n+h,j} = \frac{\sum_{i \in D_{n,n+h}} X_{ij}}{|D_{n,n+h}|}.$$

The forecast of $N_{n+h,j}$ is therefore

$$\widehat{N}_{n+h,j} = \widehat{X}^2_{n+h,j} - 1/4 = \left(\frac{\sum_{i \in D_{n,n+h}} X_{ij}}{|D_{n,n+h}|} \right)^2 - \frac{1}{4}.$$

It is worthwhile to point out that, comparing with the HA approach applied to the raw data, applying the square-root transformation prior to the HA approach does not make much difference in terms of forecasting error in our real data example.

5.2 Day-to-day Forecasting

Our approach for one-day-ahead forecasting as described in Section 3.2 relies on a dimension reduction by SVD. The original large dimensional vector time series is represented using a small number K of basis vectors. In this study, we consider $K = 1, 2, 3$ and denote the methods as TS1, TS2, TS3 respectively. These methods are compared with the benchmark HA method.

Table 1 gives summary statistics of the MRE and RMSE of the forecasts from the HA method, along with the ratios of the corresponding statistics between HA and the other three methods. The best forecasting method is highlighted in bold for each summary statistic separately. Both TS2 and TS3 show substantial improvement over the benchmark HA,

Table 1: Summary statistics (mean, median, lower quartile Q1, upper quartile Q3) of MRE and RMSE in a rolling forecast exercise. The forecasting set contains 60 days. TS3 performs the best.

	MRE (%)				RMSE			
	Q1	Median	Mean	Q3	Q1	Median	Mean	Q3
HA	4.79	6.35	6.48	7.93	40.76	61.60	66.23	85.13
TS1/HA	1.01	0.96	1.08	1.07	1.07	0.81	0.90	0.80
TS2/HA	1.00	0.86	0.94	0.87	**0.99**	0.77	0.86	0.76
TS3/HA	**0.95**	**0.81**	**0.89**	**0.79**	1.02	**0.76**	**0.85**	**0.71**

while TS3 gives the best overall performance. We omit the results using MAE since they are similar to those using RMSE.

Figure 2 plots the averages of RE and AE as functions of time-of-day for the four competing methods, where the averages are taken over the 60 days in the forecasting set. Using both measures, the three methods TS1, TS2 and TS3 all outperform significantly over HA during 9:00AM-5:00PM; typical improvement in AE is around 10-20 calls every quarter hour. The inferior performance of HA is caused by the fact that it ignores the inter-day dependence of the call volume profiles. TS1 does not perform well in the beginning (before 8:00AM) and at the end (after 7:00PM) of the day, indicating that one intra-day feature vector is not enough in representing the intra-day profiles. TS3 gives the best overall performance while its improvement over TS2 is marginal. These results can be explained by the extracted intra-day features from the SVD. The first intra-day feature vector captures the average shape of the intra-day profiles, while the second and third intra-day feature vectors describe the differences among the intra-day profiles of different weekdays, especially in early mornings and evenings.

5.3 Dynamic Within-day Updating

For each day in the forecasting set, we consider dynamically updating the forecast for the rest of the current day at every quarter hour, using available information up to that quarter. The intra-day updating starts at 8:00AM, corresponding to $m_0 = 4$, in order to have enough data to estimate the parameters. The benchmark HA method does not utilize the new information from the current day and is implemented the same way as in the day-to-day forecasting.

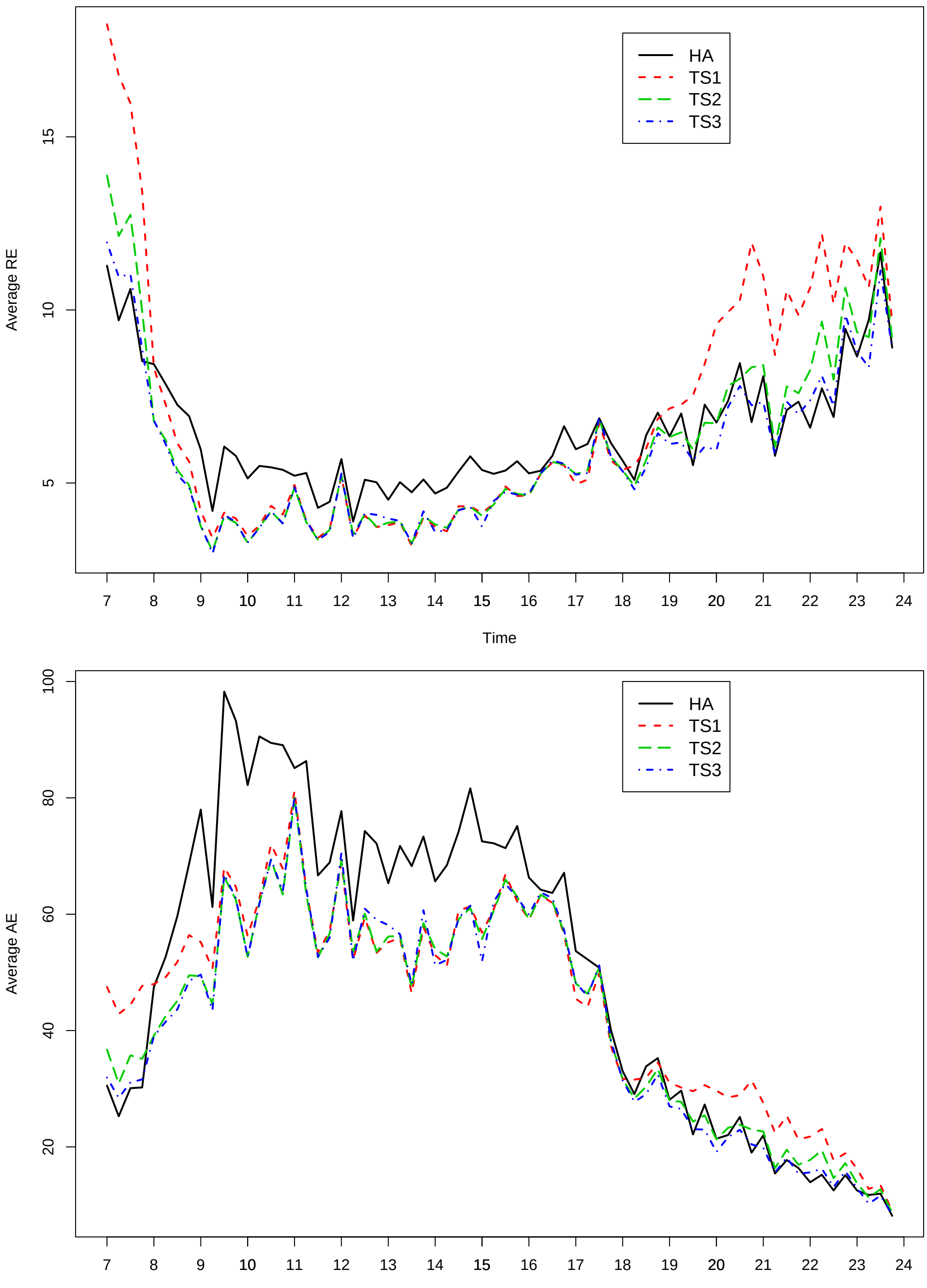

Figure 2: Comparison of Average RE and Average AE for the four one-day-ahead forecasting methods. TS3 performs the best.

We will consider the three updating methods described in Section 3.3, namely TS, LS, and PLS; see (4), (6) and (9). We expect that the PLS method should give the best forecast performance. This is confirmed empirically below. To implement these three methods, we need to choose the number of intra-day feature vectors for dimension reduction. In light of the one-day-ahead forecasting performance presented above in Section 5.2, we decide to use $K = 3$ and the three methods are therefore denoted as TS3, LS3, and PLS3. For each day in the forecasting set, we use its 150 preceding days to derive the time series forecast TS3. We employ the current day's incoming call volume information in addition to the 150 preceding days to produce the LS3 and PLS3 forecasts.

To use the PLS updating approach, we need to decide on the value of the penalty parameter λ at each updating point. To this end, we use the beginning 150 days in our data set as the training set. Note that this training set does not overlap with the forecasting set we initially hold out for out-of-sample forecast evaluation. We use the last one third (i.e., 50 days) of the 150 days in the training set as a rolling hold-out sample. For a given day in this hold-out sample, we use its preceding 100 days to extract the intra-day feature vectors and generate the time series forecast TS3. The extracted intra-day feature vectors and the obtained TS3 forecast are used for all updating points. Now fix an updating point. For each given λ, we construct the PLS updating for the given day. Pick a performance measure such as RMSE, MAE, or MRE. Compute the performance measure for all days in the hold-out sample and calculate the average value. Select λ by minimizing this average performance measure over a grid of candidate values. In this study, we choose λ from $\{0, 10, \ldots, 10^9\}$. When $\lambda = 0$, the PLS approach is the same as the LS approach. As λ goes to infinity, the PLS approach should converge to the TS approach, which is essentially achieved when $\lambda = 10^9$ in our study. The selected λ is used to generate the PLS forecast at the corresponding updating point for all days in the forecasting set.

Figure 3 plots the selected λ as functions of the updating points. We see that, for every selection criterion, the selected λ has a decreasing trend as the updating time progresses into the later part of the day. Note that a small λ value would let the PLS forecast put more weight on the LS forecast. Thus the observed decreasing trend for the selected λ makes intuitive sense. Because, as time progresses, more information is available about the current day's call arrival pattern, and the forecaster needs less influence from the TS forecast.

Figure 4 plots the averages of MRE and RMSE as functions of the updating points for the four approaches, where the averages are taken over the 60 days in the forecasting set.

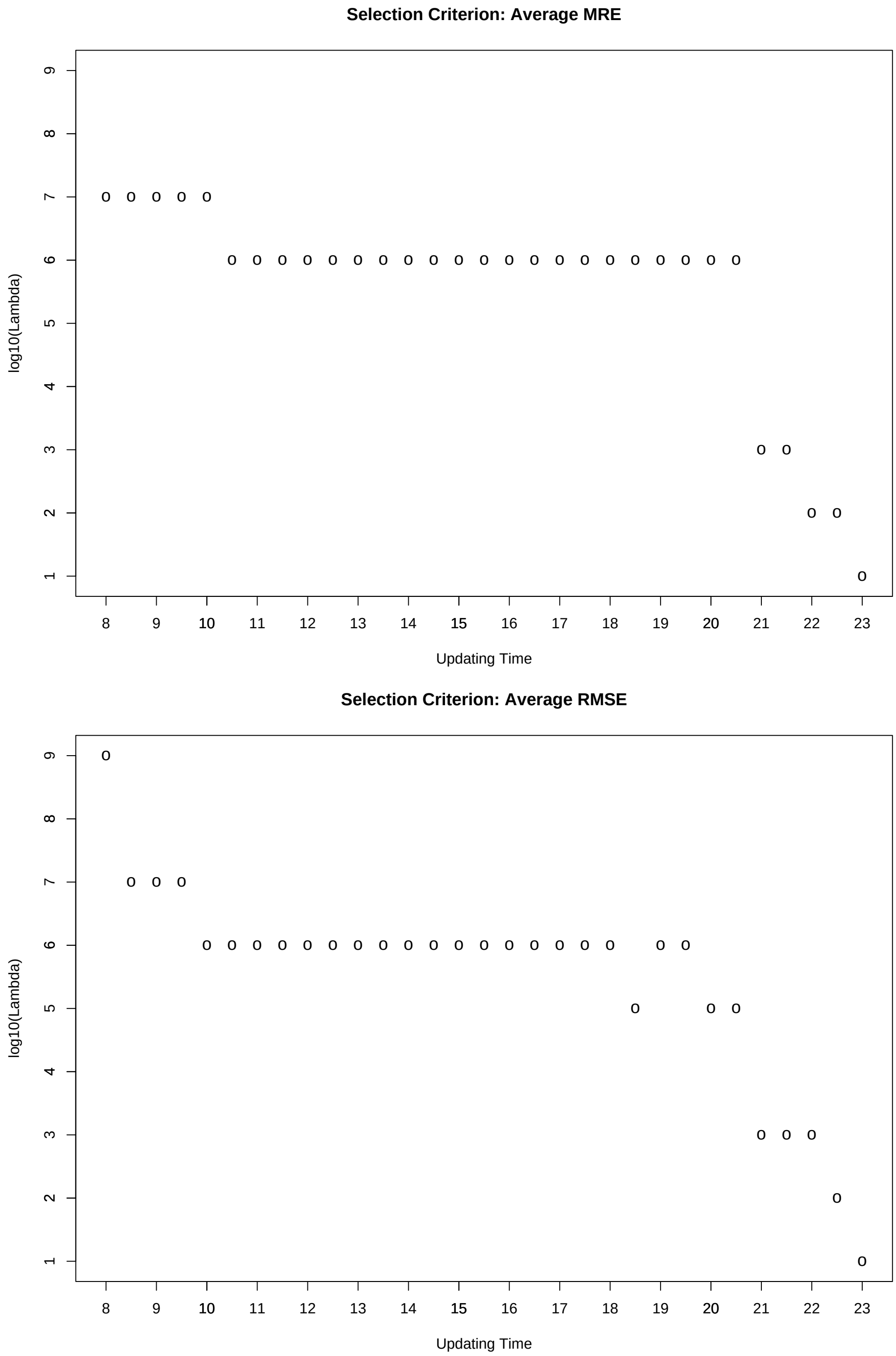

Figure 3: Penalty parameter selection for PLS3 using MRE and RMSE. The selected λ decreases as the time progresses, which means the TS forecast has less influence on the PLS forecast. This makes intuitive sense.

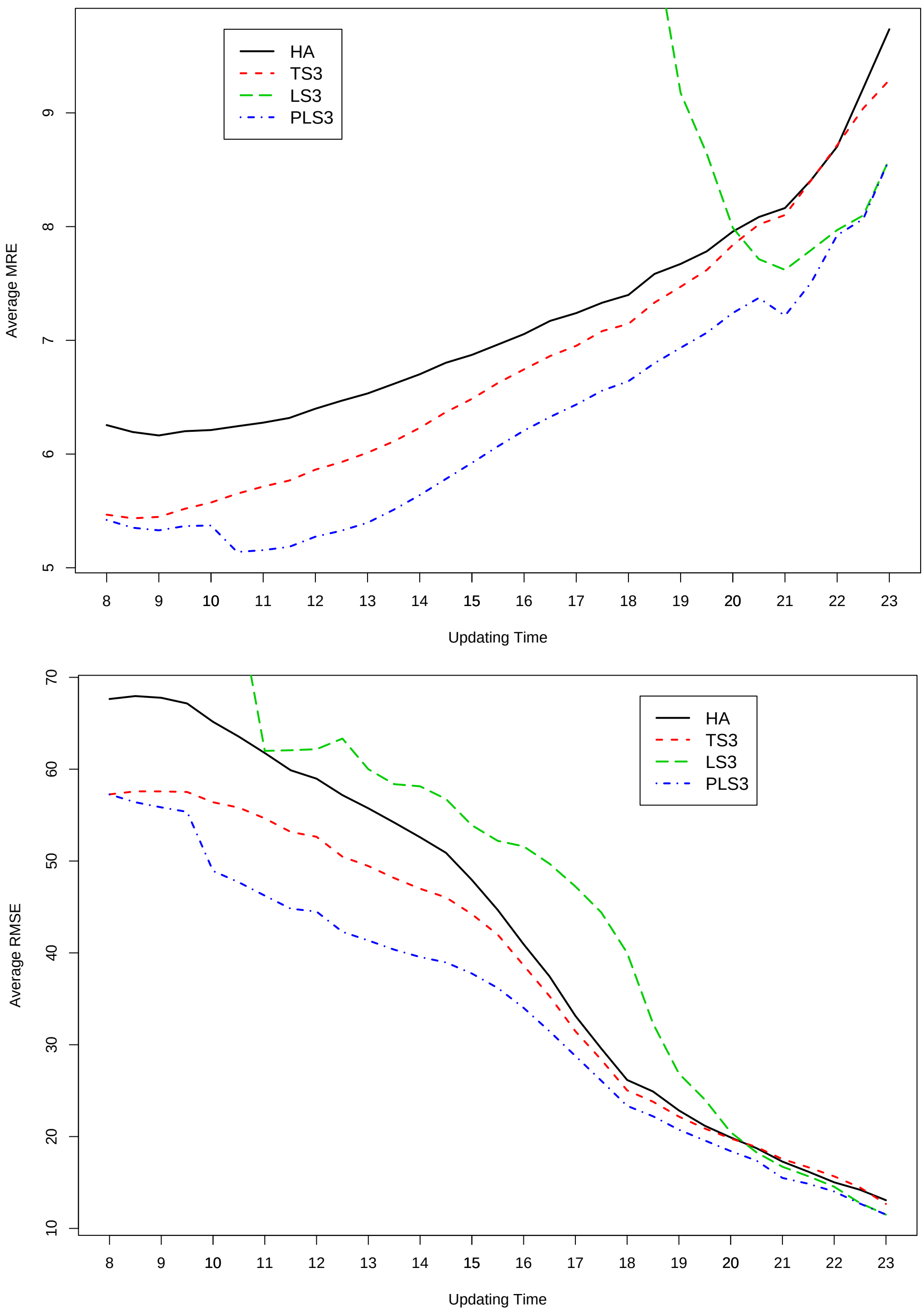

Figure 4: Comparison of Average MRE and Average RMSE as functions of the updating points. The averages are calculated over the 60 days in the forecasting set. PLS3 performs the best, and improves greatly over HA.

Note that MRE and RMSE are calculated at each updating point since we are interested in the performance of the updated forecasts at all time periods after the updating point. For PLS3, both MRE and RMSE are used to select λ, and give almost identical results. The reported results in Figure 4 are based on the λ selected using MRE if MRE is the out-of-sample forecasting performance measure, and similarly for RMSE. From these plots, PLS3 is clearly the winner, which has the smallest error measure at all updating points. One can also see that LS3 performs rather badly at the beginning of the day, especially for RMSE, which is due to the high collinearity between the initial parts of the second and third intra-day feature vectors. This also justifies the necessity for penalization. Note that TS3 does not perform dynamic updating. It just uses the information up to the end of the previous day and does not employ the current day's information. Thus, the significant improvement of PLS3 over TS3 shows the benefit of dynamic updating using the current day's information.

Table 2: Summary statistics (mean, median, lower quartile Q1 and upper quartile Q3) of MRE and RMSE of the four approaches for the 10:00AM and 12:00PM updatings. PLS3 improves greatly over the other methods.

	MRE (%)				RMSE			
	Q1	Median	Mean	Q3	Q1	Median	Mean	Q3
10:00AM								
HA	4.69	5.91	6.21	7.61	41.25	57.55	65.16	85.98
TS3/HA	0.94	0.86	0.90	0.76	0.97	0.83	0.87	0.74
LS3/HA	1.77	1.96	2.09	2.06	1.35	1.33	1.26	1.13
PLS3/HA	**0.93**	**0.83**	**0.86**	**0.75**	**0.91**	**0.77**	**0.75**	**0.60**
12:00PM								
HA	4.93	6.00	6.40	7.89	39.23	51.61	58.97	75.60
TS3/HA	0.95	0.89	0.92	0.79	0.93	0.88	0.89	0.78
LS3/HA	1.33	1.82	1.77	1.74	1.19	1.07	1.05	0.94
PLS3/HA	**0.89**	**0.81**	**0.82**	**0.71**	**0.91**	**0.77**	**0.75**	**0.58**

Below, we look at the 10:00AM updating and the 12:00PM updating individually, in order to get some idea about how different approaches perform for individual updating. Table 2 presents summary statistics of the MRE and RMSE of the forecasts from the HA method, along with the ratios of the corresponding statistics between HA and the other three methods, for the two updatings respectively. For the four competing approaches, Figures 5 and 6 plot the averages of RE and AE for each time period during the remaining part of the day after the corresponding updating. The averages are calculated over the 60 days in the forecasting set. The superior performance of PLS3 over the other methods is again quite clear. The improvement of PLS3 over the benchmark HA is substantial.

To illustrate the effects of the LS and PLS updating methods, Figure 7 plots the actual call volume profile for September 3 (a Wednesday), along with the updated forecasts from the four updating methods after the 12:00PM updating. We observe a dramatic shift from the HA/TS3 updates to the LS3/PLS3 updates. After incorporating the same-day volumes prior to 12:00PM, the LS3/PLS3 forecasts are much closer to the actual volumes. The reason is the following: September 3 is the second day after the Labor Day, which has a strong lagging holiday effect and leads to unusually high call volumes in September 3. Consequently, the TS3 forecasts of the inter-day features for September 3 fall below the actual values, as well as the volume forecasts. The same problem occurs for the HA approach. By balancing between the TS3 and LS3 forecasts, the PLS3 forecasts automatically adjust for the unusual arrival pattern, and are closer to the actual call volumes in the late afternoon and evening.

6. Conclusion

Our approach to forecasting call arrivals is based on the viewpoint that the intra-day call volume profiles form a vector time series. This point of view represents clearly that the data have a two-way temporal structure, that is, the inter-day and intra-day variations. The SVD, used as a dimensionality reduction tool, effectively separates out these two types of variations. The intra-day variation is summarized by a few intra-day feature vectors that can be extracted from historical data, while the inter-day variation is modelled by a few inter-day feature time series (see Section 3.1). The SVD is also efficient in our data example: three pairs of inter-day and intra-day features are sufficient to generate good forecasts for a 68-dimensional time series.

Assuming that the intra-day features won't change in the future, the problem of forecasting intra-day profile reduces to the problem of forecasting the inter-day features. We propose to use univariate time series techniques to forecast the inter-day feature series in order to generate one- or multi-day-ahead forecasts. For dynamic within-day updating, we have considered three ways to update the inter-day feature forecasts: 1. Apply the day-to-day forecast, using information available up to the end of the previous day (i.e. no updating); 2. Fit a least squares regression using the observed portion of the current day's volume profile on the corresponding components of the intra-day feature vectors (i.e. LS updating);

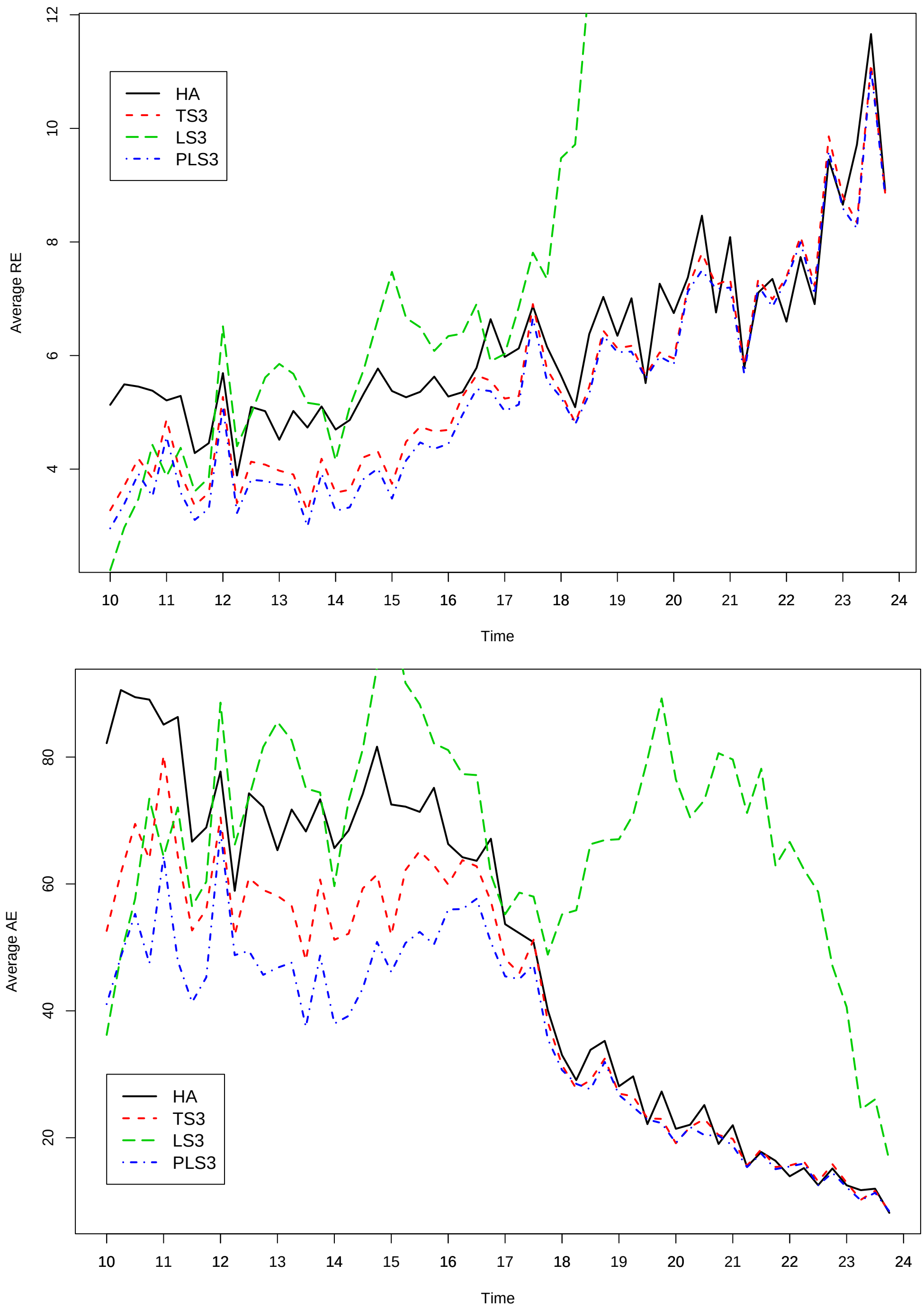

Figure 5: Comparison of Average RE and Average AE of the four approaches for the 10:00AM updating. The averages are calculated over the 60 days in the forecasting set. PLS3 performs the best.

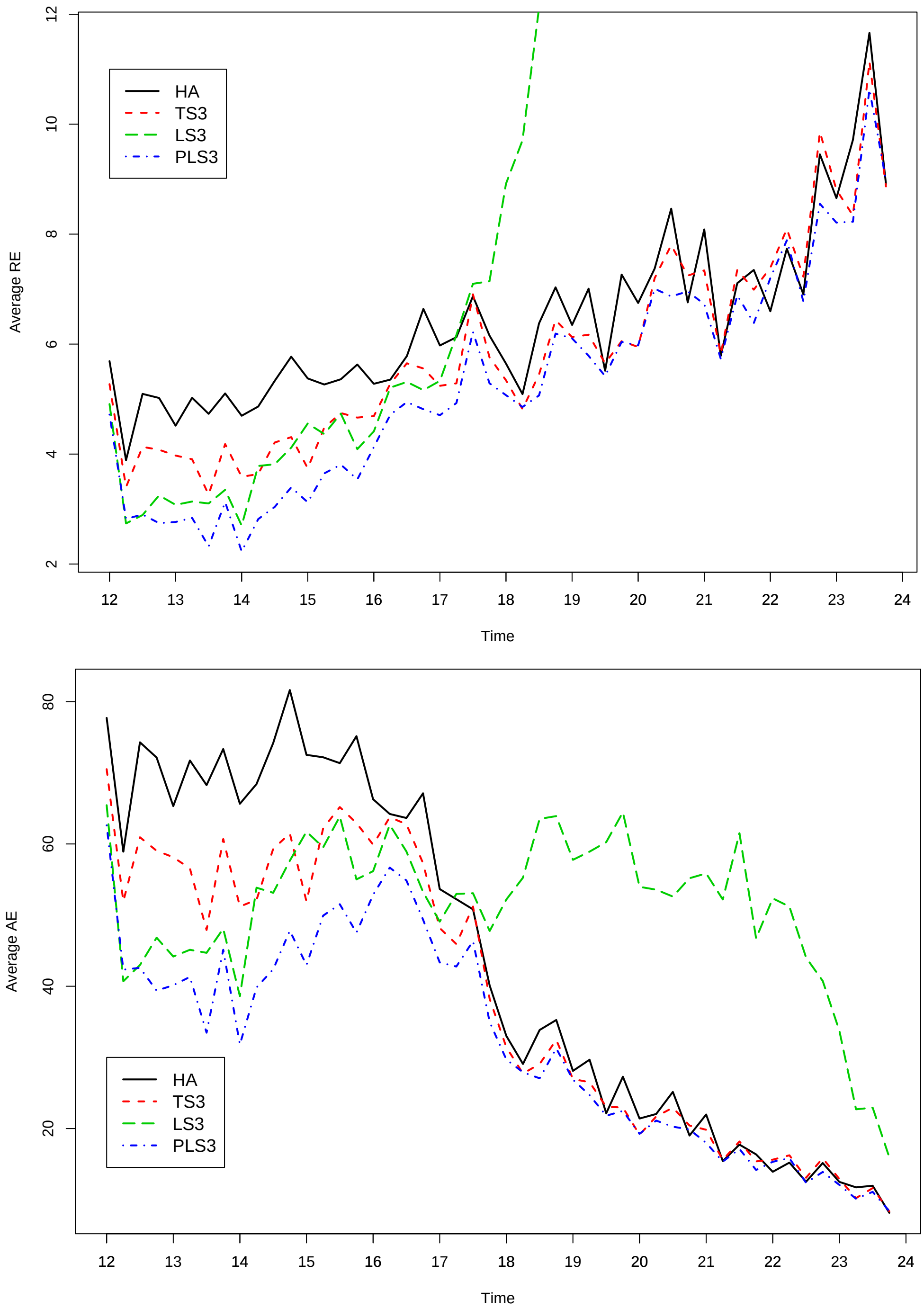

Figure 6: Comparison of Average RE and Average AE of the four approaches for the 12:00PM updating. The averages are calculated over the 60 days in the forecasting set. PLS3 performs the best.

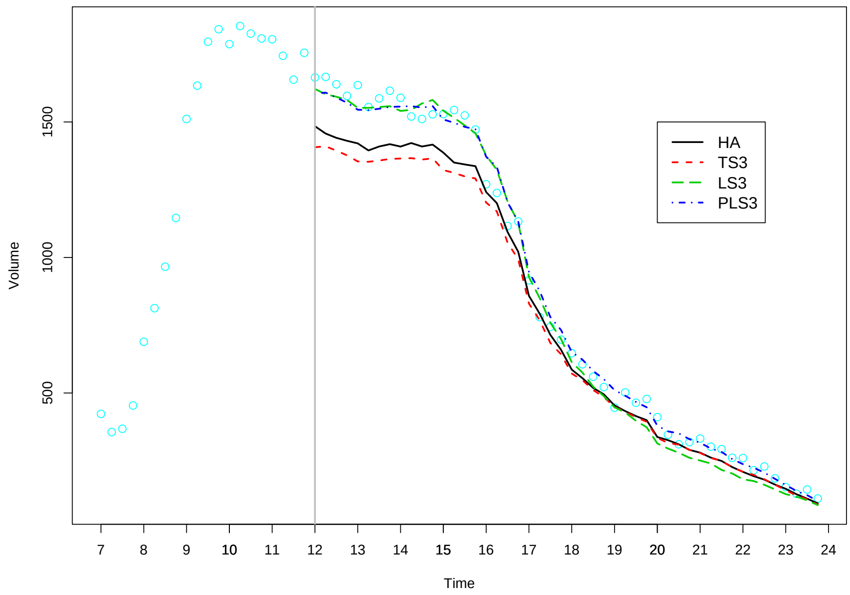

Figure 7: Comparison of the Updated Forecasts after the 12:00PM updating for Wednesday, September 3, 2003. The circles plot the actual call volume during the quarter hour intervals. This shows pictorially how LS3 and PLS3 work.

3. Combine the current day's available information with the previous day's information using penalized least squares (i.e. PLS updating). Not surprisingly, the third method (PLS) performs the best and should be the recommended method.

Our methods show significant improvements over the existing industry standard in an out-of-sample forecasting comparison. Despite the good performance, our methods are not much more complicated than the industry standard. They are easy to implement. SVD can be obtained using any software that does matrix computation such as MATLAB and SPLUS/R; and solving penalized least squares is as straightforward as performing a ridge regression. The R codes for performing our forecasting and updating methods are available from the authors upon request.

We have employed a square-root transformation to overcome the non-constant variance problem. Other variance stabilization transformation can also be used with our forecasting methods and might be more appropriate for other data sets. However, the root-unroot method of Brown et al. (2005) can be justified by assuming a Poisson distribution for the call volume within a short time interval. This assumption works well in our data set as an approximation, and is reasonable in many practical situations. In fact, when applying the standard queuing methodology, call center practitioners commonly assume that call arrivals follow a Poisson process with constant rates for short time intervals, which in turn implies Poisson distributions for the corresponding call volumes.

A natural direction for future research is to combine our dimensionality reduction idea with stochastic modelling, such as modelling the arrival process as an inhomogeneous Poisson process (Brown et al., 2005). In stochastic modelling, data transformation is not necessary and a distribution based approach for estimation is more appealing. Such an approach will be inevitably more complex than the approach proposed in the current paper. It would be interesting to investigate whether the distributional approach will yield better forecasts.

References

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

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

Bianchi, L., J. Jarrett, R. C. Hanumara. 1993. Forecasting incoming calls to telemarketing centers. *Journal of Business Forecasting Methods and Systems* **12** 3–12.

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

Box, G. E. P., G. M. Jenkins, G. C. Reinsel. 1994. *Time Series Analysis: Forecasting and Control*. Third Edition. Prentice Hall, New Jersey.

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

Draper, N. R., H. Smith. 1998. *Applied Regression Analysis*. 3rd ed. Wiley-Interscience, New York.

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

Harvey, A. C. 1990. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, New York.

Harville, D. A. 1997. *Matrix Algebra From a Statistician's Perspective*. Springer-Verlag: New York.

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

Reinsel, G. C. 2003. *Elements of Multivariate Time Series Analysis*. 2nd ed. Springer.

Shen, H., J. Z. Huang. 2005. Analysis of call centre arrival data using singular value decomposition. *Applied Stochastic Models in Business and Industry* **21** 251–263.

Soyer, R., M. M. Tarimcilar. 2004. Modeling and analysis of call center arrival data: A Bayesian approach. *Submitted to Management Science* .

Steckley, S. G., S. G. Henderson, V. Mehrotra. 2004. Service system planning in the presence of a random arrival rate. Tech. rep., Cornell University.

Weinberg, J., L. D. Brown, J. R. Stroud. 2005. Bayesian forecasting of an inhomegeneous Poisson process with applications to call center data. *Submitted to Journal of the American Statistical Association* .