

On Modeling and Forecasting Time Series of Smooth Curves

Haipeng Shen *

October 16, 2008

Abstract

We consider modeling a time series of smooth curves and develop methods for forecasting such curves and dynamically updating the forecasts. The research problem is motivated by efficient operations management of telephone customer service centers, where forecasts of daily call arrival rate profiles are needed for service agent staffing and scheduling purposes. Our methodology has three components: dimension reduction through a smooth factor model, time series modeling and forecasting of the factor scores, and dynamic updating using penalized least squares. The proposed methods are illustrated via the motivating application and two simulation studies.

KEY WORDS: dimension reduction; factor models; functional data analysis; penalization; shrinkage; singular value decomposition.

*Corresponding address: Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. Email: shenhaipeng@gmail.com.

1 Introduction

Time series of curves, or functional time series, exist in many disciplines. For example, in demography and epidemiology, researchers observe age-specific mortality rate or fertility rate curves over many years, and are interested in forecasting future mortality/fertility rate curves (Hyndman and Ullah, 2007; Erbas et al., 2007). Other examples include electricity system load (Smith, 2000), *El Niño* series (Besse et al., 2000; Valderrama et al., 2002), and television audience rates (Antoniadis et al., 2006). In this paper we propose a methodology for forecasting time series of *smooth* curves and performing dynamic within-curve updating. Our proposal combines the idea of nonparametric smoothing with time series factor models.

Our interest in time series of curves is motivated by a modern business application of forecasting future call arrival rate profiles to telephone customer service centers. Such centers have become the primary service channel of many service providers. For effective management of such centers, managers rely on accurate forecasts of the underlying arrival rates in order to match up call center workforce with future workload. See Section 2 for a detailed description of this motivating application. Briefly, such centers have been traditionally analyzed as queueing systems using queueing theoretical models; more recently, service engineering researchers have been proposing to complement the queueing models with carefully chosen statistical inference techniques (Gans et al., 2003), which to some extent motivates our current research.

As one of the key primitives of the queueing models, the call arrival process is usually assumed as a time-varying Poisson process with a smooth rate function (Brown et al., 2005; Weinberg et al., 2007), and the Poisson rates during short time intervals such as quarter hours are then used as input parameters to various workforce management softwares. However the rates are unobservable; hence one needs to forecast future rate profiles based on historical call volumes during such intervals. The volumes

can be viewed as an observed noisy version of the underlying arrival rates. Following the general framework of functional data analysis (Ramsay and Silverman, 2002), it is convenient to treat the daily arrival rate profile as being discretely sampled from some underlying smooth arrival rate function (or curve). Hence, our research question is that we observe some noisy discrete realizations from a time series of smooth curves, and are interested in forecasting future smooth curves. Imposing smoothness on the rate forecasts has operational benefits as discussed later.

Curve forecasting is a problem of infinite dimension, as the quantities of interest are future curves. To achieve dimension reduction, we introduce a low-dimensional factor model that constraints the factors to be *smooth*. To obtain the smooth factors and associated factor scores, we combine low-rank matrix approximation with smoothness-inducing penalties. Following dimension reduction, each curve in the time series is represented as a linear combination of a small number of factors plus noise, see (2) in Section 3.1. Low-dimensional time series models can then be built on the factor score series to forecast future scores. The resulting forecasts are subsequently combined with the extracted factors to produce forecasts of future curves. For dynamic within-curve updating, we introduce a penalization approach that combines information from the historical curves and the early part of the current curve. It updates the existing time series forecasts of the factor scores for the current curve to best fit the newly-observed early segment of the curve, while appropriately penalizing a large departure from the existing forecasts. This penalization approach is closely related to ridge regression and also has a Bayesian interpretation. Distributional forecasts and updates can be obtained within our framework through bootstrapping.

The current paper extends and improves the methods of Shen and Huang (2005, 2008) where singular value decomposition (SVD) of the data matrix was used for dimension reduction. The authors focused on forecasting future call volumes instead of rates, and their approach ignores the *smooth* nature of the underlying arrival rate

curve. Introducing smoothness improves the forecast accuracy when the noise level is high (see Section 5.1). Moreover, a smooth arrival rate curve forecast has beneficial effects on the follow-up agent staffing and scheduling, because the staffing level won't change dramatically between consecutive time periods, and therefore the scheduling is simpler and more efficient. In another related work, Weinberg et al. (2007) developed a Bayesian model for forecasting call center arrivals that specifically assumes different weekday has its own smooth arrival rate profile. However, the implementation of their Bayesian approach needs special expertise with Markov Chain Monte Carlo. In contrast, the methods proposed in this paper are easier to implement that enables wider dissemination.

Various proposals have been introduced for the curve-forecasting problem in the functional data analysis (FDA) literature motivated by different applications; however none of the papers have studied the problem of dynamic within-curve updating. Using a two-step procedure similar to ours, Valderrama et al. (2002) used natural cubic splines to interpolate the observed discrete data when forecasting annual *El Niño* series, while Hyndman and Ullah (2007) and Erbas et al. (2007) first smoothed each discrete curve nonparametrically before decomposing the smoothed curves via a basis function expansion, while considering forecasting mortality and fertility rates. Different from these works, our approach does not require pre-smoothing of the data and naturally combines nonparametric smoothing with factor models. Moreover, our approach allows different smoothness for different smooth factors while a single smoothing parameter has been used in the curve forecasting literature so far. Under a different school of thinking, Besse et al. (2000) and Antoniadis et al. (2006) extended the idea of nonparametric kernel regression for forecasting univariate time series to functional data, forecasting the entire annual cycle of *El Niño* Southern Oscillation series, electrical power consumption and television audience rates.

The roadmap of the paper is the following. Section 2 introduces the motivating

application and the data. Section 3.1 describes the proposed smooth factor model and the estimation technique through penalized low-rank approximation, as well as several related methodological details. Our methods for curve forecasting and dynamic within-curve updating are discussed in Sections 3.2 and 3.3. To illustrate our methods, Section 4 presents the call center application, while Section 5 considers two simulation studies.

2 One Motivating Application

Our primary motivating application is to forecast *smooth* arrival rate profiles of daily customer service calls to a US banking call center, and to update existing forecasts using additional information during a business day. We are also interested in understanding dominant call arrival patterns.

Call centers have become an increasingly important component of service providers in our modern business world (Gans et al., 2003). Call center agents represent 4% of the workforce in US, 1.8% in Israel and 1.5% in UK (AFII, 2004). Every FORTUNE 500 company has at least one call center, which employs on average 4,500 agents; more than \$300 billion is spent annually on call centers worldwide (Gilson and Khandelwal, 2005).

Brown et al. (2005) state that human resource cost accounts for about 70% of the call center running expense. Hence efficient agent staffing and scheduling, which relies on accurate forecast of future call arrival rates, is one crucial operational problem. Call center managers usually need day-to-day forecasts of future call rate profiles. Recently, Weinberg et al. (2007) and Shen and Huang (2008) studied within-day updating of existing forecasts for the latter segment of a day using newly available information during the early segment. As stated by Shen and Huang, such updating

is operationally beneficial and feasible, which (if performed appropriately) results in higher efficiency and productivity: based on the revised more accurate forecasts, a manager can adjust staffing levels correspondingly, by offering overtime to agents on duty, calling in part-time or work-from-home agents, reducing cross-selling; or dismissing agents from work early, moving them to handle other activities such as email inquiries and faxes, increasing cross-selling.

We use the same data analyzed by Shen and Huang, which are collected at an inbound call center of a major northeastern US financial firm. For every weekday between January 6 and October 24, 2003, the data record the number of customer calls handled by the center during every quarter hour between 7:00AM and midnight.

Let N_{ij} denote the call volume during the j -th quarter-hour interval on the i -th day, $j = 1, \dots, 68$, $i = 1, \dots, 210$. In the call center literature, both Brown et al. (2005) and Weinberg et al. (2007) model N_{ij} using a time-varying Poisson process with a random rate function $\lambda_i(t)$ where t stands for time-of-the-day. They apply the square-root transformation to N_{ij} to stabilize the variance and make the data approximately normally distributed. The theoretical justification is recently provided by Brown et al. (2007): if $N \sim \text{Poisson}(\lambda)$, then $Y = \sqrt{N + 1/4}$ has approximately a mean of $\sqrt{\lambda}$ and a variance of $1/4$; in addition, as $\lambda \rightarrow \infty$, Y is approximately normal.

We follow their practice and henceforth work with the square-root transformed call volumes $Y_{ij} = \sqrt{N_{ij} + 1/4}$. We view the transformed volumes during day i as observations from some volume curve $y_i(t)$, and denote the *underlying* square-root-transformed arrival rate as $x_i(t) \equiv \sqrt{\lambda_i(t)}$, which we assume to *vary smoothly* over t . Our interests lie in forecasting future smooth arrival rate function $x_i(t)$ or $\lambda_i(t)$, based on the observed (transformed) volumes Y_{ij} .

For illustration purposes, the left panel of Figure 1 plots the average square-root-

transformed call volume profiles for different weekdays. Several observations can be made. First, there exists some common bi-modal within-day arrival pattern with an initial 11:00AM peak followed by a lower 2:00PM peak. Second, the underlying arrival patterns seem to be smooth, which motivates our curve-forecasting approach. Besides the common bi-modal trend, different weekdays also exhibit different arrival patterns. For example, Monday mornings start off the slowest, while Fridays begin with the fastest climb-up and finish with the quickest drop-off. The difference among weekdays is also obvious in the total daily (square-root) volumes as depicted in the right panel of Figure 1. The periodic weekly pattern is evident: Mondays are the busiest, followed by Fridays and then Tuesdays to Thursdays. Such pattern helps when building time series forecasting models in Section 4.2.

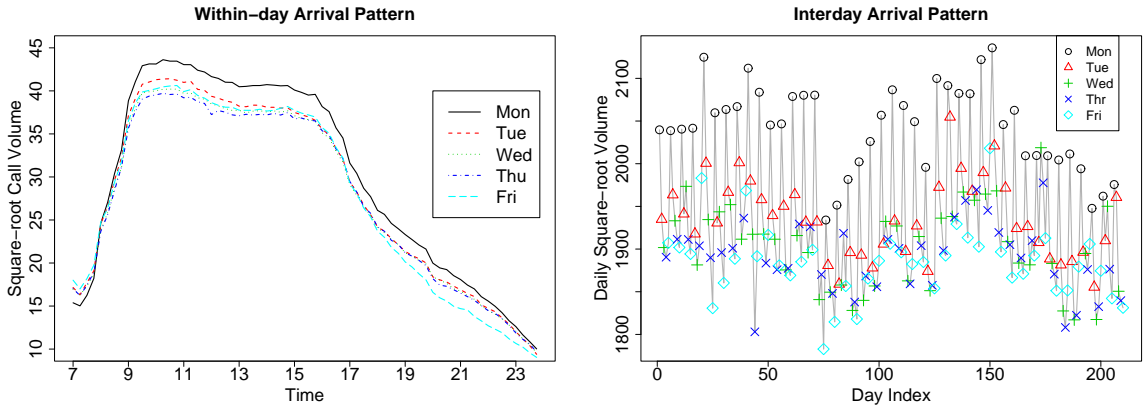


Figure 1: Average Within-day Arrival Patterns and Total Daily Volumes for Different Weekdays. Day-of-the-week effects are present in both plots.

3 Forecasting Methods

3.1 Dimension Reduction via Smooth Factor Models

Consider a time series of curves $\{y_i(t) : t \in \mathcal{T}; i = 1, \dots, n\}$. We assume that each curve consists of an underlying *smooth* curve $x_i(t)$ plus noise $\epsilon_i(t)$:

$$y_i(t) = x_i(t) + \epsilon_i(t), \quad i = 1, \dots, n. \quad (1)$$

We are interested in predicting some future smooth curve $x_{n+h}(t)$, $h > 0$. In the application reported in Section 2, $y_i(t)$ records the square-root-transformed call volumes at time t , while $x_i(t)$ is the underlying square-root arrival rate profile that is assumed to vary smoothly across time t within a day.

For dimension reduction, we represent/approximate the underlying smooth curves $x_i(t)$ by a small number of smooth factors, that is,

$$x_i(t) = \beta_{i1}f_1(t) + \dots + \beta_{iK}f_K(t), \quad (2)$$

where $f_k(t)$ are the smooth factor curves, and $\beta_{i1}, \dots, \beta_{iK}$ are the factor scores. To achieve dimension reduction, the number of factors K is chosen to be much smaller than n . Out-of-sample forecasting performance measures can be used to decide on K , as illustrated later in Section 4.2. After extracting the smooth factor curves and the associated factor scores from historical data, we apply univariate or low-dimensional vector time series models to forecast the factor score series; the forecasted scores are then combined with the extracted smooth factors to obtain a forecast for the future smooth curve $x_{n+h}(t)$, $h > 0$.

For the rest of this subsection, we focus on the method to extract the smooth factor curves and associated factor scores from the historical data. Time series forecasting of curves will be discussed in the subsequent Sections 3.2 and 3.3.

In nearly all practical situations, curves are only observed discretely. Denote the common observation times as t_1, \dots, t_m . Then, (1) and (2) together give

$$Y_{ij} \equiv y_i(t_j) = \beta_{i1}f_1(t_j) + \dots + \beta_{iK}f_K(t_j) + \epsilon_i(t_j), \quad i = 1, \dots, n; j = 1, \dots, m, \quad (3)$$

which can be rewritten in matrix expression as

$$\mathbf{Y} = \mathbf{B}\mathbf{F}^T + \mathbf{E},$$

where $\mathbf{Y} = (Y_{ij})$, $\mathbf{B} = (\beta_{ij})$, $\mathbf{F} = (f_k(t_j))$, and $\mathbf{E} = (\epsilon_i(t_j))$. We can extract \mathbf{B} and \mathbf{F} from \mathbf{Y} by minimizing the following sum of squared errors (SSE),

$$\|\mathbf{Y} - \mathbf{B}\mathbf{F}^T\|_F^2 = \text{tr}\{(\mathbf{Y} - \mathbf{B}\mathbf{F}^T)(\mathbf{Y} - \mathbf{B}\mathbf{F}^T)^T\}, \quad (4)$$

where $\|\cdot\|_F$ denotes the Frobenius norm. The solution of this minimization problem can be obtained using the singular value decomposition of \mathbf{Y} (Eckart and Young, 1936). Precisely, Let $\mathbf{Y} = \mathbf{U}\mathbf{S}\mathbf{V}^T$ denote the SVD of \mathbf{Y} , where \mathbf{U} and \mathbf{V} have orthonormal columns and \mathbf{S} is diagonal. Then the minimizing \mathbf{B} of (4) is the matrix consisting of the first K columns of $\mathbf{U}\mathbf{S}$ and the minimizing \mathbf{F} is the matrix consisting of the first K column of \mathbf{V} . This is essentially the SVD-based dimension reduction step of the forecasting methods of Shen and Huang (2008).

There are two drawbacks of the SVD-based dimension reduction for smooth curves. First, the smooth nature of the factors is ignored. Second, only values at the given grid have been obtained for each factor and thus some interpolation is needed to get the whole curve. Based on a modification of the functional PCA method by Huang et al. (2008), we propose to extract the smooth factors by introducing a roughness penalty to the SSE criterion in (4).

The first smooth factor $f_1(t)$ and the associated scores β_{i1} are obtained by minimizing the following penalized sum of squares,

$$\sum_{i=1}^n \sum_{j=1}^m \{Y_{ij} - \beta_i f(t_j)\}^2 + \omega \left(\sum_{i=1}^n \beta_i^2 \right) \int \{f''(t)\}^2 dt, \quad (5)$$

with respect to β_i and a function $f(\cdot)$ that satisfies $\int \{f''(t)\}^2 dt < \infty$. Here $\omega > 0$ is a smoothing parameter whose choice will be discussed later. Subsequent smooth factors and associated scores can be obtained sequentially from residual matrices. The sequential approach makes it easy to allow different smoothing parameters for different factors, which appears beneficial as shown in Section 5.2.

Denote $\boldsymbol{\beta} = (\beta_1, \dots, \beta_n)^T$ and $\mathbf{f} = (f_1, \dots, f_m)^T$ with $f_j \equiv f(t_j)$. According to Theorem 2.1 of Green and Silverman (1994), the minimizer $\hat{f}_1(\cdot)$ of (5) is a natural cubic spline with knots at t_j that interpolates the minimizer $\hat{\mathbf{f}}$ of the following criterion,

$$\min_{\boldsymbol{\beta}, \mathbf{f}} \{ \|\mathbf{Y} - \boldsymbol{\beta} \mathbf{f}^T\|_F^2 + \omega \boldsymbol{\beta}^T \boldsymbol{\beta} \mathbf{f}^T \boldsymbol{\Omega} \mathbf{f} \}, \quad (6)$$

where $\boldsymbol{\Omega}$ is a penalty matrix such that $\int \{f''(t)\} dt = \mathbf{f}^T \boldsymbol{\Omega} \mathbf{f}$. The explicit expression of $\boldsymbol{\Omega}$ is given in the Appendix, along with an algorithm for computing the optimizing spline.

Given $\boldsymbol{\beta}$, the optimization problem (6) can be viewed as a ridge regression; hence, the minimizer is $\mathbf{f} = (\mathbf{I} + \omega \boldsymbol{\Omega})^{-1} \mathbf{Y}^T \boldsymbol{\beta}$. This suggests a generalized cross-validation (GCV) criterion for selecting the smoothing parameter ω :

$$\text{GCV}(\omega) = \frac{1}{m} \frac{\|\mathbf{Y}^T \boldsymbol{\beta} - \mathbf{f}\|^2}{\left(1 - \frac{1}{m} \text{tr}[\mathbf{S}(\omega)]\right)^2}, \quad (7)$$

where $\mathbf{S}(\omega) = (\mathbf{I} + \omega \boldsymbol{\Omega})^{-1}$ is the smoothing matrix. We select ω by minimizing this criterion. Some computation issues in optimizing the GCV criterion are discussed in the Appendix.

Remark 3.1.1 We adopt the formulation of functional PCA of Huang et al. (2008) because of its close connection with SVD. Other formulations of functional PCA use eigen decomposition of regularized sample covariance matrix (Ramsay and Silverman, 2002). Different from functional PCA, we do not center the curves before extracting the smooth factors. This makes our approach directly connected with the

Bayesian model of Weinberg et al. (2007) and the SVD approach of Shen and Huang (2008). We also found that centering does not help improve forecasting performance.

3.2 Curve Forecasting

Suppose we have available the discretely observed historical curves $y_1(t), \dots, y_n(t)$. We are interested in forecasting the h -step-ahead underlying smooth curve $x_{n+h}(t)$ ($h > 0$) based on the historical data. We assume that the smooth factor model (2) is valid for future curves.

Using the penalized low-rank approximation procedure of Section 3.1, we can derive the first K smooth factors $\{f_k(t)\}$ and their score series $\{\beta_k\}$. For simplification purposes, we drop the hats on these notations. We then keep the extracted smooth factors (or their discrete siblings) as fixed. The model (2) suggests that forecasting a future curve $x_{n+h}(t)$ reduces to forecasting the K -dimensional score time series $\{\beta_1, \dots, \beta_K\}$.

Denote the factor score vector for the curve $x_{n+h}(t)$ as $\beta_{(n+h)} = (\beta_{n+h,1}, \dots, \beta_{n+h,K})^T$, and its time series forecast as $\hat{\beta}_{(n+h)}^{\text{TS}} = (\hat{\beta}_{n+h,1}^{\text{TS}}, \dots, \hat{\beta}_{n+h,K}^{\text{TS}})^T$. Then a point forecast of $x_{n+h}(t)$ is given by

$$\hat{x}_{n+h}(t) = \hat{\beta}_{n+h,1}^{\text{TS}} f_1(t) + \dots + \hat{\beta}_{n+h,K}^{\text{TS}} f_K(t).$$

Note the distinction between two notations β_i and $\beta_{(i)}$: β_i stands for the i -th factor score series, while $\beta_{(i)}$ denotes the vector of the scores for the i -th curve.

To obtain the score forecast $\hat{\beta}_{(n+h)}^{\text{TS}}$, one can either employ univariate time series models on β_k separately, or low-dimensional vector time series models on $\{\beta_1, \dots, \beta_K\}$ simultaneously. Because of the way how the smooth factors are obtained, we expect that β_k is approximately orthogonal to β_l for $k \neq l$. Therefore univariate time series methods seem sufficient as shown in Section 4.

Sometimes one needs distributional forecasts for $x_{n+h}(t)$. One possibility is to use a bootstrap approach to derive such forecasts. For a fixed k , we can use the fitted time series model for β_k recursively, and bootstrap the model errors from the fitted model to generate B forecasted series $\{\hat{\beta}_{n+1,k}^{\text{TS},b}, \dots, \hat{\beta}_{n+h,k}^{\text{TS},b}\}$, $1 \leq b \leq B$. For example, consider an AR(1) model for β_k , $\beta_{ik} = a_0 + a_1\beta_{i-1,k} + \epsilon_{ik}$, and denote the residuals after fitting the model as $\hat{\epsilon}_{ik} = \beta_{ik} - (\hat{a}_0 + \hat{a}_1\beta_{i-1,k})$. A bootstrap sample of $\beta_{n+1,k}$ is obtained as $\hat{\beta}_{n+1,k}^{\text{TS},b} = \hat{a}_0 + \hat{a}_1\beta_{nk} + \epsilon_{nk}^b$, $1 \leq b \leq B$, where ϵ_{nk}^b is obtained by sampling with replacement from $\{\hat{\epsilon}_{ik}\}$. Bootstrap samples of $\beta_{n+2,k}, \dots, \beta_{n+h,k}$ can be obtained similarly. The bootstrap forecasts of the factor scores lead to B forecasts of $x_{n+h}(t)$ as follows,

$$\hat{x}_{n+h}^{\text{TS},b}(t) = \hat{\beta}_{n+h,1}^{\text{TS},b}f_1(t) + \dots + \hat{\beta}_{n+h,K}^{\text{TS},b}f_K(t), \quad b = 1, \dots, B,$$

from which interval and density forecasts can be derived. In some cases, parametric bootstrapping can be performed if we can assume some parametric distribution for ϵ_{ik} . In the above AR(1) example, if ϵ_{ik} is normally distributed, then one can sample ϵ_{nk}^b from the fitted normal distribution of the model errors $\{\hat{\epsilon}_{ik}\}$.

3.3 Dynamic Curve Updating

Besides the historical curves $\{y_i(t)\}$, $1 \leq i \leq n$, suppose we also observe the initial segment of the curve $y_{n+1}(t)$, denoted as $\mathbf{y}_{n+1}^e \equiv (y_{n+1}(t_1), \dots, y_{n+1}(t_{m_0}))^T$, $1 \leq m_0 < m$. One interesting question is to update the forecast of the *unobserved* latter segment of the curve $x_{n+1}(t)$, denoted as $x_{n+1}^l(t)$ for $t > t_{m_0}$. This within-curve updating problem is motivated by the call center application (Section 2). To increase operational efficiency and quality of service, as new calls come in, a manager is interested in dynamically updating the existing arrival rate forecast for the remainder of the day. Based on the upward or downward updates, she can either call in back-up agents or arrange overtime; or schedule meetings or training sessions for excess agents (Wein-

berg et al., 2007; Shen and Huang, 2008).

To perform the updating, we note that two sets of information are available: the historical curves $\{y_i(t), 1 \leq i \leq n\}$, and the initial segment of $y_{n+1}(t)$, i.e. \mathbf{y}_{n+1}^e . A reasonable updating scheme would appropriately combine these two information sets. Depending on how early (or late) the updating occurs, they should have different influence in determining the updated forecasts.

Suppose $\hat{\boldsymbol{\beta}}_{(n+1)}^{\text{TS}}$ is a time series forecast for $\boldsymbol{\beta}_{(n+1)}$, obtained using the approach described in Section 3.2 based on the first n historical curves. Our idea is to treat $\hat{\boldsymbol{\beta}}_{(n+1)}^{\text{TS}}$ as the prior for $\boldsymbol{\beta}_{(n+1)}$, and shrink the update towards it using the new information in \mathbf{y}_{n+1}^l .

Specifically, we propose to minimize the following *penalized least squares* (PLS) criterion with respect to $\boldsymbol{\beta}_{(n+1)} = (\beta_{n+1,1}, \dots, \beta_{n+1,K})^T$,

$$\sum_{j=1}^{m_0} \left[Y_{n+1,j} - \{\beta_{n+1,1}f_1(t_j) + \dots + \beta_{n+1,K}f_K(t_j)\} \right]^2 + \lambda \sum_{k=1}^K \left(\beta_{n+1,k} - \hat{\beta}_{n+1,k}^{\text{TS}} \right)^2, \quad (8)$$

where $Y_{n+1,j} = y_{n+1}(t_j)$ and $\lambda > 0$ is a penalty parameter whose determination will be discussed later. Note that the smooth factors $\{f_k(t)\}$ are already extracted from the historical curves. In matrix form, the PLS criterion can be expressed as

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

where \mathbf{F}^e is a $m_0 \times K$ matrix whose (j, k) -th entry is $f_k(t_j)$, $1 \leq j \leq m_0$, $1 \leq k \leq K$, and $\boldsymbol{\epsilon}_{n+1}^e = (\epsilon_{n+1,1}, \dots, \epsilon_{n+1,m_0})^T$.

Minimizing the criterion (8) leads to the PLS update of $\boldsymbol{\beta}_{(n+1)}$,

$$\hat{\boldsymbol{\beta}}_{(n+1)}^{\text{PLS}} = (\mathbf{F}^{eT} \mathbf{F}^e + \lambda \mathbf{I})^{-1} (\mathbf{F}^{eT} \mathbf{y}_{n+1}^e + \lambda \hat{\boldsymbol{\beta}}_{(n+1)}^{\text{TS}}). \quad (9)$$

The update of $x_{n+1}^l(t)$ based on $\hat{\boldsymbol{\beta}}_{(n+1)}^{\text{PLS}}$ is then given by

$$\hat{x}_{n+1}^{\text{PLS}}(t) = \hat{\beta}_{n+1,1}^{\text{PLS}} f_1(t) + \dots + \hat{\beta}_{n+1,K}^{\text{PLS}} f_K(t), \quad t > t_{m_0}. \quad (10)$$

For distributional updates of $x_{n+1}^l(t)$, we can first use the bootstrap procedure aforementioned in Section 3.2 to obtain B time series forecasts $\hat{\beta}_{(n+1)}^{\text{TS},b}$. Then, B PLS forecasts $\hat{\beta}_{(n+1)}^{\text{PLS},b}$ can be calculated according to (9), which leads to B PLS updates of $x_{n+1}^l(t)$ based on (10).

The PLS criterion involves two terms. The first term is a least squares goodness-of-fit term that measures how well the model prediction matches \mathbf{y}_{n+1}^e , the newly observed segment of the current curve. The second term penalizes a large departure of $\beta_{(n+1)}$ from $\hat{\beta}_{(n+1)}^{\text{TS}}$, the time series forecast. The $\beta_{(n+1)}$ that minimizes (8) is a compromise between the two terms carefully balanced by the size of the penalty parameter λ . To choose λ using the data, we propose to minimize some out-of-sample forecasting performance measures.

Below we give a detailed description of the method of selecting λ . Fix a training set. We use a latter portion of the training set as a hold-out sample. For each curve in this sample, we use a certain number of preceding curves in the training set to extract the smooth factors and their scores. Time series forecasts of the scores can then be derived. For a particular λ , we construct the PLS updating for the given curve. Calculate some forecasting performance measure, such as root mean squared error, for all the hold-out curves, and compute the average value. Select λ from a set of grid points that minimizes this average performance measure. The selected λ is then used to generate the PLS update at the updating point for all the curves in the forecasting set. Section 4.3 contains a detailed description of the procedure through an example. Note that the performance measure mentioned here needs to be computable using the discrete observation from $y_i(t)$. The underlying smooth component $x_i(t)$ is unobservable when analyzing real data.

Remark 3.3.1 There is a close connection between the criterion (8) and ridge

regression (Hoerl and Kennard, 1970a,b). Define

$$\tilde{\mathbf{y}}_{n+1}^e = \mathbf{y}_{n+1}^e - \mathbf{F}^e \hat{\boldsymbol{\beta}}_{(n+1)}^{\text{TS}}, \quad \text{and} \quad \tilde{\boldsymbol{\beta}}_{(n+1)} = \boldsymbol{\beta}_{(n+1)} - \hat{\boldsymbol{\beta}}_{(n+1)}^{\text{TS}}.$$

Then the PLS criterion is the same minimizing criterion for a ridge regression with $\tilde{\mathbf{y}}_{n+1}^e$ as the response, \mathbf{F}^e as the design matrix, and $\tilde{\boldsymbol{\beta}}_{(n+1)}$ as the regression coefficient vector. The ridge regression shrinks $\tilde{\boldsymbol{\beta}}_{(n+1)}$ towards zero; hence our PLS approach shrinks $\boldsymbol{\beta}_{(n+1)}$ towards the TS forecast $\hat{\boldsymbol{\beta}}_{(n+1)}^{\text{TS}}$.

Remark 3.3.2 The penalized criterion can also be established from a Bayesian updating perspective. To this end, assume that the entries of $\boldsymbol{\beta}_{(n+1)}$ as independent and homoscedastic random variables. Consider the following model of \mathbf{y}_{n+1}^e ,

$$\mathbf{y}_{n+1}^e = \mathbf{F}^e \boldsymbol{\beta}_{(n+1)} + \boldsymbol{\epsilon}_{n+1}^e,$$

where $\boldsymbol{\epsilon}_{n+1}^e \sim \mathcal{N}(0, \tilde{\sigma}^2 \mathbf{I}_{m_0})$. Based on \mathbf{Y} , we can obtain some distributional forecast of $\boldsymbol{\beta}_{(n+1)}$ denoted as $\mathcal{N}(\hat{\boldsymbol{\beta}}_{(n+1)}^{\text{TS}}, \sigma^2 \mathbf{I}_K)$, which we treat as the prior distribution of $\boldsymbol{\beta}_{(n+1)}$.

Multivariate normal theory then suggests the following distribution of \mathbf{y}_{n+1}^e and $\boldsymbol{\beta}_{(n+1)}$,

$$\begin{pmatrix} \mathbf{y}_{n+1}^e \\ \boldsymbol{\beta}_{(n+1)} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mathbf{F}^e \hat{\boldsymbol{\beta}}_{(n+1)}^{\text{TS}} \\ \hat{\boldsymbol{\beta}}_{(n+1)}^{\text{TS}} \end{pmatrix}, \begin{pmatrix} \sigma^2 \mathbf{F}^e \mathbf{F}^{eT} + \tilde{\sigma}^2 \mathbf{I}_{m_0} & \sigma^2 \mathbf{F}^e \\ \sigma^2 \mathbf{F}^{eT} & \sigma^2 \mathbf{I}_K \end{pmatrix} \right).$$

Then, the posterior distribution of $\boldsymbol{\beta}_{(n+1)}$ given \mathbf{y}_{n+1}^e is normal with the following mean and variance,

$$\begin{cases} \mu_{\boldsymbol{\beta}_{(n+1)} | \mathbf{y}_{n+1}^e} = \hat{\boldsymbol{\beta}}_{(n+1)}^{\text{TS}} + \mathbf{F}^{eT} (\mathbf{F}^e \mathbf{F}^{eT} + \lambda \mathbf{I}_{m_0})^{-1} (\mathbf{y}_{n+1}^e - \mathbf{F}^e \hat{\boldsymbol{\beta}}_{(n+1)}^{\text{TS}}), \\ \Sigma_{\boldsymbol{\beta}_{(n+1)} | \mathbf{y}_{n+1}^e} = \sigma^2 [\mathbf{I}_K - \mathbf{F}^{eT} (\mathbf{F}^e \mathbf{F}^{eT} + \lambda \mathbf{I}_{m_0})^{-1} \mathbf{F}^e], \end{cases}$$

where $\lambda = \tilde{\sigma}^2 / \sigma^2$. The posterior mean $\mu_{\boldsymbol{\beta}_{(n+1)} | \mathbf{y}_{n+1}^e}$ provides a point prediction update of $\boldsymbol{\beta}_{(n+1)}$. One can prove that it is equivalent to the PLS update (9). The posterior distribution of $\boldsymbol{\beta}_{(n+1)}$ naturally leads to a distributional update for $x_{n+1}^l(t)$.

4 Call Center Forecasting Study

In this section, we illustrate our forecasting methods using the call center data described in Section 2. Section 4.1 presents some results of fitting the smooth factor model (2) to the data. An out-of-sample rolling forecasting exercise is then reported, including both one-day-ahead forecasting (Section 4.2) and dynamic within-day updating (Section 4.3).

4.1 Smooth Factors

The first four (normalized) smooth factors are plotted in Figure 2, along with the corresponding GCV plots for selecting the smoothing penalty parameter ω . The grid points for ω are $\{10^{-3}, 10^{-2}, \dots, 10^2, 10^3\}$, and the optimal values chosen by GCV vary for different factors. To highlight the effect of smoothing, we also superimpose (in grey color) the unsmoothed factors (corresponding to $\omega = 0$) that are obtained by Shen and Huang (2008).

We decide to plot the first four smooth factors because of the forecasting performance reported below in Section 4.2, where the forecasting accuracy levels off when more than four factors are used. They together explain about 99.96% of the total variation, which suggests that they are sufficient to capture the major modes of variation.

The first factor summarizes the average within-day call arrival pattern, where the rate increases sharply between 7:00AM and 9:30AM, before it starts to decrease linearly until 4:00PM, when the drop becomes much steeper until the end of the day. One can also see a slight lunch-break dip around 12:30PM. Actually, 99.91% of the total variation could be explained by this factor, which indicates that the average arrival pattern is the dominant mode of variation. The remaining smooth factors

are second-order effects that are necessary to capture the difference among arrival patterns of different weekdays, which improves forecasting accuracy as shown in Section 4.2. In particular, they present certain contrasts between arrival patterns of different within-day segments. For example, the second factor compares arrival rates between mornings and afternoons/evenings; when combined with the corresponding scores, they capture the following specific phenomenon: Fridays usually have above average rates in early mornings and below average rates in late afternoons and evenings, while the opposite is true for Mondays.

The left panel of Figure 3 plots the score series of the first factor. Different colors and symbols indicate different weekdays, revealing a strong weekly effect. This score series turns out to be highly correlated with the total daily call volume (Figure 1). The right panel shows a scatter plot between the scores of the current day and the next weekday, which appear to be highly correlated, especially when conditioning on day-of-the-week. This motivates us to consider the following varying-coefficient AR(1) model for the first score series β_1 ,

$$\beta_{i1} = a_1(d_{i-1}) + b_1\beta_{i-1,1} + \eta_i,$$

where d_{i-1} denotes the day-of-the-week of day $i - 1$, and the varying intercept a_1 depends on d_{i-1} . We also considered a more complicated model that allows the slope to vary according to day-of-the-week. However, no significant improvement is obtained over the current constant slope model. More complex models were also fitted that incorporate polynomial trends and higher order auto correlations. Again, the model performance could not be significantly improved.

Further exploratory analysis suggests that similar varying intercept AR(1) models hold for the other score series as well. Hence, we opt to use such models to forecast the scores for future curves in Section 4.2, which are denoted as

$$\beta_{i,k} = a_k(d_{i-1}) + b_k\beta_{i-1,k} + \eta_{i,k}, \quad i = 2, \dots, n, \quad k = 1, \dots, 4. \quad (11)$$

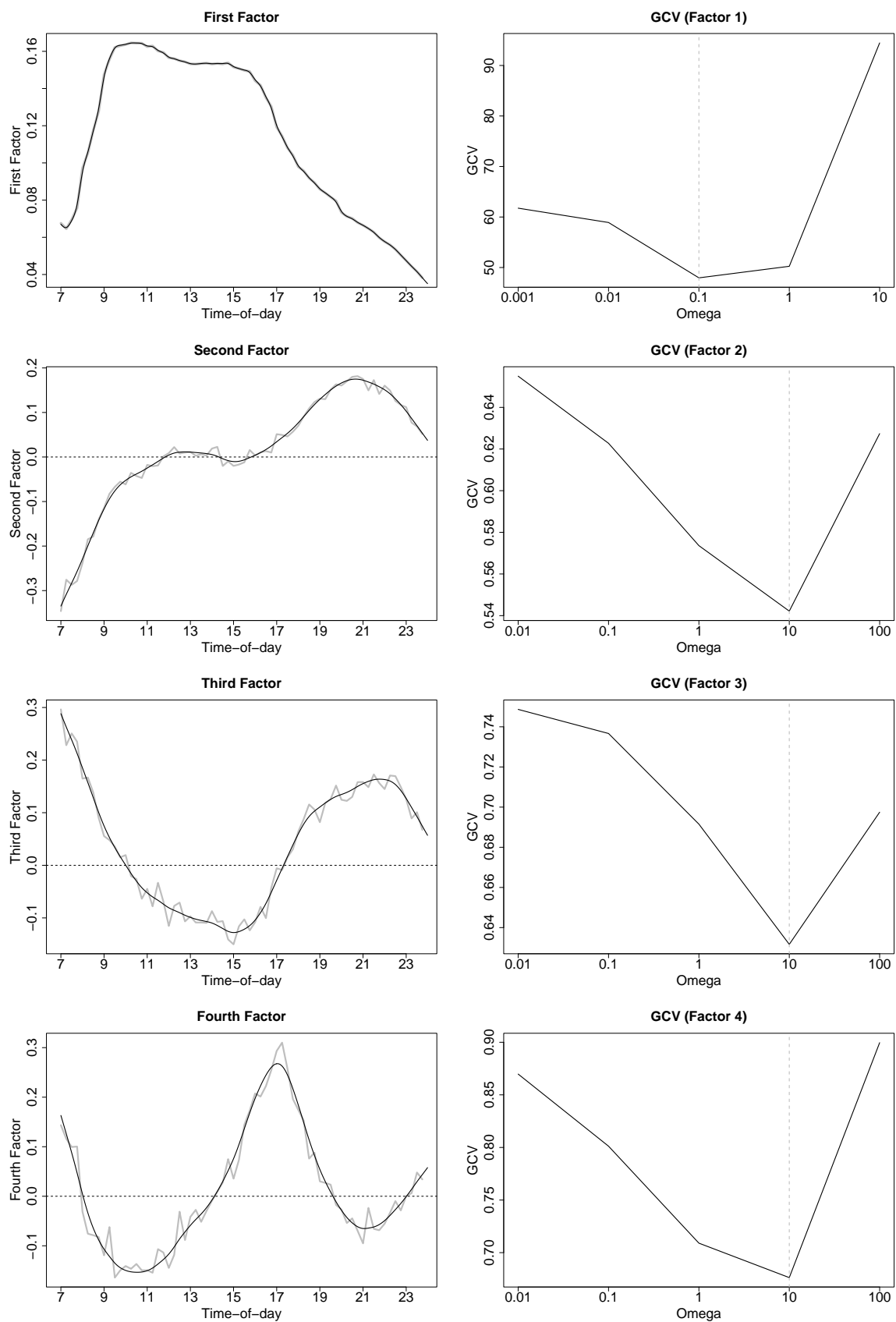


Figure 2: Plots of the First Four Smooth Factors and the Corresponding GCV Profiles.

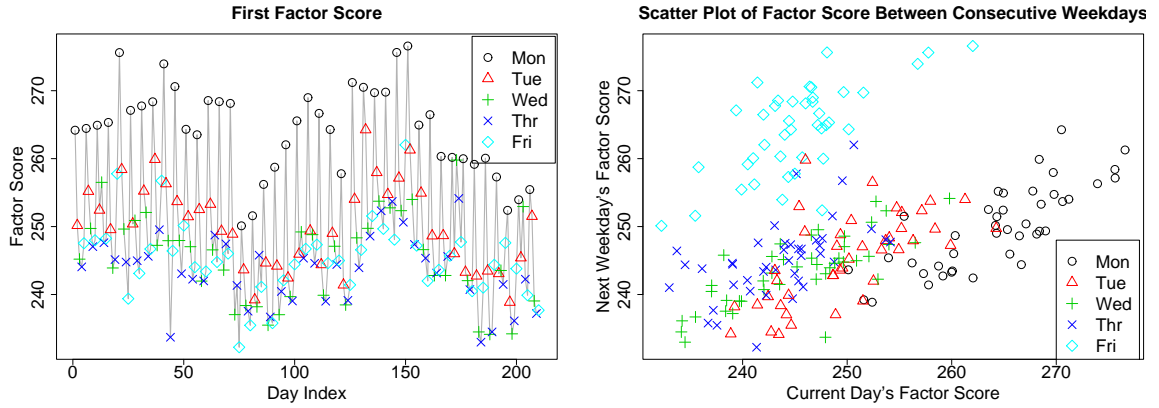


Figure 3: Plots of the First Factor Score Series. Weekly pattern and interday correlation are apparent.

4.2 One-day-ahead Forecasting

Below we use the extracted smooth factors and their scores to perform one-day-ahead forecasting. We are interested in forecasting the arrival rate profiles. However, they are unobservable; hence, the forecasting performance measures are calculated using the observed call volumes instead. The forecasts are first obtained on the square-root scale, and then back transformed to the original scale to calculate the performance measures.

The following rolling out-of-sample forecasting exercise is performed. For each of the 60 days between August 4 and October 24,

- consider the preceding 150 days as the training set;
- extract the first K smooth factors and their scores;
- perform a one-step-ahead forecast of the factor scores using the linear AR(1) model (11);
- compute the forecast root mean square error (RMSE) and average percent error

(APE) for day i as follows

$$\text{RMSE}_i = \sqrt{\frac{1}{m} \sum_{j=1}^m (N_{ij} - \widehat{N}_{ij})^2} \quad \text{and} \quad \text{APE}_i = \frac{100}{m} \sum_{j=1}^m \frac{|N_{ij} - \widehat{N}_{ij}|}{N_{ij}},$$

where $\widehat{N}_{ij} = (\widehat{x}_{ij})^2$.

In addition, we consider using historical averages (HA) of the same day-of-the-week as a forecast for the current day's *square-root-transformed* call volumes. This benchmark has been described as the industry standard (Weinberg et al., 2007; Shen and Huang, 2008).

Our factor-model based forecasting methods are more accurate than the HA approach. Table 1 gives summary statistics of the APE and RMSE of the forecasts from the HA approach, along with the ratios of the corresponding statistics between HA and our methods denoted as TS1 to TS4 depending on the number of smooth factors used. Increasing K leads to improvement in forecasting accuracy, while the improvement levels off when more than four factors are used; hence we choose to use TS4 as our final forecasting model because it reasonably balances forecasting accuracy and model complexity.

Note that the forecasting performance is similar to the one reported in Shen and Huang (2008). This is what one would expect, as the smoothing in our approach is performed on the underlying rate curves, while the forecasting errors here are calculated based on the observed call volumes that are not smooth. The advantage of the current approach over theirs is best seen in the simulation study of Section 5.1, where we know the underlying smooth rate curves, and can directly evaluate the performance of forecasting them.

Table 1: Summary Statistics of APE and RMSE for Call Volumes. The rolling forecasting exercise involves 60 days. Our approaches improve significantly over HA.

	APE (%)				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	1.02	0.97	1.08	1.07	1.13	0.69	0.83	0.64
TS2/HA	0.98	0.86	0.94	0.87	1.01	0.61	0.77	0.57
TS3/HA	0.96	0.81	0.89	0.79	1.06	0.59	0.76	0.50
TS4/HA	0.92	0.81	0.88	0.77	0.95	0.57	0.76	0.50

4.3 Dynamic Updating

Below we consider three forecasting approaches: TS4 where no updating is performed, PLS4 with an updating at 10:00AM, and PLS4 with an updating at 12:00PM. Table 2 reports a comparison of RMSE based on the call volumes. The measures are calculated only using data after 12:00PM. The forecasting gets increasingly more accurate as more observations are used for the updating.

In addition to reducing forecast errors, within-day updating also narrows the width of forecast intervals. Table 3 compares the distribution of Average Width of 95% forecast intervals for the *rate* profiles. The forecast intervals are stochastically narrower, with the intervals from the 12:00PM-updating being the shortest.

For the three approaches considered above, Figure 4 provides some graphical illustration of the forecast distributions of the arrival rates on September 3. The left panel plots the actual call volume profile (the solid circles) along with the rate profile forecasts after 12:00PM and their 95% prediction bands. The forecasts are shifted upwards dramatically after incorporating the same-day arrival information. Although

Table 2: Summary Statistics of RMSE for Call Volumes of Three Approaches: TS4 (i.e. no updating), PLS4 (10:00AM updating) and PLS4 (12:00PM updating). Updating reduces forecasting errors.

	RMSE		
	TS4	PLS4	PLS4
		10:00AM	12:00PM
Q1	35.53	33.76	31.81
Median	42.78	40.32	35.80
Mean	51.81	45.95	43.03
Q3	59.64	47.80	41.91

Table 3: Summary Statistics of Average Width of 95% Forecast Intervals for Arrival Rates of Three Approaches: TS4 (i.e. no updating), PLS4 (10:00AM updating) and PLS4 (12:00PM updating). Updating narrows the forecast intervals.

	Average Interval Width		
	TS4	PLS4	PLS4
		10:00AM	12:00PM
Q1	121.4	94.0	76.7
Median	125.9	102.8	81.6
Mean	125.9	102.8	81.7
Q3	130.7	108.1	85.7

September 3 is two days after Labor Day, the call arrival rates are still unusually high, which are under-predicted by TS4. The prediction band also narrows, which is more obvious from the right panel that shows the forecasted densities of the arrival rate at 1:00PM. The actual volume (the grey vertical line) almost falls outside the density range given by TS4. On the other hand, the density forecast from the 12:00PM updating not only centers around the actual volume, but also is more concentrated.

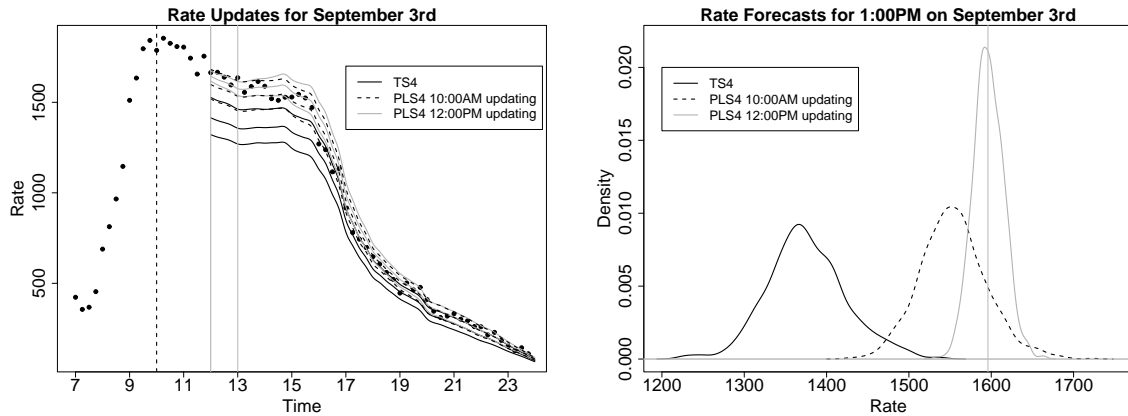


Figure 4: Comparison of the Updated Forecasts for Wednesday, September 3, 2003. Within-day updating improves both forecasting accuracy and precision.

In the aforementioned analysis, the penalty parameter λ for PLS4 is selected using the idea described in Section 3.3. We use the first 150 days of the data as the training set, and its last 50 days as the hold-out sample. For one updating point, and for each given λ , we construct the PLS4 updating for every day in this hold-out sample using its preceding 100 days. Then, we calculate the RMSE over the time periods after the updating time for each day. The λ that minimizes the average RMSE over the hold-out sample is then selected and used for all the days in the forecasting set. The candidate values for λ are $\{0, 10, \dots, 10^9\}$; and for both updating points, the procedure chooses $\lambda = 10^3$ as the “optimal” value.

5 Simulation Studies

5.1 Smoothing matters for forecasting smooth curves

In the case study of Section 4, we can not calculate the forecast errors of the underlying rate profiles because we do not observe them. Similarly, in practice, one won't observe the underlying smooth curves, which makes it impossible to calculate the forecast errors. Hence, we design the following simulation study to illustrate the benefit of taking into account the smooth nature of the curves during forecasting.

We consider the following simulation model,

$$\begin{cases} Y_{ij} = y_i(t_j) = x_i(t_j) + \epsilon_i(t_j) = X_{ij} + \epsilon_{ij}, & \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2), \\ x_i(t_j) = z_i h_{d_i}(t_j), & h_{d_i}(t) \text{ is smooth, } \sum_j h_{d_i}(t_j) = 1, \\ z_i - a_{d_i} = b(z_{i-1} - a_{d_{i-1}}) + \eta_i, & \eta_i \sim \mathcal{N}(0, \phi^2), \\ i = 1, \dots, n; & j = 1, \dots, m; \quad d_i = 1, 2, 3, 4, 5. \end{cases} \quad (12)$$

This simulation study is motivated by the call center case study. The same model has been proposed to analyze call center arrival data by Weinberg et al. (2007) under a Bayesian framework. One difference is that here we focus on forecasting the underlying smooth curve $x_i(t)$. Essentially, this model assumes $x_i(t)$ follows a multiplicative time series model, where the daily scale factor (z_i) satisfies an AR(1) model after adjusting the day-of-the-week effect (a_{d_i}), and each weekday (d_i) has its own smooth intraday arrival rate profile ($h_{d_i}(t)$).

We fit the model (12) to the call center data to suggest reasonable model parameters, some of which are given below:

$$\begin{cases} a_1 = 2045, a_2 = 1935, a_3 = 1900, a_4 = 1890, a_5 = 1885, \\ b = 0.65, \sigma = 0.5, \phi = 35, n = 210, m = 68. \end{cases}$$

The smooth intraday profiles $g_{d_i}(t)$ are fitted using cubic smoothing splines on the average within-day arrival proportion profiles for different day-of-the-week respectively.

To create scenarios with various signal-to-noise ratios, we also consider two additional cases where the error standard deviation σ is set as 1 and 1.5, respectively. For each simulation setup, 100 data sets are generated. We perform the same rolling forecasting exercise as the one in the case study of Section 4.2.

To illustrate the benefit of incorporating the smoothness constraint, we compare our approach with the SVD-based approach of Shen and Huang (2008). The discussion in Section 3.1 suggests that the SVD approach is a special case of our proposal that sets the smoothing parameter ω as zero. Shen and Huang (2008) are interested in forecasting future *discrete call volumes* instead of *smooth call rates*. For our smooth-factor-model (SFM) approach, the GCV criterion (Section 3.1) is used to select the smoothing parameters needed for each forecasting curve. As suggested by the case study of Section 4.2, we also extract $K = 4$ smooth factors in this study.

For each simulated dataset, we calculate the Mean APE and Mean RMSE of the underlying rate profiles. The forecast performance measures are defined as follows,

$$\text{APE}_i = \frac{100}{m} \sum_{j=1}^m \frac{|X_{ij}^2 - \widehat{X}_{ij}^2|}{X_{ij}^2} \quad \text{and} \quad \text{RMSE}_i = \sqrt{\frac{1}{m} \sum_{j=1}^m (X_{ij}^2 - \widehat{X}_{ij}^2)^2},$$

where $X_{ij} \equiv x_i(t_j)$ is the truth and \widehat{X}_{ij} is its forecast. Note that the model (12) is on the square-root scale, and the forecasts are back transformed to the original rate scale when calculating the performance measures.

Figure 5 then plots the empirical cumulative distribution functions (CDF) and the boxplots of the summary statistics of the performance measures over the 100 simulations. The advantage of smoothing in terms of forecasting smooth curves is clearly illustrated. In all three scenarios considered, incorporating smoothness leads to more accurate forecasts; and the improvement increases as the data become more

noisy. The CDF plots suggest that the smooth-factor approach leads to stochastically smaller performance measures.

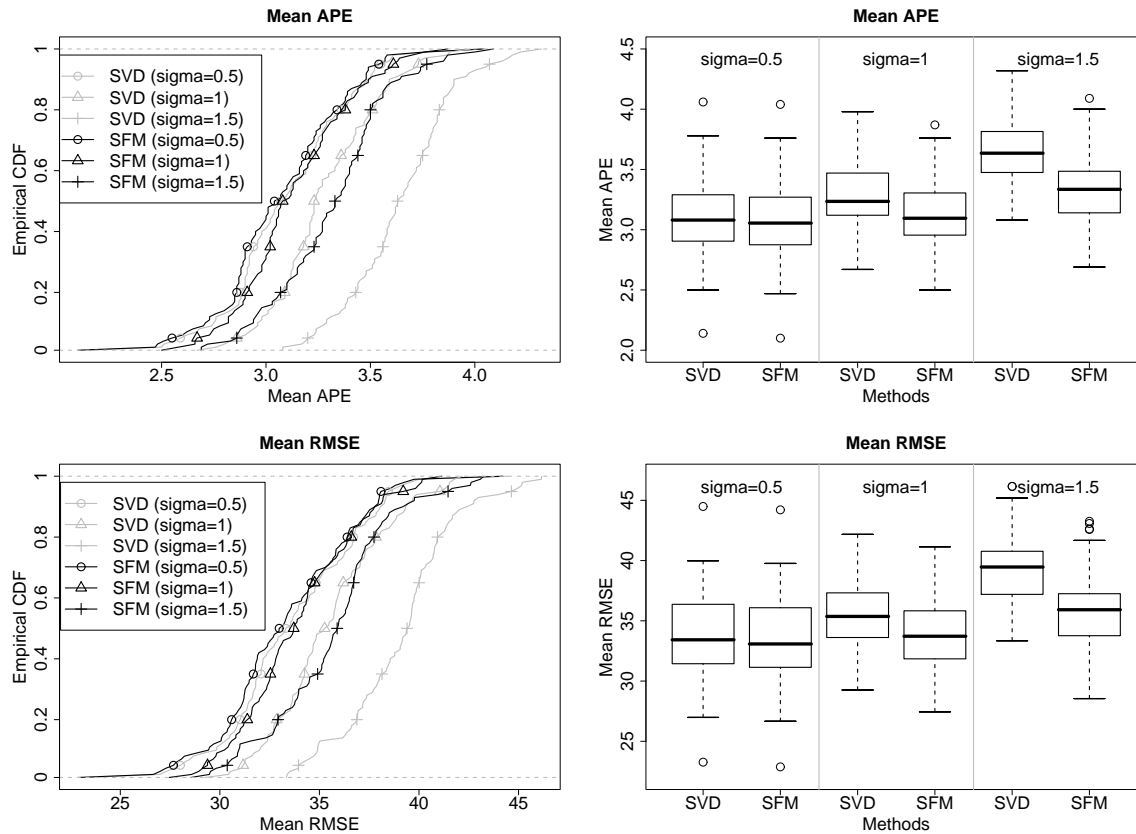


Figure 5: Comparison of Forecasting Performance between SVD and SFM. SFM leads to stochastically smaller forecasting performance measures.

5.2 Multiple smoothing parameters do matter

One advantage of our approach is that the smooth factors are estimated sequentially; hence different smoothing parameters can be used for different factors. Below we present one simulation study to illustrate the necessity and benefit of allowing multiple smoothing parameters.

More specifically, we consider the following simulation model with two smooth

factors,

$$\left\{ \begin{array}{l} Y_{ij} = y_i(t_j) = x_i(t_j) + \epsilon_i(t_j) = X_{ij} + \epsilon_{ij}, \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2), \\ x_i(t_j) = \beta_{i1}f_1(t_j) + \beta_{i2}f_2(t_j), \quad f_1(t) \text{ and } f_2(t) \text{ are smooth,} \\ \beta_{1i} - a_{1d_i} = b_1(\beta_{1(i-1)} - a_{1d_{i-1}}) + \eta_{1i}, \quad \eta_{1i} \sim \mathcal{N}(0, \sigma_1^2), \\ \beta_{2i} - a_{2d_i} = b_2(\beta_{2(i-1)} - a_{2d_{i-1}}) + \eta_{2i}, \quad \eta_{2i} \sim \mathcal{N}(0, \sigma_2^2), \\ i = 1, \dots, n; \quad j = 1, \dots, m; \quad d_i = 1, 2, 3, 4, 5. \end{array} \right. \quad (13)$$

To make the simulation study practically relevant, the above model uses the first two smooth factors extracted from the call center data in Section 4.1 as $f_1(t)$ and $f_2(t)$. We then fit the above time series models to their corresponding score series to estimate the other model parameters. Below we list the parameters (appropriately rounded) used in Model (13):

$$\left\{ \begin{array}{l} \sigma = 0.77, \quad n = 210, \quad m = 68, \\ a_{11} = 264.29, \quad a_{12} = 249.91, \quad a_{13} = 245.29, \quad a_{14} = 243.55, \quad a_{15} = 245.17, \quad b_1 = 0.69, \quad \sigma_1 = 4.35, \\ a_{21} = 4.37, \quad a_{22} = -0.27, \quad a_{23} = 0.78, \quad a_{24} = 0.81, \quad a_{25} = -5.99, \quad b_2 = 0.41, \quad \sigma_2 = 2.12. \end{array} \right.$$

We then simulate 100 data sets using the above model, and apply our method to estimate the first two smooth factors. The GCV criterion is used to select the “optimal” smoothing parameter ω from a candidate set of $\{10^{-3}, 10^{-2}, \dots, 10^2, 10^3\}$. The selected parameters are reported below in Table 4, where SP means that the parameters are restricted to be the same for the two factors, while MP means they are allowed to be different. In addition, we calculate the mean squared errors (MSE) over the grid points for each estimated smooth factor. Table 5 reports some summary statistics of the MSE ratios between SP and MP. The results suggest that different parameters are necessary for the two smooth factors; and more smoothing is needed for the second factor. The two methods give almost identical MSE summaries for $f_1(t)$ because similar smoothing parameters are chosen. However, for $f_2(t)$, the MSEs

of the SP method are much larger than the MP method with the mean ratio being 5.27. The reason is that SP selects smoothing parameters that are too small for the second factor.

Table 4: Comparison of Smoothing Parameter Selection for SP and MP. Different parameters are necessary for the two smooth factors.

	$\log_{10}(\omega)$	-3	-2	-1	0	1	2	3
SP	$f_1(t)/f_2(t)$	1	71	28	0	0	0	0
MP	$f_1(t)$	2	87	11	0	0	0	0
MP	$f_2(t)$	0	0	1	2	79	18	0

Table 5: Comparison of Summary Statistics of the MSE Ratios between SP and MP.

	$f_1(t)$				$f_2(t)$			
	Q1	Median	Mean	Q3	Q1	Median	Mean	Q3
SP/MP	1.00	1.00	1.21	1.00	3.30	4.45	5.27	6.14

Acknowledgment

Haipeng Shen’s work is partially supported by UNC Junior Faculty Development Award, and National Science Foundation (NSF) grants DMS-0606577 and CMMI-0800575.

References

AFII (2004), “A high potential for growth,” *available at*

http://www.investinfrance.org/France/KeySectors/Operations/?p=call_centers&l=en.

- Antoniadis, A., Paparoditis, E., and Sapatinas, T. (2006), “A functional wavelet-kernel approach for time series prediction,” *Journal of the Royal Statistical Society, Series B*, 837–857.
- Besse, P. C., Cardot, H., and Stephenson, D. B. (2000), “Autoregressive forecasting of some functional climatic variations,” *Scandinavian Journal of Statistics*, 27, 673–687.
- Brown, L. D., Cai, T., Zhang, R., Zhao, L., and Zhou, H. (2007), “The root-unroot algorithm for density estimation as implemented via wavelet block thresholding,” *Technical Report*.
- Brown, L. D., Gans, N., Mandelbaum, A., Sakov, A., Shen, H., Zeltyn, S., and Zhao, L. (2005), “Statistical analysis of a telephone call center: A queueing-science perspective,” *Journal of the American Statistical Association*, 100, 36–50.
- Eckart, C. and Young, G. (1936), “The approximation of one matrix by another of lower rank,” *Psychometrika*, 1, 211–218.
- Erbas, B., Hyndman, R. J., and Gertig, D. M. (2007), “Forecasting age-specific breast cancer mortality using functional data models,” *Statistics in Medicine*, 26, 458–470.
- Gans, N., Koole, G. M., and Mandelbaum, A. (2003), “Telephone call centers: Tutorial, review, and research prospects,” *Manufacturing & Service Operations Management*, 5, 79–141.
- Gilson, K. A. and Khandelwal, D. K. (2005), “Getting more from call centers,” *The McKinsey Quarterly*.
- Green, P. J. and Silverman, B. W. (1994), *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, Chapman and Hall.

- Hoerl, A. E. and Kennard, R. W. (1970a), “Ridge regression: Biased estimation for nonorthogonal problems,” *Technometrics*, 12, 55–67.
- (1970b), “Ridge regression: Application to nonorthogonal problems,” *Technometrics*, 12, 69–82.
- Huang, J. Z., Shen, H., and Buja, A. (2008), “Functional principal component analysis via penalized rank one approximation,” *Electronic Journal of Statistics*, 2, 678–695.
- Hyndman, R. J. and Ullah, M. S. (2007), “Robust forecasting of mortality and fertility rates: A functional data approach,” *Computational Statistics & Data Analysis*, 51, 4942–4956.
- Ramsay, J. O. and Silverman, B. W. (2002), *Applied Functional Data Analysis*, Springer-Verlag: New York.
- Shen, H. and Huang, J. Z. (2005), “Analysis of call centre arrival data using singular value decomposition,” *Applied Stochastic Models in Business and Industry*, 21, 251–263.
- (2008), “Interday forecasting and intraday updating of call center arrivals,” *Manufacturing & Service Operations Management*, 10, 391–410.
- Smith, M. (2000), “Modeling and short-term forecasting of New South Wales electricity system load,” *Journal of Business and Economic Statistics*, 18, 465–478.
- Valderrama, M. J., Ocana, F. A., and Aguilera, A. M. (2002), “Forecasting PC-ARIMA models for functional data,” in *Proceedings of Computational Statistics*, pp. 25–36.
- Weinberg, J., Brown, L. D., and Stroud, J. R. (2007), “Bayesian forecasting of an inhomogeneous Poisson process with applications to call center Data,” *Journal of the American Statistical Association*, 102, 1185–1199.

Appendix

A1: Computation of the Optimizing Spline of (5)

In the Appendix, we summarize some standard results for spline smoothing that are needed for implementing our methods. See Green and Silverman (1994) for more background materials on the subject.

We first give definition of the penalty matrix used in (6). Define two banded matrices Q and R as follows. Let $h_j = t_{j+1} - t_j$ for $j = 1, \dots, m-1$. Let Q be the $m \times (m-2)$ matrix with entries q_{jk} , for $j = 1, \dots, m$ and $k = 2, \dots, m-1$, given by

$$q_{k-1,k} = h_{k-1}^{-1}, \quad q_{kk} = -h_{k-1}^{-1} - h_k^{-1}, \quad q_{k+1,k} = h_k^{-1}, \quad q_{jk} = 0 \text{ for } |j-k| > 2.$$

To simplify the presentation, the columns of Q are numbered in a non-standard way, starting at $k = 2$, so that the top left element of Q is q_{12} . The symmetric matrix R is $(m-2) \times (m-2)$ with elements r_{jk} , for j and k running from 2 to $(m-1)$, given by

$$\begin{cases} r_{jj} = \frac{1}{3}(h_{j-1} + h_j) \text{ for } j = 2, \dots, m-1, \\ r_{j,j+1} = r_{j+1,j} = \frac{1}{6}h_j \text{ for } j = 2, \dots, m-2, \end{cases}$$

and $r_{jk} = 0$ for $|j-k| > 2$. The matrix R is strictly diagonal dominant and thus is strictly positive-definite. The penalty matrix $\mathbf{\Omega} = QR^{-1}Q^T$.

Suppose $\hat{\mathbf{f}}$ minimizes the criterion (6). The minimizer $\hat{f}(\cdot)$ of (5) is a natural cubic spline with knots at t_j that interpolates $\hat{\mathbf{f}}$. Define $\mathbf{g} = R^{-1}Q^T\hat{\mathbf{f}} \equiv (g_2, \dots, g_{m-1})^T$ and $g_0 = g_m = 0$. The natural cubic spline $\hat{f}(\cdot)$ that minimizing (5) is completely determined by $\hat{\mathbf{f}}$ and \mathbf{g} as follows,

$$\begin{aligned} \hat{f}(t) = & \frac{(t-t_j)\hat{f}_{j+1} + (t_{j+1}-t)\hat{f}_j}{h_j} \\ & - \frac{1}{6}(t-t_j)(t_{j+1}-t) \left\{ \left(1 + \frac{t-t_j}{h_j}\right)g_{j+1} + \left(1 + \frac{t_{j+1}-t}{h_j}\right)g_j \right\} \end{aligned}$$

for $t_j \leq t \leq t_{j+1}$, $1 \leq j \leq m - 1$; the spline outside $[t_1, t_m]$ is obtained by linear extrapolation.

A2: Efficient Computation of the GCV Criterion (7)

In practice, one would calculate the GCV criterion for multiple candidate values of ω , and choose the value that minimizes the selection criterion. The bottleneck of such an operation is the evaluation of the smoothing matrix $\mathbf{S}(\omega) = (\mathbf{I} + \omega \mathbf{\Omega})^{-1}$.

The evaluation can be made efficient via eigen decomposition of the symmetric and positive definite penalty matrix $\mathbf{\Omega}$. The eigen decomposition of $\mathbf{\Omega}$ suggests that

$$\mathbf{\Omega} = \mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Gamma}^T,$$

where $\mathbf{\Gamma}$ is an orthogonal matrix containing the eigen vectors, and $\mathbf{\Lambda}$ is a diagonal matrix of the eigen values. Hence, the eigen decomposition of $\mathbf{S}(\omega)$ follows easily as,

$$\mathbf{S}(\omega) = \mathbf{\Gamma} (\mathbf{I} + \omega \mathbf{\Lambda})^{-1} \mathbf{\Gamma}^T.$$

The above derivation suggests that, once one calculates the eigen decomposition of $\mathbf{\Omega}$, it is very easy to evaluate $\mathbf{S}(\omega)$ for any arbitrary ω .