Statistica Sinica **34** (2024), 1-26
doi:https://doi.org/10.5705/ss.202021.0223

FACTOR-AUGMENTED MODEL FOR FUNCTIONAL DATA

Yuan Gao[1], Han Lin Shang[2] and Yanrong Yang*[1]

[1]*Australian National University and* [2]*Macquarie University*

Abstract: We propose modeling raw functional data as a mixture of a smooth function and a high-dimensional factor component. The conventional approach to retrieving a smooth function from raw data is to use a smoothing technique. However, the smoothing model is unable to recover the smooth curve or capture the data variation in some situations, for example, when there is a large measurement error, the smoothing basis functions are incorrectly identified, or the step jumps in the functional mean levels are neglected. We propose a factor-augmented smoothing model to address these challenges, and implement an iterative numerical estimation approach. Including the factor model component in the proposed method solves the aforementioned problems because a few common factors often drive the variation that cannot be captured by the smoothing model. We also establish asymptotic theorems to demonstrate the effects of including factor structures on the smoothing results. Specifically, we show that the smoothing coefficients projected on the complement space of the factor loading matrix are asymptotically normal. Of independent interest, we present an estimator for the population covariance matrix of the raw data, based on the proposed model. Extensive simulation studies show that these factor adjustments are essential to improving the estimation accuracy and avoiding the curse of dimensionality. Lastly, we demonstrate the performance of our model by applying it to Australian temperature data.

Key words and phrases: Basis function misspecification, functional data smoothing, high-dimensional factor model, measurement error, statistical inference on covariance estimation.

1. Introduction

Functional data analysis (FDA) has become increasingly popular as data storage technology has improved. Functional data are realizations of smooth random objects using curves, images, and shapes. Ramsay and Silverman (2002, 2005) and Ramsay and Hooker (2017) discuss the methodology and applications of FDA; see also Ferraty and Vieu (2006), Horváth and Kokoszka (2012) Cuevas (2014), Febrero-Bande, Galeano and González-Manteiga (2017), Goia and Vieu (2016), Reiss et al. (2017) and Wang, Chiou and Müller (2016). One of the main challenges in an FDA is that we cannot observe functional curves directly, but only discrete points, which are often contaminated by measurement errors. To

*Corresponding author.

model a mixture of functional data and high-dimensional measurement errors, we introduce a factor-augmented smoothing model (FASM).

We denote a random sample of n functional data as $\mathcal{X}_i(u)$, for $i = 1, \ldots, n$, and $u \in \mathcal{I} \subset \mathbb{R}$, where $\mathcal{I}$ is a compact interval on the real line $\mathbb{R}$. In practice, the observed data are discrete points and are often contaminated by noise or measurement errors. We use Y_{ij} to represent the jth observation on the ith subject. The observed data can then be expressed as a "signal plus noise" model:

$$Y_{ij} = \mathcal{X}_i(u_j) + \eta_{ij}, \quad i = 1, \ldots, n, \ j = 1, \ldots, p.$$

We use $\mathcal{X}_i(u_j)$ to denote the realization of the jth discrete point on the curve $\mathcal{X}_i(\cdot)$, and η_{ij} is the noise or measurement error. We assume that a measurement error occurs only where the measurements are taken; thus, the error $\boldsymbol{\eta}_i = (\eta_{i1}, \ldots, \eta_{ip})$ is a multivariate term of dimension p. However, in practice, although the signal function component $\boldsymbol{\mathcal{X}}_i = \{\mathcal{X}_i(u_1), \ldots, \mathcal{X}_i(u_j)\}$ is of the same dimension p, it differs from $\boldsymbol{\eta}_i$ in nature. Although functions are potentially infinite-dimensional, we may impose smoothing assumptions on them, which usually implies that they possess one or more derivatives. This smoothness feature is used to separate the functions from the measurement errors, that is, a functional smoothing procedure.

When the variance of the noise level is a tiny fraction of the variance of the function, we say the signal-to-noise ratio is high. In this case, classic smoothing tools can be applied to the functional data, including kernel methods (e.g., Wand and Jones (1995)), local polynomial smoothing (e.g., Fan and Gijbels (1996)), and spline smoothing (e.g., Wahba (1990); Eubank (1999); Green and Silverman (1999)). Once we have pre-smoothed functions, we can obtain estimates such as the mean and covariance functions. More recent studies on functional smoothing approaches include those of Cai and Yuan (2011), Yao and Li (2013), and Zhang and Wang (2016). In this study, we apply basis smoothing to the functions $\mathcal{X}_i(u)$; that is, we represent $\mathcal{X}_i(u)$ as $\mathcal{X}_i(u) = \sum_{k=1}^{K} c_{ik}\phi_k(u)$, where $\{\phi_k(u), \ k = 1, \ldots, K\}$ are the basis functions and $\{c_{ik}, \ i = 1, \ldots, n, \ k = 1, \ldots, K\}$ are the smoothing coefficients. The smoothing model then becomes

$$Y_{ij} = \sum_{k=1}^{K} c_{ik}\phi_k(u_j) + \eta_{ij}, \quad i = 1, \ldots, n, \ j = 1, \ldots, p.$$

When the signal-to-noise level is low, smoothing tools may not adequately remove the measurement error, resulting in an inefficient estimation of the smoothing coefficients. Here we examine further the measurement error η_{ij}. In an FDA, the number of discrete points p on each subject is often large compared with the sample size n. Hence, the term $\boldsymbol{\eta}_i$ is a high-dimensional component. In this case, the observed data are a mixture of functional data and high-dimensional data. The existence of the large measurement error η_{ij} causes the curse of dimensionality problem, requiring that we apply dimension-reduction models to

η_{ij}. Here, factor models are widely used as a dimension-reduction technique for high-dimensional data; among these, factor models are widely used (e.g., Fan, Fan and Lv (2008); Lam, Yao and Bathia (2011)).

We propose using a factor model for the measurement error term. Without further information on the measurement error, the factor model is appropriate because the estimation of the latent factors does not require any observed variables. The high-dimensional measurement error is assumed to be driven by a small number of unobserved common factors:

$$\eta_{ij} = \boldsymbol{a}_j^\top \boldsymbol{f}_i + \epsilon_{ij}, \quad i = 1, \ldots, n,\ j = 1, \ldots, p,$$

where $\boldsymbol{f}_i \in \mathbb{R}^r$ are the unobserved factors, $\boldsymbol{a}_j \in \mathbb{R}^r$ are the unobserved factor loadings, r is the number of latent factors, and ϵ_{ij} are idiosyncratic errors with mean zero. Thus, the observed data Y_{ij} can be written as the sum of two components:

$$Y_{ij} = \sum_{k=1}^{K} c_{ik}\phi_k(u_j) + \boldsymbol{a}_j^\top \boldsymbol{f}_i + \epsilon_{ij}, \quad i = 1, \ldots, n,\ j = 1, \ldots, p.$$

This is a basis smoothing model with a factor-augmented form. This proposed model can be modified easily for nonparametric smoothing methods. In Section S1, we illustrate the use of spline smoothing approaches. In Section S3.3, we apply the nonparametric smoothing model to simulated data.

The following three cases motivate the proposed FASM. In each case, using the proposed model remedies the defects of the traditional smoothing model. Examples of the three cases are provided in Section 2:

1. In traditional smoothing models, the measurement error η_{ij} is assumed to be non-informative and independently and identically distributed (i.i.d.) in both directions. This is an unrealistic assumption when the measurement errors contain information. By applying the factor model, we assume that a small number of unobserved factors capture the covariance in the measurement error. This is usually reasonable in practice, because a few common factors are often the cause of systematic measurement errors.

2. When the smoothing basis functions are incorrectly identified, the smoothing model leads to an erroneous coefficient estimate and large residuals. The proposed model deals with this problem, because the unexplained variation from the misidentification can be modeled using a small number of unobserved common factors.

3. When there are step jumps in the mean levels of the functions, neglecting these mean shifts in a smoothing model results in large residuals at the points where the jumps occur. The changes in the mean levels of the

functions come from a universal source, and can be modeled by common factors.

Because the latent factors are unobserved, we propose an iterative approach to simultaneously estimate the smooth function and the factors. We use a principal component analysis (PCA) to estimate the factor model, and apply a penalized least squares estimation to construct the estimator for the smoothing coefficient c_{ik}. We establish the asymptotic theory of the smoothing coefficient estimator, and prove the consistency of the estimator. We also provide the asymptotic distribution of the projected estimator in the orthogonal complement of the space spanned by the factors $\boldsymbol{f}_i$. Here, we see the interplay between the smooth component and the factor model component.

The remainder of this paper is structured as follows. In Section 2, we elaborate on the previously mentioned three motivations, and provide examples. In Section 3, we formally state the model, and provide an iterative estimation approach. We discuss the asymptotic properties of the smoothing coefficients under various assumptions in Section 4. In Section 5, we conduct Monte Carlo simulations on the proposed model under different settings. A real-data example is given in Section 6, and Section 7 concludes the paper. In the Supplementary Material, we provide several important model extensions and data analysis, as well as proofs for the theorems.

2. Motivation

We introduce three examples to motivate the proposed model. In each case, the smoothing model is inadequate for capturing the signal information of the raw data. In the first example, a large measurement error exists, and the residuals after smoothing are large with extreme values. In the second example, the basis functions are selected incorrectly, and part of the functions' variation cannot be captured by the smoothing model. In the third example, there are step jumps in the functional data, and the residuals after smoothing contain gaps. These examples demonstrate that further modeling of the smoothing residuals is needed.

2.1. Functional data with measurement errors

Figure 1 shows rainbow plots of the average daily temperature and log precipitation at 35 locations in Canada. Because of the nature of the two kinds of data, it is reasonable to assume that temperature and log precipitation are functions over time. However, the two graphs display distinct features. Although there are some perturbations in the temperature plot, it is relatively easy to discern each curve's shape. However, in the precipitation plot, there is a tremendous amount of variability in the raw data, such that it is almost impossible to observe the underlying shape of the curves.

Smooth temperature data can be retrieved without much difficulty using

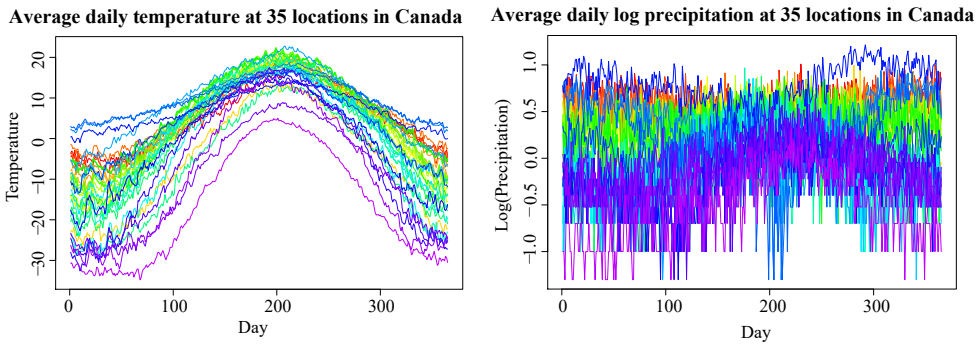


Figure 1. Average daily temperature and log precipitation from 35 Canadian weather stations, averaged over the period 1960 to 1994.

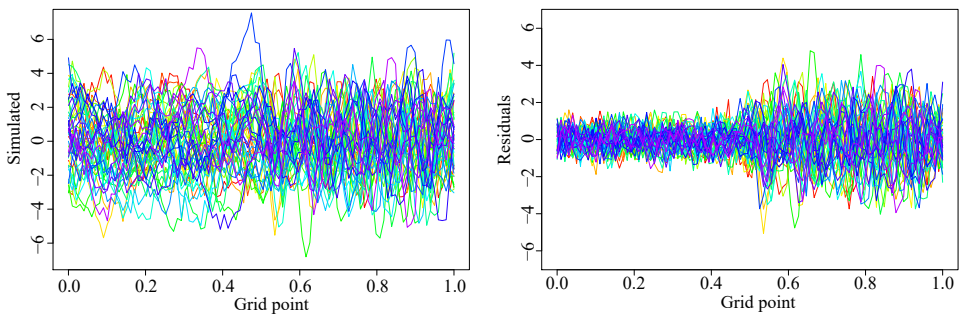
basic smoothing techniques. The residuals are small, with constant variation. On the other hand, the residuals after smoothing exhibit a high level of variation for the precipitation data, and even contain some extreme values. Our model endeavors to explain the large residuals in similar cases to the precipitation data; we show the fitting result in Section S4.

2.2. Misidentification of the basis function

It is important to choose appropriate basis functions for the smoothing method. In this example, we show the inadequacy of the smoothing model when the basis functions are misidentified. We generate functional data using basis functions with changing frequencies. The raw data are shown in Figure 2a. Fourier basis functions are used. In the second half of the data, the frequency of the Fourier basis functions increases, so the data set exhibits more variation toward the right end. Suppose we are unaware of the change in the frequencies of the basis functions, and so used the basis of the first half of the data for the complete curve. The result is shown in Figure 2b. The residuals are large in the second half of the data, but the smoothing model fails to reduce the residuals. Here, a factor model can be used to further model the signal hidden in the large residuals. The data-generating process and further analysis are provided in Section S3.4.

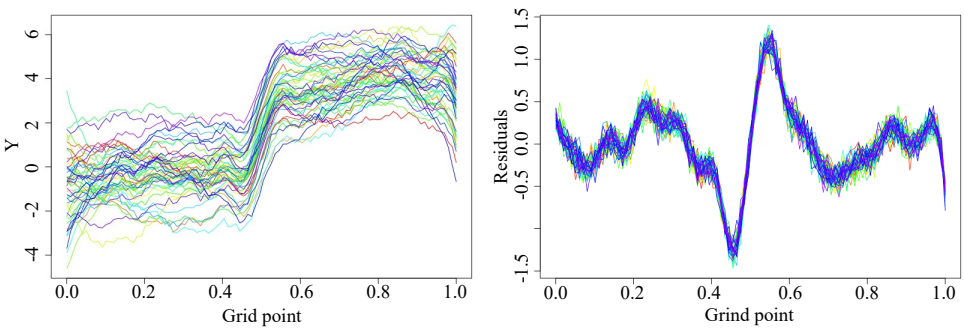
2.3. Functional data with step jumps in the mean level

Suppose we observe a sample of raw functional data, shown in Figure 3a, where there is a jump at around $u = 0.5$. The jump applies to all sample data, so this sudden shift is at the mean level. We explain how the data are generated in Section S3.5. The residuals after smoothing are presented in Figure 3b. The large residuals around the jump indicate that without measures to deal with the step jumps, smoothing itself is not enough to model these kinds of data. We show in Section S3.5 that applying the proposed model to the same data generates



(a) Raw data (b) Residuals

Figure 2. A simulated sample of functional data with changing basis functions.



(a) Raw data (b) Residuals

Figure 3. A simulated sample of functional data with step jump.

smaller residuals and has less flexibility, which is one of our main goals.

3. Model Specification and Estimation

We formally state the proposed model in Section 3.1, and present the estimation method in Section 3.2. We first show how the smoothing coefficient $\boldsymbol{c}_i$ and the latent factors $\boldsymbol{f}_i$ are estimated separately, and then introduce an iterative approach to find these estimates simultaneously.

3.1. Factor-augmented smoothing model

We consider a sample of functional data $\mathcal{X}_i(u)$, that takes values in the space $H := L^2(\mathcal{I})$ of real-valued square integrable functions on $\mathcal{I}$. The space H is a Hilbert space, equipped with the inner product $\langle x, y\rangle := \int x(u)y(u)du$. The function norm is defined as $\|x\| := \langle x, x\rangle^{1/2}$. The functional nature of $\mathcal{X}_i(u)$ allows us to represent it as a linear expansion of a set of K smooth basis functions:

$$\mathcal{X}_i(u) = \sum_{k=1}^{K} c_{ik}\phi_k(u), \quad u \in \mathcal{I},$$

where $\{\phi_k(u), k = 1, \ldots, K\}$ is a set of common basis functions and c_{ik} is the kth coefficient for the ith curve. Therefore, we can express the full model as

$$\begin{aligned} Y_{ij} &= \sum_{k=1}^{K} c_{ik}\phi_k(u_j) + \eta_{ij}, \\ \eta_{ij} &= \boldsymbol{a}_j^\top \boldsymbol{f}_i + \epsilon_{ij}, \quad i = 1, \ldots, n,\ j = 1, \ldots, p, \end{aligned}$$

where $\boldsymbol{f}_i \in \mathbb{R}^r$ are the unobserved common factors, $\boldsymbol{a}_j \in \mathbb{R}^r$ are the unobserved factor loadings, and r is the number of factors. We call this model the FASM. For the model to be identifiable, we require the following condition.

Condition 1 (Identification condition). We require

(i) $\{\mathcal{X}_i(u_j) : i = 1, \ldots, n; j = 1, \ldots, p\}$ are independent of $\{\eta_{ij} : i = 1, \ldots, n; j = 1, \ldots, p\}$;

(ii) $p^{-1}\sum_{j=1}^{p} \boldsymbol{a}_j\boldsymbol{a}_j^\top \xrightarrow{p} \Sigma_{\boldsymbol{a}} > 0$, for some $r \times r$ matrix $\Sigma_{\boldsymbol{a}}$ as $p \to \infty$; $n^{-1}\sum_{i=1}^{n} \boldsymbol{f}_i\boldsymbol{f}_i^\top \xrightarrow{p} \Sigma_{\boldsymbol{f}} > 0$, for some $r \times r$ matrix $\Sigma_{\boldsymbol{f}}$ as $n \to \infty$.

The first part of the identification condition ensures that the signal function component and the factor model component are independent. The second part ensures the existence of r factors, each of which makes a nontrivial contribution to the variance of η_{ij}, which in turn guarantees the identifiability between the factors and the error term ϵ_{ij}.

We treat the basis functions $\{\phi_k(u) : k = 1, 2, \ldots, K\}$ as known, and the number of basis functions K can be fixed or can go to infinity. There are various choices for basis functions in empirical data analysis, and the decision can be subjective. For example, Fourier bases are preferred for periodic data, whereas spline basis systems are most commonly used for nonperiodic data. Other bases include wavelet, polynomial, and ad-hoc basis functions. The number of basis functions K controls the smoothness of the predicted functions. As K increases, the estimator variance increases, but the bias decays. Cross-validation, for instance, can be used to determine K (Wahba and Word (1975)). We estimate our model using a roughness penalty approach explained in the next section. The smoothness of the functional component is switched from being determined by K to being determined by the tuning parameter of the penalty term. In practice, we can use a relatively large K and select the tuning parameter carefully. In Remark 4, we discuss common methods for selecting the tuning parameter.

3.2. Penalized estimation approach

We can write the model for the ith object as

$$\boldsymbol{Y}_i = \boldsymbol{\Phi}\boldsymbol{c}_i + \boldsymbol{A}\boldsymbol{f}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, n, \tag{3.1}$$

where

$$\boldsymbol{Y}_i = \begin{bmatrix} Y_{i1} \\ \vdots \\ Y_{ip} \end{bmatrix}, \ \boldsymbol{c}_i = \begin{bmatrix} c_{i1} \\ \vdots \\ c_{iK} \end{bmatrix}, \ \boldsymbol{\Phi} = \begin{bmatrix} \phi_1(u_1) \cdots \phi_K(u_1) \\ \vdots \qquad\quad \vdots \\ \phi_1(u_p) \cdots \phi_K(u_p) \end{bmatrix}, \ \boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1^\top \\ \vdots \\ \boldsymbol{a}_p^\top \end{bmatrix}, \ \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{ip} \end{bmatrix}.$$

Combining all the objects, we have, in matrix form,

$$\boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{C} + \boldsymbol{A}\boldsymbol{F}^\top + \boldsymbol{E}, \tag{3.2}$$

where $\boldsymbol{Y}$ is $p \times n$ and $\boldsymbol{C} = (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n)$ is a $K \times n$ matrix containing all the smoothing coefficients. The matrix $\boldsymbol{F} = (\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n)^\top$ is $n \times r$ and $\boldsymbol{E} = (\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n)$ is $p \times n$. Because $\boldsymbol{\Phi}$ is assumed to be known, we show how to estimate the parameters $\boldsymbol{C}, \boldsymbol{A}$, and $\boldsymbol{f}$.

For the latent factor estimation, there is an identification problem because $\boldsymbol{A}\boldsymbol{F}^\top = \boldsymbol{A}\boldsymbol{U}\boldsymbol{U}^{-1}\boldsymbol{F}^\top$ for any $r \times r$ invertible matrix $\boldsymbol{U}$. Thus, we impose the following normalization restriction on the matrices $\boldsymbol{A}$ and $\boldsymbol{F}$:

$$\frac{\boldsymbol{A}^\top\boldsymbol{A}}{p} = \boldsymbol{I}_r, \quad \text{and } \boldsymbol{F}^\top\boldsymbol{F} \text{ is a diagonal matrix.} \tag{3.3}$$

We propose implementing a penalized least squares, where the objective function is defined as

$$\mathrm{SSR}(\boldsymbol{c}_i, \boldsymbol{A}, \boldsymbol{f}) = \frac{1}{np}\sum_{i=1}^{n}\left\{(\boldsymbol{Y}_i - \boldsymbol{\Phi}\boldsymbol{c}_i - \boldsymbol{A}\boldsymbol{f}_i)^\top(\boldsymbol{Y}_i - \boldsymbol{\Phi}\boldsymbol{c}_i - \boldsymbol{A}\boldsymbol{f}_i) + \alpha\mathrm{PEN}_2(\mathcal{X}_i)\right\},$$

where $\mathrm{PEN}_2(\mathcal{X}_i)$ is a penalty term used for regularization, and α is the tuning parameter controlling the degree of regularization. The same α is used for all the functional observations i. This is a simplified case, where we assume a similar degree of smoothness for all curves. The tuning parameter can be chosen using cross-validation or information criteria. We penalize the "roughness" of the function term. To quantify the notion of "roughness" in a function, we use the square of the second derivative. Define the measure of roughness as

$$\mathrm{PEN}_2(\mathcal{X}_i) = \int_{\mathcal{I}} \left\{D^2\mathcal{X}_i(s)\right\}^2 ds,$$

where $D^2\mathcal{X}_i$ denotes taking the second derivative of the function $\mathcal{X}_i$. Here, a larger tuning parameter α indicates a smoother estimated function. Furthermore, we

denote

$$\boldsymbol{\Phi}(u) = \{\phi_1(u), \ldots, \phi_K(u)\}^\top. \tag{3.4}$$

Then,

$$\mathcal{X}_i(u) = \boldsymbol{c}_i^\top \boldsymbol{\Phi}(u).$$

We can re-express the roughness penalty $\mathrm{PEN}_2(\mathcal{X}_i)$ in matrix form, as follows:

$$\begin{aligned}
\mathrm{PEN}_2(\mathcal{X}_i) &= \int_{\mathcal{I}} \left\{D^2 \mathcal{X}_i(s)\right\}^2 ds \\
&= \int_{\mathcal{I}} \left\{D^2 \boldsymbol{c}_i^\top \boldsymbol{\Phi}(s)\right\}^2 ds \\
&= \int_{\mathcal{I}} \boldsymbol{c}_i^\top D^2\boldsymbol{\Phi}(s) D^2\boldsymbol{\Phi}^\top(s) \boldsymbol{c}_i ds \\
&= \boldsymbol{c}_i^\top \left\{ \int_{\mathcal{I}} D^2\boldsymbol{\Phi}(s) D^2\boldsymbol{\Phi}^\top(s) ds \right\} \boldsymbol{c}_i \\
&= \boldsymbol{c}_i^\top \boldsymbol{R} \boldsymbol{c}_i, \qquad i = 1, \ldots, n,
\end{aligned}$$

where

$$\boldsymbol{R} \equiv \int_{\mathcal{I}} D^2\boldsymbol{\Phi}(s) D^2\boldsymbol{\Phi}^\top(s) ds. \tag{3.5}$$

The matrix $\boldsymbol{R}$ is the same for all subjects, and the penalty term $\mathrm{PEN}_2(\mathcal{X}_i)$ differs for each subject only by the coefficient $\boldsymbol{c}_i$.

Remark 1. The number of smoothing coefficients $\boldsymbol{c}_i$ increases as the sample size increases. The penalty term penalizes the "roughness" of the smoothed function and mitigates the effect of the increasing number of parameters, thus controling the model flexibility.

Therefore, the objective function can be written as

$$\mathrm{SSR}(\boldsymbol{c}_i, \boldsymbol{A}, \boldsymbol{f}) = \frac{1}{np} \sum_{i=1}^{n} \left\{ (\boldsymbol{Y}_i - \boldsymbol{\Phi}\boldsymbol{c}_i - \boldsymbol{A}\boldsymbol{f}_i)^\top (\boldsymbol{Y}_i - \boldsymbol{\Phi}\boldsymbol{c}_i - \boldsymbol{A}\boldsymbol{f}_i) + \alpha \boldsymbol{c}_i^\top \boldsymbol{R} \boldsymbol{c}_i \right\},$$

subject to the constraint $\boldsymbol{A}^\top \boldsymbol{A}/p = \boldsymbol{I}_r$.

We aim to estimate the smoothing coefficient $\boldsymbol{c}_i$. We left multiply a matrix to each term in (3.1) to project the factor model term onto a zero matrix. Define the projection matrix

$$\boldsymbol{M}_{\boldsymbol{A}} \equiv \boldsymbol{I}_p - \boldsymbol{A}(\boldsymbol{A}^\top \boldsymbol{A})^{-1} \boldsymbol{A}^\top = \boldsymbol{I}_p - \frac{\boldsymbol{A}\boldsymbol{A}^\top}{p}. \tag{3.6}$$

Then,

$$\boldsymbol{M}_{\boldsymbol{A}} \boldsymbol{A} \boldsymbol{f}_i = \left(\boldsymbol{I}_p - \frac{\boldsymbol{A}\boldsymbol{A}^\top}{p} \right) \boldsymbol{A}\boldsymbol{f}_i = \left(\boldsymbol{A} - \frac{\boldsymbol{A}\boldsymbol{A}^\top \boldsymbol{A}}{p} \right) \boldsymbol{f}_i = \boldsymbol{0}.$$

Therefore, we estimate $\boldsymbol{c}_i$ from the projected equation

$$\boldsymbol{M_A Y}_i = \boldsymbol{M_A \Phi c}_i + \boldsymbol{M_A \epsilon}_i.$$

The projected objective function becomes

$$\mathrm{SSR}(\boldsymbol{c}_i, \boldsymbol{A}) = \frac{1}{np} \sum_{i=1}^{n} \left\{ (\boldsymbol{M_A Y}_i - \boldsymbol{M_A \Phi c}_i)^\top (\boldsymbol{M_A Y}_i - \boldsymbol{M_A \Phi c}_i) + \alpha \boldsymbol{c}_i^\top \boldsymbol{R c}_i \right\}. \tag{3.7}$$

By taking the derivative of $\mathrm{SSR}(\boldsymbol{c}_i, \boldsymbol{A})$ with respective to each $\boldsymbol{c}_i$ and setting it to zero, we can solve for the estimator $\widehat{\boldsymbol{c}}_i$:

$$\frac{\partial \mathrm{SSR}(\boldsymbol{c}_i \boldsymbol{A})}{\partial \boldsymbol{c}_i} = -\frac{1}{np} (\boldsymbol{M_A Y}_i - \boldsymbol{M_A \Phi c}_i)^\top (\boldsymbol{M_A \Phi}) + \frac{1}{np} \alpha \boldsymbol{c}_i^\top \boldsymbol{R}.$$

Setting the derivative to zero and rearranging the terms, we have

$$\left(\boldsymbol{\Phi}^\top \boldsymbol{M}_{\boldsymbol{A}}^\top \boldsymbol{M_A \Phi} + \alpha \boldsymbol{R} \right) \boldsymbol{c}_i = \boldsymbol{\Phi}^\top \boldsymbol{M}_{\boldsymbol{A}}^\top \boldsymbol{M_A Y}_i.$$

Using the fact that $\boldsymbol{M}_{\boldsymbol{A}}^\top \boldsymbol{M_A} = (\boldsymbol{I}_p - \boldsymbol{A}\boldsymbol{A}^\top/p)^\top (\boldsymbol{I}_p - \boldsymbol{A}\boldsymbol{A}^\top/p) = \boldsymbol{M_A}$, we obtain the least squares estimator for $\boldsymbol{c}_i$ given $\boldsymbol{A}$:

$$\widehat{\boldsymbol{c}}_i = \left(\boldsymbol{\Phi}^\top \boldsymbol{M_A \Phi} + \alpha \boldsymbol{R} \right)^{-1} \boldsymbol{\Phi}^\top \boldsymbol{M_A Y}_i.$$

Next, to estimate $\boldsymbol{A}$ and $\boldsymbol{f}_i$, we focus on the factor model

$$\boldsymbol{\eta}_i = \boldsymbol{A f}_i + \boldsymbol{\epsilon}_i,$$

expressed in matrix form as

$$\boldsymbol{Z} = \boldsymbol{A F}^\top + \boldsymbol{E},$$

where $\boldsymbol{Z} = (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_n)$. In high-dimensional cases, the unknown factors and loadings are typically estimated using least squares (i.e., the principal component analysis; see, e.g., Fan, Fan and Lv (2008), Onatski (2012)). The least squares objective function is

$$\mathrm{tr} \left\{ (\boldsymbol{Z} - \boldsymbol{A F}^\top)(\boldsymbol{Z} - \boldsymbol{A F}^\top)^\top \right\}. \tag{3.8}$$

Minimizing the objective function with respect to $\boldsymbol{F}^\top$, we have $\boldsymbol{F}^\top = (\boldsymbol{A}^\top \boldsymbol{A})^{-1} \boldsymbol{A}^\top \boldsymbol{Z} = \boldsymbol{A}^\top \boldsymbol{Z}/p$, using (3.3). Substituting in (3.8), we obtain the objective function

$$\mathrm{tr} \left\{ \left(\boldsymbol{Z} - \frac{\boldsymbol{A}\boldsymbol{A}^\top \boldsymbol{Z}}{p} \right) \left(\boldsymbol{Z} - \frac{\boldsymbol{A}\boldsymbol{A}^\top \boldsymbol{Z}}{p} \right)^\top \right\}$$

$$= \operatorname{tr}\left(\boldsymbol{Z}\boldsymbol{Z}^\top - \frac{\boldsymbol{Z}\boldsymbol{Z}^\top\boldsymbol{A}\boldsymbol{A}^\top}{p} - \frac{\boldsymbol{Z}\boldsymbol{Z}^\top\boldsymbol{A}\boldsymbol{A}^\top}{p} + \frac{\boldsymbol{A}\boldsymbol{A}^\top\boldsymbol{Z}\boldsymbol{Z}^\top\boldsymbol{A}\boldsymbol{A}^\top}{p^2}\right)$$
$$= \operatorname{tr}(\boldsymbol{Z}\boldsymbol{Z}^\top) - \frac{\operatorname{tr}(\boldsymbol{A}^\top\boldsymbol{Z}\boldsymbol{Z}^\top\boldsymbol{A})}{p},$$

where the last equality uses (3.3) and $\operatorname{tr}(\boldsymbol{Z}\boldsymbol{Z}^\top\boldsymbol{A}\boldsymbol{A}^\top) = \operatorname{tr}(\boldsymbol{A}^\top\boldsymbol{Z}\boldsymbol{Z}^\top\boldsymbol{A})$. Thus, minimizing the objective function is equivalent to maximizing $\operatorname{tr}(\boldsymbol{A}^\top\boldsymbol{Z}\boldsymbol{Z}^\top\boldsymbol{A})/p$. The estimator for $\boldsymbol{A}$ is obtained by finding the first r eigenvectors corresponding to the r largest eigenvalues of the matrix $(1/p)\boldsymbol{Z}\boldsymbol{Z}^\top$ in descending order, where

$$\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top = \frac{1}{p}\sum_{i=1}^{n}\boldsymbol{\eta}_i\boldsymbol{\eta}_i^\top = \frac{1}{p}\sum_{i=1}^{n}(\boldsymbol{Y}_i - \boldsymbol{\Phi}\boldsymbol{c}_i)(\boldsymbol{Y}_i - \boldsymbol{\Phi}\boldsymbol{c}_i)^\top.$$

Therefore, knowing $\boldsymbol{c}_i$, we solve for $\widehat{\boldsymbol{A}}$ using

$$\left\{\frac{1}{np}\sum_{i=1}^{n}(\boldsymbol{Y}_i - \boldsymbol{\Phi}\boldsymbol{c}_i)(\boldsymbol{Y}_i - \boldsymbol{\Phi}\boldsymbol{c}_i)^\top\right\}\widehat{\boldsymbol{A}} = \widehat{\boldsymbol{A}}\boldsymbol{V}_{np}, \tag{3.9}$$

where $\boldsymbol{V}_{np}$ is an $r \times r$ diagonal matrix containing the r eigenvalues of the matrix in the square brackets in decreasing order. The additional coefficient $1/n$ is used for scaling.

Remark 2. The number of factors r is selected using some criteria on the eigenvalues. There have been many studies on this topic. For example, Bai and Ng (2002) propose two model selection criteria functions, Onatski (2010) estimate the number of factors using differenced adjacent eigenvalues, and Ahn and Horenstein (2013) select the number based on the ratio of two adjacent eigenvalues. A modified ratio criterion of Ahn and Horenstein (2013) is proposed in Section S3.1. Note that the simulations and empirical applications presented here show that our model performs well under various common selection criteria in the literature.

We need $\boldsymbol{A}$ to find $\widehat{\boldsymbol{c}}_i$, and $\boldsymbol{c}_i$ to find $\widehat{\boldsymbol{A}}$. The final estimator $(\widehat{\boldsymbol{c}}_i, \widehat{\boldsymbol{A}})$ is the solution of the set of equations

$$\begin{cases}\widehat{\boldsymbol{c}}_i = \left(\boldsymbol{\Phi}^\top\boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{\Phi} + \alpha\boldsymbol{R}\right)^{-1}\boldsymbol{\Phi}^\top\boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{Y}_i, \qquad i = 1, \ldots, n \\ \left\{\frac{1}{np}\sum_{i=1}^{n}\left(\boldsymbol{Y}_i - \boldsymbol{\Phi}\widehat{\boldsymbol{c}}_i\right)\left(\boldsymbol{Y}_i - \boldsymbol{\Phi}\widehat{\boldsymbol{c}}_i\right)^\top\right\}\widehat{\boldsymbol{A}} = \widehat{\boldsymbol{A}}\boldsymbol{V}_{np}.\end{cases} \tag{3.10}$$

Because there is no closed-form expression of $\widehat{\boldsymbol{A}}$ and $\widehat{\boldsymbol{c}}_i$, we propose using numerical iterations to find the estimates.

Remark 3. In this paper, we use $\widehat{\boldsymbol{A}}^{(0)} = \boldsymbol{0}$. This means we start by ignoring the factor model component, so the initial value for the smoothing coefficient $\widehat{\boldsymbol{c}}_i^{(0)} = \left(\boldsymbol{\Phi}^\top\boldsymbol{\Phi} + \alpha\boldsymbol{R}\right)^{-1}\boldsymbol{\Phi}^\top\boldsymbol{Y}_i$, which is simply the ridge estimator. The convergence of

Algorithm 1: Iterations for estimating FASM.

1. Denote the initial value as $\widehat{\boldsymbol{A}}^{(0)}$. Using (3.10), we obtain $\widehat{\boldsymbol{c}}_i^{(0)} = (\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}^{(0)}} \boldsymbol{\Phi} + \alpha \boldsymbol{R})^{-1} \boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}^{(0)}} \boldsymbol{Y}_i$.
2. With $\widehat{\boldsymbol{c}}_i^{(t)}$, we substitute into the second equation of (3.10) to obtain $\widehat{\boldsymbol{A}}^{(t+1)} = (\widehat{\boldsymbol{a}}_1^{(t+1)}, \ldots, \widehat{\boldsymbol{a}}_r^{(t+1)})^\top$, where $\widehat{\boldsymbol{a}}_j^{(t+1)}$ is the eigenvector of the matrix $(np)^{-1} \sum_{i=1}^n (\boldsymbol{Y}_i - \boldsymbol{\Phi}\widehat{\boldsymbol{c}}_i^{(t+1)})(\boldsymbol{Y}_i - \boldsymbol{\Phi}\widehat{\boldsymbol{c}}_i^{(t+1)})^\top$ corresponding to its jth largest eigenvalue.
3. With $\widehat{\boldsymbol{A}}^{(t+1)}$, we obtain $\widehat{\boldsymbol{c}}_i^{(t+1)} = \left(\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}^{(t+1)}} \boldsymbol{\Phi} + \alpha \boldsymbol{R}\right)^{-1} \boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}^{(t+1)}} \boldsymbol{Y}_i$ using (3.10).
4. We then repeat steps 2 and 3 until $\|\widehat{\boldsymbol{c}}_i^{(t+1)} - \widehat{\boldsymbol{c}}_i^{(t)}\| < \delta$, where δ is a prescribed small positive value.

Newton's numeric iteration requires the convergence of this estimator, which, in turn, requires the factor model component η_{ij} to have an expectation of zero. The stopping criterion focuses only on $\widehat{\boldsymbol{c}}_i$, because we are interested in estimating η_{ij} as a whole.

Remark 4. Common methods for selecting the shrinkage parameter α include Akaike's information criterion (AIC; Akaike (1974)), the Bayesian information criterion (BIC; Schwarz (1978)), and cross-validation. We use the mean generalized cross-validation (mGCV) method (Golub, Heath and Wahba (1979)). We define, at step t,

$$\mathrm{mGCV}^{(t)} = \frac{1}{n} \sum_{i=1}^{n} \frac{\mathrm{pSSE}_i^{(t)}}{\{p - df^{(t)}(\alpha)\}^2}, \tag{3.11}$$

where $\mathrm{SSE}_i^{(t)}$ is the sum of squares residual for the ith object at step t, and $df^{(t)}(\alpha)$ is the equivalent degrees of freedom measure, which can be calculated as

$$\mathrm{df}^{(t)}(\alpha) = \mathrm{tr}\left\{ \boldsymbol{\Phi} \left(\boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}^{(t)}} \boldsymbol{\Phi} + \alpha \boldsymbol{R}\right)^{-1} \boldsymbol{\Phi}^\top \boldsymbol{M}_{\widehat{\boldsymbol{A}}^{(t)}} \right\}. \tag{3.12}$$

At each step of the iteration, the tuning parameter α is chosen by minimizing the $mGCV^{(t)}$.

Remark 5. Algorithm 1 iterates the ridge regression and the PCA. The convergence of this iterative algorithm is studied in Jiang et al. (2021). For instance, Theorem 2 of Jiang et al. (2021) provides sufficient conditions under which the recursive algorithm converges to the true value, or some other values. In particular, this algorithm converges to the true parameter when the regressors are independent of the common factors, or when the factors involved in the regressors are weaker than the common ones.

After we obtain the estimates $\widehat{\boldsymbol{A}}$ and $\widehat{\boldsymbol{c}}_i$, the estimated coefficient matrix $\widehat{\boldsymbol{C}}$ is constructed as $\widehat{\boldsymbol{C}} = (\widehat{\boldsymbol{c}}_1, \ldots, \widehat{\boldsymbol{c}}_n)$, and the estimated factor can be obtained using

$$\widehat{\boldsymbol{F}}^\top = \widehat{\boldsymbol{A}}^\top (\boldsymbol{Y} - \boldsymbol{\Phi}\widehat{\boldsymbol{C}}).$$

Finally, the functional component can be estimated by $\widehat{\mathcal{X}}_i(u) = \widehat{\boldsymbol{c}}_i^\top \boldsymbol{\Phi}(u)$, where $\boldsymbol{\Phi}(u)$ is defined in (3.4).

Remark 6. Although we have imposed the constraint in (3.3) and identification condition 1, $\boldsymbol{A}$ and $\boldsymbol{f}_i$ are not determined uniquely, because the model (3.1) is unchanged if we replace $\boldsymbol{A}$ and $\boldsymbol{f}_i$ with $\boldsymbol{AU}$ and $\boldsymbol{U}^\top \boldsymbol{f}_i$, for any orthogonal $r \times r$ matrix $\boldsymbol{U}$. However, the linear space spanned by the columns of $\boldsymbol{A}$ is uniquely defined. Although we are not able to estimate $\boldsymbol{A}$, we can still estimate a rotation of $\boldsymbol{A}$, which spans the same space as $\boldsymbol{A}$ does. The matrix $\boldsymbol{M_A}$ defined in (3.6) is a projecting matrix onto the orthogonal complement of the linear space spanned by the columns of $\boldsymbol{A}$. In the next section, we show that the estimator $\boldsymbol{M}_{\widehat{\boldsymbol{A}}}$ for $\boldsymbol{M_A}$ is consistent.

4. Asymptotic Theory

In this section, we study the asymptotic properties of the coefficient estimator $\widehat{\boldsymbol{c}}_i$ with growing sample size and dimension. We state the assumptions in Section 4.1 and provide the asymptotic results of $\widehat{\boldsymbol{c}}_i$ in Section 4.2.

4.1. Assumptions

We use $(\boldsymbol{c}_i^0, \boldsymbol{A}^0)$ to denote the true parameters. In this paper, we use the L^2 norm. The norm of a vector $\boldsymbol{x}$ is defined as $\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^\top \boldsymbol{x}}$, and the norm of a matrix $\boldsymbol{U}$ is defined as $\|\boldsymbol{U}\| = \sqrt{\lambda_{\max}(\boldsymbol{U}^\top \boldsymbol{U})}$, where $\lambda_{\max}(\cdot)$ represents the largest eigenvalue of a matrix. We introduce the matrix

$$\boldsymbol{D}_i(\boldsymbol{A}) \equiv \frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M_A}\boldsymbol{\Phi} - \frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M_A}\boldsymbol{\Phi}\boldsymbol{f}_i^\top \left(\frac{\boldsymbol{F}^\top \boldsymbol{F}}{n}\right)^\top \boldsymbol{f}_i. \tag{4.1}$$

This matrix plays an important role in the proof of the consistency of $\widehat{\boldsymbol{c}}_i$; see S5.1 Appendix A. The identifying condition for $\boldsymbol{c}_i^0$ is that $\boldsymbol{D}_i(\boldsymbol{A})$ is positive definite for all i, which is stated in Assumption 3.

First, we state the assumptions.

Assumption 1.

$$\frac{1}{p}\left\|\boldsymbol{\Phi}^\top \boldsymbol{\Phi}\right\| = O(1), \quad ||\boldsymbol{R}|| = O(1), \quad K = o\{\min(n, p)\}, \quad p = o\left(n^2\right).$$

In this assumption, the number of basis functions K can either be fixed or tend to infinity.

Assumption 2.

$$\|\boldsymbol{c}_i^0\| = O_p(1), \text{ for all } i = 1, 2, \ldots, n.$$

This assumption is introduced to ensure the consistency of the estimated coefficients $\widehat{\boldsymbol{c}}_i^0$. Note that the dimension of $\boldsymbol{c}_i^0$ is K, which may go to infinity, but its norm is restricted to be of constant order. Therefore, this assumption controls the flexibility of the model.

Assumption 3. *Let* $\mathcal{A} = \{\boldsymbol{A} : \boldsymbol{A}^\top \boldsymbol{A}/p = \boldsymbol{I}, \text{ and } \boldsymbol{A} \text{ independent of } \boldsymbol{\Phi}\}$. *We assume*

$$\inf_{\boldsymbol{A} \in \mathcal{A}} \boldsymbol{D}_i(\boldsymbol{A}) > 0.$$

This assumption is the identification condition for $\boldsymbol{c}_i^0$. The usual assumption for the least-squares estimator contains only the first term on the right-hand side of (4.1). The second term arises because of the unobservable matrices $\boldsymbol{F}$ and $\boldsymbol{A}$.

Assumption 4. *For some constant* $M > 0$,

1. $\mathbb{E}\|\boldsymbol{a}_j^0\|^4 \leq M$, *for* $j = 1, \ldots, p$, *and* $(1/p)\boldsymbol{A}^\top \boldsymbol{A} \xrightarrow{p} \boldsymbol{\Sigma}_{\boldsymbol{a}} > 0$, *for some* $r \times r$ *matrix* $\boldsymbol{\Sigma}_{\boldsymbol{a}}$, *as* $p \to \infty$.
2. $\mathbb{E}\|\boldsymbol{f}_i\|^4 \leq M$ *and* $(1/n)\boldsymbol{F}\boldsymbol{F}^\top \xrightarrow{p} \boldsymbol{\Sigma}_{\boldsymbol{f}} > 0$, *for some* $r \times r$ *matrix* $\boldsymbol{\Sigma}_{\boldsymbol{f}}$, *as* $n \to \infty$.

Assumption 5. *For some constant* $M > 0$, *the error terms* ϵ_{ji}, *for* $j = 1, \ldots, p$ *and* $i = 1, \ldots, n$, *are i.i.d. in both directions, with* $\mathbb{E}(\epsilon_{ji}) = 0$, $Var(\epsilon_{ji}) = \sigma^2$, *and* $\mathbb{E}|\epsilon_{ji}|^8 \leq M$.

Assumption 6. ϵ_{ji} *is independent of* ϕ_s, $\boldsymbol{f}_t$, *and* $\boldsymbol{a}_s^0$, *for all* j, i, s, t.

We require that the errors are independent in themselves and also of the functional term $\phi(u)$ and the factor model terms $\boldsymbol{f}_i$ and $\boldsymbol{a}_j^0$. To highlight the main contribution of our method, we use a simplified setting on the error terms to exclude endogeneity. Nevertheless, with simple, but tedious modifications, Assumption 5 can be relaxed, and our model can be extended to more complicated settings that allow correlations between the error term and the factor model term.

Assumption 7. *The tuning parameter* α *satisfies* $\alpha = o(p)$.

This assumption is conventional in ridge regression (see, e.g., Knight and Fu (2000)) and ensures that the estimator's asymptotic bias is zero.

Before stating the next assumption, we introduce some notation. Let $\boldsymbol{\omega}_j$, for $j = 1, \ldots, p$, denote the jth column of the $K \times p$ matrix $\boldsymbol{\Phi}^\top \boldsymbol{M}_{\boldsymbol{A}^0}$, and let ψ_{ik} denote the (i, k)th element of the matrix $\boldsymbol{M}_{\boldsymbol{F}}$, where

$$\boldsymbol{M}_{\boldsymbol{F}} \equiv \boldsymbol{I}_n - \boldsymbol{F}\left(\boldsymbol{F}^\top \boldsymbol{F}\right)\boldsymbol{F}^\top. \tag{4.2}$$

Then, for any vector $\boldsymbol{b} = (b_1, \ldots, b_n)^\top$, we can write

$$\frac{1}{\sqrt{np}}\boldsymbol{\Phi}^\top \boldsymbol{M_{A^0}} \boldsymbol{E} \boldsymbol{M_F} \boldsymbol{b} = \frac{1}{\sqrt{np}} \sum_i^n \sum_j^p \boldsymbol{\omega}_j \epsilon_{ji} \sum_k^n \psi_{ik} b_k \equiv \frac{1}{\sqrt{np}} \sum_i^n \sum_j^p \boldsymbol{x}_{ij}. \tag{4.3}$$

In (4.3), for notational simplicity, we define $\boldsymbol{x}_{ij}$ as $\boldsymbol{\omega}_j \epsilon_{ji} \sum_k^n \psi_{ik} b_k$. The matrix $\boldsymbol{\Phi}^\top \boldsymbol{M_{A^0}} \boldsymbol{E} \boldsymbol{M_F}$ is of interest because it is the main component that contributes to the asymptotic distribution of the estimators, as shown in the next section.

Let

$$\boldsymbol{L}_{np} \equiv \frac{\sigma^2}{np} \sum_i^n \sum_j^p \boldsymbol{\omega}_j \boldsymbol{\omega}_j^\top \left(\sum_k^n \psi_{ik} b_k \right)^2. \tag{4.4}$$

We make the following assumption.

Assumption 8. *When K is fixed, we assume there exists a $K \times K$ matrix $\boldsymbol{L}$ such that*

$$\boldsymbol{L} \equiv \lim_{n,p \to \infty} \boldsymbol{L}_{np}, \tag{4.5}$$

where $\boldsymbol{L}_{np}$ is defined in (4.4)*. Let ν^2 be the smallest eigenvalue of the matrix $\boldsymbol{L}$ defined in* (4.5)*. Then assume that $\nu^2 > 0$, and that, for all $\varepsilon > 0$,*

$$\lim_{n,p \to \infty} \frac{1}{np\nu^2} \sum_{i=1}^n \sum_{j=1}^p \mathbb{E}\left\{ \|\boldsymbol{x}_{ij}\|^2 \mathbb{I}\left(\|\boldsymbol{x}_{ij}\|^2 \geq \varepsilon np\nu^2 \right) \right\} = 0, \tag{4.6}$$

where $\mathbb{I}(\cdot)$ is the indicator function, defined by $\mathbb{I}(A) = 1$ if A occurs, and $\mathbb{I}(A) = 0$ otherwise.

This assumption is the multivariate Lindeberg condition, which is needed to construct the central limit theorem when K is fixed. This is by no means a strong condition; for instance, when the factor model component is ignored, $\boldsymbol{\omega}_j$ is simply $\boldsymbol{\phi}_j$, and $\boldsymbol{x}_{ij} = \boldsymbol{\phi}_j b_i \epsilon_{ji}$. Because we assume $\boldsymbol{\phi}_j = O(1)$ in Assumption 1, the Lindeberg condition in (4.6) is met.

Next, we introduce a similar assumption to Assumption 8 for the central limit theorem when K goes to infinity. Let $\widetilde{\boldsymbol{\omega}}_j$, for $j = 1, \ldots, p$, denote the jth column of the $K \times p$ matrix $((1/p)\Phi^\top \boldsymbol{M_{A^0}} \boldsymbol{\Phi})^{-1} \boldsymbol{\Phi}^\top \boldsymbol{M_{A^0}}$. Denote

$$\begin{aligned} &\boldsymbol{\gamma}^\top \left(\frac{1}{p} \Phi^\top \boldsymbol{M_{A^0}} \boldsymbol{\Phi} \right)^{-1} \frac{1}{\sqrt{np}} \boldsymbol{\Phi}^\top \boldsymbol{M_{A^0}} \boldsymbol{E} \boldsymbol{M_F} \mathbf{b} \\ &= \frac{1}{\sqrt{np}} \sum_i^n \sum_j^p \boldsymbol{\gamma}^\top \widetilde{\boldsymbol{\omega}}_j \epsilon_{ji} \sum_k^n \psi_{ik} b_k \equiv \frac{1}{\sqrt{np}} \sum_i^n \sum_j^p \widetilde{x}_{ij}. \end{aligned} \tag{4.7}$$

Assumption 9. *Assume there exists a constant $\widetilde{L} > 0$ such that*

$$\widetilde{L} \equiv \lim_{n,p,K \to \infty} \widetilde{L}_{np}, \tag{4.8}$$

where $\widetilde{L}_{np}$ *is defined as*

$$\widetilde{L}_{np} \equiv \frac{\sigma^2}{np} \sum_{i}^{n} \sum_{j}^{p} \left(\boldsymbol{\gamma}^\top \widetilde{\boldsymbol{\omega}}_j\right)^2 \left(\sum_{k}^{n} \psi_{ik} b_k\right)^2. \tag{4.9}$$

Moreover, for any small value $\varepsilon > 0$*, we assume*

$$\lim_{n,p,K\to\infty} \frac{1}{np\nu^2} \sum_{i=1}^{n} \sum_{j=1}^{p} \mathbb{E}\left\{|\widetilde{x}_{ij}|^2 \mathbb{I}\left(|\widetilde{x}_{ij}|^2 \geq \varepsilon np\nu^2\right)\right\} = 0, \tag{4.10}$$

where $\mathbb{I}(\cdot)$ *is the indicator function, defined by* $\mathbb{I}(A) = 1$ *if* A *occurs, and* $\mathbb{I}(A) = 0$ *otherwise.*

4.2. Asymptotic properties

As mentioned previously, the identification problem of the latent factor implies that we use $\widehat{\boldsymbol{A}}$ to estimate a rotation of $\boldsymbol{A}^0$. Based on the objective function (3.7) in Section 3, we use the following center-adjusted objective function:

$$S_{np}(\boldsymbol{c}_i, \boldsymbol{A}) = \frac{1}{np} \sum_{i=1}^{n} \left\{(\boldsymbol{Y}_i - \boldsymbol{\Phi}\boldsymbol{c}_i)^\top \boldsymbol{M}_{\boldsymbol{A}} (\boldsymbol{Y}_i - \boldsymbol{\Phi}\boldsymbol{c}_i) + \alpha \boldsymbol{c}_i^\top \boldsymbol{R} \boldsymbol{c}_i\right\} - \frac{1}{np} \sum_{i=1}^{n} \epsilon_i^\top \boldsymbol{M}_{\boldsymbol{A}^0} \epsilon_i, \tag{4.11}$$

where $\boldsymbol{M}_{\boldsymbol{A}}$ is defined in (3.6), satisfying $\boldsymbol{A}^\top \boldsymbol{A}/p = \boldsymbol{I}_r$. The second term on the right-hand side of (4.11) does not contain the unknowns $\boldsymbol{A}$ and $\boldsymbol{c}_i$, so including this term does not affect the optimization result. This term is only used to adjust the center. Thus, the resulting objective function has an expectation of zero. We estimate $\boldsymbol{c}_i^0$ and $\boldsymbol{A}^0$ by

$$(\widehat{\boldsymbol{c}}_i, \widehat{\boldsymbol{A}}) = \underset{\boldsymbol{c}_i, \boldsymbol{A}}{\operatorname{argmin}} S_{np}(\boldsymbol{c}_i, \boldsymbol{A}). \tag{4.12}$$

In the following, we establish the asymptotic properties for the estimated coefficient matrix $\widehat{\boldsymbol{C}}$. In Theorem 1, we prove the consistency of the matrix $\widehat{\boldsymbol{C}}$. In Theorem 2, we show the rate of convergence of $\widehat{\boldsymbol{C}}$. Theorem 3 provides the asymptotic distribution of $\widehat{\boldsymbol{C}}$.

Let $\boldsymbol{P}_{\boldsymbol{U}} = \boldsymbol{U}(\boldsymbol{U}^\top \boldsymbol{U})^{-1} \boldsymbol{U}^\top$ for a matrix $\boldsymbol{U}$.

Theorem 1. *Under Assumptions* 1-6*, as* $n, p \to \infty$*, we have the following statements:*

(i) $(1/\sqrt{n})\|\boldsymbol{C} - \widehat{\boldsymbol{C}}\| \xrightarrow{p} 0$.

(ii) $\|\boldsymbol{P}_{\widehat{\boldsymbol{A}}} - \boldsymbol{P}_{\boldsymbol{A}^0}\| \xrightarrow{p} 0$.

We start by proving the consistency of the vector $\widehat{\boldsymbol{c}}_i$. This consistency is uniform for all $i = 1, \ldots, n$. Therefore, we can combine $\boldsymbol{c}_i$, for all $i = 1, \ldots, n$,

and have the result for the coefficient matrix $\widehat{\boldsymbol{C}}$ in (i). The matrix $\widehat{\boldsymbol{C}}$ is of dimension $K \times n$, where K is fixed and the sample size n goes to infinity, so there is a $(\sqrt{n})^{-1}$ scale in the result of (i). In the second part of the theorem, note that $\boldsymbol{P_A} = \boldsymbol{I}_p - \boldsymbol{M_A}$, where $\boldsymbol{M_A}$ is the projection matrix onto the orthogonal complement of the linear space spanned by the columns of $\boldsymbol{A}$. Thus, $\boldsymbol{P}_{\widehat{\boldsymbol{A}}}$ and $\boldsymbol{P}_{\boldsymbol{A}^0}$ represent the spaces spanned by $\widehat{\boldsymbol{A}}$ and $\boldsymbol{A}^0$, respectively, and we show that they are asymptotically the same in (ii).

Next, we obtain the rate of convergence.

Theorem 2. *Under Assumptions* 1-6*, if* $p/n \to \rho > 0$ *as* $n, p \to \infty$*, then*

$$\left\| \frac{(\boldsymbol{C}^0 - \widehat{\boldsymbol{C}})}{\sqrt{n}} \boldsymbol{M_F} \right\| = O_p\left(\frac{1}{\sqrt{p}}\right),$$

where $\boldsymbol{M_F}$ *is defined in* (4.2)*.*

We study the case when the dimension p and the sample size n are comparable. We achieve rate $\sqrt{p}$ convergence, considering $(\sqrt{n})^{-1}\|\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}\|$ on the whole. We expect that the rate of convergence of the smoothing models depends on the number of discrete points p observed on each curve.

Remark 7. The asymptotic result in Theorem 2 contains a projection matrix $\boldsymbol{M_F}$. This matrix projects $\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}$ onto the space orthogonal to the factor matrix $\boldsymbol{F}$. This theorem shows the relationship between $\boldsymbol{C}^0$ and $\boldsymbol{F}$. When $\boldsymbol{C}^0$ and $\boldsymbol{F}$ are orthogonal, $(\boldsymbol{C}^0 - \widehat{\boldsymbol{C}})\boldsymbol{M_F} = \boldsymbol{C}^0 - \widehat{\boldsymbol{C}}$, and we obtain the rate of convergence of $\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}$. When $\boldsymbol{C}^0$ and $\boldsymbol{F}$ are not orthogonal, the inference on $\boldsymbol{C}^0$ is affected by the existence of the factor model component.

We further establish the limiting distribution. In S5.1 Appendix A, we show that

$$\left\| \sqrt{p}\frac{(\boldsymbol{C}^0 - \widehat{\boldsymbol{C}})}{\sqrt{n}} \boldsymbol{M_F} \right\| = \left\| \left(\frac{1}{p}\Phi^\top \boldsymbol{M}_{\boldsymbol{A}^0}\boldsymbol{\Phi}\right)^{-1} \frac{1}{\sqrt{np}} \boldsymbol{\Phi}^\top \boldsymbol{M}_{\boldsymbol{A}^0} \boldsymbol{E} \boldsymbol{M_F} \right\| + o_p(1).$$

The limiting distribution is constructed based on the first term on the right-hand side.

Theorem 3. *In addition to Assumptions* 1-8*, we assume that* K *is fixed and, as* $n, p \to \infty$*, there exists a positive-definite matrix* $\widetilde{\boldsymbol{Q}}$ *such that*

$$\boldsymbol{Q}\left(\boldsymbol{A}^0\right) \equiv \frac{1}{p}\boldsymbol{\Phi}^\top \boldsymbol{M}_{\boldsymbol{A}^0}\boldsymbol{\Phi} \xrightarrow{p} \widetilde{\boldsymbol{Q}}.$$

Then, we have, for any vector $\boldsymbol{b} \in \mathbb{R}^n$*,*

$$\sqrt{p}\left(\frac{\boldsymbol{C}^0 - \widehat{\boldsymbol{C}}}{\sqrt{n}}\right)\boldsymbol{M_F}\boldsymbol{b} \xrightarrow{d} \mathcal{N}\left(\boldsymbol{0}, \widetilde{\boldsymbol{Q}}^{-1}\boldsymbol{L}\widetilde{\boldsymbol{Q}}^{-1}\right),$$

where $\boldsymbol{M_F}$ *is defined in Theorem* 2*, and* $\boldsymbol{L}$ *is defined in* (4.5)*. Moreover, if the limit of* $\boldsymbol{\Phi}\widetilde{\boldsymbol{Q}}^{-1}\boldsymbol{L}\widetilde{\boldsymbol{Q}}^{-1}\boldsymbol{\Phi}^\top$ *exists, we have*

$$\sqrt{p}\boldsymbol{\Phi}\left(\frac{\boldsymbol{C}^0-\widehat{\boldsymbol{C}}}{\sqrt{n}}\right)\boldsymbol{M_F}\boldsymbol{b}\xrightarrow{d}\mathcal{N}\left(\boldsymbol{0},\lim_{n,p\to\infty}\boldsymbol{\Phi}\widetilde{\boldsymbol{Q}}^{-1}\boldsymbol{L}\widetilde{\boldsymbol{Q}}^{-1}\boldsymbol{\Phi}^\top\right),$$
$$\sqrt{\frac{p}{n}}\left(\boldsymbol{X}-\widehat{\boldsymbol{X}}\right)\boldsymbol{M_F}\boldsymbol{b}\xrightarrow{d}\mathcal{N}\left(\boldsymbol{0},\lim_{n,p\to\infty}\boldsymbol{\Phi}\widetilde{\boldsymbol{Q}}^{-1}\boldsymbol{L}\widetilde{\boldsymbol{Q}}^{-1}\boldsymbol{\Phi}^\top\right).$$

The vector $\boldsymbol{b}$ comes from the same vector in Lemma 1. The asymptotic bias is zero, because we assume no serial or cross-sectional correlation in the error terms. This simplified setting can be extended to allow for weak correlations in the errors in both directions. In that case, the asymptotic distribution includes a nonzero bias term.

Remark 8. Theorem 3 shows that the asymptotic distribution of the coefficient matrix $\widehat{\boldsymbol{C}}$ relies on the unobserved factor loading matrix $\boldsymbol{A}^0$. Although we are unable to consistently estimate $\boldsymbol{A}^0$ using $\widehat{\boldsymbol{A}}$, what we need is the projected matrix $\boldsymbol{M}_{\boldsymbol{A}^0}$, which can be estimated using $\boldsymbol{M}_{\widehat{\boldsymbol{A}}}$. We derive the following estimators for $\boldsymbol{Q}$ and $\boldsymbol{L}$ based on $\boldsymbol{M}_{\widehat{\boldsymbol{A}}}$:

$$\widehat{\boldsymbol{Q}}=\frac{1}{p}\boldsymbol{\Phi}^\top\boldsymbol{M}_{\widehat{\boldsymbol{A}}}\boldsymbol{\Phi}$$
$$\widehat{\boldsymbol{L}}=\frac{\sigma^2}{np}\sum_i^n\sum_j^p\widehat{\boldsymbol{\omega}}_j^\top\widehat{\boldsymbol{\omega}}_j\left(\sum_k^n\widehat{\psi}_{ik}b_k\right)^2,$$

where $\widehat{\boldsymbol{\omega}}_j$ is the jth column of the $K\times p$ matrix $\boldsymbol{\Phi}^\top\boldsymbol{M}_{\widehat{\boldsymbol{A}}}$, and $\widehat{\psi}_{ik}$ is the (i,k)th element in the matrix $\boldsymbol{M}_{\widehat{\boldsymbol{F}}}$.

When K goes to infinity, the random vector $\{(\boldsymbol{C}^0-\widehat{\boldsymbol{C}})/\sqrt{n}\}\boldsymbol{M_F}\boldsymbol{b}$ considered in Theorem 3 is a high-dimensional random vector. Then, we establish the asymptotic distribution for some linear combination of this random vector in the following theorem.

Theorem 4. *Suppose Assumptions* 1-7 *and Assumption* 9 *are satisfied. As* n,p,K *tend to infinity, we have, for any vectors* $\boldsymbol{\gamma}\in\mathbb{R}^K$ *and* $\boldsymbol{b}\in\mathbb{R}^n$*,*

$$\sqrt{p}\boldsymbol{\gamma}^\top\left(\frac{\boldsymbol{C}^0-\widehat{\boldsymbol{C}}}{\sqrt{n}}\right)\boldsymbol{M_F}\boldsymbol{b}\xrightarrow{d}\mathcal{N}\left(\boldsymbol{0},\widetilde{L}\right),\tag{4.13}$$

where $\widetilde{L}$ *is defined in* (4.8)*.*

5. Simulation Studies

In this section, we use simulated data to demonstrate the performance of the proposed model. We compare the FASM with three other models. The first is

the basis smoothing model with a penalty (Bsmooth). We use the same basis functions as those of the FASM and the L_2-penalty. Thus, the Bsmooth model differs from the FASM only in that it does not consider the augmented factor component. The second model is the local linear smoothing model (Localin). Here, we use Cross-validation to choose the bandwidth parameter. The third model is the principal component analysis with conditional expectation (PACE), proposed in Yao, Müller and Wang (2005). This method is mainly used for sparse data, where very few data points are observed on each functional curve.

5.1. Data-generating process

Setting 1. We generate simulated data Y_{ij}, where $i = 1, \ldots, n$ and $j = 1, \ldots, p$, from the following model:

$$\begin{aligned} Y_{ij} &= \mathcal{X}_i(u_j) + \eta_{ij} + \epsilon_{ji} \\ &= \sum_{k=1}^{13} c_{ik}\phi_k(u_j) + \sum_{k=1}^{4} f_{ik}a_{kj} + \epsilon_{ji}, \end{aligned}$$

where $\phi_k(u)$ are chosen as B-spline basis functions of order four and the smoothing coefficients c_{ik} are generated from $\mathcal{N}(0, 1.5^2)$. The factor loadings a_{kj} follow $\mathcal{N}(0, 0.6^2)$, and the factors $(f_{i1}, f_{i2}, f_{i3}, f_{i4})^\top \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a 4×4 covariance matrix. We set the multivariate mean term $\boldsymbol{\mu} = \mathbf{0}$ and the variance $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_4$. We adjust the value of σ^2 to control the signal-to-noise ratio. When σ^2 is large, the signal-to-noise level is low, and when σ^2 is small, the level is high. The random error terms ϵ_{ji} follow $\mathcal{N}(0, 0.5^2)$.

Setting 2. We generate simulated data Y_{ij}, where $i = 1, \ldots, n$ and $j = 1, \ldots, p$, from the following model:

$$\begin{aligned} Y_{ij} &= \mathcal{X}_i(u_j) + \eta_{ij} + \epsilon_{ji} \\ &= \sum_{k=1}^{5} c_{ik}\phi_k(u_j) + \sum_{k=1}^{4} f_{ik}a_{kj} + \epsilon_{ji}, \end{aligned}$$

where $\phi_k(u)$ are chosen as B-spline basis functions of order four and the smoothing coefficients c_{ik} are generated from $\mathcal{N}(0, \gamma_k^2)$, where $\gamma_1 = 3, \gamma_2 = 2.5, \gamma_3 = 2, \gamma_4 = 1.5$, and $\gamma_5 = 1$. The factor loadings a_{kj} follow $\mathcal{N}(0, 0.8^2)$ and the factors $(f_{i1}, f_{i2}, f_{i3}, f_{i4})^\top \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a 4 by 4 covariance matrix. We set the multivariate mean term $\boldsymbol{\mu} = \mathbf{0}$ and variance $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_4$. We adjust the value of σ^2 to control the signal-to-noise ratio. When σ^2 is large, the signal-to-noise level is low, and when σ^2 is small, the signal-to-noise level is high. The random error terms ϵ_{ji} follow $\mathcal{N}(0, 0.5^2)$.

Setting 3. We generate simulated data as in Setting 1, except that we use

$K = 21$ B-spline basis functions. We use $n = 40$ and $p = 50$. Here, we show how the model works when the number of basis functions is large compared with n and p.

5.2. Results

We repeat the simulation setup 500 times and obtain the estimated smooth function $\widehat{\mathcal{X}}_i(u) = \widehat{\boldsymbol{c}}_i^\top \boldsymbol{\Phi}(u)$. The mean integrated squared error (MISE) for the function estimation is calculated as

$$\text{MISE} = \frac{1}{n}\sum_{i=1}^{n} \int \left\{ \mathcal{X}_i(u) - \widehat{\mathcal{X}}_i(u) \right\}^2 du. \tag{5.1}$$

The results for Settings 1 and 2 are reported in Tables 1 and 2, respectively. With the same sample size n, increasing the number of points p on the curve decreases the estimation error. However, with the same value for p, increasing the sample size does not decrease the estimation error. This is consistent with the convergence rate stated in Section 4, where the estimator converges with the rate related to p.

When σ is large, that is, the signal-to-noise ratio is high, the FASM outperforms the smoothing model. When $\sigma = 0$, the actual data are generated without the factor model component. The performance of the two methods is identical. This implies that the proposed model is robust in that it improves the estimation when there is a measurement error, and is no worse than the ordinary smoothing model when there is no measurement error. This is credited to the selection criterion used in (S3.4), which always selects zero factors when there are no common factors.

In Table 3, we show the MISE result from Setting 3. Again, the proposed model shows its superiority when K is large compared with n and p. This setting corresponds to allowing the number of basis functions K to go to infinity, discussed in Section 4.

6. Application to Climatology

Here, we apply the aforementioned FASM, Bsmooth, Localin, and PACE methods to real data sets. This section examines the Friday temperature data at Adelaide airport. We provide analysis of Canadian annual temperature and precipitation data in the Supplementary Material.

We choose Adelaide because it tends to have the hottest temperature among Australia's big cities. Data from other weekdays exhibit similar features, and thus are not shown here. The data are measured every half an hour for the period 1997 to 2007. The sample size n is 508, and the number of discrete data points p from each curve is 48. A plot of the raw data is shown in Figure 4a. The data are quite noisy, with extreme values in some of the curves due to large

Table 1. Setting 1: The MISE of the function estimates with different samples sizes and dimensions. The σ^2-value is used to control the signal-to-noise ratio.

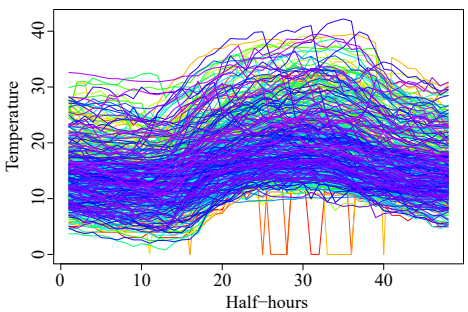
			MISE			
Sample Size	Dimension	Size of η	FASM	Bsmooth	Localin	PACE
$n = 20$	$p = 51$	$\sigma = 0$	**0.058**	**0.058**	0.075	0.238
		$\sigma = 0.5$	**0.131**	**0.131**	0.149	0.587
		$\sigma = 0.75$	**0.204**	0.208	0.226	1.026
		$\sigma = 1$	**0.297**	0.318	0.336	1.676
$n = 20$	$p = 101$	$\sigma = 0$	**0.031**	**0.031**	0.043	0.238
		$\sigma = 0.5$	**0.076**	**0.076**	0.089	0.595
		$\sigma = 0.75$	**0.115**	0.123	0.136	1.020
		$\sigma = 1$	**0.159**	0.187	0.201	1.686
$n = 50$	$p = 51$	$\sigma = 0$	**0.058**	**0.058**	0.076	0.195
		$\sigma = 0.5$	**0.131**	0.135	0.153	0.543
		$\sigma = 0.75$	**0.186**	0.214	0.233	1.013
		$\sigma = 1$	**0.248**	0.314	0.331	1.668
$n = 100$	$p = 101$	$\sigma = 0$	**0.031**	**0.031**	0.044	0.170
		$\sigma = 0.5$	**0.062**	0.074	0.089	0.463
		$\sigma = 0.75$	**0.090**	0.118	0.132	0.908
		$\sigma = 1$	**0.133**	0.181	0.194	1.570

Table 2. Setting 2: The MISE of the function estimates with different samples sizes and dimensions. The σ^2-value is used to control the signal-to-noise ratio.

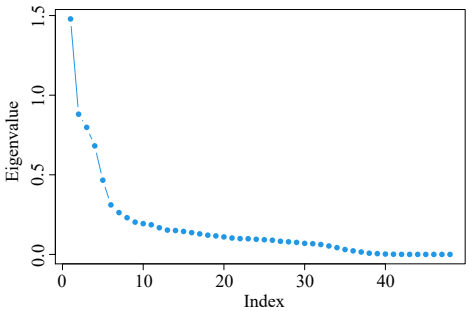
			MISE			
Sample Size	Dimension	Size of η	FASM	Bsmooth	Localin	PACE
$n = 20$	$p = 51$	$\sigma = 0$	**0.024**	**0.024**	0.070	0.116
		$\sigma = 0.5$	0.086	**0.085**	0.195	0.621
		$\sigma = 0.75$	**0.160**	**0.160**	0.340	1.431
		$\sigma = 1$	**0.272**	0.280	0.538	2.685
$n = 20$	$p = 101$	$\sigma = 0$	**0.012**	**0.012**	0.040	0.089
		$\sigma = 0.5$	0.047	**0.046**	0.117	0.660
		$\sigma = 0.75$	**0.079**	**0.079**	0.192	1.434
		$\sigma = 1$	**0.135**	0.138	0.301	2.549
$n = 50$	$p = 51$	$\sigma = 0$	**0.025**	**0.025**	0.070	0.076
		$\sigma = 0.5$	0.087	**0.086**	0.202	0.584
		$\sigma = 0.75$	**0.162**	0.169	0.348	1.435
		$\sigma = 1$	**0.250**	0.269	0.522	2.557
$n = 100$	$p = 101$	$\sigma = 0$	**0.012**	**0.012**	0.041	0.033
		$\sigma = 0.5$	**0.045**	**0.045**	0.114	0.518
		$\sigma = 0.75$	**0.085**	**0.085**	0.195	1.337
		$\sigma = 1$	**0.138**	0.139	0.299	2.618

Table 3. Setting 3: The MISE of the function estimates when K is large compared with n and p. The σ^2-value is used to control the signal-to-noise ratio.

			MISE			
Sample Size	Dimension	Size of η	FASM	Bsmooth	Localin	PACE
$n = 50$	$p = 40$	$\sigma = 0$	**0.097**	**0.097**	0.121	0.225
		$\sigma = 0.5$	**0.206**	0.207	0.235	0.590
		$\sigma = 0.75$	**0.291**	0.309	0.335	1.044
		$\sigma = 1$	**0.390**	0.457	0.466	1.747



(a) Raw temperature data



(b) Eigenvalues of the residuals in from Bsmooth

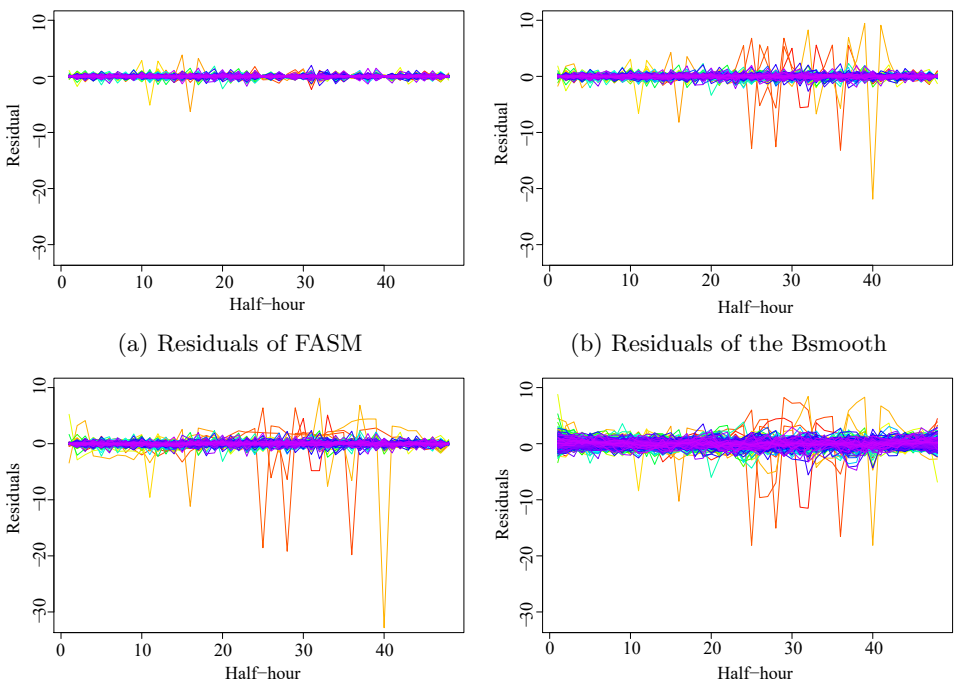
Figure 4. Half-hourly temperature data on Fridays at Adelaide airport.

measurement errors.

We use B-spline basis functions of order four with knots at every data point. We show the residuals from the four models in Figure 5. Apart from the FASM, the other models produce large residuals, with some extreme values. To examine the residual struture from Bsmooth model, we conduct a principal component analysis on the residuals. Figure 4b shows the screeplot of the eigenvalues. The residuals from Bsmooth model exhibit a "spike" structure, where the first few eigenvalues are significantly larger than the rest. This means the residuals contain information that can be captured by only a few factors, which calls for a further dimension reduction model on the residuals.

7. Conclusion

In this paper, we propose an FASM for functional data. We study raw functional data, a mixture of functional curves, and high-dimensional errors. When the measurement error is informative, a smoothing model alone is inadequate for capturing data variation and recovering the signal functional component. The proposed model incorporates a factor structure into the smoothing model to further explain the large residuals. We propose a numerical iteration approach to obtain estimates in the smoothing and the factor models



(a) Residuals of FASM

(b) Residuals of the Bsmooth

(c) Residuals of Localin

(d) Residuals of PACE

Figure 5. Half-hourly temperature data on Friday at Adelaide airport.

simultaneously. The asymptotic distribution of the estimators is given, with a proof. Our model also serves as a dimension-reduction method for functional and high-dimensional mixture data, easing the path to making inferences. We provide an example in which we construct a covariance estimator for the raw data. Furthermore, we show that the model can be applied when there is a misidentification in the data structure, two examples of which are the misspecification of the smoothing basis functions and neglecting the step jumps in the mean level of the functions. The advantages of the proposed model are demonstrated using simulation studies. We also show how our model performs by applying it to Australian temperature data.

The proposed model is a good starting point for modeling complex data structures. We deal with a mixture of smooth functional curves and high-dimensional measurement errors. The factor model component can be regarded as a "boosting component" that improves model accuracy. Extending this idea, the model can be applied to other data structures, such data that contain change points. The change point is a popular problem in many statistics and econometric topics, and has been studied extensively in multivariate settings. Previous works on change points in functional data include those of Berkes et al. (2009), Hörmann and Kokoszka (2010), and Hörmann, Kidziński and Hallin (2015). In our simulations, we show examples that our model can be used to

model functional data with change points in the cross-sectional direction. The model can be modified to account for change points in the sample direction. This is left to future research.

Supplementary Material

In the online Supplementary Material, we extend our model to nonparametric settings, introduce a covariance estimator, and present extensive simulated and real-data analysis. Here, we also provide proofs for the theorems and additional results.

Acknowledgments

We acknowledge the research funding ARC DP230102250 funded by Australian Research Council.

References

Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica* **81**, 1203–1227.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**, 716–723.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70**, 191–221.

Berkes, I., Gabrys, R., Horváth, L. and Kokoszka, P. (2009). Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **71**, 927–946.

Cai, T. T. and Yuan, M. (2011). Optimal estimation of the mean function based on discretely sampled functional data: Phase transition. *The Annals of Statistics* **39**, 2330–2355.

Cuevas, A. (2014). A partial overview of the theory of statistics with functional data. *Journal of Statistical Planning and Inference* **147**, 1–23.

Eubank, R. L. (1999). *Nonparametric Regression and Spline Smoothing*. 2nd Edition. Marcel Dekker, New York.

Fan, J., Fan, Y. and Lv, J. (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics* **147**, 186–197.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall, London.

Febrero-Bande, M., Galeano, P. and González-Manteiga, W. (2017). Functional principal component regression and functional partial least-squares regression: An overview and a comparative study. *International Statistical Review* **85**, 61–83.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Science & Business Media, New York.

Goia, A. and Vieu, P. (2016). An introduction to recent advances in high/infinite dimensional statistics. *Journal of Multivariate Analysis* **146**, 1–6.

Golub, G. H., Heath, M. and Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics* **21**, 215–223.

Green, P. J. and Silverman, B. W. (1999). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Chapman & Hall, London.

Hörmann, S., Kidziński, L. and Hallin, M. (2015). Dynamic functional principal components. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **77**, 319–348.

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *The Annals of Statistics* **38**, 1845–1884.

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer Science & Business Media, New York.

Jiang, B., Yang, Y., Gao, J. and Hsiao, C. (2021). Recursive estimation in large panel data models: Theory and practice. *Journal of Econometrics* **224**, 439–465.

Knight, K. and Fu, W. (2000). Asymptotics for Lasso-type estimators. *The Annals of Statistics* **28**, 1356–1378.

Lam, C., Yao, Q. and Bathia, N. (2011). Estimation of latent factors for high-dimensional time series. *Biometrika* **98**, 901–918.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics* **92**, 1004–1016.

Onatski, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics* **168**, 244–258.

Ramsay, J. O. and Hooker, G. (2017). *Dynamic Data Analysis: Modeling Data with Differential Equations*. Springer, New York.

Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis*. Springer, New York.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer, New York.

Reiss, P. T., Goldsmith, J., Shang, H. L. and Ogden, R. T. (2017). Methods for scalar-on-function regression. *International Statistical Review* **85**, 228–249.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* **6**, 461–464.

Wahba, G. (1990). *Spline Models for Observational Data*. SIAM.

Wahba, G. and Word, S. (1975). Periodic splines for spectral density estimation: The use of cross validation for determining the degree of smoothing. *Communications in Statistics: Theory and Methods* **4**, 125–141.

Wand, M. P. and Jones, C. M. (1995). *Kernel Smoothing*. Chapman & Hall, Boca Raton.

Wang, J.-L., Chiou, J.-M. and Müller, H.-G. (2016). Functional data analysis. *Annual Review of Statistics and Its Application* **3**, 257–295.

Yao, F., Müller, H. and Wang, J. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100**, 577–590.

Yao, W. and Li, R. (2013). New local estimation procedure for a non-parametric regression function for longitudinal data. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **75**, 123–138.

Zhang, X. and Wang, J. L. (2016). From sparse to dense functional data and beyond. *The Annals of Statistics* **44**, 2281–2321.

Yuan Gao

Research School of Finance, Actuarial Studies and Statistics, Australian National University, Canberra, ACT 0200, Australia.

E-mail: yuan.gao@anu.edu.au

Han Lin Shang

Department of Actuarial Studies and Business Analytics, Macquarie University, Sydney, NSW 2109, Australia.

E-mail: hanlin.shang@mq.edu.au

Yanrong Yang

Research School of Finance, Actuarial Studies and Statistics, Australian National University, Canberra, ACT 0200, Australia.

E-mail: yanrong.yang@anu.edu.au

(Received June 2021; accepted April 2022)