Statistica Sinica Preprint No: SS-2024-0028	
Title	Functional Linear Models with Latent Factors
Manuscript ID	SS-2024-0028
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202024.0028
Complete List of Authors	Zixuan Han, Tao Li, Jinhong You and Jiguo Cao
Corresponding Authors	Jiguo Cao
E-mails	Jiguo_cao@sfu.ca
Notice: Accepted author version.	

Statistica Sinica

Functional Linear Models with Latent Factors

Zixuan Han, Tao Li, Jinhong You

Department of Statistics and Management

Shanghai University of Finance and Economics

Jiguo Cao [1]

Department of Statistics and Actuarial Science

Simon Fraser University

Abstract: We propose a novel functional linear model incorporating latent factors, where scalar response, scalar covariates, and functional covariates have repeated measurements for each subject. Our model accounts for latent factors that may impact the response but remain unobservable. To unveil and estimate these latent factors, we propose an iterated profile estimation method. We then establish the consistency and asymptotic properties of the estimators. To demonstrate the efficacy of our proposed estimation procedure, we conduct simulation studies across various scenarios. We compare our results with estimations derived from conventional functional linear models, revealing the superior performance of our method in addressing latent factors. We further illustrate our proposed model and methodology by analyzing real data from both financial markets and air pollution datasets. In these analyses, we successfully uncover hidden factors that exert influence in these specific fields.

Key words and phrases: Functional Data Analysis; Factor Model; Penalized Spline; Functional Regression; Profile Estimation.

[1]Corresponding author. Email: jiguo_cao@sfu.ca

1. Introduction

Functional data refers to a type of data that consists of observations or measurements that are functions rather than simple numerical values or categorical variables (Ramsay, 1982). In other words, instead of representing data points as individual values or discrete categories, functional data captures information in the form of entire curves, trajectories, or continuous functions. Functional data analysis (FDA) is a statistical framework and set of techniques used to analyze and interpret functional data (Ramsay and Silverman (2005); Ferraty and Vieu (2006)). It involves treating functions as the fundamental unit of analysis and applying statistical methods specifically designed for functional data. Functional regression is one of the most popular FDA tools (Chen et al. (2021); Liu et al. (2022)). A great quantity of literature focused on this field can be divided into three categories based on whether the response variable is functional data (Jiang and Wang, 2011; Cai et al., 2022a) , the prediction variable is functional data (Hilgert et al., 2013; Xun et al., 2022), or both are functional data (Li et al., 2017; Cai et al., 2022b). Functional linear models can be used to model the relationship between functional variables and scalar response, which have been applied to many fields as financial market, biomedical science, climate and environment (Lin et al. (2017); Zhu et al. (2019); Guan et al. (2020);

Chen et al. (2022)).

Early research literature extensively explored estimation methods for functional linear models. For instance, Crambes et al. (2009) proposed a smoothing splines estimator for the functional slope parameter based on a modified roughness penalty. Yuan and Cai (2010) studied a smoothness regularization method for functional linear regression and constructed a unified treatment for both the prediction and estimation problems using a reproducing kernel Hilbert space approach. Wu et al. (2010) proposed a varying-coefficient approach where the slope parameter was regarded as a function of scalar covariates. Morris (2015) well summarized some existing forms and generalizations of functional linear models.

In recent years, research on functional linear models has been extended to more complex data structure with better interpretability. For example, Lin et al. (2017) presented a locally sparse estimator for functional coefficient based on a functional regularization technique called fSCAD. Liu et al. (2017) proposed a new functional linear mixed model where each subject shared a common slope function (fixed effect) in addition to an individual one (random effects). Liu et al. (2018) proposed a functional variable selection method via Gram–Schmidt (FGS) orthogonalization for models with multiple functional predictors. Fang et al. (2020) considered

a multiple-smooth and locally sparse (m-SLoS) estimator for functional coefficients based on the interconnections among multiple responses. Chen et al. (2022) considered the serial dependence of the functional predictors and proposed an autocovariance-based generalized method-of-moments estimator for the slope function.

The previous studies did not take into account hidden factors that could potentially influence the response as well. Ignoring the existence of hidden factors can lead to certain deviations in real data analysis. Take one stock data for example. The trading data is obtained from the China Securities Index (CSI) 300 constituent stocks from August 3rd to October 10th, 2020, with the stock price recorded every 3 seconds. The scalar response is the daily returns of stocks, defined as the change of the closing price in comparison with the opening price of the individual stock at the same day. The functional covariate is the observed individual stock prices every 3 seconds in one trade day. In addition to daily stock price trajectories, some hidden factors may exist that affect stock returns to varying degrees. Specifically, from a macro perspective, hidden factors such as macroeconomic situation, political factors, changes in the whole financial market supply and demand may bring a common shock to the return of all individual stocks. On the other hand, from a micro perspective, subjective conditions like business

performance, enterprise distribution policy, and characteristics of the industry the enterprise located may also have varying degrees of impact. These unobersevable factors cannot be expressed numerically and the influence may vary with individual stocks over time. Moreover, the individual effect and time effect of these factors may also interact multiplicatively with the response variable. The factor structure can depict the correlation among the 300 individual stocks, and integrate economic elements that are not measurable at the macro or micro level. Our objective is to identify these hidden factors and quantify their influence on the stock return.

The flourishing development of statistical theory and practical applications has made great progress in large-scale optimization and dimensional reduction techniques, and factor models is one of the core modeling methods. Fan et al. (2021) provided a selective overview of the recent developments and applications of factor models. Pesaran (2006) proposed a common correlated effect (CCE) approach, Bai (2009) developed the iterated principal component (IPC) approach, and Bai and Li (2014) studied the likelihood approach. With the popularization of potential factors in data analysis, model research and statistical analysis has gradually been promoted (Fan et al. (2011); Bai and Li (2014); Feng et al. (2018)).

In the modeling process of functional data, the existence and impact

of potential factors are rarely considered. In this article, our objective is to introduce a novel functional linear model that incorporates latent factors and functional covariates specifically designed for functional data. Specifically, we propose the following Functional Linear Models with Latent Factors (FLiF):

$$Y_{it} = \boldsymbol{\alpha}^{\tau}\boldsymbol{W}_{it} + \int_S \boldsymbol{\beta}^{\tau}(s)\boldsymbol{X}_{it}(s)ds + \boldsymbol{\lambda}_i^{\tau}\boldsymbol{F}_t + \varepsilon_{it}, \tag{1.1}$$

where Y_{it} is a scalar response for the i-th subject at time t, $i = 1, \ldots, N, t = 1, \ldots, T$, $\boldsymbol{W}_{it}$ is a $p \times 1$ vector of scalar covariates, $\boldsymbol{\alpha}$ is a $p \times 1$ vector of unknown coefficients, $\boldsymbol{X}_{it}(s) = (X_{it1}(s), \cdots, X_{itq}(s))^{\tau}$ is a $q \times 1$ vector of functional predictors, $\boldsymbol{\beta}(s) = (\beta_1(s), \cdots, \beta_q(s))^{\tau}$ is the corresponding vector of slope functions, representing the cumulative effect of $\boldsymbol{X}_{it}(s)$ on Y_{it}, S is assumed to be a compact subset of an Euclidean space such as $[0, 1]$, $\boldsymbol{F}_t$ and $\boldsymbol{\lambda}_i$ denote an $r \times 1$ vector of common factors and factor loadings for the ith subject, respectively, and ε_{it} are mutually independent and identically distributed (i.i.d.) random errors with mean 0 and variance σ^2.

The above FLiF model can be regarded as a generalized form of several conventional models. For example, if no functional covariate can be observed, i.e., $\boldsymbol{\beta}(\cdot) = 0$, and the random error is serial and cross-sectional

dependent, then the model becomes the panel data model with interactive effects studies in Bai (2009). If no underlying factors exist, the model is the conventional functional linear model in Ramsay and Silverman (2005). If the factors is assumed known, then the model becomes the functional linear mixed-effects model proposed by Liu et al. (2017). Detailed evaluation and comparison between our proposed FLiF model and the functional linear mixed-effects model can be found in the supplementary materials. Moreover, if there are no covariates including both scalar and functional types in the model, that is, without functional coefficients, the model becomes the factor-augmented smoothing model (FASM) proposed by Gao et al. (2022).

Our article make three main contributions. First, to the best of our knowledge, this is the first attempt to model functional data with latent factors. Second, we develop a profile method to estimate the slope functions $\boldsymbol{\beta}(s)$ using penalized B-splines and to identify the underlying factor structure $\boldsymbol{F}_t$ and $\boldsymbol{\lambda}_i$ using principal component analysis. Last but not least, we establish the theoretical results including the convergence rate and asymptotic normality for both scalar and functional coefficient estimators for the proposed FLiF model. We conduct numerical simulation studies under various settings of correlation between predictors and response, and compare with the conventional models. Moreover, we demonstrate our FLiF model

with applications in financial marketing and air pollution. We provide the R codes for the simulation studies and real data analysis in the website `https://github.com/statszx/FLiF`.

The paper is organized as follows. In Section 2, we introduce the model and estimation procedure for both scalar and function coefficients, as well as the factor structure. Meanwhile, the identification of the factor numbers and the optimal choice of the smoothing parameter are also provided. In Section 3, we establish the asymptotic theory of the resulting estimators and the convergency of the factor determination method under assumptions. In Section 4, numerical studies are worked to prove the efficiency of the proposed estimation method under different cases. In Section 5, two real data are analyzed to verify the application of the model and explain the unobserved potential factors regarding the data. The discussion is concluded in Section 6. Additional simulation results and technical details including auxiliary theoretical results are provided in Supplementary material.

2. Estimating the FLiF Model

In this section, we propose the estimation method for the FLiF model, as well as the implementation issues in the application to determine the optimal factor number and identify the optimal smoothing parameter.

2.1 Iterated Profile Estimation Method

We first propose a iterated profile estimation method for the FLiF model (1.1), which can be expressed in vector and matrix notations as follows

$$\boldsymbol{Y}_i = \boldsymbol{W}_i\boldsymbol{\alpha} + \int_S \boldsymbol{X}_i(s)\boldsymbol{\beta}(s)ds + \boldsymbol{F}\boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i, \qquad i = 1, \ldots, N, \tag{2.2}$$

where $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{iT})'$, $\boldsymbol{W}_i = (\boldsymbol{W}_{i1}, \ldots, \boldsymbol{W}_{iT})'$, $\boldsymbol{X}_i(s) = (\boldsymbol{X}_{i1}(s), \ldots, \boldsymbol{X}_{iT}(s))'$, $\boldsymbol{F} = (\boldsymbol{F}_1, \ldots, \boldsymbol{F}_T)'$, $\boldsymbol{\varepsilon}_i = (\boldsymbol{\varepsilon}_{i1}, \ldots, \boldsymbol{\varepsilon}_{iT})'$. We denote $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_N)^\tau$. For the identification of the FLiF model (2.2), the following two restrictions are added (Bai, 2009): $\boldsymbol{F}^\tau\boldsymbol{F}/T = \boldsymbol{I}_r$, and $\boldsymbol{\Lambda}^\tau\boldsymbol{\Lambda}$ is a diagonal matrix.

Before conducting the estimation procedure for model (2.2), we first represent $\boldsymbol{\beta}(s)$ by B-spline approximation. For each $j = 1, \cdots, q$, let $\beta_j(s)$ be the linear combination of B-spline basis function as

$$\beta_j(s) \approx \sum_{l=1}^{L} \gamma_{jl} b_l(s) = \boldsymbol{\gamma_j^\tau}\boldsymbol{b}(s), \tag{2.3}$$

where $\boldsymbol{b}(s) = (b_1(s), \cdots, b_L(s))^\tau$ is the vector of basis functions of order $d+1$ with K interior knots, where $L = K + d + 1$, and $\boldsymbol{\gamma}_j = (\gamma_{j1}, \cdots, \gamma_{jL})^\tau$ is the corresponding vector of basis coefficients. For simplicity, we assume the same basis functions are used to represent $\beta_j(s)$.

Let $\boldsymbol{B}_i = \int \boldsymbol{X}_i(s) \otimes \boldsymbol{b}^\tau(s) ds$, and $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^\tau, \ldots, \boldsymbol{\gamma}_q^\tau)^\tau$. Plug (2.3) into (2.2), we have the approximation of (2.2) as $\boldsymbol{Y}_i \approx \boldsymbol{W}_i\boldsymbol{\alpha} + \boldsymbol{B}_i\boldsymbol{\gamma} + \boldsymbol{F}\boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i$. Let $\boldsymbol{Z}_i = (\boldsymbol{W}_i, \boldsymbol{B}_i)$, $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\tau, \boldsymbol{\gamma}^\tau)^\tau$, then the approximation can be expressed as $\boldsymbol{Y}_i \approx \boldsymbol{Z}_i\boldsymbol{\theta} + \boldsymbol{F}\boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i$. We define the objective function as

$$Q(\boldsymbol{\theta}, \boldsymbol{\Lambda}, \boldsymbol{F}) = \sum_{i=1}^{N} (\boldsymbol{Y}_i - \boldsymbol{Z}_i\boldsymbol{\theta} - \boldsymbol{F}\boldsymbol{\lambda}_i)^\tau (\boldsymbol{Y}_i - \boldsymbol{Z}_i\boldsymbol{\theta} - \boldsymbol{F}\boldsymbol{\lambda}_i) + \sum_{j=1}^{q} \xi_j \int_S \left(\frac{d^2\beta_j(s)}{ds^2} \right)^2 ds, \tag{2.4}$$

subject to the identification restriction $\boldsymbol{F}^\tau\boldsymbol{F}/T = \boldsymbol{I}_r$, and $\boldsymbol{\Lambda}^\tau\boldsymbol{\Lambda}$ is a diagonal matrix. The second term in (2.4) is a roughness penalty in which ξ_j, $j = 1, \ldots, q$, is the smoothing parameter. We define $\boldsymbol{G}_{\boldsymbol{\beta}} = \mathtt{diag}(\xi_1, \ldots, \xi_q) \otimes \int_S \boldsymbol{b}^{(2)}(s)(\boldsymbol{b}^{(2)}(s))^\tau ds$, and $\boldsymbol{G}_{\boldsymbol{\theta}} = \mathtt{diag}(\boldsymbol{0}_{\boldsymbol{p}\times\boldsymbol{p}}, \boldsymbol{G}_{\boldsymbol{\beta}})$. Then the objective function (2.4) can be expressed as

$$Q(\boldsymbol{\theta}, \boldsymbol{\Lambda}, \boldsymbol{F}) = \sum_{i=1}^{N} (\boldsymbol{Y}_i - \boldsymbol{Z}_i\boldsymbol{\theta} - \boldsymbol{F}\boldsymbol{\lambda}_i)^\tau (\boldsymbol{Y}_i - \boldsymbol{Z}_i\boldsymbol{\theta} - \boldsymbol{F}\boldsymbol{\lambda}_i) + \boldsymbol{\theta}^\tau\boldsymbol{G}_{\boldsymbol{\theta}}\boldsymbol{\theta}. \tag{2.5}$$

Define the projection matrix $\boldsymbol{M}_{\boldsymbol{F}} = \boldsymbol{I}_T - \boldsymbol{F}(\boldsymbol{F}^\tau\boldsymbol{F})^{-1}\boldsymbol{F}^\tau = \boldsymbol{I}_T - \boldsymbol{F}\boldsymbol{F}^\tau/T$. For any given $\boldsymbol{\theta}$ and $\boldsymbol{F}$, we can obtain the profile estimate for $\boldsymbol{\lambda}_i$ by solving the score equation as $\widehat{\boldsymbol{\lambda}}_i(\boldsymbol{\theta}, \boldsymbol{F}) = \frac{1}{T}\boldsymbol{F}^\tau(\boldsymbol{Y}_i - \boldsymbol{Z}_i\boldsymbol{\theta})$. Then substituting $\widehat{\boldsymbol{\lambda}}_i(\boldsymbol{\theta}, \boldsymbol{F})$ into (2.5), the objective function becomes $Q(\boldsymbol{\theta}, \boldsymbol{F}) = \sum_{i=1}^{N} (\boldsymbol{Y}_i - \boldsymbol{Z}_i\boldsymbol{\theta})^\tau \boldsymbol{M}_{\boldsymbol{F}} (\boldsymbol{Y}_i - \boldsymbol{Z}_i\boldsymbol{\theta}) + \boldsymbol{\theta}^\tau\boldsymbol{G}_{\boldsymbol{\theta}}\boldsymbol{\theta}$.

Therefore, following Bai (2009), the least square estimators of $\boldsymbol{\theta}$ and $\boldsymbol{F}$ can be obtained by minimizing $Q(\boldsymbol{\theta}, \boldsymbol{F})$ with an iteration algorithm as

$$\hat{\boldsymbol{\theta}} = \left(\sum_{i=1}^{N} \boldsymbol{Z}_i^\tau \boldsymbol{M}_{\hat{\boldsymbol{F}}} \boldsymbol{Z}_i + \boldsymbol{G}_{\boldsymbol{\theta}} \right)^{-1} \sum_{i=1}^{N} \boldsymbol{Z}_i^\tau \boldsymbol{M}_{\hat{\boldsymbol{F}}} \boldsymbol{Y}_i, \tag{2.6}$$
$$\left[\frac{1}{NT} \sum_{i=1}^{N} (\boldsymbol{Y}_i - \boldsymbol{Z}_i \hat{\boldsymbol{\theta}})(\boldsymbol{Y}_i - \boldsymbol{Z}_i \hat{\boldsymbol{\theta}})^\tau \right] \hat{\boldsymbol{F}} = \hat{\boldsymbol{F}} \boldsymbol{V}_{NT},$$

where $\boldsymbol{V}_{NT}$ is the diagonal matrix of the r largest eigenvalues, arranged in decreasing order, of matrix $\sum_{i=1}^{N} (\boldsymbol{Y}_i - \boldsymbol{Z}_i \hat{\boldsymbol{\theta}})(\boldsymbol{Y}_i - \boldsymbol{Z}_i \hat{\boldsymbol{\theta}})^\tau / (NT)$, and $\hat{\boldsymbol{F}}$ is the matrix composed with the corresponding first r orthogonal eigenvectors (multiplied by $\sqrt{T}$) associated with these r largest eigenvalues. Finally, the factor loadings can be estimated by $\hat{\boldsymbol{\Lambda}} = (\hat{\boldsymbol{\lambda}}_1, \cdots, \hat{\boldsymbol{\lambda}}_N)^\tau$ with $\hat{\boldsymbol{\lambda}}_i = \frac{1}{T} \hat{\boldsymbol{F}}^\tau (\boldsymbol{Y}_i - \boldsymbol{Z}_i \hat{\boldsymbol{\theta}})$.

Because $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\tau, \boldsymbol{\gamma}^\tau)^\tau$, by decomposing (2.6), we have

$$\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\alpha}}^\tau, \hat{\boldsymbol{\gamma}}^\tau)^\tau = \begin{pmatrix} \sum_{i=1}^{N} \boldsymbol{W}_i^\tau \boldsymbol{M}_{\hat{\boldsymbol{F}}} \boldsymbol{W}_i & \sum_{i=1}^{N} \boldsymbol{W}_i^\tau \boldsymbol{M}_{\hat{\boldsymbol{F}}} \boldsymbol{B}_i \\ \sum_{i=1}^{N} \boldsymbol{B}_i^\tau \boldsymbol{M}_{\hat{\boldsymbol{F}}} \boldsymbol{W}_i & \sum_{i=1}^{N} \boldsymbol{B}_i^\tau \boldsymbol{M}_{\hat{\boldsymbol{F}}} \boldsymbol{B}_i + \boldsymbol{G}_{\boldsymbol{\theta}} \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^{N} \boldsymbol{W}_i^\tau \boldsymbol{M}_{\hat{\boldsymbol{F}}} \boldsymbol{Y}_i \\ \sum_{i=1}^{N} \boldsymbol{B}_i^\tau \boldsymbol{M}_{\hat{\boldsymbol{F}}} \boldsymbol{Y}_i \end{pmatrix}.$$

For simplicity, we define $\tilde{\boldsymbol{W}}_i = \boldsymbol{M}_{\hat{\boldsymbol{F}}} \boldsymbol{W}_i$, $\tilde{\boldsymbol{B}}_i = \boldsymbol{M}_{\hat{\boldsymbol{F}}} \boldsymbol{B}_i$, and $\tilde{\boldsymbol{Y}}_i = \boldsymbol{M}_{\hat{\boldsymbol{F}}} \boldsymbol{Y}_i$. Meanwhile, denote $\tilde{\boldsymbol{W}} = (\tilde{\boldsymbol{W}}_1^\tau, \cdots, \tilde{\boldsymbol{W}}_N^\tau)^\tau$, $\tilde{\boldsymbol{B}} = (\tilde{\boldsymbol{B}}_1^\tau, \cdots, \tilde{\boldsymbol{B}}_N^\tau)^\tau$, and $\tilde{\boldsymbol{Y}} = (\tilde{\boldsymbol{Y}}_1^\tau, \cdots, \tilde{\boldsymbol{Y}}_N^\tau)^\tau$. We can obtain the explicit expression of $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\gamma}}$,

respectively, as $\hat{\boldsymbol{\alpha}} = (\tilde{\boldsymbol{W}}^{\tau}\tilde{\boldsymbol{W}})^{-1}\tilde{\boldsymbol{W}}^{\tau}\left(\boldsymbol{I}-\tilde{\boldsymbol{B}}(\tilde{\boldsymbol{B}}^{\tau}\boldsymbol{M}_{\tilde{\boldsymbol{W}}}\tilde{\boldsymbol{B}}+\boldsymbol{G}_{\boldsymbol{\theta}})^{-1}\tilde{\boldsymbol{B}}^{\tau}\boldsymbol{M}_{\tilde{\boldsymbol{W}}}\right)\tilde{\boldsymbol{Y}}$, $\hat{\gamma} = (\tilde{\boldsymbol{B}}^{\tau}\boldsymbol{M}_{\tilde{\boldsymbol{W}}}\tilde{\boldsymbol{B}} + \boldsymbol{G}_{\boldsymbol{\theta}})^{-1}\tilde{\boldsymbol{B}}^{\tau}\boldsymbol{M}_{\tilde{\boldsymbol{W}}}\tilde{\boldsymbol{Y}}$, where $\boldsymbol{M}_{\tilde{\boldsymbol{W}}}$ is the projection matrix of $\tilde{\boldsymbol{W}}$ defined by $\boldsymbol{M}_{\tilde{\boldsymbol{W}}} = \boldsymbol{I}_{NT} - \tilde{\boldsymbol{W}}(\tilde{\boldsymbol{W}}^{\tau}\tilde{\boldsymbol{W}})^{-1}\tilde{\boldsymbol{W}}^{\tau}$. The expression of $\hat{\boldsymbol{\alpha}}$ can be simplified as $\hat{\boldsymbol{\alpha}} = \left(\sum_{i=1}^{N}\boldsymbol{W}_{\boldsymbol{i}}^{\tau}\boldsymbol{M}_{\hat{\boldsymbol{F}}}\boldsymbol{W}_{\boldsymbol{i}}\right)^{-1}\sum_{i=1}^{N}\boldsymbol{W}_{\boldsymbol{i}}^{\tau}\boldsymbol{M}_{\hat{\boldsymbol{F}}}(\boldsymbol{Y}_{\boldsymbol{i}} - \boldsymbol{B}_{\boldsymbol{i}}\hat{\boldsymbol{\gamma}})$. The corresponding estimator for $\beta_j(s)$ is $\hat{\beta}_j(s) = \hat{\boldsymbol{\gamma}}_j\boldsymbol{b}(s), \quad 1 \le j \le q$.

2.2 Determination of the Number of Factors

In the aforementioned estimation procedure, it is assumed that the number of factors r and smoothing parameter $\boldsymbol{\xi} = (\xi_1, \cdots, \xi_q)^{\tau}$ are given. But in empirical applications, they are usually unknown and should be determined by data. We derive an information criterion to estimate the number of factors, r as follows. Assume that the factor number is bounded by a finite integer r_{max}. Let $\dot{\boldsymbol{\theta}}_r, \dot{\boldsymbol{F}}_r$, and $\dot{\boldsymbol{\Lambda}}_r$ denote the least square estimator of $\boldsymbol{\theta}, \boldsymbol{F}$, and $\boldsymbol{\Lambda}$ with respect to the sum of squared residuals (without the smooth penalty term) as $\bar{Q}(\boldsymbol{\theta}, \boldsymbol{\Lambda}, \boldsymbol{F}) = \frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{Y}_i - \boldsymbol{Z}_i\boldsymbol{\theta} - \boldsymbol{F}\boldsymbol{\lambda}_i)^{\tau}(\boldsymbol{Y}_i - \boldsymbol{Z}_i\boldsymbol{\theta} - \boldsymbol{F}\boldsymbol{\lambda}_i)$. Here, the subscript r means that the estimator is the one when the number of factors is set to be r. The estimator $\dot{\boldsymbol{\theta}}_r$ can equivalently be obtained by minimizing the profile objective function that concentrates out $\boldsymbol{F}$ and $\boldsymbol{\Lambda}$ as $\dot{\boldsymbol{\theta}}_r = \arg\min_{\boldsymbol{\theta}}\sum_{j=r+1}^{T} v_j\left[\frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{Y}_i - \boldsymbol{Z}_i\boldsymbol{\theta})(\boldsymbol{Y}_i - \boldsymbol{Z}_i\boldsymbol{\theta})^{\tau}\right]$, where $v_j\left[\frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{Y}_i - \boldsymbol{Z}_i\boldsymbol{\theta})(\boldsymbol{Y}_i - \boldsymbol{Z}_i\boldsymbol{\theta})^{\tau}\right]$ is defined as the jth largest

eigenvalue of matrix $\frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{Y}_i - \boldsymbol{Z}_i\boldsymbol{\theta})(\boldsymbol{Y}_i - \boldsymbol{Z}_i\boldsymbol{\theta})^\tau$.

We proposed to choose the number of factors, r, by minimizing $BIC(r)$ defined as

$$BIC(r) = \ln V(r) + \rho r, \tag{2.7}$$

where $V(r) = \sum_{j=r+1}^{T} v_j[\sum_{i=1}^{N}(\boldsymbol{Y}_i - \boldsymbol{Z}_i\dot{\theta}_r)(\boldsymbol{Y}_i - \boldsymbol{Z}_i\dot{\boldsymbol{\theta}}_r)^\tau]/(NT)$, and the penalty coefficient ρ is set to be $\rho = (N+T)(p+q)/(NT)\ln(NT/(N+T))$, as suggested by Bai and Ng (2002), where $p+q$ is the total dimension of both scalar and functional covariates.

2.3 Choice of the Smoothing Parameter

The smoothing parameter $\boldsymbol{\xi}$ in the objective function controls the smoothness of $\boldsymbol{\beta}(s)$. We propose to choose the optimal value of the smoothing parameter by minimizing the the generalized cross-validation (GCV) criterion (Wahba, 1990) defined as

$$GCV(\xi) = \frac{SSE(\boldsymbol{\xi})}{\mathrm{tr}[\boldsymbol{I}_{NT} - \boldsymbol{S}(\boldsymbol{\xi})]^2}, \tag{2.8}$$

where $SSE(\boldsymbol{\xi}) = \sum_{i=1}^{N}(\boldsymbol{Y}_i - \boldsymbol{Z}_i\hat{\boldsymbol{\theta}}_{\boldsymbol{\xi}})^\tau(\boldsymbol{Y}_i - \boldsymbol{Z}_i\hat{\boldsymbol{\theta}}_{\boldsymbol{\xi}})$, in which the subscript $\boldsymbol{\xi}$ denotes the estimation is with the smoothing parameter $\boldsymbol{\xi}$, and $\boldsymbol{S}(\boldsymbol{\xi})$ is defined as $\boldsymbol{S}(\boldsymbol{\xi}) = \boldsymbol{Z}\left(\tilde{\boldsymbol{Z}}^\tau\tilde{\boldsymbol{Z}} + \boldsymbol{G}_{\hat{\boldsymbol{\theta}}}\right)^{-1}\tilde{\boldsymbol{Z}}^\tau$, where $\boldsymbol{Z} = (\boldsymbol{Z}_1^\tau, \ldots, \boldsymbol{Z}_N^\tau)^\tau$,

$\tilde{\boldsymbol{Z}} = (\tilde{\boldsymbol{Z}}_1^\tau, \ldots, \tilde{\boldsymbol{Z}}_N^\tau)^\tau$, $\tilde{\boldsymbol{Z}}_i^\tau = \boldsymbol{M}_{\hat{\boldsymbol{F}}}\boldsymbol{Z}_i$, $\boldsymbol{G}_{\hat{\boldsymbol{\theta}}}$ and $\boldsymbol{M}_{\hat{\boldsymbol{F}}}$ are the estimation for $\boldsymbol{G}_{\boldsymbol{\theta}}$ and $\boldsymbol{M}_{\boldsymbol{F}}$ for the given smoothing parameter $\boldsymbol{\xi}$, respectively.

We summarize the complete estimation procedure as the following iteration algorithm.

Step 1: Set the number of spline bases $L = T + d + 1$, the smoothing parameter $\boldsymbol{\xi}$ where $\xi_1 = \ldots = \xi_q = c_1 \in [a, b]$ with n grids $c_1, \ldots, c_n$, the factor number $r = 1 \leq r_{\max}$.

Step 2: Obtain the initial value of $(\hat{\boldsymbol{F}}^{(0)}, \hat{\boldsymbol{\Lambda}}^{(0)})$ without considering $\boldsymbol{\theta}$.

Step 3: Given $\hat{\boldsymbol{F}}^{(k-1)}$ and $\hat{\boldsymbol{\Lambda}}^{(k-1)}$, get $\hat{\boldsymbol{\theta}}^{(k-1)}$ as $\hat{\boldsymbol{\theta}}^{(k-1)}(\hat{\boldsymbol{\Lambda}}^{(k-1)}, \hat{\boldsymbol{F}}^{(k-1)}) = \left(\sum_{i=1}^N \boldsymbol{Z}_i^\tau \boldsymbol{Z}_i + \boldsymbol{G}_{\boldsymbol{\theta}}\right)^{-1} \sum_{i=1}^N \boldsymbol{Z}_i^\tau (\boldsymbol{Y}_i - \hat{\boldsymbol{F}}^{(k-1)} \hat{\boldsymbol{\lambda}}_i^{(k-1)})$.

Step 4: Given $\hat{\boldsymbol{\theta}}^{(k-1)}$, get $\hat{\boldsymbol{F}}^{(k)}$ as $\left[\frac{1}{NT}\sum_{i=1}^N (\boldsymbol{Y}_i - \boldsymbol{Z}_i\hat{\boldsymbol{\theta}}^{(k-1)})(\boldsymbol{Y}_i\boldsymbol{Z}_i\hat{\boldsymbol{\theta}}^{(k-1)})^\tau\right]\hat{\boldsymbol{F}}^k = \hat{\boldsymbol{F}}^k \boldsymbol{V}_{NT}$, and $\hat{\boldsymbol{\Lambda}}^{(k)}$ as $\hat{\boldsymbol{\lambda}}_i^k = \frac{1}{T}(\hat{\boldsymbol{F}}^k)^\tau(\boldsymbol{Y}_i - \boldsymbol{Z}_i\hat{\boldsymbol{\theta}}^{(k-1)})$.

Step 5: For $k = 1, \cdots, K$, repeat Step 3-4 until $\hat{\boldsymbol{\theta}}^{(K)}$ converges to $\hat{\boldsymbol{\theta}}^{(K-1)}$.

Step 6: Set $r = r + 1$, repeat Step 2-5 until $r = r_{max}$, and choose the optimal number of factors $\hat{r}$ using BIC criterion.

Step 7: Set $\boldsymbol{xi} = c_2\boldsymbol{I}_q$ with $c_2 \in [a, b]$, repeat Step 2-5 under the condition of $\hat{r}$ until $\boldsymbol{xi} = c_n\boldsymbol{I}_q$, and choose the optimal smoothing parameter $\boldsymbol{\xi}$ using GCV method.

Step 8: Get the iterated estimation $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\alpha}}^\tau, \hat{\boldsymbol{\gamma}}^\tau)^\tau$, $(\hat{\boldsymbol{F}}, \hat{\boldsymbol{\Lambda}})$, and $\hat{\beta}_j(s) = \hat{\boldsymbol{\gamma}}_j^\tau \boldsymbol{b}(s), j = 1, \cdots, q$.

3. Theoretical Results

We first give some assumptions required for the theoretical properties.

(A1) (a) The scalar regressor $\boldsymbol{W}_{it}$ has bounded support, $E||\boldsymbol{W}_{it}||^2 \leq M$, $E||\boldsymbol{W}_{it}||^4 \leq M$. The eigenvalues of matrix $\boldsymbol{W}\boldsymbol{W}^\tau$ are bounded from 0. (b) The functional regressor $\boldsymbol{X}_{it}$ satisfies that $E||\int \boldsymbol{X}_{it}(s)ds||^2 \leq M$, $E||\int \boldsymbol{X}_{it}(s)ds||^4 \leq M$. The maximal and minimal eigenvalues of $\boldsymbol{B}\boldsymbol{B}^\tau$ are bounded from 0 and ∞ as $N, T \to \infty$, where $\boldsymbol{B}$ is defined similarly as $\tilde{\boldsymbol{B}}$.

(A2) Let $\mathcal{F} : \{\boldsymbol{F} : \boldsymbol{F}^\tau\boldsymbol{F}/T = \boldsymbol{I}\}$. Define $D(\boldsymbol{F}) = \frac{1}{NT}\sum_{i=1}^N \boldsymbol{Z}_i^\tau \boldsymbol{M}_{\boldsymbol{F}}\boldsymbol{Z}_i - \frac{1}{N^2T}\sum_{i=1}^N\sum_{k=1}^N \boldsymbol{Z}_i^\tau\boldsymbol{M}_{\boldsymbol{F}}\boldsymbol{Z}_k \cdot \boldsymbol{\lambda}_i^\tau\left(\boldsymbol{\Lambda}^\tau\boldsymbol{\Lambda}/N\right)^{-1}\boldsymbol{\lambda}_k$. We assume $\inf_{\mathcal{F}} D(\boldsymbol{F}) > 0$.

(A3) $E||\boldsymbol{F}_t||^4 \leq M$, and $\sum_{t=1}^T \boldsymbol{F}_t^\tau\boldsymbol{F}_t/T \xrightarrow{P} \boldsymbol{\Sigma}_{\boldsymbol{F}} > 0$ for some $r \times r$ matrix $\boldsymbol{\Sigma}_{\boldsymbol{F}}$, as $T \to \infty$.

(A4) $E||\boldsymbol{\lambda}_i||^4 \leq M$, and $\boldsymbol{\Lambda}^\tau\boldsymbol{\Lambda}/N \xrightarrow{P} \boldsymbol{\Sigma}_{\boldsymbol{\Lambda}} > 0$ for some $r \times r$ matrix $\boldsymbol{\Sigma}_{\boldsymbol{\Lambda}}$, as $N \to \infty$.

(A5) (a) The error is stationary, $E(\varepsilon_{it}) = 0$, $E||\boldsymbol{\varepsilon}_i||^4 \leq M$, and $E||\boldsymbol{\varepsilon}_i||^8 \leq M$. (b) For every (t, s), $E[N^{-1/2}\sum_{i=1}^N \varepsilon_{it}\varepsilon_{is}]^4 \leq M$.

(A6) $\beta_j(s) \in \mathcal{H}_d$ for all $j = 1, \ldots, q$, where $\mathcal{H}_d$ is defined as the collection of all functions on the support $\mathcal{S}$ whose mth order derivative satisfies the Holder condition of order v with $d \equiv m + v$, where $0 < v < 1$.

(A7) The smoothing parameter ξ_j in roughness penalty satisfies that $\xi_j = O((NL)^{1/2}), j = 1, \ldots, q$.

Remark 1. Assumption (A1) is the mild and general stationary condition for scalar and functional regressors. Assumption (A2) is an identification condition for $\boldsymbol{\theta}$ to be uniquely determined. This assumption rules out the time-invariant and common regressors, in which cases $D(\boldsymbol{F}) = 0$. Assumptions (A3) and (A4) imply the existence of r factors. Assumption (A5) shows that ε_{it} is uniformly bounded. The random error is assumed to independent and identically distributed which can be extended to the existence of correlation according to minor motivations (Bai (2009)). Assumption (A6) is the mild condition for functions which have been wildly used in many models and practical applications. Assumption (A7) is the condition to make the bias from the roughness penalty negligible.

Theorem 1 establishes the consistency of the estimation of both scalar and functional coefficients, as well as the factors.

Theorem 1. *Suppose that assumptions (A1)∼(A6) hold, as* $N, T \to \infty$, *define* $\boldsymbol{P_A} = \boldsymbol{A}(\boldsymbol{A}^\tau \boldsymbol{A})^{-1}\boldsymbol{A}^\tau$ *for a given matrix* $\boldsymbol{A}$, *then*

(a) The estimator $\hat{\boldsymbol{\alpha}}$ *is consistent such that* $\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} \xrightarrow{P} 0$.

(b) The estimator $\hat{\beta}_j(\cdot)$, $j = 1, \ldots, q$, *are uniquely defined with probability tending to one.*

(c) The matrix $\boldsymbol{F}^\tau \hat{\boldsymbol{F}}/T$ *is invertible and* $||\boldsymbol{P}_{\hat{\boldsymbol{F}}} - \boldsymbol{P}_{\boldsymbol{F}}|| \xrightarrow{P} 0$.

Theorem 2 shows the consistency of the determination method for the

unknown factor numbers.

Theorem 2. *Suppose that assumptions (A1)∼(A6) hold, r is the true number of unobservable factor, then as $N, T \to \infty$, $P(\hat{r} = r) \to 1$.*

For ease of notation, define $\delta_{NT} = \min[\sqrt{N}, \sqrt{T}]$, and $\varsigma = \min[N^{-1}, T^{-1}, L^{-2}]$.

We derive the convergence rate for both scalar and functional coefficients in Theorem 3.

Theorem 3. *Assume the assumptions (A1)∼(A6) hold. If $\delta_{NT}^{-2} L \log L \to 0$ and $T/N \to 0$ as $N, T \to \infty$, then*

(1) $||\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}||_{L_2}^2 = Op(L^2(NT)^{-1} + L^{-2d+1} + L\varsigma^2)$.

(2) $\left|\left|\hat{\beta}_j(\cdot) - \beta_j(\cdot)\right|\right|_{L_2}^2 = Op(L(NT)^{-1} + L^{-2d} + \varsigma^2), \quad j = 1, \ldots, q.$

Define $\boldsymbol{V}_i = \boldsymbol{Z}_i^\tau \boldsymbol{M}_{\boldsymbol{F}} - \frac{1}{N}\sum_{k=1}^N a_{ik} \boldsymbol{Z}_k^\tau \boldsymbol{M}_{\boldsymbol{F}}$, where $a_{ik} = \boldsymbol{\lambda}_i^\tau (\boldsymbol{\Lambda}^\tau \boldsymbol{\Lambda}/N)^{-1} \boldsymbol{\lambda}_k$. Let $\boldsymbol{c}_{\boldsymbol{\alpha}} = (\boldsymbol{I}_p, \boldsymbol{0}_{p \times qL})$ denote the $p \times (p + qL)$-dimension matrix, then $\hat{\boldsymbol{\alpha}} = \boldsymbol{c}_{\boldsymbol{\alpha}} \hat{\boldsymbol{\theta}}$. As $N, T \to \infty$ simultaneously, and $\mathcal{D} = \{(\boldsymbol{W}_{it}, \boldsymbol{X}_{it}, \boldsymbol{\lambda}_i, \boldsymbol{f}_t), i = 1, \ldots, N, t = 1, \ldots, T\}$, the conditional variance matrix $\boldsymbol{\Phi}_{\boldsymbol{\alpha}} = Var(\hat{\boldsymbol{\alpha}}|\mathcal{D})$ of $\hat{\boldsymbol{\alpha}}$ conditioning on $\mathcal{D}$ is the limit in probability of $\boldsymbol{\Phi}_{\boldsymbol{\alpha}}^* = \boldsymbol{c}_{\boldsymbol{\alpha}} (\sum_{i=1}^N \boldsymbol{V}_i^\tau \boldsymbol{V}_i + \boldsymbol{G}_{\boldsymbol{\theta}})^{-1} (\sum_{i=1}^N \sigma^2 \boldsymbol{V}_i^\tau \boldsymbol{V}_i (\sum_{i=1}^N \boldsymbol{V}_i^\tau \boldsymbol{V}_i + \boldsymbol{G}_{\boldsymbol{\theta}})^{-1} \boldsymbol{c}_{\boldsymbol{\alpha}}{}^\tau$. Similarly, let $\boldsymbol{c}_{\boldsymbol{\gamma}} = (\boldsymbol{0}_{qL \times p}, \boldsymbol{I}_{qL})$ denote the $qL \times (p + qL)$-dimension matrix, then $\hat{\boldsymbol{\gamma}} = \boldsymbol{c}_{\boldsymbol{\gamma}} \hat{\boldsymbol{\theta}}$. As $N, T \to \infty$ simultaneously, the conditional variance matrix $\boldsymbol{\Phi}_{\boldsymbol{\gamma}} = Var(\hat{\boldsymbol{\gamma}}|\mathcal{D})$ of $\hat{\boldsymbol{\gamma}}$ conditioning on $\mathcal{D}$ is the limit in probability of $\boldsymbol{\Phi}_{\boldsymbol{\gamma}}^* = \boldsymbol{c}_{\boldsymbol{\gamma}} (\sum_{i=1}^N \boldsymbol{V}_i^\tau \boldsymbol{V}_i +$

$\boldsymbol{G_\theta})^{-1}(\sum_{i=1}^{N}\sigma^2\boldsymbol{V}_i^\tau\boldsymbol{V}_i(\sum_{i=1}^{N}\boldsymbol{V}_i^\tau\boldsymbol{V}_i+\boldsymbol{G_\theta})^{-1}\boldsymbol{c_\gamma}^\tau$.

The conditional variance matrix of $\hat{\boldsymbol{\beta}}(s)$ conditioning on $\mathcal{D}$ is $Var(\hat{\boldsymbol{\beta}}(s)|\mathcal{D}) = \boldsymbol{B}(s)\boldsymbol{\Phi_\gamma}\boldsymbol{B}^\tau(s)$, where $\boldsymbol{B}(s) = \boldsymbol{b}^\tau(s)\otimes\boldsymbol{I}_q$. Let $\boldsymbol{w}_j$ denote the unit vector in $\mathbb{R}^q$ with 1 in the jth coordinate and 0 in all other coordinates for $j = 1,\ldots,q$, then the conditional variance of $\hat{\beta}_j(s)$ is $Var(\hat{\beta}_j(s)|\mathcal{D}) = \boldsymbol{w}_j^\tau Var(\hat{\boldsymbol{\beta}}(s)|\mathcal{D})\boldsymbol{w}_j = \boldsymbol{w}_j^\tau\boldsymbol{B}(s)\boldsymbol{\Phi_\gamma}\boldsymbol{B}^\tau(s)\boldsymbol{w}_j$. Let $\bar{\boldsymbol{\beta}}(s) = (\bar{\beta}_1(s),\ldots,\bar{\beta}_q(s))^\tau$, where $\bar{\beta}_j(s) = E(\hat{\beta}_j(s)|\mathcal{D})$ is the mean of $\hat{\beta}_j(s)$ conditioning on $\mathcal{D}$. Then, we have the following asymptotic result of the estimator.

The asymptotic results of both scalar and functional coefficients are given in Theorem 4 and 5.

Theorem 4. *Assume the assumptions (A1)∼(A6) hold. If* $\delta_{NT}^{-2}L\log L\to 0$ *and* $T/N\to 0$ *as* $N,T\to\infty$, *then*

(1) $(Var(\hat{\boldsymbol{\alpha}}|\mathcal{D}))^{-1/2}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})\xrightarrow{L}N(\boldsymbol{0},\boldsymbol{I})$.

(2) $(Var(\hat{\boldsymbol{\beta}}(s)|\mathcal{D}))^{-1/2}(\hat{\boldsymbol{\beta}}(s)-\bar{\boldsymbol{\beta}}(s))\xrightarrow{L}N(\boldsymbol{0},\boldsymbol{I})$.

In particular, we have $(Var(\hat{\beta}_j(s)|\mathcal{D}))^{-1/2}(\hat{\beta}_j(s)-\bar{\beta}_j(s))\xrightarrow{L}N(0,1), j = 1,\ldots,q$, *where* $\xrightarrow{L}$ *denotes the convergence in distribution.*

Theorem 5. *Assume the assumptions (A1)∼(A6) hold. If* $L^{2d+1}/NT\to\infty$ *as* $N,T\to\infty$, *then* $\sup_{s\in\mathcal{S}}|(Var(\hat{\beta}_j(s)|\mathcal{D}))^{-1/2}(\bar{\beta}_j(s)-\beta_j(s))| = o_p(1), j = 1,\ldots,q$.

Theorem 4 and 5 establish the asymptotic properties of the estima-

tor $\hat{\beta}_j(\cdot)$ of the coefficient functions $\beta_j(\cdot)$ under assumptions. Moreover, compared with the variance term, as $N, T \to \infty$, the bias term of the estimator $\beta_j(\cdot)$ are asymptotically negligible from the proof of Theorem 5. Similar results can be found in Huang et al. (2004). Therefore, we can obtain the asymptotic pointwise confidence interval of $\beta_j(s)$ as $\hat{\beta}_j(s) \pm z_{\alpha/2}\hat{Var}(\hat{\beta}_j(s)|\mathcal{D})^{-1/2} \quad j = 1, \ldots, q$, where $\hat{Var}(\hat{\beta}_j(s)|\mathcal{D})$ is the estimator of $Var(\hat{\beta}_j(s)|\mathcal{D})$, and $z_{\alpha/2}$ is the $(1-\alpha/2)$ of the standard normal distribution.

4. Simulation Studies

In this section, we conduct simulations to evaluate the estimation performance of the proposed method. The simulated data are generated from the proposed model $Y_{it} = \boldsymbol{\alpha}^{\tau}\boldsymbol{W_{it}} + \int_0^1 \boldsymbol{\beta}^{\tau}(s)\boldsymbol{X_{it}}(s)ds + \boldsymbol{\lambda_i^{\tau}}\boldsymbol{F_t} + \varepsilon_{it}$. For the part of interactive effects, $\boldsymbol{\lambda_i} = (\lambda_{1i}, \lambda_{2i})^{\tau}$ and $\boldsymbol{F_t} = (F_{1t}, F_{2t})^{\tau}$ are two dimensional vectors simulated as $\boldsymbol{\lambda_i} \sim N(0, \boldsymbol{I_2})$ and $\boldsymbol{F_t} \sim N(0, 0.5\boldsymbol{I_2})$. We consider the scalar regressors $\boldsymbol{W_{it}}$ with the form as $(W_{1it} + c_1\boldsymbol{\lambda}_i^{\tau}\boldsymbol{F}_t, W_{2it})^{\tau}$, where c_1 is a constant indicating the correlation of scalar covariates with the hiding factors, W_{1it} and W_{2it} are generated from the exponential distribution and the uniform distribution, respectively: $W_{1it} \sim Exp(2)$, $W_{2it} \sim U(0, 1)$. Meanwhile, the coefficient $\boldsymbol{\alpha}$ is a two dimensional vector specified as $\boldsymbol{\alpha} = (1, 0.5)^{\tau}$. We define the smooth function $\boldsymbol{\beta}(\cdot) = (\beta_1(\cdot), \beta_2(\cdot))^{\tau}$ to be

$\beta_1(s) = 2 + 3s + e^{2s}$, $\beta_2(s) = 5 + 3\mathbf{sin}(2\pi s) + 2\cos(2\pi s)$,

The corresponding functional predictors $\boldsymbol{X_{it}}(\cdot) = (X_{1it}(\cdot), X_{2it}(\cdot))^\tau$ is set as $X_{1it}(s) = 1 + c_2 \cdot \boldsymbol{\lambda}_i^\tau \boldsymbol{F}_t + \delta_{1it} \cdot s$, $X_{2it}(s) = c_2 \cdot \boldsymbol{\lambda}_i^\tau \boldsymbol{F}_t + \delta_{2it} \cdot \sin(2\pi s)$, where c_2 is another constant representing the correlation of functional covariates with the hiding factors, δ_{1it} is simulated from the uniform distribution $U(-1, 1)$, which is independent of δ_{2it}, while δ_{2it} is simulated from the normal distribution $N(0, 2)$. The regression error ε_{it} are generated $i.i.d$ from the normal distribution $N(0, 1)$. The optimal smoothing parameter are chosen by minimizing the GCV criterion defined as (2.8) with a grid search in the range of 10^{-5} to 10^5.

4.1 Model Evaluation

To show the impact of the factors on the estimation, Figure 1 displays the true slope functions $\boldsymbol{\beta}(s) = (\beta_1(s), \beta_2(s))^\tau$ and our proposed iterated spline estimation in comparison with the estimations without considering the latent factors from 200 simulation replicates under the setting of the sample size $N = 100$ and the number of observations $T = 100$. In the first case, $c_1 = c_2 = 0$, which means both the scalar and functional regressors are independent with the responses and the latent factor, while the other cases $c_1 = c_2 = 0.5$ or 1 consider the existence of dependency. Figure 1 shows

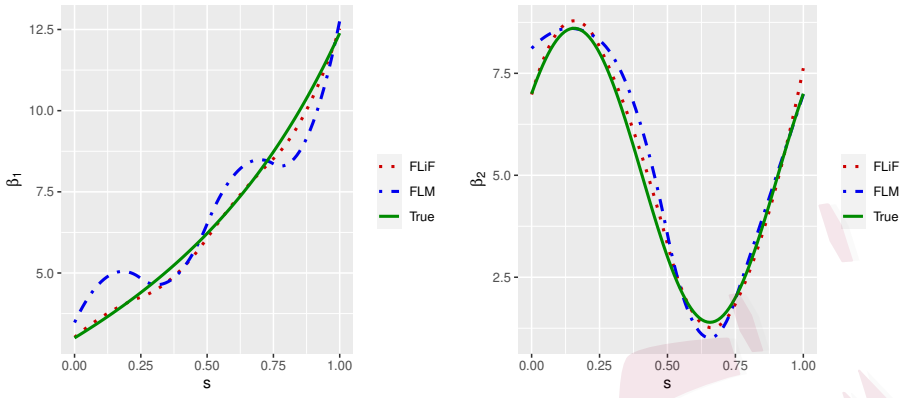


Figure 1: The estimate for the slope functions $\boldsymbol{\beta}(s) = (\beta_1(s), \beta_2(s))^\tau$ using the functional linear model with latent factors (FLiF) and the conventional functional linear model (FLM) in the case $c_1 = c_2 = 1$ obtained from one Monte Carlo run with the sample sizes $N = 100$ and the number of observations $T = 100$, where c_1 and c_2 are two constants indicating the correlation of scalar and functional covariates with the hiding factors, respectively.

that our proposed estimation procedure for $\boldsymbol{\beta}(s)$ can roughly approximate the true functions, which is superior to that without considering the factor structure when $c_1 = c_2 = 1$. It indicates that the latent factors have an influence on the estimation of the model, and our proposed iteration estimation procedure indeed achieves accurate estimates for the slope functions. The performance under the other two correlation cases and point-wise bias of the estimations are provided in the supplementary material.

To investigate the overall performance of the proposed estimation procedure and the determination method for the number of factors, we set the sample size $N = 50, 100$ and the number of observation points $T = 50, 100$

		True Number of Factors $R_0 = 0$					True Number of Factors $R_0 = 2$				
		Estimated Number of Factors R					Estimated Number of Factors R				
N	T	0	1	2	3	$\geq$4	0	1	2	3	$\geq$4
50	50	91.5%	4.0%	2.0%	1.5%	1.0%	1.0%	1.0 %	93.0 %	3.5 %	1.5%
	100	93.5%	3.0%	1.5%	1.0%	1.0%	0.5%	1.5%	95.0%	2.0 %	1.0%
100	50	96.0%	2.0%	1.0%	0.5%	0.5%	0.5%	1.0%	96.5%	1.5 %	0.5%
	100	98.0%	1.0%	0.5%	0.5%	0.0%	0.0%	0.1%	99.0%	0.5%	0.0%

Table 1: Percentage of the estimated number of factors R among 100 simulation replicates when varying the true number of factors $R_0 = 0$ or 2, the sample sizes $N = 50, 100$ and the number of observations $T = 50, 100$.

in the simulations. For each setting, the simulation repeated 100 times. We use the root mean integrated squared errors (RMISEs) and standard deviation (SD) to measure the accuracy of estimations repectively, which is defined as $\texttt{RMISE}(\hat{\beta}_l) = \left(\frac{1}{B}\sum_{b=1}^{B}\int_S\left(\hat{\beta}_l^b(s) - \beta_l(s)\right)^2 ds\right)^{1/2}$, $\texttt{SD}(\hat{\beta}_l) = \left(\frac{1}{B}\sum_{b=1}^{B}\int_S\left(\hat{\beta}_l^b(s) - \bar{\hat{\beta}}_l^B(s)\right)^2 ds\right)^{1/2}$, where $\hat{\beta}_l^b(s)$, $l = 1, 2$, are the estimations of $\beta_l(s)$ from the b-th simulation replicate, and $\bar{\hat{\beta}}_l^B(s)$ is the average estimation of $\beta_l(s)$ from the total B replicates. Meanwhile, we denote the estimation of $\beta_l(s)$ when ignoring the factor structure as $\tilde{\beta}_l(s)$.

We first evaluate the accuracy of determine the number of factors. We consider two scenarios (i) no factors exist (ii) two factors exist, both under the first setting where $c_1 = c_2 = 0$. Table 1 displays the percentage of correctly estimating the number of factors R in two scenarios. It shows that the accuracy to identify the number of factors under both scenarios

increases with the sample size N or observation number T. When $N = 100$ and $T = 100$, the number of factors are correctly determined in almost all simulation replicates.

RMISE and Standard Deviation (SD) of the Estimates									
Sample Size		RMISE				SD			
		FLM		FLiF		FLM		FLiF	
N	T	$\tilde{\beta}_1$	$\tilde{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\tilde{\beta}_1$	$\tilde{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
$c_1 = c_2 = 0$									
50	50	0.495	0.457	0.292	0.274	0.281	0.253	0.268	0.251
50	100	0.448	0.426	0.243	0.219	0.224	0.217	0.206	0.195
100	100	0.402	0.374	0.187	0.159	0.179	0.152	0.153	0.121
$c_1 = c_2 = 0.5$									
50	50	0.581	0.547	0.344	0.319	0.348	0.321	0.306	0.285
50	100	0.527	0.479	0.281	0.252	0.283	0.258	0.257	0.213
100	100	0.477	0.441	0.234	0.205	0.237	0.209	0.209	0.171
$c_1 = c_2 = 1$									
50	50	0.664	0.638	0.415	0.402	0.429	0.408	0.365	0.359
50	100	0.621	0.579	0.356	0.324	0.367	0.331	0.298	0.272
100	100	0.569	0.521	0.303	0.256	0.293	0.271	0.255	0.219

Table 2: The root mean integrated squared errors (RMISE) and standard deviations (SD) of the estimate for the slope functions $\beta_1(s)$ and $\beta_2(s)$ using the functional linear model with latent factors (FLiF) and the conventional functional linear model (FLM) when varying the sample sizes $N = 50, 100$ and the number of observations $T = 50, 100$, where c_1 and c_2 are two constants indicating the correlation of scalar and functional covariates with the hiding factors, respectively.

Table 2 presents the RMISEs and standard deviations of estimations of $\boldsymbol{\beta}(s) = (\beta_1(s), \beta_2(s))^\tau$ under three different cases of correlation. It shows

that our proposed estimation procedure has a smaller RMISE of the estimate for the functional coefficients compared to the one without considering factor structure. Moreover, the RMISEs and the standard deviations of the estimated slope functions are decreasing with the sample size N and the number of observations T in all cases. In addition, the RMISE and the standard deviation in the case when $c_1 = c_2 = 0$ are generally smaller than the other two cases, which indicates that the correlation between the regressors and responses affects the estimation accuracy to some extent. Moreover, the result shows an obvious difference between the estimations from the three correlation cases. The estimation of $\boldsymbol{\beta}(s)$ in the first case, where no correlation exists between regressors and responses, performs better than the other two with the dependency as expected. This means that the estimations obtained in the first case is unbiased, while in the presence of correlation in the other two cases, we can only get the biased estimations. The RMSE of the factor structure under three cases are presented in the supplementary material. It also proves that the existence of correlation has a certain degree of impact on estimations.

4.2 Comparison with the Panel Data Model

In this simulation study, we compare the proposed functional linear model with latent factors (FLiF) (1.1) model with the conventional panel data model $Y_{it} = \boldsymbol{\alpha}^{\tau}\boldsymbol{W_{it}} + \boldsymbol{\beta}^{\tau}\overline{\boldsymbol{X}}_{it} + \boldsymbol{\lambda}_{\boldsymbol{i}}^{\tau}\boldsymbol{F_t} + \varepsilon_{it}$, where $\overline{\boldsymbol{X}}_{it}$ is the mean value of $\boldsymbol{X_{it}}(s)$ over the domain S. We choose RMSE to measure the accuracy of the estimated response variable under the two different models $\texttt{RMSE}(\widehat{\boldsymbol{Y}}) = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T}(\hat{Y}_{it} - Y_{it})^2\right]^{1/2}$, where $\widehat{\boldsymbol{Y}}$ and $\hat{Y}_{it}$ are the estimations for the response variable $\boldsymbol{Y}$ and Y_{it}, respectively. The simulated data are generated based on the FLiF model with the sample size $N = 50$ and the number of observations per subject $T = 50, 100$.

Average RMSEs for the estimated response $\widehat{\boldsymbol{Y}}$							
Sample Size		$c_1 = c_2 = 0$		$c_1 = c_2 = 0.5$		$c_1 = c_2 = 1$	
N	T	Panel	FLiF	Panel	FLiF	Panel	FLiF
50	50	0.467	0.327	0.535	0.376	0.648	0.462
50	100	0.402	0.257	0.469	0.313	0.557	0.405
100	100	0.352	0.193	0.417	0.235	0.493	0.349

Table 3: Average Root Mean Squared Errors (RMSEs) for the estimated response $\widehat{\boldsymbol{Y}}$ over 100 simulation replicates using the proposed functional linear model with latent factors (FLiF) model and the conventional panel data model when varying the sample sizes $N = 50, 100$ and the number of observations $T = 50, 100$, where c_1 and c_2 are two constants indicating the correlation of scalar and functional covariates with the hiding factors.

Table 3 shows the average RMSEs of the estimated response $\widehat{\boldsymbol{Y}}$ of two models under three different correlation settings. It shows that the estima-

tions in the FLiF model have smaller RMSE than those in the panel data model in all settings. The RMSE of estimations in both models decreases with the sample size N and T in both cases.

5. Applications

In this section, we illustrate the efficiency of the proposed functional linear model with latent factors (FLiF) and estimation procedure through the analysis of two real data sets.

5.1 Stock Return Analysis

The objective of this study is to investigate how the daily returns of stocks are affected by the temporal stock price trajectories with the existence of hidden factors. The data we considered contains the daily stock price every three seconds of CSI 300 constituent shares from August 3 to October 30 in 2020, as well as the average opening and the average closing price of each stock in this period. The data is available on the official website of the Shanghai and Shenzhen stock exchange (www.sse.gov.cn and www.szse.gov.cn). To present the daily fluctuation of the stock price more clearly, Figure 2 depicts the daily stock price every three seconds of two chosen constituent stocks Huichuan Inovance and Shiji Information in three

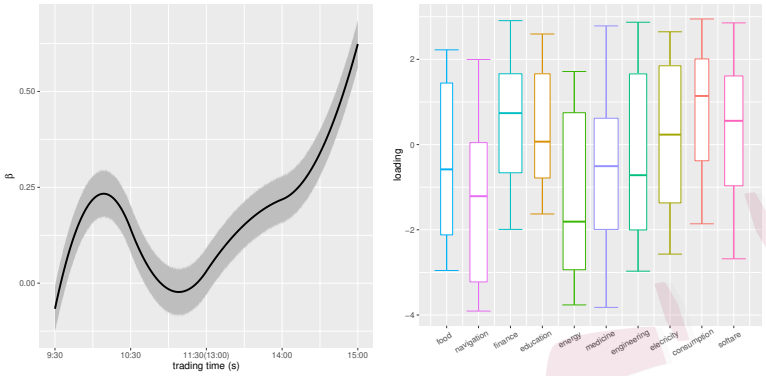
different days, respectively.

Figure 2: The stock price every three seconds of two constituent shares Huichuan Inovance and Shiji Information on three days: August 26, September 20, October 23, 2020.

We consider the functional linear model with latent factors (1.1) with $N = 300, T = 59$, where Y_{it} is the daily return of the i-th CSI 300 constituent share in the day t, W_{it} is the corresponding average daily opening price, and $X_{it}(s)$ is the trajectory of the stock price every three seconds of each stock. We define the daily return of stocks, Y_{it}, as the logarithm of the ratio of the daily closing price and the opening price of each stock.

Figure 3 (a) displays the estimated slope function $\hat{\beta}(s)$. It shows that the stock price around the time 10am has the largest positive impact on the stock return in the morning. The stock price in the afternoon has an increasing positive impact on the stock return.

One hidden factor is determined to have an impact on stock yields besides the average opening price and daily stock price trajectories by min-



(a) (b)

Figure 3: (a) The estimated slope function $\hat{\beta}(s)$ for quantifying the influence of daily stock price on the daily return of CSI 300 constituent shares. The grey area indicates the 95% point-wise confidence intervals for $\beta(s)$. (b) Boxplot of the estimated loading on the hidden factors for CSI 300 constituent shares, which is divided into ten blocks.

imizing the BIC criterion defined in (2.7). In order to understand the impact of the hidden factor on the stock yield, we classify the CSI 300 constituent shares into ten sectors as 'consumption', 'education', 'electricity', 'energy', 'engineering', 'finance', 'food', 'medicine', 'navigation', and 'software', which is based on the industry and related attributes of each enterprise. Figure 3 (b) presents the boxplot of the estimated loadings sorted according to their mean values in each sector.

The boxplot reveals that the average loadings of the 'navigation' and 'energy' sectors are relatively low, whereas the 'consumption', 'finance', and 'software' sectors have relatively high average loadings. Specifically,

the factor has an interquartile range of [-3.16, 0.13] with a median of -1.86 for the 'energy' sector, and an interquartile range of [-2.96, 0.32] with a median of -1.75 for the 'software' sector, indicating that these sectors are more susceptible to the negative impact of the hidden factor. Meanwhile, the hidden factor has an interquartile range of [-0.51, 2.01] with a median of 1.14 for the 'consumption' sector, an interquartile range of [-0.66, 1.46] with a median of 0.74 for the 'finance' sector, and an interquartile range of [-1.25, 1.76] with a median of 0.51 for the 'software' sector, suggesting a more positive impact on these sectors. Moreover, the 'education' and 'electricity' sectors, as well as the 'food' and 'medicine' sectors, are similarly affected respectively.

In the context of 2020, the trade war initiated by the United States against China escalated rapidly, increasing technical trade barriers. The U.S. tariff policy towards China restricted China's import and export trade. Besides imposing high tariffs on Chinese exports, the U.S. intensified the trade war at the technical level, particularly targeting technology-intensive products and energy sectors like chips and cybersecurity. Additionally, the U.S. imposed sanctions on Chinese technology companies such as Huawei, ZTE, ByteDance, and Tencent, citing national security concerns. In response, China implemented stronger anti-sanction policies, and public sen-

timent towards the trade war fostered a more positive trend in the domestic market economy and national consumption. Aside from external environmental shocks, factors like operating performance, distribution policies, and industry characteristics also impact individual stock returns. Therefore, the hidden factor can be seen as a combination of these aspects.

5.2 Air Pollution Data Analysis

The air quality index (AQI) is an all-encompassing measure that quantifies the level of air pollution in a relative and dimensionless manner. It condenses the monitored air composition into a single conceptual index value by considering the proportion of different components present in the air. The AQI assesses the severity of air pollution and the prevailing air quality conditions, making it an effective tool for capturing the short-term status and trend of air quality changes. Usually, particulate matter 2.5 (PM 2.5), sulfur dioxide ($\mathbf{SO}_2$), nitrogen dioxide ($\mathbf{NO}_2$), carbon monoxide (CO), particulate matter 10 (PM 10) and ozone ($\mathbf{Q}_3$) are used as accounting factors. In this article, we analyze the effect of daily temperature on the PM 2.5 index. The data is obtained at the website of the China Meteorological Administration (http://www.cma.gov.cn).

We consider the data containing the daily temperature from 2015 to

2020 from meteorological stations in 31 provincial capitals of China, as well as the average monthly PM 2.5 index and humidity in this period of each station. Figure 4 presents the daily temperature curves of two cities Tianjin and Nanjing in three different months during the period, respectively.

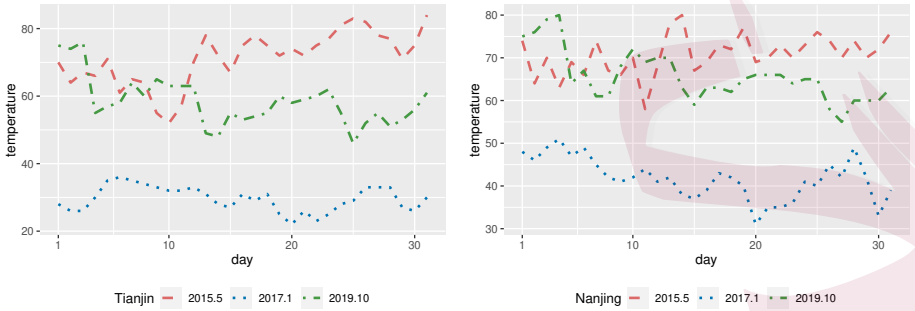


Figure 4: Daily temperature curves of two cities Tianjin (left panel) and Nanjing (right panel) in three different months May 2015, January 2017 and October 2019, respectively.

The functional linear model with latent factors of interest is setting as (1.1) with $N = 31, T = 72$, where Y_{it} is the average monthly PM 2.5 of the $i-$th weather station in the month t, W_{it} is the corresponding average monthly humidity of each weather station, and $X_{it}(s)$ is the daily temperature of each weather station.

Two hidden factors are determined to have an impact on the air pollution in addition to humidity and daily temperature by minimizing the BIC criterion defined in (2.7). The existence of the factors also shows that other common interactive effects have a certain impact that have a certain

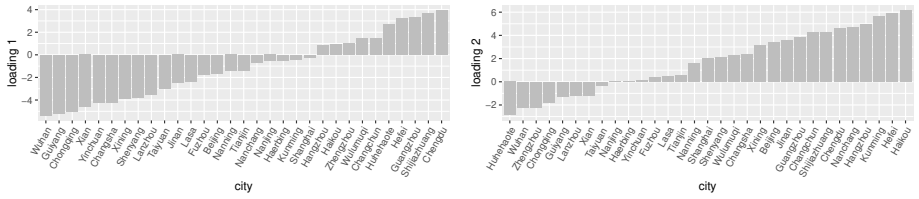


Figure 5: The estimated loadings on two hidden factors in 31 provincial capital cities assorted based on their loading values in the functional linear model with latent factors (FLiF) from the air pollution data.

impact on the air pollution besides the average humidity. To further study the impact of potential factors on the different cities, Figure 5 depicts the factor loadings of each provincial capital city and sorts them according to their values. Specifically, the first loading has the most significant negative impact on the cities of Wuhan, Guiyang and Chongqing, while showing a strong positive impact on cities such as Guangzhou, Shijiahzuang and Chengdu. Meanwhile, the second loading has a more relatively negative impact on Huhehaote, Wuhan and Zhengzhou, while cities such as Kunming, Hefei and Haikou will be more positively affected. By analyzing cities with similar impacts from each factor separately, the first factor can be roughly regarded as the influence of geographical morphology on the speed and intensity of air flow propagation, while the second factor can be viewed as the impact of geographical vegetation area or extreme weather conditions of cities.

Meanwhile, Figure 6 also plots the map of estimated loadings in 31

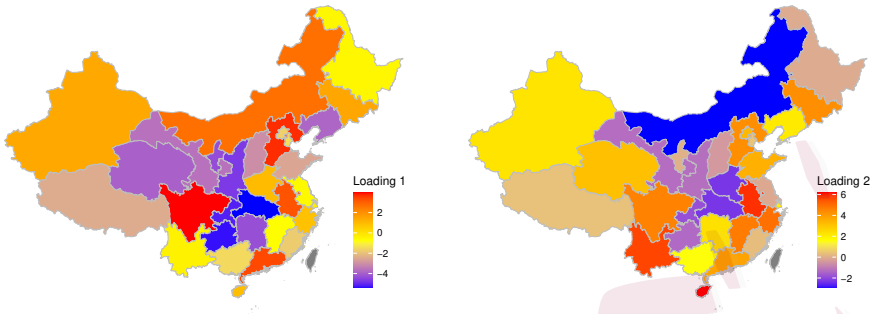


Figure 6: The map of estimated loadings on two hidden factors for 31 provincial capital cities in the functional linear model with latent factors (FLiF) from the air pollution data. For visualization purposes, the color of each province represents the estimated loading for the respective provincial capital city.

capital cities respectively. First, we can find that the impact degree of two factors on the same region is roughly different. Next, for the loading of first factor, the areas with deeper negative impacts are mainly concentrated in the inland areas of the central plains, while the impact on the northern border areas is relatively similar. However, for the loading of second factor, the performance of the northwest region is relatively similar, while the impact on the southern region is generally positive. Considering the actual situation of each city, the presence of potential factors are not unexpected. Generally speaking, there are many different factors including pollutant emissions, atmospheric diffusion capacity, meteorological conditions such

as wind direction, wind speed, inversion stratification, precipitation, meteorological conditions, topography, etc. These possible influencing elements can be explained by the latent factor structure.

6. Conclusions and Discussion

In this article, we present a novel functional linear model incorporating latent factors. Our proposed iterated profile estimation method allows us to estimate both the hidden factors and functional coefficients effectively. The first step involves employing the B-spline expansion to approximate the slope function. Then, we implement an iterative profile estimation technique to derive estimators for both the functional coefficients and the hidden factors. To determine the number of hidden factors, we introduce a BIC (Bayesian Information Criterion) criterion. Furthermore, we establish the theoretical properties, including the convergence and asymptotic normality of the estimators.

The efficiency of our estimation procedure is validated through comprehensive simulation studies conducted under diverse scenarios. We systematically evaluate the performance of our novel iterated profile estimation method under various correlation conditions between the response and covariates. To gauge its effectiveness, we compare the results obtained

from our approach with those derived from the conventional functional linear model, which disregards the hidden factors. The comparative analysis demonstrates that our proposed model and estimation method outperform the conventional functional linear model in different scenarios. The improvement in performance signifies the significance of incorporating hidden factors in the model to achieve more accurate estimations.

To illustrate the practical applicability of our proposed model, we conduct analyses on real-world data from both the financial market and air pollution. By doing so, we successfully identify hidden factors that play a crucial role in influencing the response variables in these domains. This real-data application further strengthens the validity and usefulness of our approach in gaining insights from functional data analysis.

A key limitation of our proposed model lies in the underlying assumption that the influence of the functional covariate on the response remains static across time. This constraint may limit its practical application in scenarios where the functional covariate and the response variable exhibit a dynamic, time-varying relationship. To address this limitation, we aim to extend our research in the future by exploring models that can capture such temporal variations in the interplay between the covariate and the response variable. There is also a large amount of follow-up research that

can be carried out in the future for this topic. For example, the quantile model characterizes the different relationships that variables exhibit in different quantile situations (Zheng et al. (2018); Zhang et al. (2019); Peng and Wang (2022)), so the model proposed in the article can be extended in this field, such as studying the functional quantile model with factors.

Acknowledgments

The authors would like to thank the editor, the associate editor, and two referees for many insightful comments. These comments are very helpful for us to improve our work. T. Li's research was supported by the Humanities and Social Science Fund of Ministry of Education of China (21YJA910001). J. You's research was supported by the National Natural Science Foundation of China (Grant No.11971291) and Innovative Research Team of Shanghai University of Finance and Economics. J. Cao's research is supported by a Discovery grant (RGPIN-2023-04057) from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canada Research Chairs Program.

Supplementary Materials

The supplementary document includes the additional numerical results and detailed proofs for theoretical results. We also provide the R codes for the simulation studies and real data analysis on the website `https://github.com/statszx/FLiF`.

References

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77:1229–1279.

Bai, J. and Li, K. (2014). Theory and methods of panel data models with interactive fixed effects. *The Annals of Statistics*, 42(1):142–170.

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

Cai, X., Xue, L., and Cao, J. (2022a). Robust estimation and variable selection for function-on-scalar regression. *Canadian Journal of Statistics*, 50:162–179.

Cai, X., Xue, L., and Cao, J. (2022b). Variable selection for multiple function-on-function linear regression. *Statistica Sinica*, 32:1435–1465.

Chen, C., Guo, S., and Qiao, X. (2022). Functional linear regression: dependence and error contamination. *Journal of Business and Economic Statistics*, 40(1):444–457.

Chen, J., Li, D., Wei, L., and Zhang, W. (2021). Nonparametric homogeneity pursuit in functional-coefficient models. *Journal of Nonparametric Statistics*, 33(3-4):387–416.

Crambes, C., Kneip, A., and Sarda, P. (2009). Smoothing splines estimators for functional linear regression. *The Annals of Statistics*, 37(1):35–72.

Fan, J., Li, K., and Liao, Y. (2021). Recent developments on factor models and its applications in econometric learning. *Annual Review of Financial Economics*, 13:401–430.

Fan, J., Liao, Y., and Mincheva, M. (2011). High dimensional covariance matrix estimation in approximate factor models. *The Annals of Statistics*, 39:3320–3356.

Fang, K., Zhang, X., Ma, S., and Zhang, Q. (2020). Smooth and locally sparse estimation for multiple-output functional linear regression. *Journal of Statistical Computation and Simulation*, 90(2):341–354.

Feng, S., Li, G., Peng, H., and Tong, T. (2018). Varying coefficient

panel data model with interactive fixed effects. *arXiv preprint*, page arXiv:1803.02714.

Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis*. Springer, New York.

Gao, Y., Shang, H., and Yang, Y. (2022). Factor-augmented model for functional data. *Statistica Sinica: Preprint*, doi:10.5705/ss.202021.022.

Guan, T., Lin, Z., and Cao, J. (2020). Estimating truncated functional linear models with a nested group bridge approach. *Journal of Computational and Graphical Statistics*, 29(3):620–628.

Hilgert, N., Mas, A., and Verzelen, N. (2013). Minimax adaptive tests for the functional linear model. *The Annals of Statistics*, 41(2):838–869.

Huang, J. Z., Wu, C. O., and Zhou, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statistica Sinica*, 14:763–788.

Jiang, C. and Wang, J. (2011). Functional single index models for longitudinal data. *The Annals of Statistics*, 39(1):362–388.

Li, J., Huang, C., and Zhu, H. (2017). A functional varying-coefficient

single-index model for functional response data. *Journal of the American Statistical Association*, 112(519):1169–1181.

Lin, Z., Cao, J., Wang, L., and Wang, H. (2017). Locally sparse estimator for functional regression models. *Journal of Computational and Graphical Statistics*, 26(2):306–318.

Liu, B., Wang, L., and Cao, J. (2017). Estimating functional linear mixed-effects regression models. *Computational Statistics and Data Analysis*, 106:153–164.

Liu, R., Wang, H., and Wang, S. (2018). Functional variable selection via Gram-Schmidt orthogonalization for multiple linear regression. *Journal of Statistical Computation and Simulation*, 88(18):3664–3680.

Liu, X., Ma, S., and Chen, K. (2022). Multivariate functional regression via nested reduced-rank regularization. *Journal of Computational and Graphic Statistics*, 31:231–240.

Morris, J. (2015). Funcitional regression. *Annual Review of Statistics and Its Application*, 2:321–359.

Peng, X. and Wang, H. (2022). A generalized quantile tree method for sub-

group identification. *Journal of Computational and Graphical Statistics*, 31(3):824–834.

Pesaran, M. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74:967–1012.

Ramsay, J. (1982). When the data are functions. *Psychometrika*, 47(4):379–396.

Ramsay, J. and Silverman, B. (2005). *Functional data analysis*. Springer Series in Statistics, New York: Springer, 2nd ed.

Wahba, G. (1990). *Spline models for observational data*. SIAM, Philadelphia.

Wu, Y., Fan, J., and Müller, H. (2010). Varying-coefficient functional linear regression. *Bernoulli*, 16:730–758.

Xun, X., Guan, T., and Cao, J. (2022). Sparse estimation of historical functional linear models with a nested group bridge approach. *Canadian Journal of Statistics*, 50:1254–1269.

Yuan, M. and Cai, T. (2010). A reproducing kernel Hilbert space approach to functional linear regression. *The Annals of Statistics*, 38(6):3412–3444.

Zhang, Y., Wang, H., and Zhu, Z. (2019). Quantile-regression-based clustering for panel data. *Journal of Econometrics*, 213(1):54–67.

Zheng, Y., Zhu, Q., Li, G., and Xiao, Z. (2018). Hybrid quantile regression estimation for time series models with conditional heteroscedasticity. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 80(5):975–993.

Zhu, H., Chen, K., Luo, X., Yuan, Y., and Wang, J. (2019). Fmem: Functional mixed effects models for longitudinal functional responses. *Statistica Sinica*, 29(4):2007–2033.

Department of Statistics and Management, Shanghai University of Finance and Economics

E-mail: h_zxuan@163.com

Department of Statistics and Management, Shanghai University of Finance and Economics

E-mail: li.tao@mail.shufe.edu.cn

Department of Statistics and Management, Shanghai University of Finance and Economics

E-mail: johnyou07@163.com

Department of Statistics and Actuarial Science, Simon Fraser University

E-mail: jiguo_cao@sfu.ca