Sparsity learning via structured functional factor augmentation

Hanteng Ma,

School of Statistics and Data Science, Shanghai University of Finance and Economics

Ziliang Shen,

School of Statistics and Data Science, Shanghai University of Finance and Economics

Xingdong Feng,

School of Statistics and Data Science, Shanghai University of Finance and Economics,

ORCID ID: 0000-0002-6091-8169

and

Xin Liu*

School of Statistics and Data Science, Shanghai University of Finance and Economics,

ORCID ID: 0001-8779-2618.

April 22, 2026

Abstract

As one of the most powerful tools for examining the association between functional covariates and a response, the functional regression model has been widely adopted in various interdisciplinary studies. Usually, a limited number of functional covariates are assumed in a functional linear regression model. Nevertheless, correlations may exist between functional covariates in high-dimensional functional linear regression models, which brings significant statistical challenges to statistical inference and functional variable selection. In this article, a novel functional factor augmentation structure (fFAS) is introduced for multivariate functional series, and a multivariate functional factor augmentation selection model (fFASM) is further proposed to deal with issues arising from variable selection of correlated functional covariates. Theoretical justifications for the proposed fFAS are provided, and statistical inference results of the proposed fFASM are established. Numerical investigations support the superb performance of the novel fFASM model in terms of estimation accuracy and selection consistency.

Keywords: correlated functional covariates, functional factor augmentation structure, functional variable selection, factor augmentation regression.

*Dr. Xin Liu is the corresponding author whose email address is liu.xin@mail.shufe.edu.cn

1 Introduction

Functional data, usually referred to as a sequential collection of instances over time with serial dependence, have been widely observed in various scenarios across different scientific disciplines, such as earth, medical and social sciences (Peng et al. 2005, Centofanti et al. 2021). Instead of employing conventional time series modeling techniques, an underlying trajectory is usually assumed for a functional process and sampled at a set of time points in some interval, so that statistical modeling and inference are feasible, including estimation of the mean process and its intrinsic covariance structure, trajectory recovery, and prediction of a functional; see, for example, Ramsay & Silverman (2005), Yao et al. (2005), Ferraty (2006), Hall & Horowitz (2007), Hörmann & Kokoszka (2010), Cuevas (2014), Hörmann et al. (2015), Aneiros, Horová, Hušková & Vieu (2022), Zhenhua Lin & Müller (2023), Petersen (2024), among others.

Conventionally, functional data analysis has focused on a single or fixed number of separate functional series, while recent studies have explored associations between multiple functional series. The most widely adopted model may be the functional linear model, which assumes a linear link between between a response and a bunch of covariates, with at least one functional covariate, using unknown functional coefficient curves. Such a model offers a statistical tool to infer linear association between the response and functional covariates. Several studies have attempted along this path, employing parametric, non-parametric, and semi-parametric models; see, for instance, Cardot et al. (2003), Yao et al. (2005), Li & Hsing (2007), Müller & Yao (2008), Yuan & Cai (2010), Chen et al. (2011), Kong et al. (2016), Chen et al. (2022), Lin & Yao (2020), among others.

However, two main issues arise in multivariate functional data analysis in practice. One is that functional series may be correlated with each other, especially in high dimensions or less dense observation cases, while the current functional linear model analysis may typically assume independence between multiple functional covariates. To address this problem, several methods have been proposed by explicitly assuming pairwise covariance structures for functional covariates (Chiou & Müller 2016, Lin & Wang 2022). However, these methods may seem less robust to model misspecification and lower estimation accuracy, primarily due to their limited flexibility in capturing complex covariance structures and computational inefficiencies when dealing with high-dimensional functional data. Alternatively, one can describe such associations by using a factor model with lower ranks (Bai 2003, Fan et al. 2020), where all functional covariates are assumed to share a finite number of certain latent functional factors (Castellanos et al. 2015), without assuming any explicit correlation structure for multivariate functional covariates. A few studies have touched the so-called functional factor models, e.g., Hays et al. (2012), Chen et al. (2021), Yang & Ling (2024), while these methods provide limited exploration of common functional factors shared by functional covariates. To the best of our knowledge, very limited research has focused on how to capture common associations for multivariate functional series efficiently and effectively with inferential justifications from a statistical perspective.

Another issue is how to select useful functional covariates in multivariate functional linear regression models when correlations between functional covariates exist, which is also frequently needed in practice, such as Internet of Things (IoT) data analysis (Gonzalez-Vidal et al. 2019), antibiotic resistance prediction (Jiménez et al. 2020), non-small cell lung cancer (NSCLC) clinical study (Fang et al. 2015), and stock market forecasting (Nti et al. 2019, Htun et al. 2023). A common way to achieve functional variable selection is to impose a penalty on the corresponding functional coefficient curves in a group-wise manner, using popular penalties such as the Lasso (Tibshirani 1996), SCAD (Fan & Li

2001) and MCP (Zhang 2010). A few studies have touched on functional variable selection in certain scenarios; see, for example, Matsui & Konishi (2011), Kong et al. (2016), Aneiros, Horová, Hušková & Vieu (2022), among others. However, when correlation exists between functional covariates, such a strategy turns out to be less accurate in functional coefficient estimation and fails to capture the truly useful ones, as demonstrated in the simulation studies in this article. Consequently, it remains statistically challenging to select useful functional covariates with selection consistency in multivariate functional linear models with correlated functional covariates in high dimensions.

Inspired by the challenges above, we firstly propose a novel functional factor augmentation structure (fFAS) to capture associations for correlated multivariate functional data, and further propose a functional factor augmentation selection model (fFASM) to achieve selection of functional covariates in high dimensions when correlations exist between functional covariates using the penalized method. Not only is the correlation addressed without assuming an explicit covariance structure, but theoretical properties of the estimated functional factors are also established. Further, pertaining to the correlated functional covariates, the proposed fFASM method successfully captures the useful functional covariates simultaneously in the context of functional factor models. Numerical studies on both simulated and real datasets support the superior performance of the proposed fFASM. The main contributions of the proposed method are two-fold as follows.

- We propose a feasible time-independent functional factor augmentation structure (fFAS) for functional data, and establish theoretical justifications for statistical inference, revealing valuable insights into current literature. Also, the impact from truncation inherent in functional data analysis to the proposed fFAS is discussed, and solutions are provided to enhance robustness and applicability of our model. A key result is how the difference may be addressed and controlled by comparing the true factor model and the estimated one when using truncated expansion.

- A multivariate functional linear regression model with correlated functional covariates is proposed, by employing the proposed functional structural factor. As correlations often exist in multiple functional covariates in high dimensions but may be difficult to estimate, the proposed fFASM method decomposes the functional covariates into two parts with low correlations with each other, and simultaneously expands the dimensions of parameters to be estimated. By this way our approach improves the performance of functional variable selection.

The rest of the paper is organized as follows. Section 2 introduces a novel fFAS for functional processes with its statistical properties, and the fFASM is further proposed in Section 3 for correlated functional covariates and functional variable selection in detail with theoretical justifications. Section 4 employs simulated data to examine the proposed method in various scenarios, and Section 5 presents its applications on two sets of real-world datasets. Section 6 concludes the article with discussions.

Notation: $\boldsymbol{I}_n$ denotes the $n \times n$ identity matrix; $\mathbf{0}$ refers to a general $n \times m$ zero matrix; $\mathbf{0}_n$ and $\mathbf{1}_n$ represent the all-zero and all-one vectors in $\mathbb{R}^n$, respectively. For a vector $\boldsymbol{a}$, $\|\boldsymbol{a}\|_2$ denotes its Euclidean norm. For a vector $\boldsymbol{v} \in \mathbb{R}^p$ and $S \subseteq [p]$, denote $\boldsymbol{v}_S = (\boldsymbol{v}_i)_{i \in S}$ as its sub-vector. For a matrix $\boldsymbol{M} \in \mathbb{R}^{n \times m}, I \subseteq [n]$ and $J \subseteq [m]$, define $\boldsymbol{M}_{IJ} = (\boldsymbol{M}_{ij})_{i \in I, j \in J}, \boldsymbol{M}_{I.} = (\boldsymbol{M}_{ij})_{i \in I, j \in [m]}$ and $\boldsymbol{M}_{.J} = (\boldsymbol{M}_{ij})_{i \in [n], j]}$, and its matrix entry-wise max norm is denoted as $\|\boldsymbol{M}\|_{\max} = \max_{i,j} |M_{ij}|$, and $\|\boldsymbol{M}\|_F$ and $\|\boldsymbol{M}\|_p$ as its Frobenius and induced p-norms, respectively. Denote $\lambda_{\min}(\boldsymbol{M})$ as the minimum eigenvalue of $\boldsymbol{M}$ if it is symmetric. Let ∇ and ∇^2 be the gradient and Hessian operators. For $f : \mathbb{R}^p \to \mathbb{R}$ and $I, J \subseteq [p]$, define $\nabla_I f(x) = (\nabla f(x))_I$ and $\nabla^2_{IJ} f(x) = (\nabla^2 f(x))_{IJ}$. $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ refers to the normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

2 A novel functional factor augmentation structure

2.1 Functional observations and its expansion

To start with, suppose that G functional processes are observed and collected sequentially in a sample $\{W_i^{(g)}(t_{ij}^{(g)}), t_{ij}^{(g)} \in \mathcal{T};\ g = 1, \ldots, G,\ i = 1, \ldots, n,\ j = 1, \ldots, n_i^{(g)}\}$ for the g-th functional process from the i-th subject at time t_{ij} in a certain time interval $\mathcal{T}$, and we use $W_i^{(g)}(t_j)$ for abbreviation without confusion. Usually in the analysis of a single functional process, $W_i^{(g)}(t_j)$ is assumed to consist of the underlying process $X_i^{(g)}(t_j)$ and an independent noise $\varepsilon_i^{(g)}(t_j)$, i.e., $W_i^{(g)}(t_j) = X_i^{(g)}(t_j) + \varepsilon_i^{(g)}(t_j)$, where $X_i^{(g)}(\cdot)$ and $\varepsilon_i^{(g)}(\cdot)$ are assumed to be identically and independently distributed (i.i.d.) with a mean function and a covariance structure, respectively, and $\mathbb{E}(\varepsilon_i^{(g)}(t)) = 0, \mathbb{E}(X_i^{(g)}(t)) = \mu(t)$. For convenience, we centralize these functional processes and still use the notation $X_i^{(g)}(\cdot)$ to represent them. To recover the functional trajectory $X_i^{(g)}(t)$, it is usually smoothed by assuming an expansion over a set of pre-specified orthogonal basis functions $\{\phi_{0j}^{(g)}(\cdot), j = 1, \ldots, m_g\}$ as

$$X_i^{(g)}(t) = \sum_{j=1}^{m_g} a_{0,ij}^{(g)} \phi_{0j}^{(g)}(t) + e_{0i}^{(g)}(t) = \boldsymbol{a}_{0,i}^{(g)\top} \boldsymbol{\phi}_0^{(g)}(t) + e_{0i}^{(g)}(t), \tag{1}$$

where $\boldsymbol{a}_{0,i}^{(g)} = (a_{0,i1}^{(g)}, \ldots, a_{0,im_g}^{(g)})^\top$ are the time-invariant coefficients, $\boldsymbol{\phi}_0^{(g)}(t) = (\phi_{01}^{(g)}(t), \ldots, \phi_{0m_g}^{(g)}(t))^\top$, and $e_{0i}^{(g)}(\cdot)$ is the residual orthogonal to $\phi_{0j}^{(g)}(\cdot)$. As in practice, $\boldsymbol{\phi}_0^{(g)}(\cdot)$ is usually unknown, an identifiable estimator $\widehat{\boldsymbol{a}}_{0,i}^{(g)}$ can be obtained under certain conditions, as will be shown later in Lemma 1.

Without loss of generality, we assume that the functional covariates $\{X^{(g)}(\cdot), g = 1, \ldots, G\}$ are centered with a zero mean. In practice, the true trajectories are unobservable and are contaminated by discrete measurement errors, yielding observations $W_i^{(g)}(t_{ij}) =$

$X_i^{(g)}(t_{ij})+\epsilon_i^{(g)}(t_{ij})$. To recover the underlying trajectories, we employ a local linear smoother yielding $\widehat{X}_i^{(g)}(\cdot)$. To establish the asymptotic properties of the proposed estimator, certain regularity conditions are required concerning the functional processes and the observation design. The core conditions are informally summarized to clarify the sampling regimes covered by our framework (Kong et al. 2016), and detailed mathematical formulations of these broadly applicable conditions (B1–B4) are deferred to the Supplementary Material.

- **Process Regularity and Smoothness (Conditions B1 & B2):** We assume that the underlying functional trajectories $X^{(g)}(\cdot)$ are sufficiently smooth (specifically, twice continuously differentiable) with bounded moments. The functional principal component scores are also assumed to possess bounded moments, ensuring that the process does not exhibit extreme, singular behaviors.

- **Observation Grid (Condition B3):** The measurement time points $t_{ij}^{(g)}$ are assumed to be generated from regular, uniformly bounded density functions over the domain $\mathcal{T}$. This ensures that the observation points are well-distributed without exhibiting excessively large gaps, which is crucial for uniform smoothing.

- **Dense Sampling Regime (Condition B4):** Our asymptotic analysis operates under the *dense functional data* framework. Specifically, we require the number of observations per curve, denoted by $n^{(g)}$, to grow at a rate strictly faster than $n^{5/4}$ (where n is the sample size). Under this dense sampling regime, combined with an optimal bandwidth choice of $h^{(g)} \sim (n^{(g)})^{-1/5}$, the local linear smoothers $\widehat{X}_i^{(g)}$ can reconstruct the true trajectories so accurately that the smoothing error becomes asymptotically negligible. Consequently, the estimated trajectories can serve as perfect substitutes for the true unobserved $X_i^{(g)}$ in our subsequent factor augmentation and variable selection procedures.

Lemma 1 *Under Conditions B1–B4, for $X_i^{(g)}(\cdot)$ which has the following properties,*

- $\mathbb{E}(X_i^{(g)}(t)) = \mathbb{E}(a_{0,ij}^{(g)}) = \mathbb{E}(a_{0,ij}^{(g)}a_{0,ik}^{(g)}) = \mathbb{E}(e_{0,i}^{(g)}(t)) = \mathbb{E}(a_{0,ij}^{(g)}e_{0i}^{(g)}(t)) = 0$, *where* $k \neq j$,
- $\mathbb{E}(X_i^{(g_1)}(s)e_{0i}^{(g_2)}(t)) = 0$ *for* $g_1 \neq g_2$, $g_1, g_2 = 1, \ldots, G$, $s, t \in \mathcal{T}$,
- $\mathrm{Var}(a_{0,i1}^{(g)}) > \mathrm{Var}(a_{0,i2}^{(g)}) > \ldots > \mathrm{Var}(a_{0,im_g}^{(g)})$,
- *all the eigenvalues of* $\mathrm{Cov}(e_{0,i}^{(g)}(s), e_{0,i}^{(g)}(t))$ *are less than* $\lambda_{\min}(\mathrm{Cov}(\boldsymbol{a}_{0,i}^{(g)}))$, *where* $\lambda_{\min}(\mathrm{Cov}(\boldsymbol{a}_{0,i}^{(g)}))$ *is the smallest eigenvalue of* $\mathrm{Cov}(\boldsymbol{a}_{0,i}^{(g)})$,

where the j-th eigenfunction $\gamma_j^{(g)}(\cdot)$ of $\mathrm{Cov}(X_i^{(g)}(s), X_i^{(g)}(t))$ is $\phi_{0j}^{(g)}(\cdot)$, with the j-th functional score $a_{ij}^{(g)} := \int_{\mathcal{T}} X_i^{(g)}(t)\gamma_j^{(g)}(t)dt = a_{0,ij}^{(g)}$.

Remark 1 *Lemma 1 states $\boldsymbol{a}_i^{(g)} = (a_{i1}^{(g)}, \ldots, a_{im_g}^{(g)})^\top$ is a reasonable approximation of $\boldsymbol{a}_{0,i}^{(g)}$. All eigenvalues of $\mathrm{Cov}(e_{0,i}^{(g)}(s), e_{0,i}^{(g)}(t))$ are required to be less than $\lambda_{\min}(\mathrm{Cov}(\boldsymbol{a}_{0,i}^{(g)}))$, indicating that segregation can be conducted based on contribution to the variance of $X_i^{(g)}(t)$. Furthermore, the orthogonality assumption across different functional covariates, $\mathbb{E}(X_i^{(g_1)}(s)e_{0i}^{(g_2)}(t)) = 0$ for $g_1 \neq g_2$, serves as a standard identification condition in factor-augmented models (Fan et al. 2020). It formally posits that once the pervasive common factors driving the strong cross-covariate correlations are extracted into the leading scores, the remaining high-frequency, covariate-specific idiosyncratic noises are mutually uncorrelated.*

The whole recovery process can be achieved by the popular functional principal component analysis (fPCA) with the Karhunen-Loéve (KL) expansion (Yao et al. 2005).

2.2 A functional factor augmentation structure

As correlations may exist between multivariate functional processes, we propose a functional factor augmentation structure (fFAS) to address the issue. Consider a simplified scenario with only two correlated functional processes, $X_i^{(1)}(t)$ and $X_i^{(2)}(t)$, generated by

$$X_i^{(1)}(t) = \boldsymbol{a}_{0,i}^{(1)\top} \cdot \boldsymbol{\phi}_0^{(1)}(t) + e_{0i}^{(1)}(t), \tag{2}$$
$$X_i^{(2)}(t) = \boldsymbol{a}_{0,i}^{(2)\top} \cdot \boldsymbol{\phi}_0^{(2)}(t) + e_{0i}^{(2)}(t),$$

where $e_{0i}^{(1)}(t)$ and $e_{0i}^{(2)}(t)$ are independent of each other. Assume $\boldsymbol{a}_{0,i}^{(g_1)}$ and $e_{0i}^{(g_2)}(t)$ are uncorrelated for $g_1, g_2 = 1, \ldots, G$, where $G = 2$, and hence the correlation between $X_i^{(1)}(t)$ and $X_i^{(2)}(t)$ arises only from that between $\boldsymbol{a}_{0,i}^{(1)}$ and $\boldsymbol{a}_{0,i}^{(2)}$. To capture such a correlation, it is assumed that each $\boldsymbol{a}_{0,i}^{(g)}$ shares K common underlying factors using a linear combination

$$\begin{pmatrix} \boldsymbol{a}_{0,i}^{(1)} \\ \boldsymbol{a}_{0,i}^{(2)} \end{pmatrix} = \begin{pmatrix} B^{(1)} \\ B^{(2)} \end{pmatrix} \boldsymbol{f}_i + \begin{pmatrix} \boldsymbol{u}_i^{(1)} \\ \boldsymbol{u}_i^{(2)} \end{pmatrix} = \boldsymbol{B}\boldsymbol{f}_i + \boldsymbol{u}_i, \tag{3}$$

where $\boldsymbol{B} = \left(B^{(1)\top}, B^{(2)\top}\right)^\top$ is a $(m_1+m_2)\times K$ factor loading matrix, $\boldsymbol{f}_i$ is a $K\times 1$ vectorized latent factor, and $\boldsymbol{u}_i$ is an idiosyncratic component independent of $\boldsymbol{f}_i$ which carries a weak correlation. Note that the covariance $C_{\boldsymbol{a}_0} := \mathrm{Cov}(\boldsymbol{a}_{0,i}) = \mathrm{Cov}(\boldsymbol{B}\boldsymbol{f}_i+\boldsymbol{u}_i) = \boldsymbol{B}\cdot\mathrm{Cov}(\boldsymbol{f}_i)\cdot\boldsymbol{B}^\top+$ $\mathrm{Cov}(\boldsymbol{u}_i)$. For model identifiability, it is assumed that $\mathrm{Cov}(\boldsymbol{f}_i) = \boldsymbol{I}_K$ and $\mathrm{Cov}(\boldsymbol{u}_i) = \omega\Lambda_u$ with a $p\times p$ matrix Λ_u, where $\|\Lambda_u\|_{\max} \le C_u$, a constant depending on the distribution of $\boldsymbol{u}_i$. By the spectral decomposition, $\boldsymbol{B} = \left(\sqrt{\lambda_1^{(B)}}\boldsymbol{\xi}_1^{(B)}, \ldots, \sqrt{\lambda_K^{(B)}}\boldsymbol{\xi}_K^{(B)}\right)$ with the eigenvalues $\lambda_1^{(B)} > \ldots > \lambda_K^{(B)}$ of $\boldsymbol{B}\boldsymbol{B}^\top$ and the corresponding eigenvectors $\{\boldsymbol{\xi}_k^{(B)}, k = 1, \ldots, K\}$. In this way, $X_i^{(1)}(t)$ and $X_i^{(2)}(t)$ in (2) can be expanded as

$$X_i^{(1)}(t) = \boldsymbol{\phi}_0^{(1)}(t)^\top(B^{(1)}\boldsymbol{f}_i + \boldsymbol{u}_i^{(1)}) + e_{0i}^{(1)}(t) = \tilde{\boldsymbol{\phi}}_0{}^{(1)}(t)^\top\boldsymbol{f}_i + \boldsymbol{\phi}_0^{(1)}(t)^\top\boldsymbol{u}_i^{(1)} + e_{0i}^{(1)}(t), \tag{4}$$
$$X_i^{(2)}(t) = \boldsymbol{\phi}_0^{(2)}(t)^\top(B^{(2)}\boldsymbol{f}_i + \boldsymbol{u}_i^{(2)}) + e_{0i}^{(2)}(t) = \tilde{\boldsymbol{\phi}}_0{}^{(2)}(t)^\top\boldsymbol{f}_i + \boldsymbol{\phi}_0^{(2)}(t)^\top\boldsymbol{u}_i^{(2)} + e_{0i}^{(2)}(t),$$

where $\tilde{\boldsymbol{\phi}}_0{}^{(g)}(t)^\top = \boldsymbol{\phi}_0^{(g)}(t)^\top B^{(g)}$. This indicates $X_i^{(g)}(t)$ can be decomposed into two correlated parts, namely, the functional factor part $\boldsymbol{f}_i$ and the weakly correlated part $\boldsymbol{\phi}_0^{(g)}(t)^\top\boldsymbol{u}_i^{(g)}$, plus an independent error term. We provide an example to illustrate such a fFAS when two functional covariates share a linear structure.

Example 1 *Suppose $X_i^{(1)}(\cdot)$ and $X_i^{(2)}(\cdot)$ are associated with a linear structure as*

$$X_i^{(2)}(t) = \mathbb{E}(X_i^{(2)}(t)|X_i^{(1)}(t)) + \epsilon_i(t) = \int_{\mathcal{T}} \beta(s,t) X_i^{(1)}(s)ds + \epsilon_i(t),$$

with $\beta(s,t) = \sum_{k=1}^{\infty}\sum_{m=1}^{\infty} \frac{\mathbb{E}(a_{0,im}^{(1)} a_{0,ik}^{(2)})}{\mathbb{E}(a_{0,im}^{(1)\,2})} \phi_{0k}^{(2)}(t)\phi_{0m}^{(1)}(s)$. Then it is obtained that

$$\begin{aligned}\int_{\mathcal{T}} \beta(s,t) X_i^{(1)}(s)ds &= \int_{\mathcal{T}} \beta(s,t) \sum_{j=1}^{\infty} a_{0,ij}^{(1)} \phi_{0j}^{(1)}(s)ds \\ &= \int_{\mathcal{T}} \sum_{k=1}^{\infty}\sum_{m=1}^{\infty} \frac{\mathbb{E}(a_{0,im}^{(1)} a_{0,ik}^{(2)})}{\mathbb{E}(a_{0,im}^{(1)\,2})} \phi_{0k}^{(2)}(t)\phi_{0m}^{(1)}(s) \sum_{j=1}^{\infty} a_{0,ij}^{(1)} \phi_{0j}^{(1)}(s)ds \\ &= \sum_{k=1}^{\infty} \left(\sum_{j=1}^{\infty} \frac{\mathbb{E}(a_{0,ij}^{(1)} a_{0,ik}^{(2)})}{\mathbb{E}(a_{0,ij}^{(1)\,2})} a_{0,ij}^{(1)} \right) \phi_{0k}^{(2)}(t),\end{aligned}$$

so that $a_{0,ik}^{(2)} \approx \sum_{j=1}^{\infty} \frac{\mathbb{E}(a_{0,ij}^{(1)} a_{0,ik}^{(2)})}{\mathbb{E}(a_{0,ij}^{(1)\,2})} a_{0,ij}^{(1)} = \sum_{j=1}^{m_1} \frac{\mathbb{E}(a_{0,ij}^{(1)} a_{0,ik}^{(2)})}{\mathbb{E}(a_{0,ij}^{(1)\,2})} a_{0,ij}^{(1)} + \sum_{j=m_1+1}^{\infty} \frac{\mathbb{E}(a_{0,ij}^{(1)} a_{0,ik}^{(2)})}{\mathbb{E}(a_{0,ij}^{(1)\,2})} a_{0,ij}^{(1)}$. With a truncation of the first m_1 and m_2 scores in $X_i^{(1)}(t)$ and $X_i^{(2)}(t)$, and combining them as $\boldsymbol{a}_{0,i} = (\boldsymbol{a}_{0,i}^{(1)\top}, \boldsymbol{a}_{0,i}^{(2)\top})^\top$, and a matrix $\boldsymbol{E}$ with elements $E_{kj} = \frac{\mathbb{E}(a_{0,ij}^{(1)} a_{0,ik}^{(2)})}{\mathbb{E}(a_{0,ij}^{(1)\,2})}$ and a diagonal matrix Λ with elements $\Lambda_{jj} = \mathbb{E}(a_{0,ij}^{(1)\,2})$, it is easily obtained that

$$\begin{pmatrix} \boldsymbol{a}_{0,i}^{(1)} \\ \boldsymbol{a}_{0,i}^{(2)} \end{pmatrix} \approx \begin{pmatrix} \boldsymbol{I}_{m_1} \\ \boldsymbol{E} \end{pmatrix} \boldsymbol{a}_{0,i}^{(1)} = \left(\begin{pmatrix} \boldsymbol{I}_{m_1} \\ \boldsymbol{E} \end{pmatrix} \Lambda^{\frac{1}{2}} \right) \left(\Lambda^{-\frac{1}{2}} \boldsymbol{a}_{0,i}^{(1)} \right) = \left(\begin{pmatrix} \boldsymbol{I}_{m_1} \\ \boldsymbol{E} \end{pmatrix} \Lambda^{\frac{1}{2}} \boldsymbol{P} \right) (\boldsymbol{P}^\top \Lambda^{-\frac{1}{2}} \boldsymbol{a}_{0,i}^{(1)}),$$

where $\boldsymbol{P}$ is an orthogonal matrix, so that $\left(\begin{pmatrix} \boldsymbol{I}_{m_1} \\ \boldsymbol{E} \end{pmatrix} \Lambda^{\frac{1}{2}} \boldsymbol{P} \right)^\top \left(\begin{pmatrix} \boldsymbol{I}_{m_1} \\ \boldsymbol{E} \end{pmatrix} \Lambda^{\frac{1}{2}} \boldsymbol{P} \right)$ is diagonal, and $\boldsymbol{f}_i = \boldsymbol{P}^\top \Lambda^{-\frac{1}{2}} \boldsymbol{a}_{0,i}^{(1)}$ with $\mathrm{Cov}(\boldsymbol{f}_i) = \boldsymbol{I}_{m_1}$. Note that no matter what values of m_1 and m_2 are in practice, one can always obtain such a functional factor augmentation structure.

Next, we consider a more general case where the structure $\boldsymbol{a}_{0,i}^{(g)}$ contains correlations for $g = 1, \ldots, G$, where $\mathbb{E}(a_{0,ij}^{(g)} a_{0,ik}^{(g)}) \neq 0$ (for all $0 < j < k < m_g$). Then, Lemma 2 shows the relation between $\boldsymbol{a}_i^{(g)}$ and $\boldsymbol{a}_{0,i}^{(g)}$ in this circumstance that $\boldsymbol{a}_i^{(g)}$ is still an approximation of $\boldsymbol{a}_{0,i}^{(g)}$ by imposing an orthogonal rotation induced by the basis functions in K-L expansion.

Lemma 2 *Under the conditions in Lemma 1 without* $\mathbb{E}(a_{0,ij}^{(g)} a_{0,ik}^{(g)}) = 0$, *there exists an orthogonal matrix* $P^{(g)}$, *such that* $\boldsymbol{\phi}_0^{(g)}(\cdot) = P^{(g)}(\gamma_1^{(g)}(\cdot), \ldots, \gamma_{m_g}^{(g)}(\cdot))^\top$, *and*

$$\begin{pmatrix} a_{0,i1}^{(g)} \\ a_{0,i2}^{(g)} \\ \vdots \\ a_{0,im_g}^{(g)} \end{pmatrix} = P^{(g)} \begin{pmatrix} a_{i1}^{(g)} \\ a_{i2}^{(g)} \\ \vdots \\ a_{im_g}^{(g)} \end{pmatrix}.$$

Furthermore, denote $\boldsymbol{a}_i^\top = (\boldsymbol{a}_i^{(1)\top}, \ldots, \boldsymbol{a}_i^{(G)\top})$ *and* $\boldsymbol{a}_{0,i}^\top = (\boldsymbol{a}_{0,i}^{(1)\top}, \ldots, \boldsymbol{a}_{0,i}^{(G)\top})$. *Then*

$$\mathrm{Cov}(\boldsymbol{a}_i) = \begin{pmatrix} P^{(1)} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & P^{(2)} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & P^{(G)} \end{pmatrix}^\top \mathrm{Cov}(\boldsymbol{a}_{0,i}) \begin{pmatrix} P^{(1)} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & P^{(2)} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & P^{(G)} \end{pmatrix},$$

where

$$\mathrm{Cov}(\boldsymbol{a}_i) = \begin{pmatrix} \Sigma_{m_1} & \Sigma_{m_1 m_2} & \cdots & \Sigma_{m_1 m_G} \\ \Sigma_{m_2 m_1} & \Sigma_{m_2} & \cdots & \Sigma_{m_2 m_G} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{m_G m_1} & \Sigma_{m_G m_2} & \cdots & \Sigma_{m_G} \end{pmatrix},$$

and

$$\mathrm{Cov}(\boldsymbol{a}_{0,i}) = \begin{pmatrix} \Sigma_{0,m_1} & \Sigma_{0,m_1 m_2} & \cdots & \Sigma_{0,m_1 m_G} \\ \Sigma_{0,m_2 m_1} & \Sigma_{0,m_2} & \cdots & \Sigma_{0,m_2 m_G} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{0,m_G m_1} & \Sigma_{0,m_G m_2} & \cdots & \Sigma_{0,m_G} \end{pmatrix}.$$

Remark 2 *Lemma 2 indicates that if there is a fFAS on* $\boldsymbol{a}_{0,i}$, *there will also be a fFAS on*

$\boldsymbol{a}_i$ *by the fact that*

$$\boldsymbol{a}_i = \begin{pmatrix} P^{(1)} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & P^{(2)} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & P^{(G)} \end{pmatrix}^\top \boldsymbol{a}_{0,i} = \boldsymbol{P}^\top \boldsymbol{a}_{0,i} = \boldsymbol{P}^\top \boldsymbol{B} \boldsymbol{f}_i + \boldsymbol{P}^\top \boldsymbol{u}_i, \qquad (5)$$

and treating $\boldsymbol{P}^\top \boldsymbol{B}$ *and* $\boldsymbol{P}^\top \boldsymbol{u}_i$ *as the updated loading matrix and the idiosyncratic component in* (3)*, respectively, where* $\boldsymbol{P}$ *is still an orthogonal matrix. More specifically, if* $\boldsymbol{a}_{0,i}$ *has a fFAS induced by* $\boldsymbol{f}_i$*,* $\boldsymbol{a}_i$ *also has such a fFAS with the same factor* $\boldsymbol{f}_i$*. Consequently, even if the functional covariates are correlated with each other, we can still use the functional scores* $\boldsymbol{a}_i$ *to estimate* $\boldsymbol{a}_{0,i}$ *as if they were uncorrelated.*

To further get the estimates of the loading matrix $\boldsymbol{B}$ and the functional factors $\boldsymbol{F} = (\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n)^\top \in \mathbb{R}^{n\times K}$ after obtaining the score matrix $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)^\top$ with $\boldsymbol{a}_i$ in (5), we employ the principal component analysis (Bai 2003) on the covariance of $\boldsymbol{A}$, and further obtain that $\widehat{\boldsymbol{U}} = \boldsymbol{A} - \widehat{\boldsymbol{F}}\widehat{\boldsymbol{B}}^\top$, where $\widehat{\boldsymbol{U}} = (\widehat{\boldsymbol{u}}_1, \ldots, \widehat{\boldsymbol{u}}_n)^\top \in \mathbb{R}^{n\times p}$. More specifically, the columns of $\widehat{\boldsymbol{F}}/\sqrt{n}$ are the eigenvectors of $\boldsymbol{A}\boldsymbol{A}^\top$ corresponding to the top K eigenvalues, and $\widehat{\boldsymbol{B}} = n^{-1}\boldsymbol{A}^\top \widehat{\boldsymbol{F}}$. This is similar to the case $\widehat{\boldsymbol{B}} = \left(\sqrt{\lambda_1}\boldsymbol{\xi}_1, \ldots, \sqrt{\lambda_K}\boldsymbol{\xi}_K\right)$ and $\widehat{\boldsymbol{F}} = \boldsymbol{A}\widehat{\boldsymbol{B}}\,\mathrm{diag}\left(\lambda_1^{-1} \ldots, \lambda_K^{-1}\right)$, where $\{\lambda_j\}_{j=1}^K$ and $\left\{\boldsymbol{\xi}_j\right\}_{i=1}^K$ are top K eigenvalues in descending order and their associated eigenvectors of the sample covariance matrix.

A practical issue is how to determine the number of factors K. Given that latent factors, loadings, and idiosyncratic components are all unobservable, a conditional sparsity-based Eigenvalue Ratio (ER) criterion is adopted (Ahn & Horenstein 2013, Fan et al. 2020). Specifically, let $\lambda_k\left(\boldsymbol{A}^\top\boldsymbol{A}\right)$ denote the k-th largest eigenvalue of $\boldsymbol{A}^\top\boldsymbol{A}$, $K_{\max}$ be a prespecified upper bound for K, and C_n be a constant dependent on n and $p = \sum_{g=1}^G m_g$.

The number of factors K is then determined by

$$\widehat{K} = \underset{k \leq K_{\max}}{\operatorname{argmin}} \frac{\lambda_{k+1}\left(\boldsymbol{A}^\top \boldsymbol{A}\right) + C_n}{\lambda_k\left(\boldsymbol{A}^\top \boldsymbol{A}\right) + C_n} \tag{6}$$

for a given C_n and $K_{\max}$. Alternative approaches include the information criteria proposed by Bai & Ng (2002) and Fan et al. (2013), referred to as IC, respectively. As will be shown in the section F.3 of the Supplementary Material, the traditional fraction of variance explained (FVE, e.g., Ramsay & Silverman 2005) may suffer from severe over-truncation as high-frequency noise increases, although it performs reliably under low noise. Similarly, traditional likelihood-based criteria, such as AIC (Akaike 1974) and BIC (Schwarz 1978), may completely fail due to insufficient penalization of the continuous noise spectrum. In contrast, the adopted spectral data-driven methods (ER and IC) robustly capture the true 'eigengap'.

Furthermore, the consistency of $\widehat{K}$ in this functional score context is empirically guaranteed. Since the functional principal component scores $\widehat{\boldsymbol{A}}$ maintain a bounded approximation error (Lemma 4), the sample eigenvalues effectively preserve the underlying principal eigengap. Extensive simulation studies confirm that the empirical probability of selecting the true factor number approaches 1 as the sample size n increases, corroborating the consistency of the ER criterion in our functional framework.

2.3 Properties of the functional factor augmentation structure

To establish the properties of the functional factor structure, we consider a spiked covariance framework. Specifically, let $\boldsymbol{\Sigma}_A = \mathbb{E}(\boldsymbol{a}_i \boldsymbol{a}_i^\top) = \boldsymbol{B}\boldsymbol{B}^\top + \boldsymbol{\Sigma}_u$ be the population covariance matrix of the scores, where $\boldsymbol{\Sigma}_u = \mathbb{E}(\boldsymbol{u}_i \boldsymbol{u}_i^\top)$. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ be the eigenvalues of $\boldsymbol{\Sigma}_A$. We assume the first K eigenvalues (factor strengths) scale with the dimension p, creating a significant separation from the idiosyncratic noise spectrum. We define the principal eigengap as:

$$\delta_n = \lambda_K(\mathbf{\Sigma}_A) - \lambda_{K+1}(\mathbf{\Sigma}_A). \tag{7}$$

By utilizing the Davis-Kahan theorem, the estimation error of the factor space is governed by the ratio of the sample noise to this spectral gap, rather than a scaling parameter ω. Let $\widehat{\mathbf{\Sigma}}_A = n^{-1}\widehat{\boldsymbol{A}}^\top\widehat{\boldsymbol{A}}$ denote the sample covariance matrix of the unobservable scores. The estimation error of the factor space depends on the ratio of the empirical covariance perturbation to the principal eigengap δ_n, rather than the absolute magnitude of the idiosyncratic variance. This approach ensures the fFASM framework remains theoretically valid in practical settings where idiosyncratic variance is non-negligible.

Lemma 3 *Suppose the principal eigengap satisfies $\delta_n > c > 0$ for some constant c. Let $\widehat{\boldsymbol{f}}_i, \widehat{\boldsymbol{B}}$ be the estimated factor and loading matrix. There exists an orthogonal rotation matrix $\boldsymbol{V}_0 \in \mathbb{R}^{K\times K}$ such that the estimation error is bounded by:*

$$\|\widehat{\boldsymbol{B}} - \boldsymbol{B}\boldsymbol{V}_0\|_F^2 = O_P\left(\frac{\|\widehat{\mathbf{\Sigma}}_A - \mathbf{\Sigma}_A\|_2^2}{\delta_n^2}\right),$$

$$\|\widehat{\boldsymbol{f}}_i - \boldsymbol{V}_0^\top\boldsymbol{f}_i\|_2^2 = O_P\left(\frac{\|\widehat{\mathbf{\Sigma}}_A - \mathbf{\Sigma}_A\|_2^2}{\delta_n^2} + \|\boldsymbol{u}_i^\top\boldsymbol{B}(\boldsymbol{B}^\top\boldsymbol{B})^{-1}\|_2^2\right).$$

Remark 3 *Lemma 3 clarifies that there are two fundamental error sources: (i) the spectral perturbation error, which is controlled by the eigengap δ_n, and (ii) the projection error of $\boldsymbol{u}_i$ onto the factor space. This framework remains valid even when the idiosyncratic variance is non-negligible, provided that the signal-to-noise ratio maintains the eigengap.*

In real applications, the scores $\boldsymbol{a}_i$ are not directly observable and must be estimated via fPCA. We establish the properties of these estimated scores $\widehat{\boldsymbol{A}}$ under the following regularity assumptions regarding the decay of functional eigenvalues.

Assumption 1 *(A1) For each $g = 1, \ldots, G$, the eigenvalues $\tau_k^{(g)}$ of the covariance kernel satisfy $\tau_k^{(g)} - \tau_{k+1}^{(g)} \geq Ck^{-a-1}$ for $k \geq 1$ and some $a > 1$.*

(A2) The truncation parameter s_n is chosen such that $(s_n^{2a+2} + s_n^{a+4})/n = o(1)$.

Assumption 2 *The functional processes $X_i^{(g)}(t)$ satisfy Conditions (B1)–(B4) in the Supplementary Material, ensuring the uniform consistency of the PACE-based trajectory estimation.*

Lemma 4 *For a fixed number G of trajectory types, under Assumptions 1 and 2, the empirical inner product of estimated functional scores concentrates around the population counterpart:*

$$\|n^{-1}\widehat{\boldsymbol{A}}^\top \widehat{\boldsymbol{A}} - \boldsymbol{\Sigma}_A\|_{\max} = O_p(1/\sqrt{n}).$$

Further, the individual score estimation error satisfies $\|\widehat{\boldsymbol{a}}_i - \boldsymbol{a}_i\|_2^2 \le O_P(1/n)$.

By combining the spectral perturbation bounds with the functional score concentration, we establish the high-probability error bounds for the functional structure factors.

Theorem 1 *Let δ_n be the principal eigengap. Under Assumptions 1 and 2, the estimated factors $\widehat{\boldsymbol{f}}_i$ and idiosyncratic components $\widehat{\boldsymbol{u}}_i$ satisfy:*

$$\begin{cases} \|\widehat{\boldsymbol{f}}_i - \boldsymbol{V}_0 \boldsymbol{f}_i\|_2^2 = O_P\left(\frac{p}{n\delta_n^2} + \frac{1}{n}\right), \\ \|\widehat{\boldsymbol{u}}_i - \boldsymbol{u}_i\|_2^2 = O_P\left(\frac{p}{n\delta_n^2} + \frac{1}{n}\right), \end{cases}$$

where $p = \sum_{g=1}^G m_g$ is the total dimension of the extracted functional scores.

Remark 4 *The error bounds established in Theorem 1 are derived under the regime where the number of functional covariates G and truncation levels m_g are fixed constants ($p = \sum m_g = O(1)$), while the sample size $n \to \infty$. It is worth noting that in a double-asymptotic high-dimensional setting where the total dimension $p \to \infty$ alongside n, the convergence rate would explicitly absorb a p-dependence. Based on matrix perturbation theory, the variance of the idiosyncratic noise would inflate the bound, roughly scaling the $O_P(1/n)$ term* *to $O_P(p/n)$ under standard strong-factor assumptions. Investigating the optimal variable selection rates under such double-divergent regimes poses a significant challenge and is left as a promising direction for future research.*

Remark 5 *While the fFASM treats latent factors primarily as a tool to absorb correlations for accurate variable selection, these factors themselves offer rich practical interpretability if one examines the estimated loading matrix $\widehat{\boldsymbol{B}}$. For example, in macroeconomic applications, a factor with high loadings on employment and national income reflects an underlying "economic vitality index," whereas a factor driven by literacy and fertility rates represents a "societal development index." Similar structural interpretations apply to industrial or medical data, analogous to interpretable functional factors like yield curve components (Hays et al. 2012) or biomechanical synergies (Castellanos et al. 2015).*

2.4 Truncation analysis and the relationship between K and m_g

A practical issue when modeling functional data with sample instances is how to determine the number of basis functions using truncation, i.e., m_g for $X_i^{(g)}(t)$, and how this truncation would affect the estimate of functional factors $\boldsymbol{f}_i$ in (3). Usually, m_g would be determined by the cumulative contribution to the variance from the corresponding functional scores, which may lead to an overestimated value of m_g, namely, $\widehat{m}_g > m_g$. To illustrate, assume $\widehat{m}_1 > m_1$ and $\widehat{m}_g = m_g$ for $g = 2, \ldots, G$, and define $\tilde{\boldsymbol{a}}_i = (\boldsymbol{a}_i^{(1)\top}, \tilde{\boldsymbol{a}}_i^{(1)\top}, \boldsymbol{a}_i^{(2)\top}, \ldots, \boldsymbol{a}_i^{(G)\top})^\top$ where $\tilde{\boldsymbol{a}}_i^{(g)\top}$ is a redundant variable for the factor model. In this case,

$$\mathrm{Cov}((\boldsymbol{a}_i^{(g)\top}, \tilde{\boldsymbol{a}}_i^{(g)\top})^\top) = \begin{pmatrix} \Sigma_{m_g} & \mathbf{0} \\ \mathbf{0} & \tilde{\Lambda}_0^{(g)} \end{pmatrix},$$

and

$$\mathrm{Cov}(\tilde{\boldsymbol{a}}_i) = \begin{pmatrix} \Lambda_{m_1} & \mathbf{0} & \Sigma_{m_1 m_2} & \cdots & \Sigma_{m_1 m_G} \\ \mathbf{0} & \tilde{\Lambda}_0^{(1)} & \mathbf{0} & \cdots & \mathbf{0} \\ \Sigma_{m_2 m_1} & \mathbf{0} & \Lambda_{m_2} & \cdots & \Sigma_{m_2 m_G} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Sigma_{m_G m_1} & \mathbf{0} & \Sigma_{m_G m_2} & \cdots & \Lambda_{m_G} \end{pmatrix}.$$

If $\xi^\star = (\xi_1^{\star\top}, \xi_2^{\star\top})^\top$ is one of the first K eigenvectors of $\mathrm{Cov}(\boldsymbol{a}_i)$, $(\xi_1^{\star\top}, \mathbf{0}, \xi_2^{\star\top})^\top$ is an eigenvector of $\mathrm{Cov}(\tilde{\boldsymbol{a}}_i)$ with the same eigenvalue, and the position of $\mathbf{0}$ in $(\xi_1^{\star\top}, \mathbf{0}, \xi_2^{\star\top})^\top$ corresponds to that of $\tilde{\Lambda}_0^{(1)}$ in $\mathrm{Cov}(\tilde{\boldsymbol{a}}_i)$. As $\widehat{\boldsymbol{F}} = \boldsymbol{A}\widehat{\boldsymbol{B}} \,\mathrm{diag}\left(\lambda_1^{-1} \ldots, \lambda_K^{-1}\right)$, $\tilde{\boldsymbol{a}}_i^{(g)}$ will not affect $\widehat{\boldsymbol{F}}$ since the row of $\boldsymbol{B}$ corresponding to $\tilde{\boldsymbol{a}}_i^{(g)}$ is $\mathbf{0}$ when $\widehat{K} = K$. Actually, in this overestimating situation, we can treat

$$\tilde{\boldsymbol{a}}_i = \begin{pmatrix} B_1 \\ \mathbf{0} \\ B_2 \end{pmatrix} \boldsymbol{f}_i + \begin{pmatrix} \boldsymbol{u}_i^{(1)} \\ \tilde{\boldsymbol{a}}_i^{(1)} \\ \boldsymbol{u}_i^{(C)} \end{pmatrix} = B^\star \boldsymbol{f}_i + \tilde{\boldsymbol{u}}_i,$$

as the real model with $\boldsymbol{u}_i = (\boldsymbol{u}_i^{(1)\top}, \boldsymbol{u}_i^{(C)\top})^\top$ and $\boldsymbol{B} = \left(B_1^\top, B_2^\top\right)^\top$.

Another case is that m_g may be underestimated for some specific $g_0 \leq G$, especially when ω and K are relatively small. In this case, a majority amount of variance of $X_i(t)$ is concentrated in $\tilde{\boldsymbol{\phi}}_0{}^{(g_0)}(t)^\top \boldsymbol{f}_i$ by (4). When $\tilde{\boldsymbol{\phi}}_0^{(g_0)}(t)$ are linearly independent for $g = 1, \ldots, G$, $\widehat{m}_{g_0}$ is determined as K by the variance contribution criteria. However, when they are correlated, an underestimation of m_{g_0} may not significantly affect the estimation of $\widehat{\boldsymbol{f}}_i$. To illustrate, consider that the covariance matrix of $\boldsymbol{a}_i$ with $G = 2$ as

$$\mathrm{Cov}(\boldsymbol{a}_i) = \begin{pmatrix} \Lambda_{m_1} & \Sigma_{m_1 m_2} \\ \Sigma_{m_2 m_1} & \Lambda_{m_2} \end{pmatrix},$$

with $K < m_1, m_2$, and for convenience, we assume $\Lambda_u = \boldsymbol{I}_p$. Since $B^{(g)}$ is an $m_g \times K$ matrix, $B^{(g)} B^{(g)\top}$ has at most K non-zero eigenvalues, denoted as $\nu_j^{(g)}, j = 1, \ldots, K$. By plugging in $\Lambda_{m_g} = P^{(g)\top} B^{(g)} B^{(g)\top} P^{(g)} + \omega \boldsymbol{I}_{m_g} = \mathrm{diag}\{\nu_1^{(g)} + \omega, \ldots, \nu_K^{(g)} + \omega, \omega, \ldots, \omega\}$, $\mathrm{Var}(\boldsymbol{a}_i)$ is written as

$$\mathrm{Cov}(\boldsymbol{a}_i) = \begin{pmatrix} \Lambda_{m_{1(K)}} & 0 & \Sigma_{m_{1(K)}m_{2(K)}} & \Sigma_{m_{1(K)}m_{2(C)}} \\ 0 & \Lambda_{m_{1(C)}} & \Sigma_{m_{1(C)}m_{2(K)}} & \Sigma_{m_{1(C)}m_{2(C)}} \\ \Sigma_{m_{2(K)}m_{1(K)}} & \Sigma_{m_{2(K)}m_{1(C)}} & \Lambda_{m_{2(K)}} & 0 \\ \Sigma_{m_{2(C)}m_{1(K)}} & \Sigma_{m_{2(C)}m_{1(C)}} & 0 & \Lambda_{m_{2(C)}} \end{pmatrix}$$
$$\approx \begin{pmatrix} \Lambda_{m_{1(K)}} & 0 & \Sigma_{m_{1(K)}m_{2(K)}} & 0 \\ 0 & 0 & 0 & 0 \\ \Sigma_{m_{2(K)}m_{1(K)}} & 0 & \Lambda_{m_{2(K)}} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

with $\Lambda_{m_g(K)} = \mathrm{diag}\{\tau_1^{(g)} + \omega, \ldots, \tau_K^{(g)} + \omega\}$ and $\Lambda_{m_g(C)} = \mathrm{diag}\{\omega, \ldots, \omega\}$. Consequently, when $\omega \approx 0$, underestimating m_g will not significantly affect $\widehat{\boldsymbol{f}_i}$ as long as $\widehat{m}_g \geq K$, which is easy to achieve in practice.

Regarding the practical selection of the FPCA truncation parameter $\hat{m}_g$, a key advantage of the fFASM framework is its theoretical robustness to over-truncation. Since the subsequent factor extraction step naturally filters out redundant idiosyncratic noise via the principal eigengap, under-truncation poses the primary risk of signal loss. Consequently, practitioners are advised to intentionally over-truncate by specifying a sufficiently large integer (e.g., $\hat{m}_g = 30$) to safely capture all structural functional signals. If data-driven variance thresholds are preferred, highly inclusive criteria (e.g., $\geq 99\%$ FVE) should be utilized rather than imposing restrictive caps, as the fFASM structure inherently absorbs and neutralizes the excess finite-sample noise..

3 A functional factor augmentation selection model

3.1 The functional linear regression model

In this section, we address a multivariate functional linear regression model with correlated functional covariates, using the proposed fFAS in Section 2. To start with, consider a functional linear regression model where a scalar response Y with $\mathbb{E}(Y) = \mu_Y$ is generated by a group of G functional covariates $\{X^{(1)}(t), X^{(2)}(t), \ldots, X^{(G)}(t); t \in \mathcal{T}\}$ as

$$Y = \mu_Y + \sum_{g=1}^{G} \int_{\mathcal{T}} \beta^{*(g)}(t) \left(X^{(g)}(t) - \mu^{(g)}(t)\right) dt + \epsilon, \tag{8}$$

where $\beta^{*(g)}(\cdot)$ is the square-integrable regression parameter function, and ϵ is a random noise with a zero mean and a constant variance σ^2. We note $\beta_0^* = \mu_Y$ as the intercept term. Thus, with given i.i.d samples $\{y_i, X_i^{(g)}(t), g = 1, \ldots, G; t \in \mathcal{T}, i = 1, \ldots, n\}$ which have the same distribution as $\{Y, X^{(g)}(t), g = 1, \ldots, G; t \in \mathcal{T}\}$, the sample functional linear regression model is described as

$$y_i = \beta_0^* + \sum_{g=1}^{G} \int_{\mathcal{T}} \beta^{*(g)}(t) \left(X_i^{(g)}(t) - \mu^{(g)}(t)\right) dt + \epsilon_i. \tag{9}$$

To detect active functional covariates with correlation, we develop a functional factor augmentation selection model (fFASM) as follows. Without loss of generalizability, we use $X_i^{(g)}(\cdot)$ and y_i as the centered functional covariates and scalar response variable, respectively, and accordingly (9) is equivalent to

$$y_i = \sum_{g=1}^{G} \int_{\mathcal{T}} \beta^{*(g)}(t) X_i^{(g)}(t) dt + \epsilon_i, \tag{10}$$

and can be further expanded by K-L expansion as

$$y_i = \sum_{g=1}^{G} \sum_{j=1}^{m_g} a_{ij}^{(g)} \eta_j^{*(g)} + \sum_{g=1}^{G} \int_{\mathcal{T}} \beta^{*(g)}(t) e_i^{(g)}(t) dt + \epsilon_i = \boldsymbol{H}^{*\top} \cdot \boldsymbol{a}_i + \tilde{\epsilon}_i,$$

where $\boldsymbol{H}^* = \left(\boldsymbol{\eta}^{*(1)\top}, \ldots, \boldsymbol{\eta}^{*(G)\top}\right)^\top \in \mathbb{R}^p$, and $\eta_j^{*(g)} = \int_{\mathcal{T}} \beta^{*(g)}(t)\gamma_j^{(g)}(t)dt$, with $\gamma_j^{(g)}(t)$ being the j-th eigenfunction of the covariance function for the g-th functional covariate ($j = 1, \ldots, m_g$). The modified error term is given by $\tilde{\epsilon}_i = \epsilon_i + \sum_{g=1}^G \int_{\mathcal{T}} \beta^{*(g)}(t)e_i^{(g)}(t)dt$, where the integral $\int_{\mathcal{T}} \beta^{*(g)}(t)e_i^{(g)}(t)dt$ has a zero mean for all $g = 1, \ldots, G$. Consequently, to select useful functional covariates $X_i^{(g)}(t)$ is to find $\beta^{*(g)}(\cdot)$ such that $\beta^{*(g)}(\cdot) \neq 0$, which is further assumed as $\boldsymbol{\eta}^{*(g)} \neq \mathbf{0}$. Plugging in the proposed fFAS, one easily obtains

$$y_i = \boldsymbol{H}^{*\top}\left(\boldsymbol{B}\boldsymbol{f}_i + \boldsymbol{u}_i\right) + \tilde{\epsilon}_i,$$

or equivalently,

$$\boldsymbol{y} = \boldsymbol{F}\boldsymbol{B}^\top\boldsymbol{H}^* + \boldsymbol{U}\boldsymbol{H}^* + \tilde{\boldsymbol{\epsilon}}, \tag{11}$$

and with $\widehat{\boldsymbol{F}}$, $\widehat{\boldsymbol{B}}$ and $\widehat{\boldsymbol{U}}$ obtained by the fFAS, and defining $\tilde{\boldsymbol{e}} = (\tilde{e}_1, \ldots, \tilde{e}_n)^\top = \left(\boldsymbol{A} - \widehat{\boldsymbol{A}} + \mathbb{E}(\widehat{\boldsymbol{A}})\right)\boldsymbol{H}^*$, (11) is further equivalent to

$$\left(\boldsymbol{I}_n - \boldsymbol{P}_{\widehat{\boldsymbol{F}}}\right)\boldsymbol{y} = \left(\boldsymbol{I}_n - \boldsymbol{P}_{\widehat{\boldsymbol{F}}}\right)\widehat{\boldsymbol{U}}\boldsymbol{H}^* + \left(\boldsymbol{I}_n - \boldsymbol{P}_{\widehat{\boldsymbol{F}}}\right)\left(\tilde{\boldsymbol{e}} + \tilde{\boldsymbol{\epsilon}}\right), \tag{12}$$

where $\boldsymbol{P}_{\widehat{\boldsymbol{F}}} = \widehat{\boldsymbol{F}}(\widehat{\boldsymbol{F}}^\top\widehat{\boldsymbol{F}})^{-1}\widehat{\boldsymbol{F}}^\top$ is the orthogonal projection matrix onto the column space $C(\widehat{\boldsymbol{F}})$. Consequently to select useful functional covariates, the penalized loss function

$$L_n\left(\boldsymbol{y}, \widehat{\boldsymbol{U}}\boldsymbol{H}, \widehat{\boldsymbol{F}}\right) = \frac{1}{n}\left\|(\boldsymbol{I} - P_{\widehat{\boldsymbol{F}}})(\boldsymbol{y} - \widehat{\boldsymbol{U}}\boldsymbol{H})\right\|_2^2 + \sum_{g=1}^{G}\sum_{k=1}^{m_g}\lambda\left\|\eta_k^{(g)}\right\|_1$$

is minimized with respect to $\boldsymbol{H}$, where $\|\cdot\|_1$ is a Lasso penalty controlled by the parameter λ and it can be set as the popular penalties such as SCAD or MCP with λ being selected using cross-validation. It is worth noting that before minimizing the penalized loss function, data preprocessing is necessary. Explicitly, the response variable $\boldsymbol{y}$ is centered, while the estimated functional scores (or equivalently, the decomposed factors $\widehat{\boldsymbol{F}}$ and idiosyncratic components $\widehat{\boldsymbol{U}}$) are standardized to have unit variance. Given that functional principal component scores inherently possess rapidly decaying variances, standardization ensures that the penalty parameter λ is scaled equitably across all basis components, thereby preventing scale-dependent selection bias. Hence, we successfully transform the problem from model selection with highly correlated functional covariates to model selection with weakly correlated or uncorrelated ones by lifting the space to a higher dimension. As $\widehat{\boldsymbol{H}} = \operatorname{argmin}_{\boldsymbol{H}} L_n(\boldsymbol{y}, \widehat{\boldsymbol{U}}\boldsymbol{H}, \widehat{\boldsymbol{F}})$ is obtained, $\beta^{(g)}(t)$ is estimated as

$$\widehat{\beta}^{(g)}(t) = \sum_{j=1}^{\widehat{m}_g} \widehat{\eta}_j^{(g)} \widehat{\gamma}_j^{(g)}(t).$$

Accordingly, when $\widehat{\beta}^{(g)}(t) \neq 0$, $X_i{}^{(g)}(t)$ is selected as a useful functional covariate. Note that the group selection method, such as the GM strategy by Aneiros, Novo & Vieu (2022), is not adopted in our case.

Also, this procedure may work even in the generalized linear context. Honestly,

$$\begin{aligned}
\boldsymbol{y} &= \left(\widehat{\boldsymbol{A}} - \mathbb{E}(\widehat{\boldsymbol{A}})\right)\boldsymbol{H}^* + \left(\boldsymbol{A} - \widehat{\boldsymbol{A}} + \mathbb{E}(\widehat{\boldsymbol{A}})\right)\boldsymbol{H}^* + \tilde{\boldsymbol{\epsilon}} \\
&= \widehat{\boldsymbol{F}}\widehat{\boldsymbol{B}}^\top \boldsymbol{H}^* + \widehat{\boldsymbol{U}}\boldsymbol{H}^* + \left(\boldsymbol{A} - \widehat{\boldsymbol{A}} + \mathbb{E}(\widehat{\boldsymbol{A}})\right)\boldsymbol{H}^* + \tilde{\boldsymbol{\epsilon}} \\
&= \widehat{\boldsymbol{F}}\widehat{\boldsymbol{B}}^\top \boldsymbol{H}^* + \widehat{\boldsymbol{U}}\boldsymbol{H}^* + \tilde{\boldsymbol{e}} + \tilde{\boldsymbol{\epsilon}}, \qquad (13)
\end{aligned}$$

indicating that the explanatory variables are switched from $\boldsymbol{A}$ to $\widehat{\boldsymbol{A}}$. In practice, when using the sample mean of $\widehat{\boldsymbol{A}}$ to substitute $\mathbb{E}(\widehat{\boldsymbol{A}})$, $\mathbb{E}(\widehat{\boldsymbol{a}}_i) \to 0$ under some regularization conditions (Kong et al. 2016). After centralizing $\widehat{\boldsymbol{A}}$, by $\left(\widehat{\boldsymbol{A}} - \mathbb{E}(\widehat{\boldsymbol{A}})\right)\boldsymbol{H}^* = \widehat{\boldsymbol{F}}\widehat{\boldsymbol{B}}^\top\boldsymbol{H}^* + \widehat{\boldsymbol{U}}\boldsymbol{H}^* = \widehat{\boldsymbol{F}}\boldsymbol{\gamma}^* + \widehat{\boldsymbol{U}}\boldsymbol{H}^*$, the unknown parameters are transformed into $(\boldsymbol{H}, \boldsymbol{\gamma})$, so that their corresponding covariates $\widehat{\boldsymbol{U}}$ and $\widehat{\boldsymbol{F}}$ are weakly correlated by introducing $\boldsymbol{\gamma}$. Note that in linear cases, $\boldsymbol{\gamma}^*$ is further eliminated by using the projection matrix in (12). Consequently, for generalized linear models and samples without centralization, the loss function is updated as

$$L_n\left(\boldsymbol{y}, \widehat{\boldsymbol{W}}\boldsymbol{\theta}\right) = \frac{1}{n}\sum_{i=1}^{n}\left[-y_i\left(\beta_0 + \widehat{\boldsymbol{f}}_i^\top\boldsymbol{\gamma} + \widehat{\boldsymbol{u}}_i^\top\boldsymbol{H}\right) + b\left(\beta_0 + \widehat{\boldsymbol{f}}_i^\top\boldsymbol{\gamma} + \widehat{\boldsymbol{u}}_i^\top\boldsymbol{H}\right)\right] + \sum_{g=1}^{G}\sum_{k=1}^{m_g}\lambda\left\|\eta_k^{(g)}\right\|_1 \qquad (14)$$

where $\widehat{\boldsymbol{w}}_i = \left(1, \widehat{\boldsymbol{u}}_i^\top, \widehat{\boldsymbol{f}}_i^\top\right)^\top$, $\widehat{\boldsymbol{W}} = (\widehat{\boldsymbol{w}}_1, \ldots, \widehat{\boldsymbol{w}}_n)$ and $\boldsymbol{\theta} = \left(\beta_0, \boldsymbol{H}^\top, \boldsymbol{\gamma}^\top\right)^\top$, β_0 is the intercept term and $b(\cdot)$ is a known function, where $b(z) = z^2/2$ in linear models. The estimate is given by $\left(\widehat{\beta}_0, \widehat{\boldsymbol{H}}^\top, \widehat{\boldsymbol{\gamma}}^\top\right)^\top = \operatorname{argmin}_{\boldsymbol{\theta}} L_n\left(\boldsymbol{y}, \widehat{\boldsymbol{W}}\boldsymbol{\theta}\right)$.

Based on the functional factor augmentation structure (fFAS) and the truncation analysis discussed above, we propose a computational procedure for the Functional Factor Augmentation Selection Model (fFASM). The procedure consists of three main stages: (1) performing marginal FPCA on each functional covariate to obtain the empirical truncation $\widehat{m}_g$ and the corresponding functional scores; (2) extracting the latent functional factors $\widehat{\boldsymbol{f}}_i$ from the stacked score vector to capture the common variations; and (3) solving the factor-augmented penalized regression problem to identify significant functional predictors. The detailed implementation is summarized in Algorithm 1.

Note that the functional goal of the proposed fFARM method tries to select the significant functional covariates to the response in the functional linear regression model, rather than the shared factors f_i or covariate-specific residual components $u_i^{(g)}$. After our FPCA and factor analysis decomposition into the shared factor component f_i and covariate-specific residual components $u_i^{(g)}$, this is equivalent to extracting non-zero sub-vectors of the coefficient vector $\boldsymbol{H}^*$, and such a step can be achieved by introducing group penalty over rows of $\boldsymbol{H}^*$. On one hand, if both f_i and $u_i^{(g)}$ contribute to the response, the covariate $\boldsymbol{X}^{(g)}$ will be selected for sure, as $\beta^{*(g)}(t) \neq 0$ regardless of f_i or $u_i^{(g)}$. On the other hand, if only $u_i^{(g)}$ contributes but not the shared parts, the covariate $\boldsymbol{X}^{(g)}$ will still be selected because $\beta^{*(g)}(t) \neq 0$. Thus, the proposed method will correctly identify that this functional predictor contains useful information for predicting the response variable y.

Remark 6 *(Computational Scalability) A major computational advantage of the proposed fFASM algorithm is its inherent scalability for extremely high-dimensional functional data.*

Algorithm 1 Estimation Procedure for fFASM

Require: Data $\{(Y_i, \{X_i^{(g)}(t)\}_{g=1}^{G})\}_{i=1}^{n}$, variance threshold τ, and tuning parameters (default $K_{\max} = 15$, $C_n = 0.01 \log(\min(n,p))/\min(n,p)$).

Ensure: Estimated functional coefficients $\widehat{\beta}_g(t)$ and selected active set $\widehat{\mathcal{S}}$.

Step 1: Marginal FPCA

for $g = 1, \ldots, G$ **do**

  Perform FPCA on $X_i^{(g)}(t)$ and determine truncation m_g via variance contribution τ.

  Extract the vector of functional scores $\widehat{\boldsymbol{a}}_i^{(g)} \in \mathbb{R}^{m_g}$.

end for

Step 2: Latent Factor Extraction (fFAS)

Stack all score vectors into $\widehat{\boldsymbol{a}}_i = ((\widehat{\boldsymbol{a}}_i^{(1)})^\top, \ldots, (\widehat{\boldsymbol{a}}_i^{(G)})^\top)^\top$.

Perform PCA on the stacked scores $\widehat{\boldsymbol{a}}_i$ to determine factor number $\widehat{K}$.

Compute the estimated latent factors $\widehat{\boldsymbol{f}}_i \in \mathbb{R}^{\widehat{K}}$.

Step 3: Augmented Variable Selection

Center the response y_i and standardize the functional scores $\widehat{\boldsymbol{a}}_i^{(g)}$ and factors $\widehat{\boldsymbol{f}}_i$.

Construct an augmented regression of y_i on factors $\widehat{\boldsymbol{f}}_i$ (unpenalized) and scores $\widehat{\boldsymbol{a}}_i^{(g)}$ (penalized).

Solve the optimization using a group Lasso to select significant g.

return Reconstructed functional coefficients $\widehat{\beta}_g(t)$ and the estimated set $\widehat{\mathcal{S}} = \{g : \widehat{\beta}_g(t) \not\equiv 0\}$.

By avoiding the construction and eigen-decomposition of a joint dense covariance matrix across all G functional covariates (which suffers from the curse of dimensionality), fFASM performs marginal functional principal component analysis and extracts latent factors on a strictly dimension-reduced space. This design ensures that the algorithm scales linearly with the number of functional covariates G. For a detailed step-by-step theoretical complexity analysis and extensive empirical runtime evaluations, please refer to Section F.4 in the Supplementary Material.

3.2 Theoretical justifications for functional variable selection

In this section, the proposed method is theoretically investigated under the general linear model context in (14), and hence the linear model will be covered as a special case. To start with, some notations and assumptions are introduced. Recall $p = \sum_{g=1}^{G} m_g$, and define $p_1 = 1 + p$, $\boldsymbol{H}_1 = (\beta_0, \boldsymbol{H}^\top)^\top$, $\boldsymbol{H}_1^* = (\beta_0^*, \boldsymbol{H}^{*\top})^\top$, $\boldsymbol{\theta}^* = (\beta_0^*, \boldsymbol{H}^{*\top}, (\boldsymbol{B}^\top \boldsymbol{H}^*)^\top)^\top$, $S = \mathrm{supp}(\boldsymbol{\theta}^*)$, $S_1 = \mathrm{supp}(\boldsymbol{H}_1^*)$, and $S_2 = [p_1 + K] \backslash S$. Suppose $\widehat{\boldsymbol{F}}$ and $\widehat{\boldsymbol{U}}$ are obtained given $\boldsymbol{A}$. We write

$$X_i^{(g)}(t) = X_{f,i}^{(g)}(t) + X_{u,i}^{(g)}(t) + e_{0i}^{(g)}(t), \tag{15}$$

with $X_{f,i}^{(g)}(t) = \boldsymbol{\gamma}^{(g)}(t)^\top B^{(g)} \boldsymbol{f}_i$ and $X_{u,i}^{(g)}(t) = \boldsymbol{\gamma}^{(g)}(t)^\top \boldsymbol{u}_i^{(g)}$.

Assumption 3 *(Smoothness). $b(z) \in C^3(\mathbb{R})$, i.e., for some constants M_2 and M_3, $0 \le b''(z) \le M_2$ and $|b'''(z)| \le M_3, \forall z$.*

Assumption 4 *(Restricted strong convexity and irrepresentable condition). For $\boldsymbol{W} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n)$ where $\boldsymbol{w}_i = \left(1, \boldsymbol{u}_i^\top, \boldsymbol{f}_i^\top\right)^\top$, there exist $\kappa_2 > \kappa_\infty > 0$ and $\tau \in (0, 0.5)$ such that*

$$\textit{(Convexity)} \left\| \left[\nabla_{SS}^2 L_n\left(\boldsymbol{y}, \boldsymbol{W}\boldsymbol{\theta}^*\right)\right]^{-1} \right\|_\ell \le \frac{1}{4\kappa_\ell}, \qquad \textit{for } \ell = 2 \textit{ and } \infty, \tag{16}$$

$$\textit{(Irrepresentable condition)} \left\| \nabla_{S_2 S}^2 L_n\left(\boldsymbol{y}, \boldsymbol{W}\boldsymbol{\theta}^*\right) \left[\nabla_{SS}^2 L_n\left(\boldsymbol{y}, \boldsymbol{W}\boldsymbol{\theta}^*\right)\right]^{-1} \right\|_\infty \le 1 - 2\tau. \tag{17}$$

Assumption 5 *(Estimation of factor model).* $\|\gamma_j^{(g)}(\cdot)\|_\infty \leq M_\gamma$ *for* $j = 1, \ldots, m_g,\ g = 1, \ldots, G,$ *and* $\|X_{f,i}^{(g)}(\cdot)\|_\infty \leq \frac{1}{2} M_\gamma M_0 \sqrt{\lambda_1^{(B)}}$ *and* $\|X_{u,i}^{(g)}(\cdot)\|_\infty \leq \frac{1}{2} M_\gamma M_0$ *for some constant* $M_0 > 2$*. In addition, there exists a* $K \times K$ *nonsingular matrix* $\boldsymbol{V}_0$*, and* $\boldsymbol{V} = \begin{pmatrix} \boldsymbol{I}_{p_1} & \mathbf{0}_{p_1 \times K} \\ \mathbf{0}_{K \times p_1} & \boldsymbol{V}_0 \end{pmatrix}$ *such that for* $\overline{\boldsymbol{W}} = \widehat{\boldsymbol{W}} \boldsymbol{V}$*, we have* $\|\overline{\boldsymbol{W}} - \boldsymbol{W}\|_{\max} \leq \frac{M_0}{2}$ *and* $\max_{j \in [p_1+K]} \left(\frac{1}{n} \sum_{i=1}^n |\bar{w}_{ij} - w_{ij}|^2 \right)^{1/2} \leq \frac{2\kappa_\infty \tau}{3 M_0 M_2 |S|}$*, with* w_{ij} *and* $\bar{w}_{ij}$ *being the* (i,j)*-th element of* $\mathbf{W}$ *and* $\overline{\mathbf{W}}$*.*

Assumption 3 holds for a large family of generalized linear models. For example, linear model has $b(z) = \frac{1}{2} z^2$, $M_2 = 1$, and $M_3 = 0$; logistic model has $b(z) = \log(1 + e^z)$ and finite M_2, M_3. Assumption 4 is easily satisfied with a small matrix and holds with high probability as long as $\mathbb{E}\left[\nabla^2 L_n(\boldsymbol{y}, \boldsymbol{W}\boldsymbol{\theta}^*)\right]$ satisfies similar conditions by standard concentration inequalities (Merlevède et al. 2011). Assumption 5 indicates the norm $\|\boldsymbol{W}\|_{\max}$ when $X_i^{(g)}(\cdot)$ and $\varepsilon_i^{(g)}(\cdot)$ is bounded, and Lemma 3 will be satisfied with a high probability when n is large.

Assumption 4 holds at the population level with high probability because the extracted functional factors $\boldsymbol{f}_i$ and the idiosyncratic residual components $\boldsymbol{u}_i$ are mutually uncorrelated by construction. This orthogonality guarantees a well-conditioned population Hessian matrix, thus satisfying the Restricted Strong Convexity (RSC) and Irrepresentable Condition (IRC). Furthermore, $\boldsymbol{V}_0$ in Assumption 5 represents the standard PCA rotation matrix used to align the estimated latent space with the population factor space. Throughout our asymptotic analysis, we consider the regime where the number of functional covariates G and the total dimension of extracted scores p are fixed constants (i.e., $G, p = O(1)$), while the sample size $n \to \infty$. If the dimension p were allowed to diverge ($p \to \infty$), the error bound in the factor estimation would correspondingly inflate to $O_P(p/n)$. Furthermore, with $J(\cdot) = |\cdot|$ and $\widehat{\boldsymbol{H}}$ obtained by minimizing (14), the following result is established.

Theorem 2 *Suppose Assumptions 3-5 holds. Define* $M = M_0^3 M_3 |S|^{3/2}$, *and*

$$\varepsilon^* = \max_{j \in [p_1+K]} \left| \frac{1}{n} \sum_{i=1}^{n} \bar{w}_{ij} \left[-y_i + b' \left(\boldsymbol{a}_i^T \boldsymbol{H}^* + \beta_0^* \right) \right] \right|.$$

If $\frac{7\varepsilon^*}{\tau} < \lambda < \frac{\kappa_2 \kappa_\infty \tau}{12M\sqrt{|S|}}$, *then* $\mathrm{supp}(\widehat{\boldsymbol{H}}) \subseteq \mathrm{supp}(\boldsymbol{H}^*)$ *and*

$$\left\| \widehat{\boldsymbol{H}} - \boldsymbol{H}^* \right\|_\infty \le \frac{6\lambda}{5\kappa_\infty}, \quad \left\| \widehat{\boldsymbol{H}} - \boldsymbol{H}^* \right\|_2 \le \frac{4\lambda\sqrt{|S|}}{\kappa_2}, \quad \left\| \widehat{\boldsymbol{H}} - \boldsymbol{H}^* \right\|_1 \le \frac{6\lambda|S|}{5\kappa_\infty}.$$

In addition, if $\varepsilon^* < \frac{\kappa_2 \kappa_\infty \tau^2}{12CM\sqrt{|S|}}$ *and* $\min\left\{ \left| \boldsymbol{H}_{1j}^* \right| : \boldsymbol{H}_{1j}^* \neq 0, j \in [p_1] \right\} > \frac{6C\varepsilon^*}{5\kappa_\infty \tau}$ *hold for some* $C > 7$, *where* $\boldsymbol{H}_{1j}^*$ *means the j-th element of* $\boldsymbol{H}_1^*$, *then by taking* $\lambda \in \left(\frac{7}{\tau}\varepsilon^*, \frac{C}{\tau}\varepsilon^* \right)$ *the sign consistency* $\mathrm{sign}(\widehat{\boldsymbol{H}}) = \mathrm{sign}(\boldsymbol{H}^*)$ *is achieved.*

While Theorem 2 explicitly establishes selection consistency under the Group Lasso penalty (which relies on a generalized irrepresentable condition), the theoretical properties of the fFASM framework are fundamentally robust across other penalty forms, such as Group SCAD and Group MCP. These non-convex alternatives can relax the irrepresentable condition and achieve the oracle property, offering varying trade-offs between computational convexity and estimation bias. A comprehensive theoretical discussion, accompanied by empirical sensitivity analyses and practical guidelines for penalty selection, is provided in Section F.5 of the Supplementary Material.

Theorem 2 guarantees the selection consistency of the functional covariates by developing sign consistency under some mild conditions. Furthermore, it demonstrates the relationship between $\left\| \widehat{\boldsymbol{H}} - \boldsymbol{H}^* \right\|$ and λ whose value depends on ε^* and τ, where ε^* comes from the first-order partial derivative of the empirical loss function (without penalty) when the parameters are known and τ satisfying a generalized irrepresentable condition from (17) (Lee et al. 2015). When ε^* is small, selecting an appropriate λ leads to satisfactory performance in the model estimation. Furthermore, when $\boldsymbol{A}$ is not available, $\widehat{\boldsymbol{A}}$ will be estimated and employed. Accordingly, the sign consistency is further guaranteed by showing that the loss function (14) based on $\widehat{\boldsymbol{A}}$ will have an asymptotic property as follows.

Theorem 3 *Let* $L_n\left(\boldsymbol{y}, \widehat{\boldsymbol{W}}\boldsymbol{\theta}\right)|_{\boldsymbol{A}}$ *be the loss function in* (14)*, based on which* $\widehat{\boldsymbol{f}}_i$ *and* $\widehat{\boldsymbol{u}}_i$ *are obtained using the true* $\boldsymbol{A}$*, and* $L_n\left(\boldsymbol{y}, \widehat{\boldsymbol{W}}\boldsymbol{\theta}\right)|_{\widehat{\boldsymbol{A}}}$ *using* $\widehat{\boldsymbol{A}}$*. Then for any fixed* $\boldsymbol{\theta}$*, by Assumption 3 and Lemma 4,*

$$\left|L_n\left(\boldsymbol{y}, \widehat{\boldsymbol{W}}\boldsymbol{\theta}\right)|_{\boldsymbol{A}} - L_n\left(\boldsymbol{y}, \widehat{\boldsymbol{W}}\boldsymbol{\theta}\right)|_{\widehat{\boldsymbol{A}}}\right| \to 0 \text{ in probability.}$$

Corollary 1 *Suppose* $L_n\left(\boldsymbol{y}, \widehat{\boldsymbol{W}}\boldsymbol{\theta}\right)$ *has a unique global minimum. Then* $L_n\left(\boldsymbol{y}, \widehat{\boldsymbol{W}}\boldsymbol{\theta}\right)|_{\boldsymbol{A}}$ *and* $L_n\left(\boldsymbol{y}, \widehat{\boldsymbol{W}}\boldsymbol{\theta}\right)|_{\widehat{\boldsymbol{A}}}$ *will have an identical global minimizer when* n *is sufficiently large, so that the estimates from* $L_n\left(\boldsymbol{y}, \widehat{\boldsymbol{W}}\boldsymbol{\theta}\right)|_{\widehat{\boldsymbol{A}}}$ *will also have sign consistency.*

4 Simulation Studies

In this section, the performance of the proposed fFASM estimator is examined for correlated functional data using simulation studies. To start with, the functional covariates $\{X_i^{(g)}(t)\}_{g=1}^G$ are generated by $X_i^{(g)}(t) = \boldsymbol{a}_{0,i}^{(g)\top} \cdot \boldsymbol{\phi}_{0,m_g}^{(g)}(t) + \varepsilon_i(t)$ with $\varepsilon_i(\cdot)$ independently and identically (i.i.d) drawn from $N(0, 0.25)$, and sampled over 51 uniformly distributed grid points for t from $\mathcal{T} = [0,1]$ with interval lengths of 0.02, where G is set as 20, 50, 100 and 150, respectively. The basis functions $\boldsymbol{\phi}_{0,m_g}^{(g)}(t) = \boldsymbol{\phi}_m(t) = (\phi_1(t), \ldots, \phi_m(t))^\top$ are Fourier basis with m set as 10. In all subsequent estimations, the truncation parameters are consistently selected via the Eigenvalue Ratio (ER) criterion, with the maximum search bound set to $K_{\max} = 15$ and the numerical stabilizer practically chosen as $C_n = 0.01 \log(\min(n,p))/\min(n,p)$. To blend correlations into functional covariates $X_i^{(g)}(t)$, two scenarios are considered when $\boldsymbol{a}_i$ is generated:

- Scenario I: the factor model case, i.e., $\boldsymbol{a}_{0,i} = \left(\boldsymbol{a}_{0,i}^{(1)\top}, \ldots, \boldsymbol{a}_{0,i}^{(G)\top}\right)^\top = \boldsymbol{B}\boldsymbol{f}_i + \boldsymbol{u}_i$, where the elements of $\boldsymbol{f}_i \in \mathbb{R}^K$, $\boldsymbol{u}_i \in \mathbb{R}^{Gm}$ and $\boldsymbol{B} \in \mathbb{R}^{Gm\times K}$ are generated from $N(0, 25)$,

$N(0,1)$, and $N(0,1)$, respectively. The true number of factors K is set as $1,2,\ldots,6$, respectively;

- Scenario II: the equal correlation case, i.e., $\boldsymbol{a}_{0,i}$ are generated from a multivariate normal distribution $\boldsymbol{a}_{0,i} \sim N(\mathbf{0}, \Sigma_{Equal})$ where Σ_{Equal} has diagonal elements as 1 and off-diagonal elements as ρ. The true value of ρ is set as $0, 0.1, 0.2 \ldots, 0.9$, respectively.

Furthermore, the true functional coefficients $\beta^{*(g)}$ in (10) are generated $\beta^{*(1)}(t) = \beta^{*(2)}(t) = \beta^{*(3)}(t) = \beta^{*(4)}(t) = (1, 1/2^2, \ldots, 1/m^2) \cdot \boldsymbol{\phi}_m(t)$ (the Harmonic attenuation type), $\beta^{*(5)}(t) = \beta^{*(6)}(t) = (0, 1/2^2, 0, \ldots, 0) \cdot \boldsymbol{\phi}_m(t)$ (the weak single signal type), and $\beta^{*(g)}(t) = 0$ for $g = 7, 8, \ldots, G$. Accordingly, the response y_i is generated by (10) with $\epsilon_i \sim N(0, 0.1)$.

We compare the performance of the proposed fFASM method (Lasso penalty) with other functional variable selection methods, including the Lasso (`Lasso`) and the group Lasso (`grLasso`) methods on the synthetic data in different scenarios. To evaluate the performance, three different measurements are adopted, i.e, the model size defined as the cardinality of the set $\left\{g : \widehat{\beta}^{(g)}(t) \neq 0\right\}$, the integrated mean squared error (IMSE)

$$IMSE = \sum_{g=1}^{G} \mathbb{E}\left(\left\|\widehat{\beta}^{(g)}(t) - \beta^{*(g)}(t)\right\|_{L^2}^2\right) = \sum_{g=1}^{G} \mathbb{E}\int_{\mathcal{T}} \left(\widehat{\beta}^{(g)}(t) - \beta^{*(g)}(t)\right)^2 dt,$$

and the true positive rate (TPR)

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}},$$

where TP represents the number of correctly predicted nonzero instances of $\widehat{\beta}^{(g)}(t)$ (events that are actually positive and predicted as positive), and FN for that of instances that are actually nonzero but are predicted as zero. Note that a model with the TPR of 1 and the model size of 6 indicates perfect recovery. The proposed fFASM method employs the Lasso regularization, and the hyperparameters are tuned using cross-validation by minimizing the cross-validation error with a randomly selected subset of $[n/3]$ training samples, where the sample size n is set as 100. The whole experiment is repeated for 500 times, and the averaged results are reported.

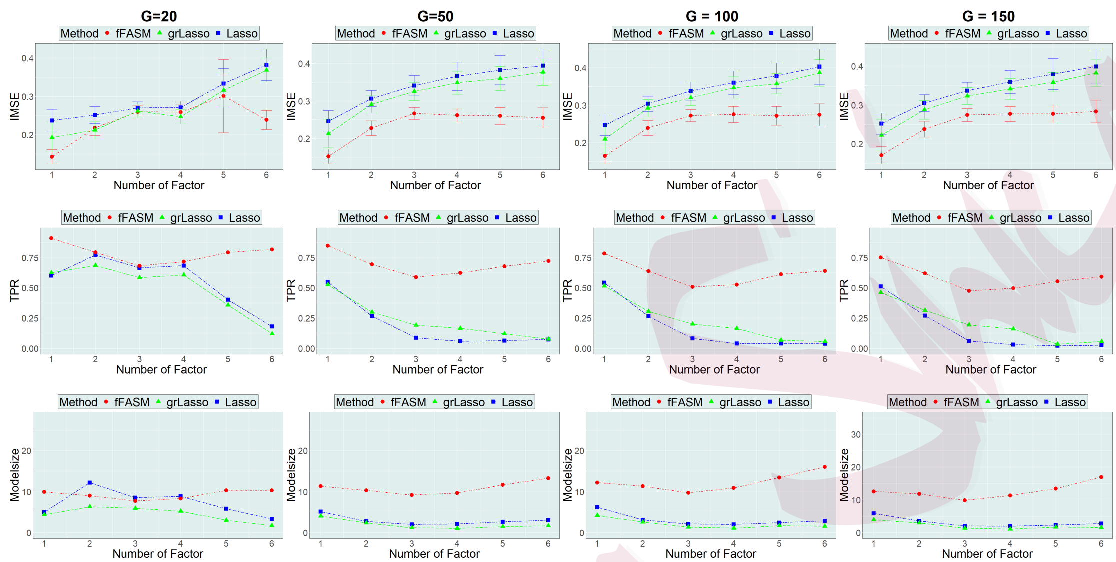

Figure 1: Estimation and selection performance of three methods with different numbers of factors K for Scenario I. In each row, four subfigures are displayed with $G = 20, 50, 100, 150$, respectively, and in each column with a given K, the IMSE (with 0.5 standard deviations), the TPR, and model size are presented, respectively.

Figure 1 shows the estimation and selection performance for Scenario I (the factor model structure case) using different methods. As is easily observed, the proposed fFASM method significantly and consistently outperforms Lasso and grLasso in the sense that the proposed fFASM has the smallest IMSE for each K, demonstrating a more accurate estimation of the functional coefficients. Furthermore, the TPR for the proposed fFASM method is always greater than those of Lasso and grLasso as K increases, and Lasso and grLasso even show dramatic drops when K increases. In terms of model size, the proposed fFASM tends to select more functional covariates into the model, slightly greater than the true size of 6, while the Lasso and grLasso methods tend to select irrelevant functional covariates into the model. When looking deeper at the selection results, the grLasso method tends to be more conservative in selecting functional covariates than Lasso.

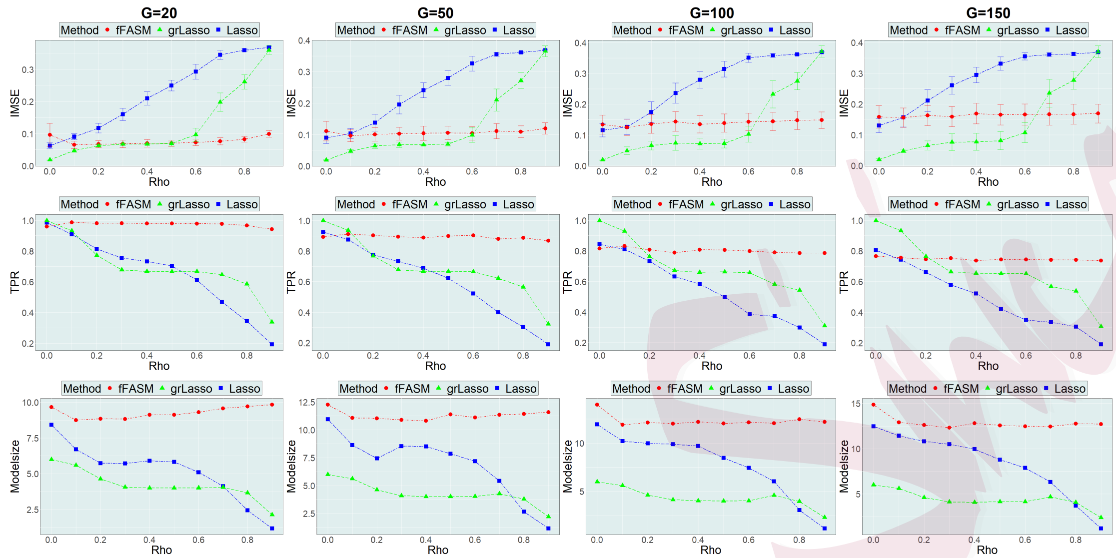

Figure 2: Estimation and selection performance of three methods with different correlations for Scenario II. In each row, four subfigures are displayed with $G = 20, 50, 100, 150$, respectively, and in each column with a given ρ, the IMSE (with 0.5 standard deviations), the TPR, and model size are presented, respectively.

Furthermore, Figure 2 shows the estimation and selection results in Scenario II (the equal correlation case). The IMSE values of the proposed fFASM method remain at a low level when the correlation ρ increases in each subfigure in the first row, indicating robust estimation performance from fFASM, while Lasso and grLasso tend to increase, demonstrating less satisfactory estimation performance. Although the two methods show slightly better results when functional covariates have small correlation with each other, this is expected since equal correlation indicates a small number of factors, which may be captured easily. In terms of TPR, the fFASM method is again consistently close to 1 when the correlation varies, while TPRs for the competitor methods drop dramatically. Also, when the number of functional covariates G increases, the TPRs from all three methods turn out to drop without surprise. When looking at the model size, the proposed fFASM remains stable, which may slightly overestimate the model size, while the competitor models show a much less stable model size with little consistency.

Additionally, the selection frequencies of the two types of functional coefficients $\beta^{*(g)}(t)$ are examined, namely the Harmonic attenuation type (`Type1`) and the weak single signal type (`Type2`), displayed in Figure 3 and 4, respectively. Note that the number of true `Type1` functional covariates is 4 and that of `Type2` is 2 in both scenarios. On one hand, Figure 3 shows that the selection frequency of the proposed fFASM method decreases first and then increases when the number of factors K increases in both two types of functional covariates, while those for Lasso and grLasso tend to be 0, especially in the `Type2` setting where the two competitor methods show lower starting points. Note that the selection frequencies of the proposed method for both two types are not extraordinarily high, since the functional linear model is somewhat lack of fit due to the complexity of the true model. On the other hand, in Scenario II, Figure 4 shows a much improved selection consistency of the proposed method than grLasso and Lasso for both two types, even though `Type2` may still select fewer covariates because of slight violation of norm requirement in Theorem 2.

5 Real Data Application

5.1 Effects of macroeconomic covariates on lifetime expectancy

In this section, the effects of macroeconomic covariates on national average lifetime expectancy are explored using the open-sourced EPS data platform [1]. The data have been

[1] https://www.epsnet.com.cn/

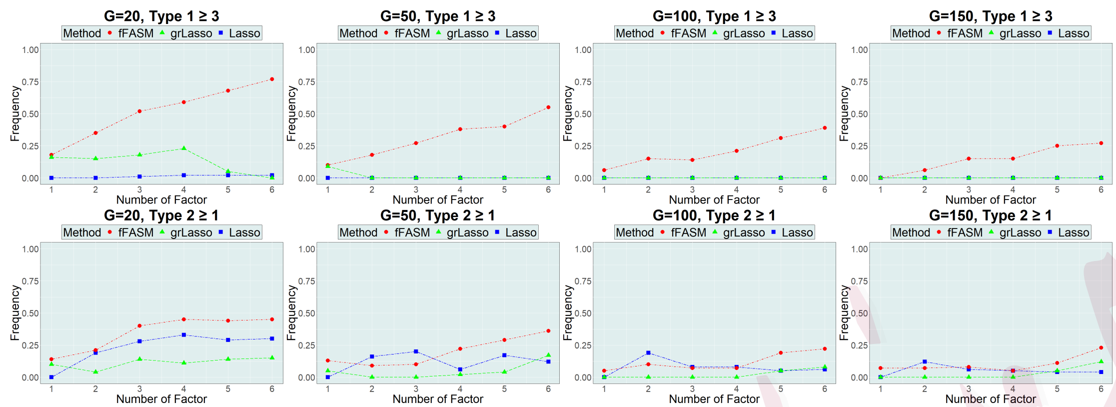

Figure 3: The selection frequencies of truly useful functional covariates in Scenario I with different K and G values. The first row displays the selection frequency that at least three out of four truly useful ones are selected for `Type1`, and the second displays the frequency that at least one out of two for `Type2`.

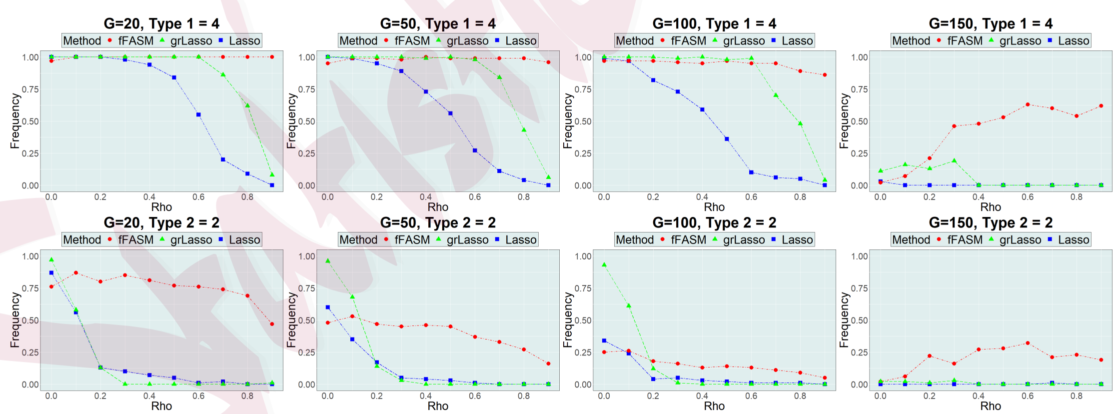

Figure 4: The selection frequencies of truly useful functional covariates in Scenario II with different ρ and G values. The first and second rows display the frequency that all truly useful ones are selected for `Type1` and `Type2`, respectively.

collected and documented annually over a span of 21 years, from 2000 to 2020, for 40 countries or regions. Our focus is to predict the average lifetime expectancy as a scalar dependent variable against 33 macroeconomic functional covariates, such as gross domestic product, healthcare expenditure, employment, and educational attainment. Functional covariates are intrinsically correlated with each other. For instance, higher levels of educational attainment often correlate with higher gross national income (GNI), as a more educated workforce typically contributes to greater economic productivity.

To handle the high-dimensional functional covariates and explicitly address the strong correlations among them, our proposed fFASM (Factor-Adjusted Sparse Model, executed via FarmSelect) is employed and compared against three baseline approaches: the standard Lasso, the Group Lasso (grLasso), and the true Multivariate Functional Principal Component Analysis (MFPCA) proposed by Chiou et al. (2014). Note that the joint MFPCA fuses all functional variables into a global dynamic system, which eliminates the ability to select single specific variables but serves as the theoretical upper bound for the predictive capability of the dataset.

During the study, 30 randomly selected countries (or regions) out of 40 are used as the training set to build the models, while the remaining 10 are used as the out-of-sample test set. To ensure the robustness of the evaluation given the highly volatile nature of macroeconomic indicators, the whole procedure is repeated 200 times. The worst 20 iterations (extreme outliers due to specific small-sample draws) are trimmed, and the average out-of-sample R^2, and model sizes are reported in Table 1. The top five most frequently selected functional covariates are presented in Table 1 of the Supplementary Material. As is presented, even though the joint MFPCA achieves the highest predictive accuracy, variable selection is less interpretable. In comparison, among the variable-selection methods, both the fFASM and standard Lasso exhibit highly stable and competitive predictive power, while the grLasso model completely collapses in out-of-sample prediction, primarily because forcing the inclusion or exclusion of entire functional covariate structures introduces heavy noise in small-sample settings. Consequently, the necessity and advantages of the proposed fFASM approach may be supported.

In terms of variable selection, the proposed fFASM successfully balances predictive power with model sparsity. By extracting and stripping away latent common macroeconomic factors, fFASM yields a slightly sparser average model size (11.89) compared to the standard Lasso (12.86) without sacrificing any prediction accuracy. From the selection frequencies, both fFASM and Lasso successfully identify core socioeconomic drivers, consistently selecting “Employment”, “Adult literacy rate”, “Total fertility rate”, “Deposit rate”, and “GNI per capita”. Meanwhile, the over-penalized grLasso yields an average model size of only 4.46 and strictly focuses on “GNI per capita”, “Employment”, and “Undernourishment rate”, missing several key nuanced indicators. Ultimately, the fFASM framework demonstrates its superiority in identifying purely heterogeneous functional drivers while stripping away global macroeconomic co-movements.

Method	**Mean Out-of-sample R^2**	**Average Model Size**
Baseline Lasso	0.2781	12.86
fFASM	0.2792	11.89
grLasso	-0.8850	4.46
MFPCA (Chiou & Müller 2016)	0.4607	-

Table 1: The average out-of-sample prediction performance (R^2 and MSE) and model sizes across 200 random splits for the EPS macroeconomic dataset.

5.2 Prediction of sales areas of commercial houses

In this section, we focus on the prediction of the annual average sales area of commercial houses in each province in China. Data are collected from the National Bureau of Statistics of China[2] on a monthly basis for 31 administrative regions of mainland China in the years 2022 and 2023. The dependent variable is the annual average sales area of commercial houses against 60 functional covariates, including cement, medium tractors, engines, aluminum materials, and other variables related to industrial output. To recover the functional trajectories, 24 grid points are predetermined on a monthly basis from 2022 to 2023.

To predict the response and select the most informative industrial drivers, the proposed fFASM, the standard Lasso, and the grLasso methods are employed. Furthermore, the true joint MFPCA is also evaluated as a global baseline. During the study, the collected data for 20 randomly selected provinces are used as the training set to build the model, and the remaining 11 as the test set to evaluate the prediction performance, with 200 repetitions of the whole procedure. The robust average out-of-sample R^2 and the average model sizes (evaluated on the aligned non-extreme iterations) are reported in Table 2, and the most frequently selected functional covariates are provided in Table 2 of the Supplementary Material. It is found that the proposed fFASM method shows significantly better performance than Lasso and grLasso, achieving the highest average out-of-sample while maintaining the lowest average model size. Conversely, the joint MFPCA approach yields a negative out-of-sample R^2, indicating that fusing all 60 functional covariates into a global dynamic system causes severe overfitting on this dataset, thereby highlighting the necessity of effective variable selection from the proposed method.

[2] https://data.stats.gov.cn/

In Table 2 of the Supplementary Material, all three variable-selection methods identify "Cement" as the most frequently selected functional covariate out of 200 repetitions, which is expected as it is the most critical raw material required for house construction. Similarly, the functional covariate "Aluminum Materials" is selected among the top five. Specifically, the fFASM and Lasso methods accurately select the functional covariates "Medium Tractors" and "Engine", which are closely related to the construction machinery used in the house building industry. In contrast, grLasso selects variables such as "Mechanized Paper", "Hydropower Generation", and "Phosphate Rock", which intuitively have limited direct connections with house construction. This structural over-penalization may explain why grLasso yields a larger model size (2.61) but a noticeably inferior average out-of-sample prediction performance ($R^2 = 0.420$).

Method	**Average out-of-sample R^2**	**Average model size**
fFASM (Proposed)	0.647	1.56
Lasso	0.583	2.46
grLasso	0.420	2.61
MFPCA (Chiou & Müller 2016)	-0.197	-

Table 2: The average out-of-sample prediction performance and model sizes across 200 random splits.

6 Conclusion

In this article, a novel functional factor augmentation structure (fFAS) is proposed to capture associations for correlated functional processes, and further a functional factor augmentation selection model (fFASM) is developed to select useful functional covariates in high dimensions with correlated functional covariates. Note only is the correlation between functional covariates addressed without assuming an explicit covariance structure, theoretical properties of the estimated functional factors are established. We primarily discuss the rationale for constructing a fFAS, how to estimate the fFAS and its estimation error, and the impact of truncating functional data on the validity and estimation of the factor model. Due to the unique characteristics of functional data, the assumed factor model and the actually estimated factor model may differ. Numerical investigations on both simulated and real datasets support the superior performance of the proposed fFASM method. It is found that our proposed method performs better than general functional data variable selection methods when dealing with the variable selection problem of correlated multivariate functional covariates. A practical issue may be how to determine the model size, as our method may slightly select more functional covariates in simulation studies, which may be a trade-off for modeling functional processes with correlations in high dimensions.

The proposed fFASM framework may be extended to higher-dimensional cases, where empirical studies already demonstrate its excellent selection accuracy. Another promising extension is generalized linear models (GLMs), such as functional logistic regression. While our two-stage procedure remains conceptually applicable via an appropriate link function, nonlinear parameter coupling prevents the direct orthogonal projection of unpenalized latent factors, necessitating joint estimation and computationally intensive iterative algorithms. Finally, future studies will focus on establishing theoretical justifications for selection consistency and optimal selection rates in these expanded settings.

References

Ahn, S. C. & Horenstein, A. R. (2013), 'Eigenvalue ratio test for the number of factors', *Econometrica* **81**(3), 1203–1227.

Akaike, H. (1974), 'A new look at the statistical model identification', *IEEE transactions on automatic control* **19**(6), 716–723.

Aneiros, G., Horová, I., Hušková, M. & Vieu, P. (2022), 'On functional data analysis and related topics', *Journal of Multivariate Analysis* **189**, 104861.

Aneiros, G., Novo, S. & Vieu, P. (2022), 'Variable selection in functional regression models: A review', *Journal of Multivariate Analysis* **188**, 104871.

Bai, J. (2003), 'Inferential theory for factor models of large dimensions', *Econometrica* **71**(1), 135–171.

Bai, J. & Ng, S. (2002), 'Determining the number of factors in approximate factor models', *Econometrica* **70**(1), 191–221.

Cardot, H., Ferraty, F. & Sarda, P. (2003), 'Spline estimators for the functional linear model', *Statistica Sinica* pp. 571–591.

Castellanos, L., Vu, V. Q., Perel, S., Schwartz, A. B. & Kass, R. E. (2015), 'A multivariate gaussian process factor model for hand shape during reach-to-grasp movements', *Statistica Sinica* **25**(1), 5.

Centofanti, F., Lepore, A., Menafoglio, A., Palumbo, B. & Vantini, S. (2021), 'Functional regression control chart', *Technometrics* **63**(3), 281–294.

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

Chen, D., Hall, P. & Müller, H.-G. (2011), 'Single and multiple index functional regression models with nonparametric link', *The Annals of Statistics* **39**(3), 1720–1747.

Chen, L., Wang, W. & Wu, W. B. (2021), 'Dynamic semiparametric factor model with structural breaks', *Journal of Business & Economic Statistics* **39**(3), 757–771.

Chiou, J.-M. & Müller, H.-G. (2016), 'A pairwise interaction model for multivariate functional and longitudinal data', *Biometrika* **103**(2), 377–396.

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

Fan, J., Ke, Y. & Wang, K. (2020), 'Factor-adjusted regularized model selection', *Journal of Econometrics* **216**(1), 71–85.

Fan, J. & Li, R. (2001), 'Variable selection via nonconcave penalized likelihood and its oracle properties', *Journal of the American Statistical Association* **96**(456), 1348–1360.

Fan, J., Liao, Y. & Mincheva, M. (2013), 'Large covariance estimation by thresholding principal orthogonal complements', *Journal of the Royal Statistical Society Series B: Statistical Methodology* **75**(4), 603–680.

Fang, L., Zhao, H., Wang, P., Yu, M., Yan, J., Cheng, W. & Chen, P. (2015), 'Feature selection method based on mutual information and class separability for dimension reduction in multidimensional time series for clinical data', *Biomedical Signal Processing and Control* **21**, 82–89.

Ferraty, F. (2006), *Nonparametric functional data analysis*, Springer.

Gonzalez-Vidal, A., Jimenez, F. & Gomez-Skarmeta, A. F. (2019), 'A methodology for energy multivariate time series forecasting in smart buildings based on feature selection', *Energy and Buildings* **196**, 71–82.

Hall, P. & Horowitz, J. L. (2007), 'Methodology and convergence rates for functional linear regression', *The Annals of Statistics* **35**(1), 70–91.

Hays, S., Shen, H. & Huang, J. Z. (2012), 'Functional dynamic factor models with application to yield curve forecasting', *The Annals of Applied Statistics* pp. 870–894.

Hörmann, S., Kidziński, Ł. & Hallin, M. (2015), 'Dynamic functional principal components', *Journal of the Royal Statistical Society Series B: Statistical Methodology* **77**(2), 319–348.

Hörmann, S. & Kokoszka, P. (2010), 'Weakly dependent functional data', *The Annals of Statistics* **38**(3), 1845–1884.

Htun, H. H., Biehl, M. & Petkov, N. (2023), 'Survey of feature selection and extraction techniques for stock market prediction', *Financial Innovation* **9**(1), 26.

Jiménez, F., Palma, J., Sánchez, G., Marín, D., Francisco Palacios, M. & Lucía López, M. (2020), 'Feature selection based multivariate time series forecasting: An application to antibiotic resistance outbreaks prediction', *Artificial Intelligence in Medicine* **104**, 101818.

Kong, D., Xue, K., Yao, F. & Zhang, H. H. (2016), 'Partially functional linear regression in high dimensions', *Biometrika* **103**(1), 147–159.

Lee, J. D., Sun, Y. & Taylor, J. E. (2015), 'On model selection consistency of regularized M-estimators', *Electronic Journal of Statistics* **9**(1), 608–642.

Li, Y. & Hsing, T. (2007), 'On rates of convergence in functional linear regression', *Journal of Multivariate Analysis* **98**(9), 1782–1804.

Lin, Z. & Wang, J.-L. (2022), 'Mean and covariance estimation for functional snippets', *Journal of the American Statistical Association* **117**(537), 348–360. PMID: 35757778.

Lin, Z. & Yao, F. (2020), 'Functional regression on the manifold with contamination', *Biometrika* **108**(1), 167–181.

Matsui, H. & Konishi, S. (2011), 'Variable selection for functional regression models via the l1 regularization', *Computational Statistics & Data Analysis* **55**(12), 3304–3310.

Merlevède, F., Peligrad, M. & Rio, E. (2011), 'A bernstein type inequality and moderate deviations for weakly dependent sequences', *Probability Theory and Related Fields* **151**, 435–474.

Müller, H.-G. & Yao, F. (2008), 'Functional additive models', *Journal of the American Statistical Association* **103**(484), 1534–1544.

Nti, K. O., Adekoya, A. & Weyori, B. (2019), 'Random forest based feature selection of macroeconomic variables for stock market prediction', *American Journal of Applied Sciences* **16**(7), 200–212.

Peng, R. D., Dominici, F., Pastor-Barriuso, R., Zeger, S. L. & Samet, J. M. (2005), 'Seasonal analyses of air pollution and mortality in 100 us cities', *American Journal of Epidemiology* **161**(6), 585–594.

Petersen, A. (2024), 'Mean and covariance estimation for discretely observed high-dimensional functional data: Rates of convergence and division of observational regimes', *Journal of Multivariate Analysis* **204**, 105355.

Ramsay, J. & Silverman, B. (2005), *Functional Data Analysis*, Springer Series in Statistics, Springer.

Schwarz, G. (1978), 'Estimating the dimension of a model', *The annals of statistics* **6**(2), 461–464.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society Series B: Statistical Methodology* **58**(1), 267–288.

Yang, S. & Ling, N. (2024), 'Robust estimation of functional factor models with functional pairwise spatial signs', *Computational Statistics* pp. 1–24.

Yao, F., Müller, H.-G. & Wang, J.-L. (2005), 'Functional linear regression analysis for longitudinal data', *The Annals of Statistics* **33**(6), 2873–2903.

Yuan, M. & Cai, T. T. (2010), 'A reproducing kernel Hilbert space approach to functional linear regression', *The Annals of Statistics* **38**(6), 3412–3444.

Zhang, C.-H. (2010), 'Nearly unbiased variable selection under minimax concave penalty', *Annals of Statistics* **38**(2), 894–942.

Zhenhua Lin, M. E. L. & Müller, H.-G. (2023), 'High-dimensional manova via bootstrapping and its application to functional and sparse count data', *Journal of the American Statistical Association* **118**(541), 177–191.