

## Envelope Model for Function-on-Function Linear Regression

Zhihua Su, Bing Li & Dennis Cook

**To cite this article:** Zhihua Su, Bing Li & Dennis Cook (2023) Envelope Model for Function-on-Function Linear Regression, *Journal of Computational and Graphical Statistics*, 32:4, 1624-1635, DOI: [10.1080/10618600.2022.2163652](https://doi.org/10.1080/10618600.2022.2163652)

**To link to this article:** <https://doi.org/10.1080/10618600.2022.2163652>



[View supplementary material](#)



Published online: 07 Feb 2023.



[Submit your article to this journal](#)



Article views: 339



[View related articles](#)



CrossMark

[View Crossmark data](#)



## Envelope Model for Function-on-Function Linear Regression

Zhihua Su<sup>a</sup>, Bing Li<sup>b</sup>, and Dennis Cook<sup>c</sup>

<sup>a</sup>Department of Statistics, University of Florida, Gainesville, FL; <sup>b</sup>Department of Statistics, Pennsylvania State University, University Park, PA; <sup>c</sup>School of Statistics, University of Minnesota, Minneapolis, MN

### ABSTRACT

The envelope model is a recently developed methodology for multivariate analysis that enhances estimation accuracy by exploiting the relation between the mean and eigenstructure of the covariance matrix. We extend the envelope model to function-on-function linear regression, where the response and the predictor are assumed to be random functions in Hilbert spaces. We use a double envelope structure to accommodate the eigenstructures of the covariance operators for both the predictor and the response. The central idea is to establish a one-to-one relation between the functional envelope model and the multivariate envelope model and estimate the latter using an existing method. We also developed the asymptotic theories, confidence and prediction bands, an order determination method along with its consistency, and a characterization of the efficiency gain by the proposed model. Simulation comparisons with the standard function-on-function regression and data applications show significant improvement by our method in terms of cross-validated prediction error. Supplementary materials for this article are available online.

### ARTICLE HISTORY

Received February 2022

Accepted December 2022

### KEYWORDS

Confidence and prediction bands; Covariance operator; Gaussian process in Hilbert space; Karhunen-Loeve expansion; Lattice; Order determination

## 1. Introduction

The envelope model is a recently developed multivariate analysis strategy that enhances estimation accuracy by reducing model complexity using the reducing subspaces of the covariance matrices. It was first proposed for multivariate linear regression by Cook et al. (2010), and has been extended to many settings such as partial least squares (Cook et al. 2013), generalized linear models (Cook and Zhang 2015a), tensor regression (Zhang and Li 2017; Ding and Cook 2018), quantile regression (Ding et al. 2021) and spatial models (Rekabdarkolaee et al. 2019), among others. See Cook (2018) for a comprehensive review.

In this article, we extend the envelope models from multivariate linear regression to function-on-function linear regression. This type of regression is becoming increasingly common in modern applications. See, for example, Ferraty et al. (2012), Ivanescu et al. (2015), and Luo and Qi (2017). An envelope model adapted to this setting could significantly enhance estimation accuracy by exploiting inherent model parsimony.

The development of the envelope paradigm went through several phases: the response envelope model (Cook et al. 2010), the predictor envelope model (Cook et al. 2013), and the simultaneous envelope model (Cook and Zhang 2015b). Since both the response and the predictor envelope models are special cases of the simultaneous envelope, we only need to extend the latter to function-on-function regression. To facilitate this extension, we now give an overview of the simultaneous envelope in the multivariate setting.

Let  $Y \in \mathbb{R}^r$  be the response vector and let  $X \in \mathbb{R}^p$  be the predictor vector, and suppose that they satisfy the multivariate linear regression model

$$Y = \mu + \beta X + \epsilon, \quad (1)$$

where  $\mu \in \mathbb{R}^r$  is a constant vector,  $\beta \in \mathbb{R}^{r \times p}$  is a matrix that contains the regression coefficients,  $\epsilon$  is the error vector independent of  $X$  having mean 0 and covariance matrix  $\Sigma_\epsilon$ , and the predictor vector  $X$  has mean 0 and covariance matrix  $\Sigma_X$ . Let us say that a matrix  $A$  is a basis matrix of a subspace  $\mathcal{S}$  if the columns of  $A$  form a basis of  $\mathcal{S}$ . For a finite-dimensional subspace  $\mathcal{S}$ , let  $P_{\mathcal{S}}$  and  $Q_{\mathcal{S}} = I - P_{\mathcal{S}}$  denote the projection matrices onto  $\mathcal{S}$  and its orthogonal complement  $\mathcal{S}^\perp$ , respectively, in the usual inner product. For the response envelope, we consider subspaces  $\mathcal{S}$  of  $\mathbb{R}^r$  that satisfy

$$(a) \quad Q_{\mathcal{S}} Y \perp\!\!\!\perp X, \quad (b) \quad P_{\mathcal{S}} Y \perp\!\!\!\perp Q_{\mathcal{S}} Y | X. \quad (2)$$

Conditions (a) and (b) imply that  $Q_{\mathcal{S}} Y$  carries no information about  $\beta$  directly or indirectly through its relation with  $P_{\mathcal{S}} Y$ . They hold if and only if  $Q_{\mathcal{S}} Y \perp\!\!\!\perp (P_{\mathcal{S}} Y, X)$  and thus  $P_{\mathcal{S}} Y$  carries all the regression information in (1). They hold trivially when  $\mathcal{S} = \mathbb{R}^r$  and, when  $\mathcal{S} = \{0\}$ , they imply that  $Y \perp\!\!\!\perp X$ . The response envelope is defined as the intersection of all subspaces  $\mathcal{S}$  that satisfy (a) and (b).

For the predictor envelope, we consider subspaces  $\mathcal{T}$  of  $\mathbb{R}^p$  that satisfy

$$(c) \quad Y \perp\!\!\!\perp Q_{\mathcal{T}} X | P_{\mathcal{T}} X, \quad (d) \quad P_{\mathcal{T}} X \perp\!\!\!\perp Q_{\mathcal{T}} X. \quad (3)$$

Conditions (c) and (d) imply that  $Q_{\mathcal{T}}X$  does not affect  $Y$  directly or indirectly through its relation with  $P_{\mathcal{T}}X$ . They hold if and only if  $Q_{\mathcal{T}}X \perp\!\!\!\perp (Y, P_{\mathcal{T}}X)$  and so  $P_{\mathcal{T}}X$  carries all the regression information about  $Y$ . They hold trivially when  $\mathcal{T} = \mathbb{R}^p$  and, when  $\mathcal{T} = \{0\}$ , they again imply that  $Y \perp\!\!\!\perp X$ . The predictor envelope is defined as the intersection of all subspaces  $\mathcal{T}$  that satisfy (c) and (d).

The simultaneous envelope model is the multivariate regression model (1) that requires conditions (a)–(d), and is parameterized in terms of the response and predictor envelopes. We refer to this model as the Multivariate Envelope Linear Model (MELM). The response envelope model is the special case of MELM with  $\mathcal{T} = \mathbb{R}^p$ ; the predictor envelope is the special case of MELM with  $\mathcal{S} = \mathbb{R}^r$ ; finally, the classical multivariate linear regression is the special case of MELM with both  $\mathcal{S} = \mathbb{R}^r$  and  $\mathcal{T} = \mathbb{R}^p$ .

The theoretical structure of our extension is sketched as follows. Let  $X$  and  $Y$  be random functions in Hilbert spaces  $\mathcal{H}_X$  and  $\mathcal{H}_Y$ . Assume that they satisfy the linear model

$$Y = \alpha + BX + \epsilon, \quad (4)$$

where  $\alpha$  is a fixed member in  $\mathcal{H}_Y$ ,  $\epsilon$  is a random member of  $\mathcal{H}_Y$ , and  $B: \mathcal{H}_X \rightarrow \mathcal{H}_Y$  is a linear operator. We illustrate this function-on-function regression model with an example. To assess the economic effect of Covid-19, the daily new confirmed cases and daily mobility of retail and recreation are measured for all 21 counties in June 2020 in New Jersey. The curve of new confirmed cases over the entire month serves as the predictor  $X$ . The curve of mobility, which corresponds to visits to places like restaurants, shopping centers, and movie theaters, is the response  $Y$ . For many counties, the new confirmed cases are recorded, while the mobility data is missing. Using the function-on-function regression model (4), it is possible to predict the mobility from the new confirmed cases for those counties. The goal of envelope model is to achieve a more efficient estimation of the model parameters, which leads to more accurate prediction.

Now we provide a sketch of the construction of the envelope model under (4). Let  $\Sigma_\epsilon$  and  $\Sigma_X$  be the covariance operators of  $\epsilon$  and  $X$ , respectively. If  $A: \mathcal{H} \rightarrow \mathcal{H}$  is a bounded linear operator, then a reducing subspace of  $A$  is any subspace  $\mathcal{S}$  of  $\mathcal{H}$  such that  $A\mathcal{S} \subseteq \mathcal{S}$  and  $A\mathcal{S}^\perp \subseteq \mathcal{S}^\perp$ ; if  $A$  is self-adjoint, then these two conditions are equivalent. Let  $\mathcal{S}$  be the intersection of all reducing subspaces of  $\Sigma_\epsilon$  that contains the range of  $B$ , and let  $\mathcal{T}$  be the intersection of all reducing subspaces of  $\Sigma_X$  that contains the range of  $B^*$ , the adjoint of  $B$ . The subspaces  $\mathcal{S}$  and  $\mathcal{T}$  are the functional extensions of the response and predictor envelopes defined for (1). The rationale based on the covariance operators may seem different from that used for model (1), but it will be shown in [Theorem 1](#) that these operator-based definitions lead to extensions of (2) and (3) to the functional model (4). When  $\mathcal{S}$  and  $\mathcal{T}$  are proper subspaces of  $\mathcal{H}_Y$  and  $\mathcal{H}_X$ , they offer a reduction of the complexity of the function-on-function linear regression, which, as we will show, can be very substantial in applications. Our goal is to estimate  $\mathcal{S}$ ,  $\mathcal{T}$ ,  $\alpha$  and  $B$  in the dimension reduced function-on-function linear regression. We refer to model (4) with the simplifying structure  $(\mathcal{S}, \mathcal{T})$  as the Functional Envelope Linear Model (FELM).

In the theoretical development of our approach, we rely on the assumption that the regression error is a Gaussian random element in a Hilbert space. When the Gaussian assumption is violated, the objective function used to construct our estimator is still valid if we replace the conditional independence relations by some weaker conditions such as conditional uncorrelation.

The function-on-function envelope model developed here is but one way to achieve model parsimony for functional regression. Other approaches to model parsimony have been considered previously such as the penalized function-on-function linear regression ([Ivanescu et al. 2015](#); [Scheipl and Greven 2016](#); [Sun et al. 2018](#)). Our approach uses a different philosophy to achieve this: rather than penalizing the roughness of the coefficient functions, we impose sparse structures on the spectra of the covariance operators of the predictor and response processes, and let the data tell us which parts of the spectra are important. To the best of our knowledge, this article is the first attempt to generalize the envelope model to function-on-function regression. The closest earlier work is [Zhang et al. \(2018\)](#), which extended the envelope model to sufficient dimension reduction, where the response is a scalar and the predictor is a function. This article aims at estimating the functional dimension reduction space whereas we provide the explicit regression estimator, furnished with confidence and prediction bands, as well as an order determination method.

## 2. Functional Envelope Linear Model

Let  $(\Omega, \mathcal{F}, P)$  be a probability space, let  $N \subseteq \mathbb{R}$  denote an interval and let  $\mathcal{H}_X$  and  $\mathcal{H}_Y$  denote separable Hilbert spaces of real-valued functions on  $N$ . Let  $X: \Omega \rightarrow \mathcal{H}_X$  be a random element in  $\mathcal{H}_X$  and  $Y: \Omega \rightarrow \mathcal{H}_Y$  a random element in  $\mathcal{H}_Y$ .

Recall that, if  $\mathcal{H}$  is a generic Hilbert space, and  $X$  a random element in  $\mathcal{H}$  with  $E(\|X\|_{\mathcal{H}}) < \infty$ , then the function  $T: \mathcal{H} \rightarrow \mathbb{R}$ ,  $g \mapsto E(\langle g, X \rangle_{\mathcal{H}})$  is a bounded linear functional, which has a Riesz representation  $g_0$  s.t.  $T(g) = \langle g_0, g \rangle_{\mathcal{H}}$  for each  $g \in \mathcal{H}$ . The function  $g_0$  is defined as the expectation of  $X$ , and is written as  $E(X)$ .

To define the second moment of  $X$ , recall that, if  $\mathcal{H}_1, \mathcal{H}_2$  are Hilbert spaces and  $f_1, f_2$  are members of  $\mathcal{H}_1, \mathcal{H}_2$ , respectively, then the tensor product  $f_1 \otimes f_2$  is a rank one linear operator from  $\mathcal{H}_2$  to  $\mathcal{H}_1$  such that, for each  $g \in \mathcal{H}_2$ ,  $(f_1 \otimes f_2)(g) = f_1 \langle f_2, g \rangle_{\mathcal{H}_2}$ . Let  $\mathcal{B}(\mathcal{H}_2, \mathcal{H}_1)$  denote the class of all bounded linear operators from  $\mathcal{H}_2$  to  $\mathcal{H}_1$ . Then  $\mathcal{B}(\mathcal{H}_2, \mathcal{H}_1)$  is a Banach space with respect to the operator norm. If  $A$  is a random operator in  $\mathcal{B}(\mathcal{H}_2, \mathcal{H}_1)$ , then the function  $b: \mathcal{H}_1 \times \mathcal{H}_2 \rightarrow \mathbb{R}$ ,  $(f, g) \mapsto E(\langle f, Ag \rangle_{\mathcal{H}_1})$  is a bounded bilinear form, which induces a linear operator  $B \in \mathcal{B}(\mathcal{H}_2, \mathcal{H}_1)$  such that  $b(f, g) = \langle f, Bg \rangle_{\mathcal{H}_1}$ . See, for example, ([Conway 2013](#), chap. 2). The operator  $B$  is then defined as the expectation of  $A$ , written as  $E(A)$ . Using this definition of expectation of a random operator, we define the second moment operator of  $X$  as  $E(X \otimes X)$ , and the variance operator of  $X$  as  $E[(X - E(X)) \otimes (X - E(X))]$ , which is denoted by  $\text{var}(X)$ .

Our definitions of  $E(X)$  and  $E(A)$  are different from the standard definitions of moments in the functional data analysis, which are done through the pointwise moments such as  $E[X(t)]$  and  $E[X(s)X(t)]$ . See, for example, [Ramsay and Silverman \(2007\)](#). The two definitions of  $E(X)$ , as the Riesz representation of the bounded linear functional  $f \mapsto E(X, f)_{\mathcal{H}}$  and as

the pointwise expectation function  $t \mapsto EX(t)$ , are equivalent under mild conditions, for example, they are equivalent when  $\mathcal{H}$  separable. However, the former definition does not rely on the nature of  $\mathcal{H}$ ; that is, it does not require  $\mathcal{H}$  to be a set of functions of  $t$ , whereas the latter definition does.

For a subspace  $\mathcal{U}$  of  $\mathcal{H}$ , and a self-adjoint operator  $A : \mathcal{H} \rightarrow \mathcal{H}$ , the  $A$ -envelope of  $\mathcal{U}$ , denoted by  $\mathcal{E}(\mathcal{U}; A)$ , is defined as  $\mathcal{E}(\mathcal{U}; A) = \bigcap \{\mathcal{S} : \mathcal{S} \in \text{Lat}_{\mathcal{U}}(A)\}$ , where  $\text{Lat}_{\mathcal{S}}(A)$  represents the collection of all reducing subspaces of  $A$  that contain  $\mathcal{S}$ . The symbol  $\text{Lat}$  comes from the word “lattice.” For another-bounded linear operator  $B : \mathcal{H} \rightarrow \mathcal{H}$ , let  $\text{ran}(B)$  represent the range (or image) of  $B$ , and  $\overline{\text{ran}}(B)$  the closure of  $\text{ran}(B)$ . We abbreviate  $\mathcal{E}(\overline{\text{ran}}(B); A)$  by  $\mathcal{E}(B; A)$ . In this article, we will only deal with  $\mathcal{E}(B; A)$  where  $\text{ran}(B) = \overline{\text{ran}}(B)$ .

Let  $B : \mathcal{H}_X \rightarrow \mathcal{H}_Y$  be a bounded linear operator. Let  $\epsilon$  be a random element in  $\mathcal{H}_Y$  such that  $\epsilon \perp\!\!\!\perp X$  and  $E(\epsilon) = 0$ . For simplicity, we tentatively assume that  $E(X) = 0$ . We consider the function-on-function linear regression

$$Y = \alpha + BX + \epsilon, \quad (5)$$

where  $\epsilon$  and  $X$  are Gaussian random elements in  $\mathcal{H}_Y$  and  $\mathcal{H}_X$ , respectively, satisfying

$$E(\epsilon) = 0, \quad E(X) = 0, \quad E\|\epsilon\|_{\mathcal{H}_Y}^2 < \infty, \quad E\|X\|_{\mathcal{H}_X}^2 < \infty, \quad (6)$$

and  $\alpha$  is a nonrandom member of  $\mathcal{H}_Y$ . Let  $\Sigma_\epsilon = E(\epsilon \otimes \epsilon)$ ,  $\Sigma_X = E(X \otimes X)$ , and  $\Sigma_Y = E(Y \otimes Y)$ . It is easy to see these are self-adjoint operators. For an operator  $A$ , let  $A^*$  represent the adjoint operator of  $A$ .

Next, we give a rigorous definition of FELM. Let  $\{(\lambda_i, \phi_i), i \in \mathbb{N}\}$ ,  $\{(\rho_i, \chi_i), i \in \mathbb{N}\}$  and  $\{(\tau_i, \psi_i), i \in \mathbb{N}\}$  be the sequences of eigenvalues and eigenfunctions for the linear operator  $\Sigma_Y$ ,  $\Sigma_\epsilon$  and  $\Sigma_X$ . We say that a subspace  $\mathcal{S}$  of  $\mathcal{H}_Y$  is covered by a subset  $C$  of  $\{\chi_i : i \in \mathbb{N}\}$  if  $\mathcal{S}$  is contained in the subspace spanned by  $C$ . The same applies to a subspace of  $\mathcal{H}_X$ .

**Definition 1.** A functional linear model is called a functional envelope linear model (FELM) with respect to the response envelope  $\mathcal{E}(B; \Sigma_\epsilon)$  and predictor envelope  $\mathcal{E}(B^*; \Sigma_X)$ . Furthermore:

- (i) If at least one of the envelopes is a proper subset of its ambient space, then we call the FELM a proper envelope model.
- (ii) If  $\text{ran}(B)$  is covered by a finite subset of  $\{\chi_1, \chi_2, \dots\}$  and  $\text{ran}(B^*)$  by a finite subset of  $\{\psi_1, \psi_2, \dots\}$ , then we call the FELM an eigen-sparse envelope model.

According to part (i) of this definition, any functional linear model is a FELM, but only when one or both envelopes are proper subspaces, does the FELM lead to efficiency gains. When the predictor envelope is the ambient space, the FELM reduces to the functional response envelope model; when the response envelope is the ambient space, the FELM reduces to the functional predictor envelope model. Since, in the multivariate setting, the predictor envelope model is the underlying model of the partial least squares, the FELM generalizes the partial least squares (Wold 1966; De Jong 1993; Cook et al. 2013) to the functional setting. See also Delaigle and Hall (2012).

It is also reasonable to call  $\mathcal{E}(B; \Sigma_\epsilon)$  the residual envelope, but, as we will show in [Theorem 2](#), this envelope is the same as  $\mathcal{E}(B; \Sigma_Y)$ . Hence, it is justified to call it the response envelope.

In this article, we focus on the eigen-sparse envelope model in part (ii) of [Definition 1](#), where the numbers of eigenfunctions  $\chi$ 's and  $\psi$ 's that cover the response and predictor envelopes are fixed. In principle, we can allow the numbers of the covering eigenfunctions to grow with sample size  $n$ . The situation is similar to sparse estimation, except that, here, the sparsity is imposed on the eigenstructures. However, due to limited space, the development for FELM where the numbers of the covering eigenfunctions increase with the sample size is left to future research.

We next express the eigen-sparse FELM explicitly in terms of the eigenfunctions of the covariance operators. Let  $I$  and  $J$  be the finite subsets of  $\{1, 2, \dots\}$  such that  $\{\chi_i : i \in I\}$  covers  $\text{ran}(B)$  and  $\{\psi_j : j \in J\}$  covers  $\text{ran}(B^*)$ . Then we have

$$B = \sum_{i \in I} \sum_{j \in J} b_{ij} (\chi_i \otimes \psi_j), \quad (7)$$

where  $b_{ij}$  are real numbers. Corresponding to  $I$  and  $J$ ,  $\Sigma_X$  and  $\Sigma_\epsilon$  can be decomposed as

$$\begin{aligned} \Sigma_\epsilon &= \sum_{i \in I} \rho_i (\chi_i \otimes \chi_i) + \sum_{i \notin I} \rho_i (\chi_i \otimes \chi_i), \\ \Sigma_X &= \sum_{j \in J} \tau_j (\psi_j \otimes \psi_j) + \sum_{j \notin J} \tau_j (\psi_j \otimes \psi_j). \end{aligned} \quad (8)$$

It is easy to see  $\mathcal{E}(B; \Sigma_\epsilon) = \text{span}\{\chi_i : i \in I\}$ , and  $\mathcal{E}(B^*; \Sigma_X) = \text{span}\{\psi_j : j \in J\}$ .

Let  $s$  and  $t$  denote the cardinalities of  $I$  the cardinality of  $J$ , respectively. Note that  $\mathcal{E}(B; \Sigma_\epsilon)$  is isomorphic to  $\mathbb{R}^s$  via the isomorphism  $f \in \mathcal{E}(B; \Sigma_\epsilon) \leftrightarrow \{\langle f, \chi_i \rangle_{\mathcal{H}_Y} : i \in I\} \in \mathbb{R}^s$ . Similarly,  $\mathcal{E}(B^*; \Sigma_X)$  is isomorphic to  $\mathbb{R}^t$  via the isomorphism  $f \in \mathcal{E}(B^*; \Sigma_X) \leftrightarrow \{\langle f, \psi_j \rangle_{\mathcal{H}_X} : j \in J\} \in \mathbb{R}^t$ . The dimensions of  $\mathcal{E}(B; \Sigma_\epsilon)$  and  $\mathcal{E}(B^*; \Sigma_X)$  are  $s$  and  $t$ .

Our [Definition 1](#) of FELM was made solely in geometric terms, but, as the next theorem will show, under the Gaussian assumption, they induce independence and conditional independence relations parallel to (a), (b), (c), and (d) in [\(2\)](#) and [\(3\)](#) for MELM. These relations are the true motivation behind our definition of FELM. Let  $P_{\mathcal{E}(B; \Sigma_\epsilon)}$ ,  $Q_{\mathcal{E}(B; \Sigma_\epsilon)}$ ,  $P_{\mathcal{E}(B^*; \Sigma_X)}$ , and  $Q_{\mathcal{E}(B^*; \Sigma_X)}$  be the projections onto  $\mathcal{E}(B; \Sigma_\epsilon)$ ,  $\mathcal{E}(B; \Sigma_\epsilon)^\perp$ ,  $\mathcal{E}(B^*; \Sigma_X)$  and  $\mathcal{E}(B^*; \Sigma_X)^\perp$ , then

$$\begin{aligned} P_{\mathcal{E}(B; \Sigma_\epsilon)} &= \sum_{i \in I} \chi_i \otimes \chi_i, & Q_{\mathcal{E}(B; \Sigma_\epsilon)} &= \sum_{i \notin I} \chi_i \otimes \chi_i, \\ P_{\mathcal{E}(B^*; \Sigma_X)} &= \sum_{j \in J} \psi_j \otimes \psi_j, & Q_{\mathcal{E}(B^*; \Sigma_X)} &= \sum_{j \notin J} \psi_j \otimes \psi_j. \end{aligned} \quad (9)$$

The Karhunen-Lo  e expansions of  $\epsilon$  and  $X$  are

$$\begin{aligned} \epsilon &= \sum_{i=1}^{\infty} \langle \epsilon, \chi_i \rangle_{\mathcal{H}_Y} \chi_i = \sum_{i=1}^{\infty} \rho_i^{1/2} v_i \chi_i, \\ X &= \sum_{i=1}^{\infty} \langle X, \psi_i \rangle_{\mathcal{H}_X} \psi_i = \sum_{i=1}^{\infty} \tau_i^{1/2} \xi_i \psi_i, \end{aligned} \quad (10)$$

where by the Gaussian assumption,  $v_1, v_2, \dots$  are iid  $N(0, 1)$ , and  $\xi_1, \xi_2, \dots$  are iid  $N(0, 1)$ .

**Theorem 1.** Under the assumptions in [Definition 1](#), we have

- (a)  $Q_{\mathcal{E}(B; \Sigma_\epsilon)} Y \perp\!\!\!\perp X$ ,
- (b)  $P_{\mathcal{E}(B; \Sigma_\epsilon)} Y \perp\!\!\!\perp Q_{\mathcal{E}(B; \Sigma_\epsilon)} Y | X$ ,
- (c)  $Y \perp\!\!\!\perp Q_{\mathcal{E}(B^*; \Sigma_X)} X | P_{\mathcal{E}(B^*; \Sigma_X)} X$ ,
- (d)  $P_{\mathcal{E}(B^*; \Sigma_X)} X \perp\!\!\!\perp Q_{\mathcal{E}(B^*; \Sigma_X)} X$ .

Relationships (11a) and (11b) are the functional counterparts of (2a) and (2b), and (11c) and (11d) are the functional counterparts of (3c) and (3d). Following the finite-dimensional analog, (11a) and (11b) hold if and only if  $Q_{\mathcal{E}(B; \Sigma_\epsilon)} Y \perp\!\!\!\perp (X, P_{\mathcal{E}(B; \Sigma_\epsilon)} Y)$ , and (11c) and (11d) hold if and only if  $(Y, P_{\mathcal{E}(B^*; \Sigma_X)} X) \perp\!\!\!\perp Q_{\mathcal{E}(B^*; \Sigma_X)} X$ . In consequence, if we knew the envelopes  $\mathcal{E}(B; \Sigma_\epsilon)$  and  $\mathcal{E}(B^*; \Sigma_X)$ , then, to estimate  $B$ , all we would need to do is to regress  $P_{\mathcal{E}(B; \Sigma_\epsilon)} Y$  on  $P_{\mathcal{E}(B^*; \Sigma_X)} X$ , while leaving  $Q_{\mathcal{E}(B; \Sigma_\epsilon)} Y$  and  $Q_{\mathcal{E}(B^*; \Sigma_X)} X$  solely for the estimation of the corresponding parts of the covariance operator  $\Sigma_\epsilon$  and  $\Sigma_X$ . This is the underlying mechanism that drives the efficiency gains of the envelope model. Of course, at the sample level, the envelopes themselves have to be estimated as well, and  $Q_{\mathcal{E}(B; \Sigma_\epsilon)} Y$  and  $Q_{\mathcal{E}(B^*; \Sigma_X)} X$  do participate in the estimation of them, thus, being indirectly involved in the estimation of  $B$ .

The next theorem shows that  $\mathcal{E}(B; \Sigma_\epsilon)$  and  $\mathcal{E}(B; \Sigma_Y)$  are in fact the same, which generalizes a result to the functional setting (see, [Cook et al. 2010](#), Proposition 3.1).

**Theorem 2.** Under the functional linear model (5),  $\mathcal{E}(B; \Sigma_\epsilon) = \mathcal{E}(B; \Sigma_Y)$ .

Next we derive the Karhunen-Loëve expansion of  $Y$ . In preparation, let  $i_1, \dots, i_s$  be the members of  $I$ . For  $k \in I$ , let  $U_k = \sum_{j \in I} b_{kj} \tau_j^{1/2} \xi_j + \rho_k^{1/2} v_k$ , and  $U = (U_{i_1}, \dots, U_{i_s})^\top$ . Let  $\Sigma_U = \text{var}(U)$ , and  $QDQ^\top$  the spectral decomposition of  $\Sigma_U$ , where  $Q \in \mathbb{R}^{s \times s}$  is an orthogonal matrix and  $D \in \mathbb{R}^{s \times s}$  is a diagonal matrix. For a sequence  $a_1, a_2, \dots$ , and a subset  $A \subseteq \{1, 2, \dots\}$ , we denote  $\{a_i : i \in A\}$  by  $a(A)$ . Using this notation, we have

$$\mathcal{E}(B; \Sigma_\epsilon) = \text{span}\{\chi(I)\} \quad \text{and} \quad \mathcal{E}(B^*; \Sigma_X) = \text{span}\{\psi(J)\}.$$

In the following, if  $A$  is a matrix and  $a$  is a vector, we use  $\text{diagmv}(A)$  to represent the diagonal vector of  $A$ , and  $\text{diagvm}(a)$  the diagonal matrix with  $a$  as its diagonal.

**Theorem 3.** Under the assumptions in [Definition 1](#), the Karhunen-Loëve expansion of  $Y$  has the form  $Y = \alpha + \sum_{i=1}^{\infty} \lambda_i^{1/2} \zeta_i \phi_i$ , where  $\{\phi_i : i = 1, 2, \dots\}$ ,  $\{\zeta_i : i = 1, 2, \dots\}$ , and  $\{\chi_i : i = 1, 2, \dots\}$  are sequences such that

$$\begin{aligned} \phi(I) &= Q^\top \chi(I), & \zeta(I) &= D^{-1/2} Q^\top U, & \lambda(I) &= \text{diagmv}(D), \\ \phi(I^c) &= \chi(I^c), & \zeta(I^c) &= v(I^c), & \lambda(I^c) &= \rho(I^c). \end{aligned}$$

This implies that there is an isomorphism between the eigenfunctions  $\{\phi_1, \phi_2, \dots\}$  of  $\Sigma_Y$  and the eigenfunctions  $\{\chi_1, \chi_2, \dots\}$  of  $\Sigma_\epsilon$ . In particular, for the indices in  $I$ ,

$$(\phi_{i_1}, \dots, \phi_{i_s})^\top = Q^\top (\chi_{i_1}, \dots, \chi_{i_s})^\top. \quad (12)$$

Using this relation we can reexpress  $B$  in terms of  $\{\phi_i\}$  and  $\{\psi_i\}$  as follows:

$$B = \sum_{i \in I} \sum_{j \in J} b_{ij} (\chi_i \otimes \psi_j) = \sum_{j \in J} \left[ \left( \sum_{i \in I} b_{ij} \chi_i \right) \otimes \psi_j \right].$$

By (12),  $\sum_{i \in I} b_{ij} \chi_i = (b_{i_1 j}, \dots, b_{i_s j}) Q Q^\top (\chi_{i_1}, \dots, \chi_{i_s})^\top = \sum_{i \in I} \tilde{b}_{ij} \phi_i$ , where  $(\tilde{b}_{i_1 j}, \dots, \tilde{b}_{i_s j}) = (b_{i_1 j}, \dots, b_{i_s j}) Q$ . Hence

$$B = \sum_{i \in I} \sum_{j \in J} \tilde{b}_{ij} (\phi_i \otimes \psi_j). \quad (13)$$

The exposition so far in this section is coordinate-free: the entire system is built on invariant objects such as subspaces, lattices, and linear operators. To make the presentation less abstract, we now give an alternative, and somewhat intuitive, construction of the FELM based on orthonormal bases. Let  $\{e_i^{(X)} : i \in \mathbb{N}\}$  and  $\{e_i^{(Y)} : i \in \mathbb{N}\}$  be orthonormal bases of  $\mathcal{H}_X$  and  $\mathcal{H}_Y$ , respectively. Then  $X$ ,  $Y$ , and  $\epsilon$  can be expanded as

$$X = \sum_{i \in \mathbb{N}} \langle X, e_i^{(X)} \rangle_{\mathcal{H}_X} e_i^{(X)}, \quad Y = \sum_{i \in \mathbb{N}} \langle Y, e_i^{(Y)} \rangle_{\mathcal{H}_Y} e_i^{(Y)}, \\ \epsilon = \sum_{i \in \mathbb{N}} \langle \epsilon, e_i^{(Y)} \rangle_{\mathcal{H}_Y} e_i^{(Y)}.$$

Let  $X^\circ$  and  $Y^\circ$  and  $\epsilon^\circ$  represent the sequences

$$\begin{aligned} \{\langle X, e_i^{(X)} \rangle_{\mathcal{H}_X} : i \in \mathbb{N}\}, \quad \{\langle Y, e_i^{(Y)} \rangle_{\mathcal{H}_Y} : i \in \mathbb{N}\}, \\ \{\langle \epsilon, e_i^{(Y)} \rangle_{\mathcal{H}_Y} : i \in \mathbb{N}\}. \end{aligned}$$

These establish one-to-one correspondences between  $\mathcal{H}_X$  and  $\ell_2$ , and between  $\mathcal{H}_Y$  and  $\ell_2$ , where  $\ell_2$  is the collection of all square summable sequences. In this framework the linear operators  $\Sigma_X$  and  $\Sigma_\epsilon$  correspond to the  $\infty \times \infty$  matrices

$$\Sigma_{X^\circ} = \{\text{cov}(X_i^\circ, X_j^\circ) : i, j \in \mathbb{N}\}, \quad \Sigma_{\epsilon^\circ} = \{\text{cov}(\epsilon_i^\circ, \epsilon_j^\circ) : i, j \in \mathbb{N}\}.$$

The eigenvectors of these  $\infty \times \infty$  matrices are sequences in  $\ell_2$ . Let  $\{\psi_i^\circ : i \in \mathbb{N}\}$  and  $\{\chi_i^\circ : i \in \mathbb{N}\}$  be the eigenvectors of  $\Sigma_{X^\circ}$  and  $\Sigma_{\epsilon^\circ}$ , respectively, and let

$$X^* = \{\langle X^\circ, \psi_i^\circ \rangle_{\ell_2} : i \in \mathbb{N}\}, \quad Y^* = \{\langle Y^\circ, \chi_i^\circ \rangle_{\ell_2} : i \in \mathbb{N}\}.$$

Note that  $X^*$  and  $X$  are different representations of the same random elements; the same can be said of  $Y$  and  $Y^*$ . The envelope model can be transparently explained in terms of  $X^*$  and  $Y^*$ : it simply assumes that only finite number of components of  $X^*$  and  $Y^*$  participate in the regression between  $X$  and  $Y$ ; that is, there exist finite subsets  $I$  and  $J$  of  $\mathbb{N}$  such that  $\{Y_i^* : i \in I\}$  and  $\{X_j^* : j \in J\}$  follow a multivariate regression model. Furthermore, the theoretical properties of the envelope structure guarantee that there is no additional regression relation left in the rest of the components. Of course, the regression in terms of  $Y^*$  versus  $X^*$  needs to be translated back to the original spaces for  $Y$  and  $X$  through the isomorphisms  $\mathcal{H}_X$  and  $\ell_2$  and  $\mathcal{H}_Y$  and  $\ell_2$ . The numerical procedures in [Sections 5](#) and [6](#) are simply the implementation of the above construction at the sample level.

### 3. From FELM to MELM

Our basic idea for estimating FELM consists of three steps: first, find its one-to-one relation with MELM, second, use the available methods to estimate the MELM, and third, transform the results back to FELM via the one-to-one relation. In this and the next sections we develop such a one-to-one relation.

Before proceeding further, we need to review the coordinate form of the MELM. Recall from the Introduction that the multivariate linear regression model (1) is said to follow MELM if Assumptions (a), (b), (c), and (d) hold with  $\mathcal{S} = \mathcal{E}(\beta; \Sigma_\epsilon)$

and  $\mathcal{T} = \mathcal{E}(\beta^\top; \Sigma_X)$ . Also recall that  $\Gamma$  is an  $r \times u$  matrix representing an orthonormal basis matrix of  $\mathcal{E}(\beta; \Sigma_\epsilon)$  and  $\Phi$  is a  $p \times d$  matrix representing an orthonormal basis matrix of  $\mathcal{E}(\beta^\top; \Sigma_X)$ . Similarly,  $\Gamma_0 \in \mathbb{R}^{r \times (r-u)}$  and  $\Phi_0 \in \mathbb{R}^{p \times (p-d)}$  are matrices representing orthonormal basis matrices of  $\mathcal{E}(\beta; \Sigma_\epsilon)^\perp$  and  $\mathcal{E}(\beta^\top; \Sigma_X)^\perp$ , respectively. The coordinate form of MELM is

$$Y = \mu + \Gamma \eta \Phi^\top X + \epsilon, \quad \Sigma_\epsilon = \Gamma \Omega \Gamma^\top + \Gamma_0 \Omega_0 \Gamma_0^\top, \\ \Sigma_X = \Phi \Delta \Phi^\top + \Phi_0 \Delta_0 \Phi_0^\top, \quad (14)$$

where  $\beta = \Gamma \eta \Phi^\top$ ,  $\Omega \in \mathbb{R}^{u \times u}$  and  $\Omega_0 \in \mathbb{R}^{(r-u) \times (r-u)}$  carry the coordinates of  $\Sigma_\epsilon$  with respect to  $\Gamma$  and  $\Gamma_0$ , and  $\Delta \in \mathbb{R}^{d \times d}$  and  $\Delta_0 \in \mathbb{R}^{(p-d) \times (p-d)}$  carry the coordinates of  $\Sigma_X$  with respect to  $\Phi$  and  $\Phi_0$ . The matrix  $\eta \in \mathbb{R}^{u \times d}$  is the regression coefficient matrix for the reduced multivariate linear model where  $Y$  is replaced by  $\Gamma^\top Y$  and  $X$  by  $\Phi^\top X$ .

Let  $\{b_1, b_2, \dots\}$  and  $\{c_1, c_2, \dots\}$  be orthonormal bases for  $\mathcal{H}_X$  and  $\mathcal{H}_Y$ , respectively. We need the following assumption.

**Assumption 1.** There are integers  $k > 0$  and  $l > 0$  such that

$$\mathcal{E}(B^*; \Sigma_X) \subseteq \text{span}\{b_1, \dots, b_k\}, \quad \mathcal{E}(B; \Sigma_\epsilon) \subseteq \text{span}\{c_1, \dots, c_l\}.$$

By construction,  $t \leq k$  and  $s \leq l$ . In the following, let  $b$  and  $c$  be the vectors of functions  $(b_1, \dots, b_k)^\top$  and  $(c_1, \dots, c_l)^\top$ .

We need to introduce some additional notations. For a Hilbert space  $\mathcal{H}$ ,  $u = (u_1, \dots, u_m) \in \mathcal{H}^m$ , and  $v = (v_1, \dots, v_n) \in \mathcal{H}^n$ , define the  $m \times n$  inner product matrix between  $u$  and  $v$  as  $\langle u, v^\top \rangle_{\mathcal{H}} = (\langle u_i, v_j \rangle_{\mathcal{H}})_{i=1, \dots, m, j=1, \dots, n}$ . Let  $\Gamma \in \mathbb{R}^{l \times s}$  and  $\Phi \in \mathbb{R}^{k \times t}$  be matrices  $\Gamma = \langle c, \phi(I)^\top \rangle_{\mathcal{H}_Y}$  and  $\Phi = \langle b, \psi(J)^\top \rangle_{\mathcal{H}_X}$ . As implied by the notation, these will be the envelope basis matrices, analogous to those in (14), that arise from mapping FELM to MELM. Because  $\{c_1, \dots, c_l\}$  and  $\{b_1, \dots, b_k\}$  are orthonormal sets, we have  $\phi(I) = \Gamma^\top c$ , and  $\psi(J) = \Phi^\top b$ , which imply

$$\langle \phi(I), \epsilon \rangle_{\mathcal{H}_Y} = \Gamma^\top \langle c, \epsilon \rangle_{\mathcal{H}_Y} \quad \text{and} \quad \langle \psi(J), X \rangle_{\mathcal{H}_X} = \Phi^\top \langle b, X \rangle_{\mathcal{H}_X}. \quad (15)$$

If  $\mathcal{B}_1 = \{a_1, \dots, a_m\}$  and  $\mathcal{B}_2 = \{b_1, \dots, b_n\}$  are finite sets of vectors, and  $A : \text{span}(\mathcal{B}_1) \rightarrow \text{span}(\mathcal{B}_2)$  is a linear mapping, then  $\mathcal{B}_2[A]\mathcal{B}_1$  represents the coordinate of  $A$ ; that is, for any  $x = c_1 a_1 + \dots + c_m a_m \in \text{span}(\mathcal{B}_1)$ ,  $y = c'_1 b_1 + \dots + c'_n b_n \in \text{span}(\mathcal{B}_2)$ , we have  $c' = (\mathcal{B}_2[A]\mathcal{B}_1)c$ , where  $c = (c_1, \dots, c_m)^\top$ ,  $c' = (c'_1, \dots, c'_n)^\top$ . In this notation,  $\phi(I)[B]\psi(J)$  is an  $s \times t$  matrix

$$\phi(I)[B]\psi(J) = \begin{pmatrix} \tilde{b}_{i_1 j_1} & \dots & \tilde{b}_{i_1 j_t} \\ \vdots & \ddots & \vdots \\ \tilde{b}_{i_s j_1} & \dots & \tilde{b}_{i_s j_t} \end{pmatrix}.$$

The next theorem describes how one goes from FELM to MELM, which is the theoretical foundation for the two estimation methods to be developed in [Sections 5 and 6](#). Let

$$\tilde{X} = \langle b, X \rangle_{\mathcal{H}_X}, \quad \tilde{Y} = \langle c, Y \rangle_{\mathcal{H}_Y}, \quad \tilde{\epsilon} = \langle c, \epsilon \rangle_{\mathcal{H}_Y}, \quad (16)$$

where  $b$  and  $c$  are as defined immediately following [Assumption 1](#).

**Theorem 4.** If  $(X, Y)$  follows the FELM in [Definition 1](#) with response envelope  $\mathcal{E}(B; \Sigma_\epsilon)$  and predictor envelope  $\mathcal{E}(B^*; \Sigma_X)$ , then  $(\tilde{X}, \tilde{Y})$  follows the MELM

$$\tilde{Y} = \mu + \Gamma(\phi(I)[B]\psi(J))\Phi^\top \tilde{X} + \tilde{\epsilon},$$

with response envelope  $\mathcal{E}(\phi(I)[B]\psi(J); \Sigma_{\tilde{\epsilon}}) = \text{span}(\Gamma)$ , predictor envelope  $\mathcal{E}(\phi(I)[B]^\top \psi(J); \Sigma_{\tilde{X}}) = \text{span}(\Phi)$ , and  $\mu = \langle c, \alpha \rangle_{\mathcal{H}_Y}$ .

From [Theorem 4](#), we obtain a corollary where  $\{b_1, \dots, b_k\}$  and  $\{c_1, \dots, c_l\}$  are eigenfunctions of  $X$  and  $Y$ ; that is,  $b_i = \psi_i$ ,  $c_j = \phi_j$ , for  $i = 1, \dots, k$ ,  $j = 1, \dots, l$ . We denote  $(\phi_1, \dots, \phi_l)$  by  $\phi(1 : l)$  and  $(\psi_1, \dots, \psi_k)$  by  $\psi(1 : k)$ . Define  $X^\circ = \langle \psi(1 : k), X \rangle_{\mathcal{H}_X}$ ,  $Y^\circ = \langle \phi(1 : l), Y \rangle_{\mathcal{H}_Y}$ , and  $\epsilon^\circ = \langle \phi(1 : l), \epsilon \rangle_{\mathcal{H}_Y}$ . The next corollary shows that, if  $Y$  and  $X$  follow FELM, then  $Y^\circ$  and  $X^\circ$  follow MELM in [Theorem 4](#) with  $\Gamma, \Phi$  taking a special form. This fact is the theoretical foundation for estimation method developed in [Section 6](#).

**Corollary 1.** If  $(X, Y)$  follows FELM with response envelope  $\mathcal{E}(B; \Sigma_\epsilon)$  and predictor envelope  $\mathcal{E}(B^*; \Sigma_{\tilde{X}})$ , then  $(X^\circ, Y^\circ)$  follows MELM in [Theorem 4](#) with

$$\Gamma = \langle \phi(1 : l), \phi(I)^\top \rangle_{\mathcal{H}_Y}, \quad \Phi = \langle \psi(1 : k), \psi(J)^\top \rangle_{\mathcal{H}_X}. \quad (17)$$

If  $l = s$  and  $k = t$  then  $\Gamma = I_s$ ,  $\Phi = I_t$  and the MELM in [Theorem 4](#) reduces to a version of the multivariate linear regression model (1). We refer to this as the full model.

An alternative way to write the MELM in [Theorem 4](#) is through the following parameterization. Let  $\Phi_0 \in \mathbb{R}^{k \times (k-t)}$  be a matrix whose columns form an orthonormal basis of  $\mathcal{E}(\Phi(\phi(I)[B]\psi(J))^\top \Gamma^\top; \Sigma_{\tilde{X}})^\perp$ , and let  $\Gamma_0 \in \mathbb{R}^{l \times (l-s)}$  be a matrix whose columns form an orthonormal basis of  $\mathcal{E}(\Gamma(\phi(I)[B]\psi(J))\Phi^\top; \Sigma_{\tilde{\epsilon}})^\perp$ . Furthermore, let  $\Omega = \Gamma^\top \Sigma_{\tilde{\epsilon}} \Gamma \in \mathbb{R}^{s \times s}$ ,  $\Omega_0 = \Gamma_0^\top \Sigma_{\tilde{\epsilon}} \Gamma_0 \in \mathbb{R}^{(l-s) \times (l-s)}$ ,  $\Delta = \Phi^\top \Sigma_{\tilde{X}} \Phi \in \mathbb{R}^{t \times t}$ , and  $\Delta_0 = \Phi_0^\top \Sigma_{\tilde{X}} \Phi_0 \in \mathbb{R}^{(k-t) \times (k-t)}$ . In this parameterization, the covariance matrices of  $\tilde{\epsilon}$  and  $\tilde{X}$  can be rewritten as

$$\Sigma_{\tilde{X}} = \Phi \Delta \Phi^\top + \Phi_0 \Delta_0 \Phi_0^\top, \quad \Sigma_{\tilde{\epsilon}} = \Gamma \Omega \Gamma^\top + \Gamma_0 \Omega_0 \Gamma_0^\top,$$

and the MELM in [Theorem 4](#) can be restated as  $\tilde{Y}|\tilde{X} \sim N(\mu_{\tilde{Y}|\tilde{X}}, \Sigma_{\tilde{Y}|\tilde{X}})$ , where

$$\mu_{\tilde{Y}|\tilde{X}} = \mu + \Gamma(\phi(I)[B]\psi(J))\Phi^\top \tilde{X}, \\ \Sigma_{\tilde{Y}|\tilde{X}} = \Gamma(\phi(I)[B]\psi(J))\Phi^\top \Sigma_{\tilde{X}} \Phi(\phi(I)[B]\psi(J))^\top \Gamma^\top + \Sigma_{\tilde{\epsilon}}.$$

The parameterized form also applies to the MELM in [Corollary 1](#) with  $\Gamma$  and  $\Phi$  taking the special form (17). [Theorem 4](#) tells us that, under the conditions assumed therein, a FELM can always be converted into a MELM. This means we can use available methodologies and softwares to estimate the converted MELM.

#### 4. From MELM to FELM

Suppose  $(X, Y)$  obeys the FELM in [Definition 1](#). Let  $(\tilde{X}, \tilde{Y})$  be as defined in the last section.

**Theorem 5.** Under the conditions in the last paragraph, if  $(\tilde{X}, \tilde{Y})$  follows the MELM

$$\tilde{Y} = \mu + (\Gamma \eta \Phi^\top) \tilde{X} + \tilde{\epsilon}, \quad (18)$$

with  $\mathcal{E}(\Gamma \eta \Phi^\top; \Sigma_{\tilde{\epsilon}}) = \text{span}(\Gamma)$ ,  $\mathcal{E}(\Phi \eta^\top \Gamma^\top; \Sigma_{\tilde{X}}) = \text{span}(\Phi)$ , then  $(X, Y)$  follows the FELM with  $\phi(I)[B]\psi(J) = \eta$ , where  $\phi(I) = \Gamma^\top c$ ,  $\psi(J) = \Phi^\top b$ .

In the next two sections we will implement the theoretical results of Sections 3 and 4 to develop sample estimates of FELM. In the multivariate setting, since  $X$  and  $Y$  are random vectors, it is natural to build the envelope model directly on  $X$  and  $Y$  themselves. In particular, there seems no reason to build the envelope model, say, on the principal components of  $X$  and  $Y$ . The situation is entirely different in the functional setting. Since we do not have random vectors to begin with, we face two apparent options:

1. building the envelope model directly on the coordinates of  $X$  and  $Y$  with respect to the bases in  $\mathcal{H}_X$  and  $\mathcal{H}_Y$  that we happen to choose. This seems to be the most direct generalization of the original idea of the envelope model;
2. building the envelope model on the coefficients of Karhunen-Loëve expansions of  $X$  and  $Y$ .

We next develop estimation procedures via both routes. In application we do not observe the entire functions  $Y$  and  $X$ , but only at a finite number of points. So we need a method to connect the points. As a pragmatic approach, we assume  $\mathcal{H}_Y$  and  $\mathcal{H}_X$  have finite bases.

## 5. Direct Estimation

We focus on the first approach in this section. To construct orthonormal bases for  $\mathcal{H}_X$  and  $\mathcal{H}_Y$ , we start with any finite bases  $b = (b_1, \dots, b_k)^\top$  for  $\mathcal{H}_X$ , and  $c = (c_1, \dots, c_l)^\top$  for  $\mathcal{H}_Y$ . In practice,  $k$  and  $l$  are picked according to the shape of the functions. For example, if cubic splines are used,  $k$  and  $l$  are decided by the number of knots. More knots are needed for a function with more fluctuation. As an illustration, let  $N = [0, 1]$ , and  $\{t_0, \dots, t_{m_1}\}$  and  $\{s_0, \dots, s_{m_2}\}$  be two sets of nodes in  $[0, 1]$  for  $\mathcal{H}_X$  and  $\mathcal{H}_Y$ , respectively, with  $t_0 = s_0 = 0$ ,  $t_{m_1} = s_{m_2} = 1$ . Let  $\{b_1, \dots, b_k\}$  and  $\{c_1, \dots, c_l\}$  be the cubic spline bases determined by the nodes  $t_0, \dots, t_{m_1}$  and  $s_0, \dots, s_{m_2}$ , respectively. In this case,  $k = m_1 + 3$  and  $l = m_2 + 3$ . The specific forms of these bases can be found, for example, in Wang (2011). We use the  $L^2([0, 1])$  inner product for  $\mathcal{H}_X$  and  $\mathcal{H}_Y$ . For example,  $\langle b_i, b_j \rangle_{\mathcal{H}_X} = \int_0^1 b_i(t) b_j(t) dt$ , which can be written down explicitly for cubic splines. To turn  $b = (b_1, \dots, b_k)^\top$  and  $c = (c_1, \dots, c_l)^\top$  into orthonormal bases, let  $G_b$  and  $G_c$  denote the Gram matrices  $G_b = \{\langle b_i, b_j \rangle_{\mathcal{H}_X}\}_{i,j=1}^{m_1}$  and  $G_c = \{\langle c_i, c_j \rangle_{\mathcal{H}_Y}\}_{i,j=1}^{m_2}$ . The orthonormal bases  $b^*$  and  $c^*$  are then calculated by  $b^* = G_b^{-1/2}b$ , and  $c^* = G_c^{-1/2}c$ . For notational simplicity, we reset  $(b^*, c^*)$  to  $(b, c)$ .

For a function  $h \in \mathcal{H}_X$ , we use  $[h]_b$  to denote the coordinate of  $h$  with respect to the basis  $b$ . That is,  $[h]_b = (\langle b_1, h \rangle_{\mathcal{H}_X}, \dots, \langle b_m, h \rangle_{\mathcal{H}_X})^\top$ , and  $h = [h]_b^\top b$ . Let  $(X_1, Y_1), \dots, (X_n, Y_n)$  be an iid sample from  $(X, Y)$ , and let  $\tilde{X}_i = [X_i]_b$ ,  $\tilde{Y}_i = [Y_i]_c$ ,  $\tilde{X} = [X]_b$ , and  $\tilde{Y} = [Y]_c$ . The coordinates  $[X_i]_b$  and  $[Y_i]_c$  can be estimated by least squares. Let  $\{t_0^*, \dots, t_{m_1^*}^*\}$  and  $\{s_0^*, \dots, s_{m_2^*}^*\}$  be the points at which  $X_i(t)$  and  $Y_i(t)$  are observed. For brevity of notations, we assume that these observed points do not change with  $i$ . The method can be applied directly if  $(X_i, Y_i)$  are observed at different points, see the simulation in Section 9.3 of the supplementary materials. Since  $X_i = [X_i]_b^\top b$ , we obtain  $[X_i]_b$  by estimating the least squares coefficients

from the regression of  $X_i(t_0^*), \dots, X_i(t_{m_1^*}^*)$  on  $b(t_0^*), \dots, b(t_{m_1^*}^*)$ . Specifically, let  $T_X^* = \{t_0^*, \dots, t_{m_1^*}^*\}$ , and let

$$X_i(T_X^*) = \begin{pmatrix} X_i(t_0^*) \\ \vdots \\ X_i(t_{m_1^*}^*) \end{pmatrix}, \quad B(T_X^*) = \begin{pmatrix} b_1(t_0^*) & \dots & b_k(t_0^*) \\ \vdots & \ddots & \vdots \\ b_1(t_{m_1^*}^*) & \dots & b_k(t_{m_1^*}^*) \end{pmatrix}.$$

Then coordinate  $[X_i]_b$  is computed from  $[X_i]_b = [B(T_X^*)^\top B(T_X^*)]^{-1} B(T_X^*)^\top X_i(T_X^*)$ . Note that the above is exactly the spline estimate of the curve  $\{X_i(t) : t \in [0, 1]\}$  as expressed in orthonormal basis. The coordinate of  $Y_i$  with respect to  $c$  is computed similarly. Note that this process does not assume  $t$  to be equally spaced, thus, it is applicable if  $t$  is irregularly distributed. When the observed points are sparse, for example,  $m_1^* < k$ , we can use regularized regression such as ridge regression or lasso to obtain  $[X_i]_b$ . But further investigation will be needed for its theoretical justification. We can also reduce the number of basis functions which is itself a kind of regularization, for instance, if spline functions are used as basis, we can reduce the number of knots. Wang et al. (2016) has more detailed discussion on the various situations where functional data is observed in practice, sparsely or densely observed functional data.

Based on Theorem 4, we fit a MELM with  $\tilde{X}$  as the predictor and  $\tilde{Y}$  as the response

$$\begin{aligned} \tilde{Y} &= \mu + \Gamma \eta \Phi^\top \tilde{X} + \tilde{\epsilon}, & \Sigma_{\tilde{\epsilon}} &= \Gamma \Omega \Gamma^\top + \Gamma_0 \Omega_0 \Gamma_0^\top, \\ \Sigma_{\tilde{X}} &= \Phi \Delta \Phi^\top + \Phi_0 \Delta_0 \Phi_0^\top. \end{aligned} \quad (19)$$

Let  $\beta = \Gamma \eta \Phi^\top$  denote the regression coefficients. Estimation of the parameters in (19) is performed by maximizing the log-likelihood function for  $(\tilde{X}, \tilde{Y})$ . We first estimate the envelopes  $\mathcal{E}(\beta; \Sigma_{\tilde{\epsilon}})$  and  $\mathcal{E}(\beta^\top; \Sigma_{\tilde{X}})$ . Since both  $\mathcal{E}(\beta; \Sigma_{\tilde{\epsilon}})$  and  $\mathcal{E}(\beta^\top; \Sigma_{\tilde{X}})$  are subspaces, the estimation involves optimization over Grassmann manifold. Let  $a, b$  be two positive integers, and  $a > b$ . An  $a \times b$  Grassmann manifold is the set of all  $b$ -dimensional subspace of an  $a$ -dimensional space. An R package *Renvlp* for estimation for envelope models can be found on CRAN, see Lee and Su (2022). Details can be found in Section 6 of the supplementary materials.

Once we have the estimated envelopes  $\hat{\mathcal{E}}(\beta; \Sigma_{\tilde{\epsilon}})$  and  $\hat{\mathcal{E}}(\beta^\top; \Sigma_{\tilde{X}})$ ,  $\hat{\Gamma}$  and  $\hat{\Phi}$  can be taken as any orthonormal basis of  $\hat{\mathcal{E}}(\beta; \Sigma_{\tilde{\epsilon}})$  and  $\hat{\mathcal{E}}(\beta^\top; \Sigma_{\tilde{X}})$ , and  $\hat{\eta} = \hat{\Gamma}^\top \hat{\Sigma}_{\tilde{Y} \tilde{X}} \hat{\Phi} (\hat{\Phi}^\top \hat{\Sigma}_{\tilde{X}} \hat{\Phi})^{-1}$ , where  $\hat{\Sigma}_{\tilde{Y} \tilde{X}}$  denotes the sample covariance matrix between  $\tilde{Y}$  and  $\tilde{X}$ . The regression coefficient is then estimated by  $\hat{\beta} = \hat{\Gamma} \hat{\eta} \hat{\Phi}^\top$ . The MLE of the rest of the parameters are  $\hat{\Omega} = \hat{\Gamma}^\top \hat{\Sigma}_{\tilde{Y} \hat{\Phi}^\top \tilde{X}} \hat{\Gamma}$ ,  $\hat{\Omega}_0 = \hat{\Gamma}_0^\top \hat{\Sigma}_{\tilde{Y}} \hat{\Gamma}_0$ ,  $\hat{\Delta} = \hat{\Phi}^\top \hat{\Sigma}_{\tilde{X}} \hat{\Phi}$ ,  $\hat{\Delta}_0 = \hat{\Phi}_0^\top \hat{\Sigma}_{\tilde{X}} \hat{\Phi}_0$ , and  $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \tilde{Y}_i - \hat{\beta} \frac{1}{n} \sum_{i=1}^n \tilde{X}_i$ .

In Section 8, we will develop the confidence and prediction bands of the FELM regression estimate, which requires the asymptotic distribution of the above estimate. The asymptotic distribution and efficiency gain of the MELM estimator are known (Cook and Zhang 2015b), which we now outline for later use. Let  $\text{vec}(\cdot)$  denote the vector operator that stacks the columns of a matrix into a vector, and  $\text{vech}(\cdot)$  denote the vector-half operator that stacks the lower half (including the diagonal) of a symmetric matrix into a vector. The vector of all parameters of interest under model (19) is

$v_1 = (\text{vec}(\beta)^\top, \text{vech}(\Sigma_{\tilde{\epsilon}})^\top, \text{vech}(\Sigma_{\tilde{X}})^\top, \mu^\top)^\top$ . The vector of all constituent parameters is

$$v_2 = (\text{vec}(\Gamma)^\top, \text{vec}(\eta)^\top, \text{vec}(\Phi)^\top, \text{vech}(\Omega)^\top, \text{vech}(\Omega_0)^\top, \text{vech}(\Delta)^\top, \text{vech}(\Delta_0)^\top, \mu^\top)^\top.$$

Let  $G = \partial v_1 / \partial v_2^\top$  and  $J$  the Fisher information for  $v_1$  under the full model. The explicit forms of  $G$  and  $J$  are in Section 7 of the supplementary materials. Let  $\hat{v}_{1,\text{full}}$  be the maximum likelihood estimator of  $v_1$  under the full model. Then  $\sqrt{n}(\hat{v}_{1,\text{full}} - v_1) \xrightarrow{d} N(0, J^{-1})$ , where  $\xrightarrow{d}$  denotes convergence in distribution. The next theorem is due to Cook and Zhang (2015b).

**Theorem 6.** Assume that  $(\tilde{X}, \tilde{Y})$  follows the MELM (19). Let  $\hat{v}_{1,\text{melm}}$  be the maximum likelihood estimator of  $v_1$  under MELM (19), then

$$\sqrt{n}(\hat{v}_{1,\text{melm}} - v_1) \xrightarrow{d} N(0, V_{\text{melm}}), \quad V_{\text{melm}} = G(G^\top J G)^\dagger G^\top,$$

where  $\dagger$  denotes Moore-Penrose generalized inverse. Furthermore, we have  $G(G^\top J G)^\dagger G^\top \leq J^{-1}$ , which means that  $\hat{v}_{1,\text{melm}}$  is asymptotically more efficient than or as efficient as  $\hat{v}_{1,\text{full}}$ .

## 6. Karhunen-Loève Expansion Based Estimation

We now turn to the second approach outlined at the end of Section 4. Let  $b$  and  $c$  be any finite bases and  $G_b$  and  $G_c$  their Gram matrices. Let  $[X_i]_b$  and  $[Y_i]_c$  be the least squares approximations of the coordinates of  $X_i$  and  $Y_i$ , as explained in Section 5. Let  $(\hat{\tau}_r, \hat{\psi}_r)$  be the  $r$ th eigenvalue-eigenfunction pair of  $\hat{\Sigma}_X$ . As shown in Solea and Li (2020),  $[\hat{\psi}_r]_b = (G_b^{1/2})^\dagger v_r$ , where  $v_r$  is the  $r$ th eigenvector of the matrix  $G_b^{1/2} E_n \{([X]_b - E_n [X]_b)([X]_b - E_n [X]_b)^\top\} G_b^{1/2}$ , and  $\hat{\tau}_r$  is the  $r$ th eigenvalue. The  $r$ th eigenfunction of  $\hat{\Sigma}_X$  is then  $\hat{\psi}_r = [\hat{\psi}_r]_b^\top b$ . The empirical Karhunen-Loève expansion of  $X_i$  is then

$$X_i - E_n X_i = \sum_{r=1}^k \langle X_i - E_n X_i, \hat{\psi}_r \rangle_{\mathcal{H}_X} \hat{\psi}_r = \sum_{r=1}^k \hat{\tau}_r^{1/2} \hat{\xi}_{ir} \hat{\psi}_r,$$

where  $\langle X_i - E_n X_i, \hat{\psi}_r \rangle_{\mathcal{H}_X} = \hat{\tau}_r^{1/2} \hat{\xi}_{ir} = [X_i - E_n X_i]_b^\top G_b [\hat{\psi}_r]_b = [X_i - E_n X_i]_b^\top G_b^{1/2} v_r$ . Similarly, the empirical Karhunen-Loève expansion of  $Y_i$  is

$$Y_i - E_n Y_i = \sum_{r=1}^l \langle Y_i - E_n Y_i, \hat{\phi}_r \rangle_{\mathcal{H}_Y} \hat{\phi}_r = \sum_{r=1}^l \hat{\lambda}_r^{1/2} \hat{\zeta}_{ir} \hat{\phi}_r,$$

where  $\langle Y_i - E_n Y_i, \hat{\phi}_r \rangle_{\mathcal{H}_Y} = [Y_i - E_n Y_i]_c^\top G_c [\hat{\phi}_r]_c = [Y_i - E_n Y_i]_c^\top G_c^{1/2} \hat{w}_r$ , and  $(\hat{\lambda}_r, \hat{w}_r)$  is the  $r$ th eigenvalue-eigenvector pair of the matrix  $G_c^{1/2} E_n \{([Y]_c - E_n [Y]_c)([Y]_c - E_n [Y]_c)^\top\} G_c^{1/2}$ . Let

$$\tilde{X} = ((X, \hat{\psi}_1)_{\mathcal{H}_X}, \dots, (X, \hat{\psi}_k)_{\mathcal{H}_X})^\top, \quad \text{and}$$

$$\tilde{Y} = ((Y, \hat{\phi}_1)_{\mathcal{H}_Y}, \dots, (Y, \hat{\phi}_l)_{\mathcal{H}_Y})^\top.$$

Note that  $\tilde{X}$  and  $\tilde{Y}$  follow MELM (19) from Corollary 1. Estimation of the parameters in MELM is the same as discussed in Section 5.

## 7. Order Determination

In both the direct estimation in Section 5 and the Karhunen-Loève expansion based estimation in Section 6, the dimension  $s$  of the response envelope and the dimension  $t$  of the predictor envelope need to be selected. Let  $(s_0, t_0)$  denote the true dimensions of the response and predictor envelopes, and let  $(s, t)$  denote a generic pair of dimensions that is varied in the set for optimization. To estimate  $(s_0, t_0)$ , we compute the BIC value over a grid of  $(s, t)$  from  $(1, 1)$  to  $(l, k)$ , and find the pair which minimizes BIC. Specifically, for the dimension pair  $(s, t)$ ,  $\text{BIC}(s, t) = -2\hat{\ell}(s, t) + \log(n)K(s, t)$ , where

$$\hat{\ell}(s, t) = \hat{\ell}(\hat{\Gamma}, \hat{\Phi}, \hat{\Omega}, \hat{\Omega}_0, \hat{\Delta}, \hat{\Delta}_0 \hat{\eta}, \hat{\mu}) \quad (20)$$

is the maximum of the likelihood for a fixed  $(s, t)$ , and  $K(s, t) = \frac{1}{2}l(l+1) + \frac{1}{2}k(k+1) + st + l$  is the total number of model parameters when the dimensions of the response and predictor envelopes are  $s$  and  $t$ . Note that all the hatted parameters in (20) depend on  $(s, t)$ .

We estimate  $(s_0, t_0)$  by minimizing the above BIC-type criterion, assuming that we search over the grid of  $\Xi = \{1, \dots, l\} \times \{1, \dots, k\}$ , and denote the estimate by  $(\hat{s}, \hat{t})$ . Note that the consistency of order determination has not yet been proved for the envelope model even in the multivariate setting. So the next result is novel to the envelope model in general.

**Theorem 7.** If  $(\tilde{X}, \tilde{Y})$  satisfies MELM (19), then  $P((\hat{s}, \hat{t}) = (s_0, t_0)) \rightarrow 1$ .

When implementing the dimension selection procedure, we make the following adjustment to speed up the computation. Let  $r$  denote the rank of  $\beta$ . Since  $r$  is no greater than  $\min(s, t)$ , if  $r$  were known, we would only need to compute the BIC value over a grid of  $(s, t)$  from  $(r, r)$  to  $(l, k)$ . Following Cook and Zhang (2015b),  $r$  can be estimated from a Chi-squared test developed in Bura and Cook (2003). The test statistic is  $\Lambda_d = n \sum_{i=d+1}^{\min(k, l)} e_i^2$ , where  $e_1 \geq \dots \geq e_{\min(k, l)}$  are singular values of the matrix  $\hat{\beta}_{\text{std}} = [(n-p-1)/n]^{1/2} \hat{\Sigma}_{\tilde{X}}^{1/2} \hat{\beta}_{\text{OLS}}^\top \hat{\Sigma}_{\tilde{Y}|\tilde{X}}^{-1/2}$ , and  $\hat{\beta}_{\text{OLS}} = \hat{\Sigma}_{\tilde{Y}|\tilde{X}} \hat{\Sigma}_{\tilde{X}}^{-1}$ . According to Bura and Cook (2003), the asymptotic distribution of  $\Lambda_d$  is  $\chi_{(k-d)(l-d)}^2$  under the null hypothesis that  $r = d$ . Let  $\hat{r}$  be the smallest  $d$  for which this null hypothesis is not rejected. Instead of minimizing  $\text{BIC}(s, t)$  over the grid  $\{1, \dots, l\} \times \{1, \dots, k\}$ , we minimize it over  $\{\hat{r}, \dots, l\} \times \{\hat{r}, \dots, k\}$ , which works well in our examples.

## 8. Confidence Band and Prediction Band

In this section, we construct the confidence band for the mean of a new response and the prediction band for the new response itself. Specifically, let  $X_{\text{new}}$  be a new observation on  $X$ . We are interested in constructing the confidence interval for  $E[Y_{\text{new}}(t)]$  and prediction interval for  $Y_{\text{new}}(t)$ , for each  $t \in N$ .

Let  $\hat{Y}_{\text{new}}$  denote the estimate of  $E(Y_{\text{new}})$ , which is also the prediction of  $Y_{\text{new}}$ . We estimate  $E(Y_{\text{new}})$  as follows. First, we compute the coordinates  $[X_{\text{new}}]_b$  of  $X_{\text{new}}$  relative to the basis  $b$ . Then, the coordinates of  $\hat{Y}_{\text{new}}$  with respect to the basis  $c$  in  $\mathcal{H}_Y$  is  $[\hat{Y}_{\text{new}}]_c = \hat{\mu} + \hat{\beta}[X_{\text{new}}]_b$ . Finally, the prediction itself is calculated as  $\hat{Y}_{\text{new}} = [\hat{Y}_{\text{new}}]_c^\top c$ .

**Table 1.** Comparison on mean squared prediction errors.

<i>n</i>	25	50	100	200	400	25	50	100	200	400
Method	Fixed dimension					Selected dimension				
Direct	FELM	6.38	5.81	5.68	5.61	5.56	6.56	5.83	5.68	5.61
	FFFR	7.92	6.39	5.92	5.72	5.61	7.92	6.39	5.92	5.61
	PCR	12.99	12.02	11.69	11.63	11.54	9.78	7.19	5.98	5.74
	PLS	13.01	11.44	10.50	9.57	8.62	12.58	11.82	11.45	11.23
K-L expansion	FELM	6.38	5.81	5.68	5.61	5.56	6.59	5.83	5.68	5.61
	FFFR	7.92	6.39	5.92	5.72	5.61	7.92	6.39	5.92	5.61
	PCR	12.99	12.02	11.69	11.63	11.54	9.78	7.19	5.98	5.74
	PLS	13.01	11.44	10.50	9.57	8.62	12.58	11.82	11.45	11.23

We next approximate the variance of  $\hat{Y}_{\text{new}}(t_0)$  at some  $t_0 \in T$ . Since  $\hat{Y}_{\text{new}} = c^\top [\hat{Y}_{\text{new}}]_c$ , we have  $\hat{Y}_{\text{new}}(t_0) = c(t_0)^\top [\hat{Y}_{\text{new}}]_c$ . Consequently,  $\text{var}[\hat{Y}_{\text{new}}(t_0)] = c(t_0)^\top \text{var}([\hat{Y}_{\text{new}}]_c) c(t_0)$ . Since  $[\hat{Y}_{\text{new}}]_c = \hat{\mu} + \hat{\beta} [X_{\text{new}}]_b = \hat{\mu} + ([X_{\text{new}}]_b^\top \otimes I_l) \text{vec}(\hat{\beta})$ , we have

$$\begin{aligned} \text{var}[\hat{Y}_{\text{new}}(t_0)] &= c(t_0)^\top \{\text{var}(\hat{\mu}) + ([X_{\text{new}}]_b^\top \otimes I_l) \} c(t_0) \\ &\quad \text{var}[\text{vec}(\hat{\beta})] ([X_{\text{new}}]_b \otimes I_l) \} c(t_0). \end{aligned}$$

We approximate  $\text{var}[\text{vec}(\hat{\beta})]$  by  $n^{-1}$  times the asymptotic variance of  $\sqrt{n}[\text{vec}(\hat{\beta}) - \text{vec}(\beta)]$ , which is the upper left  $kl \times kl$  block of  $V_{\text{melm}}$  in [Theorem 6](#). We approximate  $\text{var}(\hat{\mu})$  by  $n^{-1}$  times the asymptotic variance of  $\sqrt{n}(\hat{\mu} - \mu)$ , which is the lower right  $l \times l$  block of  $V_{\text{melm}}$ . In symbols, we use

$$\begin{aligned} \sigma_{\text{est}}^2[\hat{Y}_{\text{new}}(t_0)] &= n^{-1} c(t_0)^\top \{ [V_{\text{melm}}(\hat{\beta}, \hat{\Sigma}_{\hat{\epsilon}}, \hat{\Sigma}_{\hat{X}}, \hat{\mu})]_{44} \\ &\quad + ([X_{\text{new}}]_b^\top \otimes I_l) [V_{\text{melm}}(\hat{\beta}, \hat{\Sigma}_{\hat{\epsilon}}, \hat{\Sigma}_{\hat{X}}, \hat{\mu})]_{11} \\ &\quad ([X_{\text{new}}]_b \otimes I_l) \} c(t_0) \end{aligned}$$

to approximate  $\text{var}[\hat{Y}_{\text{new}}(t_0)]$ , where  $[V_{\text{melm}}(\hat{\beta}, \hat{\Sigma}_{\hat{\epsilon}}, \hat{\Sigma}_{\hat{X}}, \hat{\mu})]_{11}$  is the upper left  $kl \times kl$  block of  $V_{\text{melm}}(\hat{\beta}, \hat{\Sigma}_{\hat{\epsilon}}, \hat{\Sigma}_{\hat{X}}, \hat{\mu})$  and  $[V_{\text{melm}}(\hat{\beta}, \hat{\Sigma}_{\hat{\epsilon}}, \hat{\Sigma}_{\hat{X}}, \hat{\mu})]_{44}$  is the lower right  $l \times l$  block of  $V_{\text{melm}}(\hat{\beta}, \hat{\Sigma}_{\hat{\epsilon}}, \hat{\Sigma}_{\hat{X}}, \hat{\mu})$ . The asymptotic  $(1 - \alpha)$ -confidence interval for  $E[\hat{Y}_{\text{new}}(t_0)]$  is then

$$\begin{aligned} &\left( \hat{Y}_{\text{new}}(t_0) - z_{1-\alpha/2} \sigma_{\text{est}}[\hat{Y}_{\text{new}}(t_0)], \right. \\ &\quad \left. \hat{Y}_{\text{new}}(t_0) + z_{1-\alpha/2} \sigma_{\text{est}}[\hat{Y}_{\text{new}}(t_0)] \right), \end{aligned}$$

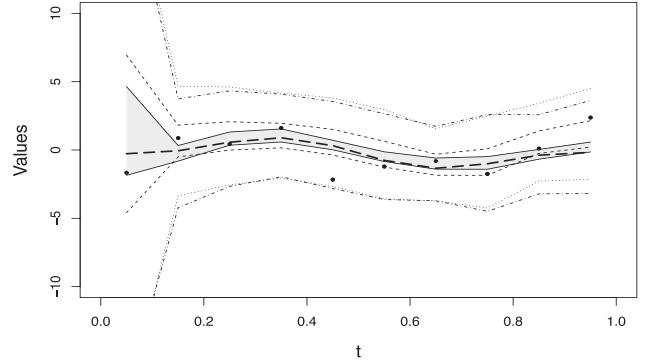
where  $z_{1-\alpha/2}$  is the  $1 - \alpha/2$  percentile of standard normal distribution.

To compute the prediction interval for  $Y_{\text{new}}(t_0)$ , let  $\sigma_{\text{pred}}^2[\hat{Y}_{\text{new}}(t_0)] = \sigma_{\text{est}}^2[\hat{Y}_{\text{new}}(t_0)] + c(t_0)^\top \hat{\Sigma}_{\hat{\epsilon}} c(t_0)$ . The asymptotic  $(1 - \alpha)$ -prediction interval for  $Y_{\text{new}}(t_0)$  is then

$$\begin{aligned} &\left( \hat{Y}_{\text{new}}(t_0) - z_{1-\alpha/2} \sigma_{\text{pred}}^2[\hat{Y}_{\text{new}}(t_0)], \right. \\ &\quad \left. \hat{Y}_{\text{new}}(t_0) + z_{1-\alpha/2} \sigma_{\text{pred}}^2[\hat{Y}_{\text{new}}(t_0)] \right). \end{aligned}$$

## 9. Simulations

We generate  $n$  independent Gaussian random functions using Fourier basis  $X(t) = \sum_{i=1}^{11} \tau_i^{1/2} \xi_i \psi_i(t)$ , and  $\epsilon(t) = \sum_{i=1}^{11} 0.2 \rho_i^{1/2} v_i \chi_i(t)$ , where both  $\chi_i$ 's and  $\psi_i$ 's are basis functions on  $[0, 1]$ :



**Figure 1.** Confidence and prediction bands: solid lines indicate confidence bands of FELM; dashed lines indicate confidence bands of FFFR; the dash-dotted lines represent prediction bands of FELM; the dotted lines represent prediction bands of FFFR. The thick long dashed line in the middle represents true mean response  $E[Y(t)]$ . The black dots represent the observed response at  $0.05, \dots, 0.95$ . For better visibility, the confidence band by FELM is shaded gray.

$$\{1, \sqrt{2} \sin(2\pi t), \sqrt{2} \cos(2\pi t), \sqrt{2} \sin(4\pi t), \sqrt{2} \cos(4\pi t), \dots, \sqrt{2} \sin(10\pi t), \sqrt{2} \cos(10\pi t)\},$$

$\xi_i$ 's and  $v_i$ 's are independent standard normal random variables, and  $\tau_i$ 's and  $\rho_i$ 's are constants. In general  $X$  and  $\epsilon$  need not be in the same Hilbert space; the above construction is for simplicity. The parameters  $\tau_i$ 's and  $\rho_i$ 's are chosen as

$$\begin{aligned} \{\tau_1, \dots, \tau_{11}\} &= \{6.4^2, 1.5.6^2, 2^2, 4^2, 3.2^2, 2.4^2, 1.6^2, 0.8^2, 0.5^2, 0.3^2\}, \\ \{\rho_1, \dots, \rho_{11}\} &= \{1^2, 0.5^2, 3^2, 1^2, 4^2, 2^2, 2.5^2, 3^2, 3.5^2, 4^2, 5^2\}. \end{aligned}$$

The  $b_{ij}$ 's in the linear operator  $B$  are all zero except entries  $b_{22} = -1.25$ ,  $b_{24} = -1$ , and  $b_{42} = b_{44} = 0.4$ . Under this setting,  $\mathcal{H}_X$  and  $\mathcal{H}_Y$  are finite dimensional. The linear operators  $\Sigma_X$  and  $\Sigma_\epsilon$  have the structure  $\Sigma_X = \sum_{i=1}^{11} \tau_i (\psi_i \otimes \psi_i)$  and  $\Sigma_\epsilon = \sum_{i=1}^{11} \rho_i (\chi_i \otimes \chi_i)$ . We set  $I = \{2, 4\}$ ,  $J = \{2, 4\}$ . The predictor envelope  $\mathcal{E}(B^*; \Sigma_X) = \text{span}\{\sqrt{2} \sin(2\pi t), \sqrt{2} \sin(4\pi t)\}$ , the response envelope  $\mathcal{E}(B; \Sigma_\epsilon) = \text{span}\{\sqrt{2} \sin(2\pi t), \sqrt{2} \sin(4\pi t)\}$  and the dimensions of the envelopes are  $s = 2$  and  $t = 2$ . The observed  $t$  are 10 evenly spaced points in  $N = [0, 1]$ , that is,  $\{t_1, \dots, t_{10}\} = \{0.1, \dots, 1\}$ , and they are the same for all the observations. The sample size varies from 50 to 400, and 100 repetitions were generated for each sample size.

In the implementation of FELM, we use the cubic spline basis with five knots 0, 0.25, 0.5, 0.75, and 1 for both  $\mathcal{H}_X$  and  $\mathcal{H}_Y$ . We compare FELM with three existing methods: the full function-on-function regression model (FFFR), the principal component regression (PCR), and the partial least squares

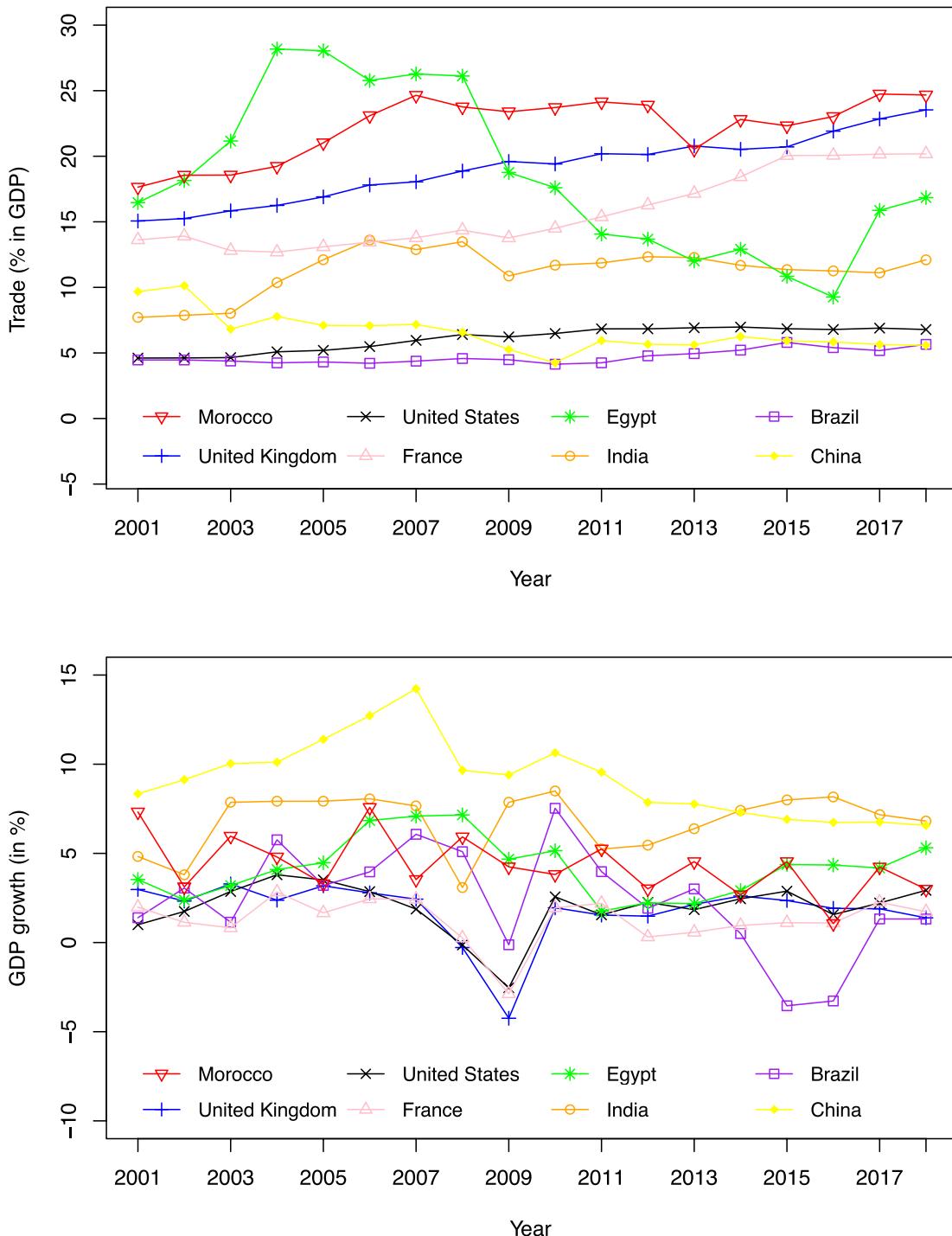


Figure 2. Plot of trade and GDP growth for eight countries.

Table 2. Mean squared prediction errors reduced by FELM as percentages of those of FFFR.

	Africa	America	Asia	Europe
Envelope dimensions $(s, t)$	(1, 2)	(2, 2)	(3, 3)	(5, 2)
Prediction error reduction	9.79%	32.15%	24.22%	24.96%

regression (PLS). Similar to FELM, FFFR, PCR and PLS can be performed either directly on the coordinates of  $X$  and  $Y$  as in Section 5, or on the empirical Karhunen-Lo  e expansions of  $X$  and  $Y$  as in Section 6. Take the direct estimation

as an illustration, we first compute the coordinates  $[X_i]_b$  and  $[Y_i]_c$  for the sample  $(X_1, Y_1), \dots, (X_n, Y_n)$ . FFFR estimates the coefficients  $\beta$  by performing a standard multivariate regression on  $([X_1]_b, [Y_1]_c), \dots, ([X_n]_b, [Y_n]_c)$ . PCR performs a standard multivariate regression of  $[Y_i]_c$  on the first few principal components of  $[X_i]_b$ , and then post-multiply the coefficients by the transpose of the loading matrix of the principal components. PLS is implemented with the R-function *plsr*, as applied to  $([X_1]_b, [Y_1]_c), \dots, ([X_n]_b, [Y_n]_c)$ . In the simulation we used two sets of tuning parameters: (i) we took  $s$  and  $t$  to be the true envelope dimension 2, and chose the numbers of components

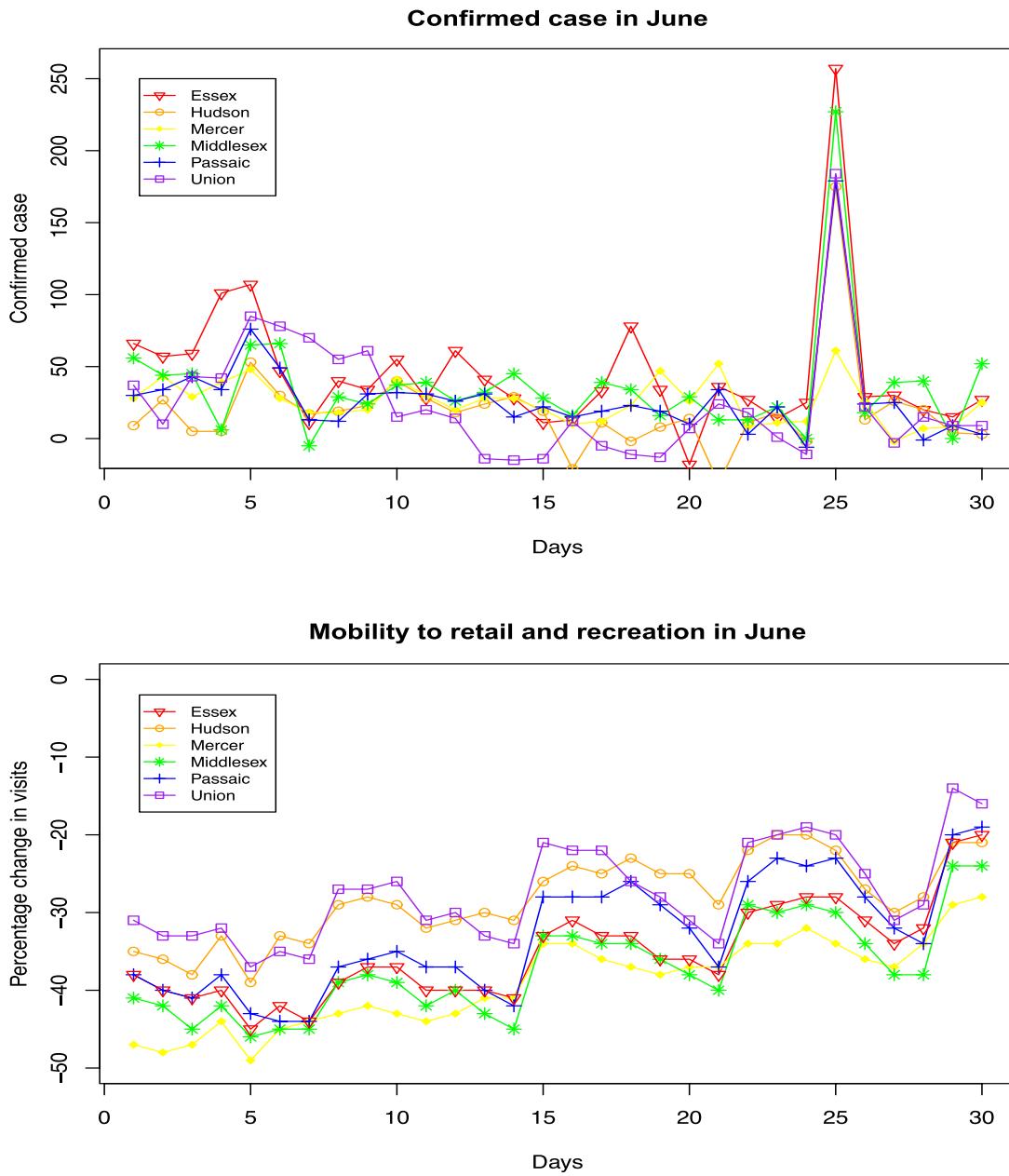


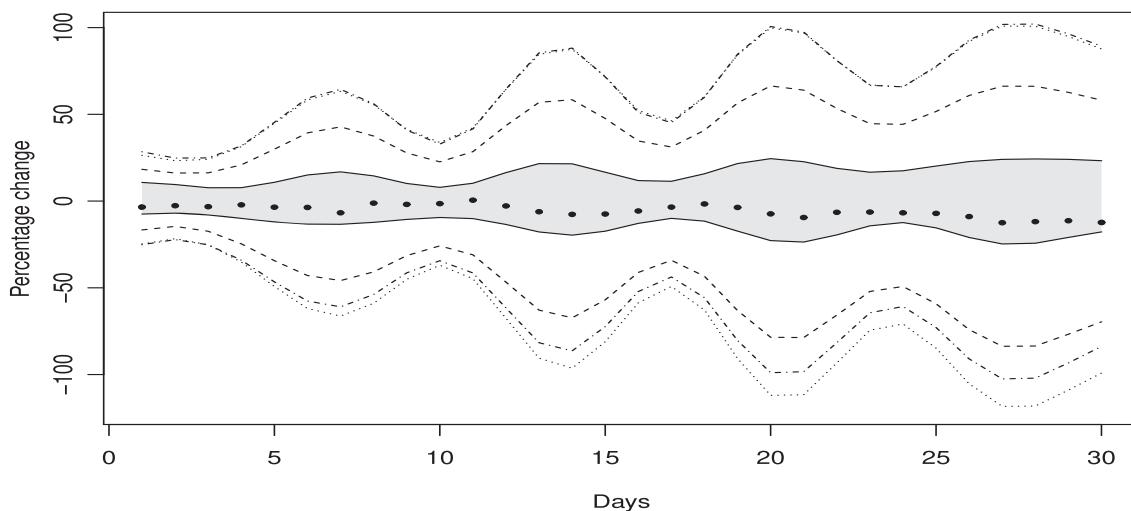
Figure 3. Plot of new confirmed case and mobility to retail and recreation in June.

of PCR and PLS also as 2; (ii) we used the BIC-type criterion developed in Section 7 to estimate  $s$  and  $t$ , and used 5-fold cross-validation to select the numbers of components in PCR and PLS.

We compare the four methods using prediction errors, which are computed by performing 5-fold cross-validation on each of the 100 simulated samples. Table 1 shows the above mean squared error averaged over the 100 simulated samples. Comparing the four methods, we see that FELM significantly outperforms the three other methods. The second best performer is FFFR, while PCR and PLS trail behind. Note that FFFR contains the true envelope model as a submodel, and is therefore asymptotically unbiased. Its difference from FELM is caused completely by the smaller asymptotic variance of FELM. That explains why the difference decreases with the sample size. In comparison, since PCR and PLS each takes only two leading

components, they do not contain the true envelope model as a submodel, and are therefore asymptotically biased. This explains their large prediction errors, and the fact that these errors do not significantly decrease with the sample size. We also note that the results obtained by the direct methods do not significantly differ from those obtained by the Karhunen-Loeve expansion methods.

Due to space limitation, additional results on larger envelope dimensions and summary of computation time are included in Sections 9.1 and 9.2 of the supplementary materials, respectively. FELM also performs well with irregular grid. An additional simulation with each sample  $(X_i, Y_i)$  observed in different random points is included in Section 9.3 of the supplementary materials. During implementation, representation error may occur due to violation of Assumption 1. Thus extra variation is involved



**Figure 4.** Confidence and prediction bands: solid lines indicate confidence bands of FELM; dashed lines indicate confidence bands of FFFR; the dash-dotted lines represent prediction bands of FELM; the dotted lines represent prediction bands of FFFR. The dots represent the observed response in each day of June. For better visibility, the confidence band by FELM is shaded gray.

in estimation of  $\tilde{X}$  and  $\tilde{Y}$ . To investigate its effect on FELM, we include a simulation in Section 9.4 of the supplementary materials.

We now turn to confidence and prediction bands using  $n = 100$  as an illustration. The confidence and prediction intervals were constructed at  $t = 0.05, 0.15, 0.25, \dots, 0.95$ , which are different from the points at which the functional data are observed. Because Table 1 shows no significant difference between the direct and the Karhunen-Loeve expansion based methods, here we only compare the intervals based on the direct method. Figure 1 shows the confidence and prediction bands computed by FELM and FFFR. Comparing the confidence intervals, we see that FELM's intervals are much narrower than the FFFR's intervals, and at the same time has better coverage (note that the blue band does not cover the black curve in the right portion of the plot, whereas the red band does). The prediction intervals of FELM and FFFR have about the same length, which is because the efficiency gain is eclipsed by the noise level. The envelope dimensions in FELM were set at the true values. If the envelope dimensions were estimated by BIC, the confidence and prediction bands were almost identical to those in Figure 1.

## 10. Data Analysis

### 10.1. Economic Data

The World Bank website (<https://data.worldbank.org/>) contains complete data on trade and gross domestic product (GDP) growth for 141 countries from 2001 and 2018. The Trade variable of a country is defined as the sum of exports and imports of goods and services measured as a percentage of GDP of that country. GDP growth is the annual percentage growth rate of GDP at market prices based on constant local currency. Both Trade and GDP growth are functional data, as observed over the 18 years. We treat Trade as the predictor and GDP growth as the response in a function-on-function regression problem, and applied FELM to the data for each continent separately.

Since South America has data for only nine countries, we combined it with North America. The continent of Australia has measurements for only seven countries, we omitted it from our analysis. The sample sizes for Africa, America, Asia, and Europe are 32, 30, 32, and 40, respectively. Spaghetti plots of Trade and GDP growth for eight countries (two for each region) are presented in Figure 2. The dimensions of the envelopes are selected by BIC. The prediction errors are measured by 5-fold cross-validation with 100 random splits. The percentages of the reduction of prediction error by FELM as compared with FFFR for the four regions are shown in Table 2. Overall, FELM achieves very significant improvement over FFFR. Moreover, it is interesting to note that the improvements are especially strong for America.

A possible explanation for this discrepancy in the levels of improvement is that the GDP growths for countries in Africa are more associated with main-stream products such as minerals, crude oil, and agricultural products (coffee, cocoa, cotton, and so on), which are related to the leading eigenvectors of the responses. In this situation, the envelope components and the principal components are somewhat aligned, leading to milder (nonetheless significant) efficiency gains. In comparison, the GDP growths for countries in America are more associated with less main-stream products such as microchips, medicines, vaccines, or airplanes, which are more related to the non-leading eigenvectors of the responses. In this situation, the envelope components are largely orthogonal to the leading principal components, yielding substantial gains in efficiency and prediction. For more discussions on when an envelope would achieve most of its gains, see Cook (2018).

### 10.2. Covid-19 Data

Covid-19 is a global pandemic of the coronavirus disease, which has up-to-date 212 million reported cases in 220 countries and territories. The Open Covid-19 Dataset (<https://github.com/open-covid-19/data#open-covid-19-dataset>) collects daily data for more than 50 countries around the world. We focus

on the daily new confirmed cases and mobility data in June, 2020 in 21 counties in the state of New Jersey. We took the new confirmed cases as the predictor and mobility of retail and recreation as the response. The data for mobility of retail and recreation records percentage change in visits to places like restaurants, shopping centers, museums, and movie theaters compared to a baseline value. The baseline is the median value for the corresponding day of the week during the five week period from January 3 to February 6 in 2020. Figure 3 shows the spaghetti plots for the predictor (upper panel) and response (lower panel) in six counties in New Jersey. We fit FELM to the data. The dimensions of the response envelope and predictor envelope were estimated as 3 and 2 by BIC, respectively. We computed the mean squared prediction errors for FFFR and FELM with 5-fold cross-validation with 100 random splits. The predicted results for Camden County and Atlantic County are in Figures 2 and 3 of the supplementary materials. Compared with FFFR, FELM reduces the prediction error by 62.3%. The 95% confidence and prediction bands for the daily mobility in Hudson County are displayed in Figure 4. Notice that the confidence bands of both FELM and FFFR have good coverage, and FELM has significantly narrower bands than FFFR. We also performed analysis taking the daily new confirmed cases in May as the predictor, and the daily new confirmed cases in June as the response. The dimensions for the response and predictor envelopes are inferred to be 5 and 4 by BIC. FELM reduces the prediction error by 11.66% compared to FFFR.

## Supplementary Materials

The supplementary materials contain proof of theorems, lemmas, and propositions, details on estimation and asymptotics as well as additional simulations.

## Acknowledgments

The authors are grateful for the careful review and helpful suggestions from two anonymous reviewers and Associate Editor.

## Funding

The research of Zhihua Su is partly supported by Simon's Foundation grant 632688. The research of Bing Li supported is partly by National Science Foundation grant DMS-2210775.

## References

Bura, E., and Cook, R. D. (2003), “Rank Estimation in Reduced-Rank Regression,” *Journal of Multivariate Analysis*, 87, 159–176. [\[1630\]](#)

Conway, J. B. (2013), *A Course in Functional Analysis*, New York: Springer. [\[1625\]](#)

Cook, R. D. (2018), *An Introduction to Envelopes: Dimension Reduction for Efficient Estimation in Multivariate Statistics*, Hoboken, NJ: Wiley. [\[1624,1634\]](#)

Cook, R. D., Helland, I., and Su, Z. (2013), “Envelopes and Partial Least Squares Regression,” *Journal of the Royal Statistical Society, Series B*, 75, 851–877. [\[1624,1626\]](#)

Cook, R. D., Li, B., and Chiaromonte, F. (2010), “Envelope Models for Parsimonious and Efficient Multivariate Linear Regression,” *Statistica Sinica*, 20, 927–1010. [\[1624,1627\]](#)

Cook, R. D., and Zhang, X. (2015a), “Foundations for Envelope Models and Methods,” *Journal of the American Statistical Association*, 110, 599–611. [\[1624\]](#)

——— (2015b), “Simultaneous Envelopes for Multivariate Linear Regression,” *Technometrics*, 57, 11–25. [\[1624,1629,1630\]](#)

De Jong, S. (1993), “SIMPLS: An Alternative Approach to Partial Least Squares Regression,” *Chemometrics and Intelligent Laboratory Systems*, 18, 251–263. [\[1626\]](#)

Delaigle, A., and Hall, P. (2012), “Methodology and Theory for Partial Least Squares Applied to Functional Data,” *The Annals of Statistics*, 40, 322–352. [\[1626\]](#)

Ding, S., and Cook, R. D. (2018), “Matrix Variate Regressions and Envelope Models,” *Journal of the Royal Statistical Society, Series B*, 80, 387–408. [\[1624\]](#)

Ding, S., Su, Z., Zhu, G., and Wang, L. (2021), “Envelope Quantile Regression,” *Statistica Sinica*, 31, 79–106. [\[1624\]](#)

Ferraty, F., Van Keilegom, I., and Vieu, P. (2012), “Regression When both Response and Predictor are Functions,” *Journal of Multivariate Analysis*, 109, 10–28. [\[1624\]](#)

Ivanescu, A. E., Staicu, A.-M., Scheipl, F., and Greven, S. (2015), “Penalized Function-on-Function Regression,” *Computational Statistics*, 30, 539–568. [\[1624,1625\]](#)

Lee, M., and Su, Z. (2022), *Renvlp: Computing Envelope Estimators*. R package version 3.1. [\[1629\]](#)

Luo, R., and Qi, X. (2017), “Function-on-Function Linear Regression by Signal Compression,” *Journal of the American Statistical Association*, 112, 690–705. [\[1624\]](#)

Ramsay, J. O., and Silverman, B. W. (2007), *Applied Functional Data Analysis: Methods and Case Studies*, New York: Springer. [\[1625\]](#)

Rekabdarkolaee, H. M., Wang, Q., Naji, Z., and Fuentes, M. (2019), “New Parsimonious Multivariate Spatial Model: Spatial Envelope,” *Statistica Sinica*, 30, 1583–1604. [\[1624\]](#)

Scheipl, F., and Greven, S. (2016), “Identifiability in Penalized Function-on-Function Regression Models,” *Electronic Journal of Statistics*, 10, 495–526. [\[1625\]](#)

Solea, E., and Li, B. (2020), “Copula Gaussian Graphical Models for Functional Data,” *Journal of the American Statistical Association*, 117, 781–793. [\[1630\]](#)

Sun, X., Du, P., Wang, X., and Ma, P. (2018), “Optimal Penalized Function-on-Function Regression Under a Reproducing Kernel Hilbert Space Framework,” *Journal of the American Statistical Association*, 113, 1601–1611. [\[1625\]](#)

Wang, J.-L., Chiou, J.-M., and Müller, H.-G. (2016), “Functional Data Analysis,” *Annual Review of Statistics and Its Application*, 3, 257–295. [\[1629\]](#)

Wang, Y. (2011), *Smoothing Splines: Methods and Applications*, Boca Raton, FL: CRC Press. [\[1629\]](#)

Wold, H. (1966), “Estimation of Principal Components and Related Models by Iterative Least Squares,” in *Multivariate Analysis* (Vol. 59), ed. P. Krishnaiah, pp. 391–420, New York: Academic Press. [\[1626\]](#)

Zhang, X., and Li, L. (2017), “Tensor Envelope Partial Least-Squares Regression,” *Technometrics*, 59, 426–436. [\[1624\]](#)

Zhang, X., Wang, C., and Wu, Y. (2018), “Functional Envelope for Model-Free Sufficient Dimension Reduction,” *Journal of Multivariate Analysis*, 163, 37–50. [\[1625\]](#)