Statistica Sinica Preprint No: SS-2023-0366	
Title	Functional Varying-Coefficient Model Under Heteroskedasticity with Application to DTI Data
Manuscript ID	SS-2023-0366
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202023.0366
Complete List of Authors	Pratim Guha Niyogi, Ping-Shou Zhong and Xiaohong Joe Zhou
Corresponding Authors	Pratim Guha Niyogi
E-mails	pgniyogi@gmail.com

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

FUNCTIONAL VARYING-COEFFICIENT MODEL UNDER HETEROSKEDASTICITY WITH APPLICATION TO DTI DATA

Pratim Guha Niyogi, Ping-Shou Zhong and Xiaohong Joe Zhou

Johns Hopkins University and University of Illinois at Chicago

Abstract: In this paper, we develop a multi-step estimation procedure to simultaneously estimate the varying-coefficient functions using a local-linear generalized method of moments (GMM) based on continuous moment conditions. To incorporate spatial dependence, the continuous moment conditions are first projected onto eigen-functions and then combined by weighted eigen-values, thereby, solving the challenges of using an inverse covariance operator directly. We propose an optimal instrument variable that minimizes the asymptotic variance function among the class of all local-linear GMM estimators, and it outperforms the initial estimates which do not incorporate the spatial dependence. Our proposed method significantly improves the accuracy of the estimation under heteroskedasticity and its asymptotic properties have been investigated. Extensive simulation studies illustrate the finite sample performance, and the efficacy of the proposed method is confirmed by real data analysis.

Key words and phrases: Diffusion tensor imaging; Heteroskedasticity; Local-linear

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

GMM; Moment conditions; Multi-step estimation procedure; Varying-coefficient model.

1. Introduction

Due to modern advancements in technology, varying-coefficient models in functional data have become popular to analyze data coming from several imaging technologies such as magnetic resonance imaging (MRI), diffusion tensor imaging (DTI), etc. We consider the problem of estimating non-parametric coefficient function $\boldsymbol{\beta}(s)$ which is defined on the functional domain (for example, space) $\mathcal{S}$ to understand the relationship between functional response $Y(s)$ and real-valued covariates denoted by $\mathbf{X} = (X_1, \cdots, X_p)^{\mathrm{T}}$, which takes the following form,

$$Y(s) = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}(s) + U(s), \tag{1.1}$$

where $\boldsymbol{\beta}(s) = (\beta_1(s), \cdots, \beta_p(s))^{\mathrm{T}}$ is a p-dimensional vector of unknown smooth functions, and it is assumed that $\boldsymbol{\beta}(\cdot)$ is twice-differentiable with continuous second-order derivatives. The random error $\{U_i(s) : s \in \mathcal{S}\}$ is assumed to be a stochastic process indexed by $s \in \mathcal{S}$ and it characterizes the within-curve dependence with mean zero and an unknown covariance function $\Sigma_{\mathbf{X}}(s, s') = \mathrm{cov}\{U(s), U(s')|\mathbf{X}\}$. To speed up theoretical exploration

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

and facilitate fast computation, this paper mainly focuses on $\mathcal{S}$ in a one-dimensional domain. The extension to a multivariate domain $\mathcal{S}$ is provided in Section S2 of the supplemental file.

The model (1.1) allows heteroskedasticity in the covariance function so that $\Sigma_{\mathbf{x}}(s, s')$ depends on $\mathbf{X}$. There exists limited research on heteroskedastic functional data. For example, Chiou et al. (2003); Jiang and Wang (2011); Ding et al. (2021) considered covariates-dependent functional principal component analysis. However, to the best of our knowledge, no existing inference for the varying coefficient model (VCM) with heteroskedastic functional data has been developed so far. The aim of this paper is to develop an efficient estimator for VCM with heteroskedastic functional data. The model in Equation (1.1) allows its regression coefficient to vary over some predictors of interest. It was introduced in the literature by Hastie and Tibshirani (1993). Because of the wide applicability of VCM, there exists abundant literature on the same, for example, to name a few of them, Fan and Zhang (1999); Wu and Chiang (2000); Fan et al. (2003); Chiou et al. (2004); Ramsay and Silverman (2005); Wang et al. (2008); Zhu et al. (2014). There is a long list of literature on VCM and the aforementioned list is by no means to be exhaustive. A more comprehensive literature review on VCM can be found in Fan et al. (2008). The main difference between a standard VCM and functional

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

VCM is in the error process $U(s)$. The standard VCM typically assumes that $U(s)$ are independent errors so that $U(s)$ and $U(s')$ are independent for $s \neq s'$, while $U(s)$ is a dependent stochastic process in the functional VCM. One important challenge is to consider dependence in the functional VCM. As noted by Lin and Carroll (2001), commonly-used kernel methods are not able to make use of the dependence. Various progress has been made in incorporating dependence into estimation and statistical inference for sparse longitudinal data. For example, Wang (2003) developed an innovative marginal kernel method to incorporate correlation and control bias. Further study in Wang et al. (2005) showed that the method in Wang (2003) achieves the semi-parametric efficient bound. Li (2011) further extended the method to include non-parametric covariance estimation. Qu and Li (2006) developed an estimation method based on penalized spline and quadratic inference. However, these above-mentioned methods are mostly designed for sparse functional data with a small number of repeated measurements. The focus of the current paper is to develop a method to incorporate dependence for dense functional data with heteroskedastic dependence.

In this paper, we develop a functional generalized method of moments (GMM) estimation procedure for such VCM, which does not require distributional assumption and can accommodate heteroskedasticity of unknown

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

form. There exists rich literature on applying GMM to varying coefficient models without functional data and heteroskedasticity. For example, Cai and Li (2008) proposed a one-step local-linear GMM estimator that corresponds to the local-linear GMM discussed in Su et al. (2013) with an identity weight matrix. Tran and Tsionas (2009) provided a local constant two-step GMM estimator with a specified weight matrix by minimizing the asymptotic variance. Su et al. (2013) developed a local-linear GMM estimator procedure of functional-coefficient instrument variable (IV) models with a general weight matrix under exogenous conditions. Cai et al. (2006) proposed a two-step local-linear estimation procedure to estimate the functional coefficient which includes the estimation of high-dimensional non-parametric model in the first step and later estimates the functional coefficients using the first-step non-parametric estimates as a generated regressor. As opposed to the classical GMM, for non-parametric local-linear GMM estimator, the integrated mean square error increases as the number of IVs increases for its arbitrary choice (Bravo, 2021).

The current work is motivated by the problem encountered in diffusion tensor imaging (DTI) where multiple diffusion properties are measured along common major white matter fiber tracts across multiple individuals to characterize the structure and orientation of white matter in the human brain.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

Recently a study has been performed to understand white matter structural alternation using DTI for obstructive sleep apnea patients (Xiong et al., 2017). As an illustration, we present smoothed functional data to analyze the efficiency properties of the network generated by the diffusion properties of the water molecules. In Figure 1, we plot the graphical characteristics of one of the diffusion properties called fractional anisotropy (FA) over different significant levels to obtain the graphical connectivity from 29 patients. Scientists are often interested to know the individual association of average path length (APL) of the network generated from FA with a set of covariates of interests such as age and lapses score. Moreover, in this data, there is sufficient evidence of heteroskedasticity in the covariates. Details about the data-set and associated variables are described in Section 6. We therefore need an estimation procedure which (1) does not need knowledge of the distribution, (2) can handle the heteroskedasticity of covariates, (3) can estimate the non-parametric coefficient functions from VCM, and (4) has a systematic technique for computing an efficient estimator.

In this article, we develop a local-linear GMM estimation procedure for VCM. For given IVs, we propose an optimal local-linear GMM estimator motivated by Lu and Wooldridge (2020). However, the key difference in our approach from the later is that we model the variance of integrated squared

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

error using a non-parametric function of covariates whereas they assume a parametric form in case of classical regression. Therefore, we can ensure that the proposed estimator is at least as efficient as local-linear estimates (initial estimator) and more efficient than that under the presence of heteroskedasticity.

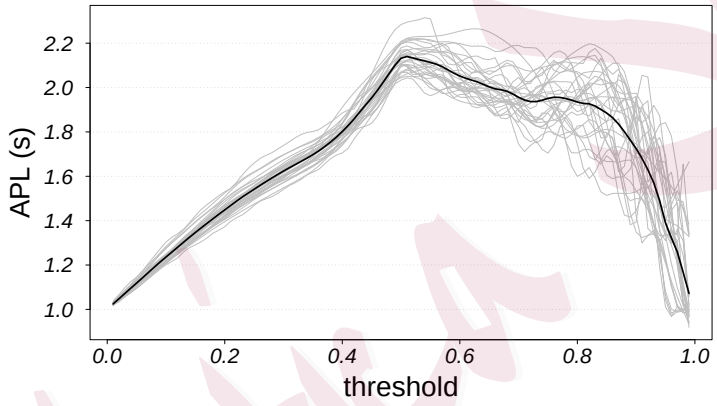

Figure 1: *Apnea-data*: Smoothed average path length (APL) from 29 patients over different thresholds(s). The black solid line indicates the mean of APL over thresholds.

This paper is organized as follows. In Section 2, we introduce our varying-coefficient model and propose a local-linear GMM estimator. In Section 3, we present a multi-step estimation procedure. We establish asymptotic results in Section 4. We perform a set of simulations studied to understand the finite

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

sample performance of the proposed estimator and present those in Section 5. In Section 6, we apply the proposed method in a real imaging data-set on obstructive sleep apnea (OSA). In Section 7, we conclude this article with some discussion. The extension of the proposed method, additional simulation results, and all technical details are provided in the supplementary material.

2. Varying-coefficient functional model and moment conditions

In this section, we first introduce heteroskedastic conditions for SVC model and thereafter, propose a mean-zero function for constructing the GMM estimator.

2.1 Model

Let $\{Y_i(s), \mathbf{X}_i\}$ for $i = 1, \cdots, n$ be independent copies of $\{Y(s), \mathbf{X}\}$. Instead of observing the entire functional trajectory, one can observe $Y(s)$ only on the discrete spatial grid $\{s_1, \cdots, s_r\}$ on the functional domain $\mathcal{S}$. Data can be Gaussian or non-Gaussian and homoskedastic or heteroskedastic depending upon the real applications. Therefore, the observed data for the i-th individual are $\{s_j, Y_i(s_j), \mathbf{X}_i : j = 1, \cdots, r\}$. For simplifying the notation, define $Y_{ij} = Y_i(s_j)$ and $U_{ij} = U_i(s_j)$. Considering the functional principal

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

component analysis (FPCA) model for $U_i(s)$, we assume that $U_i(s)$ is square-integrable and admits the Karhunen-Loève expansion (**?**Loève, 1946). Let $\omega_1(\mathbf{X}) \geq \omega_2(\mathbf{X}) \geq \cdots \geq 0$ be ordered eigen-values of the linear operator determined by $\Sigma_{\mathbf{x}}$ with $\sum_{k=1}^{\infty} \omega_k(\mathbf{X})$ being finite and $\psi_k(s)$'s being the corresponding orthonormal eigen-functions or principal components. Thus, the spectral decomposition (J Mercer, 1909) is given by $\Sigma_{\mathbf{x}}(s, s') = \sum_{k=1}^{\infty} \omega_k(\mathbf{X})\psi_k(s)\psi_k(s')$. Therefore, $U_i(s)$ admits the Karhunen-Loève expansion as follows.

$$U_i(s) = \sum_{k=1}^{\infty} \xi_k(\mathbf{X}_i)\psi_k(s), \tag{2.1}$$

where $\xi_k(\mathbf{X}_i) = \int_{\mathcal{S}} U_i(s)\psi_i(s)ds$, which is termed as the k-th functional principal score for i-th individual. The $\xi_k(\mathbf{X}_i)$ are uncorrelated over k with $\mathbb{E}\{\xi_k(\mathbf{X}_i)|\mathbf{X}_i\} = 0$ and $\text{Var}\{\xi_k(\mathbf{X}_i)|\mathbf{X}_i\} = \omega_k(\mathbf{X}_i)$, $k \geq 1$. Furthermore, assume that the eigen-values vary with $\mathbf{X}_i$ such that $\omega_k(\mathbf{X}_i) = \theta_k\sigma^2(\mathbf{X}_i)$ for some unknown function $\sigma(\mathbf{X}) \geq 0$ and $\theta_1 \geq \theta_2 \geq \cdots \geq 0$. For identifiability, we need some restrictions on θ_ks, such as known or fixed θ_1. Therefore, the above assumption on eigen-values for spectral decomposition allows us to incorporate heteroskedasticity into the model. To the best of our knowledge, this is the first attempt to model SVC with unknown heteroskedasticity.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

2.2 Local-linear mean-zero function

Let us reiterate our main objective: we want to efficiently estimate the varying-coefficient functions based on GMM for the case of continuum moment conditions together with infinite-dimensional parameters. Therefore, we need to construct a mean-zero function which will be described in this sub-section.

Since $\boldsymbol{\beta}(\cdot)$ in model (1.1) is twice continuously differentiable, we can apply the Taylor series expansion to $\boldsymbol{\beta}(s_j)$ around an interior point s_0 and get $\boldsymbol{\beta}(s_j) = \boldsymbol{\beta}(s) + \dot{\boldsymbol{\beta}}(s_0)(s_j - s_0) + \ddot{\boldsymbol{\beta}}(s^*)(s_j - s_0)^2/2$, where s^* lies between s_j and s_0 for all $j = 1, \cdots, r$ and $\dot{\boldsymbol{\beta}}$ and $\ddot{\boldsymbol{\beta}}$ denote the gradients of $\boldsymbol{\beta}$ and $\dot{\boldsymbol{\beta}}$ with respect to s. Thus, $\boldsymbol{\beta}(s_j)$ can be approximated as $\beta_k(s_j) \approx \beta(s_0) + \partial\beta_k(s_0)/\partial s \times (s_j - s_0)$. So in matrix notation, the first-order Taylor series expansion of the coefficient functions becomes

$$\boldsymbol{\beta}(s_j) \approx \mathbf{A}(s_0)\mathbf{z}_h(s_j - s_0), \tag{2.2}$$

where $\mathbf{z}_h(s_j - s_0) = (1, (s_j - s_0)/h)^{\mathrm{T}}$ and $\mathbf{A}(s_0) = [\boldsymbol{\beta}(s_0), h\dot{\boldsymbol{\beta}}(s_0)]$ which is a $p \times 2$ matrix. Hence, applying the approximation procedure in Equation (2.2), we can rewrite model (1.1) as

$$Y_{ij} \approx \mathbf{X}_i^{\mathrm{T}} \{\mathbf{A}(s_0)\mathbf{z}_h(s_j - s_0)\} + U_{ir} = \mathbf{W}_{ij}(s_0)^{\mathrm{T}}\boldsymbol{\gamma}(s_0) + U_{ij}, \tag{2.3}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

such that s_j are sufficiently close to s_0, where $\mathbf{W}_{ij}(s_0) = [\mathbf{z}_h(s_j - s_0) \otimes \mathbf{X}_i]$ and $\boldsymbol{\gamma}(s_0) = (\boldsymbol{\beta}(s_0)^{\mathrm{T}}, h\dot{\boldsymbol{\beta}}(s_0)^{\mathrm{T}})^{\mathrm{T}}$, both of which are vectors of length $2p \times 1$.

Let $K(\cdot)$ be a symmetric probability density function which is used as kernel and $h > 0$ be the bandwidth; thus, the re-scaled kernel function is defined as $K_h(\cdot) = h^{-1}K(\cdot)$. It is easy to see that for a given location $s_0 \in \mathcal{S}$, we can construct a least squares estimator of $\boldsymbol{\gamma}(s)$ defined in Equation (2.3) by minimizing the sample version of the mean squared error $\mathbb{E}\{[Y_{ij} - \mathbf{W}_{ij}^{\mathrm{T}}(s_0)\boldsymbol{\gamma}(s_0)]^2 | \mathbf{X}_i\}$. Let $\mathfrak{M}(\mathbf{X})$ be a q-dimensional IV with $q \geq p$; the moment condition can be written as $\mathbb{E}\left\{ r^{-1} \sum_{j=1}^{r} K_h(s_j - s_0)\boldsymbol{\Delta}_{ij}(s_0) \right\} = \mathbf{0}_q$ where $\boldsymbol{\Delta}_{ij}(s_0) = \mathfrak{M}(\mathbf{X}_i)\left\{ Y_{ij} - \mathbf{W}_{ij}(s_0)^{\mathrm{T}}\boldsymbol{\gamma}(s_0) \right\}$ is a zero mean stochastic process with dimension q. There exist abundance of literature on constructing IVs for optimizing parameter estimations in semi-parametric models with homoskedastic or heteroscedastic (known or unknown) error distributions (e.g., Newey (1994); Amemiya (1977); Ai (1997); Ma et al. (2006); Ghosh et al. (2023)) or parameters defined by moment conditions with or without nuisance unknown non-parametric functions (e.g., Newey (1990); Ai and Chen (2003); Chen and Pouzo (2009)). However, due to the focus of the paper being on estimating non-parametric functions, and the existence of functional dependence and heteroskedasticity of an unknown form, these existing approaches can not be directly undertaken for the model we considered. There

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

are also some papers discussing choosing IVs for optimizing non-parametric function estimators for independent errors (e.g., Cai and Li (2008); Su et al. (2013)). We take this opportunity to investigate the choice of IVs in our framework. For details, please refer to the Remark below Theorem 2.

Motivated by the idea of local-linear estimator and local GMM methods in Cai and Li (2008) and Su et al. (2013), we consider the local-linear IVs $\mathbf{Q}_{ij}(s_0) = (\mathfrak{M}(\mathbf{X}_i), \mathfrak{M}(\mathbf{X}_i)(s_j - s_0)/h)^{\mathrm{T}}$. Therefore, consider the following non-parametrically localizing augmented orthogonal moment conditions for estimating $\boldsymbol{\beta}(s)$.

$$\mathbf{g}_i\{\boldsymbol{\gamma}(s_0)\} = r^{-1}\sum_{j=1}^{r} K_h(s_j - s_0)\mathbf{z}_h(s_j - s_0) \otimes \boldsymbol{\Delta}_{ij}(s_0), \tag{2.4}$$

and note that $\{\mathbf{g}_i(\boldsymbol{\gamma}(s))\} : i = 1, \cdots, n\}$ are independent and $\mathbb{E}\{\mathbf{g}_i(\boldsymbol{\gamma}(s))\} = \mathbf{0}_{2q\times 1}$ for $s \in \mathcal{S}$.

Most of the VCMs that exist in the literature assume homoskedasticity in covariates and are limited to weakly dependent non-parametric models (Su et al., 2013; Sun, 2016), which differs significantly from our model assumptions. In contrast, we assume a spatially VCM under heteroskedasticity of unknown form.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

3. Multi-step estimation procedure

This section develops a multi-step estimation procedure to estimate $\boldsymbol{\beta}(s)$ simultaneously across all $s \in \mathcal{S}$. Essentially, the multi-step procedure can be broken down as, Step-I: an initial estimation; Step-II: estimation of the variance function, mean zero function, and eigen-components and Step-III: GMM estimation. The key ideas of each step are described below.

Step-I. Calculate the least squares estimates of $\boldsymbol{\beta}(s)$ as initial estimates, denoted by $\breve{\boldsymbol{\beta}}(s)$ across all $s \in \mathcal{S}$.

Step-II. Estimate the conditional variance of integrated square residuals nonparametrically and subsequently estimate the covariance of mean-zero function. Estimate the eigen-components using multivariate FPCA.

Step-III. Project the continuous moment conditions onto eigen-functions and then combine them by weighted eigenvalues to incorporate spatial dependence and thus obtain the updated estimate of $\boldsymbol{\beta}(s)$, denoted by $\widehat{\boldsymbol{\beta}}(s)$ across all $s \in \mathcal{S}$.

3.1 Step-I: Initial least squares estimates

We consider a local-linear smoother (Fan and Gijbels, 1996) to obtain an initial estimator of $\boldsymbol{\beta}(\cdot)$ ignoring functional dependencies. In this case, the

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

non-linear least squares function of the model 1.1 can be defined as an objective function given by $\mathcal{J}_{init}\{\boldsymbol{\beta}(\cdot)\} = (nr)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{r}\{Y_{ij} - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}(s_j)\}^2$. By the local-linear smoothing method we estimate $\boldsymbol{\gamma}$ at functional point s_0, by minimizing

$$\mathcal{J}_{init}\{\boldsymbol{\gamma}(s_0)\} = (nr)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)\left\{Y_{ij} - \mathbf{W}_{ij}(s_0)^{\mathrm{T}}\boldsymbol{\gamma}(s_0)\right\}^2. \quad (3.1)$$

The solution of the above least-squares problem can be expressed as

$$\breve{\boldsymbol{\gamma}}(s_0) = \left\{(nr)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)\mathbf{W}_{ij}(s_0)\mathbf{W}_{ij}(s_0)^{\mathrm{T}}\right\}^{-1} \times \left\{(nr)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)\mathbf{W}_{ij}(s_0)Y_{ij}\right\}. \quad (3.2)$$

Consequently, the estimator of the coefficient function vector $\boldsymbol{\beta}(s)$ at s_0 is $\breve{\boldsymbol{\beta}}(s_0) = [(1,0)\otimes\mathbf{I}_p]\breve{\boldsymbol{\gamma}}(s_0)$. We determine the tuning parameter h by using some data-driven techniques such as cross-validation and generalized cross-validation.

3.2 Step-II: Intermediate steps

Step-II consists of two important steps in determining the class of GMM estimator. First, in Step-II.A, we propose a method to obtain optimal IVs and therefore estimate the eigen-components which are used in the local-linear GMM objective function in Step-III. To estimate eigen-components,

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

we essentially need to use a multivariate version of FPCA which is quite uncommon in the literature. We borrow the method proposed by Wang (2008).

Step-II.A: Choice of instrument variables (IVs)

Choosing IVs is critical, and the required identification condition is $q \geq p$, which ensures that the dimension of $\mathbf{Q}_{ij}(s_0)$ is at least equal to the dimension of $\boldsymbol{\gamma}(s_0)$.In our model as discussed in Section 2, the error term has a potential heteroskedasticity of unknown form. We define a set of independent and identically distributed random variables $R_1, R_2, \cdots, R_n$ for n individuals where $R_i = \int U_i^2(s)ds$ for each i, termed as the integrated square of residuals, and $\mathbb{E}\{R_i|\mathbf{X}_i\} = \sigma^2(\mathbf{X}_i)\sum_{k=1}^{\infty}\theta_k$. Therefore, consider the following non-parametric regression problem.

$$\log R_i = \log \sigma^2(\mathbf{X}_i) + \epsilon_i, \tag{3.3}$$

where ϵ_i is the mean zero random variable with constant variance. The above model in Equation (3.3) boils down to the problem of estimation of $\log \sigma^2(\mathbf{X}_i)$ by regressing the logarithmic value of the integrated squared residuals variable on the covariates $\mathbf{X}_i$. This approach is along the lines of Yu and Jones (2004); Wasserman (2006), although used in a different context. Since U_is are not observable, we replace U_i by an efficient estimate that is obtained

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

from Step-I, viz., $\breve{U}_i(s) = Y_i(s) - \mathbf{X}_i^{\mathrm{T}}\breve{\boldsymbol{\beta}}(s)$ for all $s \in \mathcal{S}$. For application, this step can easily be implemented using "`gam`" function available in `mgcv` package in R to get an estimate of the non-parametric mean function, denoted by $\widehat{\mu}(\mathbf{X})$ and therefore $\widehat{\sigma}^2(\mathbf{X}) = \exp\{\widehat{\mu}(\mathbf{X})\}$. Given the estimate of $\sigma(\cdot)$, we can, therefore, choose IVs as $\mathfrak{M}(\mathbf{X}_i) = (\mathbf{X}_i, \mathbf{X}_i/\widehat{\sigma}^2(\mathbf{X}_i))^{\mathrm{T}}$.

Step-II.B: Estimation of eigen-components

Without loss of generality, assume for simplicity that the spectrum of functional domain $\mathcal{S} = [0, 1]$, and the dimension of mean-zero function $\mathbf{g}\{\boldsymbol{\gamma}(s)\} = (g_1\{\boldsymbol{\gamma}(s)\}, \cdots, g_{2q}\{\boldsymbol{\gamma}(s)\})^{\mathrm{T}}$ is $2q$. Note that $\mathbf{g}\{\boldsymbol{\gamma}(s)\}$ in (2.4) is defined on an interval $[0, 1]$ such that $\sum_{l=1}^{2q} \int \mathbb{E}\{g_l^2\{\gamma(s)\}\}ds$ is finite and the covariance function $\mathbf{C}(s, s') = \mathbb{E}\left[\mathbf{g}\{\boldsymbol{\gamma}(s)\}\mathbf{g}\{\boldsymbol{\gamma}(s)\}^{\mathrm{T}}\right]$. Under condition (C6) mentioned in Section 4, using the lining-up method in (Wang, 2008, Chapter 5) and Ramsay and Silverman (2005), define a new stochastic process $e(s_*)$ on the interval $[0, 2q]$ with eigen-function ϕ_e such that, $e(s_*) = g_l\{\boldsymbol{\gamma}(s_* - (l-1))\}$ and $\phi_e(s_*) = \phi_l\{\boldsymbol{\gamma}(s_* - (l-1))\}$ for $l - 1 \leq s_* < l$, $l = 1, \cdots 2q$, where we define the eigen-function for each g_l as ϕ_l for $l = 1, \cdots, 2q$. Therefore, the covariance function between $e(s_*)$ and $e(s'_*)$ can be expressed as $C_{l,l'}(s_*, s'_*) = \mathrm{cov}\{e(s_*), e(s'_*)\}$ for $l-1 \leq s_* < l$ and $l'-1 \leq s'_* < l'$; $l, l' = 1, \cdots, 2q$. Note that, for $2q$-dimensional processes, the Fredholm integral equation is equiv-

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

alent to $2q$-simultaneous integral equations, where each of them corresponds to a specific functional interval of $e(s_*)$. For $l-1 \leq s_* < l$; $l = 1, \cdots, 2q$, the Fredholm integral equation yields $\int_0^{2q} \text{cov}\{e(s_*), e(s'_*)\}\phi_e(s_*)ds_* = \lambda\phi_e(s_*)$. Now, for $0 < s < 1$, the above relation is equivalent to the following.

$$\sum_{l'=1}^{2q} \int_0^1 \text{cov}\{g_l\{\boldsymbol{\gamma}(s)\}, g_{l'}\{\boldsymbol{\gamma}(s')\}\}\phi_{l'}(s')ds' = \lambda\phi_l(s). \tag{3.4}$$

In a multivariate setting, the orthogonality condition becomes

$$\mathbf{1}(l = l') = \int_0^{2q} \phi_{e,l}(s_*)\phi_{e,l'}(s_*)ds_* = \sum_{k=1}^{2q} \int_0^1 \phi_{k,l}(s)\phi_{k,l'}(s)ds. \tag{3.5}$$

Using the generalized Mercer's theorem (J Mercer, 1909), the results for the covariance function can be briefly shown using the lining-up method. Assume that the covariance function is continuous after the lining up processes, so for $(l-1) \leq s_* < l$ and $(l'-1) \leq s'_* < l'$; $l, l' = 1, \cdots, 2q$, the covariance function between $g_l(s)$ and $g_{l'}(s')$ can be expressed as

$$C_{l,l'}(s, s') = \sum_{k=1}^{\infty} \lambda_k \phi_{k,l}\{s_* - (l-1)\}\phi_{k,l'}\{s'_* - (l'-1)\}. \tag{3.6}$$

Therefore, using the above argument, we can define the multivariate spectral decomposition $\mathbf{C}(s, s') = \sum_{k=1}^{\infty} \lambda_k \boldsymbol{\phi}_k(s)\boldsymbol{\phi}_k(s')^{\mathrm{T}}$ for $s, s' \in [0, 1]$ with the orthogonality condition (3.5). After the lining-up process, data are univariate and hence we can adapt the existing techniques of estimating functional eigen-values and eigen-functions in the literature (Yao et al., 2005; Müller

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

and Yao, 2010; Li and Hsing, 2010) to estimate λ and $\phi_e(s)$, and hence can estimate $\boldsymbol{\phi}(s)$ by stacking all components for aligned eigen-functions $\phi_e(s)$.

3.3 Step-III: Final estimates

Finally, we demonstrate our proposed estimator based on local-linear GMM where the proposed mean-zero function can be projected onto eigen-function and then combined by the weighted eigen-values. For any positive α, the objective function is given by

$$\begin{aligned}\mathcal{J}\{\boldsymbol{\gamma}(s_0)\} &= \sum_{k=1}^{\infty} \frac{\widehat{\lambda}_k}{\widehat{\lambda}_k^2+\alpha} \left\{\overline{\mathbf{g}}(\boldsymbol{\gamma}(s_0))^{\mathrm{T}}\widehat{\boldsymbol{\phi}}_k(s_0)\right\}^2 \\ &= \sum_{k=1}^{\infty} \frac{\widehat{\lambda}_k}{\widehat{\lambda}_k^2+\alpha} \left\{(nr)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j-s_0)\widehat{\boldsymbol{\phi}}_k(s_0)^{\mathrm{T}}\mathbf{Q}_{ij}(s_0)\left[Y_{ij}-\mathbf{W}_{ij}(s_0)^{\mathrm{T}}\boldsymbol{\gamma}(s_0)\right]\right\}^2. \end{aligned} \tag{3.7}$$

By minimizing the above objective function, we obtain

$$\begin{aligned}&\sum_{k=1}^{\infty} \frac{\widehat{\lambda}_k}{\widehat{\lambda}_k^2+\alpha} \left\{(nr)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j-s_0)\widehat{\boldsymbol{\phi}}_k(s_0)^{\mathrm{T}}\mathbf{Q}_{ij}(s_0)\mathbf{W}_{ij}(s_0)\right\} \\ &\qquad \times \left\{(nr)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j-s_0)\widehat{\boldsymbol{\phi}}_k(s_0)^{\mathrm{T}}\mathbf{Q}_{ij}(s_0)\left[Y_{ij}-\mathbf{W}_{ij}(s_0)^{\mathrm{T}}\boldsymbol{\gamma}(s_0)\right]\right\} \\ &:= \sum_{k=1}^{\infty} \frac{\widehat{\lambda}_k}{\widehat{\lambda}_k^2+\alpha} \mathcal{X}_k(s_0)\left\{\mathcal{Y}_k(s_0)-\mathcal{X}_k(s_0)^{\mathrm{T}}\boldsymbol{\gamma}(s_0)\right\}, \end{aligned} \tag{3.8}$$

where $\mathcal{X}_k(s_0) = (nr)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j-s_0)\mathbf{W}_{ij}(s_0)\mathbf{Q}_{ij}(s_0)^{\mathrm{T}}\widehat{\boldsymbol{\phi}}_k(s_0)$ and $\mathcal{Y}_k(s_0) = (nr)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j-s_0)\mathbf{Q}_{ij}(s_0)^{\mathrm{T}}\widehat{\boldsymbol{\phi}}_k(s_0)Y_{ij}$. Therefore, the final estimate of

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

the coherent function is $\widehat{\boldsymbol{\beta}}(s_0) = [(1,0) \otimes \mathbf{I}_p]\widehat{\boldsymbol{\gamma}}(s_0)$ where

$$\widehat{\boldsymbol{\gamma}}(s_0) = \left\{ \sum_{k=1}^{\infty} \frac{\widehat{\lambda}_k}{\widehat{\lambda}_k^2 + \alpha} \mathcal{X}_k(s_0) \mathcal{X}_k(s_0)^{\mathrm{T}} \right\}^{-1} \left\{ \sum_{k=1}^{\infty} \frac{\widehat{\lambda}_k}{\widehat{\lambda}_k^2 + \alpha} \mathcal{X}_k(s_0) \mathcal{Y}_k(s_0) \right\}. \quad (3.9)$$

The Algorithm 1 in the supplementary material summarizes the proposed method. For demonstration purposes, we choose the tuning parameters using cross-validation as discussed in the algorithm. In the proposed algorithm, α controls the number of eigen-values, and can be chosen so that condition (C8) defined in Section 4 is valid. Furthermore, it is essential to establish a continuity criterion for alignment to provide theoretical validation. Even when there is a lack of continuity in ϕ_e, empirical studies suggest that the final outcomes remain suitable for practical application. We also discuss the extension of the proposed method to the multivariate domain with $s \in [0,1]^d$ in Section S2 of the supplementary material.

4. Asymptotic results

In this section, we provide some assumptions and then study the asymptotic properties of the local-linear GMM estimator. Here, we allow the sample size n and the number of functional domains r to grow to infinity. Detailed technical proofs are provided in the online Supplementary Material.

Let $\boldsymbol{\beta}_0(s_0)$ be the true value of $\boldsymbol{\beta}(s_0)$ at location s_0. Consider the follow-

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

ing conditions that will be useful in asymptotic results.

(C1) Kernel function $K(\cdot)$ is a symmetric density function defined on the bounded support $[-1, 1]$ and is Lipschitz continuous.

(C2) Density function f_T of T is bounded above and away from infinity, and also below and away from zero. Moreover, f is differentiable and the derivative is continuous.

(C3) $\mathbb{E}\{\|X\|^a\} < \infty$ and $\mathbb{E}\{\sup_{s\in\mathcal{S}} |U(s)|^a\} < \infty$ for some positive $a > 1$. Define, $\mathbb{E}\{\mathfrak{M}(\mathbf{X})\mathbf{X}^{\mathrm{T}}\} = \mathbf{\Omega}$ with rank p.

(C4) The true coefficient function $\boldsymbol{\beta}_0(s)$ is three-times continuously differentiable and $\Sigma_{\mathbf{x}}(s, s')$ are twice continuously differentiable.

(C5) $\{U(s) : s \in [0, 1]\}$ and $\{\mathfrak{M}(\mathbf{X})U(s) : s \in [0, 1]\}$ are Donsker class, where $\mathbf{X} \subset \mathfrak{M}(\mathbf{X})$.

(C6) (a) $\lim_{s\searrow 1} \mathbb{E}\{|g_l\{\boldsymbol{\gamma}(s-1)\} - g_l\{\boldsymbol{\gamma}(0)\}|^2\} = 0$ for $l = 1, \cdots, 2q$

(b) $\lim_{s\nearrow 1} \mathbb{E}\{|g_{l-1}\{\boldsymbol{\gamma}(s)\} - g_l\{\boldsymbol{\gamma}(0)\}|^2\} = 0$ for $l = 2, \cdots, 2q$.

(C7) All second-order partial derivatives of $\mathbf{C}(s, s')$ exist and are bounded on the support of the functional domain.

(C8) For some $\kappa_0 \geq 1$ and $\alpha^{-1} = o\left(\sum_{k=1}^{\kappa_0} \lambda_k^{-1} / \sum_{k=\kappa_0+1}^{\infty} \lambda_k\right)$.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

(C9) The numbers of individuals and functional grid-points are growing to infinity such that $h \to 0$ and $rh \to \infty$ as $n \to \infty$. For $a > 2$, $h^{-4}(\log n/n)^{1-2/a} \to 0$ and $|\log h|^{1-2/a}/h \leq r^{1-2/a}$ for $a \in (2, 4)$.

Remark 1. Conditions (C1) and (C2) are commonly used in the literature of non-parametric regression. The bound condition for the density function in (C2) of functional points is standard for random design. Similar results can be obtained for fixed design where the grid-points are pre-fixed according to the design density $\int_0^{s_j} f(s)dt = j/r$ for $j = 1, \cdots, r$, for $r \geq 1$. The condition (C3) is similar to that in Li and Hsing (2010) which requires the bound on the higher order moment of $\mathbf{X}$. Moreover, the condition on the rank of $\mathbf{\Omega}$ is required for the identification of the functional coefficient and its first-order derivatives (Su et al., 2013).

To obtain the asymptotic expression of $\widehat{\boldsymbol{\beta}}(s)$, observed for fixed sample size, there exists κ_0 such that $k \leq \kappa_0$, λ_k^2 is much larger than α, thus, the ratio $\lambda_k/(\lambda_k^2 + \alpha) \approx \lambda_k^{-1}$. On the other hand, if $k > \kappa_0$, $\lambda_k^2 << \alpha$, as a result, the fraction $\lambda_k/(\lambda_k^2 + \alpha)$ can be approximately written as λ_k/α. Therefore, by the assumption (C8), we can write, for $s \in \mathcal{S}$, $\sum_{k=1}^{\kappa_0} \lambda_k^{-1}\boldsymbol{\phi}_k(s)\boldsymbol{\phi}_k(s')^{\mathrm{T}} + \alpha^{-1}\sum_{k=\kappa_0+1}^{\infty} \lambda_k\boldsymbol{\phi}_k(s)\boldsymbol{\phi}_k(s')^{\mathrm{T}} = \sum_{k=1}^{\kappa_0} \lambda_k^{-1}\boldsymbol{\phi}_k(s)\boldsymbol{\phi}_k(s')^{\mathrm{T}}\{1 + o(1)\}$. Condition (C9) provides the range of bandwidth. Under the fixed sampling design, this condition can be weakened, see Zhu et al. (2012). Due to the limited space,

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

detailed remarks on conditions (C4) and (C6) are included in Section S3 of the supplemental file.

The following Theorem provides weak convergence of the initial estimates in the above Step-I. Define $\mathbf{\Sigma}^*_{\mathbf{x}}(s_0, s_0) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\{\mathbf{X}_i \mathbf{X}_i^{\mathrm{T}} \Sigma_{\mathbf{x}_i}(s_0, s_0)\}$.

Theorem 1. *Let* $\nu_{a,b} = \int t^a K^b(t)dt$. *Under conditions (C1)-(C5), and (C9),*

$$\left\{ \sqrt{n} \left(\breve{\boldsymbol{\beta}}(s_0) - \boldsymbol{\beta}_0(s_0) - 0.5h^2 \nu_{21} \ddot{\boldsymbol{\beta}}_0(s_0) \right) : s_0 \in \mathcal{S} \right\}$$

weakly converges to a mean zero Gaussian process with a covariance matrix $\mathbf{\Omega}_{\mathbf{x}}^{-1} \mathbf{\Sigma}^*_{\mathbf{x}}(s_0, s_0) \mathbf{\Omega}_{\mathbf{x}}^{-1}$ *where* $\mathbf{\Omega}_{\mathbf{x}} = \mathbb{E}\{\mathbf{X}\mathbf{X}^T\}$.

Next, we study the convergence rates of the estimated eigen-components based on the proposed lining-up method. For simplicity, define, $\delta_{n1}(h) = \{(1 + (hr)^{-1}) \log n/n\}^{1/2}$ and $\delta_{n2}(h) = \{(1 + (hr)^{-1} + (hr)^{-2}) \log n/n\}^{1/2}$. The following lemma is the output of the asymptotic expansion of eigen-components of an estimated covariance function.

Lemma 1. *Under assumptions (C1)-(C3), (C6)-(C9), the following convergence holds almost surely for any finite-dimensional mean-zero function* $\mathbf{g}(s)$.

1. $\left| \widehat{\lambda}_k - \lambda_k \right| = O\{h^2 + \delta_{n1}(h) + \delta_{n2}(h)\}$
2. $\sup_{s_0 \in \mathcal{S}} \left| \widehat{\boldsymbol{\phi}}_k(s_0) - \boldsymbol{\phi}_k(s_0) \right| = O\{h^2 + \delta_{n1}(h) + \delta^2_{n2}(h)\}$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

for all $k = 1, \cdots, \kappa_0$.

We skip the proof of the above lemma as it is well-developed in the literature of functional data analysis including Hall (2004); Hall and Hosseini-Nasab (2006); Li and Hsing (2010). If $r^{-1} = O(\{n/\log n\}^{1/4})$ and $h = O(\{n/\log n\}^{-1/4})$, then Lemma 1 implies that both eigenvalues $\widehat{\lambda}_k$ and eigenfunctions $\widehat{\boldsymbol{\phi}}_k(s_0)$ converge at the order of $O\{(\log n/n)^{1/2}\}$. Next, we show the asymptotic results of the proposed estimation.

Theorem 2. *Let* $\nu_{a,b} = \int t^a K^b(t)dt$. *Under the conditions (C1)-(C9), for the proposed local-linear GMM estimator* $\widehat{\boldsymbol{\beta}}(s)$, *the following results hold.*

$$\left\{\sqrt{n}\left(\widehat{\boldsymbol{\beta}}(s) - \boldsymbol{\beta}_0(s) - 0.5h^2\nu_{21}\ddot{\boldsymbol{\beta}}(s)\right) : s \in \mathcal{S}\right\}$$

weakly converges to a mean zero Gaussian process with a covariance function $\mathcal{A}(s_0, s_0) = \mathcal{B}^{-1}(s_0, s_0)\boldsymbol{\Omega}^T\mathbf{C}^{-1}_{\kappa_0,11}(s_0, s_0)\boldsymbol{\Sigma}^*_{\mathfrak{M}}(s_0, s_0)\mathbf{C}^{-1}_{\kappa_0,11}(s_0, s_0)\boldsymbol{\Omega}\mathcal{B}^{-1}(s_0, s_0)$, *where* $\mathcal{B}(s_0, s_0) = \boldsymbol{\Omega}^T\mathbf{C}^{-1}_{\kappa_0,11}(s_0, s_0)\boldsymbol{\Omega}$, $\boldsymbol{\Omega}$ *is defined in condition (C3),* $\mathbf{C}^{-1}_{\kappa_0,11}$ *is given by*

$$\mathbf{C}^{-1}_{\kappa_0}(s, s') = \sum_{k=1}^{\kappa_0}\lambda_k^{-1}\boldsymbol{\phi}_k(s)\boldsymbol{\phi}_k(s')^T = \begin{pmatrix}\mathbf{C}^{-1}_{\kappa_0,1,1}(s, s') & 0\\ 0 & \mathbf{C}^{-1}_{\kappa_0,2,2}(s, s')\end{pmatrix}, \quad (4.1)$$

and

$$\boldsymbol{\Sigma}^*_{\mathfrak{M}}(s_0, s_0) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\{\mathfrak{M}(\mathbf{X}_i)\mathfrak{M}(\mathbf{X}_i)^T\Sigma_{\mathbf{x}_i}(s_0, s_0)\}. \quad (4.2)$$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

Remark 2. The asymptotic variance-covariance of $\widehat{\boldsymbol{\beta}}(s)$ depends on the choices of IVs in $\boldsymbol{\Omega}$. The suggested choice of IVs in Step-II.A of Section 3.2 may be optimal in the sense that it minimizes the variance-covariance matrix of $\hat{\boldsymbol{\beta}}(s)$ among the class of all local-linear GMM estimators. Please refer to the supplemental file for a detailed discussion.

5. Simulation studies

We conduct numerical studies to compare finite sample performance under different correlation structures and heterogeneity conditions. Data are generated from the model $Y_i(s) = X_i\boldsymbol{\beta}(s) + U_i(s)$, where we generate trajectories observed at r spatial locations for the i-th curve, $i = 1, \cdots, n$. Assume that the functional fixed effect is $\beta(s) = \cos(2\pi s)$ and the corresponding fixed effect covariate is generated from a normal distribution with unit mean and variance. The error process is generated as $U_i(s) = \xi_1(X_i)\psi_1(s) + \xi_2(X_i)\psi_2(s)$, where $\xi_1(X_i)$ and $\xi_2(X_i)$ are independent central normal random variables with variance $3\sigma^2(X_i)\theta_0^2$ and $1.5\sigma^2(X_i)\theta_0^2$. Here, θ_0 is determined by the relative importance of the random error signal-to-noise ratio, denoted as SNR_θ which is interpreted as the ratio of the standard deviation of the additive prediction without noise divided by the standard error of the random noise function. For example, $\text{SNR}_\theta = 0.5$ means that

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

the contribution of each functional random noise to the variability in $Y(s)$ is about twice of that of the fixed effect (Scheipl et al., 2015). Here, we use scaled orthonormal functions $\psi_1(s) \propto (1.5 - \sin(2\pi s) - \cos(2\pi s))$ and $\psi_2(s) \propto \sin(4\pi s)$; due to orthonormality, the proportionality constants can easily be determined. Contributions to the conditional variances in $\xi_k(X)$ are specified as (S.0) $\sigma^2(x) = 1$ (homoskedastic), (S.1) $\sigma^2(x) = (1 + x^2/2)^2$, (S.2) $\sigma^2(x) = \exp(1 + x^2/2)$, (S.3) $\sigma^2(x) = \exp(1 + |x| + x^2)$ and (S.4) $\sigma^2(x) = (1 + |x|/2)^2$. We sample the trajectories at r equidistant spatial points $\{s_1, \cdots, s_r\}$ on $[0, 1]$. Let $s_i = (j - 0.5)/r$ for $j = 1, \cdots, r$ for i-th curve. The number of spatial points is assumed to be $r = 200$ for each case. We set number of trajectories $n \in \{30, 50, 100, 200, 500\}$ and the controlling parameter θ_0 is determined using signal-to-noise ratio, SNR_θ which is assumed to be either 0.5 or 1. Here, we perform 500 simulation replicates. To make it consistent with theoretical results and numerical examples, we use "`FPCA`" function in R which is available in `fdapace` package (Gajardo et al., 2021) to estimate the eigen-functions. Bandwidths are selected using five-fold generalized cross-validation in all situations and for estimation, the Epanechnikov kernel $K(x) = 0.75(1 - x^2)_+$ is used; where $(a)_+ = \max(a, 0)$. Accuracy of the parameter estimation is assessed using integrated mean square error and integrated mean absolute error, which for

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

the b-th replication is defined as $\text{IMSE}_b = \left[\sum_{j=1}^{r}\left(\widehat{\boldsymbol{\beta}}_b(s_j) - \boldsymbol{\beta}(s_j)\right)^2 \Delta(s_r)\right]$ and $\text{IMAE}_b = \left[\sum_{j=1}^{r}|\widehat{\boldsymbol{\beta}}_b(s_j) - \boldsymbol{\beta}(s_j)|\Delta(s_r)\right]$ respectively, with $\Delta(s_j) = s_j - s_{j-1}$ where $s_0 = 0$ and $s_1 < \cdots < s_r$ are the observed points over the support set of observational points. Let h^* be the optimal bandwidth obtained from five-fold cross-validation, which when multiplied by a constant within a certain range provides improved results. According to the comments under Lemma 1, undersmoothing is needed, so we use $\widehat{\boldsymbol{\beta}}$ corresponding to bandwidth $0.75h^*$ for our numerical studies. Similar strategies have also been applied in Cai et al. (2006); Wei and Sun (2017); Wang et al. (2017). The constant 0.75 is not an optimal choice but it was recommended based on our numerical experiments.

We present Tables 1 here and S1 in the supplementary document where IMSEs and IMAEs are averaged over 500 replications for each situation. In parentheses, the corresponding standard deviations are reported. We denote by `LLE` and `LLGMM` the local-linear smoothing estimator described in Step-I and local-linear GMM with weight matrix proposed in Step-III in Section 3 respectively. In addition, we have compared the proposed method with that of Wei and Sun (2017), setting the spatial autoregressive parameter to zero. This approach is referred to as `LLWS` in Tables 1 and S1. As expected, for

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

all situations, the IMSE and IMAE are significantly reduced if we increase sample size and/or SNR_θ. For the homoskedastic case, the error rates of `LLE` are similar for `LLGMM` but under the presence of heteroskedasticity of unknown form, our proposed method outperforms. More simulation results with multiple covariates are included in Section S4.2 of the supplementary material.

6. Real data analysis

For illustrating the application of our proposed method and the estimation procedure therein, we use *Apnea-data* to understand white matter structural alterations using diffusion tensor imaging (DTI) in obstructive sleep apnea (OSA) patients (Xiong et al., 2017). The data consists of 29 male patients (age range: 30-55 years) who underwent the study for the diagnosis of continuous positive airway pressure (CPAP) therapy. DTI was performed at 3T, followed by the analysis using tract-based spatial statistics to investigate the difference in fractional anisotropy (FA) and other diffusion properties between the groups based on lapses. FA measures the degree of anisotropy of a diffusion process. The image acquisition is as follows: Images are recorded on a 3T MRI scanner using a commercial 32-channel head coil. An axial T1-weighted image of the brain (3D-BRAVO) is collected with repetition time

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

Table 1: Comparison among the proposed LLGMM with the local linear estimator (LLE) and Wei and Sun (2017)'s approach (LLWS) for $\mathrm{SNR}_\theta = 0.5$.

	n = 30		n = 50		n = 100		n = 200		n = 500	
Method	IMSE	IMAE	IMSE	IMAE	IMSE	IMAE	IMSE	IMAE	IMSE	IMAE
					Case: S.0					
LLE	0.0626	0.1864	0.0372	0.1429	0.0189	0.1016	0.0099	0.0737	0.0041	0.0472
	(0.0619)	(0.0951)	(0.0398)	(0.0757)	(0.0200)	(0.0540)	(0.0108)	(0.0391)	(0.0044)	(0.0256)
LLWS	0.0630	0.1865	0.0375	0.1435	0.0191	0.1022	0.0099	0.0737	0.0041	0.0471
	(0.0621)	(0.1168)	(0.0399)	(0.1435)	(0.0205)	(0.1022)	(0.0109)	(0.0737)	(0.0044)	(0.0471)
LLGMM	0.0630	0.1865	0.0388	0.1460	0.0198	0.1038	0.0100	0.0740	0.0042	0.0474
	(0.0627)	(0.1865)	(0.0408)	(0.1460)	(0.0208)	(0.1038)	(0.0109)	(0.0740)	(0.0046)	(0.0474)
					Case: S.1					
LLE	0.1513	0.2909	0.0939	0.2271	0.0516	0.1679	0.0261	0.1189	0.0106	0.0766
	(0.1509)	(0.1528)	(0.1019)	(0.1225)	(0.0541)	(0.0906)	(0.0288)	(0.0647)	(0.0109)	(0.0402)
LLWS	0.1366	0.2754	0.0816	0.2123	0.0443	0.1556	0.0227	0.1109	0.0091	0.0708
	(0.1223)	(0.2033)	(0.0848)	(0.2123)	(0.0461)	(0.1556)	(0.0249)	(0.1109)	(0.0094)	(0.0708)
LLGMM	0.1187	0.2579	0.0585	0.1820	0.0292	0.1262	0.0135	0.0867	0.0050	0.0517
	(0.1107)	(0.2579)	(0.0574)	(0.1820)	(0.0308)	(0.1262)	(0.0143)	(0.0867)	(0.0053)	(0.0517)
					Case: S.2					
LLE	0.2026	0.3407	0.1381	0.2810	0.0812	0.2169	0.0468	0.1632	0.0209	0.1094
	(0.1854)	(0.1732)	(0.1308)	(0.1423)	(0.0727)	(0.1047)	(0.0420)	(0.0787)	(0.0174)	(0.0516)
LLWS	0.3817	0.3520	0.0804	0.2105	0.0372	0.1409	0.0164	0.0902	0.0048	0.0486
	(0.1551)	(0.2373)	(0.0792)	(0.2105)	(0.0397)	(0.1409)	(0.0202)	(0.0902)	(0.0059)	(0.0486)
LLGMM	0.1427	0.2812	0.0557	0.1462	0.0134	0.0817	0.0045	0.0471	0.0015	0.0262
	(0.1407)	(0.2812)	(0.3642)	(0.1462)	(0.0169)	(0.0817)	(0.0061)	(0.0471)	(0.0031)	(0.0262)
					Case: S.3					
LLE	0.2569	0.3996	0.1762	0.3330	0.1018	0.2538	0.0581	0.1913	0.0265	0.1291
	(0.1933)	(0.1679)	(0.1240)	(0.1299)	(0.0688)	(0.0950)	(0.0365)	(0.0644)	(0.0163)	(0.0458)
LLWS	0.0781	0.1763	0.0328	0.1069	0.0126	0.0619	0.0055	0.0376	0.0023	0.0243
	(0.0738)	(0.1860)	(0.0506)	(0.1069)	(0.0211)	(0.0619)	(0.0100)	(0.0376)	(0.0039)	(0.0243)
LLGMM	0.0746	0.1275	0.1067	0.0600	0.0021	0.0201	0.0004	0.0093	0.0003	0.0064
	(0.1178)	(0.1275)	(0.2031)	(0.0600)	(0.0125)	(0.0201)	(0.0019)	(0.0093)	(0.0018)	(0.0064)
					Case: S.4					
LLE	0.0971	0.2330	0.0588	0.1798	0.0309	0.1298	0.0158	0.0928	0.0065	0.0596
	(0.0962)	(0.1198)	(0.0633)	(0.0953)	(0.0332)	(0.0694)	(0.0176)	(0.0500)	(0.0068)	(0.0317)
LLWS	0.0958	0.2306	0.0577	0.1782	0.0303	0.1285	0.0155	0.0920	0.0063	0.0589
	(0.0951)	(0.2231)	(0.0614)	(0.1782)	(0.0327)	(0.1285)	(0.0172)	(0.0920)	(0.0067)	(0.0589)
LLGMM	0.0958	0.2306	0.0576	0.1792	0.0303	0.1287	0.0153	0.0914	0.0063	0.0585
	(0.0960)	(0.2306)	(0.0584)	(0.1792)	(0.0319)	(0.1287)	(0.0170)	(0.0914)	(0.0066)	(0.0585)

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

(TR) = 12ms, echo time (TE) = 5.2ms, flip angle = 13°, inversion time = 450 ms, matrix = 384×256, voxel size = $1.2 \times 0.57 \times 0.69$mm and scan time = 2 min 54 sec. DTI are obtained in the axial plane using a spin-echo echo planner imaging sequence with TR = 4500ms, TE = 89.4ms, field of view = $20 \times 20\text{cm}^2$, matrix size = 160×132, slice thickness = 3mm, slice spacing = 1mm, b-values = 0, 1000 s/mm^2.

FA varies systematically along the trajectory of each white matter fascicle. Several pre- and post-processing steps were performed by the FSL software. The brain was extracted using brain segmentation tools. After generating FA maps using FMRIB diffusion toolbox, images from all individuals were aligned to an FA standard template through non-linear co-registration. The Johns Hopkins University (JHU) white matter tractography atlas was used as a standard template for white matter parcellation with 50 regions of interest (ROIs). All imaging parameters were calculated by averaging the voxel values in each ROI. See Xiong et al. (2017) for more details on the data and the preprocessing steps.

For each subject, we calculate the similarity matrix $\mathbf{C}$ with dimension 50×50. The (k, l)-th element of the matrix $\mathbf{C}$ is defined as $c_{kl} = |y_k - y_l|$ where y_k is the measure of FA associated with k-th ROI. For simplicity, we scale the similarity matrix such that the range of the elements of the matrix is $[0, 1]$. To

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

create the network, we threshold each similarity matrix to build an adjacency matrix $\mathbf{G}$ with elements $\{1, 0\}$ depending on whether the similarity values exceed the threshold or not. Since this threshold controls the topology of the data, we contract the adjacent matrix over a set of threshold parameters from 0.01 to 0.99, and this set is denoted as $\mathcal{S}$ with cardinality 99. The role of the threshold is to investigate the graph networks formed by the ROIs with different degrees of anisotropy. When the threshold value is small, most ROIs are connected with edges in the graph. In this case, the focus is more on global and entire brain regions. When the threshold value is high, there will be fewer ROIs connected with edges. In this case, the focus is more on a few brain regions whose degrees of anisotropy are more dissimilar from each other. In summary, the threshold allows us to tune our focus on different collections of connected ROIs, and it allows one to zoom in and zoom out.

A popular measure of connectivity in a given graph is average path length (APL) which is defined as the average number of steps along the shortest path for all possible pairs of the network nodes. Therefore it measures the efficacy of the information on a network (Albert and Barabási, 2002). For a series of threshold parameters (s), we observe APL for FA as shown in Figure 1. Scientists are often interested to know the association of APL of the network generated from FA with a set of covariates of interests such as age and lapses

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

score.

We fit the model (1.1) to APL that is collected over continuous spatial domains (viz, thresholds) from all individuals in which $\mathbf{X}_i$ included the clinical variables such as lapses, and age. We discard the subjects from the analysis with missing clinical variables and therefore the sample size $n = 27$. Here we used three-fold cross-validation to obtain the tuning parameters and the fraction of variance explained (FVE) is set at 0.99. In Figure 2, we present the estimated coefficient functions corresponding to age and number of lapses associated with APL where it can be observed that the coefficient of network property is negative with age but positive with lapses counts. Moreover, the effect of the APL is found to be increasing when the threshold is small to moderate and decreasing at moderate to large threshold; whereas, the effect of APL is more-or-less similar up to the larger values of threshold, and after that, it significantly decreases.

In Xiong et al. (2017), the authors firstly divided patients into a nonsleepy group (lapses ≤ 5) and a sleepy group (lapses > 5), and then compared FA values between the sleepy and nonsleepy group using two sample tests for all the ROIs. The authors found that "the alterations in FA of individual fiber tracts occurred mainly in the internal/external capsule, corona radiata, corpus callosum, and sagittal stratum regions". Our finding in this paper is

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

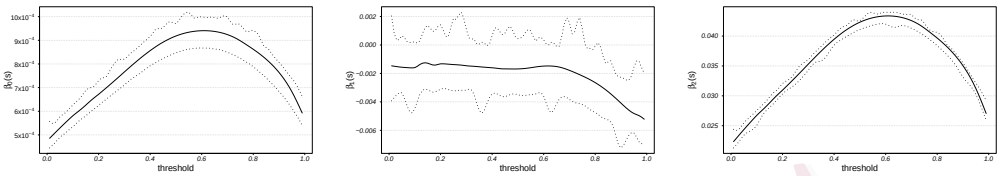

Figure 2: *Apnea-data* analysis: Plots of estimated intercept (left), coefficient functions of age (middle) and number of lapses (right) for average path length associated with Fractional Anisotropy (FA) in DTI analysis. The solid lines are the estimates of the functions, and the dotted lines are point-wise 95% confidence intervals.

consistent with that in Xiong et al. (2017). First, the coefficient function for the number of lapses is significant for all the threshold values. This indicates the association between the number of lapses and FA values. Second, we observe that the coefficient function for the number of lapses achieves its maximum when the threshold is around 0.65. The brain regions contributing most to the APL for the brain network when the threshold is 0.65, including the corpus callosum, cerebral peduncle, internal/external capsule, corona radiata, cingulum hippocampus, and tapetum, have the largest correlation with the number of lapses. These brain regions found by the proposed method include those found in Xiong et al. (2017).

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

7. Discussion

In this article, we propose an efficient estimation procedure for varying-coefficient model which is commonly used in neuroimaging and econometrics. Our procedure stands out for its efficiency in handling the integration of heteroskedasticity within the realm of functional data analysis. To the best of our knowledge, this is the first attempt to incorporate such conditions into the model. Such a model is therefore equipped with a more complex relationship between the functional response and real-valued covariates. Additionally, our method is easy to implement in a wide range of applications due to the multi-stage structure of the algorithm. The applicability of the proposed method is illustrated by numerical studies.

Supplementary material

The online supplementary material contains the proposed algorithm, the extension of the proposed method for the multivariate functional domain, comments on assumptions, additional simulation results, proofs of the theorems presented in Section 4, and a discussion on the choice of IVs.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

Acknowledgments

The research was partially supported by an NIH grant R03NS128450 and an NSF grant FRG-2152070. The authors thank Editor Prof. John Stufken, the Associate Editor, and the two referees for their constructive comments, which significantly improved the paper.

References

Ai, C. (1997). A semiparametric maximum likelihood estimator. *Econometrica 65*(4), 933–963.

Ai, C. and X. Chen (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica 71*(6), 1795–1843.

Albert, R. and A.-L. Barabási (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics 74*(1), 47–97.

Amemiya, T. (1977). The maximum likelihood and the nonlinear three-stage least squares estimator in the general nonlinear simultaneous equation model. *Econometrica 45*(4), 955–968.

Bravo, F. (2021). Second order expansions of estimators in nonparametric moment conditions models with weakly dependent data. *Econometric Reviews*, 1–24.

Cai, Z., M. Das, H. Xiong, and X. Wu (2006). Functional coefficient instrumental variables models. *Journal of Econometrics 133*(1), 207–241.

Cai, Z. and Q. Li (2008). Nonparametric estimation of varying coefficient dynamic panel data

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

models. *Econometric Theory 24*(5), 1321–1342.

Chen, X. and D. Pouzo (2009). Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals. *Journal of Econometrics 152*(1), 46–60. Recent Adavances in Nonparametric and Semiparametric Econometrics: A Volume Honouring Peter M. Robinson.

Chiou, J.-M., H.-G. Müller, and J.-L. Wang (2003). Functional quasi-likelihood regression models with smooth random effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65*(2), 405–423.

Chiou, J.-M., H.-G. Müller, and J.-L. Wang (2004). Functional response models. *Statistica Sinica*, 675–693.

Ding, X., D. Yu, Z. Zhang, and D. Kong (2021). Multivariate functional response low-rank regression with an application to brain imaging data. *Canadian Journal of Statistics 49*(1), 150–181.

Fan, J. and I. Gijbels (1996). *Local polynomial modelling and its applications*. Chapman & Hall/CRC.

Fan, J., Q. Yao, and Z. Cai (2003). Adaptive varying-coefficient linear models. *Journal of the Royal Statistical Society: series B (statistical methodology) 65*(1), 57–80.

Fan, J. and W. Zhang (1999). Statistical estimation in varying coefficient models. *The annals of Statistics 27*(5), 1491–1518.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

Fan, J., W. Zhang, and J.-T. Zhang (2008). Statistical methods with varying coefficient models. *Statistics and Its Interface 1*(1), 179–195.

Gajardo, A., C. Carroll, Y. Chen, X. Dai, J. Fan, P. Z. Hadjipantelis, K. Han, H. Ji, H.-G. Mueller, and J.-L. Wang (2021). *fdapace: Functional Data Analysis and Empirical Dynamics*. R package version 0.5.7.

Ghosh, T., Y. Ma, R. Song, and P. Zhong (2023). Flexible inference of optimal individualized treatment strategy in covariate adjusted randomization with multiple covariates. *Electronic Journal of Statistics 17*(1), 1344 – 1370.

Hall, A. R. (2004). *Generalized method of moments*. OUP Oxford.

Hall, P. and M. Hosseini-Nasab (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68*(1), 109–126.

Hastie, T. and R. Tibshirani (1993). Varying-coefficient models. *Journal of the Royal Statistical Society: Series B (Methodological) 55*(4), 757–779.

J Mercer, B. (1909). Xvi. functions of positive and negative type, and their connection the theory of integral equations. *Phil. Trans. R. Soc. Lond. A 209*(441-458), 415–446.

Jiang, C.-R. and J.-L. Wang (2011). Functional single index models for longitudinal data.

Li, Y. (2011). Efficient semiparametric regression for longitudinal data with nonparametric covariance estimation. *Biometrika*, 355–370.

Li, Y. and T. Hsing (2010). Uniform convergence rates for nonparametric regression and principal

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

component analysis in functional/longitudinal data. *The Annals of Statistics 38*(6), 3321–3351.

Lin, X. and R. J. Carroll (2001). Semiparametric regression for clustered data using generalized estimating equations. *Journal of the American Statistical Association 96*(455), 1045–1056.

Loève, M. (1946). Functions aleatoire de second ordre. *Revue science 84*, 195–206.

Lu, C. and J. M. Wooldridge (2020). A gmm estimator asymptotically more efficient than ols and wls in the presence of heteroskedasticity of unknown form. *Applied Economics Letters 27*(12), 997–1001.

Ma, Y., J.-M. Chiou, and N. Wang (2006). Efficient semiparametric estimator for heteroscedastic partially linear models. *Biometrika 93*(1), 75–84.

Müller, H.-G. and F. Yao (2010). Empirical dynamics for longitudinal data. *The Annals of Statistics 38*(6), 3458–3486.

Newey, W. K. (1990). Efficient instrumental variables estimation of nonlinear models. *Econometrica: Journal of the Econometric Society*, 809–837.

Newey, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica 62*(6), 1349–1382.

Qu, A. and R. Li (2006). Quadratic inference functions for varying-coefficient models with longitudinal data. *Biometrics 62*(2), 379–391.

Ramsay, J. O. and B. W. Silverman (2005). *Functional data analysis*. Springer series in statistics.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

Scheipl, F., A.-M. Staicu, and S. Greven (2015). Functional additive mixed models. *Journal of Computational and Graphical Statistics 24*(2), 477–501.

Su, L., I. Murtazashvili, and A. Ullah (2013). Local linear gmm estimation of functional coefficient iv models with an application to estimating the rate of return to schooling. *Journal of Business & Economic Statistics 31*(2), 184–207.

Sun, Y. (2016). Functional-coefficient spatial autoregressive models with nonparametric spatial weights. *Journal of Econometrics 195*(1), 134–153.

Tran, K. C. and E. G. Tsionas (2009). Local gmm estimation of semiparametric panel data with smooth coefficient models. *Econometric Reviews 29*(1), 39–61.

Wang, H., P.-S. Zhong, Y. Cui, and Y. Li (2017, 09). Unified Empirical Likelihood Ratio Tests for Functional Concurrent Linear Models and the Phase Transition from Sparse to Dense Functional Data. *Journal of the Royal Statistical Society Series B: Statistical Methodology 80*(2), 343–364.

Wang, L. (2008). *Karhunen-Loéve expansions and their applications*. London School of Economics and Political Science (United Kingdom).

Wang, L., H. Li, and J. Z. Huang (2008). Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. *Journal of the American Statistical Association 103*(484), 1556–1569.

Wang, N. (2003). Marginal nonparametric kernel regression accounting for within-subject corre-

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

lation. *Biometrika 90*(1), 43–52.

Wang, N., R. J. Carroll, and X. Lin (2005). Efficient semiparametric marginal estimation for longitudinal/clustered data. *Journal of the American Statistical Association 100*(469), 147–157.

Wasserman, L. (2006). *All of nonparametric statistics*. Springer Science & Business Media.

Wei, H. and Y. Sun (2017). Heteroskedasticity-robust semi-parametric gmm estimation of a spatial model with space-varying coefficients. *Spatial Economic Analysis 12*(1), 113–128.

Wu, C. O. and C.-T. Chiang (2000). Kernel smoothing on varying coefficient models with longitudinal dependent variable. *Statistica Sinica*, 433–456.

Xiong, Y., X. J. Zhou, R. A. Nisi, K. R. Martin, M. M. Karaman, K. Cai, and T. E. Weaver (2017). Brain white matter changes in cpap-treated obstructive sleep apnea patients with residual sleepiness. *Journal of Magnetic Resonance Imaging 45*(5), 1371–1378.

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

Yu, K. and M. Jones (2004). Likelihood-based local linear estimation of the conditional variance function. *Journal of the American Statistical Association 99*(465), 139–144.

Zhu, H., J. Fan, and L. Kong (2014). Spatially varying coefficient model for neuroimaging data with jump discontinuities. *Journal of the American Statistical Association 109*(507), 1084–1098.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0366

Zhu, H., R. Li, and L. Kong (2012). Multivariate varying coefficient model for functional responses. *The Annals of Statistics 40*(5), 2634–2666.

Department of Biostatistics, Johns Hopkins University, 615 North Wolfe Street, Baltimore, MD 21205, USA

E-mail: (pnyogi1@jhmi.edu)

Department of Mathematics, Statistics and Computer Science, University of Illinois at Chicago, 851 S. Morgan Street, Chicago, Illinois 60607, USA

E-mail: (pszhong@uic.edu)

Center for MR Research and Departments of Radiology, Neurosurgery, and Bioengineering, University of Illinois at Chicago, 1801 West Taylor St., Chicago, Illinois 60612, USA

E-mail: (xjzhou@uic.edu)