SUPPLEMENTARY MATERIAL FOR "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: The online supplementary material contains the proposed algorithm, the extension of the proposed method for multi-dimensional functional domain, additional simulation results, comments on assumptions, and theoretical proofs of the theorems presented in Section 4. In addition, a discussion on the choice of instrumental variables (IV) is provided.

S1 Algorithm

The outline of the proposed method in Section 3 of the main article is presented in the following Algorithm 1.

Algorithm 1 Estimation of $\boldsymbol{\beta}(s) : s \in \mathcal{S}$

Require: $\{Y_i(s_j), X_i, s_j\}$, for $j = 1, \cdots, r; i = 1, \cdots, n$

1: Calculate optimal h: $\widehat{h}_{init} \leftarrow \arg\min_{h\in\mathcal{H}} (nr)^{-1} \sum_{i=1}^{n} \sum_{r=1}^{r} \left\{ Y_{ij} - \mathbf{X}_i^{\mathrm{T}} \breve{\boldsymbol{\beta}}^{-i}(s_r; h) \right\}^2$.

2: Calculate $\breve{\boldsymbol{\gamma}}(s; \widehat{h}_{init})$.

3: Compute

$$\mathbf{g}_i\{\boldsymbol{\gamma}(s)\} = r^{-1} \sum_{j=1}^{r} K_{\widehat{h}_{init}}(s_j - s) \mathbf{Q}_{ij}(s; \widehat{h}_{init}) \{Y_{ij} - \mathbf{W}_{ij}(s; \widehat{h}_{init})^{\mathrm{T}} \breve{\boldsymbol{\gamma}}(s; \widehat{h}_{init})\}.$$

4: Determine the instrument variables $\mathfrak{M}(\mathbf{X})$.

5: Compute eigen-components $\widehat{\lambda}_k, \widehat{\boldsymbol{\phi}}_k(s)$ and get the value of α using the condition (C8).

6: Calculate optimal h: $\widehat{h}_{opt} \leftarrow \arg\min_{h\in\mathcal{H}} (nr)^{-1} \sum_{i=1}^{n} \sum_{r=1}^{r} \left\{ Y_{ij} - \mathbf{X}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}}^{-i}(s_r; h) \right\}^2$.

7: Calculate $\widehat{\boldsymbol{\beta}}(s; h_{opt})$.

S2 Outline of the proposed method in a multi-dimension spatial domain

In the manuscript, we focus on a functional varying coefficient model on a univariate spatial domain $\mathcal{S} \in [0, 1]$. In this section, we will extend the proposed method to a multi-dimensional spatial domain defined by $\mathcal{S} = [0, 1]^d = \left\{ \overrightarrow{\mathbf{s}} = (s_1, \cdots, s_d)^{\mathrm{T}} : s_k \in [0, 1] \ \forall\ k = 1, \cdots, d \right\}$. Let $\{Y_i(\overrightarrow{\mathbf{s}}), \mathbf{X}_i\}$ for $i = 1, \cdots, n$ be independent copies of $\{Y(\overrightarrow{\mathbf{s}}), \mathbf{X}\}$. For each curve, we observe $Y(\overrightarrow{\mathbf{s}})$ on the discrete spatial grid $\{\overrightarrow{\mathbf{s}}_1, \cdots, \overrightarrow{\mathbf{s}}_r\} \in [0,1]^d$ on the functional domain $\mathcal{S}$.

To obtain an initial estimation for the coefficient function vector $\boldsymbol{\beta}(\overrightarrow{\mathbf{s}}) = (\beta_1(\overrightarrow{\mathbf{s}}), \cdots, \beta_p(\overrightarrow{\mathbf{s}}))^{\mathrm{T}}$, we approximate $\beta_k(\overrightarrow{\mathbf{s}})$ at $\overrightarrow{\mathbf{s}}_0$ by the following the Taylor expansion of $\beta_k(\overrightarrow{\mathbf{s}})$ at $\overrightarrow{\mathbf{s}}_0$, for any $\overrightarrow{\mathbf{s}}_l$ in a neighborhood of $\overrightarrow{\mathbf{s}}$,

$$\beta_k(\overrightarrow{\mathbf{s}}_l) \approx \beta_k(\overrightarrow{\mathbf{s}}_0) + \sum_{a=1}^{d} \frac{\partial \beta_k(\overrightarrow{\mathbf{s}})}{\partial s_a}(s_{la} - s_{0a}) \quad \text{for } k = 1, \cdots, p,$$

where $\overrightarrow{\mathbf{s}}_l = (s_{l1}, \cdots, s_{lp})^{\mathrm{T}}$ and $\overrightarrow{\mathbf{s}}_0 = (s_{01}, \cdots, s_{0p})^{\mathrm{T}}$. Let $\mathbf{h}_1 = (h_{11}, \cdots, h_{1d})^{\mathrm{T}}$. In matrix notations, the above Taylor series expansion of the coefficient function becomes,

$$\boldsymbol{\beta}(\overrightarrow{\mathbf{s}}) \approx \mathbf{A}(\overrightarrow{\mathbf{s}})\mathbf{z}_{\mathbf{h}_1}(\overrightarrow{\mathbf{s}}_l - \overrightarrow{\mathbf{s}}) \tag{S2.1}$$

where

$$\mathbf{z}_{\mathbf{h}_1}(\overrightarrow{\mathbf{s}}_l - \overrightarrow{\mathbf{s}}) = \{1, (s_{l1} - s_{01})/h_{11}, \cdots, (s_{ld} - s_{0d})/h_{1d}\}^{\mathrm{T}}$$
$$\mathbf{A}(\overrightarrow{\mathbf{s}}_0) = \left[\boldsymbol{\beta}(\overrightarrow{\mathbf{s}}_0), h_{11}\dot{\boldsymbol{\beta}}_1(\overrightarrow{\mathbf{s}}_0), \cdots, h_{1d}\dot{\boldsymbol{\beta}}_d(\overrightarrow{\mathbf{s}}_0)\right].$$

Here $\mathbf{A}(\mathbf{s})$ is a $p \times (d+1)$ matrix and $\dot{\boldsymbol{\beta}}_a = (\frac{\partial \beta_1(\overrightarrow{\mathbf{s}})}{\partial s_a}, \cdots, \frac{\partial \beta_p(\overrightarrow{\mathbf{s}})}{\partial s_a})^{\mathrm{T}}$ denotes the partial gradient of $\boldsymbol{\beta}_a$ with respect to s_a. Define $\mathbf{W}_{ij}(\overrightarrow{\mathbf{s}}_0) = [\mathbf{z}_{h_1}(\overrightarrow{\mathbf{s}}_j - \overrightarrow{\mathbf{s}}_0) \otimes \mathbf{X}_i]$ and $\boldsymbol{\gamma}(\overrightarrow{\mathbf{s}}_0) = (\boldsymbol{\beta}(\overrightarrow{\mathbf{s}}_0)^{\mathrm{T}}, h_1\dot{\boldsymbol{\beta}}_1(s_0)^{\mathrm{T}}, \cdots, h_d\dot{\boldsymbol{\beta}}_d(s_0)^{\mathrm{T}})^{\mathrm{T}}$, both are vectors of length $p(d+1)$. The functional varying coefficient model can be written as

$$Y_{ij} \approx \mathbf{W}_{ij}(s_0)^{\mathrm{T}}\boldsymbol{\gamma}(s_0) + U_{ir}, \tag{S2.2}$$

such that $\overrightarrow{\mathbf{s}}_j$ are sufficiently close to $\overrightarrow{\mathbf{s}}_0$.

Define $K_{\mathbf{h}_1}(\overrightarrow{\mathbf{s}}_j - \overrightarrow{\mathbf{s}}_0) = \prod_{k=1}^{d} K_{h_{1k}}(s_{jk} - s_{0k})$. Based on the local linear smoother, we obtain the initial estimates $\tilde{\boldsymbol{\beta}}(\overrightarrow{\mathbf{s}}_0) = [(1, \mathbf{0}_d^{\mathrm{T}}) \otimes \mathbf{I}_p]\tilde{\boldsymbol{\gamma}}(\overrightarrow{\mathbf{s}}_0)$ where $\mathbf{0}_d$ is a column vector with all the d element 0 and

$$\breve{\boldsymbol{\gamma}}(\overrightarrow{\mathbf{s}}_0) = \left\{(nr)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{r} K_{\mathbf{h}_1}(\overrightarrow{\mathbf{s}}_j - \overrightarrow{\mathbf{s}}_0)\mathbf{W}_{ij}(\overrightarrow{\mathbf{s}}_0)\mathbf{W}_{ij}(\overrightarrow{\mathbf{s}}_0)^{\mathrm{T}}\right\}^{-1} \times \left\{(nr)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{r} K_{\mathbf{h}_1}(\overrightarrow{\mathbf{s}}_j - \overrightarrow{\mathbf{s}}_0)\mathbf{W}_{ij}(\overrightarrow{\mathbf{s}}_0)Y_{ij}\right\}. \quad \text{(S2.3)}$$

Built on the discussion in Section 3.2, we can employ the same non-parametric regression framework as presented in Equation (3.8)

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

where $R_i = \int U_i^2(\overrightarrow{\mathbf{s}})d\overrightarrow{\mathbf{s}}$. Similar to the univariate spatial domain, we can replace $U_i(s)$ in R_i by $\breve{U}_i(\overrightarrow{\mathbf{s}}) = Y_i(\overrightarrow{\mathbf{s}}) - \mathbf{X}_i^{\mathrm{T}}\breve{\boldsymbol{\beta}}(\overrightarrow{\mathbf{s}})$. Then, using a similar approach in Section 3.2 of the main text, we can obtain an estimate of $\sigma^2(\mathbf{X}_i)$ as $\widehat{\sigma}^2(\mathbf{X}_i)$.

Similar to main paper, assume that the instrument variable with dimension $q \geq p$ is

$$\mathbf{Q}_{ij}(\overrightarrow{\mathbf{s}}_0) = (\mathfrak{M}(\mathbf{X}_i), \mathfrak{M}(\mathbf{X}_i)(\overrightarrow{\mathbf{s}}_j - \overrightarrow{\mathbf{s}}_0)/h)^{\mathrm{T}},$$

where $\mathfrak{M}(\mathbf{X}_i) = (\mathbf{X}_i, \mathbf{X}_i/\widehat{\sigma}^2(\mathbf{X}_i))^{\mathrm{T}}$. Then we can define

$$\begin{aligned} \mathbf{g}_i\{\boldsymbol{\gamma}(\overrightarrow{\mathbf{s}}_0)\} &= r^{-1}\sum_{j=1}^{r} K_{\mathbf{h}}(\overrightarrow{\mathbf{s}}_j - \overrightarrow{\mathbf{s}}_0)\mathbf{Q}_{ij}(\overrightarrow{\mathbf{s}}_0)\left\{Y_{ij} - \mathbf{W}_{ij}(\overrightarrow{\mathbf{s}}_0)^{\mathrm{T}}\boldsymbol{\gamma}(\overrightarrow{\mathbf{s}}_0)\right\} \\ &= r^{-1}\sum_{j=1}^{r} K_{\mathbf{h}}(\overrightarrow{\mathbf{s}}_j - \overrightarrow{\mathbf{s}}_0)\mathbf{z}_h(\overrightarrow{\mathbf{s}}_j - \overrightarrow{\mathbf{s}}_0) \otimes \boldsymbol{\Delta}_{ij}(\overrightarrow{\mathbf{s}}_0) \qquad \text{(S2.4)} \end{aligned}$$

Note that the above $\mathbf{g}_i\{\boldsymbol{\gamma}(\overrightarrow{\mathbf{s}})\} = (g_{i1}\{\boldsymbol{\gamma}(\overrightarrow{\mathbf{s}})\}, \cdots, g_{i(2q)}\{\boldsymbol{\gamma}(\overrightarrow{\mathbf{s}})\})^{\mathrm{T}}$ is a $2q$-dimension vector of mean zero functions.

The next step is to perform eigen-function decomposition of $\mathbf{C}(\overrightarrow{\mathbf{s}}, \overrightarrow{\mathbf{s}}') = \mathbb{E}\{\mathbf{g}\{\boldsymbol{\gamma}(\overrightarrow{\mathbf{s}})\}\mathbf{g}\{\boldsymbol{\gamma}(\overrightarrow{\mathbf{s}}')\}^{\mathrm{T}}\}$. Unlike the univariate spatial domain, for every $i \in \{1, \cdots, n\}$ and a set of chosen evenly distributed grid points $\overrightarrow{\mathbf{s}}_1, \cdots, \overrightarrow{\mathbf{s}}_M \in [0,1]^d$, the M-dimension vector $g_{ij}\{\boldsymbol{\gamma}(\overrightarrow{\mathbf{s}}_1)\}, \cdots, g_{ij}\{\boldsymbol{\gamma}(\overrightarrow{\mathbf{s}}_M)\}$ (for $j = 1, \cdots, 2q$) cannot be ordered in a line naturally because $\overrightarrow{\mathbf{s}}_1, \cdots, \overrightarrow{\mathbf{s}}_M$ are in a d-dimension spatial domain. So the method for univariate spatial domain can not be applied directly. To overcome the difficulty, we will first apply the stringing method proposed in Chen et al. (2011) to transform the M-dimension vector $g_{ij}\{\boldsymbol{\gamma}(\overrightarrow{\mathbf{s}}_1)\}, \cdots, g_{ij}\{\boldsymbol{\gamma}(\overrightarrow{\mathbf{s}}_M)\}$ on d-dimension spatial domain to functional data on 1-dimension spatial domain so that it becomes a functional data $\{e^*_{ij}(\xi_1), \cdots, e^*_{ij}(\xi_{M^d})\}$ on a univariate domain in $\xi \in [0,1]$. Then, we applied the fast algorithm in Xiao et al. (2014) procedure to functional data $\mathbf{e}^*_i(\xi) = \{e^*_{i1}(\xi), \cdots, e^*_{i(2q)}(\xi)\}$ on a univariate spatial domain $\xi \in [0,1]$ to obtain the corresponding eigenvalues $\widehat{\lambda}_k$ and its eigenfunctions $\widehat{\boldsymbol{\phi}}_k(\xi)$. Lastly, one can convert them back to the original space to obtain eigenvalues $\widehat{\lambda}_k$ and eigenfunctions $\widehat{\boldsymbol{\phi}}_k(\overrightarrow{\mathbf{s}})$.

Finally, we will incorporate the covariance function into our estimation. For any positive α and given spatial location $\overrightarrow{\mathbf{s}}_0 \in [0,1]^d$, an estimate of $\boldsymbol{\gamma}(\overrightarrow{\mathbf{s}}_0)$ is given by minimizing the following objective function:

$$\mathcal{J}\{\boldsymbol{\gamma}(\overrightarrow{\mathbf{s}}_0)\} = \sum_{k=1}^{\infty} \frac{\widehat{\lambda}_k}{\widehat{\lambda}_k^2 + \alpha} \left\{ \overline{\mathbf{g}}(\boldsymbol{\gamma}(\overrightarrow{\mathbf{s}}_0))^{\mathrm{T}} \widehat{\boldsymbol{\phi}}_k(\overrightarrow{\mathbf{s}}_0) \right\}^2 .$$

By minimizing the above objective function, we obtain

$$\begin{aligned}
&\sum_{k=1}^{\infty} \frac{\widehat{\lambda}_k}{\widehat{\lambda}_k^2 + \alpha} \left\{ (nr)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{r} K_{\mathbf{h}}(\overrightarrow{\mathbf{s}}_j - \overrightarrow{\mathbf{s}}_0) \widehat{\boldsymbol{\phi}}_k(\overrightarrow{\mathbf{s}}_0)^{\mathrm{T}} \mathbf{Q}_{ij}(\overrightarrow{\mathbf{s}}_0) \mathbf{W}_{ij}(\overrightarrow{\mathbf{s}}_0) \right\} \\
&\times \left\{ (nr)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{r} K_{\mathbf{h}}(\overrightarrow{\mathbf{s}}_j - \overrightarrow{\mathbf{s}}_0) \widehat{\boldsymbol{\phi}}_k(\overrightarrow{\mathbf{s}}_0)^{\mathrm{T}} \mathbf{Q}_{ij}(\overrightarrow{\mathbf{s}}_0) \left[Y_{ij} - \mathbf{W}_{ij}(\mathbf{s}_0)^{\mathrm{T}} \boldsymbol{\gamma}(\mathbf{s}_0) \right] \right\} \\
&:= \sum_{k=1}^{\infty} \frac{\widehat{\lambda}_k}{\widehat{\lambda}_k^2 + \alpha} \mathcal{X}_k(\overrightarrow{\mathbf{s}}_0) \left\{ \mathcal{Y}_k(\overrightarrow{\mathbf{s}}_0) - \mathcal{X}_k(\overrightarrow{\mathbf{s}}_0)^{\mathrm{T}} \boldsymbol{\gamma}(\overrightarrow{\mathbf{s}}_0) \right\}, \qquad \text{(S2.5)}
\end{aligned}$$

where $\mathcal{X}_k(\overrightarrow{\mathbf{s}}_0) = (nr)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{r} K_h(\overrightarrow{\mathbf{s}}_j - \overrightarrow{\mathbf{s}}_0) \mathbf{W}_{ij}(\overrightarrow{\mathbf{s}}_0) \mathbf{Q}_{ij}(\overrightarrow{\mathbf{s}}_0)^{\mathrm{T}} \widehat{\boldsymbol{\phi}}_k(\overrightarrow{\mathbf{s}}_0)$ and $\mathcal{Y}_k(\overrightarrow{\mathbf{s}}_0) = (nr)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{r} K_h(\overrightarrow{\mathbf{s}}_j - \overrightarrow{\mathbf{s}}_0) \widehat{\boldsymbol{\phi}}_k(\overrightarrow{\mathbf{s}}_0)^{\mathrm{T}} \mathbf{Q}_{ij}(\overrightarrow{\mathbf{s}}_0) Y_{ij}$. Therefore, the final estimate of the coefficient function is $\widehat{\boldsymbol{\beta}}(\overrightarrow{\mathbf{s}}_0) = [(1, \mathbf{0}_d^{\mathrm{T}}) \otimes \mathbf{I}_p] \widehat{\boldsymbol{\gamma}}(\mathbf{s}_0)$ where $\mathbf{0}_d$ is a column vector with all the d element 0 and

$$\widehat{\boldsymbol{\gamma}}(\overrightarrow{\mathbf{s}}_0) = \left\{ \sum_{k=1}^{\infty} \frac{\widehat{\lambda}_k}{\widehat{\lambda}_k^2 + \alpha} \mathcal{X}_k(\overrightarrow{\mathbf{s}}_0) \mathcal{X}_k(\overrightarrow{\mathbf{s}}_0)^{\mathrm{T}} \right\}^{-1} \left\{ \sum_{k=1}^{\infty} \frac{\widehat{\lambda}_k}{\widehat{\lambda}_k^2 + \alpha} \mathcal{X}_k(\overrightarrow{\mathbf{s}}_0) \mathcal{Y}_k(\overrightarrow{\mathbf{s}}_0) \right\}.$$

S3 Some remarks on conditions

Remark S1. Condition (C4) is also common in functional data analysis literature (Wang et al., 2016). This condition allows us to perform Taylor series expansion. Condition in (C5) avoids the smoothness condition of the sample path (Zhu et al., 2012, 2014) which is commonly assumed in Hall and Hosseini-Nasab (2006); Zhang and Chen (2007); Cardot et al. (2013). The smoothness of the coefficient functions may be checked by comparing nonparametric functions estimated by the wavelet (Amato et al., 2020; Antoniadis, 2007) approach and the proposed approach using the test procedure in Hardle and Mammen (1993). The wavelet method is popular for functions with a few discontinuities, sharp spikes, and abrupt changes (Amato et al. (2020)). To check the smoothness of covariance functions, one could conduct a hypothesis test to compare two covariance matrices on observed grid points $s_1, \cdots, s_r$. Due to the large number of repeated measurements r, conventional sample covariance is not consistent and hence can not be applied for testing. Similar to Chen et al. (2010), the test can be constructed based on an estimator of a Frobenius norm between $\mathbf{\Sigma}(s, s)$ and its smoothed version.

Remark S2. To speed up the computation for the eigenfunction decomposition for multivariate covariance function with a large number of repeated measurements, conditions (C6) impose the continuity in the mean-zero function, which is equivalent to checking the mean square continuity of the process after lining up (Hadinejad-Mahram et al., 2002). Here, (a) shows the limits from right and remains always right; therefore, it involves only one process. A similar, but opposite, phenomenon occurs in (b). Moreover, if the vector process $\mathbf{g}(s)$ is mean-square continuous then both approaches are equivalent, as a result, the covariance function is continuous after lining up the process.

This assumption facilitates the computation of the proposed method so that the standard packages in functional data analysis can be immediately applied. However, this assumption could be removed if we use fast algorithms for eigenfunction decomposition with a large number of repeated measurements (e.g. Xiao et al. (2014); Zhong (2023)).

S4 Additional simulation results

S4.1 Additional simulation results with single covariates

Based on the simulation setup described in Section 5 of the main manuscript, we have presented in Table S1. This includes the comparison of the proposed method with the local linear estimator and Wei and Sun (2017)'s approaches for signal-to-noise ratio as unity. Similar to Table 1, we have seen that, the IMSE and IMAE are significantly reduced if we increase the sample size.

S4.2 Additional simulation results with multiple covariates

In this simulation, the data are generated from the model

$$Y_i(s) = X_{i1}\boldsymbol{\beta}_1(s) + X_{i2}\boldsymbol{\beta}_2(s) + X_{i3}\boldsymbol{\beta}_3(s) + U_i(s) \tag{S4.6}$$

where we generate trajectories that are observed at r spatial locations for i-th curve, $i = 1, \cdots, n$. Assume the functional fixed effect $\beta_1(s) = 1 + \cos(2\pi s)$, $\beta_2(s) = 2 + \sin(2\pi s)$ and $\beta_3(s) = 3 + s^4$. The corresponding fixed effect covariates X_1 and $(X_2, X_3)^{\mathrm{T}}$ generated from uniform distribution

Table S1: Comparison among the proposed LLGMM with the local linear estimator (LLE) and Wei and Sun (2017)'s approach (LLWS) for $\text{SNR}_\theta = 1$.

	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	0.0165	0.0962	0.0097	0.0729	0.0048	0.0515	0.0025	0.0372	0.0010	0.0237
	(0.0157)	(0.0472)	(0.0101)	(0.0380)	(0.0051)	(0.0273)	(0.0027)	(0.0198)	(0.0011)	(0.0129)
LLWS	0.0172	0.0980	0.0099	0.0738	0.0051	0.0526	0.0025	0.0373	0.0010	0.0238
	(0.0128)	(0.0094)	(0.0104)	(0.0738)	(0.0053)	(0.0526)	(0.0028)	(0.0373)	(0.0011)	(0.0238)
LLGMM	0.0172	0.0980	0.0101	0.0741	0.0052	0.0532	0.0025	0.0374	0.0010	0.0238
	(0.0164)	(0.0980)	(0.0109)	(0.0741)	(0.0054)	(0.0532)	(0.0028)	(0.0374)	(0.0011)	(0.0238)
Case: S.1										
LLE	0.0394	0.1487	0.0248	0.1169	0.0135	0.0860	0.0068	0.0608	0.0027	0.0387
	(0.0370)	(0.0743)	(0.0253)	(0.0598)	(0.0138)	(0.0455)	(0.0073)	(0.0329)	(0.0027)	(0.0204)
LLWS	0.0280	0.1238	0.0142	0.0887	0.0070	0.0617	0.0034	0.0430	0.0013	0.0269
	(0.0832)	(0.1255)	(0.0142)	(0.0887)	(0.0074)	(0.0617)	(0.0037)	(0.0430)	(0.0014)	(0.0269)
LLGMM	0.0270	0.1211	0.0126	0.0836	0.0062	0.0573	0.0029	0.0403	0.0012	0.0253
	(0.0265)	(0.1211)	(0.0126)	(0.0836)	(0.0066)	(0.0573)	(0.0031)	(0.0403)	(0.0013)	(0.0253)
Case: S.2										
LLE	0.0518	0.1726	0.0363	0.1440	0.0215	0.1117	0.0124	0.0842	0.0055	0.0560
	(0.0440)	(0.0820)	(0.0321)	(0.0687)	(0.0184)	(0.0522)	(0.0107)	(0.0399)	(0.0045)	(0.0264)
LLWS	0.0215	0.1069	0.0103	0.0734	0.0046	0.0480	0.0019	0.0304	0.0006	0.0172
	(0.0226)	(0.1590)	(0.0108)	(0.0734)	(0.0056)	(0.0480)	(0.0025)	(0.0304)	(0.0007)	(0.0172)
LLGMM	0.0165	0.0918	0.0069	0.0589	0.0029	0.0376	0.0012	0.0239	0.0004	0.0142
	(0.0186)	(0.0918)	(0.0083)	(0.0589)	(0.0038)	(0.0376)	(0.0020)	(0.0239)	(0.0005)	(0.0142)
Case: S.3										
LLE	0.0658	0.2020	0.0466	0.1705	0.0274	0.1314	0.0157	0.0994	0.0071	0.0669
	(0.0442)	(0.0771)	(0.0297)	(0.0616)	(0.0169)	(0.0461)	(0.0091)	(0.0319)	(0.0041)	(0.0232)
LLWS	0.0086	0.0547	0.0050	0.0398	0.0025	0.0273	0.0011	0.0168	0.0004	0.0101
	(0.0461)	(0.0601)	(0.0079)	(0.0398)	(0.0044)	(0.0273)	(0.0021)	(0.0168)	(0.0007)	(0.0101)
LLGMM	0.0021	0.0261	0.0006	0.0133	0.0003	0.0078	0.0002	0.0069	0.0001	0.0052
	(0.0055)	(0.0261)	(0.0014)	(0.0133)	(0.0013)	(0.0078)	(0.0006)	(0.0069)	(0.0003)	(0.0052)
Case: S.4										
LLE	0.0257	0.1200	0.0155	0.0924	0.0080	0.0662	0.0041	0.0472	0.0016	0.0300
	(0.0243)	(0.0593)	(0.0161)	(0.0475)	(0.0084)	(0.0349)	(0.0044)	(0.0253)	(0.0017)	(0.0160)
LLWS	0.0268	0.1213	0.0141	0.0881	0.0073	0.0631	0.0036	0.0447	0.0015	0.0285
	(0.0224)	(0.1231)	(0.0143)	(0.0881)	(0.0078)	(0.0631)	(0.0040)	(0.0447)	(0.0016)	(0.0285)
LLGMM	0.0254	0.1185	0.0142	0.0883	0.0074	0.0634	0.0037	0.0449	0.0015	0.0285
	(0.0245)	(0.1185)	(0.0142)	(0.0883)	(0.0078)	(0.0634)	(0.0039)	(0.0449)	(0.0016)	(0.0285)

on $[1,2]$ and bivariate normal distribution with mean $(1,2)^{\mathrm{T}}$ and variance $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, respectively. To check the impact of dependence on the proposed method, we choose $\rho \in \{0, 0.5\}$. The different choices of true conditional variance functions are described as follows.

T.0 $\sigma^2(x_1, x_2, x_3) = 1$ (homoskedastic)

T.1 $\sigma^2(x_1, x_2, x_3) = x_2^2 + x_3^2$

T.2 $\sigma^2(x_1, x_2, x_3) = x_2^2 \times x_3^2$

T.3 $\sigma^2(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2$

T.4 $\sigma^2(x_1, x_2, x_3) = x_1^2 \times x_2^2 \times x_3^2$

T.5 $\sigma^2(x_1, x_2, x_3) = 1 + |x_2| + x_3^2/2$

The definition of U_is remains the same as in Section 5. We choose the signal-to-noise ratio of 0.5. Based on 500 replications, we obtain IMSE, IMAE, and their standard deviation in Tables S2, S3, S4 for $\rho = 0$ and S5, S6, S7 for $\rho = 0.5$. We observe that, for all simulation situations, IMSE and IMAE of the proposed method are smaller than that of the local linear estimator and the method discussed in Wei and Sun (2017).

Table S2: Comparison among the proposed LLGMM with the local linear (LLE) and Wei and Sun (2017)'s estimators (LLWS) for $\beta_1(\cdot)$ with $\rho = 0$.

	n = 30		n = 50		n = 100		n = 200		n = 500	
Method	IMSE	IMAE	IMSE	IMAE	IMSE	IMAE	IMSE	IMAE	IMSE	IMAE
Case: T.0										
LLE	3.2568	1.3777	2.1728	1.1536	1.1894	0.8615	0.7139	0.6803	0.3897	0.5033
	(3.8318)	(0.7851)	(2.3065)	(0.5904)	(1.0737)	(0.3757)	(0.5511)	(0.2417)	(0.2525)	(0.1533)
LLWS	3.2272	1.3731	2.1738	1.1523	1.1804	0.8577	0.7093	0.6782	0.3878	0.5007
	(3.7593)	(0.7781)	(2.3238)	(0.5955)	(1.0617)	(0.3734)	(0.5494)	(0.2415)	(0.2544)	(0.1533)
LLGMM	2.9915	1.3574	2.0021	1.1306	1.1635	0.8682	0.7095	0.6762	0.3894	0.5017
	(3.6036)	(0.8246)	(2.1672)	(0.6111)	(1.0652)	(0.3843)	(0.5202)	(0.2257)	(0.2411)	(0.1434)
Case: T.1										
LLE	2.6659	1.2523	1.7693	1.0422	1.0754	0.8235	0.6567	0.6568	0.3930	0.5100
	(2.9965)	(0.6912)	(1.8162)	(0.5034)	(0.9595)	(0.3356)	(0.4589)	(0.2059)	(0.2118)	(0.1323)
LLWS	2.6047	1.2398	1.7710	1.0425	1.0686	0.8204	0.6560	0.6544	0.3973	0.5138
	(2.9044)	(0.6797)	(1.8124)	(0.5057)	(0.9652)	(0.3368)	(0.4692)	(0.2090)	(0.2108)	(0.1304)
LLGMM	2.3945	1.2079	1.6664	1.0235	1.0646	0.8203	0.6511	0.6488	0.3215	0.5079
	(2.8063)	(0.6940)	(1.7940)	(0.5143)	(0.9408)	(0.3407)	(0.4271)	(0.1872)	(0.2020)	(0.1289)
Case: T.2										
LLE	3.2029	1.3848	2.0038	1.0999	1.2863	0.8933	0.7828	0.7084	0.4607	0.5539
	(3.5392)	(0.7711)	(2.1383)	(0.5643)	(1.3877)	(0.4003)	(0.6149)	(0.2484)	(0.2607)	(0.1487)
LLWS	3.1422	1.3703	2.0269	1.1057	1.2776	0.8916	0.7822	0.7067	0.4630	0.5553
	(3.4934)	(0.7670)	(2.1689)	(0.5675)	(1.3508)	(0.3972)	(0.6288)	(0.2523)	(0.2553)	(0.1486)
LLGMM	2.0304	1.3647	1.7920	1.0462	1.2212	0.8442	0.7740	0.5648	0.1772	0.2466
	(2.0394)	(1.1751)	(2.0206)	(0.5583)	(1.7906)	(0.4287)	(4.2720)	(0.3577)	(0.6441)	(0.2205)
Case: T.3										
LLE	2.8732	1.2986	1.9363	1.0928	1.1326	0.8439	0.6834	0.6685	0.3980	0.5119
	(3.2852)	(0.7229)	(1.9551)	(0.5328)	(1.0146)	(0.3569)	(0.4968)	(0.2207)	(0.2278)	(0.1396)
LLWS	2.8318	1.2914	1.9370	1.0917	1.1278	0.8420	0.6825	0.6665	0.3984	0.5121
	(3.2071)	(0.7140)	(1.9604)	(0.5363)	(1.0216)	(0.3569)	(0.5059)	(0.2232)	(0.2302)	(0.1396)
LLGMM	2.5389	1.2471	1.7758	1.0585	1.1097	0.8452	0.6544	0.6910	0.3691	0.5008
	(3.0573)	(0.7282)	(1.9194)	(0.5519)	(1.0182)	(0.3669)	(0.4746)	(0.2052)	(0.2282)	(0.1444)
Case: T.4										
LLE	2.8747	1.3048	1.7945	1.0425	1.1675	0.8531	0.7128	0.6792	0.4244	0.5323
	(3.3002)	(0.7290)	(1.8842)	(0.5206)	(1.2541)	(0.3689)	(0.5366)	(0.2251)	(0.2256)	(0.1361)
LLWS	2.8387	1.2946	1.8209	1.0509	1.1658	0.8529	0.7099	0.6759	0.4257	0.5344
	(3.3117)	(0.7298)	(1.9036)	(0.5229)	(1.2509)	(0.3688)	(0.5450)	(0.2291)	(0.2211)	(0.1343)
LLGMM	2.6306	1.2161	1.6996	1.0124	1.0941	0.8075	0.5686	0.5456	0.1445	0.2421
	(4.3648)	(0.7294)	(1.9075)	(0.5220)	(1.3083)	(0.3978)	(0.7412)	(0.3043)	(0.2579)	(0.2091)
Case: T.5										
LLE	2.8335	1.2865	1.8836	1.0774	1.1054	0.8340	0.6714	0.6632	0.3898	0.5063
	(3.2835)	(0.7223)	(1.8885)	(0.5218)	(0.9825)	(0.3485)	(0.4825)	(0.2164)	(0.2228)	(0.1386)
LLGMM	2.7871	1.2783	1.8836	1.0763	1.1008	0.8317	0.6707	0.6615	0.3912	0.5072
	(3.2161)	(0.7137)	(1.8877)	(0.5250)	(0.9924)	(0.3490)	(0.4904)	(0.2184)	(0.2249)	(0.1381)
LLGMM	2.5307	1.2377	1.7261	1.0476	1.0885	0.8319	0.6619	0.6542	0.3291	0.5042
	(3.1371)	(0.7243)	(1.8136)	(0.5321)	(0.9762)	(0.3580)	(0.4642)	(0.2059)	(0.2141)	(0.1329)

Table S3: Comparison among the proposed LLGMM with the local linear (LLE) and Wei and Sun (2017)'s estimators (LLWS) for $\beta_2(\cdot)$ with $\rho = 0$.

	n = 30		n = 50		n = 100		n = 200		n = 500	
Method	IMSE	IMAE	IMSE	IMAE	IMSE	IMAE	IMSE	IMAE	IMSE	IMAE
Case: T.0										
LLE	1.5635	0.9661	0.9508	0.7626	0.5528	0.5889	0.3634	0.4887	0.2420	0.4098
	(1.7718)	(0.5549)	(0.9587)	(0.3780)	(0.4757)	(0.2383)	(0.2113)	(0.1304)	(0.0903)	(0.0649)
LLWS	1.5540	0.9640	0.9511	0.7630	0.5532	0.5880	0.3619	0.4880	0.2422	0.4102
	(1.7481)	(0.5530)	(0.9529)	(0.3775)	(0.4764)	(0.2389)	(0.2114)	(0.1297)	(0.0886)	(0.0633)
LLGMM	2.2415	1.0462	1.0377	0.8200	0.5784	0.6090	0.3349	0.4676	0.2144	0.3807
	(1.7499)	(0.7163)	(0.9056)	(0.3582)	(0.4715)	(0.2469)	(0.2236)	(0.1484)	(0.1006)	(0.0853)
Case: T.1										
LLE	1.7851	1.0324	1.1363	0.8286	0.6596	0.6402	0.4156	0.5200	0.2695	0.4289
	(2.0175)	(0.5972)	(1.1973)	(0.4346)	(0.5973)	(0.2816)	(0.2688)	(0.1605)	(0.1207)	(0.0834)
LLWS	1.7721	1.0285	1.1367	0.8285	0.6590	0.6395	0.4138	0.5190	0.2707	0.4303
	(2.0068)	(0.5942)	(1.2065)	(0.4361)	(0.5963)	(0.2815)	(0.2668)	(0.1592)	(0.1191)	(0.0831)
LLGMM	1.6199	1.0147	1.1238	0.8029	0.6201	0.6304	0.4014	0.5178	0.2388	0.3961
	(1.9936)	(0.5702)	(1.1356)	(0.4096)	(0.6133)	(0.2755)	(0.2816)	(0.1745)	(0.1369)	(0.1096)
Case: T.2										
LLE	2.2957	1.1815	1.5187	0.9548	0.8927	0.7432	0.5187	0.5766	0.3155	0.4600
	(2.5891)	(0.6742)	(1.7133)	(0.5301)	(0.8275)	(0.3519)	(0.3873)	(0.2140)	(0.1750)	(0.1143)
LLWS	2.2731	1.1737	1.5223	0.9543	0.8945	0.7440	0.5157	0.5748	0.3153	0.4604
	(2.5888)	(0.6738)	(1.7277)	(0.5298)	(0.8212)	(0.3516)	(0.3859)	(0.2141)	(0.1755)	(0.1148)
LLGMM	2.7597	1.2749	1.6352	1.0209	1.3148	0.8134	0.5747	0.5636	0.1945	0.3224
	(3.6677)	(0.7070)	(1.6155)	(0.4973)	(6.3750)	(0.4270)	(1.1856)	(0.2576)	(0.2241)	(0.1846)
Case: T.3										
LLE	1.7381	1.0186	1.0925	0.8137	0.6343	0.6281	0.4019	0.5121	0.2641	0.4253
	(1.9767)	(0.5890)	(1.1372)	(0.4213)	(0.5718)	(0.2729)	(0.2535)	(0.1531)	(0.1128)	(0.0785)
LLWS	1.7257	1.0149	1.0912	0.8134	0.6349	0.6278	0.4004	0.5110	0.2641	0.4256
	(1.9553)	(0.5847)	(1.1357)	(0.4212)	(0.5709)	(0.2728)	(0.2526)	(0.1522)	(0.1107)	(0.0779)
LLGMM	1.8121	1.0767	1.1705	0.8681	0.6859	0.6651	0.3872	0.5020	0.2303	0.3884
	(1.8376)	(0.5605)	(1.0582)	(0.3965)	(0.5585)	(0.2717)	(0.2689)	(0.1723)	(0.1281)	(0.1091)
Case: T.4										
LLE	2.2278	1.1600	1.4627	0.9377	0.8537	0.7259	0.4973	0.5655	0.3058	0.4529
	(2.5512)	(0.6650)	(1.6777)	(0.5198)	(0.8077)	(0.3449)	(0.3576)	(0.2008)	(0.1649)	(0.1084)
LLWS	2.2169	1.1566	1.4706	0.9389	0.8546	0.7266	0.4944	0.5642	0.3051	0.4539
	(2.5526)	(0.6649)	(1.6887)	(0.5194)	(0.7984)	(0.3427)	(0.3564)	(0.2009)	(0.1619)	(0.1063)
LLGMM	2.2150	1.1312	0.8492	0.0109	0.7889	0.7242	0.4557	0.5557	0.1915	0.3234
	(2.1676)	(0.6740)	(1.7830)	(0.4886)	(1.3157)	(0.3630)	(0.7496)	(0.2499)	(0.2053)	(0.1788)
Case: T.5										
LLE	1.6286	0.9876	1.0141	0.7833	0.5971	0.6111	0.3820	0.5003	0.2543	0.4185
	(1.8335)	(0.5654)	(1.0469)	(0.4011)	(0.5250)	(0.2561)	(0.2303)	(0.1414)	(0.1014)	(0.0718)
LLWS	1.6170	0.9843	1.0130	0.7832	0.5967	0.6101	0.3805	0.4994	0.2548	0.4194
	(1.8176)	(0.5622)	(1.0455)	(0.4010)	(0.5245)	(0.2559)	(0.2291)	(0.1403)	(0.0995)	(0.0713)
LLGMM	1.5404	0.9530	1.0002	0.7422	0.5874	0.6165	0.3689	0.4913	0.2210	0.3833
	(1.7433)	(0.5378)	(0.9761)	(0.3753)	(0.5155)	(0.2538)	(0.2442)	(0.1597)	(0.1133)	(0.0979)

Table S4: Comparison among the proposed LLGMM with the local linear (LLE) and Wei and Sun (2017)'s estimators (LLWS) for $\beta_3(\cdot)$ with $\rho = 0$.

	n = 30		n = 50		n = 100		n = 200		n = 500	
Method	IMSE	IMAE	IMSE	IMAE	IMSE	IMAE	IMSE	IMAE	IMSE	IMAE
Case: T.0										
LLE	1.2156	0.8453	0.7101	0.6318	0.3443	0.4523	0.1829	0.3219	0.0665	0.1954
	(1.6491)	(0.5471)	(0.9631)	(0.4298)	(0.4159)	(0.2796)	(0.2299)	(0.2096)	(0.0773)	(0.1154)
LLWS	1.2100	0.8442	0.7132	0.6326	0.3467	0.4534	0.1812	0.3202	0.0668	0.1959
	(1.6346)	(0.5435)	(0.9781)	(0.4327)	(0.4180)	(0.2795)	(0.2258)	(0.2066)	(0.0779)	(0.1159)
LLGMM	1.2128	0.8505	0.6865	0.6443	0.3608	0.4777	0.2078	0.3522	0.0885	0.2302
	(1.7499)	(0.7163)	(0.9056)	(0.3582)	(0.4715)	(0.2469)	(0.2236)	(0.1484)	(0.1006)	(0.0853)
Case: T.1										
LLE	1.5328	0.9466	0.8888	0.7080	0.4609	0.5207	0.2327	0.3639	0.0906	0.2284
	(2.0942)	(0.6205)	(1.2402)	(0.4815)	(0.5625)	(0.3289)	(0.3106)	(0.2364)	(0.1049)	(0.1365)
LLWS	1.5122	0.9410	0.8860	0.7069	0.4628	0.5216	0.2298	0.3616	0.0911	0.2295
	(2.0843)	(0.6183)	(1.2433)	(0.4822)	(0.5633)	(0.3293)	(0.3084)	(0.2336)	(0.1059)	(0.1369)
LLGMM	1.4442	0.9293	0.8817	0.7256	0.4611	0.5237	0.2237	0.3563	0.0821	0.2192
	(1.9936)	(0.5702)	(1.1356)	(0.4096)	(0.6133)	(0.2755)	(0.2816)	(0.1745)	(0.1369)	(0.1096)
Case: T.2										
LLE	1.4960	0.9446	0.8248	0.6864	0.4740	0.5181	0.2352	0.3605	0.0923	0.2292
	(1.7908)	(0.6054)	(1.1366)	(0.4589)	(0.6504)	(0.3507)	(0.3445)	(0.2452)	(0.1104)	(0.1371)
LLWS	1.4706	0.9351	0.8295	0.6884	0.4743	0.5189	0.2303	0.3565	0.0927	0.2301
	(1.7664)	(0.5993)	(1.1397)	(0.4600)	(0.6314)	(0.3483)	(0.3390)	(0.2415)	(0.1080)	(0.1364)
LLGMM	1.4194	0.9387	0.8176	0.6898	0.4477	0.5100	0.2265	0.3576	0.0939	0.2240
	(3.6677)	(0.7070)	(1.6155)	(0.4973)	(6.3750)	(0.4270)	(1.1856)	(0.2576)	(0.2241)	(0.1846)
Case: T.3										
LLE	1.4474	0.9208	0.8396	0.6883	0.4305	0.5031	0.2210	0.3550	0.0845	0.2206
	(1.9870)	(0.6003)	(1.1473)	(0.4683)	(0.5282)	(0.3183)	(0.2930)	(0.2296)	(0.0977)	(0.1316)
LLWS	1.4327	0.9172	0.8364	0.6868	0.4336	0.5044	0.2187	0.3531	0.0851	0.2215
	(1.9797)	(0.5991)	(1.1544)	(0.4697)	(0.5315)	(0.3192)	(0.2904)	(0.2268)	(0.0992)	(0.1322)
LLGMM	1.3549	0.9076	0.8076	0.6837	0.4337	0.5017	0.2103	0.3524	0.1075	0.2144
	(1.8376)	(0.5605)	(1.0582)	(0.3965)	(0.5585)	(0.2717)	(0.2689)	(0.1723)	(0.1281)	(0.1091)
Case: T.4										
LLE	1.3624	0.8970	0.7165	0.6409	0.4215	0.4887	0.2065	0.3364	0.0802	0.2126
	(1.6941)	(0.5814)	(0.9798)	(0.4261)	(0.5797)	(0.3288)	(0.3120)	(0.2306)	(0.0959)	(0.1274)
LLWS	1.3458	0.8887	0.7241	0.6441	0.4251	0.4903	0.2029	0.3330	0.0801	0.2128
	(1.6794)	(0.5794)	(0.9879)	(0.4280)	(0.5770)	(0.3287)	(0.3080)	(0.2284)	(0.0936)	(0.1264)
LLGMM	1.3221	0.8893	0.7254	0.6640	0.4199	0.4383	0.2080	0.3276	0.0805	0.2100
	(1.1676)	(0.6740)	(1.7830)	(0.4886)	(1.3157)	(0.3630)	(0.7496)	(0.2499)	(0.2053)	(0.1788)
Case: T.5										
LLE	1.4437	0.9181	0.8421	0.6886	0.4304	0.5041	0.2213	0.3559	0.0840	0.2203
	(1.9620)	(0.6026)	(1.1569)	(0.4696)	(0.5233)	(0.3160)	(0.2908)	(0.2292)	(0.0971)	(0.1308)
LLWS	1.4250	0.9135	0.8414	0.6880	0.4323	0.5048	0.2190	0.3539	0.0845	0.2210
	(1.9473)	(0.6006)	(1.1661)	(0.4715)	(0.5262)	(0.3169)	(0.2874)	(0.2260)	(0.0980)	(0.1313)
LLGMM	1.3465	0.9059	0.8026	0.6921	0.4370	0.5044	0.2197	0.3534	0.0860	0.2234
	(1.7433)	(0.5378)	(0.9761)	(0.3753)	(0.5155)	(0.2538)	(0.2442)	(0.1597)	(0.1133)	(0.0979)

Table S5: Comparison among the proposed LLGMM with the local linear (LLE) and Wei and Sun (2017)'s estimators (LLWS) for $\beta_1(\cdot)$ with $\rho = 0.5$

	n = 30		n = 50		n = 100		n = 200		n = 500	
Method	IMSE	IMAE	IMSE	IMAE	IMSE	IMAE	IMSE	IMAE	IMSE	IMAE
Case: T.0										
LLE	4.0246	1.5247	2.6216	1.2612	1.4204	0.9363	0.8433	0.7371	0.4629	0.5528
	(4.9122)	(0.9052)	(2.9427)	(0.6768)	(1.3627)	(0.4336)	(0.6828)	(0.2761)	(0.2973)	(0.1649)
LLWS	3.9945	1.5218	2.6262	1.2607	1.4116	0.9333	0.8394	0.7359	0.4608	0.5509
	(4.8446)	(0.8989)	(2.9628)	(0.6818)	(1.3366)	(0.4287)	(0.6809)	(0.2762)	(0.2967)	(0.1636)
LLGMM	3.5110	1.4698	2.3892	1.2275	1.3597	0.9358	0.8577	0.7527	0.4726	0.5601
	(4.5283)	(0.9100)	(2.7873)	(0.6987)	(1.3248)	(0.4440)	(0.6519)	(0.2681)	(0.2974)	(0.1753)
Case: T.1										
LLE	3.1961	1.3651	2.0250	1.1116	1.2346	0.8790	0.7430	0.6958	0.4547	0.5528
	(3.7864)	(0.7877)	(2.1379)	(0.5571)	(1.1571)	(0.3746)	(0.5339)	(0.2237)	(0.2337)	(0.1371)
LLWS	3.1227	1.3519	2.0374	1.1141	1.2178	0.8743	0.7443	0.6944	0.4589	0.5556
	(3.6681)	(0.7729)	(2.1405)	(0.5591)	(1.1468)	(0.3711)	(0.5508)	(0.2286)	(0.2351)	(0.1369)
LLGMM	2.6968	1.2834	1.3503	1.0721	1.1188	0.8659	0.7406	0.6908	0.4524	0.5457
	(3.3766)	(0.7542)	(2.1055)	(0.5701)	(1.1064)	(0.3743)	(0.5045)	(0.2103)	(0.2378)	(0.1451)
Case: T.2										
LLE	3.6110	1.4553	2.2306	1.1547	1.4132	0.9351	0.8637	0.7427	0.5159	0.5895
	(4.2514)	(0.8505)	(2.4709)	(0.6127)	(1.5061)	(0.4266)	(0.6772)	(0.2632)	(0.2649)	(0.1443)
LLWS	3.5239	1.4347	2.2264	1.1516	1.4034	0.9323	0.8631	0.7406	0.5259	0.5948
	(4.2236)	(0.8441)	(2.4801)	(0.6157)	(1.4922)	(0.4262)	(0.6898)	(0.2657)	(0.2857)	(0.1498)
LLGMM	2.9741	1.3315	2.2427	1.1195	1.3349	0.9111	0.8003	0.7105	0.4527	0.5103
	(3.8570)	(0.7977)	(4.9414)	(0.6636)	(1.3713)	(0.4105)	(0.6655)	(0.2702)	(0.3940)	(0.2311)
Case: T.3										
LLE	3.4697	1.4212	2.2654	1.1808	1.3171	0.9072	0.7842	0.7134	0.4631	0.5559
	(4.1178)	(0.8256)	(2.3604)	(0.5930)	(1.2270)	(0.3999)	(0.5918)	(0.2435)	(0.2574)	(0.1470)
LLWS	3.4199	1.4134	2.2759	1.1826	1.3060	0.9036	0.7843	0.7118	0.4654	0.5576
	(4.0162)	(0.8133)	(2.3668)	(0.5964)	(1.2214)	(0.3971)	(0.6033)	(0.2481)	(0.2604)	(0.1462)
LLGMM	3.0901	1.3661	2.0644	1.1429	1.2589	0.9031	0.7828	0.7126	0.4615	0.5515
	(3.9654)	(0.8356)	(2.2492)	(0.6111)	(1.1749)	(0.4036)	(0.5555)	(0.2328)	(0.2628)	(0.1605)
Case: T.4										
LLE	3.2962	1.3851	2.0131	1.0992	1.2988	0.8984	0.7966	0.7155	0.4796	0.5698
	(3.9847)	(0.8087)	(2.1997)	(0.5720)	(1.3619)	(0.3994)	(0.5973)	(0.2399)	(0.2249)	(0.1307)
LLWS	3.2194	1.3693	2.0182	1.0988	1.2880	0.8951	0.7870	0.7103	0.5172	0.5822
	(3.9197)	(0.8011)	(2.2185)	(0.5759)	(1.3443)	(0.3970)	(0.6039)	(0.2416)	(0.5771)	(0.1889)
LLGMM	2.7381	1.2654	1.9922	1.0898	1.2332	0.8774	0.7609	0.6962	0.4069	0.4914
	(3.5198)	(0.7577)	(2.3459)	(0.5875)	(1.2710)	(0.3877)	(0.5922)	(0.2339)	(0.3120)	(0.2181)
Case: T.5										
LLE	3.4186	1.4066	2.2018	1.1641	1.2788	0.8948	0.7688	0.7076	0.4536	0.5501
	(4.1108)	(0.8256)	(2.2746)	(0.5801)	(1.1756)	(0.3887)	(0.5693)	(0.2374)	(0.2515)	(0.1453)
LLWS	3.3813	1.4002	2.2097	1.1641	1.2703	0.8916	0.7688	0.7055	0.4567	0.5524
	(4.0530)	(0.8188)	(2.2873)	(0.5856)	(1.1762)	(0.3873)	(0.5828)	(0.2422)	(0.2536)	(0.1446)
LLGMM	2.9336	1.3323	2.0072	1.1272	1.2227	0.8891	0.7678	0.7047	0.4528	0.5504
	(3.7628)	(0.8167)	(2.1937)	(0.5943)	(1.1367)	(0.3929)	(0.5466)	(0.2296)	(0.2522)	(0.1521)

Table S6: Comparison among the proposed LLGMM with the local linear (LLE) and Wei and Sun (2017)'s estimators (LLWS) for $\beta_2(\cdot)$ with $\rho = 0.5$

	n = 30		n = 50		n = 100		n = 200		n = 500	
Method	IMSE	IMAE	IMSE	IMAE	IMSE	IMAE	IMSE	IMAE	IMSE	IMAE
Case: T.0										
LLE	3.5540	1.4543	2.0623	1.1037	1.0305	0.7908	0.6318	0.6268	0.3416	0.4745
	(4.5008)	(0.9030)	(2.4326)	(0.6529)	(1.0705)	(0.4072)	(0.5851)	(0.2762)	(0.2075)	(0.1268)
LLWS	3.5376	1.4503	2.0661	1.1055	1.0308	0.7901	0.6310	0.6268	0.3424	0.4752
	(4.5203)	(0.8983)	(2.4252)	(0.6513)	(1.0676)	(0.4057)	(0.5772)	(0.2730)	(0.2072)	(0.1265)
LLGMM	3.4935	1.4702	2.0134	1.1347	1.0884	0.8348	0.6630	0.6577	0.3270	0.4626
	(4.6278)	(0.9365)	(2.1829)	(0.6307)	(1.0182)	(0.3937)	(0.5646)	(0.2735)	(0.2264)	(0.1460)
Case: T.1										
LLE	3.8754	1.5072	2.3521	1.1766	1.1843	0.8451	0.7093	0.6634	0.3776	0.4953
	(4.9489)	(0.9466)	(2.7671)	(0.7136)	(1.2803)	(0.4565)	(0.6832)	(0.3037)	(0.2537)	(0.1496)
LLWS	3.8569	1.5019	2.3575	1.1790	1.1853	0.8455	0.7067	0.6622	0.3775	0.4953
	(4.9550)	(0.9431)	(2.7801)	(0.7137)	(1.2717)	(0.4564)	(0.6788)	(0.3017)	(0.2537)	(0.1499)
LLGMM	3.5657	1.4722	2.2346	1.1643	1.1852	0.8355	0.6910	0.6524	0.3752	0.4944
	(4.8701)	(0.9426)	(2.7355)	(0.6987)	(1.2366)	(0.4327)	(0.6451)	(0.2765)	(0.2822)	(0.1677)
Case: T.2										
LLE	3.9219	1.4984	2.6610	1.2653	1.3586	0.9023	0.8066	0.7048	0.4164	0.5174
	(5.2307)	(0.9739)	(2.9867)	(0.7500)	(1.5197)	(0.5028)	(0.8089)	(0.3421)	(0.3123)	(0.1760)
LLWS	3.8904	1.4846	2.6496	1.2623	1.3594	0.9028	0.8053	0.7039	0.4154	0.5167
	(5.3176)	(0.9768)	(2.9675)	(0.7460)	(1.5175)	(0.5040)	(0.8099)	(0.3433)	(0.3122)	(0.1756)
LLGMM	3.8468	1.4146	2.5303	1.2154	1.3000	0.8976	0.7953	0.7028	0.4144	0.5154
	(5.6730)	(0.9750)	(3.3317)	(0.7419)	(1.5252)	(0.4698)	(0.7880)	(0.3133)	(0.3823)	(0.2244)
Case: T.3										
LLE	3.8508	1.5034	2.2722	1.1590	1.1428	0.8297	0.6894	0.6541	0.3706	0.4916
	(4.9925)	(0.9442)	(2.6654)	(0.6950)	(1.2399)	(0.4462)	(0.6550)	(0.2971)	(0.2425)	(0.1445)
LLWS	3.8340	1.4997	2.2762	1.1609	1.1434	0.8300	0.6877	0.6534	0.3705	0.4916
	(4.9855)	(0.9402)	(2.6605)	(0.6935)	(1.2262)	(0.4453)	(0.6506)	(0.2950)	(0.2423)	(0.1445)
LLGMM	3.7622	1.4137	2.2216	1.1576	1.2392	0.8290	0.6704	0.6513	0.3701	0.4912
	(5.1293)	(0.9651)	(2.4444)	(0.6786)	(1.1947)	(0.4277)	(0.6291)	(0.2795)	(0.2670)	(0.1639)
Case: T.4										
LLE	4.0527	1.5166	2.6907	1.2700	1.3481	0.8962	0.8114	0.7087	0.4206	0.5197
	(5.5110)	(1.0066)	(3.0123)	(0.7582)	(1.5300)	(0.5026)	(0.7705)	(0.3398)	(0.3113)	(0.1762)
LLWS	4.0469	1.5112	2.6848	1.2677	1.3505	0.8982	0.8081	0.7069	0.4182	0.5187
	(5.5581)	(1.0080)	(3.0041)	(0.7545)	(1.5160)	(0.5023)	(0.7699)	(0.3393)	(0.3099)	(0.1751)
LLGMM	3.9732	1.5366	2.7430	1.3131	1.5087	0.9630	0.9692	0.7872	0.4665	0.5290
	(3.6553)	(1.0011)	(2.9492)	(0.7331)	(1.5978)	(0.4878)	(0.7624)	(0.3092)	(0.3807)	(0.2279)
Case: T.5										
LLE	3.6722	1.4702	2.1675	1.1302	1.0886	0.8118	0.6598	0.6403	0.3562	0.4830
	(4.7035)	(0.9218)	(2.5478)	(0.6788)	(1.1530)	(0.4285)	(0.6201)	(0.2855)	(0.2249)	(0.1358)
LLWS	3.6558	1.4658	2.1686	1.1320	1.0893	0.8114	0.6580	0.6398	0.3566	0.4834
	(4.7013)	(0.9181)	(2.5422)	(0.6778)	(1.1447)	(0.4279)	(0.6144)	(0.2830)	(0.2248)	(0.1361)
LLGMM	3.5403	1.4808	2.1296	1.1557	1.1967	0.8707	0.7413	0.6980	0.3502	0.4790
	(4.5039)	(0.9236)	(2.4036)	(0.6701)	(1.1644)	(0.4136)	(0.5899)	(0.2691)	(0.2418)	(0.1520)

Table S7: Comparison among the proposed LLGMM with the local linear (LLE) and Wei and Sun (2017)'s estimators (LLWS) for $\beta_3(\cdot)$ with $\rho = 0.5$

	n = 30		n = 50		n = 100		n = 200		n = 500	
Method	IMSE	IMAE	IMSE	IMAE	IMSE	IMAE	IMSE	IMAE	IMSE	IMAE
Case: T.0										
LLE	2.9058	1.2975	1.6299	0.9565	0.7949	0.6843	0.4162	0.4843	0.1481	0.2909
	(4.1942)	(0.8670)	(2.2539)	(0.6614)	(1.0063)	(0.4405)	(0.5441)	(0.3296)	(0.1835)	(0.1829)
LLWS	2.8981	1.2965	1.6386	0.9585	0.8005	0.6855	0.4134	0.4823	0.1489	0.2914
	(4.2126)	(0.8643)	(2.2978)	(0.6654)	(1.0124)	(0.4399)	(0.5371)	(0.3257)	(0.1855)	(0.1842)
LLGMM	2.6330	1.2561	1.5015	0.9467	0.7889	0.7044	0.4367	0.5220	0.1934	0.3460
	(4.6278)	(0.9365)	(2.1829)	(0.6307)	(1.0182)	(0.3937)	(0.5646)	(0.2735)	(0.2264)	(0.1460)
Case: T.1										
LLE	3.4259	1.3982	1.8563	1.0235	0.9818	0.7578	0.4790	0.5206	0.1879	0.3277
	(5.1517)	(0.9554)	(2.6287)	(0.7038)	(1.2684)	(0.4953)	(0.6689)	(0.3498)	(0.2318)	(0.2071)
LLWS	3.3861	1.3901	1.8511	1.0212	0.9856	0.7586	0.4735	0.5166	0.1903	0.3300
	(5.1381)	(0.9525)	(2.6382)	(0.7059)	(1.2680)	(0.4941)	(0.6656)	(0.3467)	(0.2361)	(0.2086)
LLGMM	2.8195	1.2590	1.6503	0.9728	0.9681	0.7458	0.4685	0.5111	0.1873	0.3248
	(4.8701)	(0.9426)	(2.7355)	(0.6987)	(1.2366)	(0.4327)	(0.6451)	(0.2765)	(0.2822)	(0.1677)
Case: T.2										
LLE	3.1823	1.3546	1.6713	0.9685	0.9461	0.7306	0.4524	0.5015	0.1776	0.3176
	(4.5371)	(0.9171)	(2.3302)	(0.6752)	(1.2673)	(0.5037)	(0.6760)	(0.3464)	(0.2194)	(0.1989)
LLWS	3.1297	1.3359	1.6620	0.9624	0.9451	0.7302	0.4483	0.4987	0.1805	0.3206
	(4.5271)	(0.9165)	(2.3358)	(0.6779)	(1.2495)	(0.5023)	(0.6776)	(0.3435)	(0.2218)	(0.1998)
LLGMM	2.5181	1.1670	1.5578	0.9149	0.9383	0.7282	0.4452	0.4877	0.1764	0.3129
	(5.6730)	(0.9750)	(3.3317)	(0.7419)	(1.5252)	(0.4698)	(0.7880)	(0.3133)	(0.3823)	(0.2244)
Case: T.3										
LLE	3.2804	1.3775	1.7902	1.0073	0.9352	0.7377	0.4639	0.5127	0.1775	0.3186
	(4.9544)	(0.9220)	(2.4596)	(0.6888)	(1.2176)	(0.4858)	(0.6410)	(0.3446)	(0.2183)	(0.2009)
LLWS	3.2564	1.3723	1.7889	1.0060	0.9413	0.7393	0.4611	0.5101	0.1791	0.3200
	(4.9482)	(0.9206)	(2.4811)	(0.6920)	(1.2208)	(0.4857)	(0.6405)	(0.3420)	(0.2225)	(0.2025)
LLGMM	2.9361	1.2996	1.6279	0.9844	0.9232	0.7368	0.4619	0.5101	0.1720	0.3109
	(5.1293)	(0.9651)	(2.4444)	(0.6786)	(1.1947)	(0.4277)	(0.6291)	(0.2795)	(0.2670)	(0.1639)
Case: T.4										
LLE	3.0292	1.3058	1.5242	0.9301	0.8834	0.7045	0.4178	0.4810	0.1616	0.3016
	(4.6230)	(0.9132)	(2.0874)	(0.6382)	(1.1701)	(0.4869)	(0.6464)	(0.3337)	(0.1967)	(0.1898)
LLWS	2.9687	1.2889	1.5212	0.9246	0.8851	0.7044	0.4124	0.4769	0.1745	0.3084
	(4.5720)	(0.9076)	(2.1073)	(0.6417)	(1.1668)	(0.4875)	(0.6462)	(0.3312)	(0.2883)	(0.2074)
LLGMM	2.4915	1.1307	1.4612	0.8890	0.8843	0.7021	0.4109	0.4692	0.1675	0.3023
	(3.6553)	(1.0011)	(2.9492)	(0.7331)	(1.5978)	(0.4878)	(0.7624)	(0.3092)	(0.3807)	(0.2279)
Case: T.5										
LLE	3.2795	1.3741	1.8026	1.0088	0.9343	0.7399	0.4655	0.5147	0.1765	0.3180
	(4.8809)	(0.9269)	(2.4980)	(0.6933)	(1.1991)	(0.4807)	(0.6362)	(0.3441)	(0.2173)	(0.2000)
LLWS	3.2583	1.3698	1.8043	1.0081	0.9409	0.7414	0.4621	0.5117	0.1783	0.3196
	(4.8913)	(0.9288)	(2.5296)	(0.6977)	(1.2058)	(0.4814)	(0.6342)	(0.3409)	(0.2212)	(0.2016)
LLGMM	2.8400	1.2891	1.6410	0.9832	0.9131	0.7406	0.4613	0.5125	0.1788	0.3171
	(4.5039)	(0.9236)	(2.4036)	(0.6701)	(1.1644)	(0.4136)	(0.5899)	(0.2691)	(0.2418)	(0.1520)

S5 Technical details

In this section, we provide technical details of the proposed theorems in Section 4. We prove theorems 1 and 2 by proving the following lemmas.

S5.1 Some useful lemmas

Lemma 1. *Under the conditions (C5)* $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i(s_0)\mathfrak{M}(\mathbf{X}_i)$ *is tight.*

Proof. Consider the class of function $\mathcal{C} = \{U(s_0)\mathfrak{M}(\mathbf{X}_i) : s_0 \in [0,1]\}$. Therefore, due to the assumption (C5), $\mathcal{C}$ is a P-Donsker class. Therefore, $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i(s_0)\mathfrak{M}(\mathbf{X}_i)$ is tight. □

Lemma 2. *Under the assumptions (C1), (C2) and (C9), the following holds for any power* $c \geq 0$.

$$\sup_{s\in[0,1]} \left| \int K_h(t-s)\left\{(t-s)/h\right\}^c d\Pi(t) - \Pi(t) \right| = O(1/(rh)^{-1/2}). \tag{S5.7}$$

The above bound can be replaced by $O(1/rh)$ *for fixed design case.*

Proof. This can be proved by using the empirical process techniques by observing that the class $\{K((\cdot - s/h))((\cdot - s/h))^c : s \in [0,1]\}$ is a P-Donsker class (Zhu et al., 2012). For the balanced case, the results can be shown using Tayler's series expansion. □

Lemma 3. *Define* $\mathbf{I}(s_0) = \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)\mathbf{Q}_{ij}(s_0)\mathbf{W}_{ij}(s_0)^T$. *Under the conditions (C1), (C2), (C3) and (C9)* $\mathbf{I}(s_0) = f(s_0)diag(1, \nu_{21}) \otimes \mathbf{\Omega} + O(h + \delta_{n1}(h))$ *almost surely, where* $\mathbf{\Omega} = \mathbb{E}\{\mathfrak{M}(\mathbf{X})\mathbf{X}^T\}$.

Proof. Observe the following.

$$
\begin{aligned}
\mathbf{I}(s_0) &= \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)\mathbf{Q}_{ij}(s_0)\mathbf{W}_{ij}(s_0)^{\mathrm{T}} \\
&= \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)\left\{\mathbf{z}_h(s_j - s_0) \otimes \mathfrak{M}(\mathbf{X}_i)\right\}\left\{\mathbf{z}_h(s_j - s_0) \otimes \mathbf{X}_i\right\}^{\mathrm{T}} \\
&= \frac{1}{nR}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s)\left\{\mathbf{z}_h(s_j - s_0)^{\otimes 2} \otimes \mathfrak{M}(\mathbf{X}_i)\mathbf{X}_i^{\mathrm{T}}\right\} \\
&= \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)\begin{pmatrix} \mathfrak{M}(\mathbf{X}_i)\mathbf{X}_i^{\mathrm{T}} & \mathfrak{M}(\mathbf{X}_i)\mathbf{X}_i^{\mathrm{T}}(s_j - s_0)/h \\ \mathfrak{M}(\mathbf{X}_i)\mathbf{X}_i^{\mathrm{T}}(s_j - s_0)/h & \mathfrak{M}(\mathbf{X}_i)\mathbf{X}_i^{\mathrm{T}}((s_j - s_0)/h)^2 \end{pmatrix} \\
&:= \begin{pmatrix} \mathbf{I}_{11}(s_0) & \mathbf{I}_{12}(s_0) \\ \mathbf{I}_{21}(s_0) & \mathbf{I}_{22}(s_0) \end{pmatrix}. \qquad \text{(S5.8)}
\end{aligned}
$$

Let us define $\mathbf{I}_{a,b} = \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)(s_j - s_0)^{a+b}\mathfrak{M}(\mathbf{X}_i)\mathbf{X}_i^{\mathrm{T}}$. Assume that ν_{41} is finite and due to condition (C2), for general index c, we can derive the uniform bound of for all $s_0 \in \mathcal{S}$.

$$
\begin{aligned}
\mathbb{E}\{\mathbf{I}_{a,b}(s_0)\} &= \mathbb{E}\left\{\frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)((s_j - s_0)/h)^c\mathfrak{M}(\mathbf{X}_i)\mathbf{X}_i^{\mathrm{T}}\right\} \\
&= \mathbf{\Omega}\mathbb{E}\left\{\frac{1}{r}\sum_{j=1}^{r} K_h(s_j - s_0)((s_j - s_0)/h)^c\right\} \\
&= \mathbf{\Omega}\int K_h(u - s_0)((u - s_0)/h)^c f(u)du \\
&= \mathbf{\Omega}\int K(u)u^c f(s_0 + hu)du
\end{aligned}
$$

$$= \mathbf{\Omega} \int K(u) u^c \left\{ f(s_0) + huf'(s_0) + 0.5h^2u^2 f''(s_0) + \cdots \right\} du$$

$$= \mathbf{\Omega} \begin{cases} f(s_0) + O(h^2) & c = 0, \text{if } \nu_{21} < \infty,\ f'' \text{ exists and finite} \\ O(h) & c = 1, \text{if } \nu_{21} < \infty,\ f' \text{ exists and finite} \\ f(s_0)\nu_{21} + O(h^2) & c = 2, \text{if } \nu_{41} < \infty,\ f'' \text{ exists and finite} \\ O(h) & c = 3, \text{if } \nu_{41} < \infty,\ f' \text{ exists and finite.} \end{cases} \tag{S5.9}$$

Moreover, under the condition (C3), we have $\mathbb{E}\|\mathbf{X}\|^a$ is finite for some $a > 2$ and can define, $b_n = h^2 + h/r$ where $h \to 0$ such that $b_n^{-1}(\log n/n)^{1-2/a} = o(1)$. Thus, $\delta_{n1}(h) = \{b_n \log n/nh^2\}^{1/2}$. Now to establish the uniform bound for $\mathbf{I}(s_0)$, by using Lemma 2 in Li and Hsing (2010) for each of $\mathbf{I}_{a,b}(s_0)$ for $a, b = 1, 2$, we have

$$\mathbf{I}(s_0) = f(s_0)(\text{diag}(1, \nu_{21})) \otimes \mathbf{\Omega} + O(h + \delta_{n1}(h)) \qquad \text{almost surely.} \tag{S5.10}$$

□

Lemma 4. *Define,* $\mathbf{J}(s_0) = \frac{1}{nr} \sum_{i=1}^{n} \sum_{j=1}^{r} K_h(s_j - s_0) \mathbf{Q}_{ij}(s_0) \mathbf{X}_i^T \boldsymbol{\beta}_0(s_j)$*. Thus, under the conditions (C1), (C2), (C4), (C3) and (C9),* $\mathbf{J}(s_0) - \mathbf{I}(s_0)\boldsymbol{\gamma}_0(s_0) = 0.5h^2\{f(s_0)(\nu_{21}, 0)^T \otimes \mathbf{\Omega}\}\ddot{\boldsymbol{\beta}}(s_0) + O(\delta_{n1}(h) + h)$ *almost surely, where* $\boldsymbol{\gamma}_0(s_0) = (\boldsymbol{\beta}_0(s_0)^T, h\dot{\boldsymbol{\beta}}_0(s_0)^T)^T$*. Moreover,*

$$\mathbf{T}(s_0) = \frac{1}{nr} \sum_{i=1}^{n} \sum_{j=1}^{r} K_h(s_j - s_0) \mathbf{Q}_{ij}(s_0) U_{ij} = O(\delta_{n1}(h))$$

almost surely.

Proof. Observe that, because of condition (C4), using Taylor's series expansion,

$$\begin{aligned}
\mathbf{J}(s_0) &= \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)\mathbf{Q}_{ij}(s_0)\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0(s_0) \\
&= \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)\mathbf{Q}_{ij}(s_0) \\
&\times \left\{\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0(s_0) + (s_j - s_0)\mathbf{X}_i^{\mathrm{T}}\dot{\boldsymbol{\beta}}_0(s_0) + 0.5(s_j - s_0)^2\mathbf{X}_i^{\mathrm{T}}\ddot{\boldsymbol{\beta}}_0(s_0)\right\} + o(h^2) \\
&= \mathbf{I}(s_0)\boldsymbol{\gamma}_0(s_0) + 0.5h^2\mathbf{I}_{21}(s_0)\ddot{\boldsymbol{\beta}}_0(s_0) + o(h^2). \qquad \text{(S5.11)}
\end{aligned}$$

Using similar arguments, due to Lemma 2 in (Li and Hsing, 2010), under the conditions (C1), (C2) and (C3), with ν_{41} being finite, we have

$$\begin{aligned}
\mathbf{I}_{21}(s_0) &= \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)\begin{pmatrix}((s_j - s_0)/h)^2 \\ ((s_j - s_0)/h)^3\end{pmatrix}\mathfrak{M}(\mathbf{X}_i)\mathbf{X}_i^{\mathrm{T}} \\
&= f(s_0)(\nu_{21}, 0)^{\mathrm{T}} \otimes \boldsymbol{\Omega} + O(\delta_{n1}(h) + h) \qquad \text{almost surely} \quad \text{(S5.12)}
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{T}(s_0) &= \begin{pmatrix}\frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)\mathfrak{M}(\mathbf{X}_i)U_{ij} \\ \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)((s_j - s_0)/h)\mathfrak{M}(\mathbf{X}_i)U_{ir}\end{pmatrix} \\
&= O(\delta_{n1}(h)) \qquad \text{almost surely.} \qquad \text{(S5.13)}
\end{aligned}$$

□

Lemma 5. *Under conditions (C1),(C2), (C3), (C5), (C9),* $\sqrt{n}\mathbf{T}(s_0)(1+o_{a.s.}(1)) \xrightarrow{d} N(0, f^2(s_0)(\mathcal{U}\otimes\boldsymbol{\Sigma}^*_{\mathfrak{M}}(s_0,s_0))$ *where* $\mathbf{T}(s_0)$ *is defined in Lemma 4, where the element of* (l,l') *of the matrix* $\mathcal{U}$ *is* $\nu_{l-1}\nu_{l'-1}$ *and* $\boldsymbol{\Sigma}^*_{\mathfrak{M}}(s_0,s_0) = \lim\limits_{n\to\infty}\frac{1}{n}\sum\limits_{i=1}^{n}\mathbb{E}\{\mathfrak{M}(\mathbf{X}_i)\mathfrak{M}(\mathbf{X}_i)^T\Sigma_{\mathbf{X}_i}(s_0,s_0)\}$.

Proof. Note that

$$\sqrt{n}\mathbf{T}(s_0) = \frac{1}{\sqrt{n}r}\sum_{i=1}^{n}\sum_{j=1}^{r}K_h(s_j-s_0)\left[\mathbf{z}_h(s_j-s_0)\otimes\mathfrak{M}(\mathbf{X}_i)\right]U_{ij} \quad \text{(S5.14)}$$

Therefore, the variance of the above quantity is

$$\begin{aligned}
&\mathrm{Var}\{\sqrt{n}\mathbf{T}(s_0)\}\\
&= \frac{1}{n}\mathbb{E}\left\{\sum_{i=1}^{n}\sum_{j=1}^{r}\sum_{j'=1}^{r}K_h(s_j-s_0)K_h(s_{j'}-s_0)\left[\mathbf{z}_h(s_j-s_0)\mathbf{z}_h(s_{j'}-s_0)^{\mathrm{T}}\otimes\mathfrak{M}(\mathbf{X}_i)^{\otimes 2}\right]U_{ij}U_{ij'}\right\}\\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left\{\frac{1}{r^2}\sum_{j=1}^{r}\sum_{j'=1}^{r}K_h(s_j-s_0)K_h(s_{j'}-s_0)\left[\mathbf{z}_h(s_j-s_0)\mathbf{z}_h(s_{j'}-s_0)\otimes\mathfrak{M}(\mathbf{X}_i)^{\otimes 2}\right]\Sigma_{\mathbf{X}_i}(s_j,s_{j'})\right\}\\
&= \mathbb{E}\left\{\frac{1}{r^2}\sum_{j=1}^{r}\sum_{j'=1}^{r}K_h(s_j-s_0)K_h(s_{j'}-s_0)\mathbf{z}_h(s_j-s_0)\mathbf{z}_h(s_j-s_0)^{\mathrm{T}}\otimes\boldsymbol{\Sigma}^*_{\mathfrak{M}}(s_j,s_{j'})\right\}\\
&= \mathbb{E}\{\mathbf{D}_1(s_0)\}+\mathbb{E}\{\mathbf{D}_2(s_0)\}, \qquad \text{(S5.15)}
\end{aligned}$$

where $\boldsymbol{\Sigma}^*_{\mathfrak{M}}(s,s') = \lim\limits_{n\to\infty}\frac{1}{n}\sum\limits_{i=1}^{n}\mathbb{E}\{\mathfrak{M}(\mathbf{X}_i)^{\otimes 2}\Sigma_{\mathbf{x}_i}(s,s')\}$, $\mathbf{D}_1(s_0) = \frac{1}{r^2}\sum_{j=1}^{n}K_h^2(s_j-s_0)\mathbf{z}_h(s_j-s_0)^{\otimes 2}\otimes\boldsymbol{\Sigma}^*_{\mathfrak{M}}(s_j,s_j)$ and $\mathbf{D}_2(s_0) = \frac{1}{r(r-1)}\sum_{\substack{j=1\\ j\neq j'}}^{n}\sum_{j'=1}^{r}K_h(s_j-s_0)K_h(s_{j'}-s_0)\mathbf{z}_h(s_j-s_0)\mathbf{z}_h(s_{j'}-s_0)^{\mathrm{T}}\otimes\boldsymbol{\Sigma}^*_{\mathfrak{M}}(s_j,s_{j'})$. Note that

$$\mathbb{E}\{\mathbf{D}_1(s_0)\} = \mathbb{E}\left\{\frac{1}{r^2}\sum_{j=1}^{r}K^2(s_j-s_0)\mathbf{z}_h(s_j-s_0)^{\otimes 2}\otimes\boldsymbol{\Sigma}^*_{\mathfrak{M}}(s_j,s_j)\right\}$$

$$= \frac{1}{r}\int K_h^2(t-s_0)\mathbf{z}_h(t-s_0)^{\otimes 2} \otimes \mathbf{\Sigma}^*_{\mathfrak{M}}(t,t)f(t)dt$$
$$= \frac{1}{hr}\int K^2(t)\begin{pmatrix}1 & t\\ t & t^2\end{pmatrix} \otimes \mathbf{\Sigma}^*_{\mathfrak{M}}(s_0+hu, s_0+hu)f(s_0+ht)dt$$
$$= \frac{1}{hr}\left\{f(s_0)\text{diag}(\nu_{02},\nu_{22}) \otimes \mathbf{\Sigma}^*_{\mathfrak{M}}(s_0,s_0) + O(h)\right\}. \quad \text{(S5.16)}$$

Now assume that $\Theta(s_0) = \mathbb{E}\{\mathbf{D}_2(s_0)\}$ with (l,l')-th entry $\theta_{l,l'}$ and $\mathcal{P}(t) = \int_{\mathcal{S}} K_h(t-s_0)K_h(t'-s_0)\mathbf{z}_h(t-s_0)\mathbf{z}_h(t'-s_0) \otimes \mathbf{\Sigma}^*_{\mathfrak{M}}(t,t')f(t')dt'$ with (l,l')-the block element $\mathcal{P}_{l,l'}$. Therefore, using Hájek projection (Vaart and Wellner, 1996), we have

$$\mathbf{D}_{2,l.l'}(s_0) = \theta_{l,l'}(s_0) + \frac{2}{r}\sum_{j=1}^{r}\{\mathcal{P}_{l,l'}(s_j) - \theta_{l,l'}(s_0)\} + \tilde{\epsilon}_{l,l'}(s_0), \quad \text{(S5.17)}$$

where $\frac{2}{r}\sum_{j=1}^{r}\{\mathcal{P}_{l,l'}(s_j) - \theta_{l,l'}(s_0)\}$ is the projection on $\mathbf{D}_{2,l,l'}(s_0) - \theta_{l,l'}(s_0)$ onto the set of all statistics of the linear order form. Thus, it is easy to see $\text{Var}\{\tilde{\epsilon}\} = O(1/(rh)^2)$ (Zhu et al., 2012). Since the Taylor series expansion for small $h \to 0$, we have $\theta_{l,l'}(s_0) = f(s_0)^2\nu_{l-1,}\nu_{l'-1,1}\mathbf{\Sigma}^*_{\mathfrak{M}}(s_0,s_0)$. Therefore, in summary, we have $\text{Var}\{\sqrt{n}\mathbf{T}(s_0)\} = f^2(s_0)\mathcal{U} \otimes \mathbf{\Sigma}^*_{\mathfrak{M}}(s_0,s_0)$, where the element (l,l') of the matrix $\mathcal{U}$ is $\nu_{l-1}\nu_{l'-1}$.

To hold the above asymptotic results, we need to show that $\sqrt{n}\mathbf{T}(s_0)$ be tight asymptotically. Therefore, consider the following, for suitable choice of $\underline{l} < \bar{l}$ after change of variables,

$$\sqrt{n}\mathbf{T}(s_0) = \frac{1}{\sqrt{n}r}\sum_{i=1}^{n}\sum_{r=1}^{r} K_h(s_j-s_0)[\mathbf{z}_h(s_j-s_0) \otimes \mathfrak{M}(\mathbf{X}_i)]U_{ij}$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{1}{r} \sum_{j=1}^{r} K_h(s_j - s_0) \mathbf{z}_h(s_j - s_0) U_{ij} - \int_0^1 K_h(t - s_0) \mathbf{z}_h(t - s_0) U_i(t) f(t) dt \right\} \otimes \mathfrak{M}(\mathbf{X}_i) \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i(s_0) \int_{\underline{l}}^{\bar{l}} K(t)(1,t)^{\mathrm{T}} f(s_0 + ht) dt \otimes \mathfrak{M}(\mathbf{X}_i) \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_{\underline{l}}^{\bar{l}} K(t)(1,t)^{\mathrm{T}} \left\{ U_i(s_0 + ht) - U_i(s_0) \right\} f(s_0 + ht) dt \otimes \mathfrak{M}(\mathbf{X}_i) \\
&:= \mathbf{T}_1(s_0) + \mathbf{T}_2(s_0) + \mathbf{T}_3(s_0). \qquad \text{(S5.18)}
\end{aligned}
$$

Note that,

$$
\begin{aligned}
&\mathbf{T}_1(s_0) \\
&= \frac{1}{r} \sum_{r=1}^{r} K_h(s_j - s_0) \mathbf{z}_h(s_j - s_0) \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_{ij} \otimes \mathfrak{M}(\mathbf{X}_i) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i(t) \otimes \mathfrak{M}(\mathbf{X}_i) \right\} \\
&\quad + \left\{ \frac{1}{r} \sum_{j=1}^{r} K_h(s_j - s_0) \mathbf{z}_h(s_j - s_0) - \int_{\underline{l}}^{\bar{l}} K_h(t - s_0) \mathbf{z}_h(t - s_0) f(t) dt \right\} \\
&\qquad \times \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i(t) \otimes \mathfrak{M}(\mathbf{X}_i) \right\} \\
&\quad + \int_{\underline{l}}^{\bar{l}} K_h(t - s_0) \mathbf{z}_h(t - s_0) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ U_i(s_0) - U_i(s) \right\} f(t) dt \otimes \mathfrak{M}(\mathbf{X}_i) \\
&:= \mathbf{T}_{11}(s_0) + \mathbf{T}_{12}(s_0) + \mathbf{T}_{13}(s_0). \qquad \text{(S5.19)}
\end{aligned}
$$

Due to the Donsker Theorem, we have $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathfrak{M}(\mathbf{X}_i) U_i(s)$ weekly converges to a centered Gaussian process and $\sup_{s \in [0,1]} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathfrak{M}(\mathbf{X}_i) U_i(s) \right| = O_p(1)$ (Vaart and Wellner, 1996). Therefore,

$$
|\mathbf{T}_{11}(s_0)| \leq \frac{1}{r} \sum_{j=1}^{r} K_h(s_j - s_0) \|\mathbf{z}_h(s_j - s_0)\|_2 \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_{ij} \otimes \mathfrak{M}(\mathbf{X}_i) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i(s) \otimes \mathfrak{M}(\mathbf{X}_i) \right|
$$

$$
\begin{aligned}
&\leq \frac{1}{r}\sum_{j=1}^{r} K_h(s_j - s_0)\|\mathbf{z}_h(s_j - s_0)\|_2 \\
&\quad \times \sup_{|s-s_0|\leq h}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i(s)\otimes \mathfrak{M}(\mathbf{X}_i) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i(s_0)\otimes \mathfrak{M}(\mathbf{X}_i)\right| \\
&= o_P(1). \qquad \text{(S5.20)}
\end{aligned}
$$

$$
\begin{aligned}
&|\mathbf{T}_{12}(s_0)| \\
&\leq \left|\frac{1}{r}\sum_{j=1}^{r} K_h(s_j - s_0)\mathbf{z}_h(s_j - s_0) - \int_{\underline{l}}^{\bar{l}} K_h(t - s_0)\mathbf{z}_h(t - s_0)f(t)dt\right| \\
&\quad \times \sup_{t\in[0,1]}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i(t)\otimes \mathfrak{M}(\mathbf{X}_i)\right\} \\
&= O_P(1/\sqrt{rh})O_P(1) = o_P(1). \qquad \text{(S5.21)}
\end{aligned}
$$

The above bound holds for Lemma 2 and Condition (C9) so that $mh \to \infty$.

$$
\begin{aligned}
&|\mathbf{T}_{13}(s_0)| \\
&\leq \sup_{|s-s_0|\leq h}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i(s)\otimes \mathfrak{M}(\mathbf{X}_i) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i(s_0)\otimes \mathfrak{M}(\mathbf{X}_i)\right| \\
&\times \int_{\underline{l}}^{\bar{l}} K_h(s_j - s_0)\|\mathbf{z}_h(s_j - s_0)\|_2 f(s)ds \\
&= O_P(1). \qquad \text{(S5.22)}
\end{aligned}
$$

By combining the above three bounds, due to conditions (C1),(C2), (C3), (C5), (C9), we obtain $\mathbf{T}_1(s_0) = o_P(1)$. Now, rewrite $\mathbf{T}_3(s_0)$ as

$$
\mathbf{T}_3(s) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{\underline{l}}^{\bar{l}} K(t)(1,t)^{\mathrm{T}}\left\{U_i(s_0 + ht) - U_i(s_0)\right\} f(s_0 + ht)dt \otimes \mathfrak{M}(\mathbf{X}_i)
$$

$$= \int_{\underline{l}}^{\bar{l}} K(t)(1,t)^{\mathrm{T}} \otimes \{U_i(s_0+ht) - U_i(s_0)\} \mathfrak{M}(\mathbf{X}_i) f(s_0+ht)dt. \tag{S5.23}$$

Since, $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathfrak{M}(\mathbf{X}_i) U_i(s_0)$ is asymptotically tight, for any $h \to 0$, we have the following (Vaart and Wellner, 1996).

$$\sup_{s_0 \in [0,1]:|t|\leq 1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathfrak{M}(\mathbf{X}_i) \{U_i(s_0+ht) - U_i(s_0)\} = o_P(1). \tag{S5.24}$$

Now it is enough to show that $\mathbf{T}_2(s_0)$ is tight. First, observe that

$$\begin{aligned}
&(1,0) \int_{\underline{l}}^{\bar{l}} K(t) \mathrm{diag}(1, \nu_{21}^{-1})(1,t)^{\mathrm{T}} f(s_0+ht)dt \\
&= \int_{\underline{l}}^{\bar{l}} K(t) f(s_0+ht)dt \\
&= \int_{\underline{l}}^{\bar{l}} K(t) \{f(s_0) + htf'(s_0) + \cdots\} \\
&= f(s_0) + o(h).
\end{aligned} \tag{S5.25}$$

Therefore, $\mathbf{T}_2(s_0)(1+o_P(h)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i(s_0) \otimes \mathfrak{M}(\mathbf{X}_i)$. By assumption (C5), $\mathbf{T}_2(s_0)$ is tight. □

S5.2 Proof of Theorem 1

Under the initial estimates, by considering $\mathfrak{M}(\mathbf{X}) = \mathbf{X}$, $\mathbf{\Omega}$ can be replaced by $\mathbf{\Omega_x}$ in Equation (S5.26) and inverse of $\mathbf{\Omega_x}$ exits. Therefore, by using Lemma 3, it is easy to observe that, almost surely

$$\mathbf{I}(s_0)^{-1} = f(s_0)^{-1}(\mathrm{diag}(1, \nu_{21})^{-1}) \otimes \mathbf{\Omega}_{\mathbf{x}}^{-1} + O(h + \delta_{n1}(h)). \tag{S5.26}$$

Similarly, for the numerator, we have the following.

$$
\begin{aligned}
&\frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r}K_h(s_j-s_0)\left\{\mathbf{z}_h(s_j-s_0)\otimes\mathbf{X}_i\right\}Y_{ij}\\
&=\frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r}K_h(s_j-s_0)\left\{\mathbf{z}_h(s_j-s_0)\otimes\mathbf{X}_i\right\}\left\{\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0(s_j)+U_{ij}\right\}\\
&=\mathbf{I}(s_0)\boldsymbol{\gamma}_0(s_0)+0.5h^2\mathbf{I}_{21}(s_0)\ddot{\boldsymbol{\beta}}_0(s_0)+\mathbf{T}(s_0)+o(h^2). \qquad \text{(S5.27)}
\end{aligned}
$$

Thus, using Equation (S5.26) and (S5.27), we can derive,

$$
\begin{aligned}
&\breve{\boldsymbol{\beta}}(s_0)=[(1,0)\otimes\mathbf{I}_p]\mathbf{I}(s_0)^{-1}\left\{\mathbf{I}(s_0)\boldsymbol{\gamma}_0(s_0)+0.5h^2\mathbf{I}_{21}(s_0)\ddot{\boldsymbol{\beta}}_0(s_0)+\mathbf{T}(s_0)+o(h^2)\right\}\\
&=\boldsymbol{\beta}_0(s_0)+[(1,0)\otimes\mathbf{I}_p]f(s_0)^{-1}\left\{\mathrm{diag}(1,\nu_{21})^{-1}\otimes\boldsymbol{\Omega}_{\mathbf{x}}^{-1}\right\}\left\{f(s_0)(\nu_{21},0)\otimes\boldsymbol{\Omega}_{\mathbf{x}}\right\}0.5h^2\ddot{\boldsymbol{\beta}}_0(s_0)\\
&\qquad+O(\delta_{n1}(h)+h)\\
&=\boldsymbol{\beta}_0(s_0)+0.5h^2\nu_{21}\ddot{\boldsymbol{\beta}}_0(s_0)+O(\delta_{n1}(h)+h)\\
&=\boldsymbol{\beta}_0(s_0)+O(\delta_{n1}(h)+h) \qquad \text{almost surely.} \qquad \text{(S5.28)}
\end{aligned}
$$

Therefore, $\sup_{s_0\in\mathcal{S}}\left|\breve{\boldsymbol{\beta}}(s_0)-\boldsymbol{\beta}_0(s_0)\right| = O(\delta_{n1}+h)$ almost surely. Furthermore, observe that the bias of the initial estimator is

$$
\mathbb{E}\{\breve{\boldsymbol{\beta}}(s_0)\}-\boldsymbol{\beta}_0(s_0)=0.5h^2\nu_{21}\ddot{\boldsymbol{\beta}}_0(s_0)\left\{1+O_P(\delta_{n1}(h)+h)\right\}. \qquad \text{(S5.29)}
$$

Now, to calculate the variance, note that

$$
\begin{aligned}
&\sqrt{n}\{\breve{\boldsymbol{\beta}}(s_0)-\boldsymbol{\beta}(s_0)-0.5h^2\nu_{21}\ddot{\beta}_0(s_0)\}(1+o_{a.s.}(1))\\
&=[(1,0)\otimes\mathbf{I}_p]f(s_0)\left\{\mathrm{diag}(1,\nu_{21})^{-1}\otimes\boldsymbol{\Omega}_{\mathbf{x}}^{-1}\right\}\sqrt{n}\mathbf{T}(s_0). \qquad \text{(S5.30)}
\end{aligned}
$$

By Lemma 5, we have the variance of the above quantity $\mathbf{\Omega}_{\mathbf{x}}^{-1}\mathbf{\Sigma}_{\mathbf{x}}^{*}(s_0, s_0)\mathbf{\Omega}_{\mathbf{x}}^{-1}$, where $\mathbf{\Sigma}_{\mathbf{x}}^{*}(s_0, s_0) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\{\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\Sigma_{\mathbf{x}_i}(s_0, s_0)\}$.

S5.3 Proof of Theorem 2

Define $\mathbf{C}_{\kappa_0}(s, s') = \sum_{k=1}^{\kappa_0} \lambda_k \boldsymbol{\phi}_k(s)\boldsymbol{\phi}_k(s')^{\mathrm{T}}$ and hence, we can define $\mathbf{C}_{\kappa_0}^{-1}(s, s')$ with possible block matrix

$$\mathbf{C}_{\kappa_0}^{-1}(s, s') = \sum_{k=1}^{\kappa_0} \lambda_k^{-1}\boldsymbol{\phi}_k(s)\boldsymbol{\phi}_k(s')^{\mathrm{T}} = \begin{pmatrix} \mathbf{C}_{\kappa_0,1,1}^{-1}(s, s') & 0 \\ 0 & \mathbf{C}_{\kappa_0,2,2}^{-1}(s, s') \end{pmatrix}. \tag{S5.31}$$

Also, define,

$$\widehat{\gamma}_{\kappa_0}(s_0) = \left\{\sum_{k=1}^{\kappa_0} \frac{\lambda_k}{\lambda_k^2 + \alpha}\mathcal{X}_k(s_0)\mathcal{X}_k(s_0)^{\mathrm{T}}\right\}^{-1} \left\{\sum_{k=1}^{\kappa_0} \frac{\lambda_k}{\lambda_k^2 + \alpha}\mathcal{X}_k(s_0)\mathcal{Y}_k(s_0)\right\}. \tag{S5.32}$$

where

$$\mathcal{X}_k(s_0) = \frac{1}{nr}\sum_{j=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)\mathbf{W}_{ij}(s_0)\mathbf{Q}_{ij}(s_0)^{\mathrm{T}}\boldsymbol{\phi}_k(s_0) + O(\delta(h)) \tag{S5.33}$$

and

$$\mathcal{Y}_k(s_0) = \frac{1}{nr}\sum_{j=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)\boldsymbol{\phi}_k(s_0)^{\mathrm{T}}\mathbf{Q}_{ij}(s_0)Y_{ij} + O(\delta(h)) \tag{S5.34}$$

almost everywhere. Therefore, we have the following.

$$\sum_{k=1}^{\kappa_0} \frac{\lambda_k}{\lambda_k^2 + \alpha}\mathcal{X}_k(s_0)\mathcal{X}_k(s_0)^{\mathrm{T}} + O(\delta(h))$$

$$
\begin{aligned}
&= \left\{ \frac{1}{nr} \sum_{i=1}^{n} \sum_{j=1}^{r} K_h(s_j - s_0) \mathbf{W}_{ij}(s_0) \mathbf{Q}_{ij}(s_0)^{\mathrm{T}} \right\} \sum_{k=1}^{\kappa_0} \lambda_k^{-1} \boldsymbol{\phi}_k(s_0) \boldsymbol{\phi}_k(s_0)^{\mathrm{T}} \\
&\quad \times \left\{ \frac{1}{nr} \sum_{i=1}^{n} \sum_{j=1}^{r} K_h(s_j - s_0) \mathbf{W}_{ij}(s_0) \mathbf{Q}_{ij}(s_0) \right\}^{\mathrm{T}} \\
&= \mathbf{I}(s_0)^{\mathrm{T}} \mathbf{C}_{\kappa_0}^{-1}(s_0, s_0) \mathbf{I}(s_0) \\
&= f^2(s_0) \left[\mathrm{diag}(1, \nu_{21}) \otimes \boldsymbol{\Omega}^{\mathrm{T}} \right] \mathbf{C}_{\kappa_0}^{-1}(s_0, s_0) \left[\mathrm{diag}(1, \nu_{21}) \otimes \boldsymbol{\Omega} \right] + O(\delta(h)) \\
&= \mathcal{V}(s_0) + O(\delta(h)), \qquad \text{(S5.35)}
\end{aligned}
$$

where we define $\mathcal{V}(s_0) = f^2(s_0)\mathrm{diag}\left(\boldsymbol{\Omega}^{\mathrm{T}} \mathbf{C}_{\kappa_0,1,1,}^{-1}(s_0, s_0)\boldsymbol{\Omega}, \nu_{21}^2 \boldsymbol{\Omega}^{\mathrm{T}} \mathbf{C}_{\kappa_0,2,2,}^{-1}(s_0, s_0)\boldsymbol{\Omega}\right)$

and

$$
\begin{aligned}
&\sum_{k=1}^{\kappa_0} \frac{\lambda_k}{\lambda_k^2 + \alpha} \mathcal{X}_k(s_0) \mathcal{Y}_k(s_0) \\
&= \left\{ \frac{1}{nr} \sum_{i=1}^{n} \sum_{j=1}^{r} K_h(s_j - s_0) \mathbf{W}_{ij}(s_0) \mathbf{Q}_{ij}(s_0)^{\mathrm{T}} \right\} \sum_{k=1}^{\kappa_0} \lambda_k^{-1} \boldsymbol{\phi}_k(s_0) \boldsymbol{\phi}_k(s_0)^{\mathrm{T}} \\
&\quad \times \left\{ \frac{1}{nr} \sum_{i=1}^{n} \sum_{j=1}^{r} K_h(s_j - s_0) \mathbf{Q}_{ij}(s_0) Y_{ij} \right\} \\
&= \mathbf{I}(s_0)^{\mathrm{T}} \mathbf{C}_{\kappa_0}^{-1}(s_0, s_0) \left\{ \mathbf{I}(s_0) \boldsymbol{\gamma}_0(s_0) + 0.5h^2 \mathbf{I}_{21}(s_0) \ddot{\boldsymbol{\beta}}(s_0) + \mathbf{T}(s_0) + o(h^2) \right\}. \qquad \text{(S5.36)}
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&\widehat{\boldsymbol{\beta}}(s_0) - \boldsymbol{\beta}_0(s_0) \\
&= 0.5h^2 [(1, 0) \otimes \mathbf{I}_p] \mathcal{V}(s_0)^{-1} \mathbf{I}(s_0)^{\mathrm{T}} \mathbf{C}_{\kappa_0}^{-1}(s_0, s_0) \mathbf{I}_{21}(s_0) \ddot{\boldsymbol{\beta}}(s_0) \\
&\quad + [(1, 0) \otimes \mathbf{I}_p] \mathcal{V}(s_0)^{-1} \mathbf{I}(s_0)^{\mathrm{T}} \mathbf{C}_{\kappa_0}^{-1}(s_0, s_0) \mathbf{T}(s_0) + O(\delta(h))
\end{aligned}
$$

$$= 0.5h^2 f^2(s_0)[(1,0)\otimes \mathbf{I}_p]\mathcal{V}(s_0)^{-1}[\text{diag}(1,\nu_{21})\otimes \mathbf{\Omega}^{\mathrm{T}}]\mathbf{C}_{\kappa_0}^{-1}(s_0,s_0)[(\nu_{21},0)^{\mathrm{T}}\otimes \mathbf{\Omega}]\ddot{\boldsymbol{\beta}}(s_0)$$

$$+ f^2(s_0)[(1,0)\otimes \mathbf{I}_p]\mathcal{V}(s_0)^{-1}[diag(1,\nu_{21})\otimes \mathbf{\Omega}^{\mathrm{T}}]\mathbf{C}_{\kappa_0}^{-1}(s_0,s_0)\mathbf{T}(s_0) + O(\delta(h))$$

$$= 0.5h^2\nu_{21}\ddot{\boldsymbol{\beta}}(s_0) + \mathcal{C}(s_0)\mathbf{T}(s_0) + O(\delta(h)), \tag{S5.37}$$

where, $\mathcal{C}(s_0) = f^2(s_0)[(1,0)\otimes\mathbf{I}_p]\mathcal{V}(s_0)^{-1}[diag(1,\nu_{21})\otimes\mathbf{\Omega}^{\mathrm{T}}]\mathbf{C}_{\kappa_0}^{-1}(s_0,s_0)$. Thus, in order to obtain the asymptotic variance, consider, using Lemmas 3, 4 and 5, we have $\sqrt{n}\{\widehat{\boldsymbol{\beta}}(s_0) - \boldsymbol{\beta}_0(s_0) - 0.5h^2\nu_{21}\ddot{\boldsymbol{\beta}}(s_0)\} \xrightarrow{d} N(0,\mathcal{A}(s_0,s_0))$ where $\mathcal{A}(s_0,s_0) = \mathcal{B}(s_0,s_0)^{-1}\mathbf{\Omega}^{\mathrm{T}}\mathbf{C}_{\kappa_0,11}(s_0,s_0)^{-1}\mathbf{\Sigma}^*_{\mathfrak{M}}(s_0,s_0)\mathbf{C}_{\kappa_0,11}(s_0,s_0)^{-1}\mathbf{\Omega}\mathcal{B}(s_0,s_0)^{-1}$, with $\mathcal{B}(s_0,s_0) = \mathbf{\Omega}^{\mathrm{T}}\mathbf{C}_{\kappa_0,11}^{-1}(s_0,s_0)\mathbf{\Omega}$.

S6 Discussion on the choice of IV

Let $\mathfrak{M}^*(\mathbf{X}) = \mathbf{X}/\sigma^2(\mathbf{X})$ and $s_a = \mathbb{E}\{\sigma^{2a}(\mathbf{X})\}$ for $a = 1,2$. Thus, based on the $\mathfrak{M}^*(\mathbf{X})$, $\mathbf{\Omega} = \mathbb{E}\{\mathfrak{M}^*(\mathbf{X})\mathbf{X}^{\mathrm{T}}\} = \mathbf{\Omega}_{\mathbf{x}}/s_1\{1+o(1)\}$, $\mathcal{B}^*(s_0,s_0) = \mathbf{\Omega}_{\mathbf{x}}\mathbf{C}_{\kappa_0,11}^{-1}(s_0,s_0)\mathbf{\Omega}_{\mathbf{x}}/s_1^2\{1+o(1)\}$ where $\mathbf{\Omega}_{\mathbf{x}} = E(\mathbf{X}\mathbf{X}^{\mathrm{T}})$. Then, the asymptotic variance under the choice of $\mathfrak{M}^*$ is

$$\mathcal{A}^*(s_0,s_0)$$

$$= s_1^{-2}s_2^{-1}\mathcal{B}^{*-1}(s_0,s_0)\mathbf{\Omega}_{\mathbf{x}}^{\mathrm{T}}\mathbf{C}_{\kappa_0,11}^{-1}(s_0,s_0)\mathbf{\Sigma}^*_{\mathbf{x}}(s_0,s_0)\mathbf{C}_{\kappa_0,11}^{-1}(s_0,s_0)\mathbf{\Omega}_{\mathbf{x}}\mathcal{B}^{*-1}(s_0,s_0)$$

$$= s_2^{-1}\mathbf{\Omega}_{\mathbf{x}}^{-1}\mathbf{\Sigma}^*_{\mathbf{x}}(s_0,s_0)\mathbf{\Omega}_{\mathbf{x}}^{-1}.$$

Now, define,

$$\mathfrak{D}_i(s_0) = \boldsymbol{\Omega}^{\mathrm{T}}\mathbf{C}_{\kappa_0,11}(s_0,s_0)^{-1}\left\{r^{-1}\sum_{j=1}^{r}K_h(s_j-s_0)\mathfrak{M}(\mathbf{X})U_{ij}\right\} \text{ and}$$

$$\mathfrak{D}_i^*(s_0) = r^{-1}\sum_{j=1}^{r}K_h(s_j-s_0)\mathbf{X}U_{ij}/\sigma^2(\mathbf{X}).$$

Therefore, by some calculation, it is not difficult to show the following:

$$\mathbb{E}\{\mathfrak{D}_i(s_0)\mathfrak{D}_i(s_0)^{\mathrm{T}}\} = \boldsymbol{\Omega}^{\mathrm{T}}\mathbf{C}_{\kappa_0,11}(s_0,s_0)^{-1}\boldsymbol{\Sigma}^*_{\mathfrak{M}}(s_0,s_0)\mathbf{C}_{\kappa_0,11}(s_0,s_0)^{-1}\boldsymbol{\Omega}\{1+o(1)\};$$

$$\mathbb{E}\{\mathfrak{D}_i(s_0)\mathfrak{D}_i^*(s_0)^{\mathrm{T}}\} = \boldsymbol{\Omega}^{\mathrm{T}}\mathbf{C}_{\kappa_0,11}(s_0,s_0)^{-1}\boldsymbol{\Omega}/s_1\{1+o(1)\};$$

$$\mathbb{E}\{\mathfrak{D}_i^*(s_0)\mathfrak{D}_i^*(s_0)^{\mathrm{T}}\} = \boldsymbol{\Omega}_{\mathbf{x}}/s_2\{1+o(1)\}.$$

Therefore,

$$\begin{aligned}
&\mathcal{A}(s_0,s_0) - \mathcal{A}^*(s_0,s_0)\\
&= \mathcal{B}(s_0,s_0)^{-1}\boldsymbol{\Omega}^{\mathrm{T}}\mathbf{C}_{\kappa_0,11}(s_0,s_0)^{-1}\boldsymbol{\Sigma}^*_{\mathfrak{M}}(s_0,s_0)\mathbf{C}_{\kappa_0,11}(s_0,s_0)^{-1}\boldsymbol{\Omega}\mathcal{B}(s_0,s_0)^{-1}\\
&\qquad - s_2^{-1}\boldsymbol{\Omega}_{\mathbf{x}}^{-1}\boldsymbol{\Sigma}^*_{\mathbf{x}}(s_0,s_0)\boldsymbol{\Omega}_{\mathbf{x}}^{-1}\\
&= [\mathbb{E}\{\mathfrak{D}(s_0,s_0)\mathfrak{D}^*(s_0,s_0)^{\mathrm{T}}\}]^{-1}\\
&\qquad \times \big(\mathbb{E}\{\mathfrak{D}(s_0,s_0)\mathfrak{D}(s_0,s_0)^{\mathrm{T}}\}\\
&\qquad - \mathbb{E}\{\mathfrak{D}(s_0,s_0)\mathfrak{D}^*(s_0,s_0)^{\mathrm{T}}\}[\mathbb{E}\{\mathfrak{D}^*(s_0,s_0)\mathfrak{D}^*(s_0,s_0)^{\mathrm{T}}\}]^{-1}\mathbb{E}\{\mathfrak{D}^*(s_0,s_0)\mathfrak{D}(s_0,s_0)^{\mathrm{T}}\}\big)\\
&\qquad [\mathbb{E}\{\mathfrak{D}^*(s_0,s_0)\mathfrak{D}(s_0,s_0)^{\mathrm{T}}\}]^{-1}\\
&= \mathbb{E}\{\mathfrak{R}(s_0,s_0)\mathfrak{R}^{\mathrm{T}}(s_0,s_0)\} \geq 0,
\end{aligned}$$

where

$$\mathfrak{R}(s_0, s_0) = [\mathbb{E}\{\mathfrak{D}(s_0, s_0)\mathfrak{D}^*(s_0, s_0)^{\mathrm{T}}\}]^{-1}\Big\{\mathfrak{D}(s_0, s_0) - \mathbb{E}\{\mathfrak{D}(s_0, s_0)\mathfrak{D}^*(s_0, s_0)^{\mathrm{T}}\}[\mathbb{E}\{\mathfrak{D}^*(s_0, s_0)\mathfrak{D}^*(s_0, s_0)^{\mathrm{T}}\}]^{-1}\mathfrak{D}^*(s_0, s_0)\Big\}.$$

Thus, the chosen IV estimator is optimal among the class of all local linear GMM estimators of the varying coefficient model.

References

Amato, U., A. Antoniadis, and I. D. Feis (2020). Flexible, boundary adapted, nonparametric methods for the estimation of univariate piecewise-smooth functions. *Statistics Surveys 14*(none), 32 – 70.

Antoniadis, A. (2007). Wavelet methods in statistics: some recent developments and their applications. *Statistics Surveys 1*(none), 16 – 55.

Cardot, H., D. Degras, and E. Josserand (2013). Confidence bands for horvitz–thompson estimators using sampled noisy functional data. *Bernoulli 19*(5A), 2067–2097.

Chen, K., K. Chen, H.-G. Müller, and J.-L. Wang (2011). Stringing high-dimensional data for functional analysis. *Journal of the American Statistical Association 106*(493), 275–284.

Chen, S. X., L.-X. Zhang, and P.-S. Zhong (2010). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association 105*(490), 810–819.

Hadinejad-Mahram, H., D. Dahlhaus, and D. Blomker (2002). Karhunen-loéve expansion of vector random processes. Technical report, Technical Report No. IKTNT 1019, Communications Technological Laboratory.

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.

Hardle, W. and E. Mammen (1993). Comparing Nonparametric Versus Parametric Regression Fits. *The Annals of Statistics 21*(4), 1926 – 1947.

Li, Y. and T. Hsing (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics 38*(6), 3321–3351.

Vaart, A. W. and J. A. Wellner (1996). Weak convergence. In *Weak convergence and empirical processes*. Springer.

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

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.

Xiao, L., V. Zipunnikov, D. Ruppert, and C. Crainiceanu (2014). Fast covariance estimation for high-dimensional functional data. *Stat Comput. 26*(1), 409–421.

Zhang, J.-T. and J. Chen (2007). Statistical inferences for functional data. *The Annals of Statistics 35*(3), 1052–1079.

Zhong, P.-S. (2023). Homogeneity tests of covariance for high-dimensional functional data with applications to event segmentation. *Biometrics 79*(4), 3332–3344.

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.

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