

SFB 649 Discussion Paper 2014-002

A Simultaneous Confidence Corridor for Varying Coefficient Regression with Sparse Functional Data

Lijie Gu*
Li Wang**
Wolfgang Karl Härdle***
Lijian Yang*/****



* Soochow University, China
** University of Georgia, USA
*** Humboldt-Universität zu Berlin, Germany
**** Michigan State University, USA

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

<http://sfb649.wiwi.hu-berlin.de>
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin



SFB 649 ECONOMIC RISK BERLIN

A Simultaneous Confidence Corridor for Varying
Coefficient Regression with Sparse Functional Data

Lijie GU

Center for Advanced Statistics and Econometrics Research

Soochow University

Suzhou 215006, China email: gulijie@suda.edu.cn

Li WANG

Department of Statistics

University of Georgia

Athens, GA 30602 email: lilywang@uga.edu

Wolfgang K. HÄRDLE

C.A.S.E. – Center for Applied Statistics and Economics

Humboldt-Universität zu Berlin

Unter den Linden 6

10099 Berlin, Germany email: haerdle@wiwi.hu-berlin.de

and

Lee Kong Chian School of Business, Singapore Management University

Lijian YANG

Center for Advanced Statistics and Econometrics Research

Soochow University

Suzhou 215006, China email: yanglijian@suda.edu.cn

and

Department of Statistics and Probability

Michigan State University

East Lansing, MI 48824 email: yang@stt.msu.edu

Author's Footnote:

Lijie Gu is Ph.D. student, Center for Advanced Statistics and Econometrics Research, Soochow University, Suzhou 215006, China (E-mail: gulijie@suda.edu.cn). Li Wang is Associate Professor, Department of Statistics, University of Georgia, Athens, GA 30602 (E-mail: lilywang@uga.edu). Wolfgang K. Härdle is Professor, C.A.S.E. – Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany, and Distinguished Visiting Professor, Lee Kong Chian School of Business, Singapore Management University (E-mail: haerdle@wiwi.hu-berlin.de). Lijian Yang is Director, Center for Advanced Statistics and Econometrics Research, Soochow University, Suzhou 215006, China, and Professor, Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824 (E-mail: yanglijian@suda.edu.cn; yang@stt.msu.edu). This work is supported in part by the Deutsche Forschungsgemeinschaft through the CRC 649 “Economic Risk”, the US National Science Foundation awards DMS 0905730, 1007594, 1106816, 1309800, Jiangsu Specially-Appointed Professor Program SR10700111, Jiangsu Province Key-Discipline Program (Statistics) ZY107002, National Natural Science Foundation of China award 11371272, and Research Fund for the Doctoral Program of Higher Education of China award 20133201110002.

Abstract

We consider a varying coefficient regression model for sparse functional data, with time varying response variable depending linearly on some time independent covariates with coefficients as functions of time dependent covariates. Based on spline smoothing, we propose data driven simultaneous confidence corridors for the coefficient functions with asymptotically correct confidence level. Such confidence corridors are useful benchmarks for statistical inference on the global shapes of coefficient functions under any hypotheses. Simulation experiments corroborate with the theoretical results. An example in CD4/HIV study is used to illustrate how inference is made with computable p -values on the effects of smoking, pre-infection CD4 cell percentage and age on the CD4 cell percentage of HIV infected patients under treatment.

KEYWORDS: B spline, confidence corridor, Karhunen-Loève L^2 representation, knots, functional data, varying coefficient.

JEL Classification: C14, C23

1. INTRODUCTION

Functional data are commonly encountered in biomedical studies, epidemiology and social science, where information is collected over a time period for each subject. In many longitudinal studies, repeated measurements are often collected at few irregular time points. Data of this type are frequently referred to as sparse longitudinal or sparse functional data. See, for example, James, Hastie and Sugar (2000), James and Sugar (2003), Yao, Müller and Wang (2005a), Hall, Müller and Wang (2006), and Zhou, Huang and Carroll (2008).

In longitudinal study, often, interest lies in studying the association between the covariates and the response variable. In recent years, there has been an increasing interest in nonparametric analysis of longitudinal data to enhance flexibility, see e.g., Yao and Li (2013). The varying coefficient model (VCM) proposed by Hastie and Tibshirani (1993) strikes a delicate balance between the simplicity of linear regression and the flexibility of multivariate nonparametric regression and has been widely applied in various settings, for instance, the Cobb-Douglas model for GDP growth in Liu and Yang (2010), and the longitudinal model for CD4 cell percentages in AIDS patients in Wu and Chiang (2000), Fan and Zhang (2000) and Wang, Li and Huang (2008). See Fan and Zhang (2008) for an extensive literature review of VCM.

To examine whether the association changes over time, Hoover et al. (1998) proposed the following varying coefficient model

$$Y(t) = \beta_0(t) + \mathbf{X}(t)^\top \boldsymbol{\beta}(t) + \varepsilon(t), \quad t \in \mathcal{T}, \quad (1)$$

where $\mathbf{X}(t) = (X_1(t), \dots, X_d(t))^\top$ are covariates at time t , $\boldsymbol{\beta}(t) = (\beta_1(t), \dots, \beta_d(t))^\top$ are functions of t , and $\varepsilon(t)$ is a mean zero process. Model (1) is a special case of functional linear models, see Ramsay and Silverman (2005) and Wu, Fan and Müller (2010).

The coefficient functions $\beta_l(t)$'s in model (1) can be estimated by, for example, kernel method in Hoover et al. (1998), basis function approximation method in Huang, Wu and Zhou (2002), polynomial spline method in Huang, Wu and Zhou (2004) and smoothing spline method in Brumback and Rice (1998). Fan and Zhang (2000) proposed a two-step method

to overcome the computational burden of the smoothing spline method.

For some longitudinal studies, the covariates are independent of time, and their observations are cross-sectional. Take for instance the longitudinal CD4 cell percentage data among HIV seroconverters. This dataset contains 1817 observations of CD4 cell percentages on 283 homosexual men infected with the HIV virus. Three of the covariates are observed at the time of HIV infection and hence by nature independent of the measurement time and frequency: X_{i1} , the i -th patient's smoking status; X_{i2} , the i -th patient's centered pre-infection CD4 percentage; and X_{i3} the i -th patient's centered age at the time of HIV infection. A fourth predictor, however, is time dependent: T_{ij} , the time (in years) of the j -th measurement of CD4 cell on the i -th patient after HIV infection; while the response Y_{ij} is also time dependent: the j -th measurement of the i -th patient's CD4 cell percentage at time T_{ij} . Wu and Chiang (2000), Fan and Zhang (2000) and Wang, Li and Huang (2008) all contain detailed descriptions and analysis of this dataset.

A feasible VCM for multivariate functional data such as the above takes the form

$$Y_{ij} = \sum_{l=1}^d \eta_{il}(T_{ij}) X_{il} + \sigma(T_{ij}) \varepsilon_{ij}, \quad 1 \leq i \leq n, \quad 1 \leq j \leq N_i, \quad (2)$$

where the measurement errors $(\varepsilon_{ij})_{i=1, j=1}^{n, N_i}$ satisfy $\mathbf{E}(\varepsilon_{ij}) = 0$, $\mathbf{E}(\varepsilon_{ij}^2) = 1$, and $\{\eta_{il}(t), t \in \mathcal{T}\}$ are i.i.d copies of a L^2 process $\{\eta_l(t), t \in \mathcal{T}\}$, i.e., $\mathbf{E} \int_{\mathcal{T}} \eta_l^2(t) dt < +\infty$, $l = 1, \dots, d$. The common mean function of processes $\{\eta_l(t), t \in \mathcal{T}\}$ is denoted as $m_l(t) = \mathbf{E}\{\eta_l(t)\}$, $l = 1, \dots, d$. The actual data set consists of $\{\mathbf{X}_i, T_{ij}, Y_{ij}\}$, $1 \leq i \leq n$, $1 \leq j \leq N_i$, in which the i -th subject is observed N_i times, the time independent covariates for the i -th subject are $\mathbf{X}_i = (X_{il})_{l=1}^d$, $1 \leq i \leq n$, and the random measurement time $T_{ij} \in \mathcal{T} = [a, b]$. The aforementioned data example is called sparse functional as the number of measurements N_i for the i -th subject is relatively low. (In the above CD4 example actually at most 14). In contrast, for a dense functional data $\lim_{n \rightarrow \infty} \min_{1 \leq i \leq n} N_i = \infty$.

For the CD4 cell percentage data, we introduce a fourth time independent covariate, the baseline $X_{i0} \equiv 1$, and denote by $m_l(t)$, $l = 0, 1, 2, 3$, the coefficient functions for baseline CD4 percentage, smoking status, centered pre-infection CD4 percentage and centered age,

respectively. Figures 2-5 contain spline estimates of the $m_l(t)$, $0 \leq l \leq 3$, and simultaneous confidence corridors (SCC) at various confidence levels.

In previous works the theoretical focus has mainly been on consistency and asymptotic normality of the estimators of the coefficient functions of interest, and the construction of pointwise confidence intervals. However, as demonstrated in Fan and Zhang (2000), this is unsatisfactory as investigators are often interested in testing whether some coefficient functions are significantly nonzero or varying, for which a SCC is needed. Take for instance, Figure 3, which shows both the 95% and 20.277% SCC of $m_1(t)$ contain the zero line completely, thus with a very high p -value of 0.79723 the null hypothesis of $m_1(t) \equiv 0, t \in \mathcal{T}$ is not rejected. More details are in Section 6.

Construction of computationally simple SCCs with exact coverage probability is known to be difficult even with independent cross-sectional data; see, Wang and Yang (2009) and related earlier work Härdle and Luckhaus (1984) on uniform consistency. Most earlier methods proposed in the literature restrict to asymptotic conservative SCCs. Wu, Chiang and Hoover (1998) developed asymptotic SCCs for the unknown coefficients based on local polynomial methods, which are computationally intensive, as the kernel estimator requires solving an optimization problem at every point. Huang, Wu and Zhou (2004) proposed approximating each coefficient function by a polynomial spline and developed spline SCCs, which are simpler to construct, while Xue and Zhu (2007) proposed maximum empirical likelihood estimators and constructed SCCs for the coefficient functions. All these SCCs are Bonferroni-type variability bands according to Hall and Titterton (1988). The idea is to invoke pointwise confidence intervals on a very fine grid of $[a, b]$, then adjust the level of these confidence intervals by the Bonferroni method to obtain uniform confidence bands, and finally bridge the gaps between the grid points via smoothness conditions on the coefficient curve. However, to use these bands in practice, one must have a priori bounds on the magnitude of the bias on each subinterval as well as a choice for the number of grid points. Chiang, Rice and Wu (2001) proposed a bootstrap procedure to construct confidence intervals. However, theoretical properties of their procedures have not yet been developed.

In this paper, we derive SCCs with exact coverage probability for the coefficient functions $m_l(t)$, $l = 1, \dots, d$, in (3) via extreme value theory of Gaussian processes and approximating coefficient functions by piecewise-constant splines. The results represent the first attempt at developing exact SCCs for the coefficient functions in VCM for sparse functional data. Our simulation studies indicate the proposed SCCs are computationally efficient and have the right coverage probability for finite samples. Our work parallels Zhu, Li and Kong (2012) which established asymptotic theory of SCC in the case of VCM for dense functional data. It is important to mention as well that the linear covariates in Zhu, Li and Kong (2012) are time dependent, which does not complicate the problem as they work with dense data instead of the sparse data we concentrate on.

We organize our paper as follows. Section 2 describes spline estimators, and establish their asymptotic properties for sparse longitudinal data. Section 3.1 proposes asymptotic pointwise confidence intervals and SCCs constructed from piecewise constant splines. Section 3.2 describes actual steps to implement the proposed SCCs. In Section 4 we provide further insights into the estimation error structure of spline estimators. Section 5 reports findings from a simulation study. A real data example appears in Section 6. Proofs of technical lemmas are in the Appendix and Supplementary Materials.

2. SPLINE ESTIMATION AND ASYMPTOTIC PROPERTIES

For a functional data $\{\mathbf{X}_i, T_{ij}, Y_{ij}\}$, $1 \leq i \leq n$, $1 \leq j \leq N_i$, denote the eigenvalues and eigenfunctions sequences of its covariance operator $G_l(s, t) = \text{cov}\{\eta_l(s), \eta_l(t)\}$ as $\{\lambda_{k,l}\}_{k=1}^\infty$, $\{\psi_{k,l}(t)\}_{k=1}^\infty$, in which $\lambda_{1,l} \geq \lambda_{2,l} \geq \dots \geq 0$, $\sum_{k=1}^\infty \lambda_{k,l} < \infty$, and $\{\psi_{k,l}\}_{k=1}^\infty$ form an orthonormal basis of $L^2(\mathcal{T})$. It follows from spectral theory that $G_l(s, t) = \sum_{k=1}^\infty \lambda_{k,l} \psi_{k,l}(s) \psi_{k,l}(t)$. For any $l = 1, \dots, d$, the i -th trajectory $\{\eta_{il}(t), t \in \mathcal{T}\}$ allows the Karhunen-Loève L^2 representation (Yao, Müller and Wang, 2005b): $\eta_{il}(t) = m_l(t) + \sum_{k=1}^\infty \xi_{ik,l} \phi_{k,l}(t)$, where the random coefficients $\xi_{ik,l}$ are uncorrelated with mean 0 and variances 1, and the functions $\phi_{k,l} = \sqrt{\lambda_{k,l}} \psi_{k,l}$, thus $G_l(s, t) = \sum_{k=1}^\infty \phi_{k,l}(s) \phi_{k,l}(t)$, and the response measurements (2) can

be represented as follows

$$Y_{ij} = \sum_{l=1}^d m_l(T_{ij}) X_{il} + \sum_{l=1}^d \sum_{k=1}^{\infty} \xi_{ik,l} \phi_{k,l}(T_{ij}) X_{il} + \sigma(T_{ij}) \varepsilon_{ij}. \quad (3)$$

Without loss of generality, we take $\mathcal{T} = [a, b]$ to be $[0, 1]$. Following Xue and Yang (2006), we approximate each coefficient function by the spline smoothing method. To describe the spline functions, one can divide the finite interval $[0, 1]$ into $(N_s + 1)$ equal subintervals $\chi_J = [v_J, v_{J+1})$, $J = 0, \dots, N_s - 1$, $\chi_{N_s} = [v_{N_s}, 1]$. A sequence of equally-spaced points $\{v_J\}_{J=1}^{N_s}$, called interior knots, are given as $v_0 = 0 < v_1 < \dots < v_{N_s} < 1 = v_{N_s+1}$. Let $v_J = Jh_s$ for $0 \leq J \leq N_s + 1$, where $h_s = 1/(N_s + 1)$ is the distance between neighboring knots. We denote by $G^{(-1)} = G^{(-1)}[0, 1]$ the space of functions that are constant on each subinterval χ_J , and the B-spline basis of $G^{(-1)}$, as $\{b_J(t)\}_{J=0}^{N_s}$, which are simply indicator functions of intervals χ_J , $b_J(t) = I_{\chi_J}(t)$, $J = 0, 1, \dots, N_s$. For any $t \in [0, 1]$, define its location index as $J(t) = J_n(t) = \min\{\lceil t/h_s \rceil, N_s\}$ so that $t \in \chi_{J(t)}$.

Next we define the space of spline coefficient functions on $\mathcal{T} \times \mathbb{R}^d$ as

$$\mathcal{M} = \left\{ g(t, \mathbf{x}) = \sum_{l=1}^d g_l(t) x_l : g_l(t) \in G^{(-1)}, t \in \mathcal{T}, \mathbf{x} = (x_1, \dots, x_d)^\top \in \mathbb{R}^d \right\},$$

and propose estimating the multivariate function $\sum_{l=1}^d m_l(t) x_l$ by

$$\hat{m}(t, \mathbf{x}) = \sum_{l=1}^d \hat{m}_l(t) x_l = \underset{g \in \mathcal{M}}{\operatorname{argmin}} \sum_{i=1}^n \sum_{j=1}^{N_i} \{Y_{ij} - g(T_{ij}, \mathbf{X}_i)\}^2. \quad (4)$$

Let $\sigma_Y^2(t, \mathbf{x})$ be the conditional variance of \mathbf{Y} given $T = t$ and $\mathbf{X} = \mathbf{x} = (x_1, \dots, x_d)^\top \in \mathbb{R}^d$

$$\sigma_Y^2(t, \mathbf{x}) = \operatorname{Var}(Y | T = t, \mathbf{X} = \mathbf{x}) = \sum_{l=1}^d G_l(t, t) x_l^2 + \sigma^2(t).$$

Next for any $t \in [0, 1]$, let

$$\begin{aligned} \mathbf{\Gamma}_n(t) &= c_{J(t),n}^{-2} \{n \mathbb{E}(N_1)\}^{-1} \mathbb{E} \mathbf{X} \mathbf{X}^\top \left[\int_{\chi_{J(t)}} \sigma_Y^2(u, \mathbf{X}) f(u) du \right. \\ &\quad \left. + \frac{\mathbb{E}\{N_1(N_1 - 1)\}}{\mathbb{E} N_1} \sum_{l=1}^d X_l^2 \int_{\chi_{J(t)} \times \chi_{J(t)}} G_l(u, v) f(u) f(v) dudv \right], \end{aligned} \quad (5)$$

where

$$c_{J,n} = \mathbf{E} b_J^2(T) = \int_0^1 b_J^2(t) f(t) dt, \quad J = 0, \dots, N_s. \quad (6)$$

Further denote

$$\boldsymbol{\Sigma}_n(t) = \mathbf{H}^{-1} \boldsymbol{\Gamma}_n(t) \mathbf{H}^{-1} = \{\sigma_{n,l'l'}^2(t)\}_{l,l'=1}^d, \quad (7)$$

where $\sigma_{n,l'l'}^2(t)$ are later shown to be the asymptotic covariances between $\hat{m}_l(t)$ and $\hat{m}_{l'}(t)$.

THEOREM 1. *Under Assumptions (A1)–(A6) in Appendix A, for any $t \in [0, 1]$, as $n \rightarrow \infty$,*

$$\boldsymbol{\Sigma}_n^{-1/2}(t) \{\hat{\mathbf{m}}(t) - \mathbf{m}(t)\} \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}_{d \times d}),$$

where $\hat{\mathbf{m}}(t) = (\hat{m}_1(t), \dots, \hat{m}_d(t))^\top$ is the estimate of $\mathbf{m}(t) = (m_1(t), \dots, m_d(t))^\top$. Furthermore, for any $l = 1, \dots, d$ and $\alpha \in (0, 1)$,

$$\lim_{n \rightarrow \infty} P \{\sigma_{n,ll}^{-1}(t) |\hat{m}_l(t) - m_l(t)| \leq Z_{1-\alpha/2}\} = 1 - \alpha.$$

REMARK 1. *Note that $\boldsymbol{\Sigma}_n(t) = \{\sigma_{n,l'l'}^2(t)\}_{l,l'=1}^d$ in (7) is complicated to compute in practice. The next proposition suggests that, for any $t \in [0, 1]$, $\boldsymbol{\Gamma}_n(t)$ in (5) can be simplified by*

$$\tilde{\boldsymbol{\Gamma}}_n(t) \equiv \mathbf{E} \left[\mathbf{X} \mathbf{X}^\top \frac{\sigma_Y^2(t, \mathbf{X})}{f(t) h_s n \mathbf{E}(N_1)} \left\{ 1 + \frac{\mathbf{E} N_1 (N_1 - 1) \sum_{l=1}^d X_l^2 G_l(t, t) f(t) h_s}{\mathbf{E} N_1 \sigma_Y^2(t, \mathbf{X})} \right\} \right]. \quad (8)$$

Denote the supremum norm of a function ϕ on $[a, b]$ by $\|\phi\|_\infty = \sup_{t \in [a, b]} |\phi(t)|$. For any matrix $\mathbf{A} = (a_{ij})$, define $\|\mathbf{A}\|_\infty = \max |a_{ij}|$, where the maximum is taken over all the elements of \mathbf{A} , while for a matrix function $\mathbf{A}(t) = (a_{ij}(t))$, $\|\mathbf{A}\|_\infty = \sup_{t \in [a, b]} \|\mathbf{A}(t)\|_\infty$.

PROPOSITION 1. *Under Assumptions (A2)–(A6) in Appendix A, there exists a constant $c > 0$ such that as $n \rightarrow \infty$, $\|\boldsymbol{\Gamma}_n(t) - \tilde{\boldsymbol{\Gamma}}_n(t)\|_\infty = \mathcal{O}(n^{-1} h_s^{r-1}) = \mathcal{O}(n^{-c})$.*

To derive the maximal deviation distribution of estimators $\hat{m}_l(t)$, $l = 1, \dots, d$, let

$$Q_{N_s+1}(\alpha) = b_{N_s+1} - a_{N_s+1}^{-1} \log \left\{ -\frac{1}{2} \log(1 - \alpha) \right\}, \quad \alpha \in (0, 1) \quad (9)$$

$$a_{N_s+1} = \{2 \log(N_s + 1)\}^{1/2}, \quad b_{N_s+1} = a_{N_s+1} - \frac{\log(2\pi a_{N_s+1}^2)}{2a_{N_s+1}}. \quad (10)$$

THEOREM 2. *Under Assumptions (A1)–(A6) in Appendix A, for $l = 1, \dots, d$ and any $\alpha \in (0, 1)$,*

$$\lim_{n \rightarrow \infty} P \left\{ \sup_{t \in [0,1]} \sigma_{n,l}^{-1}(t) |\hat{m}_l(t) - m_l(t)| \leq Q_{N_s+1}(\alpha) \right\} = 1 - \alpha,$$

where $\sigma_{n,l}(t)$ and $Q_{N_s+1}(\alpha)$ are given in (7) and (9), respectively.

3. ASYMPTOTIC CONFIDENCE REGIONS

In this section we construct the confidence regions for functions $m_l(t)$, $l = 1, \dots, d$.

3.1 Asymptotic Confidence Intervals and SCCs

Theorems 1 and 2 allow one to construct pointwise confidence intervals and SCCs for components $\hat{m}_l(t)$, $l = 1, \dots, d$. The next corollary provides the theoretical underpinning upon which SCCs can be actually implemented, see subsection 3.2.

COROLLARY 1. *Under Assumptions (A1)–(A6) in Appendix A, for any $l = 1, \dots, d$ and $\alpha \in (0, 1)$, as $n \rightarrow \infty$,*

(i) *an asymptotic $100(1 - \alpha)\%$ pointwise confidence interval for $m_l(t)$, $t \in [0, 1]$, is $\hat{m}_l(t) \pm \sigma_{n,l}(t)Z_{1-\alpha/2}$, with $\sigma_{n,l}(t)$ given in (7), while $Z_{1-\alpha/2}$ is the $100(1 - \alpha/2)^{\text{th}}$ percentile of the standard normal distribution.*

(ii) *an asymptotic $100(1 - \alpha)\%$ SCC for $m_l(t)$, with $Q_{N_s+1}(\alpha)$ given in (9), is $\hat{m}_l(t) \pm \sigma_{n,l}(t)Q_{N_s+1}(\alpha)$, $t \in [0, 1]$.*

3.2 Implementation

In the following we describe procedures to construct the SCCs and the pointwise intervals given in Corollary 1. For any data set $(T_{ij}, Y_{ij}, X_{il})_{i=1, j=1, l=1}^{n, N_i, d}$ from model (3), the spline estimators $\hat{m}_l(t)$, $l = 1, \dots, d$, are obtained by (4), and the number of interior knots is taken to be $N_s = \lceil cN_T^{1/3}(\log(n)) \rceil$, in which $N_T = \sum_{i=1}^n N_i$ is the total sample size, $\lceil a \rceil$ denotes the integer part of a , and c is a positive constant.

To construct the SCCs, one needs to evaluate the functions $\sigma_{n,l}^2(t)$, $l = 1, \dots, d$, which are the diagonal elements of matrix $\Sigma_n(t)$ in (7). Based on Proposition 1, one can estimate each unknowns $f(t)$, $\sigma_Y^2(t, \mathbf{x})$, $G_l(t, t)$ and matrix \mathbf{H} and then plug these estimators into the formula of the SCCs; see Wang and Yang (2009).

The number of interior knots for pilot estimation of $f(t)$, $\sigma_Y^2(t, \mathbf{x})$, and $G_l(t, t)$ is taken to be $N_s^* = \lceil 0.5n^{1/3} \rceil$, and $h_s^* = 1/(1 + N_s^*)$. The histogram pilot estimator of the density function $f(t)$ is $\hat{f}(t) = N_T^{-1} h_s^{*-1} \sum_{i=1}^n \sum_{j=1}^{N_i} b_{J(t)}(T_{ij})$.

We now discuss the estimation of $\Gamma_n(t)$ in (5). Defining $R_{ij} = \left(Y_{ij} - \sum_{l=1}^d \hat{m}(T_{ij}) X_{il} \right)^2$, $1 \leq j \leq N_i$, $1 \leq i \leq n$, the estimator of $\sigma_Y^2(t, \mathbf{x})$ is

$$\hat{\sigma}_Y^2(t, \mathbf{x}) = \sum_{l=1}^d \sum_{J=0}^{N_s^*} \hat{\rho}_{J,l} b_J(t) x_l^2 + \sum_{J=0}^{N_s^*} \hat{\mu}_J b_J(t) = \sum_{l=1}^d \hat{G}_l(t, t) x_l^2 + \hat{\sigma}^2(t),$$

where $\{\hat{\rho}_{0,1}, \dots, \hat{\rho}_{N_s^*,d}, \hat{\mu}_0, \dots, \hat{\mu}_{N_s^*}\}^\top$ are solutions of the following least squares problem:

$$\begin{aligned} & (\hat{\rho}_{0,1}, \dots, \hat{\rho}_{N_s^*,d}, \hat{\mu}_0, \dots, \hat{\mu}_{N_s^*})^\top = \\ & \underset{(\rho_{0,1}, \dots, \mu_{N_s^*})^\top \in \mathbb{R}^{(N_s^*+1)(d+1)}}{\operatorname{argmin}} \sum_{i=1}^n \sum_{j=1}^{N_i} \left\{ R_{ij} - \sum_{l=1}^d \sum_{J=0}^{N_s^*} \rho_{J,l} b_J(T_{ij}) X_{il}^2 - \sum_{J=0}^{N_s^*} \mu_J b_J(T_{ij}) \right\}^2. \end{aligned}$$

The matrix $\Gamma_n(t)$ is estimated by substituting $f(t)$, $G_l(t, t)$ and $\sigma_Y^2(t, \mathbf{x})$ with $\hat{f}(t)$, $\hat{G}_l(t, t)$ and $\hat{\sigma}_Y^2(t, \mathbf{x})$. Define

$$\begin{aligned} \hat{\Gamma}_n(t) & \equiv \left[n^{-1} \sum_{i=1}^n X_{il} X_{il'} \hat{\sigma}_Y^2(t, \mathbf{X}_i) \left\{ \hat{f}(t) h_s N_T \right\}^{-1} \right. \\ & \quad \left. \times \left\{ 1 + \left(\frac{\sum_{i=1}^n N_i^2}{N_T} - 1 \right) \frac{\sum_{l=1}^d \hat{G}_l(t, t) X_{il}^2}{\hat{\sigma}_Y^2(t, \mathbf{X}_i)} \hat{f}(t) h_s \right\} \right]_{l,l'=1}^d. \end{aligned}$$

The following proposition provides the consistent rate of $\hat{\Gamma}_n(t)$ to $\Gamma_n(t)$.

PROPOSITION 2. *Under Assumptions (A1)-(A6) in Appendix A, there exists a constant $c > 0$ such that as $n \rightarrow \infty$, $\left\| \hat{\Gamma}_n(t) - \Gamma_n(t) \right\|_\infty = \mathcal{O}_p(n^{-c})$.*

Proposition 2 implies that $\Gamma_n(t)$ can be replaced by $\hat{\Gamma}_n(t)$ with a negligible error. Define a $d \times d$ matrix $\hat{\mathbf{H}} = \{n^{-1} \sum_{i=1}^n X_{il} X_{il'}\}_{l,l'=1}^d$, then $\Sigma_n(t)$ can be estimated well by $\hat{\Sigma}_n(t) =$

$\{\hat{\sigma}_{n,l,l'}^2(t)\}_{l,l'=1}^d = \hat{\mathbf{H}}^{-1} \hat{\mathbf{\Gamma}}_n(t) \hat{\mathbf{H}}^{-1}$. Therefore, as $n \rightarrow \infty$, $l = 1, \dots, d$, the SCCs

$$\hat{m}_l(t) \pm \hat{\sigma}_{n,l}(t) Q_{N_s+1}(\alpha), \quad (11)$$

with $Q_{N_s+1}(\alpha)$ given in (9), and the pointwise intervals $\hat{m}_l(t) \pm \hat{\sigma}_{n,l}(t) Z_{1-\alpha/2}$ have asymptotic confidence level $1 - \alpha$.

4. DECOMPOSITION

In this section, we describe the representation of the spline estimators $\hat{m}_l(t)$, $l = 1, \dots, d$, in (4), then break the estimation error $\hat{m}_l(t) - m_l(t)$ into three terms by the decomposition of Y_{ij} in model (3). Although such representation is not needed for applying the procedure describe in Section 3.2 to analyze data, it sheds insights into the proof of the main theoretical results in Section 2.

We consider the following rescaled B-spline basis $\{B_J(t)\}_{J=0}^{N_s}$ for $G^{(-1)}$:

$$B_J(t) \equiv b_J(t) (c_{J,n})^{-1/2}, \quad J = 0, \dots, N_s. \quad (12)$$

It is easily verified that $\mathbf{E}\{B_J(T)\}^2 = 1$ for $J = 0, 1, \dots, N_s$, and $B_J(t)B_{J'}(t) \equiv 0$ for $J \neq J'$.

By simple linear algebra, the spline estimator $\hat{m}_l(t)$ defined in (4) equals

$$\hat{m}_l(t) = \sum_{J=0}^{N_s} \hat{\gamma}_{J,l} B_J(t) = c_{J(t),n}^{-1/2} \hat{\gamma}_{J(t),l}, \quad l = 1, \dots, d, \quad (13)$$

where the coefficients $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\gamma}}_0^\top, \dots, \hat{\boldsymbol{\gamma}}_{N_s}^\top)^\top$ with $\hat{\boldsymbol{\gamma}}_J = (\hat{\gamma}_{J,1}, \dots, \hat{\gamma}_{J,d})^\top$ being the solution of the following least squares problem

$$\hat{\boldsymbol{\gamma}} = \underset{\boldsymbol{\gamma} = (\gamma_{0,1}, \dots, \gamma_{N_s,d})^\top \in \mathbb{R}^{d(N_s+1)}}{\operatorname{argmin}} \sum_{i=1}^n \sum_{j=1}^{N_i} \left\{ Y_{ij} - \sum_{l=1}^d \sum_{J=0}^{N_s} \gamma_{J,l} B_J(T_{ij}) X_{il} \right\}^2. \quad (14)$$

In the following let $\mathbf{Y} = (Y_{11}, \dots, Y_{1N_1}, \dots, Y_{n1}, \dots, Y_{nN_n})^\top$ be the collection of all the Y_{ij} 's. Let $\mathbf{B}(t) = (B_0(t), \dots, B_{N_s}(t))^\top$ and $\mathbf{X}_i = (X_{i1}, \dots, X_{id})^\top$ be two vectors of dimension $(N_s + 1)$ and d , respectively. Denote

$$\mathbf{D} = (\mathbf{B}(T_{11}) \otimes \mathbf{X}_1, \dots, \mathbf{B}(T_{1N_1}) \otimes \mathbf{X}_1, \dots, \mathbf{B}(T_{n1}) \otimes \mathbf{X}_n, \dots, \mathbf{B}(T_{nN_n}) \otimes \mathbf{X}_n)^\top, \quad (15)$$

a $N_T \times ((N_s + 1)d)$ matrix, where “ \otimes ” denotes the Kronecker product. Solving the least squares problem in (14), we obtain

$$\hat{\boldsymbol{\gamma}} = (\mathbf{D}^\top \mathbf{D})^{-1} (\mathbf{D}^\top \mathbf{Y}). \quad (16)$$

Denote $\mathbf{x} = (x_1, \dots, x_d)^\top$, thus equation (4) can be rewritten as

$$\sum_{l=1}^d \hat{m}_l(t) x_l = (\mathbf{B}(t) \otimes \mathbf{x})^\top (\mathbf{D}^\top \mathbf{D})^{-1} (\mathbf{D}^\top \mathbf{Y}). \quad (17)$$

According to (15), one has $\mathbf{D}^\top \mathbf{D} = \sum_{i=1}^n \sum_{j=1}^{N_i} \{\mathbf{B}(T_{ij}) \mathbf{B}(T_{ij})^\top \otimes \mathbf{X}_i \mathbf{X}_i^\top\}$, in which matrix $\mathbf{B}(T_{ij}) \mathbf{B}(T_{ij})^\top = \text{diag} \{B_0^2(T_{ij}), \dots, B_{N_s}^2(T_{ij})\}$. So matrix $\mathbf{D}^\top \mathbf{D}$ should be a block diagonal matrix, and we write $N_T^{-1} \mathbf{D}^\top \mathbf{D} = \text{diag} \{\hat{\mathbf{V}}_0, \dots, \hat{\mathbf{V}}_{N_s}\}$, where

$$\hat{\mathbf{V}}_J = \left\{ N_T^{-1} \sum_{i=1}^n \sum_{j=1}^{N_i} B_J^2(T_{ij}) X_{il} X_{il'} \right\}_{l, l'=1}^d. \quad (18)$$

On the other hand, we have $\mathbf{D}^\top \mathbf{Y} = \sum_{i=1}^n \sum_{j=1}^{N_i} \{\mathbf{B}(T_{ij}) \otimes \mathbf{X}_i\} Y_{ij}$. Thus, $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_0^\top, \dots, \hat{\gamma}_{N_s}^\top)^\top$ can be easily calculated using

$$\hat{\boldsymbol{\gamma}}_J = \hat{\mathbf{V}}_J^{-1} \left\{ N_T^{-1} \sum_{i=1}^n \sum_{j=1}^{N_i} B_J(T_{ij}) X_{il} Y_{ij} \right\}_{l=1}^d, \quad J = 0, \dots, N_s. \quad (19)$$

Then the functions $\mathbf{m}(t) = (m_1(t), \dots, m_d(t))^\top$ can be simply estimated by

$$\hat{\mathbf{m}}(t) = (\hat{m}_1(t), \dots, \hat{m}_d(t))^\top = c_{J(t), n}^{-1/2} (\hat{\gamma}_{J(t), 1}, \dots, \hat{\gamma}_{J(t), d})^\top = c_{J(t), n}^{-1/2} \hat{\boldsymbol{\gamma}}_{J(t)}. \quad (20)$$

Projecting the relationship in model (3) onto the space of spline coefficient functions on $\mathcal{T} \times \mathbb{R}^d$ as \mathcal{M} , we obtain the following important decomposition:

$$\sum_{l=1}^d \hat{m}_l(t) x_l = \sum_{l=1}^d \tilde{m}_l(t) x_l + \sum_{l=1}^d \tilde{\xi}_l(t) x_l + \sum_{l=1}^d \tilde{\varepsilon}_l(t) x_l, \quad (21)$$

where for any $l = 1, \dots, d$,

$$\tilde{m}_l(t) = \sum_{J=0}^{N_s} \tilde{\gamma}_{J, l} B_J(t) = c_{J(t), n}^{-1/2} \tilde{\gamma}_{J(t), l}, \quad (22)$$

$$\tilde{\xi}_l(t) = \sum_{J=0}^{N_s} \tilde{\alpha}_{J,l} B_J(t) = c_{J(t),n}^{-1/2} \tilde{\alpha}_{J(t),l}, \quad \tilde{\varepsilon}_l(t) = \sum_{J=0}^{N_s} \tilde{\theta}_{J,l} B_J(t) = c_{J(t),n}^{-1/2} \tilde{\theta}_{J(t),l}, \quad (23)$$

and the vectors $(\tilde{\gamma}_{J,l}, J = 0, \dots, N_s, l = 1, \dots, d)^\top$, $(\tilde{\alpha}_{J,l}, J = 0, \dots, N_s, l = 1, \dots, d)^\top$, and $(\tilde{\theta}_{J,l}, J = 0, \dots, N_s, l = 1, \dots, d)^\top$ are solutions to (14) with Y_{ij} replaced by $\sum_{l=1}^d m_l(T_{ij}) X_{il}$, $\sum_{l=1}^d \sum_{k=1}^\infty \xi_{ik,l} \phi_{k,l}(T_{ij}) X_{il}$, and $\sigma(T_{ij}) \varepsilon_{ij}$, respectively.

Furthermore, under Assumption (A5) we can decompose $\hat{m}_l(t)$ as

$$\hat{m}_l(t) = \tilde{m}_l(t) + \tilde{\xi}_l(t) + \tilde{\varepsilon}_l(t), \quad l = 1, \dots, d. \quad (24)$$

The next two propositions concern the functions $\tilde{m}_l(t)$, $\tilde{\xi}_l(t)$, $\tilde{\varepsilon}_l(t)$, $l = 1, \dots, d$, given in (22) and (23). Proposition 3 gives the uniform convergence rate of $\tilde{m}_l(t)$ to $m_l(t)$. Proposition 4 provides the asymptotic distribution for the maximum of the normalized error terms.

PROPOSITION 3. *Under Assumptions (A1), (A2) and (A4)–(A6) in Appendix A, the functions $\tilde{m}_l(t)$, $l = 1, \dots, d$ satisfy $\sup_{t \in [0,1]} \sup_{1 \leq l \leq d} |\tilde{m}_l(t) - m_l(t)| = \mathcal{O}_p(h_s)$.*

PROPOSITION 4. *Under Assumptions (A2)–(A6) in Appendix A, for $\tau \in \mathbb{R}$, and $\sigma_{n,ul}(t)$, a_{N_s+1} , b_{N_s+1} as given in (7) and (9),*

$$\lim_{n \rightarrow \infty} P \left\{ \sup_{t \in [0,1]} \sigma_{n,ul}^{-1}(t) \left| \tilde{\xi}_l(t) + \tilde{\varepsilon}_l(t) \right| \leq \tau / a_{N_s+1} + b_{N_s+1} \right\} = \exp(-2e^{-\tau}).$$

5. SIMULATION

To illustrate the finite-sample performance of the spline approach, we generate data from the following model

$$Y_{ij} = \left\{ m_1(T_{ij}) + \sum_{k=1}^2 \xi_{ik,1} \phi_{k,1}(T_{ij}) \right\} X_{i1} + \left\{ m_2(T_{ij}) + \sum_{k=1}^3 \xi_{ik,2} \phi_{k,2}(T_{ij}) \right\} X_{i2} \\ + \sigma(T_{ij}) \varepsilon_{ij}, \quad 1 \leq i \leq n, 1 \leq j \leq N_i,$$

where $T \sim U[0, 1]$, $X_1 \sim N(0, 1)$, $X_2 \sim \text{Binomial}[1, 0.5]$, $\xi_{k,1} \sim N(0, 1)$, $k = 1, 2$, $\xi_{k,2} \sim N(0, 1)$, $k = 1, 2, 3$, $\varepsilon \sim N(0, 1)$, and N_i is generated from a discrete uniform distribution from $2, \dots, 14$, for $1 \leq i \leq n$. For the first component, we take $m_1(t) = \sin\{2\pi(t - 1/2)\}$,

$\phi_{1,1}(t) = -2 \cos \{\pi (t - 1/2)\} / \sqrt{5}$, $\phi_{2,1}(t) = \sin \{\pi (t - 1/2)\} / \sqrt{5}$, thus $\lambda_{1,1} = 2/5$, $\lambda_{2,1} = 1/10$. For the second component, we take $m_2(t) = 5(t - 0.6)^2$, $\phi_{1,2}(t) = 1$, $\phi_{2,2}(t) = \sqrt{2} \sin(2\pi t)$, $\phi_{3,2}(t) = \sqrt{2} \cos(2\pi t)$, thus $\lambda_{1,2} = \lambda_{2,2} = \lambda_{3,2} = 1$. The noise level is chosen to be $\sigma = 0.5, 1.0$, and the number of subjects n is taken to be 200, 400, 600, 800.

We consider the confidence levels $1 - \alpha = 0.95$ and 0.99 . Table 1 reports the coverage of the SCCs as the percentage out of the total 500 replications for which the true curve was covered by (11) at the 101 points $\{k/100, k = 0, \dots, 100\}$.

[Table 1 about here.]

In the above SCC construction, the number of interior knots N_s is determined by the sample size n and a tuning constant c as described in Section 3.2. We have experimented with $c = 0.3, 0.5, 0.8, 1.0$ in this simulation study. The simulation results in Table 1 reflect that the coverage percentages depend on the choice of c , however, the dependency becomes weaker when sample sizes increase. For large sample sizes $n = 600, 800$, the effect of the choice of c on the coverage percentages is insignificant. Because N_s varies with N_i , for $1 \leq i \leq n$, the data-driven selection of an ‘‘optimal’’ N_s remains an open problem. At all noise levels, the coverage percentages for the SCC (11) are very close to the nominal confidence levels 0.95 and 0.99 for $c = 0.5$. Note that since $\mathbb{E} N_1 = 8$, the total sample size $N_T \approx 8 \times 200, 8 \times 400, 8 \times 600, 8 \times 800$ which explains the closeness of coverage percentages in Table 1 to the nominal levels. These large N_T ’s are realistic as we believe they are common for real data. For instance, the CD4 cell percentage data in Section 6 has $N_T = 1817$.

For visualization of actual function estimates, Figure 1 shows the true curve, the estimated curve, the asymptotic 95% SCC and the pointwise confidence intervals at $\sigma = 0.5$ with $n = 200$. The same plot for $n = 600$ has shown significantly narrower SCC and pointwise confidence intervals as expected, but is not included to save space.

[Figure 1 about here.]

6. REAL DATA ANALYSIS

To illustrate our method, we return to the CD4 cell percentage data discussed in Section 1 for further analysis. Since the actual visit times T_{ij} are irregularly spaced and vary from year 0 to year 6, we first transform the times by $Z_{ij} = F_{N_T}(T_{ij})$, where F_{N_T} is the empirical cdf of times $\{T_{ij}\}_{i=1, j=1}^{n, N_i}$. Then the Z_{ij} -values are distributed fairly uniformly. We have set a slightly smaller number of interior knots $N_s = \lceil 0.3N_T^{1/3}(\log(n)) \rceil$ to avoid singularity in solving the least squares problem.

The left plots of Figures 2, 3, 4 and 5 depict the spline estimates, the asymptotic 95% SCCs, the pointwise confidence intervals for $m_l(t)$, $l = 0, 1, 2, 3$, respectively. The horizontal solid line represents zero. Based on the shape of the SCCs, we are interested in testing the following hypotheses:

$$H_{00} : m_0(t) \equiv a + bt, \text{ for some } a, b \in \mathbb{R} \text{ v.s. } H_{10} : m_0(t) \neq a + bt, \text{ for any } a, b \in \mathbb{R};$$

$$H_{01} : m_1(t) \equiv 0 \text{ v.s. } H_{11} : m_1(t) \neq 0, \text{ for some } t \in [0, 6];$$

$$H_{02} : m_2(t) \equiv c \text{ for some } c > 0 \text{ v.s. } H_{12} : m_2(t) \neq c, \text{ for any } c > 0;$$

$$H_{03} : m_3(t) \equiv 0 \text{ v.s. } H_{13} : m_3(t) \neq 0, \text{ for some } t \in [0, 6].$$

[Figure 2 about here.]

[Figure 3 about here.]

Asymptotic p -values are calculated for each pair of hypotheses as $\hat{\alpha}_0 = 0.99072$, $\hat{\alpha}_1 = 0.79723$, $\hat{\alpha}_2 = 0.25404$, $\hat{\alpha}_3 = 0.10775$. Apparently, none of the null hypothesis is rejected.

The right plots of Figures 2, 3, 4 and 5 show the spline estimates, the $100(1 - \hat{\alpha}_l)\%$ SCCs and the pointwise confidence intervals, and estimates of $m_l(t)$ under H_{0l} , $l = 0, 1, 2, 3$. From these figures, one can see the baseline CD4 percentage of the population is a decreasing linear function of time and greater than zero over the range of time. The effects of smoking status and age at HIV infection are insignificant, while the pre-infection CD4 percentage is positively proportional to the post-infection CD4 percentage. These findings are consistent with the observations in Wu and Chiang (2000), Fan and Zhang (2000) and Wang, Li and

Huang (2008), but are put on rigorous standing due to the quantification of type I errors by computing asymptotic p -values relative to the SCCs.

APPENDIX A

Throughout this section, $a_n \sim b_n$ means $\lim_{n \rightarrow \infty} b_n/a_n = c$, where c is some nonzero constant. For functions $a_n(t)$, $b_n(t)$, $a_n(t) = u\{b_n(t)\}$ means $a_n(t)/b_n(t) \rightarrow 0$ as $n \rightarrow \infty$ uniformly for $t \in [0, 1]$, and $a_n(t) = \mathcal{U}\{b_n(t)\}$ means $a_n(t)/b_n(t) = \mathcal{O}(1)$ as $n \rightarrow \infty$ uniformly for $t \in [0, 1]$. We use $u_p(\cdot)$ and $\mathcal{U}_p(\cdot)$ if the convergence is in the sense of uniform convergence in probability.

A.1 Technical Assumptions

We define the modulus of continuity of a continuous function ϕ on $[a, b]$ by $\omega(\phi, \delta) = \max_{t, t' \in [a, b], |t - t'| \leq \delta} |\phi(t) - \phi(t')|$. For any $r \in (0, 1]$, denote the collection of order r Hölder continuous function on $[0, 1]$ by

$$C^{0,r}[0, 1] = \left\{ \phi : \|\phi\|_{0,r} = \sup_{t \neq t', t, t' \in [0, 1]} \frac{|\phi(t) - \phi(t')|}{|t - t'|^r} < +\infty \right\},$$

in which $\|\phi\|_{0,r}$ is the $C^{0,r}$ -seminorm of ϕ . Let $C[0, 1]$ be the collection of continuous function on $[0, 1]$. Clearly, $C^{0,r}[0, 1] \subset C[0, 1]$ and, if $\phi \in C^{0,r}[0, 1]$, then $\omega(\phi, \delta) \leq \|\phi\|_{0,r} \delta^r$.

The following regularity assumptions are needed for the main results.

(A1) *The regression functions $m_l(t) \in C^{0,1}[0, 1]$, $l = 1, \dots, d$.*

(A2) *The set of random variables $(T_{ij}, \varepsilon_{ij}, N_i, \xi_{ik,l}, X_{il})_{i=1, j=1, k=1, l=1}^{n, N_i, \infty, d}$ is a subset of variables $(T_{ij}, \varepsilon_{ij}, N_i, \xi_{ik,l}, X_{il})_{i=1, j=1, k=1, l=1}^{\infty, \infty, \infty, d}$ consisting of independent random variables, in which all T_{ij} 's i.i.d with $T_{ij} \sim T$, where T is a random variable with probability density function $f(t)$; X_{il} 's i.i.d for each $l = 1, \dots, d$; N_i 's i.i.d with $N_i \sim N$, where $N > 0$ is a positive integer-valued random variable with $\mathbf{E}\{N^{2r}\} \leq r!c_N^r$, $r = 2, 3, \dots$, for some constant $c_N > 0$. Variables $(\xi_{ik,l})_{i=1, k=1, l=1}^{\infty, \infty, d}$ and $(\varepsilon_{ij})_{i=1, j=1}^{\infty, \infty}$ are i.i.d $N(0, 1)$.*

(A3) *The functions $f(t)$, $\sigma(t)$ and $\phi_{k,l}(t) \in C^{0,r}[0, 1]$ for some $r \in (0, 1]$ with $f(t) \in [c_f, C_f]$, $\sigma(t) \in [c_\sigma, C_\sigma]$, $t \in [0, 1]$, for constants $0 < c_f \leq C_f < +\infty$, $0 < c_\sigma \leq C_\sigma < +\infty$.*

(A4) For $l = 1, \dots, d$, $\sum_{k=1}^{\infty} \|\phi_{k,l}\|_{\infty} < +\infty$, and $G_l(t, t) \in [c_{G,l}, C_{G,l}]$, $t \in [0, 1]$, for constants $0 < c_{G,l} \leq C_{G,l} < +\infty$.

(A5) There exist constants $0 < c_{\mathbf{H}} \leq C_{\mathbf{H}} < +\infty$ and $0 < c_{\eta} \leq C_{\eta} < +\infty$, such that $c_{\mathbf{H}} I_{d \times d} \leq \mathbf{H} = \{H_{ll'}\}_{l,l'=1}^d = \mathbf{E}(\mathbf{X}\mathbf{X}^{\top}) \leq C_{\mathbf{H}} I_{d \times d}$. For some $\eta > 4$, $l = 1, \dots, d$, $c_{\eta} \leq \mathbf{E}|X_l|^{8+\eta} \leq C_{\eta}$.

(A6) As $n \rightarrow \infty$, the number of interior knots $N_s = o(n^{\vartheta})$ for some $\vartheta \in (1/3, 1/2)$ while $N_s^{-1} = o\{n^{-1/3}(\log(n))^{-1/3}\}$. The subinterval length $h_s \sim N_s^{-1}$.

Assumptions (A1)–(A3) are common conditions used in the literature; see for example, Ma, Yang and Carroll (2012). Assumption (A1) controls the rate of convergence of the spline approximation \hat{m}_l , $l = 1, \dots, d$. The requirement of N_i in Assumption (A2) ensures that the observation times are randomly scattered, reflecting sparse and irregular designs. Assumption (A4) guarantees that the random variable $\sum_{k=1}^{\infty} \xi_{ik,l} \phi_{k,l}(t)$ absolutely uniformly converges. Assumption (A5) is analog to Assumption (A2) in Liu and Yang (2010), ensuring that the X_{il} 's are not multicollinear. Assumption (A6) describes the requirement of the growth rate of the dimension of the spline spaces relative to the sample size.

A.2 Proofs of Propositions 1–4

PROOF OF PROPOSITION 1. By Assumption (A3) on the continuity of functions $\phi_{k,l}(t)$, $\sigma^2(t)$ and $f(t)$ on $[0, 1]$ and Assumption (A4), for any $t, u \in [0, 1]$ satisfying $|t - u| \leq h_s$,

$$|G_l(t, t) - G_l(u, u)| \leq \sum_{k=1}^{\infty} |\phi_{k,l}^2(t) - \phi_{k,l}^2(u)| \leq 2 \sum_{k=1}^{\infty} \|\phi_{k,l}\|_{\infty} \omega(\phi_{k,l}, h_s) \leq Ch_s^r.$$

Furthermore,

$$\begin{aligned} \left| \int_{\mathcal{X}_{J(t)}} \{G_l(t, t)f(t) - G_l(u, u)f(u)\} du \right| &\leq Ch_s^{1+r} = \mathcal{O}(h_s^{1+r}), \\ \left| \int_{\mathcal{X}_{J(t)} \times \mathcal{X}_{J(t)}} \{G_l(t, t)f^2(t) - G_l(u, v)f(u)f(v)\} dudv \right| &\leq Ch_s^{2+r} = \mathcal{O}(h_s^{2+r}), \\ \left| \int_{\mathcal{X}_{J(t)}} \{\sigma^2(t)f(t) - \sigma^2(u)f(u)\} du \right| &\leq Ch_s^{1+r} = \mathcal{O}(h_s^{1+r}). \end{aligned}$$

According to the definition of $C_{J,n}$ in (6),

$$C_{J,n} = \int_{[v_J, v_{J+1}]} f(x) dx = f(v_J)h_s + \int_{[v_J, v_{J+1}]} \{f(x) - f(v_J)\} dx, \quad (\text{A.1})$$

thus, $|C_{J,n} - f(v_J)h_s| \leq w(f, h_s)h_s$ for all $J = 0, \dots, N_s$. Therefore,

$$\begin{aligned} \mathbf{\Gamma}_n(t) &= \left\{ f(t)h_s + \mathcal{U}(h_s^{1+r}) \right\}^{-2} (n \mathbf{E} N_1)^{-1} \mathbf{E} \left[\left\{ \sigma_Y^2(t, \mathbf{X}) f(t) h_s + \mathcal{U}_p(h_s^{1+r}) \right. \right. \\ &\quad \left. \left. + \frac{\mathbf{E} \{N_1(N_1 - 1)\}}{\mathbf{E} N_1} \sum_{l=1}^d X_l^2 G_l(t, t) f^2(t) h_s^2 + \mathcal{U}_p(h_s^{2+r}) \right\} \mathbf{X} \mathbf{X}^\top \right] \\ &= \mathbf{E} \left[\mathbf{X} \mathbf{X}^\top \sigma_Y^2(t, \mathbf{X}) \{f(t)h_s n \mathbf{E} N_1\}^{-1} \left\{ 1 + \frac{\mathbf{E} \{N_1(N_1 - 1)\}}{\mathbf{E} N_1} \right. \right. \\ &\quad \left. \left. \times \frac{\sum_{l=1}^d X_l^2 G_l(t, t) f(t) h_s}{\sigma_Y^2(t, \mathbf{X})} \right\} \{1 + \mathcal{U}_p(h_s^r)\} \right] = \tilde{\mathbf{\Gamma}}_n(t) + \mathcal{U}(n^{-1}h_s^{r-1}), \end{aligned}$$

establishing the proposition. \square

PROOF OF PROPOSITION 2. The result follows from standard theory of kernel and spline smoothing, as in Wang and Yang (2009), thus omitted. \square

PROOF OF PROPOSITION 3. According to the result on page 149 of de Boor (2001), there exist functions $g_l \in G^{(-1)}[0, 1]$ that satisfies $\|m_l - g_l\|_\infty = \mathcal{O}(h_s)$ for $l = 1, \dots, d$. By the definition of $\tilde{m}_l(t)$ in (22),

$$\tilde{\mathbf{m}}(t) = (\tilde{m}_1(t), \dots, \tilde{m}_d(t))^\top = c_{J(t),n}^{-1/2} (\tilde{\gamma}_{J(t),1}, \dots, \tilde{\gamma}_{J(t),d})^\top = c_{J(t),n}^{-1/2} \tilde{\gamma}_{J(t)},$$

where $\tilde{\gamma}_J = \hat{\mathbf{V}}_J^{-1} \left\{ N_T^{-1} \sum_{i=1}^n \sum_{j=1}^{N_i} B_J(T_{ij}) X_{il} \sum_{l'=1}^d m_{l'}(T_{ij}) X_{il'} \right\}_{l=1}^d$ for $\hat{\mathbf{V}}_J$ defined in (18).

Let $\tilde{\mathbf{g}}(t) = (\tilde{g}_1(t), \dots, \tilde{g}_d(t))^\top$, then we have

$$\tilde{\mathbf{m}}_l(t) - \tilde{\mathbf{g}}_l(t) = c_{J(t),n}^{-1/2} \hat{\mathbf{V}}_{J(t)}^{-1} \left[\frac{1}{N_T} \sum_{i=1}^n \sum_{j=1}^{N_i} B_{J(t)}(T_{ij}) X_{il} \sum_{l'=1}^d \{m_{l'}(T_{ij}) - g_{l'}(T_{ij})\} X_{il'} \right]_{l=1}^d.$$

Observing that $\tilde{g}_l \equiv g_l$ as $g_l \in G^{(-1)}[0, 1]$, there is a decomposition similar to (24), $\tilde{m}_l(t) = \tilde{m}_l(t) - \tilde{g}_l(t) + g_l(t)$, $l = 1, \dots, d$.

By (A.1), one has $\sup_{t \in [0,1]} |c_{J(t),n}| = \mathcal{O}(h_s)$. Next $\mathbf{E} |B_J(T_{ij})| = C_{J,n}^{-1/2} \int b_J(x) f(x) dx \sim h_s^{1/2}$, thus $\sup_{t \in [0,1]} |B_{J(t)}(T_{ij})| = \mathcal{O}_p(h_s^{1/2})$. Then it is easy to show that $\|\tilde{m}_l - \tilde{g}_l\|_\infty = \mathcal{O}_p(h_s^{-1/2} h_s^{1/2} h_s) = \mathcal{O}_p(h_s)$. Hence, for $l = 1, \dots, d$, $\|\tilde{m}_l - m_l\|_\infty \leq \|\tilde{m}_l - \tilde{g}_l\|_\infty + \|m_l - g_l\|_\infty = \mathcal{O}_p(h_s)$, which completes the proof. \square

Note that $B_J(t) \equiv b_J c_{J,n}^{-1/2}$, $t \in [0, 1]$, so the terms $\tilde{\xi}_l(t)$ and $\tilde{\varepsilon}_l(t)$, $l = 1, \dots, d$, defined in (23) are

$$\tilde{\boldsymbol{\xi}}(t) = \left(\tilde{\xi}_1(t), \dots, \tilde{\xi}_d(t) \right)^\top = c_{J(t),n}^{-1/2} \left(\tilde{\alpha}_{J(t),1}, \dots, \tilde{\alpha}_{J(t),d} \right)^\top = c_{J(t),n}^{-1/2} \tilde{\boldsymbol{\alpha}}_{J(t)}, \quad (\text{A.2})$$

$$\tilde{\boldsymbol{\varepsilon}}(t) = \left(\tilde{\varepsilon}_1(t), \dots, \tilde{\varepsilon}_d(t) \right)^\top = c_{J(t),n}^{-1/2} \left(\tilde{\theta}_{J(t),1}, \dots, \tilde{\theta}_{J(t),d} \right)^\top = c_{J(t),n}^{-1/2} \tilde{\boldsymbol{\theta}}_{J(t)}, \quad (\text{A.3})$$

where

$$\tilde{\boldsymbol{\alpha}}_J = \hat{\mathbf{V}}_J^{-1} \left\{ N_T^{-1} \sum_{i=1}^n \sum_{j=1}^{N_i} B_J(T_{ij}) X_{il} \sum_{l''=1}^d \sum_{k=1}^{\infty} \xi_{ik,l''} \phi_{k,l''}(T_{ij}) X_{il''} \right\}_{l=1}^d,$$

$$\tilde{\boldsymbol{\theta}}_J = \hat{\mathbf{V}}_J^{-1} \left\{ N_T^{-1} \sum_{i=1}^n \sum_{j=1}^{N_i} B_J(T_{ij}) X_{il} \sigma(T_{ij}) \varepsilon_{ij} \right\}_{l=1}^d.$$

According to Lemma B.3, the inverse of the random matrix $\hat{\mathbf{V}}_J$ can be approximated by that of a deterministic matrix $\mathbf{H} = \mathbf{E}(\mathbf{X}\mathbf{X}^\top)$. Substituting $\hat{\mathbf{V}}_J$ with \mathbf{H} in (A.2) and (A.3), we define the random vectors

$$\hat{\boldsymbol{\xi}}(t) = c_{J(t),n}^{-1/2} \mathbf{H}^{-1} \left\{ \frac{1}{N_T} \sum_{i=1}^n \sum_{j=1}^{N_i} B_{J(t)}(T_{ij}) X_{il} \sum_{l''=1}^d \sum_{k=1}^{\infty} \xi_{ik,l''} \phi_{k,l''}(T_{ij}) X_{il''} \right\}_{l=1}^d, \quad (\text{A.4})$$

$$\hat{\boldsymbol{\varepsilon}}(t) = c_{J(t),n}^{-1/2} \mathbf{H}^{-1} \left\{ \frac{1}{N_T} \sum_{i=1}^n \sum_{j=1}^{N_i} B_{J(t)}(T_{ij}) X_{il} \sigma(T_{ij}) \varepsilon_{ij} \right\}_{l=1}^d. \quad (\text{A.5})$$

PROOF OF PROPOSITION 4. Given $(T_{ij}, N_i, X_{il})_{i=1, j=1, l=1}^{n, N_i, d}$, let $\sigma_{\xi_l, n}^2(t)$ and $\sigma_{\varepsilon_l, n}^2(t)$ be the conditional variances of $\hat{\xi}_l(t)$ and $\hat{\varepsilon}_l(t)$ defined in (A.4) and (A.5), respectively. Define

$$\eta_l(t) = \left\{ \sigma_{\xi_l, n}^2(t) + \sigma_{\varepsilon_l, n}^2(t) \right\}^{-1/2} \left\{ \hat{\xi}_l(t) + \hat{\varepsilon}_l(t) \right\}. \quad (\text{A.6})$$

By Lemma B.7, $\eta_l(t)$ is a Gaussian process consisting of $(N_s + 1)$ standard normal variables $\{\eta_{J,l}\}_{J=0}^{N_s}$ such that $\eta_l(t) = \eta_{J(t),l}$ for $t \in [0, 1]$. Thus, for any $\tau \in \mathbb{R}$,

$$P \left(\sup_{t \in [0,1]} |\eta_l(t)| \leq \tau/a_{N_s+1} + b_{N_s+1} \right) = P \left(|\max\{\eta_{0,l}, \dots, \eta_{N_s,l}\}| \leq \tau/a_{N_s+1} + b_{N_s+1} \right).$$

By Theorem 1.5.3 in Leadbetter, Lindgren and Rootzén (1983), if ξ_0, \dots, ξ_{N_s} are i.i.d. standard normal r.v.'s, then for $\tau \in \mathbb{R}$

$$P \left(|\max\{\xi_0, \dots, \xi_{N_s}\}| \leq \tau/a_{N_s} + b_{N_s} \right) \rightarrow \exp(-2e^{-\tau}).$$

Next by Lemma 11.1.2 in Leadbetter, Lindgren and Rootzén (1983),

$$\begin{aligned} P(|\max\{\eta_{0,l}, \dots, \eta_{N_s,l}\}| \leq \tau/a_{N_s+1} + b_{N_s+1}) &- P(|\max\{\xi_0, \dots, \xi_{N_s}\}| \leq \tau/a_{N_s+1} + b_{N_s+1}) \\ &\leq \frac{4}{2\pi} \sum_{0 \leq J < J' \leq N_s} |\mathbf{E} \eta_{J,l} \eta_{J',l}| (1 - |\mathbf{E} \eta_{J,l} \eta_{J',l}|^2)^{-1/2} \exp \left\{ \frac{-(\tau/a_{N_s+1} + b_{N_s+1})^2}{1 + \mathbf{E} \eta_{J,l} \eta_{J',l}} \right\}. \end{aligned}$$

According to Lemma B.7, there exists a constant $C > 0$ such that $\sup_{0 \leq J \neq J' \leq N_s} |\mathbf{E} \eta_{J,l} \eta_{J',l}| \leq Ch_s$ for large n . Thus, as $n \rightarrow \infty$,

$$P(|\max\{\eta_{0,l}, \dots, \eta_{N_s,l}\}| \leq \tau/a_{N_s+1} + b_{N_s+1}) - P(|\max\{\xi_0, \dots, \xi_{N_s}\}| \leq \tau/a_{N_s+1} + b_{N_s+1}) \rightarrow 0.$$

Therefore, for any $\tau \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} P \left(\sup_{t \in [0,1]} |\eta_l(t)| \leq \tau/a_{N_s+1} + b_{N_s+1} \right) = \exp(-2e^{-\tau}). \quad (\text{A.7})$$

By Lemma B.8, we have

$$\begin{aligned} a_{N_s+1} \left(\sup_{t \in [0,1]} \sigma_{n,l}^{-1}(t) \left| \hat{\xi}_l(t) + \hat{\varepsilon}_l(t) \right| - \sup_{t \in [0,1]} |\eta_l(t)| \right) \\ = \mathcal{O}_p \left\{ \log(N_s + 1) (nh_s)^{-1/2} (\log(n))^{1/2} \right\} = \mathcal{O}_p(1). \end{aligned}$$

On the other hand, Lemma B.4 ensures that

$$\begin{aligned} a_{N_s+1} \left(\sup_{t \in [0,1]} \sigma_{n,l}^{-1}(t) \left| \tilde{\xi}_l(t) + \tilde{\varepsilon}_l(t) \right| - \sup_{t \in [0,1]} \sigma_{n,l}^{-1}(t) \left| \hat{\xi}_l(t) + \hat{\varepsilon}_l(t) \right| \right) \\ = \mathcal{O}_p \left\{ (\log(N_s + 1) nh_s)^{1/2} n^{-1} h_s^{-3/2} \log(n) \right\} \\ = \mathcal{O}_p \left\{ n^{-1/2} h_s^{-1} (\log(N_s + 1))^{1/2} \log(n) \right\} = \mathcal{O}_p(1). \end{aligned}$$

Then the proof follows from (A.7) and Slutsky's Theorem. \square

A.3 Proof of Theorem 1

For any vector $\mathbf{a} = (a_1, \dots, a_d)^\top \in \mathbb{R}^d$, $\mathbf{E} \left[\sum_{l=1}^d a_l \left\{ \hat{\xi}_l(t) + \hat{\varepsilon}_l(t) \right\} \right] = 0$. Using the conditional independence of $\hat{\xi}_l(t)$, $\hat{\varepsilon}_l(t)$ on $(T_{ij}, N_i, X_{il})_{i=1, j=1, l=1}^{n, N_i, d}$, we have

$$\begin{aligned} \text{Var} \left[\sum_{l=1}^d a_l \left\{ \hat{\xi}_l(t) + \hat{\varepsilon}_l(t) \right\} \middle| (T_{ij}, N_i, X_{il})_{j=1, i=1, l=1}^{N_i, n, d} \right] \\ = \sum_{l=1}^d \sum_{l'=1}^d a_l a_{l'} \mathbf{E} \left\{ \hat{\xi}_l(t) \hat{\xi}_{l'}(t) + \hat{\varepsilon}_l(t) \hat{\varepsilon}_{l'}(t) \middle| (T_{ij}, N_i, X_{il})_{j=1, i=1, l=1}^{N_i, n, d} \right\} \\ = \mathbf{a}^\top \{ \Sigma_{\hat{\xi}, n}(t) + \Sigma_{\hat{\varepsilon}, n}(t) \} \mathbf{a}. \end{aligned}$$

Meanwhile, Assumption (A2) entails that for any $t \in [0, 1]$, given $(T_{ij}, N_i, X_{il})_{j=1, i=1, l=1}^{N_i, n, d}$, the conditional distribution of $[\mathbf{a}^\top \{\boldsymbol{\Sigma}_{\xi, n}(t) + \boldsymbol{\Sigma}_{\varepsilon, n}(t)\} \mathbf{a}]^{-1/2} \sum_{l=1}^d a_l \{\hat{\xi}_l(t) + \hat{\varepsilon}_l(t)\}$ is a standard normal distribution. So we have

$$[\mathbf{a}^\top \{\boldsymbol{\Sigma}_{\xi, n}(t) + \boldsymbol{\Sigma}_{\varepsilon, n}(t)\} \mathbf{a}]^{-1/2} \sum_{l=1}^d a_l \{\hat{\xi}_l(t) + \hat{\varepsilon}_l(t)\} \sim N(0, 1).$$

Using (B.9), we have as $n \rightarrow \infty$

$$[\mathbf{a}^\top \boldsymbol{\Sigma}_n(t) \mathbf{a}]^{-1/2} \sum_{l=1}^d a_l \{\hat{\xi}_l(t) + \hat{\varepsilon}_l(t)\} \xrightarrow{\mathcal{L}} N(0, 1).$$

Therefore $[\mathbf{a}^\top \boldsymbol{\Sigma}_n(t) \mathbf{a}]^{-1/2} \sum_{l=1}^d a_l \{\hat{m}_l(t) - m_l(t)\} \xrightarrow{\mathcal{L}} N(0, 1)$ follows from (24), Proposition 3, Lemma B.4 and Slutsky's Theorem. Applying Cramér-Wold's device, we obtain $\boldsymbol{\Sigma}_n^{-1/2}(t) \{\hat{m}_l(t) - m_l(t)\}_{l=1}^d \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}_{d \times d})$, and consequently, $\sigma_{n, l}^{-1}(t) \{\hat{m}_l(t) - m_l(t)\} \xrightarrow{\mathcal{L}} N(0, 1)$ for any $t \in [0, 1]$ and $l = 1, \dots, d$. \square

A.4 Proof of Theorem 2

By Proposition 3, $\|\tilde{m}_l - m_l\|_\infty = \mathcal{O}_p(h_s)$, $l = 1, \dots, d$, so

$$a_{N_s+1} \left\{ \sup_{t \in [0, 1]} \sigma_{n, l}^{-1}(t) |\tilde{m}_l(t) - m_l(t)| \right\} = \mathcal{O}_p \left\{ (nh_s)^{1/2} (\log(N_s + 1))^{1/2} h_s \right\} = \mathcal{o}_p(1).$$

According to (24), it is easy to show that

$$a_{N_s+1} \left\{ \sup_{t \in [0, 1]} \sigma_{n, l}^{-1}(t) |\hat{m}_l(t) - m_l(t)| - \sup_{t \in [0, 1]} \sigma_{n, l}^{-1}(t) |\tilde{\xi}_l(t) + \tilde{\varepsilon}_l(t)| \right\} = \mathcal{o}_p(1).$$

Meanwhile, Proposition 4 entails that, for any $\tau \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} P \left\{ a_{N_s+1} \left(\sup_{t \in [0, 1]} \sigma_{n, l}^{-1}(t) |\tilde{\xi}_l(t) + \tilde{\varepsilon}_l(t)| - b_{N_s+1} \right) \leq \tau \right\} = \exp(-2e^{-\tau}).$$

Thus Slutsky's Theorem implies that

$$\lim_{n \rightarrow \infty} P \left\{ a_{N_s+1} \left(\sup_{t \in [0, 1]} \sigma_{n, l}^{-1}(t) |\hat{m}_l(t) - m_l(t)| - b_{N_s+1} \right) \leq \tau \right\} = \exp(-2e^{-\tau}).$$

Let $\tau = -\log\{-\frac{1}{2} \log(1 - \alpha)\}$, the definition of $Q_{N_s+1}(\alpha)$ in (9) entails

$$\begin{aligned} & \lim_{n \rightarrow \infty} P \{m_l(t) \in \hat{m}_l(t) \pm \sigma_{n, l}(t) Q_{N_s+1}(\alpha), \forall t \in [0, 1]\} \\ &= \lim_{n \rightarrow \infty} P \left\{ \sup_{t \in [0, 1]} \sigma_{n, l}^{-1}(t) |\hat{m}_l(t) - m_l(t)| \leq Q_{N_s+1}(\alpha) \right\} = 1 - \alpha. \end{aligned}$$

Theorem 2 is proved. \square

B. SUPPLEMENTARY MATERIALS

Supplement to “A Simultaneous Confidence Corridor for Varying Coefficient Regression with Sparse Functional Data”: Supplement containing the details of theoretical proofs referenced in the main article.

vcmfdaband.xpl: XploRe-package containing code to perform estimations and SCCs for the coefficient functions.

REFERENCES

- [1] Brumback, B. and Rice, J. A. (1998), “Smoothing Spline Models for the Analysis of Nested and Crossed Samples of Curves (with Discussion),” *Journal of the American Statistical Association*, 93, 961–994.
- [2] Chiang, C.-T., Rice, J. A. and Wu, C. O. (2001), “Smoothing Spline Estimation for Varying Coefficient Models with Repeatedly Measured Dependent Variables,” *Journal of the American Statistical Association*, 96, 605–619.
- [3] de Boor, C. (2001). *A Practical Guide to Splines*, New York: Springer-Verlag.
- [4] Fan, J. and Zhang, J. T. (2000), “Functional Linear Models for Longitudinal Data,” *Journal of the Royal Statistical Society Series B*, 62, 303–322.
- [5] Fan, J. and Zhang, W. Y. (2000), “Simultaneous Confidence Bands and Hypothesis Testing in Varying-coefficient Models,” *Scandinavian Journal of Statistics*, 27, 715–731.
- [6] Fan, J. and Zhang, W. Y. (2008), “Statistical Methods with Varying Coefficient Models,” *Statistics and its Interface*, 1, 179–195.
- [7] Hall, P., Müller, H. G., and Wang, J. L. (2006), “Properties of Principal Component Methods for Functional and Longitudinal Data Analysis,” *Annals of Statistics*, 34, 1493–1517.

- [8] Hall, P. and Titterton, D. M. (1988), “On Confidence Bands in Nonparametric Density Estimation and Regression,” *Journal of Multivariate Analysis*, 27, 228–254.
- [9] Härdle, W and Luckhaus, S. (1984), “Uniform Consistency of a Class of Regression Function Estimators,” *Annals of Statistics*, 12, 612–623.
- [10] Hastie, T. and Tibshirani, R. (1993), “Varying-coefficient Models,” *Journal of the Royal Statistical Society Series B*, 55, 757–796.
- [11] Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, L.-P. (1998), “Nonparametric Smoothing Estimates of Time-varying Coefficient Models with Longitudinal Data,” *Biometrika*, 85, 809–822.
- [12] Huang, J. Z., Wu, C. O. and Zhou, L. (2002), “Varying-coefficient Models and Basis Function Approximations for the Analysis of Repeated Measurements,” *Biometrika*, 89, 111–128.
- [13] Huang, J. Z., Wu, C. O. and Zhou, L. (2004), “Polynomial Spline Estimation and Inference for Varying Coefficient Models with Longitudinal Data,” *Statistica Sinica*, 14, 763–788.
- [14] James, G. M., Hastie, T., and Sugar, C. (2000), “Principal Component Models for Sparse Functional Data,” *Biometrika*, 87, 587–602.
- [15] James, G. M. and Sugar, C. A. (2003), “Clustering for Sparsely Sampled Functional Data,” *Journal of the American Statistical Association*, 98, 397–408.
- [16] Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983), *Extremes and Related Properties of Random Sequences and Processes*, New York: Springer-Verlag.
- [17] Liu, R. and Yang, L. (2010), “Spline-backfitted Kernel Smoothing of Additive Coefficient Model,” *Econometric Theory*, 26, 29–59.

- [18] Ma, S., Yang, L. and Carroll, R. J. (2012), “A Simultaneous Confidence Band for Sparse Longitudinal Regression,” *Statistica Sinica*, 22, 95–122.
- [19] Ramsay, J. O. and Silverman, B. W. (2005), *Functional Data Analysis, Second Edition*, New York: Springer.
- [20] Wang, L., Li, H., and Huang, J. Z. (2008), “Variable Selection in Nonparametric Varying-coefficient Models for Analysis of Repeated Measurements,” *Journal of the American Statistical Association*, 103, 1556–1569.
- [21] Wang, L. and Yang, L. (2009), “Polynomial Spline Confidence Bands for Regression Curves,” *Statistica Sinica*, 19, 325–342.
- [22] Wu, C. O. and Chiang, C.-T. (2000), “Kernel Smoothing on Varying Coefficient Models with Longitudinal Dependent Variable,” *Statistica Sinica*, 10, 433–456.
- [23] Wu, C. O., Chiang, C.-T. and Hoover, D. R. (1998), “Asymptotic Confidence Regions for Kernel Smoothing of a Varying-coefficient Model with Longitudinal Data,” *Journal of the American Statistical Association*, 93, 1388–1402.
- [24] Wu, Y., Fan, J. and Müller, H. G. (2010), “Varying-coefficient Functional Linear Regression,” *Bernoulli*, 16, 730—758.
- [25] Xue, L. and Yang, L. (2006), “Additive Coefficient Modelling via Polynomial Spline,” *Statistica Sinica*, 16, 1423–1446.
- [26] Xue, L. and Zhu, L. (2007), “Empirical Likelihood for a Varying Coefficient Model with Longitudinal Data,” *Journal of the American Statistical Association*, 102, 642–654.
- [27] Yao, W. and Li, R. (2013), “New Local Estimation Procedure for a Non-parametric Regression Function for Longitudinal Data,” *Journal of the Royal Statistical Society Series B*, 75, 123–138.

- [28] Yao, F., Müller, H. G., and Wang, J. L. (2005a), “Functional Linear Regression Analysis for Longitudinal Data,” *Annals of Statistics*, 33, 2873–2903.
- [29] Yao, F., Müller, H. G., and Wang, J. L. (2005b), “Functional Data Analysis for Sparse Longitudinal Data,” *Journal of the American Statistical Association*, 100, 577–590.
- [30] Zhou, L., Huang, J., and Carroll, R. J. (2008), “Joint Modelling of Paired Sparse Functional Data Using Principal Components,” *Biometrika*, 95, 601–619.
- [31] Zhu, H., Li, R. and Kong, L. (2012), “Multivariate Varying Coefficient Model for Functional Responses. *Annals of Statistics*, 40, 2634–2666.

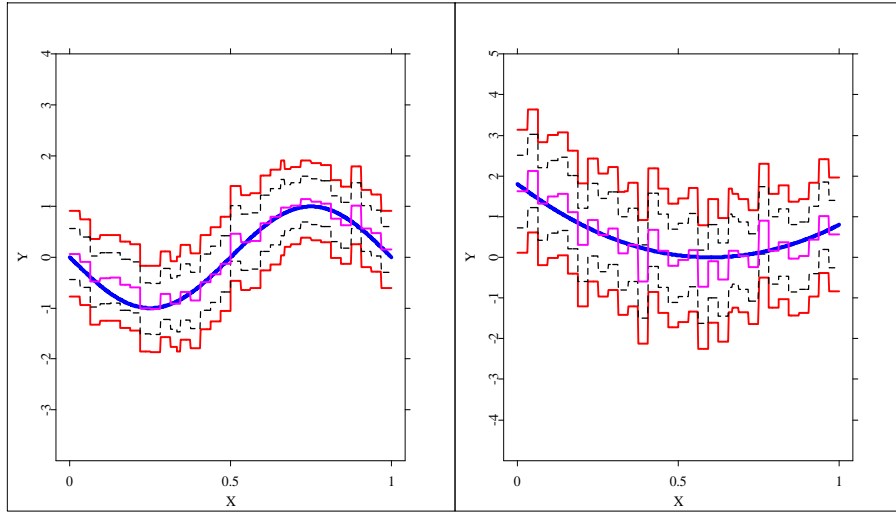


Figure 1: Plots of 95% SCC (11) (upper and lower solid), pointwise confidence intervals (dashed), the spline estimator (thin), and the true function (middle thick) at $\sigma = 0.5$, $n = 200$ for m_1 (left) and m_2 (right).

Figure 2: Plots of (a) 95% SCC (upper and lower solid), pointwise confidence intervals (dashed) and the spline estimator \hat{m}_0 (middle solid) for baseline effect; and (b) the same except with confidence level $1 - \hat{\alpha}_0$ and the estimated m_0 under H_{00} (solid linear).

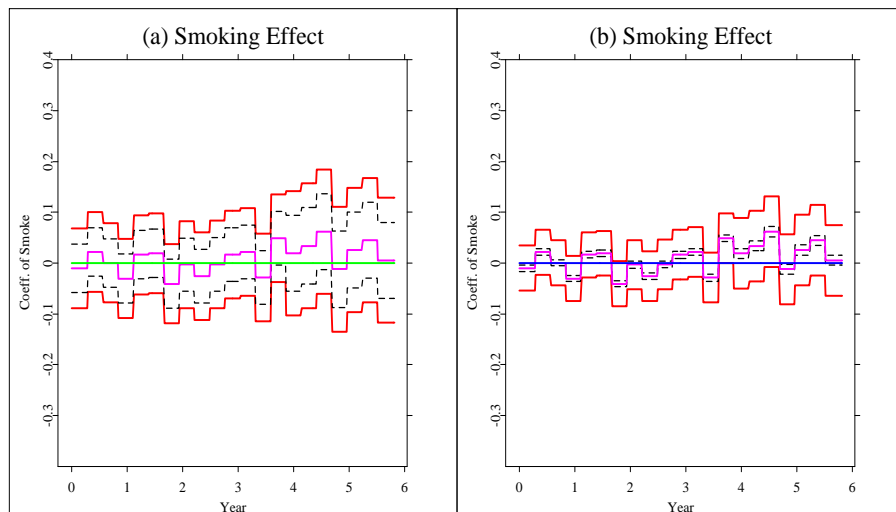
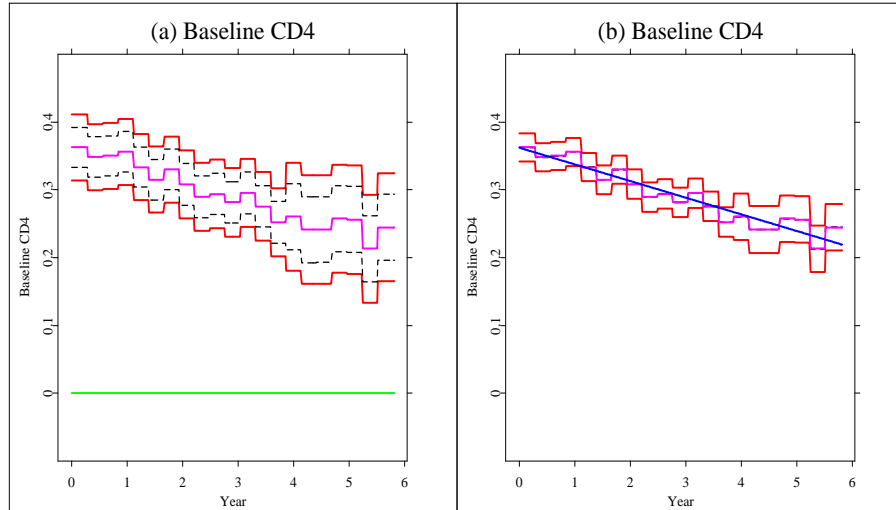


Figure 3: Plots of (a) 95% SCC (upper and lower solid), pointwise confidence intervals (dashed) and the spline estimator \hat{m}_1 (middle solid) for smoking effect; and (b) the same except with confidence level $1 - \hat{\alpha}_1$ and the estimated m_1 under H_{01} (solid linear).

Figure 4: Plots of (a) 95% SCC (upper and lower solid), pointwise confidence intervals (dashed) and the spline estimator \hat{m}_2 (middle solid) for pre-infection CD4 effect; and (b) the same except with confidence level $1 - \hat{\alpha}_2$ and the estimated m_2 under H_{02} (solid linear).

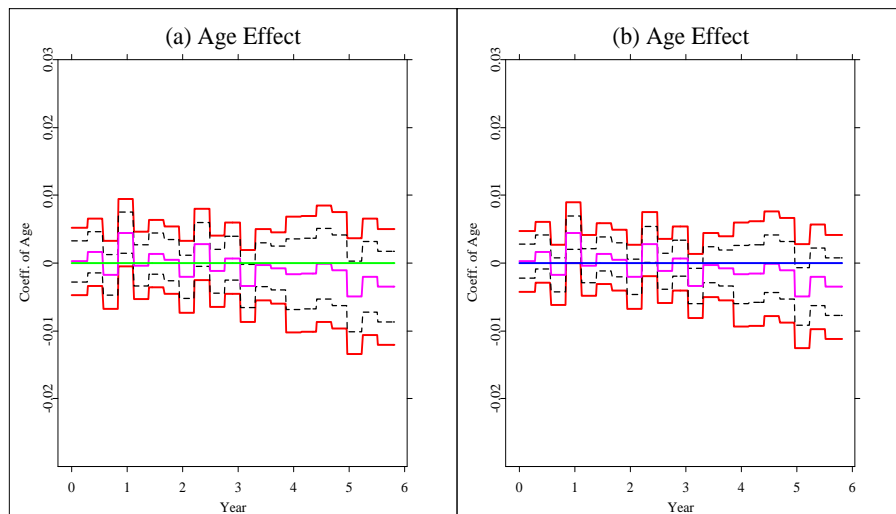
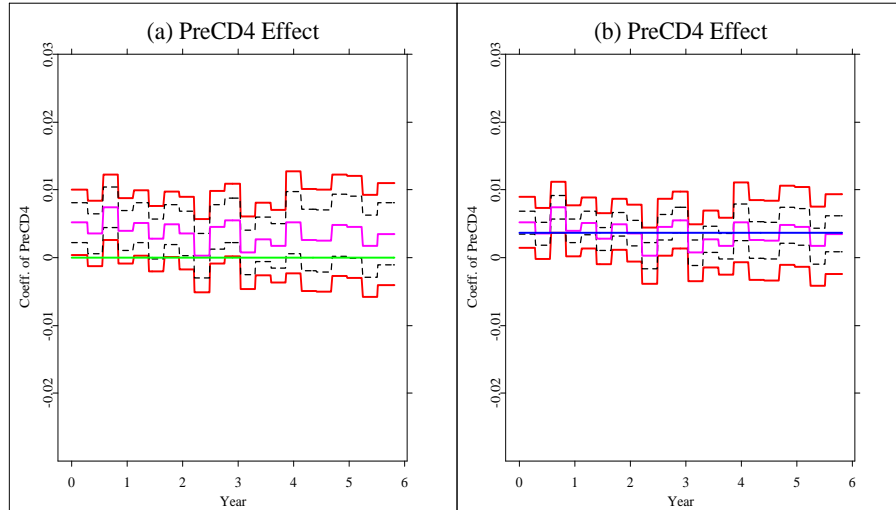


Figure 5: Plots of (a)95% SCC (upper and lower solid), pointwise confidence intervals (dashed) and the spline estimator \hat{m}_3 (middle solid) for age effect; and (b) the same except with confidence level $1 - \hat{\alpha}_3$ and the estimated m_3 under H_{03} (solid linear).

Table 1: Coverage percentages of the SCCs for functions m_1 (left) and m_2 (right), based on 500 replications.

σ	n	$1 - \alpha$	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 1$
1.0	200	0.950	0.950, 0.952	0.944, 0.948	0.920, 0.904	0.886, 0.884
		0.990	0.990, 0.998	0.990, 0.990	0.976, 0.984	0.968, 0.974
	400	0.950	0.944, 0.948	0.950, 0.930	0.922, 0.912	0.908, 0.904
		0.990	0.996, 0.984	0.990, 0.988	0.984, 0.988	0.974, 0.966
	600	0.950	0.934, 0.962	0.954, 0.946	0.930, 0.952	0.930, 0.924
		0.990	0.992, 0.996	0.992, 0.986	0.988, 0.990	0.984, 0.990
	800	0.950	0.936, 0.934	0.960, 0.966	0.950, 0.964	0.956, 0.934
		0.990	0.998, 0.996	0.994, 0.994	0.986, 0.992	0.988, 0.988
0.5	200	0.950	0.936, 0.948	0.952, 0.942	0.916, 0.900	0.912, 0.890
		0.990	0.988, 0.994	0.992, 0.990	0.972, 0.974	0.972, 0.972
	400	0.950	0.916, 0.930	0.936, 0.932	0.928, 0.916	0.904, 0.898
		0.990	0.994, 0.984	0.992, 0.988	0.996, 0.988	0.978, 0.976
	600	0.950	0.924, 0.948	0.952, 0.954	0.926, 0.958	0.936, 0.938
		0.990	0.996, 0.994	0.994, 0.986	0.984, 0.990	0.990, 0.994
	800	0.950	0.942, 0.900	0.950, 0.960	0.942, 0.962	0.960, 0.938
		0.990	0.996, 0.998	0.996, 0.994	0.990, 0.996	0.992, 0.988

Supplement to “A Simultaneous Confidence Corridor for Varying Coefficient Regression with Sparse Functional Data”

¹Lijie Gu, ³Li Wang, ^{4,5}Wolfgang K. Härdle and ^{1,2}Lijian Yang

¹Soochow University, ²Michigan State University, ³University of Georgia

⁴Humboldt-Universität zu Berlin, ⁵Singapore Management University

In this document, we have collected a number of technical lemmas and their proofs. The technical lemmas are used in the proofs of Propositions 1–4 in the paper.

LEMMA B.1. (*Bosq (1998), Theorem 1.2*). Suppose that $\{\xi_i\}_{i=1}^n$ are i.i.d with $\mathbf{E}(\xi_1) = 0$, $\sigma^2 = \mathbf{E} \xi_1^2$, and there exists $c > 0$ such that for $r = 3, 4, \dots$, $\mathbf{E} |\xi_1|^r \leq c^{r-2} r! \mathbf{E} \xi_1^2 < +\infty$. Then for each $n > 1$, $t > 0$, $P(|S_n| \geq \sqrt{n}\sigma t) \leq 2 \exp\left(-t^2 (4 + 2ct/\sqrt{n}\sigma)^{-1}\right)$, in which $S_n = \sum_{i=1}^n \xi_i$.

LEMMA B.2. Under Assumptions (A2)–(A6), we have

$$A_{n,1} = \sup_{0 \leq J \leq N_s, 1 \leq l, l' \leq d} \frac{\left| \langle B_J X_l, B_J X_{l'} \rangle_{N_T} - \langle B_J X_l, B_J X_{l'} \rangle \right|}{\sqrt{\langle B_J X_l, B_J X_l \rangle} \sqrt{\langle B_J X_{l'}, B_J X_{l'} \rangle}} = \mathcal{O}_p \left(\sqrt{\frac{\log(n)}{nh_s}} \right),$$

where for any $J = 0, \dots, N_s$ and $l, l' = 1, \dots, d$,

$$\begin{aligned} \langle B_J X_l, B_J X_{l'} \rangle_{N_T} &= N_T^{-1} \sum_{i=1}^n \sum_{j=1}^{N_i} B_J^2(T_{ij}) X_{il} X_{il'}, \\ \langle B_J X_l, B_J X_{l'} \rangle &= \mathbf{E} \{ B_J^2(T_{ij}) X_{il} X_{il'} \} = H_W. \end{aligned}$$

PROOF. Let $\omega_{i,J} = \omega_{i,J,l,l'} = \sum_{j=1}^{N_i} B_J^2(T_{ij}) X_{il} X_{il'}$, then $\mathbb{E} \omega_{i,J} = \mathbb{E} N_1 H_{ll'} \sim 1$ and $\mathbb{E} (\omega_{i,J})^2 = \mathbb{E} \left\{ \sum_{j=1}^{N_i} B_J^2(T_{ij}) \right\}^2 \mathbb{E} (X_{il} X_{il'})^2 \sim h_s^{-1}$. Next define a sequence $D_n = n^\alpha$ with $\alpha(4 + \eta/2) > 1$ and $\sqrt{\log(n)} D_n n^{-1/2} h_s^{-1/2} \rightarrow 0$, $n^{1/2} h_s^{1/2} D_n^{-(3+\eta/2)} \rightarrow 0$, which necessitates $\eta > 2$ according to Assumption (A5). We make use of the following truncated and tail decomposition

$$X_{ill'} = X_{il} X_{il'} = X_{ill',1}^{D_n} + X_{ill',2}^{D_n},$$

where $X_{ill',1}^{D_n} = X_{il} X_{il'} I \{|X_{il} X_{il'}| > D_n\}$, $X_{ill',2}^{D_n} = X_{il} X_{il'} I \{|X_{il} X_{il'}| \leq D_n\}$. Define correspondingly the truncated and tail parts of $\omega_{i,J}$ as $\omega_{i,J,m} = B_J^2(T_{ij}) X_{ill',m}^{D_n}$, $m = 1, 2$. According to Assumption (A5), for any $l, l' = 1, \dots, d$,

$$\sum_{n=1}^{\infty} P \{|X_{nl} X_{nl'}| > D_n\} \leq \sum_{n=1}^{\infty} \frac{\mathbb{E} |X_{nl} X_{nl'}|^{4+\eta/2}}{D_n^{4+\eta/2}} \leq C_\eta \sum_{n=1}^{\infty} D_n^{-(4+\eta/2)} < \infty.$$

By Borel-Cantelli Lemma, one has $\sum_{j=1}^{N_i} B_J^2(T_{ij}) X_{ill',1}^{D_n} = 0$, *a.s.* So we obtain

$$\sup_{J,l,l'} \left| n^{-1} \sum_{i=1}^n \omega_{i,J,1} \right| = \mathcal{O}_{a.s.} (n^{-k}), \quad k \geq 1,$$

and

$$\begin{aligned} \mathbb{E} \omega_{i,J,1} &= \mathbb{E} (X_{ill',1}^{D_n}) \mathbb{E} \left\{ \sum_{j=1}^{N_i} B_J^2(T_{ij}) \right\} \\ &\leq D_n^{-(3+\eta/2)} \mathbb{E} |X_{il} X_{il'}|^{4+\eta/2} \mathbb{E} N_1 \mathbb{E} B_J^2(T_{ij}) \leq c D_n^{-(3+\eta/2)}. \end{aligned}$$

Next we considerate the truncated part $\omega_{i,J,2}$. For large n , $\mathbb{E} (\omega_{i,J,2}) = \mathbb{E} (\omega_{i,J}) - \mathbb{E} (\omega_{i,J,1}) \sim 1$, $\mathbb{E} (\omega_{i,J,2})^2 = \mathbb{E} (\omega_{i,J})^2 - \mathbb{E} (\omega_{i,J,1})^2 \sim h_s^{-1}$. Define $\omega_{i,J,2}^* = \omega_{i,J,2} - \mathbb{E} (\omega_{i,J,2})$, then $\mathbb{E} \omega_{i,J,2}^* = 0$, and

$$\begin{aligned} \mathbb{E} (\omega_{i,J,2}^*)^2 &= \mathbb{E} (\omega_{i,J,2})^2 - (\mathbb{E} \omega_{i,J,2})^2 = \mathbb{E} \left\{ \sum_{j=1}^{N_i} B_J^2(T_{ij}) X_{ill',2}^{D_n} \right\}^2 - \mathcal{U}(1) \\ &= \mathbb{E} (X_{ill',2}^{D_n})^2 \mathbb{E} \left\{ \sum_{j=1}^{N_i} B_J^2(T_{ij}) \right\}^2 - \mathcal{U}(1). \end{aligned}$$

Note that

$$\begin{aligned} \mathbb{E} (X_{ill',2}^{D_n})^2 \mathbb{E} \left\{ \sum_{j=1}^{N_i} B_J^2(T_{ij}) \right\}^2 &\geq \left\{ \mathbb{E} (X_{ill'})^2 - \mathbb{E} (X_{ill',1}^{D_n})^2 \right\} \mathbb{E} \left\{ \sum_{j=1}^{N_i} B_J^4(T_{ij}) \right\} \\ &\geq \left\{ \mathbb{E} (X_{ill'})^2 - u(1) \right\} \mathbb{E} N_1 \mathbb{E} B_J^4(T_{ij}). \end{aligned}$$

Thus, there exists c_ω such that for large n , $\mathbf{E} (\omega_{i,J,2}^*)^2 \geq c_\omega \mathbf{E} (X_{ill'})^2 h_s^{-1}$. Next for any $r > 2$

$$\begin{aligned} \mathbf{E} |\omega_{i,J,2}^*|^r &= \mathbf{E} |\omega_{i,J,2} - \mathbf{E} (\omega_{i,J,2})|^r \leq 2^{r-1} (\mathbf{E} |\omega_{i,J,2}|^r + |\mathbf{E} (\omega_{i,J,2})|^r) \\ &= 2^{r-1} \left\{ \mathbf{E} |X_{ill',2}^{D_n}|^r \mathbf{E} \left| \sum_{j=1}^{N_i} B_J^2(T_{ij}) \right|^r + \mathcal{U}(1) \right\} \\ &= 2^{r-1} \left[\mathbf{E} |X_{ill',2}^{D_n}|^r \mathbf{E} \left\{ \sum_{0 \leq r_1, \dots, r_{N_i} \leq r}^{r_1 + \dots + r_{N_i} = r} \binom{r}{r_1 \dots r_{N_i}} \prod_{j=1}^{N_i} \mathbf{E} B_J^{2r_j}(T_{ij}) \right\} + \mathcal{U}(1) \right], \end{aligned}$$

then there exists $C_\omega > 0$ such that for any $r > 2$ and large n ,

$$\begin{aligned} \mathbf{E} |\omega_{i,J,2}^*|^r &\leq 2^{r-1} \left[D_n^{r-2} \mathbf{E} (X_{ill'})^2 \mathbf{E} \left\{ N_1^r \max_{j=1}^{N_i} \prod_{j=1}^{N_i} \mathbf{E} B_J^{2r_j}(T_{ij}) \right\} + \mathcal{U}(1) \right] \\ &\leq 2^{r-1} [D_n^{r-2} \mathbf{E} (X_{ill'})^2 (\mathbf{E} N_1^r) C_B h_s^{1-r} + \mathcal{U}(1)] \\ &\leq 2^r D_n^{r-2} (c_N^r r!)^{1/2} C_B h_s^{2-r} c_\omega^{-1} \mathbf{E} (\omega_{i,J,2}^*)^2 \\ &\leq (C_\omega D_n h_s^{-1})^{r-2} r! \mathbf{E} (\omega_{i,J,2}^*)^2, \end{aligned}$$

which implies that $\{\omega_{i,J,2}^*\}_{i=1}^n$ satisfies Cramér's condition. Applying Lemma B.1 to $\sum_{i=1}^n \omega_{i,J,2}^*$, for $r > 2$ and any large enough $\delta > 0$, $P \left\{ \left| n^{-1} \sum_{i=1}^n \omega_{i,J,2}^* \right| \geq \delta (nh_s)^{-1/2} (\log(n))^{1/2} \right\}$ is bounded by

$$2 \exp \left\{ \frac{-\delta^2 (\log(n))}{4 + 2C_\omega D_n h_s^{-1} \delta (\log(n))^{1/2} n^{-1/2} h_s^{1/2}} \right\} \leq 2n^{-8}.$$

Hence

$$\sum_{n=1}^{\infty} P \left\{ \sup_{0 \leq J \leq N_s, 1 \leq l, l' \leq d} \left| n^{-1} \sum_{i=1}^n \omega_{i,J,2}^* \right| \geq \delta (nh_s)^{-1/2} (\log(n))^{1/2} \right\} < \infty.$$

Thus, $\sup_{J,l,l'} \left| n^{-1} \sum_{i=1}^n \omega_{i,J,2}^* \right| = \mathcal{O}_{a.s.} \left\{ (nh_s)^{-1/2} (\log(n))^{1/2} \right\}$ as $n \rightarrow \infty$ by Borel-Cantelli Lemma. Furthermore,

$$\begin{aligned} &\sup_{J,l,l'} \left| n^{-1} \sum_{i=1}^n \omega_{i,J} - \mathbf{E} \omega_{i,J} \right| \\ &\leq \sup_{J,l,l'} \left| n^{-1} \sum_{i=1}^n \omega_{i,J,1} \right| + \sup_{J,l,l'} \left| n^{-1} \sum_{i=1}^n \omega_{i,J,2}^* \right| + \sup_{J,l,l'} |\mathbf{E} \omega_{i,J,1}| \\ &= \mathcal{U}_{a.s.} (n^{-k}) + \mathcal{O}_{a.s.} \left\{ (nh_s)^{-1/2} (\log(n))^{1/2} \right\} + \mathcal{U} (D_n^{-(3+\eta/2)}) \\ &= \mathcal{O}_{a.s.} \left\{ (nh_s)^{-1/2} (\log(n))^{1/2} \right\}. \end{aligned}$$

Finally, we notice that

$$\begin{aligned}
& \sup_{J,l,l'} \left| \langle B_J X_l, B_J X_{l'} \rangle_{N_T} - \langle B_J X_l, B_J X_{l'} \rangle \right| = \sup_{J,l,l'} \left| (nN_T^{-1}) n^{-1} \sum_{i=1}^n \omega_{i,J} - (\mathbf{E} N_1)^{-1} \mathbf{E} \omega_{i,J} \right| \\
& \leq \sup_{J,l,l'} (\mathbf{E} N_1)^{-1} |(n \mathbf{E} N_1) N_T^{-1} - 1| \left| n^{-1} \sum_{i=1}^n \omega_{i,J} \right| + \sup_{J,l,l'} (\mathbf{E} N_1)^{-1} \left| n^{-1} \sum_{i=1}^n \omega_{i,J} - \mathbf{E} \omega_{i,J} \right| \\
& = \mathcal{O}_p(n^{-1/2}) + \mathcal{O}_{a.s.} \left\{ (nh_s)^{-1/2} (\log(n))^{1/2} \right\} = \mathcal{O}_p \left\{ (nh_s)^{-1/2} (\log(n))^{1/2} \right\},
\end{aligned}$$

and $\langle B_J X_l, B_J X_l \rangle = H_{ll} = \mathcal{U}(1)$. Hence, $A_{n,1} = \mathcal{O}_p \left\{ (nh_s)^{-1/2} (\log(n))^{1/2} \right\}$. \square

For the random matrix $\hat{\mathbf{V}}_J$ defined in (18), the lemma below shows that its inverse can be approximated by the inverse of a deterministic matrix $\mathbf{H} = \mathbf{E}(\mathbf{X}\mathbf{X}^\top)$.

LEMMA B.3. *Under Assumptions (A2) and (A4)–(A6), for any $J = 0, \dots, N_s$, we have*

$$\hat{\mathbf{V}}_J^{-1} = \mathbf{H}^{-1} + \mathcal{O}_p \left\{ (nh_s)^{-1/2} (\log(n))^{1/2} \right\}. \quad (\text{B.1})$$

PROOF. By Lemma B.2, we have

$$\left\| \hat{\mathbf{V}}_J - \mathbf{H} \right\|_\infty = \mathcal{O}_p \left\{ (nh_s)^{-1/2} (\log(n))^{1/2} \right\}.$$

Using the fact that for any matrices \mathbf{A} and \mathbf{B} ,

$$(\mathbf{A} + h\mathbf{B})^{-1} = \mathbf{A}^{-1} - h\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1} + \mathcal{O}(h^2),$$

we obtain (B.1). \square

The next lemma implies that the difference between $\tilde{\boldsymbol{\xi}}(t)$ and $\hat{\boldsymbol{\xi}}(t)$ and the difference between $\tilde{\boldsymbol{\varepsilon}}(t)$ and $\hat{\boldsymbol{\varepsilon}}(t)$ are both negligible uniformly over $t \in [0, 1]$.

LEMMA B.4. *Under Assumption (A2)–(A6), for $\tilde{\boldsymbol{\xi}}(t)$, $\tilde{\boldsymbol{\varepsilon}}(t)$ given in (A.2), (A.3) and $\hat{\boldsymbol{\xi}}(t)$, $\hat{\boldsymbol{\varepsilon}}(t)$ given in (A.4), (A.5), as $n \rightarrow \infty$, we have*

$$\sup_{t \in [0,1]} \left\| \tilde{\boldsymbol{\xi}}(t) - \hat{\boldsymbol{\xi}}(t) \right\|_\infty = \mathcal{O}_p \left\{ n^{-1} h_s^{-3/2} \log(n) \right\}, \quad (\text{B.2})$$

$$\sup_{t \in [0,1]} \left\| \tilde{\boldsymbol{\varepsilon}}(t) - \hat{\boldsymbol{\varepsilon}}(t) \right\|_\infty = \mathcal{O}_p \left\{ n^{-1} h_s^{-3/2} \log(n) \right\}. \quad (\text{B.3})$$

PROOF. Comparing the equations of $\tilde{\xi}(t)$ and $\hat{\xi}(t)$ given in (A.2) and (A.4), we let

$$\frac{1}{N_T} \sum_{i=1}^n \sum_{j=1}^{N_i} B_J(T_{ij}) X_{il} \sum_{l''=1}^d \sum_{k=1}^{\infty} \xi_{ik,l''} \phi_{k,l''}(T_{ij}) X_{il''} = \frac{n}{N_T} \sum_{l''=1}^d \sum_{i=1}^n \Omega_{i,J,l'',l}.$$

where $\Omega_{i,J,l'',l} = \Omega_i = n^{-1} \left[X_{il} X_{il''} \sum_{k=1}^{\infty} \left\{ \sum_{j=1}^{N_i} B_J(T_{ij}) \phi_{k,l''}(T_{ij}) \right\} \xi_{ik,l''} \right]$. Note that $E \Omega_i = 0$ and

$$\begin{aligned} \sigma_{\Omega_i,n}^2 &= E \left(\Omega_i^2 \middle| (T_{ij}, N_i, X_{il})_{i=1,j=1,l=1}^{n,N_i,d} \right) \\ &= n^{-2} \left[X_{il} X_{il''} \sum_{k=1}^{\infty} \left\{ \sum_{j=1}^{N_i} B_J(T_{ij}) \phi_{k,l''}(T_{ij}) \right\}^2 \right] \\ &\leq n^{-2} \left\{ X_{il}^2 X_{il''}^2 \sum_{k=1}^{\infty} N_i \sum_{j=1}^{N_i} B_J^2(T_{ij}) \phi_{k,l''}^2(T_{ij}) \right\} \\ &= n^{-2} \left\{ X_{il}^2 X_{il''}^2 N_i \sum_{j=1}^{N_i} B_J^2(T_{ij}) G_{l''}(T_{ij}, T_{ij}) \right\} \\ &\leq C n^{-2} h_s^{-1} X_{il}^2 X_{il''}^2 N_i^2. \end{aligned}$$

Given $(T_{ij}, N_i, X_{il})_{i=1,j=1,l=1}^{n,N_i,d}$, $\{\sigma_{\Omega_i,n}^{-1} \Omega_i\}_{i=1}^n$ are i.i.d $N(0, 1)$. It is easy to show that for any large enough $\delta > 0$,

$$P \left\{ \frac{|\sum_{i=1}^n \Omega_i|}{\sqrt{\sum_{i=1}^n \sigma_{\Omega_i,n}^2}} \geq \delta \sqrt{\log(n)} \middle| (T_{ij}, N_i, X_{il})_{i=1,j=1,l=1}^{n,N_i,d} \right\} \leq 2 \exp \left\{ -\frac{1}{2} \delta^2 \log(n) \right\} \leq 2n^{-8},$$

$$P \left[\left| \sum_{i=1}^n \Omega_i \right| \geq \delta \left\{ \frac{C \log(n)}{nh_s} n^{-1} \sum_{i=1}^n X_{il}^2 X_{il''}^2 N_i^2 \right\}^{1/2} \middle| (T_{ij}, N_i, X_{il})_{i=1,j=1,l=1}^{n,N_i,d} \right] \leq 2n^{-8}.$$

Note that $n^{-1} \sum_{i=1}^n X_{il}^2 X_{il''}^2 N_i^2 = \mathcal{O}_p(1)$, hence

$$\sum_{n=1}^{\infty} P \left\{ \sup_{0 \leq J \leq N_s, 1 \leq l, l'' \leq d} \left| \sum_{i=1}^n \Omega_{i,J,l'',l} \right| \geq \delta (nh_s)^{-1/2} (\log(n))^{1/2} \right\} < \infty.$$

Thus, $\sup_{J,l,l''} |\sum_{i=1}^n \Omega_{i,J,l'',l}| = \mathcal{O}_p \left\{ (nh_s)^{-1/2} (\log(n))^{1/2} \right\}$ as $n \rightarrow \infty$ by Borel-Cantelli Lemma. Furthermore, $\sup_{J,l} \left| n N_T^{-1} \sum_{l''=1}^d \sum_{i=1}^n \Omega_{i,J,l'',l} \right| = \mathcal{O}_p \left\{ (nh_s)^{-1/2} (\log(n))^{1/2} \right\}$. Finally, according to Lemma B.1, we obtain (B.2). (B.3) is proved similarly. \square

Denote the inverse matrix of \mathbf{H} by $\mathbf{H}^{-1} = \{z_{l'l'}\}_{l,l'=1}^d$. For any $l = 1, \dots, d$, we rewrite the l -th element of $\hat{\xi}_l(t)$ and $\hat{\varepsilon}_l(t)$ in (A.4) and (A.5) as the following

$$\hat{\xi}_l(t) = c_{J(t),n}^{-1/2} N_T^{-1} \sum_{l''=1}^d \sum_{i=1}^n \sum_{k=1}^{\infty} R_{ik,\xi,J(t),l'',l} \xi_{ik,l''}, \quad (\text{B.4})$$

$$\hat{\varepsilon}_l(t) = c_{J(t),n}^{-1/2} N_T^{-1} \sum_{i=1}^n \sum_{j=1}^N R_{ij,\varepsilon,J(t),l} \varepsilon_{ij}, \quad (\text{B.5})$$

where for any $0 \leq J \leq N_s$,

$$R_{ik,\xi,J,l'',l} = \left(\sum_{l'=1}^d z_{l'l'} X_{il'} X_{il''} \right) \left\{ \sum_{j=1}^{N_i} B_J(T_{ij}) \phi_{k,l''}(T_{ij}) \right\}, \quad (\text{B.6})$$

$$R_{ij,\varepsilon,J,l} = \left(\sum_{l'=1}^d z_{l'l'} X_{il'} \right) B_J(T_{ij}) \sigma(T_{ij}). \quad (\text{B.7})$$

Further denote

$$S_{il''} = \left(\sum_{l'=1}^d z_{l'l'} X_{il'} X_{il''} \right)^2, \quad s_{l''} = \mathbf{E}(S_{il''}), \quad 1 \leq l, l'' \leq d. \quad (\text{B.8})$$

LEMMA B.5. *Under Assumptions (A2)-(A6), for $R_{ik,\xi,J,l'',l}$, $R_{ij,\varepsilon,J,l}$ in (B.6), (B.7),*

$$\begin{aligned} \mathbf{E} \left(\sum_{k=1}^{\infty} R_{ik,\xi,J,l'',l}^2 \right) &= c_{J,n}^{-1} s_{l''} \left[(\mathbf{E} N_1) \int b_J(u) G_{l''}(u, u) f(u) du \right. \\ &\quad \left. + \mathbf{E} \{N_1(N_1 - 1)\} \int b_J(u) b_J(v) G_{l''}(u, v) f(u) f(v) dudv \right], \\ \mathbf{E} R_{ij,\varepsilon,J,l}^2 &= c_{J,n}^{-1} z_{ll} \int b_J(u) \sigma^2(u) f(u) du, \end{aligned}$$

for $0 \leq J \leq N_s$ and $0 \leq l, l'' \leq d$. In addition, there exist $0 < c_R < C_R < \infty$, such that

$$c_R s_{l''} \leq \mathbf{E} \left(\sum_{k=1}^{\infty} R_{ik,\xi,J,l'',l}^2 \right) \leq C_R s_{l''}, \quad c_R \leq \mathbf{E} R_{ij,\varepsilon,J,l}^2 \leq C_R,$$

for $0 \leq J \leq N_s$, $0 \leq l, l'' \leq d$, and as $n \rightarrow \infty$

$$\begin{aligned} A_{n,\xi} &= \sup_{J,l''} \left| n^{-1} \sum_{i=1}^n \sum_{k=1}^{\infty} R_{ik,\xi,J,l'',l}^2 - \mathbf{E} \left(\sum_{k=1}^{\infty} R_{ik,\xi,J,l'',l}^2 \right) \right| = \mathcal{O}_{a.s.} \left\{ (nh_s)^{-1/2} (\log(n))^{1/2} \right\}, \\ A_{n,\varepsilon} &= \sup_{J,l} \left| N_T^{-1} \sum_{i=1}^n \sum_{j=1}^{N_i} R_{ij,\varepsilon,J,l}^2 - \mathbf{E} R_{ij,\varepsilon,J,l}^2 \right| = \mathcal{O}_{a.s.} \left\{ (nh_s)^{-1/2} (\log(n))^{1/2} \right\}. \end{aligned}$$

PROOF. By independence of $\{T_{ij}\}_{j=1}^{\infty}$, $\{X_{il}\}_{l=1}^d$, N_i , the definition of B_J and (B.8),

$$\begin{aligned} \mathbb{E} \left(\sum_{k=1}^{\infty} R_{ik,\xi,J,l''}^2 \right) &= \mathbb{E}(S_{il''}) \mathbb{E} \sum_{k=1}^{\infty} \left\{ \sum_{j=1}^{N_i} B_J(T_{ij}) \phi_{k,l''}(T_{ij}) \right\}^2 \\ &= s_{l''} \mathbb{E} \sum_{j=1}^{N_i} \sum_{j'=1}^{N_i} B_J(T_{ij}) B_J(T_{ij'}) G_{l''}(T_{ij}, T_{ij'}) \\ &= s_{l''} c_{J,n}^{-1} \left\{ (\mathbb{E} N_1) \int b_J(u) G_{l''}(u, u) f(u) du \right. \\ &\quad \left. + \mathbb{E} \{N_1(N_1 - 1)\} \int b_J(u) b_J(v) G_{l''}(u, v) f(u) f(v) dudv \right\}, \end{aligned}$$

thus there exist constants $0 < c_R < C_R < \infty$ such that $c_R s_{l''} \leq \mathbb{E} \left(\sum_{k=1}^{\infty} R_{1k,\xi,J,l''}^2 \right) \leq C_R s_{l''}$, $0 \leq J \leq N_s$, $0 \leq l, l'' \leq d$.

If $s_{l''} = 0$, one has $S_{il''} = 0$, almost surely. Hence $n^{-1} \sum_{i=1}^n \sum_{k=1}^{\infty} R_{ik,\xi,J,l''}^2 = 0$, almost surely. In the case of $s_{l''} > 0$, let $\zeta_{i,J} = \zeta_{i,J,l''} = \sum_{k=1}^{\infty} R_{ik,\xi,J,l''}^2$ for brevity. Under Assumption (A5), it is easy to verify that

$$0 < s_{l''}^2 \leq \mathbb{E}(S_{il''})^2 \leq d^3 \sum_{l'=1}^d \mathbb{E} |z_{l'} X_{il'} X_{il''}|^4 \leq d^3 \sum_{l'=1}^d z_{l'} \{ \mathbb{E} |X_{il'}|^8 \mathbb{E} |X_{il''}|^8 \}^{1/2} < \infty.$$

So for large n ,

$$\begin{aligned} \mathbb{E}(\zeta_{i,J})^2 &= \mathbb{E} \left\{ (S_{il''})^2 \left(\sum_{j=1}^{N_i} \sum_{j'=1}^{N_i} B_J(T_{ij}) B_J(T_{ij'}) G_{l''}(T_{ij}, T_{ij'}) \right)^2 \right\} \\ &\geq \mathbb{E}(S_{il''})^2 \frac{1}{4} c_{G,l''}^2 \mathbb{E} \left\{ \sum_{j=1}^{N_i} B_J(T_{ij}) \right\}^4 \geq c \mathbb{E} \sum_{j=1}^{N_i} B_J^4(T_{ij}) \geq c h_s^{-1}, \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}(\zeta_{i,J})^2 &\leq \mathbb{E}(S_{il''})^2 4C_{G,l''}^2 \mathbb{E} \left\{ \sum_{j=1}^{N_i} B_J(T_{ij}) \right\}^4 \\ &\leq c \mathbb{E} \left[N_1^3 \sum_{j=1}^{N_i} \mathbb{E} B_J^4(T_{ij}) \middle| N_1 \right] \leq c \mathbb{E} N_1^4 \mathbb{E} B_J^4(T_{ij}) \leq c h_s^{-1}. \end{aligned}$$

Define a sequence $D_n = n^\alpha$ that satisfies $\alpha(2 + \eta/4) > 1$, $D_n n^{-1/2} h_s^{-1/2} (\log(n))^{1/2} \rightarrow 0$, $n^{1/2} h_s^{1/2} D_n^{-(1+\eta/4)} \rightarrow 0$, which requires $\eta > 4$ provided by Assumption (A5). We make use of

the following truncated and tail decomposition

$$S_{il''} = \sum_{l'=1}^d \sum_{l'''=1}^d z_{l'} z_{l'''} X_{il'} X_{il'''} X_{il''}^2 = S_{il'',1}^{D_n} + S_{il'',2}^{D_n},$$

where

$$\begin{aligned} S_{il'',1}^{D_n} &= \sum_{l'=1}^d \sum_{l'''=1}^d z_{l'} z_{l'''} X_{il'} X_{il'''} X_{il''}^2 I \{ |X_{il'} X_{il'''} X_{il''}^2| > D_n \}, \\ S_{il'',2}^{D_n} &= \sum_{l'=1}^d \sum_{l'''=1}^d z_{l'} z_{l'''} X_{il'} X_{il'''} X_{il''}^2 I \{ |X_{il'} X_{il'''} X_{il''}^2| \leq D_n \}. \end{aligned}$$

Define correspondingly the truncated and tail parts of $\zeta_{i,J}$ as

$$\zeta_{i,J,m} = S_{il'',m}^{D_n} \sum_{j=1}^{N_i} \sum_{j'=1}^{N_i} B_J(T_{ij}) B_J(T_{ij'}) G_{l''}(T_{ij}, T_{ij'}), \quad m = 1, 2.$$

According to Assumption (A5), for any $l', l'', l''' = 1, \dots, d$,

$$\sum_{n=1}^{\infty} P \{ |X_{nl'} X_{nl'''} X_{nl''}^2| > D_n \} \leq \sum_{n=1}^{\infty} \frac{\mathbf{E} |X_{nl'} X_{nl'''} X_{nl''}^2|^{2+\eta/4}}{D_n^{2+\eta/4}} \leq C_{\eta} \sum_{n=1}^{\infty} D_n^{-(2+\eta/4)} < \infty.$$

Borel-Cantelli Lemma implies that

$$P \{ \omega \mid \exists N(\omega), |X_{nl'} X_{nl'''} X_{nl''}^2(\omega)| \leq D_n \text{ for } n > N(\omega) \} = 1,$$

$$P \{ \omega \mid \exists N(\omega), |X_{il'} X_{il'''} X_{il''}^2(\omega)| \leq D_n, i = 1, \dots, n \text{ for } n > N(\omega) \} = 1,$$

$$P \{ \omega \mid \exists N(\omega), I \{ |X_{il'} X_{il'''} X_{il''}^2(\omega)| > D_n \} = 0, i = 1, \dots, n \text{ for } n > N(\omega) \} = 1.$$

Furthermore, one has

$$n^{-1} \sum_{i=1}^n \left\{ S_{il'',1}^{D_n} \sum_{j=1}^{N_i} \sum_{j'=1}^{N_i} B_J(T_{ij}) B_J(T_{ij'}) G_{l''}(T_{ij}, T_{ij'}) \right\} = 0, \quad a.s.$$

Therefore, one has

$$\sup_{J,l,l''} \left| n^{-1} \sum_{i=1}^n \zeta_{i,J,1} \right| = \mathcal{O}_{a.s.}(n^{-k}), \quad k \geq 1.$$

Notice that

$$\begin{aligned} \mathbf{E} (S_{il'',1}^{D_n}) &= \mathbf{E} \left[\sum_{l'=1}^d \sum_{l'''=1}^d z_{l'} z_{l'''} X_{il'} X_{il'''} X_{il''}^2 I \{ |X_{il'} X_{il'''} X_{il''}^2| > D_n \} \right] \\ &\leq D_n^{-(1+\eta/4)} \sum_{l'=1}^d \sum_{l'''=1}^d z_{l'} z_{l'''} \mathbf{E} |X_{il'} X_{il'''} X_{il''}^2|^{2+\eta/4} \\ &\leq c D_n^{-(1+\eta/4)}. \end{aligned}$$

So for large n ,

$$\begin{aligned}
\mathbb{E}(\zeta_{i,J,1}) &= \mathbb{E}(S_{ill',1}^{D_n}) \mathbb{E} \left\{ \sum_{j=1}^{N_i} \sum_{j'=1}^{N_i} B_J(T_{ij}) B_J(T_{ij'}) G_{l''}(T_{ij}, T_{ij'}) \right\} \\
&\leq cD_n^{-(1+\eta/4)} 2C_{G,l''} \mathbb{E} \left\{ \sum_{j=1}^{N_i} B_J(T_{ij}) \right\}^2 \\
&\leq cD_n^{-(1+\eta/4)} \mathbb{E}(N_1^2) \mathbb{E} B_J^2(T_{ij}) \\
&\leq cD_n^{-(1+\eta/4)}.
\end{aligned}$$

Next we considerate the truncated part $\zeta_{i,J,2}$. For large n , $\mathbb{E}(\zeta_{i,J,2}) = \mathbb{E}(\zeta_{i,J}) - \mathbb{E}(\zeta_{i,J,1}) \sim 1$, $\mathbb{E}(\zeta_{i,J,2})^2 = \mathbb{E}(\zeta_{i,J})^2 - \mathbb{E}(\zeta_{i,J,1})^2 \sim h_s^{-1}$. Define $\zeta_{i,J,2}^* = \zeta_{i,J,2} - \mathbb{E}(\zeta_{i,J,2})$, then $\mathbb{E}\zeta_{i,J,2}^* = 0$, and there exist $c_\zeta, C_\zeta > 0$ such that for $r > 2$ and large n ,

$$\begin{aligned}
\mathbb{E}(\zeta_{i,J,2}^*)^2 &= \mathbb{E}|S_{ill',2}^{D_n}|^2 \mathbb{E} \left| \sum_{j=1}^{N_i} \sum_{j'=1}^{N_i} B_J(T_{ij}) B_J(T_{ij'}) G_{l''}(T_{ij}, T_{ij'}) \right|^2 - (\mathbb{E}\zeta_{i,J,2})^2 \\
&\geq \left\{ \mathbb{E}|S_{ill''}|^2 - \mathbb{E}|S_{ill',1}^{D_n}|^2 \right\} \frac{1}{4} c_{G,l''}^2 \mathbb{E} \left\{ \sum_{j=1}^{N_i} B_J(T_{ij}) \right\}^4 - \mathcal{U}(1) \\
&\geq \left\{ \mathbb{E}|S_{ill''}|^2 - \mathcal{U}(1) \right\} \frac{1}{4} c_{G,l''}^2 \mathbb{E} \left\{ \sum_{j=1}^{N_i} B_J^4(T_{ij}) \right\} - \mathcal{U}(1) \\
&\geq \frac{1}{2} \mathbb{E}|S_{ill''}|^2 \frac{1}{4} c_{G,l''}^2 \mathbb{E} N_1 \mathbb{E} B_J^4(T_{ij}) - \mathcal{U}(1) \\
&\geq c_\zeta \mathbb{E}|S_{ill''}|^2 h_s^{-1},
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}|\zeta_{i,J,2}^*|^r &= \mathbb{E}|\zeta_{i,J,2} - \mathbb{E}(\zeta_{i,J,2})|^r \leq 2^{r-1} (\mathbb{E}|\zeta_{i,J,2}|^r + |\mathbb{E}(\zeta_{i,J,2})|^r) \\
&= 2^{r-1} \left\{ \mathbb{E}|S_{ill',2}^{D_n}|^r \mathbb{E} \left| \sum_{j=1}^{N_i} \sum_{j'=1}^{N_i} B_J(T_{ij}) B_J(T_{ij'}) G_{l''}(T_{ij}, T_{ij'}) \right|^r + \mathcal{U}(1) \right\} \\
&\leq 2^{r-1} \left[(cD_n)^{r-2} \mathbb{E}|S_{ill''}|^2 (2C_{G,l''})^r \mathbb{E} \left\{ \sum_{j=1}^{N_i} B_J(T_{ij}) \right\}^{2r} + \mathcal{U}(1) \right] \\
&\leq 2^{r-1} [(cD_n)^{r-2} \mathbb{E}|S_{ill''}|^2 (2C_{G,l''})^r (\mathbb{E} N_1^{2r}) C_B h_s^{1-r} + \mathcal{U}(1)] \\
&\leq 2^r (cD_n)^{r-2} (2C_{G,l''})^r c_N^r r! C_B h_s^{2-r} c_\zeta^{-1} \mathbb{E}(\zeta_{i,J,2}^*)^2 \\
&\leq (C_\zeta D_n h_s^{-1})^{r-2} r! \mathbb{E}(\zeta_{i,J,2}^*)^2,
\end{aligned}$$

which implies that $\{\zeta_{i,J,2}^*\}_{i=1}^n$ satisfies Cramér's condition. Applying Lemma B.1 to $\sum_{i=1}^n \zeta_{i,J,2}^*$, for $r > 2$ and any large enough $\delta > 0$,

$$\begin{aligned} P \left\{ \left| n^{-1} \sum_{i=1}^n \zeta_{i,J,2}^* \right| \geq \delta (nh_s)^{-1/2} (\log(n))^{1/2} \right\} \\ \leq 2 \exp \left\{ \frac{-\delta^2 \log(n)}{4 + 2C_\zeta D_n h_s^{-1} \delta (\log(n))^{1/2} n^{-1/2} h_s^{1/2}} \right\} \leq 2n^{-8}. \end{aligned}$$

Hence

$$\sum_{n=1}^{\infty} P \left\{ \sup_{J,l''} \left| n^{-1} \sum_{i=1}^n \zeta_{i,J,2}^* \right| \geq \delta (nh_s)^{-1/2} (\log(n))^{1/2} \right\} < \infty.$$

Thus, $\sup_{J,l''} |n^{-1} \sum_{i=1}^n \zeta_{i,J,2}^*| = \mathcal{O}_{a.s.} \left\{ (nh_s)^{-1/2} (\log(n))^{1/2} \right\}$ as $n \rightarrow \infty$ by the Borel-Cantelli lemma. Furthermore, we have

$$\begin{aligned} A_{n,\xi} &\leq \sup_{J,l,l''} \left| n^{-1} \sum_{i=1}^n \zeta_{i,J,1} \right| + \sup_{J,l''} \left| n^{-1} \sum_{i=1}^n \zeta_{i,J,2}^* \right| + \sup_{J,l''} |\mathbb{E}(\zeta_{i,J,1})| \\ &= \mathcal{U}_{a.s.} (n^{-k}) + \mathcal{O}_{a.s.} \left\{ (nh_s)^{-1/2} (\log(n))^{1/2} \right\} + \mathcal{U} (D_n^{-(1+\eta/4)}) \\ &= \mathcal{O}_{a.s.} \left\{ (nh_s)^{-1/2} (\log(n))^{1/2} \right\}. \end{aligned}$$

The properties of $R_{ij,\varepsilon,J,l}$ are obtained similarly. \square

Next define two $d \times d$ matrices

$$\begin{aligned} \mathbf{\Gamma}_{\xi,n}(t) &= c_{J(t),n}^{-1} N_{\mathbb{T}}^{-2} \sum_{l''=1}^d \sum_{i=1}^n \sum_{k=1}^{\infty} \left\{ \sum_{j=1}^{N_i} B_{J(t)}(T_{ij}) \phi_{k,l''}(T_{ij}) \right\}^2 X_{il''}^2 \mathbf{X}_i \mathbf{X}_i^{\mathbb{T}}, \\ \mathbf{\Gamma}_{\varepsilon,n}(t) &= c_{J(t),n}^{-1} N_{\mathbb{T}}^{-2} \sum_{i=1}^n \sum_{j=1}^{N_i} B_{J(t)}^2(T_{ij}) \sigma^2(T_{ij}) \mathbf{X}_i \mathbf{X}_i^{\mathbb{T}}. \end{aligned}$$

LEMMA B.6. *For any $t \in \mathbb{R}$, the conditional covariance matrices of $\hat{\boldsymbol{\xi}}(t)$ and $\hat{\boldsymbol{\varepsilon}}(t)$ on $(T_{ij}, N_i, X_{il})_{i=1, j=1, l=1}^{n, N_i, d}$ are*

$$\begin{aligned} \boldsymbol{\Sigma}_{\xi,n}(t) &= \mathbb{E} \left\{ \hat{\boldsymbol{\xi}}(t) \hat{\boldsymbol{\xi}}^{\mathbb{T}}(t) \middle| (T_{ij}, N_i, X_{il})_{i=1, j=1, l=1}^{n, N_i, d} \right\} = \mathbf{H}^{-1} \mathbf{\Gamma}_{\xi,n}(t) \mathbf{H}^{-1}, \\ \boldsymbol{\Sigma}_{\varepsilon,n}(t) &= \mathbb{E} \left\{ \hat{\boldsymbol{\varepsilon}}(t) \hat{\boldsymbol{\varepsilon}}^{\mathbb{T}}(t) \middle| (T_{ij}, N_i, X_{il})_{i=1, j=1, l=1}^{n, N_i, d} \right\} = \mathbf{H}^{-1} \mathbf{\Gamma}_{\varepsilon,n}(t) \mathbf{H}^{-1}, \end{aligned}$$

and with $\boldsymbol{\Sigma}_n(t)$ defined in (7),

$$\sup_{t \in [0,1]} \|\{\boldsymbol{\Sigma}_{\xi,n}(t) + \boldsymbol{\Sigma}_{\varepsilon,n}(t)\} - \boldsymbol{\Sigma}_n(t)\|_{\infty} = \mathcal{O}_{a.s.} \left\{ n^{-3/2} h_s^{-3/2} (\log(n))^{1/2} \right\}. \quad (\text{B.9})$$

PROOF. Note that

$$\begin{aligned} \hat{\boldsymbol{\xi}}(t) \hat{\boldsymbol{\xi}}^\top(t) &= c_{J(t),n}^{-1} \mathbf{H}^{-1} \left\{ \frac{1}{N_T^2} \sum_{i=1}^n \sum_{j=1}^{N_i} B_{J(t)}(T_{ij}) X_{il} \sum_{l''=1}^d \sum_{k=1}^{\infty} \xi_{ik,l''} \phi_{k,l''}(T_{ij}) X_{il''} \right. \\ &\quad \left. \times \sum_{i=1}^n \sum_{j=1}^{N_i} B_{J(t)}(T_{ij}) X_{il'} \sum_{l''=1}^d \sum_{k=1}^{\infty} \xi_{ik,l''} \phi_{k,l''}(T_{ij}) X_{il''} \right\}_{l,l'=1}^d \mathbf{H}^{-1}. \end{aligned}$$

Thus,

$$\begin{aligned} \boldsymbol{\Sigma}_{\xi,n}(t) &= \mathbf{E} \left\{ \hat{\boldsymbol{\xi}}(t) \hat{\boldsymbol{\xi}}^\top(t) \middle| (T_{ij}, N_i, X_{il})_{i=1, j=1, l=1}^{n, N_i, d} \right\} = c_{J(t),n}^{-1} \mathbf{H}^{-1} \\ &\quad \times \left[N_T^{-2} \sum_{l''=1}^d \sum_{i=1}^n \sum_{k=1}^{\infty} \left\{ \sum_{j=1}^{N_i} B_{J(t)}(T_{ij}) \phi_{k,l''}(T_{ij}) \right\}^2 X_{il''}^2 \mathbf{X}_i \mathbf{X}_i^\top \right] \mathbf{H}^{-1} \\ &= \mathbf{H}^{-1} \boldsymbol{\Gamma}_{\xi,n}(t) \mathbf{H}^{-1}. \end{aligned}$$

Similarly, we can derive the conditional covariance matrix of $\hat{\boldsymbol{\varepsilon}}(t)$. Next let

$$\begin{aligned} \Psi_{ik,\xi,J,l,l',l''} &= \left\{ \sum_{j=1}^{N_i} B_J(T_{ij}) \phi_{k,l''}(T_{ij}) \right\}^2 X_{il''}^2 X_{il} X_{il'}, \\ \Psi_{ij,\varepsilon,J,l,l'} &= B_J^2(T_{ij}) \sigma^2(T_{ij}) X_{il} X_{il'}. \end{aligned}$$

Similar to the proof of Lemma B.5,

$$\begin{aligned} \mathbf{E} \left(\sum_{k=1}^{\infty} \Psi_{ik,\xi,J,l,l',l''} \right) &= c_{J,n}^{-1} \mathbf{E} (X_{il''}^2 X_{il} X_{il'}) \left[(\mathbf{E} N_1) \int_{\chi_J} G_{l''}(u, u) f(u) du + \right. \\ &\quad \left. + \mathbf{E} \{N_1(N_1 - 1)\} \int_{\chi_J \times \chi_J} G_{l''}(u, v) f(u) f(v) dudv \right], \\ \mathbf{E} \Psi_{ij,\varepsilon,J,l,l'} &= c_{J,n}^{-1} \mathbf{E} (X_{il} X_{il'}) \int_{\chi_J} \sigma^2(u) f(u) du, \end{aligned}$$

and as $n \rightarrow \infty$,

$$\begin{aligned} \sup_{J,l,l',l''} \left| n^{-1} \sum_{i=1}^n \sum_{k=1}^{\infty} \Psi_{ik,\xi,J,l,l',l''} - \mathbf{E} \left(\sum_{k=1}^{\infty} \Psi_{ik,\xi,J,l,l',l''} \right) \right| &= \mathcal{O}_{a.s.} \left\{ (nh_s)^{-1/2} (\log(n))^{1/2} \right\}, \\ \sup_{J,l,l'} \left| N_T^{-1} \sum_{i=1}^n \sum_{j=1}^{N_i} \Psi_{ij,\varepsilon,J,l,l'} - \mathbf{E} \Psi_{ij,\varepsilon,J,l,l'} \right| &= \mathcal{O}_{a.s.} \left\{ (nh_s)^{-1/2} (\log(n))^{1/2} \right\}. \end{aligned}$$

Furthermore,

$$\begin{aligned}
& \sup_{J,l,l',l''} \left| N_T^{-2} \sum_{i=1}^n \sum_{k=1}^{\infty} \Psi_{ik,\xi,J,l,l',l''} - n^{-1} (\mathbf{E} N_1)^{-2} \mathbf{E} \left(\sum_{k=1}^{\infty} \Psi_{ik,\xi,J,l,l',l''} \right) \right| \\
& \leq \sup_{J,l,l',l''} n^{-1} (\mathbf{E} N_1)^{-2} \left\{ \left| \left(\frac{n \mathbf{E} N_1}{N_T} \right)^2 - 1 \right| \left| n^{-1} \sum_{i=1}^n \sum_{k=1}^{\infty} \Psi_{ik,\xi,J,l,l',l''} \right| \right. \\
& \quad \left. + \left| n^{-1} \sum_{i=1}^n \sum_{k=1}^{\infty} \Psi_{ik,\xi,J,l,l',l''} - \mathbf{E} \left(\sum_{k=1}^{\infty} \Psi_{ik,\xi,J,l,l',l''} \right) \right| \right\} \\
& = \mathcal{O}_{a.s.} \left\{ n^{-3/2} h_s^{-1/2} (\log(n))^{1/2} \right\},
\end{aligned}$$

and

$$\begin{aligned}
& \sup_{J,l,l'} \left| N_T^{-2} \sum_{i=1}^n \sum_{j=1}^{N_i} \Psi_{ik,\varepsilon,J,l,l'} - (n \mathbf{E} N_1)^{-1} \mathbf{E} \Psi_{ik,\varepsilon,J,l,l'} \right| \\
& \leq \sup_{J,l,l'} (n \mathbf{E} N_1)^{-1} \left\{ \left| \frac{n \mathbf{E} N_1}{N_T} - 1 \right| \left| N_T^{-1} \sum_{i=1}^n \sum_{j=1}^{N_i} \Psi_{ik,\varepsilon,J,l,l'} \right| \right. \\
& \quad \left. + \left| N_T^{-1} \sum_{i=1}^n \sum_{j=1}^{N_i} \Psi_{ik,\varepsilon,J,l,l'} - \mathbf{E} \Psi_{ik,\varepsilon,J,l,l'} \right| \right\} \\
& = \mathcal{O}_{a.s.} \left\{ n^{-3/2} h_s^{-1/2} (\log(n))^{1/2} \right\}.
\end{aligned}$$

Notice that

$$\begin{aligned}
\Sigma_n(t) &= \mathbf{H}^{-1} c_{J(t),n}^{-1} (n \mathbf{E} N_1)^{-1} \left\{ (\mathbf{E} N_1)^{-1} \mathbf{E} \left(\sum_{l''=1}^d \sum_{k=1}^{\infty} \Psi_{ik,\xi,J(t),l,l',l''} \right) + \mathbf{E} \Psi_{ij,\varepsilon,J(t),l,l'} \right\}_{l,l'=1}^d \\
& \quad \times \mathbf{H}^{-1},
\end{aligned}$$

$$\Sigma_{\xi,n}(t) + \Sigma_{\varepsilon,n}(t)$$

$$= \mathbf{H}^{-1} c_{J(t),n}^{-1} N_T^{-2} \left\{ \sum_{l''=1}^d \sum_{i=1}^n \sum_{k=1}^{\infty} \Psi_{ik,\xi,J(t),l,l',l''} + \sum_{i=1}^n \sum_{j=1}^{N_i} \Psi_{ij,\varepsilon,J(t),l,l'} \right\}_{l,l'=1}^d \mathbf{H}^{-1},$$

and (A.1) implies $\sup_{t \in [0,1]} |c_{J(t),n}| = \mathcal{O}(h_s)$. Hence (B.9) holds. \square

Given $(T_{ij}, N_i, X_{il})_{i=1, j=1, l=1}^{n, N_i, d}$, let $\sigma_{\xi_l, n}^2(t)$ and $\sigma_{\varepsilon_l, n}^2(t)$ be the conditional variances of $\hat{\xi}_l(t)$ and $\hat{\varepsilon}_l(t)$ defined in (A.4) and (A.5), respectively. Lemma B.6 implies that

$$\sup_{t \in [0,1]} \left| \sigma_{\xi_l, n}^2(t) + \sigma_{\varepsilon_l, n}^2(t) - \sigma_{n, ll}^2(t) \right| = \mathcal{O}_{a.s.} \left\{ n^{-3/2} h_s^{-3/2} (\log(n))^{1/2} \right\}. \quad (\text{B.10})$$

LEMMA B.7. Under Assumptions (A2)-(A6), for $l = 1, \dots, d$, $\eta_l(t)$ defined in (A.6) is a Gaussian process consisting of $(N_s + 1)$ standard normal variables $\{\eta_{J,l}\}_{J=0}^{N_s}$ such that $\eta_l(t) = \eta_{J(t),l}$ for $t \in [0, 1]$, and there exists a constant $C > 0$ such that for large n , $\sup_{0 \leq J \neq J' \leq N_s} |\mathbf{E} \eta_{J,l} \eta_{J',l}| \leq Ch_s$.

PROOF. For any fixed $l = 1, \dots, d$ and $0 \leq J \leq N_s$, $\mathcal{L} \left\{ \eta_{J,l} \middle| (T_{ij}, N_i, X_{il})_{i=1, j=1, l=1}^{n, N_i, d} \right\} = N(0, 1)$ by Assumption (A2), so $\mathcal{L} \{ \eta_{J,l} \} = N(0, 1)$, for $0 \leq J \leq N_s$.

Next we derive the upper bound for $\sup_{0 \leq J \neq J' \leq N_s} |\mathbf{E} \eta_{J,l} \eta_{J',l}|$. Let

$$\bar{R}_{\xi, J(t), l} = N_T^{-1} \sum_{l''=1}^d \sum_{i=1}^n \sum_{k=1}^{\infty} R_{ik, \xi, J(t), l'', l}^2, \quad \bar{R}_{\varepsilon, J(t), l} = N_T^{-1} \sum_{i=1}^n \sum_{j=1}^{N_i} R_{ij, \varepsilon, J(t), l}^2,$$

then we have

$$\sigma_{\xi_l, n}(t) = \left\{ c_{J(t), n}^{-1} N_T^{-2} \sum_{l''=1}^d \sum_{i=1}^n \sum_{k=1}^{\infty} R_{ik, \xi, J(t), l'', l}^2 \right\}^{1/2} = \left\{ c_{J(t), n}^{-1} N_T^{-1} \bar{R}_{\xi, J(t), l} \right\}^{1/2},$$

$$\sigma_{\varepsilon_l, n}(t) = \left\{ c_{J(t), n}^{-1} N_T^{-2} \sum_{i=1}^n \sum_{j=1}^{N_i} R_{ij, \varepsilon, J(t), l}^2 \right\}^{1/2} = \left\{ c_{J(t), n}^{-1} N_T^{-1} \bar{R}_{\varepsilon, J(t), l} \right\}^{1/2}.$$

For $J \neq J'$, by (B.7) and the definition of B_J ,

$$R_{ij, \varepsilon, J, l} R_{ij, \varepsilon, J', l} = \left(\sum_{l''=1}^d z_{l''} X_{il''} \right)^2 B_J(T_{ij}) B_{J'}(T_{ij}) \sigma^2(T_{ij}) = 0,$$

along with the conditional independence of $\hat{\xi}_l(t), \hat{\varepsilon}_l(t)$ on $(T_{ij}, N_i, X_{il})_{i=1, j=1, l=1}^{n, N_i, d}$, and independence of $\xi_{ik, l}, T_{ij}, N_i, \{X_{il}\}_{l=1}^d$, $1 \leq j \leq N_i$, $1 \leq i \leq n$, $k = 1, 2, \dots$,

$$\begin{aligned} \mathbf{E}(\eta_{J,l} \eta_{J',l}) &= \mathbf{E} \left[(\bar{R}_{\xi, J, l} + \bar{R}_{\varepsilon, J, l})^{-1/2} (\bar{R}_{\xi, J', l} + \bar{R}_{\varepsilon, J', l})^{-1/2} \right. \\ &\quad \times N_T^{-1} \mathbf{E} \left\{ \left(\sum_{l''=1}^d \sum_{i=1}^n \sum_{k=1}^{\infty} R_{ik, \xi, J, l'', l} \xi_{ik, l''} \right) \left(\sum_{l''=1}^d \sum_{i=1}^n \sum_{k=1}^{\infty} R_{ik, \xi, J', l'', l} \xi_{ik, l''} \right) \right. \\ &\quad \left. \left. + \left(\sum_{i=1}^n \sum_{j=1}^{N_i} R_{ij, \varepsilon, J, l} \varepsilon_{ij} \right) \left(\sum_{i=1}^n \sum_{j=1}^{N_i} R_{ij, \varepsilon, J', l} \varepsilon_{ij} \right) \middle| (T_{ij}, N_i, X_{il})_{i=1, j=1, l=1}^{n, N_i, d} \right\} \right] \\ &= \mathbf{E} C_{n, J, J', l}, \end{aligned}$$

in which

$$C_{n, J, J', l} = (\bar{R}_{\xi, J, l} + \bar{R}_{\varepsilon, J, l})^{-1/2} (\bar{R}_{\xi, J', l} + \bar{R}_{\varepsilon, J', l})^{-1/2} \left\{ N_T^{-1} \sum_{l''=1}^d \sum_{i=1}^n \sum_{k=1}^{\infty} R_{ik, \xi, J, l'', l} R_{ik, \xi, J', l'', l} \right\}.$$

Note that according to definitions of $R_{ik,\xi,J,l''}, R_{ij,\varepsilon,J,l}$, and Lemma B.5, for $0 \leq J \leq N_s$

$$\bar{R}_{\xi,J(t),l} + \bar{R}_{\varepsilon,J(t),l} \geq \bar{R}_{\varepsilon,J(t),l} \geq ER_{ij,\varepsilon,J,l}^2 - A_{n,\varepsilon} \geq c_R - A_{n,\varepsilon},$$

$$P \left[\inf_{0 \leq J \neq J' \leq N_s} \{(\bar{R}_{\xi,J,l} + \bar{R}_{\varepsilon,J,l})(\bar{R}_{\xi,J',l} + \bar{R}_{\varepsilon,J',l})\} \geq \left(c_R - \delta \sqrt{\frac{\log(n)}{nh_s}} \right)^2 \right] \geq 1 - 2n^{-8}.$$

Thus for large n , with probability $\geq 1 - 2n^{-8}$, the denominator of $C_{n,J,J',l}$ is uniformly greater than $c_R^2/4$. On the other hand, we consider the numerator of $C_{n,J,J',l}$.

$$\begin{aligned} \mathbb{E} \left(N_T^{-1} \sum_{l''=1}^d \sum_{i=1}^n \sum_{k=1}^{\infty} R_{ik,\xi,J,l''} R_{ik,\xi,J',l''} \right) &= \mathbb{E} \left\{ N_T^{-1} \sum_{l''=1}^d \sum_{i=1}^n \left(\sum_{l'=1}^d z_{l'l''} X_{il'} X_{il''} \right)^2 \right. \\ &\quad \left. \times \left(\sum_{j=1}^{N_i} \sum_{j'=1}^{N_i} B_J(T_{ij}) B_{J'}(T_{ij'}) G_{l''}(T_{ij}, T_{ij'}) \right) \right\} \sim h_s. \end{aligned}$$

Applying Bernstein's inequality, there exists $C_0 > 0$ such that, for large n ,

$$P \left(\sup_{0 \leq J \neq J' \leq N_s} \left| N_T^{-1} \sum_{l''=1}^d \sum_{i=1}^n \sum_{k=1}^{\infty} R_{ik,\xi,J,l''} R_{ik,\xi,J',l''} \right| \leq C_0 h_s \right) \geq 1 - 2n^{-8}.$$

Putting the above together, for large n , $C_1 = C_0 (c_R^2/4)^{-1}$,

$$P \left(\sup_{0 \leq J \neq J' \leq N_s} |C_{n,J,J',l}| \leq C_1 h_s \right) \geq 1 - 4n^{-8}.$$

Note that as a continuous random variable, $\sup_{0 \leq J \neq J' \leq N_s} |C_{n,J,J',l}| \in [0, 1]$, thus

$$\mathbb{E} \left(\sup_{0 \leq J \neq J' \leq N_s} |C_{n,J,J',l}| \right) = \int_0^1 P \left(\sup_{0 \leq J \neq J' \leq N_s} |C_{n,J,J',l}| > u \right) du.$$

For large n , $C_1 h_s < 1$ and then $\mathbb{E} \left(\sup_{0 \leq J \neq J' \leq N_s, l} |C_{n,J,J',l}| \right)$ is

$$\begin{aligned} &\int_0^{C_1 h_s} P \left\{ \sup_{0 \leq J \neq J' \leq N_s, l} |C_{n,J,J',l}| > u \right\} du + \int_{C_1 h_s}^1 P \left\{ \sup_{0 \leq J \neq J' \leq N_s, l} |C_{n,J,J',l}| > u \right\} du \\ &\leq \int_0^{C_1 h_s} 1 du + \int_{C_1 h_s}^1 4n^{-8} du \leq C_1 h_s + 4n^{-8} \leq C h_s \end{aligned}$$

for some $C > 0$ and large enough n . The lemma now follows from

$$\sup_{0 \leq J \neq J' \leq N_s} |\mathbb{E}(C_{n,J,J',l})| \leq \mathbb{E} \left(\sup_{0 \leq J \neq J' \leq N_s} |C_{n,J,J',l}| \right) \leq C h_s.$$

This completes the proof of the lemma. \square

LEMMA B.8. Under Assumptions (A2)-(A6), for $\eta_l(t), \sigma_{n,ll}(t), l = 1, \dots, d$, defined in (A.6) and (7), one has $\left| \sigma_{n,ll}(t)^{-1} \left\{ \hat{\xi}_l(t) + \hat{\varepsilon}_l(t) \right\} - \eta_l(t) \right| = |r_{n,l}(t) - 1| |\eta_l(t)|$, where $r_{n,l}(t) = \sigma_{n,ll}^{-1}(t) \left\{ \sigma_{\xi_l,n}^2(t) + \sigma_{\varepsilon_l,n}^2(t) \right\}^{1/2}$, and as $n \rightarrow \infty$,

$$\sup_{t \in [0,1]} \{a_{N_s+1} |r_{n,l}(t) - 1|\} = \mathcal{O}_{a.s.} \left\{ (nh_s)^{-1/2} (\log(N_s + 1) \log(n))^{1/2} \right\}.$$

PROOF. By Lemma B.5, $\sigma_{n,ll}^2(t)$ in (7) can be rewritten as

$$\begin{aligned} \sigma_{n,ll}^2(t) &= c_{J(t),n}^{-1} (n \mathbb{E} N_1)^{-1} \left\{ (\mathbb{E} N_1)^{-1} \sum_{l''=1}^d \mathbb{E} \left(\sum_{k=1}^{\infty} R_{ik,\xi,J(t),l'',l}^2 \right) + \mathbb{E} R_{ij,\varepsilon,J(t),l}^2 \right\} \\ &\sim n^{-1} h_s^{-1}. \end{aligned}$$

Hence, according to (B.10) and (10),

$$\begin{aligned} \sup_{t \in [0,1]} \{a_{N_s+1} |r_{n,l}(t) - 1|\} &= \sup_{t \in [0,1]} \left\{ a_{N_s+1} \left| \sigma_{n,ll}^{-1}(t) \left\{ \sigma_{\xi_l,n}^2(t) + \sigma_{\varepsilon_l,n}^2(t) \right\}^{1/2} - 1 \right| \right\} \\ &\leq \sup_{t \in [0,1]} \left\{ a_{N_s+1} \left| \sigma_{n,ll}^{-2}(t) \left\{ \sigma_{\xi_l,n}^2(t) + \sigma_{\varepsilon_l,n}^2(t) \right\} - 1 \right| \right\} \\ &= \sup_{t \in [0,1]} \left\{ a_{N_s+1} \sigma_{n,ll}^{-2}(t) \left| \sigma_{\xi_l,n}^2(t) + \sigma_{\varepsilon_l,n}^2(t) - \sigma_{n,ll}^2(t) \right| \right\} \\ &= \mathcal{O}_{a.s.} \left\{ (nh_s)^{-1/2} (\log(N_s + 1) \log(n))^{1/2} \right\}. \end{aligned}$$

This completes the proof. \square

SFB 649 Discussion Paper Series 2014

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 001 "Principal Component Analysis in an Asymmetric Norm" by Ngoc Mai Tran, Maria Osipenko and Wolfgang Karl Härdle, January 2014.
- 002 "A Simultaneous Confidence Corridor for Varying Coefficient Regression with Sparse Functional Data" by Lijie Gu, Li Wang, Wolfgang Karl Härdle and Lijian Yang, January 2014.

SFB 649, Spandauer Straße 1, D-10178 Berlin
<http://sfb649.wiwi.hu-berlin.de>

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

