A SIMPLE METHOD TO CONSTRUCT CONFIDENCE BANDS IN FUNCTIONAL LINEAR REGRESSION

Masaaki Imaizumi and Kengo Kato

Institute of Statistical Mathematics and Cornell University

Supplementary Material

The supplementary material contains an extension of the methodology to cases with additional regressors, additional numerical experiments, and all the proofs.

S1 Extension to cases with additional regressors

In some applications, we may want to include a finite dimensional vector regressor $Z = (Z_1, \dots, Z_d)^T \in \mathbb{R}^d$ which we assume to include the constant $Z_1 \equiv 1$, in addition to a functional regressor X (cf. Shin, 2009; Kong et al., 2016). Consider the model

$$Y = Z^T \gamma + \int_I b(t) X(t) dt + \varepsilon, \tag{S1.1}$$

where ε is independent of (Z, X) with mean zero and variance $\sigma^2 \in (0, \infty)$, and $\gamma \in \mathbb{R}^d$ and $b \in L^2(I)$ are unknown parameters. We shall discuss how to modify our methodology to construct a confidence band for b in the model (S1.1). In the following discussion, we will assume that $\mathrm{E}(Z_j^2) < \infty$ for all $j = 2, \dots, d$, $E(\|X\|^2) < \infty$, and the matrix $\mathrm{E}(ZZ^T)$ is invertible.

The idea here is to partial out the effect of Z from X. To this end, consider $X^c(t) = X(t) - Z^T \Upsilon(t)$ with $\Upsilon(t) = \{\mathrm{E}(ZZ^T)\}^{-1}\mathrm{E}\{ZX(t)\}$, and observe that $Y = Z^T\gamma^c + \int_I b(t)X^c(t)dt + \varepsilon$ where $\gamma^c = \gamma + \{\mathrm{E}(ZZ^T)\}^{-1}\mathrm{E}(Z\langle b, X\rangle)$. Let K denote the covariance function of X^c, namely, $K(s,t) = \mathrm{E}\{X^c(s)X^c(t)\}$ for $s, t \in I$ (note that $E\{X^c(t)\} = 0$ since $Z_1 \equiv 1$), and assume that the integral operator from $L^2(I)$ into itself with kernel K is injective, so that K admits the spectral expansion $K(s,t) = \sum_j \kappa_j \phi_j(s)\phi_j(t)$ where $\kappa_1 \ge \kappa_2 \ge \cdots > 0$ and $\{\phi_j\}$ is an orthonormal basis of $L^2(I)$. Expanding b and X^c as $b = \sum_j b_j\phi_j$ and $X^c = \sum_j \xi_j \phi_j$ with $b_j = \langle b, \phi_j\rangle$ and $\xi_j = \langle X^c, \phi_j\rangle$, we have

$$Y = Z^T\gamma^c + \sum_j b_j\xi_j + \varepsilon.$$

Importantly, each ξ_j and Z are uncorrelated, namely, $\mathrm{E}(\xi_j Z) = 0$, so that we have $b_j = E(\xi_j Y)/\kappa_j$ as before.

To estimate b, we shall first estimate K. Let $(Y_1, Z_1, X_1), \dots, (Y_n, Z_n, X_n)$ be independent copies of (Y, Z, X), and estimate $X_i^c(t)$ by $\widehat{X}_i^c(t) = X_i(t) -$

$Z_i^T \widehat{\Upsilon}(t)$ with

$$\widehat{\Upsilon}(t) = \left\{ n^{-1} \sum_{j=1}^{n} Z_j Z_j^T \right\}^{-1} \left\{ n^{-1} \sum_{j=1}^{n} Z_j X_j(t) \right\}.$$

Now, we estimate K by $\widehat{K}(s,t) = n^{-1} \sum_{i=1}^{n} \widehat{X}_i^c(s) \widehat{X}_i^c(t)$ for $s, t \in I$, and let $\widehat{K}(s,t) = \sum_j \widehat{\kappa}_j \widehat{\phi}_j(s) \widehat{\phi}_j(t)$ be the spectral expansion of $\widehat{K}$ where $\widehat{\kappa}_1 \geq \widehat{\kappa}_2 \geq \cdots \geq 0$ and $\{\widehat{\phi}_j\}$ is an orthonormal basis of $L^2(I)$. Under this notation, the rest of the procedure is the same as before (replace $X_i - \overline{X}$ by $\widehat{X}_i^c$). Namely, estimate each b_j by $\widehat{b}_j = n^{-1} \sum_{i=1}^{n} Y_i \widehat{\xi}_{i,j} / \widehat{\kappa}_j$ with $\widehat{\xi}_{i,j} = \langle \widehat{X}_i^c, \widehat{\phi}_j \rangle$, and estimate b by $\widehat{b} = \sum_{j=1}^{m_n} \widehat{b}_j \widehat{\phi}_j$. In construction of confidence bands, estimation of the error variance σ^2 is needed. We propose to estimate σ^2 by $\widehat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} (Y_i - Z_i^T \widehat{\gamma}^c - \langle \widehat{b}, \widehat{X}_i^c \rangle)^2$, where $\widehat{\gamma}^c = \{n^{-1} \sum_{i=1}^{n} Z_i Z_i^T\}^{-1} \{n^{-1} \sum_{i=1}^{n} Z_i Y_i\}$.

S2 Additional numerical experiments

In this appendix, we present additional experiments for comparison of our band with the MS band. In the additional experiments, we assume that the eigenfunctions ϕ_j are known to reduce the effect of estimating the eigenfunctions ϕ_j. In addition, we select $\widehat{m}_n$ according to an AIC-type criterion suggested in Müller & Stadmüller (2005). To ensure undersmoothing, we also examine $\widehat{m}_n + 1$ for the cutoff level. Coverage probabilities of the experiments are plotted in Figure 1. UCPs and MCPs of both bands are (overall) better than the previous case of unknown ϕ_j, but still our band performs better than the MS band in terms of coverage probabilities.

S3 Proofs

S3.1 Proof of Theorem 1

We first prove the following technical lemma, which is concerned with concentration and anti-concentration of a weighted sum of independent $\chi^2(1)$ random variables.

Lemma 1. *Let $\eta_1, \dots, \eta_m$ be independent $\chi^2(1)$ random variables, and let $a_1, \dots, a_m$ be nonnegative constants such that $\sigma_a^2 = \sum_{j=1}^m a_j^2 > 0$.*

(i) (Anti-concentration). For every $z > 0$ and $h > 0$,

$$\mathrm{P}\left(\left| \sum_{j=1}^m a_j \eta_j - z \right| \le h \right) \le 2\sqrt{2h/(\sigma_a \pi)},$$

where $\sigma_a = \sqrt{\sigma_a^2}$.

(ii) (Concentration). For every $c > 0$ and $r > 0$,

$$\mathrm{P}\left\{ \sum_{j=1}^m a_j \eta_j \ge (1+c) \sum_{j=1}^m a_j + 2(1+c^{-1}) a_{\max} r \right\} \le e^{-r},$$

where $a_{\max} = \max_{1 \le j \le m} a_j$.

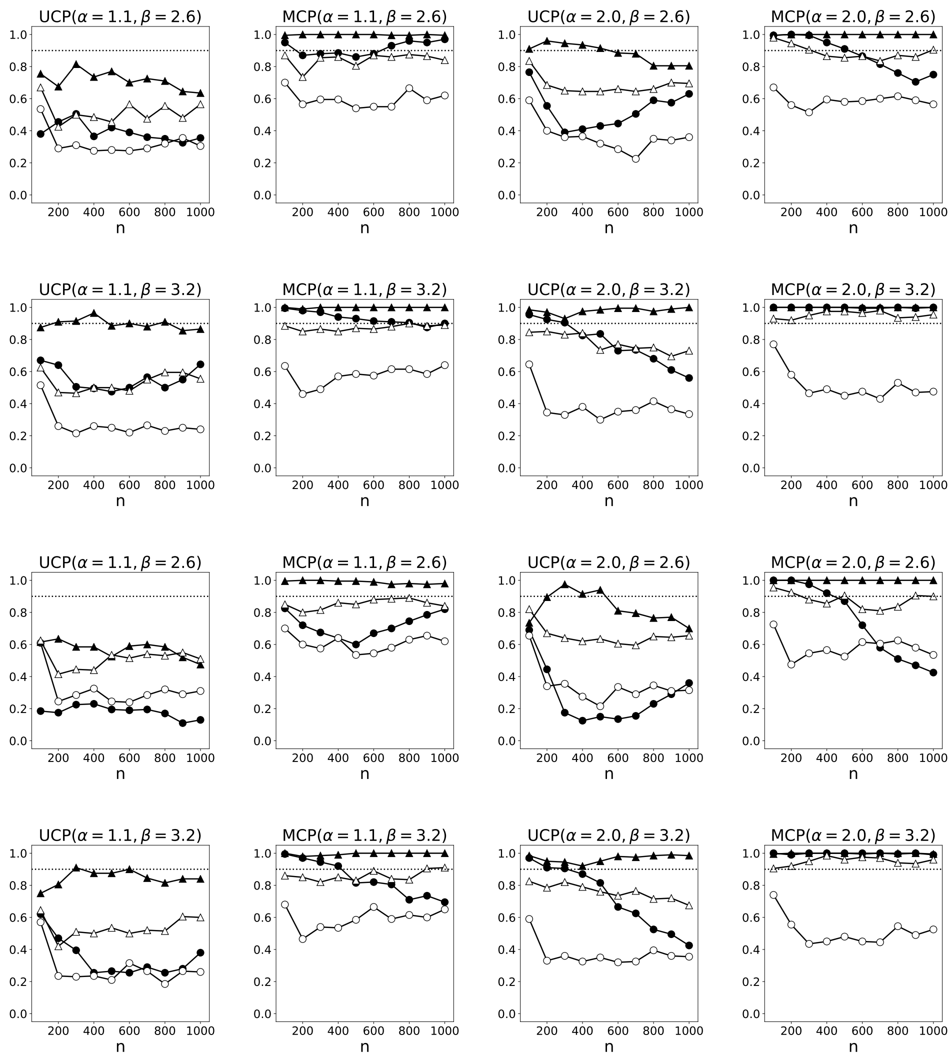

Figure 1: Coverage probabilities with Gaussian noise (upper 2 rows) and χ^2 noise (lower 2 rows). The black markers show coverage probabilities of our band, and the white markers show those of the MS band with AIC. Circles correspond to cases with cutoff level $\max\{\widehat{m}_n, 2\}$, triangles to those with $\widehat{m}_n+1$. The dashed line shows the value $1 - \tau_1 = 0.90$.

Proof. Part (i) follows from Lemma 7.2 in Xu, Zhang & Wu (2014), and Part (ii) is derived from the Gaussian concentration inequality. For the sake of completeness, we provide their proofs.

Part (i). Since $\sup_{z>0} \mathrm{P}(|\sum_{j=1}^m a_j\eta_j - z| \le h) = \sup_{z>0} \mathrm{P}(|\sum_{j=1}^m (a_j/\sigma_a)\eta_j - z| \le h/\sigma_a)$ for every $h > 0$, it suffices to prove the desired inequality when $\sigma_a = 1$. Furthermore, without loss of generality, we may assume that $a_1 = \max_{1\le j\le m} a_j$. Let $V = \sum_{j=1}^m a_j\eta_j$. If $a_1 \le 1/2$, then from the proof of Lemma 7.2 in Xu, Zhang & Wu (2014), the density of V is bounded by 1, so that $\mathrm{P}(|V - z| \le h) \le 2h$. Consider the case where $a_1 > 1/2$, and let $V_{-1} = \sum_{j=2}^m a_j\eta_j$ (if $m = 1$, then let $V_{-1} = 0$). Since η_1 and V_{-1} are independent, we have for every $z > 0$ and $h > 0$,

$$
\begin{aligned}
&\mathrm{P}(|V - z| \le h) \\
&= \mathrm{P}(|\eta_1 - (z - V_{-1})/a_1| \le h/a_1) \le \mathrm{P}(|\eta_1 - (z - V_{-1})/a_1| \le 2h) \\
&\quad = \mathrm{E}[\mathrm{P}(|\eta_1 - (z - V_{-1})/a_1| \le 2h \mid V_{-1})] \le \sup_{z'\in\mathbb{R}} \mathrm{P}(|\eta_1 - z'| \le 2h).
\end{aligned}
$$

Pick any $z' \in \mathbb{R}$. Suppose first that $z' - 2h > 0$. Since $\eta_1 \sim \chi^2(1)$, we have that

$$
\begin{aligned}
&\mathrm{P}(|\eta_1 - z'| \le 2h) \\
&= \sqrt{1/(2\pi)} \int_{z'-2h}^{z'+2h} w^{-1/2} e^{-w/2} dw \le \sqrt{1/(2\pi)} \int_{z'-2h}^{z'+2h} w^{-1/2} dw \\
&\quad = \sqrt{2/\pi}(\sqrt{z' + 2h} - \sqrt{z' - 2h}) \le 2\sqrt{2h/\pi}.
\end{aligned}
$$

On the other hand, if $z' - 2h \le 0$, then $\mathrm{P}(|\eta_1 - z'| \le 2h) \le \mathrm{P}(\eta_1 \le 4h) \le 2\sqrt{2h/\pi}$.

Therefore, in either case of $a_1 \le 1/2$ or $a_1 > 1/2$, we have $\sup_{z>0} \mathrm{P}(|V - z| \le h) \le 2\max\{h, \sqrt{2h/\pi}\}$ for every $h > 0$. This inequality is meaningful only if $h \le 1/2$ since otherwise the upper bound is larger than 1, but if $0 < h \le 1/2$, then $h \le \sqrt{2h/\pi}$. This completes the proof of Part (i).

Part (ii). Let $Z = (Z_1, \ldots, Z_m)^T$ be a standard normal random vector in $\mathbb{R}^m$, and let $F(Z) = \sqrt{\sum_{j=1}^m a_j Z_j^2}$. Then F is Lipschitz continuous with Lipschitz constant bounded by $\sqrt{a_{\max}}$, and $\mathrm{E}\{F(Z)\} \le \sqrt{\mathrm{E}\{F^2(Z)\}} = \sqrt{\sum_{j=1}^m a_j}$. The Gaussian concentration inequality (cf. Boucheron, Lugosi & Massart, 2013, Theorem 5.6) then yields that

$$\mathrm{P}\left\{F(X) \ge \sqrt{\sum_{j=1}^m a_j} + \sqrt{a_{\max}}\, r\right\} \le e^{-r^2/2}$$

for every $r > 0$. The desired conclusion follows from the fact that $F^2(Z)$ has the same distribution as $\sum_{j=1}^m a_j \eta_j$, and the simple inequality $2xy \le cx^2 + c^{-1}y^2$ for any $x, y \in \mathbb{R}$ and $c > 0$. □

Proof of Theorem 1. We will use the following notation. Let P_ε and E_ε denote the probability and expectation with respect to ε_i's only. The notation $\lesssim$ signifies that the left hand side is bounded by the right-hand side up to a constant that depends only on α, β, and C_1. We first note that $\widehat{b}$ is invariant with respect to choices of signs of $\widehat{\phi}_j$'s, and so without loss of generality, we may assume that $\int_I \widehat{\phi}_j(t)\phi_j(t)dt \geq 0$ for all $j = 1, 2, \ldots$. Lemma 4.2 in Bosq (2000) yields that $\sup_{j\geq 1} |\widehat{\kappa}_j - \kappa_j| \leq \widehat{\Delta} := |||\widehat{K} - K|||$. Since $\mathrm{E}\{\|X - \mathrm{E}(X)\|^4\} = \mathrm{E}\{(\sum_{j=1}^{\infty} \xi_j^2)^2\} = \sum_{j,k} \mathrm{E}(\xi_j^2 \xi_k^2) \leq \sum_{j,k} \sqrt{\mathrm{E}(\xi_j^4)}\sqrt{\mathrm{E}(\xi_k^4)} \lesssim (\sum_j \kappa_j)^2 \lesssim 1$ (which follows from the assumption that $\mathrm{E}(\xi_j^4) \lesssim \kappa_j^2$), we have that $\widehat{\Delta} = O_{\mathrm{P}}(n^{-1/2})$. Observe that for $1 \leq j \leq m_n$, $|\widehat{\kappa}_j/\kappa_j - 1| \lesssim j^{\alpha}|\widehat{\kappa}_j - \kappa_j| \leq m_n^{\alpha}\widehat{\Delta} = o_{\mathrm{P}}(1)$, from which we have $\max_{1\leq j\leq m_n} |\widehat{\kappa}_j/\kappa_j - 1| = o_{\mathrm{P}}(1)$. Furthermore, observe that, whenever $1 \leq j \leq m_n$ and $j \neq k$, $|\kappa_j - \kappa_k| \geq \min\{\kappa_{j-1} - \kappa_j, \kappa_j - \kappa_{j+1}\} \geq C_1^{-1} j^{-\alpha-1} \geq C_1^{-1} m_n^{-\alpha-1}$, and since $n^{-1/2} = o(m_n^{-\alpha-1})$, we have that $\mathrm{P}\{|\widehat{\kappa}_j - \kappa_k| \geq |\kappa_j - \kappa_k|/\sqrt{2}$, for all (j, k) such that $1 \leq j \leq m_n, k \neq j\} \to 1$. Now, following the arguments used in Hall & Horowitz (2007, p.83-84), we have that with probability approaching one,

$$
\begin{aligned}
&(1 - C m_n^{2\alpha+2}\widehat{\Delta}_n^2)\|\widehat{\phi}_j - \phi_j\|^2 \\
&\quad \leq \underbrace{8 \sum_{k:k\neq j} (\kappa_j - \kappa_k)^{-2} \left[\int \{\widehat{K}(s,t) - K(s,t)\}\phi_j(s)\phi_k(t)dsdt\right]^2}_{=\widehat{u}_j^2},
\end{aligned}
$$

for all j such that $1 \leq j \leq m_n$. Here, C is a constant that depends only on C_1, and $\mathrm{E}(\widehat{u}_j^2) \lesssim j^2/n$. Since $m_n^{2\alpha+2}\widehat{\Delta}^2 = o_{\mathrm{P}}(1)$, we conclude that

$$
\|\widehat{\phi}_j - \phi_j\|^2 \leq 8\{1 + o_{\mathrm{P}}(1)\}\widehat{u}_j^2 \quad \text{and} \quad \mathrm{E}(\widehat{u}_j^2) \lesssim j^2/n, \tag{S3.2}
$$

where $o_{\mathrm{P}}(1)$ is uniform in $1 \le j \le m_n$. We divide the rest of the proof into several steps.

Step 1. In this step, we shall verify the expansion (3.5). Since $\{\widehat{\phi}_j\}_{j=1}^{\infty}$ is an orthonormal basis of $L^2(I)$, expand b as $b = \sum_j \breve{b}_j \widehat{\phi}_j$ with $\breve{b}_j = \int_I b(t)\widehat{\phi}_j(t)dt$. Arguing as in the proof of Theorem 1 in Imaizumi & Kato (2016), we have that $\widehat{b}_j = \breve{b}_j + n^{-1}\sum_{i=1}^n \varepsilon_i \widehat{\xi}_{i,j}/\widehat{\kappa}_j$ and $\sum_{j=1}^{m_n}(\breve{b}_j - b_j)^2 = O_{\mathrm{P}}(n^{-1})$. Now, observe that

$$
\begin{aligned}
&\widehat{b} - b \\
&= \sum_{j=1}^{m_n}\left(n^{-1}\sum_{i=1}^n \varepsilon_i \widehat{\xi}_{i,j}/\widehat{\kappa}_j\right)\widehat{\phi}_j + \sum_{j=1}^{m_n}(\breve{b}_j - b_j)\widehat{\phi}_j + \sum_{j=1}^{m_n} b_j(\widehat{\phi}_j - \phi_j) + \sum_{j>m_n} b_j\phi_j \\
&=: I_n + II_n + III_n + IV_n.
\end{aligned}
$$

Since

$$
\begin{aligned}
&\mathrm{E}_{\varepsilon}(\|I_n\|^2) \\
&= \sum_{j=1}^{m_n}\mathrm{E}_{\varepsilon}\left\{\left(n^{-1}\sum_{i=1}^n \varepsilon_i\widehat{\xi}_{i,j}/\widehat{\kappa}_j\right)^2\right\} = (\sigma^2/n)\sum_{j=1}^{m_n}\left(n^{-1}\sum_{i=1}^n \widehat{\xi}_{i,j}^2/\widehat{\kappa}_j^2\right) \\
&= (\sigma^2/n)\sum_{j=1}^{m_n}\widehat{\kappa}_j^{-1} = O_{\mathrm{P}}\left(n^{-1}\sum_{j=1}^{m_n}\kappa_j^{-1}\right) = O_{\mathrm{P}}(m_n^{\alpha+1}/n),
\end{aligned}
$$

we have that $\|I_n\|^2 = O_{\mathrm{P}}(m_n^{\alpha+1}/n)$. Furthermore, observe that

$$
\|II_n\|^2 = \sum_{j=1}^{m_n}(\breve{b}_j - b_j)^2 = O_{\mathrm{P}}(n^{-1}),\ \|IV_n\|^2 \lesssim \sum_{j>m_n} j^{-2\beta} = O(m_n^{-2\beta+1}),\ \text{and}
$$

$$
\|III_n\|^2 \lesssim m_n\sum_{j=1}^{m_n} j^{-2\beta}\|\widehat{\phi}_j - \phi_j\|^2 = O_{\mathrm{P}}\left\{(m_n/n)\sum_{j=1}^{m_n} j^{-2\beta+2}\right\} = O_{\mathrm{P}}(m_n/n).
$$

Therefore, we have

$$\begin{aligned}\|\widehat{b}-b\|^2 &= \|I_n\|^2 + 2\langle I_n, II_n + III_n + IV_n\rangle + \|II_n + III_n + IV_n\|^2 \\ &= \|I_n\|^2 + O_{\mathrm{P}}(m_n^{\alpha/2+1}/n + n^{-1/2}m_n^{-\beta+\alpha/2+1} + m_n^{-2\beta+1}).\end{aligned}$$

This leads to the expansion (3.5).

Step 2. In this step, we shall show that for any fixed $\tau \in (0,1)$,

$$\mathrm{P}\{n\|\widehat{b}-b\|^2 \le \sigma^2 \widehat{c}_n^2(1-\tau)\} = 1-\tau+o(1).$$

Define $R_n = n(\|\widehat{b}-b\|^2 - \|I_n\|^2)$, and observe that $R_n = o_{\mathrm{P}}(m_n^{\alpha+1/2})$. So there exists a sequence of constants $\delta_n \downarrow 0$ such that $\mathrm{P}(|R_n| > \delta_n m_n^{\alpha+1/2}) \to 0$. Now, observe that

$$\begin{aligned}&\mathrm{P}_\varepsilon\left\{n\|\widehat{b}-b\|^2 \le \sigma^2\widehat{c}_n^2(1-\tau)\right\} \\ &\quad \le \mathrm{P}_\varepsilon\left\{n\|I_n\|^2 \le \sigma^2\widehat{c}_n^2(1-\tau) + \delta_n m_n^{\alpha+1/2}\right\} + \mathrm{P}_\varepsilon(|R_n| > \delta_n m_n^{\alpha+1/2}),\end{aligned}$$

and conditionally on X_1^n, $n\|I_n\|^2$ has the same distribution as $\sigma^2\sum_{j=1}^{m_n}\eta_j/\widehat{\kappa}_j$, where $\eta_1,\dots,\eta_{m_n}$ are independent $\chi^2(1)$ random variables independent of X_1^n. Lemma 1 (i) then yields that

$$\mathrm{P}_\varepsilon\left\{n\|I_n\|^2 \le \sigma^2\widehat{c}_n^2(1-\tau) + \delta_n m_n^{\alpha+1/2}\right\} - (1-\tau) \lesssim \left\{\frac{\delta_n m_n^{\alpha+1/2}}{(\sum_{j=1}^{m_n}\widehat{\kappa}_j^{-2})^{1/2}}\right\}^{1/2}.$$

Since $\sum_{j=1}^{m_n}\widehat{\kappa}_j^{-2} \ge \{1-o_{\mathrm{P}}(1)\}\sum_{j=1}^{m_n}\kappa_j^{-2} \gtrsim \{1-o_{\mathrm{P}}(1)\}m_n^{2\alpha+1}$, the right-hand side on the above displayed equation is $o_{\mathrm{P}}(1)$. This yields that $\mathrm{P}_\varepsilon\{n\|\widehat{b}-$

$b\|^2 \le \sigma^2\widehat{c}_n^2(1-\tau)\} \le 1-\tau+o_{\mathrm{P}}(1)$. Likewise, we have $\mathrm{P}_\varepsilon\{n\|\widehat{b}-b\|^2 \le \sigma^2\widehat{c}_n^2(1-\tau)\} \ge 1-\tau-o_{\mathrm{P}}(1)$, so that

$$\mathrm{P}_\varepsilon\left\{n\|\widehat{b}-b\|^2 \le \sigma^2\widehat{c}_n^2(1-\tau)\right\} = 1-\tau+o_{\mathrm{P}}(1).$$

Finally, Fubini's theorem and the dominated convergence theorem yield that $\mathrm{P}\{n\|\widehat{b}-b\|^2 \le \sigma^2\widehat{c}_n^2(1-\tau)\} = 1-\tau+o(1)$.

Step 3. In this step, we shall show that $\widehat{\sigma}^2 = \sigma^2 + o_{\mathrm{P}}(m_n^{-1/2})$. Observe that

$$Y_i - \overline{Y} - \sum_{j=1}^{m_n}\widehat{b}_j\widehat{\xi}_{i,j} = \int_I \{X_i(t)-\overline{X}(t)\}\{b(t)-\widehat{b}(t)\}dt + \varepsilon_i - \overline{\varepsilon},$$

where $\overline{\varepsilon} = n^{-1}\sum_{i=1}^n \varepsilon_i$, so that

$$\begin{aligned}\widehat{\sigma}^2 &= n^{-1}\sum_{i=1}^n(\varepsilon_i-\overline{\varepsilon})^2 + 2\int_I\left[n^{-1}\sum_{i=1}^n(\varepsilon_i-\overline{\varepsilon})\{X_i(t)-\overline{X}(t)\}\right]\{b(t)-\widehat{b}(t)\}dt\\ &\quad + n^{-1}\sum_{i=1}^n\left[\int_I\{X_i(t)-\overline{X}(t)\}\{b(t)-\widehat{b}(t)\}dt\right]^2\end{aligned}$$

From Step 1, it is seen that $\|\widehat{b}-b\|^2 = O_{\mathrm{P}}(m_n^{\alpha+1}/n)$, so that by the Cauchy-Schwarz inequality, the second and third terms on the right-hand side are $O_{\mathrm{P}}(m_n^{\alpha/2+1/2}/n)$ and $O_{\mathrm{P}}(m_n^{\alpha+1}/n)$, respectively. Furthermore, $n^{-1}\sum_{i=1}^n(\varepsilon_i - \overline{\varepsilon})^2 = \sigma^2 + O_{\mathrm{P}}(n^{-1/2})$. The conclusion of this step follows from the fact that $n^{-1/2} + m_n^{\alpha+1}/n = o(m_n^{-1/2})$.

Step 4. In this step, we shall show that for any fixed $\tau \in (0,1)$,

$$\mathrm{P}\{n\|\widehat{b}-b\|^2 \le \widehat{\sigma}^2\widehat{c}_n^2(1-\tau)\} = 1-\tau+o(1). \tag{S3.3}$$

By Lemma 1 (ii), we have $\widehat{c}_n^2(1-\tau) \lesssim \sum_{j=1}^{m_n} \widehat{\kappa}_j^{-1} + \widehat{\kappa}_{m_n}^{-1} \log(1/\tau) = O_{\mathrm{P}}(m_n^{\alpha+1})$, so that

$$\widehat{\sigma}^2 \widehat{c}_n^2(1-\tau) = \sigma^2 \widehat{c}_n^2(1-\tau) + (\widehat{\sigma}^2 - \sigma^2)\widehat{c}_n^2(1-\tau) = \sigma^2 \widehat{c}_n^2(1-\tau) + o_{\mathrm{P}}(m_n^{\alpha+1/2}).$$

Hence, arguing as in Step 2, we obtain the result (S3.3).

In view of the discussion in Section 2.2, the result (2.9) follows directly from (S3.3). Finally, the width of the band $\widehat{\mathcal{C}}$ is $\lesssim \widehat{\sigma}\widehat{c}_n(1-\tau_1)/\sqrt{n} = O_{\mathrm{P}}(\sqrt{m_n^{\alpha+1}/n})$. This completes the proof. □

S3.2 Proof of Theorem 2

The proof of Theorem 2theorem.2 relies on the following multi-dimensional version of the Berry-Esseen bound, due to Bentkus (2005). Let $\|\cdot\|_2$ denote the standard Euclidean norm.

Theorem 1 (Bentkus (2005))**.** *Let $W_1, \dots, W_n$ be independent random vectors in $\mathbb{R}^m$ with mean zero, and suppose that the covariance matrix Σ of $S_n = \sum_{i=1}^n W_i$ is invertible. Then there exists a universal constant $c > 0$ such that*

$$\sup_{A \in \mathcal{C}} |\mathrm{P}(S_n \in A) - \gamma_\Sigma(A)| \le c m^{1/4} \sum_{i=1}^n \mathrm{E}(\|\Sigma^{-1/2} W_i\|_2^3),$$

where $\mathcal{C}$ is the class of all Borel measurable convex sets in $\mathbb{R}^m$, and $\gamma_\Sigma = N(0, \Sigma)$.

We will also use the following well-known inequality.

Lemma 2. *Let* $\zeta_1, \ldots, \zeta_n$ *be random variables such that* $\mathrm{E}(|\zeta_i|^r) < \infty$ *for all* $i = 1, \ldots, n$ *for some* $r \geq 1$. *Then,*

$$\mathrm{E}\left(\max_{1 \leq i \leq n} |\zeta_i|\right) \leq n^{1/r} \max_{1 \leq i \leq n} \{\mathrm{E}(|\zeta_i|^r)\}^{1/r}.$$

This inequality follows from the observation that

$$\mathrm{E}(\max_{1 \leq i \leq n} |\zeta_i|) \leq \{\mathrm{E}(\max_{1 \leq i \leq n} |\zeta_i|^r)\}^{1/r} \leq \left\{\sum_{i=1}^n \mathrm{E}(|\zeta_i|^r)\right\}^{1/r} \leq n^{1/r} \max_{1 \leq i \leq n} \{\mathrm{E}(|\zeta_i|^r)\}^{1/r}.$$

We are now in position to prove Theorem 2theorem.2.

Proof of Theorem 2. We follow the notation used in the proof of Theorem 1theorem.1. In view of the proof of Theorem 1theorem.1, it suffices to show that

$$\sup_{z>0} \left| \mathrm{P}_\varepsilon(n\|I_n\|^2/\sigma^2 \leq z) - \mathrm{P}_\eta\left(\sum_{j=1}^{m_n} \eta_j/\widehat{\kappa}_j \leq z\right) \right| \xrightarrow{\mathrm{P}} 0,$$

where P_η denotes the probability with respect to η_j's only. To this end, let $W_i = \{\varepsilon_i \widehat{\xi}_{i,j}/(\sigma\sqrt{n}\widehat{\kappa}_j)\}_{j=1}^{m_n}$ for $i = 1, \ldots, n$. Observe that the covariance matrix of $\sum_{i=1}^n W_i$ conditionally on X_1^n is $\Lambda_n = \mathrm{diag}(1/\widehat{\kappa}_1, \ldots, 1/\widehat{\kappa}_{m_n})$, and $n\|I_n\|^2/\sigma^2 = \|\sum_{i=1}^n W_i\|_2^2$. For $z > 0$, let $B_z = \{w \in \mathbb{R}^{m_n} : \|w\|_2^2 \leq z\}$, and observe that $\mathrm{P}_\eta(\sum_{j=1}^{m_n} \eta_j/\widehat{\kappa}_j \leq z) = \gamma_{\Lambda_n}(B_z)$. Therefore, the problem reduces to proving that

$$\sup_{z>0} \left| \mathrm{P}_\varepsilon\left(\sum_{i=1}^n W_i \in B_z\right) - \gamma_{\Lambda_n}(B_z) \right| \xrightarrow{\mathrm{P}} 0,$$

but in view of Theorem 1, the left hand side is $\lesssim m_n^{1/4}\sum_{i=1}^n \mathrm{E}_\varepsilon(\|\Lambda_n^{-1/2}W_i\|_2^3)$.

Observe that

$$\sum_{i=1}^n \mathrm{E}_\varepsilon(\|\Lambda_n^{-1/2}W_i\|_2^3) = \mathrm{E}(|\varepsilon/\sigma|^3)n^{-3/2}\sum_{i=1}^n\left(\sum_{j=1}^{m_n}\widehat{\xi}_{i,j}^2/\widehat{\kappa}_j\right)^{3/2}$$
$$\leq O(m_n n^{-1/2})\max_{1\leq i\leq n}\left(\sum_{j=1}^{m_n}\widehat{\xi}_{i,j}^2/\widehat{\kappa}_j\right)^{1/2} \leq O_{\mathrm{P}}(m_n n^{-1/2})\max_{1\leq i\leq n}\left(\sum_{j=1}^{m_n}\widehat{\xi}_{i,j}^2/\kappa_j\right)^{1/2}.$$

We have to bound $\max_{1\leq i\leq n}\sum_{j=1}^{m_n}\widehat{\xi}_{i,j}^2/\kappa_j$, to which end it is without loss of generality to assume that $\mathrm{E}\{X(t)\} = 0$ for all $t \in I$. Let $\xi_{i,j} = \int_I X_i(t)\phi_j(t)dt$, and observe that

$$\widehat{\xi}_{i,j} = \int_I\{X_i(t)-\overline{X}(t)\}\widehat{\phi}_j(t)dt = \xi_{i,j}+\int_I X_i(t)\{\widehat{\phi}_j(t)-\phi_j(t)\}dt-\int_I\overline{X}(t)\widehat{\phi}_j(t)dt.$$

From this decomposition, we have

$$\max_{1\leq i\leq n}\sum_{j=1}^{m_n}\widehat{\xi}_{i,j}^2/\kappa_j$$
$$\lesssim \max_{1\leq i\leq n}\sum_{j=1}^{m_n}\xi_{i,j}^2/\kappa_j + \left(\max_{1\leq i\leq n}\|X_i\|^2\right)\sum_{j=1}^{m_n}\kappa_j^{-1}\|\widehat{\phi}_j-\phi_j\|^2 + \|\overline{X}\|^2\sum_{j=1}^{m_n}\kappa_j^{-1}$$
$$= \max_{1\leq i\leq n}\sum_{j=1}^{m_n}\xi_{i,j}^2/\kappa_j + \left(\max_{1\leq i\leq n}\|X_i\|^2\right)O_{\mathrm{P}}\left(\sum_{j=1}^{m_n}j^{\alpha+2}/n\right) + O_{\mathrm{P}}(m_n^{\alpha+1}/n)$$
$$= \max_{1\leq i\leq n}\sum_{j=1}^{m_n}\xi_{i,j}^2/\kappa_j + \left(\max_{1\leq i\leq n}\|X_i\|^2\right)O_{\mathrm{P}}(m_n^{\alpha+3}/n) + O_{\mathrm{P}}(m_n^{\alpha+1}/n),$$

where we have used (S3.2). Condition (3.13) together with Lemma 2 yield that

$$\mathrm{E}\left(\max_{1\leq i\leq n}\sum_{j=1}^{m_n}\xi_{i,j}^2/\kappa_j\right) \leq \sum_{j=1}^{m_n}\mathrm{E}\left\{\max_{1\leq i\leq n}(\xi_{i,j}^2/\kappa_j)\right\} \leq m_n n^{1/q}C_1^{1/q}.$$

Furthermore, a repeated application of Hölder's inequality yields that

$$\mathrm{E}\{(\xi_{j_1}^2/\kappa_{j_1})\cdots(\xi_{j_q}^2/\kappa_{j_q})\} \le [\mathrm{E}\{(\xi_{j_1}^2/\kappa_{j_1})^q\}]^{1/q}\cdots[\mathrm{E}\{(\xi_{j_q}^2/\kappa_{j_q})^q\}]^{1/q} \le C_1,$$

from which we have

$$\begin{aligned}
&\mathrm{E}(\|X\|^{2q}) \\
&= \mathrm{E}\left\{\left(\sum_{j=1}^{\infty}\xi_j^2\right)^q\right\} = \sum_{j_1=1}^{\infty}\cdots\sum_{j_q=1}^{\infty}(\kappa_{j_1}\cdots\kappa_{j_q})\mathrm{E}\{(\xi_{j_1}^2/\kappa_{j_1})\cdots(\xi_{j_q}^2/\kappa_{j_q})\} \\
&\le C_1\sum_{j_1=1}^{\infty}\cdots\sum_{j_q=1}^{\infty}\kappa_{j_1}\cdots\kappa_{j_q} = C_1\left(\sum_{j=1}^{\infty}\kappa_j\right)^q < \infty.
\end{aligned}$$

This implies that $\mathrm{E}(\max_{1\le i\le n}\|X_i\|^2) = O(n^{1/q})$ by Lemma 2. Therefore, we conclude that $\max_{1\le i\le n}\sum_{j=1}^{m_n}\widehat{\xi}_{i,j}^2/\kappa_j = O_{\mathrm{P}}(m_n n^{1/q})$, so that

$$m_n^{1/4}\sum_{i=1}^{n}\mathrm{E}_{\varepsilon}(\|\Lambda_n^{-1/2}W_i\|_2^3) = O_{\mathrm{P}}\{m_n^{7/4}/n^{1/2-1/(2q)}\},$$

which is $o_{\mathrm{P}}(1)$ under Condition (3.14). This completes the proof. □

References

Bentkus, V. (2005). A Lyapunov-type bound in $\mathbb{R}^d$. *Theory Probab. Appl.* **49**, 311-323.

Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications.* Springer.

Boucheron, S., Lugosi, G. & Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press.

Hall, P. & Horowitz, J.L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35**, 70-91.

Imaizumi, M. & Kato, K. (2016). PCA-based estimation for functional linear regression with functional responses. arXiv:1609.00286.

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

Müller, H.-G. & Stadtmüller, U. (2005). Generalized functional linear models. *Ann. Statist.* **33**, 774-805.

Shin, H. (2009). Partial functional linear regression. *J. Stat. Plan. Infer.* **139**, 3405-3418.

Xu, M., Zhang, D. & Wu, W.B. (2014). L^2 asymptotics for high-dimensional data. arXiv:1405.7244.