Supplemental Materials for "Two-sample Functional Linear Models"

Wenchao Xu[1], Riquan Zhang[1] and Hua Liang[2]

[1] *East China Normal University and* [2] *George Washington University*

In this document, we present a simulation example continued from Section 6, the detailed proofs of the Proposition 1 and Theorems 1–4, and the derivation of the efficient score given in (3.13).

S1 Simulation Study for Sparse and Irregular Data

We use the same simulation setting in Section 6.1 but $\alpha = 1.1$ and $\sigma = 0.5$. We generate longitudinal data from (5.16), in which the time points T_{ij} are i.i.d. from a uniform distribution on $[0, 1]$ and measurement errors ϵ_{ij} are i.i.d. from a zero-mean Gaussian error with variance $\sigma_\epsilon^2 = 0.1^2$. We let $N_i = N$ for all i and compare the estimation results for $N = 5$ and 15, and $n = 200, 350, 500$, and 800. Note that the cases of $N = 5$ may be viewed as representing sparse functional data. For each configuration, we repeated 1000 times.

Table 1 presents the average values and standard deviation of the estimated $\widetilde{\theta}$. For each N, the average of $\widetilde{\theta}$ gets closer to the true value and the standard deviation decreases as n increases. For each n, the performance of $\widetilde{\theta}$ improves as N increases.

Table 1: The results of the simulation study (sparse data). The average and standard deviation (SD) of $\widetilde{\theta}$ given $\theta = 1.5$.

N	$n = 200$		$n = 350$		$n = 500$		$n = 800$	
	mean	SD	mean	SD	mean	SD	mean	SD
5	1.777	0.702	1.626	0.453	1.570	0.323	1.547	0.243
15	1.634	0.410	1.578	0.290	1.537	0.234	1.531	0.178

The MISE and associated standard derivations of $\widetilde{b}(t)$ and $\widehat{b}^{\mathrm{FMR}}(t)$ are displayed in Table 2. For each N, the MISE and the standard deviation of $\widetilde{b}(t)$ decrease as n increases. For each n, the performance of $\widetilde{b}(t)$ improves as N increases. The MISE and the standard deviations of $\widehat{b}^{\mathrm{FMR}}(t)$ are consistently larger than that of $\widetilde{b}(t)$. This means that $\widetilde{b}(t)$ outperforms its competitor $\widehat{b}^{\mathrm{FMR}}(t)$.

Table 2: The results of simulation study (sparse data). The MISE and corresponding standard deviations (SD) of the estimated slope functions $\widetilde{b}(t)$ and $\widehat{b}^{\mathrm{FMR}}(t)$.

		$n = 200$		$n = 350$		$n = 500$		$n = 800$	
N		MISE	SD	MISE	SD	MISE	SD	MISE	SD
5	$\widetilde{b}$	0.552	0.544	0.344	0.305	0.273	0.248	0.199	0.162
	$\widehat{b}^{\mathrm{FMR}}$	1.196	1.986	0.769	1.221	0.760	2.323	0.487	1.360
15	$\widetilde{b}$	0.248	0.154	0.175	0.101	0.139	0.079	0.108	0.052
	$\widehat{b}^{\mathrm{FMR}}$	0.832	0.985	0.737	1.270	0.633	0.773	0.612	1.073

S2 Proof of Proposition 1

Without loss of generality, assume $E(X) = 0$. Then (2.3) is equivalent to

$$P\left\{(1-U)\delta + U\delta' = 0\right\} = 1,$$

where $\delta = (a - a_1) + \int X(b - b_1)$ and $\delta' = \theta a - \theta_1 a_1 + \int X(\theta b - \theta_1 b_1)$. This implies that $E\{(1-U)\delta + U\delta'\}^2 = (1-\pi)E\delta^2 + \pi E\delta'^2 = 0$. Then, we have $a = a_1, \theta a = \theta_1 a_1$, and

$$\sum_{j=1}^{\infty} \lambda_j \left\{\int (b - b_1)\phi_j\right\}^2 = 0, \quad \sum_{j=1}^{\infty} \lambda_j \left\{\int (\theta b - \theta_1 b_1)\phi_j\right\}^2 = 0.$$

These two equations hold if and only if $\lambda_j^{1/2} \int (b - b_1)\phi_j = 0$ and $\lambda_j^{1/2} \int (\theta b - \theta_1 b_1)\phi_j = 0$ for each j. Since K is of full rank, $b = b_1$ and $\theta b = \theta_1 b_1$ almost everywhere on $\mathcal{I}$.

S3 Proof of Theorem 1

Since π is estimated by $\widehat{\pi}$ with a parametric rate, we shall assume that π is known in our proofs for simple notation. Given an univariate function f, let $\|f\|_2 = \{\int_{\mathcal{I}} f^2(t)\, dt\}^{1/2}$ be the standard norm for $L_2(\mathcal{I})$.

Let $\widehat{\Delta}^2 = \iint_{\mathcal{I}^2} (\widehat{K} - K)^2$ and $\delta_j = \min_{k \le j}(\lambda_k - \lambda_{k+1})$. It can be shown from the results of Bhatia, Davis, and McIntosh (1983) that

$$\sup_{j \ge 1} |\widehat{\lambda}_j - \lambda_j| \le \widehat{\Delta}, \quad \sup_{j \ge 1} \delta_j \|\widehat{\phi}_j - \phi_j\|_2 \le 8^{1/2} \widehat{\Delta}. \tag{S3.1}$$

Define the event $\mathcal{E}_{m_n} = \{\widehat{\Delta} \le 2^{-1}\lambda_{m_n}\}$; i.e., the set of all realizations such that, for sample size n, $2^{-1}\lambda_{m_n} \ge \widehat{\Delta}$. A standard moment calculation can be used to prove that $E(\widehat{\Delta}^2) = O(n^{-1})$ as $n \to \infty$ (Hall and Horowitz, 2007). Using this result, Markov's inequality and Assumption (A4), we have $P(\mathcal{E}_{m_n}) \to 1$. Thus, it suffices to work with bounds on $\mathcal{E}_{m_n}$. This strategy was adapted by Hall and Horowitz (2007) as well.

Define

$$S_{j1} = \int (\widehat{g} - g)\phi_j, S_{j2} = \int g(\widehat{\phi}_j - \phi_j), S_{j3} = \int (\widehat{g} - g)(\widehat{\phi}_j - \phi_j);$$
$$T_{j1} = \int (\widehat{h} - h)\phi_j, T_{j2} = \int h(\widehat{\phi}_j - \phi_j), T_{j3} = \int (\widehat{h} - h)(\widehat{\phi}_j - \phi_j).$$

Then

$$\widehat{g}_j = S_{j1} + S_{j2} + S_{j3} + g_j, \quad \widehat{h}_j = T_{j1} + T_{j2} + T_{j3} + h_j. \tag{S3.2}$$

The key step in the proof of Theorem 1 relies on the bounds of S_{jk} and T_{jk} for $k = 1, 2, 3$, which are given in the following lemma.

Lemma 1. *Under Assumptions (A1)–(A4), we have*

$$\left.\begin{aligned} S_{j1} &= O_p(n^{-1/2}\lambda_j^{1/2}), \\ S_{j2} &= O_p(n^{-1/2}j^{-\alpha/2}) + O_p(n^{-1}\sqrt{\log n} + n^{-1}j^{\alpha-\beta+2}) \\ &\qquad + O_p(n^{-1}j) + O_p(n^{-1/2}j^{-\alpha-\beta+1}), \\ S_{j3} &= O_p(n^{-1}j), \end{aligned}\right\} \tag{S3.3}$$

where $O_p(\cdot)$'s in these three equations are uniform in $1 \le j \le m_n$. Note that $S_{j2} = \theta T_{j2}$. The same results hold for T_{jk} for $k = 1, 2, 3$, respectively.

Proof. Let $\xi_{ij} = \int (X_i - \mu_X)\phi_j$ and $\bar{\xi}_j = n^{-1}\sum_i \xi_{ij}$. Thus,

$$S_{j1} = \frac{\theta}{n\pi}\sum_{i=1}^{n}\left[U_i\left(a + \int X_i b\right)\xi_{ij} - E\left\{U_i\left(a + \int X_i b\right)\xi_{ij}\right\} - U_i\left(a + \int X_i b\right)\bar{\xi}_j\right] + \frac{1}{n\pi}\sum_{i=1}^{n}(\varepsilon_i U_i \xi_{ij} - \varepsilon_i U_i \bar{\xi}_j).$$

It can be proved that

$$nES_{j1}^2 \le \text{const}\cdot\left[\text{var}\left\{U\left(a + \int Xb\right)\xi_j\right\} + \text{var}(\varepsilon U \xi_j)\right] \le \text{const}\cdot(E\xi_j^4)^{1/2} \le \text{const}\cdot\lambda_j,$$

where the constants do not depend on j nor n. The second inequality uses the Cauchy–Schwarz inequality and the last inequality is due to Assumption (A1). This completes the proof of the first equation of (S3.3).

For any $1 \le j \le m_n$ and $k \ne j$, by Assumption (A2), we have $|\lambda_j - \lambda_k| \ge \min\{\lambda_j - \lambda_{j+1}, \lambda_{j-1} - \lambda_j\} \ge C^{-1}j^{-\alpha-1} \ge C^{-1}m_n^{-\alpha-1}$. Observe that,

$$\max_{1\le j\le m_n}\left|\frac{|\widehat{\lambda}_j - \lambda_k|}{|\lambda_j - \lambda_k|} - 1\right| \le \max_{1\le j\le m_n}\frac{|\widehat{\lambda}_j - \lambda_j|}{|\lambda_j - \lambda_k|} \le C\widehat{\Delta}m_n^{\alpha+1} = O_p(n^{-1/2}m_n^{\alpha+1}). \tag{S3.4}$$

Then, Assumption (A4) ensures that the left-hand side (S3.4) converges to 0 in probability. Therefore, $|\widehat{\lambda}_j - \lambda_k| = |\lambda_j - \lambda_k|\{1 + o_p(1)\}$, where $o_p(1)$ is uniform in $1 \le j \le m_n$. Combining the result with similar techniques as Hall and Horowitz (2007, pp85–86), we complete the proof of the second equation of (S3.3).

Using equation (20) in Imaizumi and Kato (2018) and the Cauchy–Schwarz inequality, the proof of the third equation of (S3.3) is completed.

□

Before presenting the proof of Theorem 1, we give two technical lemmas.

Lemma 2. *Under Assumptions (A1)–(A4),*

$$\sum_{j=1}^{m_n} \widehat{\lambda}_j^{-1}\widehat{g}_j^2 \xrightarrow{p} \theta^2 E\left\{\int (X-\mu_X)b\right\}^2 \quad \textit{and} \quad \sum_{j=1}^{m_n} \widehat{\lambda}_j^{-1}\widehat{g}_j\widehat{h}_j \xrightarrow{p} \theta E\left\{\int (X-\mu_X)b\right\}^2.$$

Proof. We only prove the first part. The proof of the second part is similar and thus omitted. Note that $\theta^2 E\{\int (X-\mu_X)b\}^2 = \sum_{j=1}^{\infty}\lambda_j^{-1}g_j^2$. Using the triangle inequality, we have

$$\left|\sum_{j=1}^{m_n}\widehat{\lambda}_j^{-1}\widehat{g}_j^2 - \sum_{j=1}^{\infty}\lambda_j^{-1}g_j^2\right| \le \left|\sum_{j=1}^{m_n}\widehat{\lambda}_j^{-1}\widehat{g}_j^2 - \sum_{j=1}^{m_n}\lambda_j^{-1}g_j^2\right| + \left|\sum_{j=m_n+1}^{\infty}\lambda_j^{-1}g_j^2\right| =: B_1 + B_2.$$

B_2 converges to 0 as $n\to\infty$ because $\sum_{j=1}^{\infty}\lambda_j^{-1}g_j^2<\infty$.

$$\begin{aligned} B_1 &\le 3\sum_{k=1}^{3}\sum_{j=1}^{m_n}\widehat{\lambda}_j^{-1}S_{jk}^2 + 6\sum_{k=1}^{3}\sum_{j=1}^{m_n}\widehat{\lambda}_j^{-1}|g_j||S_{jk}| + 3\sum_{j=1}^{m_n}|\widehat{\lambda}_j^{-1}-\lambda_j^{-1}|g_j^2 \\ &\le 6\sum_{k=1}^{3}\sum_{j=1}^{m_n}\lambda_j^{-1}S_{jk}^2 + 12\theta\sum_{k=1}^{3}\sum_{j=1}^{m_n}|b_j||S_{jk}| + 6\theta^2\sum_{j=1}^{m_n}|\widehat{\lambda}_j-\lambda_j|b_j^2, \end{aligned} \tag{S3.5}$$

where the first inequality is obtained by using the first part of (S3.2) and some fundamental inequalities. The second inequality is obtained by using the first part of (S3.1) and is valid on $\mathcal{E}_{m_n}$.

Using the results given in (S3.3) for S_{jk}, we have

$$\left.\begin{aligned}
&\sum_{j=1}^{m_n} \lambda_j^{-1} S_{j1}^2 = O_p(n^{-1}m_n), \quad \sum_{j=1}^{m_n} \lambda_j^{-1} S_{j3}^2 = O_p(n^{-2}m_n^{\alpha+3}),\\
&\sum_{j=1}^{m_n} |b_j||S_{j1}| = O_p(n^{-1/2}), \quad \sum_{j=1}^{m_n} |b_j||S_{j3}| = O_p(n^{-1}m_n),\\
&\sum_{j=1}^{m_n} \lambda_j^{-1} S_{j2}^2 = O_p(n^{-1}m_n) + O_p(n^{-2}m_n^{\alpha+1}\log n + n^{-2}m_n^{3\alpha-2\beta+5})\\
&\qquad\qquad + O_p(n^{-2}m_n^{\alpha+3}) + O_p(n^{-1}),\\
&\sum_{j=1}^{m_n} |b_j||S_{j2}| = O_p(n^{-1/2}) + O_p(n^{-1}\sqrt{\log n} + n^{-1}m_n) + O_p(n^{-1}m_n).
\end{aligned}\right\} \tag{S3.6}$$

Recall $m_n = o(n^{1/(2\alpha+2)})$ and Assumptions (A3)–(A4). It can be shown that all of summations of (S3.6) equal $o_p(1)$. Using the first part of (S3.1) and $\widehat{\Delta} = O_p(n^{-1/2})$, we have

$$\sum_{j=1}^{m_n} |\widehat{\lambda}_j - \lambda_j| b_j^2 \leq \widehat{\Delta} \sum_{j=1}^{m_n} b_j^2 = O_p(n^{-1/2}).$$

Combining the results given in (S3.6) with (S3.5) yields $B_1 = o_p(1)$. We complete the proof of Lemma 2. □

Lemma 3. *Under Assumptions (A1)-(A4),*

$$\sum_{j=1}^{m_n} \widehat{\lambda}_j^{-1} \widehat{g}_j(\widehat{g}_j - \theta\widehat{h}_j) = \theta \int (\widehat{g} - \theta\widehat{h}) b + o_p(n^{-1/2}).$$

Proof. Using (S3.2), we know that

$$\widehat{g}_j - \theta\widehat{h}_j = S_{j3} - \theta T_{j3} + \int (\widehat{g} - \theta\widehat{h})\phi_j.$$

The following inequality is due to (S3.2), and holds on $\mathcal{E}_{m_n}$

$$
\begin{aligned}
&\left|\sqrt{n}\sum_{j=1}^{m_n}\widehat{\lambda}_j^{-1}\widehat{g}_j(\widehat{g}_j-\theta\widehat{h}_j)-\sqrt{n}\theta\int(\widehat{g}-\theta\widehat{h})b\right| \\
&\leq \left|\sqrt{n}\sum_{j=1}^{m_n}\widehat{\lambda}_j^{-1}\widehat{g}_j(\widehat{g}_j-\theta\widehat{h}_j)-\sqrt{n}\sum_{j=1}^{m_n}\lambda_j^{-1}g_j\int(\widehat{g}-\theta\widehat{h})\phi_j\right|+\left|\sqrt{n}\sum_{j=m_n+1}^{\infty}\lambda_j^{-1}g_j\int(\widehat{g}-\theta\widehat{h})\phi_j\right| \\
&\leq \sqrt{n}\sum_{j=1}^{m_n}\lambda_j^{-1}(7S_{j1}^2+3S_{j2}^2+7S_{j3}^2+4\theta^2T_{j1}^2+4\theta^2T_{j3}^2)+2\theta\sqrt{n}\sum_{j=1}^{m_n}|b_j|(|S_{j3}|+\theta|T_{j3}|) \\
&\quad +2\theta\sqrt{n}\sum_{j=1}^{m_n}\lambda_j^{-1}|\widehat{\lambda}_j-\lambda_j||b_j|(|S_{j1}|+\theta|T_{j1}|)+\theta\sqrt{n}\sum_{j=m_n+1}^{\infty}|b_j|(|S_{j1}|+\theta|T_{j1}|) \\
&=: D_1+D_2+D_3+D_4.
\end{aligned}
$$

Using the results in (S3.3), we have

$$
\begin{aligned}
D_1 &= O_p(n^{-1/2}m_n)+O_p(n^{-3/2}m_n^{\alpha+1}\log n+n^{-3/2}m_n^{3\alpha-2\beta+5}) \\
&\quad +O_p(n^{-3/2}m_n^{\alpha+3})+O_p(n^{-1/2}),
\end{aligned}
$$

and $D_2=O_p(n^{-1/2}m_n)$. Recall $m_n=o(n^{1/(2\alpha+2)})$ and Assumptions (A3)–(A4). We have $D_1=o_p(1)$ and $D_2=o_p(1)$. Using the first part of (S3.1), and the fact $\widehat{\Delta}=O_p(n^{-1/2})$, we have

$$
D_3=O_p\left(n^{-1/2}\sum_{j=1}^{m_n}j^{\alpha/2-\beta}\right)=O_p(n^{-1/2}).
$$

Using Assumptions (A2) and (A3), we have

$$
D_4\leq C\theta\sum_{j=m_n+1}^{\infty}j^{-\beta}\lambda_j^{1/2}=O_p\left(\sum_{j=m_n+1}^{\infty}j^{-\beta-\alpha/2}\right)=o_p(1).
$$

This completes the proof of Lemma 3. □

Proof of Theorem 1. A direct calculation shows that

$$\widehat{\mu}_1 - \theta\widehat{\mu}_0 = \frac{1}{n}\sum_{i=1}^{n} W_i, \quad \widehat{g} - \theta\widehat{h} = \frac{1}{n}\sum_{i=1}^{n} W_i(X_i - \bar{X}), \tag{S3.7}$$

where

$$W_i = \theta\left(a + \int X_i b\right)\left(\frac{U_i}{\widehat{\pi}} - \frac{1-U_i}{1-\widehat{\pi}}\right) + \varepsilon_i\left(\frac{U_i}{\widehat{\pi}} - \theta\frac{1-U_i}{1-\widehat{\pi}}\right).$$

By the law of large numbers, $\widehat{\mu}_0 \to \mu_0$ and $\widehat{\mu}_1 \to \mu_1$ in probability. Using Lemma 2 and the fact $\mu_1 = \theta\mu_0$, it is ready to see that $\widehat{\theta}$ is a consistent estimator of θ. This completes the proof of the first part of Theorem 1.

We now establish the asymptotic normality of $\widehat{\theta}$. Note that

$$\sqrt{n}(\widehat{\theta} - \theta) = \frac{\sqrt{n}\{\sum_{j=1}^{m_n} \widehat{\lambda}_j^{-1}\widehat{g}_j(\widehat{g}_j - \theta\widehat{h}_j) + \widehat{\mu}_1(\widehat{\mu}_1 - \theta\widehat{\mu}_0)\}}{\sum_{j=1}^{m_n} \widehat{\lambda}_j^{-1}\widehat{g}_j\widehat{h}_j + \widehat{\mu}_0\widehat{\mu}_1}. \tag{S3.8}$$

The numerator can be further decomposed as

$$\begin{aligned}
&\sqrt{n}\left\{\sum_{j=1}^{m_n} \widehat{\lambda}_j^{-1}\widehat{g}_j(\widehat{g}_j - \theta\widehat{h}_j) + \widehat{\mu}_1(\widehat{\mu}_1 - \theta\widehat{\mu}_0)\right\} \\
&\quad = \sqrt{n}\left\{\theta\int(\widehat{g} - \theta\widehat{h})b + \widehat{\mu}_1(\widehat{\mu}_1 - \theta\widehat{\mu}_0)\right\} + o_p(1) \\
&\quad = n^{-1/2}\sum_{i=1}^{n}\left\{\theta\int(X_i - \bar{X})b + \widehat{\mu}_1\right\}W_i + o_p(1) \\
&\quad = n^{-1/2}\sum_{i=1}^{n}\theta W_i(a + \int X_i b) + o_p(1) \\
&\quad = n^{-1/2}u_2\theta\sum_{i=1}^{n}\psi(\theta; Y_i, X_i, U_i) + o_p(1),
\end{aligned} \tag{S3.9}$$

where the first two equalities are due to Lemma 3 and (S3.7), respectively. The third equality is proceeded by the two expressions: $\widehat{\mu}_1 = \theta\mu_0 + o_p(1)$

and $n^{-1/2}\sum_{i=1}^n W_i\int(X_i-\bar{X})b = n^{-1/2}\sum_{i=1}^n W_i\int(X_i-\mu_X)b+o_p(1)$. The first one is easy to see and the second one is derived as follows.

$$
\begin{aligned}
n^{-1/2}\sum_{i=1}^n & W_i\int(X_i-\bar{X})b - n^{-1/2}\sum_{i=1}^n W_i\int(X_i-\mu_X)b \\
&= n^{-1/2}\sum_{i=1}^n W_i\left(n^{-1}\sum_{k=1}^n\int X_k b - E\int Xb\right) \\
&= O_p(1)o_p(1) = o_p(1).
\end{aligned}
$$

The last equality is easy to be proved. By the law of large numbers and Lemma 2, the denominator of (S3.8) converges in probability to

$$
\theta E\left\{\int(X-\mu_X)b\right\}^2+\mu_0\mu_1 = \theta E(a+\int Xb)^2 = u_2\theta. \qquad \text{(S3.10)}
$$

Combining (S3.8)–(S3.10), and using the central limit theorem and Slutsky's theorem, we complete the proof of the second part of Theorem 1. □

S4 Proof of Theorem 2

Proof of (2.11). Recall that $\widehat{b}_j = \widehat{\lambda}_j^{-1}\{(1-\pi)\widehat{h}_j + \pi\widehat{\theta}\widehat{g}_j\}/(1-\pi+\pi\widehat{\theta}^2)$ and $b_j = \lambda_j^{-1}h_j$. We have $\widehat{b}_j - b_j = R_1(\widehat{\theta})\widehat{\lambda}_j^{-1}(\widehat{h}_j - h_j) + R_2(\widehat{\theta})\widehat{\lambda}_j^{-1}(\widehat{g}_j - g_j) - R_2(\widehat{\theta})(\widehat{\theta}-\theta)\widehat{\lambda}_j^{-1}h_j + (\widehat{\lambda}_j^{-1} - \lambda_j^{-1})h_j$, where

$$R_1(\widehat{\theta}) = \frac{1-\pi}{1-\pi+\pi\widehat{\theta}^2}, \quad R_2(\widehat{\theta}) = \frac{\pi\widehat{\theta}}{1-\pi+\pi\widehat{\theta}^2}. \tag{S4.1}$$

Thus, we have

$$\begin{aligned}
\sum_{j=1}^{m_n}(\widehat{b}_j - b_j)^2 &\le 4R_1^2(\widehat{\theta})\sum_{j=1}^{m_n}\widehat{\lambda}_j^{-2}(\widehat{h}_j - h_j)^2 + 4R_2^2(\widehat{\theta})\sum_{j=1}^{m_n}\widehat{\lambda}_j^{-2}(\widehat{g}_j - g_j)^2 \\
&\quad + 4R_2^2(\widehat{\theta})(\widehat{\theta}-\theta)^2\sum_{j=1}^{m_n}\widehat{\lambda}_j^{-2}h_j^2 + 4\sum_{j=1}^{m_n}(\widehat{\lambda}_j^{-1} - \lambda_j^{-1})^2h_j^2 \\
&\le 12R_1^2(\widehat{\theta})\sum_{k=1}^{3}\sum_{j=1}^{m_n}\widehat{\lambda}_j^{-2}T_{jk}^2 + 12R_2^2(\widehat{\theta})\sum_{k=1}^{3}\sum_{j=1}^{m_n}\widehat{\lambda}_j^{-2}S_{jk}^2 \\
&\quad + 4R_2^2(\widehat{\theta})(\widehat{\theta}-\theta)^2\sum_{j=1}^{m_n}\widehat{\lambda}_j^{-2}h_j^2 + 4\sum_{j=1}^{m_n}(\widehat{\lambda}_j^{-1} - \lambda_j^{-1})^2h_j^2 \\
&\le 48R_1^2(\widehat{\theta})\sum_{k=1}^{3}\sum_{j=1}^{m_n}\lambda_j^{-2}T_{jk}^2 + 48R_2^2(\widehat{\theta})\sum_{k=1}^{3}\sum_{j=1}^{m_n}\lambda_j^{-2}S_{jk}^2 \\
&\quad + 16R_2^2(\widehat{\theta})(\widehat{\theta}-\theta)^2\sum_{j=1}^{m_n}\lambda_j^{-2}h_j^2 + 2\sum_{j=1}^{m_n}(\widehat{\lambda}_j^{-1} - \lambda_j^{-1})^2h_j^2,
\end{aligned}$$

where the first inequality holds. The second inequality is due to (S3.2), and the third inequality is valid on $\mathcal{E}_{m_n}$.

Recall that $m_n \asymp n^{1/(\alpha+2\beta)}$. It is easy to show that

$$\sum_{j=1}^{m_n} \lambda_j^{-2} T_{j1}^2 = O_p(n^{-(2\beta-1)/(\alpha+2\beta)}),$$
$$\sum_{j=1}^{m_n} \lambda_j^{-2} T_{j2}^2 = o_p(n^{-(2\beta-1)/(\alpha+2\beta)}),$$
$$\sum_{j=1}^{m_n} \lambda_j^{-2} T_{j3}^2 = O_p(n^{-(2\beta-1)/(\alpha+2\beta)}).$$

Similar results hold for S_{jk}. From Theorem 1, $(\widehat{\theta} - \theta)^2 = O_p(n^{-1}) = o_p(n^{-(2\beta-1)/(\alpha+2\beta)})$, and $\sum_{j=1}^{m_n} \lambda_j^{-2} h_j^2 = O(1)$. From the proof of (Hall and Horowitz, 2007, pp82), we have $\sum_{j=1}^{m_n} (\widehat{\lambda}_j^{-1} - \lambda_j^{-1})^2 h_j^2 = o_p(n^{-(2\beta-1)/(\alpha+2\beta)})$. These results indicate that $R_1(\widehat{\theta}) < 1$ and $R_2(\widehat{\theta}) = O_p(1)$, and so

$$\sum_{j=1}^{m_n} (\widehat{b}_j - b_j)^2 = O_p(n^{-(2\beta-1)/(\alpha+2\beta)}).$$

According to Assumption (A3), $|b_j| \le Cj^{-\beta}$ and $\sum_{j=m_n+1}^{\infty} b_j^2 = O(m^{-(2\beta-1)})$ and $\|\widehat{\phi}_j - \phi_j\|_2^2 = O_p(n^{-1}j^2)$ uniformly on $1 \le j \le m_n$. These results with the fact $m_n \asymp n^{1/(\alpha+2\beta)}$ yield

$$\begin{aligned}
\int (\widehat{b} - b)^2 &= \int \left\{ \sum_{j=1}^{m_n} (\widehat{b}_j - b_j)\widehat{\phi}_j + \sum_{j=1}^{m_n} b_j(\widehat{\phi}_j - \phi_j) - \sum_{j=m_n+1}^{\infty} b_j \phi_j \right\}^2 \\
&\le 3 \left\{ \sum_{j=1}^{m_n} (\widehat{b}_j - b_j)^2 + m_n \sum_{j=1}^{m_n} b_j^2 \|\widehat{\phi}_j - \phi_j\|_2^2 + \sum_{j=m_n+1}^{\infty} b_j^2 \right\} \\
&\le 3 \sum_{j=1}^{m_n} (\widehat{b}_j - b_j)^2 + O_p(m_n n^{-1}) + O(m_n^{-(2\beta-1)}) \\
&= O_p(n^{-(2\beta-1)/(\alpha+2\beta)}).
\end{aligned}$$

This completes the proof of first part of Theorem 2. □

Next, we prove (2.12) by using Assouad's Lemma (Assouad, 1983), whose proof can be found on pages 347–348 of Van der Vaart (2000).

Lemma 4. [Assouad] *Let $X \sim \mathbb{P}_\omega$ with $\omega \in \Omega = \{0,1\}^r$, where $\mathbb{P}_\omega$ is a distribution. Let T be an estimator of $\psi(\omega)$ based on X. Then, for all $s > 0$,*

$$\max_\omega 2^s E_\omega d^s(T, \psi(\omega)) \geq \min_{\rho(\omega,\omega')\geq 1} \frac{d^s(\psi(\omega), \psi(\omega'))}{\rho(\omega, \omega')} \frac{r}{2} \min_{\rho(\omega,\omega')=1} \|\mathbb{P}_\omega \wedge \mathbb{P}_{\omega'}\|,$$

where $\rho(\omega, \omega') = \sum_{i=1}^r |\omega_i - \omega_i'|$ is the Hamming distance, and $\|\mathbb{P} \wedge \mathbb{Q}\| = \int p \wedge q \, d\mu$ for two probability measure $\mathbb{P}$ and $\mathbb{Q}$ with densities p and q.

Proof of (2.12). Note that a lower bound for a specific case yields a lower bound for general cases. Thus, it suffices to consider the case when $\varepsilon \sim N(0, \sigma^2)$. Let $\mu_X(t) \equiv 0$ and L_n be the smallest integer greater than $C_1 n^{1/(\alpha+2\beta)}$ for the constant $C_1 > 0$. For a $\omega = (\omega_{L_n+1}, \ldots, \omega_{2L_n}) \in \{0,1\}^{L_n}$, let

$$b_\omega(t) = \sum_{j=L_n+1}^{2L_n} C_1 j^{-\beta} \omega_j \phi_j(t).$$

Denote by P_ω the joint distribution of $\{(Y_i, X_i, U_i) : i = 1, \ldots, n\}$ with $b(\cdot) = b_\omega(\cdot)$. By Lemma 4 with $s = 2$ and $r = L_n$, we have

$$\sup_\omega E \int_{\mathcal{I}} \{\bar{b}(t) - b_\omega(t)\}^2 \, dt \geq C_1^2 \frac{L_n}{8} \frac{\sum_{j=L_n+1}^{2L_n} j^{-2\beta} (\omega_j - \omega_j')^2}{\rho(\omega, \omega')} \min_{\rho(\omega,\omega')=1} \|P_\omega \wedge P_{\omega'}\|$$
$$\geq \frac{C_1^2}{8} 2^{-2\beta} L_n^{-(2\beta-1)} \min_{\rho(\omega,\omega')=1} \|P_\omega \wedge P_{\omega'}\|.$$

For any $\omega, \omega' \in \{0,1\}^{L_n}$,

$$\begin{aligned}\log(P_\omega/P_{\omega'}) =& \frac{1}{\sigma^2}\sum_{i=1}^{n}\left\{Y_i - (1-U_i+U_i\theta)\left(a+\int X_i b_{\omega'}\right)\right\}(1-U_i+U_i\theta)\int X_i(b_\omega - b_{\omega'}) \\ &- \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left\{\int X_i(b_\omega - b_{\omega'})\right\}^2(1-U_i+U_i\theta)^2.\end{aligned}$$

Therefore, the Kullback–Leibler distance between P_ω and $P_{\omega'}$ can be bounded by

$$\begin{aligned}KL(P_\omega, P_{\omega'}) &= \frac{n}{2\sigma^2}C_2\iint_{\mathcal{I}^2}\{b_\omega(t)-b_{\omega'}(t)\}K(s,t)\{b_\omega(s)-b_{\omega'}(s)\}\,dsdt \\ &= \frac{n}{2\sigma^2}C_1^2C_2\sum_{j=L_n+1}^{2L_n} j^{-2\beta}\lambda_j(\omega_j-\omega_j')^2 \\ &\le \frac{n}{2\sigma^2}CC_1^2C_2(L_n+1)^{-(\alpha+2\beta)}\rho(\omega,\omega') \\ &\le C_3\rho(\omega,\omega'),\end{aligned}$$

where constant $C_2 = 1-\pi+\pi\theta^2$ and $C_3 = CC_1^{2-\alpha-2\beta}C_2/(2\sigma^2) > 0$ is independent of n. By LeCam's inequality (Tsybakov, 2009, Lemma 2.3), we know

$$\|P_\omega \wedge P_{\omega'}\| \ge \frac{1}{2}\left(\int\sqrt{d\,P_\omega d\,P_{\omega'}}\right)^2 = \frac{1}{2}\left(1-\frac{H^2(P_\omega,P_{\omega'})}{2}\right)^2,$$

where $H(P_\omega, P_{\omega'})$ is the Hellinger distance between P_ω and $P_{\omega'}$. Since the Hellinger distance $H(P_\omega, P_{\omega'})$ satisfies $H^2(P_\omega,P_{\omega'}) \le KL(P_\omega,P_{\omega'})$,

$$\min_{\rho(\omega,\omega')=1}\|P_\omega \wedge P_{\omega'}\| \ge c_0,$$

where $c_0 = 2^{-1}(1-C_3/2)^2$ depends on C_1 and C_2, which is independent of n. We can choose C_1 such that $c_0 > 0$. We complete the proof of the second part of Theorem 2. □

S5 Semiparametrically Efficient Score

Without loss of generality, assume σ^2 is known and $a = 0$. Then, the log-likelihood for a single sample $\{Y, X(\cdot), U\}$ is

$$\begin{aligned} l(\theta, b, \varphi) = &- \frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left\{Y - (1 - U + U\theta)\int Xb\right\}^2 \\ &+ U\log(\pi) + (1-U)\log(1-\pi) + \log\varphi(X), \end{aligned} \tag{S5.1}$$

where φ is the density function of the functional predictor X. Thus, the ordinary score function for θ is given by

$$\dot{l}_\theta = \frac{\varepsilon}{\sigma^2}U\int Xb,$$

where $\varepsilon \equiv Y - (1 - U + U\theta)\int Xb$.

Consider a parametric and smooth sub-model $\{\varphi_{(t)} : t \in \mathbb{R}\}$ satisfying $\varphi_{(0)} = \varphi$ and

$$\frac{\partial \log \varphi_{(t)}}{\partial t}(X)\big|_{t=0} = \eta(X),$$

where η is a functional satisfied $E\eta(X) = 0$ and $E\{\eta(X)\}^2 < \infty$. Let $r_{(t)}(X) = r(X) + t\int Xf$, for $t \in \mathbb{R}$ and $f \in L_2(\mathcal{I})$. Therefore, $r_{(0)}(X) = r(X) = \int Xb$ and

$$\frac{\partial r_{(t)}}{\partial t}(X)\big|_{t=0} = \int Xf.$$

Define $\mathcal{P}_1 = \{l(\theta, b, \varphi)$ given in (S5.1) for a given $\theta\}$. Since

$$\frac{\partial l}{\partial t}(\theta, b + tf, \varphi_{(t)})\big|_{t=0} = \frac{\varepsilon}{\sigma^2}(1 - U + U\theta)\int Xf + \eta(X),$$

then the tangent space of $\mathcal{P}_1$ is

$$\dot{\mathcal{P}}_1 = \left\{\varepsilon(1 - U + U\theta)\int Xf + \eta(X): \text{ for all } f \in L_2(\mathcal{I}) \text{ and } E\eta(X) = 0, E\{\eta(X)\}^2 < \infty\right\}.$$

By Theorem 3.4.1 of Bickel et al. (1998), the efficient score function of θ is the projection of $\dot{l}_\theta$ into the orthogonal complement of the linear space $\dot{\mathcal{P}}_1$; that is, $\dot{l}^*_\theta = \dot{l}_\theta - \Pi(\dot{l}_\theta | \dot{\mathcal{P}}_1)$. A direct calculation can verify (3.13).

S6 Proof of Theorem 3

Lemma 5. *Under the assumptions of Theorem 3, we have*

$$\frac{1}{n}\sum_{i=1}^n \{\widehat{r}(X_i) - r(X_i)\}^2 = O_p(n^{-(2\beta-1)/(\alpha+2\beta)}) \quad \textit{and} \quad \widehat{u}_2 - u_2 = O_p(n^{-1/2}).$$

Proof. Note that

$$\begin{aligned}\widehat{a} - a = {} & R_1(\widehat{\theta})(\widehat{\mu}_0 - \mu_0) + R_2(\widehat{\theta})\{(\widehat{\mu}_1 - \mu_1) - (\widehat{\theta} - \theta)\mu_0\} \\ & - \int \bar{X}(\widehat{b} - b) - \frac{1}{n}\sum_{i=1}^n \left(\int X_i b - \int \mu_X b\right),\end{aligned}$$

where $R_1(\widehat{\theta})$ and $R_2(\widehat{\theta})$ are defined in (S4.1). By the central limit theorem, Theorems 1 and 2, and Cauchy–Schwarz inequality, we have $(\widehat{a} - a)^2 = O_p(n^{-(2\beta-1)/(\alpha+2\beta)})$. Then, we have

$$\frac{1}{n}\sum_{i=1}^n \{\widehat{r}(X_i) - r(X_i)\}^2 \le 2(\widehat{a} - a)^2 + \frac{2}{n}\sum_{i=1}^n \|X_i\|_2 \|\widehat{b} - b\|_2 = O_p(n^{-(2\beta-1)/(\alpha+2\beta)}).$$

Similarly, $n^{-1}\sum_{i=1}^{n}\{\widehat{r}(X_i)-r(X_i)\}r(X_i)=O_p(n^{-(2\beta-1)/(\alpha+2\beta)})$. Since $\beta > \alpha/2+1$,

$$\begin{aligned}\widehat{u}_2-u_2 &= \frac{1}{n}\sum_{i=1}^{n}\{\widehat{r}(X_i)-r(X_i)\}^2+\frac{1}{n}\sum_{i=1}^{n}r^2(X_i)-E\{r^2(X_i)\}\\ &\quad+\frac{2}{n}\sum_{i=1}^{n}\{\widehat{r}(X_i)-r(X_i)\}r(X_i)\\ &= O_p(n^{-1/2}).\end{aligned}$$

This completes the proof of Lemma 5. □

Proof of Theorem 4. It suffices to show that

$$\widehat{\theta}^*=\widehat{\theta}+\frac{1}{n}\sum_{i=1}^{n}\left(\frac{U_i}{\pi}-\widehat{\theta}\frac{1-U_i}{1-\pi}\right)\frac{r(X_i)}{u_2}\{Y_i-(1-U_i+U_i\widehat{\theta})r(X_i)\}+o_p(n^{-1/2}). \tag{S6.1}$$

Since if (S6.1) holds, replacing Y_i in (S6.1) with $(1-U_i+U_i\theta)r(X_i)+\varepsilon_i$ and by a simple calculation, we have

$$\widehat{\theta}^*-\theta=(\widehat{\theta}-\theta)Z_n+\frac{1}{n}\sum_{i=1}^{n}\psi^*(\theta;Y_i,X_i,U_i)+o_p(n^{-1/2}),$$

where

$$Z_n=1-\frac{1}{n}\sum_{i=1}^{n}\frac{U_i}{\pi u_2}r^2(X_i)-\frac{1}{n}\sum_{i=1}^{n}\frac{1-U_i}{1-\pi}\frac{r(X_i)}{u_2}\varepsilon_i.$$

Theorem 1 and the law of large numbers imply that $\widehat{\theta}-\theta=O_p(n^{-1/2})$ and $Z_n=o_p(1)$. This completes the proof of Theorem 3.

We now show (S6.1). A direct manipulation yields

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left(\frac{U_i}{\pi}-\widehat{\theta}\frac{1-U_i}{1-\pi}\right)\widehat{r}(X_i)\widehat{\varepsilon}_i \\
&\quad= (\theta-\widehat{\theta})\frac{1}{n\pi}\sum_{i=1}^{n}U_i\widehat{r}(X_i)r(X_i)+\frac{1}{n}\sum_{i=1}^{n}\left(\frac{U_i}{\pi}-\widehat{\theta}\frac{1-U_i}{1-\pi}\right)\widehat{r}(X_i)\varepsilon_i \\
&\qquad+\frac{\widehat{\theta}}{n}\sum_{i=1}^{n}\left(\frac{U_i}{\pi}-\frac{1-U_i}{1-\pi}\right)\widehat{r}(X_i)\{\widehat{r}(X_i)-r(X_i)\} \qquad \text{(S6.2)}\\
&\quad=: I+II+III,
\end{aligned}$$

and

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left(\frac{U_i}{\pi}-\widehat{\theta}\frac{1-U_i}{1-\pi}\right)r(X_i)\{Y_i-(1-U_i+U_i\widehat{\theta})r(X_i)\} \\
&\quad= (\theta-\widehat{\theta})\frac{1}{n\pi}\sum_{i=1}^{n}U_i r^2(X_i)+\frac{1}{n}\sum_{i=1}^{n}\left(\frac{U_i}{\pi}-\widehat{\theta}\frac{1-U_i}{1-\pi}\right)r(X_i)\varepsilon_i \qquad \text{(S6.3)}\\
&\quad=: IV+V.
\end{aligned}$$

Let X^* be a copy of X independent of the training data X_i's, and E^* stand for the expectation taken over X^* only. Using a similar routine to that of Schick (1993), we can establish the following results.

$$\frac{1}{n}\sum_{i=1}^{n}U_i\{\widehat{r}(X_i)-r(X_i)\}r(X_i)=\widehat{\pi}E^*[\{\widehat{r}(X^*)-r(X^*)\}r(X^*)]+o_p(n^{-1/2}),$$

and

$$\frac{1}{n}\sum_{i=1}^{n}(1-U_i)\{\widehat{r}(X_i)-r(X_i)\}r(X_i)=(1-\widehat{\pi})E^*[\{\widehat{r}(X^*)-r(X^*)\}r(X^*)]+o_p(n^{-1/2}).$$

From these results and Lemma 5, we immediately know

$$\begin{aligned} I - IV &= (\theta - \widehat{\theta})\frac{1}{n\pi}\sum_{i=1}^{n} U_i\{\widehat{r}(X_i) - r(X_i)\}r(X_i) \\ &= (\theta - \widehat{\theta})\frac{\widehat{\pi}}{\pi}E^*[\{\widehat{r}(X^*) - r(X^*)\}r(X^*)] + o_p(n^{-1/2}) \\ &= o_p(n^{-1/2}), \end{aligned}$$

$$\begin{aligned} III &= \left|\frac{\widehat{\theta}}{n}\sum_{i=1}^{n}\left(\frac{U_i}{\pi} - \frac{1-U_i}{1-\pi}\right)\{\widehat{r}(X_i) - r(X_i)\}^2 \right. \\ &\quad \left. + \frac{\widehat{\theta}}{n}\sum_{i=1}^{n}\left(\frac{U_i}{\pi} - \frac{1-U_i}{1-\pi}\right)\{\widehat{r}(X_i) - r(X_i)\}r(X_i)\right| \\ &\leq |\widehat{\theta}| \max\left\{\frac{1}{\pi}, \frac{1}{1-\pi}\right\}\frac{1}{n}\sum_{i=1}^{n}\{\widehat{r}(X_i) - r(X_i)\}^2 + o_p(n^{-1/2}) \\ &= O_p(n^{-(2\beta-1)/(\alpha+2\beta)}) + o_p(n^{-1/2}) \\ &= o_p(n^{-1/2}). \end{aligned}$$

Using the Cauchy–Schwarz inequality and the central limit theorem, we have

$$\begin{aligned} &\left|\frac{1}{n}\sum_{i=1}^{n} U_i\{\widehat{r}(X_i) - r(X_i)\}\varepsilon_i\right| \\ &\leq \left|(\widehat{a} - a)\frac{1}{n}\sum_{i=1}^{n} U_i\varepsilon_i\right| + \left|\frac{1}{n}\sum_{i=1}^{n} U_i\varepsilon_i \int_{\mathcal{I}} X_i(t)(\widehat{b}(t) - b(t))\,dt\right| \\ &\leq o_p(n^{-1/2}) + \left\|\frac{1}{n}\sum_{i=1}^{n} U\varepsilon_i X_i\right\|_2 \|\widehat{b} - b\|_2 \\ &= o_p(n^{-1/2}). \end{aligned}$$

Similarly, we can prove that $n^{-1}\sum_{i=1}^{n}(1-U_i)\{\widehat{r}(X_i)-r(X_i)\}\varepsilon_i = o_p(n^{-1/2})$.

Therefore

$$II - V = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{U_i}{\pi} - \widehat{\theta}\frac{1-U_i}{1-\pi}\right)\{\widehat{r}(X_i) - r(X_i)\}\varepsilon_i = o_p(n^{-1/2}).$$

Combining (S6.2) and (S6.3), and recalling the orders of $I - IV$, III and $II - V$, and Lemma 5, we complete the proof of (S6.1). □

S7 The Proof of Theorem 4

Let $\widetilde{\Delta}^2 = \iint_{\mathcal{I}^2}(\widetilde{K} - K)^2$. Then based on Assumptions (B1)–(B5), we have $\widetilde{\Delta} = O_p(\rho_{n3})$ (Zhang and Wang, 2016). Using the same arguments as in the proof of Theorem 4.1 in Zhang and Wang (2016), we have $\|\widetilde{g}-g\|_2 = O_p(\rho_{n1})$ and $\|\widetilde{h}-h\|_2 = O_p(\rho_{n2})$. Then by using the Cauchy–Schwarz inequality and (S3.1),

$$\begin{aligned} |\widetilde{g}_j - g_j| &= |\int(\widetilde{g} - g)\widetilde{\phi}_j + \int g(\widetilde{\phi}_j - \phi_j)| \\ &\le \|\widetilde{g} - g\|_2 + \|g\|_2\|\widetilde{\phi}_j - \phi_j\|_2 \\ &= O_p(\rho_{n1} + \delta_j^{-1}\rho_{n3}) \end{aligned} \tag{S7.1}$$

uniformly in j. Similarly, $\widetilde{h}_j - h_j = O_p(\rho_{n2} + \delta_j^{-1}\rho_{n3})$ uniformly in j. Therefore,

$$\sum_{j=1}^{m_n}\lambda_j^{-1}(\widetilde{g}_j - g_j)^2 = O_p(\rho_{n1}^2\sum_{j=1}^{m_n}\lambda_j^{-1} + \rho_{n3}^2\sum_{j=1}^{m_n}\lambda_j^{-1}\delta_j^{-2}),$$

$$\sum_{j=1}^{m_n}\lambda_j^{-1}g_j(\widetilde{g}_j - g_j) = O_p(\rho_{n1} + \rho_{n3}\sum_{j=1}^{m_n}b_j\delta_j^{-1}).$$

Since $\delta_j \geq C^{-1}j^{-\alpha-1}$ by Assumption (A2), then Assumptions (A3) and (B7) imply that $\sum_{j=1}^{m_n} \lambda_j^{-1}(\widetilde{g}_j - g_j)^2 = o_p(1)$ and $\sum_{j=1}^{m_n} \lambda_j^{-1} g_j(\widetilde{g}_j - g_j) = o_p(1)$. Thus, it can be shown that Lemma 2 with $\widehat{\lambda}_j, \widehat{g}_j$ and $\widehat{h}_j$ replaced by $\widetilde{\lambda}_j, \widetilde{g}_j$ and $\widetilde{h}_j$ holds, respectively. This completes the proof of consistency of $\widetilde{\theta}$.

Next, using Assumption (B7) and (S7.1), we can prove

$$\sum_{j=1}^{m_n} \lambda_j^{-2}(\widetilde{g}_j - g_j)^2 = O_p(m_n^{2\alpha+1}\rho_{n1}^2 + m_n^{4\alpha+3}\rho_{n3}^2) = o_p(1).$$

Similarly, $\sum_{j=1}^{m_n} \lambda_j^{-2}(\widetilde{h}_j - h_j)^2 = o_p(1)$, $\sum_{j=1}^{m_n}(\widetilde{\lambda}_j^{-1} - \lambda_j^{-1})^2 h_j^2 = O_p(\rho_{n3}^2 \sum_{j=1}^{m_n} b_j^2 \lambda_j^{-2}) = o_p(1)$ and $m_n \sum_{j=1}^{m_n} b_j^2 \|\widetilde{\phi}_j - \phi_j\|_2^2 = O_p(\rho_{n3}^2 m_n \sum_{j=1}^{m_n} b_j^2 \delta_j^{-2}) = o_p(1)$. From the proof of Theorem 2, this completes the proof of the second part of Theorem 4.

Bibliography

Assouad, P. (1983). Deux remarques sur l'estimation. *Comptes rendus des séances de l'Académie des sciences. Série 1, Mathématique* **296**, 1021-1024.

Bhatia, R., Davis, C. and McIntosh, A. (1983). Perturbation of spectral subspaces and solution of linear operator equations. *Linear Algebra Appl.* **52**, 45-67.

Bickel, P. J., Klaasen, C. A. J., Ritov, Y. and Wellner, J. A. (1998). *Efficient and Adaptive Estimation for Semiparametric Models.* Springer, New York.

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

Imaizumi, M. and Kato, K. (2018). PCA-based estimation for functional linear regression with functional responses. *J. Multivariate Anal.* **163**, 15-36.

Schick, A. (1993). On efficient estimation in regression models. *Ann. Statist.* **21**, 1486-1521.

Tsybakov, A. B. (2009). *Introduction to Nonparametric Estimation.* Springer, New York.

Van der Vaart, A. W. (2000). *Asymptotic Statistics*, Volume 3. Cambridge University Press.

Zhang, X. and Wang, J.-L. (2016). From sparse to dense functional data and beyond. *Ann. Statist.* **44**, 2281-2321.