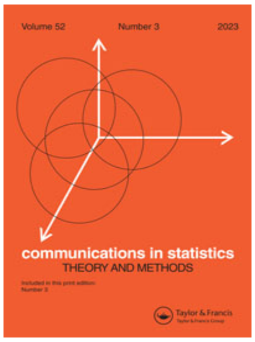

Communications in Statistics - Theory and Methods

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/lsta20

Smoothed average variance estimation for dimension reduction with functional data

Mètolidji Moquilas Raymond Affossogbe, Guy Martial Nkiet & Carlos Ogouyandjou

COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
2023, VOL. 52, NO. 3, 806–829
https://doi.org/10.1080/03610926.2021.1931330

Taylor & Francis
Taylor & Francis Group

Smoothed average variance estimation for dimension reduction with functional data

Mètolidji Moquilas Raymond Affossogbe[a], Guy Martial Nkiet[b], and Carlos Ogouyandjou[a]

[a]Institut de Mathématiques et de Sciences Physiques, Département de Mathématiques, Porto Novo, Bénin; [b]Université des Sciences et Techniques de Masuku, Département de Mathématiques et Informatique, Franceville, Gabon

ABSTRACT
We propose an estimation method, named functional average variance estimation (FAVE), for estimating the EDR space in functional semiparametric regression model, based on kernel estimates of density and regression. Consistency results are then established for the estimator of the interest operator, and for the directions of EDR space. A simulation study that shows that the proposed approach performs better than traditional ones is presented.

ARTICLE HISTORY
Received 21 July 2020
Accepted 11 May 2021

KEYWORDS
FAVE; kernel estimator; functional data; asymptotic study

1. Introduction

In recent years, much attention has been given to functional statistics, which can be described as the set of statistical methods for processing data having the form of curves considered as observations of functions belonging to given functional spaces. Among the references in this field, there are the books by Ramsay and Silverman (1997) for the applied aspects, Bosq (2000) for the theoretical aspects, Horvath and Kokoszka (2012) and Ferraty and Vieu (2016) for recent developments. Many works in this field deal with problems that appear in the general framework of functional regression models which are usually used to find the best link between a real random variable Y and a random curve X whose values belong to $\mathcal{H} = L^2([0,1])$, the set of square integrable functions from $[0,1]$ to $\mathbb{R}$. An abundant literature has examined cases of parametric functional regression models (e.g., Ramsay and Silverman (1997); Cardot, Ferraty, and Sarda (1999); Hall and Horowitz (2007); Yao and Müller (2010)) described by the relation $Y = f_\theta(X, \varepsilon)$, where f_θ belongs to a well-known family of functions parameterized by the unknow parameter θ which is to be estimated, and ε is an error term. In contrast to this, some works deal with a nonparametric model $Y = f(X) + \varepsilon$ where f is an unknown and arbitrary function to estimate, and have introduced nonparametric estimation approaches, such as methods based on kernel estimators (Ferraty and Vieu 2002, 2016). Alternatively, between these two different approaches, a semiparametric regression model

$$Y = f(< \beta_1, X >_{\mathcal{H}}, < \beta_2, X >_{\mathcal{H}}, \ldots, < \beta_K, X >_{\mathcal{H}}, \varepsilon) \quad (1)$$

CONTACT Guy Martial Nkiet gnkiet@hotmail.com Université des Sciences et Techniques de Masuku, Département de Mathématiques et Informatique, Franceville, Gabon.

was considered (Dauxois, Ferré, and Yao 2001; Ferré and Yao 2003, 2005; Lian and Li 2014). In the model (1), $< \cdot, \cdot >_{\mathcal{H}}$ denotes the inner product of $\mathcal{H}$ defined for all g_1 and g_2 belonging to $\mathcal{H}$ by $< g_1, g_2 >_{\mathcal{H}} = \int_0^1 g_1(t)g_2(t)dt$, and $\beta_1, ..., \beta_K$ are elements of $\mathcal{H}$ to be estimated. This model just is an extension in the functional case of the model introduced by Li (1991) in the multivariate context and which has been intensively studied since then. It expresses the fact that the information in X about Y depends only on the projection of X onto the subspace spanned by $\{\beta_1, ..., \beta_K\}$, called effective dimension-reduction (EDR) space. Li (1991) showed that the problem of estimating the EDR space comes down, under a fairly general condition, to the spectral analysis of an operator depending on the covariance operator of the conditional expectation $\mathbb{E}(X|Y)$ of X given Y. Then, he proposed an estimation method, called sliced inverse regression (SIR), based on an estimate of an approximation of this covariance operator obtained by slicing the range of Y. Alternatively, Cook (2000) proposed another method, called sliced average variance estimation (SAVE), for estimating the EDR by using an estimate of an approximation of an operator depending on the conditional covariance operator $Var(X|Y)$ of X given Y. SIR and SAVE are the most popular methods for dimension reduction in the multivariate context, and smoothed estimation methods, based on kernel estimates, have been proposed for them respectively by Zhu and Fang (1996) and Zhu and Zhu (2007). In the functional context, SIR has been extended to functional SIR (FSIR) by Ferré and Yao (2003) who also proposed later a smoothed estimation procedure based on kernel estimates, so defining smoothed functional inverse regression (FIR). On the other hand, more recently, Lian and Li (2014) extended SAVE to functional SAVE (FSAVE) and so, they introduced an estimation method also based on slicing the range Y. A drawback of such an approach is that the estimation procedure could be sensitive to the used slicing approach and, in particular, to the number of slices which therefore becomes a parameter of the procedure. Furthermore, as it is well known, estimation based on smooth procedures such as kernel methods is generally more accurate than one based on slicing. There is therefore an interest in introducing a smoothed estimation of SAVE in the functional context. To the best of our knowledge, such an approach have not been proposed yet. Taking all this into consideration, we introduce in this paper a kernel functional average variance estimation (FAVE) method for estimating the EDR space related to model (1). The rest of the paper is organized as follows. In Section 2, we recall some basic facts about FAVE in the functional context, and we specify the interest operator to estimate. In Section 3, an estimator based on kernel estimates is proposed for this estimating this operator. Section 4 is devoted to an asymptotic study of the introduced estimator. A simulation study that permits to evaluate the performance of our proposal is presented in Section 5. The proofs of theorems are postponed in Section 6.

2. The FAVE method

Let us consider the random variables Y and X involved in the model (1); we assume, without loss of generality, that $\mathbb{E}(X) = 0$, and that $\mathbb{E}(||X||^2_{\mathcal{H}}) < +\infty$. Then, the covariance operator of X is defined by $\Gamma = \mathbb{E}(X \otimes X)$, where for any $x, y \in \mathcal{H}, x \otimes y$ denotes the linear operator from $\mathcal{H}$ to itself such that $(x \otimes y)(h) =< x, h >_{\mathcal{H}} y$ for any $h \in \mathcal{H}$. Throughout the paper, Γ will be assumed to be nonsingular and positive definite.

Letting $\mathcal{B} = (< \beta_1, X >_{\mathcal{H}}, < \beta_2, X >_{\mathcal{H}}, ..., < \beta_K, X >_{\mathcal{H}})$ and denoting by $Var(X|\mathcal{B})$ the conditional covariance operator of X given $\mathcal{B}$, we consider the following assumptions:

$(\mathscr{A}_1)$: for all $b \in \mathcal{H}$, one has $\mathbb{E}(< b, X >_{\mathcal{H}} |\mathcal{B}) = \sum_{k=1}^{K} c_k < \beta_k, X >_{\mathcal{H}}$, where $c_1, ..., c_K$ are real numbers;

$(\mathscr{A}_2)$: $Var(X|\mathcal{B})$ is nonrandom.

Lian and Li (2014) showed that under the assumptions $(\mathscr{A}_1)$ and $(\mathscr{A}_2)$, one has the inclusion

$$R(\Gamma - Var(X|Y)) \subset \Gamma\mathfrak{I}, \tag{2}$$

where $R(A)$ denotes the range of the operator A, $Var(X|Y)$ denotes the conditional covariance operator of X given Y, and $\mathfrak{I}$ is the EDR space, that is the space spanned by $\beta_1, ..., \beta_K$. Therefore, $R(\Gamma_I) \subset \mathfrak{I}$, where

$$\Gamma_I := \Gamma^{-1}\mathbb{E}(\Gamma - 2Var(X|Y) + Var(X|Y)\Gamma^{-1}Var(X|Y)).$$

An important consequence is that Γ_I is degenerate in any direction orthonormal to the β_k's $(k = 1, 2, ..., K)$. Then Γ_I is a finite rank operator whose range is contained into the EDR space. This space can, therefore, be approached by the subspace spanned by the eigenvectors of Γ_I associated with the K largest non-null eigenvalues of Γ_I in the same way as in the multivariate case. In the following we suppose that $\text{rank}(\Gamma_I) = K$. We see, therefore, that the eigenvectors associated with the K largest eigenvalues of Γ_I form a base to EDR space, which makes the EDR space identifiable. So Γ_I is the interest operator of the FAVE method. Since the domain of Γ^{-1} is not the whole $\mathcal{H}$, Γ_I may not be well-defined. Conditions under which this operator is well defined are established in Lian and Li (2014) and recalled below.

Let

$$X = \sum_{j=1}^{+\infty} \xi_j \phi_j,$$

be the well-known Karhunen-Loève expansion of X, where $\mathbb{E}[\xi_j^2] = \alpha_j$ are the eigenvalues and ϕ_j are the eigenfunctions. As usual in the functional data literature (e.g., Lian and Li 2014; Ferré and Yao 2005), we assume that $\alpha_1 > \alpha_2 > \cdots > 0$. We now introduce the assumptions:

$$(\mathscr{A}_3) : \mathbb{E}\left(||X||_{\mathcal{H}}^4\right) < +\infty;$$

$$(\mathscr{A}_4) : \mathbb{E}\left[\left(\sum_{j=1}^{+\infty} \alpha_j^{-2} \sum_{i=1}^{+\infty} Cov^2(\xi_i, \xi_j|Y)\right)^2\right] < +\infty.$$

It is known that if $(\mathscr{A}_3)$ and $(\mathscr{A}_3)$ hold, then Γ_I is well-defined (see Proposition 1 in Lian and Li 2014).

3. A kernel estimator

For performing the FAVE method Γ_I, has to be estimated. Lian and Li (2014) introduced an estimator obtained by slicing the range of Y. In this section, we propose another estimator of this operator based on kernel estimates of density and regression. Since $\Gamma = \mathbb{E}[Var(X|Y)] + Var[\mathbb{E}(X|Y)]$, we have

$$\Gamma_I = \Gamma^{-1}(2\Gamma_e + \Psi - \Gamma) \tag{3}$$

where $\Gamma_e = Var[\mathbb{E}(X|Y)]$ is the covariance operator of the conditional expectation $\mathbb{E}(X|Y)$ and $\Psi = \mathbb{E}(Var(X|Y)\Gamma^{-1}Var(X|Y))$. Ferré and Yao (2003) introduced a kernel estimator of Γ_e and showed its consistency. Here, we will use this estimator, and also a kernel estimator of Ψ together with the empirical counterpart of Γ in order to define an estimator of Γ_I. Letting f be the density of Y and putting

$$m(y) = \mathbb{E}(\mathbf{1}_{\{Y=y\}}\ X), \quad M(y) = \mathbb{E}(\mathbf{1}_{\{Y=y\}}\ X \otimes X),$$
$$r(Y) = \mathbb{E}(X|Y) = \frac{m(Y)}{f(Y)} \quad \text{and} \quad R(Y) = \mathbb{E}(X \otimes X|Y) = \frac{M(Y)}{f(Y)},$$

we have $\Psi = \mathbb{E}(C(Y)\Gamma^{-1}C(Y))$ where $C(Y) = Var(X|Y) = R(Y) - r(Y) \otimes r(Y)$. As it was done in Zhu and Fang (1996), in order to avoid the effect of the small values in the denominator, we consider $f_{e_n} = \max(f, e_n)$ instead of f, where $(e_n)_{n\in\mathbb{N}^*}$ is a sequence of real numbers which tends to zero as $n \to +\infty$. Then, we consider

$$r_{e_n}(Y) = \frac{m(Y)}{f_{e_n}(Y)}, \quad R_{e_n}(Y) = \frac{M(Y)}{f_{e_n}(Y)} \quad \text{and} \quad C_{e_n}(Y) = R_{e_n}(Y) - r_{e_n}(Y) \otimes r_{e_n}(Y)$$

instead of $r(Y)$, $R(Y)$ and $C(Y)$. The definition of Γ_I given in (3) requires to use the inverse of Γ. But since Γ is an Hilbert-Schmidt operator, even though its inverse exists it is not generally bounded. To avoid this difficulty, we consider instead the finite-rank operator $\Gamma_D = \Pi_D\Gamma\Pi_D$, where $D \in \mathbb{N}^*$ and Π_D is the projector onto the subspace S_D spanned by the system $\{\phi_1, \ldots, \phi_D\}$ consisting of the D first elements of an orthonormal basis of $\mathcal{H}$. This basis can, for example, be obtained either from principal component analysis (PCA) of X or by using B-splines basis. This operator has a bounded (pseudo-) inverse defined by $\Gamma_D^{-1} = \Pi_D\Gamma^{-1}\Pi_D$.

Let $\{(X_i, Y_i)\}_{1\le i\le n}$ be an i.i.d. sample of (X, Y); the empirical counterpart of Γ is given by $\hat{\Gamma}_n = n^{-1}\sum_{i=1}^n X_i \otimes X_i$. Considering the estimate $\hat{\Pi}_D$ of Π_D defined as the projector onto an estimate $\hat{S}_D$ of S_D, we estimate Γ_D by $\hat{\Gamma}_D = \hat{\Pi}_D\hat{\Gamma}_n\hat{\Pi}_D$. If we use PCA (resp. B-splines basis) then $\hat{S}_D$ consists of the eigenvectors associated with the D largest eigenvalues of $\hat{\Gamma}_n$ (resp. $\hat{S}_D = S_D$). For a given kernel function $K : \mathbb{R} \to \mathbb{R}_+$ and a given real $h > 0$, we consider the estimates

$$\hat{f}(y) = \frac{1}{nh}\sum_{i=1}^n K\left(\frac{Y_i - y}{h}\right), \quad \hat{m}(y) = \frac{1}{nh}\sum_{i=1}^n K\left(\frac{Y_i - y}{h}\right)\ X_i$$

and

$$\hat{M}(y) = \frac{1}{nh}\sum_{i=1}^n K\left(\frac{Y_i - y}{h}\right)\ X_i \otimes X_i$$

of f, m and M respectively. Then, putting

$$\hat{f}_{e_n}(y) = \max\{e_n, \hat{f}(y)\}, \quad \hat{r}_{e_n}(y) = \frac{\hat{m}(Y)}{\hat{f}_{e_n}(y)}, \quad \hat{R}_{e_n}(y) = \frac{\hat{M}(y)}{\hat{f}_{e_n}(y)}$$

and

$$\hat{C}_{e_n}(y) = \hat{R}_{e_n}(y) - \hat{r}_{e_n}(y) \otimes \hat{r}_{e_n}(y)$$

we consider

$$\hat{\Gamma}_{e,n} = \frac{1}{n}\sum_{i=1}^{n} \hat{r}_{e_n}(Y_i) \otimes \hat{r}_{e_n}(Y_i), \qquad \hat{\Psi}_{e_n,D} = \frac{1}{n}\sum_{i=1}^{n} \hat{C}_{e_n}(Y_i)\hat{\Gamma}_D^{-1}\hat{C}_{e_n}(Y_i)$$

and we estimate Γ_I by the random operator

$$\hat{\Gamma}_{I,n} = \hat{\Gamma}_D^{-1}(2\hat{\Gamma}_{e,n} + \hat{\Psi}_{e_n,D} - \Gamma_n).$$

This random operator determines our kernel FAVE approach for estimating the EDR space. This estimation procedure is achieved by considering the space spanned by the eigenvectors $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_K$ of $\hat{\Gamma}_{I,n}$, associated respectively with the K largest eigenvalues $\hat{\lambda}_1, \dots, \hat{\lambda}_K$.

4. Asymptotic study

In this section, we deal with asymptotics for $\hat{\Gamma}_{I,n}$. More precisely, we first establish its consistency as an estimator of Γ_I. Then we show the $\hat{\beta}_k$'s are also consistent estimators of the β_k's. We need the following assumptions:

$(\mathscr{A}_5)$: Γ is positive definite.
$(\mathscr{A}_6)$: f, r and R belong to C^k;
$(\mathscr{A}_7)$: the kernel K is of order $k > 2$, has compact support $[a, b]$, is symmetric about zero and staisfies $K \leq 1$, $\int_a^b |u|^k K(u)du < +\infty$,;

$(\mathscr{A}_8)$: there exist real numbers d_1, d_2 and d_3 such that $\sup_{y\in\mathbb{R}} |f^{(k)}(y)| \leq d_1$, $\sup_{y\in\mathbb{R}} ||m^{(k)}(y)||_{\mathcal{H}} \leq d_2$ and $\sup_{y\in\mathbb{R}} ||m^{(k)}(y)||_{hs} \leq d_3$, where $|| \cdot ||_{hs}$ denotes the Hilbert-Scmidt norm of operators;

$(\mathscr{A}_9)$: $h \sim n^{-c_1}$ and $e_n \sim n^{-c_2}$, where c_1 and c_2 are real numbers satisfying $c_1 > 0$, $0 < c_2 < \frac{k-2}{4(k+1)}$ and $\frac{c_2}{k} + \frac{1}{2k} < c_1 < \frac{1}{4} - c_2$;

$(\mathscr{A}_{10})$: $\sqrt{n}\ \mathbb{E}\left[||R(Y)||_{hs}^2 \mathbf{1}_{\{f(Y)<e_n\}}\right], \sqrt{n}\ \mathbb{E}\left[||R(Y)||_{hs}\ ||r(Y)||_H^2 \mathbf{1}_{\{f(Y)<e_n\}}\right]$ and $\sqrt{n}\ \mathbb{E}\left[||R(Y)||_H^4 \mathbf{1}_{\{f(Y)<e_n\}}\right]$ tends to 0 as $n \to +\infty$;

$(\mathscr{A}_{11})$: the function $y \mapsto \mathbb{E}\left[||X||_{\mathcal{H}}^2 | Y = y\right]$ is continuous.

Remark 1. Zhu and Fang (1996) introduced $\hat{f}_{e_n}(y) = \max(\hat{f}(y), e_n)$ to overcome technical diffiulties due to small values in the denominator of $\hat{r}(y)$. But this approach does not guarantee that we get a good estimator of f. Indeed, if we take for example $e_n = n^{-1/11}$, then until $n = 2000$ we still have $e_n > 1/2$ and, therefore, $\hat{f}_{e_n}(y) = 1/2$ for all $y \in \mathbb{R}$. To overcome this later problem, Nkou and Nkiet (2019) proposed to take $e_n = \min(a; n^{-c_2})$, where a is a fixed strictly positive number. When a is sufficiently small $\hat{f}_{e_n}(y)$ is a good estimator of f, because $\sup_{x\in\mathbb{R}} |\hat{f}_{e_n}(y) - \hat{f}(y)| \leq a$ and we still have $e_n \sim n^{-c_2}$.

For $D \in \mathbb{N}^*$, we consider

$$\Psi_D = \mathbb{E}\left[Var(X|Y)\Gamma_D^{-1} Var(X|Y)\right],$$

and denoting by t_D the minimum positive eigenvalue of Γ_D, we have:

Theorem 1. *Under assumptions* $(\mathscr{A}_1)$ *to* $(\mathscr{A}_3)$ *and* $(\mathscr{A}_7)$ *to* $(\mathscr{A}_{11})$, *if we suppose that when* $D \to +\infty$, *we have* $||\hat{\Gamma}_D - \Gamma_D||_\infty = o_p(t_D)$, *then*

$$||\Psi_D - \hat{\Psi}_{e_n,D}||_{hs} = O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{t_D\sqrt{n}}\right) + O_p\left(\frac{1}{e_n t_D}\left(h^k + \frac{\sqrt{\log(n)}}{h\sqrt{n}}\right)\right)$$
$$= O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{n^\gamma t_D}\right),$$

where γ *is a constant satisfying* $0 < \gamma < 1/4$.

Remark 2. This theorem gives an idea on the convergence rate of each component of $\hat{\Gamma}_{I,n}$ as we know the one of $\hat{\Gamma}_{e,n}$ from Ferré and Yao (2005). We cannot reach the $\sqrt{n}$-convergence, because the rate of convergence will be penalized by the one of t_D.The assumption $||\hat{\Gamma}_D - \Gamma_D||_\infty = o_p(t_D)$ was also used in Lian and Li (2014) for obtaining a similar result for the case of Functional SAVE. A justification of this asumption can be found in this paper.

In the following theorem consistency of $\hat{\Gamma}_{I,n}$ is established under some conditions.

Theorem 2. *Under the assumptions* $(\mathscr{A}_1)$ *to* $(\mathscr{A}_{11})$, *if we suppose that for some* $0 < \gamma < 1/4$, *when* $D \to +\infty, ||\hat{\Gamma}_D - \Gamma_D||_\infty = o_p(t_D), 1/(t_D\sqrt{n}) \to 0,\ 1/(n^\gamma t_D^2) \to 0$, *then* $\hat{\Gamma}_{I,n} - \Gamma_I = o_p(1)$.

Remark 3. This result only gives the convergence in probability of $\hat{\Gamma}_{I,n}$ to Γ_I without specifying the rate. For the functional SAVE, Lian and Li (2014) did not show the convergence of their estimator of Γ_I.

Now, we deal with the $\hat{\beta}_k$'s. For doing that, we assume that $\beta_1, \beta_2, ..., \beta_K$ are the K eigenvectors of Γ_I associated with the K eigenvalues $\lambda_1, ..., \lambda_K$ respectiveley, and that $\lambda_1 > \lambda_2 > \cdots > \lambda_K > 0$.

Theorem 3. *Under the assumptions* $(\mathscr{A}_1)$ *to* $(\mathscr{A}_{11})$, *if we suppose that for some* $0 < \gamma < 1/4$, *when* $D \to +\infty, ||\hat{\Gamma}_D - \Gamma_D||_\infty = o_p(t_D), 1/(t_D\sqrt{n}) \to 0,\ 1/(n^\gamma t_D^{5/2}) \to 0$, *then* $||\hat{\beta}_j - \beta_j||_{\mathcal{H}} = o_p(1)$ *for* $j = 1, 2, ..., K$.

Remark 4. This result is similar to that of FSIR obtained by Ferré and Yao (2003). It is an extension to the functional case of a property of the kernel method for sliced average variance estimation developped by Zhu and Zhu (2007) in the multivariate context.

5. Simulations

In order to observe the performance of the introduced FAVE method, and to compare it with existing methods, we made simulations within a framework corresponding to the model:

$$Y = 20 < \beta_1, X >_{\mathcal{H}}^2 + 10 < \beta_2, X >_{\mathcal{H}}^2 + \varepsilon,$$

where X is a standard brownian motion on $[0, 1]$, observed on a grid of $p = 100$ equally spaced points, $\beta_1(t) = 4t^2, \beta_2(t) = \sin(5\pi t/2)$ and $\varepsilon \sim N(0, 0.1^2)$.

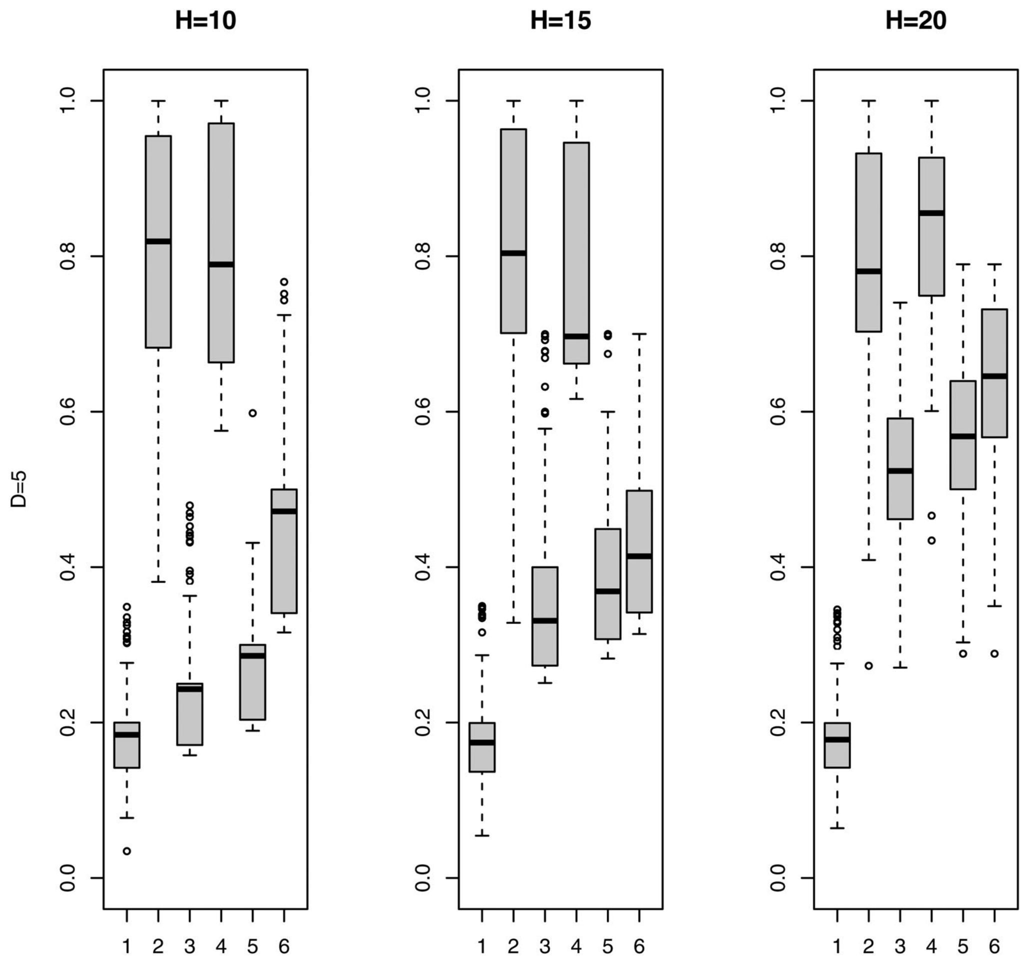

Figure 1. Boxplots showing $||P - \hat{P}||_{hs}$ from various methods. 1: FAVE; 2: FIR; 3: FSAVE with $H = 10$, 15, 20; 4: Hybrid ($\alpha = 0.2$); 5: Hybrid ($\alpha = 0.8$); 6: Hybrid ($\alpha = 0.5$). Dimension $D = 5$.

We generated $m = 100$ independent samples of size $n = 100$ from the above model; for each of these samples we computed estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ of β_1 and β_2, and then the distance between the true EDR space and its estimation is computed via $\mathscr{E} = ||P - \hat{P}||_{hs}$, where P (resp. $\hat{P}$) denotes the projector onto the space spanned by β_1 and β_2 (resp. $\hat{\beta}_1$ and $\hat{\beta}_2$). These computations were made by using the following methods: (1) FAVE; (2) the FIR method of Ferré and Yao (2005); (3) the FSAVE method of Lian and Li (2014); (4) the hybrid method of FSIR and FSAVE (see Wang et al. 2015) with $\alpha = 0.2$; (5) the hybrid method of FSIR and FSAVE with $\alpha = 0.8$; (6) the hybrid method of FSIR and FSAVE with $\alpha = 0.5$.

For computing $\hat{\Pi}_D$, we used quadratic B-spline basis based on knots that are equally spaced on $[0, 1]$; this was done by using the function "create.bspline.basis" from the R package "fda". Various dimensions D were used, $D = 5$, 10, 15. For FAVE and FIR methods, we used the gaussian kernel and the bandwidth h was selected by cross-validation approach. FSAVE was performed with number of slices $H = 10$, 15, 20.

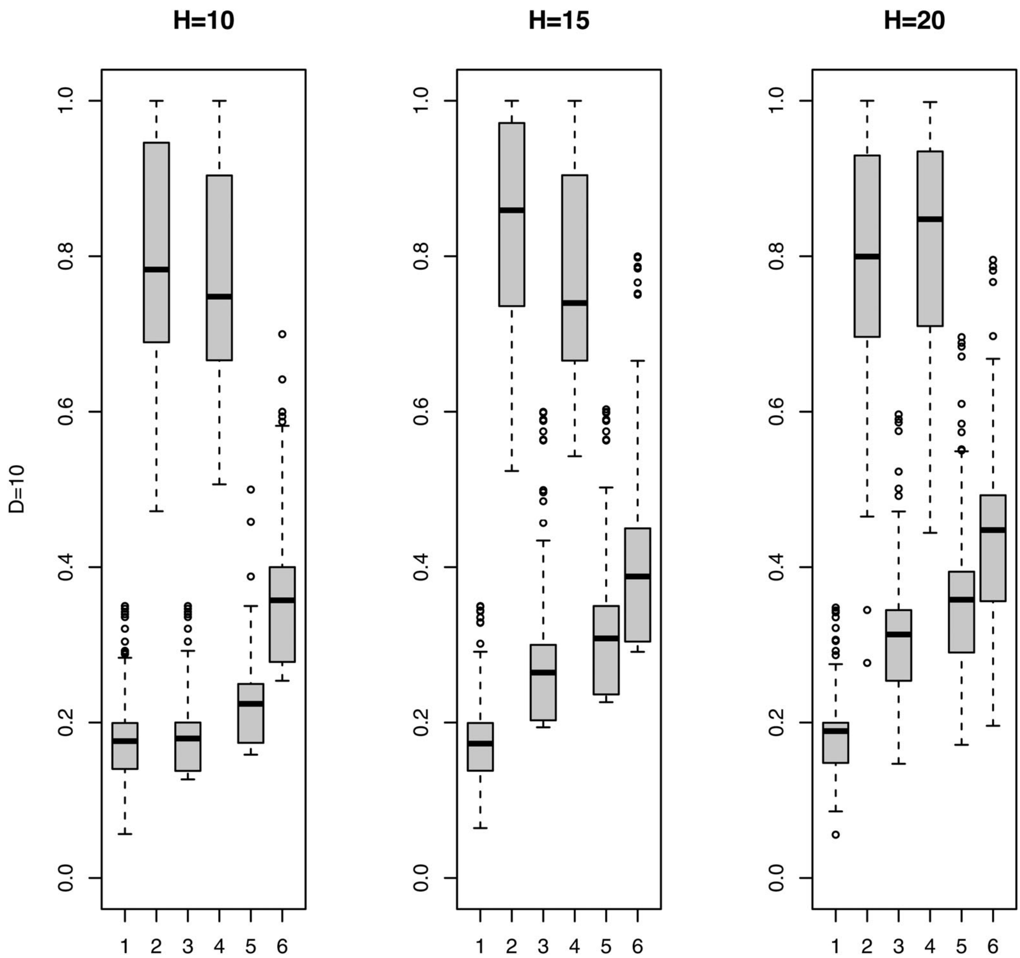

Figure 2. Boxplots showing $||P-\hat{P}||_{hs}$ from various methods. 1: FAVE; 2: FIR; 3: FSAVE with $H=10$, 15, 20; 4: Hybrid ($\alpha=0.2$); 5: Hybrid ($\alpha=0.8$); 6: Hybrid ($\alpha=0.5$). Dimension $D=10$.

Figures 1–3 show the boxplots of $||P-\hat{P}||_{hs}$ for all methods using different values of D and H. It is seen that the FAVE method outperforms all the other methods. In particular, these results show that FAVE brings an improvement to FSAVE and gives much better results than the FIR method. However, the gap between FAVE and FSAVE is smaller than that between FAVE and FIR.

6. Proofs

6.1. Preliminary results

In this section we will give some lemmas necessary to get the proofs of the previous theorems.

Lemma 1. *Under assumption* $(\mathscr{A}_6)$ *to* $(\mathscr{A}_9)$, *we have:*

$$\sup_{y\in\mathbb{R}} ||\hat{M}(y)-M(y)||_{hs} = O_p\left(h^k+\frac{\sqrt{\log(n)}}{h\sqrt{n}}\right).$$

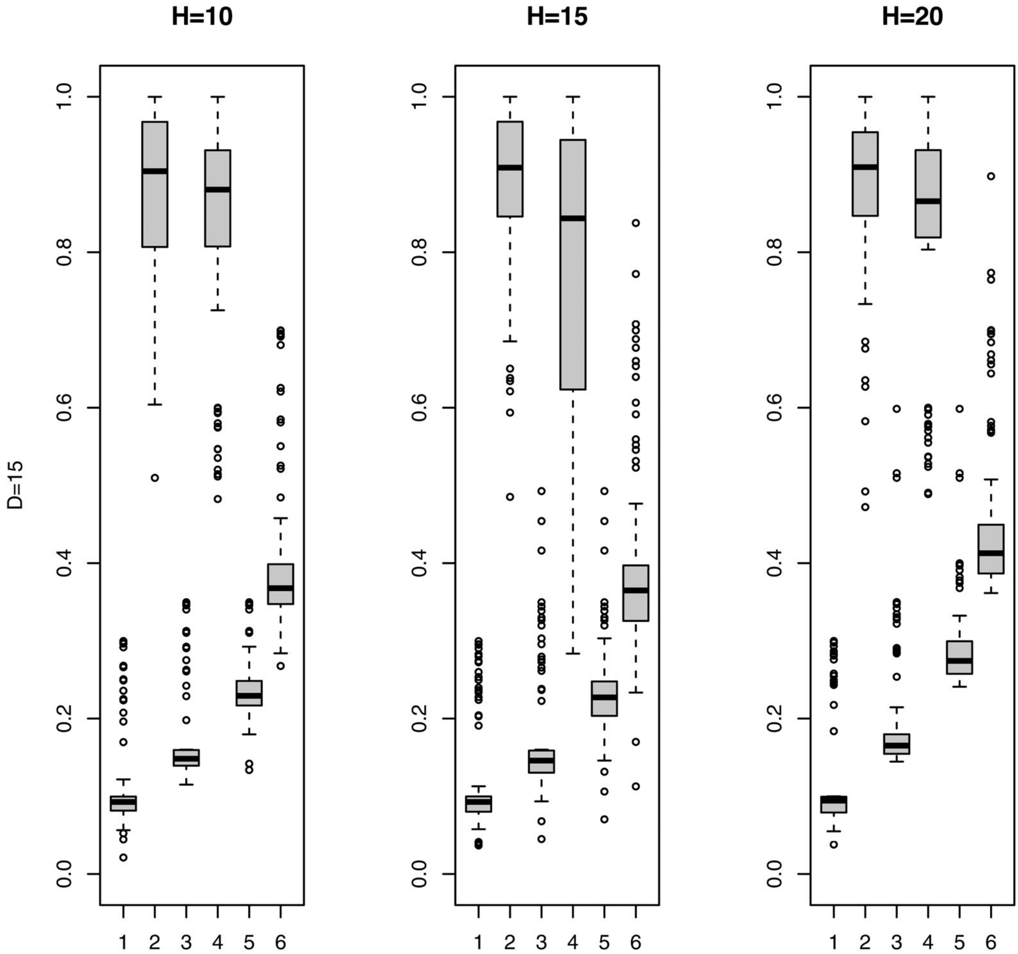

Figure 3. Boxplots showing $||P - \hat{P}||_{hs}$ from various methods. 1: FAVE; 2: FIR; 3: FSAVE with $H = 10$, 15, 20; 4: Hybrid ($\alpha = 0.2$); 5: Hybrid ($\alpha = 0.8$); 6: Hybrid ($\alpha = 0.5$). Dimension $D = 15$.

Proof. It is easy to check that for all $y \in \mathbb{R}$, one has

$$\mathbb{E}\left[\hat{M}(y)\right] = \frac{M * K_h(y)}{h}.$$

Then

$$\begin{aligned}
\mathbb{E}\left[\hat{M}(y)\right] - M(y) &= \frac{1}{h}\int_{\mathbb{R}} R(v)K_h(v-y)f(v)dv - M(y) \\
&= \frac{1}{h}\int_{\mathbb{R}} \left[M(v) - M(y)\right]K_h(v-y)dv \\
&= \int_{\mathbb{R}} \left[M(y+hw) - M(y)\right]K(w)dw \\
&= \int_{\mathbb{R}} \left[\sum_{j=1}^{k-1} \frac{(wh)^j}{j!} M^{(j)}(y) + \frac{(wh)^k}{k!} M^{(k)}(y + \theta hw)\right] K(w)dw \\
&= \frac{h^k}{k!}\int_a^b w^k M^{(k)}(y+\theta hw)K(w)dw.
\end{aligned}$$

Hence

$$||\mathbb{E}\big[\hat{M}(y)\big] - M(y)||_{hs} \leq \frac{h^k}{k!}\sup_{y\in I}||M^{(k)}(y)||_{hs}\int_a^b |w|^k K(w)dw = Ch^k,$$

that is $\sup_{y\in I}||\frac{M*K_h(y)}{h} - M(y)||_{hs} = O(h^k)$. We deduce that

$$\begin{aligned}\sup_{y\in\mathbb{R}}||\hat{M}(y) - M(y)||_{hs} &\leq \sup_{y\in\mathbb{R}}||\hat{M}(y) - \mathbb{E}\big[\hat{M}(y)\big]||_{hs} + \sup_{y\in\mathbb{R}}||\mathbb{E}\big[\hat{M}(y)\big] - M(y)||_{hs}\\ &= D_1 + D2.\end{aligned}$$

Let $\varepsilon > 0$ and $(a_n)_{n\in\mathbb{N}}$, a sequence of non-negative reals numbers converging to $+\infty$. We have:

$$\begin{aligned}P(D_1 > \varepsilon) = P\Big(&\sup_{y\in\mathbb{R}}||\hat{M}(y) - \mathbb{E}\big[\hat{M}(y)\big]||_{hs} > \varepsilon\Big)\\ &\leq P\Big(\sup_{y\in\mathbb{R}}||\hat{M}(y) - \mathbb{E}[\hat{M}(y)]||_{hs} > \varepsilon; ||X\otimes X||_{hs} \leq a_n\Big)\\ &+ P\Big(\sup_{y\in\mathbb{R}}||\hat{M}(y) - \mathbb{E}\big[\hat{M}(y)\big]||_{hs} > \varepsilon; ||X\otimes X||_{hs} > a_n\Big)\\ &\leq P\Big(\sup_{y\in\mathbb{R}}||\hat{M}(y) - \mathbb{E}\big[\hat{M}(y)\big]||_{hs} > \varepsilon; ||X\otimes X||_{hs} \leq a_n\Big)\\ &+P\big(||X\otimes X||_{hs} > a_n\big).\end{aligned}$$

Since $K \leq 1$, we have for all $y \in \mathbb{R}$,

$$\begin{aligned}||\hat{M}(y) - \mathbb{E}\big[\hat{M}(y)\big]||_{hs} &= \left|\left|\frac{1}{nh}\sum_{i=1}^n\left[X_i\otimes X_i K\left(\frac{Y_i - y}{h}\right) - \mathbb{E}\left[X_i\otimes X_i K\left(\frac{Y_i - y}{h}\right)\right]\right]\right|\right|_{hs}\\ &\leq \frac{1}{nh}\sum_{i=1}^n\{||X_i\otimes X_i||_{hs} + \mathbb{E}\big[||X_i\otimes X_i||_{hs}\big]\}.\end{aligned}$$

Thus

$$\sup_{y\in\mathbb{R}}||\hat{M}(y) - \mathbb{E}\big[\hat{M}(y)\big]||_{hs} \leq \frac{1}{nh}\sum_{i=1}^n\{||X_i\otimes X_i||_{hs} + \mathbb{E}\big[||X_i\otimes X_i||_{hs}\big]\}$$

and

$$\begin{aligned}&P\Big(\sup_{y\in\mathbb{R}}||\hat{M}(y) - \mathbb{E}\big[\hat{M}(y)\big]||_{hs} > \varepsilon; ||X\otimes X||_{hs} \leq a_n\Big)\\ &\leq P\left(\frac{1}{nh}\sum_{i=1}^n\big(||X_i\otimes X_i||_{hs} + \mathbb{E}\big[||X_i\otimes X_i||_{hs}\big]\big) > \varepsilon; ||X\otimes X||_{hs} \leq a_n\right)\\ &\leq P\left(\frac{1}{nh}\sum_{i=1}^n\big(||X_i\otimes X_i||_{hs} + \mathbb{E}\big[||X_i\otimes X_i||_{hs}\big]\big)\mathbf{1}_{\{||X\otimes X||_{hs}\leq a_n\}} > \varepsilon\right).\end{aligned}$$

However, for any $i \in \{1, ..., n\}$, one has

$$\frac{1}{h}\left(||X_i \otimes X_i||_{hs} + \mathbb{E}\left[||X_i \otimes X_i||_{hs}\right]\right)\mathbf{1}_{\{||X \otimes X||_{hs} \leq a_n\}} \leq \frac{2a_n}{h}.$$

Then using Bernstein inequality, we get

$$P\left(\sup_{y \in \mathbb{R}} ||\hat{M}(y) - \mathbb{E}\left[\hat{M}(y)\right]||_{hs} > \varepsilon; ||X \otimes X||_{hs} \leq a_n\right) \leq 2\exp\left(-\frac{n\varepsilon^2 h^2}{16a_n^2}\right),$$

from what we deduce

$$\begin{aligned} P(D_2 > \varepsilon) &\leq P\left(||X \otimes X||_{hs} > a_n\right) + 2\exp\left(-\frac{n\varepsilon^2 h^2}{16a_n^2}\right) \\ &\leq \frac{\mathbb{E}\left(||X \otimes X||_{hs}^2\right)}{a_n^2} + 2\exp\left(-\frac{n\varepsilon^2 h^2}{16a_n^2}\right) \\ &= \frac{\mathbb{E}\left(||X||_H^4\right)}{a_n^2} + 2\exp\left(-\frac{n\varepsilon^2 h^2}{16a_n^2}\right). \end{aligned}$$

Taking $\varepsilon = \frac{\varepsilon_0}{h}\sqrt{\frac{\log(n)}{n}}$, where $\varepsilon > 0$, and $a_n = (\log(n))^{1/4}$, we have:

$$P\left(D_2 > \frac{\varepsilon_0}{h}\sqrt{\frac{\log(n)}{n}}\right) \leq \frac{\mathbb{E}\left(||X||_H^4\right)}{(\log(n))^{1/2}} + 2\exp\left(-\frac{\varepsilon_0(\log(n))^{1/2}}{16}\right),$$

and since

$$\lim_{n \to +\infty}\left(\frac{\mathbb{E}\left(||X||_H^4\right)}{(\log(n))^{1/2}} + 2\exp\left(-\frac{\varepsilon_0(\log(n))^{1/2}}{16}\right)\right) = 0,$$

we conclude that $D_2 = O_p(h^{-1}n^{-1/2}(\log(n))^{1/2})$ and, consequently, that

$$\sup_{y \in \mathbb{R}} ||\hat{M}(y) - M(y)||_{hs} = O_p\left(h^k + \frac{\sqrt{\log(n)}}{h\sqrt{n}}\right).$$

□

Lemma 2. We have:

$$\left|\frac{\hat{f}_{e_n}(Y_j) - \hat{f}(Y_j)}{f_{e_n}(Y_j)}\right| \leq 2\left[\mathbf{1}_{\{f(Y_j) < 2e_n\}} + \frac{\left(\sup_{y \in \mathbb{R}} |\hat{f}(y) - f(y)|\right)^2}{e_n^2}\right].$$

Proof. Since

$$\begin{aligned}|\hat{f}_{e_n}(Y_j) - \hat{f}(Y_j)| &= |e_n \mathbf{1}_{\{\hat{f}(Y_j) < e_n\}} + \hat{f}(Y_j)\mathbf{1}_{\{\hat{f}(Y_j) \geq e_n\}} - \hat{f}(Y_j)| \\ &= |e_n \mathbf{1}_{\{\hat{f}(Y_j) < e_n\}} - \hat{f}(Y_j)\mathbf{1}_{\{\hat{f}(Y_j) < e_n\}}| \\ &\leq \left(e_n + \hat{f}(Y_j)\right)\mathbf{1}_{\{\hat{f}(Y_j) < e_n\}},\end{aligned}$$

we obtain

$$\left|\frac{\hat{f}_{e_n}(Y_j) - \hat{f}(Y_j)}{f_{e_n}(Y_j)}\right| \leq \left(\frac{e_n}{f_{e_n}(Y_j)} + \frac{\hat{f}(Y_j)}{f_{e_n}(Y_j)}\right)\mathbf{1}_{\{\hat{f}(Y_j) < e_n\}} \leq 2\ \mathbf{1}_{\{\hat{f}(Y_j) < e_n\}}.$$

It is easy to check that

$$\mathbf{1}_{\{\hat{f}_{e_n}(Y_j) < e_n\}} \leq \mathbf{1}_{\{f(Y_j) < 2e_n\}} + \frac{\left(\sup_{y\in\mathbb{R}} |\hat{f}(y) - f(y)|\right)^2}{e_n^2}.$$

Thus

$$\left|\frac{\hat{f}_{e_n}(Y_j) - \hat{f}(Y_j)}{f_{e_n}(Y_j)}\right| \leq 2\left[\mathbf{1}_{\{f(Y_j) < 2e_n\}} + \frac{\left(\sup_{y\in\mathbb{R}} |\hat{f}(y) - f(y)|\right)^2}{e_n^2}\right].$$

□

Lemma 3. *Under the assumption* $(\mathscr{A}_3)$, *if we suppose* $||\hat{\Gamma}_D - \Gamma_D||_\infty = o_p(t_D)$, *we have:*

$$A_{2n} = \frac{1}{n}\sum_{j=1}^{n} C(Y_j)\Gamma_D^{-1}C(Y_j) - \frac{1}{n}\sum_{j=1}^{n} C(Y_j)\hat{\Gamma}_D^{-1}C(Y_j) = O_p\left(\frac{1}{t_D\sqrt{n}}\right).$$

Proof.

$$\begin{aligned}||A_{2n}||_{hs} &= \left|\left|\frac{1}{n}\sum_{j=1}^{n} C(Y_j)\left[\Gamma_D^{-1} - \hat{\Gamma}_D^{-1}\right]C(Y_j)\right|\right|_{hs} \\ &= \left|\left|\frac{1}{n}\sum_{j=1}^{n} C(Y_j)\left[\hat{\Gamma}_D^{-1}(\hat{\Gamma}_D - \Gamma_D)\Gamma_D^{-1}\right]C(Y_j)\right|\right|_{hs} \\ &\leq ||\frac{1}{n}\sum_{j=1}^{n} C(Y_j)\left[\hat{\Gamma}_D^{-1}(\hat{\Gamma}_D - \Gamma_D)\Gamma^{-1}\right]C(Y_j)||_{hs} \\ &\leq ||\hat{\Gamma}_D^{-1}(\hat{\Gamma}_D - \Gamma_D)||_\infty \frac{1}{n}\sum_{j=1}^{n} ||C(Y_j)||_{hs}\ ||\Gamma^{-1}C(Y_j)||_{hs}.\end{aligned}$$

Since $||\hat{\Gamma}_D - \Gamma_D||_\infty = o_p(t_D)$ and $||\hat{\Gamma}_D^{-1}||_\infty = O_p\left(\frac{1}{t_D}\right)$, we deduce from the preceding inequality that $||A_{2n}||_{hs} = o_p\left(\frac{1}{t_D\sqrt{n}}\right)$. □

Lemma 4. *Under assumptions* $(\mathscr{A}_3)$ *and* $(\mathscr{A}_{10})$, *if we suppose* $||\hat{\Gamma}_D - \Gamma_D||_\infty = o_p(t_D)$, *we have:*

$$A_{3n} = \frac{1}{n}\sum_{j=1}^{n} C(Y_j)\hat{\Gamma}_D^{-1} C(Y_j) - \frac{1}{n}\sum_{j=1}^{n} C_{e_n}(Y_j)\hat{\Gamma}_D^{-1} C_{e_n}(Y_j) = O_p\left(\frac{1}{t_D\sqrt{n}}\right).$$

Proof. We can write

$$\begin{aligned} A_{3n} &= \frac{1}{n}\sum_{j=1}^{n}\left[C(Y_j) - C_{e_n}(Y_j)\right]\hat{\Gamma}_D^{-1} C(Y_j) - \frac{1}{n}\sum_{j=1}^{n} C_{e_n}(Y_j)\hat{\Gamma}_D^{-1}\left[C_{e_n}(Y_j) - C(Y_j)\right] \\ &= A_{31n} - A_{32n}. \end{aligned}$$

First, we deal with A_{31n}. We have

$$\begin{aligned} A_{31n} &= \frac{1}{n}\sum_{j=1}^{n}\left[R(Y_j) - R_{e_n}(Y_j)\right]\hat{\Gamma}_D^{-1} C(Y_j) \\ &\quad + \frac{1}{n}\sum_{j=1}^{n}\left[r_{e_n}(Y_j)\otimes r_{e_n}(Y_j) - r(Y_j)\otimes r(Y_j)\right]\hat{\Gamma}_D^{-1} C(Y_j) \\ &= A_{311n} + A_{312n}. \end{aligned}$$

Further,

$$\begin{aligned} \sqrt{n}||A_{311n}||_{hs} &= \sqrt{n}\left|\left| \frac{1}{n}\sum_{j=1}^{n}\left[R(Y_j) - R_{e_n}(Y_j)\right]\hat{\Gamma}_D^{-1} C(Y_j)\right|\right|_{hs} \\ &\le \frac{||\hat{\Gamma}_D^{-1}||_\infty\sqrt{n}}{n}\sum_{j=1}^{n}||M(Y_j)||_{hs}\ ||C(Y_j)||_{hs}\ \left|\frac{1}{f(Y_j)} - \frac{1}{f_{e_n}(Y_j)}\right|, \end{aligned}$$

and since

$$\left|\frac{1}{f(Y_j)} - \frac{1}{f_{e_n}(Y_j)}\right| \le \frac{1}{f(Y_j)}\mathbf{1}_{\{f(Y_j)<e_n\}},$$

it follows

$$\begin{aligned} \sqrt{n}||A_{311n}||_{hs} &\le \frac{||\hat{\Gamma}_D^{-1}||_\infty\sqrt{n}}{n}\sum_{j=1}^{n}||M(Y_j)||_{hs}\ ||C(Y_j)||_{hs}\frac{1}{f(Y_j)}\mathbf{1}_{\{f(Y_j)<e_n\}} \\ &= \frac{||\hat{\Gamma}_D^{-1}||_\infty\sqrt{n}}{n}\sum_{j=1}^{n}||R(Y_j)||_{hs}\ ||C(Y_j)||_{hs}\mathbf{1}_{\{f(Y_j)<e_n\}} \\ &\le \frac{||\hat{\Gamma}_D^{-1}||_\infty\sqrt{n}}{n}\sum_{j=1}^{n}||R(Y_j)||_{hs}\ ||R(Y_j) - r(Y_j)\otimes r(Y_j)||_{hs}\mathbf{1}_{\{f(Y_j)<e_n\}} \\ &\le \frac{||\hat{\Gamma}_D^{-1}||_\infty\sqrt{n}}{n}\sum_{j=1}^{n}\left[||R(Y_j)||_{hs}^2 + ||R(Y_j)||_{hs}\ ||r(Y_j)||_{hs}^2\right]\mathbf{1}_{\{f(Y_j)<e_n\}}. \end{aligned}$$

Thus

$$\mathbb{E}\left[\frac{\sqrt{n}||A_{311n}||_{hs}}{||\hat{\Gamma}_D^{-1}||_\infty}\right] \le \sqrt{n}\mathbb{E}\left[||R(Y)||_{hs}^2 \mathbf{1}_{\{f(Y)<e_n\}}\right] + \sqrt{n}\mathbb{E}\left[||R(Y)||_{hs}\ ||r(Y)||_{hs}^2 \mathbf{1}_{\{f(Y)<e_n\}}\right],$$

and since $||\hat{\Gamma}_D^{-1} - \Gamma_D^{-1}||_\infty = o_p(t_D), ||\hat{\Gamma}_D^{-1}||_\infty = O_p\left(\frac{1}{t_D}\right)$, we deduce from the preceding inequality, assumption $(\mathscr{A}_{10})$ and Markov inequality that $A_{311n} = o_p\left(\frac{1}{t_D\sqrt{n}}\right)$. On the other hand,

$$\begin{aligned}
\sqrt{n}||A_{312n}||_{hs} &\le \frac{||\hat{\Gamma}_D^{-1}||_\infty\sqrt{n}}{n}\sum_{j=1}^n ||m(Y_j)\otimes m(Y_j)\left[\frac{1}{f^2(Y_j)} - \frac{1}{f_{e_n}^2(Y_j)}\right]||_{hs}\ ||C(Y_j)||_{hs} \\
&\le \frac{||\hat{\Gamma}_D^{-1}||_\infty\sqrt{n}}{n}\sum_{j=1}^n ||m(Y_j)\otimes m(Y_j)||_{hs}\ ||C(Y_j)||_{hs}\left|\frac{1}{f^2(Y_j)} - \frac{1}{f_{e_n}^2(Y_j)}\right| \\
&\le \frac{||\hat{\Gamma}_D^{-1}||_\infty\sqrt{n}}{n}\sum_{j=1}^n ||m(Y_j)\otimes m(Y_j)||_{hs}\ ||C(Y_j)||_{hs}\frac{1}{f^2(Y_j)}\mathbf{1}_{\{f(Y_j)<e_n\}} \\
&\le \frac{||\hat{\Gamma}_D^{-1}||_\infty\sqrt{n}}{n}\sum_{j=1}^n ||r(Y_j)\otimes r(Y_j)||_{hs}\ ||C(Y_j)||_{hs}\mathbf{1}_{\{f(Y_j)<e_n\}} \\
&\le \frac{||\hat{\Gamma}_D^{-1}||_\infty\sqrt{n}}{n}\sum_{j=1}^n \left[||r(Y_j)||_H^2\ ||R(Y_j)||_{hs} + ||r(Y_j)||_H^4\right]\mathbf{1}_{\{f(Y_j)<e_n\}}.
\end{aligned}$$

Thus

$$\mathbb{E}\left[\frac{\sqrt{n}||A_{312n}||_{hs}}{||\hat{\Gamma}_D^{-1}||_\infty}\right] \le \sqrt{n}\mathbb{E}\left[||r(Y)||_H^2\ ||R(Y)||_{hs}\mathbf{1}_{\{f(Y)<e_n\}}\right] + \sqrt{n}\mathbb{E}\left[||r(Y)||_H^4 \mathbf{1}_{\{f(Y)<e_n\}}\right],$$

and since $||\hat{\Gamma}_D^{-1}||_\infty = O_p\left(\frac{1}{t_D}\right)$, we deduce from the preceding inequality, assumption $(\mathscr{A}_{10})$ and Markov inequality that $A_{312n} = o_p\left(\frac{1}{t_D\sqrt{n}}\right)$. This permits to conclude that $A_{31n} = A_{311n} + A_{312n} = o_p\left(\frac{1}{t_D\sqrt{n}}\right)$. Now, we deal with A_{32n}; we have

$$\begin{aligned}
A_{32n} &= \frac{1}{n}\sum_{j=1}^n C_{e_n}(Y_j)\hat{\Gamma}_D^{-1}\left[R_{e_n}(Y_j) - R(Y_j)\right] - \frac{1}{n}\sum_{j=1}^n C_{e_n}(Y_j)\hat{\Gamma}_D^{-1}\left[r_{e_n}(Y_j)\otimes r_{e_n}(Y_j) - r(Y_j)\otimes r(Y_j)\right] \\
&= A_{321n} - A_{322n}
\end{aligned}$$

and

$$\begin{aligned}
A_{321n} &= \frac{1}{n}\sum_{j=1}^n C_{e_n}(Y_j)\hat{\Gamma}_D^{-1}\left[R_{e_n}(Y_j) - R(Y_j)\right] \\
&= \frac{1}{n}\sum_{j=1}^n R_{e_n}(Y_j)\hat{\Gamma}_D^{-1}\left[R_{e_n}(Y_j) - R(Y_j)\right] - \frac{1}{n}\sum_{j=1}^n r_{e_n}(Y_j)\otimes r_{e_n}(Y_j)\hat{\Gamma}_D^{-1}\left[R_{e_n}(Y_j) - R(Y_j)\right] \\
&= A_{3211n} - A_{3212n}.
\end{aligned}$$

Moreover

$$\begin{aligned}\sqrt{n}||A_{3211n}||_{hs} &\leq \frac{||\hat{\Gamma}_D^{-1}||_\infty\sqrt{n}}{n}\sum_{j=1}^{n}||R_{e_n}(Y_j)||_{hs}\ \ ||\left[R_{e_n}(Y_j)-R(Y_j)\right]||_{hs}\\ &\leq \frac{||\hat{\Gamma}_D^{-1}||_\infty\sqrt{n}}{n}\sum_{j=1}^{n}||R_{e_n}(Y_j)||_{hs}\ \ ||m(Y_j)||_{hs}\left|\frac{1}{f_{e_n}(Y_j)}-\frac{1}{f(Y_j)}\right|\\ &\leq \frac{||\hat{\Gamma}_D^{-1}||_\infty\sqrt{n}}{n}\sum_{j=1}^{n}||R_{e_n}(Y_j)||_{hs}\ \ ||R(Y_j)||_{hs}\mathbf{1}_{\{f(Y_j)<e_n\}}\\ &\leq \frac{||\hat{\Gamma}_D^{-1}||_\infty\sqrt{n}}{n}\sum_{j=1}^{n}||R(Y_j)||_{hs}^2\mathbf{1}_{\{f(Y_j)<e_n\}}.\end{aligned}$$

Hence

$$\mathbb{E}\left[\frac{\sqrt{n}||A_{3211n}||_{hs}}{||\hat{\Gamma}_D^{-1}||_\infty}\right]\leq\sqrt{n}\mathbb{E}\left[||R(Y)||_{hs}^2\mathbf{1}_{\{f(Y)<e_n\}}\right];$$

then using $||\hat{\Gamma}_D^{-1}||_\infty=O_p\left(\frac{1}{t_D}\right)$, assumption $(\mathscr{A}_{10})$ and Markov inequality, we conclude that $A_{3211n}=o_p\left(\frac{1}{t_D\sqrt{n}}\right)$. Furthermore, we have

$$\begin{aligned}\sqrt{n}||A_{3212n}||_{hs} &\leq \frac{||\hat{\Gamma}_D^{-1}||_\infty\sqrt{n}}{n}\sum_{j=1}^{n}||r(Y_j)\otimes r(Y_j)||_{hs}\ \ ||M(Y_j)||_{hs}\left|\frac{1}{f_{e_n}(Y_j)}-\frac{1}{f(Y_j)}\right|\\ &\leq \frac{||\hat{\Gamma}_D^{-1}||_\infty\sqrt{n}}{n}\sum_{j=1}^{n}||r(Y_j)||_H^2\ \ ||R(Y_j)||_{hs}\mathbf{1}_{\{f(Y_j)<e_n\}},\end{aligned}$$

what implies

$$\mathbb{E}\left[\frac{\sqrt{n}||A_{3212n}||_{hs}}{||\hat{\Gamma}_D^{-1}||_\infty}\right]\leq\sqrt{n}\mathbb{E}\left[||r(Y)||_H^2\ \ ||R(Y)||_{hs}\mathbf{1}_{\{f(Y)<e_n\}}\right]=o_p(1).$$

Thus, $A_{3212n}=o_p\left(\frac{1}{t_D\sqrt{n}}\right)$ and we can then conclude that $A_{321n}=o_p\left(\frac{1}{t_D\sqrt{n}}\right)$. It remains to treat A_{322n}. We have:

$$\begin{aligned}A_{322n} &= \frac{1}{n}\sum_{j=1}^{n}R_{e_n}(Y_j)\hat{\Gamma}_D^{-1}\left[r_{e_n}(Y_j)\otimes r_{e_n}(Y_j)-r(Y_j)\otimes r(Y_j)\right]\\ &\quad-\frac{1}{n}\sum_{j=1}^{n}r_{e_n}(Y_j)\otimes r_{e_n}(Y_j)\hat{\Gamma}_D^{-1}\left[r_{e_n}(Y_j)\otimes r_{e_n}(Y_j)-r(Y_j)\otimes r(Y_j)\right]\\ &= A_{3221n}-A_{3222n},\end{aligned}$$

and since

$$\sqrt{n}||A_{3221n}||_{hs}\leq||\hat{\Gamma}_D^{-1}||_\infty\sqrt{n}\mathbb{E}\left[||r(Y)||_H^2\ \ ||R(Y)||_{hs}\mathbf{1}_{\{f(Y)<e_n\}}\right],$$

we obtain from assumption $(\mathscr{A}_{10})$ that $A_{3221n}=o_p\left(\frac{1}{t_D\sqrt{n}}\right)$. Further,

$$\sqrt{n}\mathbb{E}[||A_{3222n}]||_{hs} \leq ||\hat{\Gamma}_D^{-1}||_\infty \sqrt{n}\mathbb{E}\left[||r(Y)||_H^4 \mathbf{1}_{\{f(Y)<e_n\}}\right],$$

and from assumtion (A_8) and Markov inequality we deduce that $A_{3222n} = o_p\left(\frac{1}{t_D\sqrt{n}}\right)$. Consequently, $A_{322n} = o_p\left(\frac{1}{t_D\sqrt{n}}\right)$ and $A_{32n} = o_p\left(\frac{1}{t_D\sqrt{n}}\right)$. All of the above permit to conclude that $A_{3n} = o_p\left(\frac{1}{t_D\sqrt{n}}\right)$. □

Lemma 5. *Under assumptions* $(\mathscr{A}_3), (\mathscr{A}_6)$ *to* $(\mathscr{A}_9)$ *if we suppose that* $||\hat{\Gamma}_D - \Gamma_D||_\infty = o_p(t_D)$, *we have:*

$$\begin{aligned} A_{4n} &= \frac{1}{n}\sum_{j=1}^n C_{e_n}(Y_j)\hat{\Gamma}_D^{-1} C_{e_n}(Y_j) - \frac{1}{n}\sum_{j=1}^n \hat{C}_{e_n}(Y_j)\hat{\Gamma}_D^{-1}\hat{C}_{e_n}(Y_j) \\ &= O_p\left(\frac{1}{t_D e_n}\left(h^k + \frac{\sqrt{\log(n)}}{h\sqrt{n}}\right)\right) + O_p\left(\frac{1}{t_D\sqrt{n}}\right) \\ &= O_p\left(\frac{1}{t_D n^\gamma}\right), \end{aligned}$$

where γ *is a real constant satisfying* $0 < \gamma < 1/4$.

Proof.

$$A_{4n} = \frac{1}{n}\sum_{j=1}^n \left[C_{e_n}(Y_j) - \hat{C}_{e_n}(Y_j)\right]\hat{\Gamma}_D^{-1} C_{e_n}(Y_j) - \frac{1}{n}\sum_{j=1}^n \left[\hat{C}_{e_n}(Y_j) - C_{e_n}(Y_j)\right]\hat{\Gamma}_D^{-1}\left[\hat{C}_{e_n}(Y_j) - C_{e_n}(Y_j)\right]$$
$$+ \frac{1}{n}\sum_{j=1}^n C_{e_n}(Y_j)\hat{\Gamma}_D^{-1}\left[C_{e_n}(Y_j) - \hat{C}_{e_n}(Y_j)\right] = A_{41n} - A_{42n} + A_{43n}.$$

First, we deal with A_{41n}; we have:

$$\begin{aligned} A_{41n} &= \frac{1}{n}\sum_{j=1}^n \left[R_{e_n}(Y_j) - \hat{R}_{e_n}(Y_j)\right]\hat{\Gamma}_D^{-1} C_{e_n}(Y_j) \\ &\quad + \left[\hat{r}_{e_n}(Y_j) \otimes \hat{r}_{e_n}(Y_j) - r_{e_n}(Y_j) \otimes r_{e_n}(Y_j)\right]\hat{\Gamma}_D^{-1} C_{e_n}(Y_j) \\ &= \frac{1}{n}\sum_{j=1}^n \left[R_{e_n}(Y_j) - \hat{R}_{e_n}(Y_j)\right]\hat{\Gamma}_D^{-1} C_{e_n}(Y_j) \\ &\quad + \frac{1}{n}\sum_{j=1}^n \left(\hat{r}_{e_n}(Y_j) - r_{e_n}(Y_j)\right) \otimes \left(\hat{r}_{e_n}(Y_j) - r_{e_n}(Y_j)\right)\hat{\Gamma}_D^{-1} C_{e_n}(Y_j) \\ &\quad + \frac{1}{n}\sum_{j=1}^n \left(\hat{r}_{e_n}(Y_j) - r_{e_n}(Y_j)\right) \otimes r_{e_n}(Y_j)\hat{\Gamma}_D^{-1} C_{e_n}(Y_j) \\ &\quad + \frac{1}{n}\sum_{j=1}^n r_{e_n}(Y_j) \otimes \left(\hat{r}_{e_n}(Y_j) - r_{e_n}(Y_j)\right)\hat{\Gamma}_D^{-1} C_{e_n}(Y_j) \\ &= A_{411n} + A_{412n} + A_{413n} + A_{414n} \end{aligned}$$

and

$$
\begin{aligned}
||A_{411n}||_{hs} &\leq ||\hat{\Gamma}_D^{-1}||_\infty \frac{1}{n}\sum_{j=1}^{n} ||R_{e_n}(Y_j) - \hat{R}_{e_n}(Y_j)||_{hs} \; ||C(Y_j)||_{hs} \\
&\leq ||\hat{\Gamma}_D^{-1}||_\infty \frac{1}{n}\sum_{j=1}^{n} ||R_{e_n}(Y_j) - \hat{R}_{e_n}(Y_j)||_{hs} \; ||C(Y_j)||_{hs} \\
&\leq ||\hat{\Gamma}_D^{-1}||_\infty \frac{1}{n}\sum_{j=1}^{n} ||\frac{R_{e_n}(Y_j)}{\hat{f}_{e_n}(Y_j)}\left[f(Y_j) - \hat{f}(Y_j)\right] + \frac{1}{\hat{f}_{e_n}(Y_j)}\left[\hat{M}(Y_j) - M(Y_j)\right]||_{hs} \; ||C(Y_j)||_{hs} \\
&\leq \frac{||\hat{\Gamma}_D^{-1}||_\infty}{e_n} \frac{1}{n}\sum_{j=1}^{n} ||R(Y_j)||_{hs} \; ||C(Y_j)||_{hs} \sup_{y\in\mathbb{R}} |f(y) - \hat{f}(y)| \\
&+ \frac{||\hat{\Gamma}_D^{-1}||_\infty}{e_n} \frac{1}{n}\sum_{j=1}^{n} \; ||C(Y_j)||_{hs} \sup_{y\in\mathbb{R}} ||M(y) - \hat{M}(y)||_{hs}.
\end{aligned}
$$

It is known from Prakasa Rao (1983) that

$$
\sup_{y\in\mathbb{R}} |\hat{f}(y) - f(y)| = O_p\left(h^k + \frac{\sqrt{\log(n)}}{h\sqrt{n}}\right); \tag{4}
$$

then, this property together with Lemma 1, assumption $(\mathscr{A}_9)$ and the preceding inequality imply

$$
A_{411n} = O_p\left(\frac{1}{t_D e_n}\left(h^k + \frac{\sqrt{\log(n)}}{h\sqrt{n}}\right)\right) = O_p\left(\frac{1}{t_D n^\gamma}\right).
$$

A similar reasoning, but by using instead of Lemma 1 the following result from Yao and Müller (2010):

$$
\sup_{y\in\mathbb{R}} ||\hat{m}(y) - m(y)||_H = O_p\left(h^k + \frac{\sqrt{\log(n)}}{h\sqrt{n}}\right)
$$

permits to obtain

$$
A_{413n} = O_p\left(\frac{1}{t_D e_n}\left(h^k + \frac{\sqrt{\log(n)}}{h\sqrt{n}}\right)\right) = O_p\left(\frac{1}{t_D n^\gamma}\right)
$$

and

$$
A_{414n} = O_p\left(\frac{1}{t_D e_n}\left(h^k + \frac{\sqrt{\log(n)}}{h\sqrt{n}}\right)\right) = O_p\left(\frac{1}{t_D n^\gamma}\right).
$$

On the other hand,

$$
\begin{aligned}
||A_{412n}||_{hs} &= \frac{1}{n}\sum_{j=1}^{n} ||\left(\hat{r}_{e_n}(Y_j) - r_{e_n}(Y_j)\right) \otimes \left(\hat{r}_{e_n}(Y_j) - r_{e_n}(Y_j)\right) \hat{\Gamma}_D^{-1} C_{e_n}(Y_j)||_{hs} \\
&\leq ||\hat{\Gamma}_D^{-1}||_\infty \frac{1}{n}\sum_{j=1}^{n} ||\hat{r}_{e_n}(Y_j) - r_{e_n}(Y_j)||_{hs}^2 ||C(Y_j)||_{hs}.
\end{aligned}
$$

Similar developpements as previously done for $||A_{411n}||_{hs}$ permit to obtain $\frac{1}{n}\sum_{j=1}^{n}||\hat{r}_{e_n}(Y_j) - r_{e_n}(Y_j)||_{hs}^2||C(Y_j)||_{hs} = O_p(\frac{1}{\sqrt{n}})$, and since $||\hat{\Gamma}_D^{-1}||_{\infty} = O_p(\frac{1}{t_D})$, we conclude that $A_{412n} = O_p(\frac{1}{t_D\sqrt{n}})$. Therefore,

$$A_{41n} = O_p\left(\frac{1}{t_D e_n}\left(h^k + \frac{\sqrt{\log(n)}}{h\sqrt{n}}\right)\right) + O_p\left(\frac{1}{t_D\sqrt{n}}\right) = O_p\left(\frac{1}{t_D n^{\gamma}}\right).$$

Further,

$$\begin{aligned}
A_{42n} =& \frac{1}{n}\sum_{j=1}^{n}\left[\hat{R}_{e_n}(Y_j) - R_{e_n}(Y_j)\right]\hat{\Gamma}_D^{-1}\left[\hat{R}_{e_n}(Y_j) - R_{e_n}(Y_j)\right] \\
&- \frac{1}{n}\sum_{j=1}^{n}\left[\hat{R}_{e_n}(Y_j) - R_{e_n}(Y_j)\right]\hat{\Gamma}_D^{-1}\left[\hat{r}_{e_n}(Y_j)\otimes\hat{r}_{e_n}(Y_j) - r_{e_n}(Y_j)\otimes r_{e_n}(Y_j)\right] \\
&- \frac{1}{n}\sum_{j=1}^{n}\left[\hat{r}_{e_n}(Y_j)\otimes\hat{r}_{e_n}(Y_j) - r_{e_n}(Y_j)\otimes r_{e_n}(Y_j)\right]\hat{\Gamma}_D^{-1}\left[\hat{R}_{e_n}(Y_j) - R_{e_n}(Y_j)\right] \\
&+ \frac{1}{n}\sum_{j=1}^{n}\left[\hat{r}_{e_n}(Y_j)\otimes\hat{r}_{e_n}(Y_j) - r_{e_n}(Y_j)\otimes r_{e_n}(Y_j)\right]\hat{\Gamma}_D^{-1}\left[\hat{r}_{e_n}(Y_j)\otimes\hat{r}_{e_n}(Y_j) - r_{e_n}(Y_j)\otimes r_{e_n}(Y_j)\right] \\
=& A_{421n} - A_{422n} - A_{423n} + A_{424n}
\end{aligned}$$

and

$$\begin{aligned}
\sqrt{n}||A_{421n}||_{hs} \leq& \frac{||\hat{\Gamma}_D^{-1}||_{\infty}\sqrt{n}}{n}\sum_{j=1}^{n}||\hat{R}_{e_n}(Y_j) - R_{e_n}(Y_j)||_{hs}^2 \\
\leq& \frac{||\hat{\Gamma}_D^{-1}||_{\infty}\sqrt{n}}{n}\sum_{j=1}^{n}\left|\left|\frac{R_{e_n}(Y_j)}{\hat{f}_{e_n}(Y_j)}\left[f_{e_n}(Y_j) - \hat{f}_{e_n}(Y_j)\right] - \frac{1}{\hat{f}_{e_n}(Y_j)}\left[\hat{M}(Y_j) - M(Y_j)\right]\right|\right|_{hs}^2 \\
=& \frac{||\hat{\Gamma}_D^{-1}||_{\infty}\sqrt{n}}{n}\sum_{j=1}^{n}\left|\left|\frac{R_{e_n}(Y_j)}{\hat{f}_{e_n}(Y_j)}\left[f_{e_n}(Y_j) - \hat{f}_{e_n}(Y_j)\right]\right|\right|_{hs}^2 \\
&+ \frac{||\hat{\Gamma}_D^{-1}||_{\infty}\sqrt{n}}{n}\sum_{j=1}^{n}\left|\left|\frac{1}{\hat{f}_{e_n}(Y_j)}\left[\hat{M}(Y_j) - M(Y_j)\right]\right|\right|_{hs}^2 \\
&- \frac{2||\hat{\Gamma}_D^{-1}||_{\infty}\sqrt{n}}{n}\sum_{j=1}^{n}\left\langle\frac{R_{e_n}(Y_j)}{\hat{f}_{e_n}(Y_j)}\left[f_{e_n}(Y_j) - \hat{f}_{e_n}(Y_j)\right], \frac{1}{\hat{f}_{e_n}(Y_j)}\left[\hat{M}(Y_j) - M(Y_j)\right]\right\rangle_{hs} \\
\leq& \frac{||\hat{\Gamma}_D^{-1}||_{\infty}\sqrt{n}}{ne_n^2}\sum_{j=1}^{n}||R_{e_n}(Y_j)||_{hs}^2\left|f_{e_n}(Y_j) - \hat{f}_{e_n}(Y_j)\right|^2 \\
&+ \frac{||\hat{\Gamma}_D^{-1}||_{\infty}\sqrt{n}}{ne_n^2}\sum_{j=1}^{n}||\hat{M}(Y_j) - M(Y_j)||_{hs}^2 \\
&+ \frac{2||\hat{\Gamma}_D^{-1}||_{\infty}\sqrt{n}}{ne_n^2}\sum_{j=1}^{n}||R_{e_n}(Y_j)||_{hs}\ ||\hat{M}(Y_j) - M(Y_j)||_{hs}\ \left|f_{e_n}(Y_j) - \hat{f}_{e_n}(Y_j)\right| \\
\leq& \frac{||\hat{\Gamma}_D^{-1}||_{\infty}\sqrt{n}}{ne_n^2}\sum_{j=1}^{n}||R(Y_j)||_{hs}^2\left(\sup_{y\in\mathbb{R}}\left|\hat{f}(y) - f(y)\right|\right)^2 \\
&+ \frac{||\hat{\Gamma}_D^{-1}||_{\infty}\sqrt{n}}{ne_n^2}\sum_{j=1}^{n}\left(\sup_{y\in\mathbb{R}}||\hat{M}(y) - M(y)||_{hs}\right)^2 \\
&+ \frac{2||\hat{\Gamma}_D^{-1}||_{\infty}\sqrt{n}}{ne_n^2}\sum_{j=1}^{n}||R(Y_j)||_{hs}\ \sup_{y\in\mathbb{R}}||\hat{M}(y) - M(y)||_{hs}\ \sup_{y\in\mathbb{R}}\left|\hat{f}(y) - f(y)\right|.
\end{aligned}$$

From the weak law of large numbers we obtain $\frac{1}{n}\sum_{j=1}^{n}||R(Y_j)||_{hs}=O_p(1)$ and $\frac{1}{n}\sum_{j=1}^{n}||R(Y_j)||_{hs}^2=O_p(1)$. Then using (4), Lemma 1, the assumption $(\mathscr{A}_9)$ and the fact that $||\hat{\Gamma}_D^{-1}||_\infty=O_p\left(\frac{1}{t_D}\right)$, we obtain

$$\begin{aligned}\sqrt{n}||A_{421n}||_{hs}&=O_p\left[\frac{1}{t_D}n^{1/2+2c_2}\left(h^k+\frac{1}{h}\sqrt{\frac{\log(n)}{n}}\right)^2\right]\\&=O_p\left[\frac{1}{t_D}\left(n^{c_2-kc_1+1/4}+n^{c_1+c_2-1/4}\sqrt{\log(n)}\right)^2\right]\\&=O_p\left(\frac{1}{t_D}\right)\end{aligned}$$

from what we deduce that $A_{421n}=O_p\left(\frac{1}{t_D\sqrt{n}}\right)$. In the same way, we show that $A_{422n}=O_p\left(\frac{1}{t_D\sqrt{n}}\right)$, $A_{423n}=O_p\left(\frac{1}{t_D\sqrt{n}}\right)$, $A_{424n}=O_p\left(\frac{1}{t_D\sqrt{n}}\right)$. Thus, $A_{42n}=O_p\left(\frac{1}{t_D\sqrt{n}}\right)$. Now, we deal with A_{43n}; since $A_{43n}=(A_{41n})^*$, we also have

$$A_{43n}=O_p\left(\frac{1}{t_De_n}\left(h^k+\frac{\sqrt{\log(n)}}{h\sqrt{n}}\right)\right)+O_p\left(\frac{1}{t_D\sqrt{n}}\right)=O_p\left(\frac{1}{t_Dn^\gamma}\right).$$

Finally, we obtain

$$\begin{aligned}A_{4n}&=\frac{1}{n}\sum_{j=1}^{n}C_{e_n}(Y_j)\hat{\Gamma}_D^{-1}C_{e_n}(Y_j)-\frac{1}{n}\sum_{j=1}^{n}\hat{C}_{e_n}(Y_j)\hat{\Gamma}_D^{-1}\hat{C}_{e_n}(Y_j)\\&=O_p\left(\frac{1}{t_De_n}\left(h^k+\frac{\sqrt{\log(n)}}{h\sqrt{n}}\right)\right)+O_p\left(\frac{1}{t_D\sqrt{n}}\right)\\&=O_p\left(\frac{1}{t_Dn^\gamma}\right).\end{aligned}$$

6.2. Proof of Theorem 1

Since

$$\begin{aligned}&\mathbb{E}\left[Var(X|Y)\Gamma_D^{-1}Var(X|Y)\right]-\frac{1}{n}\sum_{j=1}^{n}\hat{C}_{e_n}(Y_j)\hat{\Gamma}_D^{-1}\hat{C}_{e_n}(Y_j)\\&=\left(\mathbb{E}\left[Var(X|Y)\Gamma_D^{-1}Var(X|Y)\right]-\frac{1}{n}\sum_{j=1}^{n}C(Y_j)\Gamma_D^{-1}C(Y_j)\right)\\&\quad+\left(\frac{1}{n}\sum_{j=1}^{n}C(Y_j)\Gamma_D^{-1}C(Y_j)-\frac{1}{n}\sum_{j=1}^{n}C(Y_j)\hat{\Gamma}_D^{-1}C(Y_j)\right)\\&\quad+\left(\frac{1}{n}\sum_{j=1}^{n}C(Y_j)\hat{\Gamma}_D^{-1}C(Y_j)-\frac{1}{n}\sum_{j=1}^{n}C_{e_n}(Y_j)\hat{\Gamma}_D^{-1}C_{e_n}(Y_j)\right)\\&\quad+\left(\frac{1}{n}\sum_{j=1}^{n}C_{e_n}(Y_j)\hat{\Gamma}_D^{-1}C_{e_n}(Y_j)-\frac{1}{n}\sum_{j=1}^{n}\hat{C}_{e_n}(Y_j)\hat{\Gamma}_D^{-1}\hat{C}_{e_n}(Y_j)\right)\\&=A_{1n}+A_{2n}+A_{3n}+A_{4n}.\end{aligned}$$

From the central limit theorem we have $A_{1n} = O_p\left(\frac{1}{\sqrt{n}}\right)$; then the required result is obtained by applying lemmas 2 to 4.

6.3. Proof of Theorem 2

Putting

$$G = 2\Gamma_e + \Psi - \Gamma, \quad G_D = 2\Gamma_e + \mathbb{E}\left[Var(X|Y)\Gamma_D^{-1}Var(X|Y)\right] - \Gamma \tag{5}$$

and

$$\hat{G} = 2\hat{\Gamma}_{e,n} + \frac{1}{n}\sum_{j=1}^{n}\hat{C}_{e_n}(Y_j)\hat{\Gamma}_D^{-1}\hat{C}_{e_n}(Y_j) - \Gamma_n = 2\hat{\Gamma}_{e,n} + \hat{\Psi}_{e_n,D} - \Gamma_n, \tag{6}$$

we have:

$$\begin{aligned} ||\hat{\Gamma}_{I,n} - \Gamma_I||_{hs} \le\ & ||\Gamma^{-1}G - \Gamma^{-1}G_D||_{hs} + ||\Gamma^{-1}G_D - \Gamma_D^{-1}G_D||_{hs} + ||\Gamma_D^{-1}G_D - \hat{\Gamma}_D^{-1}G_D||_{hs} \\ & + ||\hat{\Gamma}_D^{-1}G_D - \hat{\Gamma}_D^{-1}\hat{G}||_{hs} \\ =\ & K_{1n} + K_{2n} + K_{3n} + K_{4n}. \end{aligned} \tag{7}$$

First,

$$\lim_{D\to+\infty} K_{1n} = \lim_{D\to+\infty} ||\mathbb{E}\left[\Gamma^{-1}Var(X|Y)(\Gamma^{-1} - \Gamma_D^{-1})Var(X|Y)\right]||_{hs} = 0,$$

and since $K_{2n} = ||\Gamma^{-1}G_D - \Gamma_D^{-1}G_D||_{hs} \le ||(\Gamma^{-1} - \Gamma_D^{-1})G||_{hs}$, we also have $\lim_{D\to+\infty} K_{2n} = 0$. Further,

$$\begin{aligned} K_{3n} &= ||\Gamma_D^{-1}G_D - \hat{\Gamma}_D^{-1}G_D||_{hs} \le ||(\Gamma_D^{-1} - \hat{\Gamma}_D^{-1})G||_{hs} \\ &\le ||\hat{\Gamma}_D^{-1}(\hat{\Gamma}_D - \Gamma_D)\Gamma_D^{-1}G||_{hs} \le ||\hat{\Gamma}_D^{-1}||_{hs}\ ||\hat{\Gamma}_D - \Gamma_D||_{hs}\ ||\Gamma^{-1}G||_{hs}, \end{aligned}$$

then since $||\hat{\Gamma}_D^{-1}||_{hs} = O_p(1/t_D)$ and $||\hat{\Gamma}_D - \Gamma_D||_{hs} = o_p(t_D)$, we deduce that $K_{3n} = o_p(1)$. On the other hand

$$\begin{aligned} K_{4n} &= ||\hat{\Gamma}_D^{-1}G_D - \hat{\Gamma}_D^{-1}\hat{G}||_{hs} \\ &\le ||\hat{\Gamma}_D^{-1}\left(\hat{\Gamma}_{e,n} - \Gamma_e\right)\Big|\Big|_{hs} + ||\hat{\Gamma}_D^{-1}\left\{\mathbb{E}\left[Var(X|Y)\Gamma_D^{-1}Var(X|Y)\right] - \Psi_{e,n}\right\}\Big|\Big|_{hs} \\ &\quad + ||\hat{\Gamma}_D^{-1}(\Gamma_n - \Gamma)||_{hs}. \end{aligned}$$

Since $||\hat{\Gamma}_{e,n} - \Gamma_e||_{hs} = O_p\left(1/\sqrt{n}\right)$ (see Ferré and Yao 2003), we deduce from the preceding inequality that

$$K_{4n} = O_p\left(\frac{1}{t_D\sqrt{n}}\right) + O_p\left(\frac{1}{t_D^2 n^{\gamma}}\right) = o_p(1).$$

Then using (7) and the previous results we obtain: $\hat{\Gamma}_{I,n} - \Gamma_I = o_p(1)$.

6.4. Proof of Theorem 3

Denoting by $(\hat{\beta}_k)_{1\le k\le K}$ the orthonormal eigenvectors associated with the K largest eigenvalues $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_K > 0$ of $\hat{\Gamma}_D^{-1}\hat{G}$ and by $(\beta_k)_{1\le k\le K}$ the orthonormal eigenvectors associated with the K largest eigenvalues $\lambda_1 > \lambda_2 > \cdots > \lambda_K > 0$ of $\Gamma^{-1}G$, where G and $\hat{G}$ are defined in (5) and (6), we will only show the convergence of the vector $\hat{\beta}_1$ as the proof for the others are the same. Clearly, $\beta_1 = \lambda_1^{-1}\Gamma_2\eta$ and $\hat{\beta}_1 = \hat{\lambda}_1^{-1}\hat{\Gamma}_2\hat{\eta}$ where $\Gamma_2 = \Gamma^{-1}\{2\Gamma_e + \Psi - \Gamma\}\Gamma^{-1/2}, \eta = \Gamma^{1/2}\beta_1$ and $\hat{\eta} = \hat{\Gamma}_D^{1/2}\hat{\beta}_1$ with

$$\hat{\Gamma}_2 \frac{1}{\hat{\lambda}_1}\hat{\Gamma}_D^{-1}\left[2\hat{\Gamma}_{e,n} + \hat{\Psi}_{e_n} - \Gamma_n\right]\hat{\Gamma}_D^{-1/2}.$$

Hence

$$\begin{aligned}||\hat{\beta}_1 - \beta_1||_H &\le ||\frac{1}{\hat{\lambda}_1}(\hat{\Gamma}_2 - \Gamma_2)\hat{\eta}||_H + ||\frac{1}{\lambda_1}\Gamma_2(\hat{\eta} - \eta)||_H + ||\left(\frac{1}{\hat{\lambda}_1} - \frac{1}{\lambda_1}\right)\Gamma_2\hat{\eta}||_H \\ &\le \frac{||\hat{\eta}||_H}{|\hat{\lambda}_1|}||\hat{\Gamma}_2 - \Gamma_2||_\infty + \frac{||\Gamma_2||_\infty}{|\lambda_1|}\ ||\hat{\eta} - \eta||_H + \frac{|\hat{\lambda}_1 - \lambda_1|}{|\lambda_1\hat{\lambda}_1|}||\Gamma_2\hat{\eta}||_H.\end{aligned}$$

Then from Lemma 1 in Ferré and Yao (2003) we obtain the inequalities

$$|\hat{\lambda}_1 - \lambda_1| \le ||\hat{\Gamma}_D^{1/2}\hat{\Gamma}_2 - \Gamma^{1/2}\Gamma_2||_\infty \quad \text{and} \quad ||\hat{\eta} - \eta||_H \le C_9||\hat{\Gamma}_D^{1/2}\hat{\Gamma}_2 - \Gamma^{1/2}\Gamma_2||_\infty, \tag{8}$$

where C_9 is an appropriate positive constant. Then, putting $L_n = ||\hat{\Gamma}_2 - \Gamma_2||_\infty$ and $M_n = ||\Gamma_n^{1/2}\hat{\Gamma}_2 - \Gamma^{1/2}\Gamma_2||_\infty$, we have

$$||\hat{\beta}_1 - \beta_1||_H \le \frac{||\hat{\eta}||_H}{|\hat{\lambda}_1|}L_n + \left(C_{10} + \frac{C_{11}||\hat{\eta}||_H}{|\hat{\lambda}_1|}\right) M_n. \tag{9}$$

Let us verify that $L_n = o_p(1)$ and $M_n = o_p(1)$. First,

$$\begin{aligned}L_n &= ||\Gamma^{-1}G\Gamma^{-1/2} - \hat{\Gamma}_D^{-1}\hat{G}\hat{\Gamma}_D^{-1/2}||_\infty \\ &\le ||\Gamma^{-1}G\Gamma^{-1/2} - \Gamma_D^{-1}G\Gamma_D^{-1/2}||_\infty + ||\Gamma_D^{-1}G\Gamma_D^{-1/2} - \hat{\Gamma}_D^{-1}G\hat{\Gamma}_D^{-1/2}||_\infty + ||\hat{\Gamma}_D^{-1}(G - \hat{G})\hat{\Gamma}_D^{-1/2}||_\infty \\ &= L_{1n} + L_{2n} + L_{3n}.\end{aligned}$$

We know that $\Gamma_D^{-1}G\Gamma_D^{-1/2} = \Pi_D\Gamma^{-1}G\Gamma^{-1/2}\Pi_D$ and putting $\Pi_D^{\perp} = I - \Pi_D$ we have

$$\begin{aligned}\Gamma_D^{-1}G\Gamma_D^{-1/2} - \Gamma^{-1}G\Gamma^{-1/2} &= \Pi_D\Gamma^{-1}G\Gamma^{-1/2}\Pi_D - \Gamma^{-1}G\Gamma^{-1/2} \\ &= \Pi_D\Gamma^{-1}G\Gamma^{-1/2}\Pi_D - \Pi_D^{\perp}\Gamma^{-1}G\Gamma^{-1/2} - \Pi_D\Gamma^{-1}G\Gamma^{-1/2} \\ &= \Pi_D\Gamma^{-1}G\Gamma^{-1/2} - \Pi_D\Gamma^{-1}G\Gamma^{-1/2}\Pi_D^{\perp} - \Pi_D^{\perp}\Gamma^{-1}G\Gamma^{-1/2} \\ &\quad - \Pi_D\Gamma^{-1}G\Gamma^{-1/2} \\ &= -\Pi_D\Gamma^{-1}G\Gamma^{-1/2}\Pi_D^{\perp} - \Pi_D^{\perp}\Gamma^{-1}G\Gamma^{-1/2}.\end{aligned}$$

Thus

$$L_{1n} \le ||\Pi_D \Gamma^{-1} G \Gamma^{-1/2} \Pi_D^{\perp}||_{\infty} + ||\Pi_D^{\perp} \Gamma^{-1} G \Gamma^{-1/2}||_{\infty}$$
$$\le ||\Gamma^{-1} G \Gamma^{-1/2} \Pi_D^{\perp}||_{\infty} + ||\Pi_D^{\perp} \Gamma^{-1} G \Gamma^{-1/2}||_{\infty}$$

and, consequently, $\lim_{D\to+\infty} L_{1n} = 0$ because $\lim_{D\to+\infty} \Pi_D^{\perp} = 0$. On the other hand,

$$L_{2n} \le ||(\Gamma_D^{-1} - \hat{\Gamma}_D^{-1}) G \Gamma_D^{-1/2}||_{\infty} + ||\Gamma_D^{-1} G (\Gamma_D^{-1/2} - \hat{\Gamma}_D^{-1/2})||_{\infty}$$
$$+ ||(\Gamma_D^{-1} - \hat{\Gamma}_D^{-1}) G (\Gamma_D^{-1/2} - \hat{\Gamma}_D^{-1/2})||_{\infty}$$

and

$$\begin{aligned}
||(\Gamma_D^{-1} - \hat{\Gamma}_D^{-1}) G \Gamma_D^{-1/2}||_{\infty} &= ||\hat{\Gamma}_D^{-1} (\hat{\Gamma}_D - \Gamma_D) \Gamma_D^{-1} G \Gamma_D^{-1/2}||_{\infty} \\
&\le ||\hat{\Gamma}_D^{-1} (\hat{\Gamma}_D - \Gamma_D) \Gamma^{-1} G \Gamma^{-1/2}||_{\infty} \\
&\le ||\hat{\Gamma}_D^{-1}||_{\infty} \; ||\hat{\Gamma}_D - \Gamma_D||_{\infty} \; ||\Gamma^{-1} G \Gamma^{-1/2}||_{\infty} \\
&= O_p\left(\frac{1}{t_D \sqrt{n}}\right) \\
&= o_p(1).
\end{aligned}$$

Using the following properties of operators (see, e.g., Fukumizu, Bach, and Gretton 2007):

$$A^{-1/2} - B^{-1/2} = A^{-1/2}(B^{3/2} - A^{3/2}) B^{-3/2} + (A - B) B^{-3/2} \quad \text{and} \quad ||A^{3/2} - B^{3/2}||_{\infty} \le C_{12} ||A - B||_{\infty}$$

we obtain:

$$||\Gamma_D^{-1} G (\Gamma_D^{-1/2} - \hat{\Gamma}_D^{-1/2})||_{\infty} = O_p\left(\frac{1}{t_D^{3/2} \sqrt{n}}\right)$$

and

$$||(\Gamma_D^{-1} - \hat{\Gamma}_D^{-1}) G (\Gamma_D^{-1/2} - \hat{\Gamma}_D^{-1/2})||_{\infty} = O_p\left(\frac{1}{t_D^{5/2} n}\right).$$

Therefore, $L_{2n} = o_p(1)$. For dealing with the last term L_{3n} we consider the operator $G_D = 2\Gamma_e + \mathbb{E}[Var(X|Y)\Gamma_D^{-1} Var(X|Y)] - \Gamma$ and we have

$$
\begin{aligned}
||\hat{\Gamma}_D^{-1}(G-G_D)\hat{\Gamma}_D^{-1/2}||_{hs} &\leq ||(\hat{\Gamma}_D^{-1}-\Gamma_D^{-1})(G-G_D)(\hat{\Gamma}_D^{-1/2}-\Gamma_D^{-1/2})||_{hs} \\
&\quad + ||(\hat{\Gamma}_D^{-1}-\Gamma_D^{-1})(G-G_D)\Gamma_D^{-1/2}||_{hs} \\
&\quad + ||\Gamma_D^{-1}(G-G_D)(\hat{\Gamma}_D^{-1/2}-\Gamma_D^{-1/2})||_{hs} + ||\Gamma_D^{-1}(G-G_D)\Gamma_D^{-1/2}||_{hs} \\
&\leq ||\hat{\Gamma}_D^{-1}-\Gamma_D^{-1}||_{hs}\ ||G-G_D||_{hs}\ ||\hat{\Gamma}_D^{-1/2}-\Gamma_D^{-1/2}||_{hs} \\
&\quad + ||\hat{\Gamma}_D^{-1}-\Gamma_D^{-1}||_{hs}\ ||(G-G_D)\Gamma^{-1/2}||_{hs} \\
&\quad + ||\Gamma^{-1}(G-G_D)||_{hs}\ ||\hat{\Gamma}_D^{-1/2}-\Gamma_D^{-1/2}||_{hs} + ||\Gamma^{-1}(G-G_D)\Gamma^{-1/2}||_{hs} \\
&\leq ||\hat{\Gamma}_D^{-1}-\Gamma_D^{-1}||_{hs}\ ||G||_{hs}\ ||\hat{\Gamma}_D^{-1/2}-\Gamma_D^{-1/2}||_{hs} + ||\hat{\Gamma}_D^{-1}-\Gamma_D^{-1}||_{hs}\ ||G\Gamma^{-1/2}||_{hs} \\
&\quad + ||\Gamma^{-1}G||_{hs}\ ||\hat{\Gamma}_D^{-1/2}-\Gamma_D^{-1/2}||_{hs} + ||\Gamma^{-1}(G-G_D)\Gamma^{-1/2}||_{hs} \\
&= O_p\left(\frac{1}{t_D\sqrt{n}}\right) + O_p\left(\frac{1}{t_D^{3/2}\sqrt{n}}\right) + O_p\left(\frac{1}{t_D^{5/2}n}\right) + o_p(1) \\
&= o_p(1).
\end{aligned}
$$

Thus

$$
\begin{aligned}
L_{3n} &\leq ||\hat{\Gamma}_D^{-1}(G-G_D)\hat{\Gamma}_D^{-1/2}||_{hs} + ||\hat{\Gamma}_D^{-1}(G_D-\hat{G})\hat{\Gamma}_D^{-1/2}||_{hs} \\
&\leq ||\hat{\Gamma}_D^{-1}(\Gamma-\Gamma_n)\hat{\Gamma}_D^{-1/2}||_\infty + 2||\hat{\Gamma}_D^{-1}(\Gamma_e-\hat{\Gamma}_{e,n})\hat{\Gamma}_D^{-1/2}||_\infty \\
&\quad + ||\hat{\Gamma}_D^{-1}\left(\mathbb{E}\left[Var(X|Y)\Gamma_D^{-1}Var(X|Y)\right] - \frac{1}{n}\sum_{j=1}^{n}\hat{C}_{e_n}(Y_j)\hat{\Gamma}_D^{-1}\hat{C}_{e_n}(Y_j)\right)\hat{\Gamma}_D^{-1/2}||_\infty + o_p(1) \\
&= O_p\left(\frac{1}{t_D^{3/2}\sqrt{n}}\right) + O_p\left(\frac{1}{t_D^{5/2}n^\gamma}\right) + o_p(1) \\
&= o_p(1)
\end{aligned}
$$

From all what precedes we deduce that $L_n = o_p(1)$. From similar reasoning we also obtain $M_n = o_p(1)$. Then from (8) and what precedes, we deduce that $\hat{\lambda}_1^{-1} = O_p(1)$ and $||\hat{\eta}||_H = O_p(1)$. Therefore, (9) allows to conclude that $||\hat{\beta}_1 - \beta_1||_H = o_p(1)$.

References

Bosq, D. 2000. *Linear processes in function spaces. Theory and applications*. New York: Springer-Verlag.

Cardot, H., F. Ferraty, and P. Sarda. 1999. Functional linear model. *Statistics & Probability Letters* 45 (1):11–22. doi:10.1016/S0167-7152(99)00036-X.

Cook, R. D. 2000. SAVE: A method for dimension reduction and graphics in regression. *Communications in Statistics: Theory and Methods* 29:2109–21.

Dauxois, J., L. Ferré, and A. F. Yao. 2001. Un modèle semi-paramétrique pour variable aléatoire hilbertienne. *Comptes Rendus de L'académie Des Sciences - Series I - Mathematics* 333 (10): 947–52. doi:10.1016/S0764-4442(01)02163-2.

Fukumizu, K., F. Bach, and A. Gretton. 2007. Statistical consistency of kernel canonical correlation analysis. *Journal of Machine Learning Research* 8:361–83.

Ferraty, F., and P. Vieu. 2002. The functional nonparametric model and application to spectrometric data. *Computational Statistics* 17 (4):545–64. doi:10.1007/s001800200126.

Ferraty, F., and P. Vieu. 2016. *Nonparametric functional data analysis: Theory and practice*. New York: Springer.

Ferré, L., and A. F. Yao. 2003. Functional sliced inverse regression analysis. *Statistics* 37 (6): 475–88. doi:10.1080/0233188031000112845.

Ferré, L., and A. F. Yao. 2005. Smoothed functional inverse regression. *Statistica Sinica* 15: 665–83.

Hall, P., and J. L. Horowitz. 2007. Methodology and convergence rates for functional linear regression. *The Annals of Statistics* 35 (1):70–91. doi:10.1214/009053606000000957.

Horvath, L., and P. Kokoszka. 2012. *Inference for functional data with applications*. New York: Springer.

Li, K. C. 1991. Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association* 86 (414):316–42. doi:10.1080/01621459.1991.10475035.

Lian, H., and G. Li. 2014. Series expansion for functional sufficent dimension reduction. *Journal of Multivariate Analysis* 124:150–65. doi:10.1016/j.jmva.2013.10.019.

Nkou, E. D. D., and G. M. Nkiet. 2019. Strong consistency of kernel estimator in a semiparametric regression model. *Statistics* 53 (6):1289–305. doi:10.1080/02331888.2019.1656723.

Prakasa Rao, B. L. S. 1983. *Nonparametric functional estimation*. Orlando, FL: Academic Press.

Ramsay, J. O., and B. W. Silverman. 1997. *Functional data analysis*. New York: Springer.

Wang, G., G. Zhou, Y. Feng, and B. X. N. Zhang. 2015. The hybrid method of FSIR and FSAVE for functional effective dimension reduction. *Computational Statistics and Data Analysis* 91: 64–77. doi:10.1016/j.csda.2015.05.011.

Yao, F., and H. G. Müller. 2010. Functional quadratic regression. *Biometrika* 97 (1):49–64. doi:10.1093/biomet/asp069.

Zhu, L. X., and K. T. Fang. 1996. Asymptotics for kernel estimate of sliced inverse regression. *Annals of Statistics* 24:1053–68.

Zhu, L. P., and L. X. Zhu. 2007. On kernel method for sliced average variance estimation. *Journal of Multivariate Analysis* 98 (5):970–91. doi:10.1016/j.jmva.2006.11.005.