Theoretical advances for cross-validation procedures in structural analysis of functional data

Frédéric Ferraty[a] Adela Martínez–Calvo[b]

Philippe Vieu[a]

[a] Université Paul Sabatier, 31062 Toulouse Cédex 9, France

[b] Universidade de Santiago de Compostela,

15782 Santiago de Compostela, Spain

Abstract

This document is a companion paper to Ferraty et al. (2013) in which an exploratory tool has been developed for detecting structural changes in some functional dataset. Its aim is to develop asymptotic theory for the method described in the previous paper. However, it is written in some self-contained way in order to be used for people who are only interested with technical aspects linked with cross-validation theopry with functional variables. The proofs are presented with all details, with main hope to overpass the specific problem of the paper (structural change detection) in order to be helpful for the future in any situation involving cross-validation and infinite dimensional random variables.

Key Words: Functional data; Asymptotics; Cross-validation; Hidden structure detection.

1 INTRODUCTION

In recent literature, many works have focused on the functional regression model with scalar response from both parametric viewpoint (Ramsay and Silverman 2005) and nonparametric one (Ferraty and Vieu 2006). In this paper, the general nonparametric framework has been chosen and the regression model given by

$$Y = r(X) + \epsilon,$$

has been studied, where Y is a real random variable, X is a random variable valued in a separable Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle)$, $r : \mathcal{H} \to \mathbb{R}$ is the regression operator, and ϵ is a real random variable such that $\mathbb{E}(\epsilon|X) = 0$ and $\mathbb{E}(\epsilon^2|X) = \sigma_\epsilon^2(X) < \infty$. Sometimes, one is confronted with complex regression structures, which are unlikely detectable using standard graphical or descriptive techniques, such as the existence of different subsamples of functional covariates or different regression models in the sample. In Ferraty et al. (2013) an exploratory tool, based on cross-validation ideas, has been developed and its nice behaviour on finite curve datasets (simulated or real) have been highlighted. The aim of this report is to complete the understanding of this method by studying its asymptotic behaviour.

Section 2 recalls the methodology while some specific notations are staetd in Section 3. The method is based on the behaviour of some specific nonparametric estimate constructed by taking in consideration some structural change. So, the main theoretical advances are divided into two parts. Firstly, in Section 4, one states as preliminary results some asymptotic theorems (with rates) for this new estimate. Then, in Section 5, we will validate the exploratory structural analysis tool given in Ferraty et al. (2013). The end of the paper contains the proofs.

2 RECALLING THE METHODOLOGY

2.1 Kernel nonparametric estimates

From now on, let $\{(X_i, Y_i)\}_{i=0}^n$ be a sample of independent and identically distributed (i.i.d.) pairs as (X, Y). The key of the procedure is to rewrite the regression operator given by $r(x) = \mathbb{E}(Y|X = x)$ as a finite sum of operators as follows. First of all, let Ψ be a function such that $\Psi : \mathcal{H} \times \mathbb{R} \to \mathcal{E}$, being $\mathcal{E}$ a beforehand fixed space, and let $\{(\mathcal{E}_1^\upsilon, \ldots, \mathcal{E}_{N_\mathcal{E}}^\upsilon)\}_{\upsilon \in \Upsilon}$ be an indexed family of sets such that $N_\mathcal{E}$ is a fixed integer such that $1 < N_\mathcal{E} < \infty$ and, for all $\upsilon \in \Upsilon$,

$$\begin{cases} \mathcal{E}_s^\upsilon \subset \mathcal{E} \quad \forall s \in S, \quad \mathcal{E}_{s_1}^\upsilon \cap \mathcal{E}_{s_2}^\upsilon = \emptyset \quad \forall s_1, s_2 \in S \text{ such that } s_1 \neq s_2, \\ \mathbb{P}(\Psi(X, Y) \in \mathcal{E}_s^\upsilon) > 0 \quad \forall s \in S, \quad \text{and} \quad \mathbb{P}(\Psi(X, Y) \in \bigcup_{s \in S} \mathcal{E}_s^\upsilon) = 1, \end{cases}$$

where $S = \{1, \ldots, N_\mathcal{E}\}$. From now on such a function Ψ will be called *structural function*.

For each $s \in S$, the next definitions can be introduced $Y_s^\upsilon = Y \mathbb{I}_{\{\Psi(X,Y) \in \mathcal{E}_s^\upsilon\}}$ being $\mathbb{I}$ the indicator function, $r_s^\upsilon(x) = \mathbb{E}(Y_s^\upsilon | X = x)$, and $\epsilon_s^\upsilon = \epsilon \mathbb{I}_{\{\Psi(X,Y) \in \mathcal{E}_s^\upsilon\}}$. Thus, one gets that $Y = \sum_{s \in S} Y_s^\upsilon$, $r(x) = \sum_{s \in S} r_s^\upsilon(x)$, and $\epsilon = \sum_{s \in S} \epsilon_s^\upsilon$. Consequently, the regression model can be expressed as

$$\sum_{s \in S} Y_s^\upsilon = \sum_{s \in S} r_s^\upsilon(X) + \sum_{s \in S} \epsilon_s^\upsilon.$$

Once the regression operator $r(\cdot)$ is written as the sum of the operators r_s^υ, each component r_s^υ can be estimated separately, using the sample $\{(X_i, Y_{i,s}^\upsilon)\}_{i=1}^n$, where $Y_{i,s}^\upsilon = Y_i \mathbb{I}_{\{\Psi(X_i, Y_i) \in \mathcal{E}_s^\upsilon\}}$ for $i = 1, \ldots, n$. Thus, an indexed family of estimates can be built by means of

$$\hat{r}^{\upsilon}(x) = \sum_{s \in S} \hat{r}_s^{\upsilon}(x), \quad \forall \upsilon \in \Upsilon.$$

Specifically, for each $s \in S$, the following kernel–type estimator

$$\hat{r}_s^{\upsilon}(x) = \frac{\sum_{i=1}^{n} Y_{i,s}^{\upsilon} K(h_s^{-1} \|X_i - x\|)}{\sum_{i=1}^{n} K(h_s^{-1} \|X_i - x\|)}$$

has been taken, where $\|\cdot\| = \langle \cdot, \cdot \rangle^{1/2}$ is the induced norm of $\mathcal{H}$, $K(\cdot)$ is a kernel function, and $h_s = h_s(n)$ is a sequence of strictly positive real bandwidths such that $h_s \in H_n \subset \mathbb{R}^+$ for all $s \in S$. Hence, the estimator $\hat{r}^{\upsilon}$ previously introduced can be expressed as

$$\hat{r}^{\upsilon}(x) = \sum_{s \in S} \frac{\sum_{i=1}^{n} Y_{i,s}^{\upsilon} K(h_s^{-1} \|X_i - x\|)}{\sum_{i=1}^{n} K(h_s^{-1} \|X_i - x\|)}. \tag{2.1}$$

Note that, when the same bandwidth is selected for all the components $\hat{r}_s^{\upsilon}$ in the estimator $\hat{r}^{\upsilon}$, that is, when exists $h \in H_n$ such that $h_s = h$ for all $s \in S$, the proposed estimator (2.1) is just the standard kernel–type estimator given by

$$\hat{r}_h(x) = \frac{\sum_{i=1}^{n} Y_i K(h^{-1} \|X_i - x\|)}{\sum_{i=1}^{n} K(h^{-1} \|X_i - x\|)}, \tag{2.2}$$

which has been studied in detail during the last few years (see, for instance, Ferraty and Vieu 2006; Ferraty et al. 2007, 2010).

2.2 A special case

To fix the idea let us consider the simple situation when $\mathcal{H} = L^2[a, b]$ and when the hidden structure is just acting as a splitting of the data into two subsamples. This can be modelled by means of strutural function Ψ in which:

$$\mathcal{E} = \mathbb{R}, \Upsilon \subset \mathbb{R} \text{ and } N_{\mathcal{E}} = 2.$$

In this case the indexed family of pairs $\{(\mathcal{E}_1^{\upsilon}, \mathcal{E}_2^{\upsilon})\}_{\upsilon \in \Upsilon}$ is given by $\mathcal{E}_1^{\upsilon} = (-\infty, \upsilon]$ and $\mathcal{E}_2^{\upsilon} = (\upsilon, +\infty)$ for each $\upsilon \in \Upsilon$.

2.3 The Cross-validation structural detection method

. From now on, the following notation is going to be used: $(\{h_s\}_{s\in S}) \equiv (h_1, \ldots, h_{N_{\mathcal{E}}})$, $(\upsilon, \{h_s\}_{s\in S}) \equiv (\upsilon, h_1, \ldots, h_{N_{\mathcal{E}}})$, $H_n^{N_{\mathcal{E}}} \equiv H_n \times \overset{N_{\mathcal{E}}}{\cdots} \times H_n$, and $\Upsilon \times H_n^{N_{\mathcal{E}}} \equiv \Upsilon \times H_n \times \overset{N_{\mathcal{E}}}{\cdots} \times H_n$. The idea is to look at the predictive performance of the estimator $\hat{r}^{\upsilon}$. At this end, one of the most widespread techniques in the literature is a cross-the validation method. This technique has been firstly investigated (and theoretically motivated) by Härdle and Marron 1985 for bandwidth selection in multivariate problems and extended to functional data in Benhenni et al. 2007 and to other parameters than bandwidths in Ait-Saïdi et al. 2008. In this case, the aim is to find $(\upsilon, \{h_s\}_{s\in S}) \in \Upsilon \times H_n^{N_{\mathcal{E}}}$ minimizing the criterion

$$\mathrm{CV}(\upsilon, \{h_s\}_{s\in S}) = \frac{1}{n}\sum_{j=1}^{n} (Y_j - \hat{r}^{\upsilon,(-j)}(X_j))^2, \tag{2.3}$$

where $\hat{r}^{\upsilon,(-j)}(x) = \sum_{s\in S} \hat{r}_s^{\upsilon,(-j)}(x)$ being

$$\hat{r}_s^{\upsilon,(-j)}(x) = \frac{\sum_{i\neq j} Y_{i,s}^{\upsilon} K(h_s^{-1}\|X_i - x\|)}{\sum_{i\neq j} K(h_s^{-1}\|X_i - x\|)}.$$

Minimizing the CV criterion defined in (2.3), the model parameters $(\upsilon, \{h_s\}_{s\in S})$ in the estimator (2.1) will be estimated by

$$(\upsilon_{\rm CV}, \{h_{s,{\rm CV}}\}_{s\in S}) = \arg \min_{(\upsilon,\{h_s\}_{s\in S})\in\Upsilon\times H_n^{N_\mathcal{E}}} {\rm CV}(\upsilon, \{h_s\}_{s\in S}).$$

This criterion serves as exploratory tool for dtecting (or not) some possible changes in the structure in the data, as illustrated in Ferraty et al. (2013). The main aim our paper is to study its assymptotic optimality property (see Section 5).

3 SOME NOTATIONS

3.1 The general setting.

Let $\tilde{K}_{s,i}$ be $\tilde{K}_{s,i}(x) = K(h_s^{-1}\|X_i - x\|)$, and introduce the following terms

$$\hat{r}^{\upsilon}_{s,N}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{Y^{\upsilon}_{i,s}\tilde{K}_{s,i}(x)}{\mathbb{E}(\tilde{K}_{s,0}(x))} \quad \text{and} \quad \hat{r}^{\upsilon}_{s,D}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{\tilde{K}_{s,i}(x)}{\mathbb{E}(\tilde{K}_{s,0}(x))}.$$

Hence, the estimator $\hat{r}^{\upsilon}$ can be expressed as

$$\hat{r}^{\upsilon}(x) = \sum_{s\in S} \frac{\sum_{i=1}^{n} Y^{\upsilon}_{i,s}\tilde{K}_{s,i}(x)}{\sum_{i=1}^{n}\tilde{K}_{s,i}(x)} = \sum_{s\in S}\frac{\hat{r}^{\upsilon}_{s,N}(x)}{\hat{r}^{\upsilon}_{s,D}(x)}.$$

Analogously, adopting the following notation

$$\hat{r}^{\upsilon,(-j)}_{s,N}(x) = \frac{1}{n-1}\sum_{i\neq j} \frac{Y^{\upsilon}_{i,s}\tilde{K}_{s,i}(x)}{\mathbb{E}(\tilde{K}_{s,0}(x))} \quad \text{and} \quad \hat{r}^{\upsilon,(-j)}_{s,D}(x) = \frac{1}{n-1}\sum_{i\neq j} \frac{\tilde{K}_{s,i}(x)}{\mathbb{E}(\tilde{K}_{s,0}(x))},$$

then $\hat{r}^{v,(-j)}$ can be built by means of the next expression

$$\hat{r}^{v,(-j)}(x) = \sum_{s\in S} \frac{\sum_{i\neq j} Y_{i,s}^{v} \tilde{K}_{s,i}(x)}{\sum_{i\neq j} \tilde{K}_{s,i}(x)} = \sum_{s\in S} \frac{\hat{r}_{s,N}^{v,(-j)}(x)}{\hat{r}_{s,D}^{v,(-j)}(x)}.$$

In the following C will denote a generic positive constant which may take on different values even in the same formula.

3.2 A particular scenario

Suppose that

(H.1) there exists a compact set $\mathcal{C}$ of $\mathcal{H}$ such that $\mathbb{P}(X \in \mathcal{C}) = 1$.

Moreover, assume that $\mathcal{E}$ is a metric space, provided with a metric $\rho_{\mathcal{E}}(\cdot,\cdot)$, and Ψ is only related to the covariate X such that

(H.2) $\Psi : \mathcal{H} \times \mathbb{R} \to \mathcal{E}$ with $\Psi(x,y) = \tilde{\Psi}(x)$ for all $(x,y) \in \mathcal{H} \times \mathbb{R}$, where $\tilde{\Psi} : \mathcal{H} \to \mathcal{E}$ is continuous on $\mathcal{C}$.

Note that when the change only depends on X

$$Y_s^v = Y \mathbb{I}_{\{\tilde{\Psi}(X)\in\mathcal{E}_s^v\}} \quad \text{and} \quad r_s^v(X) = \mathbb{E}(Y_s^v|X) = \mathbb{E}(Y|X)\mathbb{I}_{\{\tilde{\Psi}(X)\in\mathcal{E}_s^v\}} = r(X)\mathbb{I}_{\{\tilde{\Psi}(X)\in\mathcal{E}_s^v\}}. \tag{3.1}$$

Given that $\tilde{\Psi}$ is continuous on a compact set $\mathcal{C}$, $\tilde{\Psi}$ is uniformly continuous on $\mathcal{C}$ by Heine–Cantor theorem. Thus, for any $\varepsilon > 0$, there is a $\delta > 0$ such that

$$\forall x_1, x_2 \in \mathcal{C} \text{ satisfying } \|x_1 - x_2\| < \delta, \text{ the inequality } \rho_{\mathcal{E}}(\tilde{\Psi}(x_1), \tilde{\Psi}(x_2)) < \varepsilon \text{ holds.} \tag{3.2}$$

On the other hand, let $\{(\mathcal{E}_1^v, \ldots, \mathcal{E}_{N_\mathcal{E}}^v)\}_{v\in\Upsilon}$ be an indexed family of sets such that $N_\mathcal{E}$ is finite and

$$\text{(H.3)} \quad \begin{cases} \mathcal{E}_s^v \subset \mathcal{E} \quad \forall s \in S, \quad \mathcal{E}_{s_1}^v \cap \mathcal{E}_{s_2}^v = \emptyset \quad \forall s_1, s_2 \in S \text{ such that } s_1 \neq s_2, \\ \mathbb{P}(\tilde{\Psi}(X) \in \mathcal{E}_s^v) \geq c_0 > 0 \quad \forall s \in S, \quad \text{and} \quad \mathbb{P}(\tilde{\Psi}(X) \in \bigcup_{s\in S} \mathcal{E}_s^v) = 1, \end{cases}$$

which also satisfies the following condition

(H.4) there exists $D > 0$ such that

$$D = \min_{s_1,s_2\in S, s_1\neq s_2} \rho_\mathcal{E}(\mathcal{E}_{s_1}^v, \mathcal{E}_{s_2}^v), \quad \text{where} \quad \rho_\mathcal{E}(\mathcal{E}_{s_1}^v, \mathcal{E}_{s_2}^v) = \inf_{e_1\in\mathcal{E}_{s_1}^v, e_2\in\mathcal{E}_{s_2}^v} \rho_\mathcal{E}(e_1, e_2).$$

It is important to highlight that, since $\tilde{\Psi}$ is uniformly continuous (see (3.2)) and $D > 0$, there exists $\delta_D > 0$ such that for all $x_1, x_2 \in \mathcal{C}$ verifying $\|x_1 - x_2\| < \delta_D$, $\rho_\mathcal{E}(\tilde{\Psi}(x_1), \tilde{\Psi}(x_2)) < D$, and as result, $\tilde{\Psi}(x_1)$ and $\tilde{\Psi}(x_2)$ belongs to the same subset of $\{(\mathcal{E}_1^v, \ldots, \mathcal{E}_{N_\mathcal{E}}^v)\}_{v\in\Upsilon}$. Hence,

$$\forall x_1, x_2 \in \mathcal{C} \text{ satisfying } \|x_1 - x_2\| < \delta_D, \quad \mathbb{I}_{\{\tilde{\Psi}(x_1)\in\mathcal{E}_s^v\}} = \mathbb{I}_{\{\tilde{\Psi}(x_2)\in\mathcal{E}_s^v\}}, \quad \forall s \in S. \tag{3.3}$$

Therefore, in the situation that has just been described, if the bandwidth h_s satisfies $h_s < \delta_D$, (3.1) and (3.3) imply that $\hat{r}_s^v(x)$ is indeed

$$\hat{r}_s^v(x) = \frac{\sum_{i=1}^n Y_i \mathbb{I}_{\{\tilde{\Psi}(X_i)\in\mathcal{E}_s^v\}} \tilde{K}_{s,i}(x)}{\sum_{i=1}^n \tilde{K}_{s,i}(x)} = \frac{\sum_{i=1}^n Y_i \tilde{K}_{s,i}(x)}{\sum_{i=1}^n \tilde{K}_{s,i}(x)} \mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_s^v\}} = \hat{r}_{h_s}(x) \mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_s^v\}}, \tag{3.4}$$

where $\hat{r}_{h_s}$ is the standard kernel estimator with bandwidth h_s (recall that the standard kernel estimator is defined as

$$\hat{r}_h(x) = \frac{\sum_{i=1}^{n} Y_i K(h^{-1}\|X_i - x\|)}{\sum_{i=1}^{n} K(h^{-1}\|X_i - x\|)}, \tag{3.5}$$

where h is the smoothing parameter or bandwidth). In addition, since there is only a $s_x \in S$ such that $\tilde{\Psi}(x) \in \mathcal{E}_{s_x}^v$, $\hat{r}_{s_x}^v(x)$ is the only non–null component of $\hat{r}^v(x)$ and, consequently, $\hat{r}^v(x) = \sum_{s \in S} \hat{r}_s^v(x) = \hat{r}_{s_x}^v(x) = \hat{r}_{h_{s_x}}(x)$.

Remark 3.1. In a certain sense, $\hat{r}^v(x)$ can be seen as a kernel–type estimate with "local" bandwith which depends on the value $\tilde{\Psi}(x)$: for all $x \in \mathcal{C}$ such that $\tilde{\Psi}(x) \in \mathcal{E}_1^v$, the kernel estimate is computed using the bandwidth h_1; for all $x \in \mathcal{C}$ such that $\tilde{\Psi}(x) \in \mathcal{E}_2^v$, the kernel estimate is computed using the bandwidth h_2,...

4 Mean square convergence of kernel estimate

In the following, let $x \in \mathcal{C}$ be a fixed element, such that s_x denotes the only $s_x \in S$ such that $\tilde{\Psi}(x) \in \mathcal{E}_{s_x}^v$. Expectation and variance of each component $\hat{r}_s^v$ of the proposed estimator $\hat{r}^v$ are given in this section for the scenario described above. Hence, the expectation and the variance of $\hat{r}^v$ can be obtained as a corollary.

Next, before formulating the theoretical results of this section, the following functions are introduced

$$\psi_x(t) = \mathbb{E}\left((r(X) - r(x)) \mid \|X - x\| = t\right), \quad \forall t \in \mathbb{R},$$

and

$$\tau_{x,h}(t) = \frac{\varphi_x(ht)}{\varphi_x(h)}, \quad \forall t \in [0,1],$$

where φ_x is the small ball probability given by $\varphi_x(h) = \mathbb{P}(\|X - x\| \leq h)$. Moreover, the following assumptions are also required for stating the results below:

(H.5) $\varphi_x(0) = 0$, and $\tau_{x,h}(t) \to \tau_{x,0}(t)$ as $h \to 0$ for all $t \in [0, 1]$.

(H.6) $r(\cdot)$ and $\sigma^2_\epsilon(\cdot)$ are continuous in a neighborhood of x.

(H.7) $\psi'_x(0)$ exists.

(H.8) For all $s \in S$, the sequence of bandwiths h_s verifies that $\lim_{n\to+\infty} h_s = 0$, $\lim_{n\to+\infty} n\varphi_x(h_s) = +\infty$, and $h_s < \delta_D$ (with δ_D defined as in (3.3)).

(H.9) $K(\cdot)$ is a kernel supported on $[0, 1]$ with a continuous derivative on $[0, 1)$ such that $K(1) > 0$ and $K'(t) < 0$.

Theorem 4.1. *Under (H.1)–(H.4), if (H.5)–(H.9) are satisfied, then for all $s \in S$*

$$\mathbb{E}(\hat{r}^v_s(x)) = \left(r(x) + \psi'_x(0)\frac{M_{x,0}}{M_{x,1}}h_s + O\left(\frac{1}{n\varphi_x(h_s)}\right) + o(h_s)\right)\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}^v_s\}},$$

$$Var(\hat{r}^v_s(x)) = \left(\sigma^2_\epsilon(x)\frac{M_{x,2}}{M^2_{x,1}}\frac{1}{n\varphi_x(h_s)} + o\left(\frac{1}{n\varphi_x(h_s)}\right)\right)\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}^v_s\}},$$

$$Cov(\hat{r}^v_{s_1}(x), \hat{r}^v_{s_2}(x)) = 0,$$

with $M_{x,0} = K(1) - \int_0^1 (tK(t))'\tau_{x,0}(t)dt$, $M_{x,1} = K(1) - \int_0^1 K'(t)\tau_{x,0}(t)dt$ and $M_{x,2} = K^2(1) - \int_0^1 (K^2)'(t)\tau_{x,0}(t)dt$.

Corollary 4.1. *Under the assumptions of Theorem 4.1, one gets*

$$\mathbb{E}(\hat{r}^{\upsilon}(x)) = r(x) + \psi_x'(0)\frac{M_{x,0}}{M_{x,1}}\sum_{s\in S} h_s \mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_s^{\upsilon}\}} + \sum_{s\in S}\left(O\left(\frac{1}{n\varphi_x(h_s)}\right) + o(h_s)\right)\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_s^{\upsilon}\}},$$

$$Var(\hat{r}^{\upsilon}(x)) = \sigma_\epsilon^2(x)\frac{M_{x,2}}{M_{x,1}^2}\sum_{s\in S}\frac{1}{n\varphi_x(h_s)}\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_s^{\upsilon}\}} + \sum_{s\in S}\left(o\left(\frac{1}{n\varphi_x(h_s)}\right)\right)\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_s^{\upsilon}\}},$$

with $M_{x,0} = K(1) - \int_0^1 (tK(t))'\tau_{x,0}(t)dt$, $M_{x,1} = K(1) - \int_0^1 K'(t)\tau_{x,0}(t)dt$ *and* $M_{x,2} = K^2(1) - \int_0^1 (K^2)'(t)\tau_{x,0}(t)dt$. *In particular, if* s_x *denotes the only* $s_x \in S$ *such that* $\tilde{\Psi}(x) \in \mathcal{E}_{s_x}^{\upsilon}$, *then*

$$\mathbb{E}(\hat{r}^{\upsilon}(x)) = r(x) + \psi_x'(0)\frac{M_{x,0}}{M_{x,1}}h_{s_x} + O\left(\frac{1}{n\varphi_x(h_{s_x})}\right) + o(h_{s_x}),$$

$$Var(\hat{r}^{\upsilon}(x)) = \sigma_\epsilon^2(x)\frac{M_{x,2}}{M_{x,1}^2}\frac{1}{n\varphi_x(h_{s_x})} + o\left(\frac{1}{n\varphi_x(h_{s_x})}\right).$$

5 ASYMPTOTIC STUDY OF CV CRITERION

5.1 Introduction

Recall the CV criterion that was proposed in Section 2 in order to choose the parameters involved in the proposed estimator: $(\upsilon, \{h_s\}_{s\in S})$. The theoretical results below are focused on showing the optimality of this data-driven selection regarding to the mean integrated squared error given by

$$\text{MISE}(\upsilon, \{h_s\}_{s\in S}) = \mathbb{E}((r(X) - \hat{r}^{\upsilon}(X))^2),$$

which depends on the unknown regression operator and it is incalculable in practice. The first optimality result ensures that $(\upsilon_{\text{CV}}, \{h_{s,\text{CV}}\}_{s\in S})$ approximates the optimal choice in terms of MISE criterion, that is, $(\upsilon_{\text{CV}}, \{h_{s,\text{CV}}\}_{s\in S})$ approximates

$$(\upsilon^*, \{h_s^*\}_{s\in S}) = \arg \min_{(\upsilon,\{h_s\}_{s\in S})\in\Upsilon\times H_n^{N_\mathcal{E}}} \text{MISE}(\upsilon, \{h_s\}_{s\in S}).$$

5.2 Technical assumptions

First of all, it is necessary to introduce the definition of the Kolmogorov's ζ–entropy and certain assumptions. Given a subset $\mathcal{S} \subset \mathcal{H}$ and $\zeta > 0$, the Kolmogorov's ζ–entropy of $\mathcal{S}$ is defined as $H_\mathcal{S}(\zeta) = \log N_\zeta(\mathcal{S})$, where $N_\zeta(\mathcal{S})$ is the minimal number of open balls in $\mathcal{H}$ of radius ζ such that $\mathcal{S}$ is covered, that is,

$$N_\zeta(\mathcal{S}) = \min\{N \in \mathbb{N} : \exists(x_1, \ldots, x_N) \in \mathcal{H}\times \overset{N}{\cdots} \times\mathcal{H} \text{ such that } \mathcal{S} \subset \bigcup_{k=1}^{N} B(x_k, \zeta)\},$$

with $B(x_k, \zeta) = \{x \in \mathcal{H} : \|x - x_k\| \leq \zeta\}$. Besides, the conditions which are required are the following:

(H.10) $\varphi_x(0) = 0$ and $\tau_{x,h}(t) \to \tau_{x,0}(t)$ as $h \to 0$ for all $t \in [0, 1]$, for all $x \in \mathcal{C}$. Furthermore, for $h > 0$, $0 < c_1\phi(h) \leq \varphi_x(h) \leq c_2\phi(h) < +\infty$ for all $x \in \mathcal{C}$, being $c_1, c_2 > 0$ and ϕ a bijective increasing function satisfying that $\exists c_3 > 0$ and $\exists h_0 > 0$ such that $\phi'(h) < c_3$ for all $h < h_0$.

(H.11) For all $s \in S$, there exist $c_4, c_5 > 0$ such that $c_4\, n^{-\nu_2} < \phi(h_s) < c_5\, n^{-\nu_1}$, with $0 < \nu_1 < \nu_2 < 1$ (thus, $\lim_{n\to+\infty} n\phi(h_s) = +\infty$).

(H.12) There exist $c_6 > 0$ and $\beta > 0$ such that $|r(x_1) - r(x_2)| \leq c_6\|x_1 - x_2\|^\beta$, for all $x_1, x_2 \in \mathcal{C}$.

(H.13) For all $p \geq 1$, $\mathbb{E}(|Y|^p\, |X = x) \leq c_7 < +\infty$ for all $x \in \mathcal{C}$.

(H.14) There exists $c_8 > 0$ such that $\mathbb{E}(Y^2|X = x) = \sigma(x) \geq c_8$, with σ continuous on $\mathcal{C}$.

(H.15) $K(\cdot)$ is an asymmetric, bounded and Lipschitz kernel supported on $[0, 1]$ with a continuous derivative on $[0, 1)$ such that $K(1) > 0$ and $K'(t) < 0$ for all $t \in [0, 1)$.

(H.16) For all $s \in S$, the sequence of bandwiths h_s verifies that $\lim_{n\to+\infty} h_s = 0$ and $h_s < \delta_D$ (with δ_D defined as in (3.3)).

(H.17) For n large enough, $(\log n)^2/(n\phi(h_s)) < H_{\mathcal{C}}(\log n/n) < (n\phi(h_s))/\log n$, for all $s \in S$ (note that this fact implies that $\lim_{n\to+\infty} H_{\mathcal{C}}(\log n/n)/(n\phi(h_s)) = 0$ and $\lim_{n\to+\infty} \log n/(n\phi(h_s)) = 0$). Furthermore, the Kolmogorov's entropy of $\mathcal{C}$ verifies for some $c_9 > 1$ that

$$\sum_{n=1}^{+\infty} \exp\left\{(1 - c_9)H_{\mathcal{C}}\left(\frac{\log n}{n}\right)\right\} < +\infty.$$

(H.18) $card(\Upsilon \times H_n^{N\varepsilon}) = n^\alpha$ with $\alpha > 0$.

5.3 Asymptotic optimality of the cross-validated parameters

The first result is the main one of this paper since it states the asymptotic optimality, with respect to Mean Square Error, of the parameters obtained by the cross-validation procedure.

Theorem 5.1. *Under (H.1)–(H.4) and (H.10)–(H.18), one gets*

$$\frac{MISE(\upsilon^*, \{h_s^*\}_{s\in S})}{MISE(\upsilon_{CV}, \{h_{s,CV}\}_{s\in S})} \to 1 \quad a.s.$$

5.4 Additional asymptotics

In addition, if $(\{h_s^*(\upsilon)\}_{s\in S})$ is defined as

$$(\{h_s^*(\upsilon)\}_{s\in S}) = \arg \min_{(\{h_s\}_{s\in S})\in H_n^{N_\mathcal{E}}} \mathrm{MISE}(\upsilon, \{h_s\}_{s\in S}),$$

then Theorem 5.2 indicates that $(\{h_s^*(\upsilon)\}_{s\in S})$ can be approximated by $(\{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S})$, whereas Theorem 5.3 shows that both CV and MISE criteria have similar profile.

Theorem 5.2. *Under hypotheses of Theorem 5.1, one has*

$$\frac{MISE(\upsilon, \{h_s^*(\upsilon)\}_{s\in S})}{MISE(\upsilon, \{h_{s,CV}(\upsilon)\}_{s\in S})} \to 1 \quad a.s.$$

for each $\upsilon \in \Upsilon$.

Theorem 5.3. *Under hypotheses of Theorem 5.1, one has*

$$\sup_{\upsilon\in\Upsilon} \left| \frac{CV(\upsilon, \{h_{s,CV}(\upsilon)\}_{s\in S}) - MISE(\upsilon, \{h_s^*(\upsilon)\}_{s\in S}) - \hat{\sigma}_\epsilon^2}{MISE(\upsilon, \{h_s^*(\upsilon)\}_{s\in S})} \right| \to 0 \quad a.s.$$

where $\hat{\sigma}_\epsilon^2 = n^{-1}\sum_{j=1}^n \epsilon_j^2$.

6 PROOFS OF RESULTS OF SECTION 4

6.1 Proof of Theorem 4.1

As commented in (3.4), $\hat{r}_s^\upsilon(x) = \hat{r}_{h_s}(x)\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_s^\upsilon\}}$, where $\hat{r}_{h_s}$ is the standard kernel estimator (2.2). The required assumptions in this theorem guarantee that Theorem 1 by Ferraty et al. (2007) can be applied, which stated the following asymptotics for the standard kernel estimator

$$\mathbb{E}(\hat{r}_{h_s}(x)) = r(x) + \psi'_x(0)\frac{M_{x,0}}{M_{x,1}}h_s + O\left(\frac{1}{n\varphi_x(h_s)}\right) + o(h_s), \tag{6.1}$$

and

$$\mathrm{Var}(\hat{r}_{h_s}(x)) = \sigma^2_\epsilon(x)\frac{M_{x,2}}{M^2_{x,1}}\frac{1}{n\varphi_x(h_s)} + o\left(\frac{1}{n\varphi_x(h_s)}\right), \tag{6.2}$$

for all $s \in S$, with $M_{x,0} = K(1) - \int_0^1 (tK(t))'\tau_{x,0}(t)dt$, $M_{x,1} = K(1) - \int_0^1 K'(t)\tau_{x,0}(t)dt$ and $M_{x,2} = K^2(1) - \int_0^1 (K^2)'(t)\tau_{x,0}(t)dt$. Therefore, the expectation of $\hat{r}^v_s(x)$ comes from (6.1) as follows

$$\begin{aligned}\mathbb{E}(\hat{r}^v_s(x)) &= \mathbb{E}\left(\hat{r}_{h_s}(x)\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}^v_s\}}\right) = \mathbb{E}(\hat{r}_{h_s}(x))\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}^v_s\}}\\ &= \left(r(x) + \psi'_x(0)\frac{M_{x,0}}{M_{x,1}}h_s + O\left(\frac{1}{n\varphi_x(h_s)}\right) + o(h_s)\right)\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}^v_s\}},\end{aligned}$$

whereas the expression for variance is obtained using (6.2)

$$\begin{aligned}\mathrm{Var}(\hat{r}^v_s(x)) &= \mathrm{Var}(\hat{r}_{h_s}(x)\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}^v_s\}}) = \mathrm{Var}(\hat{r}_{h_s}(x))\left(\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}^v_s\}}\right)^2\\ &= \left(\sigma^2_\epsilon(x)\frac{M_{x,2}}{M^2_{x,1}}\frac{1}{n\varphi_x(h_s)} + o\left(\frac{1}{n\varphi_x(h_s)}\right)\right)\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}^v_s\}}.\end{aligned}$$

Furthermore, for all $s_1, s_2 \in S$ such that $s_1 \neq s_2$, one has

$$\begin{aligned}&\mathrm{Cov}(\hat{r}^v_{s_1}(x), \hat{r}^v_{s_2}(x)) = \mathrm{Cov}\left(\hat{r}_{h_{s_1}}(x)\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}^v_{s_1}\}}, \hat{r}_{h_{s_2}}(x)\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}^v_{s_2}\}}\right)\\ &= \mathrm{Cov}(\hat{r}_{h_{s_1}}(x), \hat{r}_{h_{s_2}}(x))\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}^v_{s_1}\}}\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}^v_{s_2}\}} = \mathrm{Cov}(\hat{r}_{h_{s_1}}(x), \hat{r}_{h_{s_2}}(x))\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}^v_{s_1}\cap\mathcal{E}^v_{s_2}\}} = 0,\end{aligned}$$

since $\mathcal{E}^v_{s_1} \cap \mathcal{E}^v_{s_2} = \emptyset$ for all $s_1 \neq s_2$ due to (H.3).

6.2 Proof of Corollary 4.1

Given that $\hat{r}^v(x) = \sum_{s\in S} \hat{r}^v_s(x)$, one has that $\mathbb{E}(\hat{r}^v(x)) = \sum_{s\in S} \mathbb{E}(\hat{r}^v_s(x))$ and

$$\mathrm{Var}(\hat{r}^v(x)) = \sum_{s\in S} \mathrm{Var}(\hat{r}^v_s(x)) + \sum_{s_1\in S}\sum_{s_1\neq s_2} \mathrm{Cov}(\hat{r}^v_{s_1}(x), \hat{r}^v_{s_2}(x)).$$

Hence, applying Theorem 4.1, one has that

$$\begin{aligned}
&\mathbb{E}(\hat{r}^v(x)) = \sum_{s\in S} \mathbb{E}(\hat{r}^v_s(x)) \\
&= \sum_{s\in S} \left(r(x) + \psi'_x(0)\frac{M_{x,0}}{M_{x,1}} h_s + O\left(\frac{1}{n\varphi_x(h_s)}\right) + o(h_s)\right) \mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}^v_s\}} \\
&= r(x) + \psi'_x(0)\frac{M_{x,0}}{M_{x,1}} \sum_{s\in S} h_s \mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}^v_s\}} + \sum_{s\in S}\left(O\left(\frac{1}{n\varphi_x(h_s)}\right) + o(h_s)\right) \mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}^v_s\}},
\end{aligned}$$

and for its variance one gets

$$\begin{aligned}
\mathrm{Var}(\hat{r}^v(x)) &= \sum_{s\in S} \mathrm{Var}(\hat{r}^v_s(x)) + \sum_{s_1\in S}\sum_{s_1\neq s_2} \mathrm{Cov}(\hat{r}^v_{s_1}(x), \hat{r}^v_{s_2}(x)) \\
&= \sum_{s\in S}\left(\sigma^2_\epsilon(x)\frac{M_{x,2}}{M^2_{x,1}}\frac{1}{n\varphi_x(h_s)} + o\left(\frac{1}{n\varphi_x(h_s)}\right)\right) \mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}^v_s\}} \\
&= \sigma^2_\epsilon(x)\frac{M_{x,2}}{M^2_{x,1}} \sum_{s\in S} \frac{1}{n\varphi_x(h_s)} \mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}^v_s\}} + \sum_{s\in S}\left(o\left(\frac{1}{n\varphi_x(h_s)}\right)\right) \mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}^v_s\}}.
\end{aligned}$$

Furthermore, if s_x denotes the only $s_x \in S$ such that $\tilde{\Psi}(x) \in \mathcal{E}^v_{s_x}$, then

$$\mathbb{E}(\hat{r}^v(x)) = r(x) + \psi'_x(0)\frac{M_{x,0}}{M_{x,1}} h_{s_x} + O\left(\frac{1}{n\varphi_x(h_{s_x})}\right) + o(h_{s_x}),$$

and

$$\mathrm{Var}(\hat{r}^v(x)) = \sigma^2_\epsilon(x)\frac{M_{x,2}}{M^2_{x,1}}\frac{1}{n\varphi_x(h_{s_x})} + o\left(\frac{1}{n\varphi_x(h_{s_x})}\right).$$

7 PROOFS OF RESULTS OF SECTION 5

7.1 Proof of Theorem 5.1

The theorem will be proven by showing that

$$\left|\frac{\mathrm{MISE}(\upsilon_{\mathrm{CV}}, \{h_{s,\mathrm{CV}}\}_{s\in S}) - \mathrm{MISE}(\upsilon^*, \{h_s^*\}_{s\in S})}{\mathrm{MISE}(\upsilon_{\mathrm{CV}}, \{h_{s,\mathrm{CV}}\}_{s\in S})}\right| \to 0 \quad a.s.$$

Let $\hat{\sigma}_\epsilon^2$ be defined as $\hat{\sigma}_\epsilon^2 = n^{-1}\sum_{j=1}^n \epsilon_j^2$, and note that $\mathrm{MISE}(\upsilon_{\mathrm{CV}}, \{h_{s,\mathrm{CV}}\}_{s\in S}) \geq \mathrm{MISE}(\upsilon^*, \{h_s^*\}_{s\in S})$ and $\mathrm{CV}(\upsilon^*, \{h_s^*\}_{s\in S}) \geq \mathrm{CV}(\upsilon_{\mathrm{CV}}, \{h_{s,\mathrm{CV}}\}_{s\in S})$. Thus, one has that

$$\begin{aligned}|\mathrm{MISE}(\upsilon_{\mathrm{CV}}, \{h_{s,\mathrm{CV}}\}_{s\in S}) - \mathrm{MISE}(\upsilon^*, \{h_s^*\}_{s\in S})| &\leq | - \mathrm{CV}(\upsilon_{\mathrm{CV}}, \{h_{s,\mathrm{CV}}\}_{s\in S}) \\ + \mathrm{MISE}(\upsilon_{\mathrm{CV}}, \{h_{s,\mathrm{CV}}\}_{s\in S}) + \hat{\sigma}_\epsilon^2 + \mathrm{CV}(\upsilon^*, \{h_s^*\}_{s\in S}) &- \mathrm{MISE}(\upsilon^*, \{h_s^*\}_{s\in S}) - \hat{\sigma}_\epsilon^2|.\end{aligned}$$

As a result, one gets

$$\begin{aligned}&\left|\frac{\mathrm{MISE}(\upsilon_{\mathrm{CV}}, \{h_{s,\mathrm{CV}}\}_{s\in S}) - \mathrm{MISE}(\upsilon^*, \{h_s^*\}_{s\in S})}{\mathrm{MISE}(\upsilon_{\mathrm{CV}}, \{h_{s,\mathrm{CV}}\}_{s\in S})}\right| \\ &\leq \left|\frac{\mathrm{CV}(\upsilon_{\mathrm{CV}}, \{h_{s,\mathrm{CV}}\}_{s\in S}) - \mathrm{MISE}(\upsilon_{\mathrm{CV}}, \{h_{s,\mathrm{CV}}\}_{s\in S}) - \hat{\sigma}_\epsilon^2}{\mathrm{MISE}(\upsilon_{\mathrm{CV}}, \{h_{s,\mathrm{CV}}\}_{s\in S})}\right| \\ &\quad + \left|\frac{\mathrm{CV}(\upsilon^*, \{h_s^*\}_{s\in S}) - \mathrm{MISE}(\upsilon^*, \{h_s^*\}_{s\in S}) - \hat{\sigma}_\epsilon^2}{\mathrm{MISE}(\upsilon^*, \{h_s^*\}_{s\in S})}\right| \left|\frac{\mathrm{MISE}(\upsilon^*, \{h_s^*\}_{s\in S})}{\mathrm{MISE}(\upsilon_{\mathrm{CV}}, \{h_{s,\mathrm{CV}}\}_{s\in S})}\right| \\ &\leq 2 \sup_{(\upsilon,\{h_s\}_{s\in S})\in\Upsilon\times H_n^{N_\mathcal{E}}} \left|\frac{\mathrm{CV}(\upsilon, \{h_s\}_{s\in S}) - \mathrm{MISE}(\upsilon, \{h_s\}_{s\in S}) - \hat{\sigma}_\epsilon^2}{\mathrm{MISE}(\upsilon, \{h_s\}_{s\in S})}\right|,\end{aligned}$$

where the last inequality is true since $\mathrm{MISE}(\upsilon^*, \{h_s^*\}_{s\in S}) \leq \mathrm{MISE}(\upsilon_{\mathrm{CV}}, \{h_{s,\mathrm{CV}}\}_{s\in S})$. Hence, the convergence is deduced from Lemma 7.1.

Remark 7.1. Ait-Saïdi et al. (2008) showed the asymptotic optimality of the cross–validation techniques in the single–functional index model. The procedure and reasonings which they proposed in their paper were mimicked in the proof of the technical lemmas which are necessary to get Theorem 5.1 and obtain the theoretical properties of the cross–validation method.

7.2 Proof of Theorem 5.2

This theorem can be proven mimicking the proof of Theorem 5.1 as follows. Given that $\mathrm{MISE}(\upsilon, \{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S}) \geq \mathrm{MISE}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S})$ and $\mathrm{CV}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S}) \geq \mathrm{CV}(\upsilon, \{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S})$, it can be shown that

$$
\begin{aligned}
&|\mathrm{MISE}(\upsilon, \{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S}) - \mathrm{MISE}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S})| \leq | - \mathrm{CV}(\upsilon, \{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S}) \\
&\quad + \mathrm{MISE}(\upsilon, \{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S}) + \hat{\sigma}_\epsilon^2 + \mathrm{CV}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S}) - \mathrm{MISE}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S}) - \hat{\sigma}_\epsilon^2|
\end{aligned}
$$

with $\hat{\sigma}_\epsilon^2 = n^{-1}\sum_{j=1}^n \epsilon_j^2$ as usual. Then

$$
\begin{aligned}
&\left|\frac{\mathrm{MISE}(\upsilon, \{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S}) - \mathrm{MISE}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S})}{\mathrm{MISE}(\upsilon, \{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S})}\right| \\
&\leq \left|\frac{\mathrm{CV}(\upsilon, \{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S}) - \mathrm{MISE}(\upsilon, \{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S}) - \hat{\sigma}_\epsilon^2}{\mathrm{MISE}(\upsilon, \{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S})}\right| \\
&\quad + \left|\frac{\mathrm{CV}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S}) - \mathrm{MISE}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S}) - \hat{\sigma}_\epsilon^2}{\mathrm{MISE}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S})}\right| \left|\frac{\mathrm{MISE}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S})}{\mathrm{MISE}(\upsilon, \{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S})}\right| \\
&\leq 2 \sup_{(\upsilon, \{h_s\}_{s\in S}) \in \Upsilon \times H_n^{N\varepsilon}} \left|\frac{\mathrm{CV}(\upsilon, \{h_s\}_{s\in S}) - \mathrm{MISE}(\upsilon, \{h_s\}_{s\in S}) - \hat{\sigma}_\epsilon^2}{\mathrm{MISE}(\upsilon, \{h_s\}_{s\in S})}\right|.
\end{aligned}
$$

Consequently, the theorem is proven due to Lemma 7.1.

7.3 Proof of Theorem 5.3

In this case, one has that

$$
\begin{aligned}
&|\mathrm{CV}(\upsilon, \{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S}) - \mathrm{MISE}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S}) - \hat{\sigma}_\epsilon^2| = |\mathrm{MISE}(\upsilon, \{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S}) \\
&\quad - \mathrm{MISE}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S}) + \mathrm{CV}(\upsilon, \{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S}) - \mathrm{MISE}(\upsilon, \{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S}) - \hat{\sigma}_\epsilon^2|.
\end{aligned}
$$

Therefore,

$$\begin{aligned}
&\left|\frac{\mathrm{CV}(\upsilon,\{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S})-\mathrm{MISE}(\upsilon,\{h_s^*(\upsilon)\}_{s\in S})-\hat{\sigma}_\epsilon^2}{\mathrm{MISE}(\upsilon,\{h_s^*(\upsilon)\}_{s\in S})}\right| \\
&\le \left|\frac{\mathrm{MISE}(\upsilon,\{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S})}{\mathrm{MISE}(\upsilon,\{h_s^*(\upsilon)\}_{s\in S})}-1\right| \\
&\quad+\left|\frac{\mathrm{CV}(\upsilon,\{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S})-\mathrm{MISE}(\upsilon,\{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S})-\hat{\sigma}_\epsilon^2}{\mathrm{MISE}(\upsilon,\{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S})}\right| \\
&\quad\cdot\left|\frac{\mathrm{MISE}(\upsilon,\{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S})}{\mathrm{MISE}(\upsilon,\{h_s^*(\upsilon)\}_{s\in S})}\right| \\
&\le \left|\frac{\mathrm{MISE}(\upsilon,\{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S})}{\mathrm{MISE}(\upsilon,\{h_s^*(\upsilon)\}_{s\in S})}-1\right| \\
&\quad+\sup_{(\upsilon,\{h_s\}_{s\in S})\in\Upsilon\times H_n^{N_{\mathcal{E}}}}\left|\frac{\mathrm{CV}(\upsilon,\{h_s\}_{s\in S})-\mathrm{MISE}(\upsilon,\{h_s\}_{s\in S})-\hat{\sigma}_\epsilon^2}{\mathrm{MISE}(\upsilon,\{h_s\}_{s\in S})}\right| \\
&\quad\cdot\left|\frac{\mathrm{MISE}(\upsilon,\{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S})}{\mathrm{MISE}(\upsilon,\{h_s^*(\upsilon)\}_{s\in S})}\right|.
\end{aligned}$$

Hence, Theorem 5.2 and Lemma 7.1 allow to finish the proof.

7.4 Technical lemmas

7.4.1 Formulation and Proof of Lemma 7.1

The main aim of the next lemma is to allow to show the optimality of the CV procedure with respect to the MISE criterion. For proving this lemma, one needs to introduce some other quadratic distances such as the average squared error

$$\mathrm{ASE}(\upsilon,\{h_s\}_{s\in S})=\frac{1}{n}\sum_{j=1}^{n}(r(X_j)-\hat{r}^{\upsilon}(X_j))^2 \tag{7.1}$$

and the following two terms

$$\widetilde{\mathrm{ASE}}(\upsilon,\{h_s\}_{s\in S})=\frac{1}{n}\sum_{j=1}^{n}(r(X_j)-\hat{r}^{\upsilon,(-j)}(X_j))^2,\quad\text{and} \tag{7.2}$$

$$\mathrm{CT}(\upsilon, \{h_s\}_{s\in S}) = \frac{1}{n}\sum_{j=1}^{n} \epsilon_j(\hat{r}^{\upsilon,(-j)}(X_j) - r(X_j)). \tag{7.3}$$

Lemma 7.1. *Under hypotheses of Theorem 5.1,*

$$\sup_{(\upsilon,\{h_s\}_{s\in S})\in\Upsilon\times H_n^{N_\mathcal{E}}} \left| \frac{CV(\upsilon, \{h_s\}_{s\in S}) - MISE(\upsilon, \{h_s\}_{s\in S}) - \hat{\sigma}_\epsilon^2}{MISE(\upsilon, \{h_s\}_{s\in S})} \right| \to 0 \quad a.s.$$

where $\hat{\sigma}_\epsilon^2 = n^{-1}\sum_{j=1}^{n} \epsilon_j^2$.

Proof. First of all, note that the CV criterion can be expressed as

$$\begin{aligned}
&\mathrm{CV}(\upsilon, \{h_s\}_{s\in S}) = \frac{1}{n}\sum_{j=1}^{n}(Y_j - \hat{r}^{\upsilon,(-j)}(X_j))^2 = \frac{1}{n}\sum_{j=1}^{n}((r(X_j) - \hat{r}^{\upsilon,(-j)}(X_j)) + \epsilon_j)^2 \\
&= \frac{1}{n}\sum_{j=1}^{n}(r(X_j) - \hat{r}^{\upsilon,(-j)}(X_j))^2 + \frac{2}{n}\sum_{j=1}^{n}(r(X_j) - \hat{r}^{\upsilon,(-j)}(X_j))\epsilon_j + \frac{1}{n}\sum_{j=1}^{n}\epsilon_j^2 \\
&= \widetilde{\mathrm{ASE}}(\upsilon, \{h_s\}_{s\in S}) - 2\mathrm{CT}(\upsilon, \{h_s\}_{s\in S}) + \hat{\sigma}_\epsilon^2 \\
&= \widetilde{\mathrm{ASE}}(\upsilon, \{h_s\}_{s\in S}) - \mathrm{ASE}(\upsilon, \{h_s\}_{s\in S}) + \mathrm{ASE}(\upsilon, \{h_s\}_{s\in S}) - 2\mathrm{CT}(\upsilon, \{h_s\}_{s\in S}) + \hat{\sigma}_\epsilon^2.
\end{aligned}$$

Therefore, one has

$$\begin{aligned}
&|\mathrm{CV}(\upsilon, \{h_s\}_{s\in S}) - \mathrm{MISE}(\upsilon, \{h_s\}_{s\in S}) - \hat{\sigma}_\epsilon^2| \le |\widetilde{\mathrm{ASE}}(\upsilon, \{h_s\}_{s\in S}) - \mathrm{ASE}(\upsilon, \{h_s\}_{s\in S})| \\
&\quad + |\mathrm{ASE}(\upsilon, \{h_s\}_{s\in S}) - \mathrm{MISE}(\upsilon, \{h_s\}_{s\in S})| + 2|\mathrm{CT}(\upsilon, \{h_s\}_{s\in S})|.
\end{aligned}$$

Taking into account this fact, one gets

$$\begin{aligned}
&\sup_{(\upsilon,\{h_s\}_{s\in S})\in\Upsilon\times H_n^{N\mathcal{E}}}\left|\frac{\mathrm{CV}(\upsilon,\{h_s\}_{s\in S})-\mathrm{MISE}(\upsilon,\{h_s\}_{s\in S})-\hat{\sigma}_\epsilon^2}{\mathrm{MISE}(\upsilon,\{h_s\}_{s\in S})}\right| \\
\leq &\sup_{(\upsilon,\{h_s\}_{s\in S})\in\Upsilon\times H_n^{N\mathcal{E}}}\left|\frac{\widetilde{\mathrm{ASE}}(\upsilon,\{h_s\}_{s\in S})-\mathrm{ASE}(\upsilon,\{h_s\}_{s\in S})}{\mathrm{MISE}(\upsilon,\{h_s\}_{s\in S})}\right| \\
&+\sup_{(\upsilon,\{h_s\}_{s\in S})\in\Upsilon\times H_n^{N\mathcal{E}}}\left|\frac{\mathrm{ASE}(\upsilon,\{h_s\}_{s\in S})-\mathrm{MISE}(\upsilon,\{h_s\}_{s\in S})}{\mathrm{MISE}(\upsilon,\{h_s\}_{s\in S})}\right| \\
&+2\sup_{(\upsilon,\{h_s\}_{s\in S})\in\Upsilon\times H_n^{N\mathcal{E}}}\left|\frac{\mathrm{CT}(\upsilon,\{h_s\}_{s\in S})}{\mathrm{MISE}(\upsilon,\{h_s\}_{s\in S})}\right|.
\end{aligned}$$

Then, the theorem is proven due to Lemma 7.2, Lemma 7.7 and Lemma 7.9. □

7.4.2 Formulation and Proof of Lemma 7.2

Recall that C will denote a generic positive constant which may take on different values even in the same formula.

Lemma 7.2. *Under hypotheses of Theorem 5.1,*

$$\sup_{(\upsilon,\{h_s\}_{s\in S})\in\Upsilon\times H_n^{N\mathcal{E}}}\left|\frac{\widetilde{ASE}(\upsilon,\{h_s\}_{s\in S})-ASE(\upsilon,\{h_s\}_{s\in S})}{MISE(\upsilon,\{h_s\}_{s\in S})}\right|\to 0 \quad a.s.$$

where $\widetilde{ASE}(\upsilon,\{h_s\}_{s\in S})$ *and* $ASE(\upsilon,\{h_s\}_{s\in S})$ *are defined at* (7.2) *and* (7.1)*, respectively.*

Proof. This proof is analogous to the proof of Lemma 3 by Ait-Saïdi et al. (2008).

By Lemma 7.3, the following expression for $\mathrm{ASE}(\upsilon,\{h_s\}_{s\in S})$ can be obtained

$$\begin{aligned}
\mathrm{ASE}(\upsilon,\{h_s\}_{s\in S}) &= \frac{1}{n}\sum_{j=1}^{n}(r(X_j)-\hat{r}^{\upsilon}(X_j))^2 = \frac{1}{n}\sum_{j=1}^{n}\left(\sum_{s\in S}(r_s^{\upsilon}(X_j)-\hat{r}_s^{\upsilon}(X_j))\right)^2\\
&= \frac{1}{n}\sum_{j=1}^{n}\left(\sum_{s\in S}(\hat{r}_{s,D}^{\upsilon}(X_j)(r_s^{\upsilon}(X_j)-\hat{r}_s^{\upsilon}(X_j))+(1-\hat{r}_{s,D}^{\upsilon}(X_j))(r_s^{\upsilon}(X_j)-\hat{r}_s^{\upsilon}(X_j)))\right)^2\\
&= \frac{1}{n}\sum_{j=1}^{n}\left(\sum_{s\in S}(\hat{r}_{s,D}^{\upsilon}(X_j)r_s^{\upsilon}(X_j)-\hat{r}_{s,N}^{\upsilon}(X_j))\right)^2 + o_{a.co.}(\mathrm{ASE}(\upsilon,\{h_s\}_{s\in S}))\\
&= \mathrm{ASE}^*(\upsilon,\{h_s\}_{s\in S}) + o_{a.co.}(\mathrm{ASE}(\upsilon,\{h_s\}_{s\in S})),
\end{aligned}$$

where $\mathrm{ASE}^*(\upsilon,\{h_s\}_{s\in S}) = n^{-1}\sum_{j=1}^{n}\left(\sum_{s\in S}(\hat{r}_{s,D}^{\upsilon}(X_j)r_s^{\upsilon}(X_j)-\hat{r}_{s,N}^{\upsilon}(X_j))\right)^2$. Analogously, it can be seen that

$$\widetilde{\mathrm{ASE}}(\upsilon,\{h_s\}_{s\in S}) = \widetilde{\mathrm{ASE}}^*(\upsilon,\{h_s\}_{s\in S}) + o_{a.co.}(\widetilde{\mathrm{ASE}}(\upsilon,\{h_s\}_{s\in S})),$$

with $\widetilde{\mathrm{ASE}}^*(\upsilon,\{h_s\}_{s\in S}) = n^{-1}\sum_{j=1}^{n}\left(\sum_{s\in S}(\hat{r}_{s,D}^{\upsilon,(-j)}(X_j)r_s^{\upsilon}(X_j)-\hat{r}_{s,N}^{\upsilon,(-j)}(X_j))\right)^2$. In order to finish the proof of the lemma, the equivalence between $\mathrm{ASE}^*(\upsilon,\{h_s\}_{s\in S})$ and $\widetilde{\mathrm{ASE}}^*(\upsilon,\{h_s\}_{s\in S})$ can be found by means of a similar procedure to that given by Härdle and Marron (1985) as follows.

First of all, note that

$$\hat{r}_{s,N}^{\upsilon,(-j)}(X_j) = \frac{1}{n-1}\sum_{i\neq j}\frac{Y_{i,s}^{\upsilon}\tilde{K}_{s,i}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))} = \frac{n}{n-1}\hat{r}_{s,N}^{\upsilon}(X_j) - \frac{1}{n-1}\frac{Y_{j,s}^{\upsilon}K(0)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))},$$

$$\hat{r}_{s,D}^{\upsilon,(-j)}(X_j) = \frac{1}{n-1}\sum_{i\neq j}\frac{\tilde{K}_{s,i}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))} = \frac{n}{n-1}\hat{r}_{s,D}^{\upsilon}(X_j) - \frac{1}{n-1}\frac{K(0)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))}.$$

Hence, one has that

$$\widetilde{\mathrm{ASE}}^*(\upsilon, \{h_s\}_{s\in S}) = \frac{1}{n}\sum_{j=1}^{n}\left(\sum_{s\in S}(\hat{r}_{s,D}^{\upsilon,(-j)}(X_j) r_s^{\upsilon}(X_j) - \hat{r}_{s,N}^{\upsilon,(-j)}(X_j))\right)^2$$

$$= \frac{1}{n}\sum_{j=1}^{n}\left(\frac{n}{n-1}\sum_{s\in S}(\hat{r}_{s,D}^{\upsilon}(X_j) r_s^{\upsilon}(X_j) - \hat{r}_{s,N}^{\upsilon}(X_j)) + \frac{K(0)}{n-1}\sum_{s\in S}\frac{Y_{j,s}^{\upsilon} - r_s^{\upsilon}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))}\right)^2$$

$$= \frac{n^2}{(n-1)^2}\mathrm{ASE}^*(\upsilon, \{h_s\}_{s\in S})$$

$$+ 2\frac{n}{(n-1)^2}K(0)\frac{1}{n}\sum_{j=1}^{n}\left(\sum_{s\in S}(\hat{r}_{s,D}^{\upsilon}(X_j) r_s^{\upsilon}(X_j) - \hat{r}_{s,N}^{\upsilon}(X_j))\right)$$

$$\cdot\left(\sum_{s\in S}\frac{Y_{j,s}^{\upsilon} - r_s^{\upsilon}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))}\right) + \frac{1}{(n-1)^2}K^2(0)\frac{1}{n}\sum_{j=1}^{n}\left(\sum_{s\in S}\frac{Y_{j,s}^{\upsilon} - r_s^{\upsilon}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))}\right)^2,$$

and, consequently,

$$|\widetilde{\mathrm{ASE}}^*(\upsilon, \{h_s\}_{s\in S}) - \mathrm{ASE}^*(\upsilon, \{h_s\}_{s\in S})| \leq \frac{2n-1}{(n-1)^2}\mathrm{ASE}^*(\upsilon, \{h_s\}_{s\in S})$$
$$+ 2\frac{n}{(n-1)^2}K(0)\left|\frac{1}{n}\sum_{j=1}^{n}\left(\sum_{s\in S}(\hat{r}_{s,D}^{\upsilon}(X_j) r_s^{\upsilon}(X_j) - \hat{r}_{s,N}^{\upsilon}(X_j))\right)\left(\sum_{s\in S}\frac{Y_{j,s}^{\upsilon} - r_s^{\upsilon}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))}\right)\right|$$
$$+ \frac{1}{(n-1)^2}K^2(0)\frac{1}{n}\sum_{j=1}^{n}\left(\sum_{s\in S}\frac{Y_{j,s}^{\upsilon} - r_s^{\upsilon}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))}\right)^2. \tag{7.4}$$

Using the Cauchy–Schwarz inequality, it can be found that

$$\left|\frac{1}{n}\sum_{j=1}^{n}\left(\sum_{s\in S}(\hat{r}_{s,D}^{\upsilon}(X_j) r_s^{\upsilon}(X_j) - \hat{r}_{s,N}^{\upsilon}(X_j))\right)\left(\sum_{s\in S}\frac{Y_{j,s}^{\upsilon} - r_s^{\upsilon}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))}\right)\right|$$
$$\leq (\mathrm{ASE}^*(\upsilon, \{h_s\}_{s\in S}))^{1/2}\left(\frac{1}{n}\sum_{j=1}^{n}\left(\sum_{s\in S}\frac{Y_{j,s}^{\upsilon} - r_s^{\upsilon}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))}\right)^2\right)^{1/2}. \tag{7.5}$$

In addition, the SLLN ensures that

$$\frac{1}{n}\sum_{j=1}^{n}\left(\sum_{s\in S}\frac{Y_{j,s}^{\upsilon}-r_s^{\upsilon}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))}\right)^2 \to \mathbb{E}\left(\left(\sum_{s\in S}\frac{Y_s^{\upsilon}-r_s^{\upsilon}(X)}{\mathbb{E}(\tilde{K}_{s,0}(X))}\right)^2\right) \quad a.s., \tag{7.6}$$

whereas Lemma 7.4 states that

$$\mathbb{E}\left(\left(\sum_{s\in S}\frac{Y_s^{\upsilon}-r_s^{\upsilon}(X)}{\mathbb{E}(\tilde{K}_{s,0}(X))}\right)^2\right) \le C\left(\sum_{s\in S}\frac{1}{\phi(h_s)}\right)^2. \tag{7.7}$$

Therefore, (7.4), (7.5), (7.6) and (7.7) ensure that

$$\begin{aligned}|\widetilde{\mathrm{ASE}}^*(\upsilon,\{h_s\}_{s\in S})-\mathrm{ASE}^*(\upsilon,\{h_s\}_{s\in S})| &\le \frac{2n-1}{(n-1)^2}\mathrm{ASE}^*(\upsilon,\{h_s\}_{s\in S})\\ &+2C^{1/2}\frac{n}{(n-1)^2}K(0)(\mathrm{ASE}^*(\upsilon,\{h_s\}_{s\in S}))^{1/2}\left(\sum_{s\in S}\frac{1}{\phi(h_s)}\right)(1+o_{a.s.}(1))\\ &+C\frac{1}{(n-1)^2}K^2(0)\left(\sum_{s\in S}\frac{1}{\phi(h_s)}\right)^2(1+o_{a.s.}(1)).\end{aligned}$$

By the previous expression and Lemma 7.6, it can be found that

$$\sup_{(\upsilon,\{h_s\}_{s\in S})\in\Upsilon\times H_n^{N\varepsilon}}\left|\frac{\widetilde{\mathrm{ASE}}^*(\upsilon,\{h_s\}_{s\in S})-\mathrm{ASE}^*(\upsilon,\{h_s\}_{s\in S})}{\mathrm{MISE}(\upsilon,\{h_s\}_{s\in S})}\right| \to 0 \quad a.s.$$

which leads to Lemma 7.2. □

7.4.3 Formulation and Proof of Lemma 7.3

For the lemma below, it is necessary to introduce the definition of the almost completely convergence. Let $\{Z_n\}_{n\in\mathbb{N}}$ be a sequence of real random variables, and let $\{u_n\}_{n\in\mathbb{N}}$ be a deterministic sequence of positive real numbers. Then $Z_n = O_{a.co.}(u_n)$ if and only if $\exists\varepsilon > 0$ such that $\sum_{n\in\mathbb{N}}\mathbb{P}(|Z_n| > \varepsilon u_n) < \infty$. In addition, note that

Borel–Cantelli Lemma ensures that if $Z_n = O_{a.co.}(u_n)$ then $Z_n = O_{a.s.}(u_n)$.

Lemma 7.3. *Under (H.1), (H.10), (H.15) and (H.17), one gets*

$$\sup_{x\in\mathcal{C}} |\hat{r}^{\upsilon}_{s,D}(x) - 1| = O_{a.co.}\left(\sqrt{\frac{H_{\mathcal{C}}(\log n/n)}{n\phi(h_s)}}\right), \quad \forall s \in S,$$

and

$$\sup_{x\in\mathcal{C}} |\hat{r}^{\upsilon,(-j)}_{s,D}(x) - 1| = O_{a.co.}\left(\sqrt{\frac{H_{\mathcal{C}}(\log n/n)}{n\phi(h_s)}}\right), \quad \forall s \in S.$$

Proof. Ferraty et al. (2010) studied rates of uniform consistency for a generalized nonparametric regression context in terms of almost completely convergence. In particular, the first statement in the lemma corresponds to Lemma 8 in Ferraty et al. (2010). Furthermore, the result still hold if $\hat{r}^{\upsilon}_{s,D}$ is replaced with $\hat{r}^{\upsilon,(-j)}_{s,D}$, since the second statement can be seen as a corollary of the first one. □

7.4.4 Formulation and Proof of Lemma 7.4

Recall that C will denote a generic positive constant which may take on different values even in the same formula.

Lemma 7.4. *Under hypotheses of Theorem 5.1,*

(i) for $p = 1, 2, \ldots$, there exists $c_{10,p} > 0$ such that

$$\mathbb{E}\left(\left|\sum_{s\in S} \frac{Y^{\upsilon}_s - r^{\upsilon}_s(X)}{\mathbb{E}(\tilde{K}_{s,0}(X))}\right|^p\right) \le c_{10,p}\left(\sum_{s\in S}\frac{1}{\phi(h_s)}\right)^p.$$

(ii) for $p = 1, 2, \ldots$, there exists $c_{11,p} > 0$ such that

$$\mathbb{E}\left(\left|\sum_{s\in S} \frac{(Y^{\upsilon}_{i,s} - r^{\upsilon}_s(X_i))\tilde{K}_{s,i}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right|^p |X_i\right) \le c_{11,p}\left(\sum_{s\in S}\frac{1}{\phi(h_s)}\right)^p \quad a.s.$$

(iii) for $p = 1, 2, \ldots$, there exists $c_{12,p} > 0$ such that for all $i \neq j$

$$\mathbb{E}\left(\left|\sum_{s\in S}\frac{(Y^v_{j,s} - r^v_s(X_i))\tilde{K}_{s,j}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right|^p |X_i\right) \leq c_{12,p}\left(\sum_{s\in S}\frac{1}{\phi(h_s)}\right)^{p-1} \quad a.s.$$

(iv) there exists $c_{13} > 0$ such that for all $i \neq j$

$$\mathbb{E}\left(\left(\sum_{s\in S}\frac{(Y^v_{j,s} - r^v_s(X_i))\tilde{K}_{s,j}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right)^2 |X_i\right) \geq c_{13}\sum_{s\in S}\frac{\mathbb{I}_{\{\tilde{\Psi}(X_i)\in\mathcal{E}^v_s\}}}{\phi(h_s)} \quad a.s.$$

Proof. Proof of item (i). First of all, note that (3.1) ensures that

$$Y^v_s - r^v_s(X) = Y\mathbb{I}_{\{\tilde{\Psi}(X)\in\mathcal{E}^v_s\}} - r(X)\mathbb{I}_{\{\tilde{\Psi}(X)\in\mathcal{E}^v_s\}} = (Y - r(X))\mathbb{I}_{\{\tilde{\Psi}(X)\in\mathcal{E}^v_s\}} = \epsilon\mathbb{I}_{\{\tilde{\Psi}(X)\in\mathcal{E}^v_s\}}. \tag{7.8}$$

Using (7.8), and the fact that $\mathcal{E}^v_{s_1} \cap \mathcal{E}^v_{s_2} = \emptyset$ for all $s_1 \neq s_2$, one gets

$$\begin{aligned}\mathbb{E}\left(\left|\sum_{s\in S}\frac{Y^v_s - r^v_s(X)}{\mathbb{E}(\tilde{K}_{s,0}(X))}\right|^p\right) &= \mathbb{E}\left(|\epsilon|^p\left|\sum_{s\in S}\frac{\mathbb{I}_{\{\tilde{\Psi}(X)\in\mathcal{E}^v_s\}}}{\mathbb{E}(\tilde{K}_{s,0}(X))}\right|^p\right)\\ &= \mathbb{E}\left(\mathbb{E}(|\epsilon|^p|X)\sum_{s\in S}\frac{\mathbb{I}_{\{\tilde{\Psi}(X)\in\mathcal{E}^v_s\}}}{(\mathbb{E}(\mathbb{E}(\tilde{K}_{s,0}(X)|X))^p}\right).\end{aligned}$$

Then, by assumption (H.13) and Lemma 7.5,

$$\begin{aligned}&\mathbb{E}\left(\left|\sum_{s\in S}\frac{Y^v_s - r^v_s(X)}{\mathbb{E}(\tilde{K}_{s,0}(X))}\right|^p\right) \leq C\mathbb{E}\left(\sum_{s\in S}\frac{\mathbb{I}_{\{\tilde{\Psi}(X)\in\mathcal{E}^v_s\}}}{(\phi(h_s))^p}\right) = C\sum_{s\in S}\frac{\mathbb{E}(\mathbb{I}_{\{\tilde{\Psi}(X)\in\mathcal{E}^v_s\}})}{(\phi(h_s))^p}\\ &= C\sum_{s\in S}\frac{\mathbb{P}(\tilde{\Psi}(X)\in\mathcal{E}^v_s)}{(\phi(h_s))^p} \leq C\sum_{s\in S}\frac{1}{(\phi(h_s))^p} \leq C\left(\sum_{s\in S}\frac{1}{\phi(h_s)}\right)^p.\end{aligned}$$

Proof of item (ii). By (3.1), (H.13) and the fact that $\mathcal{E}^v_{s_1} \cap \mathcal{E}^v_{s_2} = \emptyset$ for all $s_1 \neq s_2$, one has

$$\mathbb{E}\left(\left|\sum_{s\in S}\frac{(Y_{i,s}^{v}-r_{s}^{v}(X_i))\tilde{K}_{s,i}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right|^p|X_i\right)=\mathbb{E}\left(|\epsilon_i|^p\left(\sum_{s\in S}\frac{\mathbb{I}_{\{\tilde{\Psi}(X_i)\in\mathcal{E}_s^v\}}K(0)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right)^p|X_i\right)$$
$$\leq C\mathbb{E}(|\epsilon_i|^p|X_i)\left(\sum_{s\in S}\frac{\mathbb{I}_{\{\tilde{\Psi}(X_i)\in\mathcal{E}_s^v\}}}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right)^p\leq C\sum_{s\in S}\frac{1}{(\mathbb{E}(\tilde{K}_{s,0}(X_i)))^p}$$
$$=C\sum_{s\in S}\frac{1}{(\mathbb{E}(\mathbb{E}(\tilde{K}_{s,0}(X_i)|X_i)))^p}.$$

Thus, Lemma 7.5 leads to

$$\mathbb{E}\left(\left|\sum_{s\in S}\frac{(Y_{i,s}^{v}-r_{s}^{v}(X_i))\tilde{K}_{s,i}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right|^p|X_i\right)\leq C\sum_{s\in S}\frac{1}{(\phi(h_s))^p}\leq C\left(\sum_{s\in S}\frac{1}{\phi(h_s)}\right)^p.$$

Proof of item (iii). If $i\neq j$ and $\|X_i-X_j\|\leq h_s$, then (3.1), (3.3), (7.8) and assumption (H.12) lead to

$$\begin{aligned}|Y_{j,s}^{v}-r_s^v(X_i)|&\leq|Y_{j,s}^{v}-r_s^v(X_j)|+|r_s^v(X_j)-r_s^v(X_i)|\\&=|\epsilon_j|\mathbb{I}_{\{\tilde{\Psi}(X_j)\in\mathcal{E}_s^v\}}+|r(X_j)-r(X_i)|\mathbb{I}_{\{\tilde{\Psi}(X_j)\in\mathcal{E}_s^v\}}\leq(|\epsilon_j|+Ch_s^{\beta})\mathbb{I}_{\{\tilde{\Psi}(X_j)\in\mathcal{E}_s^v\}}.\end{aligned}\tag{7.9}$$

Using (7.9), (H.13), (H.16) and the fact that $\mathcal{E}_{s_1}^v\cap\mathcal{E}_{s_2}^v=\emptyset$ for all $s_1\neq s_2$, it can be shown that

$$
\begin{aligned}
&\mathbb{E}\left(\left|\sum_{s\in S}\frac{(Y^v_{j,s}-r^v_s(X_i))\tilde{K}_{s,j}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right|^p|X_i\right)\\
&\leq \mathbb{E}\left((|\epsilon_j|+C)^p\left(\sum_{s\in S}\frac{\mathbb{I}_{\{\tilde{\Psi}(X_j)\in\mathcal{E}^v_s\}}\tilde{K}_{s,j}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right)^p|X_i\right)\\
&\leq \mathbb{E}\left(\mathbb{E}((|\epsilon_j|+C)^p|X_j)\left(\sum_{s\in S}\frac{\mathbb{I}_{\{\tilde{\Psi}(X_j)\in\mathcal{E}^v_s\}}\tilde{K}_{s,j}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right)^p|X_i\right)\\
&\leq C\mathbb{E}\left(\left(\sum_{s\in S}\frac{\mathbb{I}_{\{\tilde{\Psi}(X_j)\in\mathcal{E}^v_s\}}\tilde{K}_{s,j}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right)^p|X_i\right)\leq C\mathbb{E}\left(\sum_{s\in S}\frac{\tilde{K}^p_{s,j}(X_i)}{(\mathbb{E}(\tilde{K}_{s,0}(X_i)))^p}|X_i\right)\\
&= C\sum_{s\in S}\frac{\mathbb{E}(\tilde{K}^p_{s,j}(X_i)|X_i)}{(\mathbb{E}(\tilde{K}_{s,0}(X_i)))^p}=C\sum_{s\in S}\frac{\mathbb{E}(\tilde{K}^p_{s,j}(X_i)|X_i)}{(\mathbb{E}(\mathbb{E}(\tilde{K}_{s,0}(X_i)|X_i)))^p}.
\end{aligned}
$$

Hence, using Lemma 7.5,

$$
\begin{aligned}
\mathbb{E}\left(\left|\sum_{s\in S}\frac{(Y^v_{j,s}-r^v_s(X_i))\tilde{K}_{s,j}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right|^p|X_i\right)&\leq C\sum_{s\in S}\frac{\phi(h_s)}{(\phi(h_s))^p}=C\sum_{s\in S}\frac{1}{(\phi(h_s))^{p-1}}\\
&\leq C\left(\sum_{s\in S}\frac{1}{\phi(h_s)}\right)^{p-1}.
\end{aligned}
$$

Proof of item (iv). Using the reasonings presented in (7.9), (3.3), the fact that

$\mathcal{E}_{s_1}^v \cap \mathcal{E}_{s_2}^v = \emptyset$ for all $s_1 \neq s_2$, and (H.14), then

$$\begin{aligned}
&\mathbb{E}\left(\left(\sum_{s\in S}\frac{(Y_{j,s}^v - r_s^v(X_i))\tilde{K}_{s,j}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right)^2 |X_i\right)\\
&= \mathbb{E}\left(\sum_{s\in S}\frac{(\epsilon_j + (r(X_j) - r(X_i)))^2\mathbb{I}_{\{\tilde{\Psi}(X_i)\in\mathcal{E}_s^v\}}\tilde{K}_{s,j}^2(X_i)}{(\mathbb{E}(\tilde{K}_{s,0}(X_i)))^2}|X_i\right)\\
&= \sum_{s\in S}\frac{\mathbb{E}(\mathbb{E}((\epsilon_j + (r(X_j) - r(X_i)))^2|X_j)\tilde{K}_{s,j}^2(X_i)|X_i)\mathbb{I}_{\{\tilde{\Psi}(X_i)\in\mathcal{E}_s^v\}}}{(\mathbb{E}(\tilde{K}_{s,0}(X_i)))^2}\\
&= \sum_{s\in S}\frac{\mathbb{E}((\mathbb{E}(\epsilon_j^2|X_j) + (r(X_j) - r(X_i))^2)\tilde{K}_{s,j}^2(X_i)|X_i)\mathbb{I}_{\{\tilde{\Psi}(X_i)\in\mathcal{E}_s^v\}}}{(\mathbb{E}(\tilde{K}_{s,0}(X_i)))^2}\\
&\geq C\sum_{s\in S}\frac{\mathbb{E}(\tilde{K}_{s,j}^2(X_i)|X_i)\mathbb{I}_{\{\tilde{\Psi}(X_i)\in\mathcal{E}_s^v\}}}{(\mathbb{E}(\mathbb{E}(\tilde{K}_{s,0}(X_i)|X_i)))^2}.
\end{aligned}$$

Therefore, using Lemma 7.5

$$\mathbb{E}\left(\left(\sum_{s\in S}\frac{(Y_{j,s}^v - r_s^v(X_i))\tilde{K}_{s,j}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right)^2 |X_i\right) \geq C\sum_{s\in S}\frac{\mathbb{I}_{\{\tilde{\Psi}(X_i)\in\mathcal{E}_s^v\}}}{\phi(h_s)}.$$

□

7.4.5 Formulation and Proof of Lemma 7.5

Lemma 7.5. *Under (H.1), (H.10) and (H.15), for all $\gamma > 0$ and for all $i \neq j$, there exist $c_{14,\gamma}, c_{15,\gamma} > 0$ such that*

$$c_{14,\gamma}\phi(h_s) \leq \mathbb{E}(\tilde{K}_{s,j}^\gamma(X_i)|X_i) \leq c_{15,\gamma}\phi(h_s) \quad a.s., \quad \forall s \in S.$$

Proof. Note that (H.15) ensures that there exist $c, c' > 0$ such that

$$c\mathbb{I}_{\{t\in[0,1]\}} \leq K(t) \leq c'\mathbb{I}_{\{t\in[0,1]\}}, \quad \forall t \in [0,1]. \tag{7.10}$$

For all $\gamma > 0$, using (7.10) with $t = h_s^{-1}\|X_j - X_i\|$, one gets

$$c^\gamma \mathbb{I}_{\{\|X_j - X_i\| \leq h_s\}} \leq \tilde{K}_{s,j}^\gamma(X_i) \leq (c')^\gamma \mathbb{I}_{\{\|X_j - X_i\| \leq h_s\}}.$$

Therefore, applying the conditional expectation, one has

$$c^\gamma \mathbb{P}(\|X_j - X_i\| \leq h_s|X_i) \leq \mathbb{E}(\tilde{K}_{s,j}^\gamma(X_i)|X_i) \leq (c')^\gamma \mathbb{P}(\|X_j - X_i\| \leq h_s|X_i),$$

and, consequently,

$$c^\gamma c_1 \phi(h_s) \leq \mathbb{E}(\tilde{K}_{s,j}^\gamma(X_i)|X_i) \leq (c')^\gamma c_2 \phi(h_s),$$

since (H.10) holds. Hence, the proof is finished by taking $c_{14,\gamma} = c^\gamma c_1$ and $c_{15,\gamma} = (c')^\gamma c_2$. □

7.4.6 Formulation and Proof of Lemma 7.6

Recall that C will denote a generic positive constant which may take on different values even in the same formula.

Lemma 7.6. *Under hypotheses of Theorem 5.1,*

$$MISE(\upsilon, \{h_s\}_{s\in S}) \geq c_{16} \sum_{s\in S} \frac{1}{n\phi(h_s)}.$$

Proof. This proof is analogous to the proof of Lemma 1 by Ait-Saïdi et al. (2008). For each $(\upsilon, \{h_s\}_{s\in S}) \in \Upsilon \times H_n^{N\varepsilon}$, one gets

$$\begin{aligned}\mathrm{MISE}(\upsilon, \{h_s\}_{s\in S}) &= \mathbb{E}((r(X) - \hat{r}^\upsilon(X))^2) = \mathbb{E}(\mathbb{E}((r(X) - \hat{r}^\upsilon(X))^2|X)) \\ &= \mathbb{E}((r(X) - \mathbb{E}(\hat{r}^\upsilon(X)|X))^2) + \mathbb{E}(\mathrm{Var}(\hat{r}^\upsilon(X)|X)) \geq \mathbb{E}(\mathrm{Var}(\hat{r}^\upsilon(X)|X)).\end{aligned}$$

Note that hypotheses of Theorem 5.1 ensure that Corollary 4.1 can be applied. Thus, Corollary 4.1, assumptions (H.3), (H.10), (H.14) and (H.15) lead to

$$
\begin{aligned}
\mathbb{E}(\mathrm{Var}(\hat{r}^{\upsilon}(X)|X)) &\geq C \sum_{s\in S} \frac{1}{n\phi(h_s)} \mathbb{E}\left(\mathbb{I}_{\{\tilde{\Psi}(X)\in\mathcal{E}_s^{\upsilon}\}}\right) = C \sum_{s\in S} \frac{1}{n\phi(h_s)} \mathbb{P}(\tilde{\Psi}(X) \in \mathcal{E}_s^{\upsilon}) \\
&\geq C \sum_{s\in S} \frac{1}{n\phi(h_s)},
\end{aligned}
$$

which completes the proof of Lemma 7.6. □

7.4.7 Formulation and Proof of Lemma 7.7

Lemma 7.7. *Under hypotheses of Theorem 5.1,*

$$
\sup_{(\upsilon,\{h_s\}_{s\in S})\in\Upsilon\times H_n^{N_\mathcal{E}}} \left| \frac{ASE(\upsilon,\{h_s\}_{s\in S}) - MISE(\upsilon,\{h_s\}_{s\in S})}{MISE(\upsilon,\{h_s\}_{s\in S})} \right| \to 0 \quad a.s.
$$

where $ASE(\upsilon,\{h_s\}_{s\in S})$ *is defined at* (7.1).

Proof. This lemma is analogous to Lemma 2 by Ait-Saïdi et al. (2008). Recall that it was shown in the proof of Lemma 7.2 that

$$
\mathrm{ASE}(\upsilon,\{h_s\}_{s\in S}) = \mathrm{ASE}^*(\upsilon,\{h_s\}_{s\in S}) + o_{a.co.}(\mathrm{ASE}(\upsilon,\{h_s\}_{s\in S})),
$$

with $\mathrm{ASE}^*(\upsilon,\{h_s\}_{s\in S}) = n^{-1}\sum_{j=1}^{n}\left(\sum_{s\in S}\left(\hat{r}_{s,D}^{\upsilon}(X_j) r_s^{\upsilon}(X_j) - \hat{r}_{s,N}^{\upsilon}(X_j)\right)\right)^2$. Similar calculations and Lemma 7.3 allow to obtain the following expression for MISE

$$\text{MISE}(\upsilon, \{h_s\}_{s\in S}) = \mathbb{E}((r(X) - \hat{r}^{\upsilon}(X))^2) = \mathbb{E}\left(\left(\sum_{s\in S}(r_s^{\upsilon}(X) - \hat{r}_s^{\upsilon}(X))\right)^2\right)$$

$$= \mathbb{E}\left(\left(\sum_{s\in S}(\hat{r}_{s,D}^{\upsilon}(X)(r_s^{\upsilon}(X) - \hat{r}_s^{\upsilon}(X)) + (1 - \hat{r}_{s,D}^{\upsilon}(X))(r_s^{\upsilon}(X) - \hat{r}_s^{\upsilon}(X)))\right)^2\right)$$

$$= \mathbb{E}\left(\left(\sum_{s\in S}(\hat{r}_{s,D}^{\upsilon}(X)r_s^{\upsilon}(X) - \hat{r}_{s,N}^{\upsilon}(X))\right)^2\right) + o_{a.co.}(\text{MISE}(\upsilon, \{h_s\}_{s\in S}))$$

$$= \text{MISE}^*(\upsilon, \{h_s\}_{s\in S}) + o_{a.co.}(\text{MISE}(\upsilon, \{h_s\}_{s\in S})),$$

where $\text{MISE}^*(\upsilon, \{h_s\}_{s\in S}) = \mathbb{E}\left(\left(\sum_{s\in S}(\hat{r}_{s,D}^{\upsilon}(X)r_s^{\upsilon}(X) - \hat{r}_{s,N}^{\upsilon}(X))\right)^2\right)$. Therefore, the lemma can be proven by showing the equivalence between $\text{ASE}^*(\upsilon, \{h_s\}_{s\in S})$ and $\text{MISE}^*(\upsilon, \{h_s\}_{s\in S})$. Specifically, it is enough to show that

$$\sup_{(\upsilon,\{h_s\}_{s\in S})\in\Upsilon\times H_n^{N_{\mathcal{E}}}}\left|\frac{\text{ASE}^*(\upsilon, \{h_s\}_{s\in S}) - \text{MISE}^*(\upsilon, \{h_s\}_{s\in S})}{\text{MISE}^*(\upsilon, \{h_s\}_{s\in S})}\right| \to 0 \quad a.s. \tag{7.11}$$

For that, first of all, assume that $s_0 \in S$ is selected, and $\upsilon \in \Upsilon$ and $h_s \in H_n$ for all $s \in S$ with $s \neq s_0$ are fixed. In addition, consider $\lambda \in \Lambda = \{1/\phi(h_{s_0}) + \sum_{s\neq s_0} 1/\phi(h_s) : h_{s_0} \in H_n\}$. Then, a delta sequence estimator $\hat{g}_\lambda : \mathcal{H} \to \mathbb{R}$ can be defined as follows

$$\hat{g}_\lambda(x) = \frac{1}{n}\sum_{i=1}^{n}\delta_\lambda(x, X_i, Y_i),$$

where

$$\delta_\lambda(x, X_i, Y_i) = \frac{(r_{s_0}^{\upsilon}(x) - Y_{i,s_0}^{\upsilon})K(f(\lambda)^{-1}\|X_i - x\|)}{\mathbb{E}(K(f(\lambda)^{-1}\|X_0 - x\|))} + \sum_{s\neq s_0}\frac{(r_s^{\upsilon}(x) - Y_{i,s}^{\upsilon})\tilde{K}_{s,i}(x)}{\mathbb{E}(\tilde{K}_{s,0}(x))},$$

with $f(\lambda) = \phi^{-1}((\lambda - (\sum_{s\neq s_0} 1/\phi(h_s)))^{-1})$, being ϕ^{-1} the inverse function of ϕ (note that (H.10) ensures that ϕ is a bijective function, so there is a unique inverse function ϕ^{-1} which is also a bijection). Given that $f(\lambda) = h_{s_0}$, in fact, $\delta_\lambda(x, X_i, Y_i)$ is

$$\delta_\lambda(x, X_i, Y_i) = \sum_{s\in S} \frac{(r_s^\upsilon(x) - Y_{i,s}^\upsilon)\tilde{K}_{s,i}(x)}{\mathbb{E}(\tilde{K}_{s,0}(x))}. \tag{7.12}$$

In this situation, $\hat{g}_\lambda(x) = \sum_{s\in S} (\hat{r}_{s,D}^\upsilon(x) r_s^\upsilon(x) - \hat{r}_{s,N}^\upsilon(x))$, and it may be considered that $\hat{g}_\lambda$ estimates the operator $g : \mathcal{H} \to \mathbb{R}$ defined as $g(x) = 0$ for all $x \in \mathcal{H}$. Besides, computing the mean integrated squared error and the average squared error for $\hat{g}_\lambda$ (denoted by $\mathrm{MISE}_{\hat{g}}(\lambda)$ and $\mathrm{ASE}_{\hat{g}}(\lambda)$, respectively), one has

$$\mathrm{MISE}_{\hat{g}}(\lambda) = \mathbb{E}((g(X) - \hat{g}_\lambda(X))^2) = \mathbb{E}((\hat{g}_\lambda(X))^2) = \mathrm{MISE}^*(\upsilon, \{h_s\}_{s\in S}), \tag{7.13}$$

and

$$\mathrm{ASE}_{\hat{g}}(\lambda) = \frac{1}{n}\sum_{j=1}^{n} (g(X_j) - \hat{g}_\lambda(X_j))^2 = \frac{1}{n}\sum_{j=1}^{n} (\hat{g}_\lambda(X_j))^2 = \mathrm{ASE}^*(\upsilon, \{h_s\}_{s\in S}). \tag{7.14}$$

On the other hand, (H.11), (H.18) and Lemma 7.8 indicate that the assumptions of the theoretical results for delta sequence estimators in Marron and Härdle (1986) hold. Thus, applying Theorem 2 by Marron and Härdle (1986), one gets

$$\sup_{\lambda\in\Lambda} \left| \frac{\mathrm{ASE}_{\hat{g}}(\lambda) - \mathrm{MISE}_{\hat{g}}(\lambda)}{\mathrm{MISE}_{\hat{g}}(\lambda)} \right| \to 0 \quad a.s. \tag{7.15}$$

Taking into account (7.13) and (7.14), it can be seen

$$\left| \frac{\mathrm{ASE}^*(\upsilon, \{h_s\}_{s\in S}) - \mathrm{MISE}^*(\upsilon, \{h_s\}_{s\in S})}{\mathrm{MISE}^*(\upsilon, \{h_s\}_{s\in S})} \right| \leq \sup_{\lambda\in\Lambda} \left| \frac{\mathrm{ASE}_{\hat{g}}(\lambda) - \mathrm{MISE}_{\hat{g}}(\lambda)}{\mathrm{MISE}_{\hat{g}}(\lambda)} \right|.$$

Hence, (7.11) is verified due to (7.15), and consequently the proof of the lemma is complete. ☐

7.4.8 Formulation and Proof of Lemma 7.8

Recall that C will denote a generic positive constant which may take on different values even in the same formula.

Lemma 7.8. *Under hypotheses of Theorem 5.1,*

(i) $\forall p = 1, 2, \ldots$, $\forall q = 2, \ldots, 2p$,

$$\left|\mathbb{E}\left(\prod_{i=1}^{q}\prod_{j=1}^{q}\left(\delta_\lambda(X_i, X_j, Y_j)\right)^{a_{ij}}\right)\right| \leq c_{17,p}\left(\sum_{s\in S}\frac{1}{\phi(h_s)}\right)^{p-q/2},$$

where $a_{ij} \in \{0, \ldots, p\}$, $\sum_{i=1}^{q}\sum_{j=1}^{q} a_{ij} = p$, *and, for each* $i \in \{1, \ldots, q\}$, *there exists* $j \neq i$ *such that either* $a_{ij} \neq 0$ *or* $a_{ji} \neq 0$,

(ii) $\left|\mathbb{E}\left(\left(\mathbb{E}\left(\delta_\lambda(X_3, X_1, Y_1)\delta_\lambda(X_3, X_2, Y_2)|X_1, X_2\right)\right)^2\right)\right| \leq c_{18}\sum_{s\in S}\frac{1}{\phi(h_s)}$,

(iii) $\left|\mathbb{E}\left(\delta_\lambda(X_3, X_1, Y_1)\delta_\lambda(X_3, X_2, Y_2)\right)\right| \leq c_{19}$,

(iv) $\mathbb{E}\left(\left(\delta_\lambda(X_1, X_2, Y_2)\right)^2\right) \geq c_{20}\sum_{s\in S}\frac{1}{\phi(h_s)}$,

(v) $\forall p = 1, 2, \ldots$, $\mathbb{E}\left(\left(\mathbb{E}\left(\delta_\lambda(X_1, X_2, Y_2)|X_1\right)\right)^{2p}\right) \leq c_{21,p}$,

(vi) $\forall p = 1, 2, \ldots$, $\mathbb{E}\left(\left(\delta_\lambda(X_1, X_1, Y_1)\right)^{2p}\right) \leq c_{22,p}\left(\sum_{s\in S}\frac{1}{\phi(h_s)}\right)^{2p}$,

where the operator δ_λ *is defined at* (7.12).

Proof. This lemma is analogous to Lemma 6 by Ait-Saïdi et al. (2008).

Proof of item (i). By Jensen's inequality, (7.9), the assumption (H.13) and the fact that $\mathcal{E}_{s_1}^v \cap \mathcal{E}_{s_2}^v = \emptyset$ for all $s_1 \neq s_2$, it can be seen that

$$\begin{aligned}
&\left|\mathbb{E}\left(\prod_{i=1}^{q}\prod_{j=1}^{q}(\delta_\lambda(X_i,X_j,Y_j))^{a_{ij}}\right)\right| \leq \mathbb{E}\left(\prod_{i=1}^{q}\prod_{j=1}^{q}|\delta_\lambda(X_i,X_j,Y_j)|^{a_{ij}}\right)\\
&\leq \mathbb{E}\left(\prod_{i=1}^{q}\prod_{j=1}^{q}\left(\sum_{s\in S}\frac{|r_s^v(X_i)-Y_{j,s}^v|\tilde{K}_{s,j}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right)^{a_{ij}}\right)\\
&\leq \mathbb{E}\left(\prod_{i=1}^{q}\prod_{j=1}^{q}\left(\sum_{s\in S}\frac{(|\epsilon_j|+C)\mathbb{I}_{\{\tilde{\Psi}(X_j)\in\mathcal{E}_s^v\}}\tilde{K}_{s,j}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right)^{a_{ij}}\right)\\
&\leq C\mathbb{E}\left(\prod_{i=1}^{q}\prod_{j=1}^{q}\left(\sum_{s\in S}\frac{\mathbb{I}_{\{\tilde{\Psi}(X_j)\in\mathcal{E}_s^v\}}\tilde{K}_{s,j}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right)^{a_{ij}}\right) \qquad (7.16)\\
&\leq C\mathbb{E}\left(\prod_{i=1}^{q}\prod_{j=1}^{q}\left(\sum_{s\in S}\frac{\mathbb{I}_{\{\tilde{\Psi}(X_j)\in\mathcal{E}_s^v\}}\tilde{K}_{s,j}^{a_{ij}}(X_i)}{(\mathbb{E}(\tilde{K}_{s,0}(X_i)))^{a_{ij}}}\right)\right)\\
&\leq C\mathbb{E}\left(\sum_{s\in S}\left(\prod_{i=1}^{q}\prod_{j=1}^{q}\frac{\mathbb{I}_{\{\tilde{\Psi}(X_j)\in\mathcal{E}_s^v\}}\tilde{K}_{s,j}^{a_{ij}}(X_i)}{(\mathbb{E}(\tilde{K}_{s,0}(X_i)))^{a_{ij}}}\right)\right)\\
&\leq C\sum_{s\in S}\mathbb{E}\left(\prod_{i=1}^{q}\prod_{j=1}^{q}\frac{\tilde{K}_{s,j}^{a_{ij}}(X_i)}{(\mathbb{E}(\mathbb{E}(\tilde{K}_{s,0}(X_i)|X_i)))^{a_{ij}}}\right).
\end{aligned}$$

On the other hand, Lemma 7.5 guarantees that

$$\mathbb{E}\left(\prod_{i=1}^{q}\prod_{j=1}^{q}\frac{\tilde{K}_{s,j}^{a_{ij}}(X_i)}{(\mathbb{E}(\tilde{K}_{s,0}(X_i)))^{a_{ij}}}\right) \leq C\frac{1}{(\phi(h_s))^p}\mathbb{E}\left(\prod_{i=1}^{q}\prod_{j=1}^{q}\tilde{K}_{s,j}^{a_{ij}}(X_i)\right).$$

Besides, the restrictions on the definition of the pairs (i,j) and a_{ij} imply that there are $q/2$ separated pairs (i,j) with $a_{ij} \neq 0$. This fact and Lemma 7.5 allow to deduce

$$\mathbb{E}\left(\prod_{i=1}^{q}\prod_{j=1}^{q}\frac{\tilde{K}_{s,j}^{a_{ij}}(X_i)}{(\mathbb{E}(\tilde{K}_{s,0}(X_i)))^{a_{ij}}}\right) \leq C\frac{1}{(\phi(h_s))^p}(\phi(h_s))^{q/2} = C\frac{1}{(\phi(h_s))^{p-q/2}}. \qquad (7.17)$$

Consequently, by (7.16) and (7.17), one has,

$$\left|\mathbb{E}\left(\prod_{i=1}^{q}\prod_{j=1}^{q}\left(\delta_\lambda(X_i,X_j,Y_j)\right)^{a_{ij}}\right)\right| \le C\sum_{s\in S}\frac{1}{(\phi(h_s))^{p-q/2}} \le C\left(\sum_{s\in S}\frac{1}{\phi(h_s)}\right)^{p-q/2}.$$

Proof of item (ii). It can be shown that

$$\begin{aligned}
&\left|\mathbb{E}\left(\left(\mathbb{E}\left(\delta_\lambda(X_3,X_1,Y_1)\delta_\lambda(X_3,X_2,Y_2)|X_1,X_2\right)\right)^2\right)\right| \\
&\le \mathbb{E}\left(\mathbb{E}\left(|\delta_\lambda(X_3,X_1,Y_1)\delta_\lambda(X_3,X_2,Y_2)\delta_\lambda(X_4,X_1,Y_1)\delta_\lambda(X_4,X_2,Y_2)||X_1,X_2\right)\right) \\
&= \mathbb{E}\left(|\delta_\lambda(X_3,X_1,Y_1)\delta_\lambda(X_3,X_2,Y_2)\delta_\lambda(X_4,X_1,Y_1)\delta_\lambda(X_4,X_2,Y_2)|\right).
\end{aligned}$$

Hence, due to (7.9), the fact that $\mathcal{E}^v_{s_1}\cap\mathcal{E}^v_{s_2}=\emptyset$ for all $s_1\neq s_2$, and (H.13), it can be found that

$$\begin{aligned}
&\left|\mathbb{E}\left(\left(\mathbb{E}\left(\delta_\lambda(X_3,X_1,Y_1)\delta_\lambda(X_3,X_2,Y_2)|X_1,X_2\right)\right)^2\right)\right| \\
&\le \mathbb{E}\Bigg(\sum_{s_1\in S}\sum_{s_2\in S}\sum_{s_3\in S}\sum_{s_4\in S}\Bigg(|r^v_{s_1}(X_3)-Y^v_{1,s_1}||r^v_{s_2}(X_3)-Y^v_{2,s_2}||r^v_{s_3}(X_4)-Y^v_{1,s_3}| \\
&\qquad\cdot|r^v_{s_4}(X_4)-Y^v_{2,s_4}|\frac{\tilde{K}_{s_1,1}(X_3)\tilde{K}_{s_2,2}(X_3)\tilde{K}_{s_3,1}(X_4)\tilde{K}_{s_4,2}(X_4)}{\mathbb{E}(\tilde{K}_{s_1,0}(X_3))\mathbb{E}(\tilde{K}_{s_2,0}(X_3))\mathbb{E}(\tilde{K}_{s_3,0}(X_4))\mathbb{E}(\tilde{K}_{s_4,0}(X_4))}\Bigg)\Bigg) \\
&\le \mathbb{E}\Bigg(\sum_{s_1\in S}\sum_{s_2\in S}\sum_{s_3\in S}\sum_{s_4\in S}\Bigg((|\epsilon_1|+C)^2(|\epsilon_2|+C)^2\mathbb{I}_{\{\tilde{\Psi}(X_3)\in\mathcal{E}^v_{s_1}\}}\mathbb{I}_{\{\tilde{\Psi}(X_3)\in\mathcal{E}^v_{s_2}\}}\mathbb{I}_{\{\tilde{\Psi}(X_4)\in\mathcal{E}^v_{s_3}\}} \\
&\qquad\cdot\mathbb{I}_{\{\tilde{\Psi}(X_4)\in\mathcal{E}^v_{s_4}\}}\frac{\tilde{K}_{s_1,1}(X_3)\tilde{K}_{s_2,2}(X_3)\tilde{K}_{s_3,1}(X_4)\tilde{K}_{s_4,2}(X_4)}{\mathbb{E}(\tilde{K}_{s_1,0}(X_3))\mathbb{E}(\tilde{K}_{s_2,0}(X_3))\mathbb{E}(\tilde{K}_{s_3,0}(X_4))\mathbb{E}(\tilde{K}_{s_4,0}(X_4))}\Bigg)\Bigg) \\
&\le \mathbb{E}\Bigg(\mathbb{E}((|\epsilon_1|+C)^2(|\epsilon_2|+C)^2|X_1,X_2,X_3,X_4)\sum_{s_1\in S}\sum_{s_3\in S}\Bigg(\mathbb{I}_{\{\tilde{\Psi}(X_3)\in\mathcal{E}^v_{s_1}\}}\mathbb{I}_{\{\tilde{\Psi}(X_4)\in\mathcal{E}^v_{s_3}\}} \\
&\qquad\cdot\frac{\tilde{K}_{s_1,1}(X_3)\tilde{K}_{s_1,2}(X_3)\tilde{K}_{s_3,1}(X_4)\tilde{K}_{s_3,2}(X_4)}{(\mathbb{E}(\tilde{K}_{s_1,0}(X_3)))^2(\mathbb{E}(\tilde{K}_{s_3,0}(X_4)))^2}\Bigg)\Bigg) \\
&\le C\mathbb{E}\left(\sum_{s_1\in S}\sum_{s_3\in S}\left(\mathbb{I}_{\{\tilde{\Psi}(X_3)\in\mathcal{E}^v_{s_1}\}}\mathbb{I}_{\{\tilde{\Psi}(X_4)\in\mathcal{E}^v_{s_3}\}}\frac{\tilde{K}_{s_1,1}(X_3)\tilde{K}_{s_1,2}(X_3)\tilde{K}_{s_3,1}(X_4)\tilde{K}_{s_3,2}(X_4)}{(\mathbb{E}(\tilde{K}_{s_1,0}(X_3)))^2(\mathbb{E}(\tilde{K}_{s_3,0}(X_4)))^2}\right)\right).
\end{aligned}$$

Given that $\tilde{K}_{s_3,2}(X_4) \le C$ for all $s \in S$ (since (H.15) holds), $\tilde{K}_{s,i}(X_j) = \tilde{K}_{s,j}(X_i)$, the indicator functions are bounded, and Lemma 7.5 can be applied, one gets

$$\begin{aligned}
&\left|\mathbb{E}\left((\mathbb{E}\left(\delta_\lambda(X_3,X_1,Y_1)\delta_\lambda(X_3,X_2,Y_2)|X_1,X_2\right))^2\right)\right| \\
&\le C\sum_{s_1\in S}\sum_{s_3\in S}\frac{\mathbb{E}(\tilde{K}_{s_1,1}(X_3)\tilde{K}_{s_1,2}(X_3)\tilde{K}_{s_3,1}(X_4))}{(\mathbb{E}(\mathbb{E}(\tilde{K}_{s_1,0}(X_3)|X_3)))^2(\mathbb{E}(\mathbb{E}(\tilde{K}_{s_3,0}(X_4)|X_4)))^2} \\
&\le C\sum_{s_1\in S}\sum_{s_3\in S}\frac{\mathbb{E}(\tilde{K}_{s_1,1}(X_3)\mathbb{E}(\tilde{K}_{s_1,2}(X_3)|X_1,X_3,X_4)\tilde{K}_{s_3,1}(X_4))}{(\phi(h_{s_1}))^2(\phi(h_{s_3}))^2} \\
&\le C\sum_{s_1\in S}\sum_{s_3\in S}\frac{\mathbb{E}(\tilde{K}_{s_1,1}(X_3)\mathbb{E}(\tilde{K}_{s_3,1}(X_4)|X_1,X_3))}{\phi(h_{s_1})(\phi(h_{s_3}))^2} \le C\sum_{s_1\in S}\sum_{s_3\in S}\frac{\mathbb{E}(\tilde{K}_{s_1,1}(X_3))}{\phi(h_{s_1})\phi(h_{s_3})} \\
&\le C\sum_{s\in S}\frac{1}{\phi(h_s)}.
\end{aligned}$$

Proof of item (iii). Regarding to this item, Jensen's inequality, (3.3), (7.9), the fact that $\mathcal{E}^v_{s_1}\cap\mathcal{E}^v_{s_2}=\emptyset$ for all $s_1\neq s_2$, and (H.13) imply

$$\begin{aligned}
&|\mathbb{E}\left(\delta_\lambda(X_3,X_1,Y_1)\delta_\lambda(X_3,X_2,Y_2)\right)| \le \mathbb{E}\left(|\delta_\lambda(X_3,X_1,Y_1)\delta_\lambda(X_3,X_2,Y_2)|\right) \\
&\le \mathbb{E}\left(\sum_{s_1\in S}\sum_{s_2\in S}\frac{|r^v_{s_1}(X_3)-Y^v_{1,s_1}||r^v_{s_2}(X_3)-Y^v_{2,s_2}|\tilde{K}_{s_1,1}(X_3)\tilde{K}_{s_2,2}(X_3)}{\mathbb{E}(\tilde{K}_{s_1,0}(X_3))\mathbb{E}(\tilde{K}_{s_2,0}(X_3))}\right) \\
&\le \mathbb{E}\left(\sum_{s\in S}\frac{(|\epsilon_1|+C)(|\epsilon_2|+C)\mathbb{I}_{\{\tilde{\Psi}(X_3)\in\mathcal{E}^v_s\}}\tilde{K}_{s,1}(X_3)\tilde{K}_{s,2}(X_3)}{(\mathbb{E}(\tilde{K}_{s,0}(X_3)))^2}\right) \\
&=\sum_{s\in S}\frac{\mathbb{E}(\mathbb{E}((|\epsilon_1|+C)(|\epsilon_2|+C)|X_1,X_2,X_3)\mathbb{I}_{\{\tilde{\Psi}(X_3)\in\mathcal{E}^v_s\}}\tilde{K}_{s,1}(X_3)\tilde{K}_{s,2}(X_3))}{(\mathbb{E}(\tilde{K}_{s,0}(X_3)))^2} \\
&\le C\sum_{s\in S}\frac{\mathbb{E}(\mathbb{I}_{\{\tilde{\Psi}(X_3)\in\mathcal{E}^v_s\}}\tilde{K}_{s,1}(X_3)\tilde{K}_{s,2}(X_3))}{(\mathbb{E}(\tilde{K}_{s,0}(X_3)))^2} \\
&= C\sum_{s\in S}\frac{\mathbb{E}(\mathbb{I}_{\{\tilde{\Psi}(X_3)\in\mathcal{E}^v_s\}}\mathbb{E}(\tilde{K}_{s,1}(X_3)|X_3)\mathbb{E}(\tilde{K}_{s,2}(X_3)|X_3))}{(\mathbb{E}(\mathbb{E}(\tilde{K}_{s,0}(X_3)|X_3)))^2}.
\end{aligned}$$

Hence, the application of Lemma 7.5 leads to

$$|\mathbb{E}\left(\delta_\lambda(X_3, X_1, Y_1)\delta_\lambda(X_3, X_2, Y_2)\right)| \leq C \sum_{s\in S} \mathbb{P}(\tilde{\Psi}(X_3) \in \mathcal{E}_s^v) \leq C N_{\mathcal{E}} \leq C.$$

Proof of item (iv). This item comes from Lemma 7.4 and (H.3) as follows

$$\begin{aligned}
\mathbb{E}\left((\delta_\lambda(X_1, X_2, Y_2))^2\right) &= \mathbb{E}\left(\left(\sum_{s\in S} \frac{(r_s^v(X_1) - Y_{2,s}^v)\tilde{K}_{s,2}(X_1)}{\mathbb{E}(\tilde{K}_{s,0}(X_1))}\right)^2\right) \\
&= \mathbb{E}\left(\mathbb{E}\left(\left(\sum_{s\in S} \frac{(r_s^v(X_1) - Y_{2,s}^v)\tilde{K}_{s,2}(X_1)}{\mathbb{E}(\tilde{K}_{s,0}(X_1))}\right)^2 |X_1\right)\right) \\
&\geq C \sum_{s\in S} \frac{\mathbb{P}(\tilde{\Psi}(X_1) \in \mathcal{E}_s^v)}{\phi(h_s)} \geq C \sum_{s\in S} \frac{1}{\phi(h_s)}.
\end{aligned}$$

Proof of item (v). By Jensen's inequality and Lemma 7.4,

$$\begin{aligned}
|\mathbb{E}\left(\delta_\lambda(X_1, X_2, Y_2)|X_1\right)| &\leq \mathbb{E}\left(|\delta_\lambda(X_1, X_2, Y_2)|\,|X_1\right) \\
&= \mathbb{E}\left(\left|\sum_{s\in S} \frac{(r_s^v(X_1) - Y_{2,s}^v)\tilde{K}_{s,2}(X_1)}{\mathbb{E}(\tilde{K}_{s,0}(X_1))}\right| |X_1\right) \leq C.
\end{aligned}$$

Thus, $\mathbb{E}\left((\mathbb{E}\left(\delta_\lambda(X_1, X_2, Y_2)|X_1\right))^{2p}\right) \leq C$.

Proof of item (vi). This item is a direct consequence of Lemma 7.4 given that

$$\begin{aligned}
\mathbb{E}\left((\delta_\lambda(X_1, X_1, Y_1))^{2p}\right) &= \mathbb{E}\left(\left(\sum_{s\in S} \frac{(r_s^v(X_1) - Y_{1,s}^v)K(0)}{\mathbb{E}(\tilde{K}_{s,0}(X_1))}\right)^{2p}\right) \\
&\leq C\mathbb{E}\left(\left(\sum_{s\in S} \frac{r_s^v(X_1) - Y_{1,s}^v}{\mathbb{E}(\tilde{K}_{s,0}(X_1))}\right)^{2p}\right) \leq C\left(\sum_{s\in S} \frac{1}{\phi(h_s)}\right)^{2p}.
\end{aligned}$$

□

7.4.9 Formulation and Proof of Lemma 7.9

Recall that C will denote a generic positive constant which may take on different values even in the same formula.

Lemma 7.9. *Under hypotheses of Theorem 5.1,*

$$\sup_{(\upsilon,\{h_s\}_{s\in S})\in\Upsilon\times H_n^{N\mathcal{E}}} \left| \frac{CT(\upsilon,\{h_s\}_{s\in S})}{MISE(\upsilon,\{h_s\}_{s\in S})} \right| \to 0 \quad a.s.$$

where $CT(\upsilon,\{h_s\}_{s\in S})$ *is defined at* (7.3)*.*

Proof. The proof of this lemma can be obtained following the proof of Lemma 4 in Ait-Saïdi et al. (2008), which in turn is based on the ideas proposed by Härdle and Marron (1985), as follows. By the second statement in Lemma 7.3, one can see that

$$\begin{aligned}
&\mathrm{CT}(\upsilon,\{h_s\}_{s\in S}) = \frac{1}{n}\sum_{j=1}^{n}\epsilon_j(\hat{r}^{\upsilon,(-j)}(X_j) - r(X_j)) \\
&= \frac{1}{n}\sum_{j=1}^{n}\epsilon_j\left(\sum_{s\in S}(\hat{r}_s^{\upsilon,(-j)}(X_j) - r_s^{\upsilon}(X_j))\right) \\
&= \frac{1}{n}\sum_{j=1}^{n}\epsilon_j\left(\sum_{s\in S}(\hat{r}_{s,D}^{\upsilon,(-j)}(X_j)(\hat{r}_s^{\upsilon,(-j)}(X_j) - r_s^{\upsilon}(X_j)))\right) \\
&\quad + \frac{1}{n}\sum_{j=1}^{n}\epsilon_j\left(\sum_{s\in S}((1-\hat{r}_{s,D}^{\upsilon,(-j)}(X_j))(\hat{r}_s^{\upsilon,(-j)}(X_j) - r_s^{\upsilon}(X_j)))\right) \\
&= \frac{1}{n}\sum_{j=1}^{n}\epsilon_j\left(\sum_{s\in S}(\hat{r}_{s,N}^{\upsilon,(-j)}(X_j) - \hat{r}_{s,D}^{\upsilon,(-j)}(X_j)r_s^{\upsilon}(X_j))\right) + o_{a.co}(\mathrm{CT}(\upsilon,\{h_s\}_{s\in S})) \\
&= \mathrm{CT}^*(\upsilon,\{h_s\}_{s\in S}) + o_{a.co}(\mathrm{CT}(\upsilon,\{h_s\}_{s\in S})),
\end{aligned}$$

with $\mathrm{CT}^*(\upsilon,\{h_s\}_{s\in S}) = n^{-1}\sum_{j=1}^{n}\epsilon_j\left(\sum_{s\in S}(\hat{r}_{s,N}^{\upsilon,(-j)}(X_j) - \hat{r}_{s,D}^{\upsilon,(-j)}(X_j)r_s^{\upsilon}(X_j))\right)$. This fact and Lemma 7.6 allow to deduce that it is enough to show

$$\sup_{(v,\{h_s\}_{s\in S})\in\Upsilon\times H_n^{N_{\mathcal{E}}}} \left| \left(\sum_{s\in S} \frac{1}{n\phi(h_s)} \right)^{-1} \mathrm{CT}^*(v, \{h_s\}_{s\in S}) \right| \to 0 \quad a.s.$$

in order to prove the lemma. In addition, note that if $i \neq j$ and $\|X_i - X_j\| \leq h_s$, (3.1) and (3.3) imply that

$$(Y_{i,s}^v - r_s^v(X_j)) = (Y_{i,s}^v - r_s^v(X_i)) + (r_s^v(X_i) - r_s^v(X_j)) = (\epsilon_i + (r(X_i) - r(X_j)))\, \mathbb{I}_{\{\tilde{\Psi}(X_j)\in\mathcal{E}_s^v\}}.$$

Therefore,

$$\begin{aligned}
|\mathrm{CT}^*(v, \{h_s\}_{s\in S})| &= \left| \frac{1}{n} \sum_{j=1}^n \epsilon_j \left(\sum_{s\in S} (\hat{r}_{s,N}^{v,(-j)}(X_j) - \hat{r}_{s,D}^{v,(-j)}(X_j) r_s^v(X_j)) \right) \right| \\
&= \left| \sum_{s\in S} \left(\frac{1}{n(n-1)} \sum_{j=1}^n \sum_{i\neq j} \frac{\epsilon_j (Y_{i,s}^v - r_s^v(X_j)) \tilde{K}_{s,i}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))} \right) \right| \\
&\leq \sum_{s\in S} \left(\frac{1}{n(n-1)} \sum_{j=1}^n \sum_{i\neq j} \frac{|\epsilon_j||\epsilon_i| \mathbb{I}_{\{\tilde{\Psi}(X_i)\in\mathcal{E}_s^v\}} \tilde{K}_{s,i}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))} \right) \\
&\quad + \sum_{s\in S} \left(\frac{1}{n(n-1)} \sum_{j=1}^n \sum_{i\neq j} \frac{|\epsilon_j||r(X_i) - r(X_j)| \mathbb{I}_{\{\tilde{\Psi}(X_j)\in\mathcal{E}_s^v\}} \tilde{K}_{s,i}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))} \right) \\
&\leq \sum_{s\in S} \left(\frac{1}{n(n-1)} \sum_{j=1}^n \sum_{i\neq j} U_{i,j,s}^v + \frac{1}{n(n-1)} \sum_{j=1}^n \sum_{i\neq j} V_{i,j,s}^v \right),
\end{aligned}$$

where

$$U_{i,j,s}^v = \frac{|\epsilon_j||\epsilon_i| \tilde{K}_{s,i}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))} \quad \text{and} \quad V_{i,j,s}^v = \frac{|\epsilon_j||r(X_i) - r(X_j)| \tilde{K}_{s,i}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j)),}$$

for $i, j \in \{1, \ldots, n\}$ such that $i \neq j$. Furthermore, note that

$$\left|\left(\sum_{s\in S}\frac{1}{n\phi(h_s)}\right)^{-1}\mathrm{CT}^*(\upsilon,\{h_s\}_{s\in S})\right| \leq \left|\left(\frac{1}{n\phi(h_s)}\right)^{-1}\mathrm{CT}^*(\upsilon,\{h_s\}_{s\in S})\right|$$
$$\leq \sum_{s\in S}\left(\left|\left(\frac{1}{n\phi(h_s)}\right)^{-1}\frac{1}{n(n-1)}\sum_{j=1}^{n}\sum_{i\neq j}U_{i,j,s}^{\upsilon}\right|\right.$$
$$\left.+\left|\left(\frac{1}{n\phi(h_s)}\right)^{-1}\frac{1}{n(n-1)}\sum_{j=1}^{n}\sum_{i\neq j}V_{i,j,s}^{\upsilon}\right|\right).$$

Consequently, the lemma will be established as soon as, for each $s \in S$, one states that

$$\sup_{(\upsilon,h_s)\in\Upsilon\times H_n}\left|\left(\frac{1}{n\phi(h_s)}\right)^{-1}\frac{1}{n(n-1)}\sum_{j=1}^{n}\sum_{i\neq j}U_{i,j,s}^{\upsilon}\right| \to 0 \quad a.s. \tag{7.18}$$

and

$$\sup_{(\upsilon,h_s)\in\Upsilon\times H_n^{N\mathcal{E}}}\left|\left(\frac{1}{n\phi(h_s)}\right)^{-1}\frac{1}{n(n-1)}\sum_{j=1}^{n}\sum_{i\neq j}V_{i,j,s}^{\upsilon}\right| \to 0 \quad a.s. \tag{7.19}$$

In order to prove (7.18), note that by (H.18) and Chebyshev's inequality, given $\eta > 0$ and for all $p = 1, 2, \ldots$, one has

$$\mathbb{P}\left(\sup_{(\upsilon,h_s)\in\Upsilon\times H_n}\left|\left(\frac{1}{n\phi(h_s)}\right)^{-1}\frac{1}{n(n-1)}\sum_{j=1}^{n}\sum_{i\neq j}U_{i,j,s}^{\upsilon}\right| > \eta\right)$$
$$\leq \eta^{-2p} card(\Upsilon\times H_n)\sup_{(\upsilon,h_s)\in\Upsilon\times H_n}\mathbb{E}\left(\left(\left(\frac{1}{n\phi(h_s)}\right)^{-1}\frac{1}{n(n-1)}\sum_{j=1}^{n}\sum_{i\neq j}U_{i,j,s}^{\upsilon}\right)^{2p}\right)$$
$$\leq \eta^{-2p}n^{\alpha}\sup_{(\upsilon,h_s)\in\Upsilon\times H_n}\left\{\frac{\phi(h_s)^{2p}}{(n-1)^{2p}}\mathbb{E}\left(\left(\sum_{j=1}^{n}\sum_{i\neq j}U_{i,j,s}^{\upsilon}\right)^{2p}\right)\right\}.$$

Hence, it is enough to show that, for p large enough,

$$\sum_{n=1}^{\infty}n^{\alpha}\sup_{(\upsilon,h_s)\in\Upsilon\times H_n}\left\{\frac{\phi(h_s)^{2p}}{(n-1)^{2p}}\mathbb{E}\left(\left(\sum_{j=1}^{n}\sum_{i\neq j}U_{i,j,s}^{\upsilon}\right)^{2p}\right)\right\} < \infty, \tag{7.20}$$

to prove (7.18) due to Borel–Cantelli Lemma. Analogously, it can be found that (7.19) can be verified by showing that

$$\sum_{n=1}^{\infty} n^{\alpha} \sup_{(v,h_s)\in\Upsilon\times H_n} \left\{ \frac{\phi(h_s)^{2p}}{(n-1)^{2p}} \mathbb{E}\left(\left(\sum_{j=1}^{n} \sum_{i\neq j} V_{i,j,s}^{v} \right)^{2p} \right) \right\} < \infty. \tag{7.21}$$

To obtain (7.20), note that using Lemma 7.5 it can be seen that

$$\begin{aligned} \mathbb{E}\left(\left(\sum_{j=1}^{n} \sum_{i\neq j} U_{i,j,s}^{v} \right)^{2p} \right) &= \sum_{I_{2p}} \mathbb{E}\left(\prod_{l=1}^{2p} U_{i_l,j_l,s}^{v} \right) \\ &\leq C\phi(h_s)^{-2p} \sum_{q=2}^{4p} \sum_{J_q} \mathbb{E}\left(\prod_{l=1}^{q} |\epsilon_{r_l}|^{a_l} \tilde{K}_{s,u_l}^{b_l}(X_{w_l}) \right), \end{aligned}$$

where $I_{2p} = \{(i_1, \ldots, i_{2p}, j_1, \ldots, j_{2p}) \in \{1, \ldots, n\}^{2p},$ such that $i_1 \neq j_1, \ldots, i_{2p} \neq j_{2p}\}$, $J_q \subset I_{2p}$ is the subset which contains the elements of I_{2p} with only q different integers, and $\sum_{l=1}^{q} a_l = 4p$ ($a_l \geq 1$) and $\sum_{l=1}^{q} b_l = 2p$. It can be shown that

$$\begin{aligned} \mathbb{E}\left(\prod_{l=1}^{q} |\epsilon_{r_l}|^{a_l} \tilde{K}_{s,u_l}^{b_l}(X_{w_l}) \right) &= \mathbb{E}\left(\mathbb{E}\left(\prod_{l=1}^{q} |\epsilon_{r_l}|^{a_l} \tilde{K}_{s,u_l}^{b_l}(X_{w_l}) | X_{r_1}, \ldots, X_{r_q} \right) \right) \\ &= \mathbb{E}\left(\prod_{l=1}^{q} \mathbb{E}(|\epsilon_{r_l}|^{a_l} | X_{r_l}) \prod_{l=1}^{q} \tilde{K}_{s,u_l}^{b_l}(X_{w_l}) \right). \end{aligned}$$

This last quantity vanishes when $q > 2p$. Using this fact and Lemma 7.5, and taking into account that there are $q/2$ separated pairs with q different integers, one can obtain

$$\mathbb{E}\left(\left(\sum_{j=1}^{n} \sum_{i\neq j} U_{i,j,s}^{v} \right)^{2p} \right) \leq C\phi(h_s)^{-2p} \sum_{q=2}^{2p} n^{q} \phi(h_s)^{q/2} \leq C n^{2p} \phi(h_s)^{-p},$$

where the last inequality is due to (H.11). Therefore, using (H.11) again, one has

$$\sum_{n=1}^{\infty} n^{\alpha} \sup_{(\upsilon,h_s)\in\Upsilon\times H_n} \left\{ \frac{\phi(h_s)^{2p}}{(n-1)^{2p}} \mathbb{E}\left(\left(\sum_{j=1}^{n} \sum_{i\neq j} U_{i,j,s}^{\upsilon} \right)^{2p} \right) \right\} \leq C \sum_{n=1}^{\infty} n^{\alpha-\nu_1 p},$$

so (7.20) holds for p large enough, and consequently (7.18) is proven.

Analogously, it can be checked (7.21) as follows in order to show the convergence for the term related to $V_{i,j,s}^{\upsilon}$. Firstly, Lemma 7.5 and (H.12) can be used to get

$$\begin{aligned}
\mathbb{E}\left(\left(\sum_{j=1}^{n} \sum_{i\neq j} V_{i,j,s}^{\upsilon} \right)^{2p} \right) &= \sum_{I_{2p}} \mathbb{E}\left(\prod_{l=1}^{2p} V_{i_l,j_l,s}^{\upsilon} \right) \\
&\leq C\phi(h_s)^{-2p} \sum_{q=2}^{4p} \sum_{J_q} \mathbb{E}\left(\prod_{l=1}^{q} |\epsilon_{r_l}|^{a_l} \tilde{K}_{s,u_l}^{b_l}(X_{w_l}) \right),
\end{aligned}$$

with I_{2p} and J_q defined as above, and $\sum_{l=1}^{q} a_l = \sum_{l=1}^{q} b_l = 2p$. Then, similar arguments to those used for the case $U_{i,j,s}^{\upsilon}$ above lead to

$$\sum_{n=1}^{\infty} n^{\alpha} \sup_{(\upsilon,h_s)\in\Upsilon\times H_n} \left\{ \frac{\phi(h_s)^{2p}}{(n-1)^{2p}} \mathbb{E}\left(\left(\sum_{j=1}^{n} \sum_{i\neq j} V_{i,j,s}^{\upsilon} \right)^{2p} \right) \right\} \leq C \sum_{n=1}^{\infty} n^{\alpha-\nu_1 p}.$$

Consequently, one gets (7.21) for p large enough, and thus (7.19) is shown. □

REFERENCES

Ait-Saïdi, A., Ferraty, F., Kassa, R., and Vieu, P. (2008), "Cross–validated estimations in the single–functional index model," *Statistics: A Journal of Theoretical and Applied Statistics*, 42, 475–494.

Benhenni, K., Ferraty, F., Rachdi, M., and Vieu, P. (2007), "Local smoothing regression with functional data," *Computational Statistics*, 22, 353–369.

Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. (2010), "Rate of uniform consistency for nonparametric estimates with functional variables," *Journal of Statistical Planning and Inference*, 140, 335–352.

Ferraty, F., Martinez-Calvo, A., and Vieu, P. (2013), "Exploring structures in functionaldata," *Preprint*.

Ferraty, F., Mas, A., and Vieu, P. (2007), "Nonparametric regression on functional data: inference and practical aspects," *Australian & New Zealand Journal of Statistics*, 49, 267–286.

Ferraty, F., and Vieu, P. (2006), *Nonparametric functional data analysis: theory and practice* (1st ed.), New York: Springer Series in Statistics.

Härdle, W., and Marron, J. (1985), "Optimal bandwidth selection in nonparametric regression function estimation," *Annals of Statistics*, 13, 1465–1481.

Marron, J., and Härdle, W. (1986), "Random Approximations to some measures of accuracy in nonparametric curve estimation," *Journal of Multivariate Analysis*, 20, 91–113.

Ramsay, J., and Silverman, B. (2005), *Functional Data Analysis* (2nd ed.), New York: Springer–Verlag.