On a Nadaraya-Watson Estimator with Two Bandwidths
Fabienne Comte, Nicolas Marie

To cite this version:

HAL Id: hal-02457079
https://hal.archives-ouvertes.fr/hal-02457079v2
Submitted on 25 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
On a Nadaraya-Watson Estimator with Two Bandwidths

Fabienne COMTE† and Nicolas MARIE⋄
†Laboratoire MAP5, Université de Paris, Paris, France
e-mail: fabienne.comte@parisdescartes.fr
⋄Laboratoire MODAL’X, Université Paris Nanterre, Nanterre, France
e-mail: nmarie@parisnanterre.fr

Abstract: In a regression model, we write the Nadaraya-Watson estimator of the regression function as the quotient of two kernel estimators, and propose a bandwidth selection method for both the numerator and the denominator. We prove risk bounds for both data driven estimators and for the resulting ratio. The simulation study confirms that both estimators have good performances, compared to the ones obtained by cross-validation selection of the bandwidth. However, unexpectedly, the single-bandwidth cross-validation estimator is found to be much better than the ratio of the previous two good estimators, in the small noise context. However, the two methods have similar performances in models with large noise.

MSC2020 subject classifications: Primary 62G08; secondary 62G05.

Keywords and phrases: Bandwidth selection, Nonparametric kernel estimator, Quotient estimator, Regression model.

Contents

1 Introduction .................................................. 2
2 Bound on the MISE of the 2bNW estimator ...................... 4
3 A bandwidth selection procedure for the 2bNW estimator based on the GL method ................................. 6
4 A bandwidths selection procedure for the 2bNW estimator based on the PCO method .................................. 8
5 Simulation study .................................................. 11
  5.1 Estimation of \( b_f \) ........................................ 11
  5.2 Estimation of \( b \) ........................................... 14
    5.2.1 Small noise case .................................... 14
    5.2.2 Large noise case .................................... 17
6 Proofs ......................................................... 20
  6.1 Proof of Proposition 2.4 .................................. 20
  6.2 Proof of Theorem 3.2 .................................... 20
  6.3 Proof of Corollary 3.3 ................................... 24
  6.4 Proof of Theorem 4.2 .................................... 24
    6.4.1 Steps of the proof. .................................. 25
    6.4.2 Proof of Lemma 6.2 ................................. 28
    6.4.3 Proof of Lemma 6.3 ................................. 35
1. Introduction

Consider $n \in \mathbb{N} \setminus \{0\}$ independent random variables $X_1, \ldots, X_n$ having the same probability density $f$ with respect to Lebesgue’s measure. Consider also the random variables $Y_1, \ldots, Y_n$ defined by

$$Y_i = b(X_i) + \varepsilon_i; \quad i \in \{1, \ldots, n\},$$

where $b$ is a measurable function from $\mathbb{R}$ into itself and $\varepsilon_1, \ldots, \varepsilon_n$ are $n$ i.i.d. centered random variables with variance $\sigma^2 > 0$ and independent of $X_1, \ldots, X_n$. Since Nadaraya [15] and Watson [20], a lot of consideration has been given to the estimator of $b$ defined by

$$\hat{b}_{n,h}(x) := \frac{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) Y_i}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)}; \quad x \in \mathbb{R},$$

where $K : \mathbb{R} \to \mathbb{R}$ is a kernel, and $h > 0$ is the bandwidth. This estimator has been dealt with as a weighted estimator, for $K \geq 0$:

$$\hat{b}_{n,h}(x) = \sum_{i=1}^n u_{n,h}^{(i)}(x) Y_i, \quad u_{n,h}^{(i)}(x) = \frac{K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)},$$

and is often called ”local average regression”. It is studied e.g. in Jones and Wand [11], Győrfi et al. [8] or defined in Tsybakov [19]. Recent papers still propose methods to improve the estimation, see Chang et al. [3]. Several strategies have been proposed to select the bandwidth in a data driven way. Cross-validation based on leave-one-out principle is one of the most standard methods to perform this choice (see Győrfi et al. [8]), even if a lot of refinements have been proposed. Optimal rates depend on the regularity of the function $b$ and have been first established by Stone [18]: roughly speaking, they are of order $O(n^{-p/(2p+1)})$ for $b$ admitting $p$ derivatives. From theoretical point of view, the rates of the adaptive final estimator are not always given, nor proved.

In this paper, we re-write the Nadaraya-Watson as the quotient of two estimators, an estimator of $bf$ divided by an estimator of $f$:

$$\hat{b}_{n,h}(x) := \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) Y_i$$

and

$$\hat{f}_{n,h'}(x) := \frac{1}{nh'} \sum_{i=1}^n K\left(\frac{X_i - x}{h'}\right).$$
Clearly, \( \hat{f}_{n,h'} \) is the Parzen-Rosenblatt estimator of \( f \) (see Rosenblatt [17] and Parzen [16]). The question we are interested in is the following one: can we choose separately the two bandwidths in an adaptive way and obtain good performance for each, and then for the ratio? This is why we study the estimator

\[
\hat{b}_{n,h,h'}(x) := \frac{\hat{b}f_{n,h}(x)}{\hat{f}_{n,h'}(x)} ; \ x \in \mathbb{R}
\]

as an estimator of the regression function \( b \), where \( h, h' > 0 \) and \( K : \mathbb{R} \to \mathbb{R} \) is a (not necessarily nonnegative) kernel. Thus, \( \hat{b}_{n,h,h} = \hat{b}_{n,h} \) is the initial Nadaraya-Watson estimator of \( b \) with single bandwidth \( h \). For this reason, the estimator studied in this paper is called the two bandwidths Nadaraya-Watson (2bNW) estimator.

Adaptive estimation of the density has been widely studied recently. A bandwidth selection method has been proposed by Goldenschluger and Lepski [7], and proved to reach the adequate bias-variance compromise. Implementation of this method revealed to be difficult due to the choice of two constants involved in the procedure, the intuition of which is not obvious. This is why the question was further investigated by Lacour et al. [13]: they improve and modify the strategy by using specific theoretical tools for their proofs. Precisely, thanks to a deviation inequality for U-statistics proved by Houdré and Reynaud-Bouret [10], they bound the Mean Integrated Square Error of their final estimator, which they call PCO (Penalised Comparison to Overfitting) estimator. Numerically, the good performance of their proposal has been illustrated in a naive way and for high order kernels in Comte and Marie [5], and through a systematic numerical study in Lacour et al. [14], including the multivariate case. These two methods and the associated results are dedicated to the selection of \( h' \) for \( \hat{f}_{n,h'}(x) \), and we can use them. Unfortunately, the theoretical results do not apply to \( \hat{b}f_{n,h}(x) \), mainly because they hold under a boundedness assumption: in our context, this would lead to assume that the \( Y_i \)'s are bounded. We do not want to require such an assumption as it would exclude the case of Gaussian errors \( \varepsilon_i \), for instance. Thus, we give moment assumptions under which the Goldenshluger and Lepski method on the one hand (see Section 3) and the PCO estimator on the other hand (see Section 4) can be applied to the estimation of \( bf \). When gathering the results for the numerator and the denominator, we can bound the risk of the quotient estimator of \( b \).

Concretely, we implement the PCO method for \( bf \) and compare it with a cross-validation (CV) strategy: in our examples, PCO almost always performs slightly better than CV. Therefore, the PCO adaptive estimation strategies for \( f \) and for \( bf \) are clearly good. However, unexpectedly, for small noise (\( \sigma = 0.1 \)), the quotient fails systematically to beat the specific regression CV method. Even if we compare the classical single-bandwidth CV regression estimator to the ratio of the oracles estimators of the numerator and the denominator, the former wins, and we obtain a quotient with two bandwidths which is in mean much less good than the CV estimator with single bandwidth. In practice, the bandwidth selected by the CV algorithm in that case is very small, and associated to
quite bad estimators of the numerator and of the denominator. This remark is of important interest for practitioners. In a second time, we increased the noise ($\sigma = 0.7$), and finally obtained results indicating that the two methods can have similar Mean Integrated Squared Errors (MISE) in this more difficult context. This, together with the fact we establish a theoretical risk bound on the PCO adaptive 2bNW estimator, imply that the PCO method, for both numerator and denominator, remains an interesting bandwidth selection method. Moreover, we believe that both positive but also negative results are of interest, and detailed tables, explanations and discussion are given in Section 5.

Notations:

1. For every square integrable functions $f, g : \mathbb{R} \rightarrow \mathbb{R}$,
   $$(f * g)(x) := \int_{-\infty}^{\infty} f(x - y)g(y)dy ; x \in \mathbb{R}.$$  

2. $K_\varepsilon := 1/\varepsilon K(\cdot/\varepsilon)$ for every $\varepsilon > 0$.

2. Bound on the MISE of the 2bNW estimator

First, we state some simple risk bound results in the case of a fixed bandwidth.

Consider $\beta > 0$ and $\ell := \lfloor \beta \rfloor$, where $\lfloor \beta \rfloor$ denotes the largest integer smaller than $\beta$. In the sequel, the kernel $K$ and the density function $f$ fulfill the following assumption.

**Assumption 2.1.**

(i) The map $K$ belongs to $L^2(\mathbb{R}, dy)$, $K$ is bounded and $\int_{\mathbb{R}} K(y)dy = 1$.

(ii) The density function $f$ is bounded.

Under this assumption, a suitable control of the MISE of $\hat{bf}_{n,h}$ has been established in Comte [4], Proposition 4.2.1.

**Proposition 2.2.** Under Assumption 2.1,

$$\mathbb{E}(\|\hat{bf}_{n,h} - bf\|^2) \leq \|bf - (bf)_h\|^2 + \frac{c_{K,Y}}{nh}$$

where $(bf)_h := K_h * (bf)$ and $c_{K,Y} := \|K\|_2^2 \mathbb{E}(Y^2)$.

In order to provide a suitable control on the MISE of the 2bNW estimator, we assume that $b$ and $f$ fulfill the following assumption.

**Assumption 2.3.** The function $b^2 f$ is bounded by a constant $c_{b,f} > 0$.

Note that this assumption does not require that $b$ is bounded and is satisfied in most classical examples.

Moreover, for any $S \in B(\mathbb{R})$, consider the norm $\|\cdot\|_{2,f,S}$ on $L^2(S, f(x)dx)$ defined by

$$\|\varphi\|_{2,f,S} := \left(\int_S \varphi(x)^2 f(x)dx\right)^{1/2} ; \forall \varphi \in L^2(S, f(x)dx).$$
Proposition 2.4. Let $m_n$ be a positive real number and consider
\[ S_n := \{ x \in \mathbb{R} : f(x) \geq m_n \}. \]

Under Assumptions 2.1 and 2.3,
\[
\mathbb{E}(\|n_{h,h'} - b\|^2_{f,S_n}) \leq \frac{8c_f}{m_n^2} \left( \|bf - (bf)\|_2 + \frac{c_{K,Y}}{nh} + 2c_{b,f} \left( \|f - f_{h'}\|^2_2 + \frac{c_{K}}{nh'} \right) \right)
\]
where $(bf)_h := K_h \ast (bf)$, $f_{h'} = K_{h'} \ast f$, $c_f := \|f\|_2^2 \vee 1$ and $c_K := \int \mathbb{R} K(y)^2 dy$.

The idea behind Proposition 2.4 is that we cannot pretend to accurately estimate $b$ on domains where few $X_i$’s are observed. Such domains correspond to small level of the density. For small $m_n$, the set $S_n$ excludes these cases.

Proposition 2.4 gives a decomposition of the risk of the quotient estimator as the sum of the risks of the estimators of the numerator $bf$ and of the denominator $f$, up to the multiplicative constant $8c_f/m_n^2$. Therefore, the rate of the quotient estimator is, in the best case, the worst rate of the two estimators used to define it (see also Remark 2.5 below). The factor $1/m_n^2$ may imply a global loss with respect to this rate. Clearly, the smaller is $m_n$, the larger is the loss.

For instance, if $f$ is lower bounded by a known constant $f_0$ on a given compact set $A$, then we can take $S_n = A$ and $m_n = f_0$. In that case, no loss occurs. If $f_0$ is unknown, we still can bound the risk with $S_n = A$ and $1/m_n^2 = \log(n)$ for $n$ large enough. A log-loss occurs then in the rate.

Remark 2.5. We consider, for $\beta, L > 0$, the Nikol’ski ball $\mathcal{H}(\beta, L)$, defined as the set of $\ell = [\beta]$ times continuously derivable functions $\varphi : \mathbb{R} \to \mathbb{R}$ such that $\varphi(t)$ satisfies
\[
\left[ \int_{-\infty}^{\infty} (\varphi(t)(x + t) - \varphi(t)(x))^2 dx \right]^{1/2} \leq L|t|^{\beta-t}; \forall t \in \mathbb{R}.
\]

For instance, for $p \in \mathbb{N}$, any function $\varphi \in C^{p+1}(\mathbb{R})$ such that $\text{supp}(\varphi^{(p)}) = [0, 1]$ and $\|\varphi^{(p+1)}\|_{\infty} \leq L$ belongs to $\mathcal{H}(p + 1, L)$. Indeed, for every $t \in \mathbb{R}_+$,
\[
\int_{-\infty}^{\infty} (\varphi(t)(x + t) - \varphi(t)(x))^2 dx \leq t \int_{-t}^{t} \int_{x}^{x+t} \varphi^{(p+1)}(z)1_{[0,1]}(z)dzdx \\
\leq L^2t \int_{-t}^{t} ((x + t) \wedge 1 - x \vee 0)dx \\
= L^2t \left( \int_{0}^{1+t} (x \wedge 1)dx - \int_{0}^{1} xdx \right) = L^2t^2.
\]

More subtly, $\psi : x \mapsto e^{-x}1_{\mathbb{R}_+}(x)$ belongs to $\mathcal{H}(1/2, 1)$. Indeed, for every $t \in \mathbb{R}_+$,
\[
\int_{-\infty}^{\infty} (\psi(x + t) - \psi(x))^2 dx = \int_{-t}^{\infty} e^{-2(x+t)}dx - 2 \int_{0}^{\infty} e^{-t-2x}dx + \int_{0}^{\infty} e^{-2x}dx \\
= \lim_{x \to \infty} -\frac{1}{2} [e^{-2t}(e^{-2x} - e^{2t}) - 2e^{-t}(e^{-2x} - 1) + e^{-2x} - 1] = 1 - e^{-t} \leq t.
\]
Now, assume that $bf$ belongs to $H(\beta_1, L)$ and $f$ to $H(\beta_2, L)$. We also assume that the kernel $K$ satisfies Assumption 2.1 and is of order $\ell = \lceil \max(\beta_1, \beta_2) \rceil$, that is
\[
\int_{-\infty}^{\infty} |u^k K(u)| du < \infty \quad \text{and} \quad \int_{-\infty}^{\infty} u^k K(u) du = 0 \; \forall k \in \{1, \ldots, \ell\}.
\]
Then, it follows from Tsybakov [19], Chapter 1, that
\[
\|bf - (bf)_{h_\text{opt}}\|_2^2 \leq C(\beta_1, L)h^{2\beta_1} \quad \text{and} \quad \|f - f_{h'}\|_2^2 \leq C'(\beta_2, L)(h')^{2\beta_2}.
\]
This implies that choosing $h_\text{opt} = c_1 n^{1/(2\beta_1+1)}$ in Proposition 2.2 yields
\[
\mathbb{E}(\|bf_{n,h,h'}_{n,h'} - bf\|_2^2) \lesssim n^{-2\beta_1/(2\beta_1+1)},
\]
which is a standard optimal rate of estimation on Nikol’ski balls. The same rate holds for the estimation of $f$ under our assumptions, with $\beta_1$ replaced by $\beta_2$, and $h'_{\text{opt}} = c_2 n^{1/(2\beta_2+1)}$. This implies that
\[
\|bf - (bf)_{h_{\text{opt}}}\|_2^2 + \frac{c_{K,Y}}{nh_{\text{opt}}^2} + 2c_{\nu,f} \left( \|f - f_{h'_{\text{opt}}}\|_2^2 + \frac{c_{K}}{nh'_{\text{opt}}^2} \right) \lesssim \max(n^{-2\beta_1/(2\beta_1+1)}, n^{-2\beta_2/(2\beta_2+1)}).
\]
So, the rate is optimal if $\beta = \min(\beta_1, \beta_2)$ is the regularity of $b$.
However, such bandwidth choices are not possible in practice, as they depend on unknown regularity parameters. Data driven bandwidth selection methods are settled to automatically reach a squared bias-variance compromise, inducing the optimal rate if the function under estimation does belong to a regularity space.

3. A bandwidth selection procedure for the 2bNW estimator based on the GL method

The bound on the MISE of $\hat{b}_{n,h,h'}$ obtained in Proposition 2.4 suggests to select $h$ and $h'$ separately, so that both bounds are minimal. The Goldenshluger-Lepski method (see Goldenshluger and Lepski [7]) allows to do this for $\hat{f}_{n,h}$, but requires to be extended to the estimator of $bf$. In particular, extensions of the proof are required as we do not wish to assume that the $Y_i$’s are bounded.

Consider the collection of bandwidths $\mathcal{H}_n := \{h_1, \ldots, h_{N(n)}\} \subset [0, 1]$, where $N(n) \in \{1, \ldots, n\}$ and
\[
\frac{1}{n} < h_1 < \cdots < h_{N(n)}.
\]
Moreover, we will need the following conditions.

**Assumption 3.1.** There exists $m > 0$, not depending on $n$, such that
\[
\frac{1}{n} \sum_{i=1}^{N(n)} \frac{1}{h_i} \leq m.
\]
and for every $c > 0$, there exists $m(c) > 0$, not depending on $n$, such that

$$\sum_{i=1}^{N(n)} \frac{1}{\sqrt{h_i}} \exp \left( -\frac{c}{\sqrt{h_i}} \right) \leq m(c).$$

**Example.** Consider the dyadic bandwidths defined by

$$h_i := 2^{-i}; \forall i = 0, 1, \ldots, \left\lfloor \frac{\log(n)}{\log(2)} \right\rfloor.$$

Then,

$$\frac{1}{n} \sum_{i=1}^{\left\lfloor \frac{\log(n)}{\log(2)} \right\rfloor} 2^{i} \leq \frac{2n - 1}{n} \leq 2$$

and

$$\sum_{i=1}^{\left\lfloor \frac{\log(n)}{\log(2)} \right\rfloor} 2^i \exp(-c2^{i/2}) \leq \sum_{i=1}^{\left\lfloor \frac{\log(n)}{\log(2)} \right\rfloor} \sqrt{i} \exp(-c\sqrt{i}) \leq m(c) < \infty.$$

Thus, Assumption 3.1 is fulfilled.

Consider also

$$\hat{bf}_{n,h,\eta}(x) := (K_\eta * \hat{bf}_{n,h})(x) = \frac{1}{n} \sum_{i=1}^{n} Y_i (K_\eta * K_h)(X_i - x).$$

We apply the Goldenshluger-Lepski bandwidth selection method to $\hat{bf}_{n,h}$ by solving the minimization problem

$$\min_{h \in H_n} \{A_n(h) + V_n(h)\}$$

(1)

where

$$A_n(h) := \sup_{\eta \in H_n} (\|\hat{bf}_{n,h,\eta} - \hat{bf}_{n,\eta}\|^2_2 - V_n(\eta))_+ \quad \text{and} \quad V_n(h) := \frac{\mu}{nh} \|K\|^2_1,$$

with $\mu > 0$ not depending on $n$ and $h$, and $\mu_{K,Y} = \|K\|^2_2 \mathbb{E}(Y^2_1)$. In the sequel, the solution to the minimization Problem (1) is denoted by $\hat{h}_n$.

The idea behind the criterion is that $A_n(h)$ is an estimate of the squared bias term $\|(bf)_h - bf\|^2_2$ and $V_n(h)$ an estimate of the variance. So, $\hat{h}_n$ makes the compromise. See more details about the heuristics in Chagny [2], Section 4.4.

**Theorem 3.2.** Under Assumptions 2.1 and 3.1, if $\mathbb{E}(Y^6_1) < \infty$, then there exist two deterministic constants $c, \tau > 0$, not depending on $n$, such that

$$\mathbb{E}(\|\hat{bf}_{n,\hat{h}_n} - bf\|^2_2) \leq c \cdot \inf_{h \in H_n} \{\|(bf)_h - bf\|^2_2 + V_n(h)\} + \frac{\mu}{n} \log(n)^2.$$
Theorem 3.2 states that \( \hat{b}_{n,h_n} \) automatically leads to a compromise between the squared bias (\( \| (b f)_h - b f \|_2^2 \)) and the variance (\( V_n(h) \)) terms. The multiplicative constant \( c \), which is larger than one, is the price of the method but preserves the rate. Lastly, the additive quantity \( \log(n)^2/n \) is negligible with respect to the possible rate of convergence (see Remark 2.5).

We recall now a version of the result proved by Goldenshluger and Lepski [7], which is available for the estimator of \( f \). See also a simplified proof in Comte [4], Section 4.2. Let us consider

\[
\hat{h}'_n \in \arg \min_{h' \in \mathcal{H}_n} \{ A'_n(h') + V'_n(h') \},
\]

where

\[
A'_n(h') := \sup_{\eta \in \mathcal{H}_n} \| K \eta * \hat{f}_{n,h'} - \hat{f}_{n,\eta} \|_2^2 - V'_n(h') \quad \text{and} \quad V'_n(h') := \chi \frac{\| K \|_2^2 \| K \|_1^{\frac{1}{2}} \| K \|_2^{\frac{1}{2}}}{nh'}
\]

with \( \chi > 0 \) not depending on \( n \) and \( h' \). Under Assumptions 2.1 and 3.1, there exist two deterministic constants \( \epsilon', \chi' > 0 \), not depending on \( n \), such that

\[
E(\| \hat{f}_{n,h_n} - f \|_2^2) \leq \epsilon' \cdot \inf_{h' \in \mathcal{H}_n} \{ \| f_{h'} - f \|_2^2 + V'_n(h') \} + \frac{\epsilon'}{n} \tag{2}
\]

Gathering (2) and Theorem 3.2 yields a Corollary similar to Proposition 2.4.

**Corollary 3.3.** Let \( m_n \) be a positive real number and consider

\[
S_n := \{ x \in \mathbb{R} : f(x) \geq m_n \}.
\]

Under Assumptions 2.1, 2.3 and 3.1, if \( E(Y_1^6) < \infty \), then

\[
E(\| \hat{b}_{n,h_n,\hat{h}_n} - b \|_2^2, S_n) \leq \mathcal{C}_n \inf_{(h,h') \in \mathcal{H}_n^2} \{ \| (b f)_h - b f \|_2^2 + \| f_{h'} - f \|_2^2 + V_n(h) + V'_n(h') \} + \mathcal{T}_n \frac{\log(n)^2}{n},
\]

where

\[
\mathcal{C}_n := \frac{8\epsilon f}{m_n^2} (\epsilon \vee (2\epsilon_{b,f} \epsilon')) \quad \text{and} \quad \mathcal{T}_n := \frac{8\epsilon f}{m_n^2} (\epsilon + 2\epsilon_{b,f} \epsilon').
\]

The comments following Proposition 2.4 and in Remark 2.5 apply here.

**4. A bandwidths selection procedure for the 2bNW estimator based on the PCO method**

The Goldenshluger–Lepski method is mathematically very nice and provides a rigorous risk bound for the adaptive estimator with random bandwidth. However, it has been acknowledged as being difficult to implement, due to the square grid in \( h, \eta \) required to compute intermediate versions of the criterion and to the lack of intuition to guide the choice of the constants \( \nu \) and \( \chi \) which should
be calibrated from preliminary simulation experiments, see e.g. Comte and Rebačka [6]. This is the reason why Lacour et al. [13] investigated and proposed a simplified criterion (PCO) relying on deviation inequalities for $U$-statistics due to Houdré and Reynaud-Bouret [10]. This inequality applies in our more complicated context and Lacour-Massart-Rivoirard’s result can be extended here as follows.

Let us recall that $K_h(\cdot) = 1/h K(\cdot/h)$ and

$$(bf)_h = E(\hat{bf}_{n,h}) = K_h * (bf)$$

(see Lemma 6.1). Let $h_{\text{min}}$ be the smallest bandwidth value in $\mathcal{H}_n$ and consider

$$\text{crit}(h) := \|\hat{bf}_{n,h} - \hat{bf}_{n,h_{\text{min}}}\|_2^2 + \text{pen}(h)$$

with

$$\text{pen}(h) := \frac{2(K_{h_{\text{min}}} K_h)^2}{n^2} \sum_{i=1}^n Y_i^2.$$ 

Then, let us define

$$\tilde{h}_n \in \arg\min_{h \in \mathcal{H}_n} \text{crit}(h).$$

The idea behind the proposal of Lacour et al. (2017) is that, instead of comparing estimators $\hat{bf}_{n,h}$ to a collection of estimators $\hat{bf}_{n,h_\eta}$ for different bandwidths $\eta$, it is sufficient to compare them to the same single estimator, corresponding to the smallest bandwidth. See their Section 3.1 for more heuristic elements. This implies a faster and more efficient numerical procedure.

In the sequel, in addition to Assumption 2.1, the kernel $K$, the functions $b$ and $f$, the distribution of $Y_1$ and $h_{\text{min}}$ fulfill the following assumption.

**Assumption 4.1.** The kernel $K$ is symmetric and $K(0) > 0,

$$\frac{1}{nh_{\text{min}}} \leq 1,$$

$bf$ is bounded, and there exists $\alpha > 0$ such that $E(\exp(\alpha|Y_1|)) < \infty$.

As for Assumption 2.3, we can note that assuming $bf$ bounded does not require $b$ to be bounded, since most densities decrease fast at infinity. Moreover, the moment condition here is $E(\exp(\alpha|Y_1|)) < \infty$ and is stronger than for the Goldenschluger and Lepski method ($E(Y_1^6) < \infty$).

**Theorem 4.2.** Consider $\theta \in (0,1)$. Under Assumptions 2.1 and 4.1, there exist two deterministic constants $a, b > 0$, not depending on $n$, $h_{\text{min}}$ and $\theta$, such that

$$E(\|\hat{bf}_{n,\tilde{h}_n} - bf\|_2^2) \leq (1+\theta) \inf_{h \in \mathcal{H}_n} E(\|\hat{bf}_{n,h} - bf\|_2^2) + \frac{a}{\theta} \|bf\|_{h_{\text{min}}}^2 - bf\|_2^2 + \frac{b}{\theta} \log(n)^5 \frac{1}{n}. $$
Theorem 4.2 states that the estimator \( \hat{bf}_{n,h_n} \) has performance of order of the best estimator of the collection \( \inf_{h \in H_n} \mathbb{E}(\|\hat{bf}_{n,h} - bf\|_2^2) \) up to a factor \((1 + \vartheta)\). Indeed, the two other terms can be considered as negligible. If \( bf \) is in the Nikol’ski ball \( H(\beta_1, L) \) as in Remark 2.5, then the first right-hand-side term is of order \( n^{-2\beta_1/(2\beta_1+1)} \). Since for \( h_{\text{min}} = 1/n \), \( \|bf\|_{h_{\text{min}}} - bf \) is of order \( n^{-2\beta_1} \), both this term and the last residual term \( \log(n)^5/n \) are negligible compared to the first one.

Now, we state the result that can be deduced from Lacour et al. [13] for the estimator of \( f \). Let us consider \( \tilde{h}'_n \in \text{arg min}_{h' \in H_n} \text{crit}'(h') \), where
\[
\text{crit}'(h') := \|\hat{f}_{n,h'} - \hat{f}_{n,h_{\text{min}}}\|_2^2 + \text{pen}'(h') \quad \text{and} \quad \text{pen}'(h') := \frac{2(K_{h_{\text{min}}}, K_h)}{n}.
\]
By Lacour et al. [13], Theorem 2, there exists two deterministic constants \( a', b' > 0 \), not depending on \( n \) and \( h_{\text{min}} \), such that for every \( \vartheta \in (0, 1) \),
\[
\mathbb{E}(\|\hat{f}_{n,h_n} - f\|_2^2) \leq (1 + \vartheta) \inf_{h' \in H_n} \mathbb{E}(\|\hat{f}_{n,h'} - f\|_2^2) + \frac{a'}{\vartheta} \|f_{h_{\text{min}}} - f\|_2^2 + \frac{b'}{\vartheta n}.
\]
Again, we can gather this last result and Theorem 4.2 to get the following Corollary.

**Corollary 4.3.** Let \( m_n \) be a positive real number and consider
\[
S_n := \{x \in \mathbb{R} : f(x) \geq m_n\}.
\]
Consider also \( \vartheta \in (0, 1) \). Under Assumptions 2.1, 2.3 and 4.1,
\[
\mathbb{E}(\|\hat{bf}_{n,h_n,h'_n} - bf\|_2^2, f, S_n) \leq (1 + \vartheta) C_n(1, 1) \inf_{(h,h') \in H_n^2} \{\mathbb{E}(\|\hat{bf}_{n,h} - bf\|_2^2) + \mathbb{E}(\|\hat{f}_{n,h'} - f\|_2^2) \}
\]
\[
+ \frac{C_n(a,a')}{\vartheta} (\|(bf)_{h_{\text{min}}} - bf\|_2^2 + \|f_{h_{\text{min}}} - f\|_2^2)
\]
\[
+ \frac{C_n(b,b')}{\vartheta} \cdot \frac{\log(n)^5}{n},
\]
where
\[
C_n(u, v) := \frac{8c_f}{m_n^2} (u \lor (2c_f v)) \; ; \forall u, v \in \mathbb{R}.
\]
The proof of Corollary 4.3 relies to the same arguments as the proof of Corollary 3.3 provided in Section 3.3, and is therefore omitted.
5. Simulation study

For the noise, we consider \(\varepsilon \sim \sigma N(0,1)\), with \(\sigma = 0.1\) and \(\sigma = 0.7\). For the signal, we take either \(X \sim N(0,1)\) or \(X \sim \gamma(3, 2)/5\) (where the factor 5 is set to keep the variance of \(X\) of order 1, as in the first case). For the function \(b\), we took functions with different features and regularities:

- \(b_1(x) = \exp(-x^2/2)\),
- \(b_2(x) = x^2/4 - 1\),
- \(b_3(x) = \sin(\pi x)\),
- \(b_4(x) = \exp(-|x|)\).

We illustrate in Figures 1 and 2 the difference between a sample generated with \(\sigma = 0.1\) (small noise) and with \(\sigma = 0.7\) (large noise), compared with the functions to estimate. We can see that the first case is easy and that the second one is very difficult. Notice that the vertical scales are different.

![Observations](image1.png)

**Fig 1.** Observations \((X_i, Y_i)_{1 \leq i \leq n}\) for \(n = 1000\) in the four cases of functions \(b_1\) to \(b_4\), with small noise \(\sigma = 0.1\), and true regression function in bold red.

5.1. Estimation of \(bf\)

The PCO method is implemented for \(f\) and \(bf\) with a kernel of order 7 (i.e. \(\int x^k K(x) dx = 0\) for \(k = 1\) to 7), defined by \(K(x) = 4n_1(x) - 6n_2(x) + 4n_3(x) - n_4(x)\), where \(n_j(.)\) is a Gaussian density with mean 0 and variance \(j\). Note that,
Fig 2. Observations \((X_i, Y_i)\)\(_{1 \leq i \leq n}\) for \(n = 1000\) in the four cases of functions \(b_1\) to \(b_4\), with large noise \(\sigma = 0.7\), and true regression function in bold red.

for \(n_{j,h}(x) := 1/hn_j(x/h)\), it holds that

\[
\langle n_i, h_1, n_{j,h_2} \rangle_2 = \int_{-\infty}^{\infty} n_{i,h_1}(x)n_{j,h_2}(x)dx = \frac{1}{\sqrt{2\pi}} \times \frac{1}{\sqrt{ih_1^2 + jh_2^2}} \quad (3)
\]

The bandwidth is selected among \(M = 75\) equispaced values in between 0.01 and 1. All functions (true or estimated) are computed at 100 equispaced points in the interquantile interval corresponding to the 2% and 98% quantiles of \(X\). The bandwidth is selected via the PCO criterion, where \(h_{\text{min}} = 0.01\), and

\[
\text{crit}(h) := \|\hat{f}_{n,h} - \hat{f}_{n,h_{\text{min}}}\|_2^2 + \text{pen}(h) \quad \text{with} \quad \text{pen}(h) := \frac{2(K_{h_{\text{min}}}K_h)^2}{n^2} \sum_{i=1}^{n} Y_i^2.
\]

Note that the bandwidth of the density estimator is selected as in Comte and Marie [5], by minimizing

\[
\text{crit}'(h) := \|\hat{f}_{n,h} - \hat{f}_{n,h_{\text{min}}}\|_2^2 + 2\text{pen}'(h) \quad \text{with} \quad \text{pen}'(h) := \frac{2(K_{h_{\text{min}}}K_h)^2}{n}.
\]

The \(L^2\)-norm is computed as a Riemann sum on the interquantile interval, while the penalty is explicit and exact, thanks to Formula (3). The cross-validation (CV) criterion for selecting the bandwidth of \(\hat{f}_{n,h}\) is com-
puted as follows:

$$CV(h) := \int \hat{b}f_h(x)^2 dx - \frac{2}{n(n-1)}h \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} Y_i Y_j N\left(\frac{X_i - X_j}{h}\right),$$

where $N(.)$ is the Gaussian kernel, also used to compute the estimator $\hat{b}f_{n,h}$ in this case. It provides an estimation of $\|\hat{b}f_h\|_2^2 - 2\langle \hat{b}f_h, b f \rangle_2$ relying on the idea that the empirical for $\langle t, b f \rangle_2$ is $1/n \sum_{i=1}^{n} Y_i t(X_i)$. The chosen bandwidth is the minimizer of $CV(h)$ in the same collection as previously.

### Table 1

<table>
<thead>
<tr>
<th>n</th>
<th>PCO</th>
<th>CV</th>
<th>Or</th>
<th>PCO</th>
<th>CV</th>
<th>Or</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>0.33</td>
<td>0.37</td>
<td>0.16</td>
<td>0.32</td>
<td>0.38</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>(0.28)</td>
<td>(0.82)</td>
<td>(0.14)</td>
<td>(0.28)</td>
<td>(1.01)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>500</td>
<td>0.17</td>
<td>0.29</td>
<td>0.08</td>
<td>0.17</td>
<td>0.26</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>(0.13)</td>
<td>(0.86)</td>
<td>(0.07)</td>
<td>(0.14)</td>
<td>(1.10)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>1000</td>
<td>0.09</td>
<td>0.21</td>
<td>0.05</td>
<td>0.09</td>
<td>0.21</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.57)</td>
<td>(0.04)</td>
<td>(0.07)</td>
<td>(0.60)</td>
<td>(0.03)</td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>n</th>
<th>PCO</th>
<th>CV</th>
<th>Or</th>
<th>PCO</th>
<th>CV</th>
<th>Or</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>0.45</td>
<td>0.42</td>
<td>0.31</td>
<td>0.45</td>
<td>0.40</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>(0.25)</td>
<td>(0.32)</td>
<td>(0.17)</td>
<td>(0.24)</td>
<td>(0.81)</td>
<td>(0.17)</td>
</tr>
<tr>
<td>500</td>
<td>0.23</td>
<td>0.35</td>
<td>0.15</td>
<td>0.20</td>
<td>0.39</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.88)</td>
<td>(0.08)</td>
<td>(0.13)</td>
<td>(0.84)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>1000</td>
<td>0.12</td>
<td>0.24</td>
<td>0.09</td>
<td>0.11</td>
<td>0.19</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.55)</td>
<td>(0.05)</td>
<td>(0.07)</td>
<td>(0.40)</td>
<td>(0.04)</td>
</tr>
</tbody>
</table>

MISE obtained for 200 repetitions and sample sizes 250, 500 and 1000, for the estimation of $b f$ with PCO and CV methods, for $\sigma = 0.1$ (Table 1) and $\sigma = 0.7$ (Table 2). The column "Or" gives the mean of the minimal squared errors for each sample, which requires to use the unknown true function and represents what could be obtained at best (that is if the best possible bandwidth was chosen for each sample). We postpone results with $X \sim \gamma(3, 2)/5$ in Appendix A since they are similar. We can see that the PCO method is globally better than the CV, with no important difference, and the oracle shows that we are in the right orders even if not at best.

Table 3 presents the mean of the selected bandwidths in each case PCO and CV, and allows to compare it with the oracle bandwidth, for the same paths and configurations as previously. The conclusion here is that, in mean, the PCO method over-estimates the oracle bandwidth, while the CV method slightly under-evaluates it. Clearly, the too-large choice gives better results.
Table 2

100*MISE (with 100*std in parenthesis below) for the estimation of $b_1f$ corresponding to the
four examples $b_1,\ldots, b_4$, 200 repetitions, $X \sim N(0,1)$ and $\sigma = 0.7$. Columns PCO and CV
correspond to the two competing methods. "Or" is for "oracle" and gives the average error
of the best possible estimator of the collection, computed for each sample.

<table>
<thead>
<tr>
<th>n</th>
<th>PCO</th>
<th>CV</th>
<th>Or</th>
<th>PCO</th>
<th>CV</th>
<th>Or</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>0.56</td>
<td>0.70</td>
<td>0.30</td>
<td>0.51</td>
<td>0.92</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>(0.40)</td>
<td>(1.44)</td>
<td>(0.22)</td>
<td>(0.33)</td>
<td>(2.83)</td>
<td>(0.22)</td>
</tr>
<tr>
<td>500</td>
<td>0.30</td>
<td>0.32</td>
<td>0.16</td>
<td>0.29</td>
<td>0.33</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>(0.23)</td>
<td>(0.68)</td>
<td>(0.13)</td>
<td>(0.23)</td>
<td>(0.60)</td>
<td>(0.13)</td>
</tr>
<tr>
<td>1000</td>
<td>0.14</td>
<td>0.29</td>
<td>0.08</td>
<td>0.15</td>
<td>0.30</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.91)</td>
<td>(0.06)</td>
<td>(0.15)</td>
<td>(1.14)</td>
<td>(0.07)</td>
</tr>
</tbody>
</table>

5.2. Estimation of $b$

Now, we present the results for the estimation of the regression function $b$, obtained either with a single-bandwidth estimator, or with the ratio of two adaptive PCO estimators of $b_1f$ and $f$.

The PCO estimators are the ones studied above, which proved to be good estimators (see also the study for the estimation of $f$ in Comte and Marie [5]).

We simply take a point by point ratio of the two adaptive PCO estimators. The oracle we refer to is computed with the estimator of $b$ obtained as a quotient of the two oracles of $b_1f$ and $f$ for each path. It is the best performance we can expect with a PCO-ratio strategy.

For the one-bandwidth Nadaraya-Watson estimator $\hat{b}_{n,h}$, it is computed with the Gaussian kernel $N(.)$. The leave-one-out cross-validation criterion which is minimized for the bandwidth selection is

$$CV_{NW}(h) := \sum_{i=1}^{n} (Y_i - \hat{b}_{n,h}^{(-i)}(X_i))^2$$

with

$$\hat{b}_{n,h}^{(-i)}(x) := \frac{\sum_{j=1, j\neq i}^{n} N((X_j - x)/h)}{\sum_{k=1, k\neq i}^{n} N((X_k - x)/h)} Y_j.$$

5.2.1. Small noise case

We’ve started the study with $\sigma = 0.1$, which in our mind was an easy case (see Figure 1). Table 4 presents the results for the estimation of $b$, either with
\begin{table}
\centering
\begin{tabular}{|c|ccc|ccc|}
\hline
\textbf{n} & \hline
250 & 0.71 & 0.54 & 0.54 & 0.78 & 0.62 & 0.62 \\
 & (0.14) & (0.15) & (0.08) & (0.15) & (0.17) & (0.09) \\
500 & 0.62 & 0.47 & 0.51 & 0.70 & 0.54 & 0.28 \\
 & (0.13) & (0.14) & (0.08) & (0.15) & (0.18) & (0.08) \\
1000 & 0.55 & 0.41 & 0.47 & 0.63 & 0.50 & 0.54 \\
 & (0.11) & (0.14) & (0.07) & (0.12) & (0.17) & (0.07) \\
\hline
\end{tabular}
\caption{Means of selected bandwidths (with std in parenthesis below) for the estimation of $bf$, 200 repetitions, $X \sim N(0,1)$, $\sigma = 0.1$.}
\end{table}

\begin{table}
\centering
\begin{tabular}{|c|ccc|ccc|}
\hline
\textbf{n} & \hline
250 & 0.35 & 0.28 & 0.31 & 0.57 & 0.39 & 0.37 \\
 & (0.04) & (0.05) & (0.03) & (0.17) & (0.14) & (0.07) \\
500 & 0.32 & 0.26 & 0.28 & 0.48 & 0.32 & 0.33 \\
 & (0.04) & (0.06) & (0.03) & (0.13) & (0.14) & (0.07) \\
1000 & 0.30 & 0.24 & 0.26 & 0.39 & 0.28 & 0.28 \\
 & (0.03) & (0.07) & (0.03) & (0.11) & (0.11) & (0.06) \\
\hline
\end{tabular}
\caption{100*MISE (with 100*std in parenthesis below) for the estimation of $b_i$, $i = 1, \ldots, 4$, 200 repetitions, $X \sim N(0,1)$, $\sigma = 0.1$. CV and PCO are the two competing methods. Column “Or” gives the average of ISE for the ratio of the two best estimators of $bf$ and $f$ in the collection.}
\end{table}
Table 5
Mean of selected bandwidth (with std in parenthesis below) with the CV method for NW-single bandwidth estimator of $b$, $\sigma = 0.1$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$b_3$</th>
<th>$b_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>0.13</td>
<td>0.13</td>
<td>0.06</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.03)</td>
<td>(0.01)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>500</td>
<td>0.12</td>
<td>0.11</td>
<td>0.05</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>1000</td>
<td>0.11</td>
<td>0.09</td>
<td>0.05</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
</tbody>
</table>

Fig 3. Small bandwidth effect, the ratio of two bad estimators is a good estimator. For bandwidth $h = 0.01$ and $n = 1000$, true (bold) and estimated $f$ (left), $b_1 f$ (middle), and $b_4$ (right).
the $CV_{NW}$ criterion, with ratio of PCO of $bf$ and $f$, or with the ratio of the best estimators of $bf$ and $f$ in the collection. More precisely, the column "Or" gives here the MISE computed with the estimator of $b$ obtained as a quotient of the two oracles of $bf$ and $f$ in each example and for each sample path. Clearly, the performance of the Nadaraya-Watson cross-validation criterion is much better, within a multiplicative factor from 2 and up to 6. The variance of the quotient estimators (oracle and PCO) are large, which shows that the mean performance is probably deteriorated by a few very bad results. However, the result is puzzling: even the ratio of the two best estimators of the numerator and denominator does not reach the good performance of the single-bandwidth CV method. Table 5 shows in addition that the selected bandwidths are in mean very small. We can check that the ratio of this bad numerator divided by a bad denominator fits well to the $b$ quotient function: this is illustrated by Figure 3. It is likely that both imply a compensation resulting in a locally, and thus also globally, better estimate. We can notice that the selected bandwidth also decrease more slowly when $n$ increases (see Table 5) than for the estimator of $bf$ (see Table 3). Our explanation (see the heuristic Remark 5.1 below) is that the risk of the Nadaraya-Watson estimator behaves as $C(h^{2\alpha} + \sigma^2/(nh))$, for some $\alpha > 0$ related to the regularity of $b$, like in the projection least-squares method (see e.g. Baraud [1]). In the small noise case, $\sigma^2$ makes the variance term negligible, so that the bandwidth selection method aims at having small bias term $h^{2\alpha}$. On the other hand, the risk decomposition of the estimator of $bf$ involves a variance term of order $\|K\|^2 E(Y^2)/(#nh)$, and in all our examples, empirical evaluations of $E(Y^2)$ is in the range [0.34, 0.70], making the ratio with $\sigma^2$ between 34 and 70. In other words, the variance term for this estimator is 34 to 70 times larger. This is why it is important to investigate large noise case and a less favorable signal to noise ratio.

5.2.2. Large noise case

When setting $\sigma = 0.7$, the empirical order of $E(Y^2)$ for the four models is between 0.91 and 1.31, which divided by $\sigma^2$ gives now a value between 1.85 and 2.67. This is much smaller than previously. This corresponds to a more difficult estimation problem, as can be seen from Figure 2.

We now comment the results given in Table 6. The MISE are quite larger, but in Figure 4, we show examples of estimated curves in this case, and the associated orders of MISEs, computed for 25 repetitions; they are not as good as for small noise, but still reasonable. The results in Table 6 show that the MISE have now the same orders, and the oracles can be much better than the results of the Nadaraya-Watson estimator. The selected bandwidths are larger and decreasing with $n$ (see Table 7 in Appendix).

The conclusion of this study is that adaptive estimation of functions with kernel estimators and bandwidth selection relying on the PCO method proposed by Lacour et al. [13] gives very good results in theory and practice, not only for
density estimation. However, for regression function estimation, one bandwidth selected with a criterion directly suited to the regression function is safer than the two different bandwidths selected when considering the Nadaraya-Watson estimator as a quotient of two functions that may be estimated separately. The results are not bad, but the strategy must be devoted to more complicated contexts where direct estimators of $b$ are not feasible.

**Remark 5.1.** Note that if $h = h'$, which means that $\hat{b}_{n,h,h'}$ is the usual Nadaraya-Watson estimator $\hat{b}_{n,h}$,

$$\hat{b}_{n,h}(x) = \sum_{i=1}^{n} w_{n,h}(i,x) \varepsilon_i \quad \text{with} \quad w_{n,h}(i,x) := \frac{K((X_i - x)/h)}{\sum_{j=1}^{n} K((X_j - x)/h)}.$$

Then,

$$\hat{b}_{n,h}(x) - b(x) = \sum_{i=1}^{n} w_{n,h}(i,x)(b(X_i) - b(x)) + \sum_{i=1}^{n} w_{n,h}(i,x)\varepsilon_i,$$

and for a nonnegative kernel with compact support $[-1, 1]$, if the regression function $b$ is Lipschitz continuous, then

$$\mathbb{E}[(\hat{b}_{n,h}(x) - b(x))^2] \leq Ch^2 + \sigma^2 \mathbb{E} \left( \sum_{i=1}^{n} w_{n,h}(i,x)^2 \right),$$

Moreover,

$$\mathbb{E} \left( \sum_{i=1}^{n} w_{n,h}(i,x)^2 \right) = \frac{1}{nh} \mathbb{E} \left( \frac{1}{\frac{1}{nh} \sum_{i=1}^{n} K((X_i - x)/h)^2} \right),$$

<table>
<thead>
<tr>
<th>$n$</th>
<th>CV</th>
<th>$b_1$</th>
<th>PCO</th>
<th>Or</th>
<th>CV</th>
<th>$b_2$</th>
<th>PCO</th>
<th>Or</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>7.81</td>
<td>9.02</td>
<td>6.54</td>
<td>8.32</td>
<td>8.61</td>
<td>5.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.32)</td>
<td>(5.86)</td>
<td>(5.13)</td>
<td>(6.02)</td>
<td>(6.67)</td>
<td>(4.87)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>4.33</td>
<td>4.34</td>
<td>3.25</td>
<td>4.57</td>
<td>4.83</td>
<td>3.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.09)</td>
<td>(2.71)</td>
<td>(2.15)</td>
<td>(2.86)</td>
<td>(4.07)</td>
<td>(2.58)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>2.26</td>
<td>2.19</td>
<td>1.72</td>
<td>2.38</td>
<td>2.13</td>
<td>1.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.37)</td>
<td>(1.30)</td>
<td>(1.16)</td>
<td>(1.39)</td>
<td>(2.04)</td>
<td>(1.15)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n$</th>
<th>CV</th>
<th>$b_3$</th>
<th>PCO</th>
<th>Or</th>
<th>CV</th>
<th>$b_4$</th>
<th>PCO</th>
<th>Or</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>16.2</td>
<td>18.5</td>
<td>14.8</td>
<td>7.84</td>
<td>9.36</td>
<td>7.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(8.47)</td>
<td>(10.3)</td>
<td>(9.06)</td>
<td>(4.94)</td>
<td>(4.62)</td>
<td>(4.94)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>8.54</td>
<td>9.28</td>
<td>7.58</td>
<td>4.41</td>
<td>5.01</td>
<td>4.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.58)</td>
<td>(5.26)</td>
<td>(4.41)</td>
<td>(2.83)</td>
<td>(2.31)</td>
<td>(2.27)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>4.54</td>
<td>4.62</td>
<td>3.87</td>
<td>2.40</td>
<td>2.85</td>
<td>2.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.81)</td>
<td>(2.35)</td>
<td>(2.18)</td>
<td>(1.28)</td>
<td>(1.42)</td>
<td>(1.20)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 6**

100*MISE (with 100*std in parenthesis below) for the estimation of $b_i$, $i = 1, \ldots, 4$, 200 repetitions, $X \sim N(0, 1)$, $\sigma = 0.7$. CV and PCO are the two competing methods. Column "Or" gives the average of ISE for the ratio of the two best estimators of $b_f$ and $f$ in the collection.
and for a fixed \( h > 0 \), by the law of large numbers,

\[
\frac{1}{nh} \sum_{i=1}^{n} K\left( \frac{X_i - x}{h} \right)^2 \xrightarrow{\text{a.s.}} \frac{1}{h} \mathbb{E} \left[ K\left( \frac{X_1 - x}{h} \right)^2 \right] = \int_{-\infty}^{\infty} K(u)^2 f(x + uh) du
\]

and

\[
\left[ \frac{1}{nh} \sum_{i=1}^{n} K\left( \frac{X_i - x}{h} \right) \right]^2 \xrightarrow{\text{a.s.}} \mathbb{E} \left[ \frac{1}{h} K\left( \frac{X_1 - x}{h} \right) \right]^2 = \left[ \int_{-\infty}^{\infty} K(u) f(x + uh) du \right]^2.
\]

Then, for small \( h \), the first limit has order \( \| K \|^2 f(x) \) and the second one has order \( f^2(x) \). To sum up, the risk of \( \hat{b}_{n,h}(x) \) is heuristically of order \( Ch^2 + \sigma^2 \| K \|^2 f(x)/(nh) \). This explains why, for small \( \sigma^2 \), the variance term gets small and the estimator can choose small bandwidth to make the bias as small as possible.

---

**Fig 4.** Example of 25 estimated \( b \) with large noise \( \sigma = 0.7 \), \( n = 1000 \), true \( b \) in bold red, estimated \( b \) with sign-bandwidth CV selection (dotted green, left) and with ratio of PCO (dotted green, right) for functions 1 and 3. 100*MSE(100 std) are 2.62(1.64) and 2.87(2.41) (top); 4.71(1.74) and 5.49(2.85).
6. Proofs

6.1. Proof of Proposition 2.4

On the one hand, by Comte [4], Proposition 3.3.1,
\[
\mathbb{E}(\|\hat{f}_{n,h'} - f\|_2^2) \leq \|f - f_{h'}\|_2^2 + \frac{c_K}{nh'}(4)
\]
and, by Proposition 2.2,
\[
\mathbb{E}(\|\hat{b}_f_{n,h} - b_f\|_2^2) \leq \|b_f - (b_f)_{h}\|_2^2 + \frac{c_{K,Y}}{nh}(5)
\]
For the proof of Inequality (4), the reader can also refer to Tsybakov [19]. On the other hand,
\[
\hat{b}_{n,h,h'} - b = \left(\frac{\hat{b}_f_{n,h} - b_f}{\hat{f}_{n,h'}} + b_f \left(\frac{1}{\hat{f}_{n,h'}} - \frac{1}{f}\right)\right) \mathbf{1}_{|\hat{f}_{n,h'}(\cdot)| > m/n} - b \mathbf{1}_{|\hat{f}_{n,h'}(\cdot)| \leq m/n}.
\]
Then,
\[
\|\hat{b}_{n,h,h'} - b\|_{2,f,S_n}^2 \leq \frac{8\epsilon_1}{m_n^2} \left(\|\hat{f}_{n,h} - b_f\|_2^2 + \int_{-\infty}^{\infty} b(x)^2 f(x) |\hat{f}_{n,h'}(x) - f(x)|^2 dx\right)
\]
\[
+ 2 \int_{S_n} b(x)^2 f(x) \mathbf{1}_{|\hat{f}_{n,h'}(x) - f(x)| > m/n} dx
\]
with \(\epsilon_1 := \|f\|_\infty \vee \|f\|_2^2\).

By Markov’s inequality,
\[
\mathbb{E}(\|\hat{b}_{n,h,h'} - b\|_{2,f,S_n}^2) \leq \frac{8\epsilon_1}{m_n^2} \left(\mathbb{E}(\|\hat{f}_{n,h} - b_f\|_2^2) + \epsilon_{b,f} \mathbb{E}(\|\hat{f}_{n,h'} - f\|_2^2)\right)
\]
\[
+ 2\epsilon_{b,f} \int_{S_n} \mathbb{P}(|\hat{f}_{n,h'}(x) - f(x)| > m/n^2) dx
\]
\[
\leq \frac{8(\epsilon_1 \vee 1)}{m_n^2} (\mathbb{E}(\|\hat{f}_{n,h} - b_f\|_2^2) + 2\epsilon_{b,f} \mathbb{E}(\|\hat{f}_{n,h'} - f\|_2^2)).
\]
Inequalities (4) and (5) allow to conclude.

6.2. Proof of Theorem 3.2

First, let us prove the following lemma.

Lemma 6.1. Consider
\[
(b_f)_{h} := K_h * (b_f) \quad \text{and} \quad (b_f)_{h,\eta} := K_\eta * K_h * (b_f).
\]
Then,
\[
\mathbb{E}(\hat{b}_f_{n,h}(x)) = (b_f)_{h}(x) \quad \text{and} \quad \mathbb{E}(\hat{b}_f_{n,h,\eta}(x)) = (b_f)_{h,\eta}(x).
\]
Proof. Since $\mathbb{E}(\varepsilon_k) = 0$ and $X_k$ and $\varepsilon_k$ are independent for every $k \in \{1, \ldots, n\}$,
\[
\mathbb{E}(\hat{b}f_{n,h}(x)) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}(b(X_k)K_h(x_k - x)) + \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}(\varepsilon_k)\mathbb{E}(K_h(x_k - x))
\]
\[
= \int_{-\infty}^{\infty} K_h(y - x)b(y)f(y)dy = (bf)_h(x)
\]
and
\[
\mathbb{E}(\hat{b}f_{n,h,\eta}(x)) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}(b(X_k)(K_{\eta} * K_h)(x_k - x))
\]
\[
+ \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}(\varepsilon_k)\mathbb{E}((K_{\eta} * K_h)(x_k - x))
\]
\[
= \int_{-\infty}^{\infty} (K_{\eta} * K_h)(y - x)b(y)f(y)dy = (bf)_{h,\eta}(x).
\]

Since
\[
\hat{h}_n \in \arg \min_{h \in \mathcal{H}_n} \{A_n(h) + V_n(h)\},
\]
for every $h \in \mathcal{H}_n$,
\[
\mathbb{E}(\|\hat{b}f_{n,\hat{h}_n} - bf\|_2^2) \leq 3\mathbb{E}(\|\hat{b}f_{n,h} - bf\|_2^2) + 6V_n(h) + 6\mathbb{E}(A_n(h)) \tag{6}
\]
Let us find a suitable control of $\mathbb{E}(A_n(h))$. First of all, for any $h, \eta \in \mathcal{H}_n$,
\[
\|\hat{b}f_{n,h,\eta} - \hat{b}f_{n,\eta}\|_2^2 \leq 3\|\hat{b}f_{n,h,\eta} - (bf)_{h,\eta}\|_2^2
\]
\[
+ \|\hat{b}f_{n,\eta} - (bf)_{\eta}\|_2^2 + \|(bf)_{h,\eta} - (bf)_{\eta}\|_2^2.
\]
Then,
\[
A_n(h) \leq 3 \left[ \sup_{\eta \in \mathcal{H}_n} \left( \|\hat{b}f_{n,h,\eta} - (bf)_{h,\eta}\|_2^2 - \frac{V_n(\eta)}{6} \right) \right. + \left. \sup_{\eta \in \mathcal{H}_n} \left( \|\hat{b}f_{n,\eta} - (bf)_{\eta}\|_2^2 - \frac{V_n(\eta)}{6} \right) + \left. \|(bf)_{h,\eta} - (bf)_{\eta}\|_2^2 \right]. \tag{7}
\]
On the one hand,
\[
\|(bf)_{h,\eta} - (bf)_{\eta}\|_2 = \|K_{\eta} * (K_h * (bf) - bf)\|_2 \leq \|K\|_1 \|bf - (bf)_h\|_2.
\]
On the other hand, let $\mathcal{C}$ be a countable and dense subset of the unit sphere of $L^2(\mathbb{R}, dx)$ and consider $m(n) > 0$. Then, by Lemma 6.1,
\[
\mathbb{E} \left[ \sup_{\eta \in \mathcal{H}_n} \left( \|\hat{b}f_{n,\eta} - (bf)_{\eta}\|_2^2 - \frac{V_n(\eta)}{6} \right) \right] \leq \sum_{\eta \in \mathcal{H}_n} \mathbb{E} \left( \left( \sup_{\psi \in \mathcal{C}} 2V_{n,\eta}(\psi) - \frac{V_n(\eta)}{6} \right) \right) + 2 \sum_{\eta \in \mathcal{H}_n} \mathbb{E}(\mathbf{W}_{n,\eta})
\]
where, for any $\psi \in C$,

$$V_{n, \eta}(\psi) := \frac{1}{n} \sum_{k=1}^{n} (v_{\psi, n, \eta}(X_k, Y_k) - \mathbb{E}(v_{\psi, n, \eta}(X_k, Y_k)))$$

with

$$v_{\psi, n, \eta}(x, y) := y \mathbf{1}_{|y| \leq m(n)} \int_{-\infty}^{\infty} \psi(u) K_\eta(x - u) du; \quad \forall (x, y) \in \mathbb{R}^2,$$

and

$$W_{n, \eta} := \frac{1}{n} \sum_{k=1}^{n} \int_{-\infty}^{\infty} |Y_k \mathbf{1}_{|Y_k| > m(n)} K_\eta(X_k - u) - \mathbb{E}(Y_k \mathbf{1}_{|Y_k| > m(n)} K_\eta(X_k - u))|^2 du.$$ 

In order to apply Talagrand’s inequality (see Klein and Rio [12]), we compute bounds.

- For every $\psi \in C$, $x \in \mathbb{R}$ and $y \in [-m(n), m(n)]$,

  $$|v_{\psi, n, \eta}(x, y)| \leq |y| \int_{-\infty}^{\infty} |\psi(u)| \cdot |K_\eta(u - x)| du$$

  $$\leq |y| \cdot \|K_\eta(\cdot - x)\|_2 \leq \frac{m(n)\|K\|_2}{\sqrt{\eta}}.$$ 

  Then,

  $$\sup_{\psi \in C} \|v_{\psi, n, \eta}\|_\infty \leq m_1(n, \eta) := \frac{m(n)\|K\|_2}{\sqrt{\eta}}.$$

- By Proposition 2.4 and Lemma 6.1,

  $$\mathbb{E}\left(\sup_{\psi \in C} V_{n, \eta}(\psi)^2\right) \leq \int_{-\infty}^{\infty} \text{var}(\hat{b}f_{n, \eta}(u)) du$$

  $$\leq m_2(n, \eta) := \frac{c_{K,Y}}{n\eta}.$$ 

- For any $\psi \in C$ and $k \in \{1, \ldots, n\},$

  $$\text{var}(v_{\psi, n, \eta}(X_k, Y_k))$$

  $$\leq \mathbb{E}\left(\left|Y_k \int_{-\infty}^{\infty} \psi(u) K_\eta(X_k - u) du\right|^2\right)$$

  $$\leq \mathbb{E}(g_\eta * \psi(X_1)^4)^{1/2} \mathbb{E}(Y_1^4)^{1/2} \leq ||f||_{L_\infty}^{1/2} ||K_\eta * \psi||_2^2 \mathbb{E}(Y_1)^{1/2}.$$ 

  By Young’s inequality, $||K_\eta * \psi||_4 \leq ||\psi||_2 ||K_\eta||_{4/3}$. So,

  $$\text{var}(v_{\psi, n, \eta}(X_k, Y_k)) \leq m_3 := \frac{m_{f,K}}{\sqrt{\eta}}$$

  with $m_{f,K} := ||f||_{L_\infty}^{1/2} ||K||_{4/3}^2 \mathbb{E}(Y_1)^{1/2}$. 

By applying Talagrand’s inequality to \((u_{\psi, n, \eta})_{\psi \in \mathcal{C}}\) and to the independent random variables \((X_1, Y_1), \ldots, (X_n, Y_n)\), there exist three constants \(c_1, c_2, c_3 > 0\), not depending on \(n\) and \(\eta\), such that

\[
\mathbb{E}\left( \left( \sup_{\psi \in \mathcal{C}} V_{n, \eta}(\psi)^2 - 4m_2(n, \eta) \right)_+ \right) \\
\leq c_1 \left( m_3 \frac{\exp \left( -c_2 \frac{m_2(n, \eta)}{m_3} \right)}{n^2} + m_1(n, \eta)^2 \frac{\exp \left( -c_3 \frac{nm_2(n, \eta)^{1/2}}{m_1(n, \eta)} \right)}{n^2} \right) \\
= c_1 \left[ \frac{m_{f,K}}{n^{3/2}} \exp \left( -\frac{c_2 \mathbb{E} \sup_{\eta \in H} \left[ |K|^2 \right]}{m_{f,K} \sqrt{n}} \right) + \frac{1}{n^{2\eta}} \mathbb{E} \left[ \mathbb{E} \left( Y_1^2 \right)^{1/2} \right] \right].
\]

By taking \(m(n) := c_3 \mathbb{E} \left( Y_1^2 \right)^{1/2} n^{1/2} / \log(n)^{1/2}\),

\[
\mathbb{E}\left( \left( \sup_{\psi \in \mathcal{C}} V_{n, \eta}(\psi)^2 - 4m_2(n, \eta) \right)_+ \right) \\
\leq c_1 \left[ \frac{m_{f,K}}{n^{3/2}} \exp \left( -\frac{c_2 \mathbb{E} \sup_{\eta \in H} \left[ |K|^2 \right]}{m_{f,K} \sqrt{n}} \right) + \frac{c_3^2 \mathbb{E} \left( Y_1^2 \right)^{1/2}}{n \log(n)} \right].
\]

By the conditional Markov inequality,

\[
\mathbb{E}(W_{n, \eta}) \leq \int_{-\infty}^{\infty} \mathbb{E} \left( Y_1^2 \mathbb{1}_{|Y_1| > m(n)} \right) K_\eta(X_1 - z)^2 dz \\
= \frac{\mathbb{E}(Y_1^2 \mathbb{1}_{|Y_1| > m(n)} Y_1)}{\eta} \\
\leq \frac{||K||^2}{\eta m(n)^2} \mathbb{E}(Y_1^6) = c_4^2 \mathbb{E}(Y_1^2)^{-2} \mathbb{E}(Y_1^6) ||K||^2 \frac{\log(n)^2}{n^2 \eta}.
\]

Finally, for \(v \geq 48\),

\[
\frac{V_n(\eta)}{12} \geq 4m_2(n, \eta).
\]

Then, since

\[
\frac{1}{n} \sum_{\eta \in \mathcal{H}_n} \frac{1}{\eta} \leq m, \quad \text{and} \quad \sum_{\eta \in \mathcal{H}_n} \frac{1}{\sqrt{\eta}} \exp \left( -\frac{c}{\sqrt{\eta}} \right) \leq m(c) ; \forall c > 0,
\]

there exists a constant \(c_4 > 0\), not depending on \(n\), such that

\[
\mathbb{E}\left[ \sup_{\eta \in \mathcal{H}_n} \left( ||\hat{b}_{f,n,\eta} - (bf)||_2^2 - \frac{V_n(\eta)}{6} \right)_+ \right] \\
\leq 2 \sum_{\eta \in \mathcal{H}_n} \mathbb{E}\left( \left( \sup_{\psi \in \mathcal{C}} V_{n, \eta}(\psi)^2 - 4m_2(n, \eta) \right)_+ \right) + 2 \sum_{\eta \in \mathcal{H}_n} \mathbb{E}(W_{n, \eta}) \\
\leq c_4 \frac{\log(n)^2}{n} \quad (8)
\]
The same ideas give that there exists a constant \( c_5 > 0 \), not depending on \( n \) and \( h \), such that

\[
E \left[ \sup_{\eta \in H_n} \left( \| \hat{f}_{n,h,\eta} - (bf)_{n,\eta} \|^2 - V_n(\eta) \right) \right] + \leq \frac{c_5 (\log(n))^2}{n} \quad (9)
\]

Therefore, by Inequalities (6)–(9), there exist two deterministic constants \( c, \tilde{c} > 0 \), not depending on \( n \), such that

\[
E(\| \hat{f}_{n,h_n} - bf \|^2) \leq c \cdot \inf_{h \in H_n} \{ \| (bf) - bf \|^2 + V_n(h) \} + \frac{\log(n)^2}{n}.
\]

**6.3. Proof of Corollary 3.3**

As established in the proof of Proposition 2.4,

\[
\| \hat{f}_{n,h_n} - bf \|^2_{f,S_n} \leq \frac{8c_1}{m_n^2} \left( \| \hat{f}_{n,h_n} - bf \|^2 + \epsilon_{b,f} \int_{-\infty}^{\infty} \| \hat{f}_{n,h_n}(x) - f(x) \|^2 dx \right)
+ 2\epsilon_{b,f} \int_{S_n} P(\| \hat{f}_{n,h_n}(x) - f(x) \| > m_n/2) dx
\]

with \( c_1 := \| f \|_\infty \vee \| f \|_\infty^2 \). By Markov’s inequality,

\[
E(\| \hat{f}_{n,h_n} - bf \|^2_{f,S_n}) \leq \frac{8c_1}{m_n^2} \left( E(\| \hat{f}_{n,h_n} - bf \|^2) + \epsilon_{b,f} E(\| \hat{f}_{n,h_n} - f \|^2) \right)
+ 2\epsilon_{b,f} \int_{S_n} P(\| \hat{f}_{n,h_n}(x) - f(x) \| > m_n/2) dx
\leq \frac{8(c_1 \vee 1)}{m_n^2} \left( E(\| \hat{f}_{n,h_n} - bf \|^2) + 2\epsilon_{b,f} E(\| \hat{f}_{n,h_n} - f \|^2) \right)
\]

Theorem 3.2 and Inequality (2) allow to conclude.

**6.4. Proof of Theorem 4.2**

The proof relies on three lemmas, which are stated first.

**Lemma 6.2.** Consider the U-statistic

\[
U_n(h, h_{\min}) := \sum_{k \neq l} \{ Y_kK_h(X_k - \cdot) - (bf)_{h}, Y_lK_{h_{\min}}(X_l - \cdot) - (bf)_{h_{\min}} \}.
\]
Under Assumption 2.1, if there exists $\alpha > 0$ such that $\mathbb{E}(\exp(\alpha |Y_1|)) < \infty$, then there exists a deterministic constant $c_U > 0$, not depending on $n$ and $h_{\min}$, such that for every $\vartheta \in (0,1)$,

$$
\mathbb{E} \left( \sup_{h \in \mathcal{H}_n} \left\{ \frac{|U_n(h, h_{\min})|}{n^2} - \vartheta \|K\|^2 \mathbb{E}(Y_1^2) \right\} \right) \leq c_U \frac{\log(n)^5}{\vartheta n}.
$$

Lemma 6.3. For every $\eta, \eta' \in \mathcal{H}_n$, consider

$$
V_n(\eta, \eta') := \langle \hat{b}f_{n, \eta} - (bf)_{\eta'}, (bf)_{\eta'} - bf \rangle_2.
$$

Under Assumption 2.1, if there exists $\alpha > 0$ such that $\mathbb{E}(\exp(\alpha |Y_1|)) < \infty$ and $bf$ is bounded, then there exists a deterministic constant $c_V > 0$, not depending on $n$ and $h_{\min}$, such that for every $\vartheta \in (0,1)$,

$$
\mathbb{E} \left( \sup_{\eta, \eta' \in \mathcal{H}_n} (|V_n(\eta, \eta')| - \vartheta \|bf\|_2^2) \right) \leq c_V \frac{\log(n)^3}{\vartheta n}.
$$

Lemma 6.4. Under Assumption 2.1, if $bf$ is bounded and if there exists $\alpha > 0$ such that $\mathbb{E}(\exp(\alpha |Y_1|)) < \infty$, then there exists a deterministic constant $c_L > 0$, not depending on $n$ and $h_{\min}$, such that for every $\vartheta \in (0,1)$,

$$
\mathbb{E} \left( \sup_{h \in \mathcal{H}_n} \left\{ \|bf\h_n - bf\|_2^2 + \frac{c_KY}{n} - \frac{1}{1 - \vartheta} \|\hat{b}f_{n,h} - bf\|_2^2 \right\} \right) \leq \frac{c_L}{\vartheta(1 - \vartheta)} \cdot \frac{\log(n)^5}{n}.
$$

6.4.1. Steps of the proof.

The proof of Theorem 4.2 is dissected in three steps.

Step 1. In this step, a suitable decomposition of

$$
\|\hat{b}f_{n, \tilde{h}_n} - bf\|_2^2
$$

is provided. On the one hand,

$$
\|\hat{b}f_{n, \tilde{h}_n} - bf\|_2^2 + \text{pen} (\tilde{h}_n)
= \|\hat{b}f_{n,h_n} - \hat{b}f_{n,h_{\min}}\|_2^2 + \text{pen}(h_n)
+ \|\hat{b}f_{n,h_{\min}} - bf\|_2^2 - 2 \langle \hat{b}f_{n, h_{\min}} - bf, \hat{b}f_{n, h_{\min}} - bf \rangle_2.
$$

Since

$$
\tilde{h}_n \in \arg \min_{h \in \mathcal{H}_n} \text{crit}(h) \quad \text{with} \quad \text{crit}(h) = \|\hat{b}f_{n,h} - \hat{b}f_{n,h_{\min}}\|_2^2 + \text{pen}(h),
$$

for any $h \in \mathcal{H}_n$,

$$
\|\hat{b}f_{n, \tilde{h}_n} - bf\|_2^2 \leq \|\hat{b}f_{n,h} - bf\|_2^2 + \text{pen}(h) - 2\psi_n(h) - (\text{pen}(\tilde{h}_n) - 2\psi_n(\tilde{h}_n)).
$$

(10)
with
\[
\psi_n(h) := (\hat{f}_{n,h_{\text{min}}} - f, \hat{f}_{n,h} - f)_{\mathbb{L}^2}.
\]

On the other hand,
\[
\psi_n(h) = \psi_{1,n}(h) + \psi_{2,n}(h) + \psi_{3,n}(h)
\]
where
\[
\psi_{1,n}(h) := \frac{(K_{h_{\text{min}}}, K_h)_{\mathbb{L}^2}}{n^2} \sum_{k=1}^{n} Y_k^2 + \frac{U_n(h, h_{\text{min}})}{n^2},
\]
\[
\psi_{2,n}(h) := - \frac{1}{n^2} \left( \sum_{k=1}^{n} Y_k (K_{h_{\text{min}}}(X_k - \cdot), (bf)_h)_{\mathbb{L}^2} + \frac{1}{n} \langle (bf)_{h_{\text{min}}}, (bf)_h \rangle_{\mathbb{L}^2} \right) + \frac{1}{n} \langle (bf)_h - f, (bf)_{h_{\text{min}}} - f \rangle_{\mathbb{L}^2}
\]
\[
\psi_{3,n}(h) := V_n(h, h_{\text{min}}) + V_n(h_{\text{min}}, h) + \langle (bf)_h - f, (bf)_{h_{\text{min}}} - f \rangle_{\mathbb{L}^2}.
\]

**Step 2.** In this step, let us provide some suitable controls of
\[\mathbb{E}(\psi_{i,n}(h))\] and \[\mathbb{E}(\psi_{i,n}(\hat{h}_n))\] ; \(i = 1, 2, 3\).

1. Consider
\[
\bar{\psi}_{1,n}(h) := \psi_{1,n}(h) - \frac{(K_{h_{\text{min}}}, K_h)_{\mathbb{L}^2}}{n^2} \sum_{k=1}^{n} Y_k^2 = \frac{U_n(h, h_{\text{min}})}{n^2}.
\]
By Lemma 6.2,
\[
\mathbb{E}(\bar{\psi}_{1,n}(h)) \leq \frac{\theta \|K\|_{\mathbb{L}^2}^2}{n} \mathbb{E}(Y_1^2) + \frac{2\varepsilon U}{\theta} \frac{\log(n)^{5}}{n}.
\]
and
\[
\mathbb{E}(\bar{\psi}_{1,n}(\hat{h}_n)) \leq \mathbb{E}(\bar{\psi}_{1,n}(h)) \leq \mathbb{E}(\bar{\psi}_{1,n}(h)) + \frac{\theta \|K\|_{\mathbb{L}^2}^2}{n} \mathbb{E}(Y_1^2) + \frac{2\varepsilon U}{\theta} \frac{\log(n)^{5}}{n}.
\]

2. On the one hand, for every \(\eta, \eta' \in \mathcal{H}_n\), consider
\[
\Psi_{2,n}(\eta, \eta') := \frac{1}{n} \sum_{k=1}^{n} Y_k \langle K_\eta(X_k - \cdot), (bf)_{\eta'} \rangle_{\mathbb{L}^2}.
\]
Then,
\[
\mathbb{E} \left( \sup_{\eta, \eta' \in \mathcal{H}_n} |\Psi_{2,n}(\eta, \eta')| \right) \leq \mathbb{E} \left( |Y_1| \sup_{\eta, \eta' \in \mathcal{H}_n} \int_{\mathbb{R}} |K_\eta(X_1 - u)(bf)_{\eta'}(u)| du \right) \leq \mathbb{E}(Y_1^2)^{1/2} \|K\|_{\mathbb{L}^2} \|bf\|_{\mathbb{L}^2}.
On the other hand,
\[
\sup_{\eta, \eta' \in \mathcal{H}_n} |\langle (bf)_{\eta}, (bf)_{\eta'} \rangle | \leq \sup_{\eta, \eta' \in \mathcal{H}_n} \|K_{\eta} \ast (bf)\|_2 \|K_{\eta'} \ast (bf)\|_2 \\
\leq |K_{\eta}^2 \|bf\|_2^2 \leq \mathbb{E}(Y_1^2)^{1/2} \|K_{\eta}^2 \|bf\|_\infty.
\]
Then,
\[
\mathbb{E}(|\psi_{2,n}(h)|) \leq \frac{3}{n} \mathbb{E}(Y_1^2)^{1/2} \|K_{\eta}^2 \|bf\|_\infty
\]
and
\[
\mathbb{E}(|\psi_{2,n}(\tilde{h}_n)|) \leq \frac{3}{n} \mathbb{E}(Y_1^2)^{1/2} \|K_{\eta}^2 \|bf\|_\infty.
\]

3. By Lemma 6.3,
\[
\mathbb{E}(|\psi_{n,3}(h)|) \leq \theta \left( \|\langle (bf)_{\eta} \rangle & - bf \|_2^2 + \|\langle (bf)_{h_{\min}} \rangle \|bf \|_2^2 \right) + 4c_V \frac{\log(n)^3}{\theta n}
\]
\[
+ \left( \frac{\theta}{2} \right)^{1/2} \|\langle (bf)_{\eta} \rangle \|bf \|_2 \times \left( \frac{2}{\theta} \right)^{1/2} \|\langle (bf)_{h_{\min}} \rangle \|bf \|_2
\]
\[
\leq \theta \|\langle (bf)_{\eta} \rangle \|bf \|_2^2 + \left( \frac{\theta}{2} \right)^{1/2} \|\langle (bf)_{h_{\min}} \rangle \|bf \|_2^2
\]
\[
\quad + 4c_V \frac{\log(n)^3}{\theta n}
\]
and
\[
\mathbb{E}(|\psi_{n,3}(\tilde{h}_n)|) \leq \theta \mathbb{E}(\|\langle (bf)_{\tilde{h}_n} \rangle \|bf \|_2^2)
\]
\[
+ \left( \frac{\theta}{2} + \frac{2}{\theta} \right) \|\langle (bf)_{h_{\min}} \rangle \|bf \|_2^2 + 4c_V \frac{\log(n)^3}{\theta n}.
\]

**Step 3.** Consider
\[
\tilde{\psi}_n(h) := \psi_n(h) - \frac{\langle K_{h_{\min}}, K_{\eta} \rangle}{n^2} \sum_{k=1}^n Y_k^2.
\]
By Step 2, there exists a deterministic constant $c_{U,V} > 0$, not depending on $n$, $h$ and $h_{\min}$, such that
\[
\mathbb{E}(|\tilde{\psi}_n(h)|) \leq \theta \left( \|\langle (bf)_{\eta} \rangle \|bf \|_2^2 + \frac{\xi_{K,Y}}{n\theta h} \right)
\]
\[
+ \frac{c_{U,V}}{\theta} \cdot \frac{\log(n)^5}{n} + \left( \frac{\theta}{2} + \frac{2}{\theta} \right) \|\langle (bf)_{h_{\min}} \rangle \|bf \|_2^2
\]
and
\[
\mathbb{E}(|\tilde{\psi}_n(\tilde{h}_n)|) \leq \theta \left[ \mathbb{E}(\|\langle (bf)_{\tilde{h}_n} \rangle \|bf \|_2^2) + \mathbb{E} \left( \frac{\xi_{K,Y}}{nH_n} \right) \right]
\]
\[
+ \frac{c_{U,V}}{\theta} \cdot \frac{\log(n)^5}{n} + \left( \frac{\theta}{2} + \frac{2}{\theta} \right) \|\langle (bf)_{h_{\min}} \rangle \|bf \|_2^2.
\]
Then, by Lemma 6.4,
\[
\mathbb{E}(|\tilde{\psi}_n(h)|) \leq \frac{\theta}{1 - \theta} \mathbb{E}((\tilde{bf}_{n,h} - bf)^2) + \left(\frac{\theta}{2} + \frac{2}{\theta}\right) \|bf\|_{h_{\text{min}}} - bf^2 + \left(\frac{c_{U,V}}{\theta} + \frac{c_L}{1 - \theta}\right) \log(n)^5
\]
and
\[
\mathbb{E}(|\tilde{\psi}_n(h_n)|) \leq \frac{\theta}{1 - \theta} \mathbb{E}((\tilde{bf}_{n,h_n} - bf)^2) + \left(\frac{\theta}{2} + \frac{2}{\theta}\right) \|bf\|_{h_{\text{min}}} - bf^2 + \left(\frac{c_{U,V}}{\theta} + \frac{c_L}{1 - \theta}\right) \log(n)^5.
\]
By Inequality (10), there exist two deterministic constant \(c_1, c_2 > 0\), not depending on \(n, h\) and \(h_{\text{min}}\), such that
\[
\mathbb{E}((\tilde{bf}_{n,h_n} - bf)^2) \leq \mathbb{E}((\tilde{bf}_{n,h} - bf)^2) + 2(\mathbb{E}(|\tilde{\psi}_n(h)|) + \mathbb{E}(|\tilde{\psi}_n(h_n)|))
\leq \left(1 + \frac{2\theta}{1 - \theta}\right) \mathbb{E}((\tilde{bf}_{n,h} - bf)^2)
+ \frac{2\theta}{1 - \theta} \mathbb{E}((\tilde{bf}_{n,h_n} - bf)^2)
+ \frac{c_1}{\theta} \|bf\|_{h_{\text{min}}} - bf^2 + \frac{c_2}{\theta(1 - \theta)} \cdot \log(n)^5.
\]
This concludes the proof.

6.4.2. Proof of Lemma 6.2
Consider
\[
\Delta_n := \{(k, l) \in \{1, \ldots, n\} : 2 \leq k \text{ and } l < k\}
\]
and \(Z_k := (X_k, Y_k)\) for every \(k \in \{1, \ldots, n\}\).

On the one hand, consider \(n \in \mathbb{N}\) such that \(m(n) := 4 \log(n)/\alpha \geq 1\) and
\[
U_{1,n}(h, h_{\text{min}}) := \sum_{k=2}^{n} \sum_{l<k} (G_{n,h,h_{\text{min}}}(Z_k, Z_l) + G_{n,h_{\text{min}},h}(Z_k, Z_l))
\]
where, for every \(\eta, \eta' \in \{h, h_{\text{min}}\}\) and \(z, z' \in \mathbb{R}^2\),
\[
G_{n,\eta,\eta'}(z, z') :=
\langle z \mathbf{1}_{|z| \leq m(n)} K_\eta(z_1 - \cdot) - (bf)_{n,\eta}, z' \mathbf{1}_{|z'| \leq m(n)} K_{\eta'}(z'_1 - \cdot) - (bf)_{n,\eta'} \rangle_2
\]
and
\[
(bf)_{n,\eta} := \mathbb{E}(Y_1 \mathbf{1}_{|Y_1| \leq m(n)} K_\eta(X_1 - \cdot)).
\]
For every $\eta, \eta' \in \{h, h_{\min}\}$ and $(k, l) \in \Delta_n$,
\[
E(G_{n, \eta, \eta'}(Z_k, Z_l) | Z_k) = \int_{-\infty}^{\infty} (Y_k 1_{|Y_k| \leq m(n)} K_{\eta}(X_k - z) - (bf)_{n, \eta}(z)) \times E(Y_l 1_{|Y_l| \leq m(n)} K_{\eta'}(X_l - z) - (bf)_{n, \eta'}(z)) \, dz = 0.
\]
So, by Houdré and Reynaud-Bouret [10], Theorem 3.4, there exists a universal sequel.
\[
P(|U_{1,n}(h, h_{\min})| \geq c_n \lambda^{1/2} + \sigma_n \lambda + b_n \lambda^{3/2} + a_n \lambda^2) \leq 5.54 e^{-\lambda}
\]
where the constants $a_n$, $b_n$, $c_n$ and $\sigma_n$ will be defined and controlled in the sequel.

- **The constant $a_n$.** Consider
\[
a_n := \sup_{(z, z') \in \mathbb{R}^2 \times \mathbb{R}^2} A_n(z, z'),
\]
where
\[
A_n(z, z') := |G_{n, h, h_{\min}}(z, z') + G_{n, h_{\min}, h}(z, z')| ; \forall z, z' \in \mathbb{R}^2.
\]
First, note that for every $\eta \in \mathcal{H}_n$,
\[
\|(bf)_{n, \eta}\|_1 \leq E(|Y_1 1_{|Y_1| \leq m(n)}| \|K\|_1 \leq m(n) \|K\|_1
\]
and
\[
\|(bf)_{n, \eta}\|_{\infty} \leq \frac{m(n) \|K\|_{\infty}}{\eta}.
\]
For any $z, z' \in \mathbb{R} \times [-m(n), m(n)]$,
\[
A_n(z, z') \leq \langle z_2 K_h(z_1 - :) - (bf)_{n, h}, z_2' K_{h_{\min}}(z_1' - :) - (bf)_{n, h_{\min}} \rangle + \langle z_2 K_{h_{\min}}(z_1 - :) - (bf)_{n, h_{\min}}, z_2' K_{h}(z_1' - :) - (bf)_{n, h} \rangle
\]
\[
\leq 2(m(n) \|K\|_{\infty}) + (bf)_{n, h_{\min}} \|\|_{\infty} \times (m(n) \|K\|_1 + \|(bf)_{n, h}\|_1)
\]
\[
\leq \frac{8 \|K\| \|K\|_{\infty} m(n)^2}{h_{\min}}.
\]
Therefore,
\[
\frac{a_n \lambda^2}{n^2} \leq \frac{8 \|K\|_1 \|K\|_{\infty} m(n)^2 \lambda^2}{n^2 h_{\min}}.
\]
- **The constant $b_n$.** Consider
\[
b_n^2 := n \max \left\{ \sup_{z \in \mathbb{R}^2} E(G_{n, h, h_{\min}}(z, Z_1)^2) ; \sup_{z \in \mathbb{R}^2} E(G_{n, h_{\min}, h}(z, Z_1)^2) \right\}.
\]
First, note that for every $\eta \in \mathcal{H}_n$,
\[
\|(bf)_{n, \eta}\|_2^2 \leq \frac{m(n)^2 \|K\|_2^2}{\eta}.
\]
For any $\eta, \eta' \in \{h, h_{\text{min}}\}$ and $z \in \mathbb{R} \times [-m(n), m(n)]$,
\[
\mathbb{E}(G_{n,\eta,\eta'}(z, Z_1)^2) \leq \|z_2 K_\eta(z_1 - \cdot) - (bf)_{n,\eta}\|_2^2 \\
\times \int_{-\infty}^{\infty} \mathbb{E}(\{Y_1 1|Y_1| \leq m(n)\} K_{\eta'}(X_1 - u) - (bf)_{n,\eta'}(u))^2) \, du \\
\leq \frac{2\|K\|_{2}^2 m(n)^2}{\eta} \int_{-\infty}^{\infty} \mathbb{v}(\{Y_1 1|Y_1| \leq m(n)\} K_{\eta'}(X_1 - u)) \, du \\
\leq \frac{2\|K\|_{2}^2}{\eta\eta'} \mathbb{E}(Y_1^2) m(n)^2.
\]
Therefore, for any $\theta \in (0, 1)$,
\[
\frac{b_n \lambda_{3/2}}{n^2} \leq \sqrt{\frac{2}{n^{3/2} (hh_{\text{min}})^{1/2}}} \mathbb{E}(Y_1^2)^{1/2} m(n) \lambda_{3/2} \\
\leq 2 \left( \frac{3e}{\theta} \right)^{1/2} \frac{\|K\|_2^2 n^{2} m(n) \lambda_{3/2}^2}{\theta h_{\text{min}}^{1/2}} \frac{\|K\|_2}{n^{3/2} h^{1/2}} \mathbb{E}(Y_1^2)^{1/2} \\
\leq \frac{3e \|K\|_2^2 \mathbb{E}(Y_1^2) m(n)^2 \lambda_{3}^2}{\theta h_{\text{min}}^2} + \frac{\theta \|K\|_2^2 \mathbb{E}(Y_1^2)}{3e n h}.
\]
\textbf{The constant $c_n$. Consider}
\[
c_n^2 := \sum_{(k,l) \in \Delta_n} \mathbb{E}(|G_{n,h,h_{\text{min}}}(Z_k, Z_l) + G_{n,h_{\text{min}},h}(Z_k, Z_l)|^2).
\]
First, note that for every $\eta \in \mathcal{H}_n$,
\[
\|(bf)_{n,\eta}\|_\infty \leq m(n) \|f\|_\infty \|K\|_1.
\]
For any $\eta, \eta' \in \{h, h_{\text{min}}\}$ and $(k,l) \in \Delta_n$,
\[
\mathbb{E}(G_{n,\eta,\eta'}(Z_k, Z_l)^2) \leq 4(m(n))^2 \mathbb{E}((K_\eta(X_k - \cdot) - K_{\eta'}(X_l - \cdot))^2 Y_1^2) \\
+ \|(bf)_{n,\eta}\|_\infty^2 \mathbb{E}(Y_1^2) \|K_{\eta'}(X_l - \cdot)\|_1^2 \\
+ \|(bf)_{n,\eta'}\|_\infty^2 \mathbb{E}(Y_1^2) \|K_{\eta'}(X_l - \cdot)\|_1^2 \\
+ \|(bf)_{n,\eta'}\|_\infty^2 \|(bf)_{n,\eta'}\|_1^2 \\
\leq 4(m(n))^2 (\mathbb{E}((K_\eta(X_k - \cdot) - K_{\eta'}(X_l - \cdot))^2 Y_1^2) \\
+ 3\|f\|_\infty^2 \|K\|_1^2 \mathbb{E}(Y_1^2)).
\]
Moreover,
\[
\mathbb{E}((K_\eta(X_k - \cdot) - K_{\eta'}(X_l - \cdot))^2 Y_1^2) = \sigma^2 \mathbb{E}((K_{\eta} * K_{\eta'})(X_k - X_l)^2) \\
+ \mathbb{E}((K_{\eta} * K_{\eta'})(X_k - X_l)^2 \sigma^2) \\
\leq \sigma^2 \|f\|_\infty \|K_{\eta} * K_{\eta'}\|_2^2 \\
+ \|f\|_\infty \mathbb{E}(b(X_1)^2) \|K_{\eta} * K_{\eta'}\|_2^2 \\
\leq \frac{\|f\|_\infty \|K\|_2^2 \|K\|_2^2 \mathbb{E}(Y_1^2)}{\eta}.
\]
Then, there exists a universal constant $c_1 > 0$ such that
\[ c_1 n^2 \|f\|_\infty^2 \|K\|_2^2 m(n) \mathbb{E}(Y_1^2) \left( \frac{\|K\|_2^2}{h} + 3\|f\|_\infty \|K\|_1^2 \right). \]

Therefore, since $m(n)$ is larger than 1, there exists a universal constant $c_2 > 0$ such that
\[ \frac{c_2 n^{1/2}}{n^2} \leq \frac{c_1}{3\pi n} \|f\|_\infty \|K\|_1^2 m(n)^2 (\lambda^{1/2} + \lambda). \]

- **The constant $d_n$.** Consider
\[ d_n := \sup_{(\alpha, \beta) \in S} \sum_{(k, l) \in \Delta_n} \mathbb{E}((G_{h, h_{\min}}(Z_k, Z_l) + G_{h_{\min}, h}(Z_k, Z_l)) \alpha_k(Z_k) \beta_l(Z_l)), \]

where
\[ S := \left\{ (\alpha, \beta) : \sum_{k=2}^n \mathbb{E}(\alpha_k(Z_k)^2) \leq 1 \text{ and } \sum_{l=1}^{n-1} \mathbb{E}(\beta_l(Z_l)^2) \leq 1 \right\}. \]

For any $(\alpha, \beta) \in S$,
\[ \sum_{(k, l) \in \Delta_n} \mathbb{E}(G_{h, h_{\min}}(Z_k, Z_l) \alpha_k(Z_k) \beta_l(Z_l)) \leq D_2(\alpha, \beta) \sup_{u \in \mathbb{R}} D_1(\alpha, \beta, u), \]

with, for every $u \in \mathbb{R}$,
\[ D_1(\alpha, \beta, u) := \sum_{k=2}^n \mathbb{E}(|\alpha_k(Z_k)(Y_k 1_{|Y_k| \leq m(n)} K_h(X_k - u) - (bf)_{n,h}(u))|) \]
\[ \leq \mathbb{E} \left[ \left( \sum_{k=2}^n \alpha_k(Z_k)^2 \right)^{1/2} \times \left( \sum_{k=2}^n |Y_k 1_{|Y_k| \leq m(n)} K_h(X_k - u) - (bf)_{n,h}(u)|^2 \right)^{1/2} \right] \]
\[ \leq \left( \sum_{k=2}^n \mathbb{E}(\alpha_k(Z_k)^2) \right)^{1/2} \times \left( \sum_{k=2}^n \mathbb{E}(Y_k^2 1_{|Y_k| \leq m(n)} K_h(X_k - u)^2) \right)^{1/2} \]
\[ \leq \frac{\|f\|_\infty^2 \|K\|_2^2}{h^{1/2}} n^{1/2} m(n). \]
and
\[
\mathbf{D}_2(\alpha, \beta) := \sum_{l=1}^{n-1} \mathbb{E} \left( |\beta_l(Z_l)| \int_{-\infty}^{\infty} |Y_l| |y| \leq m(n) k_{h_{min}}(X_l - u) - (bf)_{n,h_{min}}(u) |du\right)
\]
\[
\leq \sqrt{2} \left( \sum_{l=1}^{n-1} \mathbb{E}(\beta_l(Z_l)^2) \right)^{1/2}
\times \left( \sum_{l=1}^{n-1} \mathbb{E}(Y_l^2 \| K_{h_{min}}(X_l - \cdot) \|_1^2) + \|(bf)_{n,h_{min}}\|_1^2 \right)^{1/2}
\leq \sqrt{2} \cdot \|K\|_1 \mathbb{E}(Y_1^2)^{1/2} n^{-1/2}.
\]

Then,
\[
\mathcal{D}_n \leq 2n \frac{\|K\|_2 \|K\|_1 \|f\|_\infty^{1/2}}{h^{1/2}} \mathbb{E}(Y_1^2)^{1/2} m(n).
\]

Therefore,
\[
\frac{\mathcal{D}_n \lambda}{n^2} \leq 2 \times \left( \frac{\theta}{3\epsilon} \right)^{1/2} \frac{\|K\|_2^{1/2}}{(nh)^{1/2}} \mathbb{E}(Y_1^2)^{1/2} \times \left( \frac{3\epsilon}{\theta} \right)^{1/2} \frac{\|K\|_1 \|f\|_\infty^{1/2}}{n^{1/2}} m(n) \lambda
\leq \frac{\theta \|K\|_2^2}{3nh} \mathbb{E}(Y_1^2) + \frac{3\epsilon \|K\|_2^2 \|f\|_\infty}{\theta n} m(n)^2 \lambda^2.
\]

So, there exist two universal constants \(c_3, c_4 > 0\) such that, with probability larger than \(1 - 5.54e^{-\lambda}\),
\[
\frac{|U_{1,n}(h, h_{min})|}{n^2} \leq \frac{\theta \|K\|_2^2}{nh} \mathbb{E}(Y_1^2)
\leq \frac{c_3 \|K\|_1 \|K\|_\infty \lambda^2}{n^2 h_{min}} m(n)^2 \left( \frac{\lambda^3}{\theta} + \lambda^2 \right)
\leq \frac{\theta \|K\|_2^2}{3nh} \mathbb{E}(Y_1^2)
\leq \frac{c_3 \|K\|_1 \|K\|_\infty \lambda^2}{n^2 h_{min}} + \frac{\|f\|_\infty \|K\|_2^2}{n} m(n)^2 (1 + \lambda)^3.
\]

Then, with probability larger than \(1 - 5.54|\mathcal{H}_n|e^{-\lambda}\),
\[
S_n(h_{min}) \leq \frac{c_4}{\theta} \left( \frac{\|K\|_1 \|K\|_\infty}{n^2 h_{min}} + \frac{\|f\|_\infty \|K\|_2^2}{n} \right) m(n)^2 (1 + \lambda)^3
\]

where
\[
S_n(h_{min}) := \sup_{h \in \mathcal{H}_n} \left\{ \frac{|U_{1,n}(h, h_{min})|}{n^2} - \frac{\theta \|K\|_2^2}{nh} \mathbb{E}(Y_1^2) \right\}.
\]
For every $s \in \mathbb{R}_+$, consider
\[
\lambda(s) := -1 + \left( \frac{s}{m(n, h_{\min}, \theta)} \right)^{1/3},
\]
where
\[
m(n, h_{\min}, \theta) := c_4 \left( \frac{\|K\|_1}{n^2 h_{\min}} + \frac{\|f\|_\infty}{n} \right) m(n)^2.
\]
Then, for any $A > 0$,
\[
E(S_n(h_{\min})) \leq A + \int_{A}^{\infty} \mathbb{P}(S_n(h_{\min}) \geq s) ds
\]
\[
\leq A + 5.54 c_5 |H_n| m(n, h_{\min}, \theta) \exp \left( - \frac{A^{1/3}}{2m(n, h_{\min}, \theta)^{1/3}} \right)
\]
where
\[
c_5 := \int_{0}^{\infty} e^{-s^{1/3}/2} ds.
\]
Since there exists a deterministic constant $c_6 > 0$, not depending on $n$ and $h_{\min}$ such that
\[
m(n, h_{\min}, \theta) \leq \frac{\log(n)^2}{n},
\]
by taking $A := 2^3 c_6 \log(n)^5/n$,
\[
E(S_n(h_{\min})) \leq 2^3 c_6 \frac{\log(n)^5}{n} + 5.54 c_5 m(n, h_{\min}, \theta) \frac{|H_n|}{n}.
\]
Therefore, since $|H_n| \leq n$, there exists a deterministic constant $c_7 > 0$, not depending on $n$ and $h_{\min}$, such that
\[
E \left( \sup_{h \in \mathcal{H}_n} \left\{ \frac{|U_{1,n}(h, h_{\min})|}{n^2} - \frac{\theta \|K\|^2}{nh} E(Y_1^2) \right\} \right) \leq \frac{c_7}{\theta} \cdot \frac{\log(n)^5}{n}.
\]
On the other hand,
\[
U_n(h, h_{\min}) = \sum_{i=1}^{4} U_{i,n}(h, h_{\min})
\]
where, for $i = 2, 3, 4$,
\[
U_{i,n}(h, h_{\min}) := \sum_{k \neq l} g^i_{n,h,h_{\min}}(Z_k, Z_l)
\]
with
\[
g^2_{n,h,h_{\min}}(z, z') := (z_2 \mathbf{1}_{|z_2| \leq m(n)} K_h(z_1 - \cdot), z'_2 \mathbf{1}_{|z'_2| > m(n)} K_{h_{\min}}(z'_1 - \cdot))_2,
\]
\[
g^3_{n,h,h_{\min}}(z, z') := (z_2 \mathbf{1}_{|z_2| > m(n)} K_h(z_1 - \cdot), z'_2 \mathbf{1}_{|z'_2| \leq m(n)} K_{h_{\min}}(z'_1 - \cdot))_2 \text{ and}
\]
\[
g^4_{n,h,h_{\min}}(z, z') := (z_2 \mathbf{1}_{|z_2| > m(n)} K_h(z_1 - \cdot), z'_2 \mathbf{1}_{|z'_2| > m(n)} K_{h_{\min}}(z'_1 - \cdot))_2.
\]
For every \( z, z' \in \mathbb{R}^2 \) Consider \( k, l \in \{1, \ldots, n\} \) such that \( k \neq l \). By Markov’s inequality,

\[
\begin{align*}
\mathbb{E} \left( \sup_{h \in H_n} |g^2_{n,h,h_{min}}(Z_k, Z_l)| \right) & \leq m(n) \sum_{h \in H_n} \int_{-\infty}^{\infty} \mathbb{E}(\|K_h(X_k - u)\|)\mathbb{E}(\|Y_1|Y_1|>m(n)|K_{h_{min}}(X_l - u))du \\
& \leq m(n)|H_n| \cdot \|f\|_\infty \|K\|_2^2 \mathbb{E}(Y_1^2)^{1/2} \mathbb{P}(\exp(\alpha|Y_1|) > n^4)^{1/2} \\
& \leq \|f\|_\infty \|K\|_2^2 \mathbb{E}(Y_1^2)^{1/2} \mathbb{E}(\exp(\alpha|Y_1|))^{1/2} \frac{m(n)}{n^2} |H_n|.
\end{align*}
\]

Then, there exists a deterministic constant \( c_8 > 0 \), not depending on \( n \) and \( h_{min} \), such that

\[
\mathbb{E} \left( \sup_{h \in H_n} \frac{|U_{2,n}(h, h_{min})|}{n^2} \right) \leq c_8 \frac{\log(n)}{n}.
\]

The same ideas give that there exists a deterministic constant \( c_9 > 0 \), not depending on \( n \) and \( h_{min} \), such that

\[
\mathbb{E} \left( \sup_{h \in H_n} \frac{|U_{3,n}(h, h_{min})|}{n^2} \right) \leq c_9 \frac{\log(n)}{n}.
\]

For \( i = 4 \), by Markov’s inequality,

\[
\begin{align*}
\mathbb{E} \left( \sup_{h \in H_n} |g^4_{n,h,h_{min}}(Z_k, Z_l)| \right) & \leq \sum_{h \in H_n} \int_{-\infty}^{\infty} \mathbb{E}(\|K_h(X_k - u)\|)\mathbb{E}(\|Y_1|Y_1|>m(n)|K_{h_{min}}(X_l - u))du \\
& \leq \|K\|_{h_{min}} \mathbb{E}(\|Y_1|Y_1|>m(n)) \\
& \quad \times \sum_{h \in H_n} \int_{-\infty}^{\infty} \mathbb{E}(\|K_h(X_k - u)\|)du \\
& \leq \|K\|_2^2 \mathbb{E}(Y_1^2)^{1/2} \mathbb{E}(\exp(\alpha|Y_1|))^{1/2} \frac{1}{n^4 h_{min}} |H_n|.
\end{align*}
\]

Then, there exists a deterministic constant \( c_{10} > 0 \), not depending on \( n \) and \( h_{min} \), such that

\[
\mathbb{E} \left( \sup_{h \in H_n} \frac{|U_{4,n}(h, h_{min})|}{n^2} \right) \leq c_{10} \frac{\log(n)}{n^3 h_{min}}.
\]

Therefore,

\[
\mathbb{E} \left( \sup_{h \in H_n} \left\{ \frac{|U_{n}(h, h_{min})|}{n^2} - \frac{\theta \|K\|_2^2}{n^2} \mathbb{E}(Y_1^2) \right\} \right) \leq \frac{c_U}{\theta} \cdot \frac{\log(n)^5}{n}.
\]
6.4.3. Proof of Lemma 6.3

Consider $m(n) := 4 \log(n)/\alpha$. For any $\eta, \eta' \in \mathcal{H}_n$,

$$V_n(\eta, \eta') = V_{1,n}(\eta, \eta') + V_{2,n}(\eta, \eta')$$

where

$$V_{i,n}(\eta, \eta') := \frac{1}{n} \sum_{k=1}^{n} (g_{\eta,\eta'}^i(X_k, Y_k) - E(g_{\eta,\eta'}(X_k, Y_k))) ; \quad i = 1, 2$$

with, for every $x, y \in \mathbb{R}$,$$
g_{\eta,\eta'}^1(x, y) := \langle y K_\eta(x - \cdot), (bf)_{\eta'} - bf \rangle_2 1_{|y| \leq m(n)}$$

and

$$g_{\eta,\eta'}^2(x, y) := \langle y K_\eta(x - \cdot), (bf)_{\eta'} - bf \rangle_2 1_{|y| > m(n)}.$$ In order to apply Bernstein’s inequality to $g_{\eta,\eta'}^1(X_k, Y_k)$, $k = 1, \ldots, n$, let us find suitable controls of

$$\epsilon_{\eta,\eta'} := \frac{\|g_{\eta,\eta'}^1\|_\infty}{3} \quad \text{and} \quad v_{\eta,\eta'} := \mathbb{E}(g_{\eta,\eta'}^1(X_1, Y_1)^2).$$

On the one hand, since $\|K\|_1 \geq 1$ and $bf$ is bounded,

$$\epsilon_{\eta,\eta'} = \frac{1}{3} \sup_{x, y \in \mathbb{R}} |\langle y K_\eta(x - \cdot), (bf)_{\eta'} - bf \rangle_2 1_{|y| \leq m(n)}|$$

$$\leq \frac{m(n)}{3} \|bf\|_\infty \sup_{x \in \mathbb{R}} \|K_\eta(x - \cdot)\|_1$$

$$\leq \frac{m(n)}{3} \|K\|_1 (\|K\|_1 + 1) \|bf\|_\infty \leq \frac{2}{3} \frac{m(n)}{m(n)} \|K\|_1^3 \|bf\|_\infty.$$ On the other hand,

$$v_{\eta,\eta'} = \mathbb{E}(Y_1^2 K_\eta(X_1 - \cdot), (bf)_{\eta'} - bf \|_2^2 1_{|Y_1| \leq m(n)})$$

$$= \mathbb{E}\left(Y_1^2 1_{|Y_1| \leq m(n)} \left| \int_{-\infty}^{\infty} K_\eta(X_1 - u)((bf)_{\eta'}(u) - (bf)(u))du \right|^2 \right)$$

$$\leq m(n)^2 \|f\|_\infty \|K\|_1^2 \|bf\|_2^2.$$ So, by Bernstein’s inequality, there exists a universal constant $c_1 > 0$ such that with probability larger than $1 - 2e^{-\lambda}$,

$$|V_{1,n}(\eta, \eta')| \leq \sqrt{2\lambda \mathbb{E}(v_{\eta,\eta'}) + \frac{\lambda}{n} \mathbb{E}(\epsilon_{\eta,\eta'})}$$

$$\leq \theta \|bf\|_2^2 + c_1 \frac{m(n)^2}{\theta m} \|K\|_1^2 (\|f\|_\infty + \|bf\|_\infty) \lambda.$$
Then, with probability larger than $1 - 2|\mathcal{H}_n|^2 e^{-\lambda}$,

$$S_n \leq c_1 \frac{m(n)^2}{\theta n} \| K \|_2^2 (\| f \|_\infty + \| bf \|_\infty) \lambda$$

where

$$S_n := \sup_{\eta, \eta' \in \mathcal{H}_n} \{ |V_{1,n}(\eta, \eta')| - \theta \| (bf)_{\eta'} - bf \|_2^2 \}.$$ 

For every $s \in \mathbb{R}_+$, consider

$$\lambda(s) := s \frac{m(n, \theta)}{m(n, \theta)} \text{ with } m(n, \theta) := c_1 \frac{m(n)^2}{\theta n} \| K \|_2^2 (\| f \|_\infty + \| bf \|_\infty).$$

Then, for any $A > 0$,

$$E(S_n) \leq A + \int_A^{\infty} \mathbb{P}(S_n \geq s) ds$$

$$\leq A + 2c_2 |\mathcal{H}_n|^2 m(n, \theta) \exp \left( - \frac{A}{2m(n, \theta)} \right)$$

where $\int_0^{\infty} e^{-s/2} ds = 2$. Since there exists a deterministic constant $c_3 > 0$, not depending on $n$ and $h_{\min}$ such that

$$m(n, \theta) \leq c_3 \frac{\log(n)^2}{n},$$

by taking $A := 4c_3 \log(n)^3/n$,

$$E(S_n) \leq 4c_3 \frac{\log(n)^3}{n} + 2c_2 m(n, \theta) \frac{|\mathcal{H}_n|}{n^2}.$$ 

Therefore, since $|\mathcal{H}_n| \leq n$, there exists a deterministic constant $c_4 > 0$, not depending on $n$ and $h_{\min}$, such that

$$E \left( \sup_{\eta, \eta' \in \mathcal{H}_n} \{ |V_{1,n}(\eta, \eta')| - \theta \| (bf)_{\eta'} - bf \|_2^2 \} \right) \leq c_4 \frac{\log(n)^3}{n}.$$ 

Now, let us find a suitable control of

$$v_{2,n} := E \left( \sup_{\eta, \eta' \in \mathcal{H}_n} |V_{2,n}(\eta, \eta')| \right).$$

By Markov’s inequality,

$$v_{2,n} \leq 2E \left( \sup_{\eta \in \mathcal{H}_n} |Z_{2,1}(\eta, \eta')| \right)$$

$$\leq 2E(Y_1^2 |Y_1 > m(n))^{1/2}$$

$$\times E \left( \sup_{\eta, \eta' \in \mathcal{H}_n} \left| \int_{-\infty}^{\infty} K_\eta(X_1 - u)((bf)_{\eta'}(u) - (bf)(u)) du \right|^2 \right)^{1/2}$$

$$\leq 2E(Y_1^4)^{1/4} \mathbb{P}(\exp(\alpha|Y_1|) > n^4)^{1/4} \left( \| K \|_1 \sup_{\eta' \in \mathcal{H}_n} \| (bf)_{\eta'} - bf \|_\infty \right)$$

$$\leq 2E(Y_1^4)^{1/4} \mathbb{E}(\exp(\alpha|Y_1|))^{1/4} \| K \|_1^2 \| bf \|_\infty \frac{1}{n}.$$
Therefore,

\[ \mathbb{E} \left( \sup_{\eta, \eta' \in U_n} \{ |V_n(\eta, \eta')| - \theta \|(bf)_{\eta'} - bf\|_2^2 \} \right) \leq \epsilon_V \frac{\log(n)^3}{\theta n}. \]

### 6.4.4. Proof of Lemma 6.4

First of all,

\[ \|(bf)_h - bf\|_2^2 = \|\hat{bf}_{n,h} - bf\|_2^2 - \|\hat{bf}_{n,h} - (bf)_h\|_2^2 + 2V_n(h, h). \]

Then, for any \( \theta \in (0, 1/2) \),

\[ (1 - 2\theta) \left( \|(bf)_h - bf\|_2^2 + \frac{c_{KY}}{nh} \right) - \|\hat{bf}_{n,h} - bf\|_2^2 \leq 2(V_n(h, h)) - \theta \|(bf)_h - bf\|_2^2 + \Lambda_n(h) - 2\theta \frac{c_{KY}}{nh} \]

where

\[ \Lambda_n(h) := \left| \frac{\|\hat{bf}_{n,h} - (bf)_h\|_2^2 - c_{KY}}{nh} \right| \]

\[ = \left| \frac{U_n(h, h) - W_n(h)}{n} - \frac{1}{n} \|(bf)_h\|_2^2 \right| \]

with

\[ W_n(h) := \frac{1}{n} \sum_{k=1}^{n} (Z_k(h) - \mathbb{E}(Z_k(h))) \]

and

\[ Z_k(h) := \|Y_k K_h(X_k - \cdot) - (bf)_h\|_2^2 \quad \forall k \in \{1, \ldots, n\}, \]

because

\[ \mathbb{E}(Z_1(h)) = \sigma^2 \int_{-\infty}^{\infty} \mathbb{E}(K_h(X_1 - u)^2) du + \int_{-\infty}^{\infty} \mathbb{E}(b(X_1)^2 K_h(X_1 - u)^2) du \]

\[ - 2 \int_{-\infty}^{\infty} \mathbb{E}(b(X_1) K_h(X_1 - u)(bf)_h(u)) du + \int_{-\infty}^{\infty} (bf)_h(u)^2 du \]

\[ = \frac{\|K\|_2^2}{h} (\sigma^2 + \mathbb{E}(b(X_1)^2)) - \|(bf)_h\|_2^2 = \frac{c_{KY}}{h} - \|(bf)_h\|_2^2. \]

Consider \( m(n) := 2 \log(n)/\alpha \) and note that \( W_n(h) = W_{1,n}(h) + W_{2,n}(h) \), where

\[ W_{i,n}(h) := \frac{1}{n} \sum_{k=1}^{n} (g^i_h(X_k, Y_k) - \mathbb{E}(g^i_h(X_k, Y_k))) \quad i = 1, 2 \]

with, for every \( x, y \in \mathbb{R} \),

\[ g^i_h(x, y) := \|y K_h(x - \cdot) - (bf)_h\|_2^2 \mathbf{1}_{|y| \leq m(n)} \]
and
\[ g_h^2(x, y) := \|yK_h(x - \cdot) - (bf)_h\|^2 \mathbf{1}_{|y| > m(n)}. \]

Note also that
\[ \|bf\|_2 \leq \|K\|_2 \int_{-\infty}^{\infty} |b(x)| f(x) dx \leq \frac{\|K\|_2}{h^{1/2}} E(|b(X_1)|) \leq \left( \frac{\epsilon_{K,Y}}{h} \right)^{1/2} \]
and
\[ \|bf\|_2 \leq \|K\|_1 \left( \int_{-\infty}^{\infty} b(x)^2 f(x)^2 dx \right)^{1/2} \leq \|K\|_1 \|f\|_\infty^{1/2} E(b(X_1)^2)^{1/2}. \]

In order to apply Bernstein's inequality to \( g_h^1(X_k, Y_k), k = 1, \ldots, n \), let us find suitable controls of
\[ c_h := \frac{\|g_h^1\|_\infty}{3} \quad \text{and} \quad v_h := \mathbb{E}(g_h^1(X_1, Y_1)^2). \]

On the one hand,
\[ c_h = \frac{1}{3} \sup_{x, y \in \mathbb{R}} \|yK_h(x - \cdot) - (bf)_h\|_2^2 \mathbf{1}_{|y| \leq m(n)} \leq \frac{2}{3} \left( m(n)^2 \frac{\|K\|_2^2}{h} + \frac{\epsilon_{K,Y}}{h} \right). \]

On the other hand,
\[ v_h \leq 2 \mathbb{E}(Z_1(h)(\|Y_1K_h(X_1 - \cdot)\|_2^2 \mathbf{1}_{|Y_1| \leq m(n)} + \|bf\|_2^2)) \leq \frac{2}{h} E(Z_1(h))(\|K\|_2^2 m(n)^2 + c_{K,Y}) \leq 2(\|K\|_2^2 + c_{K,Y}) \frac{\epsilon_{K,Y}}{h} m(n)^2. \]

So, by Bernstein's inequality, there exists a universal constant \( c_1 > 0 \) such that with probability larger than \( 1 - 2e^{-\lambda} \),
\[ |W_{1,n}(h)| \leq \sqrt{\frac{2\lambda}{n} v_h + \frac{\lambda}{n} c_h} \leq \theta \frac{\epsilon_{K,Y}}{h} + c_1 \frac{m(n)^2}{\theta n h_{\min}} (\|K\|_2^2 + c_{K,Y}) \lambda. \]

Then, with probability larger than \( 1 - 2|\mathcal{H}_n|e^{-\lambda} \),
\[ S_n(h_{\min}) \leq c_1 \frac{m(n)^2}{\theta n^2 h_{\min}} (\|K\|_2^2 + c_{K,Y}) \lambda \]
where
\[ S_n(h_{\min}) := \sup_{h \in \mathcal{H}_n} \left\{ \left| W_{1,n}(h) \right| - \theta \frac{\epsilon_{K,Y}}{n h} \right\}. \]
For every $s \in \mathbb{R}_+$, consider

$$
\lambda(s) := \frac{s}{m(n, h_{\min}, \theta)} \quad \text{with} \quad m(n, h_{\min}, \theta) := c_1 \frac{m(n)^2}{\theta n^2 h_{\min}} (\|K\|_2^2 + c_{K,Y}).
$$

Then, for any $A > 0$,

$$
\mathbb{E}(S_n(h_{\min})) \leq A + \int_A^\infty \mathbb{P}(S_n(h_{\min}) \geq s) \, ds \\
\leq A + 2c_2 |\mathcal{H}_n| m(n, h_{\min}, \theta) \exp \left( -\frac{A}{2m(n, h_{\min}, \theta)} \right)
$$

where $c_2 := \int_0^\infty e^{-s/2} \, ds = 2$. Since there exists a deterministic constant $c_3 > 0$, not depending on $n$ and $h_{\min}$, such that

$$
m(n, \theta) \leq \frac{\log(n)^2}{n},
$$

by taking $A := 2c_3 \log(n)^3/n$,

$$
\mathbb{E}(S_n(h_{\min})) \leq 2c_3 \frac{\log(n)^3}{n} + 2c_2 m(n, h_{\min}, \theta) \frac{|\mathcal{H}_n|}{n}.
$$

Therefore, since $|\mathcal{H}_n| \leq n$, there exists a deterministic constant $c_4 > 0$, not depending on $n$ and $h_{\min}$, such that

$$
\mathbb{E} \left( \sup_{h \in \mathcal{H}_n} \left\{ \left| \frac{W_{1,n}(h)}{n} - \theta \frac{c_{K,Y}}{nh} \right| \right\} \right) \leq \frac{c_4}{\theta} \frac{\log(n)^3}{n}.
$$

Now, by Markov's inequality,

$$
\mathbb{E} \left( \sup_{h \in \mathcal{H}_n} \left| \frac{W_{2,n}(h)}{n} \right| \right) \leq \frac{2}{n} \mathbb{E} \left( \sup_{h \in \mathcal{H}_n} |Z_1(h)| 1_{|Y_1| > m(n)} \right) \\
\leq \frac{4}{n} \mathbb{E} \left( \sup_{h \in \mathcal{H}_n} (\|Y_1 X_h - \cdot\|_2^2 + \|(bf)_h\|_2^2) 1_{|Y_1| > m(n)} \right) \\
\leq \frac{4}{n h_{\min}} (\|K\|_2^2 \mathbb{E}(Y_1^4)^{1/2} + c_{K,Y}) \mathbb{P}(|Y_1| > m(n))^{1/2} \\
\leq 4(\|K\|_2^2 \mathbb{E}(Y_1^4)^{1/2} + c_{K,Y}) \mathbb{E}((\exp(\alpha|Y_1|))^{1/2} \frac{1}{\theta n^2 h_{\min}}).
$$

Then, there exists a deterministic constant $c_5 > 0$, not depending on $n$ and $h_{\min}$, such that

$$
\mathbb{E} \left( \sup_{h \in \mathcal{H}_n} \left\{ \left| \frac{W_n(h)}{n} \right| - \theta \frac{c_{K,Y}}{nh} \right\} \right) \leq \frac{c_5}{\theta} \frac{\log(n)^3}{n}.
$$
Therefore, by Lemma 6.2, there exists a deterministic constant \( \epsilon_6 > 0 \), not depending on \( n \) and \( h_{\text{min}} \), such that

\[
\mathbb{E} \left( \sup_{h \in \mathcal{H}_n} \{ \Lambda_n(h) - 2\theta \frac{\epsilon_{K,Y} n h}{n} \} \right) \leq \frac{\epsilon_V}{\theta} \frac{\log(n)^5}{n} + \frac{\epsilon_5}{\theta} \frac{\log(n)^3}{n} + \frac{1}{n} \| K \|_2^2 \| f \|_\infty \mathbb{E}(b(X_1)^2) \leq \frac{\epsilon_6}{\theta} \frac{\log(n)^5}{n}.
\]

Moreover, by Lemma 6.3,

\[
\mathbb{E} \left( \sup_{h \in \mathcal{H}_n} \{ |V_n(h, h)| - \theta \| (bf)_h - bf \|_2^2 \} \right) \leq \frac{\epsilon_V}{\theta} \frac{\log(n)^3}{n}.
\]

In conclusion, by Inequality (12),

\[
\mathbb{E} \left( \sup_{h \in \mathcal{H}_n} \left\{ \| (bf)_h - bf \|_2^2 + \frac{\epsilon_{K,Y} n h}{n} - \frac{1}{1 - 2\theta} \| \hat{bf}_{n,h} - bf \|_2^2 \right\} \right) \leq \frac{\epsilon_L}{\theta(1 - 2\theta)} \frac{\log(n)^5}{n}.
\]

References

Appendix A: Additional simulation results

<table>
<thead>
<tr>
<th>( n )</th>
<th>( b_1 )</th>
<th>( b_2 )</th>
<th>( b_3 )</th>
<th>( b_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>0.32</td>
<td>0.29</td>
<td>0.14</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.08)</td>
<td>(0.03)</td>
<td>(0.10)</td>
</tr>
<tr>
<td>500</td>
<td>0.29</td>
<td>0.25</td>
<td>0.12</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.06)</td>
<td>(0.02)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>1000</td>
<td>0.25</td>
<td>0.22</td>
<td>0.10</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.04)</td>
<td>(0.01)</td>
<td>(0.05)</td>
</tr>
</tbody>
</table>

Table 7: Mean of selected bandwidth (with std in parenthesis below) with the CV method for NW-single bandwidth estimator of \( b \), \( \sigma = 0.7 \), \( X \sim N(0,1) \), 200 repetitions.
### Table 8

100\(^*\)MISE (with 100\(^*\)std in parenthesis below) for the estimation of \(b_1f\), 200 repetitions, 
\(X \sim \gamma(3, 2)/5\) and \(\sigma = 0.1\). Same columns as in Table 1.

<table>
<thead>
<tr>
<th>(n)</th>
<th>(b_1f) PCO CV Or</th>
<th>(b_1f) PCO CV Or</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>0.39 (0.35) 0.46 (0.56) 0.20 (0.19)</td>
<td>0.51 (0.44) 0.57 (0.81) 0.25 (0.23)</td>
</tr>
<tr>
<td>500</td>
<td>0.19 (0.16) 0.27 (0.37) 0.10 (0.09)</td>
<td>0.27 (0.21) 0.32 (0.38) 0.13 (0.13)</td>
</tr>
<tr>
<td>1000</td>
<td>0.11 (0.09) 0.18 (0.30) 0.06 (0.05)</td>
<td>0.15 (0.12) 0.23 (0.44) 0.07 (0.06)</td>
</tr>
</tbody>
</table>

### Table 9

100\(^*\)MISE (with 100\(^*\)std in parenthesis below) for the estimation of \(b_1f\), 200 repetitions, 
\(X \sim \gamma(3, 2)/5\) and \(\sigma = 0.7\). Same columns as in Table 1.

<table>
<thead>
<tr>
<th>(n)</th>
<th>(b_2f) PCO CV Or</th>
<th>(b_4f) PCO CV Or</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>0.61 (0.37) 0.71 (0.70) 0.42 (0.27)</td>
<td>0.83 (0.16) 0.85 (0.26) 0.49 (0.09)</td>
</tr>
<tr>
<td>500</td>
<td>0.31 (0.18) 0.34 (0.30) 0.22 (0.14)</td>
<td>0.47 (0.08) 0.44 (0.17) 0.23 (0.04)</td>
</tr>
<tr>
<td>1000</td>
<td>0.16 (0.11) 0.24 (0.41) 0.11 (0.07)</td>
<td>0.05 (0.04) 0.08 (0.10) 0.03 (0.02)</td>
</tr>
</tbody>
</table>

### Table 1

...
Table 10
100*MISE (with 100*std in parenthesis below) for the estimation of $b_i, i = 1, \ldots, 4$, 200 repetitions, $X \sim \gamma(3, 2)/5$, $\sigma = 0.1$. CV and PCO are the two competing methods. Column "Or" gives the average of ISE for the ratio of the two best estimators of $bf$ and $f$ in the collection.

<table>
<thead>
<tr>
<th>n</th>
<th>CV</th>
<th>PCO</th>
<th>Or</th>
<th>CV</th>
<th>PCO</th>
<th>Or</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>0.22</td>
<td>0.66</td>
<td>0.39</td>
<td>0.56</td>
<td>8.75</td>
<td>1.74</td>
</tr>
<tr>
<td></td>
<td>(0.15)</td>
<td>(0.42)</td>
<td>(0.34)</td>
<td>(0.78)</td>
<td>(39.2)</td>
<td>(2.50)</td>
</tr>
<tr>
<td>500</td>
<td>0.11</td>
<td>0.30</td>
<td>0.17</td>
<td>0.43</td>
<td>1.35</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.23)</td>
<td>(0.14)</td>
<td>(0.16)</td>
<td>(2.29)</td>
<td>(0.91)</td>
</tr>
<tr>
<td>1000</td>
<td>0.07</td>
<td>0.15</td>
<td>0.09</td>
<td>0.12</td>
<td>0.38</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.09)</td>
<td>(0.09)</td>
<td>(0.08)</td>
<td>(0.41)</td>
<td>(0.30)</td>
</tr>
</tbody>
</table>

Table 11
100*MISE (with 100*std in parenthesis below) for the estimation of $b_i, i = 1, \ldots, 4$, 200 repetitions, $X \sim \gamma(3, 2)/5$, $\sigma = 0.7$. CV and PCO are the two competing methods. Column "Or" gives the average of ISE for the ratio of the two best estimators of $bf$ and $f$ in the collection.

<table>
<thead>
<tr>
<th>n</th>
<th>CV</th>
<th>PCO</th>
<th>Or</th>
<th>CV</th>
<th>PCO</th>
<th>Or</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>10.6</td>
<td>19.4</td>
<td>11.1</td>
<td>4.54</td>
<td>7.15</td>
<td>6.29</td>
</tr>
<tr>
<td></td>
<td>(10.6)</td>
<td>(19.4)</td>
<td>(11.1)</td>
<td>(4.37)</td>
<td>(7.95)</td>
<td>(9.28)</td>
</tr>
<tr>
<td>500</td>
<td>5.71</td>
<td>5.84</td>
<td>5.51</td>
<td>2.52</td>
<td>2.84</td>
<td>2.97</td>
</tr>
<tr>
<td></td>
<td>(3.45)</td>
<td>(5.26)</td>
<td>(7.10)</td>
<td>(2.16)</td>
<td>(2.07)</td>
<td>(2.89)</td>
</tr>
<tr>
<td>1000</td>
<td>3.17</td>
<td>2.70</td>
<td>2.47</td>
<td>1.50</td>
<td>1.38</td>
<td>1.41</td>
</tr>
<tr>
<td></td>
<td>(2.02)</td>
<td>(1.69)</td>
<td>(1.61)</td>
<td>(1.44)</td>
<td>(1.14)</td>
<td>(1.17)</td>
</tr>
</tbody>
</table>