# "Adaptive non-parametric estimation in the presence of dependence"

Asin, Nicolas ; Johannes, Jan

**Abstract**

We consider non-parametric estimation problems in the presence of dependent data, notably non-parametric regression with random design and non-parametric density estimation. The proposed estimation procedure is based on a dimension reduction. The minimax optimal rate of convergence of the estimator is derived assuming a sufficiently weak dependence characterized by fast decreasing mixing coefficients. We illustrate these results by considering classical smoothness assumptions. However, the proposed estimator requires an optimal choice of a dimension parameter depending on certain characteristics of the function of interest, which are not known in practice. The main issue addressed in our work is an adaptive choice of this dimension parameter combining model selection and Lepski's method. It is inspired by the recent work of Goldenshluger and Lepski [2011]. We show that this data-driven estimator can attain the lower risk bound up to a constant provided a fast decay of the mixing co...

Document type : *Document de travail (Working Paper)*

## Référence bibliographique

Asin, Nicolas ; Johannes, Jan. *Adaptive non-parametric estimation in the presence of dependence.* ISBA Discussion Paper ; 2016/07 (2016) 39 pages

INSTITUT DE STATISTIQUE
BIOSTATISTIQUE ET
SCIENCES ACTUARIELLES
(ISBA)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION
PAPER

2016/07

# Adaptive non-parametric estimation in the presence of dependence

ASIN, N. and J. JOHANNES

# Adaptive non-parametric estimation in the presence of dependence

NICOLAS ASIN[1*] JAN JOHANNES[2]

Université catholique de Louvain Ruprecht-Karls-Universität Heidelberg

**Abstract**

We consider non-parametric estimation problems in the presence of dependent data, notably non-parametric regression with random design and non-parametric density estimation. The proposed estimation procedure is based on a dimension reduction. The minimax optimal rate of convergence of the estimator is derived assuming a sufficiently weak dependence characterized by fast decreasing mixing coefficients. We illustrate these results by considering classical smoothness assumptions. However, the proposed estimator requires an optimal choice of a dimension parameter depending on certain characteristics of the function of interest, which are not known in practice. The main issue addressed in our work is an adaptive choice of this dimension parameter combining model selection and Lepski's method. It is inspired by the recent work of Goldenshluger and Lepski [2011]. We show that this data-driven estimator can attain the lower risk bound up to a constant provided a fast decay of the mixing coefficients.

Keywords: Density estimation, non-parametric regression, dependence, mixing, minimax theory, adaptation

AMS 2000 subject classifications: Primary 62G05; secondary 62G07, 62G08.

# 1 Introduction

We study the non-parametric estimation of a functional parameter of interest $f$ based on a sample of identically distributed random variables $Z_1, \dots, Z_n$. For convenience, the function

[1]Institut de statistique, biostatistique et sciences actuarielles (ISBA), Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium, e-mail: nicolas.asin@uclouvain.be

[*]Corresponding author.

[2]Institut für Angewandte Mathematik, Im Neuenheimer Feld 294, 69120 Heidelberg, Germany, e-mail: johannes@math.uni-heidelberg.de

of interest $f$ belongs to the Hilbert space $L_2 := L_2[0,1]$ of square integrable real-valued functions defined on $[0,1]$ which is endowed with its usual inner product $\langle\cdot,\cdot\rangle_{L_2}$ and its induced norm $\|\cdot\|_{L_2}$. In this paper we study the attainable accuracy of a fully data-driven estimator of $f$ for independent as well as dependent observations $Z_1,\dots,Z_n$ from a minimax point of view. The estimator is based on an orthogonal series approach where the fully data-driven selection of the dimension parameter is inspired by the recent work of Goldenshluger and Lepski [2011]. We derive conditions that allow us to bound the maximal risk of the fully data-driven estimator over suitable chosen classes $\mathcal{F}$ for $f$, which are constructed flexibly enough to characterize, in particular, differentiable or analytic functions. Considering two classical non-parametric problems, namely non-parametric density estimation and non-parametric regression with random design, we show that these conditions indeed hold true, if the identically distributed observations $Z_1,\dots,Z_n$ are independent (iid.) or weakly dependent with sufficiently fast decay of their $\beta$-mixing coefficients. Thereby, we establish the rate of convergence of the fully data-driven estimator for independent as well as weakly dependent observations. Considering iid. observations we show that these rates of convergence are minimax-optimal for a wide variety of classes $\mathcal{F}$, and hence the fully data-driven estimator is called adaptive. Replacing the independence assumption by mixing conditions the rates of convergence of the fully data-driven estimator are generally slower. A comparison, however, allows us to state conditions on the mixing coefficients which ensure that the fully data-driven estimator still attains the minimax-optimal rates for a wide variety of classes of $\mathcal{F}$, and hence, is adaptive. The adaptive non-parametric estimation based on weakly dependent observations of either a density or a regression function has been consider by Tribouley and Viennet [1998], Comte and Merlevede [2002], Comte and Rozenholc [2002], Gannaz and Wintenberger [2010], Comte et al. [2008] or Bertin and Klutchnikoff [2014], to name but a view. However, our conditions to derive rates of convergence of the fully-data driven estimator can be verified for both, non-parametric density estimation and non-parametric regression with random design. Thereby, we think that these conditions provide a promising starting point to deal with more complex non-parametric models, as for example, errors in variables model.

The paper is organized as follows: in Section 2 we introduce our basic assumptions, define the class $\mathcal{F}$ and develop the data-driven orthogonal series estimator. We present key arguments of the proofs while technical details are postponed to the Appendix. We show, in Section 3, the minimax-optimality of the data-driven estimator of a density as well as a regression function based on iid. observations. In Section 4 we briefly review elementary dependence notions and present standard coupling arguments. Considering again the non-parametric estimation of a density as well as a regression function we derive mixing conditions such that the fully data-driven estimator based on dependent observations can attain the minimax-rates for independent data. Finally, considering the framework used by Gannaz and Wintenberger [2010] and Bertin and Klutchnikoff [2014] results of a simulation study are reported in Section 5 which allow to compare the finite sample performance of different data-driven estimators of a density as well as a regression function given independent or dependent observations.

# 2 Model assumptions and notations

## 2.1 Assumptions and notations

We construct an estimator of the unknown function $f$ using an orthogonal series approach. The estimation of $f$ is based on a dimension reduction which we elaborate in the following. Let us specify an arbitrary orthonormal system $\{\phi_j\}_{j=1}^{\infty}$ of $L_2$. We denote by $\Pi_\Phi$ and $\Pi_\Phi^\perp$ the orthogonal projections on the linear subspace $\Phi$ spanned by this orthonormal system and its orthogonal complement $\Phi^\perp$ in $L_2$, respectively. Consequently, any function $h \in \Phi$ admits an expansion $h = \sum_{j=1}^{\infty}[h]_j\phi_j$ as a generalised Fourier series with coefficients $[h]_j := \langle h, \phi_j\rangle_{L_2}$ for $j \geqslant 1$. The unknown function $f \in L_2$ is thereby uniquely determined by its coefficients $([f]_j)_{j\geqslant 1}$, or $[f]$ for short, and $\Pi_\Phi^\perp f$. In what follows $\Pi_\Phi^\perp f$ is know in advance while the sequence of coefficients $[f]$ has to be estimated. Given a dimension parameter $m \geqslant 1$ we have the subspace $\mathbb{D}_m$ spanned by the first $m$ basis functions $\{\phi_j\}_{j=1}^{m}$ at our disposal. For abbreviation, we denote by $\Pi_m$ and $\Pi_m^\perp$ the orthogonal projections on the linear subspace $\mathbb{D}_m$ and its orthogonal complement $\mathbb{D}_m^\perp$ in $\Phi$, respectively. We consider the orthogonal projection $f_m := \Pi_\Phi^\perp f + \Pi_m f$ of $f$ admitting the expansion $\Pi_m f_m = \sum_{j=1}^{m}[f]_j\phi_j$ and its associated approximation error $\mathrm{bias}_m(f) := \|f_m - f\|_{L_2} = \|\Pi_m^\perp f\|_{L_2}$ where $\mathrm{bias}_m(f)$ tends to zero as $m \to \infty$ for all $f \in L_2$ due to the dominated convergence theorem. We consider an orthogonal series estimator $\widehat{f}_m$ by replacing, for $j = 1, \dots, m$, the coefficient $[f]_j$ by its empirical counterpart $\widehat{[f]}_j$, that is, $\widehat{f}_{\widehat{m}} = \sum_{j=1}^{m}\widehat{[f]}_j\phi_j$. The attainable accuracy of the proposed estimator of $f$ are basically determined by a priori conditions on $f$. These conditions are often expressed in the form $f \in \mathcal{F}$, for a suitably chosen class $\mathcal{F} \subset L_2$. This class $\mathcal{F}$ reflects prior information on the function $f$, e.g., its level of smoothness, and will be constructed flexibly enough to characterize, in particular, differentiable or analytic functions. We determine the class $\mathcal{F}$ by means of a weighted norm in $\Phi$. Given the orthonormal basis $\{\phi_j\}_{j=1}^{\infty}$ of $\Phi$ and a strictly positive sequence of weights $(\mathfrak{a}_j)_{j\geqslant 1}$, or $\mathfrak{a}$ for short, we define for $h \in \Phi$ the weighted norm $\|h\|_{\mathfrak{a}}^2 := \sum_{j\in\mathbb{N}}\mathfrak{a}_j^{-1}[h]_j^2$. Furthermore, we denote by $\Phi_{\mathfrak{a}}$ and $\Phi_{\mathfrak{a}}^r$ for a constant $r > 0$, the completion of $\Phi$ with respect to $\|\cdot\|_{\mathfrak{a}}$ and the ellipsoid $\Phi_{\mathfrak{a}}^r := \{h \in \Phi : \|h\|_{\mathfrak{a}}^2 \leqslant r^2\}$. Obviously, for a non-increasing sequence $\mathfrak{a}$ the class $\Phi_{\mathfrak{a}}^r$ is a subspace of $\Phi$. Here and subsequently, we assume that there exist a monotonically non-increasing and strictly positive sequence of weights $\mathfrak{a}$ tending to zero and a constant $r > 0$ such that the function of interest $f$ belongs to the $\mathcal{F}_{\mathfrak{a}}^r := \{f \in L_2 : \Pi_\Phi f \in \Phi_{\mathfrak{a}}^r\}$. We may emphasize that for any $f \in \mathcal{F}_{\mathfrak{a}}^r$, $\mathrm{bias}_m^2(f) =$

$\sum_{j>m}(\mathfrak{a}_j/\mathfrak{a}_j)[f]_j^2 \leqslant \mathfrak{a}_m\|\Pi_\Phi f\|_\mathfrak{a}^2 \leqslant \mathfrak{a}_m r^2$ which we use in the sequel without further reference.

Further denote by $\|h\|_\infty$ as usual the $L_\infty$ norm of a function $h \in L_2$. We require in the sequel that the orthonormal system $\{\phi\}_j$ and the sequence $\mathfrak{a}$ satisfy the following assumptions.

**(A1)** There exists a finite constant $\tau_\infty \geqslant 1$ such that $\|\sum_{j=1}^m \phi_j^2\|_\infty \leqslant \tau_\infty^2 m$ for all $m \in \mathbb{N}$.

**(A2)** The sequence $\mathfrak{a}$ is monotonically decreasing with limit zero and there exists a finite constant $\mathfrak{A} \geqslant 1$ such that $\|\sum_{j\geqslant 1} \mathfrak{a}_j\phi_j^2\|_\infty \leqslant \mathfrak{A}^2$.

According to Lemma 6 of Birgé and Massart [1997] assumption **(A1)** is exactly equivalent to following property: there exists a positive constant $\tau_\infty$ such that for any $h \in \mathbb{D}_m$ holds $\|h\|_\infty \leqslant \tau_\infty\sqrt{m}\|h\|_{L_2}$. Typical example are bounded basis, such as the trigonometric basis, or basis satisfying the assertion, that there exists a positive constant $C_\infty$ such that for any $(c_1, \dots, c_m) \in \mathbb{R}^m$, $\|\sum_{j=1}^m c_j\phi_j\|_\infty \leqslant C_\infty\sqrt{m}|c|_\infty$ where $|c|_\infty = \max_{1\leqslant j\leqslant m}|c_j|$. Birgé and Massart [1997] have shown that the last property is satisfied for piecewise polynomials, splines and wavelets. On the other hand side, in the case of a bounded basis the property **(A2)** holds for any summable weight sequence $\mathfrak{a}$, i.e., $|\mathfrak{a}|_1 := \sum_{j\geqslant 1}\mathfrak{a}_j < \infty$. More generally, under **(A1)** the additional assumption $\sum_{j\geqslant 1} j\mathfrak{a}_j < \infty$ is sufficient to ensure **(A2)**. Furthermore, under **(A2)** the elements of $\Phi_\mathfrak{a}^r$ are bounded uniformly, that is $\|h\|_\infty^2 \leqslant \|\sum_{j\geqslant 1}\mathfrak{a}_j\phi_j^2\|_\infty\|h\|_\mathfrak{a}^2 \leqslant \mathfrak{A}^2 r^2 < \infty$ for any $h \in \Phi_\mathfrak{a}^r$.

## 2.2 Observations

In this work we focus on two models, namely non-parametric regression with random design and non-parametric density estimation. The important point to note here is that in each model the identically distributed (i.d.) observations $Z_1, \dots, Z_n$ satisfy $\mathbb{E}\psi_j(Z_i) = [f]_j$ for a certain function $\psi_j$ , $j \geqslant 1$. Therefore, given an i.d. sample $\{Z_i\}_{i=1}^n$, it is natural to consider the estimator $\widehat{[f]}_j = n^{-1}\sum_{i=1}^n \psi_j(Z_i)$ of $[f]_j$.

**Non-parametric regression.** A common problem in statistics is to investigate the dependence of a real random variable $Y$ on the variation of an explanatory random variable $U$. For convenience, the regressor $U$ is supposed to be uniformly distributed on the interval $[0, 1]$, i.e., $U \sim \mathcal{U}[0, 1]$. In this paper, the dependence of $Y$ on $U$ is characterised by $Y = f(U) + \sigma\varepsilon$, for $\sigma > 0$, where $f \in L_2$ is an unknown function and $\varepsilon$ is a centred and standardised error term. Furthermore, we suppose that $\varepsilon$ and $U$ are independent. Keeping in mind the expansion $f = \sum_{j=1}^\infty [f]_j\phi_j$ with respect to the basis $\{\phi_j\}_{j=1}^\infty$ we observe that $[f]_j = \mathbb{E}(\psi_j(Y, U))$ with $\psi_j(Y, U) = Y\phi_j(U)$ for all $j \geqslant 1$.

**Non-parametric density estimation.** Let $X$ be a random variable taking its values in $[0,1]$ and admitting a density $f$ which belongs to the set $\mathcal{D}$ of all densities with support included in $[0,1]$. We focus on the non-parametric estimation of the density $f$ if it is in addition square integrable, i.e., $f \in L_2$. For convenient notations, let $\mathbb{1}(t) := 1, t \in [0,1]$ and $\{\mathbb{1}\} \cup \{\phi_j\}_{j=1}^\infty$ be an orthonormal basis of $L_2$. Keeping in mind that $f$ is a density, it admits an expansion $f = \mathbb{1} + \sum_{j=1}^\infty [f]_j \phi_j$ where $[f]_j = \mathbb{E}[\phi_j(X)]$ for all $j \geqslant 1$. In this context we notice that $\Phi^\perp$ is spanned by $\mathbb{1}$. Since $f$ is a density function we have $\Pi_\Phi^\perp f = \mathbb{1}$, which is obviously known in advance.

## 2.3 Methodology and background

For the simplicity of the presentation, we assume throughout this section that $f \in \Phi$, that is $\Pi_\Phi^\perp f = 0$. The orthogonal projection $f_m = \sum_{j=1}^m [f]_j \phi_j$ at hand let us define an orthogonal series estimator by replacing for $j = 1, \dots, m$ the unknown coefficient $[f]_j$ by its empirical mean $\widehat{[f]}_j = n^{-1} \sum_{i=1}^n \psi_j(Z_i)$, that is, $\widehat{f}_m = \sum_{j=1}^m \widehat{[f]}_j \phi_j$. We shall assess the accuracy of the estimator $\widehat{f}_m$ by its maximal integrated mean squared error with respect to the class $\mathcal{F}$, that is $\mathcal{R}\big[\widehat{f} \,|\, \mathcal{F}\big] := \sup_{f \in \mathcal{F}} \mathbb{E}\|\widehat{f} - f\|_{L_2}^2$. Considering identically and independent distributed (iid.) observation obeying the two models, non-parametric regression and density estimation, we derive a lower bound for the maximal risk over $\mathcal{F}$ for all estimators and show that it provides up to a positive constant $C$ possibly depending on the class $\mathcal{F}$ also an upper bound for the maximal risk over $\mathcal{F}$ of the orthogonal series estimator $\widehat{f}_{m_n^\star}$ with suitable chosen dimension parameter $m_n^\star \in \mathbb{N}$, i.e.,

$$\mathcal{R}\big[\widehat{f}_{m_n^\star} \,|\, \mathcal{F}\big] \leqslant C \cdot \inf_{\widetilde{f}} \mathcal{R}\big[\widetilde{f} \,|\, \mathcal{F}\big]$$

where the infimum is taken over all estimators of $f$. We thereby prove the minimax optimality of the estimator $\widehat{f}_{m_n^\star}$. Obviously, if the observations are independent or sufficiently weak dependent there exists a finite constant $C > 0$ possibly depending on the class $\mathcal{F}_\mathfrak{a}^r$ such that $\sup_{f \in \mathcal{F}_\mathfrak{a}^r} \sum_{j=1}^m \mathbb{V}\mathrm{ar}(\widehat{[f]}_j) \leqslant Cmn^{-1}$ for all $m, n \geqslant 1$. From the Pythagorean formula we obtain the identity $\|\widehat{f}_m - f\|_{L_2}^2 = \|\widehat{f}_m - f_m\|_{L_2}^2 + \mathrm{bias}_m^2(f)$ and, hence together with $\mathrm{bias}_m^2(f) \leqslant \mathfrak{a}_m r^2$ for all $f \in \mathcal{F}_\mathfrak{a}^r$ follows

$$\mathcal{R}\big[\widehat{f}_m \,|\, \mathcal{F}_\mathfrak{a}^r\big] \leqslant \mathfrak{a}_m r^2 + Cmn^{-1} = (r^2 + C) \max(\mathfrak{a}_m, mn^{-1}). \tag{2.1}$$

The upper bound in the last display depends on the dimension parameter $m$ and hence by choosing an optimal value $m_n^\star$ the upper bound will be minimized which we formalize next. For a sequence $(a_m)_{m \geqslant 1}$ with minimal value in $A$ we set $\arg\min_{m \in A} \{a_m\} := \min\{m : a_m \leqslant$

$a_k, \forall k \in A\}$ and define for all $n, m \geqslant 1$

$$\mathcal{R}_n^m := \mathcal{R}_n^m(\mathfrak{a}) := [\mathfrak{a}_m \vee mn^{-1}] := \max(\mathfrak{a}_m, mn^{-1}),$$
$$m_n^\star := m_n^\star(\mathfrak{a}) := \arg\min_{m\in\mathbb{N}} \{\mathcal{R}_n^m\} \quad \text{and} \quad \mathcal{R}_n^\star := \mathcal{R}_n^\star(\mathfrak{a}) := \mathcal{R}_n^{m_n^\star} = \min_{m\in\mathbb{N}} \mathcal{R}_n^m. \quad (2.2)$$

From (2.1) we deduce that $\mathcal{R}\Big[\widehat{f}_{m_n^\star} \,|\, \mathcal{F}_{\mathfrak{a}}^r\Big] \leqslant (r^2 + C)\mathcal{R}_n^\star$ for all $n \geqslant 1$. Moreover if it is possible to show that $\mathcal{R}_n^\star$ provides up to a constant also a lower bound of $\mathcal{R}\Big[\widehat{f}_{m_n^\star} \,|\, \mathcal{F}_{\mathfrak{a}}^r\Big]$ then the estimator $\widehat{f}_{m_n^\star}$ with optimal chosen $m_n^\star$ is minimax rate-optimal. However, $m_n^\star$ depends on the unknown regularity of $f$ and hence we will introduce below a data-driven procedure to select the dimension parameter. Let us first briefly illustrate the last definitions by stating the order of $m_n^\star$ and $\mathcal{R}_n^\star$ for typical choices of the sequence $\mathfrak{a}$.

ILLUSTRATION 1. *We will illustrate all our results considering the following two configurations for the sequence $\mathfrak{a}$. Here and subsequently, we use for two strictly positive sequences $(x_n)_{n\geqslant1}$, $(y_n)_{n\geqslant1}$ the notation $x_n \asymp y_n$ if $(x_n/y_n)_{n\geqslant1}$ is bounded away both from zero and infinity. Let,*

*(p)* $\mathfrak{a}_j = |j|^{-2p}$, $j \geqslant 1$, *with* $p > 1$, *then* $m_n^\star \asymp n^{-1/(2p+1)}$ *and* $\mathcal{R}_n^\star \asymp n^{-2p/(2p+1)}$;

*(e)* $\mathfrak{a}_j = \exp(|j|^{-2p})$, $j \geqslant 1$, *with* $p > 0$, *then* $m_n^\star \asymp (\log(n))^{1/2p}$ *and* $\mathcal{R}_n^\star \asymp n^{-1}(\log(n))^{1/2p}$.

*We note that the assumption **(A2)** and* $(\mathcal{R}_n^\star)^{-1} \min(\mathfrak{a}_{m_n^\star}, m_n^\star n^{-1}) \asymp 1$ *hold true in both cases.*

Our selection method of the dimension parameter is inspired by the work of Goldenshluger and Lepski [2011] and combines the techniques of model selection and Lepski's method. We determine the dimension parameter among a collection of admissible values by minimizing a penalized contrast function. To this end, for all $n \geqslant 1$ let $(\mathrm{pen}_1, ..., \mathrm{pen}_n)$ be a subsequence of non-negative and non-decreasing penalties. We select $\widetilde{m}$ among the collection $\{1, \dots, n\}$ such that:

$$\widetilde{m} = \arg\min_{1\leqslant m\leqslant n} \{\Upsilon_m + \mathrm{pen}_m\} \quad (2.3)$$

where the contrast is defined by $\Upsilon_m := \max_{m\leqslant k\leqslant n} \Big\{\|\widehat{f}_m - \widehat{f}_k\|_{L_2}^2 - \mathrm{pen}_k\Big\}$ for all $1 \leqslant m \leqslant n$. The data-driven estimator is now given by $\widehat{f}_{\widetilde{m}}$ and our aim is to prove an upper bound for its maximal risk $\mathcal{R}\Big[\widehat{f}_{\widetilde{m}} \,|\, \mathcal{F}_{\mathfrak{a}}^r\Big]$. We outline next the main ideas of the proof and introduce conditions which we will show below hold indeed true for the two considered non-parametric estimation problems. A key argument is the next lemma due to Comte and Johannes [2012].

LEMMA 2.1. *If* $(\mathrm{pen}_1, \dots, \mathrm{pen}_n)$ *is a non-decreasing subsequence and* $1 \leqslant m \leqslant n$, *then*

$$\|\widehat{f}_{\widetilde{m}} - f\|_{L_2}^2 \leqslant 85 \max(\mathrm{bias}_m^2(f), \mathrm{pen}_m) + 42 \max_{m\leqslant k\leqslant n} \Big(\|\widehat{f}_k - f_k\|_{L_2}^2 - \mathrm{pen}_k/6\Big)_+$$

*where* $(x)_+ := \max(x, 0)$.

Keeping in mind that $\operatorname{bias}_m^2(f) \leqslant \mathfrak{a}_m r^2$ for all $f \in \mathcal{F}_{\mathfrak{a}}^r$ we impose the following condition.

**(C1)** There exists a finite constant $\delta > 0$ possibly depending on the class $\mathcal{F}_{\mathfrak{a}}^r$ such that $\sup_{f\in\mathcal{F}_{\mathfrak{a}}^r} \max_{1\leqslant m\leqslant n}\{\operatorname{pen}_m /m\} \leqslant \delta n^{-1}$ for all $n \geqslant 1$.

Under condition **(C1)** and employing and $\mathcal{R}_n^m = \max(\mathfrak{a}_m, mn^{-1})$ we have due to Lemma 2.1 that for all $1 \leqslant m \leqslant n$

$$\sup_{f\in\mathcal{F}_{\mathfrak{a}}^r} \mathbb{E}\|\widehat{f}_{\widetilde{m}} - f\|_{L_2}^2 \leqslant 85(r^2 \vee \delta)\mathcal{R}_n^m + 42 \sup_{f\in\mathcal{F}_{\mathfrak{a}}^r} \mathbb{E} \max_{m\leqslant k\leqslant n} \left(\|\widehat{f}_k - f_k\|_{L_2}^2 - \operatorname{pen}_k/6\right)_+. \tag{2.4}$$

Keeping mind that $\mathcal{R}_n^\star = \min_{m\in\mathbb{N}} \mathcal{R}_n^m = \mathcal{R}_n^{m_n^\star}$ where $m_n^\star = \underset{m\in\mathbb{N}}{\arg\min}\{\mathcal{R}_n^m\}$ realises a variance-squared-bias compromise among all values in $\mathbb{N}$. Considering the subset $\{1,\dots,n\}$ rather than $\mathbb{N}$ we have trivially $\mathcal{R}_n^\star = \min_{1\leqslant m\leqslant n} \mathcal{R}_n^m$ if $m_n^\star \leqslant n$. On the other hand, since $\mathcal{R}_n^\star = o(1)$ as $n \to \infty$ there exists $n_\diamond \in \mathbb{N}$ with $\mathcal{R}_{n_\diamond}^\star \leqslant 1$ for all $n \geqslant n_\diamond$ which in turn implies $m_n^\star \leqslant n$ for all $n \geqslant n_\diamond$. Indeed, $m_n^\star n^{-1} \leqslant \mathcal{R}_n^\star \leqslant \mathcal{R}_{n_\diamond}^\star \leqslant 1$ for all $n \geqslant n_\diamond$ implies that $m_n^\star \leqslant n$. Thereby, we have $\mathcal{R}_n^\star = \min_{1\leqslant m\leqslant n} \mathcal{R}_n^m$ for all $n \geqslant n_\diamond$. Consequently, from (2.4) follows for all $n \geqslant n_\diamond$

$$\mathcal{R}\left[\widehat{f}_{\widetilde{m}} \,|\, \mathcal{F}_{\mathfrak{a}}^r\right] \leqslant 85(\delta \vee r^2)\mathcal{R}_n^{m_n^\star} + 42 \sup_{f\in\mathcal{F}_{\mathfrak{a}}^r} \mathbb{E} \max_{m_n^\star\leqslant m\leqslant n} \left(\|\widehat{f}_m - f_m\|_{L_2}^2 - \operatorname{pen}_m/6\right)_+. \tag{2.5}$$

The second right hand side (rhs.) term in the last display we bound using the next condition.

**(C2)** There exists a finite constant $\Delta > 0$ possibly depending on the class $\mathcal{F}_{\mathfrak{a}}^r$ such that $\sup_{f\in\mathcal{F}_{\mathfrak{a}}^r} \mathbb{E}\left\{\max_{m_n^\star\leqslant m\leqslant n} \left(\|\widehat{f}_m - f_m\|_{L_2}^2 - 1/6\operatorname{pen}_m\right)_+\right\} \leqslant \Delta n^{-1}$ for all $n \geqslant 1$.

From (2.5) together with **(C2)** it follows that

$$\mathcal{R}\left[\widehat{f}_{\widetilde{m}} \,|\, \mathcal{F}_{\mathfrak{a}}^r\right] \leqslant 85(\delta \vee r^2)\mathcal{R}_n^{m_n^\star} + 42\Delta n^{-1}, \quad \text{for all } n \geqslant 1. \tag{2.6}$$

The next assertion is an immediate consequence and hence we omit its proof.

PROPOSITION 2.2. *Let **(C1)** and **(C2)** be satisfied, then for all $n \geqslant n_\diamond$ holds*

$$\mathcal{R}\left[\widehat{f}_{\widetilde{m}} \,|\, \mathcal{F}_{\mathfrak{a}}^r\right] \leqslant 85(\delta \vee r^2)\mathcal{R}_n^\star + 42\Delta n^{-1} \leqslant 127(\delta \vee r^2 \vee \Delta)\,\mathcal{R}_n^\star, \quad \textit{for all } n \geqslant n_\diamond$$

*where $n_\diamond \in \mathbb{N}$ satisfies $\mathcal{R}_{n_\diamond}^\star \leqslant 1$.*

The last assertion establishes an upper risk bound of the estimator $\widehat{f}_{\widetilde{m}}$. We call $\widehat{f}_{\widetilde{m}}$ partially data-driven if the sequence of penalty terms still depend on unknown quantities which however, can be estimated. In this situation, let $\widehat{\operatorname{pen}}_m$ be an estimator of $\operatorname{pen}_m$ such that the subsequence of penalties $(\widehat{\operatorname{pen}}_1, \dots, \widehat{\operatorname{pen}}_n)$ is non-negative and non-decreasing. The dimension parameter $\widehat{m}$ is then selected among the collection $\{1,\dots,n\}$ as follows

$$\widehat{m} = \underset{1\leqslant m\leqslant n}{\arg\min}\left\{\widehat{\Upsilon}_m + \widehat{\operatorname{pen}}_m\right\} \tag{2.7}$$

where the contrast is defined by $\widehat{\Upsilon}_m := \max_{m\leqslant k\leqslant n}\left\{\|\widehat{f}_m - \widehat{f}_k\|^2_{L_2} - \widehat{\text{pen}}_k\right\}$ for all $1 \leqslant m \leqslant n$. Following line by line the proof of Lemma 2.1 we obtain

$$\|\widehat{f}_{\widehat{m}} - f\|^2_{L_2} \leqslant 85\max(\text{bias}^2_m(f), \widehat{\text{pen}}_m) + 42 \max_{m\leqslant k\leqslant n}\left(\|\widehat{f}_k - f_k\|^2_{L_2} - \widehat{\text{pen}}_k/6\right)_+. \tag{2.8}$$

Keeping the last bound in mind we decompose the risk with respect to an event on which the quantity $\widehat{\text{pen}}_m$ is close to its theoretical counterpart $\text{pen}_m$. More precisely, define the event

$$\Omega = \{\text{pen}_m \leqslant \widehat{\text{pen}}_m \leqslant 3\,\text{pen}_m; \quad \forall 1 \leqslant m \leqslant n\} \tag{2.9}$$

and denote by $\Omega^c$ its complement. Let us consider the following decomposition for the maximal risk :

$$\mathcal{R}\left[\widehat{f}_{\widehat{m}}\,|\,\mathcal{F}^r_{\mathfrak{a}}\right] = \sup_{f\in\mathcal{F}^r_{\mathfrak{a}}} \mathbb{E}\left(\mathbb{1}_\Omega\|\widehat{f}_{\widehat{m}} - f\|^2_{L_2}\right) + \sup_{f\in\mathcal{F}^r_{\mathfrak{a}}} \mathbb{E}\left(\mathbb{1}_{\Omega^c}\|\widehat{f}_{\widehat{m}} - f\|^2_{L_2}\right) \tag{2.10}$$

where we bound the two rhs. terms separately.

LEMMA 2.3. *Under Assumption **(C1)** and **(C2)** we have*

$$\sup_{f\in\mathcal{F}^r_{\mathfrak{a}}} \mathbb{E}\left(\mathbb{1}_{\Omega^c}\|\widehat{f}_{\widehat{m}} - f\|^2_{L_2}\right) \leqslant \Delta n^{-1} + \{r^2 + \delta\}P(\Omega^c).$$

Due to the last assertion the second rhs. term in (2.10) is bounded up to a constant by $n^{-1}$ if the probability $P(\Omega^c)$ is sufficiently small, which we precize next.

**(C3)** There exists a finite constant $\kappa > 0$ possibly depending on the class $\mathcal{F}^r_{\mathfrak{a}}$ such that $\sup_{f\in\mathcal{F}^r_{\mathfrak{a}}} nP(\Omega^c) \leqslant \kappa$ for all $n \geqslant 1$.

Considering the first rhs. term in (2.10) we employ the inequality (2.8), that is

$$\|\widehat{f}_{\widehat{m}} - f\|^2_{L_2}\,\mathbb{1}_\Omega \leqslant 255\max(\text{bias}^2_m(f), \text{pen}_m) + 42\max_{m\leqslant k\leqslant n}\left(\|\widehat{f}_k - f_k\|^2_{L_2} - \text{pen}_k\right)_+. \tag{2.11}$$

Following now line by line the proof of Proposition 2.2 the next assertion is an immediate consequence of Lemma 2.3, the condition **(C3)** and (2.11) and we omit its proof.

PROPOSITION 2.4. *Under **(C1)**, **(C2)** and **(C3)** holds*

$$\mathcal{R}\left[\widehat{f}_{\widehat{m}}\,|\,\mathcal{F}^r_{\mathfrak{a}}\right] \leqslant 255(\delta\vee r^2)\mathcal{R}^\star_n + 43\Delta n^{-1} + \kappa(r^2+\delta)n^{-1} \leqslant (298+2\kappa)(\delta\vee r^2\vee\Delta)\,\mathcal{R}^\star_n,\ \textit{for all } n \geqslant n_\diamond$$

*where $n_\diamond \in \mathbb{N}$ satisfies $\mathcal{R}^\star_{n_\diamond} \leqslant 1$.*

Considering the two models, namely non-parametric density estimation and non-parametric regression, we will show that the conditions **(C1)** and **(C2)** and **(C3)** are verified. Thereby, an upper bound for the data-driven estimator $\widehat{f}_{\widetilde{m}}$ and $\widehat{f}_{\widehat{m}}$ can be deduced from Proposition 2.2 and 2.4, respectively.

# 3 Independent observations

In this section we suppose that the identically distributed $n$-sample $\{Z_i\}_{i=1}^n$ consists of independent random variables. Considering the two non-parametric estimation problems we will show that $\mathcal{R}_n^\star$ given in (2.2) provides a lower bound of the maximal risk $\mathcal{R}\big[\widetilde{f}\,|\,\mathcal{F}_{\mathfrak{a}}^r\big]$ for all possible estimators $\widetilde{f}$. On the other hand side, $\mathcal{R}_n^\star$ will provide also an upper bound up to a constant of the maximal risk of the orthogonal series estimator $\widehat{f}_{m_n^\star} = \sum_{j=1}^{m_n^\star} \widehat{[f]}_j \phi_j$ with optimally chosen dimension parameter. Thereby, $\mathcal{R}_n^\star$ is the minimax-optimal rate of convergence and the estimator $\widehat{f}_{m_n^\star}$ is minimax-rate optimal. However, the dimension parameter $m_n^\star$ depends on the class of unknown function. In a second step we will show by applying Proposition 2.2 and 2.4, respectively, that the data-driven estimator $\widehat{f}_{\widetilde{m}}$ and $\widehat{f}_{\widehat{m}}$ can attain the minimax-optimal rate of convergence. The key argument to verify the condition **(C2)** is the following inequality, which is due to Talagrand [1996] and can be found for example in Klein and Rio [2005].

LEMMA 3.1. *(Talagrand's inequality) Let $Z_1, \dots, Z_k$ be independent $\mathcal{Z}$-valued random variables and let $\overline{\nu_t} = k^{-1}\sum_{i=1}^k [\nu_t(Z_i) - \mathbb{E}(\nu_t(Z_i))]$ for $\nu_t$ belonging to a countable class $\{\nu_t, t \in \mathcal{T}\}$ of measurable functions. Then,*

$$\mathbb{E}\Big(\sup_{t\in\mathcal{T}} |\overline{\nu_t}|^2 - 6H^2\Big)_+ \leqslant C\left[\frac{v}{k}\exp\left(\frac{-kH^2}{6v}\right) + \frac{h^2}{k^2}\exp\left(\frac{-KkH}{h}\right)\right]$$

*with numerical constants $K = (\sqrt{2}-1)/(21\sqrt{2})$ and $C > 0$ and where*

$$\sup_{t\in\mathcal{T}}\sup_{x\in\mathcal{Z}} |\nu_t(x)| \leqslant h, \qquad \mathbb{E}\left[\sup_{t\in\mathcal{T}} |\overline{\nu_t}|\right] \leqslant H, \qquad \sup_{t\in\mathcal{T}} \frac{1}{k}\sum_{i=1}^k \mathbb{V}\mathrm{ar}(\nu_t(Z_i)) \leqslant v.$$

REMARK 2. *Let us briefly reconsider the orthogonal series estimator. Introduce further the unit ball $\mathbb{B}_m := \{h \in \mathbb{D}_m : \|h\|_{L_2} \leqslant 1\}$ contained in the subspace $\mathbb{D}_m = \mathrm{lin}\{\phi_1, \dots, \phi_m\}$ which is a countable set of functions. Moreover, set $\overline{\nu_t} = n^{-1}\sum_{i=1}^n [\nu_t(Z_i) - \mathbb{E}(\nu_t(Z_i))]$ and $\nu_t(Z) = \sum_{j=1}^m [t]_j \psi_j(Z)$, then we have*

$$\|\widehat{f}_m - f_m\|_{L_2}^2 = \sup_{t\in\mathbb{B}_m} |\langle \widehat{f}_m - f_m, t\rangle|^2 = \sup_{t\in\mathbb{B}_m} |\sum_{j=1}^m (\widehat{[f]}_j - [f]_j)[t]_j|^2 = \sup_{t\in\mathbb{B}_m} |\overline{\nu_t}|^2.$$

*The last identity provides the necessary argument to link the condition **(C2)** and Talagrand's inequality. Moreover we will suppose that the ONS $\{\phi_j\}_{j\in\mathbb{N}}$ and the weight sequence $\mathfrak{a}$ used to construct the ellipsoid $\mathcal{F}_{\mathfrak{a}}^r$ satisfy the assumptions **(A1)** and **(A2)**.*

## 3.1 Non-parametric density estimation

In this paragraph we suppose that the identically distributed $n$-sample $\{X_i\}_{i=1}^n$ consists of independent random variables admitting a common density $f$ which belongs to the set $\mathcal{D}$ of all densities with support included in $[0, 1]$.

PROPOSITION 3.2 (Upper bound). *Let $\{X_i\}_{i=1}^n$ be an iid. $n-$sample. Under the assumption **(A1)** holds*

$$\mathcal{R}\Big[\widehat{f}_{m_n^\star} \,|\, \mathcal{F}_{\mathfrak{a}}^r \cap \mathcal{D}\Big] \leqslant (\tau_\infty^2 + r^2)\, \mathcal{R}_n^\star, \quad \textit{for all } n \geqslant 1. \tag{3.1}$$

PROPOSITION 3.3 (Lower bound). *Suppose $\{X_i\}_{i=1}^n$ is an iid. $n-$sample. Let the assumption **(A2)** holds true and assume further that*

$$0 < \eta := \inf_{n \geqslant 1}\{(\mathcal{R}_n^\star)^{-1} \min(\mathfrak{a}_{m_n^\star}, m_n^\star n^{-1}) \leqslant 1 \tag{3.2}$$

*then for all $n \geqslant 2$ we have*

$$\inf_{\widetilde{f}} \mathcal{R}\Big[\widetilde{f} \,|\, \mathcal{F}_{\mathfrak{a}}^r \cap \mathcal{D}\Big] \geqslant \tfrac{\eta}{8} \min(r-1, (4\mathfrak{A})^{-1})\, \mathcal{R}_n^\star \tag{3.3}$$

*where the infimum is to be taken over all possible estimators $\widetilde{f}$ of $f$.*

Note that in the configurations considered in the Illustration 1 the additional condition (3.2) is always satisfied. Comparing the upper bound (3.1) and the lower bound (3.3) we have shown that $\mathcal{R}_n^\star$ is the minimax-optimal rate of convergence and the estimator $\widehat{f}_{m_n^\star}$ is minimax-optimal.

**Fully data-driven estimator.** We consider the fully-data-driven estimator $\widehat{f}_{\widetilde{m}}$ where $\widetilde{m}$ is defined in (2.3) with $\mathrm{pen}_m := 36\tau_\infty^2 m n^{-1}$ which satisfies trivially the condition **(C1)**. The proof of the next Proposition is based on Talagrand's inequality (Lemma 3.1).

PROPOSITION 3.4. *Let $\{X_i\}_{i=1}^n$ be an iid. $n-$sample. Suppose that the assumptions **(A1)** and **(A2)** are satisfied. There exists a numerical constant $C > 0$ such that*

$$\sup_{f \in \mathcal{F}_{\mathfrak{a}}^r \cap \mathcal{D}} \mathbb{E}\Big\{\max_{1 \leqslant m \leqslant n} \Big(\|\widehat{f}_m - f_m\|_{L_2}^2 - 6\tau_\infty^2 m n^{-1}\Big)_+\Big\} \leqslant C n^{-1} \tau_\infty^2 \zeta(r\mathfrak{A}/\tau_\infty^2)$$

*where $\zeta(x) := 1 + x\sum_{m=1}^\infty \exp(-m/(6\sqrt{2}x))$, for any $x > 0$.*

By using the definition of the penalty term the last Proposition implies that the condition **(C2)** is satisfied. Thereby, the next assertion is an immediate consequence of Proposition 2.2 and we omit its proof.

THEOREM 3.5. *Suppose $\{X_i\}_{i=1}^n$ is an iid. $n-$sample. Let **(A1)** and **(A2)** be satisfied. Select the dimension parameter $\widetilde{m}$ as given by (2.3) with $\mathrm{pen}_m := 36\tau_\infty^2 m n^{-1}$. There exists a numerical constant $C > 0$ such that for all $n \geqslant n_\diamond$ with $\mathcal{R}_{n_\diamond}^\star \leqslant 1$ we have*

$$\mathcal{R}\Big[\widehat{f}_{\widetilde{m}} \,|\, \mathcal{F}_{\mathfrak{a}}^r \cap \mathcal{D}\Big] \leqslant C\,[r \vee \tau_\infty^2 \vee \tau_\infty^2 \zeta(r\mathfrak{A}/\tau_\infty^2)]\, \mathcal{R}_n^\star.$$

The last assertion establishes the minimax-optimality of the data-driven estimator $\widehat{f}_{\widetilde{m}}$ over all classes $\mathcal{F}_{\mathfrak{a}}^r \cap \mathcal{D}$ where $\mathfrak{a}$ is a monotonically non-increasing and strictly positive sequence of weights tending to zero. Therefore, the fully data-driven estimator is called adaptive.

### 3.2 Non-parametric regression

In this paragraph we suppose that the identically distributed $n$-sample $\{(Y_i, U_i)\}_{i=1}^n$ consists of independent random variables.

PROPOSITION 3.6. *Let $\{(Y_i, U_i)\}_{i=1}^n$ be an iid. $n-$sample. Under the assumption **(A1)** holds*

$$\mathcal{R}\Big[\widehat{f}_{m_n^\star} \,|\, \mathcal{F}_{\mathfrak{a}}^r\Big] \leqslant (\tau_\infty^2(\sigma^2 + r^2) + r^2)\, \mathcal{R}_n^\star, \quad \textit{for all } n \geqslant 1, \tag{3.4}$$

PROPOSITION 3.7. *Suppose $\{(Y_i, U_i)\}_{i=1}^n$ is an iid. $n-$sample. Let the error term be normally distributed and assume further that*

$$0 < \eta := \inf_{n \geqslant 1}\{(\mathcal{R}_n^\star)^{-1} \min(\mathfrak{a}_{m_n^\star}, m_n^\star n^{-1}) \leqslant 1, \tag{3.5}$$

*then for all $n \geqslant 1$ we have*

$$\inf_{\widetilde{f}} \mathcal{R}\Big[\widetilde{f} \,|\, \mathcal{F}_{\mathfrak{a}}^r\Big] \geqslant \frac{\eta}{8} \min(2r^2, \sigma^2)\, \mathcal{R}_n^\star \tag{3.6}$$

*where the infimum is to be taken over all possible estimators $\widetilde{f}$ of $f$.*

Again in the configurations considered in the Illustration 1 the condition (3.5) hold true. Combining the upper bound (3.4) and the lower bound (3.6) we have shown that $\mathcal{R}_n^\star$ is the minimax-optimal $\mathcal{R}\Big[\widehat{f}_{\widetilde{m}} \,|\, \mathcal{F}_{\mathfrak{a}}^r\Big]$ by apply the Proposition 2.2. rate of convergence and the estimator $\widehat{f}_{m_n^\star}$ is minimax-optimal.

**Partially data-driven estimator.** In this paragraph, we select the dimension parameter following the procedure sketched in (2.3) where the subsequence of non-negative and non-decreasing penalties $(\mathrm{pen}_1, \dots, \mathrm{pen}_n)$ is given by $\mathrm{pen}_m = 144\sigma_Y^2\tau_\infty^2 m n^{-1}$ with $\sigma_Y^2 = \mathbb{E}Y^2$. Since $\sigma_Y$ has to be estimated from the data, the considered selection method leads to a partially data-driven estimator of the non-parametric regression function $f$ only. In order to apply the Proposition 2.2 it remains to check the conditions **(C1)** and **(C2)**. Keeping in mind the definition of the penalties subsequence, the condition **(C1)** is obviously satisfied. The next Proposition provides our key argument to verify the condition **(C2)**.

PROPOSITION 3.8. *Let $\{(Y_i, U_i)\}_{i=1}^n$ be an iid. $n-$sample. Suppose that the assumptions **(A1)** and **(A2)** are satisfied. If $\mathbb{E}\varepsilon^6 < \infty$ then there exists a finite constant $C(r\mathfrak{A}, \sigma, \tau_\infty, \mathbb{E}\varepsilon^6)$ depending only on the quantities $r\mathfrak{A}$, $\sigma$, $\tau_\infty$ and $\mathbb{E}\varepsilon^6$ such that*

$$\sup_{f \in \mathcal{F}_{\mathfrak{a}}^r} \mathbb{E}\Big\{\max_{1 \leqslant m \leqslant n} \Big(\|\widehat{f}_m - f_m\|_{L_2}^2 - 12\tau_\infty^2\sigma_Y^2 m n^{-1}\Big)_+\Big\} \leqslant n^{-1} C(r\mathfrak{A}, \sigma, \tau_\infty, \mathbb{E}\varepsilon^6), \quad \textit{for all } n \geqslant 1.$$

Obviously, taking into account the definition of penalties sequence the last Proposition shows that the condition **(C2)** is satisfied. Thereby, the next assertion is an immediate consequence of Proposition 2.2 and we omit its proof.

PROPOSITION 3.9. *Suppose* $\{(Y_i, U_i)\}_{i=1}^n$ *is an iid.* $n-$*sample. Let assumptions* ***(A1)*** *and* ***(A2)*** *be satisfied. Select the dimension parameter* $\widetilde{m}$ *as given by* (2.3) *with* $\mathrm{pen}_m := 72\tau_\infty^2\sigma_Y^2 mn^{-1}$. *If* $\mathbb{E}\varepsilon^6 < \infty$ *then there exists a numerical constant* $C$ *and a finite constant* $\zeta(r\mathfrak{A}, \sigma, \tau_\infty, \mathbb{E}\varepsilon^6)$ *depending only on the quantities* $r\mathfrak{A}$, $\sigma$, $\tau_\infty$ *and* $\mathbb{E}\varepsilon^6$ *such that for all* $n \geqslant n_\diamond$ *with* $\mathcal{R}_{n_\diamond}^\star \leqslant 1$ *we have*

$$\mathcal{R}\Big[\widehat{f}_{\widetilde{m}} \,|\, \mathcal{F}_{\mathfrak{a}}^r\Big] \leqslant C[r^2 \vee \tau_\infty^2\sigma_Y^2 \vee \zeta(r\mathfrak{A}, \sigma, \tau_\infty, \mathbb{E}\varepsilon^6)]\ \mathcal{R}_n^\star.$$

Since $\sigma_Y^2 = \mathbb{E}Y^2$ is generally unknown, the penalty term specified in the last assertion is not feasible. However, we have a natural estimator $\widehat{\sigma}_Y^2 = n^{-1}\sum_{i=1}^n Y_i^2$ of the quantity $\sigma_Y^2$ at hand.

**Fully data-driven estimator.** In the sequel we consider the subsequence of non-negative and non-decreasing penalties $(\widehat{\mathrm{pen}}_1, \dots, \widehat{\mathrm{pen}}_n)$ given by $\widehat{\mathrm{pen}}_m = 144\widehat{\sigma}_Y^2\tau_\infty^2 mn^{-1}$. The dimension parameter $\widehat{m}$ is then selected as in (2.7). Keeping in mind the Proposition 2.4 it remains to show that the condition **(C3)** holds true. Therefore, define further the event $\mathcal{V} := \{1/2 \leqslant \widehat{\sigma}_Y^2/\sigma_Y^2 \leqslant 3/2\}$ and denote by $\mathcal{V}^c$ its complement.

LEMMA 3.10. *Let* $\{(Y_i, U_i)\}_{i=1}^n$ *be an iid.* $n$*-sample. If* $\mathbb{E}\varepsilon^4 < \infty$*, then* $\sup_{f\in\mathcal{F}_{\mathfrak{a}}^r} P(\mathcal{V}^c) \leqslant 128n^{-1}\Big((\mathbb{E}\varepsilon^4)^{1/4} + r\mathfrak{A}/\sigma\Big)^4$.

Considering the event $\Omega$ given in (2.9) it is easily seen that $\mathcal{V} \subset \Omega$ and hence, by employing the last assertion together with Proposition 3.8 the conditions **(C1)**-**(C3)** are satisfied. Thereby, the next assertion is an immediate consequence of Proposition 2.4 and we omit its proof.

THEOREM 3.11. *Suppose* $\{(Y_i, U_i)\}_{i=1}^n$ *is an iid.* $n-$*sample. Let assumptions* ***(A1)*** *and* ***(A2)*** *be satisfied. Select the dimension parameter* $\widehat{m}$ *as given by* (2.7) *with* $\widehat{\mathrm{pen}}_m := 144\tau_\infty^2\widehat{\sigma}_Y^2 mn^{-1}$. *If* $\mathbb{E}\varepsilon^6 < \infty$ *then there exists a numerical constant* $C$ *and a finite constant* $\zeta(r\mathfrak{A}, \sigma, \tau_\infty, \mathbb{E}\varepsilon^6)$ *depending only on the quantities* $r\mathfrak{A}$, $\sigma$, $\tau_\infty$ *and* $\mathbb{E}\varepsilon^6$ *such that for all* $n \geqslant n_\diamond$ *with* $\mathcal{R}_{n_\diamond}^\star \leqslant 1$ *we have*

$$\mathcal{R}\Big[\widehat{f}_{\widehat{m}} \,|\, \mathcal{F}_{\mathfrak{a}}^r\Big] \leqslant C[r^2 \vee \tau_\infty^2\sigma_Y^2 \vee \zeta(r\mathfrak{A}, \sigma, \tau_\infty, \mathbb{E}\varepsilon^6)]\ \mathcal{R}_n^\star.$$

We shall emphasise that the last assertion establishes the minimax-optimality of the fully data-driven estimator $\widehat{f}_{\widehat{m}}$ over all classes $\mathcal{F}_{\mathfrak{a}}^r$. Therefore, the estimator is called adaptive.

# 4 Dependent observations

In this section we dismiss the independence assumption and assume weakly dependent observations. More precisely, $Z_1, \dots, Z_n$ are drawn from a strictly stationary process $(Z_i)_{i\in\mathbb{Z}}$ taking still its values in $[0, 1]$. Keep in mind that a process is called strictly stationary if its finite dimensional distributions does not change when shifted in time. Consequently, the random variables $\{Z_i\}$ are identically distributed. Our aim is the non-parametric estimation of the function $f$ under some mixing conditions on the dependence of the process $(Z_i)_{i\in\mathbb{Z}}$. Let us begin with a brief review of a classical measure of dependence, leading to the notion of a stationary absolutely regular process.

Let $(\Omega, \mathscr{A}, P)$ be a probability space. Given two $\sigma$-algebras $\mathscr{U}$ and $\mathscr{V}$ of $\mathscr{A}$ we introduce next the definition and properties of the absolutely regular mixing (or $\beta$-mixing) coefficient $\beta(\mathscr{U}, \mathscr{V})$. The coefficient was introduced by Kolmogorov and Rozanov [1960] and is defined by

$$\beta(\mathscr{U}, \mathscr{V}) := \tfrac{1}{2} \sup \left\{ \sum_i \sum_j |P(U_i)P(V_i) - P(U_i \cap V_i)| \right\}$$

where the supremum is taken over all finite partitions $(U_i)_{i\in I}$ and $(V_j)_{j\in J}$, which are respectively $\mathscr{U}$ and $\mathscr{V}$ measurable. Obviously, $\beta(\mathscr{U}, \mathscr{V}) \leqslant 1$. As usual, if $Z$ and $Z'$ are two real-valued random variables, we denote by $\beta(Z, Z')$ the mixing coefficient $\beta(\sigma(Z), \sigma(Z'))$, where $\sigma(Z)$ and $\sigma(Z')$ are, respectively, the $\sigma$-fields generated by $Z$ and $Z'$. Consider a strictly stationary process $(Z_i)_{i\in\mathbb{Z}}$ then for any integer $k$ the mixing coefficient $\beta(Z_0, Z_k)$ does not change when shifted over time, i.e., $\beta(Z_0, Z_k) = \beta(Z_{0+l}, Z_{k+l})$ for all integer $l$. The next assertion follows along the lines of the proof of Theorem 2.1 in Viennet [1997] and we omit its proof.

LEMMA 4.1. *Let $(Z_i)_{i\in\mathbb{Z}}$ be a strictly stationary process of real-valued random variables. There exists a sequence $(b_k)_{k\geqslant 1}$ of measurable functions $b_k : \mathbb{R} \to [0,1]$ with $\mathbb{E}b_k(Z_0) = \beta(Z_0, Z_k)$ such that for any measurable function $h$ with $\mathbb{E}|h(Z_0)|^2 < \infty$ and any integer $n$,*

$$\mathbb{V}\mathrm{ar}(\sum_{i=1}^{n} h(Z_i)) \leqslant n\,\mathbb{E}\Big\{ |h(Z_0)|^2 \Big(1 + 4\sum_{k=1}^{n-1} b_k(Z_0)\Big)\Big\}.$$

Given $p \geqslant 2$, a non-negative sequence $w := (w_k)_{k\geqslant 0}$ and a probability measure $P$ let $\mathscr{L}(p, w, P)$ be the set of functions $b : \mathbb{R} \to [0, \infty]$ such that there exists a sequence $(b_k)_{k\geqslant 0}$ of measurable functions $b_k : \mathbb{R} \to [0,1]$, with $b_0 = \mathbb{1}$ and $\mathbb{E}_P b_k \leqslant w_k$ satisfying $b = \sum_{k=0}^{\infty} (k+1)^{p-2} b_k$. We note that the elements of $\mathscr{L}(p, w, P)$ are generally not $P$-integrable, however, whenever $\sum_{k=0}^{\infty} (k+1)^{p-2} w_k < \infty$, each function $b$ in $\mathscr{L}(p, w, P)$ is a non-negative $P$-integrable function. Moreover, reconsidering a strictly stationary process $(Z_i)_{i\in\mathbb{Z}}$ with common marginal distribution $P_{Z_0}$ and associated non-negative sequence of mixing coefficients $w = (w_k)_{k\geqslant 0}$ with $w_0 = 1$ and $w_k = \beta(Z_0, Z_k)$ an immediate consequence of Lemma 4.1 is the existence of a function $b$ belonging to $\mathscr{L}(2, \beta, P_{Z_0})$ such that for any measurable function $h$ with $\mathbb{E}|h(Z_0)|^2 < \infty$ and any integer $n$,

$$\mathbb{V}\mathrm{ar}(\sum_{i=1}^{n} h(Z_i)) \leqslant 4n\mathbb{E}(|h(Z_0)|^2 b(Z_0)). \tag{4.1}$$

Note that the assumptions stated yet do not ensure that the right hand side in the last display is finite. However, the function $b$ is $P_{Z_0}$-integrable whenever $\sum_{k\geqslant 1}\beta(Z_0, Z_k) < \infty$. Therefore, imposing in addition that $\sum_{k\geqslant 1}\beta(Z_0, Z_k) < \infty$ and, for example, that $\|h\|_\infty < \infty$ we have $\mathbb{E}(h(Z_0)|^2 b(Z_0)| \leqslant \|h\|_\infty \mathbb{E} b(Z_0) < \infty$. Obviously, given conjugate exponents $p$ and $q$ if $b$ has a finite $p$-th moment, i.e., $\mathbb{E}|b(Z_0)|^p < \infty$, and $\mathbb{E}|h(Z_0)|^{2q} < \infty$, then we have $\mathbb{E}(h(Z_0)|^2 b(Z_0)| \leqslant \{\mathbb{E}|h(Z_0)|^{2q}\}^{1/q}\{\mathbb{E}|b(Z_0)|^p\}^{1/p} < \infty$. Lemma 4.2 in Viennet [1997] provides now sufficient conditions to ensure the existence of a finite $p$-th moment of $b$ which is summarized in the next assertion.

LEMMA 4.2. *Let the sequence $w := (w_k)_{k\geqslant 0}$ be non-increasing, tending to $0$ as $k \to \infty$ with $w_0 = 1$ and such that $\sum_{k=0}^{\infty}(k+1)^{p-1}w_k < \infty$ for some $1 \leqslant p \leqslant \infty$. Then, for each $b$ in $\mathscr{L}(2, w, P)$ the function $b^p$ is $P$-integrable and $\mathbb{E}_P|b|^p \leqslant p\sum_{k=0}^{\infty}(k+1)^{p-1}w_k$.*

We will use Lemma 4.1, the estimate (4.1) together with Lemma 4.2 to derive an upper bound for the maximal risk of the non-parametric estimator with suitable choice of the dimension parameter. However, in order to control the deviation of the data-driven estimator, more precisely in order to show that the condition **(C2)** holds true, we have made use of Talagrand's inequality which is formulated for independent observations only. Inspired by the work of Comte et al. [2008] we will use coupling techniques to extend Talagrand's inequality to dependent data which we present next. We assume in the sequel that there exists a sequence of independent random variables with uniform distribution on $[0, 1]$ independent of the sequence $(Z_i)_{i\geqslant 1}$. Employing Lemma 5.1 in Viennet [1997] we construct by induction a sequence $(Z_i^\perp)_{i\geqslant 1}$ satisfying the following properties. Given an integer $q$ we introduce disjoint even and odd blocks of indices, i.e., for any $l \geqslant 1$, $\mathcal{I}_l^e := \{2(l-1)q+1, \dots, (2l-1)q\}$ and $\mathcal{I}_l^o := \{(2l-1)q+1, \dots, 2lq\}$, respectively, of size $q$. Let us further partition into blocks the random processes $(Z_i)_{i\geqslant 1} = (E_l, O_l)_{l\geqslant 1}$ and $(Z_i^\perp)_{i\geqslant 1} = (E_l^\perp, O_l^\perp)_{l\geqslant 1}$ where

$$E_l = (Z_i)_{i\in\mathcal{I}_l^e}, \qquad E_l^\perp = (Z_i^\perp)_{i\in\mathcal{I}_l^e}, \qquad O_l = (Z_i)_{i\in\mathcal{I}_l^o}, \qquad O_l^\perp = (Z_i^\perp)_{i\in\mathcal{I}_l^o}.$$

If we set further $\mathscr{F}_l^- := \sigma(Z_j, j \leqslant l)$ and $\mathscr{F}_l^+ := \sigma(Z_j, j \geqslant l)$, then the sequence $(\beta_k)_{k\geqslant 0}$ of $\beta$-mixing coefficient defined by $\beta_0 := 1$ and $\beta_k := \beta(\mathscr{F}_0^-, \mathscr{F}_k^+)$, $k \geqslant 1$, is monotonically non-increasing and satisfies trivially $\beta_k \geqslant \beta(Z_0, Z_k)$ for any $k \geqslant 1$. Based on the construction presented in Viennet [1997], the sequence $(Z_i^\perp)_{i\geqslant 1}$ can be chosen such that for any integer $l \geqslant 1$:

**(P1)** $E_l^\perp$, $E_l$, $O_l^\perp$ and $O_l$ are identically distributed,

**(P2)** $P(E_l \neq E_l^\perp) \leqslant \beta_{q+1}$, and $P(O_l \neq O_l^\perp) \leqslant \beta_{q+1}$.

**(P3)** The variables $(E_1^\perp, \dots, E_l^\perp)$ are iid. and so $(O_1^\perp, \dots, O_l^\perp)$.

We may emphasise that the random vectors $E_1^\perp, \dots, E_l^\perp$ are iid. but the components within each vector are generally not independent.

## 4.1 Non-parametric density estimation

Let us turn our attention back to the orthogonal series estimator defined in the paragraph 2.2. Keep in mind that $X_1, \dots, X_n$ are drawn from a strictly stationary process $(X_i)_{i\in\mathbb{Z}}$ with common marginal distribution admitting a density $f$. Exploiting the assumption **(A1)** and Lemma 4.1 we obtain the next assertion

PROPOSITION 4.3 (Upper bound). *Let $(X_i)_{i\in\mathbb{Z}}$ be a strictly stationary process with associated sequence of mixing coefficients $\{\beta(X_0, X_k)\}_{k\geqslant 1}$. Under assumption **(A1)** holds*

$$\mathcal{R}\Big[\widehat{f}_{m_n^\star} \,|\, \mathcal{F}_{\mathfrak{a}}^r \cap \mathcal{D}\Big] \leqslant (\tau_\infty^2\{1 + 4\sum_{k=1}^{n-1}\beta(X_0, X_k)\} + r^2)\, \mathcal{R}_n^\star, \quad \textit{for all } n \geqslant 1. \tag{4.2}$$

Let us compare briefly the last result and the upper risk bound assuming independent observations given in Proposition 3.2. We see, that this upper risk bound provides up to finite constant also an upper risk bound in the presence of dependence whenever $\sum_{k=1}^{\infty}\beta(X_0, X_k) < \infty$. However, the upper bound given in Proposition 4.3 depends on the unknown mixing coefficients $\{\beta(X_0, X_k)\}_k$. Their estimation is a demanding task, and hence, we next derive an upper bound which does not depend on the mixing coefficients at least for all sufficiently large sample sizes $n$. This upper bound relies on the next assumption which has been used, for example, in Bosq [1998].

**(D1)** For any integer $k$ the joint distribution $P_{X_0,X_k}$ of $(X_0, X_k)$ admits a density $f_{X_0,X_k}$ which is square integrable. Let $\|f_{X_0,X_k}\|^2 := \int_0^1\int_0^1 |f_{X_0,X_k}(x,y)|^2 dxdy < \infty$ with a slight abuse of notations. If we denote further by $h \otimes g : [0,1]^2 \to \mathbb{R}$ the bivariate function $[h\otimes g](x,y) := h(x)g(y)$, then let $\gamma_f := \sup_{k\geqslant 1}\|f_{X_0,X_k} - f\otimes f\| < \infty$.

LEMMA 4.4. *Let $(X_i)_{i\in\mathbb{Z}}$ be a strictly stationary process with associated sequence of mixing coefficients $\{\beta(X_0, X_k)\}_{k\geqslant 1}$. Under the assumptions **(A1)** and **(D1)** for any $n \geqslant 1$ and $K \in \{0, \dots, n-1\}$ it holds*

$$\sum_{j=1}^{m}\mathbb{V}\mathrm{ar}(\sum_{i=1}^{n}\phi_j(X_i)) \leqslant nm\{\tau_\infty^2 + 2[\gamma_f K/\sqrt{m} + 2\tau_\infty^2\sum_{k=K+1}^{n-1}\beta(X_0, X_k)]\}. \tag{4.3}$$

If we assume in addition that $\sum_{k=1}^{\infty}\beta(X_0, X_k) < \infty$ and $\gamma := \sup_{f\in\mathcal{F}_{\mathfrak{a}}^r\cap\mathcal{D}}\gamma_f < \infty$ then there exist an integer $K_o$ and an integer $n_o$ such that $\sum_{k=K_o+1}^{\infty}\beta(X_0, X_k) < 1/8$ and $K_n := \lfloor 4\tau_\infty^2\sqrt{m_n^\star}/\gamma\rfloor \geqslant K_o$ with $m_n^\star$ as given in (2.2) for all $n \geqslant n_o$. Thereby, we have for all $n \geqslant n_o$ that $\sum_{j=1}^{m_n^\star}\mathbb{V}\mathrm{ar}(\sum_{i=1}^{n}\phi_j(X_i)) \leqslant \tau_\infty^2 n\, m_n^\star$. We note that $n_o$ depends on the sequence of mixing coefficients. The next assertion is an immediate consequence and we omit its proof.

PROPOSITION 4.5 (Upper bound). *Let $(X_i)_{i\in\mathbb{Z}}$ be a strictly stationary process with associated sequence of mixing coefficients $\{\beta(X_0, X_k)\}_{k\geqslant 1}$. Under Condition **(A1)** and **(D1)** if*

$\sum_{k=1}^{\infty} \beta(X_0, X_k) < \infty$ *and* $\gamma := \sup_{f \in \mathcal{F}_{\mathfrak{a}}^r \cap \mathcal{D}} \gamma_f < \infty$ *then there exists an integer* $n_o$ *(possibly depending on the mixing coefficients and* $\gamma$*) such that*

$$\mathcal{R}\Big[\widehat{f}_{m_n^\star} \,|\, \mathcal{F}_{\mathfrak{a}}^r \cap \mathcal{D}\Big] \leqslant (\tau_\infty^2 + r^2)\, \mathcal{R}_n^\star, \quad \textit{for all } n \geqslant n_o. \tag{4.4}$$

Consequently under the condition of Proposition 4.5 the estimator $\widehat{f}_{m_n^\star}$ attains the minimax-optimal rate $\mathcal{R}_n^\star$ for independent data

**Fully data-driven estimator.** Consider the estimator $\widehat{f}_{\widetilde{m}}$ where $\widetilde{m}$ is defined in (2.3) with $\mathrm{pen}_m := 288\tau_\infty^2 m n^{-1}$. We aim to derive an upper bound for its maximal risk $\mathcal{R}\Big[\widehat{f}_{\widetilde{m}} \,|\, \mathcal{F}_{\mathfrak{a}}^r \cap \mathcal{D}\Big]$ by making use of Proposition 2.2. Therefore, it remains to check the conditions **(C1)** and **(C2)** where **(C1)** holds obviously true due to the definition of penalty term. The next assertion provides our key argument in order to verify the condition **(C2)**.

PROPOSITION 4.6. *Let* $(X_i)_{i\in\mathbb{Z}}$ *be a strictly stationary process with associated sequence of mixing coefficients* $(\beta_k)_{k\geqslant 1}$ *satisfying* $\mathfrak{B} := 2\sum_{k=0}^{\infty}(k+1)\beta_k < \infty$*. Under the assumptions (A1), (A2) and (D1), let* $\gamma := \sup_{f\in\mathcal{F}_{\mathfrak{a}}^r\cap\mathcal{D}} \gamma_f < \infty$*,* $K_n := \lfloor 4\tau_\infty^2 \sqrt{m_n^\star}/\gamma \rfloor$ *and* $\mu_n \geqslant \{3 + 8\sum_{k=K_n+1}^{\infty}\beta_k\}$*. There exists a numerical constant* $C > 0$ *such that for any integer* $q$

$$\sup_{f\in\mathcal{F}_{\mathfrak{a}}^r} \mathbb{E}\Big\{\max_{m_n^\star \leqslant m \leqslant n}\Big(\|\widehat{f}_m - f_m\|^2 - 12\tau_\infty^2 m n^{-1}\mu_n\Big)_+\Big\} \\ \leqslant C\, n^{-1}\tau_\infty^2\left\{\mu_n \Psi\left(\frac{r\mathfrak{A}\mathfrak{B}}{\tau_\infty^2\mu_n^2}\right) + nq^2\exp\left(-\frac{n^{1/2}}{q}\frac{\mu_n^{1/2}}{144}\right) + n^2\beta_{q+1}\right\} \tag{4.5}$$

*where* $\Psi(x) := \sum_{m\geqslant 1}^{\infty} x^{1/2} m^{1/2}\exp(-m^{1/2}/(48x^{1/2})) < \infty$*, for any* $x > 0$.

Note that the condition $\mathfrak{B} = 2\sum_{k=0}^{\infty}(k+1)\beta_k < \infty$ implies $\sum_{k=K_n+1}^{\infty}\beta_k \leqslant (K_n+1)^{-1}\mathfrak{B}$ and hence, $\{3 + 8\sum_{k=K_n+1}^{\infty}\beta_k\} \leqslant 4$ whenever $K_n = \lfloor 4\tau_\infty^2\sqrt{m_n^\star}/\gamma\rfloor \geqslant 8\mathfrak{B}$. Since $m_n^\star \to \infty$ as $n \to \infty$ there exists an integer $n_o$ such that for all $n \geqslant n_o$ we can chose $\mu_n = 4$. The next assertion is thus an immediate consequence of Proposition 4.6, and hence we omit its proof.

COROLLARY 4.7. *Let the assumptions of Proposition 4.6 be satisfied. Suppose that there exists an unbounded sequence of integers* $(q_n)_{n\geqslant 1}$ *and a finite constant* $L > 0$ *such that*

$$\sup_{n\geqslant 1} nq_n^2\exp\left(-\frac{n^{1/2}}{q_n}\frac{1}{72}\right) \leqslant L \quad \textit{and} \quad \sup_{n\geqslant 1} n^2\beta_{q_n+1} \leqslant L. \tag{4.6}$$

*There exist a numerical constant* $C > 0$ *and an integer* $n_o$ *such that for all* $n \geqslant n_o$

$$\sup_{f\in\mathcal{F}_{\mathfrak{a}}^r} \mathbb{E}\Big\{\max_{m_n^\star\leqslant m\leqslant n}\Big(\|\widehat{f}_m - f_m\|^2 - 48\tau_\infty^2 m n^{-1}\Big)_+\Big\} \leqslant Cn^{-1}\tau_\infty^2\left\{\Psi\left(\frac{r\mathfrak{A}\mathfrak{B}}{16\tau_\infty^2}\right) + L\right\}.$$

Is it interesting to note that an arithmetically decaying sequence of mixing coefficients $(\beta_k)_{k\geqslant 1}$ satisfies (4.6). To be more precise, consider two sequence of integers $(q_n)_{n\geqslant 1}$, $(p_n)_{n\geqslant 1}$

such that $n = 2q_np_n$ and assume additionally $\beta_k \leqslant k^{-s}$. The sequence $q_n \asymp n^{p_n}$, i.e., $(n^{-p_n}q_n)_{n\geqslant 1}$ is bounded away both from zero and infinity, and satisfies the condition (4.6) whenever $2 < p_n s$ and $1/2 > p_n$. In other words, if the sequence of mixing coefficients $(\beta_k)_{k\geqslant 1}$ is sufficiently fast decaying, that is $s > 2(2+\theta)$ for some $\theta > 0$, then the condition (4.6) holds true taking, for example, a sequence $q_n \asymp n^{1/(2+\theta)}$.

Obviously, using the penalty $\mathrm{pen}_m := 288\tau_\infty^2 mn^{-1}$ for any $m \in \mathbb{N}$ the conditions **(C1)** and **(C2)** due to Proposition 4.7 are satisfied. Thereby, the next assertion is an immediate consequence of Proposition 2.2 and we omit its proof.

THEOREM 4.8. *Under the assumptions of Proposition 4.6 and the condition (4.6) there exist a numerical constant $C > 0$ and an integer $n_o$ such that for all $n \geqslant n_o$ we have*

$$\mathcal{R}\Big[\widehat{f}_{\widehat{m}} \,|\, \mathcal{F}_{\mathfrak{a}}^r \cap \mathcal{D}\Big] \leqslant C\left[r \vee \tau_\infty^2 \vee \tau_\infty^2\Big\{\Psi\Big(\frac{r\mathfrak{A}\mathfrak{B}}{16\tau_\infty^2}\Big) + L\Big\}\right]\mathcal{R}_n^\star.$$

Note that the penalty term depends only on known quantities and, hence the $\widehat{f}_{\widehat{m}}$ is fully data-driven. The last assertion establishes the minimax-rate optimality of the fully data-driven estimator $\widehat{f}_{\widehat{m}}$ over all classes $\mathcal{F}_{\mathfrak{a}}^r \cap \mathcal{D}$. Therefore, the estimator is called adaptive.

## 4.2 Non-parametric regression

Let us turn our attention to the orthogonal series estimator defined in the paragraph 2.2. In the sequel we suppose that the explanatory variables $U_1, \dots, U_n$ are drawn from a strictly stationary process $(U_i)_{i\in\mathbb{Z}}$ with common marginal uniform distribution on the interval $[0, 1]$. Moreover, we still assume that the error terms $\{\varepsilon_i\}_{i=1}^n$ are iid. and independent to the explanatory variables. Exploiting the assumption **(A1)** and Lemma 4.1 we obtain the next assertion

PROPOSITION 4.9 (Upper bound). *Let $(U_i)_{i\in\mathbb{Z}}$ be a strictly stationary process with associated sequence of mixing coefficients $\{\beta(U_0, U_k)\}_{k\geqslant 1}$. Under **(A1)** holds*

$$\mathcal{R}\Big[\widehat{f}_{m_n^\star} \,|\, \mathcal{F}_{\mathfrak{a}}^r\Big] \leqslant (\sigma^2 + \|f\|_\infty^2\tau_\infty^2\{1 + 4\sum_{k=1}^{n-1}\beta(U_0, U_k)\} + r^2)\,\mathcal{R}_n^\star, \quad \textit{for all } n \geqslant 1. \tag{4.7}$$

Comparing the last result and Proposition 3.6 the upper risk bound assuming independent observations provides up to a finite constant also an upper risk bound in the presence of dependence whenever $\sum_{k=1}^\infty \beta(U_0, U_k) < \infty$.

**(D2)** For any integer $k$ the joint distribution $P_{U_0,U_k}$ of $(U_0, U_k)$ admits a density $f_{U_0,U_k}$ which is square integrable and satifies $\gamma := \sup_{k\geqslant 1}\|f_{U_0,U_k} - \mathbb{1}\otimes\mathbb{1}\| < \infty$.

LEMMA 4.10. *Let $(U_i)_{i\in\mathbb{Z}}$ be a strictly stationary process with associated sequence of mixing coefficients $\{\beta(U_0, U_k)\}_{k\geqslant 1}$. Under assumptions **(A1)** and **(D2)** holds for any $n \geqslant 1$ and*

$K \in \{0, \dots, n-1\}$

$$\sum_{j=1}^{m} \mathbb{V}\mathrm{ar}(\sum_{i=1}^{n} f(U_i)\phi_j(U_i)) \leqslant nm\{\tau_\infty^2 \|f\|_{L_2}^2 + 2\|f\|_\infty^2[\gamma K/\sqrt{m} + 2\tau_\infty^2 \sum_{k=K+1}^{n-1} \beta(U_0, U_k)]\}. \tag{4.8}$$

Note that supposing further assumption **(A2)** we have $\|f\|_\infty^2 \leqslant r^2\mathfrak{A}^2$ for all $f \in \mathcal{F}_\mathfrak{a}^r$. If we assume in addition that $\sum_{k=1}^\infty \beta(U_0, U_k) < \infty$ then there exists an integer $K_o$ and an integer $n_o$ such that $\sum_{k=K_o+1}^\infty \beta(U_0, U_k) < 1/(8r^2\mathfrak{A}^2)$ and $K_n := \lfloor \tau_\infty^2 \sqrt{m_n^\star}/(\gamma r^2 \mathfrak{A}^2) \rfloor \geqslant K_o$ for all $n \geqslant n_o$. Thereby, we have for all $n \geqslant n_o$ that $\sum_{j=1}^{m_n^\star} \mathbb{V}\mathrm{ar}(\sum_{i=1}^n f(U_i)\phi_j(U_i)) \leqslant (r^2+1)\tau_\infty^2 n\, m_n^\star$ for all $f \in \mathcal{F}_\mathfrak{a}^r$. We note that $n_o$ depends on the sequence of mixing coefficients and the quantity $r\mathfrak{A}$. The next assertion is an immediate consequence and we omit its proof.

PROPOSITION 4.11 (Upper bound). *Let $(U_i)_{i\in\mathbb{Z}}$ be a strictly stationary process with associated sequence of mixing coefficients $\{\beta(U_0, U_k)\}_{k\geqslant 1}$. Let assumptions **(A1)**, **(A2)**, **(D2)** and $\sum_{k=1}^\infty \beta(U_0, U_k) < \infty$ be satisfied. There exists an integer $n_o$ (possibly depending on the mixing coefficients and the quantity $r\mathfrak{A}$) such that*

$$\mathcal{R}\Big[\widehat{f}_{m_n^\star} \,|\, \mathcal{F}_\mathfrak{a}^r\Big] \leqslant (\sigma^2 + (r^2+1)\tau_\infty^2 + r^2)\, \mathcal{R}_n^\star, \quad \textit{for all } n \geqslant n_o. \tag{4.9}$$

**Partially data-driven estimator.** In this paragraph, we select the dimension parameter following the procedure sketched in (2.3) where the subsequence of non-negative and non-decreasing penalties $(\mathrm{pen}_1, \dots, \mathrm{pen}_n)$ is given by $\mathrm{pen}_m = 1152\sigma_Y^2\tau_\infty^2 mn^{-1}$ with $\sigma_Y^2 = \mathbb{E}Y^2$. Since $\sigma_Y$ has to be estimated from the data, the considered selection method leads to a partially data-driven estimator of the non-parametric regression function $f$ only. In order to apply the Proposition 2.2 it remains to check the conditions **(C1)** and **(C2)**. Keeping in mind the definition of the penalties subsequence, the condition **(C1)** is obviously satisfied. The next Proposition provides our key argument to verify the condition **(C2)**.

PROPOSITION 4.12. *Let $(U_i)_{i\in\mathbb{Z}}$ be a strictly stationary process with associated sequence of mixing coefficients $(\beta_k)_{k\geqslant 1}$ satisfying $\mathfrak{B} := 2\sum_{k=0}^\infty (k+1)\beta_k < \infty$. Under the assumptions of Proposition 4.11, let $K_n := \lfloor 4\tau_\infty^2\|f\|_{L_2}^2\sqrt{m_n^\star}/(\gamma r^2\mathfrak{A}^2)\rfloor$ and $\mu_n \geqslant 3/2 + 4\sum_{k=K_n+1}^\infty \beta_k$. If $\mathbb{E}\varepsilon^6 < \infty$, then there exist a finite constant $\zeta(r\mathfrak{A}, \sigma, \tau_\infty, \mathfrak{B}, \mathbb{E}\varepsilon^6)$ depending on the quantities $r\mathfrak{A}$, $\sigma$, $\tau_\infty$, $\mathfrak{B}$ and $\mathbb{E}\varepsilon^6$ only and a numerical constant $C > 0$ such that for any integer $q$*

$$\sup_{f\in\mathcal{F}_\mathfrak{a}^r} \mathbb{E}\Big\{\max_{m_n^\star\leqslant m\leqslant n}\Big(\|\widehat{f}_m - f_m\|^2 - 24\tau_\infty^2 mn^{-1}\sigma_Y^2\mu_n\Big)_+\Big\}$$
$$\leqslant C\, n^{-1}(\sigma+r\mathfrak{A})^2\Big\{\zeta(r\mathfrak{A}, \sigma, \tau_\infty, \mathfrak{B}, \mathbb{E}\varepsilon^6) + n^{3/2}q^2\exp\Big(-\frac{n^{1/4}}{q}\frac{1}{576(1+r\mathfrak{A}/\sigma)}\Big)\Big\} + n^2\beta_{q+1}\Big\}.$$

Note that the condition $\mathfrak{B} = 2\sum_{k=0}^\infty(k+1)\beta_k < \infty$ implies $\sum_{k=K_n+1}^\infty \beta(U_0, U_k) \leqslant \sum_{k=K_n+1}^\infty \beta_k \leqslant (K_n+1)^{-1}\mathfrak{B}$ and hence, $\{3/2 + 4\sum_{k=K_n+1}^\infty \beta(U_0, U_k)\} \leqslant 2$ whenever

$K_n = \lfloor \tau_\infty^2 \sqrt{m_n^\star}/(\gamma r^2 \mathfrak{A}^2) \rfloor \geqslant 4\mathfrak{B}$. Since $m_n^\star \to \infty$ as $n \to \infty$ there exists an integer $n_o$ such that for all $n \geqslant n_o$ we can chose $\mu_n = 2$. The next assertion is thus an immediate consequence of Corollary 4.12, and hence we omit its proof.

COROLLARY 4.13. *Let the assumptions of Proposition 4.12 be satisfied. Suppose that there exists an unbounded sequence of integers $(q_n)_{n\geqslant 1}$ and a finite constant $L > 0$ such that*

$$\sup_{n\geqslant 1} n^{3/2} q_n^2 \exp\left(-\frac{n^{1/4}}{q_n}\frac{1}{576(1+r\mathfrak{A}/\sigma)}\right) \leqslant L \quad \textit{and} \quad \sup_{n\geqslant 1} n^2 \beta_{q_n+1} \leqslant L. \tag{4.10}$$

*Then there exist a numerical constant $C > 0$ and an integer $n_o$ such that for all $n \geqslant n_o$*

$$\sup_{f\in\mathcal{F}_\mathfrak{a}^r} \mathbb{E}\Big\{\max_{m_n^\star \leqslant m \leqslant n}\Big(\|\widehat{f}_m - f_m\|^2 - 48\tau_\infty^2 m n^{-1}\sigma_Y^2\Big)_+\Big\} \\ \leqslant C n^{-1}(\sigma + r\mathfrak{A})^2\Big\{\zeta(r\mathfrak{A}, \sigma, \tau_\infty, \mathfrak{B}, \mathbb{E}\varepsilon^6) + L\Big\}.$$

Let us briefly comment on the additional condition (4.10). Consider two sequence of integers $(q_n)_{n\geqslant 1}$, $(p_n)_{n\geqslant 1}$ such that $n = 2q_n p_n$ and assume additionally a polynomial decay of the sequence of mixing coefficients $(\beta_k)_{k\geqslant 1}$, that is $\beta_k \leqslant k^{-s}$. The sequence $q_n \asymp n^{p_n}$, i.e., $(n^{-p_n} q_n)_{n\geqslant 1}$ is bounded away both from zero and infinity, satisfies then the condition (4.10) if $2 < p_n s$ and $1/4 > p_n$. In other words, if the sequence of mixing coefficients $(\beta_k)_{k\geqslant 1}$ is sufficiently fast decaying, that is $s > 2(4+\theta)$ for some $\theta > 0$, then the condition (4.10) holds true taking a sequence $q_n \asymp n^{1/(4+\theta)}$.

Obviously taking into account Proposition 4.13 the conditions **(C1)** and **(C2)** are satisfied. Thereby, the next assertion is an immediate consequence of Proposition 2.2 and we omit its proof.

PROPOSITION 4.14. *Under the assumptions of Proposition 4.13 and the condition (4.10), there exist a numerical constant $C > 0$ and exists an integer $n_o$ such that for all $n \geqslant n_o$ we have*

$$\mathcal{R}\Big[\widehat{f}_{\widetilde{m}} \,|\, \mathcal{F}_\mathfrak{a}^r \cap \mathcal{D}\Big] \leqslant C\left[r^2 \vee \tau_\infty^2 \vee (\sigma + r\mathfrak{A})^2\Big\{\zeta(r\mathfrak{A}, \sigma, \tau_\infty, \mathfrak{B}, \mathbb{E}\varepsilon^6) + L\Big\}\right] \mathcal{R}_n^\star.$$

**Fully data-driven estimator.** Note that in general $\sigma_Y^2 = \mathbb{E}Y^2$ is unknown and hence the penalty term specified in the last assertion is not feasible, but it can be estimated straightforwardly by $\widehat{\sigma}_Y^2 = n^{-1}\sum_{i=1}^n Y_i^2$. Consequently, we consider next the sub-sequence of non-negative and non-decreasing penalties $(\widehat{\text{pen}}_1, \dots, \widehat{\text{pen}}_n)$ given by $\widehat{\text{pen}}_m = 1152\tau_\infty^2 m n^{-1}\widehat{\sigma}_Y^2$. $\widehat{\sigma}_Y^2 = n^{-1}\sum_{i=1}^n Y_i^2$ of the quantity $\sigma_Y^2$ at hand. The dimension parameter $\widehat{m}$ is then selected as in (2.7). Keeping in mind the Proposition 2.4 it remains to show that the Condition **(C3)** holds true. Consider again the event $\mathcal{V} := \{1/2 \leqslant \widehat{\sigma}_Y^2/\sigma_Y^2 \leqslant 3/2\}$ and its complement $\mathcal{V}^c$.

LEMMA 4.15. *Let $(U_i)_{i\in\mathbb{Z}}$ be a strictly stationary process with associated sequence of mixing coefficients $(\beta_k)_{k\geqslant 1}$. If $\mathbb{E}\varepsilon^4 < \infty$ and $\mathfrak{B} = 2\sum_{k=0}^\infty (k+1)\beta_k < \infty$, then $\sup_{f\in\mathcal{F}_\mathfrak{a}^r} P(\Omega^c) \leqslant 91 n^{-1}\sqrt{\mathfrak{B}}\Big[(\mathbb{E}\varepsilon^4)^{1/4} + r\mathfrak{A}/\sigma\Big]^2$.*

Considering the event $\Omega$ given in (2.9) it is easily seen that $\mathcal{V} \subset \Omega$ and hence, taking into account the last assertion together with Proposition 4.13, the conditions **(C1)**, **(C2)** and **(C3)** are satisfied. Thereby, the next assertion is an immediate consequence of Proposition 2.4 and we omit its proof.

THEOREM 4.16. *Under the assumptions of Proposition 4.12 and the condition* (4.10). *Select the dimension parameter $\widehat{m}$ as given by* (2.7) *with $\widehat{\mathrm{pen}}_m := 1152\tau_\infty^2 m n^{-1}\widehat{\sigma}_Y^2$. There exists a numerical constant $C$ and a finite constant $\zeta(r\mathfrak{A}, \sigma, \tau_\infty, \mathbb{E}\varepsilon^6)$ depending only on the quantities $r\mathfrak{A}$, $\sigma$, $\tau_\infty$ and $\mathbb{E}\varepsilon^6$ such that for all $n \geqslant n_\diamond$ with $\mathcal{R}^\star_{n_\diamond} \leqslant 1$ we have*

$$\mathcal{R}\Big[\widehat{f}_{\widehat{m}} \,|\, \mathcal{F}^r_{\mathfrak{a}}\Big] \leqslant C[r^2 \vee \tau_\infty^2 \sigma_Y^2 \vee \zeta(r\mathfrak{A}, \sigma, \tau_\infty, \mathbb{E}\varepsilon^6)]\, \mathcal{R}^\star_n.$$

We shall emphasise that the last assertion establishes the minimax-optimality of the fully data-driven estimator $\widehat{f}_{\widehat{m}}$ over all classes $\mathcal{F}^r_{\mathfrak{a}}$. Therefore, the estimator is called adaptive.

# 5 Simulation study

In this section we illustrate the performance of the proposed data-driven estimation procedure by means of a simulation study. As competitors we consider two widely used approaches, namely model selection and cross-validation, which we briefly introduce next. Following a model selection approach (see for example Comte and Rozenholc [2002] in the context of dependent data) the dimension parameter is selected as following

$$\widehat{m}_{MS} := \underset{1 \leqslant m \leqslant n}{\arg\min} \Big\{ -\|\widehat{f}_m\|^2_{L_2} + cmn^{-1}\widehat{\sigma}_Y^2 \Big\}.$$

We shall emphasize that this procedure relies on the contrast $-\|\widehat{f}_m\|^2_{L_2}$ rather than $\Upsilon_m$ (see equation (2.3)) used in the approach studied in this paper. Moreover, the penalty term in both selection procedures involves a constant $c$ which has been calibrated in advance by a simulation study. A popular alternative provides a cross validation approach. Exploiting that the estimated coefficients satisfy $\widehat{[f]}_j = n^{-1}\sum_{i=1}^n \psi_j(Z_i)$, for $j \geqslant 1$, we consider the cross validation criterium given by

$$CV(m) := \int_{[0,1]} \widehat{f}^2_m(x)dx - \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^m \sum_{k \neq i} \psi_j(Z_k)\phi_j(Z_i).$$

The dimension parameter is then selected as $\widehat{m}_{CV} = \underset{1 \leqslant m \leqslant n}{\arg\min}\, CV(m)$. Considering the orthonormal series estimator $\widehat{f}_m$ we denote by $\widehat{f}_{MS} := \widehat{f}_{\widehat{m}_{MS}}$ and $\widehat{f}_{CV} := \widehat{f}_{\widehat{m}_{CV}}$ the fully data-driven estimator based on a dimension parameter choice using the model selection and the cross-validation approach, respectively. Moreover, $\widehat{f}_{\widehat{m}}$ denotes the orthogonal series estimator with $\widehat{m}$ given as in (2.7). In addition we compare the three fully data-driven estimators with the oracle estimator $\widehat{f}_O := \widehat{f}_{m_o}$ where the dimension parameter $m_n^\diamond$ minimizes the integrated squared error (ISE), that is $m_o := \arg\min_{m \geqslant 1} \|\widehat{f}_m - f\|_{L_2}$. Obviously this choice is not feasible in practice.

In the following we report the performance of the four estimation procedures given independent as well as dependent observations. Therefore we make use of the framework introduced by Gannaz and Wintenberger [2010] which has also been studied, for example, by Bertin and Klutchnikoff [2014]. In the simulations we generate observations $Z_1, \dots, Z_n$ according to the following three different weak-dependence cases with the same marginal absolutely continuous distribution $F$.

**Case 1** The $Z_i$ are given by $F^{-1}(U_i)$ for $1 \leqslant i \leqslant n$ on $[0,1]$ where the $\{U_i\}_{i=1}^n$ are i.i.d. uniform random variables on $[0,1]$.

**Case 2** The $Z_i$ are given by $F^{-1}(G(Y_i))$ where $G(y) := \frac{2}{\pi}\arcsin(\sqrt{y})$ and the $Y_i$ are defined by $Y_1 = G^{-1}(U_1)$ and recursively, for any $i \geqslant 2$, $Y_i = T(Y_{i-1})$ with $T(y) = 4y(1-y)$.

**Case 3** The $Z_i$ are given by $F^{-1}(G(Y_i))$ where $G$ is the marginal distribution of $Y_i$ (see for details Gannaz and Wintenberger [2010]) and the $Y_i$, $i \in \mathbb{Z}$ is given by

$$Y_i = 2(Y_{i-1} + Y_{i+1})/5 + 5\zeta_i/21,$$

with $\{\zeta_i\}_{i\in\mathbb{Z}}$ is an i.i.d. sequence of Bernoulli variables with parameter 1/2. The computation of $Z_i$'s variable is based on the method developed in Doukhan and Truquet [2007].

Throughout the simulation study we consider the orthogonal series estimator based on the trigonometric basis. We repeat the estimation procedure for each of the four dimension selection procedures on 501 generated samples of size $n$ =100, 1000, 10000. However we present only the results for $n = 1000$ since in the other cases the findings were similar.

## 5.1 Non-parametric density estimation

We consider the estimation of two different density functions. The first one is a mixture of two Gaussian distributions, that is

$$f_1(x) = C\left(\frac{3}{10}\phi_{0.5;0.1}(x) + \frac{1}{4}\phi_{0.7;0.06}(x)\right)\mathbb{1}_{[0,1]}$$

where $\phi_{\mu;\sigma}$ stands for the density of a normal distribution with mean $\mu$ and standard deviation $\sigma$. The second one is defined by

$$f_2(x) = C\left(4(1 + |5(x-1/2)|)\right)^{-3/2}\mathbb{1}_{[0,1]}.$$

In the both cases the numerical constant $C$ is the normalizing factor. The observations $X_1, \dots, X_n$ are generated according to the three cases of weak-dependence with the same marginal density $f_1$ or $f_2$.

Figure 1 and 2 represent the overall behaviour of the data-driven estimator $\widehat{f}_{\widehat{m}}$ of the density functions $f_1$ and $f_2$, respectively, for the three considered cases of weak-dependence. More precisely, in each figure the point-wise median and the 5% and 95% point-wise percentile are depicted. The quality of the estimator is visually reasonable. In addition Table 1 reports the empirical mean and standard deviation of the ISE over the 501 Monte-Carlo repetitions. As expected the oracle estimator $\widehat{f}_O$ outperforms the data-driven estimators. However, the increase of the estimation error for the data-driven procedures is rather small. Moreover the data-driven estimator $\widehat{f}_{\widehat{m}}$ studied in this paper and the model selection based estimator $\widehat{f}_{MS}$ perform better than the cross validation procedure for both densities and all three cases of weak-dependence. Surprisingly, the selected values $\widehat{m}$ and $\widehat{m}_{MS}$ coincided in at least four out of the 501 Monte-Carlo repetitions for each density and each of three cases of weak-dependence, which explains the identical values in Table 1.

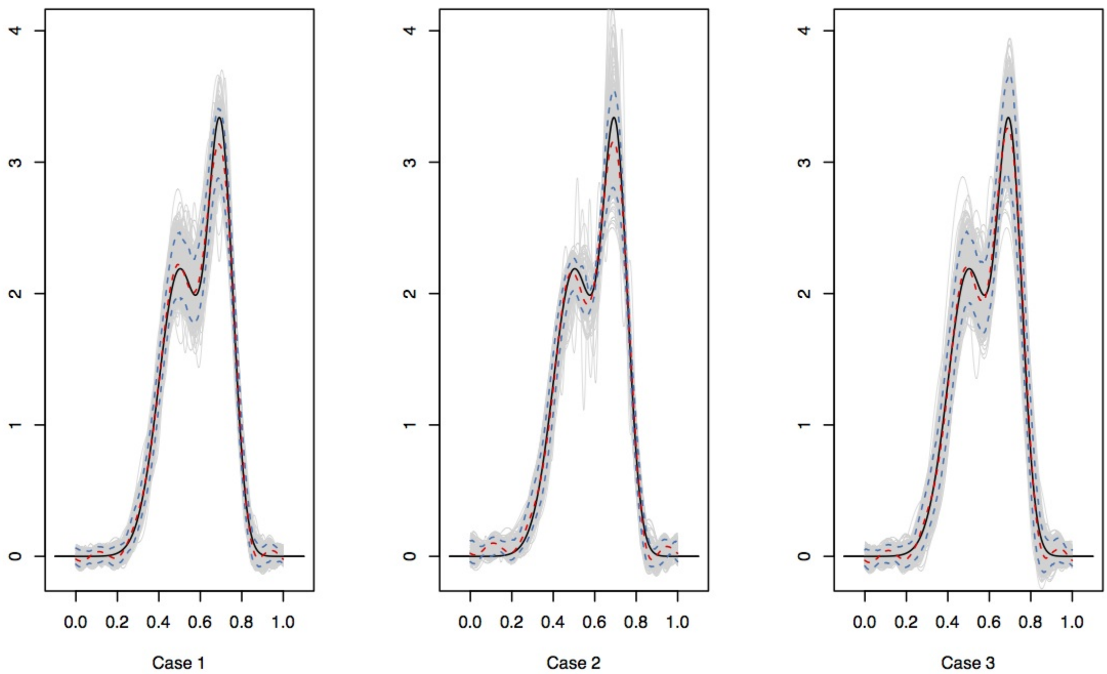


Figure 1: The grey graphs depict the Monte-Carlo realisations of the data-driven estimator $\widehat{f}_{\widehat{m}}$ for the density $f_1$ in the three cases of weak-dependence. The solid line corresponds to the true function, the red dashed line and the blue dashed lines represent, respectively, the point-wise median and the 5% and 95% point-wise percentile of the 501 replications.

| | | $\widehat{f}_O$ | $\widehat{f}_{\widehat{m}}$ | $\widehat{f}_{MS}$ | $\widehat{f}_{CV}$ |
|---|---|---|---|---|---|
| $f_1$ | Case 1 | 0.0112 (0.0065) | 0.0142 (0.0089) | 0.0142 (0.0089) | 0.0178 (0.0140) |
| | Case 2 | 0.0102 (0.0084) | 0.0129 (0.0123) | 0.0128 (0.0119) | 0.0151 (0.0155) |
| | Case 3 | 0.0188 (0.0138) | 0.0213 (0.0148) | 0.0213 (0.0148) | 0.0242 (0.0169) |
| $f_2$ | Case 1 | 0.0110 (0.0037) | 0.0153 (0.0053) | 0.0153 (0.0053) | 0.0159 (0.0076) |
| | Case 2 | 0.0123 (0.0071) | 0.0177 (0.0110) | 0.0178 (0.0108) | 0.0232 (0.0197) |
| | Case 3 | 0.0158 (0.0071) | 0.0210 (0.0087) | 0.0211 (0.0087) | 0.0223 (0.0118) |

Table 1: Empirical mean (and standard deviation) of the ISE over the 501 Monte-Carlo simulations of sample of size $n = 1000$ for the oracle and the three different data-driven estimators of the densities $f_1$ and $f_2$ in the three cases of weak-dependence.

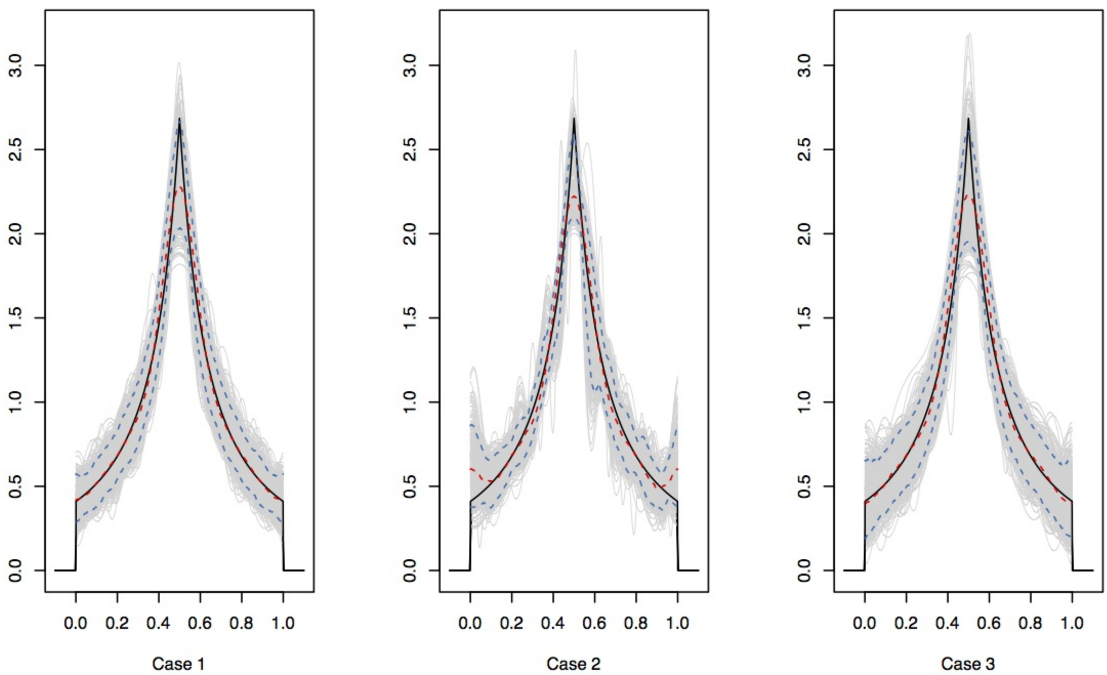


Figure 2: The grey graphs depict the Monte-Carlo realisations of the data-driven estimator $\widehat{f}_{\widehat{m}}$ for the density $f_2$ in the three cases of weak-dependence. The solid line corresponds to the true function, the red dashed line and the blue dashed lines represent, respectively, the point-wise median and the 5% and 95% point-wise percentile of the 501 replications.

## 5.2 Non-parametric regression estimation

Two different regression functions are considered. The first one is a Doppler function

$$f_1(x) = (x(1-x))^{1/2} \sin\left(\frac{2.6\pi}{x+0.3}\right) \mathbb{1}_{[0,1]}$$

and the second one is a mixture of a sinus function and a indicator function defined by

$$f_2(x) = \sin(4x)\, \mathbb{1}_{[0,1/4]} + \mathbb{1}_{]1/4,1]}\,.$$

In the both cases the error terms are independently and identically standard normally distributed and the noise level is set to $\sigma = 0.5$. The explanatory random variables $U_1, \ldots, U_n$ are generated according to the three cases of weak-dependence with identical marginal uniform distribution on the interval $[0, 1]$.

| | | $\widehat{f}_O$ | $\widehat{f}_{\widehat{m}}$ | $\widehat{f}_{MS}$ | $\widehat{f}_{CV}$ |
|---|---|---|---|---|---|
| $f_1$ | Case 1 | 0.0306 (0.0091) | 0.0369 (0.0111) | 0.0369 (0.0111) | 0.0340 (0.0099) |
| | Case 2 | 0.0309 (0.0116) | 0.0375 (0.0146) | 0.0375 (0.0146) | 0.0343 (0.0122) |
| | Case 3 | 0.0332 (0.0098) | 0.0392 (0.0109) | 0.0392 (0.0109) | 0.0370 (0.0106) |
| $f_2$ | Case 1 | 0.0251 (0.0054) | 0.0318 (0.0081) | 0.0318 (0.0081) | 0.0354 (0.0122) |
| | Case 2 | 0.0235 (0.0064) | 0.0310 (0.0098) | 0.0310 (0.0098) | 0.0366 (0.0137) |
| | Case 3 | 0.0297 (0.0091) | 0.0372 (0.0139) | 0.0372 (0.0139) | 0.0388 (0.0133) |

Table 2: Empirical mean (and standard deviation) of the ISE over the 501 Monte-Carlo simulations of sample of size $n = 1000$ for the oracle and the three different data-driven estimators of the regressions $f_1$ and $f_2$ in the three cases of weak-dependence.

Figure 3 and 4 represent the overall behaviour of the data-driven estimator $\widehat{f}_{\widehat{m}}$ of the regression functions $f_1$ and $f_2$, respectively, for the three considered cases of weak-dependence. The quality of the estimator is again visually reasonable. As in the density estimation case, the Table 2 reports the empirical mean and standard deviation of the ISE over the 501 Monte-Carlo repetitions. The findings are the same as for the density estimation problem with the only exception that for the regression function $f_1$ the cross validation approach performs slightly better than the other two data-driven procedures. We shall emphasize that again the selected values $\widehat{m}$ and $\widehat{m}_{MS}$ coincided in at least 99% of the Monte-Carlo repetitions for each regression function and each of three cases of weak-dependence. This explains the identical value in Table 2 for the model selection based estimator $\widehat{f}_{MS}$ and the data-driven estimator $\widehat{f}_{\widehat{m}}$ studied in this paper.

**Conclusions and perspectives.** In this work we present a data-driven non-parametric estimation procedure of a density and a regression function in the presence of dependent data that can attain minimax-optimal rates for independent data. Obviously, the data-driven non-parametric estimation in errors in variables models as, for example, deconvolution problems or instrumental variable regressions, are only one amongst the many interesting questions for further research and we are currently exploring this topic.

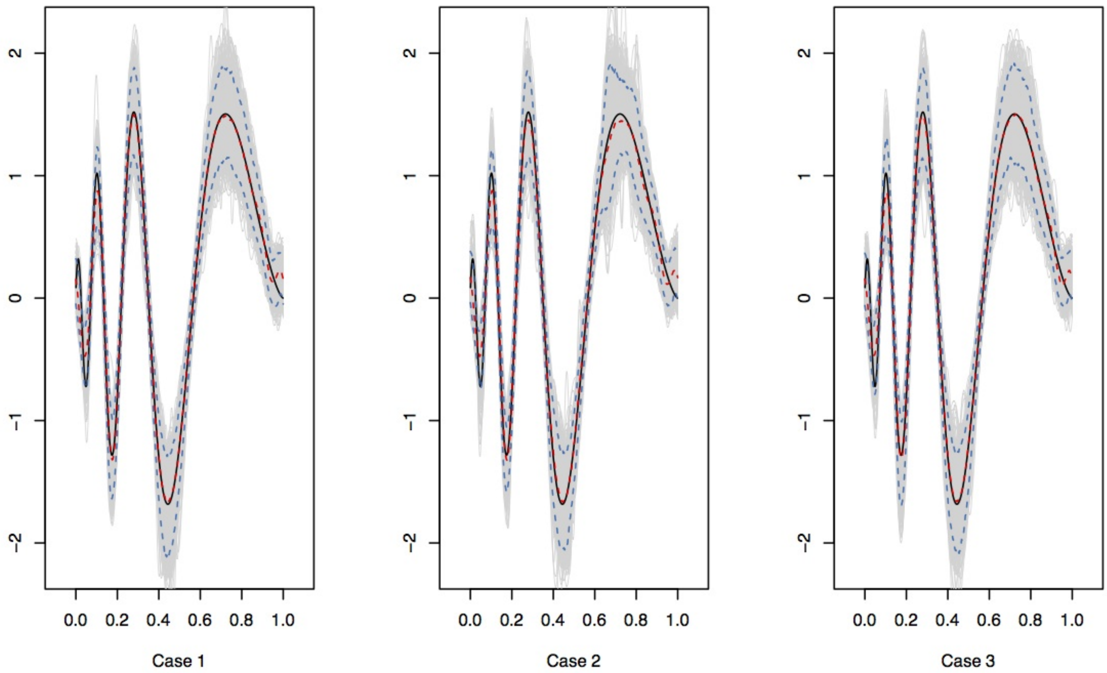


Figure 3: The grey graphs depict the Monte-Carlo realisations of the data-driven estimator $\widehat{f}_{\widehat{m}}$ for the regression $f_1$ in the three cases of weak-dependence. The solid line corresponds to the true function, the red dashed line and the blue dashed lines represent, respectively, the point-wise median and the 5% and 95% point-wise percentile of the 501 replications.

**Acknowledgments.** This work was supported by the IAP research network no. P7/06 of the Belgian Government (Belgian Science Policy), by the ”Fonds Spéciaux de Recherche” from the Université catholique de Louvain and by the ARC contract 11/16-039 of the ”Communauté française de Belgique”, granted by the Académie universitaire Louvain.

# A Appendix: Proofs of Section 2

*Proof of Lemma 2.3.* Keeping in mind the identity $\|\widehat{f}_k - f\|_{L_2}^2 = \|\widehat{f}_k - f_k\|_{L_2}^2 + \|f_k - f\|_{L_2}^2$ for any $k \in \mathbb{N}$, we obtain:

$$\mathbb{E}\left(\mathbb{1}_{\Omega^c}\|\widehat{f}_{\widetilde{m}} - f\|_{L_2}^2\right) = \mathbb{E}\left(\mathbb{1}_{\Omega^c}\left\{\|\widehat{f}_{\widetilde{m}} - f_{\widetilde{m}}\|_{L_2}^2 + \|f_{\widetilde{m}} - f\|_{L_2}^2\right\}\right)$$
$$\leqslant \mathbb{E}\left(\mathbb{1}_{\Omega^c}\left\{\|\widehat{f}_n - f_n\|_{L_2}^2 + \|f\|_{L_2}^2\right\}\right) \quad \text{(A.1)}$$

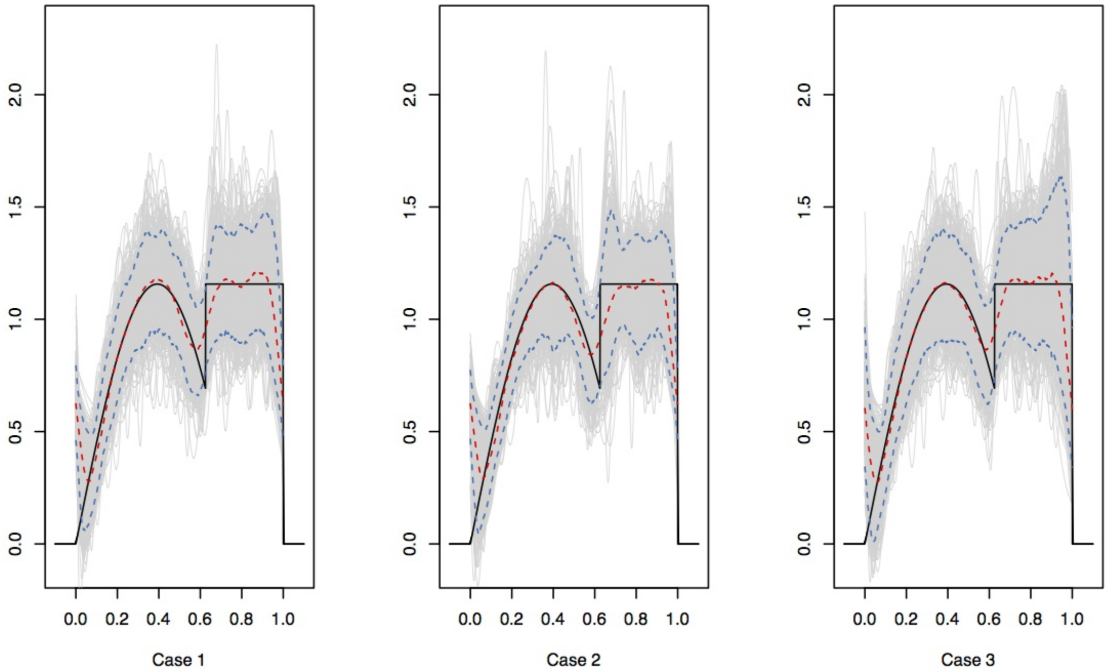


Figure 4: The grey graphs depict the Monte-Carlo realisations of the data-driven estimator $\widehat{f}_{\widehat{m}}$ for the regression $f_2$ in the three cases of weak-dependence. The solid line corresponds to the true function, the red dashed line and the blue dashed lines represent, respectively, the point-wise median and the 5% and 95% point-wise percentile of the 501 replications.

since $\|\widehat{f}_k - f_k\|_{L_2}^2 \leqslant \|\widehat{f}_n - f_n\|_{L_2}^2$ and $\|f_k - f\|_{L_2}^2 \leqslant \|f\|_{L_2}^2$ for all $1 \leqslant k \leqslant n$. Considering the first right hand side term we have

$$\mathbb{E}\left(\|\widehat{f}_n - f_n\|_{L_2}^2 \mathbb{1}_{\Omega^c}\right) \leqslant \mathbb{E}\Big\{\left(\|\widehat{f}_n - f_n\|_{L_2}^2 - \mathrm{pen}_n / 6\right)_+\Big\} + \tfrac{1}{6}\,\mathrm{pen}_n\, P(\Omega^c) \tag{A.2}$$

The assertion follows now by combination of (A.1) and (A.2) together with the conditions **(C1)** and **(C2)**, and $\|f\|_{L_2}^2 \leqslant r^2$, for all $f \in \mathcal{F}_{\mathfrak{a}}^r$, which completes the proof. □

# B Appendix: Proofs of Section 3

## B.1 Appendix: Proofs of Section 3.1

*Proof of Proposition 3.2.* In the case of independent observations it holds obviously that

$$\sum_{j=1}^{m} \mathbb{V}\mathrm{ar}\left\{\frac{1}{n}\sum_{i=1}^{n}\phi_j(X_i)\right\} = n^{-1}\sum_{j=1}^{m}\mathbb{V}\mathrm{ar}\{\phi_j(X)\} \leqslant n^{-1}\mathbb{E}\sum_{j=1}^{m}\phi_j^2(X) \leqslant n^{-1} m \tau_\infty^2 \tag{B.1}$$

where we have exploited the assumption **(A1)**. Consequently, we have for $n, m \geqslant 1$ that

$$\mathcal{R}\Big[\widehat{f}_m \,|\, \mathcal{F}_{\mathfrak{a}}^r \cap \mathcal{D}\Big] \leqslant n^{-1} m \tau_\infty^2 + \mathfrak{a}_m r^2 \leqslant (\tau_\infty^2 + r^2) \max(mn^{-1}, \mathfrak{a}_m) = (\tau_\infty^2 + r^2) \mathcal{R}_n^m.$$

Keeping in mind that the dimension parameter $m_n^\star$ given in (2.2), minimises the last upper risk bound, we get (3.1) which completes the proof. $\square$

*Proof of Proposition 3.3.* Given $\zeta := \eta \min(r-1, (4\mathfrak{A})^{-1})$ and $\alpha_n := \mathcal{R}_n^\star/(m_n^\star) \leqslant (n\eta)^{-1}$ based on the definition of $\eta$ we consider the function $f := 1 + (\zeta\alpha_n)^{1/2} \sum_{1 \leqslant j \leqslant m_n^\star} [f]_j \phi_j$. We will show that for any $\theta := (\theta_j) \in \{-1, 1\}^{m_n^\star}$, the function $f_\theta := 1 + \sum_{1 \leqslant j \leqslant m_n^\star} \theta_j [f]_j \phi_j$ belongs to $\mathcal{F}_{\mathfrak{a}}^r \cap \mathcal{D}$ and is hence a possible candidate of the density. We denote by $f_\theta^n$ the joint density of an iid. $n$-sample from $f_\theta$ and by $\mathbb{E}_\theta$ the expectation with respect to the joint density $f_\theta^n$. Furthermore, for $0 < j \leqslant m_n^\star$ and each $\theta$ we introduce $\theta^{(j)}$ by $\theta_l^{(j)} = \theta_l$ for $j \neq l$ and $\theta_j^{(j)} = -\theta_j$. The key argument of this proof is the following reduction scheme. If $\widetilde{f}$ denotes an estimator of $f$ then we conclude

$$\begin{aligned}\mathcal{R}\Big[\widetilde{f} \,|\, \mathcal{F}_{\mathfrak{a}}^r \cap \mathcal{D}\Big] &\geqslant \max_{\theta \in \{-1,1\}^{m_n^\star}} \mathbb{E}_\theta \|\widetilde{f} - f_\theta\|_{L_2}^2 \geqslant \frac{1}{2^{m_n^\star}} \sum_{\theta \in \{-1,1\}^{m_n^\star}} \mathbb{E}_\theta \|\widetilde{f} - f_\theta\|_{L_2}^2 \\ &\geqslant \frac{1}{2^{m_n^\star}} \sum_{0 < j \leqslant m_n^\star} \frac{1}{2} \sum_{\theta \in \{-1,1\}^{m_n^\star}} \Big\{ \mathbb{E}_\theta |[\widetilde{f} - f_\theta]_j|^2 + \mathbb{E}_{\theta^{(j)}} |[\widetilde{f} - f_{\theta^{(j)}}]_j|^2 \Big\}. \qquad \text{(B.2)}\end{aligned}$$

by using that for each $0 < j \leqslant m_n^\star$ and any function $F : \{-1, 1\}^{m_n^\star} \to \mathbb{R}$, it holds

$$\sum_{\theta \in \{-1,1\}^{m_n^\star}} f(\theta) = \sum_{\theta \in \{-1,1\}^{m_n^\star}} f(\theta^{(j)}).$$

Below we show furthermore that for all $n \geqslant 2$ we have

$$\Big\{ \mathbb{E}_\theta |[\widetilde{f} - f_\theta]_j|^2 + \mathbb{E}_{\theta^{(j)}} |[\widetilde{f} - f_{\theta^{(j)}}]_j|^2 \Big\} \geqslant \frac{\zeta}{8} \alpha_n. \qquad \text{(B.3)}$$

From the last lower bound and the reduction scheme, by employing the definition of $\zeta$ and $\alpha_n$, we obtain the result (3.3), that is

$$\mathcal{R}\Big[\widetilde{f} \,|\, \mathcal{F}_{\mathfrak{a}}^r \cap \mathcal{D}\Big] \geqslant \frac{1}{2^{m_n^\star}} \sum_{\theta \in \{-1,1\}^{m_n^\star}} \sum_{0 < j \leqslant m_n^\star} \frac{1}{2} \frac{\zeta}{4} \alpha_n = \frac{\zeta}{4} \alpha_n m_n^\star = \frac{\eta}{8} \min(r-1, (4\mathfrak{A}\tau_\infty^2)^{-1}) \, \mathcal{R}_n^\star.$$

To conclude the proof, it remains to check (B.3) and $f_\theta \in \mathcal{F}_{\mathfrak{a}}^r \cap \mathcal{D}$ for all $\theta \in \{-1, 1\}^{m_n^\star}$. The latter is easily verified if $f \in \mathcal{F}_{\mathfrak{a}}^r \cap \mathcal{D}$. In order to show that $f \in \mathcal{F}_{\mathfrak{a}}^r \cap \mathcal{D}$, we first notice that $f$ integrates to one. Moreover, $f$ is non-negative because $\|\sum_{0 < j \leqslant m_n^\star} [f]_j \phi_j\|_\infty \leqslant 1/2$, and $\|f\|_{\mathfrak{a}}^2 \leqslant r$, which can be realised as follows. From the assumption **(A2)** it follows

$$\|\sum_{j=1}^{m_n^\star} [f]_j \phi_j\|_\infty^2 \leqslant \|\sum_{j=1}^{m_n^\star} \mathfrak{a}_j \phi_j^2\|_\infty \left( \sum_{j=1}^{m_n^\star} \mathfrak{a}_j^{-1} [f]_j^2 \right) \leqslant \mathfrak{A}^2 \Big( \zeta \alpha_n \sum_{j=1}^{m_n^\star} \mathfrak{a}_j^{-1} \Big).$$

Since $\mathfrak{a}^{-1}$ is monotonically increasing the definition of $\zeta$, $\alpha_n$ and $\eta$ implies

$$\|\sum_{j=1}^{m_n^\star}[f]_j\phi_j\|_\infty^2 \leqslant \mathfrak{A}^2\zeta\alpha_n m_n^\star \mathfrak{a}_{m_n^\star}^{-1} \leqslant (\eta/4)\mathfrak{a}_{m_n^\star}^{-1}\alpha_n m_n^\star = \eta\mathfrak{a}_{m_n^\star}^{-1}\mathcal{R}_n^\star/4 \leqslant 1/4 \tag{B.4}$$

as well as $\|f\|_{\mathfrak{a}}^2 \leqslant 1+\zeta\mathfrak{a}_{m_n^\star}^{-1}\alpha_n m_n^\star \leqslant 1+\zeta/\eta \leqslant r$. It remains to show (B.3). Consider the Hellinger affinity $\rho(f_\theta^n, f_{\theta^{(j)}}^n) = \int\sqrt{f_\theta^n}\sqrt{f_{\theta^{(j)}}^n}$, then we obtain for any estimator $\widetilde{f}$ of $f$ that

$$\rho(f_\theta^n, f_{\theta^{(j)}}^n) \leqslant \Big(\int\frac{|[\widetilde{f}-f_{\theta^{(j)}}]_j|^2}{|[f_\theta-f_{\theta^{(j)}}]_j|^2}f_{\theta^{(j)}}^n\Big)^{1/2} + \Big(\int\frac{|[\widetilde{f}-f_\theta]_j|^2}{|[f_\theta-f_{\theta^{(j)}}]_j|^2}f_\theta^n\Big)^{1/2}.$$

Rewriting the last estimate we obtain

$$\Big\{\mathbb{E}_\theta|[\widetilde{f}-f_\theta]_j|^2 + \mathbb{E}_{\theta^{(j)}}|[\widetilde{f}-f_{\theta^{(j)}}]_j|^2\Big\} \geqslant \frac{1}{2}|[f_\theta-f_{\theta^{(j)}}]_j|^2\rho^2(f_\theta^n, f_{\theta^{(j)}}^n). \tag{B.5}$$

Next we bound from below the Hellinger affinity $\rho(f_\theta^n, f_{\theta^{(j)}}^n)$. Therefore, we consider first the Hellinger distance

$$H^2(f_\theta, f_{\theta^{(j)}}) = \int\frac{|f_\theta-f_{\theta^{(j)}}|^2}{\big(\sqrt{f_\theta}+\sqrt{f_{\theta^{(j)}}}\big)^2} \leqslant \frac{1}{2}\|f_\theta - f_{\theta^{(j)}}\|_{L_2}^2 = 2|[f]_j|^2 \leqslant \frac{2\zeta}{\eta n},$$

where we have used that $\alpha_n \leqslant (n\eta)^{-1}$ and $f_\theta \geqslant 1/2$ because $|\sum_{0<j\leqslant m_n^\star}[f_\theta]_j\phi_j| \leqslant 1/2$ (see (B.4)). Therefore, the definition of $\zeta$ implies $H^2(f_\theta, f_{\theta^{(j)}}) \leqslant 2/n$. By using the independence, i.e., $\rho(f_\theta^n, f_{\theta^{(j)}}^n) = \rho(f_\theta, f_{\theta^{(j)}})^n$, together with the identity $\rho(f_\theta, f_{\theta^{(j)}}) = 1-\frac{1}{2}H^2(f_\theta, f_{\theta^{(j)}})$ it follows $\rho(f_\theta^n, f_{\theta^{(j)}}^n) \geqslant (1-n^{-1})^n \geqslant 1/4$ for all $n \geqslant 2$. By combination of the last estimate with (B.5) we obtain (B.3) which completes the proof. □

*Proof of Proposition 3.4.* Keeping in mind Remark 2 we intend to apply Talagrand's inequality (Lemma 3.1) where we need to compute the quantities $h$, $H$ and $v$ verifying the three required inequalities. Consider first $h$ where due to the assumption **(A1)**

$$\sup_{t\in\mathbb{B}_m}\|\nu_t\|_\infty^2 = \|\sum_{j=1}^m\phi_j^2\|_\infty \leqslant \tau_\infty^2 m =: h^2. \tag{B.6}$$

Consider next $H$ where

$$\mathbb{E}\sup_{t\in\mathbb{B}_m}|\overline{\nu_t}| = \Big(\mathbb{E}\|\widehat{f}_m - f_m\|_{L_2}^2\Big)^{1/2} = \Big(\sum_{j=1}^m\mathbb{V}\mathrm{ar}([\widehat{f}]_j)\Big)^{1/2} \leqslant \Big[mn^{-1}\tau_\infty^2\Big]^{1/2} =: H. \tag{B.7}$$

Consider finally $v$. Due to assumption **(A2)** for all $f\in\mathcal{F}_{\mathfrak{a}}^r$, we have

$$\sup_{t\in\mathbb{B}_m}\mathbb{E}|\nu_t(X)|^2 = \sup_{t\in\mathbb{B}_m}\mathbb{E}|\sum_{j=1}^m[t]_j\phi_j(X)|^2 \leqslant \|f\|_\infty \leqslant r\mathfrak{A} =: v. \tag{B.8}$$

The assertion follows from Lemma 3.1 by using the quantities $h$, $H$ and $v$ given in (B.6), (B.7) and (B.8), respectively and by employing the definition of $\zeta$, which completes the proof. □

## B.2 Appendix: Proofs of Section 3.2

*Proof of Proposition 3.6.* In the case of independent observations it holds obviously that

$$\sum_{j=1}^{m} \mathbb{V}\mathrm{ar}\left\{\frac{1}{n}\sum_{i=1}^{n} Y_i\phi_j(U_i)\right\} \leqslant n^{-1}\mathbb{E}Y^2\sum_{j=1}^{m}\phi_j^2(U) \leqslant n^{-1}m\tau_\infty^2(\sigma^2+\|f\|_{L_2}^2) \tag{B.9}$$

where we have exploited assumption **(A1)** and $\sigma_Y^2 := \mathbb{E}Y^2 = \sigma^2 + \|f\|_{L_2}^2$. Keeping mind that $\|f\|_{L_2}^2 \leqslant r^2$ for all $f \in \mathcal{F}_{\mathfrak{a}}^r$ we have for $n, m \geqslant 1$, $\mathcal{R}\Big[\widehat{f}_m \,|\, \mathcal{F}_{\mathfrak{a}}^r\Big] \leqslant (\tau_\infty^2(\sigma^2+r^2)+r^2)\mathcal{R}_n^m$.

Employing further that the dimension parameter $m_n^\star$ given in (2.2) minimises the last upper risk bound, ie., the term $\mathcal{R}_n^m = \max(mn^{-1}, \mathfrak{a}_m)$, with respect to the dimension parameter, we obtain (3.4) which completes the proof. □

*Proof of Proposition 3.7.* Given $\zeta := \eta\min(r^2, \sigma^2/2)$ and $\alpha_n := \mathcal{R}_n^\star/m_n^\star \leqslant (n\eta)^{-1}$ due (3.5) we consider the function $f := (\zeta\alpha_n)^{1/2}\sum_{j=1}^{m_n^\star}\phi_j$. We will show that for any $\theta := (\theta_j)_{j=1}^{m_n^\star} \in \{-1,1\}^{m_n^\star}$, the function $f_\theta := \sum_{j=1}^{m_n^\star}\theta_j[f]_j\phi_j$ belongs to $\mathcal{F}_{\mathfrak{a}}^r$ and is hence a possible candidate of the regression function. For a fixed $\theta$ and under the hypothesis that the regression function is $f_\theta$, we denote by $P_\theta^n$ the joint distribution of the observation $\{(Y_i, U_i)\}_{i=1}^n$ and by $\mathbb{E}_\theta$ the expectation with respect to this distribution. Furthermore, for $1 \leqslant j \leqslant m_n^\star$ and each $\theta$ we introduce $\theta^{(j)}$ by $\theta_l^{(j)} = \theta_l$ for $j \neq l$ and $\theta_j^{(j)} = -\theta_j$. The key argument of this proof is the following reduction scheme (B.2) . From the lower bound (B.3) and the reduction scheme (B.2), by employing the definition of $\zeta$ and $\alpha_n$, we obtain the result (3.6), that is

$$\mathcal{R}\Big[\widetilde{f}\,|\,\mathcal{F}_{\mathfrak{a}}^r\Big] \geqslant \frac{1}{2^{m_n^\star}}\sum_{\theta\in\{-1,1\}^{m_n^\star}}\sum_{j=1}^{m_n^\star}\frac{1}{2}\frac{\zeta}{2}\alpha_n = \frac{\zeta}{4}\alpha_n m_n^\star = \frac{\eta}{8}\min(2r^2,\sigma^2)\,\mathcal{R}_n^\star.$$

To conclude the proof, it remains to check (B.3) and $f_\theta \in \mathcal{F}_{\mathfrak{a}}^r$ for all $\theta \in \{-1,1\}^{m_n^\star}$. The latter is easily verified if $f \in \mathcal{F}_{\mathfrak{a}}^r$, which can be realised as follows. By applying successively that $\mathfrak{a}$ is monotonically increasing, that $\mathcal{R}_n^\star\mathfrak{a}_{m_n^\star} \leqslant \eta^{-1}$ due (3.5) and, hence $\zeta\alpha_n m_n^\star\mathfrak{a}_{m_n^\star} = \zeta\mathcal{R}_n^\star\mathfrak{a}_{m_n^\star} \leqslant r^2$ we obtain $\|f\|_{\mathfrak{a}}^2 \leqslant \zeta\alpha_n m_n^\star\mathfrak{a}_{m_n^\star} \leqslant r^2$ which proves the claim.

Next we bound from below the Hellinger affinity $\rho(P_\theta^n, P_{\theta^{(j)}}^n)$ using the well-known relationship $\rho(P_\theta^n, P_{\theta^{(j)}}^n) \geqslant 1-(1/2)KL(P_\theta^n, P_{\theta^{(j)}}^n)$ between the Kullback-Leibler divergence and the Hellinger affinity. We will show that $KL(P_\theta^n, P_{\theta^{(j)}}^n) \leqslant 1$, and hence $\rho(P_\theta^n, P_{\theta^{(j)}}^n) \geqslant 1/2$ which together with (B.5) and $|[f_\theta - f_{\theta^{(j)}}]_j|^2 = 4[f]_j^2 = 4\zeta\alpha_n$ implies (B.3). Therefore, consider the Kullback-Leibler divergence between $P_\theta^n$ and $P_{\theta^{(j)}}^n$. Recall, that for a fixed $\theta$ and under the hypothesis that the regression function is $f_\theta$, the observations $\{Y_i\}_{i=1}^n$ are conditional independent given the regressors $\{U_i\}_{j=1}^n$ and for each $1 \leqslant i \leqslant n$ the conditional distribution of $Y_i$ given the regressor $U_i$ is normal with conditional mean $f_\theta(U_i)$ and conditional variance $\sigma^2$. Therefore, we have

$$\log\frac{dP_\theta^n(\{(Y_i,U_i)\}_{i=1}^n)}{dP_{\theta^{(j)}}^n(\{(Y_i,U_i)\}_{i=1}^n)} = \sum_{i=1}^n\frac{2\zeta\alpha_n}{\sigma^2}\phi_j^2(U_i) + \sum_{i=1}^n\frac{2\theta_j(\zeta\alpha_n)^{1/2}}{\sigma^2}\phi_j(U_i)(Y_i - f_\theta(U_i)).$$

Taking the expectation $\mathbb{E}_\theta$ with respect to $P_\theta^n$ leads to $KL(P_\theta^n, P_{\theta(j)}^n) = 2\zeta\alpha_n n/\sigma^2$. By employing that $\alpha_n n \leqslant 1/\eta$ and $\zeta/(\eta\sigma^2) \leqslant 1/2$ we obtain that $KL(P_\theta^n, P_{\theta(j)}^n) \leqslant 1$ which shows the claim and completes the proof. $\square$

*Proof of Proposition 3.8.* The key argument of the next assertion is again Talagrand's inequality. However, a direct application employing $\sup_{t\in\mathbb{B}_m}|\overline{\nu_t}|^2 = \|\widehat{f}_m - f_m\|_{L_2}^2$ with $\overline{\nu_t} = \frac{1}{n}\sum_{i=1}^n [\nu_t(\varepsilon_i, U_i) - \mathbb{E}\left(\nu_t(\varepsilon_i, U_i)\right)]$ and $\nu_t(\varepsilon, U) = \sum_{j=1}^m [t]_j(\sigma\varepsilon + f(U))\phi_j(U)$ is not possibly noting that $\varepsilon$ and hence $\nu_t$ are generally not uniformly bounded. Therefore, let us introduce $\varepsilon^b := \varepsilon\,\mathbb{1}_{\{|\varepsilon|\leqslant n^{1/4}\}} - \mathbb{E}\varepsilon\,\mathbb{1}_{\{|\varepsilon|\leqslant n^{1/4}\}}$ and $\varepsilon^u := \varepsilon - \varepsilon^b = \varepsilon\,\mathbb{1}_{\{|\varepsilon|> n^{1/4}\}} - \mathbb{E}\varepsilon\,\mathbb{1}_{\{|\varepsilon|> n^{1/4}\}}$. Setting $\overline{\nu_t^b} = \frac{1}{n}\sum_{i=1}^n\left[\nu_t(\varepsilon_i^b, U_i) - \mathbb{E}\Big(\nu_t(\varepsilon_i^b, U_i)\Big)\right]$, $\nu_t^u(\varepsilon^u, U) := \sum_{j=1}^m [t]_j\sigma\varepsilon^u\phi_j(U)$ and $\overline{\nu_t^u} = \frac{1}{n}\sum_{i=1}^n [\nu_t^u(\varepsilon_i^u, U_i) - \mathbb{E}\left(\nu_t^u(\varepsilon_i^u, U_i)\right)]$ we have obviously $\overline{\nu_t} = \overline{\nu_t^b} + \overline{\nu_t^u}$. Consequently, exploiting the elementary inequality $|\overline{\nu_t}|^2 \leqslant 2\Big\{|\overline{\nu_t^b}|^2 + |\overline{\nu_t^u}|^2\Big\}$ follows that

$$\mathbb{E}\Big(\max_{m_n^\star\leqslant m\leqslant n}\{\|\widehat{f}_m - f_m\|_{L_2}^2 - \tfrac{\mathrm{pen}_m}{6}\}\Big)_+ \leqslant 2\mathbb{E}\Big(\max_{m_n^\star\leqslant m\leqslant n}\{\sup_{t\in\mathbb{B}_m}|\overline{\nu_t^b}|^2 - \tfrac{\mathrm{pen}_m}{12}\}\Big)_+ + 2\mathbb{E}\sup_{t\in\mathbb{B}_n}|\overline{\nu_t^u}|^2. \tag{B.10}$$

We bound separately each term on the rhs. of the last display. Consider first the second right hand side term. Since $\mathbb{E}(\varepsilon^6) < \infty$ which implies that $\mathbb{E}(\varepsilon^2)\,\mathbb{1}_{\{\varepsilon^2>\eta\}} \leqslant \eta^{-2}\mathbb{E}(\varepsilon^6)$ for all $\eta > 0$, it follows from the independence assumption and **(A1)** that

$$\mathbb{E}\sup_{t\in\mathbb{B}_n}|\overline{\nu_t^u}|^2 \leqslant \sigma^2\tau_\infty^2\,\mathbb{V}\mathrm{ar}(\varepsilon^u) \leqslant \sigma^2\tau_\infty^2\mathbb{E}\Big(\varepsilon^2\,\mathbb{1}_{\{|\varepsilon|>n^{1/4}\}}\Big) \leqslant n^{-1}\sigma^2\tau_\infty^2\mathbb{E}(\varepsilon^6). \tag{B.11}$$

In order to bound the second right hand side term in (B.10), we aim to apply Talagrand's inequality (Lemma 3.1) which necessitates the computation of the quantities $h$, $H$ and $v$ verifying the required inequalities. Consider first $h$. Let $\psi_j(e^b, u) = (\sigma e^b + f(u))\phi_j(u)$ and note that $|\varepsilon^b| \leqslant 2n^{1/4}$ by construction. Hence, employing **(A1)** we have

$$\sup_{t\in\mathbb{B}_m}\|v_t\|_\infty^2 = \sum_{j=1}^m\|\psi_j^2\|_\infty \leqslant \tau_\infty^2 m(2\sigma n^{1/4} + \|f\|_\infty)^2 =: h^2. \tag{B.12}$$

Next we compute the quantity $H$, where due to assumption **(A1)**

$$\mathbb{E}\sup_{t\in\mathbb{B}_m}|\overline{v_t^b}|^2 \leqslant \frac{1}{n}\mathbb{E}\Big\{(\sigma\varepsilon_1^b + f(U_1))^2\sum_{j=1}^m\phi_j^2(U_1)\Big\} \leqslant \frac{m\tau_\infty^2}{n}\mathbb{E}(\sigma\varepsilon_1^b + f(U_1))^2.$$

Exploiting $\mathbb{V}\mathrm{ar}\,\varepsilon^b \leqslant \mathbb{E}\Big(\varepsilon^2\,\mathbb{1}_{\{|\varepsilon|>n^{1/4}\}}\Big) \leqslant \mathbb{E}\varepsilon^2 = 1$ and the independence between $\varepsilon$ and $U$ we have $\mathbb{E}(\sigma\varepsilon_1^b + f(U_1))^2 = \sigma^2\,\mathbb{V}\mathrm{ar}\,\varepsilon_1^b + \|f\|_{L_2}^2 \leqslant \sigma^2 + \|f\|_{L_2}^2 = \mathbb{E}Y^2 = \sigma_Y^2$. Combining the bounds it follows that

$$\mathbb{E}\sup_{t\in\mathbb{B}_m}|\overline{v_t^b}| \leqslant \Big(\mathbb{E}\sup_{t\in\mathbb{B}_m}|\overline{v_t^b}|^2\Big)^{1/2} \leqslant n^{-1/2}m^{1/2}\tau_\infty\sigma_Y =: H. \tag{B.13}$$

It remains to calculate the third quantity $\nu$, where due to the independence between $\varepsilon$ and $U$

$$\sup_{t\in\mathbb{B}_m}\frac{1}{n}\sum_{i=1}^{n}\mathbb{V}\mathrm{ar}(v_t(\varepsilon_i^b,U_i))\leqslant\sup_{t\in\mathbb{B}_m}\mathbb{E}(v_t(\varepsilon_1^b,U_1))^2$$
$$=\sup_{t\in\mathbb{B}_m}\{\sigma^2\,\mathbb{V}\mathrm{ar}(\varepsilon^b)\mathbb{E}\Big(\sum_{j=1}^{m}[t]_j\phi_j(U_1)\Big)^2+\mathbb{E}\Big(f(U_1)\sum_{j=1}^{m}[t]_j\phi_j(U_1)\Big)^2\}$$
$$\leqslant\sup_{t\in\mathbb{B}_m}\{\sigma^2\|t\|_{L_2}^2+\|f\|_\infty^2\|t\|_{L_2}^2\}=\sigma^2+\|f\|_\infty^2=:\nu.\quad\text{(B.14)}$$

Replacing in Lemma 3.1 the constants $h$, $H$ and $v$ by (B.12), (B.13) and (B.14) respectively, there exists a finite numerical constant $C>0$ such that

$$\mathbb{E}\bigg(\sup_{t\in\mathbb{B}_m}|\overline{v_t^b}|^2-6\tau_\infty^2\sigma_Y^2mn^{-1}\bigg)_+\leqslant C\bigg[\frac{\sigma^2+\|f\|_\infty^2}{n}\exp\bigg(-\frac{m\tau_\infty^2\sigma_Y^2}{6(\sigma^2+\|f\|_\infty^2)}\bigg)$$
$$+\frac{2\tau_\infty^2m(\sigma+\|f\|_\infty)^2}{n^{3/2}}\exp(-\frac{K}{2}n^{1/4}\frac{\sigma_Y}{\sigma+\|f\|_\infty})\bigg].$$

The last upper bound and $\frac{\sigma^2+\|f\|_\infty^2}{\sigma_Y^2}=\frac{\sigma^2+\|f\|_\infty^2}{\sigma^2+\|f\|_{L_2}^2}\leqslant2\left(\frac{\sigma+\|f\|_\infty}{\sigma+\|f\|_{L_2}}\right)^2\leqslant2(1+\|f\|_\infty/\sigma)^2$ imply together the existence of a finite numerical constant $C>0$ such that

$$\mathbb{E}\bigg(\max_{1\leqslant m\leqslant n}\{\sup_{t\in\mathbb{B}_m}|\overline{v_t^b}|^2-6\tau_\infty^2\sigma_Y^2mn^{-1}\}\bigg)_+\leqslant C\frac{\sigma^2+\|f\|_\infty^2}{n}\Big[\sum_{m=1}^{n}\exp\bigg(-\frac{m\tau_\infty^2}{12(1+\|f\|_\infty/\sigma)^2}\bigg)$$
$$+n^{3/2}\tau_\infty^2\exp(-n^{1/4}\frac{K}{2(1+\|f\|_\infty/\sigma)})\Big]$$

and hence, from $\|f\|_\infty\leqslant r\mathfrak{A}$ for all $f\in\mathcal{F}_\mathfrak{a}^r$ due to assumption **(A2)** there exists a finite constant $C(r\mathfrak{A},\sigma,\tau_\infty)$ depending only on the quantities $r\mathfrak{A}$, $\sigma$ and $\tau_\infty$ such that

$$\sup_{f\in\mathcal{F}_\mathfrak{a}^r}\mathbb{E}\bigg(\max_{1\leqslant m\leqslant n}\{\sup_{t\in\mathbb{B}_m}|\overline{v_t^b}|^2-6\tau_\infty^2\sigma_Y^2mn^{-1}\}\bigg)_+\leqslant n^{-1}C(r\mathfrak{A},\sigma,\tau_\infty),\quad\text{for all }n\geqslant1.$$

The assertion of Proposition 3.8 follows now by combination of the last bound, (B.11) and the decomposition (B.10), which completes the proof. □

*Proof of Lemma 3.10.* We start the proof with the observation that $\mathcal{V}^c\subset\left\{\left|\frac{\widehat{\sigma}_Y^2}{\sigma_Y^2}-1\right|\geqslant\frac{1}{2}\right\}$ and, hence

$$P(\mathcal{V}^c)\leqslant P\left(\left|\frac{\widehat{\sigma}_Y^2}{\sigma_Y^2}-1\right|\geqslant\frac{1}{2}\right)=P\left(\left|n^{-1}\sum_{i=1}^{n}\left(\frac{Y_i^2}{\sigma_Y^2}-1\right)\right|\geqslant\frac{1}{2}\right).$$

Since $\mathbb{E}Y_i^2=\sigma_Y^2$ and employing Tchebysheff's inequality

$$P\left(\left|n^{-1}\sum_{i=1}^{n}\left(\frac{Y_i^2}{\sigma_Y^2}-1\right)\right|\geqslant\frac{1}{2}\right)\leqslant\frac{4}{n\sigma_Y^4}\mathbb{E}Y_1^4\leqslant\frac{128}{n}\left((\mathbb{E}\varepsilon^4)^{1/4}+\|f\|_\infty/\sigma\right)^4.$$

The assertion follows now by taking into account that $\|f\|_\infty\leqslant r\mathfrak{A}$ for all $f\in\mathcal{F}_\mathfrak{a}^r$, which completes the proof. □

# C Appendix: Proofs of Section 4

## C.1 Appendix: Proofs of Section 4.1

*Proof of Lemma 4.3.* Combining the assumption **(A1)** and Lemma 4.1 we get a first bound for its variance,

$$\sum_{j=1}^{m} \mathbb{V}\mathrm{ar}(\tfrac{1}{n}\sum_{i=1}^{n} \phi_j(X_i)) \leqslant \tfrac{1}{n}\mathbb{E}(\sum_{j=1}^{m} |\phi_j(X_0)|^2\{1+4\sum_{k=1}^{n-1} b(X_0)\}) \leqslant \tau_\infty^2\{1+4\sum_{k=1}^{n-1}\beta(X_0,X_k)\}mn^{-1}.$$

Then, the assertion 4.2 is an immediate consequence. □

*Proof of Lemma 4.4.* We start the proof with the observation that for any orthonormal system $\{\phi_j\}_{j=1}^m$ we have $\|\sum_{j=1}^m \phi_j \otimes \phi_j\|_{L_2}^2 = \sum_{j=1}^m \sum_{l=1}^m |\langle \phi_j, \phi_l\rangle_{L_2}|^2 = m$. Thereby, exploiting the assumption **(D1)** it follows that

$$\Big|\sum_{j=1}^{m} \mathbb{C}\mathrm{ov}(\phi_j(X_0),\phi_j(X_k))\Big| \leqslant \|\sum_{j=1}^{m}\phi_j\otimes\phi_j\|_{L^2}\|f_{X_0,X_k} - f_{X_0}\otimes f_{X_k}\|_{L_2} \leqslant \sqrt{m}\gamma_f. \quad \text{(C.1)}$$

On the other hand side, following the proof of Lemma 4.1 there exists a function $b_k : \mathbb{R} \to [0,1]$ with $\mathbb{E}b_k(X_0) = \beta(X_0,X_k)$ such that

$$\Big|\sum_{j=1}^{m} \mathbb{C}\mathrm{ov}(\phi_j(X_0),\phi_j(X_k))\Big| \leqslant 2\mathbb{E}(b_k(X_0)\{\sum_{j=1}^{m}\phi_j^2(X_0)\}) \leqslant 2m\tau_\infty^2\beta(X_0,X_k) \quad \text{(C.2)}$$

where the last inequality follows from the assumption **(A1)**. By combination of (C.1) and (C.2) we obtain for any $0 \leqslant K \leqslant n-1$

$$\begin{aligned}\sum_{k=1}^{n-1}(n+1-k)\sum_{j=1}^{m}\mathbb{C}\mathrm{ov}(\phi_j(X_0),\phi_j(X_k)) &\leqslant \sqrt{m}\gamma_f nK + 2m\tau_\infty^2 n\sum_{k=K+1}^{n-1}\beta(X_0,X_k)\\ &= mn\{\gamma_f K/\sqrt{m} + 2\tau_\infty^2\sum_{k=K+1}^{n-1}\beta(X_0,X_k)\}.\end{aligned}$$

From the last bound and the assumption **(A1)** we conclude that

$$\begin{aligned}\sum_{j=1}^{m}\mathbb{V}\mathrm{ar}(\sum_{i=1}^{n}\phi_j(X_i)) &= \sum_{j=1}^{m}\sum_{i=1}^{n}\mathbb{V}\mathrm{ar}(\phi_j(X_i)) + 2\sum_{j=1}^{m}\sum_{i=2}^{n}(n+1-i)\,\mathbb{C}\mathrm{ov}(\phi_j(X_1),\phi_j(X_i))\\ &\leqslant n\mathbb{E}\{\sum_{j=1}^{m}\phi_j^2(X_0)\} + 2\sum_{k=1}^{n-1}(n-k)\Big|\sum_{j=1}^{m}\mathbb{C}\mathrm{ov}(\phi_j(X_0),\phi_j(X_k))\Big|\\ &\leqslant nm\tau_\infty^2 + 2mn\{\gamma_f K/\sqrt{m} + 2\tau_\infty^2\sum_{k=K+1}^{n-1}\beta(X_0,X_k)\}\end{aligned}$$

which shows the assertion and completes the proof. □

*Proof of Proposition 4.6.* Following the construction presented in Section 4 let $(X_i)_{i\geqslant 1} = (E_l, O_l)_{l\geqslant 1}$ and $(X_i^\perp)_{i\geqslant 1} = (E_l^\perp, O_l^\perp)_{l\geqslant 1}$ be random vectors satisfying the coupling properties **(P1)**, **(P2)** and **(P3)**. Let $n$, $p$ and $q$ be integers such that $n = 2pq$. Let us introduce exactly in the same way $(x_1, \dots, x_n) = (e_1, o_1, \dots, e_p, o_p)$ with $e_l = (x_i)_{i\in\mathcal{I}_l^e}$ and $o_l = (x_i)_{i\in\mathcal{I}_l^o}$, $l = 1, \dots, p$. If we set further for any $x = (x_1, \dots, x_q) \in [0,1]^q$, $\vec{v}_t(x) := (1/q)\sum_{i=1}^q v_t(x_i)$, then $\frac{1}{n}\sum_{i=1}^n \nu_t(x_i) = \frac{1}{2}\Big\{\frac{1}{p}\sum_{l=1}^p \vec{v}_t(e_l) + \frac{1}{p}\sum_{l=1}^p \vec{v}_t(o_l)\Big\}$. Thereby, it follows for $\overline{\nu_t} = (1/n)\sum_{i=1}^n [\nu_t(X_i) - \mathbb{E}(\nu_t(X_i))] = \langle t, \widehat{f}_m - f_m\rangle$ that $\overline{\nu_t} =: \frac{1}{2}\{\overline{\nu_t^e} + \overline{\nu_t^o}\}$. Considering rather than $(X_i)_{i=1}^n$ the random variables $(X_i^\perp)_{i=1}^n$ we introduce additionally

$$\overline{\nu_t}^\perp = \frac{1}{2}\Bigg\{\frac{1}{p}\sum_{l=1}^p\{\vec{v}_t(E_l^\perp) - \mathbb{E}\vec{v}_t(E_l^\perp)\} + \frac{1}{p}\sum_{l=1}^p\{\vec{v}_t(O_l^\perp) - \mathbb{E}\vec{v}_t(O_l^\perp)\}\Bigg\} =: \frac{1}{2}\Big\{\overline{\nu_t^e}^\perp + \overline{\nu_t^o}^\perp\Big\}.$$

Using successively Jensen's inequality, i.e., $|\overline{\nu_t}|^2 \leqslant \frac{1}{2}\{|\overline{\nu_t^e}|^2 + |\overline{\nu_t^o}|^2\}$, $|a|^2 \leqslant 2\{|b|^2 + |a-b|^2\}$, $\mathbb{B}_m \leqslant \mathbb{B}_n$ for all $1 \leqslant m \leqslant n$ it follows that

$$\begin{aligned}
&\mathbb{E}\Big(\max_{m_n^\star\leqslant m\leqslant n}\Big\{\|\widehat{f}_m - f_m\|^2 - \tfrac{\mathrm{pen}_m}{6}\Big\}\Big)_+ \\
&\qquad\leqslant \mathbb{E}\Bigg(\max_{m_n^\star\leqslant m\leqslant n}\Bigg\{\sup_{t\in\mathbb{B}_m}|\overline{\nu_t^e}^\perp|^2 - \tfrac{\mathrm{pen}_m}{12}\Bigg\}\Bigg)_+ + \mathbb{E}\Bigg(\sup_{t\in\mathbb{B}_n}|\overline{\nu_t^e}^\perp - \overline{\nu_t^e}|^2\Bigg)_+ \\
&\qquad\quad + \mathbb{E}\Bigg(\max_{m_n^\star\leqslant m\leqslant n}\Bigg\{\sup_{t\in\mathbb{B}_m}|\overline{\nu_t^o}^\perp|^2 - \tfrac{\mathrm{pen}_m}{12}\Bigg\}\Bigg)_+ + \mathbb{E}\Bigg(\sup_{t\in\mathbb{B}_n}|\overline{\nu_t^o}^\perp - \overline{\nu_t^o}|^2\Bigg)_+.
\end{aligned}$$

The desired assertion follows by combining the last bound and Lemma C.1 and C.2 below. □

LEMMA C.1. *Under assumptions of Proposition 4.6. Suppose that $\mathfrak{B} := 2\sum_{k=0}^\infty (k+1)\beta_k < \infty$ and set $\Psi(x) := \sum_{m\geqslant 1}^\infty x^{1/2}m^{1/2}\exp(-m^{1/2}/(48x^{1/2})) < \infty$, for any $x > 0$, and $K_n := \lfloor 4\tau_\infty^2\sqrt{m_n^\star}/\gamma_f\rfloor$ then there exists a numerical constant $C > 0$ such that for any $\mu_n \geqslant \{3 + 8\sum_{k=K_n+1}^{q-1}\beta(X_0, X_k)\}$ holds*

$$\begin{aligned}
\sup_{f\in\mathcal{F}_\mathfrak{a}^r}\mathbb{E}\Bigg(\max_{m_n^\star\leqslant m\leqslant n}\Bigg\{\sup_{t\in\mathbb{B}_m}|\overline{\nu_t^e}^\perp|^2 - 6mn^{-1}\tau_\infty^2\mu_n\Bigg\}\Bigg)_+ &\leqslant Cn^{-1}\tau_\infty^2\Bigg\{\mu_n\Psi\Bigg(\frac{r\mathfrak{A}\mathfrak{B}}{\tau_\infty^2\mu_n^2}\Bigg) \\
&\qquad + nq^2\exp\Bigg(-\frac{n^{1/2}}{q}\frac{\mu_n^{1/2}}{144}\Bigg)\Bigg\};\\
\sup_{f\in\mathcal{F}_\mathfrak{a}^r}\mathbb{E}\Bigg(\max_{m_n^\star\leqslant m\leqslant n}\Bigg\{\sup_{t\in\mathbb{B}_m}|\overline{\nu_t^o}^\perp|^2 - 6mn^{-1}\tau_\infty^2\mu_n\Bigg\}\Bigg)_+ &\leqslant Cn^{-1}\tau_\infty^2\Bigg\{\mu_n\Psi\Bigg(\frac{r\mathfrak{A}\mathfrak{B}}{\tau_\infty^2\mu_n^2}\Bigg) \\
&\qquad + nq^2\exp\Bigg(-\frac{n^{1/2}}{q}\frac{\mu_n^{1/2}}{144}\Bigg)\Bigg\}.
\end{aligned}$$

*Proof of Lemma C.1.* We prove the first assertion, the proof of the second follows exactly in the same way and, hence we omit the details. We shall emphasise that $\overline{\nu_t^e}^\perp = p^{-1}\sum_{l=1}^p \vec{v}_t(E_l)$ where $(E_l)_{l=1}^p$ are iid., which we use below without further reference. Keep in mind that

$\vec{v}_t(x) := (1/q)\sum_{i=1}^q v_t(x_i)$ and set $\vec{\phi}_j(x) := (1/q)\sum_{i=1}^q \phi_j(x_i)$ for $x \in [0,1]^q$. In order to apply Talagrand's inequality we compute the constants $h$, $H$ and $v$. Consider first $h$ where

$$\sup_{t\in\mathcal{B}_m}\|\vec{\nu}_t\|_\infty^2 = \sup_{y\in[0,1]^q}\sum_{j=1}^m |\tfrac{1}{q}\sum_{i=1}^q \phi_j(y_i)|^2 \leqslant \tau_\infty^2 m =: h^2 \tag{C.3}$$

employing the assumption **(A1)**. Consider next $H$. From property **(P3)**, follows that

$$\mathbb{E}\sup_{t\in\mathcal{B}_m}|\overline{\nu_t^e}^\perp|^2 = \sum_{j=1}^m \mathbb{V}\mathrm{ar}\{\tfrac{1}{p}\sum_{l=1}^p \vec{\phi}_j(E_l^\perp)\} = \tfrac{1}{p}\sum_{j=1}^m \mathbb{V}\mathrm{ar}\{\vec{\phi}_j(E_1^\perp)\}$$

and hence exploiting the definition of $\vec{\phi}_j$ and the property **(P1)**, we have

$$\mathbb{E}\sup_{t\in\mathcal{B}_m}|\overline{\nu_t^e}^\perp|^2 = \tfrac{1}{p}\sum_{j=1}^m \mathbb{V}\mathrm{ar}\{\vec{\phi}_j(E_1^\perp)\} = \tfrac{1}{p}\sum_{j=1}^m \mathbb{V}\mathrm{ar}\{\vec{\phi}_j(E_1)\} = \tfrac{1}{p}\sum_{j=1}^m \mathbb{V}\mathrm{ar}\{\tfrac{1}{q}\sum_{i=1}^q \phi_j(X_i)\} \tag{C.4}$$

We employ next Lemma 4.4, thereby under the assumptions **(A1)** and **(D1)** we have for all $K\in\{0,\dots,q-1\}$ and for any $q \geqslant 1$

$$\sum_{j=1}^m \mathbb{V}\mathrm{ar}\{\tfrac{1}{q}\sum_{i=1}^q \phi_j(X_i)\} \leqslant \frac{m}{q}\{\tau_\infty^2 + 2[\gamma K/\sqrt{m} + 2\tau_\infty^2\sum_{k=K+1}^{q-1}\beta(X_0,X_k)]\}.$$

Given $K_n = \lfloor 4\tau_\infty^2\sqrt{m_n^\star}/\gamma\rfloor$ we have $\sum_{j=1}^m \mathbb{V}\mathrm{ar}\{\frac{1}{q}\sum_{i=1}^q \phi_j(X_i)\} \leqslant \frac{m}{q}\tau_\infty^2\{3/2+4\sum_{k=K_n+1}^{q-1}\beta_k\}$, for all $m \geqslant m_n^\star$. Thereby, from (C.4) follows for any $\mu_n \geqslant \{3+8\sum_{k=K_n+1}^\infty \beta_k\}$ that

$$\mathbb{E}\sup_{t\in\mathcal{B}_m}|\overline{\nu_t^e}^\perp|^2 \leqslant \frac{m}{n}\tau_\infty^2\{3+8\sum_{k=K_n+1}^\infty \beta_k\} \leqslant \frac{m}{n}\tau_\infty^2\mu_n =: H^2. \tag{C.5}$$

Consider $v$. Keep in mind that $\sup_{t\in\mathcal{B}_m}\frac{1}{p}\sum_{i=1}^p \mathbb{V}\mathrm{ar}(\vec{\nu}_t(E_i^\perp)) = \sup_{t\in\mathcal{B}_m}\mathbb{V}\mathrm{ar}(\frac{1}{q}\sum_{i=1}^q v_t(X_i))$ due to **(P1)** and **(P3)**, $\sup_{t\in\mathcal{B}_m}\mathbb{E}|v_t(X_1)|^2 \leqslant r\mathfrak{A}$, and $\sup_{t\in\mathcal{B}_m}\|v_t\|_\infty \leqslant m^{1/2}\tau_\infty$ given in (B.8) and (B.6), respectively. By applying (4.1) and setting $\mathfrak{B} = 2\sum_{k=0}^\infty (k+1)\beta_k$ we have

$$\begin{aligned}\sup_{t\in\mathcal{B}_m}\frac{1}{p}\sum_{i=1}^p \mathbb{V}\mathrm{ar}(\vec{\nu}_t(E_i^\perp)) &\leqslant \tfrac{4}{q}\sup_{t\in\mathcal{B}_m}\{\mathbb{E}|v_t(X_1)|^2\}^{1/2}\|v_t\|_\infty\{2\sum_{k=0}^\infty (k+1)\beta_k\}^{1/2}\\ &\leqslant \tfrac{4}{q}(mr\mathfrak{A}\mathfrak{B})^{1/2}\tau_\infty =: v. \end{aligned}\tag{C.6}$$

The assertion follows from Lemma 3.1 by using the quantities $h$, $H$ and $v$ given in (C.3), (C.5) and (C.6), respectively, and by employing the definition of $\Psi$, which completes the proof. □

LEMMA C.2. *Under assumptions of Proposition 4.6. We have*

$$\mathbb{E}\Big(\sup_{t\in\mathbb{B}_n}|\overline{\nu_t^e} - \overline{\nu_t^e}^\perp|^2\Big)_+ \leqslant 4\tau_\infty^2 n\beta_{q+1}, \quad \textit{and,} \quad \mathbb{E}\Big(\sup_{t\in\mathbb{B}_n}|\overline{\nu_t^o} - \overline{\nu_t^o}^\perp|^2\Big)_+ \leqslant 4\tau_\infty^2 n\beta_{q+1}.$$

*Proof of Lemma C.2.* Since $\{E_l\}_{l=1}^p$ and $\{E_l^\perp\}_{l=1}^p$ are identically distributed due to **(P1)** we have $|\overline{\nu_t^e} - \overline{\nu_t^{e\perp}}| = |p^{-1}\sum_{l=1}^p \{\vec{v}_t(E_l) - \vec{v}_t(E_l^\perp)\}| \leqslant 2\|\vec{\nu}_t\|_\infty \mathbb{1}_{\{E_l \neq E_l^\perp\}}$ and hence, by using **(P2)** it follows that

$$\mathbb{E}\Big(\sup_{t\in\mathcal{B}_n} |\overline{\nu_t^e} - \overline{\nu_t^{e\perp}}|^2\Big)_+ \leqslant 4\sup_{t\in\mathcal{B}_n}\|\vec{\nu}_t\|_\infty^2 p^{-1}\sum_{l=1}^p P(E_l \neq E_l^\perp) \leqslant 4\sup_{t\in\mathcal{B}_n}\|\vec{\nu}_t\|_\infty^2 \beta_{q+1}$$

which together with (C.3) shows the first assertion. The proof of the second assertion is made exactly in the same way, and hence we omit the details, which completes the proof. □

## C.2 Appendix: Proofs of Section 4.2

*Proof of Lemma 4.9.* Exploiting the assumption **(A1)** and Lemma 4.1 we obtain,

$$\begin{aligned}\sum_{j=1}^m \mathbb{V}\mathrm{ar}(\tfrac{1}{n}\sum_{i=1}^n(\sigma\varepsilon_i + f(U_i))\phi_j(U_i)) &\leqslant \frac{\sigma^2 m}{n} + \frac{1}{n}\|f\|_\infty^2\|\sum_{j=1}^m \phi_j^2\|_\infty\{1 + 4\sum_{k=1}^{n-1}\beta(U_0,U_k)\} \\ &\leqslant [\sigma^2 + \|f\|_\infty^2\tau_\infty^2\{1+4\sum_{k=1}^{n-1}\beta(U_0,U_k)\}]mn^{-1}. \qquad (C.7)\end{aligned}$$

Replacing (B.1) by (C.7), the assertion follows as in the proof of Proposition 3.2. □

*Proof of Lemma 4.10.* We start the proof with the observation that for any orthonormal system $\{\phi_j\}_{j=1}^m$ we have $\|\sum_{j=1}^m \phi_j\otimes\phi_j\|_{L_2}^2 = \sum_{j=1}^m\sum_{l=1}^m |\langle\phi_j,\phi_l\rangle|^2 = m$. Thereby, from **(D2)** follows

$$\begin{aligned}&\Big|\sum_{j=1}^m \mathbb{C}\mathrm{ov}(f(U_0)\phi_j(U_0), f(U_k)\phi_j(U_k))\Big| \\ &\qquad \leqslant \|\sum_{j=1}^m \phi_j\otimes\phi_j\|_{L_2}\|f\otimes f\{f_{U_0,U_k} - \mathbb{1}\otimes\mathbb{1}\}\|_{L_2} \leqslant \sqrt{m}\,\|f\|_{L_2}^2\gamma \qquad (C.8)\end{aligned}$$

On the other hand side, keeping in mind **(A1)** there exists a function $b_k : \mathbb{R} \to [0,1]$ with $\mathbb{E}b_k(U_0) = \beta(U_0,U_k)$ due to Lemma 4.1 in Viennet [1997] such that

$$\begin{aligned}\Big|\sum_{j=1}^m \mathbb{C}\mathrm{ov}(f(U_0)\phi_j(U_0), f(U_k)\phi_j(U_k))\Big| &\leqslant 2\mathbb{E}(b_k(U_0)\{f^2(U_0)\sum_{j=1}^m\phi_j^2(U_0)\}) \\ &\leqslant 2m\|f\|_\infty^2\tau_\infty^2\beta(U_0,U_k)\end{aligned}$$

which together with (C.8) implies for any $0 \leqslant K \leqslant n-1$

$$\begin{aligned}&\sum_{k=1}^{n-1}(n+1-k)\sum_{j=1}^m \mathbb{C}\mathrm{ov}(f(U_0)\phi_j(U_0), f(U_k)\phi_j(U_k)) \leqslant \sqrt{m}\|f\|_\infty^2\gamma nK \\ &\qquad + 2m\|f\|_\infty^2\tau_\infty^2 n\sum_{k=K+1}^{n-1}\beta(U_0,U_k) = mn\,\|f\|_\infty^2\{\gamma K/\sqrt{m} + 2\tau_\infty^2\sum_{k=K+1}^{n-1}\beta(U_0,U_k)\}.\end{aligned}$$

From the last bound and $\sum_{j=1}^m\sum_{i=1}^n \mathbb{V}\mathrm{ar}(f(U_i)\phi_j(U_i)) \leqslant nm\tau_\infty^2\|f\|_{L_2}^2$ due to **(A1)** follows the desired assertion. □

*Proof of Proposition 4.12.* Recalling the notations given in the proof of Proposition 3.8, our proof starts with the observation that a combination of (B.10) and (B.11) leads to

$$\mathbb{E}\Big(\max_{m_n^\star\leqslant m\leqslant n}\{\|\widehat{f}_m - f_m\|_{L_2}^2 - \tfrac{\mathrm{pen}_m}{6}\}\Big)_+ \leqslant 2\mathbb{E}\Big(\max_{m_n^\star\leqslant m\leqslant n}\{\sup_{t\in\mathbb{B}_m}|\overline{\nu_t^b}|^2 - \tfrac{\mathrm{pen}_m}{12}\}\Big)_+ + 2n^{-1}\sigma^2\tau_\infty^2\mathbb{E}(\varepsilon^6). \tag{C.9}$$

In order to bound the first rhs. term we use a construction similar to that in the proof of Proposition 4.6. Let $(U_i)_{i\geqslant 1} = (E_l, O_l)_{l\geqslant 1}$ and $(U_i^\perp)_{i\geqslant 1} = (E_l^\perp, O_l^\perp)_{l\geqslant 1}$ be random vectors satisfying the coupling properties **(P1)**, **(P2)** and **(P3)**. Introduce exactly in the same manner $(\varepsilon_i^b)_{i\geqslant 1} = (\vec{\varepsilon}_l^{\,be}, \vec{\varepsilon}_l^{\,bo})_{l\geqslant 1}$. If we set $\vec{v}_t(x,y) := (1/q)\sum_{i=1}^q v_t(x_i, y_i)$, then for $n = 2pq$ it follows

$$\overline{\nu_t^b} = \tfrac{1}{2}\Big\{\tfrac{1}{p}\sum_{l=1}^p\{\vec{v}_t(\vec{\varepsilon}_l^{\,be}, E_l) - \mathbb{E}\vec{v}_t(\vec{\varepsilon}_l^{\,be}, E_l)\} + \tfrac{1}{p}\sum_{l=1}^p\{\vec{v}_t(\vec{\varepsilon}_l^{\,bo}, O_l) - \mathbb{E}\vec{v}_t(\vec{\varepsilon}_l^{\,bo}, O_l)\}\Big\} =: \tfrac{1}{2}\{\overline{\nu_t^{be}} + \overline{\nu_t^{bo}}\}.$$

Considering the random variables $(U_i^\perp)_{i\geqslant 1}$ rather than $(U_i)_{i\geqslant 1}$ we introduce in addition

$$\overline{\nu_t^b}^\perp = \tfrac{1}{2}\Big\{\tfrac{1}{p}\sum_{l=1}^p\{\vec{v}_t(\vec{\varepsilon}_l^{\,be}, E_l^\perp) - \mathbb{E}\vec{v}_t(\vec{\varepsilon}_l^{\,be}, E_l^\perp)\} + \tfrac{1}{p}\sum_{l=1}^p\{\vec{v}_t(\vec{\varepsilon}_l^{\,bo}, O_l^\perp) - \mathbb{E}\vec{v}_t(\vec{\varepsilon}_l^{\,bo}, O_l^\perp)\}\Big\} =: \tfrac{1}{2}\{\overline{\nu_t^{be}}^\perp + \overline{\nu_t^{bo}}^\perp\}.$$

As in the proof of Proposition 4.6, it follows that

$$\begin{aligned}\mathbb{E}\Big(\max_{m_n^\star\leqslant m\leqslant n}\{\sup_{t\in\mathbb{B}_m}|\overline{\nu_t^b}|^2 - \tfrac{\mathrm{pen}_m}{12}\}\Big)_+ &\leqslant \mathbb{E}\Big(\max_{m_n^\star\leqslant m\leqslant n}\{\sup_{t\in\mathbb{B}_m}|\overline{\nu_t^{be}}^\perp|^2 - \tfrac{\mathrm{pen}_m}{24}\}\Big)_+ + \mathbb{E}\Big(\sup_{t\in\mathbb{B}_n}|\overline{\nu_t^{be}}^\perp - \overline{\nu_t^{be}}|^2\Big)_+ \\ &\quad + \mathbb{E}\Big(\max_{m_n^\star\leqslant m\leqslant n}\{\sup_{t\in\mathbb{B}_m}|\overline{\nu_t^{bo}}^\perp|^2 - \tfrac{\mathrm{pen}_m}{24}\}\Big)_+ + \mathbb{E}\Big(\sup_{t\in\mathbb{B}_n}|\overline{\nu_t^{bo}}^\perp - \overline{\nu_t^{bo}}|^2\Big)_+.\end{aligned}$$

The desired assertion follows by combining (C.9), the last bound, Lemma C.2 and C.3 . □

LEMMA C.3. *Let the assumptions **(A1)**, **(A2)**, **(P1)**, **(P3)**, and **(D2)** be satisfied. Suppose that* $\mathfrak{B} := 2\sum_{k=0}^\infty (k+1)\beta_k < \infty$. *Let* $K_n := \lfloor \tau_\infty^2\|f\|_{L_2}^2\sqrt{m_n^\star}/(\gamma r^2\mathfrak{A}^2)\rfloor$ *and* $\mu_n \geqslant \{3 + 8\sum_{k=K_n+1}^\infty \beta_k\}$. *There exist a finite constant* $\zeta(r\mathfrak{A}, \sigma, \tau_\infty, \mathfrak{B})$ *depending on the quantities* $r\mathfrak{A}$, $\sigma$, $\tau_\infty$ *and* $\mathfrak{B}$ *only and a numerical constant* $C > 0$ *such that for any holds*

$$\begin{aligned}&\sup_{f\in\mathcal{F}_{\mathfrak{a}}^r}\mathbb{E}\Big(\max_{m_n^\star\leqslant m\leqslant n}\sup_{t\in\mathbb{B}_m}|\overline{\nu_t^{be}}^\perp|^2 - 6\tfrac{m}{n}\sigma_Y^2\tau_\infty^2\mu_n\Big)_+ \\ &\qquad\leqslant Cn^{-1}\tau_\infty^2(\sigma + r\mathfrak{A})^2\Big\{\zeta(r\mathfrak{A}, \sigma, \tau_\infty, \mathfrak{B}) + n^{3/2}q^2\exp\Big(-\frac{n^{1/4}}{q}\frac{1}{576(1 + r\mathfrak{A}/\sigma)}\Big)\Big\};\\ &\sup_{f\in\mathcal{F}_{\mathfrak{a}}^r}\mathbb{E}\Big(\max_{m_n^\star\leqslant m\leqslant n}\sup_{t\in\mathbb{B}_m}|\overline{\nu_t^{bo}}^\perp|^2 - 6\tfrac{m}{n}\sigma_Y^2\tau_\infty^2\mu_n\Big)_+ \\ &\qquad\leqslant Cn^{-1}\tau_\infty^2(\sigma + r\mathfrak{A})^2\Big\{\zeta(r\mathfrak{A}, \sigma, \tau_\infty, \mathfrak{B}) + n^{3/2}q^2\exp\Big(-\frac{n^{1/4}}{q}\frac{1}{576(1 + r\mathfrak{A}/\sigma)}\Big)\Big\}.\end{aligned}$$

*Proof of Lemma C.3.* We prove the first assertion, the proof of the second follows exactly in the same way and, hence we omit the details. In order to apply Talagrand's inequality given in Lemma 3.1 we need to compute the constants $h$, $H$ and $v$ which verify the three required inequalities. Keep in mind that $\overline{\nu_t^{be}}^{\perp} = \frac{1}{p}\sum_{l=1}^{p} \vec{v}_t(\vec{\varepsilon}_l^{\,be}, E_l^{\perp}) - \mathbb{E}\vec{v}_t(\vec{\varepsilon}_l^{\,be}, E_l^{\perp})$ with $\vec{v}_t(\vec{\varepsilon}_l^{\,be}, E_l^{\perp}) = \sum_{j=1}^{m}[t]_j\vec{\psi}_j(\vec{\varepsilon}_l^{\,be}, E_l^{\perp})$, $\vec{\psi}_j(\vec{\varepsilon}_l^{\,be}, E_l^{\perp}) = (1/q)\sum_{i\in\mathcal{I}_l^e}\psi_j(\varepsilon_i^b, U_i^{\perp})$ and $\psi_j(\varepsilon_i^b, U_i^{\perp}) = (\sigma\varepsilon_i^b + f(U_i^{\perp}))\phi_j(U_i^{\perp})$, where $|\vec{\varepsilon}_l^{\,be}|_\infty = \max_{i\in\mathcal{I}_l^e}|\varepsilon_i^b| \leqslant 2n^{1/4}$ and $E_l^{\perp} \in [0,1]^q$. Consider first $h$. As in (B.12), the assumption **(A1)** implies

$$\sup_{t\in\mathcal{B}_m}\|\vec{\nu_t}\|_\infty^2 = \sum_{j=1}^{m}\|\vec{\psi}_j\|_\infty^2 \leqslant \sum_{j=1}^{m}\|\psi_j^2\|_\infty \leqslant \tau_\infty^2 m(2\sigma n^{1/4} + \|f\|_\infty)^2 =: h^2. \tag{C.10}$$

Consider next $H$. Exploiting successfully property **(P3)**, the definition of $\vec{\psi}_j$ and the property **(P1)** together with the independence within $\{\varepsilon_i\}$ and between $\{\varepsilon_i\}$ and $\{U_i\}$ we have

$$\mathbb{E}\sup_{t\in\mathcal{B}_m}|\overline{\nu_t^{be}}^{\perp}|^2 \leqslant \frac{2m\sigma^2\tau_\infty^2}{n} + \frac{1}{p}\sum_{j=1}^{m}\mathbb{V}\mathrm{ar}\Big(\frac{1}{q}\sum_{i=1}^{q} f(U_i)\phi_j(U_i)\Big). \tag{C.11}$$

Given $K_n = \lfloor 4\tau_\infty^2\|f\|_{L^2}^2\sqrt{m_n^\star}/(\gamma r^2\mathfrak{A}^2)\rfloor$, Lemma 4.10, assumptions **(A1)** and **(D2)** imply together for all $m \geqslant m_n^\star$ that

$$\sum_{j=1}^{m}\mathbb{V}\mathrm{ar}\Big(\frac{1}{q}\sum_{i=1}^{q} f(U_i)\phi_j(U_i)\Big) \leqslant \frac{m}{q}\tau_\infty^2\|f\|_{L^2}^2\{3/2 + 4\sum_{k=K_n+1}^{q-1}\beta(U_0,U_k)]\}.$$

Thereby, from (C.11) follows for any $\mu_n \geqslant \{3 + 8\sum_{k=K_n+1}^{\infty}\beta_k\}$ that

$$\mathbb{E}\sup_{t\in\mathcal{B}_m}|\overline{\nu_t^{e}}^{\perp}|^2 \leqslant \frac{2m}{n}\sigma^2\tau_\infty^2 + \frac{m}{n}\tau_\infty^2\|f\|_{L^2}^2\mu_n \leqslant \frac{m}{n}\tau_\infty^2\sigma_Y^2\mu_n =: H^2. \tag{C.12}$$

Consider finally $v$. Employing successively **(P3)**, **(P1)** and (4.1) we have

$$\begin{aligned}\sup_{t\in\mathcal{B}_m}\frac{1}{p}\sum_{l=1}^{p}\mathbb{V}\mathrm{ar}(\vec{\nu_t}(\vec{\varepsilon}_l^{\,be}, E_l^{\perp})) &\leqslant \frac{\sigma^2}{q} + \sup_{t\in\mathcal{B}_m}\mathbb{V}\mathrm{ar}(\frac{1}{q}\sum_{i=1}^{q} f(U_i)\sum_{j=1}^{m}[t]_j\phi_j(U_i))\\ &\leqslant \frac{\sigma^2}{q} + \frac{4}{q}\sup_{t\in\mathcal{B}_m}\{\mathbb{E}|f(U_i)\sum_{j=1}^{m}[t]_j\phi_j(U_i)|^2\}^{1/2}\|f\sum_{j=1}^{m}[t]_j\phi\|_\infty\{2\sum_{k=0}^{\infty}(k+1)\beta_k\}^{1/2}.\end{aligned} \tag{C.13}$$

Since $\sup_{t\in\mathcal{B}_m}\mathbb{E}|f(U_i)\sum_{j=1}^{m}[t]_j\phi_j(U_i)|^2 \leqslant \|f\|_\infty^2$, $\sup_{t\in\mathcal{B}_m}\|f\sum_{j=1}^{m}[t]_j\phi_j\|_\infty^2 \leqslant m\tau_\infty^2\|f\|_\infty^2$ and $\mathfrak{B} = 2\sum_{k=0}^{\infty}(k+1)\beta_k$ it follows that

$$\sup_{t\in\mathcal{B}_m}\frac{1}{p}\sum_{l=1}^{p}\mathbb{V}\mathrm{ar}(\vec{\nu_t}(\vec{\varepsilon}_l^{\,be}, E_l^{\perp})) \leqslant \frac{m^{1/2}\tau_\infty}{q}(\sigma^2 + 4\|f\|_\infty^2\mathfrak{B}^{1/2}) =: v. \tag{C.14}$$

The assertion follows from Lemma 3.1 by using the quantities $h$, $H$ and $v$ given in (C.10), (C.12) and (C.14), respectively, and by employing $\mu_n \geqslant 3/2$, $(\sigma + \|f\|_\infty)^2/\sigma_Y^2 \leqslant 2(1 + \|f\|_\infty/\sigma)^2$, and $\|f\|_\infty \leqslant r\mathfrak{A}$ for all $f \in \mathcal{F}_{\mathfrak{a}}^r$, which completes the proof. □

*Proof of Lemma 4.15.* Since $\mathbb{E}Y_1^2 = \sigma_Y^2$ using successively the Tchebysheff inequality, the inequality (4.1), the Cauchy-Schwarz inequality and Lemma 4.2 we get

$$P\left(\left|n^{-1}\sum_{i=1}^{n}\left(\frac{Y_i^2}{\sigma_Y^2}-1\right)\right| \geqslant \frac{1}{2}\right) \leqslant 16n^{-1}(\mathbb{E}Y_1^4/\sigma_Y^4)^{1/2}(2\sum_{k=0}^{\infty}(k+1)\beta_k)^{1/2}$$

which implies with $\mathbb{E}(Y_1^4/\sigma_Y^4) \leqslant 8\frac{\sigma^4\mathbb{E}\varepsilon^4+\|f\|_\infty^4}{(\sigma^2+\|f\|^2)^2} \leqslant 32\{\frac{\sigma(\mathbb{E}\varepsilon^4)^{1/4}+\|f\|_\infty}{\sigma+\|f\|}\}^4$ the desired assertion. □

# References


K. Bertin and N. Klutchnikoff. Pointwise adaptive estimation of the marginal density of a weakly dependent process. Technical report, Université Rennes 2, 2014.

L. Birgé and P. Massart. From model selection to adaptive estimation. Pollard, David (ed.) et al., Festschrift for Lucien Le Cam: research papers in probability and statistics. New York, NY: Springer. 55-87 (1997)., 1997.

D. Bosq. Nonparametric Statistics for Stochastic Processes. Springer, New York, 1998.

F. Comte and J. Johannes. Adaptive functional linear regression. The Annals of Statistics, 40 (6):2765–2797, 2012.

F. Comte and F. Merlevede. Adaptive estimation of the stationary density of discrete and continuous time mixing processes. ESAIM: Probability and Statistics, 6:211–238, 2002.

F. Comte and Y. Rozenholc. Adaptive estimation of mean and volatility functions in (auto-) regressive models. Stochastic Processes and their Applications, 97(1):111–145, 2002.

F. Comte, J. Dedecker, and M.-L. Taupin. Adaptive density deconvolution for dependent inputs with measurement errors. Mathematical Methods of Statistics, 17(2):87–112, 2008.

P. Doukhan and L. Truquet. Weakly dependent random fields with infinite interactions-paru sous le titre" a fixed point approach to model random fields". ALEA: Latin American Journal of Probability and Mathematical Statistics, 3:111–132, 2007.

I. Gannaz and O. Wintenberger. Adaptive density estimation under weak dependence. ESAIM: Probability and Statistics, 14:151–172, 2010.

A. Goldenshluger and O. Lepski. Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. The Annals of Statistics, 39:1608–1632, 2011.

T. Klein and E. Rio. Concentration around the mean for maxima of empirical processes. The Annals of Probability, 33(3):1060–1077, 2005.

A. Kolmogorov and Y. Rozanov. On the strong mixing conditions for stationary gaussian sequences. Theory of Probability and its Applications, 5:204–207, 1960.

M. Talagrand. New concentration inequalities in product spaces. Inventiones mathematicae, 126(3):505–563, 1996.

K. Tribouley and G. Viennet. $l_p$ adaptive density estimation in a $\beta$ mixing framework. Annales de l'IHP Probabilités et statistiques, 34(2):179–208, 1998.

G. Viennet. Inequalities for absolutely regular sequences: application to density estimation. Probability theory and related fields, 107(4):467–492, 1997.