



Adaptive nonparametric regression from repeated measurements under common noise

Fabienne Comte, Bianca Neubert

► To cite this version:

Fabienne Comte, Bianca Neubert. Adaptive nonparametric regression from repeated measurements under common noise. 2026. <hal-05669936>

HAL Id: hal-05669936

<https://hal.science/hal-05669936v1>

Preprint submitted on 25 Jun 2026

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

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Adaptive nonparametric regression from repeated measurements under common noise

FABIENNE COMTE^{*}, BIANCA NEUBERT[†]

Abstract

We consider nonparametric estimation of the regression function in a model where individuals share a common noise component and repeated measurements are available for each individual. We propose a projection estimator which minimizes a least-squares contrast that accounts for the covariance structure resulting from the common noise. We analyze its risk measured either as the expectation of the empirical norm or as the expectation of the theoretical norm associated with the contrast. We discuss how the number of repeated measurements affects the estimation rates in the common noise model, and precisely characterize the dependence on the number of repetitions. In addition, we propose a data-driven projection estimator and establish risk bounds in terms of the expected empirical norm. The results are illustrated with some simulation experiments.

Keywords: Common noise, Data-driven estimation, Highly dependent data, Nonparametric regression estimation, Repeated measurements

1 Introduction

In this paper, we consider the following model

$$Y_{i,j} = b(X_{i,j}) + \sqrt{\rho}\varepsilon_{0,j} + \sqrt{1-\rho}\varepsilon_{i,j} \quad (1)$$

for $i = 1, \dots, n$ and $j = 1, \dots, J$, which means that for each individual i we have J repeated observations of the model. The $X_{i,j}$ are independent and identically distributed (i.i.d.) real-valued random variables with density f , the $\varepsilon_{i,j}$ are i.i.d. and centered, with variance $\sigma^2 > 0$ and independent of the $(X_{i,j} : i = 1, \dots, n; j = 1, \dots, J)$. The parameter $\rho \in (0, 1)$ describes the correlation structure of the noise and is assumed to be known. The regression function $b: \mathbb{R} \rightarrow \mathbb{R}$ is unknown and needs to be estimated. We assume that b is square integrable and prove an upper bound on the risk of an adaptive nonparametric projection estimator.

The model is simple and interesting in itself, and refers to a context where a structural noise $\varepsilon_{0,j}$ is common to all individuals i , in addition to the standard one $\varepsilon_{i,j}$, and repeated measurements $j = 1, \dots, J$ are available for each individual. However, the truth is that we were first interested in model (1) as a simplified version of a Stochastic Differential Equation (SDE) model intensively studied in recent literature of probability and statistical inference for stochastic processes. The model arose as a way to capture the effect of some random external force influencing simultaneously all individuals i such as environment or constraints of energetic or economic type. This idea has been developed in finance (see *e.g.* Duellmann et al. [2010]) or in the theory of mean field games and control, for models of SDEs with common noise (see *e.g.* Carmona et al. [2016], Lacker et al. [2022], Maillet [2025], Delarue et al. [2024]). Indeed, these authors consider a system of SDEs of the form:

$$dX_i(t) = b(X_i(t)) dt + \sigma(\sqrt{\rho}dW_0(t) + \sqrt{1-\rho}dW_i(t)), \quad X_i(0) = x, \quad i = 1, \dots, n,$$

where $W_0, W_1, \dots, W_n$ are $n+1$ independent Brownian motions. Thus, W_0 is common to all equations and this creates identically distributed but correlated processes.

^{*}Université Paris Cité, MAP5 UMR 8145, F-75006 Paris, France, e-mail: fabienne.comte@u-paris.fr

[†]Ruprecht-Karls-Universität Heidelberg, Institut für Mathematik, MATHEMATIKON, Im Neuenheimer Feld 205, D-69120 Heidelberg, Germany, e-mail: neubert@math.uni-heidelberg.de

Now, imagine that you have discrete time observations of the n paths, $X_i(t_j)$, for $t_j = j/J$, and $i = 1, \dots, n$. Then setting $Y_{i,j} = J(X_i(t_{j+1}) - X_i(t_j))$, for J large enough, the approximation

$$Y_{i,j} \simeq b(X_i(t_j)) + \sqrt{\rho}\sigma J(W_0(t_{j+1}) - W_0(t_j)) + \sqrt{1-\rho}\sigma J(W_i(t_{j+1}) - W_i(t_j)),$$

gives a regression model. This leads to the simplified model described in (1).

As already mentioned, Model (1) is worth being studied as an interesting extension of the standard regression problem. That is, the estimation of the regression function in the nonparametric setting has a long history, and we refer to [Tsybakov \[2009\]](#) or [Efromovich \[1999\]](#) for general presentations of the topic and the associated estimation strategies. However, the literature mainly deals with the case $\rho = 0$. The present model involves a specificity in the strong dependency involved by the common noise feature corresponding to $\rho \neq 0$, which makes results related to classical dependent settings inapplicable, see [Caron et al. \[2021\]](#) and the references therein. Indeed, in [Caron et al. \[2021\]](#), the spectral radius of the noise variance matrix must be uniformly bounded, which is not the case in our model (it has order $O(n)$, see our matrix R below in Section 2.1); or the correlation between error terms must tend to 0, which is not the case here neither; or the variance of the sum of noises must be at most of order nJ , while here, it is equal to $\sigma^2(nJ)(n\rho + 1 - \rho)$, which still tends to infinity when divided by nJ . Regression models with dependent noise have also been considered in [Csörgö and Mielniczuk \[1995\]](#), [Baraud et al. \[2001b\]](#) or [Fourdrinier and Pergamenschikov \[2007\]](#) but the present fixed dependency violates all standard assumptions and in particular the ones in these papers (decreasing correlations in [Csörgö and Mielniczuk \[1995\]](#), decreasing mixing coefficients in [Baraud et al. \[2001b\]](#) or uncorrelated noise in [Fourdrinier and Pergamenschikov \[2007\]](#)).

A previous paper considers a similar context and introduces tools which are reused here: [Comte and Genon-Catalot \[2026\]](#) study model (1) with $J = 1$, and build a nonadaptive projection estimator. In the present work, we generalize their results to $J \geq 1$. In some sense, we go from n to nJ observations and everything may at first sight be changed in accordance. But this is not exactly the case. Indeed, the risk bound is modified from a decomposition into the squared bias and a variance term of order $\dim(S_m)/n$ to the squared bias and a variance term of order $\dim(S_m)/(nJ)$, where S_m denotes the finite dimensional projection space on which the estimator is built. But the stability condition associated to the matrix to be inverted for evaluating the least-squares estimator (see [Cohen et al. \[2013\]](#)) only changes from n to $n\sqrt{J}$ and not nJ , roughly speaking. The fact that different rates arise is somehow in accordance with parametric results obtained in [Genon-Catalot and Larédo \[2026\]](#); therein, if $b = b_\theta$ and the noise is re-parameterized by $\sigma(\sqrt{\rho}dW_0(t) + \sqrt{1-\rho}dW_i(t)) = \sqrt{v}(dW_i(t) + \sqrt{c}dW_0(t))$, then the parameters (v, c, θ) are estimated at rates $(\sqrt{nJ}, \sqrt{J}, \sqrt{n})$, the variance of Brownian increments $W_i(t_{j+1}) - W_i(t_j)$ being equal to $1/J$. It is worth noticing that having J repetitions also solves in part the identifiability problem encountered in [Comte and Genon-Catalot \[2026\]](#) for $\mathbb{E}[b(X_1)] \neq 0$, see the discussion in Section 2.3. We may underline that the generalization of the deviation underlying the control of the risk of the estimator is not straightforward and requires additional work. In addition to this generalization, we propose an adaptive procedure and we prove a risk bound, which requires the study of a new empirical process. This is also novel for the case $J = 1$, for which no adaptive estimator has been examined yet.

The plan of the paper is the following. We first describe in Section 2 the notation, the projection spaces (models) and the contrast. Our nonparametric estimator is the minimizer on a model of this new contrast. These first elements allow us to discuss the identifiability of Model (1). The empirical and theoretical risks are upper bounded in Section 3. The first bound expressed with respect to a random distance is quite straightforward. The risk bound expressed in term of the integrated version of the empirical norm requires more technicalities to prove that the two norms are equivalent with probability close to 1 (see Lemma 4 as a step to obtain Theorem 1). A model selection procedure is defined in Section 4 and the resulting estimator is proved to be adaptive (see Theorem 2). It also relies on Lemma 4 but, in addition, on the study of an empirical process involving dependency (Lemma 7). Section 5 gives an account of simulation experiments that also provide interesting results and discussions. Most proofs are relegated to Section 6.

2 Model and minimal contrast estimator

2.1 General notation

We consider the model introduced in (1). Before defining the estimator of regression function b in the next section, let us introduce some notation. First, define the $\mathbb{R}^n$ -valued vectors

$$\mathbf{X}_{\bullet,j} := (X_{i,j})_{1 \leq i \leq n}, \quad \mathbf{Y}_{\bullet,j} := (Y_{i,j})_{1 \leq i \leq n}.$$

Further set

$$\mathbf{X}_{\bullet,\bullet} := \begin{pmatrix} \mathbf{X}_{\bullet,1} \\ \vdots \\ \mathbf{X}_{\bullet,J} \end{pmatrix} \in \mathbb{R}^{nJ}, \quad \mathbf{Y}_{\bullet,\bullet} := \begin{pmatrix} \mathbf{Y}_{\bullet,1} \\ \vdots \\ \mathbf{Y}_{\bullet,J} \end{pmatrix} \in \mathbb{R}^{nJ}.$$

Set $\mathbf{u}_{\bullet,j} := (\sqrt{\rho}\varepsilon_{0,j} + \sqrt{1-\rho}\varepsilon_{i,j})_{1 \leq i \leq n} \in \mathbb{R}^n$ and

$$\mathbf{u}_{\bullet,\bullet} := \begin{pmatrix} \mathbf{u}_{\bullet,1} \\ \vdots \\ \mathbf{u}_{\bullet,J} \end{pmatrix}.$$

Consequently, Model (1) can also be written as

$$\mathbf{Y}_{\bullet,\bullet} = b(\mathbf{X}_{\bullet,\bullet}) + \mathbf{u}_{\bullet,\bullet}.$$

Note that the $(\mathbf{u}_{\bullet,j})_{1 \leq j \leq n}$ are i.i.d. vectors but elements within vectors are not independent. To describe the dependence structure resulting from the common noise component, define

$$\text{var}(\mathbf{u}_{\bullet,1})/\sigma^2 := R := \begin{pmatrix} 1 & & \rho \\ & \ddots & \\ \rho & & 1 \end{pmatrix} \in \mathbb{R}^{n \times n}.$$

Then, for the matrix

$$\mathbf{R} := \begin{pmatrix} R & & 0 \\ & \ddots & \\ 0 & & R \end{pmatrix} \in \mathbb{R}^{nJ \times nJ}$$

it holds $\text{var}(\mathbf{u}_{\bullet,\bullet}) = \sigma^2 \mathbf{R}$. The eigenvalues both of R and $\mathbf{R}$ are $1-\rho$ and $1+(n-1)\rho$ with multiplicity $n-1$ and 1 in R respectively. Consequently, for any $\rho \in (0, 1)$ both matrices are positive definite. The inverse matrices are given by

$$R^{-1} := \begin{pmatrix} \alpha_n & & \beta_n \\ & \ddots & \\ \beta_n & & \alpha_n \end{pmatrix}, \quad \text{with } \alpha_n := \frac{1+(n-2)\rho}{(1-\rho)(1+(n-1)\rho)}, \quad \beta_n := -\frac{\rho}{(1-\rho)(1+(n-1)\rho)}.$$

For Id_n the $n \times n$ identity matrix and $\mathbf{1}_n$ the n -dimensional vector with all coordinates equal to 1, this matrix can be written as

$$R^{-1} = (\alpha_n - \beta_n)\text{Id}_n + \beta_n \mathbf{1}_n \mathbf{1}_n^\top = \frac{1}{1-\rho} \left(\text{Id}_n - \frac{\rho}{1+(n-1)\rho} \mathbf{1}_n \mathbf{1}_n^\top \right). \quad (2)$$

In addition, it holds that

$$\mathbf{R}^{-1} := \begin{pmatrix} R^{-1} & & 0 \\ & \ddots & \\ 0 & & R^{-1} \end{pmatrix} \in \mathbb{R}^{nJ \times nJ}.$$

The corresponding eigenvalues of R^{-1} and $\mathbf{R}^{-1}$ are

$$\alpha_n - \beta_n = \frac{1}{1-\rho} \quad \text{and} \quad c_n := \frac{1}{1+(n-1)\rho}.$$

2.2 Minimal contrast estimator

Let us introduce a minimal contrast estimator for the regression function $b \in \mathbb{L}^2(A)$ of Model (1), with $A \subset \mathbb{R}$. For this, we consider an orthonormal basis $(\varphi_k)_{k \geq 1}$ of $\mathbb{L}^2(A)$ and associate to the basis a collection of models

$$S_m := \text{span}\{\varphi_1, \dots, \varphi_m\}, \quad m \in \mathbb{N}.$$

The idea is to build an estimator $\hat{b}_m = \sum_{k=1}^m \hat{a}_k \varphi_k$ by considering a generalized least squares contrast given for $t \in S_m$ by

$$\gamma_{n,J}(t) := \frac{1}{nJ} (t(\mathbf{X}_{\bullet,\bullet})^\top \mathbf{R}^{-1} t(\mathbf{X}_{\bullet,\bullet}) - 2 \mathbf{Y}_{\bullet,\bullet}^\top \mathbf{R}^{-1} t(\mathbf{X}_{\bullet,\bullet})).$$

First, let us introduce further notation. For $j \in \{1, \dots, J\}$, set $\hat{\Phi}_{m,j} := (\varphi_k(X_{i,j}))_{1 \leq i \leq n, 1 \leq k \leq m} \in \mathbb{R}^{n \times m}$ and define

$$\hat{\Phi}_m := \begin{pmatrix} \hat{\Phi}_{m,1} \\ \vdots \\ \hat{\Phi}_{m,J} \end{pmatrix} \in \mathbb{R}^{nJ \times m}$$

and for $\hat{\Psi}_{m,j}^{(R)} := \frac{1}{n} \hat{\Phi}_{m,j}^\top R^{-1} \hat{\Phi}_{m,j} \in \mathbb{R}^{m \times m}$ we set

$$\hat{\Psi}_m^{(R)} := \frac{1}{J} \sum_{j=1}^J \hat{\Psi}_{m,j}^{(R)} = \frac{1}{nJ} \hat{\Phi}_m^\top \mathbf{R}^{-1} \hat{\Phi}_m \in \mathbb{R}^{m \times m}.$$

The standard minimization of $\gamma_{n,J}$ over S_m leads to the vector of coefficients $\hat{\mathbf{a}}_m = (\hat{a}_1, \dots, \hat{a}_m) \in \mathbb{R}^m$ such that

$$\hat{\Psi}_m^{(R)} \hat{\mathbf{a}}_m = \hat{Z}_m$$

with $\hat{Z}_m := \frac{1}{n} \hat{\Phi}_m^\top R^{-1} \mathbf{Y}_{\bullet,\bullet} \in \mathbb{R}^m$ and

$$\hat{Z}_m := \frac{1}{J} \sum_{j=1}^J \hat{Z}_{m,j} = \frac{1}{nJ} \hat{\Phi}_m^\top \mathbf{R}^{-1} \mathbf{Y}_{\bullet,\bullet} \in \mathbb{R}^m.$$

Consequently, if $\hat{\Psi}_m^{(R)}$ is invertible, we obtain

$$\hat{\mathbf{a}}_m = (\hat{\Psi}_m^{(R)})^{-1} \frac{1}{nJ} \hat{\Phi}_m^\top \mathbf{R}^{-1} \mathbf{Y}_{\bullet,\bullet}$$

which leads to the proposed estimator $\hat{b}_m := \sum_{k=1}^m \hat{a}_k \varphi_k$ of b in the regression model.

Remark 1. Observe that it holds $\hat{\Phi}_m \hat{\mathbf{a}}_m = \hat{b}_m(\mathbf{X}_{\bullet,\bullet})$ and

$$P := \frac{1}{Jn} \hat{\Phi}_m (\hat{\Psi}_m^{(R)})^{-1} \hat{\Phi}_m^\top \mathbf{R}^{-1} \in \mathbb{R}^{nJ \times nJ}$$

satisfies $P = PP$ and $P^\top \mathbf{R}^{-1} = \mathbf{R}^{-1} P$. Consequently, P defines an orthogonal projection on $\mathbb{R}^{nJ}$ with respect to the scalar product induced by $\mathbf{R}^{-1}$.

Remark 2. Note that, as more specifically developed in [Comte and Genon-Catalot \[2026\]](#), using the representation in (2), for n large enough, R^{-1} is equivalent to

$$\frac{1}{1-\rho} \left(\text{Id}_n + \frac{1}{n} \mathbf{1}_n^\top \mathbf{1}_n \right),$$

and since the matrix appears twice in the formula of $\hat{\mathbf{a}}_m$, the factor $1/(1-\rho)$ cancels. Thus, in the definition of $\hat{\mathbf{a}}_m$ the matrix R^{-1} can be replaced by $\text{Id}_n + (1/n) \mathbf{1}_n^\top \mathbf{1}_n$, which does not require the knowledge of ρ .

2.3 About level identifiability

Now, we discuss the constraints implicitly contained in the invertibility assumption on $\widehat{\Psi}_m^{(R)}$. For this, we first introduce norms adequate for the model considered in this work.

For $t \in S_m$ with $t = \sum_{k=1}^m a_k \varphi_k$ for $\mathbf{a} = (a_1, \dots, a_m)^\top \in \mathbb{R}^m$, the extended empirical norm of the problem is

$$\|t\|_{n,J,R}^2 = \frac{1}{J} \sum_{j=1}^J \|t\|_{n,R,j}^2, \quad \|t\|_{n,R,j}^2 = \frac{1}{n} t(\mathbf{X}_{\bullet,j})^\top R^{-1} t(\mathbf{X}_{\bullet,j}) = \mathbf{a}^\top \widehat{\Psi}_{m,j}^{(R)} \mathbf{a}.$$

It holds that $t(\mathbf{X}_{\bullet,\bullet}) = \widehat{\Phi}_m \mathbf{a}$. Consequently, the empirical norm can also be rewritten as

$$\|t\|_{n,J,R}^2 = \frac{1}{Jn} t(\mathbf{X}_{\bullet,\bullet})^\top \mathbf{R}^{-1} t(\mathbf{X}_{\bullet,\bullet}) = \mathbf{a}^\top \frac{1}{Jn} \widehat{\Phi}_m^\top \mathbf{R}^{-1} \widehat{\Phi}_m \mathbf{a} = \mathbf{a}^\top \widehat{\Psi}_m^{(R)} \mathbf{a}.$$

The corresponding theoretical counterpart is defined as $\|t\|_R^2 := \mathbb{E}[\|t\|_{n,J,R}^2]$. Consequently, setting $\Psi_m^{(R)} = \mathbb{E}[\widehat{\Psi}_m^{(R)}]$, it also holds that

$$\|t\|_R^2 = \mathbf{a}^\top \Psi_m^{(R)} \mathbf{a}.$$

Moreover, the following Lemma provides a link between $\widehat{\Psi}_m^{(R)}$ and the covariance matrix.

Lemma 1. *For any $t \in S_m$ with $t = \sum_{k=1}^m a_k \varphi_k$ and $\mathbf{a} = (a_1, \dots, a_m)^\top \in \mathbb{R}^m$, for all $J \geq 1$, it holds that*

$$\|t\|_{n,J,R}^2 = \mathbf{a}^\top \widehat{\Psi}_m^{(R)} \mathbf{a} \xrightarrow{n \rightarrow +\infty} \frac{1}{1-\rho} \mathbf{a}^\top \Psi_m \mathbf{a}, \quad \text{almost surely,}$$

where

$$\Psi_m := (\text{cov}(\varphi_{k_1}(X_{1,1}), \varphi_{k_2}(X_{1,1})))_{1 \leq k_1, k_2 \leq m}.$$

Note that the theoretical objects $\Psi_m^{(R)}$, $\|\cdot\|_R$ and Ψ_m are equivalent to the objects of [Comte and Genon-Catalot \[2026\]](#) with the same names. Consequently, results about these objects can be directly used. However, empirical objects depend on J and care must be taken to extend results.

Lemma 1 shows the correspondence between the empirical matrix $\widehat{\Psi}_m^{(R)}$ and Ψ_m , which is not the matrix classically considered in nonparametric regression methods¹. This covariance matrix can not be invertible if the basis contains a constant function; indeed then, the basis functions are compactly supported, the support A is chosen such that all $X_{i,j}$ are in A a.s., and thus $\text{cov}(\mathbf{1}_A(X_{1,1}), \varphi_k(X_{1,1})) \simeq 0$. This enhances a level problem in the estimation procedure: if the constant function is φ_1 the first function of the basis on $A = [\mathbf{a}_1, \mathbf{a}_2]$, i.e. $\varphi_1 = (1/\sqrt{\mathbf{a}_2 - \mathbf{a}_1}) \mathbf{1}_{[\mathbf{a}_1, \mathbf{a}_2]}$, then the first term of the development of b in the basis is

$$\langle b, \varphi_1 \rangle \varphi_1(\cdot) = \left(\frac{1}{\mathbf{a}_2 - \mathbf{a}_1} \int_{\mathbf{a}_1}^{\mathbf{a}_2} b(x) dx \right) \mathbf{1}_{[\mathbf{a}_1, \mathbf{a}_2]}(\cdot),$$

and it can not be estimated. This is the reason why in this work we consider bases of $\mathbb{L}^2(A)$ without any constant functions, which makes easy comparisons possible between our setting and the standard context, that is, $\rho = 0$.

Another way of looking at the problem is to consider $Y_{i,j} = \mu + u_{i,j}$; then we can estimate μ with

$$\widehat{\mu} = \frac{1}{nJ} \sum_{i=1}^n \sum_{j=1}^J Y_{i,j}.$$

We get an unbiased and consistent estimator of μ , but we note that the variance of $\widehat{\mu}$ is of order $1/J$ (equal to $\rho/J + (1-\rho)/(nJ)$). The level correction is mandatory in the estimation procedure, and we make the transformation $Y_{i,j} - \widehat{\mu}$ for the estimation of $b^* = b - \mu$, and then compute errors for $\widehat{b^*} + \widehat{\mu}$, leading to a consistency which could not be reached for $J = 1$ in [Comte and Genon-Catalot \[2026\]](#). This is however not completely satisfactory since, as explained above, the correct level correction is rather the mean integral of b on the domain. Further discussions are given in Section 5.

¹Indeed, usually the matrix to invert is the cross-moment $(\mathbb{E}[\varphi_{k_1}(X_{1,1})\varphi_{k_2}(X_{1,1})])_{1 \leq k_1, k_2 \leq m}$, also denoted by Ψ_m in several papers.

3 Risk bounds in empirical and theoretical norms

The aim is to bound the risk of the estimator proposed in Section 2 first in integrated empirical norm $\mathbb{E}[\|\cdot\|_{n,J,R}^2]$ and, subsequently, in integrated theoretical norm $\mathbb{E}[\|\cdot\|_R^2]$.

3.1 First bound in empirical norm

We start by considering the risk of the estimator proposed as expectation of the empirical norm $\|\cdot\|_{n,J,R}^2$.

Proposition 1. *Assume that $\widehat{\Psi}_m^{(R)}$ is invertible. The risk bound on $\widehat{b}_m$ expressed as the expectation of the empirical norm $\|\cdot\|_{n,J,R}$ satisfies*

$$\mathbb{E}[\|b - \widehat{b}_m\|_{n,J,R}^2] \leq \inf_{t \in S_m} \|b - t\|_R^2 + \sigma^2 \frac{m}{nJ}.$$

We refer to the appendix for the proof. We observe that the risk decomposition is a generalization of the one obtained in Proposition 1 of Comte and Genon-Catalot [2026], where the variance term m/n for $J = 1$ becomes $m/(nJ)$ with only the invertibility constraint on $\widehat{\Psi}_m^{(R)}$.

Since $\mathbf{R}$ is also a variance matrix with the same eigenvalues as R and its inverse $\mathbf{R}^{-1}$ is also diagonalizable with the same eigenvalues as R^{-1} (with multiplicities $J(n-1)$ and J), we can extend Lemma 1 in Comte and Genon-Catalot [2026] as follows.

Lemma 2. *The following properties hold:*

1. $\|t\|_{n,J,R}^2 \geq 0$ with

$$\frac{1}{1 + (n-1)\rho} \|t\|_{n,J}^2 \leq \|t\|_{n,J,R}^2 \leq \frac{1}{1-\rho} \|t\|_{n,J}^2$$

$$\text{where } \|t\|_{n,J}^2 := \frac{1}{nJ} t(\mathbf{X}_{\bullet,\bullet})^\top t(\mathbf{X}_{\bullet,\bullet}).$$

2. $\|t\|_R^2 \geq 0$ with

$$\frac{1}{1 + (n-1)\rho} \mathbb{E}[t^2(X_{1,1})] \leq \|t\|_R^2 \leq \frac{1}{1-\rho} \mathbb{E}[t^2(X_{1,1})].$$

If f is bounded and $b_m := \sum_{j=1}^m \langle b, \varphi_j \rangle \varphi_j$ denotes the orthogonal projection of b on S_m , then, by Lemma 2,

$$\inf_{t \in S_m} \|b - t\|_R^2 \leq \frac{1}{1-\rho} \inf_{t \in S_m} \|b - t\|_R^2 \leq \frac{\|f\|_\infty}{1-\rho} \|b_m - b\|^2,$$

where $\|\cdot\|$ denotes the integral norm in $\mathbb{L}^2(\mathbb{R})$. Clearly $\|b_m - b\|^2$ tends to zero when m increases. In contrast, the variance term m/nJ increases with m . Thus, Proposition 1 exhibits a standard bias/variance decomposition requiring a compromise, and is associated with a sample size nJ without loss despite the dependency introduced by the common noise. Clearly, under standard regularity assumptions related to a regularity order for b , a nonparametric rate of order of a negative power of nJ can be expected: this explains why a level correction with risk of order $1/J$ is not completely satisfactory, see Section 2.3.

3.2 Risk bound in theoretical norm

To obtain a result with respect to the risk in theoretical norm, we have to compare precisely the empirical norm to its expectation. We give in Lemma 3 hereafter the decomposition of the empirical norm, involving several processes, whose deviation have to be controlled in the following.

Lemma 3. *For the empirical norm the following decomposition holds.*

$$\|t\|_{n,J,R}^2 = \alpha_n Z_{n,J}(t) + c_n (\mathbb{E}[t(X_{1,1})])^2 + 2c_n \mathbb{E}[t(X_{1,1})] V_{n,J}(t) + (n-1) \beta_n U_{n,J}(t)$$

with

$$Z_{n,J}(t) := \frac{1}{J} \sum_{j=1}^J Z_{n,j}(t), \quad Z_{n,j}(t) = \frac{1}{n} \sum_{i=1}^n (t(X_{i,j}) - \mathbb{E}[t(X_{1,1})])^2 \quad (3)$$

$$V_{n,J}(t) := \frac{1}{J} \sum_{j=1}^J V_{n,j}(t), \quad V_{n,j}(t) = \frac{1}{n} \sum_{i=1}^n (t(X_{i,j}) - \mathbb{E}[t(X_{1,1})])$$

$$U_{n,J}(t) := \frac{1}{J} \sum_{j=1}^J U_{n,j}(t), \quad U_{n,j}(t) = \frac{1}{n(n-1)} \sum_{1 \leq i \neq k \leq n} (t(X_{i,j}) - \mathbb{E}[t(X_{1,1})])(t(X_{k,j}) - \mathbb{E}[t(X_{1,1})]). \quad (4)$$

The main processes of the decomposition of $\|t\|_{n,J,R}^2$ are: $Z_{n,J}(t)$, which is studied thanks to a matrix Chernov deviation inequality (Lemma 12 from [Tropp \[2012\]](#) recalled in the Appendix); and $U_{n,J}(t)$ which is a degenerated U -statistics of second order and is treated with the corresponding deviation inequality stated in [Giné and Nickl \[2016\]](#) (Theorem 3.4.8 p. 183; see also Theorem 3.4 in [Houdré and Reynaud-Bouret \[2003\]](#)). As $V_{n,J}(t)$ and $U_{n,J}(t)$ are centered, it straightforwardly holds that

$$\|t\|_R^2 = \mathbb{E}[\|t\|_{n,J,R}] = \alpha_n \text{Var}(t(X_{1,1})) + c_n (\mathbb{E}[t(X_{1,1})])^2.$$

As $\alpha_n \rightarrow_{n \rightarrow +\infty} 1/(1-\rho)$ and $c_n = 1/[1+(n-1)\rho] \rightarrow_{n \rightarrow +\infty} 0$, we find, in accordance with Lemma 1, that

$$\lim_{n \rightarrow +\infty} \|t\|_R^2 = \frac{1}{1-\rho} \text{Var}(t(X_{1,1})).$$

We note that the following equality holds:

$$\|t\|_{n,J,R}^2 = \frac{1}{1-\rho} \frac{1}{nJ} \sum_{j=1}^J \sum_{i=1}^n \left[t(X_{i,j}) - \frac{1}{n} \sum_{k=1}^n t(X_{k,j}) \right]^2 + c_n \frac{1}{J} \sum_{j=1}^J \left(\frac{1}{n} \sum_{i=1}^n t(X_{i,j}) \right)^2.$$

Therefore, $\|t\|_{n,J,R}^2 = 0$ implies $t(X_{i,j}) = 0$ a.s. for all i, j . For some orthonormal bases, such as the Hermite or Tchebychev basis, $\|t\|_{n,J,R}^2 = 0$ for $t \in S_m$ even implies that $t = 0$ for nJ larger than m , see Section 5 for details on the bases.

Next, the aim is to handle the integrated risk with respect to the squared norm $\|\cdot\|_R^2$. As discussed in [Comte and Genon-Catalot \[2026\]](#), this task is far from simple and the extension to the more general setting of this, i.e. $J > 1$, also requires an adaptation of most steps. To bound the risk in theoretical norm, the following assumption is necessary.

Assumption 1. (i) Assume that for

$$L(m) := \sup_{x \in \mathbb{R}} \sum_{k=0}^m \varphi_k^2(x) < +\infty$$

it holds $L(m) = m^s$ and the functions φ_j are such that $\sup_{x \in A} |\varphi_j(x)| \leq \theta$, for some $\theta, s > 0$.

(ii) For $m, n, J \in \mathbb{N}$ and some constant $c^* > 0$ assume that the following stability condition holds:

$$(m \vee L(m))(\|\Psi_m^{-1}\|_{\text{op}} \vee 1) \leq c^* \frac{n\sqrt{J}}{\log(n)\sqrt{1 \vee \log(J)}}, \quad (5)$$

using the notation $x \vee y = \min(x, y)$ for any $x, y \in \mathbb{R}$.

Remark 3. The results of this work also hold up to constants if there is an additional constant in the assumption on L , i.e. $L(m) = cm^s$ for some $c > 0$. Note that combining the stability condition (5) and Assumption 1 (i) results in $m^{s \vee 1} \leq n\sqrt{J}$. In addition, we point out that in contrast to the model without repeated measurements, discussed in [Comte and Genon-Catalot \[2026\]](#), the stability condition is posed on the covariance Ψ_m instead of $\Psi_m^{(R)}$. The next result states that the stability condition for $\Psi_m^{(R)}$ follows.

Proposition 2. *Under Assumption 1, we have that for $n \geq 3$*

$$L(m)(\|(\Psi_m^{(R)})^{-1}\|_{\text{op}} \vee 1) \leq c^{**} \frac{n\sqrt{J}}{\log(n)\sqrt{1 \vee \log(J)}}, \quad c^{**} = 2(1 - \rho)c^*. \quad (6)$$

We refer to section 6.3 for the proof. To control the theoretical norm of the estimator, we need to impose the stability on the estimator. Thus, the corresponding empirical version of the stability condition (6) is given by

$$\Lambda_m := \left\{ L(m)(\|(\widehat{\Psi}_m^{(R)})^{-1}\|_{\text{op}} \vee 1) \leq 4c^{**} \frac{n\sqrt{J}}{\log(n)\sqrt{1 \vee \log(J)}} \right\}.$$

Let us define now the restricted estimator

$$\tilde{b}_m := \mathbf{1}_{\Lambda_m} \hat{b}_m.$$

Further, we define the following constants. For $p \geq 1$ and $r \geq 2$ and $c_0 = 18.6$ define

$$\mathbf{c}_r := 6r^2 c_0 (1 + \theta^2) 8\sqrt{6} \quad \text{and} \quad \mathbf{c}_p^* := \frac{1 - 3 \log(3/2)}{8(p+1)}. \quad (7)$$

In the results of this work, these constants restrict the constant c^* of stability condition (5) for different choices of p and r , see Remark 4 and Lemma 4 below for more details. Then, the following result holds for the risk bound in term of the expectation of the theoretical norm.

Theorem 1. *Let Assumption 1 be satisfied. For $\mathbf{c}_r$ and $\mathbf{c}_p^*$ defined in (7) with $p = 4$ and $r = 4 + 2 \min(1/s, 1)$, assume that for the constant c^* of stability condition (5) holds*

$$c^* \leq \frac{1}{8\mathbf{c}_r} \wedge \frac{\mathbf{c}_p^*}{3}. \quad (8)$$

we have for $n \geq 192/\rho$, $J \leq n(n-1)/\log^2(n)$ and $m \leq nJ$ that

$$\mathbb{E}[\|b - \tilde{b}_m\|_R^2] \leq C_1 \left(\inf_{t \in S_m} \|b - t\|_R^2 + \sigma^2 \frac{m}{nJ} \right) + \frac{C_2}{nJ}$$

for C_1 a numerical constant ($C_1 = 34$ suits) and C_2 is a constant depending on $\mathbf{c}_r$, $\mathbf{c}_p^*$, ρ , $\|b\|_R^2$, $\mathbb{E}[Y_{1,1}^4]$, but not on m, n, J .

Let us mention that the condition $J \leq n(n-1)/\log^2(n)$ is analogous to the constraint $\sqrt{J}/n \rightarrow 0$ set in Genon-Catalot and Larédo [2026] for asymptotic normality results.

Remark 4. *Note that depending on the choices of parameters p and r , the condition (8) on the constant c^* changes. More precisely, the larger p and r , the smaller c^* , leading to a more restrictive stability condition (5). These constants appear due to the application of Lemma 4, see below, to show the result for Theorem 1 with $p = 4$ and $r = 4 + 2 \min(1/s, 1)$. For the adaptive result Theorem 2 in the next section, we apply the same Lemma for $p = 5$ and $r = 5 + 2 \min(1/s, 1)$, leading to a change in constants. We also note that, in comparison to Comte and Genon-Catalot [2026], there is a change in the constant c_0 because we applied the deviation inequality for U -statistics of Giné and Nickl [2016] instead of the one of Houdré and Reynaud-Bouret [2003]. This is due to the fact that the variables are now J -dimensional vectors instead of real random variables for $J = 1$ in Comte and Genon-Catalot [2026].*

We provide a proof of Theorem 1 showing the main steps of the proof which relies on Lemma 4; this Lemma is rather difficult to demonstrate and its proof is referred to Section 6.

Proof of Theorem 1. The first main idea for the proof of Theorem 1 is to decompose the risk on the set Ω_m on which the empirical norm and the theoretical norm are equivalent, i.e.

$$\Omega_m := \left\{ \forall t \in S_m, t \neq 0, \left| \frac{\|t\|_{n,J,R}^2}{\|t\|_R^2} - 1 \right| \leq \frac{3}{4} \right\} = \left\{ \left\| (\Psi_m^{(R)})^{-1/2} \widehat{\Psi}_m^{(R)} (\Psi_m^{(R)})^{-1/2} - \text{Id}_m \right\|_{\text{op}} \leq \frac{3}{4} \right\}.$$

and its complement Ω_m^c . The main task, here is to bound the probability of the complement, and therefore, prove the following Lemma.

Lemma 4. *Let Assumption 1 be satisfied. For $\mathbf{c}_r$ and $\mathbf{c}_p^*$ defined in (7) with arbitrary $p \geq 1$ and $r \geq 2$, assume that for the constant c^* of stability condition (5) holds*

$$c^* \leq \frac{1}{8\mathbf{c}_r} \wedge \frac{\mathbf{c}_p^*}{3}.$$

Then, we have for $n \geq 192/\rho$ and $J \leq n(n-1)/\log^2(n)$ that

$$\mathbb{P}(\Omega_m^c) \leq 12(nJ)^{-\min(p, r-2\min(1/s, 1))}.$$

For the other points of the proof, we follow the line of the proof of Theorem 2 in [Comte and Genon-Catalot \[2026\]](#), which is standard. More precisely, we get

$$\mathbb{E} \left[\|\tilde{b}_m - b\|_R^2 \right] \leq 34 \inf_{t \in S_m} \|t - b\|_R^2 + 16\sigma^2 \frac{m}{nJ} + 4\mathbb{E}[\|\hat{b}_m\|_R^2 \mathbf{1}_{\Lambda_m} \mathbf{1}_{\Omega_m^c}] + \|b\|_R^2 (4\mathbb{P}(\Omega_m^c) + \mathbb{P}(\Lambda_m^c)). \quad (9)$$

To control the third summand, we first see that

$$\|\hat{b}_m\|_R^2 = \frac{1}{(nJ)^2} \mathbf{Y}_{\bullet, \bullet}^\top \mathbf{R}^{-1} \hat{\Phi}_m (\hat{\Psi}_m^{(R)})^{-1} \Psi_m^{(R)} (\hat{\Psi}_m^{(R)})^{-1} \hat{\Phi}_m^\top \mathbf{R}^{-1} \mathbf{Y}_{\bullet, \bullet}.$$

Next, we use that on Λ_m , we have

$$\|(\hat{\Psi}_m^{(R)})^{-1}\|_{\text{op}}^2 \leq 16(c^{**})^2 \left(\frac{n\sqrt{J}}{L(m) \log(n) \sqrt{1 \vee \log(J)}} \right)^2,$$

and that $\|\mathbf{R}^{-1}\|_{\text{op}} \leq 1/(1-\rho)$, $\|\Psi_m^{(R)}\|_{\text{op}} \leq L(m)/(1-\rho)$ and $\|\hat{\Phi}_m^\top \hat{\Phi}_m\|_{\text{op}} \leq nJL(m)$. We get

$$\|\hat{b}_m\|_R^2 \leq 64 \frac{(c^*)^2}{(1-\rho)^3 \log^2(n)(1 \vee \log(J))} \|\mathbf{Y}_{\bullet, \bullet}\|^2.$$

Thus

$$\begin{aligned} \mathbb{E}[\|\hat{b}_m\|_R^2 \mathbf{1}_{\Lambda_m} \mathbf{1}_{\Omega_m^c}] &\leq 64 \frac{(c^*)^2}{(1-\rho)^3 \log^2(n)(1 \vee \log(J))} \mathbb{E}[\|\mathbf{Y}_{\bullet, \bullet}\|^2 \mathbf{1}_{\Omega_m^c}] \\ &\leq 64 \frac{(c^*)^2 \mathbb{E}^{1/2}[Y_{1,1}^4]}{1-\rho} \frac{nJ}{\log^2(n)(1 \vee \log(J))} \mathbb{P}^{1/2}(\Omega_m^c). \end{aligned}$$

Further, we have $\Lambda_m^c \subseteq \Omega_m^c$ analogously to Proposition 4 in [Comte and Genon-Catalot \[2026\]](#). Then, we apply Lemma 4 with $r = 4 + 2\min(1/s, 1)$ and $p = 4$ to obtain

$$\mathbb{P}^{1/2}(\Omega_m^c) \leq \frac{\sqrt{12}}{(nJ)^2}.$$

Plugging this into (9), the result follows, i.e.

$$\mathbb{E} \left[\|\tilde{b}_m - b\|_R^2 \right] \leq 34 \left(\inf_{t \in S_m} \|t - b\|_R^2 + \sigma^2 \frac{m}{nJ} \right) + \sqrt{12} \left(256 \frac{(c^*)^2 \mathbb{E}^{1/2}[Y_{1,1}^4]}{1-\rho} + 8\|b\|_R^2 \right) \frac{1}{nJ}.$$

□

4 Adaptive Result

The goal now is to propose an adaptive estimator and provide upper bounds on the corresponding quadratic risk. The result is new, even for $J = 1$ and requires to solve the dependency problem between the $u_{i,j}$'s when bounding the deviations of a key empirical process. For this, we define the collection of models

$$\widehat{\mathcal{M}}_{n,J} := \left\{ m \in \mathbb{N}, m \leq nJ : L(m) (\|(\hat{\Psi}_m^{(R)})^{-1}\|_{\text{op}} \vee 1) \leq \frac{4c^{**}n\sqrt{J}}{\log(n)\sqrt{\log(J) \vee 1}} \right\},$$

and we set

$$\hat{m} := \arg \min_{m \in \widehat{\mathcal{M}}_{n,J}} \left\{ -\|\widehat{b}_m\|_{n,J,R}^2 + \text{pen}(m) \right\}, \quad \text{pen}(m) = \kappa \sigma^2 \frac{m}{nJ},$$

where κ is a numerical constant. A theoretical counterpart for $\widehat{\mathcal{M}}_{n,J}$ is defined by

$$\mathcal{M}_{n,J} := \left\{ m \in \mathbb{N}, m \leq nJ : L(m)(\|(\Psi_m^{(R)})^{-1}\|_{\text{op}} \vee 1) \leq \frac{c^{**}n\sqrt{J}}{\log(n)\sqrt{\log(J)} \vee 1} \right\}.$$

We introduce the following “light tail” assumption on the noise $\varepsilon_{1,1}$.

Assumption 2. *There exist $c_1, c_2 > 0$, such that for all $K > 0$ it holds that $\mathbb{P}(|\varepsilon_{1,1}| > K) \leq c_1 e^{-c_2 K}$.*

Remark 5. *We may have considered the following sub-Gaussian assumption (see [Baraud et al. \[2001a\]](#)):*

$$\forall u \in \mathbb{R}, \mathbb{E}(e^{u\varepsilon_{1,1}}) \leq e^{u^2 s^2/2}. \quad (10)$$

This assumption is satisfied for $\varepsilon_{1,1}$ bounded by B (with $s = B$) and Gaussian $\varepsilon_{1,1}$ with $s^2 = \text{Var}(\varepsilon_{1,1})$. It is worth noting that (10) implies

$$\mathbb{P}(|\varepsilon_{1,1}| > K) \leq 2e^{-K^2/(2s^2)}. \quad (11)$$

See Section 6.7 for the proof of (11). Thus, Assumption 2 is weaker and still allows us to capture bounded and Gaussian variables, but also exponential and Gamma mixtures. However, $\log^3(nJ)$ in the last term of the Inequality of Theorem 2 under Assumption 2 would only be $\log^{3/2}(nJ)$ under (10).

Under this additional assumption, we prove the following result for the proposed adaptive estimator of the regression function in empirical norm.

Theorem 2. *Assume that the Assumptions of Theorem 1 are fulfilled with $p = 5$ and $r = 5 + 2\min(1/s, 1)$. Let Assumption 2 be satisfied and $\mathbb{E}[b^4(X_{1,1})] < +\infty$. Then, for $n \geq n_0 := 192/\rho$, there exists $\kappa_0 > 0$ such that for any $\kappa \geq \kappa_0$,*

$$\mathbb{E}(\|\widehat{b}_{\widehat{m}} - b\|_{n,J,R}^2) \leq C \inf_{m \in \mathcal{M}_{n,J}} \left\{ \inf_{t \in S_m} \|t - b\|_R^2 + \kappa \sigma^2 \frac{m}{nJ} \right\} + C' \frac{\log^3(nJ)}{nJ}.$$

where C is a numerical constant ($C = 8$ suits) and C' a constant depending on the moments of $b(X_{1,1})$ and $\varepsilon_{1,1}$. The value $\kappa_0 = 64$ suits.

The result of Theorem 2 states that the estimator $\widehat{b}_{\widehat{m}}$ automatically reaches the squared bias/variance compromise in the collection $\mathcal{M}_{n,J}$, up to the multiplicative constant C (the closer to 1 the better) and the negligible additive residuals of order $\log^3(nJ)/(nJ)$. In practice, the implementation of the procedure requires a fixed value for κ , which is done from preliminary simulation experiments. This question appears in the theory of model selection by penalization since the beginning, see [Barron et al. \[1999\]](#), and concrete methods for doing this are given in [Baudry et al. \[2012\]](#).

We give the main arguments of the proof of Theorem 2, which is quite complex and we postpone the proofs of auxiliary results to Section 6.6.

Proof of Theorem 2. Let us define

$$\mathcal{M}_{n,J}^+ := \left\{ m \in \mathbb{N}, m \leq nJ, L(m)(\|(\Psi_m^{(R)})^{-1}\|_{\text{op}} \vee 1) \leq \frac{7c^{**}n\sqrt{J}}{\log(n)\sqrt{\log(J)} \vee 1} \right\}$$

and

$$\Omega_{n,J} := \bigcap_{m \in \mathcal{M}_{n,J}^+} \Omega_m. \quad (12)$$

We now split the risk on the set $\Omega_{n,J}$ to get

$$\mathbb{E}[\|\widehat{b}_{\widehat{m}} - b\|_{n,J,R}^2] = \mathbb{E}[\|\widehat{b}_{\widehat{m}} - b\|_{n,J,R}^2 \mathbf{1}_{\Omega_{n,J}}] + \mathbb{E}[\|\widehat{b}_{\widehat{m}} - b\|_{n,J,R}^2 \mathbf{1}_{\Omega_{n,J}^c}]. \quad (13)$$

To bound the risk of the estimator on $\Omega_{n,J}$, we also define the following set

$$\Xi_{n,J} = \{\mathcal{M}_{n,J} \subset \widehat{\mathcal{M}}_{n,J} \subset \mathcal{M}_{n,J}^+\}. \quad (14)$$

Lemma 5. *Under stability condition (5) it holds for $\Omega_{n,J}$ and $\Xi_{n,J}$ defined in (12) and (14), respectively, that*

$$\Omega_{n,J} \subset \Xi_{n,J}.$$

For the contrast, it holds

$$\gamma_{n,J}(t) = \|b - t\|_{n,J,R}^2 - \|b\|_{n,J,R}^2 - 2\nu_{n,J}(t), \quad \nu_{n,J}(t) = \frac{1}{nJ} \mathbf{u}_{\bullet,\bullet}^\top \mathbf{R}^{-1} t(\mathbf{X}_{\bullet,\bullet}).$$

Therefore, writing that for all $m \in \widehat{\mathcal{M}}_{n,J}$ and $t \in S_m$

$$\gamma_{n,J}(\widehat{b}_{\widehat{m}}) + \text{pen}(\widehat{m}) \leq \gamma_{n,J}(t) + \text{pen}(m)$$

implies

$$\|\widehat{b}_{\widehat{m}} - b\|_{n,J,R}^2 \leq \|t - b\|_{n,J,R}^2 + \text{pen}(m) + 2\nu_{n,J}(\widehat{b}_{\widehat{m}} - t) - \text{pen}(\widehat{m}).$$

Now, we work on $\Omega_{n,J}$, on which it holds by Lemma 5 that $\mathcal{M}_{n,J} \subset \widehat{\mathcal{M}}_{n,J}$. Consequently, the inequality holds for any $m \in \mathcal{M}_{n,J}$, and we choose $t = \widehat{b}_m$. We obtain, on $\Omega_{n,J}$ and for all $m \in \mathcal{M}_{n,J}$

$$\begin{aligned} \|\widehat{b}_{\widehat{m}} - b\|_{n,J,R}^2 &\leq \|\widehat{b}_m - b\|_{n,J,R}^2 + \text{pen}(m) + 2\nu_{n,J}(\widehat{b}_{\widehat{m}} - \widehat{b}_m) - \text{pen}(\widehat{m}) \\ &\leq \|\widehat{b}_m - b\|_{n,J,R}^2 + \text{pen}(m) + \frac{1}{4} \|\widehat{b}_{\widehat{m}} - \widehat{b}_m\|_{n,J,R}^2 \\ &\quad + 4 \left(\sup_{t \in S_m + S_{\widehat{m}}, \|t\|_{n,J,R}=1} \nu_{n,J}^2(t) - p(m, \widehat{m}) \right) + 4p(m, \widehat{m}) - \text{pen}(\widehat{m}) \end{aligned} \quad (15)$$

with

$$p(m, m') := 8\sigma^2 \frac{m \vee m'}{nJ}. \quad (16)$$

Next, note that for $\kappa \geq 32$ it holds that $4p(m, \widehat{m}) \leq \text{pen}(m) + \text{pen}(\widehat{m})$. Further, we have that

$$\frac{1}{4} \|\widehat{b}_{\widehat{m}} - \widehat{b}_m\|_{n,J,R}^2 \leq \frac{1}{2} \|\widehat{b}_{\widehat{m}} - b\|_{n,J,R}^2 + \frac{1}{2} \|b - \widehat{b}_m\|_{n,J,R}^2.$$

Plugging both into (15) and reorganizing the terms yields for $m \in \mathcal{M}_{n,J}$ on $\Omega_{n,J}$ that

$$\|\widehat{b}_{\widehat{m}} - b\|_{n,J,R}^2 \leq 3\|\widehat{b}_m - b\|_{n,J,R}^2 + 4\text{pen}(m) + 8 \left(\sup_{t \in S_m + S_{\widehat{m}}, \|t\|_{n,J,R}=1} \nu_{n,J}^2(t) - p(m, \widehat{m}) \right) +.$$

Plugging this in turn into (13), we obtain for any $m \in \mathcal{M}_{n,J}$

$$\begin{aligned} \mathbb{E}[\|\widehat{b}_{\widehat{m}} - b\|_{n,J,R}^2] &\leq \mathbb{E}[\|\widehat{b}_{\widehat{m}} - b\|_{n,J,R}^2 \mathbf{1}_{\Omega_{n,J}^c}] + 3\mathbb{E}[\|\widehat{b}_m - b\|_{n,J,R}^2] + 4\text{pen}(m) \\ &\quad + 8\mathbb{E} \left[\left(\sup_{t \in S_m + S_{\widehat{m}}, \|t\|_{n,J,R}=1} \nu_{n,J}^2(t) - p(m, \widehat{m}) \right) \mathbf{1}_{\Omega_{n,J}} \right]. \end{aligned} \quad (17)$$

We obtain the following two Lemmas to control the first and the last summand, see Section 6.6 for the proofs.

Lemma 6. *Let Assumption 1 be satisfied. For $\mathbf{c}_r$ and $\mathbf{c}_p^*$ defined in (7) with $p = 5$ and $r = 5 + 2\min(1/s, 1)$, assume that for the constant c^* of stability condition (5) holds $c^* \leq 1/(8\mathbf{c}_r) \wedge \mathbf{c}_p^*/3$. Further, assume that $\mathbb{E}[b^4(X_{1,1})], \mathbb{E}[Y_{1,1}^4] < +\infty$. Then, for $n \geq 192/\rho$ and $J \leq n(n-1)/\log^2(n)$, we have*

$$\mathbb{E}[\|\widehat{b}_{\widehat{m}} - b\|_{n,J,R}^2 \mathbf{1}_{\Omega_{n,J}^c}] \leq \frac{C}{nJ}$$

for a constant C depending on $\rho, c^{**}, s, \mathbb{E}[Y_{1,1}^4]$ and $\mathbb{E}[b^4(X_{1,1})]$.

Lemma 7. Under Assumption 2, for $p(m, m')$ defined by (16) and any $m \in \mathcal{M}_{n,J}$ it holds that

$$\mathbb{E} \left[\left(\sup_{t \in S_m + S_{\widehat{m}}, \|t\|_{n,J,R}=1} \nu_{n,J}^2(t) - p(m, \widehat{m}) \right)_+ \mathbf{1}_{\Omega_{n,J}} \right] \leq C \frac{\log^3(nJ)}{nJ}$$

for a constant C depending on c_1, c_2 (the constants of Assumption 2), c^{**} , ρ and $\mathbb{E}(Y_{1,1}^4)$.

Finally, applying Lemma 6 and 7 and Proposition 1 on (17) we obtain

$$\mathbb{E}[\|\widehat{b}_{\widehat{m}} - b\|_{n,J,R}^2] \leq C \inf_{m \in \mathcal{M}_{n,J}} \left\{ \inf_{t \in S_m} \|t - b\|_R^2 + \kappa \sigma^2 \frac{m}{nJ} \right\} + C' \frac{\log^2(nJ)}{nJ}$$

for $C = 8$ and C' a constant depending on the moments of $b(X_{1,1})$ and $\varepsilon_{1,1}$, which yields the result. $\square$

5 Simulation experiments

In this section, we illustrate the theoretical results of this work with some simulation experiments. For this, we consider two bases, which are complete orthonormal systems, containing no constant function (see Section 2.3 for the connection to identifiability):

- A Hermite basis, with support $\mathbb{R}$, $(h_j(x) = c_j H_j(x) e^{-x^2/2})_{j \geq 0}$ where H_j is the j^{th} Hermite polynomial as in Comte and Genon-Catalot [2026], see therein.
- A Tchebychev basis $(u_j(x))_{j \geq 0}$, defined below, with compact support $[-1, 1]$, or $(v_j(x))_{j \geq 0}$ with compact support $[\mathbf{a}_1, \mathbf{a}_2]$, $\mathbf{a}_1 < \mathbf{a}_2$, $\mathbf{a}_1, \mathbf{a}_2 \in \mathbb{R}$.

Tchebychev polynomials of the first kind are recursively defined for $x \in \mathbb{R}$ by

$$U_0(x) = 1, U_1(x) = 2x, \quad U_n(x) = 2xU_{n-1}(x) - U_{n-2}(x), n \geq 2.$$

They are orthogonal with respect to the scalar product in $\mathbb{L}^2([-1, 1], (1 - x^2)^{1/2} dx)$ and have squared $\mathbb{L}^2$ -norm in this context equal to $\pi/2$. So, we can consider the orthonormal $\mathbb{L}^2([-1, 1], dx)$ -basis:

$$u_j(x) = \sqrt{\frac{2}{\pi}} U_j(x) (1 - x^2)^{1/4} \mathbf{1}_{[-1, 1]}(x).$$

The polynomials are such that $\sup_{x \in [-1, 1]} U_j(x) = j + 1$, with $U_j(1) = j + 1$. So, we get

$$\sup_{x \in [-1, 1]} \sum_{j=0}^{m-1} u_j^2(x) \leq \frac{2}{\pi} \sum_{j=0}^{m-1} (j + 1)^2 = m(m + 1)(2m + 1)/(3\pi) := L(m).$$

We also set on $[\mathbf{a}_1, \mathbf{a}_2]$, $\mathbf{a}_1 < \mathbf{a}_2$,

$$v_j(x) = \frac{2}{\mathbf{a}_2 - \mathbf{a}_1} u_j \left(\frac{2x - (\mathbf{a}_1 + \mathbf{a}_2)}{\mathbf{a}_2 - \mathbf{a}_1} \right) \mathbf{1}_{[\mathbf{a}_1, \mathbf{a}_2]}(x).$$

For the regression function b , we consider the examples of Comte and Genon-Catalot [2026], four examples of odd functions

$$b_1(x) = 2x, \quad b_2(x) = 3 \sin(0.8 \pi x), \quad b_3(x) = \frac{4x}{1 + x^2}, \quad b_4(x) = 1.75x^3 \exp(-0.5|x|) \quad (18)$$

and also some even ones

$$b_5(x) = 3.25|x|, \quad b_6(x) = 4x^2 \exp(-x^2/4), \quad b_7(x) = 5.5 \sin^2(2x), \quad (19)$$

where the multiplicative constants are chosen so that for X following a $\mathcal{N}(0, 1)$ distribution, the signal to noise ratio (snr) is of approximately 2. The snr is measured as the empirical standard deviation of $b_k(X)$ divided by the standard deviation of the noise. We consider the two cases that the $X_{i,j}$'s follow

		Hermite basis								Tchebychev basis							
		$n = 250$				$n = 500$				$n = 250$				$n = 500$			
		$J = 20$		$J = 40$		$J = 20$		$J = 40$		$J = 20$		$J = 40$		$J = 20$		$J = 40$	
		Est	Or	Est	Or	Est	Or	Est	Or	Est	Or	Est	Or	Est	Or	Est	Or
b_1	MSE	1.96	1.86	0.99	0.92	1.95	1.88	0.93	0.89	2.77	2.31	1.48	1.2	2.40	2.18	1.21	1.08
	std	2.37	2.34	1.03	1.00	2.38	2.33	1.23	1.21	2.37	2.32	1.07	1.02	2.36	2.34	1.21	1.20
	dim	13.4	13.6	14.1	14.3	14.5	15.0	15.1	15.8	15.3	15.4	15.1	18.0	19.0	19.4	19.4	22.1
b_2	MSE	3.24	3.02	1.89	1.76	3.04	2.93	1.58	1.52	3.80	3.40	2.14	1.99	3.32	3.12	1.72	1.63
	std	3.35	3.27	2.11	2.06	3.80	3.76	2.05	2.01	3.36	3.25	2.09	2.07	3.76	3.75	2.02	2.02
	dim	13.0	12.6	14.0	13.2	14.1	13.0	14.7	14.9	10.8	10.9	11.0	11.1	11.5	11.4	12.3	12.1
b_3	MSE	6.50	6.04	3.12	2.89	4.82	4.64	2.62	2.53	7.67	7.01	3.79	3.45	5.35	5.05	2.97	2.78
	std	8.10	7.98	3.64	2.58	6.83	6.73	3.42	3.38	8.11	8.07	3.59	3.56	6.61	6.58	3.33	3.33
	dim	10.4	9.93	11.2	10.8	11.2	12.2	11.6	13.7	12.1	11.9	13.0	1.1	12.9	13.1	14.3	14.9
b_4	MSE	8.19	7.54	4.51	4.20	9.70	9.35	4.47	4.31	10.2	8.76	5.81	5.00	11.1	10.2	5.31	4.84
	std	9.68	9.49	6.00	5.89	11.0	10.8	5.31	5.24	9.76	9.49	5.92	5.85	10.9	10.9	5.23	5.20
	dim	11.4	11.2	12.3	12.1	12.3	12.7	13.0	13.7	10.8	10.0	11.4	11.8	12.9	11.7	15.0	14.8
b_5	MSE	1.04	0.99	0.72	0.68	0.93	0.84	0.52	0.46	1.14	1.00	0.73	0.56	0.92	0.87	0.53	0.49
	std	0.79	0.78	0.45	0.43	0.93	0.96	0.40	0.39	0.80	0.78	0.44	0.42	0.94	0.93	0.41	0.40
	dim	12.8	12.6	13.2	12.6	14.7	17.5	15.3	18.7	18.2	21.0	16.5	23.3	22.9	24.2	20.2	28.6
b_6	MSE	0.90	0.83	0.48	0.43	0.85	0.80	0.37	0.35	1.18	1.01	0.64	0.54	1.03	0.93	0.48	0.43
	std	1.01	0.97	0.57	0.55	1.12	1.10	0.49	0.48	0.01	0.99	0.56	0.55	1.10	1.09	0.49	0.48
	dim	9.29	9.36	10.7	9.79	10.7	9.71	12.6	9.99	12.4	11.5	13.4	13.2	17.1	16.2	18.2	18.5
b_7	MSE	1.32	1.25	0.76	0.69	1.16	1.10	0.62	0.57	1.63	1.49	0.86	0.80	1.34	1.24	0.70	0.65
	std	1.51	1.46	0.74	0.73	1.42	1.39	0.75	0.73	1.50	1.48	0.73	0.71	1.35	1.34	0.73	0.73
	dim	13.5	14.4	13.9	14.7	14.2	16.2	15.0	17.2	14.9	16.4	15.3	17.0	16.7	17.2	17.6	19.5

Table 1: Results for Gaussian X with 400 repetitions, for $\rho = 0.25$, sample sizes $n = 250, 500$ and $J = 20, 40$, 'Est' the estimator, compared to the oracle ('Or'), for regression functions defined in (18)-(19), MSE is $1000 \times \text{Relative MISE}$, std is $1000 \times \text{std}$, and 'dim' is the mean of the selected dimensions.

		Hermite basis								Tchebychev basis							
		$n = 250$				$n = 500$				$n = 250$				$n = 500$			
		$J = 20$		$J = 40$		$J = 20$		$J = 40$		$J = 20$		$J = 40$		$J = 20$		$J = 40$	
		Est	Or	Est	Or	Est	Or	Est	Or	Est	Or	Est	Or	Est	Or	Est	Or
A	MSE	1.31	0.89	0.74	0.52	0.60	0.45	0.33	0.25	2.85	2.10	1.72	1.29	1.38	1.02	0.82	0.62
	std	0.79	0.53	0.46	0.26	0.33	0.24	0.18	0.13	1.61	1.05	0.85	0.69	0.73	0.55	0.35	0.28
	dim	9.49	9.53	10.6	10.6	10.7	11.5	11.3	12.8	11.5	11.4	12.3	12.9	12.2	12.6	13.3	13.8
B	MSE	1.89	1.52	1.18	0.97	0.86	0.74	0.60	0.52	3.20	2.57	1.99	1.63	1.54	1.21	0.98	0.79
	std	1.50	1.39	0.97	0.89	0.79	0.75	0.55	0.52	1.86	1.62	1.20	1.08	1.00	0.89	0.64	0.57
	dim	10.2	9.56	11.2	10.8	11.1	11.9	11.8	13.0	11.9	12.3	12.8	13.0	12.8	13.2	14.4	14.4

Table 2: Results for Gaussian X with 400 repetitions, for function b_4 in (18), $A = (\rho = 0, \hat{\mu} = 0)$, and $B = (\rho = 0.25, \hat{\mu} = \mu_0)$ and sample sizes $n = 250, 500$ and $J = 20, 40$, 'Est' the estimator, compared to the oracle ('Or'), MSE is $1000 \times \text{Relative MISE}$, std is $1000 \times \text{std}$, and 'dim' is the mean of the selected dimensions.

		Estimated level $\hat{\mu} = \bar{Y}$								True level μ_0							
		$n = 62$				$n = 125$				$n = 62$				$n = 125$			
		$J = 20$		$J = 40$		$J = 20$		$J = 40$		$J = 20$		$J = 40$		$J = 20$		$J = 40$	
		Est	Or	Est	Or	Est	Or	Est	Or	Est	Or	Est	Or	Est	Or	Est	Or
b_3	MSE	12.4	11.7	5.88	5.67	11.4	11.0	5.75	5.51	2.94	2.33	1.58	1.38	1.66	1.28	0.94	0.77
	std	13.6	13.5	6.77	6.75	14.4	14.3	6.52	6.50	1.25	1.06	0.68	0.63	0.65	0.54	0.33	0.29
	dim	10.9	10.4	10.3	11.3	13.7	11.9	13.6	14.3	11.0	10.3	10.4	11.2	13.7	11.8	13.5	14.0
b_4	MSE	20.6	19.4	10.8	10.4	20.6	19.8	10.1	9.83	4.08	2.89	2.18	1.73	2.34	1.61	1.27	1.00
	std	23.1	23.0	15.1	15.1	23.8	23.7	11.7	11.6	1.75	1.31	0.78	0.64	0.90	0.63	0.45	0.35
	dim	9.31	8.03	9.26	9.54	11.7	9.71	12.7	11.6	9.22	7.84	9.47	9.06	11.7	9.47	12.6	11.2
b_5	MSE	2.09	1.96	1.22	1.11	1.58	1.50	0.91	0.85	0.89	0.77	0.58	0.48	0.44	0.37	0.29	0.22
	std	1.95	1.94	1.05	1.04	1.96	1.95	0.96	0.95	0.41	0.39	0.19	0.18	0.17	0.17	0.09	0.08
	dim	11.9	13.3	10.8	14.1	14.8	18.3	13.9	19.8	11.9	13.3	10.8	14.3	14.9	18.3	13.9	20.1
b_6	MSE	2.26	2.11	1.23	1.17	2.03	1.94	1.16	1.12	0.74	0.60	0.43	0.37	0.38	0.30	0.21	0.18
	std	2.61	2.59	1.39	1.38	2.43	2.42	1.70	1.69	0.43	0.39	0.24	0.23	0.19	0.18	0.12	0.11
	dim	10.1	8.42	9.28	9.30	12.3	9.93	12.3	11.9	9.99	8.14	9.36	9.35	12.2	9.62	12.5	11.9

Table 3: Results for non centered uniform $X(2\sqrt{3}\mathcal{U}([0,1]) - 0.25)$ with 400 repetitions, for $\rho = 0.5$, sample sizes $n = 62, 125$ and $J = 20, 40$, 'Est' the estimator, compared to the oracle ('Or'), for functions defined in (18)-(19), MSE is $1000 \times \text{Relative MISE}$, std is $1000 \times \text{std}$, and 'dim' is the mean of the selected dimensions. Estimation with Tchebychev basis.

either a $\mathcal{N}(0, 1)$ distribution (which is symmetric, see Table 1) or a uniform density $2\sqrt{3}(\mathcal{U}([0, 1]) - 0.25)$ (see Table 3).

We implement the model selection procedure described in Section 4, and get a data-driven choice $\hat{m}$ of the dimension m performed among dimensions $m = 1, \dots, D_{\max}$, such that the constraint defined for Λ_m is fulfilled, that is,

$$\widehat{\mathcal{M}}_{n,J} = \left\{ m \in \{1, \dots, D_{\max}\}, L(m)(\|(\widehat{\Psi}_m^{(R)})^{-1}\|_{\text{op}} \vee 1) \leq \frac{c}{1 - \rho \log(n) \sqrt{1 \vee \log(J)}} \right\}, \quad c = 8 \times 64$$

with $L(m) = m(m+1)(2m+1)/(3\pi)$ for the Tchebychev basis and $L(m) = \sqrt{m}$ for the Hermite basis. Since the theoretical value of $D_{\max} = nJ$ is too large in practice, we take $D_{\max} = \lfloor \sqrt{n} \rfloor$ for Hermite basis and $D_{\max} = \lfloor 2\sqrt{n} \rfloor$ for the Tchebychev basis and we check that $D_{\max}$ is never selected, otherwise we increase it. Concretely,

$$\hat{m} = \arg \min_{m \in \widehat{\mathcal{M}}_{n,J}} \left\{ -\|\hat{b}_m\|_{n,J,R}^2 + \text{pen}(m) \right\}, \quad \text{pen}(m) = \kappa \sigma^2 \frac{m}{nJ},$$

with $\kappa = 2$ for both bases.

The use of the $n \times n$ matrix R^{-1} involves numerical difficulties which leads us to compute $\widehat{\Psi}_{m,j}^{(R)}$ with the formula relying of the fact that $R^{-1} = (\alpha_n - \beta_n)\text{Id}_n + \beta_n \mathbf{1}\mathbf{1}^\top$.

The estimators are computed at points $x_k = \mathbf{a}_1 + (\mathbf{a}_2 - \mathbf{a}_1)k/K$ for $k = 1, \dots, K$, which are regularly spaced on an interval corresponding to the 1% and 99% quantiles of the observations. In order to keep a common scale for all examples, we consider relative errors (MSE) of any estimator $\hat{b}$ of b as the mean over 400 repetitions of the ratios

$$\frac{\sum_{k=1}^K (\hat{b}(x_k) - b(x_k))^2}{\sum_{k=1}^K b^2(x_k)}.$$

Now, as explained in Section 2.3, there is a level part in the model which must be preliminary estimated, but with rate $1/J$ while the rest of the function can be estimated with a nonparametric rate associated to nJ . We experiment the procedure by applying the method to $Y - \hat{\mu}$ with $\hat{\mu} = (1/nJ) \sum_{i,j} Y_{i,j}$, leading to $\hat{b}_{\hat{m}}^*$ and take $\hat{b}_{\hat{m}} = \hat{b}_{\hat{m}}^* + \hat{\mu}$. The results of experiments done with the functions of (18)-(19) are given in Table 1. They clearly show that the improvements are obtained through increasing J , while increasing n does not significantly improve the estimation. For comparison, Table 2 shows the results for function b_4 in two cases: the i.i.d. usual noise case obtained by setting $\rho = 0$ and $\hat{\mu} = 0$,

and the same case as in Table 1 but with $\hat{\mu}$ computed as $(1/K) \sum_k b(x_k)$, that is, an approximation of

$$\mu_0 = \frac{1}{a_2 - a_1} \int_{a_1}^{a_2} b(x) dx,$$

the “good correction”, as explained in Section 2.3. Table 3 gives the results obtained with the same procedure as in Table 1 or with $\hat{\mu}$ replaced by μ_0 for more cases and smaller sample sizes. We also changed the value of ρ , and the distribution of X which is no longer symmetric. We can see several interesting things:

- From Table 2, we see that the use of μ_0 gives excellent results, rather near of the ones obtained in the independent case when $\rho = 0$; the improvement obtained when we compare the use of unknown μ_0 to the use of $\tilde{Y}$ is also really puzzling in Table 3.
- Table 3 shows that the use of μ_0 decreases the error when nJ increases, whether the increase is due to n or to J , and not only for the increase of J like we could see when using $\hat{\mu}$ in Tables 1 and 3.
- The selected dimensions are generally close to the one of the oracles, and therefore very relevantly chosen.
- The selected dimensions increase with nJ , which is expected, but rather moderately.

Globally, all the errors are small (the MSE are multiplied by 1000 in the tables), so that all the results are satisfactory, even for small sample sizes from the point of view of nonparametric estimation. This is why we used smaller values of n in Table 3 than in Tables 1 and 2. Note that from Table 1 to 3, we changed ρ and the distribution of X , but these terms do not have an important impact on the results. We do not give the results for all functions for conciseness, as they are all very similar.

6 Proofs

6.1 Proof of Lemma 1

Proof of Lemma 1. From Comte and Genon-Catalot [2026], Proposition 2, we have that, for all integers j , and vectors $\mathbf{a} \in \mathbb{R}^m$,

$$\lim_{n \rightarrow +\infty} \mathbf{a}^\top \hat{\Psi}_{m,j}^{(R)} \mathbf{a} = \mathbf{a}^\top \Psi_m \mathbf{a} \quad a.s.$$

Thus, the announced result follows by averaging over j since $\hat{\Psi}_m^{(R)} = (1/J) \sum_{j=1}^J \hat{\Psi}_{m,j}^{(R)}$. $\square$

6.2 Proof of Proposition 1

Proof of Proposition 1. To consider the quadratic risk of the estimator $\hat{b}_m$, first define

$$\mathbf{a}_m := (\hat{\Psi}_m^{(R)})^{-1} \frac{1}{nJ} \hat{\Phi}_m^\top \mathbf{R}^{-1} b(\mathbf{X}_{\bullet,\bullet}) \in \mathbb{R}^{m \times 1}.$$

Note that $\mathbf{a}_m = (a_1, \dots, a_m)$ is the orthogonal projection of $b(\mathbf{X}_{\bullet,\bullet})$ on the space spanned by $\varphi_k(\mathbf{X}_{\bullet,\bullet})$ with $k \in \{1, \dots, m\}$ with respect to the scalar product induced by $\frac{1}{nJ} \mathbf{R}^{-1}$. Then, $b_m := \sum_{k=1}^m a_k \varphi_k$ is the orthogonal projection of b on S_m with respect to the scalar product $\langle \cdot, \cdot \rangle_{n,J,R}$. We can use this to obtain

$$\begin{aligned} \|b - \hat{b}_m\|_{n,J,R}^2 &= \|b - b_m\|_{n,J,R}^2 + \|\hat{b}_m - b_m\|_{n,J,R}^2 \\ &= \inf_{t \in S_m} \|b - t\|_{n,J,R}^2 + \|\hat{b}_m - b_m\|_{n,J,R}^2. \end{aligned}$$

For the second summand we insert the definitions and get

$$\begin{aligned} \|\hat{b}_m - b_m\|_{n,J,R}^2 &= (\hat{\mathbf{a}}_m - \mathbf{a}_m)^\top \hat{\Psi}_m^{(R)} (\hat{\mathbf{a}}_m - \mathbf{a}_m) \\ &= (\mathbf{Y}_{\bullet,\bullet} - b(\mathbf{X}_{\bullet,\bullet}))^\top \mathbf{R}^{-1} \frac{1}{nJ} \hat{\Phi}_m (\hat{\Psi}_m^{(R)})^{-1} \hat{\Psi}_m^{(R)} (\hat{\Psi}_m^{(R)})^{-1} \frac{1}{nJ} \hat{\Phi}_m^\top \mathbf{R}^{-1} (\mathbf{Y}_{\bullet,\bullet} - b(\mathbf{X}_{\bullet,\bullet})) \\ &= \frac{1}{n^2 J^2} (\mathbf{Y}_{\bullet,\bullet} - b(\mathbf{X}_{\bullet,\bullet}))^\top \mathbf{R}^{-1} \hat{\Phi}_m (\hat{\Psi}_m^{(R)})^{-1} \hat{\Phi}_m^\top \mathbf{R}^{-1} (\mathbf{Y}_{\bullet,\bullet} - b(\mathbf{X}_{\bullet,\bullet})). \end{aligned}$$

It holds that $\mathbb{E}[\mathbf{u}_{\bullet,\bullet}\mathbf{u}_{\bullet,\bullet}^\top] = \text{var}(\mathbf{u}_{\bullet,\bullet}) = \sigma^2\mathbf{R}$. Finally, we obtain

$$\begin{aligned}\mathbb{E}[\|\widehat{b}_m - b_m\|_{n,J,R}^2] &= \frac{1}{n^2 J^2} \mathbb{E}[\text{Tr}(\mathbf{u}_{\bullet,\bullet}^\top \mathbf{R}^{-1} \widehat{\Phi}_m (\widehat{\Psi}_m^{(R)})^{-1} \widehat{\Phi}_m^\top \mathbf{R}^{-1} \mathbf{u}_{\bullet,\bullet})] \\ &= \frac{1}{nJ} \mathbb{E}[\text{Tr}(\mathbf{R}^{-1} \widehat{\Phi}_m (\widehat{\Phi}_m^\top \mathbf{R}^{-1} \widehat{\Phi}_m)^{-1} \widehat{\Phi}_m^\top \mathbf{R}^{-1} \mathbf{u}_{\bullet,\bullet} \mathbf{u}_{\bullet,\bullet}^\top)] \\ &= \frac{\sigma^2}{nJ} \mathbb{E}[\text{Tr}(\mathbf{R}^{-1} \widehat{\Phi}_m (\widehat{\Phi}_m^\top \mathbf{R}^{-1} \widehat{\Phi}_m)^{-1} \widehat{\Phi}_m^\top \mathbf{R}^{-1} \mathbf{R})] \\ &= \frac{\sigma^2}{nJ} \text{Tr}(\text{Id}_m) = \frac{\sigma^2 m}{nJ}\end{aligned}$$

where we used the independence of $\mathbf{u}_{\bullet,\bullet}$ and $\mathbf{X}_{\bullet,\bullet}$ and commuted matrices inside the trace. Combining all the calculations, it follows

$$\begin{aligned}\mathbb{E}[\|b - \widehat{b}_m\|_{n,J,R}^2] &= \mathbb{E}\left[\inf_{t \in S_m} \|b - t\|_{n,J,R}^2\right] + \frac{\sigma^2 m}{nJ} \\ &\leq \inf_{t \in S_m} \mathbb{E}[\|b - t\|_{n,J,R}^2] + \frac{\sigma^2 m}{nJ} \\ &= \inf_{t \in S_m} \|b - t\|_R^2 + \frac{\sigma^2 m}{nJ}.\end{aligned}$$

This concludes the proof. $\square$

6.3 Proof of Proposition 2

Proof of Proposition 2. Using the first part of Proposition 2 of [Comte and Genon-Catalot \[2026\]](#), we have that

$$\frac{1}{\alpha_n} \Psi_m^{(R)} = \Psi_m + d_n \Xi_m, \quad d_n = c_n / \alpha_n,$$

where $\Xi_m = u_m u_m^\top$ for $u_m := (\mathbb{E}[\varphi_k(X_{1,1})])_{1 \leq k \leq m}$. We apply Lemma 13 for $A = \Psi_m$, $v = v_m := d_n u_m$ and $u = u_m$. As obviously $d_n u_m^\top \Psi_m^{-1} u_m \geq 0$, it holds that $1 + d_n u_m^\top \Psi_m^{-1} u_m \neq 0$ and Lemma 13 yields that $\Psi_m^{(R)}$ is invertible with

$$\begin{aligned}\alpha_n (\Psi_m^{(R)})^{-1} &= \Psi_m^{-1} - \frac{\Psi_m^{-1} u_m v_m^\top \Psi_m^{-1}}{1 + v_m^\top \Psi_m^{-1} u_m} \\ &= \Psi_m^{-\frac{1}{2}} \left(\text{Id}_m - \frac{d_n \Psi_m^{-\frac{1}{2}} u_m u_m^\top \Psi_m^{-\frac{1}{2}}}{1 + d_n u_m^\top \Psi_m^{-1} u_m} \right) \Psi_m^{-\frac{1}{2}}.\end{aligned}$$

The matrix $\Psi_m^{-\frac{1}{2}} u_m u_m^\top \Psi_m^{-\frac{1}{2}}$ is symmetric, positive definite and has rank 1. So, its eigenvalues are 0 and

$$\text{Tr}(\Psi_m^{-\frac{1}{2}} u_m u_m^\top \Psi_m^{-\frac{1}{2}}) = \text{Tr}(u_m^\top \Psi_m^{-1} u_m) = u_m^\top \Psi_m^{-1} u_m.$$

For the operator norms it follows that

$$\begin{aligned}\alpha_n \|(\Psi_m^{(R)})^{-1}\|_{\text{op}} &\leq \|\Psi_m^{-1}\|_{\text{op}} \left\| \text{Id}_m - \frac{d_n (\Psi_m^{(R)})^{-\frac{1}{2}} u_m u_m^\top (\Psi_m^{(R)})^{-\frac{1}{2}}}{1 + d_n u_m^\top \Psi_m^{-1} u_m} \right\|_{\text{op}} \\ &= \|\Psi_m^{-1}\|_{\text{op}} \max \left(1, \left| 1 - \frac{d_n u_m^\top \Psi_m^{-1} u_m}{1 + d_n u_m^\top \Psi_m^{-1} u_m} \right| \right) \\ &= \|\Psi_m^{-1}\|_{\text{op}} \max \left(1, \frac{1}{1 + d_n u_m^\top \Psi_m^{-1} u_m} \right) = \|\Psi_m^{-1}\|_{\text{op}},\end{aligned}$$

as $d_n u_m^\top \Psi_m^{-1} u_m \geq 0$. We obtain

$$\|(\Psi_m^{(R)})^{-1}\|_{\text{op}} \leq \frac{1}{\alpha_n} \|\Psi_m^{-1}\|_{\text{op}}.$$

Combining this with Assumption 1 (ii) and as $1/\alpha_n \leq 2(1-\rho)$ for $n \geq 3$, this finally leads to

$$L(m)(\|(\Psi_m^{(R)})^{-1}\|_{\text{op}} \vee 1) \leq \frac{1}{\alpha_n} L(m)(\|\Psi_m^{-1}\|_{\text{op}} \vee 1) \leq 2(1-\rho)c^* \frac{n\sqrt{J}}{\log(n)\sqrt{1 \vee \log(J)}},$$

which ends the proof of Proposition 2. $\square$

6.4 Proof of Lemma 4

Proof of Lemma 4. The idea of the proof is to appropriately split the set Ω_m again. For this, let us define the following two sets. For $Z_{n,J}$ defined in (3) we set

$$\Omega_m^{(1)} := \left\{ \sup_{t \in S_m, \text{Var}(t(X_{1,1})) \neq 0} \left| \frac{Z_{n,J}(t)}{\text{Var}(t(X_{1,1}))} - 1 \right| \leq \frac{1}{2} \right\}$$

and for $U_{n,J}$ defined in (4)

$$\Omega_m^{(2)} := \left\{ \sup_{t \in S_m, \|t\| \neq 0} \left| \frac{U_{n,J}(t)}{\|t\|^2} \right| \leq \mathfrak{c}_r \frac{(m \vee L(m))\sqrt{\log(J) \vee 1} \log(n)}{n\sqrt{J}} \right\}.$$

We show that these sets are, under the given assumptions, indeed suitable choices to control $\mathbb{P}(\Omega_m^c)$ within two steps:

- First we prove that

$$\Omega_m^{(1)} \cap \Omega_m^{(2)} \subset \Omega_m. \quad (20)$$

- Next we bound $\mathbb{P}((\Omega_m^{(1)})^c)$ in Lemma 10 and $\mathbb{P}((\Omega_m^{(2)})^c)$ in Lemma 11.

For the first step, we define for $t \in S_m$

$$S_{n,J}(t) := \alpha_n Z_{n,J}(t) + c_n \mathbb{E}[t^2(X_{1,1})] + 2c_n \mathbb{E}[t(X_{1,1})] V_{n,J}(t) = \|t\|_{n,J,R}^2 - (n-1)\beta_n U_{n,J}(t),$$

implying, by Lemma 3, that

$$\|t\|_{n,J,R}^2 = S_{n,J}(t) + (n-1)\beta_n U_{n,J}(t). \quad (21)$$

Then the first step follows from the two following Lemmas.

Lemma 8. Assume that $n \geq 192/\rho$. Then, on $\Omega_m^{(1)}$, for all $t \in S_m$ it holds

$$\left(\frac{1}{2} - \frac{1}{8}\right) \|t\|_R^2 \leq S_{n,J}(t) \leq \left(\frac{3}{2} + \frac{1}{8}\right) \|t\|_R^2.$$

Lemma 9. Assume Condition (5) and $\mathfrak{c}_r c^* \leq 1/8$, then, on $\Omega_m^{(2)}$, for all $t \in S_m$ it holds

$$(n-1)\beta_n |U_{n,J}(t)| \leq \frac{1}{8} \|t\|_R^2.$$

Indeed, combining these two results with equality (21), one concludes that on $\Omega_m^{(1)} \cap \Omega_m^{(2)}$ it holds for $t \in S_m$ with $\|t\|^2 \neq 0$ that

$$\frac{1}{4} \|t\|_R^2 = \left(\frac{1}{2} - \frac{1}{8} - \frac{1}{8}\right) \|t\|_R^2 \leq \|t\|_{n,J,R}^2 \leq \left(\frac{3}{2} + \frac{1}{8} + \frac{1}{8}\right) \|t\|_R^2 = \frac{7}{4} \|t\|_R^2.$$

Next the two following Lemmas give the result of the second step.

Lemma 10. Let Assumption 1 be satisfied. For $p \geq 1$, assume that for the constant c^* of stability condition (5) it holds $c^* \leq \mathfrak{c}_p^*/3$ for $\mathfrak{c}_p^*$ defined in (7). Then, it holds under Assumption 1, for $n \geq 3$ and $J \leq n(n-1)/\log^2(n)$ that

$$\mathbb{P}((\Omega_m^{(1)})^c) \leq 2(nJ)^{-p}.$$

Lemma 11. *Under Assumption 1 it holds for $n \geq 3$, $J \leq n(n-1)/\log^2(n)$ and $r \geq 1$ that*

$$\mathbb{P}((\Omega_m^{(2)})^c) \leq 5.44 \frac{1}{(nJ)^{r-2\min(1/s,1)}}.$$

Finally, gathering (20), Lemma 10 and Lemma 11 gives the result, i.e.

$$\mathbb{P}(\Omega_m^c) \leq \mathbb{P}((\Omega_m^{(1)})^c) + \mathbb{P}((\Omega_m^{(2)})^c) \leq 12(nJ)^{-\min(p, r-2/s)}.$$

The proofs of Lemmas 8-11 are given in Section 6.5 below. □

6.5 Proofs of Lemmas 8-11

We use the convention $\|\Psi_m^{-1}\|_{\text{op}} = +\infty$ if Ψ_m is not invertible.

Proof of Lemma 8. Recall that we consider

$$S_{n,J}(t) = \alpha_n Z_{n,J}(t) + c_n \mathbb{E}[t^2(X_{1,1})] + 2c_n \mathbb{E}[t(X_{1,1})]V_{n,J}(t) = \|t\|_{n,J,R}^2 - (n-1)\beta_n U_{n,J}(t).$$

On $\Omega_m^{(1)}$, we have for all $t \in S_m$ that

$$\frac{1}{2}\text{Var}(t(X_{1,1}) \leq Z_{n,J}(t) \leq \frac{3}{2}\text{Var}(t(X_{1,1})).$$

Therefore, using

$$\frac{1}{2}c_n(\mathbb{E}[t(X_{1,1})])^2 \leq c_n(\mathbb{E}[t(X_{1,1})])^2 \leq \frac{3}{2}c_n(\mathbb{E}[t(X_{1,1})])^2$$

and $\alpha_n \text{Var}(t(X_{1,1})) + c_n(\mathbb{E}[t(X_{1,1})])^2 = \|t\|_R^2$, we get

$$\frac{1}{2}\|t\|_R^2 \leq \alpha_n Z_{n,J}(t) + c_n(\mathbb{E}[t(X_{1,1})])^2 \leq \frac{3}{2}\|t\|_R^2. \quad (22)$$

Now, it holds for all $j = 1, \dots, J$ that $V_{n,j}^2(t) \leq Z_{n,j}(t)$. Following the same proof as in Lemma 5 of Comte and Genon-Catalot [2026], this implies that on $\Omega_m^{(1)}$, for $n \geq 192/\rho$, it holds

$$2c_n |\mathbb{E}[t(X_{1,1})]V_{n,j}(t)| \leq \frac{\sqrt{3}}{n\rho} \|t\|_R^2 \leq \frac{1}{8} \|t\|_R^2.$$

Thus, averaging over j yields

$$2c_n |\mathbb{E}[t(X_{1,1})]V_{n,J}(t)| \leq \frac{1}{8} \|t\|_R^2.$$

As a consequence, this inequality combined with (22) imply

$$(\frac{1}{2} - \frac{1}{8})\|t\|_R^2 \leq S_{n,J}(t) \leq (\frac{3}{2} + \frac{1}{8})\|t\|_R^2,$$

which ends the proof. □

Proof of Lemma 9. On $\Omega_m^{(2)}$, for any $t \in S_m$,

$$|U_{n,J}(t)| \leq \mathfrak{c}_r \|\Psi_m^{-1}\|_{\text{op}} \frac{(m \vee L(m))\sqrt{\log(J)\log(n)}}{n\sqrt{J}} \text{Var}(t(X_{1,1})).$$

Under Condition (5), we get

$$|U_{n,J}(t)| \leq \mathfrak{c}_r c^* \text{Var}(t(X_{1,1})).$$

So, as $(n-1)\beta_n/\alpha_n \leq 1$ and $\alpha_n \text{Var}(t(X_1)) \leq \|t\|_R^2$, we get

$$(n-1)\beta_n |U_{n,J}(t)| \leq \mathfrak{c}_r c^* ((n-1)\beta_n/\alpha_n) \|t\|_R^2 \leq \mathfrak{c}_r c^* \|t\|_R^2 \leq \frac{1}{8} \|t\|_R^2,$$

as $\mathfrak{c}_r c^* \leq 1/8$, which ends the proof of Lemma 9. □

Lemma 10. The result is analogous to Lemma 3 in [Comte and Genon-Catalot \[2026\]](#). First, note that due to $J \leq n(n-1)/\log^2(n)$ it also holds that

$$\frac{\log(nJ)}{\log(n)\sqrt{\log(J) \vee 1}} \leq \frac{3\log(n)}{\log(n)\sqrt{\log(J) \vee 1}} \leq 3\log(n). \quad (23)$$

Combining this with the stability condition (5) it follows that

$$\begin{aligned} L(m)\|\Psi_m^{-1}\|_{\text{op}} &\leq c^* \frac{n\sqrt{J}}{\log(n)\sqrt{1 \vee \log(J)}} \leq \frac{\mathbf{c}_p^*}{3} \frac{nJ}{\log(nJ)} \frac{n\sqrt{J}\log(nJ)}{nJ\log(n)\sqrt{1 \vee \log(J)}} \\ &\leq \mathbf{c}_p^* \frac{nJ}{\log(nJ)}, \end{aligned} \quad (24)$$

where we recall $\Psi_m = (\text{cov}(\varphi_j(X_{1,1}), \varphi_k(X_{1,1}))_{1 \leq j, k \leq m}$. Here, $\widehat{\Psi}_m$ is defined as

$$\widehat{\Psi}_m := \left(\frac{1}{nJ} \sum_{j=1}^J \sum_{i=1}^n ((\varphi_{k_1}(X_{i,j}) - \mathbb{E}[\varphi_{k_1}(X_{1,1})])(\varphi_{k_2}(X_{i,j}) - \mathbb{E}[\varphi_{k_2}(X_{1,1})])) \right)_{1 \leq k_1, k_2 \leq m}$$

and we have a decomposition of the matrix of interest as a sum over i, j of

$$K_m(X_{i,j}) = \Psi_m^{-1/2} (((\varphi_{k_1}(X_{i,j}) - \mathbb{E}[\varphi_{k_1}(X_{1,1})])(\varphi_{k_2}(X_{i,j}) - \mathbb{E}[\varphi_{k_2}(X_{1,1})])))_{1 \leq k_1, k_2 \leq m} \Psi_m^{-1/2}.$$

So, the proof follows from Tropp Chernov inequality, see Lemma 12 in the appendix from [Tropp \[2012\]](#), as used in the proof of Lemma 3 in [Comte and Genon-Catalot \[2026\]](#), but involving nJ independent and identically distributed random matrices instead of n , under (24) implied by (5). $\square$

Proof of Lemma 11. This is result the generalization of Lemma 4 from [Comte and Genon-Catalot \[2026\]](#). First note that, with Cauchy Schwarz Inequality, for $t = \sum_{k=1}^m a_k \varphi_k$ and $\|t\|^2 = \sum_{k=1}^m a_k^2 = 1$, we get

$$U_{n,J}(t) = \sum_{1 \leq k, \ell \leq m} a_k a_\ell \left(\frac{1}{J} \sum_{j=1}^J U_{n,j}(\varphi_k, \varphi_\ell) \right) \leq \left(\sum_{1 \leq k, \ell \leq m} \left(\frac{1}{J} \sum_{j=1}^J U_{n,j}(\varphi_k, \varphi_\ell) \right)^2 \right)^{1/2}.$$

Note that using again (23) we have that

$$\begin{aligned} \mathbf{c}_r \frac{L(m)\sqrt{\log(J) \vee 1} \log(n)}{n\sqrt{J}} &\geq 3\sqrt{2}r^2 c_0 (1 + \theta^2) 8\sqrt{6} \frac{\sqrt{\log(J) \vee 1} \log(n) \sqrt{n(n-1)J}}{n\sqrt{J} \log(nJ)} \frac{L(m) \log(nJ)}{\sqrt{n(n-1)J}} \\ &\geq \sqrt{2}r^2 c_0 (1 + \theta^2) 8\sqrt{6} \frac{\sqrt{n(n-1)J}}{n\sqrt{J}} \frac{L(m) \log(nJ)}{\sqrt{n(n-1)J}} \\ &\geq r^2 c_0 (1 + \theta^2) 8\sqrt{6} \frac{L(m) \log(nJ)}{\sqrt{n(n-1)J}}. \end{aligned}$$

Consequently, we can write

$$\begin{aligned} \mathbb{P}((\Omega_m^{(2)})^c) &\leq \mathbb{P} \left(\sup_{t \in S_m, \|t\| \neq 0} \left| \frac{U_{n,J}(t)}{\|t\|^2} \right| > 8\sqrt{6}r^2 c_0 \log(nJ) \frac{L(m)(1 + \theta^2)}{\sqrt{Jn(n-1)}} \right) \\ &\leq \mathbb{P} \left(\left(\sum_{1 \leq p, q \leq m} U_{n,J}^2(\varphi_p, \varphi_q) \right)^{1/2} > 8\sqrt{6}c_0 r^2 \log(nJ) \frac{L(m)(1 + \theta^2)}{\sqrt{Jn(n-1)}} \right) \\ &= \mathbb{P} \left(\sum_{1 \leq p, q \leq m} U_{n,J}^2(\varphi_p, \varphi_q) > 2 \cdot 192c_0^2 r^4 \log^2(nJ) \frac{L^2(m)(1 + \theta^2)^2}{Jn(n-1)} \right). \end{aligned} \quad (25)$$

The idea is to apply the U-statistic concentration inequality in [Giné and Nickl \[2016\]](#) (Theorem 3.4.8 p. 183). For this, we consider for fixed $p, q \in \{1, \dots, m\}$

$$U_{J,n}(\varphi_p, \varphi_q) = \frac{1}{J} \sum_{j=1}^J U_{n,j}(\varphi_p, \varphi_q) = \frac{1}{n(n-1)} \sum_{i=2}^n \sum_{\ell=1}^{i-1} g_{p,q}(X_{i,\bullet}, X_{\ell,\bullet}),$$

with

$$g_{p,q}(X_{i,\bullet}, X_{\ell,\bullet}) := \frac{1}{J} \sum_{j=1}^J [\bar{\varphi}_p(X_{i,j}) \bar{\varphi}_q(X_{\ell,j}) + \bar{\varphi}_p(X_{\ell,j}) \bar{\varphi}_q(X_{i,j})] \quad \text{and} \quad \bar{\varphi}(x) = \varphi(x) - \int \varphi(y) f(y) dy.$$

Under Assumption 1 (i), the basis functions φ_p are bounded by θ . Consequently, we have that $U_{J,n}(\varphi_p, \varphi_q)$ is a canonical U-statistic with symmetric and bounded kernel $g_{p,q}$ with i.i.d. vector observations $(X_{i,\bullet})_{1 \leq i \leq n}$. We obtain the following quantities for [Giné and Nickl \[2016\]](#) (Theorem 3.4.8 p. 183):

$$\begin{aligned} A &= 8\theta^2, \quad C = D = 2\sqrt{\frac{n(n-1)}{J}} \sqrt{\int \varphi_p^2(x) f(x) dx \int \varphi_q^2(x) f(x) dx}, \\ B^2 &= 4\theta^2 \frac{(n-1)}{J} \left(\int \varphi_p^2(x) f(x) dx + \int \varphi_q^2(x) f(x) dx \right). \end{aligned}$$

For c_0 denoting the maximum of all the numerical constants in the deviation inequality, i.e. $c_0 = 18.6$ by choosing $\varepsilon = 2$, and setting

$$\delta_{p,q} := \frac{c_0}{n(n-1)} \left[C\sqrt{r \log(nJ)} + D(r \log(nJ)) + B(r \log(nJ))^{3/2} + A(r \log(nJ))^2 \right],$$

the deviation inequality yields for $u = r \log(nJ)$, $r \geq 1$, that

$$\mathbb{P}(|U_{J,n}(\varphi_p, \varphi_q)| \geq \delta_{p,q}) \leq 2e^{1-r \log(nJ)} \leq 5.44(nJ)^{-r}. \quad (26)$$

Since we assume $n \geq 3$ and $J \geq 1$ we have also that $nJ \geq e$ and, consequently,

$$\delta_{p,q} \leq c_0 r^2 \log(nJ) \left[4 \frac{\sqrt{\int \varphi_p^2 f} \sqrt{\int \varphi_q^2 f}}{\sqrt{n(n-1)J}} + 2\theta \sqrt{\log(nJ)} \frac{\sqrt{\int \varphi_p^2 f} + \sqrt{\int \varphi_q^2 f}}{\sqrt{Jn}\sqrt{n-1}} + \frac{8\theta^2 \log(nJ)}{n(n-1)} \right].$$

For $J \leq n(n-1)/\log^2(n)$, then $\log^2(nJ)/[n(n-1)] \leq 9/J$ and $\sqrt{\log(nJ)/n} \leq \sqrt{3} \leq 2$, thus

$$\delta_{p,q} \leq \frac{c_0 r^2 \log(nJ)}{\sqrt{n(n-1)J}} \left[4 \sqrt{\int \varphi_p^2 f} \sqrt{\int \varphi_q^2 f} + 4\theta \left(\sqrt{\int \varphi_p^2 f} + \sqrt{\int \varphi_q^2 f} \right) + 8\theta^2 \right]^2.$$

Using then that $\sum_{1 \leq p, q \leq m} (\sqrt{\int \varphi_p^2 f} + \sqrt{\int \varphi_q^2 f})^2 \leq 2mL(m) + 2mL(m) = 4L(m)m \leq 4(m \vee L(m))^2$ one obtains

$$\begin{aligned} \sum_{1 \leq p, q \leq m} \delta_{p,q}^2 &\leq \frac{c_0^2 r^4 \log^2(nJ)}{n(n-1)J} (48L^2(m) + 192\theta^2 mL(m) + 192m^2\theta^4) \\ &\leq \frac{c_0^2 r^4 \log^2(nJ)}{n(n-1)J} 2 \cdot 192(m \vee L(m))^2 (1 + \theta^2)^2. \end{aligned}$$

Inserting the last equation and (26) into (25), we obtain

$$\begin{aligned} \mathbb{P}((\Omega_m^{(2)})^c) &\leq \mathbb{P} \left(\sum_{1 \leq p, q \leq m} U_{n,J}^2(\varphi_p, \varphi_q) > \sum_{1 \leq p, q \leq m} \delta_{p,q}^2 \right) \\ &\leq \sum_{1 \leq p, q \leq m} \mathbb{P}(|U_{n,J}^2(\varphi_p, \varphi_q)| > \delta_{p,q}) \\ &\leq 5.44 \frac{m^2}{(nJ)^r} \leq 5.44 \frac{1}{(nJ)^{r-2 \min(1/s, 1)}}, \end{aligned}$$

where in the last inequality we used that $m^s \leq n\sqrt{J} \leq nJ$ if $s \geq 1$, and $m \leq nJ$ if $s < 1$, which yields $m \leq (nJ)^{\frac{1}{s} \vee 1}$. $\square$

6.6 Proofs of Lemmas 5 - 7

Proof of Lemma 5. We acknowledge Huang [2025] for a decisive improvement of the result obtained in this lemma. Consider $\omega \in \Omega_{n,J}$ and

$$\widehat{G}_m(\omega) := (\Psi_m^{(R)})^{-\frac{1}{2}} \widehat{\Psi}_m^{(R)}(\omega) (\Psi_m^{(R)})^{-\frac{1}{2}} ; \forall m \in \mathcal{M}_{n,J}^+.$$

For any $m \in \mathcal{M}_{n,J}^+$, since $\omega \in \Omega_m$,

$$\text{Sp}(\widehat{G}_m(\omega)) \subset \left[\frac{1}{4}, \frac{7}{4} \right], \quad \text{and then} \quad \text{Sp}(\widehat{G}_m^{-1}(\omega)) \subset \left[\frac{4}{7}, 4 \right].$$

Moreover,

$$(\widehat{\Psi}_m^{(R)})^{-1}(\omega) = (\Psi_m^{(R)})^{-\frac{1}{2}} \widehat{G}_m^{-1}(\omega) (\Psi_m^{(R)})^{-\frac{1}{2}}.$$

So, thanks to a well-known property of the Loewner order,

$$\frac{4}{7} \mathbf{x}^\top (\Psi_m^{(R)})^{-1} \mathbf{x} \leq \mathbf{x}^\top (\widehat{\Psi}_m^{(R)})^{-1}(\omega) \mathbf{x} \leq 4 \mathbf{x}^\top (\Psi_m^{(R)})^{-1} \mathbf{x} ; \forall \mathbf{x} \in \mathbb{R}^m,$$

leading to

$$\|(\Psi_m^{(R)})^{-1}\|_{\text{op}} \leq \frac{7}{4} \|(\widehat{\Psi}_m^{(R)})^{-1}(\omega)\|_{\text{op}} \quad \text{and} \quad \|(\widehat{\Psi}_m^{(R)})^{-1}(\omega)\|_{\text{op}} \leq 4 \|(\Psi_m^{(R)})^{-1}\|_{\text{op}}$$

almost surely. Therefore, $\omega \in \Xi_{n,J}$. □

Proof of Lemma 6. First we further decompose the risk in two terms, more precisely,

$$\mathbb{E}[\|\widehat{b}_{\widehat{m}} - b\|_{n,J,R}^2 \mathbf{1}_{\Omega_{n,J}^c}] \leq 2\mathbb{E}[\|\widehat{b}_{\widehat{m}}\|_{n,J,R}^2 \mathbf{1}_{\Omega_{n,J}^c}] + 2\mathbb{E}[\|b\|_{n,J,R}^2 \mathbf{1}_{\Omega_{n,J}^c}] := 2\mathbb{S}_1 + 2\mathbb{S}_2.$$

We first consider $\mathbb{S}_2$. By Lemma 2, it holds that $\|b\|_{n,J,R}^2 = \frac{1}{nJ} b(\mathbf{X}_{\bullet,\bullet})^\top \mathbf{R}^{-1} b(\mathbf{X}_{\bullet,\bullet}) \leq \frac{1}{1-\rho} \|b\|_{n,J}^2$. In addition, we get

$$\mathbb{E}(\|b\|_{n,J,R}^4) \leq \frac{1}{(1-\rho)^2} \mathbb{E} \left(\frac{1}{nJ} \sum_{j=1}^J \sum_{i=1}^n b^4(X_{i,j}) \right) = \frac{\mathbb{E}(b^4(X_{1,1}))}{(1-\rho)^2}.$$

Note that since by Assumption 1 (i) it holds $L(m) = m^s$, and consequently $\text{card}(\mathcal{M}_{n,J}^+) \leq (7c^{**}n\sqrt{J})^{1/s}$. Applying Lemma 4 with $p = 2 + 1/s$ and $r = 2 + 3/s$, we obtain

$$\mathbb{P}(\Omega_{n,J}^c) \leq 12 \text{card}(\mathcal{M}_{n,J}^+) (nJ)^{-\min(p,r-2/s)} \leq 12(7c^{**})^{1/s} (nJ)^{-2}. \quad (27)$$

It follows by Cauchy-Schwarz inequality that

$$2\mathbb{S}_2 \leq 2\sqrt{12}(7c^{**})^{1/(2s)} \frac{\mathbb{E}(b^4(X_{1,1}))}{(1-\rho)^2} \frac{1}{nJ}.$$

Let us now consider $\mathbb{S}_1$. First note that

$$\sup_{\mathbf{x} \in \mathbb{R}^{nJ}, \|\mathbf{x}\|_{\mathbb{R}^{nJ}}=1} \mathbf{x}^\top \widehat{\Phi}_{\widehat{m}} \widehat{\Phi}_{\widehat{m}}^\top \mathbf{x} = \sup_{\mathbf{x} \in \mathbb{R}^{nJ}, \|\mathbf{x}\|_{\mathbb{R}^{nJ}}=1} \sum_{k=1}^{\widehat{m}} \left(\sum_{j=1}^J \sum_{i=1}^n x_{i,j} \varphi_k(X_{i,j}) \right)^2 \leq (nJ)L(\widehat{m}),$$

where $\|\cdot\|_{\mathbb{R}^{nJ}}$ denotes the standard Euclidean norm on $\mathbb{R}^{nJ}$. Further, we have that $\widehat{m} \in \widehat{\mathcal{M}}_{n,J}$ and $L(\widehat{m}) \|(\widehat{\Psi}_{\widehat{m}}^{(R)})^{-1}\|_{\text{op}} \leq 4c^{**}n\sqrt{J}/(\log(n)\sqrt{\log(J)} \vee 1)$. With this we get

$$\begin{aligned} \mathbb{S}_1 &= \frac{1}{(nJ)^2} \mathbb{E} \left[\mathbf{Y}_{\bullet,\bullet}^\top \mathbf{R}^{-1} \widehat{\Phi}_{\widehat{m}} (\widehat{\Psi}_{\widehat{m}}^{(R)})^{-1} \widehat{\Phi}_{\widehat{m}}^\top \mathbf{R}^{-1} \mathbf{Y}_{\bullet,\bullet} \mathbf{1}_{\Omega_{n,J}^c} \right] \\ &\leq 4c^{**} \frac{1}{nJ} \frac{n\sqrt{J}}{\log(n)\sqrt{\log(J)} \vee 1} \mathbb{E}[\|\mathbf{R}^{-1} \mathbf{Y}_{\bullet,\bullet}\|_{\mathbb{R}^{nJ}}^2 \mathbf{1}_{\Omega_{n,J}^c}]. \end{aligned}$$

Now as $\|\mathbf{R}^{-1}\|_{\text{op}} \leq 1/(1-\rho)$ applying Cauchy-Schwarz inequality it holds that

$$\begin{aligned} \mathbb{E}[\|\mathbf{R}^{-1}\mathbf{Y}_{\bullet,\bullet}\|_{\mathbb{R}^{nJ}}^2 \mathbf{1}_{\Omega_{n,J}^c}] &\leq \frac{1}{(1-\rho)^2} \sum_{j=1}^J \sum_{i=1}^n \mathbb{E}[Y_{i,j}^2 \mathbf{1}_{\Omega_{n,J}^c}] \\ &\leq \frac{nJ}{(1-\rho)^2} \sqrt{\mathbb{E}[Y_{1,1}^4]} \sqrt{\mathbb{P}(\Omega_{n,J}^c)} \end{aligned}$$

Using again Lemma 4 with $p = 4 + 1/s$ and $r = 4 + 3/s$ analogously to (27), we obtain the upper bound $\mathbb{P}(\Omega_{n,J}^c) \leq 12(7c^{**})^{1/s} (nJ)^{-4}$ and, consequently,

$$\begin{aligned} \mathbb{S}_1 &\leq \frac{4c^{**}}{(1-\rho)^2} \frac{n\sqrt{J}}{\log(n)\sqrt{\log(J)} \vee 1} \sqrt{\mathbb{E}[Y_{1,1}^4]} \sqrt{\mathbb{P}(\Omega_{n,J}^c)} \\ &\leq 12(7c^{**})^{1/s} \frac{4c^{**}}{(1-\rho)^2} \sqrt{\mathbb{E}[Y_{1,1}^4]} \frac{1}{nJ}. \end{aligned}$$

This finally yields the result, that is,

$$\mathbb{E}[\|\widehat{b}_{\widehat{m}} - b\|_{n,J,R}^2 \mathbf{1}_{\Omega_{n,J}^c}] \leq \left(24(7c^{**})^{1/s} \frac{4c^{**}}{(1-\rho)^2} \sqrt{\mathbb{E}[Y_{1,1}^4]} + 2\sqrt{12}(7c^{**})^{1/(2s)} \frac{\mathbb{E}(b^4(X_{1,1}))}{(1-\rho)^2} \right) \frac{1}{nJ}.$$

□

Proof of Lemma 7. First, we define the $n \times (n+1)$ matrix M such that $\mathbf{u}_{\bullet,j} = M\varepsilon_{\bullet,j}$, where $\varepsilon_{\bullet,j} = (\varepsilon_{0,j}, \varepsilon_{1,j}, \dots, \varepsilon_{n,j})^\top$, more precisely,

$$M := \begin{pmatrix} \sqrt{\rho} & \sqrt{1-\rho} & 0 & 0 & \dots & 0 \\ \sqrt{\rho} & 0 & \sqrt{1-\rho} & 0 & \dots & 0 \\ \vdots & \ddots & & \ddots & \ddots & \vdots \\ \vdots & & \ddots & & \ddots & \vdots \\ \sqrt{\rho} & 0 & \dots & 0 & \sqrt{1-\rho} & 0 \\ \sqrt{\rho} & 0 & \dots & 0 & 0 & \sqrt{1-\rho} \end{pmatrix}.$$

The first column is a n -dimensional vector with coordinates all equal to $\sqrt{\rho}$, concatenated with $\sqrt{1-\rho}\text{Id}_n$. Moreover, we set $\mathbf{M} = \text{diag}(M, \dots, M) \in \mathbb{R}^{nJ \times (n+1)J}$. Then, we can write $\mathbf{u}_{\bullet,\bullet} = \mathbf{M}\varepsilon_{\bullet,\bullet}$, and

$$\nu_{n,J}(t) = \frac{1}{nJ} \mathbf{u}_{\bullet,\bullet}^\top \mathbf{R}^{-1} t(\mathbf{X}_{\bullet,\bullet}) = \frac{1}{nJ} \varepsilon_{\bullet,\bullet}^\top \mathbf{M}^\top \mathbf{R}^{-1} t(\mathbf{X}_{\bullet,\bullet}) = \frac{1}{nJ} \sum_{i=0}^n \sum_{j=1}^J \varepsilon_{i,j} \underbrace{[M^\top \mathbf{R}^{-1} t(\mathbf{X}_{\bullet,j})]_{i+1}}_{=\psi_{t,i,j}(\mathbf{X})}.$$

Conditionally to $\mathbf{X}_{\bullet,\bullet} = \mathbf{x}_{\bullet,\bullet}$, the term

$$\frac{1}{nJ} \sum_{j=1}^J \sum_{i=0}^n \varepsilon_{i,j} \underbrace{[M^\top \mathbf{R}^{-1} t(\mathbf{x}_{\bullet,j})]_{i+1}}_{=\psi_{t,i,j}(\mathbf{x}_{\bullet,j})}$$

is linear with respect to t , centered and a combination of independent variables. Thus, we intend to apply the Talagrand Inequality. But for this, we need to be in a bounded context. As the $\varepsilon_{i,j}$'s are not supposed to be bounded, we rely on Assumption 2 and we split $\varepsilon_{i,j} = \varepsilon_{i,j}^{(1)} + \varepsilon_{i,j}^{(2)}$ and accordingly $\nu_{n,J} = \nu_{n,J,1} + \nu_{n,J,2}$, with $\varepsilon_{i,j}^{(2)} := \varepsilon_{i,j} - \varepsilon_{i,j}^{(1)}$ and

$$\varepsilon_{i,j}^{(1)} := \varepsilon_{i,j} \mathbf{1}_{|\varepsilon_{i,j}| \leq K} - \mathbb{E}(\varepsilon_{i,j} \mathbf{1}_{|\varepsilon_{i,j}| \leq K}).$$

Accordingly, we denote $\varepsilon_{\bullet,\bullet} = \varepsilon_{\bullet,\bullet}^{(1)} + \varepsilon_{\bullet,\bullet}^{(2)}$ and $\mathbf{u}_{\bullet,\bullet}^{(i)} = \mathbf{M}\varepsilon_{\bullet,\bullet}^{(i)}$ for $i = 1, 2$. Then, noting that it holds $S_m + S_{\widehat{m}} = S_{m \vee \widehat{m}}$ since the spaces are nested and $m \vee \widehat{m} \in \mathcal{M}_{n,J}^+$ on $\Omega_{n,J}$ due to Lemma 5, we have

$$\begin{aligned} &\mathbb{E} \left[\left(\sup_{t \in S_m + S_{\widehat{m}}, \|t\|_{n,J,R}=1} \nu_{n,J}^2(t) - p(m, \widehat{m}) \right)_+ \mathbf{1}_{\Omega_{n,J}} \right] \\ &\leq 2 \sum_{m' \in \mathcal{M}_{n,J}^+} \mathbb{E} \left[\left(\sup_{t \in S_{m \vee m'}, \|t\|_{n,J,R}=1} \nu_{n,J,1}^2(t) - \frac{p(m, m')}{2} \right)_+ \right] + 2\mathbb{E} \left[\sup_{t \in S_m + S_{\widehat{m}}, \|t\|_{n,J,R}=1} \nu_{n,J,2}^2(t) \right]. \end{aligned}$$

Denoting by $\mathbb{E}_{\mathbf{X}}$ the conditional expectation with respect to $\mathbf{X}_{\bullet,\bullet}$, we define for $m' \in \mathcal{M}_{n,J}^+$

$$\mathbb{E}_1(m') := \mathbb{E}_{\mathbf{X}} \left[\left(\sup_{t \in S'_m, \|t\|_{n,J,R}=1} \nu_{n,J,1}^2(t) - 4\sigma^2 \frac{m'}{nJ} \right)_+ \right]$$

and

$$\mathbb{E}_2 := \mathbb{E} \left[\sup_{t \in S_m + S_{\widehat{m}}, \|t\|_{n,J,R}=1} \nu_{n,J,2}^2(t) \right].$$

Then, we have that

$$\mathbb{E} \left[\left(\sup_{t \in S_m + S_{\widehat{m}}, \|t\|_{n,J,R}=1} \nu_{n,J}^2(t) - p(m, \widehat{m}) \right)_+ \right] \leq 2 \sum_{m' \in \mathcal{M}_{n,J}^+} \mathbb{E}[\mathbb{E}_1(m' \vee m)] + 2\mathbb{E}_2.$$

Let us first bound $\mathbb{E}_2$. First, we see that

$$\mathbb{E}_2 = \frac{1}{(nJ)^2} \mathbb{E} \left[\sup_{t \in S_m + S_{\widehat{m}}, \|t\|_{n,J,R}=1} ((\varepsilon_{\bullet,\bullet}^{(2)})^\top \mathbf{M}^\top \mathbf{R}^{-1} t(\mathbf{X}_{\bullet,\bullet}))^2 \right].$$

As for $\|t\|_{n,J,R} = 1$ it holds

$$\begin{aligned} ((\varepsilon_{\bullet,\bullet}^{(2)})^\top \mathbf{M}^\top \mathbf{R}^{-1} t(\mathbf{X}_{\bullet,\bullet}))^2 &\leq \|\mathbf{R}^{-1/2} \mathbf{M} \varepsilon_{\bullet,\bullet}^{(2)}\|_{\mathbb{R}^{nJ}}^2 \|\mathbf{R}^{-1/2} t(\mathbf{X}_{\bullet,\bullet})\|_{\mathbb{R}^{nJ}}^2 \\ &= (nJ) \|\mathbf{R}^{-1/2} \mathbf{u}_{\bullet,\bullet}^{(2)}\|_{\mathbb{R}^{nJ}}^2 \leq \frac{nJ}{1-\rho} \|\mathbf{u}_{\bullet,\bullet}^{(2)}\|_{\mathbb{R}^{nJ}}^2 \end{aligned}$$

we get, since $\varepsilon_{i,j}$ are i.i.d. for all $i = 1, \dots, n$ and $j = 0, \dots, J$, that

$$\mathbb{E}_2 \leq \frac{1}{nJ(1-\rho)} \mathbb{E} \left[\sum_{j=1}^J \sum_{i=1}^n [\rho(\varepsilon_{0,j}^{(2)})^2 + (1-\rho)(\varepsilon_{i,j}^{(2)})^2] \right] \leq \frac{1}{1-\rho} \mathbb{E}[\varepsilon_{1,1}^2 \mathbf{1}_{|\varepsilon_{1,1}| > K}].$$

Therefore, by Cauchy-Schwarz and Assumption 2, we obtain

$$\mathbb{E}_2 \leq \frac{1}{1-\rho} \sqrt{\mathbb{E}(\varepsilon_{1,1}^4) \mathbb{P}(|\varepsilon_{1,1}| > K)} \leq \frac{\sqrt{c_1 \mathbb{E}(\varepsilon_{1,1}^4)}}{1-\rho} e^{-c_2 K/2}.$$

Consequently, to obtain the wished result, we need for this term that $\mathbb{E}[\varepsilon_{1,1}^4] < +\infty$ and

$$\mathbb{P}(|\varepsilon_{1,1}| > K) \leq C e^{-2 \log(nJ)},$$

and we choose

$$K = \frac{4}{c_2} \log(nJ). \quad (28)$$

Let us turn to $\mathbb{E}_1(m)$ for some $m \in \mathcal{M}_{n,J}^+$, keeping in mind that we will change m into $m \vee m'$ in the end. We apply Talagrand inequality. For this we first calculate the quantities $\mathbb{H}^2$, v and M_1 upper bounds on the terms $\mathbb{T}_1, \mathbb{T}_2, \mathbb{T}_3$ defined hereafter. The first term to bound is

$$\mathbb{T}_1 := \mathbb{E}_{\mathbf{X}} \left[\sup_{t \in S_m, \|t\|_{n,J,R}=1} \nu_{n,J,1}^2(t) \right]$$

We denote by $\mathbf{Z}_{\bullet,\bullet} := \mathbf{R}^{-1/2} \mathbf{M} \varepsilon_{\bullet,\bullet} = \mathbf{R}^{-1/2} \mathbf{u}_{\bullet,\bullet}$, with decomposition $\mathbf{Z}_{\bullet,\bullet} = \mathbf{Z}_{\bullet,\bullet}^{(1)} + \mathbf{Z}_{\bullet,\bullet}^{(2)}$ following the one of $\varepsilon_{\bullet,\bullet}$. Note that the entries $\mathbf{Z}_{i,j}^{(1)}$ of $\mathbf{Z}_{\bullet,\bullet}^{(1)}$ are centered but not uncorrelated; only the components of $\mathbf{Z}_{\bullet,\bullet}$ are, as $\mathbf{Z}_{\bullet,\bullet} = \mathbf{R}^{-1/2} \mathbf{u}_{\bullet,\bullet}$ are such that $\text{Var}(\mathbf{Z}_{\bullet,\bullet}) = \sigma^2 \text{Id}_{nJ \times nJ}$.

By Gram-Schmidt orthonormalization, we consider $\varphi_{k,R}$, an orthonormalization of the φ_k with respect to the scalar product $\langle \cdot, \cdot \rangle_{n,J,R}$, so that $t \in S_m$ with $\|t\|_{n,J,R} = 1$ can be written as $t = \sum_{k=1}^m a_{k,R} \varphi_{k,R}$

with $\|t\|_{n,J,R}^2 = \sum_{k=1}^m a_{k,R}^2 = 1$ and $\|\varphi_{k,R}\|_{n,J,R}^2 = (nJ)^{-1} \varphi_{k,R}(\mathbf{X}_{\bullet,\bullet})^\top \mathbf{R}^{-1} \varphi_{k,R}(\mathbf{X}_{\bullet,\bullet}) = 1$, for any $k \in \mathbb{N}$. We get using Cauchy-Schwarz inequality

$$\mathbb{T}_1 \leq \mathbb{E}_{\mathbf{X}} \left[\sup_{t=\sum_{k=1}^m a_{k,R} \varphi_{k,R}, \sum_{k=1}^m a_{k,R}^2=1} \left(\sum_{k=1}^m a_{n,R} \nu_{n,J,1}(\varphi_{k,R}) \right)^2 \right] \leq \sum_{k=1}^m \mathbb{E}_{\mathbf{X}} [\nu_{n,J,1}^2(\varphi_{k,R})].$$

Now, relying on the independency of the $\varepsilon_{i,j}^{(1)}$, we compute

$$\begin{aligned} \mathbb{E}_{\mathbf{X}} [\nu_{n,J,1}^2(\varphi_{k,R})] &= \mathbb{E}_{\mathbf{X}} \left[\left(\frac{1}{nJ} \sum_{i=0}^n \sum_{j=1}^J \varepsilon_{i,j}^{(1)} [M^\top R^{-1} \varphi_{k,R}(X_{\bullet,j})]_{i+1} \right)^2 \right] \\ &= \frac{1}{(nJ)^2} \sum_{i=0}^n \sum_{j=1}^J \mathbb{E}[(\varepsilon_{i,j}^{(1)})^2] [M^\top R^{-1} \varphi_{k,R}(X_{\bullet,j})]_{i+1}^2 \\ &\leq \frac{1}{(nJ)^2} \sum_{i=0}^n \sum_{j=1}^J \mathbb{E}[\varepsilon_{i,j}^2] [M^\top R^{-1} \varphi_{k,R}(X_{\bullet,j})]_{i+1}^2 \end{aligned}$$

using that $\mathbb{E}[(\varepsilon_{i,j}^{(1)})^2] \leq \mathbb{E}[\varepsilon_{i,j}^2] = \sigma^2$. Therefore

$$\begin{aligned} \mathbb{E}_{\mathbf{X}} [\nu_{n,J,1}^2(\varphi_{k,R})] &\leq \frac{1}{(nJ)^2} \sum_{i=0}^n \sum_{j=1}^J \mathbb{E}[\varepsilon_{i,j}^2] [M^\top R^{-1} \varphi_{k,R}(X_{\bullet,j})]_{i+1}^2 \\ &= \mathbb{E}_{\mathbf{X}} \left[\left(\frac{1}{nJ} \sum_{i=0}^n \sum_{j=1}^J \varepsilon_{i,j} [M^\top R^{-1} \varphi_{k,R}(X_{\bullet,j})]_{i+1} \right)^2 \right]. \end{aligned}$$

We obtain

$$\mathbb{T}_1 \leq \sum_{k=1}^m \mathbb{E}_{\mathbf{X}} \left[\left(\frac{1}{nJ} \mathbf{u}_{\bullet,\bullet}^\top \mathbf{R}^{-1} \varphi_{k,R}(\mathbf{X}_{\bullet,\bullet}) \right)^2 \right].$$

We get, by using now the non correlation property

$$\begin{aligned} \mathbb{T}_1 &\leq \sum_{k=1}^m \mathbb{E}_{\mathbf{X}} \left[\left(\frac{1}{nJ} \sum_{j=1}^J \sum_{i=1}^n \mathbf{Z}_{i,j} [R^{-1/2} \varphi_{k,R}(\mathbf{X}_{\bullet,j})]_i \right)^2 \right] \\ &= \sum_{k=1}^m \frac{1}{(nJ)^2} \sum_{j=1}^J \sum_{i=1}^n \mathbb{E}_{\mathbf{X}}(\mathbf{Z}_{i,j}^2) [R^{-1/2} \varphi_{k,R}(\mathbf{X}_{\bullet,j})]_i^2 \\ &= \frac{\sigma^2}{(nJ)^2} \sum_{k=1}^m \sum_{j=1}^J \sum_{i=1}^n [R^{-1/2} \varphi_{k,R}(\mathbf{X}_{\bullet,j})]_i^2. \end{aligned}$$

As

$$\sum_{j=1}^J \sum_{i=1}^n [R^{-1/2} \varphi_{j,R}(\mathbf{X}_{\bullet,j})]_i^2 = \varphi_{j,R}(\mathbf{X}_{\bullet,\bullet})^\top \mathbf{R}^{-1} \varphi_{j,R}(\mathbf{X}_{\bullet,\bullet}) = nJ \|\varphi_{j,R}\|_{n,J,R}^2 = nJ,$$

we obtain

$$\mathbb{T}_1 \leq \sigma^2 \frac{m}{nJ} := \mathbb{H}^2.$$

Next we consider the second term useful for applying Talagrand and use the same type of trick as

previously. Further, denote by $\text{Var}_{\mathbf{X}}$ the variance conditional on $\mathbf{X}_{\bullet,\bullet}$.

$$\begin{aligned}
\mathbb{T}_2 &:= \sup_{t \in S_m, \|t\|_{n,J,R}=1} \frac{1}{nJ} \sum_{j=1}^J \sum_{i=0}^n \text{Var}_{\mathbf{X}} \left(\varepsilon_{i,j}^{(1)} [M^\top R^{-1} t(\mathbf{X}_{\bullet,j})]_{i+1} \right) \\
&= \sup_{t \in S_m, \|t\|_{n,J,R}=1} \frac{1}{nJ} \text{Var}_{\mathbf{X}} \left(\sum_{j=1}^J \sum_{i=0}^n \varepsilon_{i,j}^{(1)} [M^\top R^{-1} t(\mathbf{X}_{\bullet,j})]_{i+1} \right) \\
&\leq \sup_{t \in S_m, \|t\|_{n,J,R}=1} \frac{1}{nJ} \text{Var}_{\mathbf{X}} (\varepsilon_{\bullet,\bullet}^\top M^\top R^{-1} t(\mathbf{X}_{\bullet,\bullet})) = \sup_{t \in S_m, \|t\|_{n,J,R}=1} \frac{1}{nJ} \text{Var}_{\mathbf{X}} (\mathbf{Z}_{\bullet,\bullet}^\top \mathbf{R}^{-1/2} t(\mathbf{X}_{\bullet,\bullet})) \\
&= \sup_{t \in S_m, \|t\|_{n,J,R}=1} \frac{1}{nJ} \text{Var}_{\mathbf{X}} \left(\sum_{j=1}^J \sum_{i=1}^n \mathbf{Z}_{i,j} [R^{-1/2} t(\mathbf{X}_{\bullet,j})]_i \right) \\
&= \sigma^2 \sup_{t \in S_m, \|t\|_{n,J,R}=1} \frac{1}{nJ} \sum_{j=1}^J \sum_{i=1}^n [R^{-1/2} t(\mathbf{X}_{\bullet,j})]_i^2 = \sigma^2 := v,
\end{aligned}$$

where we used that $\sum_{j=1}^J \sum_{i=1}^n [R^{-1/2} t(\mathbf{X}_{\bullet,j})]_i^2 = nJ \|t\|_{n,J,R}^2$. The last term is computed as follows.

$$\begin{aligned}
\mathbb{T}_3 &:= \sup_{t \in S_m, \|t\|_{n,J,R}=1} \sup_{e \in \mathbb{R}} \sup_{j=1,\dots,J} \sup_{i=0,\dots,n} |e \mathbf{1}_{|e| \leq K} - \mathbb{E}(\varepsilon_{1,1} \mathbf{1}_{|\varepsilon_{1,1}| \leq K})| [M^\top R^{-1} t(\mathbf{X}_{\bullet,j})]_{i+1}| \\
&\leq 2K \sup_{t \in S_m, \|t\|_{n,J,R}=1} \sup_{i,j} |[M^\top R^{-1} t(\mathbf{X}_{\bullet,j})]_{i+1}|
\end{aligned}$$

Now we write that

$$\begin{aligned}
|[M^\top R^{-1} t(\mathbf{X}_{\bullet,j})]_{i+1}| &\leq \sqrt{\sum_{j=1}^J \sum_{i=0}^n [M^\top R^{-1} t(\mathbf{X}_{\bullet,j})]_{i+1}^2} = \sqrt{\text{Var}_{\mathbf{X}} \left(\sum_{j=1}^J \sum_{i=0}^n \varepsilon_{i,j} [M^\top R^{-1} t(\mathbf{X}_{\bullet,j})]_{i+1} \right)} \\
&= \sqrt{\text{Var}_{\mathbf{X}} (\mathbf{Z}_{\bullet,\bullet}^\top \mathbf{R}^{-1/2} t(\mathbf{X}_{\bullet,\bullet}))} = \sqrt{t(\mathbf{X}_{\bullet,\bullet})^\top \mathbf{R}^{-1/2} \text{Var}(\mathbf{Z}_{\bullet,\bullet}) \mathbf{R}^{-1/2} t(\mathbf{X}_{\bullet,\bullet})} \\
&= \sqrt{nJ} \|t\|_{n,J,R} = \sqrt{nJ}
\end{aligned}$$

for t such that $\|t\|_{n,J,R} = 1$. Finally, we obtain the bound $\mathbb{T}_3 \leq 2K\sqrt{nJ} := M_1$.

Now we obtain with Talagrand inequality for any $m \in \mathcal{M}_n^+$

$$\mathbb{E}_1(m) \leq \frac{C_1}{nJ} \left(e^{-C_2 m} + 4K^2 e^{-C_3 \sqrt{m}/(2K)} \right).$$

Therefore, it follows for any $m \in \mathcal{M}_{n,J}$

$$\begin{aligned}
\sum_{m' \in \mathcal{M}_n^+} \mathbb{E}_1(m \vee m') &\leq \sum_{m' \in \mathcal{M}_n^+} \mathbb{E}_1(m') \leq \frac{C_1}{nJ} \left(\sum_{m \geq 0} e^{-C_2 m} + 4K^2 \sum_{m \in \mathcal{M}_n^+} e^{-C_3 \sqrt{m}/(2K)} \right) \\
&\leq \frac{C_1}{nJ} \left(\Sigma(C_2) + 4K^2 \sum_{\ell=1}^{c(nJ)^2} e^{-C_3 \ell/(2K)} \right) \\
&\leq \frac{C_1}{nJ} \left(\Sigma(C_2) + \frac{4K^2}{1 - e^{-C_3/(2K)}} \right) \leq \frac{C_1}{nJ} \left(\Sigma(C_2) + \frac{8}{C_3} K^3 \right),
\end{aligned}$$

as for $x > 0$, it holds that $1/(1 - e^{-x}) \leq 1/x$. Now, plugging the choice of K given in (28) yields

$$\sum_{m' \in \mathcal{M}_n^+} \mathbb{E}_1(m \vee m') \leq C \frac{\log^3(nJ)}{nJ},$$

for some constant C depending on C_1, C_2, C_3 , the constants of the Talagrand and c_1 and c_2 the constants of Assumption 2. $\square$

6.7 Proof of (11)

Proof of (11). We write that

$$\mathbb{P}(|\varepsilon_{1,1}| > K) = \mathbb{E}(\mathbf{1}_{|\varepsilon_{1,1}| > K}) = \mathbb{E}(\mathbf{1}_{\varepsilon_{1,1} > K} \mathbf{1}_{\varepsilon_{1,1} > 0}) + \mathbb{E}(\mathbf{1}_{\varepsilon_{1,1} < -K} \mathbf{1}_{\varepsilon_{1,1} \leq 0}) := \mathbb{M}_1 + \mathbb{M}_2.$$

Then, for all $u \geq 0$,

$$\mathbb{M}_1 \leq \mathbb{E}(e^{u\varepsilon_{1,1} - uK} \mathbf{1}_{\varepsilon_{1,1} > 0}) \leq e^{-uK + \frac{u^2 s^2}{2}}$$

by using (10). By optimization with respect to u , we choose $u = K/s^2$ and obtain $\mathbb{M}_1 \leq e^{-K^2/(2s^2)}$. Similarly, for $u < 0$, $\mathbb{M}_2 \leq e^{uK + u^2 s^2/2}$ and choosing $u = -K/s^2$ yields $\mathbb{M}_2 \leq e^{-K^2/(2s^2)}$, and the result follows. $\square$

Appendix: auxiliary results

The following result of [Tropp \[2012\]](#) is used.

Lemma 12 (Tropp Chernov). *Consider a fixed sequence $(\mathbf{X}_k)$ of independent, random, self-adjoint matrices with dimension d . Let $\lambda_{\max}(\mathbf{X}_k)$ and $\lambda_{\min}(\mathbf{X}_k)$ denote the largest and smallest eigenvalue of $\mathbf{X}_k$, respectively. Further, let $\preceq$ denote the semi-definite order on self-adjoint matrices. Assume that each random matrix satisfies*

$$\mathbf{X}_k \preceq 0 \quad \text{and} \quad \lambda_{\max}(\mathbf{X}_k) \leq R, \quad \text{almost surely.}$$

Define

$$\mu_{\min} := \lambda_{\min} \left(\sum_{k \in \mathbb{N}} \mathbb{E}[\mathbf{X}_k] \right) \quad \text{and} \quad \mu_{\max} := \lambda_{\max} \left(\sum_{k \in \mathbb{N}} \mathbb{E}[\mathbf{X}_k] \right).$$

Then,

$$\begin{aligned} \mathbb{P} \left(\lambda_{\min} \left(\sum_{k \in \mathbb{N}} \mathbf{X}_k \right) \leq (1 - \delta) \mu_{\min} \right) &\leq d \cdot \left(\frac{e^{-\delta}}{(1 - \delta)^{1-\delta}} \right)^{\mu_{\min}/R}, \quad \text{for } \delta \in [0, 1], \text{ and} \\ \mathbb{P} \left(\lambda_{\max} \left(\sum_{k \in \mathbb{N}} \mathbf{X}_k \right) \leq (1 + \delta) \mu_{\max} \right) &\leq d \cdot \left(\frac{e^{\delta}}{(1 + \delta)^{1+\delta}} \right)^{\mu_{\max}/R}, \quad \text{for } \delta \geq 0. \end{aligned}$$

For the proof of Proposition 2 we use the following helpful result, which can be found e.g. in [Bartlett \[1951\]](#).

Lemma 13 (Sherman-Morrison formula). *Let $A \in \mathbb{R}^{n \times n}$ be an invertible square matrix and $u, v \in \mathbb{R}^n$. Then $A + uv^\top$ is invertible if and only if $1 + v^\top A^{-1}u \neq 0$. In this case, it holds*

$$(A + uv^\top)^{-1} = A^{-1} - \frac{A^{-1}uv^\top A^{-1}}{1 + v^\top A^{-1}u}.$$

References

- Yannick Baraud, F. Comte, and G. Viennet. Model selection for (auto-)regression with dependent data. *ESAIM Probab. Statist.*, 5:33–49, 2001a. ISSN 1292-8100,1262-3318. doi: 10.1051/ps:2001101. URL <https://doi-org.ezproxy.math.cnrs.fr/10.1051/ps:2001101>.
- Yannick Baraud, Fabienne Comte, and Gabrielle Viennet. Adaptive estimation in autoregression or β -mixing regression via model selection. *Ann. Statist.*, 29(3):839–875, 2001b. ISSN 0090-5364,2168-8966. doi: 10.1214/aos/1009210692. URL <https://doi-org.ezproxy.math.cnrs.fr/10.1214/aos/1009210692>.
- Andrew Barron, Lucien Birgé, and Pascal Massart. Risk bounds for model selection via penalization. *Probab. Theory Related Fields*, 113(3):301–413, 1999. ISSN 0178-8051,1432-2064. doi: 10.1007/s004400050210. URL <https://doi-org.ezproxy.math.cnrs.fr/10.1007/s004400050210>.

- M. S. Bartlett. An inverse matrix adjustment arising in discriminant analysis. *Ann. Math. Statistics*, 22:107–111, 1951. ISSN 0003-4851. doi: 10.1214/aoms/1177729698.
- Jean-Patrick Baudry, Cathy Maugis, and Bertrand Michel. Slope heuristics: overview and implementation. *Stat. Comput.*, 22(2):455–470, 2012. ISSN 0960-3174,1573-1375. doi: 10.1007/s11222-011-9236-1. URL <https://doi-org.ezproxy.math.cnrs.fr/10.1007/s11222-011-9236-1>.
- René Carmona, François Delarue, and Daniel Lacker. Mean field games with common noise. *The Annals of Probability*, 44(6):3740–3803, 2016.
- Emmanuel Caron, Jérôme Dedecker, and Bertrand Michel. Gaussian linear model selection in a dependent context. *Electron. J. Stat.*, 15(2):4823–4867, 2021. ISSN 1935-7524. doi: 10.1214/21-ejs1885. URL <https://doi-org.ezproxy.math.cnrs.fr/10.1214/21-ejs1885>.
- Albert Cohen, Mark A. Davenport, and Dany Leviatan. On the stability and accuracy of least squares approximations. *Found. Comput. Math.*, 13(5):819–834, 2013. ISSN 1615-3375,1615-3383. doi: 10.1007/s10208-013-9142-3. URL <https://doi-org.ezproxy.math.cnrs.fr/10.1007/s10208-013-9142-3>.
- Fabienne Comte and Valentine Genon-Catalot. Non parametric regression function estimation in presence of common noise. *The Annals of the Institute of Statistical Mathematics*, 2026.
- Sándor Csörgö and Jan Mielniczuk. Nonparametric regression under long-range dependent normal errors. *Ann. Statist.*, 23(3):1000–1014, 1995. ISSN 0090-5364,2168-8966. doi: 10.1214/aos/1176324633. URL <https://doi-org.ezproxy.math.cnrs.fr/10.1214/aos/1176324633>.
- François Delarue, Etienne Tanré, and Raphaël Maillet. Ergodicity of some stochastic fokker-planck equations with additive common noise. *arXiv preprint arXiv:2405.09950*, 2024.
- Klaus Duellmann, Jonathan Küll, and Michael Kunisch. Estimating asset correlations from stock prices or default rates—which method is superior? *Journal of Economic Dynamics and Control*, 34(11):2341–2357, 2010.
- Sam Efromovich. *Nonparametric curve estimation*. Springer Series in Statistics. Springer-Verlag, New York, 1999. ISBN 0-387-98740-1. Methods, theory, and applications.
- Dominique Fourdrinier and Sergei Pergamenschikov. Improved model selection method for a regression function with dependent noise. *Ann. Inst. Statist. Math.*, 59(3):435–464, 2007. ISSN 0020-3157,1572-9052. doi: 10.1007/s10463-006-0063-7. URL <https://doi-org.ezproxy.math.cnrs.fr/10.1007/s10463-006-0063-7>.
- Valentine Genon-Catalot and Catherine Larédo. Parametric estimation for discretely observed systems of stochastic differential equations with common noise. *Preprint hal-05515893*, 2026.
- Evarist Giné and Richard Nickl. *Mathematical foundations of infinite-dimensional statistical models*, volume [40] of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, New York, 2016. ISBN 978-1-107-04316-9. doi: 10.1017/CBO9781107337862. URL <https://doi-org.ezproxy.math.cnrs.fr/10.1017/CBO9781107337862>.
- Christian Houdré and Patricia Reynaud-Bouret. Exponential inequalities, with constants, for U-statistics of order two. In *Stochastic inequalities and applications*, volume 56 of *Progr. Probab.*, pages 55–69. Birkhäuser, Basel, 2003. ISBN 3-7643-2197-0.
- Yichuan Huang. Non-compactly supported estimation of the diffusion function in ergodic scalar SDEs. This paper focuses on the estimation of diffusion coefficients in stochastic differential equations using high-frequency data., January 2025. URL <https://hal.science/hal-04901917>.
- Daniel Lacker, Mykhaylo Shkolnikov, and Jiacheng Zhang. Superposition and mimicking theorems for conditional mckean–vlasov equations. *Journal of the European Mathematical Society*, 25(8):3229–3288, 2022.

- Raphael Maillet. A note on the long-time behaviour of stochastic mckean-vlasov equations with common noise, 2025. URL <https://arxiv.org/abs/2306.16130>.
- Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12(4):389–434, 2012.
- Alexandre B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009. ISBN 978-0-387-79051-0. doi: 10.1007/b13794. URL <https://doi-org.ezproxy.math.cnrs.fr/10.1007/b13794>. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.