The Annals of Statistics
2012, Vol. 40, No. 6, 2765–2797
DOI: 10.1214/12-AOS1050

ADAPTIVE FUNCTIONAL LINEAR REGRESSION[1]

BY FABIENNE COMTE AND JAN JOHANNES

Université Paris Descartes and Sorbonne Paris Cité, and Université Catholique de Louvain

We consider the estimation of the slope function in functional linear regression, where scalar responses are modeled in dependence of random functions. Cardot and Johannes [*J. Multivariate Anal.* **101** (2010) 395–408] have shown that a thresholded projection estimator can attain up to a constant minimax-rates of convergence in a general framework which allows us to cover the prediction problem with respect to the mean squared prediction error as well as the estimation of the slope function and its derivatives. This estimation procedure, however, requires an optimal choice of a tuning parameter with regard to certain characteristics of the slope function and the covariance operator associated with the functional regressor. As this information is usually inaccessible in practice, we investigate a fully data-driven choice of the tuning parameter which combines model selection and Lepski's method. It is inspired by the recent work of Goldenshluger and Lepski [*Ann. Statist.* **39** (2011) 1608–1632]. The tuning parameter is selected as minimizer of a stochastic penalized contrast function imitating Lepski's method among a random collection of admissible values. This choice of the tuning parameter depends only on the data and we show that within the general framework the resulting data-driven thresholded projection estimator can attain minimax-rates up to a constant over a variety of classes of slope functions and covariance operators. The results are illustrated considering different configurations which cover in particular the prediction problem as well as the estimation of the slope and its derivatives. A simulation study shows the reasonable performance of the fully data-driven estimation procedure.

1. Introduction. In functional linear regression the dependence of a real-valued response Y on the variation of a random function X is studied. Typically the functional regressor X is assumed to be square-integrable or more generally to take its values in a separable Hilbert space $\mathbb{H}$ with inner product $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ and norm $\| \cdot \|_{\mathbb{H}}$. Furthermore, we suppose that Y and X are centered, which simplifies the notations, and that the dependence between Y and X is linear in the sense that

$$Y = \langle \beta, X \rangle_{\mathbb{H}} + \sigma \varepsilon, \qquad \sigma > 0, \tag{1.1}$$

Received December 2011; revised June 2012.

[1]Supported by the IAP research network no. P6/03 of the Belgian Government (Belgian Science Policy).

MSC2010 subject classifications. Primary 62J05; secondary 62G35, 62G20.

Key words and phrases. Adaptation, model selestion, Lepski's method, linear Galerkin approach, prediction, derivative estimation, minimax theory.

for some slope function $\beta \in \mathbb{H}$ and error term ε with mean zero and variance one. Assuming an independent and identically distributed (i.i.d.) sample of (Y, X), the objective of this paper is the construction of a fully data driven estimation procedure of the slope function β which still can attain minimax-optimal rates of convergence.

Functional linear models have become very important in a diverse range of disciplines, including medicine, linguistics, chemometrics as well as econometrics; see, for instance, [15] and [36], for several case studies, or more specific, [16] and [35] for applications in economics. The main class of estimation procedures of the slope function studied in the statistical literature is based on principal components regression; see, for example, [2, 6, 9, 17] or [31] in the context of generalized linear models. The second important class of estimators relies on minimizing a penalized least squares criterion which can be seen as generalization of the ridge regression; cf. [7] and [28]. More recently an estimator based on dimension reduction and threshold techniques has been proposed by Cardot and Johannes [8] which borrows ideas from the inverse problems community ([13] and [23]). It is worth noting that all the proposed estimation procedures rely on the choice of at least one tuning parameter, which in turn, crucially influences the attainable accuracy of the constructed estimator.

It has been shown, for example, in [8], that the attainable accuracy of an estimator of the slope β is essentially determined by a priori conditions imposed on both the slope function and the covariance operator Γ associated to the random function X (defined below). These conditions are usually captured by suitably chosen classes $\mathcal{F} \subset \mathbb{H}$ and $\mathcal{G}$ of slope functions and covariance operators, respectively. Typically, the class $\mathcal{F}$ characterizes the level of smoothness of the slope function, while the class $\mathcal{G}$ specifies the decay of the sequence of eigenvalues of Γ. For example, [5, 12] or [21] consider differentiable slope functions and a polynomial decay of the eigenvalues of Γ. Furthermore, given a weighted norm $\|\cdot\|_\omega$ and the completion $\mathcal{F}_\omega$ of $\mathbb{H}$ with respect to $\|\cdot\|_\omega$ we shall measure the performance of an estimator $\widehat{\beta}$ of β by its maximal $\mathcal{F}_\omega$-risk over a class $\mathcal{F} \subset \mathcal{F}_\omega$ of slope functions and a class $\mathcal{G}$ of covariance operators, that is,

$$\mathcal{R}_\omega[\widehat{\beta}; \mathcal{F}, \mathcal{G}] := \sup_{\beta \in \mathcal{F}} \sup_{\Gamma \in \mathcal{G}} \mathbb{E}\|\widehat{\beta} - \beta\|_\omega^2.$$

This general framework with appropriate choice of the weighted norm $\|\cdot\|_\omega$ allows us to cover the prediction problem with respect to the mean squared prediction error (see, e.g., [7] or [12]) and the estimation not only of the slope function (see, e.g., [21]) but also of its derivatives. For a detailed discussion, we refer to [8]. Having these applications in mind the additional condition $\mathcal{F} \subset \mathcal{F}_\omega$ only means that the estimation of a derivative of the slope function necessitates its existence. Assuming an i.i.d. sample of (Y, X) of size n obeying model (1.1) Cardot and Johannes [8] have derived a lower bound of the maximal weighted risk, that is,

$$\mathcal{R}_\omega^*[n; \mathcal{F}, \mathcal{G}] \leq C \inf_{\widehat{\beta}} \mathcal{R}_\omega[\widehat{\beta}; \mathcal{F}, \mathcal{G}]$$

for some finite positive constant C where the infimum is taken over all possible estimators $\widehat{\beta}$. Moreover, they have shown that a thresholded projection estimator $\widehat{\beta}_{m_n^*}$ in dependence of an optimally chosen tuning parameter $m_n^* \in \mathbb{N}$ can attain this lower bound up to a constant $C > 0$,

$$R_\omega[\widehat{\beta}_{m_n^*}; \mathcal{F}, \mathcal{G}] \le C R_\omega^*[n; \mathcal{F}, \mathcal{G}]$$

for a variety of classes $\mathcal{F}$ and $\mathcal{G}$. In other words, $R_\omega^*[n; \mathcal{F}, \mathcal{G}]$ is the minimax rate of convergence and $\widehat{\beta}_{m_n^*}$ is minimax-optimal. The optimal choice m_n^* of the tuning parameter, however, follows from a classical squared-bias-variance compromise and requires an a priori knowledge about the classes $\mathcal{F}$ and $\mathcal{G}$, which is usually inaccessible in practice.

In this paper we propose a fully data driven method to select a tuning parameter $\widehat{m}$ in such a way that the resulting data-driven estimator $\widehat{\beta}_{\widehat{m}}$ can still attain the minimax-rate $R_\omega^*[n; \mathcal{F}, \mathcal{G}]$ up to a constant over a variety of classes $\mathcal{F}$ and $\mathcal{G}$. It is interesting to note that, considering a linear regression model with infinitely many regressors, Goldenshluger and Tsybakov [19, 20] propose an optimal data-driven prediction procedure allowing sharp oracle inequalities. However, a straightforward application of their results is not obvious to us since they assume a priori standardized regressors, which in turn, in functional linear regression necessitates the covariance operator Γ to be fully known in advance. In contrast, given a jointly normally distributed regressor and error term, Verzelen [38] establishes sharp oracle inequalities for the prediction problem in case the covariance operator is not known in advance. Although, it is worth noting that considering the mean prediction error as risk eliminates the ill-posedness of the underlying problem, which in turn leads to faster minimax rates of convergences of the prediction error than, for example, the mean integrated squared error. Cai and Zhou [5] present a fully data-driven estimation procedure of the slope function which attains optimal rates of convergence with respect to the maximal mean integrated squared error. On the other hand, covering both of these two risks within the general framework discussed above, Comte and Johannes [10] consider functional linear regression with circular functional regressor which results in a partial knowledge of the associated covariance operator, that is, its eigenfunctions are known in advance, but the eigenvalues have to be estimated. In this situation, Comte and Johannes [10] have applied successfully a model selection approach which is inspired by the work of [1] now extensively discussed in [29]. In the circular case, it is possible to develop the unknown slope function in the eigenbasis of the covariance operator, which in turn, allows one to derive an orthogonal series estimator in dependence of a dimension parameter. This dimension parameter has been chosen fully data driven by a model selection approach, and it is shown that the resulting data-driven orthogonal series estimator can attain minimax-optimal rates of convergence up to a constant. Although, the proof crucially relies on the possibility to write the orthogonal series estimator as a minimizer of a contrast.

In this paper we do not impose an a priori knowledge of the eigenbasis, and hence the orthogonal series estimator is no more accessible to us. Instead, we consider the thresholded projection estimator $\widehat{\beta}_m$ as presented in [8] which we did not succeed to write as a minimizer of a contrast. Therefore, our selection method combines model selection and Lepski's method (cf. [27] and its recent review in [30]) which is inspired by a bandwidth selection method in kernel density estimation proposed recently in [18]. Selecting the dimension parameter $\widehat{m}$ as minimizer of a stochastic penalized contrast function imitating Lepski's method among a random collection of admissible values, we show that the fully data-driven estimator $\widehat{\beta}_{\widehat{m}}$ can attain the minimax-rate up to a constant $C > 0$, that is,

$$R_\omega[\widehat{\beta}_{\widehat{m}}; \mathcal{F}, \mathcal{G}] \le C \cdot R^\star_\omega[n; \mathcal{F}, \mathcal{G}] \tag{1.2}$$

for a variety of classes $\mathcal{F}$ and $\mathcal{G}$. We shall emphasize that in contrast to the result obtained in [5], we show that the proposed estimator can attain minimax-optimal rates without specifying in advance neither that the slope function belongs to a class of differentiable or analytic functions nor that the decay of the eigenvalues is polynomial or exponential. The only price for this flexibility is in term of the constant C which is asymptotically not equal to one; that is, the oracle inequality (1.2) is not sharp.

The paper is organized as follows: in Section 2 we briefly introduce the thresholded projection estimator $\widehat{\beta}_m$ as proposed in [8]. We present the data driven method to select the tuning parameter and prove a first upper risk-bound for the fully data-driven estimator $\widehat{\beta}_{\widehat{m}}$ which emphasizes the key arguments. In Section 3 we review the available minimax theory as presented in [8]. Within this general framework we derive upper risk-bounds for the fully-data driven estimator imposing additional assumptions on the distribution of the functional regressor X and the error term ε. Namely, we suppose first that X and ε are Gaussian random variables and second that they satisfy certain moment conditions. In both cases the proof of the upper risk-bound employs the key arguments given in Section 2, while more technical aspects are deferred to the Appendix. The results in this paper are illustrated considering different configurations of classes $\mathcal{F}$ and $\mathcal{G}$. We recall the minimax-rates in this situations and show that up to a constant, these rates are attained by the fully-data driven estimator. A simulation study illustrating the reasonable performance of the fully data-driven estimation procedure is available at the supplementary material archive.

2. Methodology.

Consider the functional linear model (1.1) where the random function X and the error term ε are independent. Let the centred random function X, that is, $\mathbb{E}\langle X, h\rangle_{\mathbb{H}} = 0$ for all $h \in \mathbb{H}$, have a finite second moment, that is, $\mathbb{E}\|X\|^2_{\mathbb{H}} < \infty$. Multiplying both sides in (1.1) by $\langle X, h\rangle_{\mathbb{H}}$ and taking the expectation leads to the normal equation

$$\langle g, h\rangle_{\mathbb{H}} := \mathbb{E}\big[Y\langle X, h\rangle_{\mathbb{H}}\big] = \mathbb{E}\big[\langle \beta, X\rangle_{\mathbb{H}}\langle X, h\rangle_{\mathbb{H}}\big] =: \langle \Gamma\beta, h\rangle_{\mathbb{H}} \qquad \forall h \in \mathbb{H}, \tag{2.1}$$

where g belongs to $\mathbb{H}$, and Γ denotes the covariance operator associated to the random function X. Throughout the paper we shall assume that there exists a solution $\beta \in \mathbb{H}$ of equation (2.1) and that the covariance operator Γ is strictly positive definite which ensures the identifiability of the slope function β; cf. [7]. However, due to the finite second moment of X the associated covariance operator Γ has a finite trace; that is, it is nuclear. Thereby, solving equation (2.1) is an *ill-posed inverse problem* with the additional difficulty that Γ is unknown and has to be estimated; for a detailed discussion of ill-posed inverse problems in general we refer to [14].

2.1. *Thresholded projection estimator.* In this paper, we follow [8] and consider a linear Galerkin approach to derive an estimator of the slope function β. Here and subsequently, let $\{\psi_j\}_{j\geq 1}$ be a pre-specified orthonormal basis in $\mathbb{H}$ which in general does not correspond to the eigenbasis of the operator Γ defined in (2.1). With respect to this basis, we consider for all $h \in \mathbb{H}$ the development $h = \sum_{j=1}^{\infty}[h]_j \psi_j$ where the sequence $([h]_j)_{j\geq 1}$ with generic elements $[h]_j := \langle h, \psi_j\rangle_{\mathbb{H}}$ is square-summable, that is, $\|h\|_{\mathbb{H}}^2 = \sum_{j\geq 1}[h]_j^2 < \infty$. Moreover, given any strictly positive sequence of weights $(\omega_j)_{j\geq 1}$ define the weighted norm $\|h\|_{\omega}^2 := \sum_{j=1}^{\infty} \omega_j [h]_j^2$. We will refer to any sequence as a whole by omitting its index as, for example, in "the sequence of weights ω." Furthermore, for $m \geq 1$ let $[h]_{\underline{m}} := ([h]_1, \dots, [h]_m)^t$ (where x^t is the transpose of x), and let $\mathbb{H}_m$ be the subspace of $\mathbb{H}$ spanned by $\{\psi_1, \dots, \psi_m\}$. Obviously, the norm of $h \in \mathbb{H}_m$ equals the Euclidean norm of its coefficient vector $[h]_{\underline{m}}$, that is, $\|h\|_{\mathbb{H}} = ([h]_{\underline{m}}^t [h]_{\underline{m}})^{1/2} =: \|[h]_{\underline{m}}\|$ with a slight abuse of notation.

An element $\beta^m \in \mathbb{H}_m$ satisfying $\|g - \Gamma\beta^m\|_{\mathbb{H}} \leq \|g - \Gamma\breve{\beta}\|_{\mathbb{H}}$ for all $\breve{\beta} \in \mathbb{H}_m$, is called a Galerkin solution of equation (2.1). Since the covariance operator Γ is strictly positive definite, it follows that the covariance matrix $[\Gamma]_{\underline{m}} := \mathbb{E}([X]_{\underline{m}}[X]_{\underline{m}}^t)$ associated with the m-dimensional random vector $[X]_{\underline{m}}$ is strictly positive definite too. Consequently, the Galerkin solution $\beta^m \in \mathbb{H}_m$ is uniquely determined by $[\beta^m]_{\underline{m}} = [\Gamma]_{\underline{m}}^{-1}[g]_{\underline{m}}$ and $[\beta^m]_j = 0$ for all $j > m$. Although, it does generally not correspond to the orthogonal projection of β onto the subspace $\mathbb{H}_m$ and the approximation error $\sup_{k\geq m} \|\beta^k - \beta\|_{\omega}$ does generally not converge to zero as $m \to \infty$. Here and subsequently, however, we restrict ourselves to classes $\mathcal{F}$ and $\mathcal{G}$ of slope functions and covariance operators, respectively, which ensure the convergence. Obviously, this is a minimal regularity condition for us since we aim to estimate the Galerkin solution.

Assuming a sample $\{(Y_i, X_i)\}_{i=1}^n$ of (Y, X) of size n, it is natural to consider the estimators $\widehat{g} := n^{-1}\sum_{i=1}^n Y_i X_i$ and $\widehat{\Gamma} := n^{-1}\sum_{i=1}^n \langle \cdot, X_i\rangle_{\mathbb{H}} X_i$ for g and Γ, respectively. Moreover, let $[\widehat{\Gamma}]_{\underline{m}} := n^{-1}\sum_{i=1}^n [X_i]_{\underline{m}}[X_i]_{\underline{m}}^t$ and note that $[\widehat{g}]_{\underline{m}} = n^{-1}\sum_{i=1}^n Y_i [X_i]_{\underline{m}}$. Replacing the unknown quantities by their empirical counterparts $\widetilde{\beta}^m \in \mathbb{H}_m$ denotes a Galerkin solution satisfying $\|\widehat{g} - \widehat{\Gamma}\widetilde{\beta}^m\|_{\mathbb{H}} \leq \|\widehat{g} - \widehat{\Gamma}\breve{\beta}\|_{\mathbb{H}}$ for all $\breve{\beta} \in \mathbb{H}_m$. Observe that there exists always a solution $\widetilde{\beta}^m$, but it might not be unique. Obviously, if $[\widehat{\Gamma}]_{\underline{m}}$ is nonsingular, then $[\widetilde{\beta}^m]_{\underline{m}} = [\widehat{\Gamma}]_{\underline{m}}^{-1}[\widehat{g}]_{\underline{m}}$. We shall emphasize the multiplication with the inverse of the random matrix $[\widehat{\Gamma}]_{\underline{m}}$ which may result in an unstable estimator even in case $[\Gamma]_{\underline{m}}$ is well conditioned. Let $\mathbb{1}_{\{\|[\widehat{\Gamma}]_{\underline{m}}^{-1}\|_s \le n\}}$ denote the indicator function which takes the value one if $[\widehat{\Gamma}]_{\underline{m}}$ is nonsingular with spectral norm $\|[\widehat{\Gamma}]_{\underline{m}}^{-1}\|_s := \sup_{\|z\|=1} \|[\widehat{\Gamma}]_{\underline{m}}^{-1} z\|$ of its inverse bounded by n, and the value zero otherwise. The estimator of β proposed in [8] consists of thresholding the estimated Galerkin solution, that is,

$$\widehat{\beta}_m := \widetilde{\beta}^m \mathbb{1}_{\{\|[\widehat{\Gamma}]_{\underline{m}}^{-1}\|_s \le n\}}. \tag{2.2}$$

In the next paragraph we introduce a data-driven method to select the dimension parameter $m \in \mathbb{N}$.

2.2. *Data-driven thresholded projection estimator.* Given a random integer $\widehat{M}$ and a random sub sequence of penalties $(\widehat{\mathrm{pen}}_1, \ldots, \widehat{\mathrm{pen}}_{\widehat{M}})$, we select the dimension parameter $\widehat{m}$ among the random collection of admissible values $\{1, \ldots, \widehat{M}\}$ as minimizer of a penalized contrast criterion. To be precise, setting $\arg\min_{m \in A}\{a_m\} := \min\{m : a_m \le a_{m'}, \forall m' \in A\}$ for a sequence $(a_m)_{m \ge 1}$ with minimal value in $A \subset \mathbb{N}$, we define

$$\widehat{m} := \mathop{\arg\min}_{1 \le m \le \widehat{M}} \{\Psi_m + \widehat{\mathrm{pen}}_m\}. \tag{2.3}$$

The data-driven estimator of β is now given by $\widehat{\beta}_{\widehat{m}}$, and below we derive an upper bound for its maximal $\mathcal{F}_\omega$-risk. The choice of the $\mathcal{F}_\omega$-risk as performance measure is reflected in the definition of the contrasts, that is,

$$\Psi_m := \max_{m \le k \le \widehat{M}} \{\|\widehat{\beta}_k - \widehat{\beta}_m\|_\omega^2 - \widehat{\mathrm{pen}}_k\}, \qquad 1 \le m \le \widehat{M}.$$

The construction of the random penalty sequence $\widehat{\mathrm{pen}}$ and the upper bound $\widehat{M}$ given below is guided by the key arguments used in the proof of the $\mathcal{F}_\omega$-risk bound which we present first. A central step for our reasoning is the next assertion which employs essentially the particular choice of the contrast.

LEMMA 2.1. *Consider the approximation errors* $\mathrm{bias}_m = \sup_{m \le k} \|\beta^k - \beta\|_\omega$, $m \ge 1$. *If the sub sequence* $(\widehat{\mathrm{pen}}_1, \ldots, \widehat{\mathrm{pen}}_{\widehat{M}})$ *is nondecreasing, then we have*

$$\|\widehat{\beta}_{\widehat{m}} - \beta\|_\omega^2 \le 7\widehat{\mathrm{pen}}_m + 78\mathrm{bias}_m^2 + 42 \max_{m \le k \le \widehat{M}} \left(\|\widehat{\beta}_k - \beta^k\|_\omega^2 - \frac{1}{6}\widehat{\mathrm{pen}}_k\right)_+ \tag{2.4}$$

for all $1 \le m \le \widehat{M}$, *where* $(a)_+ = \max(a, 0)$.

PROOF. From the definition of $\widehat{m}$ we deduce for all $1 \le m \le \widehat{M}$ that

$$\begin{aligned} \|\widehat{\beta}_{\widehat{m}} - \beta\|_\omega^2 &\le 3\{\Psi_m + \widehat{\mathrm{pen}}_{\widehat{m}} + \Psi_{\widehat{m}} + \widehat{\mathrm{pen}}_m + \|\widehat{\beta}_m - \beta\|_\omega^2\} \\ &\le 6\{\Psi_m + \widehat{\mathrm{pen}}_m\} + 3\|\widehat{\beta}_m - \beta\|_\omega^2. \end{aligned} \tag{2.5}$$

First, employing an elementary triangular inequality allows us to write

$$\|\widehat{\beta}_m - \beta\|_\omega^2 \le \frac{1}{3}\widehat{\mathrm{pen}}_m + 2\mathrm{bias}_m^2 + 2 \max_{m\le k\le M}\left(\|\widehat{\beta}_k - \beta^k\|_\omega^2 - \frac{1}{6}\widehat{\mathrm{pen}}_k\right)_+$$

for all $1 \le m \le \widehat{M}$. Second, since $(\widehat{\mathrm{pen}}_1, \dots, \widehat{\mathrm{pen}}_{\widehat{M}})$ is nondecreasing and $4\mathrm{bias}_m^2 \ge \max_{m\le k\le \widehat{M}} \|\beta^k - \beta^m\|_\omega^2$, $1 \le m \le \widehat{M}$, it is easily verified that

$$\Psi_m \le 6 \sup_{m\le k\le \widehat{M}}\left(\|\widehat{\beta}_k - \beta^k\|_\omega^2 - \frac{1}{6}\widehat{\mathrm{pen}}_k\right)_+ + 12\mathrm{bias}_m^2.$$

Combining the last two inequalities and (2.5), we obtain the result. □

Keeping the last assertion in mind we decompose the $\mathcal{F}_\omega$-risk with respect to an event on which the quantities $\widehat{\mathrm{pen}}_m$ and $\widehat{M}$ are close to some theoretical counterparts pen_m, M_n^- and M_n^+. More precisely, define the event

$$\mathcal{E}_n := \{\mathrm{pen}_k \le \widehat{\mathrm{pen}}_k \le 72\mathrm{pen}_k, \forall 1 \le k \le M_n^+\} \cap \{M_n^- \le \widehat{M} \le M_n^+\} \tag{2.6}$$

and the corresponding risk decomposition

$$R_\omega(\widehat{\beta}_{\widehat{m}}; \mathcal{F}, \mathcal{G}) = \sup_{\beta\in\mathcal{F}}\sup_{\Gamma\in\mathcal{G}} \mathbb{E}(\|\widehat{\beta}_{\widehat{m}} - \beta\|_\omega^2 \mathbb{1}_{\mathcal{E}_n}) + \sup_{\beta\in\mathcal{F}}\sup_{\Gamma\in\mathcal{G}} \mathbb{E}(\|\widehat{\beta}_{\widehat{m}} - \beta\|_\omega^2 \mathbb{1}_{\mathcal{E}_n^c}). \tag{2.7}$$

Consider the first right-hand side (r.h.s.) term. If $(\widehat{\mathrm{pen}}_1, \dots, \widehat{\mathrm{pen}}_{\widehat{M}})$ is nondecreasing, then we may apply Lemma 2.1 which on the event $\mathcal{E}_n$ implies

$$\begin{aligned}\|\widehat{\beta}_{\widehat{m}} - \beta\|_\omega^2 \mathbb{1}_{\mathcal{E}_n} &\le 582 \max(\mathrm{pen}_{m_n^\diamond}, \mathrm{bias}_{m_n^\diamond}^2)\\ &\quad + 42 \max_{m_n^\diamond\le k\le M_n^+}\left(\|\widehat{\beta}_k - \beta^k\|_\omega^2 - \frac{1}{6}\mathrm{pen}_k\right)_+,\end{aligned}$$

where $m_n^\diamond$ realizes a penalty-squared-bias compromise among the collection of admissible values $\{1, \dots, M_n^-\}$. Keeping in mind that $m_n^\diamond$ should mimic the value of the optimal variance-squared-bias trade-off, we wish the upper bound M_n^- to be as large as possible. In contrast, in order to control the remainder term, the second r.h.s. term, we are forced to use a rather small upper bound $M_n^+ \ge M_n^-$ to ensure that the penalty term is uniformly bounded with increasing sample size. However, we bound the remainder term by imposing the following assumption, which though holds true for a wide range of classes $\mathcal{F}$ and $\mathcal{G}$ under reasonable assumptions on the distribution of ε and X; see Propositions 3.1 and 3.3 in Section 3.

ASSUMPTION 2.1. There exists a constant K_1 such that

$$\sup_{\beta\in\mathcal{F}}\sup_{\Gamma\in\mathcal{G}} \mathbb{E}\left\{\max_{m_n^\diamond\le k\le M_n^+}\left(\|\widehat{\beta}_k - \beta^k\|_\omega^2 - \frac{1}{6}\mathrm{pen}_k\right)_+\right\} \le K_1 n^{-1} \qquad \text{for all } n \ge 1.$$

Roughly speaking, the penalty term pen_m should provide an upper bound for the estimator's variation which allows us to establish a concentration inequality for the $\|\cdot\|_\omega$-norm of the corresponding empirical process. However, under Assumption 2.1 we bound the first r.h.s. term in (2.7) by

$$\sup_{\beta\in\mathcal{F}}\sup_{\Gamma\in\mathcal{G}}\mathbb{E}\big(\|\widehat{\beta}_{\widehat{m}}-\beta\|_\omega^2\mathbb{1}_{\mathcal{E}_n}\big)\le 582\sup_{\beta\in\mathcal{F}}\sup_{\Gamma\in\mathcal{G}}\max\{\mathrm{pen}_{m_n^\diamond},\mathrm{bias}^2_{m_n^\diamond}\}+42\frac{K_1}{n}. \tag{2.8}$$

It remains to consider the second r.h.s. term. The conditions on the distribution of ε and X presented in the next section are also sufficient to show that the following assumption holds true.

ASSUMPTION 2.2. There exists a constant $K_2>0$ such that

$$\sup_{\beta\in\mathcal{F}}\sup_{\Gamma\in\mathcal{G}}\mathbb{E}\big(\|\widehat{\beta}_{\widehat{m}}-\beta\|_\omega^2\mathbb{1}_{\mathcal{E}_n^c}\big)\le K_2n^{-1}\qquad\text{for all } n\ge 1.$$

Under Assumption 2.2, $\widehat{M}$ and $\widehat{\mathrm{pen}}_m$ behave similarly to their theoretical counterparts with sufficiently high probability so as not to deteriorate the estimators risk. The next assertion provides an upper bound for the maximal $\mathcal{F}_\omega$-risk over the classes $\mathcal{F}$ and $\mathcal{G}$ of the thresholded projection estimator $\widehat{\beta}_{\widehat{m}}$ with data-driven choice $\widehat{m}$ given by (2.3).

PROPOSITION 2.2. *Suppose that* $(\widehat{\mathrm{pen}}_1,\dots,\widehat{\mathrm{pen}}_{\widehat{M}})$ *is nondecreasing. If Assumptions 2.1 and 2.2 hold true, then for all* $n\ge 1$ *we have* $\mathcal{R}_\omega[\widehat{\beta}_{\widehat{m}};\mathcal{F},\mathcal{G}]\le 582\sup_{\beta\in\mathcal{F}}\sup_{\Gamma\in\mathcal{G}}\max\{\mathrm{pen}_{m_n^\diamond},\mathrm{bias}^2_{m_n^\diamond}\}+(42K_1+K_2)n^{-1}$.

PROOF. Keeping in mind the risk decomposition (2.7) the upper bound (2.8) and Assumption 2.2 imply the result. □

REMARK 2.1. The first r.h.s. term in the last upper risk-bound is strongly reminiscent of a variance-squared-bias decomposition of the $\mathcal{F}_\omega$-risk for the estimator $\widehat{\beta}_{m_n^\diamond}$ with dimension parameter $m_n^\diamond$. Indeed, in many cases the penalty term $\mathrm{pen}_{m_n^\diamond}$ is in the same order as the variance of the estimator $\widehat{\beta}_{m_n^\diamond}$; cf. Illustration 3.1[P-P] and [E-P] below. Consequently, in this situation the upper risk bound of the data-driven estimator is essentially given by $\mathcal{R}_\omega[\widehat{\beta}_{m_n^\diamond};\mathcal{F},\mathcal{G}]$. Moreover, by balancing penalty and squared-bias $m_n^\diamond$ just realizes the optimal trade-off between variance and squared-bias which in turn in many cases means that $\mathcal{R}_\omega[\widehat{\beta}_{m_n^\diamond};\mathcal{F},\mathcal{G}]$ is of optimal order.

We complete this section by introducing our choice for the random upper bound $\widehat{M}$ and the random penalty $\widehat{\mathrm{pen}}_m$ which takes its inspiration from [10]. Let us first define some auxiliary quantities required in the construction. For $m\ge 1$, let $[\nabla_\omega]_{\underline{m}}$

denote the m-dimensional diagonal matrix with diagonal entries $(\omega_j)_{1\le j\le m}$, and for any sequence $[K] := ([K]_{\underline{k}})_{k\ge 1}$ of matrices, define

$$\Delta_m^{[K]} := \max_{1\le k\le m} \|[\nabla_\omega]_{\underline{k}}^{1/2}[K]_{\underline{k}}^{-1}[\nabla_\omega]_{\underline{k}}^{1/2}\|_s \quad \text{and}$$

(2.9)

$$\delta_m^{[K]} := m\Delta_m^{[K]}\frac{\log(\Delta_m^{[K]}\vee(m+2))}{\log(m+2)}.$$

For $n\ge 1$, set $M_n^\omega := \max\{1\le m\le \lfloor n^{1/4}\rfloor : \omega_{(m)}\le n\}$ with integer part $\lfloor n^{1/4}\rfloor$ of $n^{1/4}$ and $\omega_{(m)} := \max_{1\le k\le m}\omega_k$. For any sequence $a := (a_m)_{m\ge 1}$ let

$$M_n(a) := \min\left\{2\le m\le M_n^\omega : m\omega_{(m)}a_m > \frac{n}{1+\log n}\right\} - 1, \tag{2.10}$$

where we set $M_n(a) := M_n^\omega$ if the defining set is empty. Given the sequence of covariance matrices $[\Gamma] = ([\Gamma]_{\underline{m}})_{m\ge 1}$ associated with the regressor X, define

(2.11)

$$\mathrm{pen}_m := \kappa\sigma_m^2\delta_m^{[\Gamma]}n^{-1} \qquad \text{with } \sigma_m^2 := 2(\mathbb{E}Y^2 + [g]_{\underline{m}}^t[\Gamma]_{\underline{m}}^{-1}[g]_{\underline{m}}) \quad \text{and}$$

$$M^\Gamma := M_n(a) \qquad \text{with } a := (\|[\Gamma]_{\underline{m}}^{-1}\|_s)_{m\ge 1},$$

where κ is a positive numerical constant to be chosen below. Roughly speaking the penalty term provides an upper bound of the variance of the estimator $\widehat{\beta}_m$ and is in many cases even in the same order. Its construction, however, allows a deterioration to ensure that Assumption 2.1 can be satisfied; cf. Illustration 3.1[P-E]. Moreover, for growing sample size n the penalty sequence is uniformly bounded over the collection of admissible values $\{1,\dots,M_n^\Gamma\}$. Note that the penalty and the upper bound still depend on unknown quantities which, however, can easily be estimated, that is,

(2.12)

$$\widehat{\mathrm{pen}}_m := 14\kappa\widehat{\sigma}_m^2\delta_m^{[\widehat{\Gamma}]}n^{-1}$$

$$\text{with } \widehat{\sigma}_m^2 := 2\left(\frac{1}{n}\sum_{i=1}^n Y_i^2 + [\widehat{g}]_{\underline{m}}^t[\widehat{\Gamma}]_{\underline{m}}^{-1}[\widehat{g}]_{\underline{m}}\right) \quad \text{and}$$

$$\widehat{M} := M_n(a) \qquad \text{with } a := (\|[\widehat{\Gamma}]_{\underline{m}}^{-1}\|_s)_{m\ge 1}.$$

Note that by construction $(\widehat{\mathrm{pen}}_1,\dots,\widehat{\mathrm{pen}}_{\widehat{M}})$ is nondecreasing. Indeed, the identity $\langle\widehat{\Gamma}(\widehat{\beta}_k-\widehat{\beta}_m),(\widehat{\beta}_k-\widehat{\beta}_m)\rangle_{\mathbb{H}} = [\widehat{g}]_{\underline{k}}^t[\widehat{\Gamma}]_{\underline{k}}^{-1}[\widehat{g}]_{\underline{k}} - [\widehat{g}]_{\underline{m}}^t[\widehat{\Gamma}]_{\underline{m}}^{-1}[\widehat{g}]_{\underline{m}}$ holds true for all $1\le m\le k\le \widehat{M}$. Since $\widehat{\Gamma}$ is positive definite, $[\widehat{g}]_{\underline{m}}^t[\widehat{\Gamma}]_{\underline{m}}^{-1}[\widehat{g}]_{\underline{m}} \le [\widehat{g}]_{\underline{k}}^t[\widehat{\Gamma}]_{\underline{k}}^{-1}[\widehat{g}]_{\underline{k}}$ and $\widehat{\sigma}_m^2\le\widehat{\sigma}_k^2$ which in turn implies the assertion. Consequently, we may apply Proposition 2.2 if Assumptions 2.1 and 2.2 hold true.

3. Minimax-optimality. In this section we recall first a general framework proposed by Cardot and Johannes [8] which allows us to derive minimax-optimal rates for the maximal $\mathcal{F}_\omega$-risk, $\sup_{\beta\in\mathcal{F}}\sup_{\Gamma\in\mathcal{G}}\mathbb{E}\|\widehat{\beta}-\beta\|_\omega^2$, over the classes $\mathcal{F}$ and $\mathcal{G}$.

3.1. *Notations and basic assumptions*. The classes $\mathcal{F}$ and $\mathcal{G}$ of slope functions and covariance operators, respectively, are characterized by different weighted norms in $\mathbb{H}$ with respect to the pre-specified orthonormal basis $\{\psi_j, j \in \mathbb{N}\}$. Given a strictly positive sequence of weights b and a radius $r > 0$, let $\mathcal{F}_b$ be the completion of $\mathbb{H}$ with respect to the weighted norm $\|\cdot\|_b$ and the ellipsoid $\mathcal{F}_b^r := \{h \in \mathcal{F}_b : \|h\|_b^2 \le r\}$ be the class of possible slope functions. Furthermore, as usual in the context of ill-posed inverse problems, we link the mapping properties of the covariance operator Γ and the regularity condition $\beta \in \mathcal{F}_b^r$. Denote by $\mathcal{N}$ the set of all strictly positive nuclear operators defined on $\mathbb{H}$. Given a strictly positive sequence of weights γ and a constant $d \ge 1$ define the class of covariance operators by

$$\mathcal{G}_\gamma^d := \{T \in \mathcal{N} : d^{-2}\|f\|_{\gamma^2}^2 \le \|Tf\|^2 \le d^2\|f\|_{\gamma^2}^2, \forall f \in \mathbb{H}\},$$

where arithmetic operations on sequences are defined element-wise, for example, $\gamma^2 = (\gamma_j^2)_{j\ge1}$. Let us briefly discuss the last definition. If $T \in \mathcal{G}_\gamma^d$, then we have $d^{-1} \le \langle T\psi_j, \psi_j\rangle/\gamma_j \le d$, for all $j \ge 1$. Consequently, the sequence γ is necessarily summable, because T is nuclear. Moreover, if λ denotes the sequence of eigenvalues of T, then $d^{-1} \le \lambda_j/\gamma_j \le d$, for all $j \ge 1$. In other words the sequence γ characterizes the decay of the eigenvalues of $T \in \mathcal{G}_\gamma^d$. We do not specify the sequences of weights ω, b and γ, but impose from now on the following minimal regularity conditions.

ASSUMPTION 3.1. Let ω, b and γ be strictly positive sequences of weights with $b_1 = \omega_1 = \gamma_1 = 1$, and $\sum_{j=1}^{\infty} \gamma_j < \infty$ such that the sequences b^{-1}, ωb^{-1}, γ and $\gamma^2\omega^{-1}$ are monotonically nonincreasing and converging to zero.

The last assumption is fairly mild. For example, assuming that ωb^{-1} is nonincreasing, ensures that $\mathcal{F}_b^r \subset \mathcal{F}_\omega$. Furthermore, it is shown in [8] that the minimax rate $R_\omega^*[n; \mathcal{F}_b^r, \mathcal{G}_\gamma^d]$ is of order n^{-1} for all sequences γ and ω such that $\gamma^2\omega^{-1}$ is nondecreasing. We will illustrate all our results considering the following three configurations for the sequences ω, b and γ.

ILLUSTRATION 3.1. In all three cases, we take $\omega_j = j^{2s}$, $j \ge 1$. Moreover, let:

[P-P] $b_j = j^{2p}$ and $\gamma_j = j^{-2a}$, $j \ge 1$, with $p > 0$, $a > 1/2$ and $p > s > -2a$;
[E-P] $b_j = \exp(j^{2p} - 1)$ and $\gamma_j = j^{-2a}$, $j \ge 1$, with $p > 0$, $a > 1/2$, $s > -2a$;
[P-E] $b_j = j^{2p}$ and $\gamma_j = \exp(-j^{2a} + 1)$, $j \ge 1$, with $p > 0$, $a > 0$, and $p > s$;

then Assumption 3.1 is satisfied in all cases.

REMARK 3.1. In the configurations [P-P] and [E-P], the case $s = -a$ can be interpreted as mean-prediction error; cf. [8]. Moreover, if $\{\psi_j\}$ is the trigonometric basis and the value of s is an integer, then the weighted norm $\|h\|_\omega$ corresponds to the L^2-norm of the weak sth derivative of h; cf. [33]. In other words in this situation we consider as risk the mean integrated squared error when estimating the sth derivative of β. In the configurations [P-P] and [P-E], the additional condition $p > s$ means that the slope function has at least $p \geq s+1$ weak derivatives, while for a value $p > 1$ in [E-P], the slope function is assumed to be an analytic function; cf. [25].

3.2. *Minimax optimal estimation reviewed.* Let us first recall a lower bound of the maximal $\mathcal{F}_\omega$-risk over the classes $\mathcal{F}_b^r$ and $\mathcal{G}_\gamma^d$ due to [8]. Given an i.i.d. sample of (Y, X) of size n and sequences as in Assumption 3.1, define

$$m_n^* := \arg\min_{m\geq 1}\left\{\max\left(\frac{\omega_m}{b_m}, \sum_{j=1}^{m}\frac{\omega_j}{n\gamma_j}\right)\right\} \quad \text{and}$$

(3.1)

$$R_n^* := \max\left(\frac{\omega_{m_n^*}}{b_{m_n^*}}, \sum_{j=1}^{m_n^*}\frac{\omega_j}{n\gamma_j}\right).$$

If $\xi := \inf_{n\geq 1}\{(R_n^*)^{-1}\min(\omega_{m_n^*}b_{m_n^*}^{-1}, \sum_{j=1}^{m_n^*}\omega_j(n\gamma_j)^{-1})\} > 0$, then there exists a constant $C := C(\sigma, r, d, \xi) > 0$ depending on σ, r, d and ξ only such that

$$\inf_{\widetilde{\beta}} \mathcal{R}_\omega^*[\widetilde{\beta}; \mathcal{F}_b^r, \mathcal{G}_\gamma^d] \geq C R_n^* \qquad \text{for all } n \geq 1. \tag{3.2}$$

On the other hand, considering the dimension parameter m_n^* given in (3.1) Cardot and Johannes [8] have shown that the maximal risk $R_\omega^*[\widehat{\beta}_{m_n^*}; \mathcal{F}_b^r, \mathcal{G}_\gamma^d]$ of the estimator $\widehat{\beta}_{m_n^*}$ defined in (2.2) is bounded by R_n^* up to constant for a wide range of sequences ω, b and γ, provided the random function X and the error ε satisfy certain additional moment conditions. In other words $R_n^* = R_\omega^*[n; \mathcal{F}_b^r, \mathcal{G}_\gamma^d]$ is the minimax-rate in this situation, and the estimator $\widehat{\beta}_{m_n^*}$ is minimax optimal; although, the definition of the dimension parameter m_n^* necessitates an a priori knowledge of the sequences b and γ. In the remaining part of this paper we show that the data-driven choice of the dimension parameter constructed in Section 2 can automatically attain the minimax-rate R_n^* for a variety of sequences ω, b and γ. First, let us briefly illustrate the minimax result.

ILLUSTRATION (CONTINUED) 3.2. Considering the three configurations (see Illustration 3.1), it has been shown in [8] that the estimator $\widehat{\beta}_{m_n^*}$ with m_n^* as given below attains the rate R_n^* up to a constant. We write for two strictly positive sequences $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$ that $a_n \sim b_n$, if $(a_n/b_n)_{n\geq 1}$ is bounded away from 0 and infinity.

[P-P] If $s + a > -1/2$, then $m_n^* \sim n^{1/(2p+2a+1)}$, while $m_n^* \sim n^{1/[2(p-s)]}$ for $s + a < -1/2$. Thus, $R_n^* \sim \max(n^{-(2p-2s)/(2a+2p+1)}, n^{-1})$ for $s + a \neq -1/2$. If $s + a = -1/2$, then $m_n^* \sim (n/\log n)^{1/[2(p-s)]}$ and $R_n^* \sim \log(n)/n$.

[E-P] If $s + a > -1/2$, then $m_n^* \sim (\log n - \frac{2a+1}{2p}\log(\log n))^{1/(2p)}$ and $R_n^* \sim n^{-1}(\log n)^{(2a+1+2s)/(2p)}$, while $m_n^* \sim (\log n + (s/p)\log(\log n))^{1/(2p)}$ and $R_n^* \sim n^{-1}$ for $s + a < -1/2$ [and $R_n^* \sim \log(\log n)/n$ for $a + s = -1/2$].

[P-E] $m_n^* \sim (\log n - \frac{2p+(2a-1)_+}{2a}\log(\log n))^{1/(2a)}$ and $R_n^* \sim (\log n)^{-(p-s)/a}$.

An increasing value of the parameter a leads in all three cases to a slower rate R_n^*, and hence it is called degree of ill-posedness; cf. [32].

3.3. *Minimax-optimality of the data-driven estimation procedure*. Consider the thresholded projection estimator $\widehat{\beta}_{\widehat{m}}$ with data-driven choice $\widehat{m}$ of the dimension parameter. Supposing that the joint distribution of the random function X and the error term ε satisfies certain additional conditions, we will prove below that Assumptions 2.1 and 2.2 formulated in Section 2 hold true. These assumptions rely on the existence of sequences $(m_n^\diamond)_{n\geq 1}$ and $(M_n^+)_{n\geq 1}$ which amongst others we define now referring only to the classes $\mathcal{F}_b^r$ and $\mathcal{G}_\gamma^d$. Keep in mind the notation given in (2.9) and (2.10). For $m, n \geq 1$ and $[\nabla_\gamma] = ([\nabla_\gamma]_{\underline{m}})_{m\geq 1}$ define $\Delta_m^\gamma := \Delta_m^{[\nabla_\gamma]}$ and $\delta_m^\gamma := \delta_m^{[\nabla_\gamma]}$, set $M_n^- := M_n(16d^3\gamma^{-1})$ and $M_n^+ := M_n((4d\gamma)^{-1})$, and let

$$m_n^\diamond := \mathop{\arg\min}_{1\leq m\leq M_n^-}\left\{\max\left(\frac{\omega_m}{b_m}, \frac{\delta_m^\gamma}{n}\right)\right\} \quad \text{and} \quad R_n^\diamond := \max\left(\frac{\omega_{m_n^\diamond}}{b_{m_n^\diamond}}, \frac{\delta_{m_n^\diamond}^\gamma}{n}\right),$$

where $m_n^\diamond \leq M_n^- \leq M_n^+$. Let $\Sigma := \Sigma(\mathcal{G}_\gamma^d)$ denote a finite constant such that

$$\Sigma \geq \sum_{j\geq 1}\gamma_j \quad \text{and} \quad \Sigma \geq \sum_{m\geq 1}\Delta_m^\gamma \exp\left(-\frac{m\log(\Delta_m^\gamma \vee (m+2))}{16(1+\log d)\log(m+2)}\right), \tag{3.3}$$

which by construction always exists and depends on the class $\mathcal{G}_\gamma^d$ only. We illustrate below the last definitions by revisiting the three configurations for the sequences ω, b and γ (Illustration 3.1).

ILLUSTRATION (CONTINUED) 3.3. In the following we state the order of M_n^- and δ_m^γ which in turn are used to derive the order of $m_n^\diamond$ and $R_n^\diamond$.

[P-P] $M_n^- \sim (\frac{n}{1+\log n})^{1+2a+(2s)_+}$, $\delta_m^\gamma \sim m^{1+(2a+2s)_+}$ and for $p > (s)_+$ it follows $m_n^\diamond \sim m^{1/[1+2p-2s+(2a+2s)_+]}$ and $R_n^\diamond \sim n^{-2(p-s)/[1+2p-2s+(2a+2s)_+]}$;

[E-P] $M_n^- \sim (\frac{n}{1+\log n})^{1+2a+(2s)_+}$, $\delta_m^\gamma \sim m^{1+(2a+2s)_+}$ and for $p > 0$, $m_n^\diamond \sim (\log n - \frac{1+2(a+s)_+-2s}{2p}\log(\log n))^{1/(2p)}$ and $R_n^\diamond \sim n^{-1}(\log n)^{[1+2(a+s)_+]/(2p)}$;

[P-E] $M_n^- \sim (\log n - \frac{1+2a+2(s)_+}{2a}\log(\log n))^{1/(2a)}$, $\delta_m^\gamma \sim m^{1+2s+2a}\exp(m^{2a})$ and for $p > (s)_+$, it follows $m_n^\diamond \sim (\log n - \frac{1+2a+2p}{2a}\log(\log n))^{1/(2a)}$ and $R_n^\diamond \sim (\log n)^{-(p-s)/a}$.

We proceed by formalizing additional conditions on the joint distribution of ε and X, allowing us to prove that Assumptions 2.1 and 2.2 hold true.

Imposing a joint normal distribution. Let us first assume that X is a centred Gaussian $\mathbb{H}$-valued random variable; that is, for all $k \geq 1$ and for all finite collections $\{h_1, \dots, h_k\} \subset \mathbb{H}$ the joint distribution of the real valued random variables $\langle X, h_1\rangle_{\mathbb{H}}, \dots, \langle X, h_k\rangle_{\mathbb{H}}$ is Gaussian with zero mean vector and covariance matrix with generic elements $\mathbb{E}\langle h_j, X\rangle_{\mathbb{H}}\langle X, h_l\rangle_{\mathbb{H}}$, $1 \leq j, l \leq k$. Moreover, suppose that the error term is standard normally distributed.

ASSUMPTION 3.2. The joint distribution of X and ε is normal.

The more involved proof of the next assertion is deferred to Appendix C.

PROPOSITION 3.1. *Assume an i.i.d. n-sample of (Y, X) obeying (1.1) and Assumption 3.2. Consider sequences ω, b and γ satisfying Assumption 3.1 and set $\kappa = 96$ in the definition (2.11) and (2.12) of the penalty* pen *and* $\widehat{\text{pen}}$, *respectively. For the classes $\mathcal{F}_b^r$ and $\mathcal{G}_\gamma^d$ there exist finite constants $C_1 := C_1(d)$ and $C_2 := C_2(d)$ depending on d only such that Assumptions 2.1 and 2.2 with $K_1 := C_1(\sigma^2 + r)\Sigma$ and $K_2 := C_2(\sigma^2 + r)\Sigma$, respectively, holds true.*

By taking the value $\kappa = 96$ the random penalty and upper bound given in (2.12) depend indeed only on the data, and hence the choice $\widehat{m}$ in (2.3) is fully data-driven. Moreover, we can apply Proposition 2.2 to prove the next upper risk-bound for the data-driven thresholded projection estimator $\widehat{\beta}_{\widehat{m}}$.

THEOREM 3.2. *Let the assumptions of Proposition 3.1 be satisfied. There exists a finite constant $K := K(d)$ depending on d only such that*

$$R_\omega[\widehat{\beta}_{\widehat{m}}, \mathcal{F}_b^r, \mathcal{G}_\gamma^d] \leq K(\sigma^2 + r)\{R_n^\diamond + \Sigma n^{-1}\} \qquad \text{for all } n \geq 1.$$

PROOF. We shall provide in the Appendix among others, the two technical Lemmas B.1 and B.2 which are used in the following. Moreover, we denote by $K := K(d)$ a constant depending on d only which changes from line to line. Making use of Proposition 3.1 we intend to apply Proposition 2.2. To this end, if $\beta \in \mathcal{F}_b^r$ and $\Gamma \in \mathcal{G}_\gamma^d$, then first from (iv) in Lemma B.1 it follows that $\text{bias}^2_{m_n^\diamond} \leq 34 d^8 r \omega_{m_n^\diamond} b^{-1}_{m_n^\diamond}$ because $\gamma^2\omega^{-1}$ and ωb^{-1} are nonincreasing due to Assumption 3.1. Second, by combination of (i) and (iv) in Lemma B.2, it is easily verified that $\text{pen}_{m^\diamond} \leq K(\sigma^2 + r)\delta^\gamma_{m^\diamond} n^{-1}$. Consequently, $\sup_{\beta\in\mathcal{F}_b^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\max(\text{pen}_{m_n^\diamond}, \text{bias}^2_{m_n^\diamond}) \leq K(\sigma^2 + r)R_n^\diamond$ for all $n \geq 1$ by combination of the last two estimates and the definition of $R_n^\diamond$ which in turn together with Proposition 2.2 implies the assertion of the theorem. □

Imposing moment conditions. We now dismiss Assumption 3.2 and formalize in its place, conditions on the moments of the random function X and the error term ε. In particular we use that for all $h \in \mathbb{H}$ with $\langle \Gamma h, h\rangle = 1$, the random variable $\langle h, X\rangle$ is standardized, that is, has mean zero and variance one.

ASSUMPTION 3.3. There exist a finite integer $k \geq 16$ and a finite constant $\eta \geq 1$ such that $\mathbb{E}|\varepsilon|^{4k} \leq \eta^{4k}$ and that for all $h \in \mathbb{H}$ with $\langle \Gamma h, h\rangle = 1$ the standardized random variable $\langle h, X\rangle$ satisfies $\mathbb{E}|\langle h, X\rangle|^{4k} \leq \eta^{4k}$.

It is worth noting that for any Gaussian random function X with finite second moment, Assumption 3.3 holds true, since for all $h \in \mathbb{H}$ with $\langle \Gamma h, h\rangle = 1$ the random variable $\langle h, X\rangle$ is standard normally distributed and hence $\mathbb{E}|\langle h, X\rangle|^{2k} = (2k-1)\cdot\cdots\cdot 5\cdot 3\cdot 1$. The proof of the next assertion is again rather involved and deferred to Appendix D. It follows, however, along the general lines of the proof of Proposition 2.2 though it is not a straightforward extension. Take as an example the concentration inequality for the random variable $\|[\Gamma]_{\underline{m}}^{1/2}([\widehat{g}]_{\underline{m}} - [\widehat{\Gamma}]_{\underline{m}}[\beta^m]_{\underline{m}})\|$ in Lemma C.3 in Appendix C which due to Assumption 3.2 is shown by employing elementary inequalities for Gaussian random variables. In contrast, the proof of an analogous result under Assumption 3.3 given in Lemma D.3 in Appendix D is based on an inequality due to Talagrand [37] (Proposition D.1 in the Appendix states a version as presented in [26]).

PROPOSITION 3.3. *Assume an i.i.d. n-sample of (Y, X) obeying (1.1) and Assumption 3.3. Consider sequences as in Assumption 3.1 and set $\kappa = 288$ in the definition (2.11) and (2.12) of the penalty* pen *and* $\widehat{\text{pen}}$, *respectively. For the classes $\mathcal{F}_b^r$ and $\mathcal{G}_\gamma^d$, there exist finite constants $C_1 := C_1(\sigma, \eta, \mathcal{F}_b^r, \mathcal{G}_\gamma^d)$ depending on σ, η and the classes $\mathcal{F}_b^r$ and $\mathcal{G}_\gamma^d$ only, and $C_2 := C_2(d)$ depending on d only, such that Assumptions 2.1 and 2.2 with $K_1 := C_1\eta^{64}(\sigma^2 + r)\Sigma$ and $K_2 := C_2\eta^{64}(\sigma^2 + r)\Sigma$, respectively, hold true.*

We remark on a change only in the constants when comparing the last proposition with Proposition 3.1. Note further that we need a larger value for the constant κ than in Proposition 3.1 although it is still a numerical constant and hence the choice $\widehat{m}$ given by (2.3) is again fully data-driven. Moreover, both values for the constant κ, though convenient for deriving the theory, are far too large in practice. In our simulation study they are instead determined by means of preliminary simulations as proposed in [11], for example. The next assertion provides an upper risk-bound for the data-driven thresholded projection estimator $\widehat{\beta}_{\widehat{m}}$ when imposing moment conditions.

THEOREM 3.4. *Let the assumptions of Proposition 3.3 be satisfied. There exist finite constants $K := K(d)$ depending on d only and $K' := K'(\sigma, \eta, \mathcal{F}_b^r, \mathcal{G}_\gamma^d)$ depending on σ, η and the classes $\mathcal{F}_b^r$ and $\mathcal{G}_\gamma^d$ only such that*

$$\mathcal{R}_\omega[\widehat{\beta}_{\widehat{m}}, \mathcal{F}_b^r, \mathcal{G}_\gamma^d] \leq K(\sigma^2 + r)\{\mathcal{R}_n^\diamond + K'\eta^{64}\Sigma n^{-1}\} \qquad \textit{for all } n \geq 1.$$

PROOF. Taking into account Proposition 3.3 rather than Proposition 3.1 we follow line by line the proof of Theorem 3.2 and we omit the details. □

Minimax-optimality. A comparison of the upper bounds in both Theorems 3.2 and 3.4 with the lower bound displayed in (3.2) shows that the data-driven estimator $\widehat{\beta}_{\widehat{m}}$ attains up to a constant the minimax-rate $R_n^* = \min_{1\le m<\infty}\{\max(\frac{\omega_m}{b_m}, \sum_{j=1}^m \frac{\omega_j}{n\gamma_j})\}$ only if $R_n^\diamond = \min_{1\le m\le M_n^-}\{\max(\frac{\omega_m}{b_m}, \frac{\delta_m^\gamma}{n})\}$ has the same order as R_n^*. Note that, by construction, $\delta_m^\gamma \ge \sum_{j=1}^m \frac{\omega_j}{\gamma_j}$ for all $m \ge 1$. The next assertion is an immediate consequence of Theorems 3.2 and 3.4, and we omit its proof.

COROLLARY 3.5. *Let the assumptions of either Theorems 3.2 or 3.4 be satisfied. If* $\xi^\diamond := \sup_{n\ge1}\{R_n^\diamond/R_n^*\} < \infty$ *holds true, then* $R_\omega[\widehat{\beta}_{\widehat{m}}; \mathcal{F}_b^r, \mathcal{G}_\gamma^d] \le C \cdot \inf_{\widetilde{\beta}} R_\omega[\widetilde{\beta}; \mathcal{F}_b^r, \mathcal{G}_\gamma^d]$ *for all* $n \ge 1$ *and a finite positive constant* C, *where the infimum is taken over all possible estimators* $\widetilde{\beta}$.

REMARK 3.2. In the last assertion $\xi^\diamond = \sup_{n\ge1}\{R_n^\diamond/R_n^*\} < \infty$ is, for example, satisfied if the following two conditions hold simultaneously true: (i) $m_n^* \le M_n^-$ for all $n \ge 1$ and (ii) $\Delta_m^\gamma = \max_{1\le j\le m} \omega_j\gamma_j^{-1} \le Cm^{-1}\sum_{j=1}^m \omega_j\gamma_j^{-1}$ and $\log(\Delta_m^\gamma \vee (m+2)) \le C\log(m+2)$ for all $m \ge 1$. Observe that (ii) which implies $\delta_m^\gamma \le C\sum_{j=1}^m \frac{\omega_j}{\gamma_j}$ is satisfied in case Δ_m^γ is in the order of a power of m (e.g., Illustration 3.2[P-P] and [E-P]). If this term has an exponential order with respect to m (e.g., Illustration 3.2[P-E]), then a deterioration of the term δ_m^γ compared to the variance term $\sum_{j=1}^m \frac{\omega_j}{\gamma_j}$ is possible. However, no loss in terms of the rate may occur, that is, $\xi^\diamond < \infty$, when the squared-bias term $\omega_{m_n^\diamond} b_{m_n^\diamond}^{-1}$ dominates the variance term $n^{-1}\delta_{m_n^\diamond}^\gamma$; for a detailed discussion in a deconvolution context, we refer to [3, 4].

Let us illustrate the performance of the data-driven thresholded projection estimator revisiting the three configurations presented in Illustration 3.1.

PROPOSITION 3.6. *Assume an i.i.d. n-sample of* (Y, X) *satisfying* (1.1) *and let either Assumptions* 3.2 *or* 3.3 *hold true where we set*, *respectively*, $\kappa = 96$ *or* $\kappa = 288$ *in* (2.12). *The data-driven estimator* $\widehat{\beta}_{\widehat{m}}$ *attains the minimax-rates* R_n^*, *up to a constant*, *in the three cases given in Illustration* 3.1, *if we additionally assume* $a + s \ge 0$ *in the cases* [P-P] *and* [E-P].

PROOF. Under the stated conditions it is easily verified that the assumptions of either Theorems 3.2 or 3.4 are satisfied. Moreover, the rates R_n^* (Illustration 3.2) and $R_n^\diamond$ (Illustration 3.3) are of the same order if we additionally assume $a + s \ge 0$ in the cases [P-P] and [E-P]. Therefore, Corollary 3.5 applies, and we obtain the assertion. □

APPENDIX

This section gathers preliminary technical results and the proofs of Propositions 3.1 and 3.3.

APPENDIX A: NOTATION

We begin by defining and recalling notation to be used in all proofs. Given $m \geq 1$, $\mathbb{H}_m$ denotes the subspace of $\mathbb{H}$ spanned by the functions $\{\psi_1, \dots, \psi_m\}$. Π_m and $\Pi_m^\perp$ denote the orthogonal projections on $\mathbb{H}_m$ and its orthogonal complement $\mathbb{H}_m^\perp$, respectively. If K is an operator mapping $\mathbb{H}$ to itself and if we restrict $\Pi_m K \Pi_m$ to an operator from $\mathbb{H}_m$ to itself, then it can be represented by a matrix $[K]_{\underline{m}}$ with generic entries $\langle \psi_j, K\psi_l\rangle_{\mathbb{H}} =: [K]_{j,l}$ for $1 \leq j, l \leq m$. The spectral norm of $[K]_{\underline{m}}$ is denoted by $\|[K]_{\underline{m}}\|_s$, and the inverse matrix of $[K]_{\underline{m}}$ by $[K]_{\underline{m}}^{-1}$. Furthermore, $[\nabla_\omega]_{\underline{m}}$ and $[\mathrm{Id}]_{\underline{m}}$ denote, respectively, the m-dimensional diagonal matrix with diagonal entries $(\omega_j)_{1\leq j\leq m}$ and the identity matrix. For $h \in \mathbb{H}_m$ it follows $\|h\|_\omega^2 = [h]_{\underline{m}}^t [\nabla_\omega]_{\underline{m}} [h]_{\underline{m}} = \|[\nabla_\omega]_{\underline{m}}^{1/2} [h]_{\underline{m}}\|^2$. Keeping in mind the notation given in (2.9)–(2.12) we use for all $m \geq 1$ in addition $\Lambda_m^{[\Gamma]} := \frac{\log(\Delta_m^{[\Gamma]} \vee (m+2))}{\log(m+2)}$, $\Lambda_m^\gamma := \frac{\log(\Delta_m^\gamma \vee (m+2))}{\log(m+2)}$ and $\Lambda_m^{[\widehat{\Gamma}]} := \frac{\log(\Delta_m^{[\widehat{\Gamma}]} \vee (m+2))}{\log(m+2)}$ allowing us to write $\delta_m^{[\Gamma]} = m\Delta_m^{[\Gamma]}\Lambda_m^{[\Gamma]}$, $\delta_m^\gamma = m\Delta_m^\gamma \Lambda_m^\gamma$ and $\delta_m^{[\widehat{\Gamma}]} = m\Delta_m^{[\widehat{\Gamma}]}\Lambda_m^{[\widehat{\Gamma}]}$. Given a Galerkin solution $\beta^m \in \mathbb{H}_m$ of equation (1.2), let $Z_m := Y - \langle \beta^m, X\rangle_{\mathbb{H}} = \sigma\varepsilon + \langle \beta - \beta^m, X\rangle_{\mathbb{H}}$ and denote $\rho_m^2 := \mathbb{E}Z_m^2 = \sigma^2 + \langle \Gamma(\beta - \beta^m), (\beta - \beta^m)\rangle_{\mathbb{H}}$, $\sigma_Y^2 := \mathbb{E}Y^2 = \sigma^2 + \langle \Gamma\beta, \beta\rangle_{\mathbb{H}}$ and $\sigma_m^2 = 2(\sigma_Y^2 + [g]_{\underline{m}}^t [\Gamma]_{\underline{m}}^{-1} [g]_{\underline{m}})$ employing that ε and X are uncorrelated. Define the matrix $[\Xi]_{\underline{m}} := [\Gamma]_{\underline{m}}^{-1/2} [\widehat{\Gamma}]_{\underline{m}} [\Gamma]_{\underline{m}}^{-1/2} - [\mathrm{Id}]_{\underline{m}}$ and the vector $[W]_{\underline{m}} := [\widehat{g}]_{\underline{m}} - [\widehat{\Gamma}]_{\underline{m}} [\beta^m]_{\underline{m}}$ satisfying $\mathbb{E}[\Xi]_{\underline{m}} = 0$ and $\mathbb{E}[W]_{\underline{m}} = [\Gamma(\beta - \beta^m)]_{\underline{m}} = 0$. Let further $\widehat{\sigma}_Y^2 := n^{-1}\sum_{i=1}^n Y_i^2$ and define the events

$$\Omega_{m,n} := \{\|[\widehat{\Gamma}]_{\underline{m}}^{-1}\|_s \leq n\}, \qquad \mho_{m,n} := \{8\|[\Xi]_{\underline{m}}\|_s \leq 1\},$$

$$\mathcal{A}_n := \{1/2 \leq \widehat{\sigma}_Y^2/\sigma_Y^2 \leq 3/2\}, \qquad \mathcal{B}_n := \{\|[\Xi]_{\underline{k}}\|_s \leq 1/8, \forall 1 \leq k \leq M_n^\omega\}, \tag{A.1}$$

$$\mathcal{C}_n := \{8[W]_{\underline{k}}^t [\Gamma]_{\underline{k}}^{-1} [W]_{\underline{k}} \leq ([g]_{\underline{k}}^t [\Gamma]_{\underline{k}}^{-1} [g]_{\underline{k}} + \sigma_Y^2), \forall 1 \leq k \leq M_n^\omega\},$$

and their complements $\Omega_{m,n}^c$, $\mho_{m,n}^c$, $\mathcal{A}_n^c$, $\mathcal{B}_n^c$ and $\mathcal{C}_n^c$, respectively. Furthermore, we will denote by C universal numerical constants and by $C(\cdot)$ constants depending only on the arguments. In both cases, the values of the constants may change from line to line.

APPENDIX B: PRELIMINARY RESULTS

This section gathers results exploiting Assumption 3.1 only. The proof of the next lemma can be found in [24].

LEMMA B.1. *Let $\Gamma \in \mathcal{G}_\gamma^d$ with sequence γ as in Assumption 3.1. Then we have*:

(i) $\sup_{m\geq 1}\{\gamma_m \|[\Gamma]_{\underline{m}}^{-1}\|_s\} \leq 4d^3$;

(ii) $\sup_{m\geq 1} \|[\nabla_\gamma]_{\underline{m}}^{1/2}[\Gamma]_{\underline{m}}^{-1}[\nabla_\gamma]_{\underline{m}}^{1/2}\|_s \leq 4d^3$;

(iii) $\sup_{m\geq 1} \|[\nabla_\gamma]_{\underline{m}}^{-1/2}[\Gamma]_{\underline{m}}[\nabla_\gamma]_{\underline{m}}^{-1/2}\|_s \leq d$.

Let in addition $\beta \in \mathcal{F}_b^r$ with sequence b as in Assumption 3.1. If β^m denotes a Galerkin solution of $g = \Gamma\beta$, then for each strictly positive sequence w such that wb^{-1} is nonincreasing and for all $m \geq 1$ we obtain:

(iv) $\|\beta - \beta^m\|_w^2 \leq 34d^8 r w_m b_m^{-1} \max(1, \gamma_m^2 w_m^{-1} \max_{1\leq j\leq m} w_j \gamma_j^{-2})$;

(v) $\|\beta^m\|_b^2 \leq 34d^8 r$ *and* $\|\Gamma^{1/2}(\beta - \beta^m)\|_{\mathbb{H}}^2 \leq 34d^9 r\gamma_m b_m^{-1}$.

LEMMA B.2. *Let Assumption 3.1 be satisfied. If $\Gamma \in \mathcal{G}_\gamma^d$ and $D := 4d^3$, then*:

(i) $d^{-1} \leq \gamma_m \|[\Gamma]_{\underline{m}}^{-1}\|_s \leq D$, $d^{-1} \leq \Delta_m^{[\Gamma]}/\Delta_m^\gamma \leq D$, $(1 + \log d)^{-1} \leq \Lambda_m^{[\Gamma]}/\Lambda_m^\gamma \leq (1+\log D)$, *and* $d^{-1}(1+\log d)^{-1} \leq \delta_m^{[\Gamma]}/\delta_m^\gamma \leq D(1+\log D)$, *for all* $m \geq 1$;

(ii) $\delta_{M_n^+}^\gamma \leq n4D(1 + \log D)$ *and* $\delta_{M_n^+}^{[\Gamma]} \leq n4D^2(1 + 2\log D)$, *for all* $n \geq 1$;

(iii) $n \geq 2\max_{1\leq m\leq M_n^+} \|[\Gamma]_{\underline{m}}^{-1}\|$ *if* $n \geq 2D$ *and* $\omega_{(M_n^+)} M_n^+(1 + \log n) \geq 8D^2$;

(iv) $\rho_m^2 \leq \sigma_m^2 \leq 2(\sigma^2 + 35d^9 r)$, *for all* $m \geq 1$, *assuming in addition* $\beta \in \mathcal{F}_b^r$.

PROOF. Consider (i). From Lemma B.1(i), (iii) follows $\|[\Gamma]_{\underline{m}}^{-1}\|_s \leq 4d^3\gamma_m^{-1}$ and $\gamma_m^{-1} \leq d\|[\Gamma]_{\underline{m}}^{-1}\|_s$ which in turn imply $d^{-1} \leq \|[\Gamma]_{\underline{m}}^{-1}\|_s\gamma_m \leq D$ and $d^{-1} \leq \gamma_M \max_{1\leq m\leq M} \|[\Gamma]_{\underline{m}}^{-1}\|_s \leq D$ due to the monotonicity of γ. From these estimates we conclude (i). Consider (ii). Observe that $\Delta_{M_n^+}^\gamma \leq \omega_{(M_n^+)}\gamma_{M_n^+}^{-1}$. In case $M_n^+ = 1$ the assertion follows from $\omega_{(1)}\gamma_1^{-1} = 1$ (Assumption 3.1). Thus, let $M_n^\omega \geq M_n^+ > 1$, then $\min_{1\leq j\leq M_n^+}\{\gamma_j(j\omega_{(j)})^{-1}\} \geq (1 + \log n)(4Dn)^{-1}$, and hence $M_n^+\Delta_{M_n^+}^\gamma \leq 4Dn(1 + \log n)^{-1}$, $\Lambda_{M_n^+}^\gamma \leq (1 + \log D)(1 + \log n)$, $M_n^+\Delta_{M_n^+}^{[\Gamma]} \leq 4D^2 n(1 + \log n)^{-1}$ and $\Lambda_{M_n^+}^{[\Gamma]} \leq (1 + 2\log D)(1 + \log n)$. (ii) follows now by combination of these estimates. Consider (iii). By employing $D\gamma_{M_n^+}^{-1} \geq \max_{1\leq m\leq M_n^+} \|[\Gamma]_{\underline{m}}^{-1}\|$, (iii) follows from $\gamma_1 = 1$ if $M_n^+ = 1$, while for $M_n^+ > 1$, we use $M_n^+\omega_{(M_n^+)}\gamma_{M_n^+}^{-1} \leq 4Dn(1 + \log n)^{-1}$. Consider (iv). Since ε and X are centred the identity $[\beta^m]_{\underline{m}} = [\Gamma]_{\underline{m}}^{-1}[g]_{\underline{m}}$ implies $\rho_m^2 \leq 2(\mathbb{E}Y^2 + \mathbb{E}|\langle\beta^m, X\rangle_{\mathbb{H}}|^2) = 2(\sigma_Y^2 + [g]_{\underline{m}}^t[\Gamma]_{\underline{m}}^{-1}[g]_{\underline{m}}) = \sigma_m^2$. By applying successively the inequality $\|\Gamma^{1/2}\beta\|^2 \leq d\|\beta\|_\gamma^2$ due to [22], Assumption 3.1, that is, γ and b^{-1} are nonincreasing, and the identity $\sigma_Y^2 = \sigma^2 + \langle\Gamma\beta, \beta\rangle_{\mathbb{H}}$ follows

$$\sigma_Y^2 \leq \sigma^2 + d\|\beta\|_\gamma^2 \leq \sigma^2 + dr. \tag{B.1}$$

Furthermore, from (iii) and (v) in Lemma B.1, we obtain

$$[g]^t_{\underline{m}}[\Gamma]^{-1}_{\underline{m}}[g]_{\underline{m}} \le d\|\beta^m\|^2_\gamma \le 34d^9 r, \tag{B.2}$$

which together with (B.1) implies (iv) and completes the proof. □

LEMMA B.3. *Let $\Gamma \in \mathcal{G}^d_\gamma$ with γ as in Assumption 3.1. For all $n, m \ge 1$ holds*

$$\left\{\frac{1}{4} < \frac{\|[\widehat{\Gamma}]^{-1}_{\underline{m}}\|_s}{\|[\Gamma]^{-1}_{\underline{m}}\|_s} \le 4, \forall 1 \le m \le M^\omega_n\right\} \subset \{M^-_n \le \widehat{M} \le M^+_n\}.$$

PROOF. Let $\widehat{\tau}_m := \|[\widehat{\Gamma}]^{-1}_{\underline{m}}\|^{-1}_s$ and $\tau_m := \|[\Gamma]^{-1}_{\underline{m}}\|^{-1}_s$. We use below without further reference that $D^{-1} \le \tau_m/\gamma_m \le d$ due to Lemma B.2(i). The result of the lemma follows by combination of the next two assertions,

$$\{\widehat{M} < M^-_n\} \subset \left\{\min_{1\le m\le M^\omega_n} \frac{\widehat{\tau}_m}{\tau_m} < \frac{1}{4}\right\}, \tag{B.3}$$

$$\{\widehat{M} > M^+_n\} \subset \left\{\max_{1\le m\le M^\omega_n} \frac{\widehat{\tau}_m}{\tau_m} \ge 4\right\}. \tag{B.4}$$

Consider (B.3) which holds trivially true for $M^-_n = 1$. If $M^-_n > 1$, then

$$\min_{1\le m\le M^-_n} \frac{\gamma_m}{m\omega_{(m)}} \ge \frac{4D(1+\log n)}{n}$$

implies

$$\min_{1\le m\le M^-_n} \frac{\tau_m}{m\omega_{(m)}} \ge \frac{4(1+\log n)}{n}$$

and

$$\begin{aligned} \{\widehat{M} < M^\omega_n\} \cap \{\widehat{M} < M^-_n\} &= \bigcup_{M=1}^{M^-_n-1} \{\widehat{M} = M\} \\ &\subset \bigcup_{M=1}^{M^-_n-1} \left\{\frac{\widehat{\tau}_{M+1}}{(M+1)\omega_{(M+1)}} < \frac{1+\log n}{n}\right\} \\ &\subset \left\{\min_{1\le m\le M^-_n} \frac{\widehat{\tau}_m}{\tau_m} < 1/4\right\}, \end{aligned}$$

while $\{\widehat{M} = M^\omega_n\} \cap \{\widehat{M} < M^-_n\} = \varnothing$ which shows (B.3) because $M^-_n \le M^\omega_n$.

Consider (B.4) which holds trivially true for $M_n^+ = M_n^\omega$. If $M_n^+ < M_n^\omega$, then $\frac{\tau_{M_n^+ +1}}{(M_n^+ +1)\omega_{(M_n^+ +1)}} < \frac{(1+\log n)}{4n}$ and (B.4) follows from

$$\begin{aligned}\{\widehat{M} > 1\} \cap \{\widehat{M} > M_n^+\} &= \bigcup_{M=M_n^+ +1}^{M_n^\omega} \{\widehat{M} = M\} \\ &\subset \bigcup_{M=M_n^+ +1}^{M_n^\omega} \left\{\min_{2\le m\le M} \frac{\widehat{\tau}_m}{m\omega_{(m)}} \ge \frac{1+\log n}{n}\right\} \\ &\subset \left\{\frac{\widehat{\tau}_{M_n^+ +1}}{\tau_{M_n^+ +1}} \ge 4\right\}\end{aligned}$$

and $\{\widehat{M} = 1\} \cap \{\widehat{M} > M_n^+\} = \varnothing$ which completes the proof. □

LEMMA B.4. *Let $\mathcal{A}_n$, $\mathcal{B}_n$ and $\mathcal{C}_n$ as in (A.1). For all $n \ge 1$ it holds true that* $\mathcal{A}_n \cap \mathcal{B}_n \cap \mathcal{C}_n \subset \{\mathrm{pen}_k \le \widehat{\mathrm{pen}}_k \le 72\mathrm{pen}_k, 1 \le k \le M_n^\omega\} \cap \{M_n^- \le \widehat{M} \le M_n^+\}$.

PROOF. Let $M_n^\omega \ge k \ge 1$. If $\|[\Xi]_{\underline{k}}\|_s \le 1/8$, that is, on the event $\mathcal{B}_n$, it is easily verified that $\|([\mathrm{Id}]_{\underline{k}} + [\Xi]_{\underline{k}})^{-1} - [\mathrm{Id}]_{\underline{k}}\|_s \le 1/7$ which we exploit to conclude

$$\begin{aligned}&\frac{6}{7} \le \frac{\|[\nabla_\omega]_{\underline{k}}^{1/2}[\widehat{\Gamma}]_{\underline{k}}^{-1}[\nabla_\omega]_{\underline{k}}^{1/2}\|_s}{\|[\nabla_\omega]_{\underline{k}}^{1/2}[\Gamma]_{\underline{k}}^{-1}[\nabla_\omega]_{\underline{k}}^{1/2}\|_s} \le \frac{8}{7}, \qquad \frac{6}{7} \le \frac{\|[\widehat{\Gamma}]_{\underline{k}}^{-1}\|_s}{\|[\Gamma]_{\underline{k}}^{-1}\|_s} \le \frac{8}{7} \quad \text{and} \\ &6x^t[\Gamma]_{\underline{k}}^{-1}x \le 7x^t[\widehat{\Gamma}]_{\underline{k}}^{-1}x \le 8x^t[\Gamma]_{\underline{k}}^{-1}x \qquad \text{for all } x \in \mathbb{R}^k\end{aligned} \tag{B.5}$$

and, consequently

$$(6/7)[\widehat{g}]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[\widehat{g}]_{\underline{k}} \le [\widehat{g}]_{\underline{k}}^t[\widehat{\Gamma}]_{\underline{k}}^{-1}[\widehat{g}]_{\underline{k}} \le (8/7)[\widehat{g}]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[\widehat{g}]_{\underline{k}}. \tag{B.6}$$

Moreover, from $\|[\Xi]_{\underline{k}}\|_s \le 1/8$ we obtain after some algebra,

$$\begin{aligned}&[g]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[g]_{\underline{k}} \le (1/16)[g]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[g]_{\underline{k}} + 4[W]_{\underline{k}}[\Gamma]_{\underline{k}}^{-1}[W]_{\underline{k}} + 2[\widehat{g}]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[\widehat{g}]_{\underline{k}}, \\ &[\widehat{g}]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[\widehat{g}]_{\underline{k}} \le (33/16)[g]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[g]_{\underline{k}} + 4[W]_{\underline{k}}[\Gamma]_{\underline{k}}^{-1}[W]_{\underline{k}}.\end{aligned}$$

Combining each of these estimates with (B.6) yields

$$\begin{aligned}&(15/16)[g]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[g]_{\underline{k}} \le 4[W]_{\underline{k}}[\Gamma]_{\underline{k}}^{-1}[W]_{\underline{k}} + (7/3)[\widehat{g}]_{\underline{k}}^t[\widehat{\Gamma}]_{\underline{k}}^{-1}[\widehat{g}]_{\underline{k}}, \\ &(7/8)[\widehat{g}]_{\underline{k}}^t[\widehat{\Gamma}]_{\underline{k}}^{-1}[\widehat{g}]_{\underline{k}} \le (33/16)[g]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[g]_{\underline{k}} + 4[W]_{\underline{k}}[\Gamma]_{\underline{k}}^{-1}[W]_{\underline{k}}.\end{aligned}$$

If in addition $[W]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[W]_{\underline{k}} \le \frac{1}{8}([g]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[g]_{\underline{k}} + \sigma_Y^2)$, that is, on the event $\mathcal{C}_n$, then the last two estimates imply, respectively,

$$\begin{aligned}&(7/16)\big([g]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[g]_{\underline{k}} + \sigma_Y^2\big) \le (15/16)\sigma_Y^2 + (7/3)[\widehat{g}]_{\underline{k}}^t[\widehat{\Gamma}]_{\underline{k}}^{-1}[\widehat{g}]_{\underline{k}}, \\ &(7/8)[\widehat{g}]_{\underline{k}}^t[\widehat{\Gamma}]_{\underline{k}}^{-1}[\widehat{g}]_{\underline{k}} \le (41/16)[g]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[g]_{\underline{k}} + (1/2)\sigma_Y^2\end{aligned}$$

and hence in case $1/2 \le \widehat{\sigma}_Y^2/\sigma_Y^2 \le 3/2$, that is, on the event $\mathcal{A}_n$, we obtain

$$(7/16)\big([g]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[g]_{\underline{k}} + \sigma_Y^2\big) \le (15/8)\widehat{\sigma}_Y^2 + (7/3)[\widehat{g}]_{\underline{k}}^t[\widehat{\Gamma}]_{\underline{k}}^{-1}[\widehat{g}]_{\underline{k}},$$
$$(7/8)\big([\widehat{g}]_{\underline{k}}^t[\widehat{\Gamma}]_{\underline{k}}^{-1}[\widehat{g}]_{\underline{k}} + \widehat{\sigma}_Y^2\big) \le (41/16)[g]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[g]_{\underline{k}} + (29/16)\sigma_Y^2.$$

Combining the last two estimates we have

$$\tfrac{1}{6}\big(2[g]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[g]_{\underline{k}} + 2\sigma_Y^2\big) \le \big(2[\widehat{g}]_{\underline{k}}^t[\widehat{\Gamma}]_{\underline{k}}^{-1}[\widehat{g}]_{\underline{k}} + 2\widehat{\sigma}_Y^2\big) \le 3\big(2[g]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[g]_{\underline{k}} + 2\sigma_Y^2\big).$$

On $\mathcal{A}_n \cap \mathcal{B}_n \cap \mathcal{C}_n$ the last estimate and (B.5) hold for all $1 \le k \le M_n^\omega$, hence

$$\mathcal{A}_n \cap \mathcal{B}_n \cap \mathcal{C}_n \subset \left\{\frac{1}{6} \le \frac{\widehat{\sigma}_m^2}{\sigma_m^2} \le 3 \text{ and } \frac{6}{7} \le \frac{\Delta_m^{[\widehat{\Gamma}]}}{\Delta_m^{[\Gamma]}} \le \frac{8}{7}, \forall 1 \le m \le M_n^\omega\right\}.$$

Moreover it is easily seen that $(6/7) \le \Delta_m^{[\widehat{\Gamma}]}/\Delta_m^{[\Gamma]} \le (8/7)$ implies

$$1/2 \le \big(1 + \log(7/6)\big)^{-1} \le \Lambda_m^{[\widehat{\Gamma}]}/\Lambda_m^{[\Gamma]} \le \big(1 + \log(8/7)\big) \le 3/2.$$

Due to the last estimates the definitions of pen_m and $\widehat{\mathrm{pen}}_m$ imply

$$\mathcal{A}_n \cap \mathcal{B}_n \cap \mathcal{C}_n \subset \{\mathrm{pen}_m \le \widehat{\mathrm{pen}}_m \le 72\mathrm{pen}_m, \forall 1 \le m \le M_n^\omega\}. \tag{B.7}$$

On the other hand, by exploiting successively (B.5) and Lemma B.3, we have

$$\mathcal{A}_n \cap \mathcal{B}_n \cap \mathcal{C}_n \subset \left\{\frac{6}{7} \le \frac{\|[\widehat{\Gamma}]_{\underline{m}}^{-1}\|_s}{\|[\Gamma]_{\underline{m}}^{-1}\|_s} \le \frac{8}{7}, \forall 1 \le m \le M_n^\omega\right\} \subset \{M_n^- \le \widehat{M} \le M_n^+\}.$$

The last display and (B.7) imply the assertion of the lemma. □

LEMMA B.5. *For all $m, n \ge 1$ with $n \ge (8/7)\|[\Gamma]_{\underline{m}}^{-1}\|_s$ we have $\mho_{m,n} \subset \Omega_{m,n}$.*

PROOF. Taking into account $[\widehat{\Gamma}]_{\underline{m}} = [\Gamma]_{\underline{m}}^{1/2}\{[\mathrm{Id}]_{\underline{m}} + [\Xi]_{\underline{m}}\}[\Gamma]_{\underline{m}}^{1/2}$ observe that $\|[\Xi]_{\underline{m}}\|_s \le 1/8$ and $n \ge (8/7)\|[\Gamma]_{\underline{m}}^{-1}\|_s$ imply $\|[\widehat{\Gamma}]_{\underline{m}}^{-1}\|_s \le n$ due to a Neumann series argument. Hence, $\mho_{m,n} \subset \Omega_{m,n}$ which proves the lemma. □

APPENDIX C: PROOF OF PROPOSITION 3.1

We will suppose throughout this section that the conditions of Proposition 3.1 are satisfied which allow us to employ Lemmas B.1–B.5. First, we show technical assertions (Lemmas C.1–C.5) exploiting Assumption 3.2, that is, X and ε are jointly normally distributed. They are used below to prove that Assumptions 2.1 and 2.2 are satisfied (Propositions C.6 and C.7, resp.), which is the claim of Proposition 3.1.

We begin by recalling elementary properties due to Assumption 3.2 which are frequently used in this section. Given $f \in \mathbb{H}$ the random variable $\langle f, X\rangle_{\mathbb{H}}$ is normally distributed with mean zero and variance $\langle \Gamma f, f\rangle_{\mathbb{H}}$. Consider the Galerkin solution β^m and $h \in \mathbb{H}_m$; then $\langle \beta - \beta^m, X\rangle_{\mathbb{H}}$ and $\langle h, X\rangle_{\mathbb{H}}$ are independent. Thereby, $Z_m = Y - \langle \beta^m, X\rangle_{\mathbb{H}} = \sigma\varepsilon + \langle \beta - \beta^m, X\rangle_{\mathbb{H}}$ and $[X]_{\underline{m}}$ are independent, normally distributed with mean zero and, respectively, variance ρ_m^2 and covariance matrix $[\Gamma]_{\underline{m}}$. Consequently, $(\rho_m^{-1} Z_m, [X]_{\underline{m}}^t [\Gamma]_{\underline{m}}^{-1/2})$ is a vector with independent, standard normally distributed entries. The next assertion states elementary inequalities for Gaussian random variables and its straightforward proof is omitted.

LEMMA C.1. *Let $\{U_i, V_{ij}, 1 \le i \le n, 1 \le j \le m\}$ be independent and standard normally distributed. For all $\eta > 0$ and $\zeta \ge 4m/n$ we have*:

(i) $P(n^{-1/2}\sum_{i=1}^n (U_i^2 - 1) \ge \eta) \le \exp(-\frac{1}{8}\frac{\eta^2}{1+\eta n^{-1/2}})$;

(ii) $P(n^{-1}|\sum_{i=1}^n U_i V_{i1}| \ge \eta) \le \frac{\eta n^{1/2}+1}{\eta n^{1/2}} \exp(-\frac{n}{4}\min\{\eta^2, 1/4\})$;

(iii) $P(n^{-2}\sum_{j=1}^m |\sum_{i=1}^n U_i V_{ij}|^2 \ge \zeta) \le \exp(\frac{-n}{16}) + \exp(\frac{-\zeta n}{64})$;

and for all $c \ge 1$ and $a_1, \dots, a_m \ge 0$ we obtain:

(iv) $\mathbb{E}(\sum_{i=1}^n U_i^2 - 2cn)_+ \le 16\exp(\frac{-cn}{16})$;

(v) $\mathbb{E}(\sum_{j=1}^m |n^{-1/2}\sum_{i=1}^n U_i V_{ij}|^2 - 4cm)_+ \le 16\exp(\frac{-cm}{16}) + 32\frac{cm}{n}\exp(\frac{-n}{16})$;

(vi) $\mathbb{E}(\sum_{j=1}^m a_j |\sum_{i=1}^n U_i V_{ij}|^2)^2 = n(n+2)(\sum_{j=1}^m a_j^2 + (\sum_{j=1}^m a_j)^2)$.

LEMMA C.2. *For all $n, m \ge 1$ we have*:

(i) $n^2 \rho_m^{-4}\mathbb{E}\|[W]_{\underline{m}}\|^4 \le 6(\mathbb{E}\|X\|^2)^2$.

Furthermore, there exist a numerical constant $C > 0$ such that for all $n \ge 1$:

(ii) $n^8 \max_{1\le m \le \lfloor n^{1/4}\rfloor} P(\frac{[W]_{\underline{m}}^t [\Gamma]_{\underline{m}}^{-1}[W]_{\underline{m}}}{\rho_m^2} > \frac{1}{16}) \le C$;

(iii) $n^8 \max_{1\le m \le \lfloor n^{1/4}\rfloor} P(\|[\Xi]_{\underline{m}}\|_s > 1/8) \le C$;

(iv) $n^7 P(\{1/2 \le \widehat{\sigma}_Y^2/\sigma_Y^2 \le 3/2\}^c) \le C$.

PROOF. Denote by $(\lambda_j, e_j)_{1\le j\le m}$ an eigenvalue decomposition of $[\Gamma]_{\underline{m}}$. Define $U_i := (\sigma\varepsilon_i + \langle \beta - \beta^m, X_i\rangle_{\mathbb{H}})/\rho_m$ and $V_{ij} := (\lambda_j^{-1/2} e_j^t [X_i]_{\underline{m}})$, $1 \le i \le n$, $1 \le j \le m$, where $U_1, \dots, U_n, V_{11}, \dots, V_{nm}$ are independent and standard normally distributed. Consider (i) and (ii). Taking into account $\sum_{j=1}^m \lambda_j \le \mathbb{E}\|X\|_{\mathbb{H}}^2$ and the identities $n^4 \rho_m^{-4}\|[W]_{\underline{m}}\|^4 = (\sum_{j=1}^m \lambda_j (\sum_{i=1}^n U_i V_{ij})^2)^2$ and $([W]_{\underline{m}}^t [\Gamma]_{\underline{m}}^{-1}[W]_{\underline{m}})/\rho_m^2 = n^{-2}\sum_{j=1}^m (\sum_{i=1}^n U_i V_{ij})^2$, assertions (i) and (ii) follow, respectively, from in Lemma C.1(vi) and (iii) (with $a_j = \lambda_j$). Consider (iii). Since $n\|[\Xi]_{\underline{m}}\|_s \le m \max_{1\le j,l\le m} |\sum_{i=1}^n (V_{ij}V_{il} - \delta_{jl})|$ we obtain due to (i) and (ii) in Lemma C.1

that for all $\eta > 0$

$$
\begin{aligned}
&P(\|[\Xi]_{\underline{m}}\|_s \geq \eta) \\
&\quad \leq \sum_{1\leq j,l\leq m} P\left(\left|n^{-1}\sum_{i=1}^n (V_{ij}V_{il}-\delta_{jl})\right| \geq \eta/m\right) \\
&\quad \leq m^2 \max\left\{P\left(\left|\frac{1}{n}\sum_{i=1}^n V_{i1}V_{i2}\right| \geq \frac{\eta}{m}\right), P\left(\left|\frac{1}{n^{1/2}}\sum_{i=1}^n (V_{i1}^2-1)\right| \geq n^{1/2}\frac{\eta}{m}\right)\right\} \\
&\quad \leq m^2 \max\left\{\left(1+\frac{m}{\eta n^{1/2}}\right)\exp\left(-\frac{n}{4}\min\left\{\frac{\eta^2}{m^2},\frac{1}{4}\right\}\right), 2\exp\left(-\frac{1}{8}\frac{n\eta^2/m^2}{1+\eta/m}\right)\right\}.
\end{aligned}
$$

Keeping in mind that $1/8 = \eta \leq m/2$, the last bound implies (iii). Consider (iv). Since $\{Y_i/\sigma_Y\}_{i=1}^n$ are independent, standard, normally distributed and $\{1/2 \leq \widehat{\sigma}_Y^2/\sigma_Y^2 \leq 3/2\}^c \subset \{|n^{-1}\sum_{i=1}^n Y_i^2/\sigma_Y^2 - 1| > 1/2\}$, (iv) follows from Lemma C.1(i). □

LEMMA C.3. *We have for all $c \geq 1$ and $n, m \geq 1$*

$$
\mathbb{E}\left(\frac{n[W]_{\underline{m}}^t[\Gamma]_{\underline{m}}^{-1}[W]_{\underline{m}}}{\rho_m^2} - 4cm\right)_+ \leq 16\exp\left(\frac{-cm}{16}\right) + 32\frac{cm}{n}\exp\left(\frac{-n}{16}\right).
$$

PROOF. From $n\|[\Gamma]_{\underline{m}}^{-1/2}[W]_{\underline{m}}\|^2\rho_m^{-2} = \sum_{j=1}^m (n^{-1/2}\sum_{i=1}^n U_iV_{ij})^2$ derived in the proof of Lemma C.2 and Lemma C.1(v) follows the assertion. □

LEMMA C.4. *There is a constant $C(d)$ depending on d such that for all $n \geq 1$,*

$$
\sup_{\beta\in\mathcal{F}_b^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\sum_{k=m_n^\diamond}^{M_n^+}\Delta_k^{[\Gamma]}\mathbb{E}\left([W]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[W]_{\underline{k}} - 4\sigma_k^2\frac{k\Lambda_k^{[\Gamma]}}{n}\right)_+ \leq C(d)(\sigma^2+r)\Sigma n^{-1}.
$$

PROOF. The key argument of the proof is Lemma C.3 with $c = \Lambda_k^{[\Gamma]}$. Taking into account this bound and for all $\beta \in \mathcal{F}_\beta^r$ and $\Gamma \in \mathcal{G}_\gamma^d$ that $\Delta_k^{[\Gamma]} \leq 4d^3\Delta_k^\gamma$, $(1+\log d)^{-1}\Lambda_k^\gamma \leq \Lambda_k^{[\Gamma]}$, $\delta_{M_n^+}^{[\Gamma]} \leq nCd^6(1+\log d)$ and $\rho_k^2 \leq \sigma_k^2 \leq 2(\sigma^2 + 35d^6r)$ [Lemma B.2(i), (ii) and (iv), resp.] hold true, we obtain

$$
\begin{aligned}
&\sum_{k=m_n^\diamond}^{M_n^+}\Delta_k^{[\Gamma]}\mathbb{E}\left([W]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[W]_{\underline{k}} - 4\sigma_k^2\frac{k\Lambda_k^{[\Gamma]}}{n}\right)_+ \\
&\qquad \leq \sum_{k=1}^{M_n^+}\frac{\sigma_k^2\Delta_k^{[\Gamma]}}{n}\mathbb{E}\left(\frac{n[W]_{\underline{k}}^t[\Gamma]_{\underline{k}}^{-1}[W]_{\underline{k}}}{\rho_k^2} - 4k\Lambda_k^{[\Gamma]}\right)_+
\end{aligned}
$$

$$\leq C(d)(\sigma^2 + r) \\ \times n^{-1}\left\{\sum_{k=1}^{M_n^+} \Delta_k^\gamma \exp\left(-\frac{k\Lambda_k^\gamma}{16(1+\log d)}\right) + M_n^+ \exp(-n/16)\right\}.$$

Finally, exploiting the constant Σ satisfying (3.3) and $M_n^+ \exp(-n/16) \leq C$ for all $n \geq 1$, we obtain the assertion of the lemma. □

LEMMA C.5. *There exist a numerical constant C and a constant $C(d)$ only depending on d such that for all $n \geq 1$, we have*:

(i) $\sup_{\beta\in\mathcal{F}_b^r} \sup_{\Gamma\in\mathcal{G}_\gamma^d} \{n^6 (M_n^+)^2 \max_{1\leq m\leq M_n^+} P(\mho_{m,n}^c)\} \leq C$;
(ii) $\sup_{\beta\in\mathcal{F}_b^r} \sup_{\Gamma\in\mathcal{G}_\gamma^d} \{n M_n^+ \max_{1\leq m\leq M_n^+} P(\Omega_{m,n}^c)\} \leq C(d)$;
(iii) $\sup_{\beta\in\mathcal{F}_b^r} \sup_{\Gamma\in\mathcal{G}_\gamma^d} \{n^7 P(\mathcal{E}_n^c)\} \leq C$.

PROOF. Since $M_n^+ \leq \lfloor n^{1/4}\rfloor$ and $\mho_{m,n}^c = \{\|[\Xi]_{\underline{m}}\| > 1/8\}$ assertion (i) follows from Lemma C.2(ii). Consider (ii). Let $n_o := n_o(d) := \exp(128d^6) \geq 8d^3$, and consequently $\omega_{(M_n^+)}(M_n^+ \log n) \geq 128d^6$ for all $n \geq n_o$. We distinguish in the following the cases $n < n_o$ and $n \geq n_o$. First, let $1 \leq n \leq n_o$. Obviously, $M_n^+ \max_{1\leq m\leq M_n^+} P(\Omega_{m,n}^c) \leq M_n^+ \leq n^{-1} n_o^{5/4} \leq C(d) n^{-1}$ since $M_n^+ \leq n^{1/4}$ and n_o depends on d only. On the other hand, if $n \geq n_o$, then Lemma B.2(iii) implies $n \geq 2 \max_{1\leq m\leq M_n^+} \|[\Gamma]_{\underline{m}}^{-1}\|$, and hence $\mho_{m,n} \subset \Omega_{m,n}$ for all $1 \leq m \leq M_n^+$ by employing Lemma B.5. From (i) we conclude $M_n^+ \max_{1\leq m\leq M_n^+} P(\Omega_{m,n}^c) \leq M_n^+ \max_{1\leq m\leq M_n^+} P(\mho_{m,n}^c) \leq C n^{-3}$. By combination of the two cases we obtain (ii). It remains to show (iii). Consider the events $\mathcal{A}_n$, $\mathcal{B}_n$ and $\mathcal{C}_n$ given in (A.1), where $\mathcal{A}_n \cap \mathcal{B}_n \cap \mathcal{C}_n \subset \mathcal{E}_n$ due to Lemma B.4. We have $n^7 P(\mathcal{A}_n^c) \leq C$, $n^7 P(\mathcal{B}_n^c) \leq C$, $n^7 P(\mathcal{C}_n^c) \leq C$ due to Lemma C.2(iv), (iii), (ii), respectively (keep in mind $\lfloor n^{1/4}\rfloor \geq M_n^\omega$ and $2(\sigma_Y^2 + [g]_{\underline{k}}^t [\Gamma]_{\underline{k}}^{-1} [g]_{\underline{k}}) = \sigma_k^2 \geq \rho_k^2$). Combining these estimates implies (iii). □

PROPOSITION C.6. *Let $\kappa = 96$ in definition* (2.11) *of the penalty* pen. *There exists a constant $C(d)$ only depending on d such that for all $n \geq 1$,*

$$\sup_{\beta\in\mathcal{F}_b^r} \sup_{\Gamma\in\mathcal{G}_\gamma^d} \mathbb{E}\left\{\sup_{m_n^\diamond\leq k\leq M_n^+} \left(\|\widehat{\beta}_k - \beta^k\|_\omega^2 - \frac{1}{6}\text{pen}_k\right)_+\right\} \leq C(d)(\sigma^2 + r)\Sigma n^{-1}.$$

PROOF. Since $[\widehat{\beta}_k - \beta^k]_{\underline{k}} = [\widehat{\Gamma}]_{\underline{k}}^{-1}[W]_{\underline{k}} \mathbb{1}_{\Omega_{k,n}} - [\beta^k]_{\underline{k}} \mathbb{1}_{\Omega_{k,n}^c}$ it follows

$$\|\widehat{\beta}_k - \beta^k\|_\omega^2 = \|[\nabla_\omega]_{\underline{k}}^{1/2} [\widehat{\Gamma}]_{\underline{k}}^{-1} [W]_{\underline{k}}\|^2 \mathbb{1}_{\Omega_{k,n}} + \|\beta^k\|_\omega^2 \mathbb{1}_{\Omega_{k,n}^c}. \tag{C.1}$$

Exploiting $\|([\mathrm{Id}]_{\underline{k}} + [\Xi]_{\underline{k}})^{-1}\|_s \mathbb{1}_{\mho_{k,n}} \leq 2$, $[\widehat{\Gamma}]_{\underline{k}} = [\Gamma]_{\underline{k}}^{1/2}\{[\mathrm{Id}]_{\underline{k}} + [\Xi]_{\underline{k}}\}[\Gamma]_{\underline{k}}^{1/2}$ and the definition of $\Delta_k^{[\Gamma]}$ imply $\|[\nabla_\omega]_{\underline{k}}^{1/2} [\widehat{\Gamma}]_{\underline{k}}^{-1} [W]_{\underline{k}}\|^2 \mathbb{1}_{\mho_{k,n}} \leq 4\Delta_k^{[\Gamma]} \|[\Gamma]_{\underline{k}}^{-1/2} [W]_{\underline{k}}\|^2$.

On the other hand, we have $\|[\nabla_\omega]_{\underline{k}}^{1/2}[\widehat{\Gamma}]_{\underline{k}}^{-1}[W]_{\underline{k}}\|^2\mathbb{1}_{\Omega_{k,n}} \leq \omega_{(k)}n^2\|[W]_{\underline{k}}\|^2$. From these estimates and $\|\beta^k\|_\omega \leq \|\beta^k\|_b$ (ωb^{-1} is nonincreasing due to Assumption 3.1) we deduce for all $k \geq 1$,

$$\|\widehat{\beta}_k - \beta^k\|_\omega^2 \leq 4\Delta_k^{[\Gamma]}\|[\Gamma]_{\underline{k}}^{-1/2}[W]_{\underline{k}}\|^2 + \omega_{(k)}n^2\|[W]_{\underline{k}}\|^2\mathbb{1}_{\mho_{k,n}^c} + \|\beta^k\|_b^2\mathbb{1}_{\Omega_{k,n}^c}.$$

This upper bound and $\mathrm{pen}_k = 96\sigma_k^2 k\Delta_k^{[\Gamma]}\Lambda_k^{[\Gamma]}n^{-1}$ imply

$$\begin{aligned}
&\mathbb{E}\Big\{\sup_{m_n^\diamond \leq k \leq M_n^+}\Big(\|\widehat{\beta}_k - \beta^k\|_\omega^2 - \frac{\mathrm{pen}_k}{6}\Big)_+\Big\}\\
&\quad\leq \sum_{k=m_n^\diamond}^{M_n^+} n^3(\mathbb{E}\|[W]_{\underline{k}}\|^4)^{1/2}(P(\mho_{k,n}^c))^{1/2} + \sum_{k=m_n^\diamond}^{M_n^+}\|\beta^k\|_b^2 P(\Omega_{k,n}^c)\\
&\qquad + 4\sum_{k=m_n^\diamond}^{M_n^+}\Delta_k^{[\Gamma]}\mathbb{E}\Big(\|[\Gamma]_{\underline{k}}^{-1/2}[W]_{\underline{k}}\|^2 - 4\sigma_k^2\frac{k\Lambda_k^{[\Gamma]}}{n}\Big)_+.
\end{aligned}$$

By exploiting Lemmas B.1(v) and C.2(i) together with $\rho_m^2 \leq 2(\sigma^2 + 35d^6r)$ [Lemma B.2(iv)] the first and second r.h.s. term are bounded by

$$\begin{aligned}
&6(\sigma^2 + 35d^6r)\mathbb{E}\|X\|^2n^2M_n^+\max_{m_n^\diamond \leq k \leq M_n^+}(P(\mho_{k,n}^c))^{1/2}\\
&\qquad + 34d^8rM_n^+\max_{m_n^\diamond \leq k \leq M_n^+}P(\Omega_{k,n}^c).
\end{aligned}$$

Combining this upper bound, the property $\mathbb{E}\|X\|^2 \leq d\sum_{j\geq 1}\gamma_j \leq d\Sigma$ and the estimates given in Lemma C.5, we deduce for all $\beta \in \mathcal{F}_b^r$ and $\Gamma \in \mathcal{G}_\gamma^d$ that

$$\begin{aligned}
&\sup_{\beta\in\mathcal{F}_b^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\mathbb{E}\Big\{\sup_{m_n^\diamond \leq k \leq M_n^+}\Big(\|\widehat{\beta}_k - \beta^k\|_\omega^2 - \frac{1}{6}\mathrm{pen}_k\Big)_+\Big\}\\
&\quad\leq C(d)(\sigma^2 + r)\Sigma n^{-1}\\
&\qquad + 4\sup_{\beta\in\mathcal{F}_b^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\sum_{k=m_n^\diamond}^{M_n^+}\Delta_k^{[\Gamma]}\mathbb{E}\Big(\|[\Gamma]_{\underline{k}}^{-1/2}[W]_{\underline{k}}\|^2 - 4\sigma_k^2\frac{k\Lambda_k^{[\Gamma]}}{n}\Big)_+.
\end{aligned}$$

The result of the proposition follows now by replacing the last r.h.s. term by its upper bound given in Lemma C.4, which completes the proof. □

PROPOSITION C.7. *Let $\kappa = 96$ in definition (2.11) and (2.12) of* pen *and* $\widehat{\mathrm{pen}}$. *There exists a constant $C(d)$ only depending on d such that for all $n \geq 1$*

$$\sup_{\beta\in\mathcal{F}_b^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\mathbb{E}(\|\widehat{\beta}_{\widehat{m}} - \beta\|_\omega^2\mathbb{1}_{\mathcal{E}_n^c}) \leq C(d)(\sigma^2 + r)\Sigma n^{-1}.$$

PROOF. From the decomposition (C.1) and $\|[\nabla_\omega]_{\underline{k}}^{1/2}[\widehat{\Gamma}]_{\underline{k}}^{-1}[W]_{\underline{k}}\|^2\mathbb{1}_{\Omega_{k,n}} \leq \Delta_k^\omega n^2\|[W]_{\underline{k}}\|^2$ given in the proof of Proposition C.6 we conclude

$$\|\widehat{\beta}_k - \beta\|_\omega^2 \leq 2\Delta_k^\omega n^2 \|[W]_{\underline{k}}\|^2 + 2\|\beta^k\|_\omega^2 + 2\|\beta\|_\omega^2 \qquad \text{for all } k \geq 1.$$

By exploiting Lemma B.1(v) together with $\|\beta^k\|_\omega \leq \|\beta^k\|_b$ (ωb^{-1} is nonincreasing due to Assumption 3.1) we obtain for all $\beta \in \mathcal{F}_b^r$ and $\Gamma \in \mathcal{G}_\gamma^d$

$$\|\widehat{\beta}_k - \beta\|_\omega^2 \leq 2\Delta_k^\omega n^2 \|[W]_{\underline{k}}\|^2 + 2(34d^8 r + r) \qquad \text{for all } k \geq 1.$$

Since $1 \leq \widehat{m} \leq M_n^\omega$ and $\max_{1\leq k\leq M_n^\omega} \omega_{(k)} \leq n$ it follows that

$$\mathbb{E}(\|\widehat{\beta}_{\widehat{m}} - \beta\|_\omega^2 \mathbb{1}_{\mathcal{E}_n^c}) \leq 2n^3 M_n^\omega \max_{1\leq k\leq M_n^\omega} (\mathbb{E}\|[W]_{\underline{k}}\|^4)^{1/2} |P(\mathcal{E}_n^c)|^{1/2} + 70d^8 r M_n^\omega P(\mathcal{E}_n^c).$$

From Lemma C.2(i) together with $\rho_m^2 \leq 2(\sigma^2 + 35d^6 r)$ (Lemma B.2) and $\mathbb{E}\|X\|^2 \leq d\Sigma$, we conclude for all $\beta \in \mathcal{F}_b^r$ and $\Gamma \in \mathcal{G}_\gamma^d$ that

$$\mathbb{E}(\|\widehat{\beta}_{\widehat{m}} - \beta\|_\omega^2 \mathbb{1}_{\mathcal{E}_n^c}) \leq 12(\sigma^2 + 35d^6 r)\, d\Sigma n^2 M_n^\omega |P(\mathcal{E}_n^c)|^{1/2} + 70d^8 r M_n^\omega P(\mathcal{E}_n^c).$$

The result of the proposition follows now from $M_n^\omega \leq \lfloor n^{1/4} \rfloor$ and by replacing the probability $P(\mathcal{E}_n^c)$ by its upper bound Cn^{-7} given in Lemma C.5. □

PROOF OF PROPOSITION 3.1. The assertion follows from Propositions C.6 and C.7, and we omit the details. □

APPENDIX D: PROOF OF PROPOSITION 3.3

We assume throughout this section that the conditions of Proposition 3.3 are satisfied which allows us to employ Lemmas B.1–B.5. We formulate first preliminary results (Proposition D.1 and Lemmas D.2–D.5) relying on the moment conditions (Assumption 3.3). They are used to prove that Assumptions 2.1 and 2.2 are satisfied (Propositions D.6 and D.7, resp.), which is the claim of Proposition 3.3. We begin by gathering elementary bounds due to Assumption 3.3. Let k be given by Assumption 3.3; then for all $m \geq 1$ we have

$$\mathbb{E}|Z_m|^{4k} \leq \rho_m^2 \eta^{4k}, \qquad \mathbb{E}|Y|^{4k} \leq \sigma_Y^{4k}\eta^{4k},$$
$$\max_{1\leq j\leq m} \mathbb{E}|([\Gamma]_{\underline{m}}^{-1/2}[X]_{\underline{m}})_j|^{4k} \leq \eta^{4k},$$
$$\mathbb{E}|\langle \beta - \beta^m, X\rangle_{\mathbb{H}}|^{4k} \leq \|\Gamma^{1/2}(\beta^m - \beta)\|_{\mathbb{H}}^{4k}\eta^{4k},$$
$$\mathbb{E}|[X]_{\underline{m}}^t[\Gamma]_{\underline{m}}^{-1}[X]_{\underline{m}}|^{2k} \leq m^{2k}\eta^{4k}.$$

From $\mathbb{E}|V|\mathbb{1}_{\{|V|\geq t\}} \leq t^{-k+1}\mathbb{E}|V|^k$, $t > 0$, under Assumption 3.3 follows

$$\mathbb{E}\varepsilon^2 \mathbb{1}_{\{|\varepsilon|>n^{1/6}\}} \leq \frac{\eta^{32}}{n^5},$$

$$\begin{aligned}\mathbb{E}|\langle\beta - \beta^m, X\rangle_{\mathbb{H}}|^2 \mathbb{1}_{\{|\langle\beta-\beta^m, X\rangle_{\mathbb{H}}|>\|\Gamma^{1/2}(\beta^m-\beta)\|_{\mathbb{H}} n^{1/6}\}} \\ \leq \frac{\eta^{32}}{n^5}\|\Gamma^{1/2}(\beta^m - \beta)\|_{\mathbb{H}}^2,\end{aligned} \tag{D.1}$$

$$\mathbb{E}|[X]_{\underline{m}}^t [\Gamma]_{\underline{m}}^{-1} [X]_{\underline{m}}|^2 \mathbb{1}_{\{[X]_{\underline{m}}^t [\Gamma]_{\underline{m}}^{-1} [X]_{\underline{m}} > mn^{1/3}\}} \leq \frac{\eta^{32}}{n^{14/3}} m^2$$

for all $m, n \geq 1$, and by employing Markov's inequality

$$\begin{aligned}P(|\varepsilon| > n^{1/6}) \leq \frac{\eta^{32}}{n^{16/3}}, \\ P(|\langle\beta - \beta^m, X\rangle_{\mathbb{H}}| > \|\Gamma^{1/2}(\beta^m - \beta)\|_{\mathbb{H}} n^{1/6}) \leq \frac{\eta^{32}}{n^{16/3}}.\end{aligned} \tag{D.2}$$

We exploit these bounds in the following proofs. The key argument used in the proof of Lemma D.3 is the following inequality due to [37]; see, for example, [26].

PROPOSITION D.1 (Talagrand's inequality). *Let $T_1, \ldots, T_n$ be independent $\mathcal{T}$-valued random variables and $\nu_s^* = (1/n)\sum_{i=1}^n [\nu_s(T_i) - \mathbb{E}[\nu_s(T_i)]]$, for ν_s belonging to a countable class $\{\nu_s : s \in \mathbb{S}\}$ of measurable functions. Then, for $\varepsilon > 0$,*

$$\begin{aligned}\mathbb{E}\Big(\sup_{s\in\mathbb{S}}|\nu_s^*|^2 - 2(1+2\varepsilon)H^2\Big)_+ \\ \leq C\bigg(\frac{v}{n}\exp\Big(-K_1\varepsilon\frac{nH^2}{v}\Big) + \frac{h^2}{n^2C^2(\varepsilon)}\exp\Big(-K_2 C(\varepsilon)\sqrt{\varepsilon}\frac{nH}{h}\Big)\bigg)\end{aligned}$$

with $K_1 = 1/6$, $K_2 = 1/(21\sqrt{2})$, $C(\varepsilon) = \sqrt{1+\varepsilon} - 1$ and C a universal constant and where

$$\sup_{s\in\mathbb{S}}\sup_{t\in\mathcal{T}}|\nu_s(t)| \leq h, \qquad \mathbb{E}\Big[\sup_{s\in\mathbb{S}}|\nu_s^*|\Big] \leq H, \qquad \sup_{s\in\mathbb{S}}\frac{1}{n}\sum_{i=1}^n \mathbb{V}\mathrm{ar}(\nu_s(T_i)) \leq v.$$

LEMMA D.2. *There exist a numerical constant $C > 0$ such that for all $n \geq 1$:*

(i) $n^2 \sup_{m\geq 1} \rho_m^{-4}\mathbb{E}\|[W]_{\underline{m}}\|^4 \leq C\eta^8(\mathbb{E}\|X\|_{\mathbb{H}}^2)^2$;

(ii) $n^8 \max_{1\leq m\leq\lfloor n^{1/4}\rfloor} P(\frac{[W]_{\underline{m}}^t[\Gamma]_{\underline{m}}^{-1}[W]_{\underline{m}}}{\rho_m^2} > \frac{1}{16}) \leq C\eta^{64}$;

(iii) $n^8 \max_{1\leq m\leq\lfloor n^{1/4}\rfloor} P(\|[\Xi]_{\underline{m}}\|_s > 1/8) \leq C(\eta)$;

(iv) $n^7 P(\{1/2 \leq \widehat{\sigma}_Y^2/\sigma_Y^2 \leq 3/2\}^c) \leq C\eta^{64}$.

PROOF. Denote by $(\lambda_j, e_j)_{1\le j\le m}$ an eigenvalue decomposition of $[\Gamma]_{\underline{m}}$. Define $U_i := (\sigma\varepsilon_i + \langle\beta - \beta^m, X_i\rangle_{\mathbb{H}})/\rho_m$ and $V_{ij} := (\lambda_j^{-1/2} e_j^t [X_i]_{\underline{m}})$, $1 \le i \le n$, $1 \le j \le m$. Keep in mind that $\mathbb{E}|U_i|^{4k} \le \eta^{4k}$, $\mathbb{E}|V_{ij}|^{4k} \le \eta^{4k}$ and $\mathbb{E}|U_i V_{ij}|^{2k} \le \eta^{4k}$ for $k \ge 16$ (Assumption 3.3) and $\{U_i V_{ij}\}_{i=1}^n$ are independent, centred for $1 \le j \le m$. Consider (i), (ii) where $n^4\rho_m^{-4}\|[W]_{\underline{m}}\|^4 = (\sum_{j=1}^m \lambda_j (\sum_{i=1}^n U_i V_{ij})^2)^2$ and $([W]_{\underline{m}}^t [\Gamma]_{\underline{m}}^{-1}[W]_{\underline{m}})/\rho_m^2 = n^{-2}\sum_{j=1}^m (\sum_{i=1}^n U_i V_{ij})^2$. Applying Minkowski's (resp., Jensen's) inequality and Theorem 2.10 in [34], we have

$$n^2\rho_m^{-4}\mathbb{E}\|[W]_{\underline{m}}\|^4 \le n^{-2}\left[\sum_{j=1}^m \lambda_j\left(\mathbb{E}\left|\sum_{i=1}^n U_i V_{ij}\right|^4\right)^{1/2}\right]^2$$

$$\le C\eta^8\left[\sum_{j=1}^m \lambda_j\right]^2;$$

$$n^k m^{-k}\rho_m^{-2k}\mathbb{E}\|[\Gamma]_{\underline{m}}^{-1/2}[W]_{\underline{m}}\|^{2k} \le \frac{1}{m}\sum_{j=1}^m n^{-k}\mathbb{E}\left|\sum_{i=1}^n U_i V_{ij}\right|^{2k} \le C(k)\eta^{4k},$$

which, respectively, implies (i), since $\sum_{j=1}^m \lambda_j \le \mathbb{E}\|X\|_{\mathbb{H}}^2$, and (ii), by employing Markov's inequality. Proof of (iii). Since $\{V_{ij}V_{il} - \delta_{jl}\}_{i=1}^n$ are independent, centred with $\mathbb{E}|V_{ij}V_{il} - \delta_{jl}|^{2k} \le C\eta^{4k}$, $1 \le j, l \le m$, Theorem 2.10 in [34] implies $n^k\mathbb{E}|n^{-1}\sum_{i=1}^n (V_{ij}V_{il} - \delta_{jl})|^{2k} \le C(k)\eta^{4k}$ and $m^{-2k}n^k\mathbb{E}\|[\Xi]_{\underline{m}}\|_s^{2k} \le C(k)\eta^{4k}$ because $\|[\Xi]_m\|_s^2 \le \sum_{1\le j,l\le m}|V_{ij}V_{il} - \delta_{jl}|^2$. Applying Markov's inequality gives (iii). Proof of (iv). Since $\{Y_i^2/\sigma_Y^2 - 1\}_{i=1}^n$ are independent, centred with $\mathbb{E}|Y_i^2/\sigma_Y^2 - 1|^{2k} \le C(k)\eta^{4k}$ Theorem 2.10 in [34] implies $\mathbb{E}|n^{-1}\sum_{i=1}^n Y_i^2/\sigma_Y^2 - 1|^{2k} \le C(k)n^{-k}\eta^{4k}$ and $P(|n^{-1}\sum_{i=1}^n Y_i^2/\sigma_Y^2 - 1| > 1/2) \le Cn^{-16}\eta^{64}$ employing Markov's inequality. (iv) follows now from $\{1/2 \le \widehat{\sigma}_Y^2/\sigma_Y^2 \le 3/2\}^c \subset \{|n^{-1}\sum_{i=1}^n Y_i^2/\sigma_Y^2 - 1| > 1/2\}$. □

LEMMA D.3. *Let $\varsigma_m := \sigma + \eta^2\|\Gamma^{1/2}(\beta^m - \beta)\|_{\mathbb{H}}$, $m \ge 1$. There exists a numerical constant C such that for all $\lfloor n^{1/4}\rfloor \ge m \ge 1$ we have*

$$\mathbb{E}\left(\frac{\|[\Gamma]_{\underline{m}}^{-1/2}[W_n]_{\underline{m}}\|^2}{\varsigma_m^2} - 12\frac{m\Lambda_m^{[\Gamma]}}{n}\right)_+ \le \frac{C}{n}\left\{\exp\left(-\frac{m\Lambda_m^{[\Gamma]}}{6}\right) + \exp\left(-\frac{n^{1/6}}{100}\right) + \frac{\eta^{32}}{n^2}\right\}.$$

PROOF. Let $\mathbb{S}^m := \{z \in \mathbb{R}^m : z^t z \le 1\}$. Define $\mathcal{E}_n := \{e \in \mathbb{R} : |e| \le n^{1/6}\}$, $\mathcal{X}_{1n} := \{x \in \mathbb{H} : |\langle\beta - \beta^m, x\rangle_{\mathbb{H}}| \le \|\Gamma^{1/2}(\beta - \beta^m)\|_{\mathbb{H}} n^{1/6}\}$, $\mathcal{X}_{2n} := \{x \in \mathbb{H} : [x]_{\underline{m}}^t[\Gamma]_{\underline{m}}^{-1}[x]_{\underline{m}} \le mn^{1/3}\}$ and $\mathcal{X}_n := \mathcal{X}_{1n} \cap \mathcal{X}_{2n}$. For $e \in \mathbb{R}$, $x \in \mathbb{H}$, $s \in \mathbb{S}^m$ set

$$\nu_s(e, x) := (\sigma e + \langle\beta - \beta^m, x\rangle_{\mathbb{H}}) s^t[\Gamma]_{\underline{m}}^{-1/2}[x]_{\underline{m}}\mathbb{1}_{\{e\in\mathcal{E}_n, x\in\mathcal{X}_n\}},$$

$$R_s(e,x) := (\sigma e + \langle \beta - \beta^m, x\rangle_{\mathbb{H}}) s^t [\Gamma]_{\underline{m}}^{-1/2} [x]_{\underline{m}} (1 - \mathbb{1}_{\{e\in\mathcal{E}_n, x\in\mathcal{X}_n\}}).$$

Let $\nu_s^* := n^{-1}\sum_{i=1}^n \{\nu_s(\varepsilon_i, X_i) - \mathbb{E}\nu_s(\varepsilon_i, X_i)\}$ and $R_s^* := n^{-1}\sum_{i=1}^n \{R_s(\varepsilon_i, X_i) - \mathbb{E}R_s(\varepsilon_i, X_i)\}$, then $\|[\Gamma]_{\underline{m}}^{-1/2}[W_n]_{\underline{m}}\|^2 = \sup_{s\in\mathbb{S}^m} |\nu_s^* + R_s^*|^2$ and hence

$$\begin{aligned} &\mathbb{E}\biggl(\|[\Gamma]_{\underline{m}}^{-1/2}[W_n]_{\underline{m}}\|^2 - 12\varsigma_m^2 \frac{m\Lambda_m^{[\Gamma]}}{n}\biggr)_+ \\ &\quad\le 2\mathbb{E}\biggl(\sup_{s\in\mathbb{S}^m} |\nu_s^*|^2 - 6\varsigma_m^2\frac{m\Lambda_m^{[\Gamma]}}{n}\biggr)_+ + 2\mathbb{E}\sup_{s\in\mathbb{S}^m}|R_s^*|^2 \qquad \text{(D.3)} \\ &\quad =: 2\{T_1 + T_2\}, \end{aligned}$$

where we bound the r.h.s. terms T_1 and T_2 separately. Consider first T_1. We intend to apply Talagrand's inequality. To this end, for $e\in\mathbb{R}$, $x\in\mathbb{H}$, we have

$$\begin{aligned} \sup_{s\in\mathbb{S}^m} |\nu_s(e,x)|^2 &= (\sigma e + \langle \beta - \beta^m, x\rangle_{\mathbb{H}})^2 [x]_{\underline{m}}^t [\Gamma]_{\underline{m}}^{-1}[x]_{\underline{m}} \mathbb{1}_{\{e\in\mathcal{E}_n, x\in\mathcal{X}_n\}} \\ &\le (\sigma + \|\Gamma^{1/2}(\beta^m - \beta)\|_{\mathbb{H}})^2 n^{2/3} m \le \varsigma_m^2 n^{2/3} m =: h^2. \end{aligned} \qquad \text{(D.4)}$$

By employing the independence of ε and X it is easily seen that

$$n\mathbb{E}\sup_{s\in\mathbb{S}^m}|\nu_s^*|^2 \le \sigma^2 m + \mathbb{E}|\langle \beta - \beta^m, X\rangle_{\mathbb{H}}|^2 [X]_{\underline{m}}^t [\Gamma]_{\underline{m}}^{-1}[X]_{\underline{m}},$$

$$\sup_{s\in\mathbb{S}^m} \frac{1}{n}\sum_{i=1}^n \mathbb{V}\mathrm{ar}(\nu_s(\varepsilon_i, X_i)) \le \sigma^2 + \sup_{s\in\mathbb{S}^m} \mathbb{E}|\langle \beta - \beta^m, X\rangle_{\mathbb{H}}|^2 |s^t[\Gamma]_{\underline{m}}^{-1/2}[X]_{\underline{m}}|^2.$$

By applying the Cauchy–Schwarz inequality together with $\mathbb{E}\|[\Gamma]_{\underline{m}}^{-1/2}[X]_{\underline{m}}\|^4 \le m^2\eta^4$ and $\mathbb{E}|\langle\beta - \beta^m, X\rangle_{\mathbb{H}}|^4 \le \|\Gamma^{1/2}(\beta^m - \beta)\|_{\mathbb{H}}^4 \eta^4$ we obtain

$$\mathbb{E}\sup_{s\in\mathbb{S}^m}|\nu_s^*|^2 \le \frac{m}{n}(\sigma^2 + \|\Gamma^{1/2}(\beta - \beta^m)\|_{\mathbb{H}}^2 \eta^4) \le \varsigma_m^2 \frac{m\Lambda_m^{[\Gamma]}}{n} =: H^2, \qquad \text{(D.5)}$$

and taking into account that $\mathbb{E}|s^t[\Gamma]_{\underline{m}}^{-1/2}[X]_{\underline{m}}|^4 \le \eta^4$, $s\in\mathbb{S}^m$, we obtain

$$\sup_{s\in\mathbb{S}^m}\frac{1}{n}\sum_{i=1}^n \mathbb{V}\mathrm{ar}(\nu_s(\varepsilon_i, X_i)) \le \sigma^2 + \|\Gamma^{1/2}(\beta^m - \beta)\|_{\mathbb{H}}^2\eta^4 \le \varsigma_m^2 =: v. \qquad \text{(D.6)}$$

Due to (D.4)–(D.6) Talagrand's inequality (Lemma D.1 with $\varepsilon = 1$) implies

$$\mathbb{E}\biggl(\sup_{s\in\mathbb{S}^m}|\nu_s^*|^2 - 6\varsigma_m^2\frac{m\Lambda_m^{[\Gamma]}}{n}\biggr)_+ \le C\frac{\varsigma_m^2}{n}\biggl\{\exp\biggl(-\frac{m\Lambda_m^{[\Gamma]}}{6}\biggr) + \exp\biggl(-\frac{n^{1/6}}{100}\biggr)\biggr\}, \qquad \text{(D.7)}$$

where we used that $m \le \lfloor n^{1/4}\rfloor$. Consider T_2. By employing $[X]_{\underline{m}}[\Gamma]_{\underline{m}}^{-1}[X]_{\underline{m}} \times \mathbb{1}_{\{X\in\mathcal{X}_{2,n}\}} \le mn^{1/3}$ and $\mathcal{X}_n = \mathcal{X}_{1n}\cap\mathcal{X}_{2n}$ we have

$$\begin{aligned} n\mathbb{E}\sup_{s\in\mathbb{S}^m}|R_s^*|^2 &\le \mathbb{E}(\sigma\varepsilon + \langle\beta - \beta^m, X\rangle_{\mathbb{H}})^2[X]_{\underline{m}}[\Gamma]_{\underline{m}}^{-1}[X]_{\underline{m}}\mathbb{1}_{\{X\notin\mathcal{X}_{2,n}\}} \\ &\quad + mn^{1/3}\mathbb{E}(\sigma\varepsilon + \langle\beta - \beta^m, X\rangle_{\mathbb{H}})^2(\mathbb{1}_{\{\varepsilon\notin\mathcal{E}_n\}} + \mathbb{1}_{\{X\notin\mathcal{X}_{1n}\}}). \end{aligned}$$

Since $\mathbb{E}(\sigma\varepsilon+\langle\beta-\beta^m, X\rangle_{\mathbb{H}})^4 \le (\sigma^2+\|\Gamma^{1/2}(\beta-\beta^m)\|_{\mathbb{H}}^2)^2\eta^4$, $\mathbb{E}\varepsilon^2=1$ and $\mathbb{E}|\langle\beta-\beta^m, X\rangle_{\mathbb{H}}|^2 = \|\Gamma^{1/2}(\beta-\beta^m)\|_{\mathbb{H}}^2$ the independence of ε and X implies

$$n\mathbb{E}\sup_{s\in\mathbb{S}^m}|R_s^*|^2 \le (\sigma^2+\|\Gamma^{1/2}(\beta-\beta^m)\|_{\mathbb{H}}^2)\eta^2(\mathbb{E}|[X]_{\underline{m}}[\Gamma]_{\underline{m}}^{-1}[X]_{\underline{m}}|^2\mathbb{1}_{\{X\notin\mathcal{X}_{2,n}\}})^{1/2}$$
$$+ mn^{1/3}\{\sigma^2\mathbb{E}\varepsilon^2\mathbb{1}_{\{\varepsilon\notin\mathcal{E}_n\}} + \|\Gamma^{1/2}(\beta-\beta^m)\|_{\mathbb{H}}^2 P(\varepsilon\notin\mathcal{E}_n)$$
$$+ \sigma^2 P(X\notin\mathcal{X}_{1n}) + \mathbb{E}|\langle\beta-\beta^m, X\rangle_{\mathbb{H}}|^2\mathbb{1}_{\{X\notin\mathcal{X}_{1n}\}}\}.$$

We exploit now the estimates given in (D.1) and (D.2). Thereby, we obtain

$$n\mathbb{E}\sup_{s\in\mathbb{S}^m}|R_s^*|^2 \le C(\sigma^2+\|\Gamma^{1/2}(\beta-\beta^m)\|_{\mathbb{H}}^2)\eta^{32}mn^{-7/3} \le C\varsigma_m^2\eta^{32}n^{-2},$$

where we used that $m \le \lfloor n^{1/4}\rfloor$. Keeping in mind decomposition (D.3), the last bound and (D.7) imply together the claim of Lemma D.3. □

LEMMA D.4. *There exists a constant $K := K(\sigma, \eta, \mathcal{F}_b^r, \mathcal{G}_\gamma^d)$ depending on σ, η and the classes $\mathcal{F}_b^r$ and $\mathcal{G}_\gamma^d$ only such that for all $n \ge 1$ we have*

$$\sup_{\beta\in\mathcal{F}_b^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\sum_{m=m_n^\diamond}^{M_n^+}\Delta_m^{[\Gamma]}\mathbb{E}\left(\|[\Gamma]_{\underline{m}}^{-1/2}[W_n]_{\underline{m}}\|^2 - \frac{12\sigma_m^2 m\Lambda_m^{[\Gamma]}}{n}\right)_+ \le K\frac{\eta^{32}(\sigma^2+r)\Sigma}{n}.$$

PROOF. There exists an integer $n_o := n_o(\sigma, \eta, \mathcal{F}_b^r, \mathcal{G}_\gamma^d)$ depending on σ, η and the classes $\mathcal{F}_b^r$ and $\mathcal{G}_\gamma^d$ only such that for all $n \ge n_o$ and for all $m \ge m_n^\diamond$ we have $\varsigma_m^2 \le 2(\sigma^2+\|\Gamma^{1/2}\beta\|_{\mathbb{H}}^2 + [g]_{\underline{m}}^t[\Gamma]_{\underline{m}}^{-1}[g]_{\underline{m}}) = 2(\sigma_Y^2 + [g]_{\underline{m}}^t[\Gamma]_{\underline{m}}^{-1}[g]_{\underline{m}}) = \sigma_m^2$. Indeed, we have $1/m_n^\diamond = o(1)$ as $n\to\infty$ and $|\varsigma_m^2-\sigma^2| = o(1)$ as $m\to\infty$ because $\varsigma_m = \sigma + \eta^2\|\Gamma^{1/2}(\beta^m-\beta)\|_{\mathbb{H}}$ and $\|\Gamma^{1/2}(\beta^m-\beta)\|_{\mathbb{H}}^2 \le 34d^9r\gamma_m b_m^{-1}$ due to Lemma B.1(v). First, consider $n < n_o$. Due to Lemma D.2(i) and $\rho_m^2 \le 2(\sigma^2+35d^6r)$ [Lemma B.2(iv)] we have for all $m \ge 1$

$$\mathbb{E}\left(\|[\Gamma]_{\underline{m}}^{-1/2}[W_n]_{\underline{m}}\|^2 - \frac{12\sigma_m^2 m\Lambda_m^\Gamma}{n}\right)_+ \le \mathbb{E}\|[\Gamma]_{\underline{m}}^{-1/2}[W_n]_{\underline{m}}\|^2 \le C\frac{m}{n}\eta^4(\sigma^2+d^6r).$$

Hence, $M_n^+ \le \lfloor n^{1/4}\rfloor$ and $m\Delta_m^\Gamma \le \delta_{M_n^+}^\Gamma \le nC(d)$ [Lemma B.2(ii)] imply

$$\sup_{\beta\in\mathcal{F}_b^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\sum_{m=m_n^\diamond}^{M_n^+}\Delta_m^{[\Gamma]}\mathbb{E}\left(\|[\Gamma]_{\underline{m}}^{-1/2}[W_n]_{\underline{m}}\|^2 - 12\sigma_m^2\frac{m\Lambda_m^{[\Gamma]}}{n}\right)_+$$
$$\le C(d)\frac{n_o^{5/4}\eta^4(\sigma^2+r)}{n},$$

which proves the lemma for all $1 \le n < n_o$. Consider now $n \ge n_o$ where $\varsigma_m^2 \le \sigma_m^2$ for all $m \ge m_n^\diamond$. Thereby, we can apply Lemma D.3, which gives

$$\sup_{\beta\in\mathcal{F}_b^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\sum_{m=m_n^\diamond}^{M_n^+}\Delta_m^{[\Gamma]}\mathbb{E}\bigg(\|[\Gamma]_{\underline{m}}^{-1/2}[W_n]_{\underline{m}}\|^2 - 12\sigma_m^2\frac{m\Lambda_m^{[\Gamma]}}{n}\bigg)_+$$
$$\le C\sup_{\beta\mathcal{F}_b^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\sum_{m=m_n^\diamond}^{M_n^+}\frac{\varsigma_m^2\Delta_m^{[\Gamma]}}{n}\bigg\{\exp\bigg(-\frac{m\Lambda_m^{[\Gamma]}}{6}\bigg)+\exp\bigg(-\frac{n^{1/6}}{100}\bigg)+\frac{\eta^{32}}{n^2}\bigg\}.$$

Since $\Delta_k^{[\Gamma]} \le 4d^3\Delta_k^\gamma$, $\Lambda_k^{[\Gamma]} \ge (1+\log d)^{-1}\Lambda_k^\gamma$, $M_n^+\Delta_{M_n^+}^{[\Gamma]} \le \delta_{M_n^+}^{[\Gamma]} \le nCd^6(1+\log d)$ and $\varsigma_k^2 \le \sigma_k^2 \le 2(\sigma^2+35d^6r)$ [Lemma B.2(i), (ii), (iv), resp.] follows

$$\sup_{\beta\in\mathcal{F}_b^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\sum_{m=m_n^\diamond}^{M_n^+}\Delta_m^{[\Gamma]}\mathbb{E}\bigg(\|[\Gamma]_{\underline{m}}^{-1/2}[W_n]_{\underline{m}}\|^2 - 12\sigma_m^2\frac{m\Lambda_m^{[\Gamma]}}{n}\bigg)_+$$
$$\le C(d)(\sigma^2+r)n^{-1}$$
$$\times\sup_{\beta\in\mathcal{F}_b^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\bigg\{\sum_{m=m_n^\diamond}^{M_n^+}\Delta_m^\gamma\exp\bigg(-\frac{m\Lambda_m^\gamma}{6(1+\log d)}\bigg)+n\exp\bigg(-\frac{n^{1/6}}{100}\bigg)+\frac{\eta^{32}}{n}\bigg\}.$$

Finally, $\Sigma = \Sigma(\mathcal{G}_\gamma^d)$ as in (3.3) and $n\exp(-n^{1/6}/100) \le C$ imply for $n \ge n_o$

$$\sup_{\beta\in\mathcal{F}_b^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\sum_{m=m_n^\diamond}^{M_n^+}\Delta_m^{[\Gamma]}\mathbb{E}\bigg(\|[\Gamma]_{\underline{m}}^{-1/2}[W_n]_{\underline{m}}\|^2 - 12\sigma_m^2\frac{m\Lambda_m^{[\Gamma]}}{n}\bigg)_+$$
$$\le C(d)\frac{\eta^{32}(\sigma^2+r)\Sigma}{n}.$$

Combining the cases $n < n_o$ and $n \ge n_o$ completes the proof. □

LEMMA D.5. *There exist a numerical constant C and a constant $C(d)$ only depending on d such that for all $n \ge 1$ we have*:

(i) $\sup_{\beta\in\mathcal{F}_b^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\{n^6(M_n^+)^2\max_{1\le m\le M_n^+}P(\mho_{m,n}^c)\} \le C\eta^{64}$;

(ii) $\sup_{\beta\in\mathcal{F}_b^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\{nM_n^+\max_{1\le m\le M_n^+}P(\Omega_{m,n}^c)\} \le C(d)\eta^{64}$;

(iii) $\sup_{\beta\in\mathcal{F}_b^r}\sup_{\Gamma\in\mathcal{G}_\gamma^d}\{n^7P(\mathcal{E}_n^c)\} \le C\eta^{64}$.

PROOF. By employing Lemma D.2 rather than Lemma C.2 the proof follows along the lines of the proof of Lemma C.5, and we omit the details. □

PROPOSITION D.6. *Let $\kappa = 288$ in the definition (2.11) of the penalty* pen. *There exists a constant $K := K(\sigma,\eta,\mathcal{F}_b^r,\mathcal{G}_\gamma^d)$ depending on σ, η and the classes*

$\mathcal{F}_b^r$ *and* $\mathcal{G}_\gamma^d$ *only such that for all* $n \geq 1$, *we have*

$$\sup_{\beta \in \mathcal{F}_b^r} \sup_{\Gamma \in \mathcal{G}_\gamma^d} \mathbb{E}\Big\{ \sup_{m_n^\diamond \leq k \leq M_n^+} \Big(\|\widehat{\beta}_k - \beta^k\|_\omega^2 - \frac{1}{6}\mathrm{pen}_k \Big)_+ \Big\} \leq K \eta^{64} (\sigma^2 + r) \Sigma n^{-1}.$$

PROOF. By employing Lemmas D.2, D.4 and D.5 rather than Lemmas C.2, C.4 and C.5 the proof follows along the lines of the proof of Proposition C.6, and we omit the details. □

PROPOSITION D.7. *Let* $\kappa = 288$ *in definition* (2.11) *and* (2.12) *of* pen *and* $\widehat{\mathrm{pen}}$. *There exists a constant* $C(d)$ *only depending on* d *such that for all* $n \geq 1$,

$$\sup_{\beta \in \mathcal{F}_b^r} \sup_{\Gamma \in \mathcal{G}_\gamma^d} \mathbb{E}(\|\widehat{\beta}_{\widehat{m}} - \beta\|_\omega^2 \mathbb{1}_{\mathcal{E}_n^c}) \leq C(d) \eta^{64} (\sigma^2 + r) \Sigma n^{-1}.$$

PROOF. Taking into account Lemmas D.2(i) and D.5 rather than Lemmas C.2(i) and C.5 the proof follows along the lines of the proof of Proposition C.7, and we omit the details. □

PROOF OF PROPOSITION 3.3. The result follows from Propositions D.6 and D.7, and we omit the details. □

Acknowledgments. We are grateful to two referees and the Associate Editor for constructive criticism and clear guidelines.

SUPPLEMENTARY MATERIAL

Simulation study (DOI: 10.1214/12-AOS1050SUPP; .pdf). A simulation study illustrating the finite sample behavior of the fully data-driven estimation procedure and its good performance.

REFERENCES

[1] BARRON, A., BIRGÉ, L. and MASSART, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields* **113** 301–413. MR1679028

[2] BOSQ, D. (2000). *Linear Processes in Function Spaces*: *Theory and Applications*. *Lecture Notes in Statistics* **149**. Springer, New York. MR1783138

[3] BUTUCEA, C. and TSYBAKOV, A. B. (2007). Sharp optimality in density deconvolution with dominating bias. I. *Teor. Veroyatn. Primen.* **52** 111–128. MR2354572

[4] BUTUCEA, C. and TSYBAKOV, A. B. (2007). Sharp optimality in density deconvolution with dominating bias. II. *Teor. Veroyatn. Primen.* **52** 336–349. MR2742504

[5] CAI, T. and ZHOU, H. (2008). Adaptive functional linear regression. Technical report. Available at http://www-stat.wharton.upenn.edu/~tcai/paper/Adaptive-FLR.pdf.

[6] CARDOT, H., FERRATY, F. and SARDA, P. (1999). Functional linear model. *Statist. Probab. Lett.* **45** 11–22. MR1718346

[7] CARDOT, H., FERRATY, F. and SARDA, P. (2003). Spline estimators for the functional linear model. *Statist. Sinica* **13** 571–591. MR1997162

[8] CARDOT, H. and JOHANNES, J. (2010). Thresholding projection estimators in functional linear models. *J. Multivariate Anal.* **101** 395–408. MR2564349
[9] CARDOT, H., MAS, A. and SARDA, P. (2007). CLT in functional linear regression models. *Probab. Theory Related Fields* **138** 325–361. MR2299711
[10] COMTE, F. and JOHANNES, J. (2010). Adaptive estimation in circular functional linear models. *Math. Methods Statist.* **19** 42–63. MR2682854
[11] COMTE, F., ROZENHOLC, Y. and TAUPIN, M.-L. (2006). Penalized contrast estimator for adaptive density deconvolution. *Canad. J. Statist.* **34** 431–452. MR2328553
[12] CRAMBES, C., KNEIP, A. and SARDA, P. (2009). Smoothing splines estimators for functional linear regression. *Ann. Statist.* **37** 35–72. MR2488344
[13] EFROMOVICH, S. and KOLTCHINSKII, V. (2001). On inverse problems with unknown operators. *IEEE Trans. Inform. Theory* **47** 2876–2894. MR1872847
[14] ENGL, H. W., HANKE, M. and NEUBAUER, A. (2000). *Regularization of Inverse Problems. Mathematics and Its Applications* **375**. Kluwer Academic Press, Dordrecht. MR1408680
[15] FERRATY, F. and VIEU, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer, New York. MR2229687
[16] FORNI, M. and REICHLIN, L. (1998). Let's get real: A factor analytical approach to disaggregated business cycle dynamics. *Rev. Econom. Stud.* **65** 453–473.
[17] FRANK, I. and FRIEDMAN, J. (1993). A statistical view of some chemometrics regression tools. *Technometrics* **35** 109–148.
[18] GOLDENSHLUGER, A. and LEPSKI, O. (2011). Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39** 1608–1632. MR2850214
[19] GOLDENSHLUGER, A. and TSYBAKOV, A. (2001). Adaptive prediction and estimation in linear regression with infinitely many parameters. *Ann. Statist.* **29** 1601–1619. MR1891740
[20] GOLDENSHLUGER, A. and TSYBAKOV, A. (2003). Optimal prediction for linear regression with infinitely many parameters. *J. Multivariate Anal.* **84** 40–60. MR1965822
[21] HALL, P. and HOROWITZ, J. L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35** 70–91. MR2332269
[22] HEINZ, E. (1951). Beiträge zur Störungstheorie der Spektralzerlegung. *Math. Ann.* **123** 415–438. MR0044747
[23] HOFFMANN, M. and REISS, M. (2008). Nonlinear estimation for linear inverse problems with error in the operator. *Ann. Statist.* **36** 310–336. MR2387973
[24] JOHANNES, J. and SCHENK, R. (2010). On rate optimal local estimation in functional linear model. Univ. Catholique de Louvain. Available at http://arxiv.org/abs/0902.0645.
[25] KAWATA, T. (1972). *Fourier Analysis in Probability Theory. Probability and Mathematical Statistics* **15**. Academic Press, New York. MR0464353
[26] KLEIN, T. and RIO, E. (2005). Concentration around the mean for maxima of empirical processes. *Ann. Probab.* **33** 1060–1077. MR2135312
[27] LEPSKIĬ, O. V. (1990). A problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35** 454–466.
[28] MARX, B. D. and EILERS, P. H. (1999). Generalized linear regression on sampled signals and curves: A p-spline approach. *Technometrics* **41** 1–13.
[29] MASSART, P. (2007). *Concentration Inequalities and Model Selection. Lecture Notes in Math.* **1896**. Springer, Berlin. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, with a foreword by Jean Picard. MR2319879
[30] MATHÉ, P. (2006). The Lepskiĭ principle revisited. *Inverse Problems* **22** L11–L15. MR2235633
[31] MÜLLER, H.-G. and STADTMÜLLER, U. (2005). Generalized functional linear models. *Ann. Statist.* **33** 774–805. MR2163159
[32] NATTERER, F. (1984). Error bounds for Tikhonov regularization in Hilbert scales. *Appl. Anal.* **18** 29–37. MR0762862

[33] NEUBAUER, A. (1988). When do Sobolev spaces form a Hilbert scale? *Proc*. *Amer. Math*. *Soc*. **103** 557–562. MR0943084

[34] PETROV, V. V. (1995). *Limit Theorems of Probability Theory*: *Sequences of Independent Random Variables*. *Oxford Studies in Probability* **4**. Clarendon Press, Oxford. MR1353441

[35] PREDA, C. and SAPORTA, G. (2005). PLS regression on a stochastic process. *Comput*. *Statist*. *Data Anal*. **48** 149–158. MR2134488

[36] RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. Springer, New York. MR2168993

[37] TALAGRAND, M. (1996). New concentration inequalities in product spaces. *Invent*. *Math*. **126** 505–563. MR1419006

[38] VERZELEN, N. (2010). High-dimensional Gaussian model selection on a Gaussian design. *Ann*. *Inst*. *Henri Poincaré Probab. Stat*. **46** 480–524. MR2667707

LABORATOIRE MAP5
UMR CNRS 8145
45, RUE DES SAINTS-PÈRES
F-75270 PARIS CEDEX 06
FRANCE
fabienne.comte@parisdescartes.fr

INSTITUT DE STATISTIQUE, BIOSTATISTIQUE ET SCIENCES ACTUARIELLES (ISBA)
VOIE DU ROMAN PAYS 20
B-1348 LOUVAIN-LA-NEUVE
BELGIUM
E-MAIL: jan.johannes@uclouvain.be