INSTITUT DE STATISTIQUE
BIOSTATISTIQUE ET
SCIENCES ACTUARIELLES
(I S B A)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION PAPER

1026

ADAPTIVE NONPARAMETRIC INSTRUMENTAL REGRESSION BY MODEL SELECTION

JOHANNES, J. and M. SCHWARZ

Adaptive nonparametric instrumental regression by model selection

Jan Johannes * Maik Schwarz *

Université catholique de Louvain

March 16, 2010

We consider the problem of estimating the structural function in nonparametric instrumental regression, where in the presence of an instrument W a response Y is modeled in dependence of an endogenous explanatory variable Z. The proposed estimator is based on dimension reduction and additional thresholding. The minimax optimal rate of convergence of the estimator is derived assuming that the structural function belongs to some ellipsoids which are in a certain sense linked to the conditional expectation operator of Z given W. We illustrate these results by considering classical smoothness assumptions. However, the proposed estimator requires an optimal choice of a dimension parameter depending on certain characteristics of the unknown structural function and the conditional expectation operator of Z given W, which are not known in practice. The main issue addressed in our work is a fully adaptive choice of this dimension parameter using a model selection approach under the restriction that the conditional expectation operator of Z given W is smoothing in a certain sense. In this situation we develop a penalized minimum contrast estimator with randomized penalty and collection of models. We show that this data-driven estimator can attain the lower risk bound up to a constant over a wide range of smoothness classes for the structural function.

Keywords: Nonparametric regression, Instrumental variable, Linear Galerkin approach, Minimax theory, Orthogonal series estimation, Model selection, Adaptive estimation .
JEL classifications: Primary C14; secondary C30.

This work was supported by the IAP research network no. P6/03 of the Belgian Government (Belgian Science Policy) and by the "Fonds Spéciaux de Recherche" from the Université catholique de Louvain.

*Institut de statistique, biostatistique et sciences actuarielles (ISBA), Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgique, e-mail: {jan.johannes|maik.schwarz}@uclouvain.be

1. Introduction

Nonparametric instrumental regression models have attracted increasing attention in the econometrics and statistics literature (c.f. Florens (2003), Darolles et al. (2002), Newey and Powell (2003), Hall and Horowitz (2007) or Blundell et al. (2007) to name only a few). In instrumental regression, the dependence of a response Y on the variation of an endogenous vector Z of explanatory variables is characterized by

$$Y = \varphi(Z) + U \tag{1.1a}$$

for some error term U. Additionally, a vector of exogenous instruments W such that

$$\mathbb{E}[U|W] = 0 \tag{1.1b}$$

is supposed to be observed. The nonparametric relationship is hence modeled by the structural function φ. Typical examples are error-in-variable models, simultaneous equations or treatment models with endogenous selection. However, it is worth noting that in the presence of instrumental variables the model equations (1.1a–1.1b) are the natural generalization of a standard parametric model (see, e.g., Amemiya (1974)) to the nonparametric situation. This extension has first been introduced by Florens (2003) and Newey and Powell (2003), while its identification has been studied e.g. in Carrasco et al. (2006), Darolles et al. (2002) and Florens et al. (2010). It is interesting to note that recent applications and extensions of this approach include nonparametric tests of exogeneity (Blundell and Horowitz (2007)), quantile regression models (Horowitz and Lee (2007)), or semiparametric modeling (Florens et al. (2009)) to name but a few.

The nonparametric estimation of the structural function φ based on a sample of (Y, Z, W) has been studied in the literature. For example, Ai and Chen (2003), Blundell et al. (2007) or Newey and Powell (2003) consider sieve minimum distance estimators, while Darolles et al. (2002), Gagliardini and Scaillet (2006) or Florens et al. (2010) study penalized least squares estimators. The optimal estimation in a minimax sense has been studied by Hall and Horowitz (2005) and Chen and Reiß (2008). The authors prove a lower bound for the mean integrated squared error (MISE) and propose an estimator which can attain optimal rates. In the present work, we extend this result by considering not only the MISE of the estimation of φ but, more generally, a weighted risk (defined below), which allows us for example to consider the estimation of the derivatives of φ, too. We show a lower bound for this weighted risk and propose an estimator which can attain this lower bound up to a constant.

It has been noticed by Newey and Powell (2003) and Florens (2003) that the nonparametric estimation of the structural function φ generally leads to an ill-posed inverse problem. More precisely, consider the model equations (1.1a–1.1b). Taking the conditional expectation with respect to the instruments W on both sides in equation (1.1a) leads to the conditional moment equation

$$\mathbb{E}[Y|W] = \mathbb{E}[\varphi(Z)|W]. \tag{1.2}$$

Therefore, the estimation of the structural function φ is linked to the inversion of equation (1.2), which is not stable in general and hence an ill-posed inverse problem (for a comprehensive review of inverse problems in econometrics we refer to Carrasco et al. (2006)). To cope with this instability, one generally employs regularization techniques which involve the choice of a smoothing parameter. It is well known that the resulting estimation procedure can attain optimal rates only if this parameter is chosen in an appropriate way. This choice necessitates in general knowledge of characteristics of the structural function, such as the number of its derivatives, which are not known in practise. Thus, one of the essential problems in this theoretical framework is the fully data driven choice of smoothing parameters. In the present work, we present an adaptive method which indeed does not depend on any properties of φ, but which still necessitates that some characteristics of the underlying operator are known.

One objective of this paper is the minimax optimal nonparametric estimation of the structural function φ based on an independent and identically distributed (i.i.d.) sample of (Y, Z, W) obeying (1.1a–1.1b). After showing the lower risk bounds, we will follow an estimation approach often used in the literature. For the moment being, suppose that the structural function can be developed by using only m pre-specified functions $e_1, \dots, e_k$, say $\varphi = \sum_{j=1}^{k}[\varphi]_j e_j$, where now only the coefficients $[\varphi]_1, \dots, [\varphi]_k$ are unknown. Thereby, the conditional moment equation (1.2) reduces to a multivariate linear conditional moment equation, that is, $\mathbb{E}[Y|W] = \sum_{j=1}^{k}[\varphi]_j\mathbb{E}[e_j(Z)|W]$. Notice that solving this equation is a classical textbook problem in econometrics (c.f. Pagan and Ullah (1999)). One popular approach is to replace the conditional moment equation by an unconditional one. Therefore, given k functions $f_1, \dots, f_k$ one may consider k unconditional moment equations instead of the multivariate conditional moment equation, that is, $\mathbb{E}[Yf_l(W)] = \sum_{j=1}^{k}[\varphi]_j\mathbb{E}[e_j(Z)f_l(W)]$, $l = 1, \dots, k$. Notice that once the functions $\{f_l\}_{l=1}^{k}$ are chosen, all the unknown quantities in the unconditional moment equations can be estimated by simply replacing the theoretical expectation by its empirical counterpart. Moreover, a least squares solution of the estimated equation leads to a consistent and asymptoticly normal estimator of the parameter vector $([\varphi]_j)_{j=1}^{k}$ under very mild assumptions. The choice of the functions $\{f_l\}_{l=1}^{k}$ directly influences the asymptotic variance of the estimator and thus the question of optimal instruments arises (c.f. Newey (1990)). Nevertheless, this approach is very simple and the estimator can be calculated with most statistical software. However, it has a major defect, since in a vast majority of situations an infinite number of functions $\{e_j\}_{j\geqslant 1}$ and associated coefficients $([\varphi]_j)_{j\geqslant 1}$ is needed to develop the structural function φ. The choice of the functions $\{e_j\}_{j\geqslant 1}$ now reflects the a priori information (such as smoothness) about the structural function φ. However, if we consider also an infinite number of functions $\{f_l\}_{l\geqslant 1}$ then for each $k \geqslant 1$ we could still consider the least squares estimator described above. Notice that the dimension k plays the role of a smoothing parameter and we may hope that the estimator of the structural function φ is also consistent as k tends to infinity at a suitable rate. Unfortunately, this is not true in general. Let $\varphi_k := \sum_{j=1}^{k}[\varphi_k]_j e_j$ denote a least squares solution of the reduced unconditional moment equations, that is, the vector of coefficients $([\varphi_k]_j)_{j=1}^{k}$ minimizes the quantity $\sum_{l=1}^{k}\{\mathbb{E}[Yf_l(W)] - \sum_{j=1}^{k}\beta_j\mathbb{E}[e_j(Z)f_l(W)]\}^2$ over all $(\beta_j)_{j=1}^{k}$. Then, φ_k converges to the true structural function as k tends to infinity only under an additional assumption (defined below) on the basis $\{f_j\}_{j\geqslant 1}$. In this paper, we show that in terms of a weighted risk a least squares estimator $\widehat{\varphi}_k$ of φ based on a dimension reduction together with an additional thresholding can attain optimal rates of convergence, provided an optimal choice of the dimension parameter k. It is worth to note that all the results in this paper are obtained without any additional smoothness assumption on the joint density of (Y, Z, W). In fact we do not even impose the existence of such a density.

Our main contribution is the development of a method to choose the dimension parameter k in a fully data driven way, that is, not depending on characteristics of φ, and assuming only that the underlying conditional expectation operator is smoothing in a sense to be precised

below. The central result of the present paper states that for this automatic choice $\widehat{k}$, the least squares estimator $\widehat{\varphi}_{\widehat{k}}$ can attain the lower bound up to a constant, and is thus minimax-optimal. The adaptive choice of k is motivated by the general model selection strategy developed in Barron et al. (1999). Concretely, following Comte and Taupin (2003), $\widehat{k}$ is the minimizer of a penalized contrast. Note that Comte and Taupin (2003) consider a density deconvolution problem. We illustrate all of our results by considering the estimation of derivatives of the structural function under a smoothing conditional expectation operator. Typically, two types of such operators are distinguished in the literature, finitely or infinitely smoothing. It is interesting to note that Loubes and Marteau (2009) propose an adaptive estimator for the case where the operator is known to be finitely smoothing. They derive oracle inequalities and obtain convergence rates which differ from the optimal ones by a logarithmic factor. We underline that in contrast to this, we provide in this work a unified estimation procedure which can attain minimax-optimal rates in either of the both cases. In other words, our estimation procedure attains optimal rates without knowing in advance if the operator is finitely or infinitely smoothing.
This article is organized as follows. In the next section, we develop the minimax theory for the nonparametric instrumental regression model with respect to the weighted risk. We derive, as an illustration, the optimal convergence rates for the estimation of derivatives in the finitely and in the infinitely smoothing case. Section 3 is devoted to the construction of the adaptive estimator. An upper risk bound is shown and convergence rates for the finitely and infinitely smoothing case are found to coincide with minimax optimal ones. All proofs are deferred to the appendix.

2. Minimax optimal estimation

In this section, we develop a minimax theory for the estimation of the structural function and its derivatives in nonparametric instrumental regression models.

2.1. Basic model assumptions.

It is convenient to rewrite the moment equation (1.2) in terms of an operator between Hilbert spaces. Let us first introduce the Hilbert Spaces

$$L^2_Z = \{\varphi : \mathbb{R}^p \to \mathbb{R} \,|\, \|\varphi\|^2_Z := \mathbb{E}[\varphi^2(Z)] < \infty\},$$
$$L^2_W = \{\psi : \mathbb{R}^q \to \mathbb{R} \,|\, \|\psi\|^2_W := \mathbb{E}[\psi^2(W)] < \infty\},$$

endowed with the inner products $\langle\varphi,\tilde{\varphi}\rangle_Z = \mathbb{E}[\varphi(Z)\tilde{\varphi}(Z)]$, $\varphi,\tilde{\varphi} \in L^2_Z$, and $\langle\psi,\tilde{\psi}\rangle_W = \mathbb{E}[\psi(W)\tilde{\psi}(W)]$, $\psi,\tilde{\psi} \in L^2_W$, respectively. Then the conditional expectation of Z given W defines a linear operator $T\varphi := \mathbb{E}[\varphi(Z)|W]$, $\varphi \in L^2_Z$, which maps L^2_Z into L^2_W. In this notation, the moment equation (1.2) can be written as

$$g := \mathbb{E}[Y|W] = \mathbb{E}[\varphi(Z)|W] =: T\varphi, \tag{2.1}$$

where the function g belongs to L^2_W. Estimation of the structural function φ is thus linked to the inversion of the conditional expectation operator T and it is therefore called an inverse problem. Moreover, we suppose throughout this paper that the operator T is compact, which is the case under fairly mild assumptions (c.f. Carrasco et al. (2006)). Consequently, unlike in a multivariate linear instrumental regression model, a continuous generalized inverse of T does not exist as long as the range of the operator T is an infinite dimensional subspace of L^2_W. This corresponds to the setup of ill-posed inverse problems, with the additional difficulty that T is unknown and has to be estimated. In what follows, we always assume that there exists a unique solution $\varphi \in L^2_Z$ of equation (2.1), in other words, that g belongs to the range of T, and that T is injective. For a detailed discussion in the context of inverse problems see Chapter 2.1 in Engl et al. (2000), while in the special case of a nonparametric instrumental regression we refer to Carrasco et al. (2006).

2.2. Complexity: a lower bound

In this section we show that the obtainable accuracy of any estimator of the structural function φ is essentially determined by additional regularity conditions imposed on φ and the conditional expectation operator T. In this paper, these conditions are characterized through different weighted norms in L^2_Z with respect to a pre-specified orthonormal basis $\{e_j\}_{j\geqslant 1}$ of L^2_Z. We formalize these conditions as follows.

Minimal regularity conditions. Given a strictly positive sequence of weights $w := (w_j)_{j\geqslant 1}$, we denote by $\|\cdot\|_\omega$ the weighted norm given by

$$\|f\|_\omega := \sum_{j=1}^{\infty} w_j |\langle f, e_j\rangle_Z|^2, \qquad \forall f \in L^2_Z.$$

We shall measure the accuracy of any estimator $\widehat{\varphi}$ of the unknown structural function in terms of a weighted risk, that is $\mathbb{E}\|\widehat{\varphi} - \varphi\|^2_\omega$, for a pre-specified sequence of weights $\omega := (\omega_j)_{j\geqslant 1}$. This general approach allows as to consider not only the estimation of the structural function itself but also of its derivatives, as shown in section 2.4 below. Moreover, given a sequence of weights $\gamma := (\gamma_j)_{j\geqslant 1}$ we suppose, here and subsequently, that for some constant $\rho > 0$ the structural function φ belongs to the ellipsoid

$$\mathcal{F}^\rho_\gamma := \Big\{ f \in L^2_Z : \|\varphi\|^2_\gamma \leqslant \rho \Big\}. \tag{2.2}$$

The ellipsoid $\mathcal{F}^\rho_\gamma$ captures all the prior information (such as smoothness) about the unknown structural function φ. Furthermore, as usual in the context of ill-posed inverse problems, we specify the mapping properties of the conditional expectation operator T. Therefore, consider the sequence $(\|Te_j\|_W)_{j\geqslant 1}$, which converges to zero since T is compact. In what follows, we impose restrictions on the decay of this sequence. Denote by $\mathcal{T}$ the set of all injective compact operator mapping L^2_Z into L^2_W. Given a strictly positive sequence of weights $\lambda := (\lambda_j)_{j\geqslant 1}$ and $d \geqslant 1$, we define the subset $\mathcal{T}^d_\lambda$ of $\mathcal{T}$ by

$$\mathcal{T}^d_\lambda := \Big\{ T \in \mathcal{T} : \quad \|f\|^2_\lambda / d \leqslant \|Tf\|^2_W \leqslant d\,\|f\|^2_\lambda, \quad \forall f \in L^2_Z \Big\}. \tag{2.3}$$

Notice that for all $T \in \mathcal{T}^d_\lambda$ it follows that $d^{-1} \leqslant \|Te_j\|^2_W / \lambda_j \leqslant d$. Furthermore, let us denote by $T^* : L^2_W \to L^2_Z$ the adjoint of T which satisfies $T^*\psi = \mathbb{E}[\psi(W)|Z]$ for all $\psi \in L^2_W$. If now $T \in \mathcal{T}$ and if $\{e_j\}_{j\geqslant 1}$ are the eigenfunctions of T^*T, then the sequence λ specifies the decay of the eigenvalues of T^*T. All results of this work are derived under regularity conditions on the structural function φ and the conditional expectation operator T described by the sequences γ and λ, respectively. However, below we provide illustrations of these conditions by assuming a «regular decay» of these sequences. The next assumption summarizes our minimal regularity conditions on these sequences.

Assumption A1 *Let $\gamma := (\gamma_j)_{j\in\mathbb{N}}$, $\omega := (\omega_j)_{j\in\mathbb{N}}$ and $\lambda := (\lambda_j)_{j\in\mathbb{N}}$ be strictly positive sequences of weights with $\gamma_0 = \omega_0 = \lambda_0 = 1$ and $\Gamma := \sum_{j\in\mathbb{N}} \gamma_j^{-1} < \infty$, such that $(\omega_n/\gamma_n)_{n\in\mathbb{N}}$ and $(\lambda_n)_{n\in\mathbb{N}}$ are non-increasing, respectively.*

It is worth noting that the monotonicity assumption $(\omega_n/\gamma_n)_{n\in\mathbb{N}}$ only ensures that $\|\varphi\|_\omega$ is finite, and hence the weighted risk is a well-defined measure of accuracy for estimators of φ. Heuristically, this reflects the fact that we cannot estimate the $s+1$-th derivative if the structural function has only s derivatives. Moreover, in the illustration given in section 2.4, the additional assumption $\Gamma := \sum_{j\in\mathbb{N}} \gamma_j^{-1} < \infty$ can be interpreted as a continuity assumption on φ.

The lower bound. The next assertion provides a lower bound for the weighted risk which extends the result of Chen and Reiß (2008), who have recently shown a lower bound of the mean integrated squared error.

Theorem 2.1 *Suppose that the i.i.d. (Y, Z, W)-sample of size n obeys the model (1.1a–1.1b), that the error term U belongs to $\mathcal{U}_\sigma := \{U : \mathbb{E}U|W = 0$ and $\mathbb{E}U^4|W \leqslant \sigma^4\}$, $\sigma > 0$ and that $\sup_{j\geqslant 1} \mathbb{E}[e_j^4(Z)|W] \leqslant \eta$, $\eta \geqslant 1$. Consider sequences γ, ω and λ satisfying Assumption A1 such that the conditional expectation operator T associated to (Z, W) belongs to $\mathcal{T}_\lambda^d$, $d \geqslant 1$. Define for all $n \geqslant 1$*

$$k_n^* := k_n^*(\gamma, \lambda, \omega) := \operatorname*{argmin}_{k\in\mathbb{N}} \Big\{ \max\Big(\frac{\omega_k}{\gamma_k}, \sum_{j=1}^{k} \frac{\omega_j}{n\lambda_j} \Big) \Big\} \text{ and}$$

$$R_n^* := R_n^*(\gamma, \lambda, \omega) := \max\Big(\frac{\omega_{k_n^*}}{\gamma_{k_n^*}}, \sum_{j=1}^{k_n^*} \frac{\omega_j}{n\lambda_j} \Big). \quad (2.4)$$

If in addition $\kappa := \inf_{n\geqslant 1}\{(R_n^*)^{-1} \min(\omega_{k_n^*}\gamma_{k_n^*}^{-1}, \sum_{l=1}^{k_n^*} \omega_l (n\lambda_l)^{-1})\} > 0$ and $\sigma^4 \geqslant 8(3+2\rho^2\Gamma^2)$, then for all $n \geqslant 1$ and for any estimator $\widetilde{\varphi}$ of φ, we have

$$\sup_{U\in\mathcal{U}_\sigma} \sup_{\varphi\in\mathcal{F}_\gamma^\rho} \mathbb{E}\|\widetilde{\varphi} - \varphi\|_\omega^2 \geqslant \frac{\kappa}{4} \min\left(\rho, \frac{1}{2d}\right) R_n^*.$$

Remark 2.2 The proof of the last assertion is based on Assuoad's cube technique (c.f. Korostolev and Tsybakov (1993)), which consists in constructing $2^{k_n^*}$ candidates of structural functions which have the largest possible $\|\cdot\|_\omega$-distance but are still statistically non distinguishable. In the last theorem, the additional moment condition $\sup_{j\geqslant 1} \mathbb{E}[e_j^4(Z)|W] \leqslant \eta$ is obviously satisfied if the basis functions $\{e_j\}$ are uniformly bounded (e.g. the trigonometric basis considered in Section 2.4). However, if V denotes a Gaussian random variable with mean zero and variance one, which is moreover independent of (Z, W), then the additional condition $\sigma^4 \geqslant 8(1 + 2\rho^2\Gamma^2\eta)$ ensures that for all structural functions $\varphi \in \mathcal{F}_\gamma^r$, the error term $U := V - \varphi(Z) + [T\varphi](W)$ belongs to $\mathcal{U}_\sigma$. This specific case is only needed to simplify the calculation of the distance between distributions corresponding to different structural functions (a similar assumption has been used by Chen and Reiß (2008)). On the other hand, below we derive an upper bound assuming that the error term U belongs to $\mathcal{U}_\sigma$ and that the joint distribution of (Z, W) fulfills additional moment conditions. In this situation, Theorem 2.1 obviously provides a lower bound for any estimator as long as σ is sufficiently large. Note further that this lower bound tends only to zero if $(\omega_j/\gamma_j)_{j\geqslant 1}$ is a null sequence. In other words, in case $\gamma \equiv 1$, uniform consistency over all φ such that $\|\varphi\|_Z^2 \leqslant \rho$ can only be achieved with respect to a weighted norm weaker than the L^2_Z-norm, that is, if ω is a zero-sequence. This obviously reflects the ill-posedness of the underlying inverse problem. Finally, it is important to note that the regularity conditions imposed on the structural function φ and the conditional expectation operator T involve only the basis $\{e_j\}_{j\geqslant 1}$ in L^2_Z. Therefore, the lower bound derived in Theorem 2.1 does not capture the influence of the basis $\{f_l\}_{l\geqslant 1}$ in L^2_W used to construct the estimator. In other words, the proposed estimator of φ can only attain this lower bound if $\{f_l\}_{l\geqslant 1}$ is appropriately chosen. □

2.3. Minimax-optimal Estimation by dimension reduction and thresholding.

In addition to the basis $\{e_j\}_{j\geqslant 1}$ of L^2_Z considered in the last section, we introduce now also a basis $\{f_l\}_{l\geqslant 1}$ in L^2_W. In this section we derive the asymptotic properties of the least squares estimator under minimal assumptions on these two bases. More precisely, we suppose that the structural function φ belongs to some ellipsoid $\mathcal{F}^r_\gamma$ and that the conditional expectation satisfies a link condition, i.e., $T \in \mathcal{T}^\lambda_d$. Furthermore, we introduce an additional condition linked to the basis $\{f_l\}_{l\geqslant 1}$. Then, under slightly stronger moment conditions, we show that the proposed estimator attains the lower bound derived in the last section. All these results are illustrated under classical smoothness assumptions at the end of this section.

Matrix and operator notations. Given $k \geqslant 1$, $\mathcal{E}_k$ and $\mathcal{F}_k$ denote the subspace of L^2_Z and L^2_W spanned by the functions $\{e_j\}_{j=1}^k$ and $\{f_l\}_{l=1}^k$, respectively. E_k and $E_k^\perp$ (resp. F_k and $F_k^\perp$) denote the orthogonal projections on $\mathcal{E}_k$ (resp. $\mathcal{F}_k$) and its orthogonal complement $\mathcal{E}_k^\perp$ (resp. $\mathcal{F}_k^\perp$), respectively. Given an operator (matrix) K, the inverse operator (matrix) of K is denoted by K^{-1}, the adjoint (transposed) operator (matrix) of K by K^t. $[\varphi]$, $[\psi]$ and $[K]$ denote the (infinite) vector and matrix of the function $\varphi \in L^2_Z$, $\psi \in L^2_W$ and the operator $K : L^2_Z \to L^2_W$ with the entries $[\varphi]_j = \langle \varphi, e_j\rangle$, $[\psi]_l = \langle \psi, f_l\rangle$ and $[K]_{lj} = \langle Ke_j, f_l\rangle$, respectively. The upper k subvector and $k \times k$ submatrix of $[\varphi]$, $[\psi]$ and $[K]$ is denoted by $[\varphi]_{\underline{k}}$, $[\psi]_{\underline{k}}$ and $[K]_{\underline{k}}$, respectively. Note, that $[K^t]_{\underline{k}} = [K]^t_{\underline{k}}$. The diagonal matrix with entries v is denoted by $\mathrm{diag}(v)$ and the identity operator (matrix) is denoted by I. Clearly, $[E_k\varphi]_{\underline{k}} = [\varphi]_{\underline{k}}$ and if we restrict F_kKE_k to an operator from $\mathcal{E}_k$ into $\mathcal{F}_k$, then it has the matrix $[K]_{\underline{k}}$. Moreover, if $v \in \mathbb{R}^k$ then $\|v\|$ denotes the Euclidean norm of v and given a $(k \times k)$ matrix M let $\|M\| := \sup_{\|v\|\leqslant 1}\|Mv\|$ denote its spectral-norm and $\mathrm{tr}(M)$ its trace. Consider the conditional expectation operator T associated to the regressor Z and the instrument W. If $[e(Z)]$ and $[f(W)]$ denote the infinite random vector with entries $e_j(Z)$ and $f_j(W)$ respectively, then $[T]_{\underline{k}} = \mathbb{E}[f(W)]_{\underline{k}}[e(Z)]^*_{\underline{k}}$ which is throughout the paper assumed to be non singular for all $k \geqslant 1$ (or, at least for sufficiently large k), so that $[T]^{-1}_{\underline{k}}$ always exists. Note that it is a nontrivial problem to determine in under what precise conditions such an assumption holds (see e.g. Efromovich and Koltchinskii (2001) and references therein).

Definition of the estimator. Let $(Y_1, Z_1, W_1), \ldots, (Y_n, Z_n, W_n)$ be an i.i.d. sample of (Y, Z, W). Since $[T]_{\underline{k}} = \mathbb{E}[f(W)]_{\underline{k}}[e(Z)]^t_{\underline{k}}$ and $[g]_{\underline{k}} = \mathbb{E}Y[f(W)]_{\underline{k}}$ we construct estimators by using their empirical counterparts, that is,

$$[\widehat{T}]_{\underline{k}} := (1/n)\sum_{i=1}^n [f(W_i)]_{\underline{k}}[e(Z_i)]^t_{\underline{k}} \quad \text{and} \quad [\widehat{g}]_{\underline{k}} := (1/n)\sum_{i=1}^n Y_i[f(W_i)]_{\underline{k}}. \tag{2.5}$$

Then the estimator of the structural function φ is defined by

$$\widehat{\varphi}_k := \sum_{j=1}^{k} [\widehat{\varphi}_k]_j e_j \qquad \text{with} \qquad [\widehat{\varphi}_k]_{\underline{k}} := \begin{cases} [\widehat{T}]_{\underline{k}}^{-1}[\widehat{g}]_{\underline{k}}, & \text{if } [\widehat{T}]_{\underline{k}} \text{ is nonsingular} \\ & \text{and } \|[\widehat{T}]_{\underline{k}}^{-1}\| \leqslant \sqrt{n}, \\ 0, & \text{otherwise,} \end{cases} \tag{2.6}$$

where the dimension parameter $k = k(n)$ has to tend to infinity as the sample size n increases. In fact, the estimator $\widehat{\varphi}_k$ takes its inspiration from the linear Galerkin approach used in the inverse problem community (c.f. Efromovich and Koltchinskii (2001) or Hoffmann and Reiß (2008)).

Extended link condition. Consistency of this estimator is only possible if the least squares solution $\varphi_k = \sum_{j=1}^{k} [\varphi_k]_j e_j$ with $[\varphi_k]_{\underline{k}} = [T]_{\underline{k}}^{-1}[g]_{\underline{k}}$ converges to the structural function φ as $k \to \infty$, which is not true in general. However, the condition $\sup_{k\in\mathbb{N}} \|[T]_{\underline{k}}^{-1}[TE_k^{\perp}]_{\underline{k}}\| < \infty$ is known to be necessary to ensure convergence of φ_k. Notice that this condition involves now also the basis $\{f_l\}_{l\geqslant 1}$ in L^2_W. In what follows we introduce an alternative but stronger condition to guarantee the convergence, which extends the link condition (2.3), that is, $T \in \mathcal{T}_d^{\lambda}$. We denote by $\mathcal{T}_{d,D}^{\lambda}$ for some $D \geqslant d$ the subset of $\mathcal{T}_d^{\lambda}$ given by

$$\mathcal{T}_{d,D}^{\lambda} := \Big\{ T \in \mathcal{T}_d^{\lambda} : \quad \sup_{k\in\mathbb{N}} \|[\mathrm{diag}(\lambda)]_{\underline{k}}^{1/2}[T]_{\underline{k}}^{-1}\|^2 \leqslant D \Big\}. \tag{2.7}$$

Remark 2.3 The link condition (2.3) implies the extended link condition (2.7) for a suitable $D > 0$ if $\{e_j\}$ and $\{f_j\}$ are the eigenfunctions of T and if $[T]$ is only a small perturbation of $\mathrm{diag}(\lambda^{1/2})$, or if T is strictly positive (for a detailed discussion we refer to Efromovich and Koltchinskii (2001) and Cardot and Johannes (2010)). We underline that once both bases $\{e_j\}_{j\geqslant 1}$ and $\{f_l\}_{l\geqslant 1}$ are specified, the extended link condition (2.7) restricts the class of joint distributions of (Z, W) to those for which the least squares solution φ_k is L^2-consistent. Moreover, we show below that under the extended link condition the least squares estimator of φ given in (2.6) can attain minimax-optimal rates of convergence. In this sense, given a joint distribution of (Z, W), a basis $\{f_l\}_{l\geqslant 1}$ satisfying the extended link condition can be interpreted as a set of optimal instruments. Moreover, for each pre-specified basis $\{e_j\}_{j\geqslant 1}$, we can theoretically construct a basis $\{f_l\}_{l\geqslant 1}$ of optimal instruments such that the extended link condition is not a stronger restriction than the link condition (2.3) (see Johannes and Breunig (2009) for more details). □

The upper bound. The following theorem provides an upper bound under the extended link condition (2.7) and an additional moment condition on the bases, more specific, on the random vectors $[e(Z)]$ and $[f(W)]$. We begin this section by formalizing this additional condition.

Assumption A2 *There exists $\eta \geqslant 1$ such that the joint distribution of (Z, W) satisfies*

(i) $\sup_{j\in N} \mathbb{E}[e_j^2(Z)|W] \leqslant \eta^2$ *and* $\sup_{l\in N} \mathbb{E}[f_l^4(W)] \leqslant \eta^4$*;*

(ii) $\sup_{j,l\in\mathbb{N}} \mathbb{V}\mathrm{ar}(e_j(Z)f_l(W)) \leqslant \eta^2$ *and*
$\sup_{j,l\in\mathbb{N}} \mathbb{E}|e_j(Z)f_l(W) - \mathbb{E}[e_j(Z)f_l(W)]|^8 \leqslant 8!\eta^6\, \mathbb{V}\mathrm{ar}(e_j(Z)f_l(W))$*.*

It is worth noting that any joint distribution of (Z,W) satisfies Assumption A2 for sufficiently large η if the bases $\{e_j\}_{j\geqslant 1}$ and $\{f_l\}_{l\geqslant 1}$ are uniformly bounded. Here and subsequently, we write $a_n \lesssim b_n$ when there exists a numerical constant $C>0$ such that $a_n \leqslant C\, b_n$ for all $n\in\mathbb{N}$ and $a_n \sim b_n$ when $a_n \lesssim b_n$ and $b_n \lesssim a_n$ simultaneously.

Theorem 2.4 *Suppose that the i.i.d. (Y,Z,W)-sample of size n obeys the model (1.1a–1.1b) and that the joint distribution of (Z,W) fulfills Assumption A2 for some $\eta\geqslant 1$. Consider sequences γ, ω and λ satisfying Assumption A1 such that the conditional expectation operator T associated to (Z,W) belongs to $\mathcal{T}_{d,D}^{\lambda}$, $d,D\geqslant 1$. Let k_n^*, R_n^* and κ be as given in Theorem (2.1). If in addition $\sup_{k\in\mathbb{N}} k^3/\gamma_k =: \zeta < \infty$, then we have for all $n\in\mathbb{N}$ with $(k_n^*)^3 \geqslant 4D\zeta/\kappa$ that*

$$\sup_{U\in\mathcal{U}_\sigma}\sup_{\varphi\in\mathcal{F}_\gamma^\rho} \mathbb{E}\|\widehat{\varphi}_{k_n^*}-\varphi\|_\omega^2 \lesssim D\,\eta^4\,(\sigma^2+4\Gamma\,D\,d\,\rho)\,R_n^*$$

$$\cdot\Big\{4D\zeta/\kappa + \max\Big(1, \frac{\lambda_{k_n^*}}{\omega_{k_n^*}}\max_{1\leqslant j\leqslant k_n^*}\frac{\omega_j}{\lambda_j}\Big) + (k_n^*)^3\Big|P\Big(\|[\widehat{T}]_{\underline{k_n^*}}-[T]_{\underline{k_n^*}}\|^2 > \frac{\lambda_{k_n^*}}{4D}\Big)\Big|^{1/4}\Big\}$$

$$+\,\rho P\Big(\|[\widehat{T}]_{\underline{k_n^*}}-[T]_{\underline{k_n^*}}\|^2 > \frac{\lambda_{k_n^*}}{4D}\Big).$$

Remark 2.5 We emphasize that the bound in the last theorem is not asymptotic. Moreover, it is worth noting that the term $\max\big(1, \lambda_{k_n^*}/\omega_{k_n^*}\max_{1\leqslant j\leqslant k_n^*}\omega_j/\lambda_j\big)$ is uniformly bounded by a constant if ω/λ is non decreasing, which we suppose from now on. However, this is not the case in general. □

A comparison with the lower bound from Theorem 2.1 shows that the last assertion does not establish the minimax-optimality of the estimator. However, the upper bound in Theorem 2.4 can be improved by imposing a moment condition stronger than Assumption A2. To be more precise, consider the centered random variable $e_j(Z)f_l(W)-\mathbb{E}[e_j(Z)f_l(W)]$. Then Assumption A2 (ii) states that its 8th moment is uniformly bounded over $j,l\in\mathbb{N}$. In the next Assumption we suppose that these random variables satisfy Cramer's condition uniformly, which is known to be sufficient to obtain an exponential bound for their large deviations (c.f. Bosq (1998)).

Assumption A3 *There exists $\eta\geqslant 1$ such that the joint distribution of (Z,W) satisfies Assumption A2 and in addition*

(iii) $\sup_{j,l\in\mathbb{N}}\mathbb{E}|e_j(Z)f_l(W)-\mathbb{E}[e_j(Z)f_l(W)]|^k \leqslant \eta^{k-2}k!\,\mathbb{V}\mathrm{ar}(e_j(Z)f_l(W))$, $k=3,4,\dots$.

It is well-known that Cramer's condition is fulfilled in particular if the random variable $e_j(Z)f_l(W)-\mathbb{E}[e_j(Z)f_l(W)]$ is bounded. Whenever the bases $\{e_j\}_{j\geqslant 1}$ and $\{f_l\}_{l\geqslant 1}$ are uniformly bounded it follows thus again that any joint distribution of (Z,W) satisfies Assumption A3 for sufficiently large η. On the other hand, we show that under this additional condition the deviation probability tends to zero faster than R_n^*. Hence, the rate R_n^* is optimal and $\widehat{\varphi}_{k_n^*}$ is minimax-optimal, which is summarized in the next assertion.

Theorem 2.6 *Suppose that the assumptions of Theorem 2.4 are satisfied. In addition, assume that the joint distribution of (Z,W) fulfills Assumption A3 and that the sequence (ω/λ) is non-decreasing. For all $n\in\mathbb{N}$ with $(\log k_n^*)/k_n^* \leqslant \kappa/(280D\eta^2\zeta)$ and $(\log R_n^*)/k_n^* \geqslant -\kappa/(40D\eta^2\zeta)$ we have*

$$\sup_{U\in\mathcal{U}_\sigma}\sup_{\varphi\in\mathcal{F}_\gamma^\rho} \mathbb{E}\|\widehat{\varphi}_{k_n^*}-\varphi\|_\omega^2 \lesssim D^2\,\eta^4\,\zeta\,\kappa^{-1}(\sigma^2+\Gamma\,D\,d\,\rho)\,R_n^*.$$

Remark 2.7 From Theorems 2.1 and 2.6 follows that the estimator $\widehat{\varphi}_{k_n^*}$ attains the optimal rate R_n^* for all sequences γ, ω and λ satisfying the minimal regularity conditions from Assumption A1. Let us elaborate on the interesting role of the sequences γ, ω and λ. Theorem 2.1 and 2.6 show that the faster the sequence λ decreases, the slower the obtainable optimal rate of convergence becomes. On the other hand, a faster increase of γ or decrease of ω leads to a faster optimal rate. In other words, as expected, a structural function satisfying a stronger regularity condition can be estimated faster, and measuring the accuracy with respect to a weaker norm leads to faster rates, too. $\square$

2.4. Illustration: estimation of derivatives.

To illustrate the previous results, we will describe in this section the prior information about the unknown structural function φ by its level of smoothness. In order to simplify the presentation, we follow Hall and Horowitz (2005) (where a more detailed discussion of this assumption can be found) and suppose that the marginal distribution of the scalar regressor Z and the scalar instrument W are uniformly distributed on the interval $[0,1]$. It is worth noting that all the results below can be extended to the multivariate case in a straightforward way. In the univariate case, it follows that both Hilbert spaces L_Z^2 and L_W^2 are isomorphic to $L^2[0,1]$, endowed with the usual norm $\|\cdot\|$ and inner product $\langle\cdot,\cdot\rangle$.
In the last sections, we have seen that the choice of the basis $\{e_j\}_{j\geqslant 1}$ is directly linked to the a priori assumptions we are willing to impose on the structural function. In case of classical smoothness assumptions, it is natural to consider the Sobolev space of periodic functions $\mathcal{W}_r$, $r \geqslant 0$, which for an integer r is given by

$$\mathcal{W}_r = \Big\{ f \in H_p : f^{(j)}(0) = f^{(j)}(1), \quad j = 0, 1, \dots, r-1 \Big\},$$

where $H_r := \{f \in L^2[0,1] : f^{(r-1)}$ absolutely continuous , $f^{(r)} \in L^2[0,1]\}$ is a Sobolev space. Moreover, let us introduce the trigonometric basis

$$\psi_1 :\equiv 1,\ \psi_{2j}(s) := \sqrt{2}\cos(2\pi js),\ \psi_{2j+1}(s) := \sqrt{2}\sin(2\pi js), s \in [0,1],\ j \in \mathbb{N}.$$

It is well-known that the union $\bigcup_{n\in\mathbb{N}} \mathcal{F}_w^n$ of ellipsoids $\mathcal{F}_w^n$ in $L^2[0,1]$ defined by using the trigonometric basis $\{e_j = \psi_j\}$ and the weight sequence $w_1 = 1$, $w_j = j^{2r}$, $j \geqslant 2$ in definition (2.2) coincides with the Sobolev space of periodic functions $\mathcal{W}_r$ (c.f. Neubauer (1988a,b)). Therefore, let us denote by $\mathcal{W}_r^c := \mathcal{F}_w^c$, $c > 0$ an ellipsoid in the Sobolev space $\mathcal{W}_r$. In the remainder of this section we will suppose that the prior information about the unknown structural function φ is characterized by the Sobolev ellipsoid $\mathcal{W}_p^\rho$, $\rho > 0$, i.e., that φ is $p \geqslant 0$ times differentiable. In this illustration, we consider the estimation of derivatives of the structural function φ. Therefore, it is interesting to recall that, up to a constant, for any function $h \in \mathcal{W}_p^\rho$ the weighted norm $\|h\|_\omega$ with $\omega_0 = 1$ and $\omega_j = j^{2s}$, $j \geqslant 2$, equals the L^2-norm of the s-th weak derivative $h^{(s)}$ for each integer $0 \leqslant s \leqslant p$. By virtue of this relation, the results in the previous section imply also a lower as well as an upper bound of the L^2-risk for the estimation of the s-th weak derivative of φ. Finally, we restrict our attention to conditional expectation operator $T \in \mathcal{T}_d^\lambda$ with either

[p-$\boldsymbol{\lambda}$] a polynomially decreasing sequence λ, i.e., $\lambda_0 = 1$ and $\lambda_j = j^{-2a}$, $j \geqslant 2$, for some $a > 0$, or

[e-$\boldsymbol{\lambda}$] an exponentially decreasing sequence λ, i.e., $\lambda_0 = 1$ and $\lambda_j = \exp(-j^{2a})$, $j \geqslant 2$, for some $a > 0$.

It is easily seen that the minimal regularity conditions given in Assumption A1 are satisfied if $p > 1/2$. Roughly speaking, this means that the structural function is at least continuous. The lower bound presented in the next assertion follows now directly from Theorem 2.1. Note that the additional condition, $\sup_{j\geqslant 1}\mathbb{E}[e_j^4(Z)|W] \leqslant \eta$, $\eta \geqslant 8$, is satisfied since the trigonometric basis is bounded uniformly by two.

Proposition 2.8 *Suppose an i.i.d. sample of size n from the model (1.1a–1.1b). If $\varphi \in \mathcal{W}_p^r$, $p > 1/2$, then we have for any estimator $\widetilde{\varphi}^{(s)}$ of $\varphi^{(s)}$, $0 \leqslant s < p$,*

[p-$\boldsymbol{\lambda}$] *in the polynomial decreasing case that*

$$\sup_{U\in\mathcal{U}_\sigma}\sup_{\varphi\in\mathcal{W}_p^r}\left\{\mathbb{E}\|\widetilde{\varphi}^{(s)} - \varphi^{(s)}\|^2\right\} \gtrsim n^{-2(p-s)/(2p+2a+1)},$$

[e-$\boldsymbol{\lambda}$] *in the exponentially decreasing case that*

$$\sup_{U\in\mathcal{U}_\sigma}\sup_{\varphi\in\mathcal{W}_p^r}\left\{\mathbb{E}\|\widetilde{\varphi}^{(s)} - \varphi^{(s)}\|^2\right\} \gtrsim (\log n)^{-(p-s)/a}.$$

In this section, the basis of L_W^2 is also given by the trigonometric basis $\{f_l = \psi_l\}_{l\geqslant 1}$. In this situation, the additional moment conditions formalized in Assumption A3 are automatically fulfilled since both bases $\{e_j\}_{j\geqslant 1}$ and $\{f_l\}_{l\geqslant 1}$ are uniformly bounded. We suppose that the associated conditional expectation operator T satisfies the extended link condition (2.7), that is, $T \in \mathcal{T}_{d,D}^{\lambda}$. Thereby, we restrict the set of possible joint distributions of (Z, W) to those having the trigonometric basis as optimal instruments. As an estimator of $\varphi^{(s)}$, we shall consider the s-th weak derivative of the estimator $\widehat{\varphi}_k$ defined in (2.6). Recall that for each integer $0 \leqslant s \leqslant p$, the s-th weak derivative of the estimator $\widehat{\varphi}_k$ is

$$\widehat{\varphi}_k^{(s)}(t) = \sum_{j\in\mathbb{Z}}(2i\pi j)^s\int_0^1 \widehat{\varphi}_k(u)\exp(-2i\pi ju)du\exp(-2i\pi jt).$$

Applying Theorem 2.4, the rates of the lower bound given in the last assertion provide, up to a constant, also an upper bound of the L^2-risk of the estimator $\widehat{\varphi}_k^{(s)}$, which is summarized in the next proposition. We have thus proved that these rates are optimal and the proposed estimator $\widehat{\varphi}_k^{(s)}$ is minimax optimal in both cases.

Proposition 2.9 *Suppose that the i.i.d. (Y, Z, W)-sample of size n obeys the model (1.1a–1.1b). Let $\varphi \in \mathcal{W}_p^r$, $p \geqslant 3/2$. For $0 \leqslant s < p$ consider the estimator $\widehat{\varphi}_{k_n^*}$ given in (2.6).*

[p-$\boldsymbol{\lambda}$] *In the polynomial decreasing case with dimension parameter $k_n^* \sim n^{1/(2p+2a+1)}$,*

$$\sup_{U\in\mathcal{U}_\sigma}\sup_{\varphi\in\mathcal{W}_p^r}\left\{\mathbb{E}\|\widehat{\varphi}_{k_n^*}^{(s)} - \varphi^{(s)}\|^2\right\} \lesssim n^{-2(p-s)/(2p+2a+1)}.$$

[e-$\boldsymbol{\lambda}$] *In the exponentially decreasing case with $k_n^* \sim (\log n)^{1/(2a)}$,*

$$\sup_{U\in\mathcal{U}_\sigma}\sup_{\varphi\in\mathcal{W}_p^r}\left\{\mathbb{E}\|\widehat{\varphi}_{k_n^*}^{(s)} - \varphi^{(s)}\|^2\right\} \lesssim (\log n)^{-(p-s)/a}.$$

Remark 2.10 We emphasize the interesting role of the parameters p and a characterizing the regularity conditions imposed on φ and T respectively: As we see from Theorem 2.8 and 2.9, if the value of a increases, the obtainable optimal rate of convergence decreases. Therefore, the parameter a is often called *degree of ill-posedness* (c.f. Natterer (1984)). On the other hand, an increase of the quantity p leads to a faster optimal rate. In other words, as expected, a smoother structural function can be estimated faster. Finally, as opposed to the polynomial case, in the exponential case the smoothing parameter k_n^* does not depend on the value of p. It follows that the proposed estimator is automatically adaptive, i.e., it does not depend on an a-priori knowledge of the degree of smoothness of the structural function φ. However, the choice of the smoothing parameter depends on the properties of T, more precisely, the value of a. □

3. Adaptive estimation under smoothness assumptions

In this section, our objective is to construct a fully adaptive estimator of the structural function φ. Adaptation means that in spite of the conditional expectation operator T being unknown, the estimator should attain the optimal rate of convergence over the ellipsoid $\mathcal{F}_\gamma^\rho$ for a wide range of different weight sequences γ. However, we will suppose that the operator T is diagonal with respect to the trigonometric basis $\{\psi_j\}$. In this situation, for example, an operator with polynomially decreasing λ having a degree of ill-posedness a behaves like a-times integrating, and hence it is also called *finitely smoothing*. On the other hand, when the sequence λ is exponentially decreasing with degree of ill-posedness a, the operator behaves like integrating infinitely many times, and hence it is also called *infinitely smoothing*. Thus, this additional condition imposes in fact a smoothing condition on the unknown conditional expectation operator T. Even if we assume that the operator is smoothing, we do not impose any a-priori knowledge about the specific decay of λ.
Our starting point is the estimator given in (2.6), which in this situation is of the form

$$\widehat{\varphi}_k = \sum_{j=1}^{k} \frac{\widehat{[g]}_j}{\widehat{[T]}_{jj}} \mathbb{1}\{\inf_{1\leqslant j\leqslant k} \widehat{[T]}_{jj}^2 \geqslant 1/n\}\, \psi_j, \tag{3.1}$$

with $\widehat{[g]}_j$ and $\widehat{[T]}_{jj}$ defined in (2.5). In the last section, we have shown that this estimator is minimax-optimal provided the dimension parameter k is chosen in an optimal way. In what follows, the dimension parameter k is chosen using a model selection approach via penalization. This choice will only involve the data and none of the sequences γ and λ describing the underlying smoothness. First, we introduce some sequences which are used below.

Definition 3.1

(i) *For all $k \geqslant 1$, define $\Delta_k := \max_{1\leqslant j\leqslant k} \omega_j/\lambda_j$, $\tau_k := \max_{1\leqslant j\leqslant k}(\omega_j)_{\vee 1}/\lambda_j$ with $(q)_{\vee 1} := \max(q,1)$ and*

$$\delta_k := k\Delta_k \frac{\log(\tau_k \vee (k+2))}{\log(k+2)}.$$

Let further Σ be a non-decreasing function such that for all $C > 0$

$$\sum_{k\geqslant 1} C\, \tau_k \exp\Big(- \frac{k \log(\tau_k \vee (k+2))}{6C \log(k+2)}\Big) \leqslant \Sigma(C) < \infty \tag{3.2}$$

and $\sup_{n\in\mathbb{N}} \exp\big(- K_2\, C^{-1}\, n^{1/6} + \frac{5}{3}\log n\big) \leqslant \Sigma(C)$ with $K_2 = (\sqrt{2}-1)/(21\sqrt{2})$.

(ii) *Define a sequence N follows,*

$$N_n := N_n(\lambda, d) := \max\Big\{1 \leqslant N \leqslant n \;\Big|\; n^7 \exp\Big(-\frac{n\,\lambda_N}{288d}\Big) \leqslant \Big(\frac{2016\,d}{\lambda_1}\Big)^7 \quad \text{and} \quad \delta_N/n \leqslant 1\Big\}.$$

It is easy to see that there exists always a function Σ satisfying condition (3.2). Consider the estimator $\widehat{\varphi}_{\widetilde{k}}$ defined by choosing the dimension parameter $\widetilde{k}$ such that

$$\widetilde{k} := \operatorname*{argmin}_{1\leqslant k\leqslant N_n} \left\{ -\|\widehat{\varphi}_k\|_\omega^2 + c\,\frac{\delta_k}{n} \right\}$$

for some constant $c > 0$. It is shown in Johannes and Schwarz (2010) and Comte and Johannes (2010) that such an estimator can attain minimax-optimal rates in the context of a deconvolution problem and a functional linear model respectively. However, this estimator is only partially adaptive, since the dimension parameter is chosen using a criterion function that involves the sequences N and δ which depend on λ and d. We circumvent this problem by defining empirical versions of these sequences. The fully adaptive estimator is then defined analogously to the one above, but uses the estimated rather than the original sequences.

Definition 3.2 *Let $\widehat{\delta} := (\widehat{\delta}_k)_{k\geqslant 1}$, $\widehat{N} := (\widehat{N}_n)_{n\geqslant 1}$, be as follows.*

(i) Given $\widehat{\Delta}_k := \max_{1\leqslant j\leqslant k} \omega_j \widehat{[T]}_{jj}^{-2} \mathbb{1}\{\inf_{1\leqslant j\leqslant k} \widehat{[T]}_{jj}^2 \geqslant 1/n\}$ *and* $\widehat{\tau}_k := \max_{0\leqslant j\leqslant k} (\omega_j)_{\vee 1} \widehat{[T]}_{jj}^{-2} \mathbb{1}\{\inf_{1\leqslant j\leqslant k} \widehat{[T]}_{jj}^2 \geqslant 1/n\}$ *let*

$$\widehat{\delta}_k := k\widehat{\Delta}_k\,\frac{\log(\widehat{\tau}_k \vee (k+2))}{\log(k+2)}.$$

(ii) Given $N_n^u := \operatorname{argmax}_{1\leqslant N\leqslant n} \left\{ \max_{1\leqslant j\leqslant N} \omega_j/n \leqslant 1 \right\}$, *let*

$$\widehat{N}_n := \operatorname*{argmin}_{1\leqslant j\leqslant N_n^u} \left\{ \frac{|\widehat{[T]}_j|^2}{|j|(\omega_j)_{\vee 1}} < \frac{\log n}{n} \right\}.$$

It worth to stress that all these sequences do not involve any a-priori knowledge about neither the target function φ nor the operator T. Now, we choose the dimension parameter as

$$\widehat{k} := \operatorname*{argmin}_{1\leqslant k\leqslant \widehat{N}_n} \left\{ -\|\widehat{f}_k\|_\omega^2 + 540\,\mathbb{E}[Y^2]\,\frac{\widehat{\delta}_k}{n} \right\}. \tag{3.3}$$

Throughout the paper we do not address the issue that the value $\mathbb{E}[Y^2]$ is not known in practice. Anyway, it can easily be estimated by its empirical counterpart. Moreover the constant 540, though suitable for the theory, may probably be chosen much smaller in practice by a simulation study (cf. Comte et al. (2006) in the context of a deconvolution problem).

Our main result below needs the following Assumption.

Assumption A4 *The sequence N from Definition 3.1 (ii) satisfies the conditions*

$$\max_{j\geqslant N_n} \frac{\lambda_j}{j(\omega_j)_{\vee 1}} \leqslant \frac{\log n}{4dn} \qquad \text{and} \qquad d^{-1} \min_{1\leqslant j\leqslant N_n} \lambda_j \geqslant 2/n.$$

By construction, these conditions are satisfied for sufficiently large n. However, let us illustrate them by the particular examples introduced in section 2.4.

Remark 3.3 Recall the distinction between finitely and infinitely smoothing conditional expectation operators discussed in section 2.4. The sequences from Definition 3.1 take the following values in either of the two cases.

[fs] In the finitely smoothing case, we have

$$\Delta_k = k^{2a+2s}, \quad \delta_k \sim k^{2a+2s+1}, \quad N_n \sim n^{1/(2a+2s+1)}.$$

[is] In the infinitely smoothing case, we have

$$\Delta_k = k^{2s}\exp(k^{2a}), \quad \delta_k \sim k^{2a+2s+1}\exp(k^{2a})(\log k)^{-1},$$
$$N_n \sim \left(\log \frac{n\ \log\log n}{(\log n)^{(2a+2s+1)/(2a)}}\right)^{1/(2a)}.$$

It is easily verified that the sequence N satisfies Assumption A4 in either case . □

We are now able to state the main result of this paper providing an upper risk bound for the fully adaptive estimator.

Theorem 3.4 *Assume an n-sample of (Y, Z, W). Consider sequences ω, γ, and λ satisfying Assumption A1 such that the conditional expectation operator T associated to (Z, W) belongs to $T \in \mathcal{T}_{d,D}^{\lambda}$, $d, D \geqslant 1$ and is diagonal with respect to $\{\psi_j\}$. Let the sequences δ and N be as in Definition 3.1 and suppose that Assumption A4 holds. Define further $N_n^l := \operatorname{argmax}_{1\leqslant j\leqslant N_n}\left\{\frac{\lambda_j}{j(\omega_j)\vee 1} \geqslant \frac{4d\log n}{n}\right\}$. Consider the estimator $\widehat{\varphi}_{\widehat{k}}$ defined in (3.1) with $\widehat{k}$ given by (3.3). Then for all $n \geqslant 1$*

$$\sup_{U\in\mathcal{U}_\sigma}\sup_{\varphi\in\mathcal{F}_\gamma^\rho}\left\{\mathbb{E}\|\widehat{\varphi}_{\widehat{k}}-\varphi\|_\omega^2\right\} \lesssim (2\rho\Gamma+\sigma^2+1)^4 d\,\zeta_d\Bigg[\min_{1\leqslant k\leqslant N_n^l}\left\{\max\left(\frac{\omega_k}{\gamma_k},\frac{\delta_k}{n}\right)\right\}$$
$$+\rho\max_{j\geqslant 1}\left\{\frac{\omega_j}{\gamma_j}\min\left(1,\frac{1}{n\lambda_j}\right)\right\}+\frac{1}{n}\left\{\Sigma\left(\frac{(2\rho\Gamma+\sigma^2)\zeta_d+V_{U|Z}}{V_{U|Z}^2}\right)+1\right\}\Bigg],$$

where $V_{U|Z} := \mathbb{E}[\mathbb{V}\mathrm{ar}(U|Z)]$ and $\zeta_d := (\log 3d)/\log 3$.

Compare the last assertion with the lower bound given in Theorem 2.1. It is easily seen that if (ω/λ) is non-decreasing, the second term in the upper bound of Theorem 3.4 is always smaller than the first one. Thus, in this situation the fully adaptive estimator attains the lower bound up to a constant as long as $\sup_{k\geqslant 1}\{\delta_k/(\sum_{1\leqslant j\leqslant k}\omega_j/\lambda_j)\} < \infty$ and if the optimal dimension parameter k_n^* given in Theorem 2.1 is smaller than N_n^l, which is summarized in the next assertion.

Corollary 3.5 *Let the assumptions of Theorem 3.4 be satisfied. If in addition (ω/λ) is non-decreasing, $\sup_{k\geqslant 1}\{\delta_k/(\sum_{1\leqslant j\leqslant k}\omega_j/\lambda_j)\} < \infty$ and $\sup_{n\in\mathbb{N}}(k_n^*/N_n^l) \leqslant 1$, then*

$$\sup_{U\in\mathcal{U}_\sigma}\sup_{\varphi\in\mathcal{F}_\gamma^\rho}\left\{\mathbb{E}\|\widehat{\varphi}_{\widehat{k}}-\varphi\|_\omega^2\right\} = O(R_n^*), \qquad as \qquad n\to\infty,$$

where k_n^* and R_n^* are given in (2.4).

It is worth to note that the additional assumptions in the last assertion are sufficient to establish the order optimality of the estimator, but not necessary as it is shown in the example **[is]** below.

3.1. Illustration: estimation of derivatives (continued)

The following result shows that even without any prior knowledge on the structural function φ and for all smoothing operators T, the fully adaptive penalized estimator automatically attains the optimal rate in the finitely and in the infinitely smoothing case. Recall that the computation of the dimension parameter $\widehat{k}$ given in (3.3) involves the sequence $(N_n^u)_{n\geqslant 1}$, which in our illustration satisfies $N_n^u \sim n^{1/(2s)}$ since $\omega_j = j^{2s}$, $j \geqslant 1$.

Proposition 3.6 *Suppose that the i.i.d. (Y, Z, W)-sample of size n obeys the model (1.1a–1.1b) and that $U \in \mathcal{U}_\sigma$, $\sigma > 0$. Consider the estimator $\widehat{\varphi}_{\widehat{k}}$ given in (2.6) with $\widehat{k}$ defined by (3.3).*

[fs] *In the finitely smoothing case, we obtain*

$$\sup_{U\in\mathcal{U}_\sigma} \sup_{\varphi\in\mathcal{W}_p^\rho} \left\{ \mathbb{E}\|\widehat{\varphi}_{\widehat{k}}^{(s)} - \varphi^{(s)}\|^2 \right\} = O(n^{-2(p-s)/(2p+2a+1)}).$$

[is] *In the infinitely smoothing case, we have*

$$\sup_{U\in\mathcal{U}_\sigma} \sup_{\varphi\in\mathcal{W}_p^\rho} \left\{ \mathbb{E}\|\widehat{\varphi}_{\widehat{k}}^{(s)} - \varphi^{(s)}\|^2 \right\} = O((\log n)^{-(p-s)/a}).$$

4. Concluding remarks and perspectives

We have proposed in this work a new kind of estimation procedure for the structural function and its derivatives in nonparametric instrumental regression and proved that they can attain optimal rates of convergence. These estimators require an optimal choice of a dimension parameter depending on certain characteristics of the unknown structural function and the conditional expectation operator of Z given W, which are not known in practice. By using a penalized minimum contrast estimator with randomized penalty and collection of models we have constructed a fully adaptive choice of this dimension parameter, which can attain minimax-optimal rates if the conditional expectation operator of Z given W is finitely or infinitely smoothing. However, in case the conditional expectation operator is not smoothing anymore it is still an open question if this data driven rule leads to a minimax-optimal estimation procedure. We are currently exploring this issue.

A. Proofs

A.1. Proofs of section 2

Proof of the lower bound.

Proof of Theorem 2.1. Consider (Z, W) with associated conditional expectation operator $T \in \mathcal{T}_d^\lambda$. Given $\zeta := \kappa \min(\rho, 1/(2d))$ and $\alpha_n := R_n^* (\sum_{j=1}^{k_n^*} \omega_j/(\lambda_j n))^{-1}$ we consider the function $\varphi := (\zeta\alpha_n/n)^{1/2} \sum_{j=1}^{k_n^*} \lambda_j^{-1/2} e_j$ belonging to $\mathcal{F}_\gamma^\rho$, which can be realized as follows. Since (γ/ω) is monotonically increasing it follows $\|\varphi\|_\gamma^2 \leqslant \rho\kappa(\gamma_{k_n^*}/\omega_{k_n^*}) R_n^* \leqslant \rho$ by using successively the definition of α_n and κ. Obviously for any $\theta := (\theta_j) \in \{-1, 1\}^{k_n^*}$, the function $\varphi_\theta := \sum_{j=1}^{k_n^*} \theta_j [\varphi]_j e_j$ belongs to $\mathcal{F}_\gamma^\rho$ too, and hence it is a possible candidate of the structural function. Let V be a Gaussian random variable with mean zero and variance one ($V \sim \mathcal{N}(0, 1)$) which is independent of (Z, W). Let $U_\theta := [T\varphi_\theta](W) - \varphi_\theta(Z) + V$, then U_θ belongs to $\mathcal{U}_\sigma$ for all $\sigma^4 \geqslant 8(3 + 2\rho^2\Gamma^2\eta)$. Obviously we have $\mathbb{E}U_\theta|W = 0$. Moreover, by employing twice the Cauchy-Schwarz inequality the condition $\Gamma = \sum_{j\in\mathbb{N}} \gamma_j^{-1} < \infty$ together with $\sup_j \mathbb{E}[e_j^4(Z)|W] \leqslant \eta$ implies $|\mathbb{E}f(Z)|W|^4 \leqslant \rho^2\Gamma\sum_{j\in\mathbb{N}}\gamma_j^{-1}\mathbb{E}[e_j^4(Z)|W] \leqslant \rho^2\Gamma^2\eta$ for all $f \in \mathcal{F}_\gamma^\rho$. From this estimate we conclude $\mathbb{E}[\varphi_\theta^4(Z)|W] \leqslant \eta\rho^2\Gamma^2$ and $|[T\varphi_\theta](W)|^4 \leqslant \mathbb{E}[\varphi_\theta^4(Z)|W] \leqslant \eta\rho^2\Gamma^2$. By combination of the last two bounds we obtain $\mathbb{E}[U_\theta^4|W] \leqslant 8\{2\eta\rho^2\Gamma^2+3\}$. Consequently, for each θ i.i.d. copies (Y_i, Z_i, W_i), $1 \leqslant i \leqslant n$, of (Y, Z, W) with $Y := \varphi_\theta(Z) + U_\theta$ form an n-sample of the model (1.1a–1.1b) and we denote their joint distribution by P_θ. In case of P_θ the conditional distribution of Y_i given W_i is then Gaussian with mean $[T\varphi_\theta](W_i)$ and variance 1. Furthermore, for $j = 1, \dots, k_n^*$ and each θ we introduce $\theta^{(j)}$ by $\theta_l^{(j)} = \theta_l$ for $j \neq l$ and $\theta_j^{(j)} = -\theta_j$. Then, it is easily seen that the log-likelihood of P_θ with respect to $P_{\theta^{(j)}}$ is given by

$$\log\Big(\frac{dP_\theta}{dP_{\theta^{(j)}}}\Big) = \sum_{i=1}^n 2(Y_i - [T\varphi_\theta](W_i))\theta_j[\varphi]_j[Te_j](W_i) + 2[\varphi]_j^2\sum_{i=1}^n |[Te_j](W_i)|^2.$$

Its expectation with respect to P_θ satisfies $\mathbb{E}_{P_\theta}[\log(dP_\theta/dP_{\theta^{(j)}})] = 2n[\varphi]_j^2\|Te_j\|_W^2 \leqslant 2nd[\varphi]_j^2\lambda_j$ by using $T \in \mathcal{T}_d^\lambda$. In terms of Kullback-Leibler divergence this means $KL(P_\theta, P_{\theta^{(j)}}) \leqslant 2\,d\,n\,[\varphi]_j^2\lambda_j$. Since the Hellinger distance $H(P_\theta, P_{\theta^{(j)}})$ satisfies $H^2(P_\theta, P_{\theta^{(j)}}) \leqslant KL(P_\theta, P_{\theta^{(j)}})$ it follows by employing successively the definition of φ, the property $\alpha_n \leqslant \kappa^{-1}$ and the definition of ζ that

$$H^2(P_\theta, P_{\theta^{(j)}}) \leqslant 2\,d\,n\,[\varphi]_j^2\lambda_j \leqslant 2d\,\zeta\,\alpha_n \leqslant 1. \tag{A.1}$$

Consider the Hellinger affinity $\rho(P_\theta, P_{\theta^{(j)}}) = \int\sqrt{dP_\theta dP_{\theta^{(j)}}}$ then we obtain for any estimator $\widetilde{\varphi}$ of φ that

$$\begin{aligned}\rho(P_\theta, P_{\theta^{(j)}}) &\leqslant \int\frac{|[\widetilde{\varphi}-\varphi_{\theta^{(j)}}]_j|}{|[\varphi_\theta-\varphi_{\theta^{(j)}}]_j|}\sqrt{dP_\theta dP_{\theta^{(j)}}} + \int\frac{|[\widetilde{\varphi}-\varphi_\theta]_j|}{|[\varphi_\theta-\varphi_{\theta^{(j)}}]_j|}\sqrt{dP_\theta dP_{\theta^{(j)}}}\\ &\leqslant \Big(\int\frac{|[\widetilde{\varphi}-\varphi_{\theta^{(j)}}]_j|^2}{|[\varphi_\theta-\varphi_{\theta^{(j)}}]_j|^2}dP_{\theta^{(j)}}\Big)^{1/2} + \Big(\int\frac{|[\widetilde{\varphi}-\varphi_\theta]_j|^2}{|[\varphi_\theta-\varphi_{\theta^{(j)}}]_j|^2}dP_\theta\Big)^{1/2}.\end{aligned}$$

Rewriting the last estimate by using the identity $\rho(P_\theta, P_{\theta^{(j)}}) = 1 - \frac{1}{2}H^2(P_\theta, P_{\theta^{(j)}})$ and (A.1) we obtain

$$\Big\{\mathbb{E}_\theta|[\widetilde{\varphi}-\varphi_\theta]_j|^2 + \mathbb{E}_{\theta^{(j)}}|[\widetilde{\varphi}-\varphi_{\theta^{(j)}}]_j|^2\Big\} \geqslant \frac{1}{8}|[\varphi_\theta-\varphi_{\theta^{(j)}}]_j|^2 = \frac{1}{2}[\varphi]_j^2.$$

Combining the last lower bound and the following reduction scheme is the key argument of this proof:

$$\begin{aligned}\sup_{U\in\mathcal{U}_\sigma}\sup_{\varphi\in\mathcal{F}_\gamma^\rho}\mathbb{E}_{P_\theta}\|\widetilde{\varphi}-\varphi\|_\omega^2 &\geqslant \sup_{\theta\in\{-1,1\}^{k_n^*}}\mathbb{E}_{P_\theta}\|\widetilde{\varphi}-\varphi_\theta\|_\omega^2\\ &\geqslant \frac{1}{2^{k_n^*}}\sum_{\theta\in\{-1,1\}^{k_n^*}}\sum_{j=1}^{k_n^*}\omega_j\mathbb{E}_{P_\theta}|[\widetilde{\varphi}-\varphi_\theta]_j|^2\\ &= \frac{1}{2^{k_n^*}}\sum_{\theta\in\{-1,1\}^{k_n^*}}\sum_{j=1}^{k_n^*}\frac{\omega_j}{2}\Big\{\mathbb{E}_{P_\theta}|[\widetilde{\varphi}-\varphi_\theta]_j|^2 + \mathbb{E}_{P_{\theta^{(j)}}}|[\widetilde{\varphi}-\varphi_{\theta^{(j)}}]_j|^2\Big\}\\ &\geqslant \frac{1}{2^{k_n^*}}\sum_{\theta\in\{-1,1\}^{k_n^*}}\sum_{j=1}^{k_n^*}\frac{\omega_j}{4}[\varphi]_j^2 = \frac{\zeta\alpha_n}{4}\sum_{j=1}^{k_n^*}\frac{\omega_j}{n\lambda_j}.\end{aligned}$$

Hence, from the definition of ζ and α_n we obtain the lower bound given in the theorem. □

Proof of the upper bounds. We begin by defining and recalling notations to be used in the proofs of this section. Given $k > 0$, denote $\varphi_k := \sum_{j=1}^k [\varphi_k]_j e_j$ with $[\varphi_k]_{\underline{k}} = [T]_{\underline{k}}^{-1}[g]_{\underline{k}}$ which is well-defined since $[T]_{\underline{k}}$ is non singular. Then, the identities $[T(\varphi - \varphi_k)]_{\underline{k}} = 0$ and $[\varphi_k - E_k\varphi]_{\underline{k}} = [T]_{\underline{k}}^{-1}[TE_k^{\perp}\varphi]_{\underline{k}}$ hold true. Furthermore, let $[\Xi]_{\underline{k}} := [\widehat{T}]_{\underline{k}} - [T]_{\underline{k}}$ and define vector $[B]_{\underline{k}}$ and $[S]_{\underline{k}}$ by

$$[B]_j := \frac{1}{n}\sum_{i=1}^n U_i f_j(W_i),\ [S]_j := \frac{1}{n}\sum_{i=1}^n f_j(W_i)\{\varphi(Z_i) - [\varphi_k]_{\underline{k}}^t[e(Z_i)]_{\underline{k}}\},\ 1 \leqslant j \leqslant k,$$

where $[\widehat{g}]_{\underline{k}} - [\widehat{T}]_{\underline{k}}[\varphi_k]_{\underline{k}} = [B]_{\underline{k}} + [S]_{\underline{k}}$. Note that $\mathbb{E}[B]_{\underline{k}} = 0$ due to the mean independence, i.e., $\mathbb{E}(U|W) = 0$, and that $\mathbb{E}[S]_{\underline{k}} = [T\varphi]_{\underline{k}} - [T\varphi_k]_{\underline{k}} = 0$. Moreover, let us introduce the events

$$\Omega := \{\|[\widehat{T}]_{\underline{k}}^{-1}\| \leqslant \sqrt{n}\}, \quad \Omega_{1/2} := \{\|[\Xi]_{\underline{k}}\|\,\|[T]_{\underline{k}}^{-1}\| \leqslant 1/2\}$$
$$\Omega^c := \{\|[\widehat{T}]_{\underline{k}}^{-1}\| > \sqrt{n}\} \quad \text{and} \quad \Omega_{1/2}^c = \{\|[\Xi]_{\underline{k}}\|\,\|[T]_{\underline{k}}^{-1}\| > 1/2\}.$$

Observe that $\Omega_{1/2} \subset \Omega$ in case $\sqrt{n} \geqslant 2\|[T]_{\underline{k}}^{-1}\|$. Indeed, if $\|[\Xi]_{\underline{k}}\|\|[T]_{\underline{k}}^{-1}\| \leqslant 1/2$ then the identity $[\widehat{T}]_{\underline{k}} = [T]_{\underline{k}}\{I + [T]_{\underline{k}}^{-1}[\Xi]_{\underline{k}}\}$ implies $\|[\widehat{T}]_{\underline{k}}^{-1}\| \leqslant 2\|[T]_{\underline{k}}^{-1}\|$ by the usual Neumann series argument. Moreover, in case T satisfies the extended link condition (2.7), that is $T \in \mathcal{T}_{d,D}^{\lambda}$, then $2\|[T]_{\underline{k}}^{-1}\| \leqslant 2\|[\mathrm{diag}(\lambda)]_{\underline{k}}^{-1/2}\|\|[\mathrm{diag}(\lambda)]_{\underline{k}}^{1/2}[T]_{\underline{k}}^{-1}\| \leqslant 2\sqrt{D/\lambda_k}$ since λ is non increasing. Finally, given k_n^*, R_n^* and κ defined in Theorem 2.1 we have $\kappa^{-1}\omega_{k_n^*}\gamma_{k_n^*}^{-1} \geqslant R_n^* \geqslant \sum_{j=1}^{k_n^*}\omega_j(n\lambda_j)^{-1}$ by using successively the definition of κ and R_n^*. By combination of the last estimate and the condition $\sup_{k\in\mathbb{N}} k^3\gamma_k^{-1} \leqslant \zeta$ it follows that $(k_n^*)^3(n\lambda_{k_n^*})^{-1} \leqslant \kappa^{-1}(k_n^*)^3\gamma_{k_n^*}^{-1} \leqslant \kappa^{-1}\zeta$. Thus, for all $n \in \mathbb{N}$ with $(k_n^*)^3 \geqslant 4D\zeta\kappa^{-1}$ we have $4\|[T]_{\underline{k_n^*}}^{-1}\|^2 \leqslant 4D\lambda_{k_n^*}^{-1} \leqslant n4D\zeta\kappa^{-1}(k_n^*)^{-3} \leqslant n$, and hence $\Omega_{\mathbb{1}/2} \subset \Omega$. These notations and results will be used below without further reference.
We shall prove in the end of this section three technical lemmas (A.1 – A.3) which are used in the following proofs.

Proof of Theorem 2.4. Define $\widetilde{\varphi}_{k_n^*} := \varphi_{k_n^*}\mathbb{1}_\Omega$ and decompose the risk into two terms,

$$\mathbb{E}\|\widehat{\varphi}_{k_n^*} - \varphi\|_\omega^2 \leqslant 2\{\mathbb{E}\|\widehat{\varphi}_{k_n^*} - \widetilde{\varphi}_{k_n^*}\|_\omega^2 + \mathbb{E}\|\widetilde{\varphi}_{k_n^*} - \varphi\|_\omega^2\} =: 2\{A_1 + A_2\}, \tag{A.2}$$

which we bound separately. Consider first A_2. By combination of $\Omega^c \subset \Omega_{1/2}^c$ and the identity $\|\widetilde{\varphi}_{k_n^*} - \varphi\|_\omega^2 = \|\varphi_{k_n^*} - \varphi\|_\omega^2\mathbb{1}_\Omega + \|\varphi\|_\omega^2\mathbb{1}_{\Omega^c}$ we deduce $\mathbb{E}\|\widetilde{\varphi}_{k_n^*} - \varphi\|_\omega^2 \leqslant \|\varphi_{k_n^*} - \varphi\|_\omega^2 + \|\varphi\|_\omega^2 P(\Omega_{1/2}^c)$. Since (ω/γ) is monotonically decreasing, the last estimate together with (A.12) in Lemma A.2 implies for all $\varphi \in \mathcal{F}_\gamma^\rho$

$$\mathbb{E}\|\widetilde{\varphi}_{k_n^*} - \varphi\|_\omega^2 \leqslant 4\,D\,d\,\rho\,R_n^*\max\left(1, \frac{\lambda_{k_n^*}}{w_{k_n^*}}\max_{1\leqslant j\leqslant k_n^*}\frac{w_j}{\lambda_j}\right) + \rho P(\Omega_{1/2}^c) \tag{A.3}$$

by employing the definition of R_n^*.
Consider A_1. From the identity $[\widehat{g}]_{\underline{k_n^*}} - [\widehat{T}]_{\underline{k_n^*}}[\varphi_m]_{\underline{k_n^*}} = [B]_{\underline{k_n^*}} + [S]_{\underline{k_n^*}}$ follows

$$\begin{aligned}[\widehat{\varphi}_{k_n^*} - \widetilde{\varphi}_{k_n^*}]_{\underline{k_n^*}} &= \{[T]_{\underline{k_n^*}}^{-1} + [T]_{\underline{k_n^*}}^{-1}([T]_{\underline{k_n^*}} - [\widehat{T}]_{\underline{k_n^*}})[\widehat{T}]_{\underline{k_n^*}}^{-1}\}\{[B]_{\underline{k_n^*}} + [S]_{\underline{k_n^*}}\}\mathbb{1}_\Omega \\ &= [T]_{\underline{k_n^*}}^{-1}\{[B]_{\underline{k_n^*}} + [S]_{\underline{k_n^*}}\}\mathbb{1}_\Omega - [T]_{\underline{k_n^*}}^{-1}[\Xi]_{\underline{k_n^*}}[\widehat{T}]_{\underline{k_n^*}}^{-1}\{[B]_{\underline{k_n^*}} + [S]_{\underline{k_n^*}}\}\mathbb{1}_\Omega.\end{aligned}$$

By making use of this identity we decompose A_1 further into two terms

$$\mathbb{E}\|\widehat{\varphi}_{k_n^*}-\widetilde{\varphi}_{k_n^*}\|_\omega^2 \leqslant 2\mathbb{E}[\|[\mathrm{diag}(\omega)]_{k_n^*}^{1/2}[T]_{k_n^*}^{-1}\{[B]_{k_n^*}+[S]_{k_n^*}\}\|^2\mathbb{1}_\Omega] \\ + 2\mathbb{E}[\|[\mathrm{diag}(\omega)]_{k_n^*}^{1/2}[T]_{k_n^*}^{-1}[\Xi]_{k_n^*}[\widehat{T}]_{k_n^*}^{-1}\{[B]_{k_n^*}+[S]_{k_n^*}\}\|^2\mathbb{1}_\Omega] =: 2\{A_{11}+A_{12}\} \quad \text{(A.4)}$$

which we bound separately. In case of A_{11} we employ successively (A.11) in Lemma A.1 with $M := [\mathrm{diag}(\omega)]_{k_n^*}^{1/2}[T]_{k_n^*}^{-1}$, the elementary inequality $\mathrm{tr}(A^tB^tBA) \leqslant \|A\|^2\,\mathrm{tr}(B^tB)$ valid for all $(k\times k)$ matrices A and B and the extended link condition (2.7), that is, $\|[\mathrm{diag}(\lambda)]_{k_n^*}^{1/2}[T]_{k_n^*}^{-1}\|^2 \leqslant D$. Thereby, we obtain

$$\mathbb{E}[\|[\mathrm{diag}(\omega)]_{\underline{k_n^*}}^{1/2}[T]_{\underline{k_n^*}}^{-1}\{[B]_{\underline{k_n^*}}+[S]_{\underline{k_n^*}}\}\|^2\mathbb{1}_\Omega] \\ \leqslant (2/n)\,D\,\mathrm{tr}\left([\mathrm{diag}(\lambda)]_{\underline{k_n^*}}^{-1/2}[\mathrm{diag}(\omega)]_{\underline{k_n^*}}[\mathrm{diag}(\lambda)]_{\underline{k_n^*}}^{-1/2}\right)\{\sigma^2+\eta^2\,\Gamma\,\|\varphi-\varphi_{k_n^*}\|_\gamma^2\} \\ = 2D\{\sigma^2+\eta^2\,\Gamma\,\|\varphi-\varphi_{k_n^*}\|_\gamma^2\}\sum_{j=1}^{k_n^*}\frac{\omega_j}{n\lambda_j}. \quad \text{(A.5)}$$

Consider now A_{12}. Observe that $\|[\mathrm{diag}(\omega)]_{\underline{k_n^*}}^{1/2}[T]_{\underline{k_n^*}}^{-1}\|^2 \leqslant D\max_{1\leqslant j\leqslant k_n^*}\omega_j/\lambda_j$ for all $T\in\mathcal{T}_{d,D}^\lambda$. By employing the last inequality together with $\|[\widehat{T}]_{\underline{k_n^*}}^{-1}\|^2\mathbb{1}_{\Omega_{1/2}} \leqslant 4D/\lambda_{k_n^*}$ and $\|[\widehat{T}]_{\underline{k_n^*}}^{-1}\|^2\mathbb{1}_\Omega \leqslant n$ there exists a numerical constant $C>0$ such that

$$\mathbb{E}[\|[\mathrm{diag}(\omega)]_{\underline{k_n^*}}^{1/2}[T]_{\underline{k_n^*}}^{-1}[\Xi]_{\underline{k_n^*}}[\widehat{T}]_{\underline{k_n^*}}^{-1}\{[B]_{\underline{k_n^*}}+[S]_{\underline{k_n^*}}\}\|^2\mathbb{1}_\Omega] \\ \leqslant D\max_{1\leqslant j\leqslant k_n^*}\frac{\omega_j}{\lambda_j}\Big\{4D\lambda_{k_n^*}^{-1}\mathbb{E}\|[\Xi]_{\underline{k_n^*}}\|^2\|[B]_{\underline{k_n^*}}+[S]_{\underline{k_n^*}}\|^2\mathbb{1}_{\Omega_{1/2}}+n\mathbb{E}\|[\Xi]_{\underline{k_n^*}}\|^2\|[B]_{\underline{k_n^*}}+[S]_{\underline{k_n^*}}\|^2\mathbb{1}_{\Omega_{1/2}^c}\Big\} \\ \leqslant D\max_{1\leqslant j\leqslant k_n^*}\frac{\omega_j}{\lambda_j}\Big\{4D\lambda_{k_n^*}^{-1}\big(\mathbb{E}\|[\Xi]_{\underline{k_n^*}}\|^4\big)^{1/2}+n\big(\mathbb{E}\|[\Xi]_{\underline{k_n^*}}\|^8\big)^{1/4}P(\Omega_{1/2}^c)^{1/4}\Big\}\big(\mathbb{E}\|[B]_{\underline{k_n^*}}+[S]_{\underline{k_n^*}}\|^4\big)^{1/2} \\ \leqslant C\max_{1\leqslant j\leqslant k_n^*}\frac{\omega_j}{n\lambda_j}\,D\,\eta^4\,(\sigma^2+\Gamma\,\|\varphi-\varphi_{k_n^*}\|_\gamma^2)\Big\{4D\frac{(k_n^*)^3}{\lambda_{k_n^*}n}+(k_n^*)^3|P(\Omega_{1/2}^c)|^{1/4}\Big\}$$

where the last bound follows from (A.8), (A.9) and (A.10) in Lemma A.10. By combination of the last bound and (A.5) via the decomposition (A.4) there exists a numerical constant $C>0$ such that

$$\mathbb{E}\|\widehat{\varphi}_{k_n^*}-\widetilde{\varphi}_{k_n^*}\|_\omega^2 \leqslant C\,D\,\eta^4\,(\sigma^2+\Gamma\,\|\varphi-\varphi_{k_n^*}\|_\gamma^2)\Big\{4D\zeta/\kappa+(k_n^*)^3|P(\Omega_{1/2}^c)|^{1/4}\Big\}\sum_{j=1}^{k_n^*}\frac{\omega_j}{n\lambda_j}.$$

Furthermore, taking into account the estimate (A.12) from Lemma A.2 with $w=\gamma$ and the definition of R_n^*, the last inequality implies

$$\mathbb{E}\|\widehat{\varphi}_{k_n^*}-\widetilde{\varphi}_{k_n^*}\|_\omega^2 \leqslant C\,D\,\eta^4\,(\sigma^2+4\Gamma\,Dd\rho)\Big\{4D\zeta/\kappa+(k_n^*)^3|P(\Omega_{1/2}^c)|^{1/4}\Big\}R_n^*.$$

Finally, since $\Omega_{1/2}^c \subset \{\|[\widehat{T}]_{\underline{k_n^*}}-[T]_{\underline{k_n^*}}\|^2 > \lambda_{k_n^*}/(4D)\}$, by using the decomposition (A.2) the result of the theorem follows from the last estimate and (A.3). □

Proof of Theorem 2.6. We start our proof with the observation that under Assumption A3

$$P\Big(\|\widehat{[T]}_{\underline{k_n^*}} - [T]_{\underline{k_n^*}}\|^2 > \frac{\lambda_{k_n^*}}{4D}\Big) \leqslant 2\exp\{-\frac{n\lambda_{k_n^*}}{20D\eta^2(k_n^*)^2} + 2\log k_n^*\}$$
$$\leqslant 2\exp\{-\frac{\kappa}{20D\eta^2\zeta}k_n^* + 2\log k_n^*\}$$

by applying successively (A.14) in Lemma A.3 and the estimate $(k_n^*)^3(n\lambda_{k_n^*})^{-1} \leqslant \kappa^{-1}(k_n^*)^3\gamma_{k_n^*}^{-1} \leqslant \kappa^{-1}\zeta$. From this estimate we conclude for all $n \in \mathbb{N}$ with $(\log k_n^*)/k_n^* \leqslant \kappa/(280D\eta^2\zeta)$ and $(\log R_n^*)/k_n^* \geqslant -\kappa/(40D\eta^2\zeta)$ that

$$(k_n^*)^{12}P\Big(\|\widehat{[T]}_{\underline{k_n^*}} - [T]_{\underline{k_n^*}}\|^2 > \frac{\lambda_{k_n^*}}{4D}\Big) \leqslant 2,$$
$$(R_n^*)^{-1}P\Big(\|\widehat{[T]}_{\underline{k_n^*}} - [T]_{\underline{k_n^*}}\|^2 > \frac{\lambda_{k_n^*}}{4D}\Big) \leqslant 2.$$

By employing these estimates the assertion follows now from Theorem 2.4. □

Illustration: estimation of derivatives.
Proof of Proposition 2.8. Since for each $0 \leqslant s \leqslant p$ we have $\mathbb{E}\|\widetilde{f}^{(s)} - f^{(s)}\|^2 \sim \mathbb{E}\|\widetilde{f} - f\|_\omega^2$ we intend to apply the general result given Theorem 2.1. In both cases the additional conditions formulated in Theorem 2.1 are easily verified. Therefore, it is sufficient to evaluate the lower bound R_n^* given in (2.4). Note that the optimal dimension parameter k_n^* satisfies $R_n^* \sim \omega_{k_n^*}/\gamma_{k_n^*} \sim \sum_{l=1}^{k_n^*}\omega_l/(n\lambda_l)$ since both sequences (γ_j/ω_j) and $(\sum_{0<|l|\leqslant j}\frac{\omega_l}{n\lambda_l})$ are non-increasing.
[p-λ] The well-known approximation $\sum_{j=1}^k j^r \sim k^{r+1}$ for $r > 0$ implies
$n \sim (\gamma_{k_n^*}/\omega_{k_n^*})\sum_{l=1}^{k_n^*}\omega_l/\lambda_l \sim (k_n^*)^{2a+2p+1}$. It follows that $k_n^* \sim n^{1/(2p+2a+1)}$ and the lower bound writes $R_n^* \sim n^{-(2p-2s)/(2p+2a+1)}$.
[e-λ] Applying Laplace's Method (c.f. chapter 3.7 in Olver (1974)) we have
$n \sim (\gamma_{k_n^*}/\omega_{k_n^*})\sum_{l=1}^{k_n^*}\omega_l/\lambda_l \sim (k_n^*)^{2p}\exp(|k_n^*|^{2a})$ which implies that
$k_n^* \sim \{\log(n/(\log n)^{p/a})\}^{1/(2a)} = (\log n)^{1/(2a)}(1+o(1))$ and that the lower bound can be rewritten as $R_n^* \sim (\log n)^{-(p-s)/a}$. □

Proof of Proposition 2.9. Since in both cases the dimension parameter is chosen optimal (see the proof of Proposition 2.8) the result follows from Theorem 2.4. □

Technical assertions. The following paragraph gathers technical results used in the proofs of this section.

Lemma A.1 *Suppose that $U \in \mathcal{U}_\sigma$, $\sigma > 0$ and that the joint distribution of (Z, W) satisfies Assumption A2. If in addition $\varphi \in \mathcal{F}_\gamma^r$ with $\Gamma = \sum_{j=1}^\infty \gamma_j^{-1} < \infty$, then there exists a constant*

$C > 0$ *such that for all* $k \in \mathbb{N}$ *and for all* $z \in \mathbb{R}^k$

$$\mathbb{E}|z^t\,[B]_{\underline{k}}|^2 \leqslant (1/n)\,\|z\|^2\,\sigma^2, \tag{A.6}$$
$$\mathbb{E}|z^t\,[S]_{\underline{k}}|^2 \leqslant (1/n)\,\|z\|^2\,\eta^2\,\Gamma\,\|\varphi-\varphi_k\|_\gamma^2 \tag{A.7}$$
$$\mathbb{E}\|[B]_{\underline{k}}\|^4 \leqslant C\cdot\Big((k/n)\cdot\sigma^2\cdot\eta^2\Big)^2, \tag{A.8}$$
$$\mathbb{E}\|[S]_{\underline{k}}\|^4 \leqslant C\cdot\Big((k/n)\cdot\eta^2\cdot\Gamma\cdot\|\varphi-\varphi_k\|_\gamma^2\Big)^2, \tag{A.9}$$
$$\mathbb{E}\|[\Xi]_{\underline{k}}\|^8 \leqslant C\cdot\Big((k^2/n)\cdot\eta^2\Big)^4. \tag{A.10}$$

Moreover, given a $(k\times k)$ *matrix* M*, we have*

$$\mathbb{E}\|M\{[B]_{\underline{k}}+[S]_{\underline{k}}\}\|^2 \leqslant (2/n)\,\operatorname{tr}(M^tM)\{\sigma^2+\eta^2\,\Gamma\,\|\varphi-\varphi_k\|_\gamma^2\}. \tag{A.11}$$

Proof. The proof of (A.6) - (A.10) can be found in Johannes and Breunig (2009) and we omit the details. The estimate (A.11) follows by employing (A.6) and (A.7) from the identity $\|M\{[B]_{\underline{k}}+[S]_{\underline{k}}\}\|^2=\sum_{j=1}^k\|M_j^t\{[B]_{\underline{k}}+[S]_{\underline{k}}\}\|^2$, where M_j denotes the j-th column of M^t, which completes the proof. $\square$

Lemma A.2 *Let* $g=T\varphi$ *and for each* $k\in\mathbb{N}$ *denote* $\varphi_k := [T]_{\underline{k}}^{-1}[g]_{\underline{k}}$*. Given sequences* λ *and* γ *satisfying Assumption A1 let* $T\in\mathcal{T}_{d,D}^\lambda$ *and* $\varphi\in\mathcal{F}_\gamma^r$*. For each strictly positive sequence* $w:=(w_j)_{j\in\mathbb{N}}$ *such that* w/γ *is non increasing we obtain for all* $k\in\mathbb{N}$

$$\|\varphi-\varphi_k\|_w^2 \leqslant 4\,D\,d\,\rho\,\frac{w_k}{\gamma_k}\max\left(1,\frac{\lambda_k}{w_k}\max_{1\leqslant j\leqslant k}\frac{w_j}{\lambda_j}\right) \tag{A.12}$$

Proof. The condition $T\in\mathcal{T}_{d,D}^\lambda$, that is, $\sup_{k\in\mathbb{N}}\|[\operatorname{diag}(\lambda)]_{\underline{k}}^{1/2}[T]_{\underline{k}}^{-1}\|^2\leqslant D$ and $\|Tf\|^2\leqslant d\|f\|_\lambda^2$ for all $f\in L_Z^2$, together with the identity $[E_k\varphi-\varphi_k]_{\underline{k}}=-[T]_{\underline{k}}^{-1}[TE_k^\perp\varphi]_{\underline{k}}$ implies $\|E_k\varphi-\varphi_k\|_\lambda^2\leqslant D\|TE_k^\perp\varphi\|^2\leqslant Dd\|E_k^\perp\varphi\|_\lambda^2\leqslant Dd\gamma_k^{-1}\lambda_k\rho$ for all $\varphi\in\mathcal{F}_\gamma^\rho$ because (λ/γ) is monotonically non increasing. From this estimate we conclude

$$\begin{aligned}\|E_k\varphi-\varphi_k\|_w^2 &= \|[\operatorname{diag}(w)]_{\underline{k}}^{1/2}[E_k\varphi-\varphi_k]_{\underline{k}}\|^2\\ &\leqslant \|[\operatorname{diag}(w)]_{\underline{k}}^{1/2}[\operatorname{diag}(\lambda)]_{\underline{k}}^{-1/2}\|^2\|E_k\varphi-\varphi_k\|_\lambda^2 \leqslant Dd\rho\frac{\lambda_k}{\gamma_k}\max_{1\leqslant j\leqslant k}\frac{w_j}{\lambda_j}.\end{aligned} \tag{A.13}$$

Furthermore, since (w/γ) is non increasing, we have $\|E_k\varphi-\varphi\|_w^2\leqslant\rho w_k/\gamma_k$ for all $f\in\mathcal{F}_\gamma^\rho$. The assertion follows now by combination of the last estimate and (A.13) via a decomposition based on an elementary triangular inequality. $\square$

Lemma A.3 *Suppose that the joint distribution of* (Z,W) *satisfies Assumption A3. If in addition the sequence* λ *fulfills Assumption A1, then for all* $k\in\mathbb{N}$ *we have*

$$P(\|[\Xi]_{\underline{k}}\|^2>\frac{\lambda_k}{4D})\leqslant 2\exp\{-\frac{n\lambda_k}{k^2(20D\eta^2)}+2\log k\}. \tag{A.14}$$

Proof. The proof of the assertion can be found in Johannes and Breunig (2009) and we omit the details. $\square$

A.2. Proofs of section 3

We begin by defining and recalling notations to be used in the proof. Given $u \in L^2[0,1]$ we denote by $[u]$ the infinite vector of Fourier coefficients $[u]_j := \langle u, \psi_j \rangle$. In particular we use the notations

$$\widehat{\varphi}_k = \sum_{j=1}^{k} \frac{\widehat{[g]}_j}{\widehat{[T]}_{jj}} \mathbb{1}\{\inf_{1\leqslant j\leqslant k} \widehat{[T]}_{jj}^2 \geqslant 1/n\}\psi_j, \quad \widetilde{\varphi}_k := \sum_{j=1}^{k} \frac{\widehat{[g]}_j}{[T]_{jj}} e_j, \quad \varphi_k := \sum_{j=1}^{k} \frac{[g]_j}{[T]_{jj}} \psi_j,$$

$$\widehat{\Phi}_u := \sum_{j\in\mathbb{N}} \frac{[u]_j}{\widehat{[T]}_{jj}} \mathbb{1}\{\inf_{1\leqslant j\leqslant k} \widehat{[T]}_{jj}^2 \geqslant 1/n\}\psi_j, \quad \Phi_u := \sum_{j\in\mathbb{N}} \frac{[u]_j}{[T]_{jj}} \psi_j.$$

Furthermore, let $\widehat{g}$ be the function with Fourier coefficients $[\widehat{g}]_j := \widehat{[g]}_j$ and observe that $\mathbb{E}\widehat{g} = g$. Given $1 \leqslant k \leqslant k'$ we have then for all $t \in \mathcal{S}_k := \operatorname{span}\{\psi_1, \dots, \psi_k\}$

$$\begin{aligned}
\langle t, \widehat{\varphi}_{k'} \rangle_\omega &= \langle t, \widehat{\Phi}_{\widehat{g}} \rangle_\omega = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{k} Y_i \psi_j(W_i) \frac{\omega_j [t]_j}{\widehat{[T]}_{jj}} \mathbb{1}\{\inf_{1\leqslant j\leqslant k} \widehat{[T]}_{jj}^2 \geqslant 1/n\} = \langle t, \widehat{\varphi}_k \rangle_\omega, \\
\langle t, \widetilde{\varphi}_{k'} \rangle_\omega &= \langle t, \Phi_{\widehat{g}} \rangle_\omega = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{k} Y_i \psi_j(W_i) \frac{\omega_j [t]_j}{[T]_{jj}} = \langle t, \widetilde{\varphi}_k \rangle_\omega, \\
\langle t, \varphi_{k'} \rangle_\omega &= \langle t, \Phi_g \rangle_\omega = \sum_{j=1}^{k} \frac{\omega_j [t]_j [g]_j}{[T]_{jj}} = \sum_{j=1}^{k} \omega_j [t]_j [\varphi]_j = \langle t, \varphi \rangle_\omega.
\end{aligned} \tag{A.15}$$

Consider the contrast $\Upsilon(t) := \|t\|_\omega^2 - 2\langle t, \widehat{\Phi}_{\widehat{g}} \rangle_\omega$, for all $t \in L^2[0,1]$. Obviously it follows for all $t \in \mathcal{S}_k$ that $\Upsilon(t) = \|t - \widehat{\varphi}_k\|_\omega^2 - \|\widehat{\varphi}_k\|_\omega^2$ and, hence

$$\arg\min_{t\in\mathcal{S}_k} \Upsilon(t) = \widehat{\varphi}_k, \quad \forall\, k \geqslant 1. \tag{A.16}$$

Then, the adaptive choice $\widehat{k}$ of the dimension parameter can be rewritten as

$$\widehat{k} = \operatorname*{argmin}_{1\leqslant k\leqslant N_n} \{\Upsilon(\widehat{\varphi}_k) + \widehat{\operatorname{pen}}(k)\} \qquad \text{with} \qquad \widehat{\operatorname{pen}}(k) := 540\, \mathbb{E}[Y^2] \frac{\widehat{\delta}_k}{n}. \tag{A.17}$$

Then for all $1 \leqslant k \leqslant N_n$, we have that $\Upsilon(\widehat{\varphi}_{\widehat{k}}) + \operatorname{pen}(\widehat{k}) \leqslant \Upsilon(\widehat{\varphi}_k) + \operatorname{pen}(k) \leqslant \Upsilon(\varphi_k) + \operatorname{pen}(k)$, using first (A.17) and then (A.16). This inequality implies

$$\|\widehat{\varphi}_{\widehat{k}}\|_\omega^2 - \|\varphi_k\|_\omega^2 \leqslant 2\langle \widehat{\varphi}_{\widehat{k}} - \varphi_k, \widehat{\varphi}_{\widehat{k}} \rangle_\omega + \widehat{\operatorname{pen}}(k) - \widehat{\operatorname{pen}}(\widehat{k}),$$

which together with the identities given in (A.15) for all $1 \leqslant k \leqslant N_n$ implies

$$\begin{aligned}
\|\widehat{\varphi}_{\widehat{k}} - \varphi\|_\omega^2 &= \|\varphi - \varphi_k\|_\omega^2 + \|\widehat{\varphi}_{\widehat{k}}\|_\omega^2 - \|\varphi_k\|_\omega^2 - 2\langle \widehat{\varphi}_{\widehat{k}} - \varphi_k, \varphi \rangle_\omega \\
&\leqslant \|\varphi - \varphi_k\|_\omega^2 + \widehat{\operatorname{pen}}(k) - \widehat{\operatorname{pen}}(\widehat{k}) + 2\langle \widehat{\varphi}_{\widehat{k}} - \varphi_k, \widehat{\Phi}_{\widehat{g}} - \Phi_g \rangle_\omega
\end{aligned} \tag{A.18}$$

Consider the unit ball $\mathcal{B}_k := \{f \in \mathcal{S}_k : \|f\|_\omega \leqslant 1\}$ and, for arbitrary $\tau > 0$ and $t \in \mathcal{S}_k$, the elementary inequality

$$2|\langle t, h \rangle_\omega| \leqslant 2\|t\|_\omega \sup_{t\in\mathcal{B}_k} |\langle t, h \rangle_\omega| \leqslant \tau\|t\|_\omega^2 + \frac{1}{\tau} \sup_{t\in\mathcal{B}_k} |\langle t, h \rangle_\omega|^2 = \tau\|t\|_\omega^2 + \frac{1}{\tau} \sum_{j=1}^{k} \omega_j |[h]_j|^2.$$

Combining the last estimate with (A.18) and $\widehat{\varphi}_{\widehat{k}} - \varphi_k \in \mathcal{S}_{\widehat{k}\vee k}$ we obtain

$$\|\widehat{\varphi}_{\widehat{k}} - \varphi\|_\omega^2 \leqslant \|\varphi - \varphi_k\|_\omega^2 + \tau \, \|\widehat{\varphi}_{\widehat{k}} - \varphi_k\|_\omega^2 + \widehat{\text{pen}}(k) - \widehat{\text{pen}}(\widehat{k}) + \frac{1}{\tau} \sup_{t \in \mathcal{B}_{k\vee\widehat{k}}} |\langle t, \widehat{\Phi}_{\widehat{g}} - \Phi_g\rangle_\omega|^2.$$

Letting $\tau := 1/3$ it follows from $\|\widehat{\varphi}_{\widehat{k}} - \varphi_k\|_\omega^2 \leqslant 2\|\widehat{\varphi}_{\widehat{k}} - \varphi\|_\omega^2 + 2\|\varphi_k - \varphi\|_\omega^2$ that

$$\frac{1}{3}\|\widehat{\varphi}_{\widehat{k}} - \varphi\|_\omega^2 \leqslant \frac{5}{3}\|\varphi - \varphi_k\|_\omega^2 + \widehat{\text{pen}}(k) - \widehat{\text{pen}}(\widehat{k}) + 3 \sup_{t \in \mathcal{B}_{k\vee\widehat{k}}} |\langle t, \widehat{\Phi}_{\widehat{g}} - \Phi_g\rangle_\omega|^2.$$

Consider the functions $\widehat{\nu}$ and $\widehat{\mu}$ with Fourier coefficients $[\widehat{\nu}]_j = \frac{1}{n}\sum_{i=1}^n Y_i \mathbb{1}\{|Y_i| \leqslant \alpha_n\}\psi_j(W_i)$ and $[\widehat{\mu}]_j = \frac{1}{n}\sum_{i=1}^n Y_i \mathbb{1}\{|Y_i| > \alpha_n\}\psi_j(W_i)$ respectively, and their centered versions $\nu = \widehat{\nu} - \mathbb{E}\widehat{\nu}$ and $\mu = \widehat{\mu} - \mathbb{E}\widehat{\mu}$, then we have $\widehat{g} - g = \nu + \mu$ and

$$\begin{aligned}\frac{1}{3}\|\widehat{\varphi}_{\widehat{k}} - \varphi\|_\omega^2 &\leqslant \frac{5}{3}\|\varphi - \varphi_k\|_\omega^2 + \widehat{\text{pen}}(k) - \widehat{\text{pen}}(\widehat{k}) \\ &\quad + 6 \sup_{t \in \mathcal{B}_{k\vee\widehat{k}}} |\langle t, \Phi_\nu\rangle_\omega|^2 + 12 \sup_{t \in \mathcal{B}_{k\vee\widehat{k}}} |\langle t, \widehat{\Phi}_\nu - \Phi_\nu\rangle_\omega|^2 + 12 \sup_{t \in \mathcal{B}_{k\vee\widehat{k}}} |\langle t, \widehat{\Phi}_\mu + \widehat{\Phi}_g - \Phi_g\rangle_\omega|^2\end{aligned}$$

Decompose $|\langle t, \widehat{\Phi}_\nu - \Phi_\nu\rangle_\omega|^2 = |\langle t, \widehat{\Phi}_\nu - \Phi_\nu\rangle_\omega|^2 \mathbb{1}\{\Omega_q\} + |\langle t, \widehat{\Phi}_\nu - \Phi_\nu\rangle_\omega|^2 \mathbb{1}\{\Omega_q^c\}$ further using

$$\Omega_q := \Big\{\forall\, 1 \leqslant j \leqslant N_n \ : \ \Big|\widehat{[T]}_{jj}^{-1} - [T]_{jj}^{-1}\Big| \leqslant \frac{1}{2|[T]_{jj}|} \ \wedge \ \widehat{[T]}_{jj}^2 \geqslant 1/n\Big\}. \tag{A.19}$$

Since $\mathbb{1}\{\widehat{[T]}_{jj}^2 \geqslant 1/m\}\mathbb{1}\{\Omega_q\} = \mathbb{1}\{\Omega_q\}$, it follows that for all $1 \leqslant j \leqslant N_n$ we have

$$\left(\frac{[T]_{jj}}{\widehat{[T]}_{jj}} \mathbb{1}\{\widehat{[T]}_{jj}^2 \geqslant 1/n\} - 1\right)^2 \mathbb{1}\{\Omega_q\} = |[T]_{jj}|^2 \ \mathbb{1}\{\Omega_q\} \left|\widehat{[T]}_{jj}^{-1} - [T]_{jj}^{-1}\right|^2 \leqslant \frac{1}{4}.$$

Hence, $\sup_{t \in \mathcal{B}_k} |\langle t, \widehat{\Phi}_\nu - \Phi_\nu\rangle_\omega|^2 \ \mathbb{1}\{\Omega_q\} \leqslant \frac{1}{4}\sup_{t\in\mathcal{B}_k} |\langle t, \Phi_\nu\rangle_\omega|^2$ for all $1 \leqslant k \leqslant N_n$ and

$$\begin{aligned}\frac{1}{3}\|\widehat{\varphi}_{\widehat{k}} - \varphi\|_\omega^2 &\leqslant \frac{5}{3}\|\varphi - \varphi_k\|_\omega^2 + 9 \sup_{t \in \mathcal{B}_{k\vee\widehat{k}}} |\langle t, \Phi_\nu\rangle_\omega|^2 + \widehat{\text{pen}}(k) - \widehat{\text{pen}}(\widehat{k}) \\ &\quad + 12 \sup_{t \in \mathcal{B}_{k\vee\widehat{k}}} |\langle t, \widehat{\Phi}_\nu - \Phi_\nu\rangle_\omega|^2 \ \mathbb{1}\{\Omega_q^c\} + 12 \sup_{t \in \mathcal{B}_{k\vee\widehat{k}}} |\langle t, \widehat{\Phi}_\mu + \widehat{\Phi}_g - \Phi_g\rangle_\omega|^2.\end{aligned} \tag{A.20}$$

Define $\Delta_k^T := \max_{1\leqslant j\leqslant k} \omega_j/|[T]_{jj}|^2$, $\tau_k^T := \max_{1\leqslant j\leqslant k}(\omega_j)_{\vee 1}/|[T]_{jj}|^2$,
and $\delta_k^T := k\Delta_k^T \left\{\log(\tau_k^T \vee (k+2)) \ / \ \log(k+2)\right\}$. Then, it is easily seen that

$$\delta_k^T \leqslant \delta_k \, d \, \frac{\log(3d)}{\log 3} = \delta_k \, d \, \zeta_d \qquad \forall\, k \geqslant 1. \tag{A.21}$$

with $\zeta_d = (\log 3d)/(\log 3)$. Moreover, define the event $\Omega_{qp} := \Omega_q \cap \Omega_p$ where Ω_q is given in (A.18) and

$$\Omega_p := \Big\{N_n^l \leqslant \widehat{N}_n \leqslant N_n\Big\}.$$

Observe that on Ω_q we have $(1/2)\Delta_k^T \leqslant \widehat{\Delta}_k \leqslant (3/2)\Delta_k^T$ for all $1 \leqslant k \leqslant N_n$ and hence $(1/2)[\Delta_k^T \vee (k+2)] \leqslant [\widehat{\Delta}_k \vee (k+2)] \leqslant (3/2)[\Delta_k^T \vee (k+2)]$, which implies

$$(1/2)k\Delta_k^T\Big(\frac{\log[\Delta_k^T \vee (k+2)]}{\log(k+2)}\Big)\Big(1-\frac{\log 2}{\log(k+2)}\frac{\log(k+2)}{\log(\Delta_k^T \vee [k+2])}\Big)$$
$$\leqslant \widehat{\delta}_k \leqslant (3/2)k\Delta_k^T\Big(\frac{\log(\Delta_k^T \vee [k+2])}{\log(k+2)}\Big)\Big(1+\frac{\log 3/2}{\log(k+2)}\frac{\log(k+2)}{\log(\Delta_k^T \vee [k+2])}\Big).$$

Using $\log(\Delta_k^T \vee (k+2))/\log(k+2) \geqslant 1$, we conclude from the last estimate that

$$\delta_k^T/10 \leqslant (\log 3/2)/(2\log 3)\delta_k^T \leqslant (1/2)\delta_k^T[1-(\log 2)/\log(k+2)] \leqslant \widehat{\delta}_k$$
$$\leqslant (3/2)\delta_k^T[1+(\log 3/2)/\log(k+2)] \leqslant 3\delta_k^T.$$

Recall that $\widehat{\mathrm{pen}}(k) = 540\,\mathbb{E}[Y^2]\,\widehat{\delta}_k n^{-1}$, we define

$$\mathrm{pen}(k) := 54\,\mathbb{E}[Y^2]\,\delta_k^T n^{-1},$$

then it follows that on Ω_q we have

$$\mathrm{pen}(k) \leqslant \widehat{\mathrm{pen}}(k) \leqslant 30\,\mathrm{pen}(k) \qquad \forall\, 1 \leqslant k \leqslant N_n.$$

On $\Omega_{qp} = \Omega_q \cap \Omega_p$, we have $\widehat{k} \leqslant N_n$. Thus,

$$\Big(\mathrm{pen}(k \vee \widehat{k}) + \widehat{\mathrm{pen}}(k) - \widehat{\mathrm{pen}}(\widehat{k})\Big)\mathbb{1}\{\Omega_{qp}\} \leqslant \Big(\mathrm{pen}(k) + \mathrm{pen}(\widehat{k}) + \widehat{\mathrm{pen}}(k) - \widehat{\mathrm{pen}}(\widehat{k})\Big)\mathbb{1}\{\Omega_{qp}\}$$
$$\leqslant 31\,\mathrm{pen}(k) \qquad \forall 1 \leqslant k \leqslant N_n. \tag{A.22}$$

Furthermore, we obviously have $\widehat{\Delta}_k \leqslant n\Delta_k^T$ for every $1 \leqslant k \leqslant N_n$, which implies $\widehat{\delta}_k \leqslant n\,(1+\log n)\,\delta_k^T$. Consequently, $\widehat{\mathrm{pen}}(k) \leqslant 540\,\mathbb{E}[Y^2]\,n\,(1+\log n)$, because $\delta_k^T/n \leqslant d\zeta_d\delta_k/n \leqslant d\zeta_d$ for all $1 \leqslant k \leqslant N_n$ by (A.21) and the definition of N_n. On $\Omega_q^c \cap \Omega_p$, we have $\widehat{k} \leqslant N_n$ and hence $\mathrm{pen}(k \vee \widehat{k}) \leqslant \mathrm{pen}(N_n) \leqslant 54\,\mathbb{E}[Y^2]$, which implies

$$(\mathrm{pen}(k \vee \widehat{k}) + \widehat{\mathrm{pen}}(k) - \widehat{\mathrm{pen}}(\widehat{k}))\mathbb{1}\{\Omega_q^c \cap \Omega_p\} \leqslant 594\,\mathbb{E}[Y^2]\,n\,(1+\log n)\mathbb{1}\{\Omega_q^c \cap \Omega_p\}. \tag{A.23}$$

We note further that for all $\varphi \in \mathcal{F}_\gamma^\rho$ with $\sum_{j\in\mathbb{N}}\gamma_j^{-1} = \Gamma < \infty$ and for all $z \in [0,1]$ we have $|\varphi(z)|^2 \leqslant \rho\sum_{j\in\mathbb{N}}\gamma_j^{-1}\psi_j^2(z) \leqslant 2\rho\Gamma$ by employing Cauchy-Schwarz inequality. Thereby, given $m \geqslant 1$ such that $\mathbb{E}U^{2m}|W \leqslant \sigma^{2m}$, it follows

$$\mathbb{E}[Y^{2m}|W] \leqslant 2^{2m}(2\rho\Gamma+\sigma^2)^m \text{ and, hence } \mathbb{E}[Y^{2m}] \leqslant 2^{2m}(2\rho\Gamma+\sigma^2)^m. \tag{A.24}$$

At the end of this section we will prove three technical Lemmata (A.5, A.8 and A.7) which are used in the following proof.
Proof of Theorem 3.4. The proof is based on the decomposition

$$\mathbb{E}\|\widehat{\varphi}_{\widehat{k}} - \varphi\|_\omega^2 = \mathbb{E}\|\widehat{\varphi}_{\widehat{k}} - \varphi\|_\omega^2\mathbb{1}\{\Omega_{qp}\} + \mathbb{E}\|\widehat{\varphi}_{\widehat{k}} - \varphi\|_\omega^2\mathbb{1}\{\Omega_q^c \cap \Omega_p\} + \mathbb{E}\|\widehat{\varphi}_{\widehat{k}} - \varphi\|_\omega^2\mathbb{1}\{\Omega_p^c\}.$$

Below we show that there exist a numerical constant C such that for all $n \geqslant 1$ and all $1 \leqslant k \leqslant N_n^l$ we have

$$
\begin{aligned}
\mathbb{E}\|\widehat{\varphi}_{\widehat{k}} - \varphi\|_\omega^2 \mathbb{1}\{\Omega_{qp}\} \leqslant C\Big\{ &\|\varphi - \varphi_k\|_\omega^2 + \operatorname{pen}(k) + d\rho \max_{j\geqslant 1}\Big\{\frac{\omega_j}{\gamma_j}\min\big(1, \frac{1}{n\lambda_j}\big)\Big\} \\
&+ \frac{(2\rho\Gamma+\sigma^2)^4}{n} + \frac{(2\rho\Gamma+\sigma^2+1)d\,\zeta_d}{n}\,\Sigma\left(\frac{(2\rho\Gamma+\sigma^2)\zeta_d + V_{U|Z}}{V_{U|Z}^2}\right)\Big\},
\end{aligned} \tag{A.25}
$$

$$
\begin{aligned}
\mathbb{E}\|\widehat{\varphi}_{\widehat{k}} - \varphi\|_\omega^2 \mathbb{1}\{\Omega_q^c \cap \Omega_p\} \leqslant C\Big\{ &\|\varphi - \varphi_k\|_\omega^2 + d\rho \max_{j\geqslant 1}\Big\{\frac{\omega_j}{\gamma_j}\min\big(1, \frac{1}{n\lambda_j}\big)\Big\} \\
&+ \frac{(2\rho\Gamma+\sigma^2+1)^4 d}{n}\ \left[\zeta_d\Sigma\left(\frac{(2\rho\Gamma+\sigma^2)\zeta_d + V_{U|Z}}{V_{U|Z}^2}\right)+1\right]\Big\},
\end{aligned} \tag{A.26}
$$

$$
\mathbb{E}\|\widehat{\varphi}_{\widehat{k}} - \varphi\|_\omega^2 \mathbb{1}\{\Omega_p^c\} \leqslant \frac{C}{n}\,(2\rho\Gamma+\sigma^2). \tag{A.27}
$$

The desired upper bound follows by using (A.21), that is, $\operatorname{pen}(k) \leqslant 54\,(2\rho\Gamma+\sigma^2)\,d\,\zeta_d\,\delta_k n^{-1}$, and by employing the monotonicity of ω/γ, that is $\|\varphi-\varphi_k\|_\omega^2 \leqslant \rho\omega_k/\gamma_k$.

Proof of (A.25). By employing the estimate (A.28) and $\operatorname{pen}(k) := 54\,\mathbb{E}[Y^2]\,\delta_k^T n^{-1}$ we have

$$
\begin{aligned}
\frac{1}{3}\|\widehat{\varphi}_{\widehat{k}} - \varphi\|_\omega^2 \leqslant \frac{5}{3}\|\varphi-\varphi_k\|_\omega^2 + 9\Big(\sup_{t\in\mathcal{B}_{k\vee\widehat{k}}} |\langle t, \Phi_\nu\rangle_\omega|^2 - 6\,\frac{\mathbb{E}[Y^2]\,\delta_{k\vee\widehat{k}}^T}{n}\Big)_+ \\
+ \operatorname{pen}(k\vee\widehat{k}) + \widehat{\operatorname{pen}}(k) - \widehat{\operatorname{pen}}(\widehat{k}) \\
+ 12\sup_{t\in\mathcal{B}_{k\vee\widehat{k}}} |\langle t, \widehat{\Phi}_\nu - \Phi_\nu\rangle_\omega|^2\,\mathbb{1}\{\Omega_q^c\} + 12\sup_{t\in\mathcal{B}_{k\vee\widehat{k}}} |\langle t, \widehat{\Phi}_\mu + \widehat{\Phi}_g - \Phi_g\rangle_\omega|^2
\end{aligned}
$$

and, hence using that $\widehat{k} \leqslant N_n$ on Ω_p we obtain for all $1 \leqslant k \leqslant N_n^l$

$$
\begin{aligned}
\frac{1}{3}\|\widehat{\varphi}_{\widehat{k}} - \varphi\|_\omega^2 \mathbb{1}\{\Omega_{qp}\} &\leqslant \frac{5}{3}\|\varphi-\varphi_k\|_\omega^2 + 9\sum_{k=1}^{N_n}\Big(\sup_{t\in\mathcal{B}_k} |\langle t, \Phi_\nu\rangle_\omega|^2 - 6\frac{\mathbb{E}[Y^2]\delta_k^T}{n}\Big)_+ \\
&\quad + 12\sup_{t\in\mathcal{B}_{N_n}} |\langle t, \widehat{\Phi}_\mu + \widehat{\Phi}_g - \Phi_g\rangle_\omega|^2 + \Big(\operatorname{pen}(k\vee\widehat{k}) + \widehat{\operatorname{pen}}(k) - \widehat{\operatorname{pen}}(\widehat{k})\Big)\mathbb{1}\{\Omega_{qp}\} \\
&\leqslant \frac{5}{3}\|\varphi-\varphi_k\|_\omega^2 + 9\sum_{k=1}^{N_n}\Big(\sup_{t\in\mathcal{B}_k} |\langle t, \Phi_\nu\rangle_\omega|^2 - 6\frac{\mathbb{E}[Y^2]\delta_k^T}{n}\Big)_+ \\
&\quad + 12\sup_{t\in\mathcal{B}_{N_n}} |\langle t, \widehat{\Phi}_\mu + \widehat{\Phi}_g - \Phi_g\rangle_\omega|^2 + 31\operatorname{pen}(k),
\end{aligned}
$$

where the last inequality follows from (A.22). The second term is bounded by employing Lemma A.5. In order to control the third term, apply Lemmata A.6 and A.7. Consequently, combining these estimates proves inequality (A.25).

Proof of (A.26). On $\Omega_q^c \cap \Omega_p$, we have $N_n^l \leqslant \widehat{N}_n \leqslant N_n$. Applying (A.28), it follows from

(A.23) that for all $1 \leqslant k \leqslant N_n^l$

$$\frac{1}{3}\|\widehat{\varphi}_{\widehat{k}} - \varphi\|_\omega^2 \mathbb{1}\{\Omega_q^c \cap \Omega_p\} \leqslant \frac{5}{3}\|\varphi - \varphi_k\|_\omega^2 + 9\sum_{k=1}^{N_n}\Big(\sup_{t\in\mathcal{B}_k}|\langle t, \Phi_\nu\rangle_\omega|^2 - 6\frac{\mathbb{E}[Y^2]\delta_k^T}{n}\Big)_+$$
$$+ 594\,\mathbb{E}[Y^2]\,n\,(1+\log n)\mathbb{1}\{\Omega_q^c \cap \Omega_p\}$$
$$+ 12\sup_{t\in\mathcal{B}_{N_n}}|\langle t, \widehat{\Phi}_\nu - \Phi_\nu\rangle_\omega|^2\,\mathbb{1}\{\Omega_q^c\} + 12\sup_{t\in\mathcal{B}_{N_n}}|\langle t, \widehat{\Phi}_\mu + \widehat{\Phi}_g - \Phi_g\rangle_\omega|^2.$$

Due to Lemmas A.5,A.7, A.8, and A.9, there exists a numerical constant C such that

$$\mathbb{E}\|\widehat{\varphi}_{\widehat{k}} - \varphi\|_\omega^2 \mathbb{1}\{\Omega_q^c \cap \Omega_p\} \leqslant C\Big\{\|\varphi - \varphi_k\|_\omega^2 + d\rho\max_{j\geqslant 1}\Big\{\frac{\omega_j}{\gamma_j}\min\big(1, \frac{1}{n\lambda_j}\big)\Big\}$$
$$+ (2\rho\Gamma+\sigma^2)n(1+\log n)\mathbf{P}[\Omega_q^c] + d\mathbf{P}[\Omega_q^c]^{1/2} + \frac{(2\rho\Gamma+\sigma^2)^4}{n}$$
$$+ \frac{(2\rho\Gamma+\sigma^2+1)d\,\zeta_d}{n}\,\Sigma\Big(\frac{(2\rho\Gamma+\sigma^2)\zeta_d + V_{U|Z}}{V_{U|Z}^2}\Big)\Big\}.$$

Employing Lemma A.9 now proves (A.26).

Proof of (A.27). Let $\breve{\varphi}_k := \sum_{j=1}^k [\varphi]_j \mathbb{1}\{\widehat{[T]}_{jj}^2 \geqslant 1/n\}\psi_j$. It is easy to see that $\|\widehat{\varphi}_k - \breve{\varphi}_k\|^2 \leqslant \|\widehat{\varphi}_{k'} - \breve{\varphi}_{k'}\|^2$ for all $k' \leqslant k$ and $\|\breve{\varphi}_k - \varphi\|^2 \leqslant \|\varphi\|^2$ for all $k \geqslant 1$. Thus, using that $1 \leqslant \widehat{k} \leqslant N_n^u$, we can write

$$\mathbb{E}\|\widehat{\varphi}_{\widehat{k}} - \varphi\|_\omega^2 \mathbb{1}\{\Omega_p^c\} \leqslant 2\{\mathbb{E}\|\widehat{\varphi}_{\widehat{k}} - \breve{\varphi}_{\widehat{k}}\|_\omega^2 \mathbb{1}\{\Omega_p^c\} + \mathbb{E}\|\breve{\varphi}_{\widehat{k}} - \varphi\|_\omega^2 \mathbb{1}\{\Omega_p^c\}\}$$
$$\leqslant 2\Big\{\mathbb{E}\|\widehat{\varphi}_{N_n^u} - \breve{\varphi}_{N_n^u}\|_\omega^2 \mathbb{1}\{\Omega_p^c\} + \|\varphi\|_\omega^2\,\mathbf{P}[\Omega_p^c]\Big\}.$$

Moreover, since $\sup_{j\geqslant 1}\mathbb{E}Y^4\psi_j^4(W) \leqslant 64(2\rho\Gamma+\sigma^2)^2$ and $\mathbb{E}\psi_j^4(W)\psi_j^4(Z) \leqslant 16$ due to (A.24), it follows from Theorem 2.10 in Petrov (1995) that

$$\mathbb{E}\|\widehat{\varphi}_{N_n^u} - \breve{\varphi}_{N_n^u}\|_\omega^2 \mathbb{1}\{\Omega_p^c\}$$
$$\leqslant 2n\sum_{j=1}^{N_n^u}\omega_j\Big\{\mathbb{E}(\widehat{[g]}_j - [T]_{jj}[\varphi]_j)^2\mathbb{1}\{\Omega_p^c\} + \mathbb{E}([T]_{jj}[\varphi]_j - \widehat{[T]}_{jj}[\varphi]_j)^2\mathbb{1}\{\Omega_p^c\}\Big\}$$
$$\leqslant 2n\Big\{\sum_{j=1}^{N_n^u}\omega_j\Big[\mathbb{E}\left(\widehat{[g]}_j - [g]_j\right)^4\Big]^{1/2}\mathbf{P}[\Omega_p^c]^{1/2}$$
$$+ \sum_{j=1}^{N_n^u}\omega_j|[\varphi]_j|^2[\mathbb{E}(\widehat{[T]}_{jj} - [T]_{jj})^4]^{1/2}\mathbf{P}[\Omega_p^c]^{1/2}\Big\}$$
$$\leqslant Cn\Big\{n\,(2\rho\Gamma+\sigma^2) + (n^{-1}\|\varphi\|_\omega^2)\Big\}\,\mathbf{P}[\Omega_p^c]^{1/2},$$

where we used that $\sum_{j=1}^{N_n^u}\omega_j \leqslant n(\max_{1\leqslant j\leqslant N_n^u}\omega_j) \leqslant n^2$ due to Definition 3.2 (ii). Since (ω/γ) is non-increasing, (A.27) follows from Lemma A.10, which completes the proof. $\square$

Illustration: estimation of derivatives

Proof of Proposition 3.6. In the light of the proof of Proposition 2.8 we apply Theorem 3.4, where in both cases the additional conditions are easily verified (cf. Remark 3.3) and the result follows by an evaluation of the upper bound. Note further that (ω/λ) is in both cases non decreasing, and hence the second term in the upper bound of Theorem 3.4 is always smaller than the first one.

In case **[fs]** we have $N_n^l \sim (n/(\log n))^{1/(2a+2s+1)}$ and $k_n^* := n^{1/(2a+2p+1)}$. Note that $k_n^* \lesssim N_n^l$. Thus, the upper bound is of order $O((k_n^*)^{-2(p-s)} + n^{-1}) = O(n^{-2(p-s)/(2a+2p+1)})$.

In case **[is]** we have $N_n^l \sim \{\log(n/(\log n)^{(2p+2a+1)/(2a)})\}^{1/(2a)} = (\log n)^{1/(2a)}(1+o(1)) \sim k_n^*$. Thereby, the upper bound is of order $O((k_n^*)^{-2(p-s)} + n^{-1}) = O((\log n)^{-(p-s)/a})$, which completes the proof. $\square$

Technical assertions. In the proof of Lemma A.5 below we will need the following Lemma, which can be found in Comte et al. (2006).

Lemma A.4 (Talagrand's Inequality) *Let $T_1, \dots, T_n$ be independent random variables and $\nu_n^*(r) = (1/n)\sum_{i=1}^n \big[r(T_i) - \mathbb{E}[r(T_i)]\big]$, for r belonging to a countable class $\mathcal{R}$ of measurable functions. Then,*

$$\mathbb{E}[\sup_{r\in\mathcal{R}} |\nu_n^*(r)|^2 - 6H_2^2]_+ \leqslant C\left(\frac{v}{n}\exp(-(nH_2^2/6v)) + \frac{H_1^2}{n^2}\exp(-K_2(nH_2/H_1))\right)$$

with numerical constants $K_2 = (\sqrt{2}-1)/(21\sqrt{2})$ and C and where

$$\sup_{r\in\mathcal{R}}\|r\|_\infty \leqslant H_1, \quad \mathbb{E}\left[\sup_{r\in\mathcal{R}}|\nu_n^*(r)|\right] \leqslant H_2, \quad \sup_{r\in\mathcal{R}}\frac{1}{n}\sum_{i=1}^n \mathbb{V}\mathrm{ar}(r(T_i)) \leqslant v.$$

Lemma A.5 *There exists a numerical constant $C > 0$ such that*

$$\sum_{k=1}^{N_n}\mathbb{E}\Big[\Big(\sup_{t\in\mathcal{B}_k}|\langle t, \Phi_\nu\rangle_\omega|^2 - \frac{6\,\mathbb{E}[Y^2]\,\delta_k^T}{n}\Big)_+\Big] \leqslant \frac{C}{n}\left\{(2\rho\Gamma+\sigma^2+1)d\,\zeta_d\;\Sigma\left(\frac{(2\rho\Gamma+\sigma^2)\zeta_d + V_{U|Z}}{V_{U|Z}^2}\right)\right\}.$$

where $\Sigma(\cdot)$ is the function from Definition 3.1

Proof. For $t \in \mathcal{S}_k$ define the function $r_t(y,w) := \sum_{j=1}^k \omega_j y \mathbb{1}\{|y| \leqslant n^{1/3}\}\psi_j(w)[t]_j[T]_{jj}^{-1}$, then it is readily seen that $\langle t, \Phi_\nu\rangle_\omega = \frac{1}{n}\sum_{k=1}^n r_t(Y_k, W_k) - \mathbb{E}[r_t(Y_k, W_k)]$.

Next, we compute constants H_1, H_2, and v verifying the three inequalities required in Lemma A.4. Consider H_1 first:

$$\sup_{t\in\mathcal{B}_k}\|r_t\|_\infty^2 = \sup_{y,w}\sum_{j=1}^k \omega_j\left(y\mathbb{1}\{|y| \leqslant n^{1/3}\}[T]_{jj}^{-1}\psi_j(w)\right)^2 \leqslant 2n^{2/3}\delta_k^T =: H_1^2$$

Next, find H_2. Notice that

$$\begin{aligned}\mathbb{E}[\sup_{t\in\mathcal{B}_k}|\langle t, \Phi_\nu\rangle_\omega|^2] &= \frac{1}{n}\sum_{j=1}^k \omega_j|[T]_{jj}|^{-2}\;\mathbb{V}\mathrm{ar}(Y\mathbb{1}\{|Y| \leqslant n^{1/3}\}\psi_j(W)) \\ &\leqslant \frac{1}{n}\sum_{j=1}^k \omega_j|[T]_{jj}|^{-2}\;\mathbb{E}[\mathbb{E}[Y^2|W]\,\psi_j(W)^2] \leqslant 2\mathbb{E}[Y^2]\frac{\delta_k^T}{n} =: H_2^2\end{aligned}$$

As for v, we note that due to (A.24) for all $\varphi \in \mathcal{F}_\gamma^\rho$ the condition $U \in \mathcal{U}_\sigma$, i.e., $\mathbb{E}U^2|W \leqslant \sigma^2$, implies $\mathbb{E}Y^2|W \leqslant 2(2\rho\Gamma + \sigma^2)$, and hence

$$
\begin{aligned}
\sup_{t\in\mathcal{B}_k} \mathbb{V}\mathrm{ar}(r_t(Y,W)) &\leqslant \sup_{t\in\mathcal{B}_k} \mathbb{E}\Bigg[\bigg(Y\sum_{j=1}^{k} \frac{\omega_j[t]_j}{[T]_{jj}}\psi_j(W)\bigg)^2\Bigg] \\
&= \sup_{t\in\mathcal{B}_k} \mathbb{E}\Bigg[\mathbb{E}[Y^2|W]\,\bigg(\sum_{j=1}^{k} \frac{\omega_j[t]_j}{[T]_{jj}}\psi_j(W)\bigg)^2\Bigg] \\
&\leqslant 2(2\rho\Gamma+\sigma^2)\sup_{t\in\mathcal{B}_k}\sum_{j,j'=1}^{k} \frac{\omega_j\omega_{j'}[t]_j[t]_{j'}}{[T]_{jj}[T]_{j'j'}}\mathbb{E}[\psi_j(W)\psi_{j'}(W)] \\
&\leqslant 2(2\rho\Gamma+\sigma^2)\max_{1\leqslant j\leqslant k}\frac{\omega_j}{[T]_{jj}^2}\sup_{t\in\mathcal{B}_k}\sum_{j=1}^{k}\omega_j[t]_j^2 \leqslant 2(2\rho\Gamma+\sigma^2)\Delta_k^T =: v,
\end{aligned}
$$

By employing Lemma A.4 we conclude

$$
\begin{aligned}
&\sum_{k=1}^{N_n}\mathbb{E}\Big[\Big(\sup_{t\in\mathcal{B}_k}|\langle t,\Phi_\nu\rangle_\omega|^2 - \frac{6\,\mathbb{E}[Y^2]\,\delta_k^T}{n}\Big)_+\Big] \\
&\leqslant C\left\{\frac{\mathbb{E}[Y^2]}{n}\sum_{k=1}^{N_n}\frac{(2\rho\Gamma+\sigma^2)}{\mathbb{E}[Y^2]}\,\Delta_k^T\exp\left(-\frac{\mathbb{E}[Y^2]}{6(2\rho\Gamma+\sigma^2)}(\delta_k^T/\Delta_k^T)\right)\right. \\
&\qquad\qquad\qquad\left. + n^{2/3}\exp\left(-K_2\sqrt{\mathbb{E}[Y^2]}n^{1/6}\right)\sum_{k=1}^{N_n}\frac{\delta_k^T}{n^2}\right\}.
\end{aligned}
$$

The definition of N_n together with (A.21) implies $\sum_{k=1}^{N_n}\delta_k^T/n^2 \leqslant \zeta_d$. Thereby, using (A.21), $\Delta_k^T \leqslant d\tau_k$ and the function Σ given in Definition 3.1, there exists a numerical constant $C > 0$ such that

$$
\begin{aligned}
&\sum_{k=1}^{N_n}\mathbb{E}\Big[\Big(\sup_{t\in\mathcal{B}_k}|\langle t,\Phi_\nu\rangle_\omega|^2 - \frac{6\,\mathbb{E}[Y^2]\,\delta_k^T}{n}\Big)_+\Big] \\
&\qquad\qquad \leqslant \frac{C}{n}\left\{\mathbb{E}[Y^2]d\,\Sigma\Big(\frac{(2\rho\Gamma+\sigma^2)\zeta_d}{\mathbb{E}[Y^2]}\Big) + \zeta_d\Sigma\Big(\frac{1}{\sqrt{\mathbb{E}[Y^2]}}\Big)\right\}.
\end{aligned}
$$

Moreover, we have $\mathbb{E}[Y^2] \leqslant 2(2\rho\Gamma+\sigma^2)$ and $\inf_{\varphi\in\mathcal{F}_\gamma^\rho}\mathbb{E}[Y^2] \geqslant \inf_{\varphi\in L_Z^2}\mathbb{E}(\varphi(Z)+U)^2 \geqslant \mathbb{E}(U-\mathbb{E}[U|Z])^2 = \mathbb{E}[\mathbb{V}\mathrm{ar}(U|Z)] = V_{U|Z}^2$, which implies the result. □

Lemma A.6 *For every $n \in \mathbb{N}$ we have*

$$
\mathbb{E}\Big[\sup_{t\in\mathcal{B}_{N_n}}|\langle t,\widehat{\Phi}_\mu\rangle_\omega|^2\Big] \leqslant 2^9(2\rho\Gamma+\sigma^2)^4 n^{-1}.
$$

Proof. Since $[\mu]_j = [\widehat{\mu}]_j - \mathbb{E}[\widehat{\mu}]_j$ and $\mathbb{V}\mathrm{ar}[\widehat{\mu}]_j \leqslant n^{-1}\mathbb{E}Y^2\mathbb{1}\{|Y| > n^{1/3}\}\psi_j^2(W)$, it is easily seen that

$$\mathbb{E}\Big[\sup_{t\in\mathcal{B}_{N_n}} |\langle t,\widehat{\Phi}_\mu\rangle_\omega|^2\Big] \leqslant n\sum_{j=1}^{N_n}\omega_j\,\mathbb{V}\mathrm{ar}[\widehat{\mu}]_j$$
$$\leqslant \sum_{j=1}^{N_n}\mathbb{E}\bigg[\Big(\mathbb{E}[Y^4|W]\mathbb{E}[\mathbb{1}\{|Y|>n^{1/3}\}|W]\Big)^{1/2}\psi_j^2(W)\bigg].$$

Moreover, given $m=6$ we have $\mathbb{E}[Y^{12}|W]\leqslant 2^{12}(2\rho\Gamma+\sigma^2)^6$for all $\varphi\in\mathcal{F}_\gamma^\rho$ and $U\in\mathcal{U}_\sigma$ due to (A.24) and, hence by Markov's inequality $\mathbb{E}[\mathbb{1}\{|Y|>n^{1/3}\}|W]\leqslant 2^{12}(2\rho\Gamma+\sigma^2)^6n^{-4}$. Combining these estimates, we obtain

$$\mathbb{E}\Big[\sup_{t\in\mathcal{B}_{N_n}} |\langle t,\widehat{\Phi}_\mu\rangle_\omega|^2\Big] \leqslant \sum_{j=1}^{N_n}\mathbb{E}\Big[2^8(2\rho\Gamma+\sigma^2)^4n^{-2}\psi_j^2(W)\Big]\leqslant 2^9N_n(2\rho\Gamma+\sigma^2)^4n^{-2}.$$

The result follows now from $N_n\leqslant n$. □

Lemma A.7 *There is a numerical constant $C>0$ such that for all $\varphi\in\mathcal{F}_\gamma^\rho$ and every $k,n\in\mathbb{N}$*

$$\mathbb{E}\Big[\sup_{t\in\mathcal{B}_k}|\langle t,\widehat{\Phi}_g-\Phi_g\rangle_\omega|^2\Big]\leqslant Cd\rho\max_{j\geqslant1}\Big\{\frac{\omega_j}{\gamma_j}\min\big(1,\frac{1}{n\lambda_j}\big)\Big\}.$$

Proof. Firstly, as $\varphi\in\mathcal{F}_\gamma^\rho$, it is easily seen that

$$\mathbb{E}\Big[\sup_{t\in\mathcal{B}_k}|\langle t,\widehat{\Phi}_g-\Phi_g\rangle_\omega|^2\Big]\leqslant\sum_{j=1}^k[\varphi]_j^2\omega_j\mathbb{E}[R_j^2]\leqslant\rho\max_{j\geqslant1}\Big\{\frac{\omega_j}{\gamma_j}\mathbb{E}[R_j^2]\Big\}$$

where R_j is defined by

$$R_j:=\left(\frac{[T]_{jj}}{\widehat{[T]}_{jj}}\mathbb{1}\{\widehat{[T]}_{jj}^2\geqslant1/n\}-1\right). \tag{A.28}$$

The result follows from $\mathbb{E}R_j^2\leqslant Cd\min\big(1,\frac{1}{n\lambda_j}\big)$, which can be shown as follows. Consider the identity

$$\mathbb{E}|R_j|^2=\mathbb{E}\bigg[\,\bigg|\frac{[T]_{jj}}{\widehat{[T]}_{jj}}-1\bigg|^2\mathbb{1}\{\widehat{[T]}_{jj}^2\geqslant1/n\}\,\bigg]+\mathbf{P}[\widehat{[T]}_{jj}^2<1/n]=:R_j^I+R_j^{II}. \tag{A.29}$$

Trivially, $R_j^{II}\leqslant1$. If $1\leqslant4/(n\,[T]_{jj}^2)$, then obviously $R_j^{II}\leqslant4/(n[T]_{jj}^2)\leqslant4d/(n\lambda_j)$. Otherwise, we have $1/n<[T]_{jj}^2/4$ and hence, using Tchebychev's inequality,

$$R_j^{II}\leqslant\mathbf{P}[|\widehat{[T]}_{jj}-[T]_{jj}|>|[T]_{jj}|/2\,]\leqslant\frac{4\,\mathbb{V}\mathrm{ar}(\widehat{[T]}_{jj})}{[T]_{jj}^2}\leqslant\frac{16}{n[T]_{jj}^2}\leqslant\frac{16d}{n\lambda_j},$$

where we have used that $\mathbb{V}\mathrm{ar}(\widehat{[T]}_{jj})\leqslant4/n$ for all j. Combining both estimates we have $R_j^I\leqslant16d\min\big(1,\frac{1}{n\lambda_j}\big)$. Now consider R_j^I. We find that

$$R_j^I=\mathbb{E}\bigg[\frac{|\widehat{[T]}_{jj}-[T]_{jj}|^2}{\widehat{[T]}_{jj}^2}\;\mathbb{1}\{\widehat{[T]}_{jj}^2\geqslant1/n\}\bigg]\leqslant n\,\mathbb{V}\mathrm{ar}(\widehat{[T]}_{jj})\leqslant4. \tag{A.30}$$

Using that $\mathbb{E}[|\widehat{[T]}_{jj}-[T]_{jj}|^4] \leqslant c/n^2$ for some numerical constant $c>0$ (cf. Petrov (1995), Theorem 2.10), there exists a numerical constant $c>0$ such that

$$R_j^I \leqslant \mathbb{E}\Bigg[\frac{|\widehat{[T]}_{jj}-[T]_{jj}|^2}{\widehat{[T]}_{jj}^2}\ \mathbb{1}\{\widehat{[T]}_{jj}^2 \geqslant 1/n\}\ 2\Bigg\{\frac{|\widehat{[T]}_{jj}-[T]_{jj}|^2}{[T]_{jj}^2}+\frac{\widehat{[T]}_{jj}^2}{[T]_{jj}^2}\Bigg\}\Bigg]$$
$$\leqslant \frac{2\,n\,\mathbb{E}[|\widehat{[T]}_{jj}-[T]_{jj}|^4]}{[T]_{jj}^2}+\frac{2\ \mathbb{V}\mathrm{ar}(\widehat{[T]}_{jj})}{[T]_{jj}^2} \leqslant \frac{c}{n\,[T]_{jj}^2} \leqslant \frac{cd}{n\lambda_j}.$$

Combining with (A.30) gives $R_j^I \leqslant Cd\min\left\{1,\frac{1}{n\lambda_j}\right\}$ for some numerical constant $C>0$, which completes the proof. □

Lemma A.8 *There is a numerical constant $C>0$ such that*

$$\mathbb{E}\Bigg[\sup_{t\in\mathcal{B}_{N_n}} |\langle t,\widehat{\Phi}_\nu-\Phi_\nu\rangle_\omega \mathbb{1}\{\Omega_q^c\}|^2\Bigg] \leqslant Cd(\mathbf{P}[\Omega_q^c])^{(1/2)}.$$

Proof. Given with R_j from (A.28) we begin our proof observing that

$$\mathbb{E}\Bigg[\sup_{t\in\mathcal{B}_{M_m}} |\langle t,\widehat{\Phi}_\nu-\Phi_\nu\rangle_\omega \mathbb{1}\{\Omega_q^c\}|^2\Bigg] \leqslant \sum_{j=1}^{N_n}\frac{\omega_j}{[T]_{jj}^2}\,\mathbb{E}[[\nu]_j^2\,R_j^2\,\mathbb{1}\{\Omega_q^c\}]$$
$$\leqslant \sum_{j=1}^{N_n}\frac{\omega_j}{[T]_{jj}^2}\,\left(\mathbb{E}[[\nu]_j^8]\mathbb{E}[R_j^8]\right)^{1/4}\mathbf{P}[\Omega_q^c]^{1/2},$$

where we have applied Cauchy-Schwarz twice. By Petrov's inequality, there exists a numerical constant $c>0$ such that $E[[\nu]_j^8] \leqslant cn^{-4/3}$ and hence, because $d\delta_k \geqslant \sum_{j=1}^k \frac{\omega_j}{[T]_{jj}^2}$,

$$\mathbb{E}\Bigg[\sup_{t\in\mathcal{B}_{M_m}} |\langle t,\widehat{\Phi}_\nu-\Phi_\nu\rangle_\omega \mathbb{1}\{\Omega_q^c\}|^2\Bigg] \leqslant \mathbf{P}[\Omega_q^c]^{1/2} d\delta_k \max_{1\leqslant j\leqslant N_n}(\mathbb{E}[R_j^8])^{1/4}$$

In analogy to (A.29), we decompose the moment of R_j into two terms

$$\mathbb{E}[R_j^8] = \mathbb{E}\Bigg[\,\Bigg|\frac{[T]_{jj}-\widehat{[T]}_{jj}}{\widehat{[T]}_{jj}}\Bigg|^8 \mathbb{1}\{\widehat{[T]}_{jj}^2 \geqslant 1/n\}\,\Bigg] + \mathbf{P}[\widehat{[T]}_{jj}^2 < 1/n],$$

which we bound by a constant using Petrov's inequality. This completes the proof. □

Lemma A.9 *For the event Ω_q defined in (A.19), we have $\mathbf{P}[\Omega_q^c] \leqslant 2(2016d/\lambda_1)^7\,n^{-6}$.*

Proof. Consider the complement of Ω_q given by

$$\Omega_q^c = \Bigg\{\exists\, 1\leqslant j\leqslant N_n\ :\ \Bigg|\frac{[T]_{jj}}{\widehat{[T]}_{jj}}-1\Bigg| > \frac{1}{2}\ \vee\ \widehat{[T]}_{jj}^2 < 1/n\Bigg\}.$$

It follows from Assumption A4 (i) that $[T]_{jj}^2 \geqslant 2/n$ for all $1\leqslant j\leqslant N_n$. This yields

$$\mathbf{P}(\Omega_q^c) \leqslant \sum_{j=1}^{N_n}\mathbf{P}\Bigg[\Bigg|\frac{\widehat{[T]}_{jj}}{[T]_{jj}}-1\Bigg| > \frac{1}{3}\Bigg].$$

From Hoeffding's inequality follows

$$\mathbf{P}[|\widehat{[T]}_{jj}/[T]_{jj} - 1| > 1/3] \leqslant 2 \exp\Big(-\frac{n[T]_{jj}^2}{288}\Big),$$

which implies the result by definition of N_n. $\square$

Lemma A.10 *Consider the event Ω_p defined in* (A.19). *Then we have*

$$\mathbf{P}(\Omega_p^c) \leqslant 4\left(\frac{2016\, d}{\lambda_1}\right)^7 n^{-6}, \qquad \forall\, n \geqslant 1.$$

Proof. Let $\Omega_I := \{N_n^l > \widehat{N}_n\}$ and $\Omega_{II} := \{\widehat{N}_n > N_n\}$. Then we have $\Omega_p^c = \Omega_I \cup \Omega_{II}$. Consider Ω_I first. By definition of N_n^l, we have that $\min_{1\leqslant j\leqslant N_n^l} \frac{|[T]_j|^2}{|j|(\omega_j)_{\vee 1}} \geqslant \frac{4(\log n)}{n}$, which implies

$$\begin{aligned}\{\widehat{N}_n < N_n^l\} &\subset \left\{\exists 1 \leqslant j \leqslant N_n^l : \frac{\widehat{[T]}_{jj}^2}{|j|(\omega_j)_{\vee 1}} < \frac{\log n}{n}\right\}\\ &\subset \bigcup_{1\leqslant j\leqslant N_n^l}\left\{\frac{|\widehat{[T]}_{jj}|}{|[T]_{jj}|} \leqslant 1/2\right\} \subset \bigcup_{1\leqslant j\leqslant N_n^l}\left\{|\widehat{[T]}_{jj}/[T]_{jj} - 1| \geqslant 1/2\right\}.\end{aligned}$$

Therefore, $\Omega_I \subset \bigcup_{1\leqslant |j|\leqslant N_n}\Big\{|\widehat{[\varphi]}_j/[\varphi]_j - 1| \geqslant 1/2\Big\}$, since $N_n^l \leqslant N_n$. Hence, as in (A.21) applying Hoeffding's inequality together with the definition of N_n gives

$$\mathbf{P}[\Omega_I] \leqslant \sum_{j=1}^{N_n} 2 \exp\Big(-\frac{n\,[T]_{jj}^2}{288}\Big) \leqslant 2\left(\frac{2016\, d}{\lambda_1}\right)^7 n^{-6}. \tag{A.31}$$

Consider Ω_{II}. Recall that $\frac{\log n}{4n} \geqslant \max_{|j|\geqslant N_n} \frac{[T]_{jj}^2}{|j|(\omega_j)_{\vee 1}}$ due to Assumption A4, and hence

$$\begin{aligned}\{\widehat{N}_n > N_n\} &\subset \left\{\forall 1 \leqslant j \leqslant N_n : \frac{\widehat{[T]}_{jj}^2}{|j|(\omega_j)_{\vee 1}} \geqslant \frac{\log n}{n}\right\}\\ &\subset \left\{\frac{|\widehat{[T]}_{N_n}|}{|[T]_{N_n}|} \geqslant 2\right\} \subset \left\{|\widehat{[T]}_{N_n}/[T]_{N_n} - 1| \geqslant 1\right\}.\end{aligned}$$

Hoeffding's inequality together with the definition of N_n gives $\mathbf{P}[\Omega_{II}] \leqslant 2(2016d/\lambda_1)^7 n^{-6}$, which by combining with (A.31) implies the result. $\square$

References

Ai, C. and Chen, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica*, 71:1795–1843.

Amemiya, T. (1974). The nonlinear two-stage least square estimator. *J. Econometrics*, 2:105–110.

Barron, A., Birgé, L., and Massart, P. (1999). Risk bounds for model selection via penalization. *Probability Theory and Related Fields*, 113(3):301–413.

Blundell, R., Chen, X., and Kristensen, D. (2007). Semi-nonparametric iv estimation of shape-invariant engel curves. *Econometrica.* forthcoming.

Blundell, R. and Horowitz, J. (2007). A nonparametric test of exogeneity. *Rev. Econ. Stud.* forthcoming.

Bosq, D. (1998). *Nonparametric statistics for stochastic processes.* Springer, New York.

Cardot, H. and Johannes, J. (2010). Thresholding projection estimators in functional linear models. *Journal of Multivariate Analysis*, 101:395–408.

Carrasco, M., Florens, J.-P., and Renault, E. (2006). Linear inverse problems in structural econometrics: Estimation based on spectral decomposition and regularization. In Heckman, J. and Leamer, E., editors, *Handbook of Econometrics*, volume 6. North Holland.

Chen, X. and Reiß, M. (2008). On rate optimality for ill-posed inverse problems in econometrics. Technical report, Yale University.

Comte, F. and Johannes, J. (2010). Adaptive estimation in circular functional linear models. *To appear in Mathematical Methods of Statistics.*

Comte, F., Rozenholc, Y., and Taupin, M.-L. (2006). Penalized contrast estimator for density deconvolution. *Canadian Journal of Statistics*, 37(3).

Comte, F. and Taupin, M.-L. (2003). Adaptive density deconvolution for circular data. Prépublication map5 2003-10, Université Paris Descartes.

Darolles, S., Florens, J. P., and Renault, E. (2002). Nonparametric instrumental regression. Working paper 05-2002, CRDE.

Efromovich, S. and Koltchinskii, V. (2001). On inverse problems with unknown operators. *IEEE Transactions on Information Theory*, 47(7):2876–2894.

Engl, H. W., Hanke, M., and Neubauer, A. (2000). *Regularization of inverse problems.* Kluwer Academic, Dordrecht.

Florens, J.-P. (2003). Inverse problems and structural econometrics: The example of instrumental variables. In Dewatripont, M., Hansen, L. P., and Turnovsky, S. J., editors, *Advances in Economics and Econometrics: Theory and Applications – Eight World Congress*, volume 36 of *Econometric Society Monographs.* Cambridge University Press.

Florens, J.-P., Johannes, J., and Van Bellegem, S. (2009). Instrumental regression in partially linear models. Discussion Paper # 0537, Institut de statistique, Université catholique de Louvain (first version 2005, revised).

Florens, J. P., Johannes, J., and Van Bellegem, S. (2010). Identification and estimation by penalization in nonparametric instrumental regression. *To appear in Econometric Theory.*

Gagliardini, P. and Scaillet, O. (2006). Tikhonov regularization for functional minimum distance estimators. Swiss Finance Institute Research Paper No. 06-30.

Hall, P. and Horowitz, J. L. (2005). Nonparametric methods for inference in the presence of instrumental variables. *Annals of Statistics*, 33:2904–2929.

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *Annals of Statistics*, 35(1):70–91.

Hoffmann, M. and Reiß, M. (2008). Nonlinear estimation for linear inverse problems with error in the operator. *The Annals of Statistics*, 36(1):310–336. Discussion Paper # 990, WIAS, Berlin.

Horowitz, J. L. and Lee, S. (2007). Nonparametric instrumental variables estimation of a quantile regression model. *Econometrica*, 75:1191–1208.

Johannes, J. and Breunig, C. (2009). On rate optimal local estimation in nonparametric instrumental regression. Technical report, University Heidelberg (submitted.). http://arxiv.org/abs/0902.2103.

Johannes, J. and Schwarz, M. (2010). Adaptive circular deconvolution by model selection under unknown error distribution. Technical report, Université catholique de Louvain (submitted.).

Korostolev, A. P. and Tsybakov, A. B. (1993). *Minimax Theory for Image Reconstruction.*, volume 82 of *Lecture Notes in Statistics*. Springer-Verlag.

Loubes, J.-M. and Marteau, C. (2009). Oracle inequality for instrumental variable regression. http://arXiv.org/abs/0901.4321.

Natterer, F. (1984). Error bounds for Tikhonov regularization in Hilbert scales. *Applicable Analysis*, 18:29–37.

Neubauer, A. (1988a). An a posteriori parameter choice for Tikhonov regularization in Hilbert scales leading to optimal convergence rates. *SIAM Journal on Numerical Analysis*, 25(6):1313–1326.

Neubauer, A. (1988b). When do Sobolev spaces form a Hilbert scale? *Proceedings of the American Mathematical Society*, 103(2):557–562.

Newey, W. K. (1990). Efficient instrumental variables estimation of nonlinear models. *Econometrica*, 58:809–837.

Newey, W. K. and Powell, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, 71:1565–1578.

Olver, F. (1974). *Asymptotics and special functions.* Academic Press.

Pagan, A. and Ullah, A. (1999). *Nonparametric Econometrics.* Cambridge University Press.

Petrov, V. V. (1995). *Limit theorems of probability theory. Sequences of independent random variables.* Oxford Studies in Probability. Clarendon Press., Oxford, 4. edition.

Recent Titles

1003. HUNT, J., A short note on continuous-time Markov and semi-Markov processes
1004. COLLÉE, A., LEGRAND, C., GOVAERTS, B., VAN DER VEKEN, P., DE BOODT, F. and E. DEGRAVE, Occupational exposure to noise and the prevalence of hearing loss in a Belgian military population: a cross-sectional study. Military Medicine, under revision.
1005. MEYER, N., LEGRAND, C. and G. GIACCONE, Samples sizes in oncology trials: a survey.To be submitted in the coming weeks (British Medical Journal).
1006. HAFNER, C.M. and O. REZNIKOVA, On the estimation of dynamic conditional correlation models
1007. HEUCHENNE, C. and I. VAN KEILEGOM, Estimation of a general parametric location in censored regression
1008. DAVYDOV, Y. and S. LIU, Transformations of multivariate regularly varying tail distributions
1009. CHRISTIANSEN, M. and M. DENUIT, First-order mortality rates and safe-side actuarial calculations in life insurance
1010. EECKHOUDT, L. and M. DENUIT, Stronger measures of higher-order risk attitudes
1011. DENUIT, M., HABERMAN, S. and A. RENSHAW, Comonotonic approximations to quantiles of life annuity conditional expected present values: extensions to general ARIMA models and comparison with the bootstrap
1012. DENUIT, M. and M. MESFIOUI, Generalized increasing convex and directionally convex orders
1013. DENUIT, M. and L. EECKHOUDT, A general index of absolute risk attitude
1014. PIGEON, M. and M. DENUIT, Composite Lognormal-Pareto model with random threshold
1015. DENUIT, M., EECKHOUDT, L. and M. MENEGATTI, Correlated risks, bivariate utility and optimal choices
1016. TRUFIN, J., ALBRECHER, H. and M. DENUIT, Ruin problems in presence of underwriting cycles
1017. DENUIT, M., EECKHOUDT, L. and B. REY, Somes consequences of correlation aversion in decision science
1018. DENUIT, M. and L. EECKHOUDT, Bivariate stochastic dominance and common preferences of decision-makers with risk independent utilities
1019. TRUFIN, J., ALBRECHER, H. and M. DENUIT, Properties of risk measures derived from ruin theory
1020. DENUIT, M., EECKHOUDT, L. and M. MENEGATTI, Adding independent risks in an insurance portfolio: Which shape for the insurer's preferences?
1021. DENUIT, M. and B. REY, Prudence, temperance, edginess, and higher degree risk apportionment as decreasing correlation aversion
1022. TRUFIN, J., ALBRECHER, H. and M. DENUIT, Ruin problems under IBNR dynamics
1023. DENUIT, M., HABERMAN, S. and A. RENSHAW, Longevity-indexed life annuities
1024. DENUIT, M., EECKHOUDT, L. and O. JOKUNG, Transformations preserving stochastic dominance: Theory and applications
1025. DENUIT, M., Positive dependence of signals
1026. JOHANNES, J. and M. SCHWARZ, Adaptive nonparametric instrumental regression by model selection

See also http://www.stat.ucl.ac.be/ISpub.html