2007/83

■

A unified approach to solve ill-posed inverse problems in econometrics

Jan Johannes, Sébastien Van Bellegem and Anne Vanhems

CORE DISCUSSION PAPER
2007/83

A unified approach to solve ill-posed inverse problems in econometrics

Jan JOHANNES[1], Sébastien VAN BELLEGEM[2]
and Anne VANHEMS[3]

October 2007

Abstract

We consider the general issue of estimating a nonparametric function φ from the inverse problem $r = T\varphi$ given estimates of the function r and of the linear transform T. Two typical examples include the estimation of a probability density function from data contaminated by a noise whose distribution is unknown (blind deconvolution) and the nonparametric instrumental regression. We provide a unified framework based on Hilbert scales that synthesizes most of existing results in the econometric literature and also covers new relevant structural models. Results are given on the rate of convergence of the estimator of φ as well as of its derivatives.

Keywords: inverse problem, Hilbert scale, deconvolution, instrumental variable, nonparametric regression.

JEL Classification: primary C14, secondary C30

---

[1] Universität Heidelberg, Germany. E-mail: Johannes@statlab.uni-heidelberg.de
[2] Institut de Statistique and CORE, Université catholique de Louvain, Belgium. E-mail: vanbellegem@stat.ucl.ac.be. The author is also member of ECORE, the newly created association between CORE and ECARES.
[3] Toulouse Business School, France. E-mail: a.vanhems@esc-toulouse.fr

Financial support from the IAP research network no P6/03 of the Belgian Government (Belgian Science Policy) is gratefully acknowledged.

This paper presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsibility is assumed by the authors.

1 Introduction

A wide range of econometric problems are related to the identification and estimation of a nonparametric function φ from a structural model

$$r = T\varphi, \tag{1.1}$$

where r and T is a function and a linear transform, respectively, that are known or can be estimated from observations. The goal of this paper is to provide a general framework to construct estimates of φ that can be applied in various situations, and to derive their rates of convergence.

One particular situation that necessitates to solve that model is given when φ is the probability density function (pdf) of a random variable that is observed with an additive measurement error. Measurement error is a frequent problem in data analysis, meaning that the variable of interest, X, is not observed directly, but instead a noisy version is observed

$$Y = X + \varepsilon, \tag{1.2}$$

where ε represents some additive measurement error. In terms of densities, if f_Y, resp. f_ε, denotes the pdf of Y, resp. ε, then from (1.2) the density φ of X is the solution of the integral equation

$$f_Y(y) = \varphi \star f_\varepsilon(y) := \int \varphi(u) f_\varepsilon(y-u)du. \tag{1.3}$$

This problem is therefore a particular case of (1.1), where r is the pdf of Y and T is the transform defined by $T\varphi = \varphi \star f_\varepsilon$, also called "convolution". In practice, f_Y has to be estimated from an Y-sample, and most economic studies assume that the pdf of ε is known, i.e., the transform T is known (see Horowitz (1998), Postel-Vinay and Robin (2002), Carrasco and Florens (2002) among others). Below in this paper, we reconsider deconvolution problem in the light of our treatment of the general model (1.1) and provide new results including the case when the pdf of ε is unknown.

A second example that leads to the model (1.1) is given when φ is solution of the moment equation

$$\mathbb{E}[Y|W] = \mathbb{E}[\varphi(Z)|W], \tag{1.4}$$

where Y is a dependent variable, Z is a vector of endogenous explanatory variables and W is a vector of instruments. In the literature, φ is said to be a nonparametric instrumental regression function. Identification and estimation of φ have been the subject of many recent economic studies (Darolles, Florens, and Renault (2002), Newey and Powell (2003), Hall and Horowitz (2005), Gagliardini and Scaillet (2007), Blundell and Horowitz (2007), Blundell, Chen, and Kristensen (2007) to name but a few). In the setting of (1.1), the function r is the conditional expectation $\mathbb{E}[Y|W]$ and T is the conditional expectation operator, i.e., $T\varphi = \mathbb{E}[\varphi(Z)|W] = \int \varphi(z) f_{Z|W}(z)dz$ where $f_{Z|W}$ denotes the multivariate conditional density of Z given W. This example is also expanded below in the paper, and the resolution in the general setting (1.1) brings new results on the rate of convergence of the nonparametric instrumental regression estimator.

The common aspect of all these examples is that the function φ is not directly observed in (1.1) but only through a transform T. Therefore, an inversion of the transform T, or of its estimate, is necessary in order to recover φ and that is why the estimation of φ given in (1.1) is called an inverse problem. Moreover, because the inversion of T is not stable in general (T^{-1} is not a bounded operator in the most relevant studies), the problem is called "ill-posed" and a stabilization, or regularization step, is mandatory in the estimation procedure.

The goal of this paper is to offer a synthesis of the econometric literature on the estimation of φ through model (1.1). We provide a unified framework for the general treatment of this problem, that includes new results, in particular on the two above examples of deconvolution and nonparametric instrumental regression.

One challenging issue in economic inverse problems is the structural assumption we impose on the model in order to derive rates of convergence. Several proposals have been provided in the literature. Below we define our assumptions in term of *Hilbert scales*, that are defined and illustrated in Section 2. If the transform T is known, several approaches of regularized estimation in Hilbert scales have been considered in the numerical analysis and statistical literature (e.g. Hegland (1995), Tautenhahn (1996), Mair and Ruymgaart (1996), Engl, Hanke, and Neubauer (2000), Mathé and Pereverzev (2001), Goldenshluger and Pereverzev (2003)). However the transform T is typically unknown in econometrics, and thus has to be estimated. In that context, formulating our structural assumptions in terms of Hilbert scales is new and extremely useful. Indeed, different sufficient conditions in the literature can be written in Hilbert scales, so that our work allows to compare between various estimators and rates of convergence given in the previous economic studies. Our framework covers in particular the assumptions of Hall and Horowitz (2005), the Normal model (that is known to be difficult to solve — see Florens, Johannes, and Van Bellegem (2007)), many particular situations such as deconvolution of a Laplace density with Normal measurement error, etc. Our results extend the seminal work of Johannes and Vanhems (2005) on Hilbert scales in econometrics, and parallels the recent study of Chen and Reiss (2007).

In Section 3, we derive rates of convergence of the estimator of φ as a function of rates of convergence of the estimates of r and T. One advantage of formulating assumptions in terms of Hilbert scale is that it also allows to derive the rate of convergence of φ in the norm induced by the Hilbert scale, without any additional assumption. To illustrate that point, we consider a particular Hilbert scale that directly gives the rate of convergence of the derivatives of φ.

Section 4 applies the results of Section 3 to the two above mentioned application, that is the deconvolution with unknown error distribution and the nonparametric instrumental regression. In these examples, we use explicit estimators of r and T and derive the rates of convergence as a function of the sample size. Proof of all results are given in a technical appendix.

2 Hilbert scales

One goal of this paper is to analyse the convergence of $\hat{\varphi}$ to φ in various norms, that is we are not only interested to study the mean square convergence of $\hat{\varphi}$, but also, for instance, the mean square convergence of its derivatives. A crucial element when analyzing rates of convergence from an inverse problem is the structural assumptions we impose on φ and T. In nonparametric econometrics, smoothness assumptions are often imposed on the solution φ in order to derive these rates. These assumptions impose that φ belongs e.g. to some known Sobolev or Besov space. However, as argued in Florens, Johannes, and Van Bellegem (2007) in the specific context of nonparametric instrumental regression, the correct structural assumptions are formulated in terms of a relative measure of regularity of φ with respect to the operator T, a condition called "source condition" (see also Florens, Johannes, and Van Bellegem (2005), Johannes, Van Bellegem, and Vanhems (2007)).

Below we propose to write structural assumptions in term of *Hilbert scales*, a useful tool that unifies and extends these approaches, and allows to derive rates of convergence of estimates of φ in various norms. First we define that notion which is new in econometrics, before showing its connection to the usual smoothness structural assumptions considered in the econometric literature.

2.1 Definition

A Hilbert scale is a specific sequence of Hilbert spaces constructed from a separable Hilbert space H in the following way. Let $L^2_\mu(\Omega)$ be the Hilbert space of square integrable functions defined on a σ-finite measurable space $(\Omega, \mathfrak{B}, \mu)$ endowed with inner product $\langle f, g\rangle_{L^2_\mu(\Omega)} = \int f g d\mu$. Let $b(\cdot)$ be an unbounded measurable function defined on $(\Omega, \mathfrak{B}, \mu)$ that is finite μ-a.e. and with values in $[c_b, \infty)$ with $c_b > 0$. For any $\nu \in \mathbb{R}$ we also define the linear manifold $\mathcal{L}_\nu := \{f \in L^2_\mu(\Omega) : b^{\nu/2} f \in L^2_\mu(\Omega)\}$.

Suppose U is an unitary mapping from H to $L^2_\mu(\Omega)$. Via the unitary equivalence the manifold $\mathcal{L}_\nu$ defines a subspace H_ν of H, that is $H_\nu = \{h \in H : Uh \in \mathcal{L}_\nu\}$. It can be proven from the closed graph theorem that H_ν equipped with the inner product $\langle \phi, \psi\rangle_{H_\nu} := \langle b^{\nu/2} U\phi, b^{\nu/2} U\psi\rangle_{L^2_\mu(\Omega)}$ is a Hilbert space. Specific examples are given in the next subsections.

The two following properties are straightforward:

(i) For any $\nu, \sigma \in \mathbb{R}$ such that $\nu < \sigma$, the space H_σ is dense in H_ν

(ii) If $-\infty < q < r < s < +\infty$ and $x \in H_s$, then we have:

$$\|x\|_{H_r}^{s-q} \leq \|x\|_{H_q}^{s-r} \leq \|x\|_{H_s}^{r-q}.$$

By definition, the family $(H_\nu)_{\nu \in \mathbb{R}}$ is called a *Hilbert Scale on* H. For a complete theory on Hilbert scales we refer e.g. to Krein and Petunin (1966).

We also recall the fundamental result in functional analysis given by the spectral theorem. Define the operator B such that $UBU^{-1}f = b \cdot f$ for all $f \in \mathcal{L}_1$. Then the domain $D(B)$ of B is H_1 and the spectral theorem shows that B is a densely defined, unbounded, strictly positive, self-adjoint operator (Halmos (1963)). Analogously, any power B^ν, $\nu \in \mathbb{R}$, of the operator B is defined by: $UB^\nu U^{-1} f = b^\nu \cdot f$ for all $f \in \mathcal{L}_\nu$ and satisfies $H_\nu = D(B^\nu) = D((B^\nu)^*)$. Moreover, B fulfills $\langle B^\nu \phi, \tilde{\phi} \rangle = \langle \phi, B^\nu \tilde{\phi} \rangle$ and $\langle B^\nu \phi, \phi \rangle \geq c_b^\nu \|\phi\|^2$ for all $\phi, \tilde{\phi} \in H_\nu$. The inner product in H_ν also satisfies $\langle \phi, \tilde{\phi} \rangle_{H_\nu} = \langle B^{\nu/2} \phi, B^{\nu/2} \tilde{\phi} \rangle$ and for the norm holds $\|\phi\|_{H_\nu} = \|B^{\nu/2}\phi\|$. In the next section, we give examples of Hilbert scales and their link to usual structural assumptions in econometrics.

2.2 Connection to standard smoothness classes

Very often in nonparametric econometrics, rates of convergence are derived under the assumption that the solution belongs to some smoothness class of function, such as Sobolev classes. The following examples show the connection between Sobolev spaces and Hilbert scales.

(i) Start with the Sobolev spaces $(\mathcal{W}_\nu(\mathbb{R}))_\nu$ in $L^2(\mathbb{R})$, in which case we set $L^2_\mu(\Omega) = L^2(\mathbb{R})$ and $b(t) := (1 + t^2)$. So we consider the linear manifold $\mathcal{L}_\nu := \{f \in L^2(\mathbb{R}) : \int (1 + t^2)^\nu |f(t)|^2 dt < \infty\}$. Let $\mathcal{F}$ denote the Fourier-Plancherel unitary mapping defined from $L^2(\mathbb{R})$ into itself, in which case we suppose $H = L^2(\mathbb{R})$. Then H_ν corresponds to the space $\mathcal{W}_\nu(\mathbb{R}) := \{f \in L^2(\mathbb{R}) : \int (1 + t^2)^\nu |\mathcal{F} f(t)|^2 dt < \infty\}$. It is well-known, that for integer m we have $\mathcal{W}_m(\mathbb{R}) = \{f \in L^2(\mathbb{R}) : f^{(m)} \in L^2(\mathbb{R})\}$, where $f^{(m)}$ denotes the m-th weak-derivative of f. Moreover, the norm $\|f\|_m$ is equivalent to the norm $\|f\| + \|f^{(m)}\|$. The operator B is such that: $B^2 : \mathcal{W}_2(\mathbb{R}) \to L^2(\mathbb{R})$ and $B^2 f = f - f''$.

(ii) As a second example, consider the Sobolev spaces of periodic functions with boundary conditions $(\mathcal{W}_\nu[0,1])_\nu$ in $L^2[0,1]$, in which case we suppose, that $L^2_\mu(\Omega) = L^2(\mathbb{N})$ where μ is now the counting measure on $\mathbb{N}$, i.e. the Hilbert space of square summable sequences endowed with the inner product $\langle x, y \rangle_{L^2(\mathbb{N})} = \sum_n x_n y_n$. Thereby, considering an unbounded sequence $(b_j)_j$ we define the linear manifold $\mathcal{L}_\nu := \{x \in L^2(\mathbb{N}) : \sum_j b_j^\nu x_j^2 < \infty\}$. Let U denote an unitary mapping defined from $L^2[0,1]$ into $L^2(\mathbb{N})$, then set $\phi_j = U^{-1} e_j$ where e_j is one at the j-th position and is zero otherwise. Note that $\{\phi_j\}_j$ forms an orthonormal basis in $L^2[0,1]$. Thereby, we obtain $H_\nu = \{f \in L^2[0,1] : \sum_j b_j^\nu \langle f, \phi_j \rangle^2 < \infty\}$ which is the domain of the operator $B^{\nu/2} f := \sum_j b_j^{\nu/2} \langle f, \phi_j \rangle \phi_j$. In general, elements of H_ν cannot be characterized through differentiability conditions as it was the case in (i). A simple example is given if $\{\phi_j\}_j$ is the Haar-wavelet basis in $L^2[0,1]$. However, if $b_{2j} = b_{2j+1} = (2j)^2$, $j \in \mathbb{N}$ and $\{\phi_1 \equiv 1, \phi_{2k}(x) = \sqrt{2}\cos(2\pi kx), \phi_{2k+1}(x) = \sqrt{2}\sin(2\pi kx), k \in \mathbb{N}\}$ is the trigonometric basis, then H_ν is the Sobolev space $\mathcal{W}_\nu[0,1]$ of periodic functions. That is, for integer m we have $\mathcal{W}_m[0,1] = \{f \in L^2[0,1] : f^{(m)} \in L^2[0,1], f^{(j)}(0) = f^{(j)}(1), j = 0,\dots,(m-1)\}$ and moreover it holds $\|f\|_m = (\pi)^{-m}\|f^{(m)}\|$. For more details about Sobolev spaces $\mathcal{W}_\nu[0,1]$ we refer to Adams (1975).

(iii) As a last example consider the setting setting in (i) and (ii) but now with $b(t) = \exp(|t|^\gamma)$ or $b_j = \exp(|j|^\gamma)$ for some $\gamma > 0$. Then H_ν corresponds to the space $\mathcal{W}_\nu^\gamma$ that contains for all $\nu > 0$ only functions that are infinitely differentiable. Moreover, if $\gamma \geq 1$, then all elements of H_ν are analytical functions (cf. Kawata (1972)).

2.3 Connection to the source conditions

Consider the following relation: $T\varphi = r$ where $T : H \to G$ is a linear bounded operator. We formulate our regularity conditions on the function φ in terms of the Hilbert scale $(H_\nu, \|\cdot\|_\nu)_\nu$, that is we assume:

$$\varphi \in H_p \quad \text{for some } p > 0. \tag{2.1}$$

That assumption is however not sufficient in order to derive the rate of convergence of the risk associated to an estimator of φ in the norm $\|\cdot\|_s$ of the Hilbert space H_s $(0 \leq s \leq p)$. In order to derive this rate, we need an additional assumption that makes a link between the operator T and the Hilbert scale. Suppose the Hilbert scale is generated by an operator B and define an index function $\kappa : \mathbb{R}^+ \to \mathbb{R}^+$, which is assumed to be continuous and strictly increasing with $\kappa(0+) = 0$. Then we say that *the operator T is adapted to the Hilbert scale generated by the operator B* if there exists two constants c and C such that the inequalities

$$c\|\phi\|_{s-p} \leq \|\kappa^{1/2}(B^{-s/2}T^*TB^{-s/2})\phi\| \leq C\|\phi\|_{s-p}, \text{ for all } \phi \in H \tag{2.2}$$

hold true. The necessity of that type of structural assumption has been discussed in the context of nonparametric regression with instrumental variables in Florens, Johannes, and Van Bellegem (2007). However, both assumptions (2.1) and (2.2) imply together that the general source condition $B^{s/2}\varphi = \kappa^{1/2}(B^{-s/2}T^*TB^{-s/2})\psi$ holds for some $\psi \in H$.

Finitely smoothing case

This is the usual case studied in numerical analysis and also econometrics. It is assumed that the operator T satisfies

$$c\|\phi\|_{-a} \leq \|T\phi\| \leq C\|\phi\|_{-a}, \tag{2.3}$$

roughly speaking the operator T is a-times smoothing. The index function in this situation is $\kappa(t) = t^{(p-s)/(a+s)}$. However, if the Hilbert scale is generated by the operator $B = (T^*T)^{-1}$

then the condition (2.3) holds with $a = 1$. Moreover, in the case $s = 0$, we recover the polynomial source condition considered in Darolles, Florens, and Renault (2002) or Florens, Johannes, and Van Bellegem (2005). To be more precise, in this situation, the condition (2.1) is equivalent to the source condition $\|(T^*T)^{-p/2}\varphi\| < \infty$ and the condition (2.2) is satisfied with $\kappa(t) = t^p$. On the other hand, if we consider the Hilbert Scale of Sobolev spaces defined in Section 2.2 (ii), then we cover with $s = 0$ the setting of Hall and Horowitz (2005).

Infinitely smoothing case

This case has only been recently studied in econometrics. There it is assumed that the operator T satisfies

$$c\|\phi\|_{-a} \leq \||\log(T^*T)|^{-1/2}\phi\| \leq C\|\phi\|_{-a}, \tag{2.4}$$

roughly speaking the operator T is infinitely smoothing. The index function in this situation is $\kappa(t) = |\log(t)|^{-(p-s)/a}$. However, if the Hilbert scale is generated by the operator $B = (T^*T)^{-1}$ then the condition (2.4) holds with $a = 1$. Moreover, in the case $s = 0$, we recover the logarithmic source condition considered in Johannes, Van Bellegem, and Vanhems (2007). To be more precise, in this situation, the condition (2.1) is equivalent to the source condition $\|\,|\log(T^*T)|^{p/2}\varphi\| < \infty$ and the condition (2.2) is satisfied with $\kappa(t) = |\log(t)|^{-p}$.

Reduced form given a common unitary operator

Observe that another application of the spectral theorem allows to "diagonalize" the operator T^*T. Indeed, since T^*T is a bounded, linear, self adjoint, strictly positive operator, there exists a unitary operator $U : H \to L^2_\mu(\Omega)$ and a strictly positive function $\lambda \in L^\infty_\mu(\Omega)$ such that $UT^*TU^{-1}f = \lambda^2 f$ for all $f \in L^2_\mu(\Omega)$. Moreover, since T is injective, there exists a unitary operator $V : G \to L^2_\mu(\Omega)$, such that $VTU^{-1}f = \lambda f$ for all $f \in L^2_\mu(\Omega)$.

That last decomposition is in functional analysis also called the *singular value decomposition* of the operator T (e.g. Douglas (1966)).

Finally, consider the case where both operators T^*T and B are reduced by the same unitary operator, that is $UT^*TU^{-1}f = \lambda^2 f$ and $UBU^{-1}f = bf$, where $U : H \to L^2_\mu(\Omega)$ is unitary. Then, the condition of adaptation (2.2) can be rewritten as

$$c\|b^{(s-p)/2}f\|_{L^2_\mu(\Omega)} \leq \|\kappa(\lambda^2/b^s)^{1/2}f\|_{L^2_\mu(\Omega)} \leq C\|b^{(s-p)/2}f\|_{L^2_\mu(\Omega)}, \text{ for all } f \in L^2_\mu(\Omega)$$

Again, this last assumption with (2.1) imply together that the general source condition $b^{s/2}U\varphi = \kappa^{1/2}(\lambda^2/b^s)f$ holds for some $f \in L^2_\mu(\Omega)$.

3 Main results

Consider the solution φ of the generic problem $T\varphi = r$. Our aim is to construct an estimator $\hat{\varphi}$ and to measure its performance through the risk function in the norm of the Hilbert space

H_s, that is: $\mathbb{E}\|\hat{\varphi} - \varphi\|_s^2$.

The common method of regularization is the Tikhonov regularization (see Darolles, Florens, and Renault (2002), Hall and Horowitz (2005),) and its generalization to Hilbert scale (see Florens, Johannes, and Van Bellegem (2007)), which defines an estimator of φ as solution of the minimization problem $\min_{\phi \in H_s} \|\hat{T}\phi - \hat{r}\|^2 + \alpha\|\phi\|_s^2$, where $\alpha > 0$ is again a regularization parameter. Even if Tikhonov regularization is often studied, it has some disadvantages (see Florens, Johannes, and Van Bellegem (2007) for a discussion). Therefore, several other regularization methods are proposed, such as Landweber iteration scheme or ν-methods to name but a few (e.g. Engl, Hanke, and Neubauer (2000)).

All these methods can be unified by considering the general regularization scheme in Hilbert scales

$$\widehat{\varphi}_s = B^{-s/2} g_\alpha(B^{-s/2}\widehat{T}^\star\widehat{T}B^{-s/2})B^{-s/2}\widehat{T}^\star\widehat{r}. \tag{3.1}$$

Here, the regularization scheme $g_\alpha : (0, c] \to \mathbb{R}$ is a piecewise continuous function with the property that $\lim_{\alpha\to 0+} g_\alpha(t) = 1/t$. Different regularization methods are characterized by different functions g_α (cf Tautenhahn (1996)).

Example 3.1. **(i)** The classical Tikhonov regularization corresponds to $g_\alpha(t) = 1/(t + \alpha)$.

(ii) The Tikhonov regularization of order m is a generalization of the previous method with $g_\alpha(t) = (1 - (\alpha/(t + \alpha))^m)/t$ and $m \geq 1$. The regularized estimator $\widehat{\varphi}_s := \widehat{\varphi}_{s,m}$ can be obtained by solving the m linear operator equations

$$(\widehat{T}^\star\widehat{T} + \alpha B^s)\widehat{\varphi}_{s,j} = \widehat{T}^\star\widehat{r} + \alpha B^s\widehat{\varphi}_{s,j-1}, \quad j = 1, \dots, m, \quad \widehat{\varphi}_{s,0} = 0.$$

(iii) The spectral cut-off considers $g_\alpha(t) = 1/t$ for $t \geq \alpha$, and $g_\alpha(t) = 1/\alpha$ for $t < \alpha$ and is used for instance in Cavalier and Hengartner (2005).

(iv) The Landweber iteration procedure takes $g_\alpha(t) = (1 - (1 - t)^{1/\alpha})/t$ and is studied in the context of nonparametric instrumental regression in Johannes, Van Bellegem, and Vanhems (2007).

In what follows, we consider two cases. In the first restrictive one, we assume that both operators T^*T and B are reduced by the same unitary operator. In the second case, we do not suppose this restriction.

3.1 Risk bound when the reduced form is known

In this section, we consider the case where T^*T and B have the same unitary operator, so that $UT^*TU^{-1}f = \lambda^2 f$ and $UBU^{-1}f = bf$, where $U : H \to L^2_\mu(\Omega)$ is unitary. Moreover, we suppose that $V : G \to L^2_\mu(\Omega)$ with $VTU^{-1}f = \lambda f$ is known. In that situation the estimation of T^*T is reduced to the estimation of the function λ.

This condition is natural in the deconvolution problem (see section 4.1 below). However, despite this assumption is rather restrictive, it is sometimes used in the literature, cfr Mair and Ruymgaart (1996), Hall and Horowitz (2005), or Cavalier and Hengartner (2005).

Let us now consider the main problem of recovering $\varphi \in H_p$ and define for $0 \le s \le p$ the estimator $\hat{\varphi}_s$ by

$$U\hat{\varphi}_s := \frac{V\hat{r}}{\hat{\lambda}} \mathbb{1}\{\hat{\lambda}^2 \ge \alpha b^s\}. \tag{3.2}$$

This definition of $U\hat{\varphi}_s$ corresponds to the particular case of regularization called *spectral cut-off* (cf. Example 3.1(iii)). In (3.2), the parameter α plays the role of a smoothing parameter and is supposed to tend to zero as the sample size growths.

Under this setting, upper bound on the risk function depends among others on the convergence of $\hat{r}$ to r and on the convergence of $\hat{\lambda}$ to λ.

Theorem 3.1. *Suppose that $\varphi \in H_p$ for some $p > 0$ and that for some $0 \le s \le p$, there exists an index function κ such that the operator T satisfies (2.2). Assume that on an interval $(0, c^2]$, the inverse function Φ of κ is convex. Suppose that $\mathbb{E}\|\hat{r} - r\|^2 \le \delta$ and that $\sup_{\omega \in \Omega} \mathbb{E}|\hat{\lambda}(\omega) - \lambda(\omega)|^4 \le \gamma^2$. Consider the estimator $\hat{\varphi}$ defined in (3.2) using a threshold $\alpha = d\cdot(\delta+\gamma)/\zeta[d\cdot(\delta+\gamma)]$ for some $d > 0$, where ζ is the inverse function of $\zeta^{-1}(t) := t\Phi(t)$. Then we get the risk bound*

$$\mathbb{E}\|\hat{\varphi} - \varphi\|_s^2 \le C\zeta(\delta + \gamma),$$

where C is a stricly positive constant.

Remark 3.1. **(i)** It is shown in Nair, Pereverzev, and Tautenhahn (2005), Theorem 2.2, that if in (2.2) we have equalities, then the function ζ providing the bound in the last theorem and the modulus of continuity of the inverse of T are of the same order. It is well known that the modulus of continuity essentially describes the worst case error in estimating the solution φ varying in H_p under the condition that (2.2) is satisfied. Thereby we may argue that given some estimators $\hat{r}$ and $\hat{\lambda}$ satisfying the assumptions of Theorem 3.1, the derived bound cannot be improved. This suggests that our bound is optimal provided the estimators $\hat{r}$ and $\hat{T}$ are optimal over the set of possible r and T.

(ii) Note that the conditions (2.2) in general do not imply an optimal rate of convergence of T. To be more precise, we can in general not derive an optimal rate for $\hat{T}$ on the set of all operators satisfying (2.2). Moreover, the same holds true for $r = T\varphi$. That is, in general we are not able to provide an optimal rate of convergence of an estimator of r on the set $\{r = T\varphi : \varphi$ and T satisfies (2.1) and (2.2), respectively $\}$.

However, if we have optimal rates for r and T, we may reach the optimal rate for φ but as we will see later, even if we do not estimate optimal r and T, we may obtain an optimal rate for φ.

(iii) (The finitely smoothing case). Assume that $c\|\phi\|_{-a} \leq \|T\phi\| \leq C\|\phi\|_{-a}$. The index function in this situation is $\kappa(t) = t^{(p-s)/(a+s)}$. Then, its inverse function $\Phi(t) = t^{(a+s)/(p-s)}$ is convex only if $p - s \leq a + s$. This condition is also given in the seminal paper on Hilbert scale by Natterer (1984). Then we have $\zeta(t) = t^{(p-s)/(a+p)}$. Thus we get $\mathbb{E}\|\hat{\varphi} - \varphi\|_s^2 \leq C(\delta + \gamma)^{(p-s)/(a+p)}$. However, under slightly stronger conditions, we can relax the restriction $p - s \leq a + s$ (see Theorem 3.2 below).

(iv) (The infinitely smoothing case). Assume that $c\|\phi\|_{-a} \leq \||\log(T^*T)|^{-1/2}\phi\| \leq C\|\phi\|_{-a}$. The index function in this situation is $\kappa(t) = |\log(t)|^{-(p-s)/a}$. Then, its inverse function $\Phi(t) = e^{-s^{-1/\beta}}$ with $\beta = (p-s)/a > 0$ is convex on the interval $(0, c]$ where $c = (1+\beta)^{-\beta}$. Moreover, it was shown by Mair (1994) that $\zeta(t) = |\log(t)|^{-(p-s)/a}(1+o(1))$ if t tends to 0. Thus we get $\mathbb{E}\|\hat{\varphi} - \varphi\|_s^2 \leq C|\log(\delta + \gamma)|^{-(p-s)/a}$.

(v) In most econometric problems, δ has a nonparmetric rate whereas γ has a parametric rate of convergence. Therefore, the estimator of T is negligible compared to the estimation of r in the rate of covergence given by Theorem 3.1.

(vi) If we impose that the operator T is adapted to a scale of Sobolev spaces (cf. Section 2.2), as it is assumed in a lot of econometric studies, then the above theorem shows that this assumption does not only allow to derive the rate of convergence in the L^2 norm, but also the rate of convergence in the Sobolev norm. In other words, that assumption is also sufficient to derive the rate of convergence of *derivatives* of φ without any additional assumption.

We can improve the previous result in the finitely smoothing case.

Theorem 3.2. *Suppose that $\varphi \in H_p$ for some $p > 0$ and that the operator T is finitely smoothing, that is (2.3) is satisfied for some $a > 0$. Suppose that $\mathbb{E}\|\hat{r} - r\|^2 \leq \delta$ and that $\sup_{\omega\in\Omega} \mathbb{E}|\hat{\lambda}^2(\omega) - \lambda^2(\omega)|^{2(p+a)/(a+s)} \leq \gamma^{(p+a)/(a+s)}$. Consider the estimator $\hat{\varphi}$ defined in (3.2) using a threshold $\alpha = d \cdot (\delta^{(s+a)/(p+a)} + \gamma)$ for some $d > 0$. Then we get $\mathbb{E}\|\hat{\varphi} - \varphi\|_s^2 \leq C(\delta^{(p-s)/(p+a)} + \gamma^{1\wedge(p-s)/(a+s)})$ with $C > 0$.*

In contrast to Remark 3.1(iii) above, that theorem shows that it is possible to derive an upper bound for the risk without the restriction $p - s \leq a + s$. Observe that in the common case where δ has a nonparametric rate of convergence and γ has a parametric rate of convergence, then the upper bounds are the same in the two theorems.

3.2 Risk bound when the reduced form is unknown

In this section, we do not suppose that both operators T^*T and B are reduced by the same unitary operator. This case occurs for example in nonparametric instrumental regression. For the study of the general regularization method the following additional assumption is required, which is analogous to a corresponding assumption in Nair, Pereverzev, and Tautenhahn (2005).

Assumption 3.1. *There exist positive constants c and d such that*

(i) $\sup_{t>0} t^{1/2}|g_\alpha(t)| \le c/\sqrt{\alpha}, \quad \sup_{t>0}|tg_\alpha(t)| \le 1,$

(ii) $\sup_{t>0} t|1-tg_\alpha(t)| \le d\alpha, \quad \sup_{t>0}|1-tg_\alpha(t)| \le 1$

All methods given in examples (3.1) satisfy these assumptions. Detailed discussion can be found in Tautenhahn (1996).

Under this setting, the performance of the risk function will depend on the convergence of $\hat{r}$ to r and on the convergence of $\hat{T}$ to T.

Theorem 3.3. *Suppose that $\varphi \in H_p$ for some $p > 0$ and that for some $0 \le s \le p$, there exists an index function κ such that the operator T satisfies (2.2). Assume that on an interval $(0, c^2]$, the inverse function Φ of κ is convex. Suppose that $\mathbb{E}\|\hat{r}-r\|^2 \le \delta$ and that $\mathbb{E}\|\hat{T}-T\|^4 \le \gamma^2$. Consider the estimator $\hat{\varphi}$ defined in (3.1) using a regularization parameter $\alpha = d\cdot(\delta+\gamma)/\zeta[d\cdot(\delta+\gamma)]$ for some $d > 0$, where ζ is the inverse function of $\zeta^{-1}(t) := t\Phi(t)$. Then we get the risk bound*

$$\mathbb{E}\|\hat{\varphi}-\varphi\|_s^2 \le C\zeta(\delta+\gamma),$$

where C is a strictly positive constant.

Note that we obtain the same bound as in the previous section. The main difference is the typical rate of convergence of γ, which is here nonparametric.

We can also improve the previous result in the finitely smoothing case. To do so, we need an additional assumption on the regularization scheme.

Assumption 3.2. *There exist constants $c_\beta > 0$ and $\beta_0 \ge 1$ such that for $0 \le \beta \le \beta_0$, $\sup_{0\le t\le c} t^\beta|1-tg_\alpha(t)| \le c_\beta\alpha^\beta$.*

The parameter β_0 is called qualification of the regularization method. The examples 3.1 (iii) and (iv) satisfy the previous assumption with $\beta_0 = \infty$ while in examples (i) and (ii), we have $\beta_0 = 1$ and $\beta_0 = m$, respectively. A finite qualification as in the case of Tikhonov regularization, example (i), leads to saturation effect, studied in the context of nonparametric instrumental regression in Florens, Johannes, and Van Bellegem (2007). Saturation does not appear with other regularization methods, such as the Landweber iterative regularization.

Theorem 3.4. *Suppose that $\varphi \in H_p$ for some $p > 0$ and that the operator T is finitely smoothing, that is (2.3) is satisfied for some $a > 0$. Suppose that $\mathbb{E}\|\hat{r}-r\|^2 \le \delta$ and that $\mathbb{E}\|\hat{T}-T\|^2 \le \gamma$ and if $(p-s)/(a+s) > 1$ then $\mathbb{E}\|\hat{T}-T\|^{2(p-s)/(a+s)} \le \gamma$. Consider the estimator $\hat{\varphi}$ defined in (3.1) using a regularization parameter $\alpha = d\cdot(\delta^{(s+a)/(p+a)} + \gamma^{(s+a)/(p+a)})$ for some $d > 0$. Then, if $(p-s)/(a+s) \le \beta_0$ we get the risk bound*

$$\mathbb{E}\|\hat{\varphi}-\varphi\|_s^2 \le C(\delta^{(p-s)/(p+a)} + \gamma^{(p-s)/(a+p)}),$$

where C is a stricly positive constant.

Note, that the bound is worst than in the case of known eigenfunctions, i.e., the term in γ. However, if we suppose that $\widehat{T}^\star T$ and $\widehat{T}^\star \widehat{T}$ commute, we can give a tigther bound of the variance term, and hence we have the same bound as if the eigenfunctions were known.

Note, that we have the qualification β_0 in the result, for $(p-s)/(a+s) > \beta_0$, we only get the bound $\mathbb{E}\|\hat{\varphi} - \varphi\|_s^2 \leq C(\delta^{\beta_0/(\beta_0+1)} + \gamma^{\beta_0/(\beta_0+1)})$, i.e. we have a saturation effect.

4 Examples

In this section, we illustrate the previous results on two important examples in economic studies. The first example is the estimation of a probability density function when observations are contaminated by noise (deconvolution problem). The second example is the estimation of a nonparametric regression function using instrumental variables.

4.1 Deconvolution with estimated error density

Suppose X and ε are independent random variables with unknown density functions f_X and f_ε, respectively. The objective is to estimate nonparametrically the density function f_X based on a sample of $Y = X + \varepsilon$. In this setting the density f_Y of Y is the convolution of the density of interest, f_X, and the density f_ε of the additive noise, i.e., $f_Y = f_X \star f_\varepsilon$ (see (1.3)).

Suppose we observe an iid sample $Y_1, \dots, Y_n$ from f_Y. If the error density f_ε is known, a large literature in econometrics and statistics considers the estimation of the density f_X. The most popular approaches consist in estimating first the density function f_Y by kernel and to solve equation (1.2) in the Fourier domain (see, for example, Carroll and Hall (1988), Fan (1991, 1992)). Popular alternative methods are based on spline (Mendelsohn and Rice (1982), Koo and Park (1996)) or wavelet decomposition (Pensky and Vidakovic (1999), Fan and Koo (2002), Bigot and Van Bellegem (2006)). Generalization to panel data has also been considered (Horowitz and Markatou (1996), Hall and Yao (2003), Neumann (2006), Bonhomme and Robin (2006) or Cazals, Florens, and Simar (2007)).

However, in several applications the noise density f_ε may be unknown. In this case without any additional information the density f_X can not be recovered from the density of f_Y through (1.2), i.e., the density f_X is not identified assuming only a sample $Y_1, \dots, Y_n$ from f_Y. This section deals with the estimation of a deconvolution density f_X when f_ε is unknown, but draws of the error distribution are observed. To be precise, we observe the sample $Y_1, \dots, Y_n$ from f_Y and, additionally, a sample $\varepsilon_1, \dots, \varepsilon_m$ from f_ε. The estimation of f_X in such a situation is considered Neumann (1997) and Johannes (2007).

The density f_X is solution of the inverse problem $f_Y = C_{f_\varepsilon} f_X$, where $C_{f_\varepsilon} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ is the convolution operator given by $g \mapsto C_{f_\varepsilon} g := g \star f_\varepsilon$. Note, that f_Y and C_{f_ε} have to be estimated. However, it is well-known that if $f_\varepsilon \in L^2(\mathbb{R})$ then the convolution operator C_{f_ε} has a spectral decomposition given by the unitary Fourier transform $\mathcal{F}$ and the spectral density $\lambda = \sqrt{2\pi} \cdot \mathcal{F} f_\varepsilon$, that is $\mathcal{F} C_{f_\varepsilon} \mathcal{F}^{-1} g = \sqrt{2\pi} \cdot \mathcal{F} f_\varepsilon \cdot g$ for all $g \in L^2(\mathbb{R})$. Thereby, even if the error density f_ε is unknown, the unitary operator in the spectral decomposition of the convolution operator C_{f_ε} is known and only the spectral density $\lambda = \sqrt{2\pi} \cdot \mathcal{F} f_\varepsilon$ has to be estimated.

Using the iid sample $Y_1, \ldots, Y_n$ from f_Y we construct an estimator $\widehat{f_Y}$ of f_Y. We stick to a nonparametric kernel estimation approach, but any other density estimation procedure can of course be used at this stage. The kernel estimator of f_Y is defined by

$$\widehat{f_Y}(y) := \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{Y_j - y}{h}\right), \quad y \in \mathbb{R}, \tag{4.1}$$

where $h > 0$ is a bandwidth and K a kernel function. As usual in the context of nonparametric kernel estimation the bandwidth h has to tend to zero as the sample size n increases. In order to derive a rate of convergence we follow Parzen (1962) and consider for each $\tau > 0$ the class of kernel functions defined by

$$\mathcal{K}_\tau := \{K \in L^1(\mathbb{R}) \cap L^2(\mathbb{R}) : \lim_{t \to 0} \frac{|1 - \sqrt{2\pi}[\mathcal{F}K](t)|}{|t|^\tau} = c_\tau < \infty\}. \tag{4.2}$$

In addition based on the i.i.d. sample $\varepsilon_1, \ldots, \varepsilon_m$ from f_ε we estimate the Fourier transform $\mathcal{F} f_\varepsilon$ using its empirical counterpart, that is

$$[\widehat{\mathcal{F} f_\varepsilon}](t) := \frac{1}{m \cdot \sqrt{2\pi}} \sum_{j=1}^{m} e^{-it\varepsilon_j}, \quad t \in \mathbb{R}. \tag{4.3}$$

Following the general approach presented in Section 3.1 we suppose that the level of regularity of the deconvolution density f_X is described using a Hilbert scale $(H_s)_s$ given by the unitary operator $\mathcal{F}$ and an unbounded function $b(\cdot)$. Then we consider for $s \geq 0$ the estimator $\widehat{f_X}_s$ of f_X defined by

$$\mathcal{F}\widehat{f_X}_s(t) := \frac{\mathcal{F}\widehat{f_Y}(t) \cdot \overline{\widehat{\mathcal{F} f_\varepsilon}(t)}}{\sqrt{2\pi} \cdot |\widehat{\mathcal{F} f_\varepsilon}(t)|^2} \cdot \mathbb{1}\{|\widehat{\mathcal{F} f_\varepsilon}(t)|^2 \geq \alpha \cdot b^s(t)\}, \quad t \in \mathbb{R}, \tag{4.4}$$

where $\alpha > 0$ is a threshold decreasing to zero as the samples sizes n and m increase.

Proposition 4.1. *Suppose that $f_X \in H_p$ for some $p > 0$ and that for some $0 \leq s \leq p$, there exists an index function κ such that the operator C_{f_ε} satisfies (2.2). Assume that on an interval $(0, c^2]$, the inverse function Φ of κ is convex. Suppose that the kernel estimator of f_Y defined in (4.1) satisfies $\mathbb{E}\|\hat{f}_Y - f_Y\|^2 \leq \delta$. Consider the estimator $\widehat{f_X}_s$ defined in (4.4) using a threshold $\alpha = d \cdot (\delta + 1/m)/\zeta[d \cdot (\delta + 1/m)]$ for some $d > 0$, where ζ is the inverse function of $\zeta^{-1}(t) := t\Phi(t)$. Then we get $\mathbb{E}\|\widehat{f_X}_s - f_X\|_s^2 \leq C \cdot \zeta(\delta + 1/m)$ with $C > 0$.*

In order to illustrate the last result let us consider a more specific situation. Suppose we describe the level of smoothness of the deconvolution density f_X using the Hilbert scale $(\mathcal{W}_s(\mathbb{R}))_s$ of Sobolev spaces in $L^2(\mathbb{R})$ (see Section 2.2 (i)), i.e., $b(t) := (1 + t^2)$. Then we measure the performance of the estimator $\widehat{f_X}_s$ given in (4.4) by the $\mathcal{W}_s$-risk, that is $\mathbb{E}\|\widehat{f_X}_s - f_X\|_s^2$, provided $f_X \in \mathcal{W}_p(\mathbb{R})$ for some $p \geq s$. Note that for an integer k the Sobolev norm $\|g\|_k$ is equivalent to $\|g\| + \|g^{(k)}\|$, where $g^{(k)}$ denotes the k-th weak derivative of g. Thereby, the $\mathcal{W}_k$-risk reflects the performance of $\widehat{f_X}_k$ and $\widehat{f_X}_k^{(k)}$ as estimator of f_X and $f_X^{(k)}$, respectively. Moreover, let us suppose the convolution operator C_{f_ε} is finitely smoothing, that is C_{f_ε} satisfies the condition (2.3) for some $a > 0$. Then, after some algebra we see that this condition is equivalent to the assumption of an ordinary smooth error density f_ε, i.e., $\mathcal{F}f_\varepsilon(t) \sim (1+t^2)^{-a/2}$.

Corollary 4.2. *Suppose that $f_X \in \mathcal{W}_p(\mathbb{R})$ with $p > 0$ and that f_ε is ordinary smooth, i.e., (2.3) is satisfied for some $a > 0$. Let the kernel estimator of f_Y defined in (4.1) be constructed using a kernel $K \in \mathcal{K}_{p+a}$ (see (4.2)) and a bandwidth $h = cn^{-1/(2(p+a)+1)}$ for some $c > 0$. Consider for $0 \le s \le p$ the regularized estimator $\widehat{f_X}_s$ defined in (4.4) using a threshold $\alpha = c(n^{-2(s+a)/(2(p+a)+1)} \vee m^{-1})$ for some $c > 0$. Then we have*
$\mathbb{E}\|\widehat{f_X}_s - f_X\|_s^2 = O\Big(n^{-2(p-s)/(2(p+a)+1)} + m^{-(1\wedge(p-s)/(a+s))}\Big)$.

Remark 4.1. (i) The last result provides the minimax optimal rate of convergence over the class of all deconvolution densities f_X belonging to $\mathcal{W}_p(\mathbb{R})$ and error densities f_ε satisfying (2.3) (see Neumann (1997)). Note that, if the sample size m increases as the sample size n increases, such that $m = c \cdot n^{2[(p-s)\vee(a+s)]/(2(p+a)+1)}$ for some $c > 0$, then the last result simplifies and we obtain the order $O(n^{-2(p-s)/(2(p+a)+1)})$. Thereby we have the optimal order of the $\mathcal{W}_s$-risk as in the case of a known error density and hence the estimation of the error density is negligible.

(ii) The condition (2.3) implies, that f_X belongs to $\mathcal{W}_p(\mathbb{R})$ if and only if f_Y lies in $\mathcal{W}_{p+a}(\mathbb{R})$. Thereby, considering the assumptions of the corollary we see, that we construct an order optimal kernel estimator of f_Y. Moreover, if we use an estimator of f_Y which does not lead to an order optimal rate of convergence, then the estimator of f_X would not reach the minimax optimal rate of convergence. Hence, in this situation the optimal estimation of f_Y is necessary to obtain an optimal estimator of f_X. □

The last corollary can only be applied if the error density is ordinary smooth. In the case of a supersmooth error density, that is $\mathcal{F}f_\varepsilon(t) \sim \exp(-|t|^{2a}/2)$, the condition (2.3) cannot be satisfied. Nevertheless, the operator C_{f_ε} is then infinitely smoothing, that is condition (2.4) holds true.

Corollary 4.3. *Suppose that $f_X \in \mathcal{W}_p(\mathbb{R})$ with $p > 0$ and that f_ε is supersmooth, i.e., (2.4) is satisfied for some $a > 0$. Let the kernel estimator of f_Y defined in (4.1) be constructed using a kernel $K \in \mathcal{K}_\tau$ for some $\tau > 0$ (see (4.2)) and a bandwidth $h = cn^{-1/(2\tau+1)}$ for some $c > 0$. Consider for $0 \le s \le p$ the regularized estimator $\widehat{f_X}_s$ defined in (4.4) using a threshold $\alpha = c(n^{-\tau/(2\tau+1)} \vee m^{-1/2})$ for some $c > 0$. Then we have $\mathbb{E}\|\widehat{f_X}_s - f_X\|_s^2 = O\Big(\log(n^{2\tau/(2\tau+1)} \wedge m)^{-(p-s)/a}\Big)$.*

Remark 4.2. (i) If the sample size m increases as the sample size n increases, such that $m = cn^\tau$ for some $c > 0$ and $\tau > 0$, then the last result simplifies and we obtain the minimax optimal order $O((\log n)^{-(p-s)/a})$ over the class $\mathcal{W}_p(\mathbb{R})$ given an a-priori known error density satisfying (2.4) (Mair and Ruymgaart (1996)). Thereby, we obtain the surprising result, that the estimation of the error density is negligible as long as the sample size m increases as some polynomial of n.

(ii) The condition (2.4) implies, that f_Y lies in $\mathcal{W}_\tau(\mathbb{R})$ for all $\tau > 0$. Thereby, the assumptions of the corollary provide an order $O(n^{-2\tau/(2\tau+1)})$ of the rate of convergence of the kernel estimator of f_Y for some arbitrary $\tau > 0$. However, the choice of τ, i.e., the order of the kernel estimator of f_Y, does not influence the order of rate of convergence of the estimator of f_X. Hence, in this situation an order optimal estimation of f_Y is not necessary to obtain an order optimal estimator of f_X. □

The last two results consider situations in which the density f_X is ordinary smooth and the error density f_ε is ordinary smooth or supersmooth, respectively. However, if f_ε is a Cauchy density, i.e., has heavy tails, and f_ε is Gaussian, then both Fourier transforms are exponential. In such a situation the last two corollaries are not appropriate, but we may consider the Hilbert scale $(\mathcal{W}_s^\nu(\mathbb{R}))_s$ of infinitely differentiable functions generated by $b(t) := \exp(|t|^\nu)$ for some $\nu > 0$ (see Section 2.2 (iii)). Then a Cauchy density belongs to $\mathcal{W}_p^1(\mathbb{R})$ for some $p > 0$ and a Gaussian density is adapted to the Hilbert scale $(\mathcal{W}_s^1(\mathbb{R}))_s$, i.e., satisfies (2.2) considering an index function $\kappa(t) = \exp(-\eta\sqrt{|\log t|})$ for some $\eta > 0$. Now we may apply Theorem 3.1 to derive the rate of convergence of the estimator $\widehat{f_X}_s$ in the stronger norm $\|f\|_s := \|b^{s/2} f\|_{L^2(\mathbb{R})}$ defined with respect to $b(t) := \exp(|t|^\nu)$. Thereby, we measure the performance of $\widehat{f_X}_s$ as estimator of f_X, and in addition of any weak-derivative $\widehat{f_X}_s^{(k)}$ as estimator of $f_X^{(k)}$.

4.2 Nonparametric instrumental regression

Consider a random vector (Z, W, Y) with unknown probability distribution (pd) P and φ the interest parameter defined by:

$$\begin{cases} Y &= \varphi(Z) + U, \\ \mathbb{E}[U|W] &= 0, \end{cases} \tag{4.5}$$

where U denotes an error term. W is an instrumental variable introduced to solve the identification problem since the usual condition $\mathbb{E}[U|Z] = 0$ is not satisfied. The system (4.5) can be rewritten:

$$\mathbb{E}[Y|W] = \mathbb{E}[\varphi(Z)|W]. \tag{4.6}$$

As recalled in the Introduction to this paper, this model is the foundation of many economic studies. Even when the pd P is known, the calculation of a solution φ from equation (4.6) is an ill-posed inverse problem. However the pd P is unknown in general and has to be estimated from an iid n-sample $(Y_1, Z_1, W_1), \dots, (Y_n, Z_n, W_n)$ of (Y, Z, W).

Therefore, two main steps are necessary in order to obtain an estimator of φ. The first step is to stabilize the equation (4.6), the second step is the solving of the stabilized version of equation (4.6), where P is replaced by its estimator. To achieve the first step rewrite the model (4.6) as $r = T\phi$, where $T : H \to \tilde{H}$ with $\phi \mapsto T\phi := \mathbb{E}[\phi(Z)|W]$ is assumed to be a compact linear operator defined between Hilbert spaces H, $\tilde{H}$ and $r := \mathbb{E}[Y|W]$ is an element of $\tilde{H}$. Note, that r and T depend on the pd P and have to be estimated. A more detailed discussion of possible definitions of the underlying Hilbert spaces, the operator T and the element r can be found in Florens, Johannes, and Van Bellegem (2005). To simplify the presentation we assume H and $\tilde{H}$ to be the Hilbert space $L^2[0,1]$ and we suppose that the operator T is compact and, hence admits a discrete singular value decomposition given by $Tf = \sum_j \lambda_j \langle f, \phi_j \rangle \psi_j$, where the eigenfunctions $\{\phi_j\}_j$ and $\{\psi_j\}_j$ form an orthonormal basis in $L^2[0,1]$ and the sequence of singular values $(\lambda_j)_j$ tends to zero.

Let us first consider the simpler but artificial situation where the set of eigenfunctions $\{\phi_j\}_j$ and $\{\psi_j\}_j$ are a-priori known. Thereby, the estimation of the conditional expectation operator T reduces solely to an estimation of the sequence of singular values $(\lambda_j)_j$. Assuming an i.i.d. sample (Y_i, Z_i, W_i), $i = 1, \dots, n$ we estimate any singular value λ_j using its empirical counterpart, that is

$$\hat{\lambda}_j := \frac{1}{n} \sum_{i=1}^{n} \psi_j(W_i) \phi_j(Z_i). \tag{4.7}$$

In addition we use an orthogonal series estimator of r given by

$$\widehat{r}(\cdot) := \sum_{j=1}^{k} \widehat{r}_j \psi_j(\cdot), \quad \text{where } \widehat{r}_j := \frac{1}{n} \sum_{i=1}^{n} Y_i \psi_j(W_i),\ j = 1, \dots, k_n. \tag{4.8}$$

The number k of estimated coefficients $\widehat{r}_j$ increases as the sample size n increases and the ratio $1/k$ plays the same role than a bandwidth in the theory of nonparametric smoothing by kernel.

Following the general approach presented in Section 3.1 we suppose now that the level of regularity of the function φ is described using a Hilbert scale $(H_s)_s$ given by the orthonormal basis $\{\phi_j\}$ and some unbounded sequence $(b_j)_j$ (compare Section 2.2 (ii)). Then for $s \geq 0$ we estimate φ using

$$\widehat{\varphi}_s(\cdot) := \sum_{j=1}^{k} \frac{\widehat{r}_j}{\widehat{\lambda}_j} \cdot \mathbb{1}\{\widehat{\lambda}_j^2 \geq \alpha \cdot b_j^s\} \cdot \phi_j(\cdot), \tag{4.9}$$

where $\alpha > 0$ is a threshold decreasing to zero as the samples sizes n increases.

Proposition 4.4. *Suppose that $\varphi \in H_p$ for some $p > 0$ and that for some $0 \leq s \leq p$, there exists an index function κ such that the operator T satisfies (2.2). Assume that on* *an interval* $(0, c^2]$, *the inverse function* Φ *of* κ *is convex. Suppose that the orthogonal series estimator of* r *defined in (4.8) satisfies* $\mathbb{E}\|\hat{r} - r\|^2 \leq \delta$ *and that* $\sup_{j\in\mathbb{N}} \mathbb{E}|\widehat{\lambda}_j - \lambda_j|^4 \leq \gamma^2$*. Consider the estimator* $\widehat{\varphi}_s$ *defined in (4.9) using a threshold* $\alpha = d\cdot(\delta+\gamma)/\zeta[d\cdot(\delta+\gamma)]$ *for some* $d > 0$*, where* ζ *is the inverse function of* $\zeta^{-1}(t) := t\Phi(t)$*. Then we get* $\mathbb{E}\|\widehat{\varphi}_s - \varphi\|_s^2 \leq C \cdot \zeta(\delta + \gamma)$ *with* $C > 0$*.*

In order to illustrate the last result let us consider a more specific situation. Suppose the eigenfunctions of the operator T are given by the trigonometric basis and the level of smoothness of the function φ is characterized using the Hilbert scale $(\mathcal{W}_s[0,1])_s$ of Sobolev spaces in $L^2[0,1]$ (see Section 2.2 (ii)). Then we measure the performance of the estimator $\widehat{\varphi}_s$ given in (4.9) by the $\mathcal{W}_s$-risk defined in the Sobolev space $\mathcal{W}_s[0,1]$. That is, for an integer m the $\mathcal{W}_m$-risk reflects the performance of the m-th weak derivative $\widehat{\varphi}_m^{(m)}$ of $\widehat{\varphi}_m$ as estimator of $\varphi^{(m)}$. Moreover, let us suppose the conditional expectation operator T is finitely smoothing, that is T satisfies the condition (2.3) for some $a > 0$. Then, after some algebra we see that this condition is equivalent to the assumption $\lambda_j \sim j^{-a}$ used for example in Hall and Horowitz (2005).

Corollary 4.5. *Suppose that* $\varphi \in \mathcal{W}_p[0,1]$ *with* $p > 0$ *and that* T *is finitely smoothing, i.e., (2.3) is satisfied for some* $a > 0$*. Let the orthogonal series estimator of* r *defined in (4.8) be constructed using* $k = cn^{-1/(2(p+a)+1)}$ *for some* $c > 0$*. Consider for* $0 \leq s \leq p$ *the regularized estimator* $\widehat{\varphi}_s$ *defined in (4.9) using a threshold* $\alpha = cn^{-2(s+a)/(2(p+a)+1)}$ *for some* $c > 0$*. Then we have*

$$\mathbb{E}\|\widehat{\varphi}_s - \varphi\|_s^2 = O\Big(n^{-2(p-s)/(2(p+a)+1)}\Big).$$

Remark 4.3. (i) For $s = 0$ the last result provides the minimax optimal rate of convergence over the class of all function φ belonging to $\mathcal{W}_p[0,1]$ and conditional expectation operator T satisfying (2.3) (see Hall and Horowitz (2005)). Moreover, for all $0 \leq s \leq p$ it is the minimax optimal rate of convergence over the class of all function φ belonging to $\mathcal{W}_p[0,1]$ given a known operator T satisfying (2.3) (see Mair and Ruymgaart (1996)), i.e., the estimation of conditional expectation operator T is negligible. Thereby, we may argue that under the assumptions of the corollary the derived rate of convergence cannot be improved.

(ii) The condition (2.3) implies, that φ belongs to $\mathcal{W}_p[0,1]$ if and only if r lies in $\mathcal{W}_{p+a}[0,1]$. Thereby, considering the assumptions of the corollary we see, that we construct an order optimal estimator of r. Moreover, if we use an estimator of r which does not lead to an order optimal rate of convergence, then the estimator of φ will not reach the minimax optimal rate of convergence. Hence, in this situation the optimal estimation of r is necessary to obtain an optimal estimator of φ. □

The last corollary can only be applied if the conditional expectation operator T is finitely smoothing, which excludes the important case where (Z, W) is normally distributed. In that case the conditional expectation is infinitely smoothing, i.e., $\lambda_j \sim \exp(-j^{2a}/2)$ for some $a > 0$ and the condition (2.3) cannot be satisfied. Nevertheless, the condition (2.4) holds true.

Corollary 4.6. *Suppose that $\varphi \in \mathcal{W}_p[0,1]$ with $p > 0$ and that T is infinitely smoothing, i.e., (2.4) is satisfied for some $a > 0$. Let the estimator of r defined in (4.8) be constructed using $k = cn^{-1/(2\tau+1)}$ for some $c > 0$ and $\tau > 0$. Consider for $0 \leq s \leq p$ the regularized estimator $\widehat{\varphi}_s$ defined in (4.9) using a threshold $\alpha = cn^{-\tau/(2\tau+1)}$ for some $c > 0$. Then we have $\mathbb{E}\|\widehat{\varphi}_s - \varphi\|_s^2 = O\Big((\log n)^{-(p-s)/a}\Big)$.*

Remark 4.4. (i) The last result provides the minimax optimal rate of convergence over the class of all function φ belonging to $\mathcal{W}_p[0,1]$ given the conditional expectation operator T satisfying (2.4) is known a-priori (see (Mair and Ruymgaart (1996)), i.e., the estimation of the conditional expectation operator T is negligible. Thereby, we may again argue that under the assumptions of the corollary the derived rate of convergence cannot be improved.

(ii) The condition (2.4) implies, that r lies in $\mathcal{W}_\tau[0,1]$ for all $\tau > 0$. Thereby, the assumptions of the corollary provide an order $O(n^{-2\tau/(2\tau+1)})$ of the rate of convergence of the orthogonal series estimator of r for some arbitrary $\tau > 0$. However, the choice of τ, i.e., the order of the estimator of r, does not influence the order of the rate of convergence of the estimator of φ. Hence, in this situation an order optimal estimation of r is not necessary to obtain an order optimal estimator of φ. □

Both previous results are derived under the assumption the eigenfunctions of the conditional expectation operator are known a-priori to be the trigonometric basis. Let us now consider the natural case where this restriction is relaxed. Then given an i.i.d. sample (Y_i, Z_i, W_i), $i = 1, \dots, n$ we construct estimators of the function r and the conditional expectation operator T by projection on some orthonormal basis $\{\phi_j\}_j$ and $\{\psi_j\}_j$ not necessarily corresponding to the system of eigenfunctions of T (see e.g. Hall and Horowitz (2005), Blundell, Chen, and Kristensen (2007) or Chen (2008)). Thereby, we estimate r using the series estimator given in (4.8). In order to derive the series estimator of T let $\boldsymbol{\phi}(\cdot) = (\phi_1(\cdot), \dots, \phi_m(\cdot))'$ and $\boldsymbol{\psi}(\cdot) = (\psi_1(\cdot), \dots, \psi_k(\cdot))'$ be the vector of the first m and k basis functions, respectively. The numbers m and k increase as the sample size n increases and, moreover $1/m$ and $1/k$ play the same role than bandwidths in a kernel estimation method. Then, the series estimator of T is

$$\widehat{T}g := \boldsymbol{\psi}'\widehat{M}\langle g, \boldsymbol{\phi}\rangle, \quad \text{where } \widehat{M} := \frac{1}{n}\sum_{i=1}^n \boldsymbol{\psi}(W_i)\boldsymbol{\phi}(Z_i)' \tag{4.10}$$

and $\langle g, \boldsymbol{\phi}\rangle$ denotes the column vector $(\langle g, \phi_1\rangle, \dots, \langle g, \phi_m\rangle)'$. The estimator of $T^\star$ is the dual of $\widehat{T}$, that is $\widehat{T}^\star h := \boldsymbol{\phi}'\widehat{M}'\langle h, \boldsymbol{\psi}\rangle$ where analogously $\langle h, \boldsymbol{\psi}\rangle := (\langle h, \psi_1\rangle, \dots, \langle h, \psi_k\rangle)'$. We also define the vector $\hat{v} := \frac{1}{n}\sum_{i=1}^n Y_i\boldsymbol{\psi}(W_i)$ such that we can rewrite the series estimator of r

given in (4.8) as $\widehat{r} = \boldsymbol{\psi}'\widehat{v}$. Following the general approach presented in the previous section, we suppose now that the level of regularity of the function φ is described using a Hilbert scale $(H_s)_s$ generated by the operator $Bf := \sum_j b_j \langle f, \phi_j \rangle \phi_j$ for some unbounded sequence $(b_j)_j$ (compare with Section 2.2 (ii)). Then for $s \geq 0$ a general regularized estimator of φ in H_s is given by $\hat{\varphi}_s = B^{-s/2} g_\alpha (B^{-s/2} \hat{T}^* \hat{T} B^{-s/2}) B^{-s/2} \hat{T}^* \hat{r}$ for some $\alpha > 0$ tending to zero as the samples size n increases, where g_α denotes some regularization scheme. However, to be more precise, let us consider the Tikhonov regularization of order ℓ (see Examples 3.1 (ii)). Therefore, define the diagonal matrix $\boldsymbol{b}^s := \mathrm{Diag}[b_1^s, \dots, b_m^s]$. Then the Tikhonov regularized estimator of order ℓ,

$$\widehat{\varphi}_s := \boldsymbol{\phi}' \widehat{\varphi}_{s,\ell} \, , \tag{4.11}$$

can be obtained by solving the ℓ linear matrix equations

$$(\widehat{M}'\widehat{M} + \alpha \boldsymbol{b}^s)\widehat{\varphi}_{s,j} = \widehat{M}'\widehat{v} + \alpha \boldsymbol{b}^s \widehat{\varphi}_{s,j-1}, \quad j = 1, \dots, \ell, \quad \widehat{\varphi}_{s,0} = 0. \tag{4.12}$$

Proposition 4.7. *Suppose that $\varphi \in H_p$ for some $p > 0$ and that for some $0 \leq s \leq p$, there exists an index function κ such that the operator T satisfies (2.2). Assume that on an interval $(0, c^2]$, the inverse function Φ of κ is convex. Suppose that the series estimators of r and T defined in (4.8) and (4.10) satisfy $\mathbb{E}\|\hat{r} - r\|^2 \leq \delta$ and $\mathbb{E}\|\hat{T} - T\|^4 \leq \gamma^2$, respectively. Consider the estimator $\widehat{\varphi}_s$ defined in (4.11) using $\alpha = d \cdot \{\delta + \gamma\}/\zeta(d \cdot \{\delta + \gamma\})$ for some $d > 0$, where ζ is the inverse function of $\zeta^{-1}(t) := t\Phi(t)$. Then $\mathbb{E}\|\widehat{\varphi}_s - \varphi\|_s^2 \leq C \cdot \zeta(\delta + \gamma)$ with $C > 0$.*

Remark 4.5. Suppose the level of smoothness of the function φ is characterized using the Hilbert scale $(\mathcal{W}_s[0,1])_s$ of Sobolev spaces in $L^2[0,1]$ (see Section 2.2 (ii)), i.e., $b_j := b_{2j+1} = (2j)^2$ and ϕ_j are trigonomeric functions. Then the last result provides again a bound of the $\mathcal{W}_s$-risk defined in the Sobolev space $\mathcal{W}_s[0,1]$.

(i) If we assume in addition that the conditional expectation operator T is infinitely smoothing, i.e., condition (2.4) holds, we may show that $\mathbb{E}\|\hat{T} - T\|^2 = O(n^{-2\tau'/(2\tau'+1)})$ and $\mathbb{E}\|\hat{r} - r\|^2 = O(n^{-2\tau/(2\tau+1)})$ for some $\tau > 0$ and $\tau' > 0$. Then the bound given in Proposition 4.7 simplifies to $\mathbb{E}\|\hat{\varphi}_s - \varphi\|_s^2 = O((\log n)^{-(p-s)/a})$. Thereby the remarks of Corollary 4.6 are still valid.

(ii) Let us suppose T is finitely smoothing and, hence satisfies (2.3). Then by a straightforward adaptation of the results derived in Hall and Horowitz (2005) we may show that $\mathbb{E}\|\hat{r} - r\|^2 = O(n^{-2(p+a)/(2(p+a)+1)})$ and $\mathbb{E}\|\hat{T} - T\|^{2[1 \vee (p-s)/(a+s)]} = O(n^{-2\tau/(2\tau+1)})$ for some $\tau > 0$. Thereby, if $\tau > (p + a)$ holds then supposing the order of the Tikhonov regularization satisfies $\ell \geq (p - s)/(a + s)$ is sufficient in order to ensure $\mathbb{E}\|\hat{\varphi}_s - \varphi\|_s^2 = O(n^{-2(p-s)/(2(a+p)+1)})$, which corresponds for $s = 0$ to the result derived in Hall and Horowitz (2005). However, we provide a bound also for the estimation of the s-th weak derivative of φ with $s \leq p$. □

A Proofs of Section 3

Lemma A.1. *Let $w : \Omega \to [1, \infty)$ be a weight function and let $\hat{\lambda}$ be the estimator of λ such that $\sup_{\omega\in\Omega} \mathbb{E}|\hat{\lambda}(\omega) - \lambda(\omega)|^{2(1+\tau)} \leq \gamma^{1+\tau}$ for some $\tau \geq 0$. Then for all $\omega \in \Omega$ we have*

$$\mathbb{E}\left[\mathbb{1}\{|\hat{\lambda}(\omega)/w(\omega)|^2 \geq \alpha\}.\frac{|\hat{\lambda}(\omega) - \lambda(\omega)|^2}{|\hat{\lambda}(\omega)|^2}\right] \leq \frac{C(\tau)}{|\lambda(\omega)/w(\omega)|^{2\tau}} \cdot \left\{\frac{\gamma^{1+\tau}}{\alpha} + \frac{\gamma}{\alpha^{1-\tau\wedge 1}}\right\} \tag{A.1}$$

where $C(\tau)$ is a positive constant depending only on τ.

Proof. Consider for $\omega \in \Omega$ the elementary inequality

$$1 \leq 2^{2\tau} \cdot \Big\{\frac{|\hat{\lambda}(\omega)/w(\omega) - \lambda(\omega)/w(\omega)|^{2\tau}}{|\lambda(\omega)/w(\omega)|^{2\tau}} + \frac{|\hat{\lambda}(\omega)/w(\omega)|^{2\tau}}{|\lambda(\omega)/w(\omega)|^{2\tau}}\Big\}, \tag{A.2}$$

which together with $|\lambda(\omega)/w(\omega)| \leq c$ for all $\omega \in \Omega$ and some $c > 0$ implies

$$\mathbb{E}\left[\mathbb{1}\{|\hat{\lambda}(\omega)/w(\omega)|^2 \geq \alpha\} \cdot \frac{|\hat{\lambda}(\omega) - \lambda(\omega)|^2}{\hat{\lambda}^2(\omega)}\right]$$
$$\leq \frac{2^{2\tau}}{|\lambda(\omega)/w(\omega)|^{2\tau}} \cdot \Big\{\frac{\mathbb{E}|\hat{\lambda}(\omega)/w(\omega) - \lambda(\omega)/w(\omega)|^{2(1+\tau)}}{\alpha} + \frac{\mathbb{E}|\hat{\lambda}(\omega)/w(\omega) - \lambda(\omega)/w(\omega)|^2}{\alpha^{1-\tau\wedge 1}}\Big\},$$

and, hence $\sup_{\omega\in\Omega} \mathbb{E}|\hat{\lambda}(\omega) - \lambda(\omega)|^{2(1+\tau)} \leq \gamma^{1+\tau}$ implies the result. □

Lemma A.2. *Let $w : \Omega \to [1, \infty)$ be an arbitrary weight function and let $\hat{\lambda}$ be an estimator of $\lambda \in L^\infty_\mu(\Omega)$ such that $\sup_{\omega\in\Omega} \mathbb{E}|\hat{\lambda}(\omega) - \lambda(\omega)|^4 \leq \gamma^2$. Suppose there exists an index function κ such that $\rho := \|w \cdot U\varphi \cdot |\kappa(|\lambda/w|^2)|^{-1/2}\|_{L^2_\mu(\Omega)} < \infty$ and assume that on an interval $(0, c^2]$ the inverse function Φ of κ is convex, then*

$$\mathbb{E}\|w \cdot U\varphi \cdot \mathbb{1}\{|\widehat{\lambda}/w|^2 < \alpha\}\|^2_{L^2_\mu(\Omega)} \leq C \cdot \kappa(C'\{\alpha + \gamma\}) \cdot \rho^2; \tag{A.3}$$

where C and C' are positive constants depending only on κ.

Proof. The proof is partially motivated by techniques used in Nair, Pereverzev, and Tautenhahn (2005). Denote $\hat{\psi}_\alpha = U\varphi \cdot \mathbb{1}\{|\widehat{\lambda}/w|^2 < \alpha\}$ and let $\phi(\omega) := \kappa^{1/2}(|\lambda(\omega)/w(\omega)|^2)$, $\omega \in \Omega$, then for all $\omega \in \Omega$ we have $\kappa^{1/2}(c) \geq \phi(\omega) > 0$ for some $c > 0$. Under the assumption $\rho = \|w \cdot U\varphi \cdot |\kappa(|\lambda/w|^2)|^{-1/2}\| < \infty$ (where we omitted the subscript $L^2_\mu(\Omega)$), which may be rewritten as $\rho = \|w \cdot U\varphi/\phi\| < \infty$, we obtain due to the Cauchy-Schwarz inequality

$$\|w \cdot \hat{\psi}_\alpha\|^2 = \int_\Omega w(\omega)\hat{\psi}_\alpha(\omega)\phi(\omega)\frac{w(\omega)[U\varphi](\omega)}{\phi(\omega)}\mu(d\omega) \leq \|w \cdot \hat{\psi}_\alpha \cdot \phi\| \cdot \rho, \tag{A.4}$$

which implies

$$\mathbb{E}\|w \cdot \hat{\psi}_\alpha\|^2 \leq (\mathbb{E}\|w \cdot \hat{\psi}_\alpha \cdot \phi\|^2)^{1/2} \cdot \rho. \tag{A.5}$$

Using (A.4) together with $\alpha \geq \sup_{t\in\mathbb{R}^+} t \cdot \mathbb{1}\{t < \alpha\}$ we have

$$\|\hat{\lambda} \cdot \hat{\psi}_\alpha\|^2 = \|\hat{\lambda} \cdot \mathbb{1}\{|\hat{\lambda}/w|^2 < \alpha\} \cdot \hat{\psi}_\alpha\|^2 \leq \alpha \cdot \|w \cdot \hat{\psi}_\alpha \cdot \phi\| \cdot \rho. \tag{A.6}$$

Therefore, applying the triangular inequality together with (A.6), we obtain

$$\mathbb{E}\|\lambda \cdot \hat{\psi}_\alpha\|^2 \leq 2\mathbb{E}\|w \cdot |\lambda/w - \hat{\lambda}/w| \cdot \hat{\psi}_\alpha\|^2 + 2\alpha(\mathbb{E}\|w \cdot \hat{\psi}_\alpha \cdot \phi\|^2)^{1/2} \cdot \rho.$$

By applying first the Cauchy-Schwarz inequality and using $\sup_{\omega\in\Omega}\mathbb{E}|\hat\lambda(\omega)-\lambda(\omega)|^4 \le \gamma^2$ then we bound the first term by

$$C\cdot\gamma\cdot\int(\mathbb{E}\mathbb{1}\{|\hat\lambda(\omega)/w(\omega)|^2<\alpha\})^{1/2}\cdot w^2(\omega)\cdot|[U\varphi](\omega)|^2\mu(d\omega),$$

and, hence using once again the Cauchy-Schwarz inequality leads to

$$\mathbb{E}\|\lambda\cdot\widehat{\psi_\alpha}\|^2 \le 2\Big\{C\gamma+\alpha\Big\}\cdot(\mathbb{E}\|w\cdot\widehat{\psi}_\alpha\cdot\phi\|^2)^{1/2}\cdot\rho. \tag{A.7}$$

Let Φ be the inverse function of κ which is convex on the interval $(0,c^2]$. Define $d^2=c^2/\kappa(1)\wedge 1$. Hence, using Jensen's inequality we have

$$\Phi\Big(\frac{d^2\cdot\mathbb{E}\|w\cdot\hat\psi_\alpha\cdot\phi\|^2}{\mathbb{E}\|w\cdot\hat\psi_\alpha\|^2}\Big)\le\frac{\mathbb{E}\int_\Omega\Phi(d^2\cdot\phi^2(\omega))\cdot w^2(\omega)\cdot\hat\psi_\alpha^2(\omega)\mu(d\omega)}{\mathbb{E}\int_\Omega w^2(\omega)\cdot\hat\psi_\alpha^2(\omega)\mu(d\omega)},$$

which together with $\Phi(d^2\cdot\phi^2(\omega))\le\Phi(\phi^2(\omega))=|\lambda(\omega)|^2/w^2(\omega)$ gives

$$\Phi\Big(\frac{d^2\cdot\mathbb{E}\|w\cdot\hat\psi_\alpha\cdot\phi\|^2}{\mathbb{E}\|w\cdot\hat\psi_\alpha\|^2}\Big)\le\frac{\mathbb{E}\int_\Omega|\lambda(\omega)|^2\cdot\hat\psi_\alpha^2(\omega)\mu(d\omega)}{\mathbb{E}\|w\cdot\hat\psi_\alpha\|^2}=\frac{\mathbb{E}\|\lambda\cdot\hat\psi_\alpha\|^2}{\mathbb{E}\|w\cdot\hat\psi_\alpha\|^2}. \tag{A.8}$$

In order to combine the three estimates (A.5), (A.7) and (A.8) let us introduce a new function Ψ by $\Psi(\omega):=\Phi(\omega^2)/\omega^2$. Since Φ is convex, we conclude that Ψ is monotonically increasing on the interval $(0,c]$. Hence, by (A.5), which may be rewritten as $(\mathbb{E}\|w\cdot\hat\psi_\alpha\cdot\phi\|^2)^{1/4}/\rho^{1/2}\le(\mathbb{E}\|w\cdot\hat\psi_\alpha\cdot\phi\|^2)^{1/2}/(\mathbb{E}\|w\cdot\psi_\alpha\|^2)^{1/2}$, the monotonicity of Ψ and (A.8),

$$\Psi\Big(\frac{d\cdot(\mathbb{E}\|w\cdot\hat\psi_\alpha\cdot\phi\|^2)^{1/4}}{\rho^{1/2}}\Big)\le\Psi\Big(\frac{d\cdot(\mathbb{E}\|w\cdot\hat\psi_\alpha\cdot\phi\|^2)^{1/2}}{(\mathbb{E}\|w\cdot\hat\psi_\alpha\|^2)^{1/2}}\Big)\le\frac{\mathbb{E}\|\lambda\cdot\hat\psi_\alpha\|^2}{d^2\cdot\mathbb{E}\|w\cdot\hat\psi_\alpha\cdot\phi\|^2}.$$

Multiplying by $d^2\cdot(\mathbb{E}\|w\cdot\hat\psi_\alpha\cdot\phi\|^2)^{1/2}/\rho$ and exploiting (A.7) yields

$$\Phi\Big(\frac{d^2\cdot(\mathbb{E}\|w\cdot\hat\psi_\alpha\cdot\phi\|^2)^{1/2}}{\rho}\Big)\le\frac{\mathbb{E}\|\lambda\cdot\hat\psi_\alpha\|^2}{\rho\cdot(\mathbb{E}\|w\cdot\hat\psi_\alpha\cdot\phi\|^2)^{1/2}}\le\alpha. \tag{A.9}$$

Thereby we obtain the result by combining (A.5) and (A.9). □

Proof of Theorem 3.1. Since T satisfies (2.2) for some index function κ and $\varphi\in H_p$, $p>0$, it follows that $\rho:=\|w\cdot U\varphi\cdot|\kappa(\lambda^2/w^2)|^{-1/2}\|_{L^2_\mu(\Omega)}<\infty$ with $w(\omega):=b^{s/2}(\omega)$, for $\omega\in\Omega$.
Define $U\hat\varphi_s^\alpha:=U\varphi\cdot\mathbb{1}\{\hat\lambda^2\ge\alpha\cdot b^s\}$. The proof is based on the decomposition

$$\mathbb{E}\|\hat\varphi_s-\varphi\|_s^2\le 2\mathbb{E}\|\hat\varphi_s-\hat\varphi_s^\alpha\|_s^2+2\mathbb{E}\|\hat\varphi_s^\alpha-\varphi\|_s^2 \tag{A.10}$$

Due to Lemma A.1 with $w\equiv b^{s/2}$ we show below the following bound

$$\mathbb{E}\|\hat\varphi_s-\hat\varphi_s^\alpha\|_s^2\le 2\alpha^{-1}\cdot\delta+2C(0)\cdot\rho^2\cdot\alpha^{-1}\cdot\gamma, \tag{A.11}$$

while from Lemma A.2 we obtain

$$\mathbb{E}\|\hat\varphi_s^\alpha-\varphi\|_s^2\le C\cdot\rho^2\cdot\kappa(C'\{\alpha+\gamma\}), \tag{A.12}$$

The conditions on α which may be rewritten as $d\cdot(\delta+\gamma)=d'\cdot\alpha\cdot\kappa(d'\cdot\alpha)$, for some constants d and d', ensure the balance of the two terms in (A.10), which gives the result.

Proof of (A.11). By definition, we have

$$\mathbb{E}\|\hat\varphi_s - \hat\varphi_s^\alpha\|_s^2 = \mathbb{E}\|w\cdot \mathbb{1}\{|\widehat{\lambda}/w|^2 \geq \alpha\}\cdot (\frac{V\hat r}{\hat\lambda} - U\varphi)\|^2_{L^2_\mu(\Omega)}$$

Using the triangular inequality, we get

$$\begin{aligned}\mathbb{E}\|\hat\varphi_s - \hat\varphi_s^\alpha\|_s^2 \leq{}& 2\mathbb{E}\|\frac{w}{\hat\lambda}\cdot \mathbb{1}\{|\widehat{\lambda}/w|^2 \geq \alpha\}\cdot V(\hat r - r)\|^2_{L^2_\mu(\Omega)}\\ &+ 2\mathbb{E}\|\frac{w}{\hat\lambda}\cdot \mathbb{1}\{|\widehat{\lambda}/w|^2 \geq \alpha\}\cdot (\lambda-\hat\lambda)U\varphi\|^2_{L^2_\mu(\Omega)} \quad \text{(A.13)}\end{aligned}$$

By assumption, $\mathbb{E}\|\hat r - r\|^2 < \delta$ and $\sup_{\omega\in\Omega}\mathbb{E}|\hat\lambda(\omega)-\lambda(\omega)|^2 \leq \gamma$, the first term is bounded by $2\alpha^{-1}\cdot\delta$, while we have for the second the bound $2\cdot\rho^2\cdot\frac{\gamma}{\alpha}$.

Proof of (A.12). By definition,

$$\mathbb{E}\|\hat\varphi_s^\alpha - \varphi\|_s^2 = \mathbb{E}\|b^{s/2}\cdot U\varphi\cdot \mathbb{1}\{|\widehat{\lambda}/w|^2 < \alpha\}\|^2_{L^2_\mu(\Omega)}$$

Lemma A.2 gives the following upper bound $\mathbb{E}\|\hat\varphi_s^\alpha - \varphi\|_s^2 \leq C\cdot\kappa(C'\{\alpha+\gamma\})\cdot\rho^2$ as soon as the condition $\rho := \|w\cdot U\varphi\cdot|\kappa(\lambda^2/w^2)|^{-1/2}\|_{L^2_\mu(\Omega)} < \infty$ is fulfilled. □

Proof of Theorem 3.2. Since T satisfies (2.3) for some $a>0$ and $\varphi\in H_p$, $p>0$, it follows that $\rho := \|w\cdot U\varphi\cdot|\lambda/w|^{-\beta}\|_{L^2_\mu(\Omega)} < \infty$ with $\beta := (p-s)/(a+s)$ and $w := b^{s/2}$.
Considering the decomposition (A.10) based on Lemma A.1 we show below

$$\mathbb{E}\|\hat\varphi_s - \hat\varphi_s^\alpha\|_s^2 \leq 2\alpha^{-1}\cdot\delta + 2C(\beta)\cdot\rho^2\cdot\Big\{\frac{\gamma^{1+\beta}}{\alpha} + \frac{\gamma}{\alpha^{1-\beta\wedge 1}}\Big\}, \quad \text{(A.14)}$$

while from Lemma A.2 we conclude

$$\mathbb{E}\|\hat\varphi_s^\alpha - \varphi\|_s^2 \leq C_\beta\cdot\rho^2\cdot\{\alpha^\beta + \gamma^\beta\}. \quad \text{(A.15)}$$

The condition on α ensures then the balance of these two terms, which gives the result.

Proof of (A.14). Consider the decomposition (A.13), then the first term is bounded by $2\alpha^{-1}\cdot\delta$, while we use $\rho = \|w\cdot U\varphi\cdot|\lambda/w|^{-\beta}\| < \infty$ together with Lemma A.1 to bound the second term, which gives (A.14).

Proof of (A.15). Define $\hat\psi_\alpha := U(\varphi - \hat\varphi_s^\alpha)$, then $\hat\psi_\alpha = U\varphi\cdot\mathbb{1}\{\hat\lambda^2 < \alpha\cdot w^2\}$. Using the inequality (A.2) together with $\alpha^\gamma \geq \sup_{t\in\mathbb{R}^+} t^\gamma \mathbb{1}\{t<\alpha\}$ for all $\gamma>0$ we have

$$\begin{aligned}\|w\cdot\widehat{\psi}_\alpha\|^2 &\leq 2^{2\beta}\Big\{\|w\cdot\widehat{\psi}_\alpha\cdot\frac{|\hat\lambda/w|^\beta}{|\lambda/w|^\beta}\|^2 + \|w\cdot\widehat{\psi}_\alpha\cdot\frac{|\hat\lambda/w - \lambda/w|^\beta}{|\lambda/w|^\beta}\|^2\Big\}\\ &\leq 2^{2\beta}\Big\{\alpha^\beta\cdot\rho^2 + \|w\cdot U\varphi\cdot|\lambda/w|^{-\beta}\cdot|\hat\lambda/w - \lambda/w|^\beta\|^2\Big\},\end{aligned}$$

since $\rho = \|w\cdot U\varphi\cdot|\lambda/w|^{-\beta}\| < \infty$. Thereby, the identity $\mathbb{E}\|\hat\varphi_s^\alpha - \varphi\|_s^2 = \mathbb{E}\|w\cdot\hat\psi_\alpha\|^2_{L^2_\mu(\Omega)}$ together with Lemma A.1 imply (A.14). □

Denote $T_s := TB^{-s/2}$, $\widehat{T}_s := \widehat{T}B^{-s/2}$, $\varphi_s := B^{s/2}\varphi$ and $\widehat{\varphi}_s^\alpha := B^{-s/2}g_\alpha(\widehat{T}_s^\star\widehat{T}_s)\widehat{T}_s^\star\widehat{T}_s\varphi_s$.

Lemma A.3. *Suppose the assumptions of Theorem 3.3 are satisfied. Then*

$$\mathbb{E}\|\widehat{\varphi}_s^\alpha - \varphi\|_s^2 \leq C\cdot\kappa(C'\{\alpha + (\mathbb{E}\|(T-\widehat{T})^2\|^2)^{1/2}\})\cdot\rho^2; \quad \text{(A.16)}$$

where C and C' are positive constants depending only on κ.

Proof. Since T satisfies (2.2) for some index function κ and $\varphi \in H_p$, we conclude that $\rho := \|\kappa^{-1/2}(T_s^\star T_s)\varphi_s\| < \infty$. Let $\widehat{\psi}_\alpha := \varphi - \widehat{\varphi}_s^\alpha$ and $\widehat{R}_\alpha := [I - g_\alpha(\widehat{T}_s^\star \widehat{T}_s)\widehat{T}_s^\star \widehat{T}_s]$, then we have $\widehat{\psi}_\alpha = B^{-s/2}\widehat{R}_\alpha \varphi_s$. We use $\|\widehat{R}_\alpha^{1/2}\| \leq 1$ (Assumption 3.1 (ii)) and obtain due to the Cauchy-Schwarz inequality

$$\|\widehat{\psi}_\alpha\|_s^2 = \|\widehat{R}_\alpha\varphi_s\|^2 \leq \|\widehat{R}_\alpha^{1/2}\varphi_s\|^2 = \langle \widehat{R}_\alpha\varphi_s, \varphi_s\rangle = \langle \kappa^{1/2}(T_s^\star T_s)B^{s/2}\widehat{\psi}_\alpha, \kappa^{-1/2}(T_s^\star T_s)\varphi_s\rangle$$
$$\leq \|\kappa^{1/2}(T_s^\star T_s)B^{s/2}\widehat{\psi}_\alpha\| \cdot \rho. \quad \text{(A.17)}$$

Thereby we have

$$\mathbb{E}\|\widehat{\psi}_\alpha\|_s^2 \leq (\mathbb{E}\|\kappa^{1/2}(T_s^\star T_s)B^{s/2}\widehat{\psi}_\alpha\|^2)^{1/2}\rho. \quad \text{(A.18)}$$

Using (A.18) together with $d\alpha \geq \|(\widehat{T}_s^\star \widehat{T}_s)^{1/2}\widehat{R}_\alpha^{1/2}\|^2$ (Assumption 3.1 (ii)) we obtain

$$\|\widehat{T}\widehat{\psi}_\alpha\|^2 = \|\widehat{T}_s\widehat{R}_\alpha\varphi_s\|^2 = \|(\widehat{T}_s^\star \widehat{T}_s)^{1/2}\widehat{R}_\alpha\varphi_s\|^2 \leq d\alpha\|\widehat{R}_\alpha^{1/2}\varphi_s\|^2 \leq d\alpha\|\kappa^{1/2}(T_s^\star T_s)B^{\mathfrak{s}/2}\widehat{\psi}_\alpha\|\rho$$

and hence,

$$\mathbb{E}\|\widehat{T}\widehat{\psi}_\alpha\|^2 \leq d\alpha \cdot (\mathbb{E}\|\kappa^{1/2}(T_\mathfrak{s}^\star T_s)B^{s/2}\widehat{\psi}_\alpha\|^2)^{1/2} \cdot \rho. \quad \text{(A.19)}$$

Using (A.18) together with $\|B^{-\mathfrak{s}/2}\| \leq c$ we have due to the Cauchy Schwarz inequality

$$\mathbb{E}\|(T - \widehat{T})\widehat{\psi}_\alpha\|^2 \leq \mathbb{E}\|(T - \widehat{T})\|^2\|\widehat{\psi}_\alpha\|^2 \leq c\mathbb{E}\|T - \widehat{T}\|^2\|\widehat{R}_\alpha\varphi_s\|^2$$
$$\leq c \cdot (\mathbb{E}\|T - \widehat{T}\|^4)^{1/2}(\mathbb{E}\|\kappa^{1/2}(T_s^\star T_s)B^{s/2}\widehat{\psi}_\alpha\|^2)^{1/2} \cdot \rho. \quad \text{(A.20)}$$

Combining (A.19) and (A.20) we obtain

$$\mathbb{E}\|T\widehat{\psi}_\alpha\|^2 \lesssim \mathbb{E}\|(T - \widehat{T})\widehat{\psi}_\alpha\|^2 + \mathbb{E}\|\widehat{T}\widehat{\psi}_\alpha\|$$
$$\leq c \cdot \Big\{(\mathbb{E}\|T - \widehat{T}\|^4)^{1/2} + \alpha\Big\} \cdot (\mathbb{E}\|\kappa^{1/2}(T_s^\star T_s)B^{s/2}\widehat{\psi}_\alpha\|^2)^{1/2} \cdot \rho. \quad \text{(A.21)}$$

Let Φ be the inverse function of κ, which is assumed to be convex on the interval $(0, c^2]$. Define $d^2 = c^2/\|\kappa^{1/2}(T_s^\star T_s)\|^2 \wedge 1$. If $\{\lambda_s^2, U_s : H \to L^2_{\mu_S}(\Omega_s)\}$ denotes the spectral decomposition of $T_s^\star T_s$, then $c^2 \geq d^2\kappa(\lambda_s^2)$. Hence, using Jensen's inequality we have

$$\Phi\Big(\frac{d^2\mathbb{E}\|\kappa^{1/2}(T_s^\star T_s)B^{s/2}\widehat{\psi}_\alpha\|^2}{\mathbb{E}\|\widehat{\psi}_\alpha\|_s^2}\Big) \leq \frac{\mathbb{E}\int_{\Omega_s}\Phi(d^2\kappa(\lambda_s^2(\omega)))|U_sB^{s/2}\widehat{\psi}_\alpha|^2(\omega)\mu_s(d\omega)}{\mathbb{E}\int_{\Omega_s}|U_sB^{s/2}\widehat{\psi}_\alpha|^2(\omega)\mu_s(d\omega)}.$$

Since $\Phi(d^2\kappa(\lambda_s^2)) \leq \Phi(\kappa(\lambda_s^2)) = \lambda_s^2$ we obtain

$$\Phi\Big(\frac{d^2\mathbb{E}\|\kappa^{1/2}(T_s^\star T_s)B^{s/2}\widehat{\psi}_\alpha\|^2}{\mathbb{E}\|\widehat{\psi}_\alpha\|_s^2}\Big) \leq \frac{\mathbb{E}\|(T_s^\star T_s)^{1/2}B^{s/2}\widehat{\psi}_\alpha\|^2}{\mathbb{E}\|\widehat{\psi}_\alpha\|_s^2} = \frac{\mathbb{E}\|T\widehat{\psi}_\alpha\|^2}{\mathbb{E}\|\widehat{\psi}_\alpha\|_\mathfrak{s}^2}. \quad \text{(A.22)}$$

Combining as in the proof of Lemma A.2 the three estimates (A.17), (A.21) and (A.22) by using the function $\Psi(t) = \Phi(t^2)/t^2$ we obtain

$$\Phi\Big(\frac{d^2(\mathbb{E}\|\kappa^{1/2}(T_s^\star T_s)B^{s/2}\widehat{\psi}_\alpha\|^2)^{1/2}}{\rho}\Big) \leq c\{\alpha + (\mathbb{E}\|T - \widehat{T}\|^4)^{1/2}\}, \quad \text{(A.23)}$$

and, thereby (A.18) together with (A.23) implies the result. □

Proof of Theorem 3.3. The proof is also based on the decomposition (A.10) with $\widehat{\varphi}_s^\alpha := B^{-s/2}g_\alpha(\widehat{T}_s^\star \widehat{T}_s)\widehat{T}_s^\star \widehat{T}_s\varphi_s$. We show below that under the assumptions of the theorem the two bounds

(A.11) and (A.12) still hold with $\rho := \|\kappa^{-1/2}(T_s^\star T_s)\varphi_s\| < \infty$, since T satisfies (2.2) and $\varphi \in H_p$. Thereby, the condition on α which may be rewritten as $d\cdot(\delta+\gamma) = d'\cdot\alpha\cdot\kappa(d'\cdot\alpha)$, for some constants d and d', ensure the balance of the two terms in (A.10), which gives the result.

By definition, together with $\|g_\alpha(\widehat{T}_s^\star\widehat{T}_s)\widehat{T}_s^\star\|^2 \le c/\alpha$ (Assumption 3.1 (i)) we have

$$\mathbb{E}\|\hat{\varphi}_s - \hat{\varphi}_s^\alpha\|_s^2 = \mathbb{E}\|g_\alpha(\widehat{T}_s^\star\widehat{T}_s)\widehat{T}_s^\star(\widehat{r} - \widehat{T}\varphi)\|^2 \le c\alpha^{-1}\mathbb{E}\|\widehat{r} - \widehat{T}\varphi\|^2.$$

Thereby, the assumptions $\mathbb{E}\|\hat{r} - r\|^2 < \delta$ and $\mathbb{E}\|\hat{T} - T\|^4 \le \gamma^2$ imply the first bound (A.11), while the second bound (A.12) follows from Lemma A.3. □

Proof of Theorem 3.4. Since T satisfies (2.3) for some $a > 0$ and $\varphi \in H_p$, $p > 0$, it follows that $\rho := \|(T_s^\star T_s)^{-\beta}\varphi_s\| < \infty$ with $\beta := (p-s)/(a+s)$. Considering the decomposition (A.10) with $\widehat{\varphi}_s^\alpha := B^{-s/2}g_\alpha(\widehat{T}_s^\star\widehat{T}_s)\widehat{T}_s^\star\widehat{T}_s\varphi_s$ we bound the first term as in Theorem 3.4, that is $\mathbb{E}\|\hat{\varphi}_s - \hat{\varphi}_s^\alpha\|_s^2 \le C\{\alpha^{-1}\cdot\delta + \rho^2\cdot\alpha^{-1}\cdot\gamma\}$, while we show below that under the assumptions of the theorem the following bound of the second term holds.

$$\mathbb{E}\|\hat{\varphi}_s^\alpha - \varphi\|_s^2 \le C\cdot\rho^2\cdot\{\alpha^\beta + \gamma^{\beta\wedge 1}\}), \tag{A.24}$$

Thereby, the condition on α ensures then the balance of these two terms, which gives the result.

Proof of (A.24). Let $\widehat{\psi}_\alpha := \varphi - \widehat{\varphi}_s^\alpha$ and $\widehat{R}_\alpha := [I - g_\alpha(\widehat{T}_s^\star\widehat{T}_s)\widehat{T}_s^\star\widehat{T}_s]$, then we have

$$\|\widehat{\psi}_\alpha\|_s^2 = \|\widehat{R}_\alpha\varphi_s\|^2 \le C\cdot\{\|\widehat{R}_\alpha(\widehat{T}_s^\star\widehat{T}_s)^{\beta/2}(T_s^\star T_s)^{-\beta/2}\varphi_s\|^2 + 2\|\widehat{R}_\alpha[(\widehat{T}_s^\star\widehat{T}_s)^{\beta/2} - (T_s^\star T_s)^{\beta/2}](T_s^\star T_s)^{-\beta/2}\varphi_s\|^2.$$

Thereby, since $\|\widehat{R}_\alpha\| \le 1$ (Assumption 3.1 (ii)) and $\|\widehat{R}_\alpha(\widehat{T}_s^\star\widehat{T}_s)^{\beta/2}\|^2 \le c_\beta\alpha^\beta$ (Assumption 3.2), it follows, that

$$\|\widehat{\psi}_\alpha\|_s^2 \le C\rho^2\{\alpha^\beta + \|(\widehat{T}_s^\star\widehat{T}_s)^{\beta/2} - (T_s^\star T_s)^{\beta/2}]\|^2\}.$$

Thereby, since $\|(\widehat{T}_s^\star\widehat{T}_s)^{\beta/2} - (T_s^\star T_s)^{\beta/2}]\|^2 \le c\{\|\widehat{T} - T\|^{2(\beta\wedge 1)} + \|\widehat{T} - T\|^{2\beta}\}$ (see Egger (2005) Lemma 3.2), the assumption $\mathbb{E}\|\widehat{T} - T\|^2 \le \gamma$ and if $\beta > 1$, $\mathbb{E}\|\widehat{T} - T\|^{2\beta} \le \gamma$ together with Lyapunov's inequality implies (A.24). □

B Proofs of Section 4

Proof of Proposition 4.1. The result follows from Theorem 3.1, where we use that $\gamma = 1/m$, i.e., $\sup_{t\in\mathbb{R}}\mathbb{E}|[\widehat{\mathcal{F}f_\varepsilon}](t) - [\mathcal{F}f_\varepsilon](t)|^{2\tau} \le c(\tau)n^{-\tau}$ for some $c(\tau) > 0$ and for all $\tau > 0$ (see Johannes (2007) for a detailed proof). □

Proof of Corollary 4.2. Since, f_Y belongs to $\mathcal{W}_{p+a}(\mathbb{R})$ under the conditions of the corollary the kernel estimator $\widehat{f_Y}$ obtain the order $O(n^{-2(p+a)/(2(p+a)+1)})$ and, hence we have $\delta = n^{-2(p+a)/(2(p+a)+1)}$. Thereby, the result follows from Theorem 3.2 with $\gamma = 1/m$ (see proof of Proposition 4.1). □

Proof of Corollary 4.3. Under the conditions of the corollary the kernel estimator $\widehat{f_Y}$ provides an order $O(n^{-2\tau/(2\tau+1)})$, i.e., $\delta = n^{-2\tau/(2\tau+1)}$ and, hence the result follows from Proposition 4.1 (see also Remark 3.1). □

Proof of Proposition 4.4.The result follows directly from Theorem 3.1. □

Proof of Corollary 4.5. Since, φ belongs to $\mathcal{W}_{p+a}[0,1]$ and, hence under the conditions of the corollary r lies in $\mathcal{W}_{p+a}[0,1]$ the series estimator $\widehat{r}$ obtains a rate of convergence of the order

$O(n^{-2(p+a)/(2(p+a)+1)})$, i.e., we have $\delta = n^{-2(p+a)/(2(p+a)+1)}$. Moreover, straightforward calculus gives $\sup_{j\in\mathbb{N}} \mathbb{E}|\widehat{\lambda}_j - \lambda_j|^{2\tau} \leq c(\tau)n^{-\tau}$ for all $\tau > 0$, i.e., $\gamma = 1/n$. Thereby, the result follows from Theorem 3.2 with $\gamma = 1/n$. □

Proof of Corollary 4.6. Under the conditions of the corollary the series estimator $\widehat{r}$ provides an order $O(n^{-2\tau/(2\tau+1)})$, i.e., $\delta = n^{-2\tau/(2\tau+1)}$ and we have $\sup_{j\in\mathbb{N}} \mathbb{E}|\widehat{\lambda}_j - \lambda_j|^{2\tau} \leq c(\tau)n^{-\tau}$, i.e., $\gamma = 1/n$. Thereby the result follows from Proposition 4.4 (see also Remark 3.1). □

Proof of Proposition 4.7.The result follows directly from Theorem 3.1. □

References

Adams, R. A. (1975): *Sobolev Spaces.* Academic Press, New York.

Bigot, J., and S. Van Bellegem (2006): "Log-density deconvolution by wavelet thresholding," Discussion Paper # 0617, Institut de statistique, Université catholique de Louvain.

Blundell, R., X. Chen, and D. Kristensen (2007): "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves," *Econometrica*, Forthcoming.

Blundell, R., and J. Horowitz (2007): "A Nonparametric Test of Exogeneity," *Rev. Econ. Stud.*, forthcoming.

Bonhomme, S., and J.-M. Robin (2006): "Generalized Nonparametric Deconvolution with an Application to Earnings Dynamics," Working Paper.

Carrasco, M., and J.-P. Florens (2002): "Spectral method for deconvolving a density," Working Paper 138, IDEI, Toulouse.

Carroll, R. J., and P. Hall (1988): "Optimal rates of convergence for deconvolving a density.," *J. Am. Statist. Assoc.*, 83, 1184–1186.

Cavalier, L., and N. Hengartner (2005): "Adaptive estimation for inverse problems with noisy operators.," *Inverse Problems*, 21, 1345–1361.

Cazals, C., J.-P. Florens, and L. Simar (2007): "Nonparametric frontier models for panel data," Mimeo.

Chen, X. (2008): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6. Elsevier, to appear.

Chen, X., and M. Reiss (2007): "On rate optimality for ill-posed inverse problems in econometrics," Cowles foundation discussion paper No. 1626.

Darolles, S., J.-P. Florens, and E. Renault (2002): "Nonparametric Instrumental Regression," Working Paper # 228, IDEI, Université de Toulouse I.

Douglas, R. (1966): "On Majorization, Factorization, and Range Inclusion of Operators on Hilbert Space.," *Proc. Amer. Math. Soc.*, 17, 413–415.

Egger, H. (2005): "Accelerated Newton-Landweber iterations for regularizing nonlinear inverse problems," SFB-Report 2005-3, Austrian Academy of Sciences, Linz.

Engl, H. W., M. Hanke, and A. Neubauer (2000): *Regularization of Inverse Problems.* Kluwer Academic, Dordrecht.

Fan, J. (1991): "On the optimal rates of convergence for nonparametric deconvolution problems.," *Ann. Stat.*, 19, 1257–1272.

——— (1992): "Deconvolution with supersmooth distributions.," *Can. J. Statist.*, 20, 734–747.

Fan, J., and J. Y. Koo (2002): "Wavelet deconvolution.," *IEEE Trans. Info. Theory*, 48, 734–747.

Florens, J.-P., J. Johannes, and S. Van Bellegem (2005): "Instrumental Regression in Partially Linear Models," Discussion Paper # 0537, Institut de statistique, Université catholique de Louvain.

——— (2007): "Identification and estimation by penalization in Nonparametric Instrumental Regression.," Discussion Paper # 0721, Institut de statistique, Université catholique de Louvain.

Gagliardini, P., and O. Scaillet (2007): "A Specification Test for Nonparametric Instrumental Variable Regression," Swiss Finance Institute Research Paper No. 07-13.

Goldenshluger, A., and S. V. Pereverzev (2003): "On adaptive inverse estimation of linear functional in Hilbert scales.," *Bernoulli*, 9(5), 783–807.

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

Hall, P., and Q. Yao (2003): "Inference in components of variance models with low replication.," *Ann. Stat.*, 31(2), 414–441.

Halmos, P. R. (1963): "What does the spectral theorem say?," *Amer. Math. Monthly*, 70, 241–247.

Hegland, M. (1995): "Variable Hilbert scales and their interpolation inequalities with applications to Tikhonov regularization.," *Appl. Anal.*, 59(1-4), 207–223.

Horowitz, J. L. (1998): *Semiparametric Methods in Econometrics.* Springer, New York.

Horowitz, J. L., and M. Markatou (1996): "Semiparametric estimation of regression models for panel data," *Review of Economic Studies*, 63, 145–168.

Johannes, J. (2007): "Deconvolution with unknown error density.," Discussion Paper, University Heidelberg (in revision).

Johannes, J., S. Van Bellegem, and A. Vanhems (2007): "Projection estimation in nonparametric instrumental regression," Discussion Paper, Institut de statistique, Université catholique de Louvain.

Johannes, J., and A. Vanhems (2005): "Regularity conditions for inverse problems in econometrics.," Working Paper presented at the Econometrics World Congress, London.

Kawata, T. (1972): *Fourier Analysis in Probability Theory.* Academic Press, New York.

Koo, J. Y., and B. U. Park (1996): "B-spline deconvolution based on the EM algorithm," *J. Statist. Comp. Simul.*, 54, 275–288.

Krein, S., and Y. I. Petunin (1966): "Scales of Banach spaces," in *Russian Math. Surveys*, vol. 21, pp. 85–169.

Mair, B. A. (1994): "Tikhonov regularization for finitely and infinitely smoothing operators.," *SIAM J. Appl. Anal.*, 25, 135–147.

Mair, B. A., and F. H. Ruymgaart (1996): "Statistical inverse estimation in Hilbert scales," *SIAM J. Appl. Math.*, 56, 1424–1444.

Mathé, P., and S. V. Pereverzev (2001): "Optimal discretization of inverse problems in Hilbert scales. regularization and self-regularization of projection methods.," *SIAM J. Numer. Anal.*, 38(6), 1999–2021.

Mendelsohn, J., and J. Rice (1982): "Deconvolution of micro-fluorometric histograms with B-splines.," *J. Am. Statist. Assoc.*, 77, 748–753.

Nair, M., S. V. Pereverzev, and U. Tautenhahn (2005): "Regularization in Hilbert scales under gerneal smoothing conditions.," *Inverse Problems*, 21, 1851–1869.

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

Neumann, M. H. (1997): "On the effect of estimating the error density in nonparametric deconvolution.," *Nonparametric Statistics.*, 7, 307–330.

——— (2006): "Deconvolution from panel data with unknown error distribution.," Preprint., University Jena.

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

Parzen, E. (1962): "On Estimation of a Probability Density Function and Mode.," *Ann. Math. Stat.*, 33(3), 1065–1076.

Pensky, M., and B. Vidakovic (1999): "Adaptive wavelet estimator for nonparametric density deconvolution.," *Ann. Stat.*, 27, 2033–2053.

Postel-Vinay, F., and J.-M. Robin (2002): "Equilibrium wage dispersion with worker and employer heterogeneity.," *Econometrica*, 70, 2295–2350.

Tautenhahn, U. (1996): "Error estimates for regularization methods in Hilbert scales.," *SIAM J. Numer. Anal.*, 33(6), 2120–2130.