TSE WP working paper series

10-184

Research Group: *Econometrics and Statistics*

July 16, 2010

Iterative Regularization in Nonparametric Instrumental Regression

JAN JOHANNES, SÉBASTIEN VAN BELLEGEM AND ANNE VANHEMS

Iterative regularization in nonparametric instrumental regression*

Jan Johannes[1] Sébastien Van Bellegem[2] Anne Vanhems[3]

July 16, 2010

Abstract

We consider the nonparametric regression model with an additive error that is correlated with the explanatory variables. We suppose the existence of instrumental variables that are considered in this model for the identification and the estimation of the regression function. The nonparametric estimation by instrumental variables is an ill-posed linear inverse problem with an unknown but estimable operator. We provide a new estimator of the regression function using an iterative regularization method (the Landweber-Fridman method). The optimal number of iterations and the convergence of the mean square error of the resulting estimator are derived under both mild and severe degrees of ill-posedness. A Monte-Carlo exercise shows the impact of some parameters on the estimator and concludes on the reasonable finite sample performance of the new estimator.

Keywords: Nonparametric estimation, Instrumental variable, Ill-posed inverse problem, Iterative method, Estimation by projection

JEL classifications: Primary C14; secondary C30

*This work was supported by the "Agence National de le Recherche" under contract ANR-09-JCJC-0124-01 and by the IAP research network nr P6/03 of the Belgian Government (Belgian Science Policy).

[1]Institut de statistique, biostatistique et sciences actuarielles (Belgium)

[2]Toulouse School of Economics (France) & Center for Operations Research and Econometrics (Belgium)

[3]Toulouse Business School & Toulouse School of Economics (France)

1 Introduction

The statistical inference in the nonparametric regression model

$$Y = \varphi(Z) + \varepsilon$$

where Y is a real dependent variable, Z is a real multivariate explanatory random variable and φ is an unknown function, usually requires that the error term ε is such that $\mathbb{E}(\varepsilon|Z) = 0$. A vast literature is now available about the estimation of the nonparametric function in this setting, under various regularity assumptions on φ. Less work is however available if the error ε is allowed to be correlated with Z, a situation that is frequently encountered in the empirical studies in human sciences.

A simple situation where this correlation appears is when omitted variables influence both Y and Z but are not included in the explanatory vector of the regression model. One famous example of this situation appears when Y is a measure of the income of some individuals and Z is a measure of their level of education. It is likely that other variables, such as a measure of the social ability or the intelligence, may be influencial on both the income and the level of education of the individuals. However, this variable is rarely observed or even difficult to measure, and it is thus omitted in the regression model. In consequence, the error term ε contains an information about the omitted variable, and therefore may depend on the observed explanatory variables. This example is discussed in more details in many econometrics textbooks [e.g. Wooldridge (2008)], see also the survey written by Angrist & Krueger (2001).

One conventional approach to accomodate this problem is to measure a set of new variables W that are called "instrumental variables" and such that $\mathbb{E}(\varepsilon|W) = 0$. Taking the conditional expectation of the above regression model, the nonparametric function φ now appears to be the solution of

$$\mathbb{E}(Y|W) = \mathbb{E}(\varphi(Z)|W). \tag{1.1}$$

The choice of appropriate instruments W is a delicate question in practice. The interested reader can find examples of instrumental variables in the above econometric references.

As a statistician, the interesting and nonstandard point is that the nonparametric function φ appears to be the solution of an integral equation given by (1.1). It is thus the solution of an ill-posed inverse problem. Moreover, the involved conditional expectations must be estimated from observations of (Y, Z, W), which is a source of error in both sides of equation (1.1).

A number of papers have considered the estimation of φ in model (1.1) from the observation of (Y, Z, W) when the conditional expectations are estimated nonparametrically and by regularizing the ill-posed problem in order to recover a consistent solution. We refer to Hall & Horowitz (2005) for some methods and to our recent paper, Johannes et al. (2010), that presents a unified approach to get consistent estimators of φ with optimal rates of convergence.

One goal of this paper is to present new results on a regularization scheme that has been less studied in this context. This regularization is the so-called "Landweber-Fridman" iterative procedure that we define below. Moreover, our estimator of the conditional expectations and φ are projection estimators onto finite-dimensional vectorial spaces, with dimension increasing with the sample size. The expansion of the nonparametric function φ is provided in a basis that may be different from the basis that is used to estimate the conditional expectations. This aspect of our procedure is valuable, since the degree of regularity of φ may be very different from the degree of regularity of $\mathbb{E}(\varphi(Z)|W)$.

Under a set of minimal conditions on the choice of those bases, we prove the consistency and the mean square convergence of the estimator given by iterative regularization. Results are given for both the mildly and severely ill-posed inverse problems, and reach the optimal minimax rate of convergence in standard function spaces in both cases. The results also give optimal stopping rule for the iterative regularization of the estimator.

The paper is organized as follows. In Section 2, we introduce the necessary notations of the paper and, more importantly, we formulate the estimation problem as an ill-posed inverse problem with an unknown linear operator. In Section 3 we derive the projection estimator and apply the regularization iterative method of Landweber-Fridman. Theoretical properties of the estimators have to be found in Sections 4 and 5. Those sections include the derivation of the rate of convergence under the various sets of regularity conditions, and provides a comparison with the most recent results of the literature. Section 6 discusses the role of the parameters in the estimation procedure and shows the finite sample properties of the proposed procedure via simulations. The proofs and technical results are deferred to an Appendix.

2 Model and assumptions

Let $Z \in \mathbb{R}^p$ and $W \in \mathbb{R}^q$ be two vectors of observed variables. In this section, we will write the nonparametric function φ as a solution of an inverse problem. Define the function spaces

$$L_Z^2 = \{\phi : \mathbb{R}^p \to \mathbb{R}, \|\phi\|_Z^2 := \mathbb{E}[\phi^2(Z)] < \infty\}$$

and

$$L_W^2 = \{\psi : \mathbb{R}^q \to \mathbb{R}, \|\psi\|_W^2 := \mathbb{E}[\psi^2(W)] < \infty\} .$$

For the sake of readability, we shall denote by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ the inner product and norm for both Hilbert spaces L_Z^2 and L_W^2 when there is no possible confusion about the functional space in question. The equation (1.1) can be rewritten

$$r = T\varphi \tag{2.1}$$

if r is a function of L^2_W such that $r(\cdot) := \mathbb{E}(Y|W = \cdot)$ and T is a linear operator that maps the function φ onto the conditional expectation, i.e.

$$T : L^2_Z \to L^2_W : \phi \mapsto \mathbb{E}(\phi(Z)|W = \cdot)$$

assuming the existence of the conditional density of Z given W, here denoted by $f_{Z|W}$.

For convenience, we will assume in the following that the operator T is a compact operator. This assumption implies that T and the adjoint operator $T^\star$ can be discretized using a singular value decomposition (SVD). We recall that the compacity of T implies the existence of a singular system $\{(\lambda_k, u_k, v_k); k = 1, 2, \ldots\}$ that is such that:

1. The eigenvalues λ_k are real, strictly positive and decreasing;
2. $\{\lambda_k^2; k = 1, 2, \ldots\}$ are the nonzero eigenvalues of the self-adjoint operators $T^\star T$ and $TT^\star$;
3. $\{u_k; k = 1, 2, \ldots\}$ (resp. $\{v_k; k = 1, 2, \ldots\}$) is an orthonormal system of eigenvectors of $T^\star T$ (resp. $TT^\star$).

The existence and uniqueness of the solution from equation (2.1) needs some assumptions. A detailed discussion on identification issues can be found in the seminal work of Darolles et al. (2002). In the sequel, we assume the existence and uniqueness of the solution. Uniqueness is guaranteed if we assume the operator T to be injective. The existence assumption formally requires that the function r belongs to the range of the operator T.

From (2.1), we see that φ can be recovered by inverting the operator T. However, even if T is in general an invertible operator, it is not necessary stable or, in more technical terms, T^{-1} is not a bounded operator. In other words, since the left hand side of equation (2.1) is not observed directly but needs to be estimated by $\hat{r}$, the solution $T^{-1}\hat{r}$ does not converge to $T^{-1}r$. That phenomenon is called the “ill-posedness” of the inversion. Therefore, in order to derive a consistent estimator of the functional parameter of interest φ, we shall proceed in two steps. First, T and r depend on the distribution of (X, Y, W) and are estimated from the dataset using projection method. Second, a regularized version of (2.1) is obtained using a Landweber-Fridman iterative method.

We now describe the two steps in detail.

3 Estimation and regularization

3.1 Projection step

Define two finite dimensional subsets Φ_{d_Z} of L^2_Z and Ψ_{d_W} of L^2_W. Their dimension depend on prescribed numbers $d_Z, d_W > 0$ and we suppose that $\{\phi_1, \ldots, \phi_{d_Z}\}$ is a basis for Φ_{d_Z} and $\{\psi_1, \ldots, \psi_{d_W}\}$ is a basis for Ψ_{d_W}. Note that the each of these bases is not necessarily an orthonormal basis.

The two bases are chosen independently of the operator T. However, our theory requires a condition that relates both the bases and T. Denote by P_Z, resp. Q_W, the orthogonal projection onto Φ_{d_Z}, resp. Ψ_{d_W} and by I the identity operator. We assume from now on that the operators P_Z, Q_W and T are such that

$$\|T(I - P_Z)\| \to 0 \qquad \text{and} \qquad \|(I - Q_W)T\| \to 0 \tag{3.1}$$

as $(d_Z, d_W) \to \infty$, where the norms are the operator norm, e.g.

$$\|T\| := \sup\{\|T\psi\|_W;\ \psi \in L_Z^2\}$$

is the operator norm of T. This condition is discussed in the following remark and two examples.

REMARK 3.1. Because Φ_{d_Z} and Ψ_{d_W} are finite dimensional, the range of the operators P_Z and Q_W that we denote by $\mathcal{R}(P_Z)$ and $\mathcal{R}(Q_W)$ are finite dimensional. Using a result to be found in Plato & Vainikko (1990), a necessary and sufficient condition in order to get condition (3.1) is that (i) T is a compact operator, (ii) $P_Z \to I$ pointwise on $\mathcal{R}(T^\star)$ as $d_Z \to \infty$ and (iii) $Q_W \to I$ pointwise on $\mathcal{R}(T)$ as $d_W \to \infty$. In particular, it is not difficult to derive from the singular value decomposition that the following two inequalities hold:

$$\|T(I - P_Z)\| \geqslant \lambda_{d_Z}$$

and

$$\|(I - Q_W)T\| \geqslant \lambda_{d_W}.$$

Therefore, whatever the two bases are, the best approximation of T cannot perform better than the rate of the singular values λ_{d_Z} and λ_{d_W}. Note that the equality holds when the bases are chosen to be the eigenfunctions $\{u_k; k = 1, ..., d_Z\}$ and $\{v_k; k = 1, ..., d_W\}$, but this situation is not useful in our context since T is unknown. $\square$

EXAMPLE 3.1. To simplify this example, assume that Z and W are uniformly distributed over $[0, 1]$. Let $s_0, \ldots, s_{d_Z}$ denote an equidistant grid on $[a, b]$, that is $s_k = a + k\tau$ with $\tau = (b - a)/d_Z$. Define $\phi_k := \mathbb{I}_{[s_{k-1}, s_k]}$ for $k = 1, \ldots, d_Z$ as a basis of Φ_{d_Z}. Analogously, define a grid on $[c, d]$ with $t_k = a + k\tau$ with $\tau = (d - c)/d_W$ and $\Psi_k := \mathbb{I}_{[t_{k-1}, t_k]}$ for $k = 1, \ldots, d_W$. For this example with a sufficiently smooth joint density f_{ZW}, we can write (cf. Plato & Vainikko (1990))

$$\begin{aligned}
\|T(I - P_Z)\| &\leqslant c_1 \cdot (b - a)/d_Z, \quad c_1 = \frac{2}{3}\Big(\int_a^b \int_c^d \Big|\frac{\partial f_{ZW}(z, w)}{\partial z}\Big|^2 dwdz\Big)^{1/2}, \\
\|(I - Q_W)T\| &\leqslant c_2 \cdot (d - c)/d_W, \quad c_2 = \frac{2}{3}\Big(\int_a^b \int_c^d \Big|\frac{\partial f_{ZW}(z, w)}{\partial w}\Big|^2 dwdz\Big)^{1/2}.
\end{aligned}$$

With additional regularity assumptions on f_{ZW}, it is useful to consider more regular basis functions, such as higher order B-splines or wavelets for instance. This would lead, for example, to upper bound for $\|T(I - P_Z)\|$ of the type $((b-a)/d_Z)^\eta$. The exact value of η depends on the basis system and the smoothness of f_{ZW} and is derived from typical inequalities in approximation theory (cf. DeVore & Lorentz (1993)). $\square$

Example 3.2. An interesting example for Φ_{d_Z} is given by the basis of orthonormal wavelets. The theory of wavelets offers an appealing alternative to the Fourier analysis. A wavelet system is an orthogonal basis of $L^2(\mathbb{R}^q)$ which, in contrast to the Fourier basis, contains functions that are well localized both in the time and the frequency domain. As a consequence, they appear more appropriate to decompose functions φ that have a more irregular behavior, such as jumps or peaks. For a general introduction to this theory, we refer to Vidakovic (1999).

Suppose that Z and W are uniformly distributed over $[0,1]$. Let (ϕ, ψ) be some scaling and mother functions over $[0,1]$ and assume that ψ has κ continuous derivatives and κ vanishing moments. Let j be a negative integer. The space Φ_{d_Z} in this example is the linear space that is spanned by $\{\phi_{jk}\}_{0 \leqslant k < 2^{-j}-1}$, with $\phi_{jk}(\cdot) := 2^{-j/2}\phi(2^{-j} \cdot - k)$. Denote by P_j the orthogonal projection onto this space. For any function g belonging to the Sobolev space $W^s[0,1]$ for $0 < s < \kappa$, one can show $\|(I - P_j)g\|^2 = o(2^{2sj})$ (e.g. Mallat (1997, Theorem 9.4)). Therefore, some calcuation show that

$$\|T(I - P_j)\| \leqslant C 2^{2sj} \int \|f(\cdot, w)/\sqrt{r(w)}\|^2_{W^s} dz$$

provided that $[f(\cdot, w)/\sqrt{r(w)}] \in W^s$ and that the integral exists. A similar bound can be derived for $\|(I - Q)T\|$. $\square$

We are now in position to describe the projection step of the estimation. Find a solution $\varphi^\circ \in \Phi_{d_Z}$ of the system of equations

$$\langle T\varphi^\circ, \psi_j \rangle = \langle r, \psi_j \rangle \qquad \text{for all } j = 1 \dots d_W.$$

This system is a discretization of the problem (2.1). Since $\varphi^\circ \in \Phi_{d_Z}$, we can write $\varphi^\circ = \sum_{j=1}^{d_Z} a_j^\circ \phi_j$ which, by linearity of the operator T, leads to the equivalent system of equations

$$\tilde{M}_d a^\circ = \tilde{v}_d \tag{3.2}$$

where $a^\circ = (a_1^\circ, \dots, a_{d_Z}^\circ)'$ is the vector of parameters, $\tilde{M}_d$ is the $d_W \times d_Z$ matrix with element (i,j) equal to $\langle T\phi_i, \psi_j \rangle$ and $\tilde{v}_d$ is the column vector $(\langle r, \psi_1 \rangle, \dots, \langle r, \psi_{d_W} \rangle)'$. The inversion of the system (3.3) however leads to two important issues. The first issue is that the basis systems $\{\phi_1, \dots, \phi_{d_Z}\}$ and $\{\psi_1, \dots, \psi_{d_W}\}$ are not orthonormal. Thus the inversion of (3.3) involves the Gram matrices

$$G_\phi := (\langle \phi_i, \phi_j \rangle)_{i,j=1,\dots,d_Z}$$

and

$$G_\psi := (\langle \psi_i, \psi_j \rangle)_{i,j=1,\dots,d_W}.$$

These two matrices reduce to the identity matrix when the basis systems are orthonormal. With this correction, (3.2) becomes

$$M_d a^\circ = v_d \tag{3.3}$$

where $M_d = G_\psi^{-1/2} \tilde{M}_d G_\phi^{-1/2}$ and $v_d = G_\psi^{-1/2} \tilde{v}_d$.

The second issue is the stability of the inversion. Because we are solving an integral equation, the problem (2.1) is in general ill-posed. This implies that the matrix M_d in (3.3) is ill-conditionned and thus its inversion is numerically unstable. In particular, it implies that, even if we find a consistent estimator for M_d and v_d, the estimation of a° resulting from the inversion of (3.3) has a very slow rate of convergence in general. For this reason, we need to stabilize (regularize) the inversion in order to recover faster rates of convergence in the estimation of φ. Below we propose an iterative method for this issue. However, before presenting this method, we first introduce consistent estimators of M_d and v_d.

3.2 Estimation of M_d and v_d

Let $\{(Y_l, Z_l, W_l); l = 1, 2, \dots, n\}$ be an independent and identically distributed sample from (Y, Z, W). Let $\boldsymbol{\phi}(\cdot) = (\phi_1(\cdot), \dots, \phi_{d_Z}(\cdot))'$ and $\boldsymbol{\psi}(\cdot) = (\psi_1(\cdot), \dots, \psi_{d_W}(\cdot))'$. The estimators of M_d and v_d are respectively given by

$$\widehat{M}_d = \frac{1}{n} \sum_{l=1}^{n} G_\phi^{-1/2} \boldsymbol{\phi}(Z_l) \boldsymbol{\psi}(W_l)' G_\psi^{-1/2} \tag{3.4a}$$

and

$$\widehat{v}_d = \frac{1}{n} \sum_{l=1}^{n} Y_l G_\psi^{-1/2} \boldsymbol{\psi}(W_l) \, . \tag{3.4b}$$

We give below some asymptotic properties for these two estimators that will be useful to derive the final rates of convergence. The result is valid under the following assumption.

Assumption 3.1. *Denote $\tilde{\boldsymbol{\phi}} = G_\phi^{-1/2} \boldsymbol{\phi}$ and $\tilde{\boldsymbol{\psi}} = G_\psi^{-1/2} \boldsymbol{\psi}$. The vectors $\tilde{\boldsymbol{\phi}}$ and $\tilde{\boldsymbol{\psi}}$ are the orthogonalization of $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ by the Gram matrices. We assume: $\|\tilde{\phi}_i\|_{L_Z^\zeta}^\zeta \leqslant c \cdot d_Z^{(\zeta/2)-1}$ and $\|\tilde{\psi}_i\|_{L_W^\zeta}^\zeta \leqslant c \cdot d_W^{(\zeta/2)-1}$ for all $\zeta > 2$.*

It is not difficult to check that Assumption 3.1 holds true for the basis introduced in the two examples.

Proposition 3.1. *The estimators $\widehat{M}_d$, resp $\widehat{v}_d$, are unbiased for M_d, resp. v_d. Suppose that $\mathbb{E}Y^2 < \infty$. Then, under Assumption 3.1 it holds*[1]

$$\mathbb{E}\|\widehat{v}_d - v_d\|_2^2 \lesssim \frac{d_W}{n}, \tag{3.5}$$

$$\mathbb{E}\|\widehat{M}_d - M_d\|_2^\zeta \lesssim \frac{(d_W d_Z)^{\zeta/2}}{n^{\zeta/2}} \quad \textit{for all } \zeta \in (0,2], \tag{3.6}$$

$$\mathbb{E}\|\widehat{M}_d - M_d\|_2^\zeta \lesssim \frac{(d_W d_Z)^{\zeta-1}}{n^{\zeta/2}} \quad \textit{for all } \zeta \geqslant 2, \tag{3.7}$$

where $\|\cdot\|_2$ is the ℓ^2 norm of a matrix.

Proof. The unbiasedness of $\widehat{M}_d$ and $\widehat{v}_d$ is a straightforward result. The proof of (3.5) is similar to the proof of (3.6), therefore we skip it.

Application of Lyapunov's inequality leads to $\mathbb{E}\|\widehat{M}_d - M_d\|_2^\zeta \leqslant (\mathbb{E}\|\widehat{M}_d - M_d\|_2^2)^{\zeta/2}$ for all $\zeta \in (0,2]$. Moreover, if $M_{d;ij}$ denotes the element (i,j) of the matrix M_d, we have $\mathbb{E}\|\widehat{M}_d - M_d\|_2^2 \leqslant \sum_{i,j} \mathbb{V}\mathrm{ar}\, \widehat{M}_{d;ij}$ because $\widehat{M}_d$ is an unbiased estimator of M_d. Then we can write, by independence of the sample,

$$\mathbb{V}\mathrm{ar}\, \widehat{M}_{d;ij} \leqslant \frac{1}{n}\mathbb{E}\{\tilde{\phi}_i(Z)^2 \tilde{\psi}_j(W)^2\} \tag{3.8}$$

This last expression is finite as the functions ψ_j, ϕ_i are such that $\|\tilde{\psi}_j\|_{L^2_W} = \|\tilde{\phi}_i\|_{L^2_Z} = 1$ for all i,j. The inequality (3.6) follows by noticing that the sum over i,j contains $d_W d_Z$ elements.

We now prove (3.7). Using Jensen's inequality

$$\begin{aligned}\mathbb{E}\|\widehat{M}_d - M_d\|_2^\zeta &= (d_W d_Z)^{\zeta/2}\mathbb{E}\left\{\frac{1}{d_W d_Z}\sum_{i,j}(M_{d;ij} - \widehat{M}_{d;ij})^2\right\}^{\zeta/2} \\ &\leqslant (d_W d_Z)^{(\zeta/2)-1}\sum_{i,j}\mathbb{E}(M_{d;ij} - \widehat{M}_{d;ij})^\zeta \,.\end{aligned}$$

To simplify notations, write $\widehat{M}_{d;ij} = n^{-1}\sum_l A_{ij;l}$ where $A_{ij;l} := \tilde{\phi}_i(Z_l)\tilde{\psi}_j(W_l)$. By the inequality of Minkowski, it holds $\mathbb{E}|\sum_i X_i|^p \leqslant \{\sum_i (\mathbb{E}|X_i|^p)^{2/p}\}^{p/2}$ and we can then write

$$\begin{aligned}\mathbb{E}(M_{d;ij} - \widehat{M}_{d;ij})^\zeta &\leqslant n^{-\zeta}\{\sum_l (\mathbb{E}|A_{ij;l} - \mathbb{E}A_{ij;l}|^\zeta)^{2/\zeta}\}^{\zeta/2} \\ &\lesssim n^{-\zeta}\{\sum_l (\mathbb{E}A_{ij;l}^\zeta)^{2/\zeta}\}^{\zeta/2} \end{aligned}\tag{3.9}$$

using again Jensen's inequality. As above when we derived the upper bound of (3.8), we note that $\mathbb{E}A_{ij;l}^\zeta \lesssim \|\tilde{\psi}_j\|_{L_W^\zeta}^\zeta \cdot \|\tilde{\phi}_i\|_{L_Z^\zeta}^\zeta$. Therefore, with Assumption 3.1, (3.9) is bounded by $n^{-\zeta/2}(d_Z d_W)^{(\zeta/2)-1}$ and the result follows. □

[1] We write $A \lesssim B$ if there exists a positive constant c such that $A \leqslant cB$.

3.3 Regularization iterative step

We now present the iteration procedure used in order to stabilize the inversion of the system (3.3). It is called the *Landweber-Fridman* in the numerical literature [e.g. Engl et al. (2000)]. It is of course possible to define other regularization schemes at this stage, among which is Tikhonov regularization. The Landweber-Fridman has the advantage to be numerically very simple to implement when it is applied to the projection estimator.

The vector a° in the system (3.3) is estimated by the following way:

$$\widehat{a}_0^\circ = 0$$
$$\widehat{a}_{k+1}^\circ = \widehat{a}_k^\circ - \frac{1}{\mu^2}\widehat{M}_d'\left(\widehat{M}_d\hat{a}_k^\circ - \widehat{v}_d\right) \qquad k = 0, 1, \ldots, K-1 \tag{3.10}$$

The estimator of φ° then follows by

$$\widehat{\varphi}^\circ := \widehat{a}_K^{\circ\prime} G_\phi^{-1/2} \boldsymbol{\phi} \,. \tag{3.11}$$

The presence of the parameter μ in the iterative scheme is only technical. The convergence results established below need that the norm of the matrix used in the algorithm must be less than 1. The parameter μ then normalizes the problem such that this constraint is fulfilled. In practice and in the proof of our results, we use the random bound $\mu > \max(1, \|\widehat{M}_d\|)$.

One crucial question is to decide on the number of iterations K. It is known that a too small value of K provides unsufficient regularization, whereas a too large value of K leads to a too large regularization bias. The theoretical sections below and the empirical study provide a guidance for the choice of this regularization parameter.

4 Convergence under mild ill-posedness

In order to derive a rate of convergence for our estimator we need to specify regularity assumptions for the unknown solution φ. One convenient approach is to relate the regularity of φ to the behavior of the operator T itself. The idea is that, if φ is well adapted to the operator, then the estimation should be easier, thus the rates of convergence should be faster. The meaning of how "well-adapted" is the solution to the operator is characterized by the so-called *source condition* that we define now.

The natural operator of interest in order to define our source condition is $T^\star T$, which is by construction self-adjoint, non-negative and such that

$$T^\star T g = \sum_{k\in\mathbb{N}} \lambda_k^2 \langle g, u_k\rangle u_k \qquad \text{for all } g \in L_Z^2$$

by definition of the singular value decomposition of T.

In the following it is useful to define what is a function of $T^\star T$. Consider a function ℓ that is defined on the real line. The operator $\ell(T^\star T)$ is defined through its spectral decomposition:

$$\ell(T^\star T)g := \sum_{k\in\mathbb{N}} \ell(\lambda_k^2)\langle g, u_k\rangle u_k \tag{4.1}$$

for all $g \in L_Z^2$.

The regularity assumption imposed on the solution φ is defined next.

Assumption 4.1 (Strong source condition). *The operator T and the solution φ are such that there exists $\beta > 0$ and $\psi \in L_Z^2$ with $\varphi = (T^\star T)^{\beta/2}\psi$ and $\|\psi\| \leqslant \rho$.*

To understand this condition it is convenient to note that it is equivalent to require that the solution φ is such that $(T^\star T)^{-\beta/2}\varphi$ belongs to L_Z^2. Using the singular value decomposition of T and the representation given in (4.1) it is therefore equivalent to assume

$$\sum_{k=1}^{\infty} \frac{(\langle \varphi, u_k\rangle)^2}{\lambda_k^\beta} < \infty\ .$$

Because the eigenvalues λ_k^β tend to zero, the index β that appears in the source condition is one measure of the ill-posedness of the problem.

Before stating the convergence result, we also formalize the condition (3.1) on the operators P_Z, Q_W and T in the following assumption.

Assumption 4.2. *The projection operators P_Z, Q_W and T are such that $\|T(I - P_Z)\| \leqslant \delta_Z$ and $\|(I - Q_W)T\| \leqslant \delta_W$, where δ_Z, resp. δ_W, denote two sequences vanishing as d_Z, resp. d_W, growth.*

Theorem 4.1. *Consider the estimator (3.11) constructed using the projections P_Z and Q_W and set $\mu = C\max(1, \|\hat{M}_d\|)$ for some constant $C > 1$. Suppose that the "strong source condition" (Assumption 4.1) is satisfied, that $\mathbb{E}Y^2 < \infty$ and Assumptions 3.1 and 4.2 hold true. Assume that the number of iterations K is such that*

$$K = \left(\delta_Z^{2+2(1\wedge\beta)} + \frac{d_W(d_Z+1)}{n}\right)^{-\frac{1}{\beta+1}}. \tag{4.2}$$

Then the L^2 risk of the proposed estimator $\widehat{\varphi}^\circ$ has the rate

$$\mathbb{E}\|\hat{\varphi}^\circ - \varphi\|^2 = O\left(\left(\frac{d_Z d_W}{n}\right)^{\frac{\beta}{\beta+1}} + \delta_Z^{2(\beta\wedge 1)} + \delta_W^{2(\beta\wedge 1)}\right).$$

Proof. An upper bound for the mean square error of $\widehat{\varphi}^\circ$ under the strong source condition is derived in Lemma B.1 in the technical Appendix, and is given by

$$\mathbb{E}\|\hat{\varphi}^\circ - \varphi\|^2 \lesssim \{1 + \mathbb{E}\|\widehat{M}_d - M_d\|_2^{2\beta}\}K^{-\beta} + K\left(\mathbb{E}\|\widehat{M}_d - M_d\|_2^2 + \mathbb{E}\|\widehat{v}_d - v_d\|_2^2 + \delta_Z^{2+2(\beta\wedge 1)}\right)$$
$$+ \mathbb{E}\|\widehat{M}_d - M_d\|_2^{2(\beta\wedge 1)} + \mathbb{E}\|\widehat{M}_d - M_d\|_2^{2\beta} + \delta_Z^{2(\beta\wedge 1)} + \delta_W^{2(\beta\wedge 1)}.$$

First we note that $1+\mathbb{E}\|\widehat{M}_d - M_d\|_2^{2\beta} \lesssim 1$ by Proposition 3.1. We find the optimal number of iterations K by balancing the first two terms, which gives:

$$K \sim \left\{\mathbb{E}\|\widehat{M}_d - M_d\|^2 + \mathbb{E}\|\widehat{r}_d - r_d\|^2 + \delta_Z^{2+2(\beta\wedge 1)}\right\}^{-\frac{1}{\beta+1}}.$$

The above Proposition 3.1 gives the rate of convergence for $\mathbb{E}\|\widehat{M}_d - M_d\|^2$ and $\mathbb{E}\|\widehat{r}_d - r_d\|^2$, and they lead to the optimal rate (4.2) given in the statement of the theorem. We plug in the optimal rate for K in the MSE of $\widehat{\varphi}^\circ$, and we consider the leading terms we get

$$\mathbb{E}\|\widehat{\varphi}^\circ - \varphi\|^2 \lesssim \left(\frac{d_Z(d_W+1)}{n}\right)^{\frac{\beta}{\beta+1}} + \left(\frac{d_Z d_W}{n}\right)^{\beta\wedge 1} + \delta_Z^{\frac{2\beta(1+\beta\wedge 1)}{\beta+1}} + \delta_Z^{2(\beta\wedge 1)} + \delta_W^{2(\beta\wedge 1)}.$$

The result follows by considering the leading terms and using the inequalities $\beta/(\beta+1) \leqslant \beta\wedge 1 \leqslant (1+\beta\wedge 1)\beta/(1+\beta)$ that hold for all $\beta > 0$. □

We comment this result in the following remarks.

Remark 4.1.

1. The result is presented under mild conditions on the bases used for the projection. Now impose $m := d_Z = d_W$ and $\delta_Z = \delta_W = m^{-4\beta}$. If $\beta \leqslant 1$ then the rate of convergence of the risk reduces to $n^{\frac{-\beta}{\beta+3/2}}$ which is known to be optimal in the class of solutions that satisfy the strong source condition [Johannes et al. (2010, Proposition 4.1 with $s=0$, $a=1$ and $p=\beta$)].

2. One interesting example is given when the projection basis is given by the singular value decomposition of T. We have already argued that this case is not realistic since the eigenfunctions are unknown, but it is at least of a theoretical interest. From Remark 3.1 it follows that $\delta_Z = \lambda_{d_Z}$ and $\delta_W = \lambda_{d_W}$. We may also impose that the eigenvalues are decreasing at a polynomial rate, i.e. $\lambda_d = d^{-\varepsilon}$ for some $\varepsilon > 0$. Then the rate is given by $n^{\frac{-2\beta}{2\beta+2+1/\varepsilon}}$. This particular setting has been considered for the study of other regularization methods in Hall & Horowitz (2005). This rate is known to be optimal for mildly ill-posed inverse problems over the space of functions φ that satisfy the source condition [e.g. Chen & Reiß (2010)].

3. The discontinuity $(\beta\wedge 1)$ on the range of the exponent of δ_Z and δ_W implies that the rate is no longer optimal for $\beta > 1$. This limitation is not technical, but it is intrinsic to the Landweber-Fridman method (the mathematical explanation is given by the analogous limitation in Lemma B.1 in the Appendix below). A similar phenomenon has been observed in Tautenhahn (1996) in a purely deterministic setting. In Tautenhahn (1996) a so-called "preconditionning" treatment has been proposed to improve the rate when $\beta > 1$. We conjecture that a similar solution would lead to the same improvement of our result.

5 Convergence under severe ill-posedness

There are a number of important situations where the strong source condition is a too restrictive assumption. A prominent example is given by random variables Z and W that are normally distributed. In that situation one can show that the eigenvalues of the conditional expectation operator T are exponentially decreasing, i.e. λ_k behaves like $\exp(-k^\varepsilon)$ for some $\varepsilon > 0$. Under this setting, the functions satisfying Assumption 4.1 for an arbitrary $\beta > 0$ would be very limitated. Indeed, Assumption 4.1 would imply that the solution φ has an infinite number of derivatives (i.e. φ is an analytic function). This example show that the strong source condition may be restrictive.

We can define a weaker condition than Assumption 4.1 if we consider the function ℓ in the representation (4.1) to be logarithmic. This case has been considered in the deterministic setting [e.g. Hohage (1997); Nair et al. (2005)]. Surprisingly, it has been less studied in the context where the function r and the operator T have to be estimated [see Chen & Reiß (2010) for a related condition].

Assumption 5.1 (Weak source condition). *There exists $\psi \in L^2_Z$ such that*

$$\varphi = \left\{ -\log\left(\frac{T^\star T}{2}\right)\right\}^{-\beta/2} \psi, \quad \|\psi\| \leqslant \rho \text{ and } \beta > 0, \tag{5.1}$$

where ρ is sufficiently small.

Note that this assumption is well defined since T and $T^\star$ are conditional expectation operators and therefore they are projections and such that $\|T^\star T\| = 1$. The following theorem gives the asymptotic risk under the weak source condition.

Theorem 5.1. *Consider the estimator (3.11) constructed using the projections P_Z and Q_W and set $\mu = C\max(1, \|\hat{M}_d\|)$ for some constant $C > 1$. Suppose the weak source condition (Assumption 5.1) is satisfied and d_Z, d_W are such that $(d_Z d_W)/n^2 = O(1)$. Suppose that $\mathbb{E}Y^2 < \infty$ and Assumptions 3.1 and 4.2 hold true. If the stopping index K is chosen by*

$$K = \left\{\frac{d_Z d_W}{n} + \delta_Z^2 + \delta_W^2\right\}^{-1/2} \tag{5.2}$$

then we have

$$\mathbb{E}\|\hat{\varphi}^\circ - \varphi\|^2 \lesssim \left\{\log\left(\frac{d_W d_Z}{n} + \delta_Z^2 + \delta_W^2\right)\right\}^{-\beta}.$$

Proof. In Lemma C.1 of the technical Appendix, we have derived the mean square error of $\widehat{\varphi}^\circ$ under the weak source condition. This lemma together with Lemma A.2 gives

$$\mathbb{E}\|\hat{\varphi}^\circ - \varphi\|^2 \lesssim K\{\mathbb{E}\|\widehat{M}_d - M_d\|_2^2 + \mathbb{E}\|\widehat{v}_d - v_d\|_2^2 + \delta_Z^2 + \delta_W^2\} + 2(\log K)^{-\beta} \tag{5.3}$$

provided that K, d_W and d_Z are such that

$$K^2\mathbb{E}\|M'M - \widehat{M}_d'\widehat{M}_d\|_2^2 = O(1)\,. \tag{5.4}$$

Proposition 3.1 applied to the rate (5.3) leads to $\mathbb{E}\|\hat{\varphi}^{\circ} - \varphi\|^2 \lesssim K\{d_Z(d_W+1)/n + \delta_Z^2 + \delta_W^2\} + 2(\log K)^{-\beta}$. With K such that (5.2) holds, and considering the main terms, then the mean square rate of convergence follows. It remains to check if the constraint (5.4) is fulfilled. The norm (5.4) can be decomposed into three terms:

$$K^2\mathbb{E}\|M'M - \widehat{M}_d'\widehat{M}_d\|_2^2 \lesssim K^2\left\{\mathbb{E}\|M'(M-\widehat{M})\|_2^2 + \mathbb{E}\|(M'-\widehat{M}_d')(\widehat{M}_d - M_d)\|_2^2 \right.$$
$$\left. + \mathbb{E}\|M_d(M' - \widehat{M}_d')\|_2^2\right\}$$

Using Proposition 3.1, the first and the second term are bounded by $K^2(\delta_W^2 + \delta_Z^2 + (d_W d_Z)/n)$. Using the Cauchy-Schwarz inequality and Proposition 3.1 with $\zeta = 4$, the second term is bounded up to a constant by $K^2(\delta_W^2+\delta_Z^2)(d_W d_Z)^{3/2}/n + K^2(d_W d_Z)^3/n^2$. Therefore, with the choice of K given in (5.2), it is sufficient to satisfy the constraint

$$\frac{\frac{(d_W d_Z)^3}{n^2} + \frac{d_W d_Z}{n} + (\delta_W^2 + \delta_Z^2)\left(1 + \frac{(d_W d_Z)^{3/2}}{n}\right)}{\frac{d_Z d_W}{n} + \delta_Z^2 + \delta_W^2} = O(1).$$

The last constraint is satisfied under the condition that $(d_W d_Z)^2/n$ is finite. □

Remark 5.1.
1. The optimal number K of iterations found in this result appears to be independent from the β, that is it is independent from the level of regularity of φ given by the weak source condition.
2. Suppose we take the same number of basis functions $m := d_W = d_Z$ in both Hilbert spaces. Suppose also that the basis is such that $\delta_Z = \delta_W = \exp(-m^{2\varepsilon})$ from some positive number ε. Then if we take e.g. $m = n^{1/4}$ the final rate of convergence is $\{\log(n)\}^{-\beta}$. When T and r which are known and deterministic, this rate is known to be the optimal rate of convergence over the solutions that satisfy the weak source condition [Hohage (2000)].

6 Finite sample study

We present here the results of a Monte-Carlo study that aims to study the finite sample properties of the suggested method. The function φ is this study is designed as $\varphi(z) = (0.2+z)1_{[0,0.6]}(z) + (0.8 - 0.5(z-0.6))1_{]0.6,1]}(z)$, where $1_A(z)$ is the indicator function that is equal to 1 if $z \in A$ and 0 otherwise. The true function is continuous but it contains an elbow that has point at which the function is not differentiable. Data are generated from the model $Y = \varphi(Z) + U$ with $U \sim N(0;0.3)$ and Z is the restriction to the interval $[0,1]$ of $Z = 1 - 3W - 3W^2 + 5U + V$ with $V \sim N(0;0.1)$ and $W \sim N(0;0.1)$.

The function φ is displayed in Figure 1 (solid line) together with a generated sample of $n = 500$ points. The cloud of sample points is not exactly "centered" around the function φ, as it can be expected since the variable Z is correlated with the model error U. It is

		μ									
		1.1	1.2	1.3	1.4	1.5	2	2.5	3	3.5	4
K	1	2.14	1.37	0.92	0.68	**0.55**	0.50	0.66	0.79	0.88	0.95
	2	1.12	0.70	0.65	0.65	0.63	**0.48**	0.49	0.59	0.69	0.77
	3	2.28	1.24	0.95	0.83	0.75	0.55	**0.47**	0.50	0.57	0.65
	4	1.96	1.43	1.22	1.05	0.93	0.63	0.50	**0.48**	0.51	0.57
	5	3.30	2.03	1.60	1.33	1.14	0.72	0.55	**0.48**	**0.48**	0.52
	6	3.48	2.57	2.04	1.67	1.40	0.81	0.60	0.51	**0.48**	0.50

Table 1: Average of mean square errors at the scale 10^{-2} between the approximated solution $P_Z\varphi$ and the nonparametric estimator by iterative regularization. K is the number of iterations and μ is the rescaling parameter appearing in (3.10).

precisely the information given by the instrumental variable W that allows to correct for this endogeneity issue.

The basis systems that we consider in this study is given by Haar basis. This orthonormal basis is a wavelet system generated from the scaling function $\phi(z) = 1_{[0,1]}(z)$ and mother function $\psi(z) = 1_{[0,0.5[}(z) - 1_{[0.5,1[}(z)$ (see Example 3.2). Because the random variables are Normaly distributed, we see that the basis systems are not necessarily well adapted to the conditional expectation operator of this simulation. As for the dimension of the basis systems, we consider $d_Z = d_W = 8$ other the whole study.

The classical ordinary least square estimator of φ in this basis is the estimator that is estimating the 8 coefficients by OLS. This estimator is drawn in Figure 1 (dash-dotted line). This estimator is of course biased because of the endogeneity. It is considered as the initial estimator in the iterative regularization system (3.10). Figure 1 also shows the result of the iterative regularization after 3 iterations (dashed line). For the sake of comparison, the approximated solution in the Haar basis, $P_Z\varphi$, is also displayed on the picture (dotted line). The picture shows the correction by the iteration is more effective where the cloud of points is far from the true regression function.

A systematic Monte Carlo study has been performed in this setting with 2000 replications. In particular, we want to illustrate the sensitivity of K and μ on the resulting estimator. Table 1 shows the result of the simulations for a range of K going from 1 to 6, and a range of μ going from 1.1 to 4. The table gives the average of mean square errors at the scale 10^{-2} between the approximated solution $P_Z\varphi$ and the nonparametric estimator by iterative regularization. For $\mu = 1.5$ and higher values, the results are not very sensitive to this parameter. We also notice that it is not necessary to perform a high number of iterations in order to correct for endogeneity and to regularize the estimator.

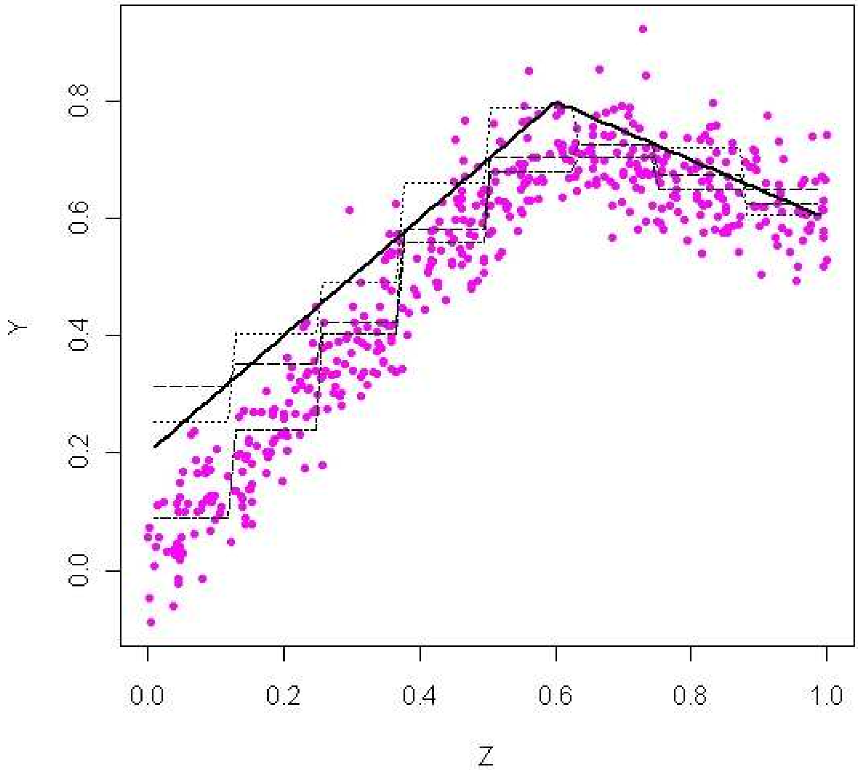

Figure 1: The solid line is true function φ and the pink points is an observed sample of size $n = 500$. The approximated solution $P_Z\varphi$ in the Haar basis is the dotted line. A standard OLS estimator of the coefficients of the projection leads to the dash-dotted line. This estimator is the initial step of the iterative algorithm. After $K = 3$ steps, the regularized estimator gives the dahsed line.

APPENDIX

A A functional definition of the estimator

In this preliminary section, we derive another writing of the estimator $\hat{\varphi}^{\circ}$. This equivalent definition relates the estimator to an empirical version of the function r and operators T and T^* defined in (2.1).

First, we introduce column vectors $\tilde{\boldsymbol{\phi}} = G_{\phi}^{-1/2}\boldsymbol{\phi}$ and $\tilde{\boldsymbol{\psi}} = G_{\psi}^{-1/2}\boldsymbol{\psi}$. These vectors are the orthogonalization of $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ by the Gram matrices. We then define $\widehat{T}_d\phi(\cdot) := \tilde{\boldsymbol{\psi}}(\cdot)'\widehat{M}_d\langle\phi,\tilde{\boldsymbol{\phi}}\rangle$ where $\langle\phi,\tilde{\boldsymbol{\phi}}\rangle$ denotes the column vector $(\langle\phi,\tilde{\phi}_1\rangle,\ldots,\langle\phi,\tilde{\phi}_{d_Z}\rangle)'$. The dual of $\widehat{T}$ is $\widehat{T}^{\star}_d\psi(\cdot) := \tilde{\boldsymbol{\phi}}(\cdot)'\widehat{M}'_d\langle\psi,\tilde{\boldsymbol{\psi}}\rangle$ where $\langle\psi,\tilde{\boldsymbol{\psi}}\rangle$ analogously denotes the column vector $(\langle\psi,\tilde{\psi}_1\rangle,\ldots,\langle\psi,\tilde{\psi}_{d_W}\rangle)'$. Finally, we define $\widehat{r}_d(\cdot) = \tilde{\boldsymbol{\psi}}(\cdot)'\widehat{v}_d$. A convenient way to write these estimators is to consider the functions

$$c : L^2_{\pi} \to \mathbb{R}^{d_Z} : \phi \mapsto \langle\phi,\tilde{\boldsymbol{\phi}}\rangle, \qquad c^* : \mathbb{R}^{d_Z} \to L^2_{\pi} : \theta \mapsto \sum_j \theta_j\tilde{\phi}_j$$

and

$$b : L^2_{\tau} \to \mathbb{R}^{d_W} : \psi \mapsto \langle\psi,\tilde{\boldsymbol{\psi}}\rangle, \qquad b^* : \mathbb{R}^{d_W} \to L^2_{\tau} : \theta \mapsto \sum_j \theta_j\tilde{\psi}_j \ .$$

With these notations, we find that $\widehat{T}_d = b^*\widehat{M}_d c$, $\widehat{T}^*_d = c^*\widehat{M}'_d b$ and $\widehat{r}_d = b^*\widehat{v}_d$. Moreover, $\widehat{T}^*_d\widehat{T}_d = c^*\widehat{M}'_d\widehat{M}_d c$ and $\widehat{T}^*_d\hat{r}_d = c^*\widehat{M}'_d\hat{v}_d$.

Now recall that the vector $\hat{a}^{\circ}_{k+1}$ was recursively defined by (3.10). From this definition, we can write $\hat{a}^{\circ}_{k+1} = R^{\mu}_{k+1}(\widehat{M}'_d\widehat{M}_d)\widehat{M}'_d\hat{v}_d$ where

$$R^{\mu}_{k+1}(A) = \frac{1}{\mu^2}\sum_{j=0}^{k}\left(I - \frac{1}{\mu^2}A\right)^j \ . \tag{A.1}$$

The final estimator (3.11) is defined as $\hat{\varphi}^{\circ} = \hat{a}^{\circ\prime}_K\tilde{\boldsymbol{\phi}}$, and the next lemma presents an equivalent definition of the estimator.

Lemma A.1. *An equivalent definition of the estimator $\hat{\varphi}^{\circ}$ is given by*

$$\hat{\varphi}^{\circ} = R^{\mu}_K(\widehat{T}^*_d\widehat{T}_d)\widehat{T}^*_d\hat{r}_d \ . \tag{A.2}$$

Proof. Apply c^* to both side of (3.10) and denote $\hat{\varphi}_k := c^*\hat{a}^{\circ}_k = \hat{a}^{\circ\prime}_k\tilde{\boldsymbol{\phi}}$. Using $cc^* = bb^* = I$, we can write $\hat{\varphi}_{k+1} = \hat{\varphi}_k - \mu^{-2}c^*\widehat{M}'_d b(b^*\widehat{M}_d c\hat{\varphi}_k - b^*\widehat{v}_d)$. By definition of $\widehat{T}_d$, $\widehat{T}^*_d$ and $\hat{r}_d$, this equation writes $\hat{\varphi}_{k+1} = \hat{\varphi}_k - \mu^{-2}\widehat{T}^*_d(\widehat{T}_d\hat{\varphi}_k - \widehat{r}_d)$. Similarly to what we argue above, this recursive formula for $\hat{\varphi}_{k+1}$ implies $\hat{\varphi}_{k+1} = R^{\mu}_{k+1}(\widehat{T}^*_d\widehat{T}_d)\widehat{T}^*_d\hat{r}_d$. This proves the result, with $\hat{\varphi}^{\circ} = \hat{\varphi}_K = \hat{a}^{\circ\prime}_K\tilde{\boldsymbol{\phi}}$. $\square$

Below we derive risk bounds for $\hat{\varphi}_K$ in terms of the operator norms $\|\hat{T}_d - T\|$ and $\|\hat{r}_d - r_d\|$. The next lemma relates these norms to the norm between matrices M_d and $\hat{M}_d$.

Lemma A.2. *Consider the estimator (3.11) constructed using the projections P_Z and Q_W. Then $M_d = Q_W T P_Z$ and*

$$\|\widehat{T}_d - T\| \leqslant \|\widehat{M}_d - M_d\|_2 + \|T(I - P_Z)\| + \|(I - Q_W)T\| \ .$$

where $\|\cdots\|_2$ is the ℓ^2 norm between matrices, and $\|\cdots\|$ is the spectral norm.

B Risk under the strong source condition assumption

The proof of the main results makes use of known results in functional analysis. It is convenient to summarize in a lemma the results we shall use.

Lemma B.1. *Let G and H be real Hilbert spaces, and $A, B : G \to H$ be linear, bounded operators with $\|A\|, \|B\| \leqslant 1$. Let $P : G \to G$ and $Q : H \to H$ be two orthogonal projections. Then, for all $\beta > 0$,*

$$\|(I - P)(A^*A)^{\beta/2}\| \leqslant \|A(I - P)\|^{\beta \wedge 1} \tag{B.1}$$

$$\|P(A^*A)^{\beta/2} - (P^*A^*Q^*QAP)^{\beta/2}\| \leqslant C_\beta \left(\|A(I - P)\|^{\beta\wedge 1} + \|(I - Q)A\|^{\beta\wedge 2}\right) \tag{B.2}$$

where C_β is a generic factor depending on β only, and for all $\beta > 0$, $\beta \neq 1$,

$$\|(A^*A)^{\beta/2} - (B^*B)^{\beta/2}\| \leqslant C_\beta \|A - B\|^{\beta\wedge 1}. \tag{B.3}$$

A proof of (B.1) can be found in Plato (1990), (B.2) is Lemma 4.4 of Plato & Vainikko (1990) and (B.3) is Lemma 3.2 of Egger (2005).

The following lemma is the key result from which we derive the results of Section 4. It gives an explicit bound for the loss of the proposed estimator $\hat{\varphi}^\circ$.

Lemma B.2. *Consider the estimator (3.11) constructed from the projectors P_Z and Q_W. Suppose that $\|T(I - P_Z)\| \leqslant \delta_Z$ and $\|(I - Q_W)T\| \leqslant \delta_W$. Set $T_d := Q_W T P_Z$ and $r_d := Q_W r$. Then, under the strong source condition (Assumption 4.1) the estimator is such that*

$$\begin{aligned}\mathbb{E}\|\hat{\varphi}^\circ - \varphi\|^2 \lesssim \{1 + \mathbb{E}\|\widehat{T}_d - T_d\|^{2\beta}\}K^{-\beta} + K\left(\mathbb{E}\|\widehat{T}_d - T_d\|^2 + \mathbb{E}\|\widehat{r}_d - r_d\|^2 + \delta_Z^{2+2(\beta\wedge 1)}\right) \\ + \mathbb{E}\|\widehat{T_d} - T_d\|^{2(\beta\wedge 1)} + \mathbb{E}\|\widehat{T_d} - T_d\|^{2\beta} + \delta_Z^{2(\beta\wedge 1)} + \delta_W^{2(\beta\wedge 1)}.\end{aligned}$$

Proof. Consider the definition of the estimator $\hat{\varphi}^\circ$ given by (A.2). The proof is based on the decomposition

$$\mathbb{E}\|\hat{\varphi}^\circ - \varphi\|^2 \lesssim \mathbb{E}\|R_K^\mu(\widehat{T}_d^\star \widehat{T}_d)\widehat{T}_d^\star\{\widehat{T}_d\varphi - \widehat{r}_d\}\|^2 + \mathbb{E}\|\{I - R_K^\mu(\widehat{T}_d^\star \widehat{T}_d)\widehat{T}_d^\star \widehat{T}_d\}\varphi\|^2 \tag{B.4}$$

We bound each term of the RHS separately.

In order to bound the first term, we bound separately the two factors $\|\mu R_K^\mu(\widehat{T}_d^\star \widehat{T}_d)\widehat{T}_d^\star\|$ and $\|\mu^{-1}(\widehat{T}_d\varphi - \widehat{r}_d)\|$. For the first factor:

$$\|\mu R_K^\mu(\widehat{T}_d^\star \widehat{T}_d)\widehat{T}_d^\star\| = \left\|R_K^1\left(\frac{\widehat{T}_d^\star}{\mu}\frac{\widehat{T}_d}{\mu}\right)\frac{\widehat{T}_d^\star}{\mu}\right\| = \sup\left\{\sqrt{\lambda}R_K^1(\lambda) \text{ s.t. } \lambda \in [0, 1]\right\}$$

because the spectrum of $\frac{\widehat{T}_d^\star}{\mu}\frac{\widehat{T}_d}{\mu}$ belongs to $[0,1]$. Using the inequality $\sqrt{\lambda}R_K(\lambda) = \lambda^{-1/2}[1-(1-\lambda)^K] \leqslant \sqrt{K}$, we get the bound

$$\|\mu R_K^\mu(\widehat{T}_d^\star \widehat{T}_d)\widehat{T}_d^\star\|^2 \leqslant K. \tag{B.5}$$

For the second factor, using the decomposition $\|\mu^{-1}(\widehat{T}_d\varphi - \widehat{r}_d)\| \lesssim \|\widehat{T}_d - T\| + \|\hat{r}_d - r\|$ that holds since $\mu > 1$, we can write

$$\mathbb{E}\|\mu^{-1}(\widehat{T}_d\varphi - \widehat{r}_d)\|^2 \lesssim \mathbb{E}\|\widehat{T}_d - T_d\|^2 + \mathbb{E}\|\widehat{r}_d - r_d\|^2 + \|r_d - T_d\varphi\|^2.$$

The last term is such that $\|r_d - T_d\varphi\|^2 = \|Q_W T\varphi - Q_W T P_Z\varphi\|^2 \leqslant \|Q_W\|^2 \cdot \|T(I-P_Z)\|^2 \cdot \|(I-P_Z)\varphi\|^2$. The strong source condition (Assumption 4.1) implies the existence of a function ψ in L^2_π such that $\varphi = (T^*T)^{\beta/2}\psi$ with $\|\psi\| \leqslant \rho$. This, together with (B.1), implies

$$\|(I-P_Z)\varphi\|^2 = \|(I-P_Z)(T^\star T)^{\beta/2}\psi\|^2 \lesssim \delta_Z^{2(\beta\wedge 1)}$$

and therefore $\|r_d - T_d\varphi\|^2 \lesssim \delta_Z^{2+2(\beta\wedge 1)}$. Finally the bound for the first term in (B.4) is

$$\mathbb{E}\|R_K^\mu(\widehat{T}_d^\star \widehat{T}_d)\widehat{T}_d^\star\{\widehat{T}_d\varphi - \widehat{r}_d\}\|^2 \lesssim K\left(\mathbb{E}\|\widehat{T}_d - T_d\|^2 + \mathbb{E}\|\widehat{r}_d - r_d\|^2 + \delta_Z^{2+2(\beta\wedge 1)}\right).$$

To treat the second term of (B.4) we separately consider the cases $\beta \neq 1$ and $\beta = 1$.

Case 1: $\beta \neq 1$. As before, the strong source condition (Assumption 4.1) implies the existence of a function ψ in L^2_π such that $\varphi = (T^*T)^{\beta/2}\psi$ with $\|\psi\| \leqslant \rho$. With this assumption the second term is up to a constant bounded by

$$\mathbb{E}\left\|\{I - R_K^\mu(\widehat{T}_d^\star\widehat{T}_d)\widehat{T}_d^\star\widehat{T}_d\}\left(\frac{\widehat{T}_d^\star}{\mu}\frac{\widehat{T}_d}{\mu}\right)^{\beta/2}\mu^\beta\psi\right\|^2 \\ + \mathbb{E}\left\|\{I - R_K^\mu(\widehat{T}_d^\star\widehat{T}_d)\widehat{T}_d^\star\widehat{T}_d\}\left\{\left(\frac{T^\star}{\mu}\frac{T}{\mu}\right)^{\beta/2} - \left(\frac{\widehat{T}_d^\star}{\mu}\frac{\widehat{T}_d}{\mu}\right)^{\beta/2}\right\}\mu^\beta\psi\right\|^2. \tag{B.6}$$

The first term of the decomposition (B.6) is bounded up to a constant by

$$\mathbb{E}\mu^{2\beta}\sup\left\{[1 - R_K(\lambda)\lambda]\lambda^{\beta/2} : \lambda \in [0,1]\right\}^2 \lesssim K^{-\beta}\mathbb{E}\mu^{2\beta}$$

because $[1 - R_K(\lambda)\lambda]\lambda^{\beta/2} = (1-\lambda)^K\lambda^{\beta/2} \leqslant C_\beta K^{-\beta/2}$ for some generic positive factor C_β depending on β only. Using that $\mu^{2\beta} \lesssim C\{(1 \vee \|T\|^{2\beta}) + \|T - \widehat{T}_d\|^{2\beta}\}$, we can finally bound the first term of (B.6):

$$\mathbb{E}\left\|\{I - R_K^\mu(\widehat{T}_d^\star\widehat{T}_d)\widehat{T}_d^\star\widehat{T}_d\}\left(\frac{\widehat{T}_d^\star}{\mu}\frac{\widehat{T}_d}{\mu}\right)^{\beta/2}\mu^\beta\psi\right\|^2 \leqslant K^{-\beta}(1 + \mathbb{E}\|T - \widehat{T}_d\|^{2\beta}) \tag{B.7}$$

up to a constant, using that $\|T\| < \infty$.

We now bound the second term of (B.6). As $\|I - R_K^\mu(\widehat{T}_d^\star\widehat{T}_d)\widehat{T}_d^\star\widehat{T}_d\| \leqslant 1$, that term is bounded up to a constant by $\mathbb{E}\|(T^*T)^{\beta/2} - (\widehat{T}_d^*\widehat{T}_d)^{\beta/2}\|^2$. In order to bound that term, we consider two subcases.

Case 1a: $\beta < 1$. Lemma B.1, inequality (B.3) allows to write with an appropriate $\tilde{\mu}$

$$\|(T^\star T)^{\beta/2} - (\widehat{T}_d^\star \widehat{T}_d)^{\beta/2}\| = \tilde{\mu}^\beta \Big\| \Big(\frac{T^\star}{\tilde{\mu}} \frac{T}{\tilde{\mu}}\Big)^{\beta/2} - \Big(\frac{\widehat{T}_d^\star}{\tilde{\mu}} \frac{\widehat{T}_d}{\tilde{\mu}}\Big)^{\beta/2} \Big\|$$

$$\leqslant C_\beta \tilde{\mu}^\beta \Big\| \frac{T}{\tilde{\mu}} - \frac{\widehat{T}_d}{\tilde{\mu}} \Big\|^\beta = C_\beta \|\widehat{T}_d - T\|^\beta. \tag{B.8}$$

Case 1b: $\beta > 1$. We proceed analogously and get the bound

$$\|(T^\star T)^{\beta/2} - (\widehat{T}_d^\star \widehat{T}_d)^{\beta/2}\| \leqslant C_\beta \tilde{\mu}^{\beta-1} \|T - \widehat{T}_d\|$$

Choosing $\tilde{\mu}^{\beta-1} \lesssim (1 \vee \|T\|^{\beta-1}) + \|\widehat{T}_d - T\|^{\beta-1}$ we can write

$$\|(T^\star T)^{\beta/2} - (\widehat{T}_d^\star \widehat{T}_d)^{\beta/2}\| \lesssim \|\widehat{T}_d - T\| + \|\widehat{T}_d - T\|^\beta, \tag{B.9}$$

Overall, (B.8) and (B.9) lead to

$$\mathbb{E}\|(T^* T)^{\beta/2} - (\widehat{T}_d^* \widehat{T}_d)^{\beta/2}\|^2 \lesssim \mathbb{E}\|\widehat{T_d} - T\|^{2(\beta \wedge 1)} + \mathbb{E}\|\widehat{T_d} - T\|^{2\beta} \qquad \text{for all } \beta \neq 1$$

and thus we get the result for all $\beta \neq 1$.

Case 2: $\beta = 1$. That case needs a slightly different technique, because (B.3) is no longer valid. We first notice that the range of the operator $(T^\star T)^{1/2}$ is the same as the range of the operator $T^\star$ (see e.g. Proposition 2.18 of Engl et al. (2000)) and thus the strong source condition implies the existence of a function $\psi \in L^2_\tau$ such that $\varphi = T^\star \psi$. Therefore the second term in (B.4) is bounded up to a constant by

$$\mathbb{E}\|\{I - R_K^\mu(\widehat{T}_d^\star \widehat{T}_d)\widehat{T}_d^\star \widehat{T}_d\}\widehat{T}_d^\star \psi\|^2 + \mathbb{E}\|\{I - R_K^\mu(\widehat{T}_d^\star \widehat{T}_d)\widehat{T}_d^\star \widehat{T}_d\}(T^\star - \widehat{T}_d^\star)\psi\|^2$$

By (B.7), the first term is bounded up to a constant by $(1 + \mathbb{E}\|T - \widehat{T}_d\|^{2\beta})K^{-\beta}$. Using again $\|I - R_K^\mu(\widehat{T}_d^\star \widehat{T}_d)\widehat{T}_d^\star \widehat{T}_d\| \leqslant 1$, the second term is bounded up to a constant by $\mathbb{E}\|T^\star - \widehat{T}_d^\star\|^2 = \mathbb{E}\|T - \widehat{T}_d\|^2$.

Combining all bounds we obtain that the second term of (B.4) is bounded up to a constant by

$$\{1 + \mathbb{E}\|\widehat{T}_d - T\|^{2\beta}\}K^{-\beta} + \mathbb{E}\|\widehat{T_d} - T\|^{2(\beta \wedge 1)} + \mathbb{E}\|\widehat{T_d} - T\|^{2\beta}.$$

By using the inequalities $\|\widehat{T}_d - T\| \leqslant \delta_Z + \delta_W + \|\widehat{T}_d - T_d\| \lesssim 1 + \|\widehat{T}_d - T_d\|$, we obtain the desired result. □

C Risk bound under the weak source condition assumption

Under the weak source condition assumption, the proof of a stochastic bound for $\|\hat{\varphi}^\circ - \varphi\|$ necessitates a different proof technique. We start with a key lemma.

Lemma C.1. *Consider the estimator (3.11) constructed using the projections P_Z and Q_W. Suppose that $\|T(I-P_Z)\| \leqslant \delta_Z$ and $\|(I-Q_W)T\| \leqslant \delta_W$. Set $T_d := Q_W T P_W$ and $r_d := Q_W r$. Then, under the weak source condition (Assumption 5.1),*

$$\mathbb{E}\|\hat{\varphi}^\circ - \varphi\|^2 \lesssim K\{\mathbb{E}\|\widehat{T}_d - T_d\|^2 + \mathbb{E}\|\widehat{r}_d - r_d\|^2 + \delta_Z^2 + \delta_W^2\} + 2(\log K)^{-\beta}$$

provided that K, h_W and h_Z are such that $K^2\mathbb{E}\|T^\star T - \widehat{T}_d^\star \widehat{T}_d\|_2^2$ is finite.

Proof. Analogously to (B.4), consider the decomposition

$$\|\hat{\varphi}^\circ - \varphi\|^2 \lesssim \|R_K^\mu(\widehat{T}_d^\star \widehat{T}_d)\widehat{T}_d^\star\{\widehat{T}_d\varphi - \widehat{r}_d\}\|^2 + \|\{I - R_K^\mu(\widehat{T}_d^\star \widehat{T}_d)\widehat{T}_d^\star \widehat{T}_d\}\varphi\|^2 \tag{C.1}$$

Using the inequality (B.5) from the previous proof, the first term is bounded by

$$\begin{aligned} K\|\{\widehat{T}_d\varphi - \widehat{r}_d\}\|^2 &\leqslant K\left\{\|\widehat{T}_d - T_d\|^2 + \|\widehat{r}_d - r_d\|^2 + \|T_d - T\|^2 + \|r_d - r\|^2\right\} \\ &\leqslant K\left\{\|\widehat{T}_d - T_d\|^2 + \|\widehat{r}_d - r_d\|^2 + \delta_Z^2 + 2\delta_W^2\right\}. \end{aligned}$$

Getting an upper bound for the second of (C.1) is more delicate. Observe that the operator $S := \{I - R_K^\mu(\widehat{T}_d^\star \widehat{T}_d)\widehat{T}_d^\star \widehat{T}_d\}$ is self-adjoint (i.e. $S^\star = S$) and such that $\|S^{1/2}\| \leqslant 1$. Therefore, the second term of (C.1) is $\|S\varphi\|^2 \leqslant \|S^{1/2}\varphi\|^2 = |\langle S\varphi, \varphi\rangle|$.

Let $\phi_\beta(u) := [-\log(u/2)]^{-\beta/2}$ and note that the operator $\phi_\beta(T^\star T)$ is also self-adjoint. This implies

$$\begin{aligned} \|S\varphi\|^2 &\leqslant |\langle S\varphi, \phi_\beta(T^\star T)\phi_\beta^{-1}(T^\star T)\varphi\rangle| \\ &= |\langle \phi_\beta(T^\star T)S\varphi, \phi_\beta^{-1}(T^\star T)\varphi\rangle| \\ &\leqslant \|\phi_\beta(T^\star T)S\varphi\| \cdot \|\phi_\beta^{-1}(T^\star T)\varphi\| \end{aligned}$$

by the Cauchy-Schwarz inequality. The weak source condition assumption implies that $\|\phi_\beta^{-1}(T^\star T)\varphi\| \leqslant \rho$. Therefore,

$$\|S\varphi\|^2 \leqslant \rho\|\phi_\beta(T^\star T)S\varphi\| \tag{C.2}$$

and the Jensen's inequality implies

$$\mathbb{E}\|S\varphi\|^2 \leqslant \rho\sqrt{\mathbb{E}\|\phi_\beta(T^\star T)S\varphi\|^2} \tag{C.3}$$

Define the function $\Gamma_\beta(u) = 2\exp(-u^{-1/\beta})$. In the technical Lemma C.2 below, we show that

$$\Gamma_\beta\left(\gamma_\beta^2 \frac{\sqrt{\mathbb{E}\|\phi_\beta(T^\star T)S\varphi\|^2}}{\rho^2}\right) \lesssim \sqrt{\mathbb{E}\|T^\star T - \widehat{T}_d^\star \widehat{T}_d\|_2^2 + \mathbb{E}\mu^4/K^2}.$$

for $\gamma_\beta = 1 \wedge 1/\{(1+\beta)^\beta \varphi_\beta(\|T\|)^2\}$. This implies

$$\mathbb{E}\|S\varphi\|^2 \leqslant \frac{1}{\gamma_\beta^2}\phi_\beta\left\{\Gamma_\beta\left(\gamma_\beta^2 \frac{\sqrt{\mathbb{E}\|\phi_\beta(T^\star T)S\varphi\|^2}}{\rho^2}\right)^2\right\}.$$

Because $1/\gamma_\beta^2$ is bounded, we can write using Lemma C.2

$$\mathbb{E}\|S\varphi\|^2 \lesssim \left\{-\log\sqrt{\mathbb{E}\|T^\star T - \widehat{T}_d^\star \widehat{T}_d\|_2^2 + \mathbb{E}\mu^4/K^2}\right\}^{-\beta}$$

Note that $\mathbb{E}\mu^4$ is finite by Proposition 3.1. Therefore, if we assume that $K^2\mathbb{E}\|T^\star T - \widehat{T}_d^\star \widehat{T}_d\|_2^2$ is finite for each K, we get the upper bound $\mathbb{E}\|S\varphi\|^2 \lesssim 2(\log K)^{-\beta}$ and the result follows. □

LEMMA C.2. *Let $\Gamma_\beta(u) := 2\exp(-u^{-1/\beta})$ and $\phi_\beta(u) := [-\log(2u)]^{-\beta/2}$. Define*

$$\gamma_\beta := 1 \wedge \frac{1}{(1+\beta)^\beta \phi_\beta(\|T\|)^2}$$

and $S := I - R_K^\mu(\widehat{T}_d^\star \widehat{T}_d)\widehat{T}_d^\star \widehat{T}_d$. Then the inequality

$$\Gamma_\beta\left(\gamma_\beta^2 \frac{\sqrt{\mathbb{E}\|\phi_\beta(T^\star T)S\varphi\|^2}}{\rho}\right) \leqslant \sqrt{2\mathbb{E}\|T^\star T - \widehat{T}_d^\star \widehat{T}_d\|_2^2 + \frac{2}{K^2}C_2'\rho^2\mathbb{E}\mu^4}$$

holds true.

Proof. We first derive three useful inequalities.

1. We first give a bound for $\|\widehat{T}_d S\varphi\|^2$. Using the bound (B.7) with $\beta = 2$, we can write

   $$\|\widehat{T}_d S\varphi\|^2 \leqslant \|\widehat{T}_d S^{1/2}\| \cdot \|S^{1/2}\varphi\|^2 \leqslant \frac{C_2'\mu^2\rho}{K} \cdot \|S^{1/2}\varphi\|^2.$$

   By (C.2), we get the bound

   $$\mathbb{E}\|\widehat{T}_d S\varphi\|^2 \leqslant \frac{C_2'\mu^2\rho^2}{K} \cdot \mathbb{E}\|\phi_\beta(T^\star T)S\varphi\|. \tag{C.4}$$

2. We derive a bound for $\mathbb{E}\|TS\varphi\|^2/\sqrt{\mathbb{E}\|\phi_\beta(T^\star T)S\varphi\|^2}$. Using (B.7), (C.2) and (C.4), we get

   $$\begin{aligned}\|TS\varphi\|^2 &= \langle T^\star TS\varphi, S\varphi\rangle \\ &= \langle (T^\star T - \widehat{T}_d^\star \widehat{T}_d)S\varphi, S\varphi\rangle + \|\widehat{T}_d S\varphi\|^2 \\ &\leqslant \|\phi_\beta(T^\star T)S\varphi\| \left\{\rho\|T^\star T - \widehat{T}_d^\star \widehat{T}_d\| + \frac{C_2'\mu^2\rho^2}{K}\right\}.\end{aligned}$$

   Taking the expectation and using the Cauchy-Schwarz's inequality, we obtain

   $$\frac{\mathbb{E}\|TS\varphi\|^2}{\sqrt{\mathbb{E}\|\phi_\beta(T^\star T)S\varphi\|^2}} \leqslant \sqrt{2\rho^2\mathbb{E}\|T^\star T - \widehat{T}_d^\star \widehat{T}_d\|^2 + \frac{2C_2'^2\mathbb{E}\mu^4\rho^4}{K^2}} \tag{C.5}$$

3. Denote by $\{(\lambda_i, u_i)\}$ the eigenvalue decomposition of the compact operator $T^\star T$, where λ_i is a decreasing sequence of positive eigenvalues in $\mathbb{R}$, and u_i is the corresponding orthonormal system of eigenfunctions. The norm is such that $\|\phi_\beta(T^\star T)S\varphi\|^2 = \sum_i \phi_\beta(\lambda_i)^2 \langle S\varphi, u_i\rangle^2$ and

$$\frac{\mathbb{E}\|\phi_\beta(T^\star T)S\varphi\|^2}{\mathbb{E}\|S\varphi\|^2} = \sum_i \phi_\beta(\lambda_i)^2 \frac{\mathbb{E}\langle S\varphi, u_i\rangle^2}{\sum_j \mathbb{E}\langle S\varphi, u_j\rangle^2} .$$

Note that the function $\Gamma_\beta(\cdot)$ is convex over the interval $(0,(1+\beta)^{-\beta}]$. Moreover, since $\gamma_\beta \leqslant$ and $\Gamma_\beta(\cdot)$ is an increasing function, we get $\Gamma_\beta(\gamma_\beta^2 \phi_\beta(\lambda_i)^2) \leqslant \Gamma_\beta(\phi_\beta(\lambda_i)^2) = \lambda_i$ for all eigenvalue λ_i in the spectrum of the operator T. Therefore, the Jensens's inequality allows to write

$$\Gamma_\beta\left(\gamma_\beta^2 \frac{\mathbb{E}\|\phi_\beta(T^\star T)S\varphi\|^2}{\mathbb{E}\|S\varphi\|^2}\right) \leqslant \sum_i \lambda_i \frac{\mathbb{E}\langle S\varphi, u_i\rangle^2}{\sum_j \mathbb{E}\langle S\varphi, u_j\rangle^2} = \frac{\mathbb{E}\|TS\varphi\|^2}{\mathbb{E}\|S\varphi\|^2} \tag{C.6}$$

To prove the result, we first note that (C.3) implies

$$\frac{(\mathbb{E}\|\phi_\beta(T^\star T)S\varphi\|^2)^{1/4}}{\sqrt{\rho}} \leqslant \frac{(\mathbb{E}\|\phi_\beta(T^\star T)S\varphi\|^2)^{1/2}}{(\mathbb{E}\|S\varphi\|^2)^{1/2}}$$

and therefore, for all monotone function g, it holds

$$g\left[\gamma_\beta \frac{(\mathbb{E}\|\phi_\beta(T^\star T)S\varphi\|^2)^{1/4}}{\sqrt{\rho}}\right] \leqslant g\left[\gamma_\beta \frac{(\mathbb{E}\|\phi_\beta(T^\star T)S\varphi\|^2)^{1/2}}{(\mathbb{E}\|S\varphi\|^2)^{1/2}}\right] .$$

If we apply that inequality with the monotone function $g(u) = \Gamma_\beta(u^2)/u^2$, we get using (C.6)

$$\Gamma_\beta\left(\gamma_\beta^2 \frac{\sqrt{\mathbb{E}\|\phi_\beta(T^\star T)S\varphi\|^2}}{\rho}\right) \leqslant \frac{\mathbb{E}\|TS\varphi\|^2}{\rho\sqrt{\mathbb{E}\|\phi_\beta(T^\star T)S\varphi\|^2}}$$

which leads to the result using (C.5). □

References

Angrist, J. D. & Krueger, A. B. (2001), 'Instrumental variables and the search for identification: From supply and demand to natural experiments', *Journal of Economic Perspectives* **15**, 69–85.

Chen, X. & Reiß, M. (2010), 'On rate optimality for ill-posed inverse problems in econometrics', *Econometric Theory* . To appear.

Darolles, S., Florens, J.-P. & Renault, E. (2002), Nonparametric instrumental regression. Working Paper # 228, IDEI, Université de Toulouse I.

DeVore, R. A. & Lorentz, G. G. (1993), *Constructive Approximation*, Springer, Berlin.

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

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

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

Hohage, T. (1997), 'Logarithmic convergence rates of the iteratively regularized Gauß-Newton method for an inverse potential and an inverse scattering problem.', *Inverse Problems* **13**, 1279–1299.

Hohage, T. (2000), 'Regularisation of exponentially ill-posed problems.', *Numer. Funct. Anal. Optim.* **21**, 439–464.

Johannes, J., Van Bellegem, S. & Vanhems, A. (2010), 'Convergence rates for ill-posed inverse problems with an unknown operator', *Econometric Theory* . To appear.

Mallat, S. (1997), *A wavelet tour of signal processing*, Academic Press.

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

Plato, R. (1990), On the discretization and regularization of selfadjoint ill-posed problems, *in* P. C. Sabatier, ed., 'Inverse Methods in Action', Inverse problems and theoretical imaging, Springer-Verlag, New York, pp. 117–121.

Plato, R. & Vainikko, G. (1990), 'On the regularization of projection methods for solving ill-posed problems', *Numer. Math.* **57**, 63–79.

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

Vidakovic, B. (1999), *Statistical modeling by wavelets*, Wiley, New York.

Wooldridge, J. (2008), *Introductory Econometrics: A Modern Approach*, 4th edn, South-Western College Pub.