Yale University

EliScholar – A Digital Platform for Scholarly Publishing at Yale

Cowles Foundation Discussion Papers | Cowles Foundation

9-1-2007

On Rate Optimality for Ill-posed Inverse Problems in Econometrics

Xiaohong Chen

Markus Reiss

Follow this and additional works at: https://elischolar.library.yale.edu/cowles-discussion-paper-series

Part of the Economics Commons

Recommended Citation

Chen, Xiaohong and Reiss, Markus, "On Rate Optimality for Ill-posed Inverse Problems in Econometrics" (2007). *Cowles Foundation Discussion Papers*. 1924.
https://elischolar.library.yale.edu/cowles-discussion-paper-series/1924

This Discussion Paper is brought to you for free and open access by the Cowles Foundation at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Cowles Foundation Discussion Papers by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.

ON RATE OPTIMALITY FOR ILL-POSED INVERSE PROBLEMS IN ECONOMETRICS

By

Xiaohong Chen and Markus Reiss

September 2007

COWLES FOUNDATION DISCUSSION PAPER NO. 1626

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

On rate optimality for ill-posed inverse problems in econometrics*

Xiaohong Chen† Markus Reiß‡

First version: March 15, 2007; this version: September 10, 2007.

Abstract

In this paper, we clarify the relations between the existing sets of regularity conditions for convergence rates of nonparametric indirect regression (NPIR) and nonparametric instrumental variables (NPIV) regression models. We establish minimax risk lower bounds in mean integrated squared error loss for the NPIR and the NPIV models under two basic regularity conditions that allow for both mildly ill-posed and severely ill-posed cases. We show that both a simple projection estimator for the NPIR model, and a sieve minimum distance estimator for the NPIV model, can achieve the minimax risk lower bounds, and are rate-optimal uniformly over a large class of structure functions, allowing for mildly ill-posed and severely ill-posed cases.

KEY WORDS: Nonparametric instrumental regression; Nonparametric indirect regression; Statistical ill-posed inverse problems; Minimax risk lower bound; Optimal rate.

JEL classifications: Primary C14; secondary C30

*We thank Joel Horowitz, Whitney Newey, Demian Pouzo, Yixiao Sun and the participants of the March 2007 Oberwolfach Workshop on Semiparametrics for very helpful discussions. X. Chen acknowledges support from the NSF/USA. The usual disclaimer applies.

†Department of Economics, Yale University, Box 208281, New Haven, CT 06520, USA. Email: xiaohong.chen@yale.edu

‡Institute of Applied Mathematics, University of Heidelberg, Germany. Email: reiss@statlab.uni-heidelberg.de

1 Introduction

Recently there is a growing interest in estimation for nonparametric instrumental variables (NPIV) regression models, see e.g., Newey and Powell (2003), Darolles, Florens and Renault (2002), Hall and Horowitz (2005), Blundell, Chen and Kristensen (2007), Gagliardini and Scaillet (2006), to name only a few. The estimators proposed in these papers belong to three broad classes: (1) the finite dimensional sieve minimum distance estimator (Newey and Powell (2003), Ai and Chen (2003) and Blundell, Chen and Kristensen (2007)); (2) the infinite dimensional kernel based Tikhonov regularized estimator (Darolles, Florens and Renault (2002), Hall and Horowitz (2005), Gagliardini and Scaillet (2006)); and (3) the finite dimensional orthogonal series Tikhonov regularized estimator (Hall and Horowitz (2005)). Each of these papers presents different sets of sufficient conditions for consistency and convergence rates of its proposed estimators. In addition, for the mildly ill-posed case (when the singular values associated with the conditional expectation operator decay to zero at a polynomial rate), Hall and Horowitz (2005) establish the minimax risk lower bound in mean integrated squared error loss for the NPIV regression model under a set of regularity conditions that are related to their estimation procedures. They also show that their proposed estimators achieve this lower bound; hence their rate is optimal for the class of structure functions they consider.

To the best of our knowledge, there is no published work that discuss the relations among the different sets of sufficient conditions imposed in these various papers. Therefore, it is unclear whether the minimax risk lower bound derived in Hall and Horowitz (2005) is still the lower bound under regularity conditions stated in the other papers. It is also unclear whether the estimators proposed in the other papers are rate optimal in a minimax framework corresponding to the conditions stated in these papers. Moreover, when the NPIV problem is severely ill-posed (for instance, when the singular value associated with the conditional expectation operator decays to zero at an exponential rate), there are no published results on minimax rates.

In this paper, we address these issues based on a general formulation of the problems. In Section 2, we first present the NPIV models. We then provide two basic regularity conditions: the approximation and the link conditions. The approximation condition is about the complexity of the class of the structural functions, which is measured as the best finite dimensional linear approximation error rate in terms of a basis expansion that may not be the eigenfunction basis of the conditional expectation operator. The link condition is about the relative smoothness of the conditional expectation operator in terms of the basis used in the first condition. We show that these two regularity conditions are natural generalizations of, and are automatically satisfied by, the so-called "general source condition", an assumption commonly imposed in the literature on ill-posed inverse problems. Our two basic regularity conditions are also implied by the ones assumed in the literature on NPIV models, such as those imposed in Darolles, Florens and Renault (2002), Hall and Horowitz (2005), and Blundell, Chen and Kristensen (2007). In Section 3, we first show that the NPIV model is no more informative than the reduced form nonparametric indirect regression (NPIR) model (actually the model assuming a known conditional expectation operator of the endogenous regressor given the instrumental variables). Under the two basic regularity conditions stated in Section 2, we derive the minimax risk lower bound in mean integrated squared error loss for the NPIR and the NPIV models, allowing for both the mildly ill-posed case and the severely ill-posed case. In Section 4, we present a simple projection estimator for the NPIR models, and establish that it achieves the lower bounds and hence is rate-optimal in the minimax sense. When restricting our conditions to various special cases, including the nonparametric mean regression models and the NPIR models under general source conditions, our results reproduce the existing known minimax optimal rates for these special cases. But more importantly, our minimax optimal rate results cover many new cases as long as their model specifications satisfy the approximation and the link conditions. We also discuss what could happen if the link condition on the relative smoothness of the conditional expectation operator is not satisfied. In Section 5, we show that the sieve minimum distance (SMD) estimator for the NPIV models is rate-optimal in the minimax sense. In fact, we show that both the projection estimator for the NPIR models and the SMD estimator for the NPIV models are rate-optimal uniformly over a large class of structure functions, allowing for arbitrarily decaying speed of the singular values of the conditional expectation operator. Section 6 provides some further discussions on the regularity conditions. Section 7 briefly concludes, and all the proofs are gathered in the Appendix.

Before we conclude this introduction, we mention closely related work in more abstract settings of linear ill-posed inverse problems. First, there exist many papers and some monographs devoted to constructing estimators and deriving optimal convergence rates in the deterministic noise framework with a known operator (or a known operator up to a deterministically perturbed error with a specified error rate). See, e.g., Engl, Hanke and Neubauer (1996), Nair, Pereverzev and Tautenhahn (2005) and the references therein. Second, there are also many results on minimax optimal rates in mean integrated squared error loss in the random white noise framework with a known operator; see, e.g., Cohen, Hoffmann and Reiß (2004), Bissantz, Hohage, Munk and Ruymgaart (2007) and the references therein. Third, there are a few recent papers on constructing estimators that achieve optimal convergence rates in the presence of a white noise and an unknown operator, but assuming the existence of an estimator of the operator with a rate. See, e.g., Efromovich and Koltchinskii (2001) and Hoffmann and Reiß (2007).

2 NPIV models and basic regularity conditions

We first specify the NPIV regression model as

$$Y_i = h_0(X_i) + U_i, \quad \mathbb{E}[U_i \,|\, W_i] = 0, \quad i = 1, \ldots, n, \tag{2.1}$$

with observations $\{(X_i, Y_i, W_i)\}_{i=1}^n$, a random sample from the unknown joint distribution of (X, Y, W). Here Y is a scalar dependent variable, X is a vector of endogenous regressors in $\mathbb{R}^d$ and W is a vector of instrumental variables in $\mathbb{R}^d$ that satisfy the property $\mathbb{E}[U \,|\, W] = 0$. (For the ease of presentation we assume that X and W do not contain any common variables, and the conditional density of X given W is well-defined). The parameter of interest is the unknown structure function $h_0(\bullet)$, while the joint law $\mathscr{L}_{UWX}$ of (U, W, X) is an unknown nuisance function.

Let us introduce the Hilbert spaces

$$L^2_X = \{h : \mathbb{R}^d \to \mathbb{R} \mid \|h\|^2_X := \mathbb{E}[h^2(X)] < \infty\},$$
$$L^2_W = \{g : \mathbb{R}^d \to \mathbb{R} \mid \|g\|^2_W := \mathbb{E}[g^2(W)] < \infty\}.$$

Since the conditional distribution of X given W is unspecified, the conditional expectation operator

$$(Kh)(w) := \mathbb{E}[h(X) \,|\, W = w]$$

is unknown, except that it is an integral operator mapping from L^2_X to L^2_W. This operator is the key in the construction of estimators of h_0 because by conditioning on W in (2.1) and using $\mathbb{E}[U \,|\, W] = 0$ we obtain

$$\mathbb{E}[Y \,|\, W] = \mathbb{E}[h_0(X) \,|\, W] + \mathbb{E}[U \,|\, W] = Kh_0(W).$$

Consequently, by regressing Y on W, estimating K and using this relationship we can hope to retrieve an estimator of h_0.

Let $\mathscr{H}$ denote a subset of L^2_X and assume $h_0 \in \mathscr{H}$. Here $\mathscr{H}$ captures all the prior information (such as the smoothness and/or shape properties) about the unknown structure function h_0. To ensure that there is a unique solution $h_0 \in \mathscr{H}$ for the NPIV model (2.1), in this paper we assume that the operator K satisfies the following restriction:

$$\{h \in \mathscr{H} : Kh = 0\} = \{0\}. \tag{2.2}$$

Depending on the choice of the function class $\mathscr{H}$, the identification condition (2.2) imposes different restrictions on the operator K (or equivalently, on the conditional density of X given W). For example, if $\mathscr{H} = L^2_X$, then condition (2.2) becomes the standard identification condition that K

is *injective*, i.e., $\mathscr{N}(K) := \{h \in L^2_X : Kh = 0\} = \{0\}$, (or equivalently, the conditional density of X given W is complete); see, e.g., Newey and Powell (2003), Darolles, Florens and Renault (2002), Carrasco, Florens and Renault (2007). If $\mathscr{H} = \{h \in L^2_X : \sup_x |h(x)| \leqslant 1\}$, then condition (2.2) corresponds to assume that the conditional density of X given W is bounded complete; see, e.g., Chernozhukov and Hansen (2005), Chernozhukov, Imbens and Newey (2007), Blundell, Chen and Kristensen (2007). For additional results on identification in semi/nonparametric models with endogeneity, see, e.g., Blundell and Powell (2003), Florens (2003), Florens, Johannes and Van Bellegem (2007) and the references therein.

2.1 Basic regularity conditions

In this paper we would like to establish a minimax risk lower bound for the NPIV model, that is, we would like to derive a result of the form: there are a finite constant $c > 0$ and a rate function $\delta_n \downarrow 0$ as $n \uparrow \infty$ such that

$$\lim_{n\to\infty} \left(\delta_n^{-1} \inf_{\hat{h}_n} \sup_{h\in\mathscr{H}} \mathbb{E}_{(\mathscr{L}_{UWX},h)}[\|\hat{h}_n - h\|_X^2]\right) \geqslant c$$

where the infimum is over all possible estimators $\hat{h}_n$ for $h \in \mathscr{H}$. Note that a NPIV model (2.1) is completely specified by prescribing the joint law $\mathscr{L}_{UWX}$ of (U, W, X) and the structure function h. This lower bound δ_n will be valid for quite general forms of $\mathscr{L}_{UWX}$, independently of knowing or not knowing it. In particular, although the mean squared error loss and the class of structure functions $\mathscr{H}$ will be defined in terms of the distribution of X, there is no need to assume any explicit properties of this distribution to derive a minimax lower bound.

We would also like to present some particular estimators that attain the lower bound rate δ_n. However, before we could establish any minimax lower and upper bounds, it is clear that we have to impose some conditions on the class of structure functions $\mathscr{H}$ and on the conditional expectation operator K. In this paper, we implicitly assume that the prior information about $\mathscr{H}$ already includes some regularity properties that could be described in terms of a Hilbert scale generated by a conveniently chosen (by the researcher) operator B. The regularizing action of the conditional expectation operator K would also be described as some smoothness relative to the known operator B. Formally, let $B : Dom(B) \subseteq L^2_X \to L^2_X$ be a densely defined self-adjoint, strictly positive definite, and *unbounded* operator (such as differential operators with boundary constraints). For the ease of presentation we assume that B has eigenvalues $\nu_k \uparrow \infty$ with corresponding L^2_X-normalized eigenfunctions $\{u_k\}$ which then form an orthonormal basis of L^2_X. For non-discrete spectrum our results will still hold, but the presentation would become more technical, using spectral measures and abstract functional calculus.

Throughout this paper we denote by $\mathscr{H}(r,R)$ a subset of $\mathscr{H} \subseteq L^2_X$, and assume the following:

2.1 Assumption (approximation condition)**.** *There are finite constants $r, R > 0$ such that $\mathscr{H}(r,R)$ consists of functions h satisfying*

$$\inf_{\{a_k\}:\sum_k a_k^2<\infty} \|h - \sum_{k=1}^{m} a_k u_k\|_X^2 \leqslant R^2 \nu_{m+1}^{-2r} \quad \textit{for all} \;\; m \in \mathbb{N}\,. \tag{2.3}$$

Note that the left-hand side of (2.3) gives the error in approximating h optimally by an element of the m-dimensional space spanned by the basis functions $\{u_1, \ldots, u_m\}$. So, Assumption 2.1 characterizes the regularity (or smoothness) of the structure functions in $\mathscr{H}(r,R)$ by the L^2_X-approximation error rates when they are approximated by the basis $\{u_k\}$ associated with B. Clearly, Assumption 2.1 will give a bound on the bias and implies that $\mathscr{H}(r,R)$ is a compact set in L^2_X. For many typical smooth function classes and basis functions like the Fourier basis, wavelets or splines the approximation error rates are well known.

For any $s > 0$ and $h \in Dom(B^s) \subseteq L^2_X$ we write $\|h\|_s := \|B^s h\|_X$. Let H^s denote the completion of $Dom(B^s)$ under the norm $\|\bullet\|_s$. $\{H^s\}_{s>0}$ is called a *Hilbert scale* generated by B (see, e.g., Engl, Hanke and Neubauer (1996) for its detailed properties). For any finite constants $r, R > 0$, we define a Sobolev-type ellipsoid as $H^r_R := \{h \in H^r, \; \|h\|_r \leqslant R\}$. Since

$$H^r_R = \Big\{h = \sum_{k=1}^{\infty} \langle h, u_k\rangle_X u_k, \; \|h\|_r^2 = \sum_{k=1}^{\infty} \nu_k^{2r} \langle h, u_k\rangle_X^2 \leqslant R^2 \Big\},$$

it is clear that H^r_R is a subset of $\mathscr{H}(r,R)$. It is also easy to see that the following hyperrectangle $\Theta^r_{R'}$ in L^2_X is a subset of $\mathscr{H}(r,R)$ for $R' > 0$ sufficiently small:

$$\Theta^r_{R'} := \Big\{h = \sum_{k=1}^{\infty} \langle h, u_k\rangle_X u_k, \; |\langle h, u_k\rangle_X| \leqslant R' \nu_k^{-\beta} \Big\}, \quad \beta = r + \tfrac{1}{2} > \tfrac{1}{2}.$$

Let us now formulate the mapping properties of the conditional expectation operator K in terms of the (generalized) Hilbert scale generated by B.

2.2 Assumption (link condition)**.** *There are a continuous increasing function $\varphi : \mathbb{R}^+ \to \mathbb{R}^+$ and a constant $M > 0$ such that $\|Kh\|_W \leqslant M\|[\varphi(B^{-2})]^{1/2} h\|_X$ for all $h \in L^2_X$.*

Assumption 2.2 is in fact equivalent to the range inclusion condition:

$$\mathrm{ran}(|K|) \subseteq \mathrm{ran}([\varphi(B^{-2})]^{1/2}) \text{ with } |K| := (K^*K)^{1/2},$$

where K^* denotes the Hilbert space (L^2_X) adjoint of K. For the NPIV models, under mild conditions, the conditional expectation operator K is a compact operator. Thus the self-adjoint compact operator K^*K has the eigenvalue-eigenfunction decomposition $\{\lambda_k, e_k\}$, where the eigenvalues are arranged in non-increasing order: $\lambda_k \geqslant \lambda_{k+1} \geqslant ... > 0$ and λ_k tends to zero as $k \uparrow \infty$. Then Assumption 2.2 can be equivalently restated in terms of two possibly different orthonormal bases $\{e_k\}$ and $\{u_k\}$ of L^2_X:

$$\sum_{k=1}^{\infty} \lambda_k \langle h, e_k\rangle_X^2 \leqslant M^2 \sum_{k=1}^{\infty} \varphi(\nu_k^{-2}) \langle h, u_k\rangle_X^2 \quad \text{for all} \quad h \in L^2_X. \tag{2.4}$$

2.3 Remark. *We can rewrite Assumptions 2.1 and 2.2 without specifying the operator B explicitly. All we require are the existence of an orthonormal basis $\{u_k\}$ in L^2_X and a sequence of increasing positive real numbers $\{\nu_k\}$ such that equations (2.3) and (2.4) hold. In fact, we can then construct the self-adjoint unbounded operator B according to*

$$Bh = \sum_{k=1}^{\infty} \nu_k \langle h, u_k\rangle_X u_k,$$

with $Dom(B) = \{h \in L^2_X : \sum_{k=1}^{\infty} \nu_k^2 \langle h, u_k\rangle_X^2 < \infty\}$.

2.4 Example. *Suppose that X is uniformly distributed on the interval $[0,1]$ and let $Bf(x) := -f''(x)$ for all $f \in L^2([0,1])$ with $f'' \in L^2([0,1])$ and with periodic boundary conditions. Then B has (complex-valued) eigenfunctions $u_k(x) = \exp(2\pi kix)$ with eigenvalues $\nu_k = (2\pi k)^2$ such that*

$$H^r = \{f \in L^2([0,1]) \,:\, \sum_{k\in\mathbb{Z}} \nu_k^{2r} |\langle f, u_k\rangle|^2 < \infty\}$$

is the classical L^2-Sobolev space H^{2r}_{per} of regularity (smoothness) $2r$ with periodic boundary conditions. See, e.g., Edmunds and Evans (1987) for many examples of generating smooth function spaces from differential operators.

For the typical choice $\varphi(t) = t^a$ for some $a > 0$, Assumption 2.2 translates to $\|Kh\|_W \leqslant M\|B^{-a}h\|_X$, which means intuitively that the operator K regularizes at least as much as B^{-a}. In the case $Bf(x) := -f''(x)$ the operator K acts like integrating at least $(2a)$-times, i.e. maps L^2 to the L^2-Sobolev space of regularity $2a$.

In the statistics literature, for the standard nonparametric mean regression model (i.e., the model in which K is the identity operator), the minimax risk lower and upper bounds have been established in mean integrated squared error loss for various classes of functions $\mathscr{H}$ such as a Sobolev ball (ellipsoid), a Hölder ball (hyperrectangle) or a Besov ball (ellipsoid or hyperrectangle or Besov body); see, e.g., Donoho, Liu and MacGibbon (1990), Yang and Barron (1999) and the references therein. As shown in these papers, what matters for minimax risk lower and upper bounds for nonparametric mean regression estimation is the complexity of the class of functions $\mathscr{H}$ that can be measured in terms of best finite dimensional approximation numbers. This motivates us to impose Assumption 2.1. However, since the basis $\{u_k\}$ (of the operator B) used to construct the best finite dimensional approximations for the class of functions $\mathscr{H}$ may differ from the eigenfunction basis $\{e_k\}$ (of the operator K^*K), we have to impose Assumption 2.2 to link these two.

We shall refer to Assumptions 2.1 and 2.2 as the two basis regularity conditions; and sometimes call Assumption 2.1 the approximation condition and Assumption 2.2 the link condition. Both assumptions are satisfied by the ones imposed in the literature, such as those in Cohen, Hoffmann and Reiß (2004), Efromovich and Koltchinskii (2001), Hoffmann and Reiß (2007), Blundell, Chen and Kristensen (2007), Chen and Pouzo (2007) and others. In the next subsection we show that these two basic regularity conditions are automatically satisfied by the so-called "general source condition", which in turn are satisfied by conditions imposed in Hall and Horowitz (2005) and all the other papers using the general source condition.

2.2 Relation to source conditions

In the numerical analysis literature on ill-posed inverse problems it is common to measure the smoothness (regularity) of the function class $\mathscr{H}$ according to the spectral representation of the operator K^*K. Denote by $\|K\| := \sup_{h:\|h\|_X \leqslant 1} \|Kh\|_W$ the operator norm. The so-called "*general source condition*" assumes that there is a continuous function ψ defined on $[0, \|K\|^2]$ with $\psi(0) = 0$ and $\lambda^{-1/2}\psi(\lambda)$ non-decreasing such that

$$\mathscr{H}_{source} := \Big\{h = \psi(K^*K)g,\ g \in L^2_X,\ \|g\|^2_X \leqslant R\Big\},\ \text{for a finite constant } R, \tag{2.5}$$

and the original "source condition" corresponds to the choice $\psi(\lambda) = \lambda^{1/2}$ (see Engl, Hanke and Neubauer (1996)). If K^*K is compact with eigenvalue-eigenfunction system $\{\lambda_k, e_k\}$, then (2.5) is equivalent to

$$\mathscr{H}_{source} = \Big\{h = \sum_{k=1}^{\infty} \langle h, e_k\rangle_X e_k,\ \sum_{k=1}^{\infty} \frac{\langle h, e_k\rangle^2_X}{\psi^2(\lambda_k)} \leqslant R^2\Big\}.$$

Therefore the general source condition implies our Assumptions 2.1 and 2.2 by setting $u_k = e_k$ and $\nu_k^{-r} = \psi(\lambda_k)$ for all $k \geqslant 1$, and $\varphi(B^{-2}) = K^*K$.

In the econometrics literature on NPIV estimation, Darolles, Florens and Renault (2002) impose a smoothness condition on the true structure function h_0 that is closely related to the source condition. Precisely, they assume $h_0 \in \mathscr{H}_{DFR}$, where

$$\mathscr{H}_{DFR} = \Big\{h \in L^2_X,\ \sum_{k=1}^{\infty} \frac{\langle h, e_k\rangle^2_X}{(\lambda_k)^a} < \infty\Big\}, \quad \text{for some } a \geqslant 1. \tag{2.6}$$

Darolles, Florens and Renault (2002) use this assumption $h_0 \in \mathscr{H}_{DFR}$ to establish the convergence rate of their kernel-based Tikhonov regularized estimator in mean squared error metric $\mathbb{E}_{h_0}[\|\hat{h} - h_0\|_X^2]$. This rate, however, will not hold uniformly over $h_0 \in \mathscr{H}_{DFR}$, since the series in (2.6) is not uniformly bounded away from infinity, which is the role of $R \in (0, \infty)$ in the definition of $\mathscr{H}_{source}$.

Hall and Horowitz (2005) assume that h_0 belongs to a hyperrectangle in L_X^2, using the eigenfunctions $\{e_k\}$ of the operator K^*K as a basis:

$$\mathscr{H}_{HH} = \Big\{ h = \sum_{k=1}^{\infty} \langle h, e_k \rangle_X e_k, \ |\langle h, e_k \rangle_X| \leqslant R' k^{-\beta} \Big\}, \tag{2.7}$$

which, when $\beta > 1/2$ plays the role of $r + 1/2$, implies our Assumptions 2.1 and 2.2 by setting $u_k = e_k$, $\nu_k = k$ for all $k \geqslant 1$, and $\varphi(B^{-2}) = K^*K$. In addition, Hall and Horowitz (2005) also assume that the eigenvalues $\{\lambda_k\}$ of the operator K^*K are such that $\lambda_k \geqslant const.k^{-\alpha}$ for some $\alpha > 1$ and $2\beta > \alpha \geqslant \beta - \frac{1}{2}$, which suggests that we could set $\varphi(t) = t^{\alpha/2}$.

3 The lower bound

In this section we shall establish a minimax risk lower bound for the NPIV model under the two basic regularity conditions stated in Section 2. We derive this result by first establishing that the NPIV model is no more informative than the reduced form nonparametric indirect regression (NPIR) model. First, the following abstract assumption ensures a certain complexity of the statistical NPIV model and permits the residuals of Y given W to be Gaussian. Recall that $\mathscr{L}_Z$ denotes the law of the random vector Z.

3.1 Assumption. *Let $\sigma_0 > 0$ be a finite constant. Let $\mathscr{C}$ be a (possibly very large) set of elements $(\mathscr{L}_{UWX}, h)$ such that the following property holds:*

- *For all $h \in \mathscr{H}$, there is a law $\mathscr{L}_{UWX}$ with $(\mathscr{L}_{UWX}, h) \in \mathscr{C}$ such that $\mathscr{L}_{WY}$ is determined by $\mathscr{L}_{UWX}$ and h, and that*

  $$V_i := Y_i - \mathbb{E}[Y_i \,|\, W_i] = h(X_i) - (Kh)(W_i) + U_i$$

  given W_i is $N(0, \sigma^2(W_i))$-distributed with $\sigma^2(W_i) \geqslant \sigma_0^2$.

3.2 Example. *A typical NPIV model (2.1) satisfying Assumption 3.1 is generated by taking W_i from an arbitrary probability distribution $\mathscr{L}_W$, then generating X_i according to a conditional density of X given W, generating V_i according to $N(0, \sigma^2(W_i))$, and defining*

$$U_i := (Kh)(W_i) - h(X_i) + V_i.$$

3.1 Reduction from NPIV model to NPIR model

For each NPIV model, we specify the reduced form NPIR model as

$$Y_i = (Kh)(W_i) + V_i, \quad i = 1, \dots, n,$$

with (W_i, V_i) i.i.d., $\mathscr{L}_{V|W=w} = N(0, \sigma^2(w))$, $h \in \mathscr{H}$ the unknown structure function, and K a *known* injective operator from L^2_X to L^2_W. The observations corresponding to the NPIR are $\{(Y_i, W_i)\}_{i=1}^n$. We shall now formally prove, that the NPIV model is statistically more demanding than an indirect regression model with known operator K. We compare statistical experiments in a decision-theoretic sense (see Le Cam and Yang (2000)), and therefore, have to ensure first that the classes of parameters are compatible.

3.3 Definition. *Let Assumption 3.1 hold. The NPIR model class $\mathscr{C}_0$ consists of all model parameters $(\mathscr{L}_{W'}, \sigma(\bullet), h)$ such that there is $(\mathscr{L}_{UWX}, h) \in \mathscr{C}$ with the following properties: $\mathscr{L}_W = \mathscr{L}_{W'}$, $\sigma^2(w) \geqslant \sigma_0^2 > 0$, the conditional law $\mathscr{L}_{X|W}$ is prescribed according to K, and $\mathscr{L}_{U|WX}$ is arbitrary among the conditions imposed in $\mathscr{C}$.*

3.4 Lemma. *The NPIR model is more informative than the NPIV model in the sense that for each estimator $\hat{h}_n$ for the NPIV model there is an estimator $\tilde{h}_n$ for the NPIR model with*

$$\sup_{(\mathscr{L}_W, \sigma(\bullet), h) \in \mathscr{C}_0} \mathbb{E}_{(\mathscr{L}_W, \sigma(\bullet), h)}[\|\tilde{h}_n - h\|_X^2] \leqslant \sup_{(\mathscr{L}_{UWX}, h) \in \mathscr{C}} \mathbb{E}_{(\mathscr{L}_{UWX}, h)}[\|\hat{h}_n - h\|_X^2].$$

3.2 The lower bound

We now formally present the minimax risk lower bound for the NPIR and the NPIV models in mean squared error loss. We establish the lower bound by considering asymptotically least favorable Bayes priors, more specifically, by applying Assouad's cube technique; see e.g. Korostelev and Tsybakov (1993) or Yang and Barron (1999). In this paper we use the notation $a_n \asymp b_n$ to mean that there is a finite positive constant c such that $ca_n \leqslant b_n \leqslant c^{-1}a_n$.

Since $\sup_{h \in \mathscr{H}(r,R)} \mathbb{E}_h[\|\hat{h}_n - h\|_X^2] \geqslant \sup_{h \in H_R^r} \mathbb{E}_h[\|\hat{h}_n - h\|_X^2]$, it suffices to establish the lower bound for functions in H_R^r, a subset of $\mathscr{H}(r, R)$.

3.5 Theorem. *Let Assumptions 2.1 and 2.2 hold. For the NPIR model we have the following minimax risk lower bound:*

$$\inf_{\hat{h}_n} \sup_{h \in H_R^r} \mathbb{E}_h[\|\hat{h}_n - h\|_X^2] \geqslant \tfrac{\sigma_0^2}{4\exp(4M)} \delta_n, \quad \delta_n := n^{-1} \sum_{k=1}^m [\varphi(\nu_k^{-2})]^{-1},$$

where the infimum runs over all possible estimators $\hat{h}_n$ based on n observations, and m is the largest possible integer satisfying

$$\sigma_0^2 n^{-1} \sum_{k=1}^{m} \nu_k^{2r} [\varphi(\nu_k^{-2})]^{-1} \leqslant R^2.$$

(1) Mildly ill-posed case: Let $\varphi(t) = t^a$ and $\nu_k \asymp k^\epsilon$ for some $a, \epsilon > 0$. If $m \asymp n^{1/(2r\epsilon+2a\epsilon+1)}$, then $\delta_n \asymp n^{-2r/(2r+2a+\epsilon^{-1})}$.

(2) Severely ill-posed case: Let $\varphi(t) = \exp(-t^{-a/2})$, $\nu_k \asymp k^\epsilon$ for some $a, \epsilon > 0$. If $m = c\log(n)^{1/a\epsilon}$ with a sufficiently small $c > 0$, then $\delta_n \asymp (\log n)^{-2r/a}$.

The next corollary follows directly from Lemma 3.4 and Theorem 3.5; hence we omit its proof.

3.6 Corollary. *Let Assumptions 2.1, 2.2 and 3.1 hold. For the NPIV model we have the same minimax risk lower bound:*

$$\inf_{\hat{h}_n} \sup_{h \in H_R^r} \mathbb{E}_h[\|\hat{h}_n - h\|_X^2] \geqslant \tfrac{\sigma_0^2}{4\exp(4M)} \delta_n, \quad \textit{with } \delta_n \textit{ given in Theorem 3.5},$$

where the infimum runs over all possible estimators $\hat{h}_n$ based on n observations.

3.7 Remark. *For the proof of the lower bound we have to consider the likelihood between the observations. This is why we require Gaussianity. Nevertheless, the proof works the same for other error densities, but bounding the Kullback-Leibler or Hellinger distance between alternatives might be more cumbersome.*

Let us also mention that the proof strategy can also yield a lower bound for convergence in probability:

$$\inf_{\hat{h}_n} \sup_{h \in H_R^r} P_h\Big(\delta_n^{-1}\|\hat{h}_n - h\|_X^2 \geqslant \tfrac{\sigma_0^2}{4\exp(4M)}\Big) \geqslant c > 0, \quad \textit{with } \delta_n \textit{ given in Theorem 3.5},$$

cf. Korostelev and Tsybakov (1993).

Note that Assumption 2.2 is automatically satisfied under the general source condition with $K^*K = \varphi(B^{-2})$. Following the proof of Theorem 3.5, we immediately obtain:

3.8 Remark. *Suppose that Assumption 2.1 is satisfied with $h \in \mathscr{H}_{source}$ and $u_k = e_k$, $\nu_k^{-r} = \psi(\lambda_k)$ for all $k \geqslant 1$. Let $\varphi(B^{-2}) = K^*K$. Then, for NPIR model and for NPIV model (under Assumption 3.1), we have the same minimax risk lower bound:*

$$\inf_{\hat{h}_n} \sup_{h \in \mathscr{H}_{source}} \mathbb{E}_h[\|\hat{h}_n - h\|_X^2] \geqslant \tfrac{\sigma_0^2}{4\exp(4M)} \delta_n, \quad \textit{with } \delta_n \textit{ given in Theorem 3.5},$$

where the infimum runs over all possible estimators $\hat{h}_n$ based on n observations. An equivalent way to determine the lower bound δ_n is to choose the largest possible integer m such that

$$\delta_n = n^{-1} \sum_{k=1}^{m} \lambda_k^{-1}, \quad \sigma_0^2 n^{-1} \sum_{k=1}^{m} [\psi(\lambda_k)]^{-2} \lambda_k^{-1} \leqslant R^2.$$

4 An upper bound for the NPIR model

We prove an upper bound for the NPIR model. The aim of this section is to convince the reader that the lower bounds given in Section 3 are rate-optimal, and to provide an easy method to attain these rates. Again we assume that B has eigenvalues $\nu_k \uparrow \infty$ with corresponding L^2_X-normalized eigenfunctions (u_k) which then form an orthonormal basis of L^2_X. For $m \geqslant 1$ we define our estimator as

$$\hat{h}_n := \sum_{k=1}^{m} \hat{\eta}_k u_k, \quad \hat{\eta}_k := \frac{1}{n} \sum_{i=1}^{n} Y_i((K^*)^{-1} u_k)(W_i). \tag{4.1}$$

This simple projection procedure using the basis $\{u_k\}$ (of B) does not seem to have been studied before. It is a natural generalization of the well-known spectral cut-off method using the eigenfunction basis $\{e_k\}$ of K^*K. Given the prior information about $\mathscr{H}(r,R)$, this is a mathematically satisfactory construction.

For the upper bound we impose the following assumptions on the NPIR model.

4.1 Assumption.
(1) There is a finite $\sigma_1 > 0$ such that $\sigma(w) \leqslant \sigma_1$ for all $w \in \operatorname{supp}(\mathscr{L}_W)$;
(2) There is a finite $S > 0$ such that $\|Kh\|_\infty = \sup_{w \in \operatorname{supp}(\mathscr{L}_W)} |(Kh)(w)| \leqslant S$ for all $h \in \mathscr{H}(r,R)$.

Assumption 4.1 is typically assumed in papers on nonparametric estimation of ill-posed indirect regression; see, e.g. Bissantz, Hohage, Munk and Ruymgaart (2007). When K is the identity operator, Assumption 4.1(2) becomes to require that $\|h\|_\infty \leqslant S$ for all $h \in \mathscr{H}(r,R)$, which is a condition imposed in Yang and Barron (1999, theorems 6 and 7) to derive their minimax rate for a standard nonparametric regression model.

4.2 Assumption (reverse link condition)**.** *There is a finite $c > 0$ such that $\|Kh\|_W \geqslant c\|[\varphi(B^{-2})]^{1/2} h\|_X$ for all $h \in L^2_X$.*

Assumption 4.2 is the reverse condition of Assumption 2.2 and is often imposed in papers on ill-posed inverse problems. We shall sometimes call Assumptions 2.2 and 4.2 together as the exact link (or exact range) condition. See Subsection 4.2 for a relaxation of this condition.

4.3 Proposition. *For the NPIR models, suppose that Assumptions 2.1, 4.1 and 4.2 hold. Then the estimator $\hat{h}_n$ defined in (4.1) satisfies*

$$\sup_{h\in\mathscr{H}(r,R)} \mathbb{E}_h[\|\hat{h}_n - h\|_X^2] \leqslant \nu_{m+1}^{-2r}R^2 + 2n^{-1}(S^2+\sigma_1^2)c^{-2}\sum_{k=1}^{m}[\varphi(\nu_k^{-2})]^{-1}.$$

If $m = m(n)$ is such that $n^{-1}\sum_{k=1}^m \nu_k^{2r}[\varphi(\nu_k^{-2})]^{-1} \asymp 1$, then, under Assumption 2.2, this estimator $\hat{h}_n$ is rate-optimal in the minimax sense: there is a finite constant $C>0$ such that

$$\sup_{h\in\mathscr{H}(r,R)} \mathbb{E}_h[\|\hat{h}_n - h\|_X^2] \leqslant C\nu_{m+1}^{-2r} \asymp n^{-1}\sum_{k=1}^{m}[\varphi(\nu_k^{-2})]^{-1} \asymp \delta_n, \quad \textit{with } \delta_n \textit{ given in Theorem 3.5.}$$

(1) Mildly ill-posed case: Let $\varphi(t) = t^a$ and $\nu_k \asymp k^\epsilon$ for some $a, \epsilon > 0$. If $m \asymp n^{1/(2r\epsilon+2a\epsilon+1)}$, then $\delta_n \asymp n^{-2r/(2r+2a+\epsilon^{-1})}$.

(2) Severely ill-posed case: Let $\varphi(t) = \exp(-t^{-a/2})$, $\nu_k \asymp k^\epsilon$ for some $a, \epsilon > 0$. If $m = c\log(n)^{1/a\epsilon}$ with a sufficiently small $c > 0$, then $\delta_n \asymp (\log n)^{-2r/a}$.

4.4 Remark. *As the proof reveals, the upper bound does not require that the errors are Gaussian, the existence of second moments suffices.*

4.5 Remark. *When K is the identity operator, the NPIR model becomes the standard nonparametric mean regression model, and Assumption 2.2 is automatically satisfied with $\varphi()$ being a constant, then Theorem 3.5 and Proposition 4.3 together reproduce the well-known minimax lower and upper bounds for the nonparametric mean regression model (see, e.g., theorem 7 of Yang and Barron (1999)), in which $\delta_n \asymp \frac{m}{n}$, and m is the largest possible integer satisfying $\nu_{m+1}^{-2r} \asymp \frac{m}{n}$.*

Comparing the minimax optimal rates in mean integrated squared error loss for the nonparametric mean regression model and for the NPIR model, we see the squared bias is of the same order (ν_{m+1}^{-2r}), but the variance blow up from $\frac{m}{n}$ for the nonparametric mean regression model to $n^{-1}\sum_{k=1}^m[\varphi(\nu_k^{-2})]^{-1}$ for the NPIR model.

Notice that the minimax optimal rate $\delta_n \asymp (\log n)^{-2r/a}$ for the severely ill-posed case is independent of ϵ (hence independent of the dimension d of X). For the mildly ill-posed case, when $\varphi(t) = t^a$ and $\nu_k \asymp k^\epsilon$ for some $a > 0$ and $\epsilon = 1/d$, Theorem 3.5 and Proposition 4.3 together give the minimax optimal rate $\delta_n = n^{-2r/(2r+2a+d)}$ for the NPIR models. This rate is well known for the special case when $\mathscr{H}(r,R)$ is a d-dimensional Sobolev ball H_R^r and the operator K is elliptic with ill-posedness degree a (i.e., $\|Kh\|_W \asymp \|B^{-a}h\|_X$ for all $h \in L_X^2$); see, e.g., Cohen, Hoffmann and Reiß (2004).

Note that Assumptions 2.2 and 4.2 are automatically satisfied under the general source condition with $K^*K = \varphi(B^{-2})$. Applying Proposition 4.3 and Remark 3.8, we immediately obtain:

4.6 Remark. *For the NPIR models, suppose that Assumption 4.1 holds, and Assumption 2.1 is satisfied with $h \in \mathscr{H}_{source}$ and $u_k = e_k$, $\nu_k^{-r} = \psi(\lambda_k)$ for all $k \geqslant 1$. Let $\varphi(B^{-2}) = K^*K$. Then the estimator $\hat{h}_n$ defined in (4.1) with $u_k = e_k$ reaches the minimax rate uniformly over $h \in \mathscr{H}_{source}$:*

$$\sup_{h \in \mathscr{H}_{source}} \mathbb{E}_h[\|\hat{h}_n - h\|_X^2] \leqslant C\delta_n, \quad \textit{with } \delta_n \textit{ and } m \textit{ given in Remark 3.8.}$$

In the literature on ill-posed inverse problems with known operator K, there are available many other estimation procedures (like Tikhonov's method) that employ source conditions; some of which lead to rate-optimal estimators only for mildly ill-posed case. See, e.g., Bissantz, Hohage, Munk and Ruymgaart (2007) and Florens, Johannes and Van Bellegem (2007) for recent results.

4.1 Relaxation of the exact link condition

For the minimax risk lower bound we impose Assumption 2.2, and for the upper bound we use Assumption 4.2. Together, these two assumptions require that the operator K satisfies

$$c\|[\varphi(B^{-2})]^{1/2}h\|_X \leqslant \|Kh\|_W \leqslant M\|[\varphi(B^{-2})]^{1/2}h\|_X \text{ for all } h \in L_X^2,$$

which is equivalent to

$$\operatorname{ran}([\varphi(B^{-2})]^{1/2}) = \operatorname{ran}(|K|). \tag{4.2}$$

This is a standard condition imposed even in books and papers on ill-posed inverse problems with deterministic errors; see, e.g., Engl, Hanke and Neubauer (1996), Nair, Pereverzev and Tautenhahn (2005) and the references therein. This condition usually holds when K acts exactly along certain function classes; see Section 6 for such an example. Moreover, this exact range condition 4.2 is automatically satisfied under the source condition with $K^*K = \varphi(B^{-2})$. However, Assumption 4.2 may fail more generally. Luckily, this assumption is not strictly necessary.

Let us indicate one possibility how Assumption 4.2 can be relaxed to requiring

$$\operatorname{ran}[\varphi(B^{-2})]^{1/2} \subseteq \operatorname{ran}|K| + L, \text{ for some finite-dimensional linear space } L.$$

To keep it simple, we consider the case that the subspace L is spanned by one eigenfunction u_ℓ of B with $u_\ell \notin \operatorname{ran}|K|$ and $1 \leqslant \ell \leqslant m$. Then the simple estimator $\hat{h}_n$ using $\hat{\eta}_\ell$ given in (4.1) is no longer well defined, but we can consider for some $v \in L_W^2$ the estimator

$$\tilde{\eta}_\ell := \frac{1}{n}\sum_{i=1}^n Y_i v(W_i), \quad \tilde{h}_n := \sum_{k=1, k\neq\ell}^m \hat{\eta}_k u_k + \tilde{\eta}_\ell u_\ell.$$

Following the bias variance decomposition in the proof of Proposition 4.3, we obtain

$$\begin{aligned}\mathbb{E}[(\tilde{\eta}_\ell - \langle h, u_\ell\rangle_X)^2] &\leqslant (\langle Kh, v\rangle_W - \langle h, u_\ell\rangle_X)^2 + 2n^{-1}(S^2 + \sigma_1^2)\|v\|_W^2 \\ &\asymp \langle h, K^* v - u_\ell\rangle_X^2 + n^{-1}\|v\|_W^2. \end{aligned} \tag{4.3}$$

The definition of H_R^r implies with some uniform constant $C > 0$

$$\sup_{h\in H_R^r} \mathbb{E}[(\tilde{\eta}_\ell - \langle h, u_\ell\rangle_X)^2] \leqslant C(R^2\|B^{-r}(K^* v - u_\ell)\|_X^2 + n^{-1}\|v\|_W^2). \tag{4.4}$$

¿From inequality (4.4), it is easy to derive that this error in estimating the coefficient $\langle h, u_\ell\rangle_X$ is minimized by

$$v = (KK^* + n^{-1}R^{-2}B^{2r})^{-1}Ku_\ell,$$

which is always well-defined. Consequently, in terms of minimax optimal rate over the class of functions H_R^r, the rate in Proposition 4.3 does not deteriorate if we use $\tilde{\eta}_\ell$ instead of $\hat{\eta}_\ell$ and its error bound

$$n^{-1}\|(KK^* + n^{-1}R^{-2}B^{2r})^{-1}Ku_\ell\|_W^2$$

is not larger than the minimax optimal rate. See Section 6 for a concrete example.

5 An upper bound for the NPIV model

We now provide an upper bound for the NPIV model. For the NPIV model additional considerations due to the unknown conditional expectation operator are necessary. It is, of course, more complex to construct an estimator that is rate-optimal for the NPIV model than for the NPIR model, which is why the approaches in the literature are more diverse and require different additional assumptions. Here, we restrict ourselves to presenting a simple estimator to illustrate that it is possible to construct a rate-optimal estimator for the NPIV model in both mildly ill-posed and severely ill-posed cases based on the SMD estimator of Newey and Powell (2003), Ai and Chen (2003) and Blundell, Chen and Kristensen (2007). First, for each integer $J \geqslant 1$, we denote by $\text{span}\{p_1, ..., p_J\}$ a J-dimensional linear subspace of L_W^2 that becomes dense in L_W^2 as $J \to \infty$. Let $P^{J_n}(w) = (p_1(w), ..., p_{J_n}(w))'$ and $\mathbf{P} = (P^{J_n}(W_1), ..., P^{J_n}(W_n))'$. We compute a sieve least squares estimator of $E[Y - h(X)|W = \bullet]$ as

$$\widehat{\mathbb{E}}[Y - h(X)|W = \bullet] = \sum_{t=1}^{n}\{Y_t - h(X_t)\}P^{J_n}(W_t)'(\mathbf{P}'\mathbf{P})^{-1}P^{J_n}(\bullet).$$

For each integer $m \geqslant 1$, we denote by $\mathcal{H}_m := \text{span}\{\psi_1, ..., \psi_m\}$ an m-dimensional linear subspace of L^2_X that becomes dense in L^2_X as $m \to \infty$. Then we compute the SMD estimator of the true structure function h_0 as

$$\hat{h}_n = \text{argmin}_{h \in \mathcal{H}_{m(n)} \cap \mathscr{H}(r,R)} \frac{1}{n} \sum_{i=1}^{n} \left\{ \widehat{\mathbb{E}}[Y - h(X)|W = W_i] \right\}^2. \tag{5.1}$$

Depending on the prior information about $\mathscr{H}(r, R)$, sometimes one may compute $\hat{h}_n$ in closed form. For example, if $\mathscr{H}(r, R) = H^r_R$ and the density of X is bounded below and above by positive constants, then

$$\hat{h}_n(x) = \sum_{k=1}^{m} \hat{\pi}_k \psi_k(x) = \psi^m(x)'\widehat{\Pi}, \tag{5.2}$$

$$\widehat{\Pi} = \left(\Psi' \mathbf{P}(\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}'\Psi + \hat{\lambda}\mathbf{C} \right)^{-1} \Psi'\mathbf{P}(\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}'\mathcal{Y}, \tag{5.3}$$

with $\Psi = (\psi^m(X_1), ..., \psi^m(X_n))'$, $\mathcal{Y} = (Y_1, ..., Y_n)'$, the penalization matrix $\mathbf{C} = \int \{[B^r\psi^m(x)][B^r\psi^m(x)]'\}dx$ and $\hat{\lambda}$ satisfies $\widehat{\Pi}'\mathbf{C}\widehat{\Pi} = R^2$.

In addition to the assumptions on the NPIR models, we impose the following:

5.1 Assumption. *The basis $\{\psi_k\}$ is a Riesz basis associated with the operator B, that is, $\sum_{k=1}^{\infty} \langle h, \psi_k \rangle^2_X \asymp \sum_{k=1}^{\infty} \langle h, u_k \rangle^2_X$ for all $h \in L^2_X$.*

Assumption 5.1 allows for the use of a Riesz basis $\{\psi_k\}$ instead of the ideal orthonormal basis $\{u_k\}$ to approximate the unknown structure function $h \in \mathscr{H}(r, R)$ with the same order of the approximation errors. Of course in applications, we need some information about the tail behavior of the density of X before we can construct such a basis. For example, if we know that the density of X is bounded above and below by finite positive constants , then we could use wavelets as the $\{\psi_k\}$.

5.2 Assumption.
(1) $\mathbb{E}[Y - \Pi_m(h(X))|W = \bullet]$ belongs to $\Lambda^{r_K}_c(\mathscr{W})$ (Hölder ball of regularity r_K) for any $\Pi_m(h) \in \mathcal{H}_m$;
(2) (i) the smallest and the largest eigenvalues of $\mathbb{E}\{P^{J_n}(W)P^{J_n}(W)'\}$ are bounded and bounded away from zero for each J_n; (ii) $P^{J_n}(W)$ is a tensor product of either a cosine series or a B-spline basis of order γ_b or a wavelet basis of order γ_b, with $\gamma_b > r_K > d/2$;
(3) the density of W is continuous and bounded away from zero over its support $\mathscr{W}$, which is a compact connected subset in $\mathbb{R}^d$ with Lipschitz continuous boundaries and non-empty interior;
(4) (i) $J_n \to \infty$ and $J_n^2/n \to 0$; (ii) $\lim_n \frac{J_n}{m(n)} = c \in (1, \infty)$ and $J_n > m(n)$.

Assumption 5.2 implies that the sieve least square estimate $\widehat{\mathbb{E}}[h(X)|W = \bullet]$ of $\mathbb{E}[h(X)|W = \bullet]$ performs well; see e.g., Blundell, Chen and Kristensen (2007) for details.

5.3 Theorem. *For the NPIV models, suppose that Assumptions 2.1, 2.2, 4.1, 4.2, 5.1 and 5.2 hold. Then the estimator* $\hat{h}_n$ *defined in (5.1) satisfies*

$$\|\hat{h}_n - h\|_X^2 \leqslant C \max\left\{\nu_{m+1}^{-2r},\ \frac{m}{n}[\varphi(\nu_m^{-2})]^{-1}\right\}$$

uniformly over $h \in \mathscr{H}(r, R)$ *except on an event whose probability tends to zero as* $n \uparrow \infty$. *If* $m = m(n)$ *is such that* $n^{-1}\sum_{k=1}^{m} \nu_k^{2r}[\varphi(\nu_k^{-2})]^{-1} \asymp 1$, *then this estimator* $\hat{h}_n$ *is rate-optimal in the minimax sense: there is a finite constant* $C > 0$ *such that*

$$\|\hat{h}_n - h\|_X^2 \leqslant C\nu_{m+1}^{-2r} \asymp \frac{m}{n}[\varphi(\nu_m^{-2})]^{-1} \asymp \delta_n, \quad \text{with } \delta_n \text{ given in Theorem 3.5},$$

uniformly over $h \in \mathscr{H}(r, R)$ *except on an event whose probability tends to zero as* $n \uparrow \infty$.

(1) Mildly ill-posed case: Let $\varphi(t) = t^a$ *and* $\nu_k \asymp k^\epsilon$ *for some* $a, \epsilon > 0$. *If* $m \asymp n^{1/(2r\epsilon+2a\epsilon+1)}$, *then* $\delta_n \asymp n^{-2r/(2r+2a+\epsilon^{-1})}$.

(2) Severely ill-posed case: Let $\varphi(t) = \exp(-t^{-a/2})$, $\nu_k \asymp k^\epsilon$ *for some* $a, \epsilon > 0$. *If* $m = c\log(n)^{1/a\epsilon}$ *with a sufficiently small* $c > 0$, *then* $\delta_n \asymp (\log n)^{-2r/a}$.

This minimax rate theorem appears to be new in the literature, and can be proved by slightly modifying the proof of Blundell, Chen and Kristensen (2007) for their theorem 2. Hall and Horowitz (2005) obtained minimax optimal rate $\sup_{h\in\mathscr{H}_{HH}} \mathbb{E}_h[\|\hat{h}_n - h\|_X^2] \leqslant C\delta_n$ for their estimators in the mildly ill-posed case for the class of functions $\mathscr{H}_{HH}$ defined in (2.7). Hoffmann and Reiß (2007) propose a wavelet estimator in the case of an unknown operator K that is elliptic with ill-posedness degree a. They assume there exists an estimator of K with specified rate, and their class of functions $\mathscr{H}(r, R)$ is a Besov ball that could be bigger than the function class defined in our Assumption 2.1, but they do not consider severely ill-posed case.

6 More on regularity conditions

In this section, we use examples to discuss the pros and cons of the approach of imposing two basic regularity conditions (the approximation and the link conditions) versus the other approach of using the general source condition. To simplify the discussion, here we assume the operator K is known. In the first class of examples, the operator K has very smooth eigenfunction basis (in the sense that its eigenfunctions are many times differentiable), while in the second class of examples, the operator K has eigenfunctions that are not differentiable.

6.1 Examples of K having infinitely times differentiable eigenfunctions

Suppose that the $\{W_i\}_{i=1}^n$ are uniformly distributed on $[0,1]$ and K is a circular convolution operator on $L^2([0,1])$: $Kh(w) = \int_0^1 k(x-w)h(x)dx$ with a 1-periodic function k that satisfies $k(-x) = k(x)$ and has Fourier coefficients $|\mathscr{F}k(m)| = |\int_0^1 k(x)\cos(mx)\,dx| \asymp (1+|m|)^{-a}$. Then K is a positive-definite self-adjoint operator which is diagonalized by the Fourier basis.

Source condition: In this canonical case the exact link between K and B is easily established with $B = K^{-1}$, $\varphi(t) = t$ hence $\|Kh\|_W = \|[\varphi(B^{-2})]^{1/2}h\|_X$ for all $h \in L^2([0,1])$. The smoothness of the unknown function h is also described using $B = K^{-1}$; hence the Hilbert scale space H^r (generated by B) is equal to the classical periodic Sobolev space H^{ra}_{per} of smoothness (or regularity) ra. Applying Remark 4.6, we obtain minimax optimal rate for this scale of periodic Sobolev spaces.

Approximation + link conditions: Suppose $\{u_k\}_{k\geqslant 1}$ is an orthonormal basis of $L^2([0,1])$ such that $\|Kg\|^2_{L^2([0,1])} \asymp \sum_{k=1}^{\infty} k^{-2a}\langle g, u_k\rangle^2$. A typical example is given by sufficiently regular periodized wavelet bases (see Cohen, Daubechies and Vial (1993)). Then we can define

$$Bg := \sum_{k\geqslant 1} k\langle g, u_k\rangle u_k,$$

and the Hilbert scale spaces H^r can be interpreted as approximation spaces for the basis (u_k). In the convolution example we obtain $\|Kg\|_W \asymp \|B^{-a}g\|_X$. Consequently, the exact link conditions (assumptions 2.2 and 4.2) between K and B hold with $\varphi(t) = t^a$. Applying Proposition 4.3, we obtain minimax optimal rate for the Hilbert scale space H^r generated by B.

The Hilbert scale of approximation spaces generated by B does not necessarily coincide with the Hilbert scale generated by K. The most pronounced example is the case $a < 1/2$, where all non-periodic wavelets on an interval still satisfy $\|Kg\|^2_{L^2([0,1])} \asymp \sum_{k=1}^{\infty} k^{-2a}\langle g, u_k\rangle^2$ (see Cohen, Daubechies and Vial (1993)). Hence, the approximation spaces for unknown true structure function need not exhibit any boundary condition. This means that a smooth, but non-periodic function on $[0,1]$ will have high regularity r in terms of the approximation space, while it is an element in periodic Sobolev spaces up to regularity $1/2$ only. If we have in mind that our true function h is smooth, but not periodic, we should therefore rather choose the approximation space approach. On the other hand, wavelets work well just to some maximal regularity and they will therefore reconstruct very smooth and periodic functions not as well as the Fourier basis.

If K is more ill-posed, that is $a \geqslant 1/2$, we can adopt the ideas explained in Subsection 4.2. We remain in the approximation space framework and use non-periodic compactly supported wavelets as basis functions $\{u_k\}$. Only the wavelets ψ_λ with support at the boundary are not in the periodic Sobolev spaces H^s_{per}, $s \geqslant 1/2$. Using some (statistical) kernel function $L_h : [-b, b] \to \mathbb{R}$ of bandwidth b, we can consider the periodically smoothed version

$$\tilde{\psi}_\lambda(x) := \int_{-b}^{b} \psi_\lambda(\{x-y\}) L_b(y)\, dy, \quad x \in [0,1],$$

where $\{z\} = z - \lfloor z \rfloor \in [0,1)$ denotes the fractional part of $z \in \mathbb{R}$. If L and ψ_λ are sufficiently often differentiable, then $\tilde{\psi}_\lambda$ lies in the range H^a_{per} of K. Using $v = K^{-1}\tilde{\psi}_\lambda$ in equation (4.3), standard kernel estimates ($h \in H^r_R$ implies $h \in H^s_{per}$ for all $s \leqslant r$ and $s < 1/2$) show that for all $h \in H^r_R$ (with adapted notation)

$$\mathbb{E}[(\tilde{\eta}_\lambda - \langle h, \psi_\lambda \rangle)^2] \leqslant C_1(\langle h, \tilde{\psi}_\lambda - \psi_\lambda \rangle^2 + n^{-1}\|K^{-1}\tilde{\psi}_\lambda\|^2) \leqslant C_2(b^{2s} + n^{-1}b^{-2a}).$$

Optimizing over b, we infer that $\langle h, \psi_\lambda \rangle$ can be estimated at rate $n^{-s/(s+a)}$, which for $r \geqslant 1/2$ is nearly $n^{-1/(2a+1)}$. Since in a wavelet approximation space of dimension 2^J only of the order J wavelets lie at the boundary, the rate in estimating h will be $n^{-s/(s+a)}\log(n) + n^{-2r/(2r+2a+1)}$, which for $r \geqslant \frac{1}{2} + \frac{1}{4a}$ is roughly $n^{-1/(2a+1)}$. If we had taken a method based on the source condition approach (like projection on eigenfunctions of K, or Tikhonov methods) the best achievable rate would have been roughly $n^{-1/(2a+2)}$.

6.2 Examples of K having non-differentiable eigenfunctions

Depending on applications, it is perfectly conceivable that the eigenfunctions of K are rough while the basis functions u_k of B are smooth (or differentiable). For example, we can use the Haar basis $\psi_{jk}(x) = \psi(2^j x - k)$ on $L^2([0,1])$ ($\psi(x) = \mathbf{1}_{[0,1/2]} - \mathbf{1}_{[1/2,1]}$, $j \in \mathbb{N}_0$, $k = 0, \dots, 2^j - 1$, and $\psi_{-1,0} = \mathbf{1}_{[0,1]}$) and define – somewhat artificially – in this Haar basis

$$K\psi_{jk} := 2^{\alpha j}\psi_{jk}.$$

Then K is self-adjoint with eigenfunctions ψ_{jk}, which are step functions. For $\alpha r < 1/2$, the Hilbert scale H^r of K (or of the Harr basis) will be a Sobolev space, whereas for any $\alpha r \geqslant 1/2$ this Hilbert scale H^r will not be described in terms of traditional smoothness. Note that this H^r will always contain piecewise constant jump functions. Nevertheless, the larger r the less complex is the function class $\mathscr{H}(r,R)$, that is the smaller the approximation error rate. As for the convolution operator we could instead define the function class $\mathscr{H}(r,R)$ in terms of a basis $\{u_k\}$ associated to B which is smoother and satisfies at the same time the link conditions of Assumptions 2.2 and 4.2.

In conclusion, we see that rate-optimal methods may behave poorly if the function of interest, the structure function h, is not regular in the setting for which the method is designed. An important part of the specification of rate optimality is therefore always the associated function class.

7 Perspectives

In this paper, we clarify the relations between the existing sets of regularity conditions for convergence rates of NPIV regression models. We establish minimax risk lower bounds in mean squared error loss for the NPIV models under two basic regularity conditions that allow for both mildly ill-posed and severely ill-posed cases. We also show that the simple SMD estimator achieves the minimax risk lower bound, hence is rate-optimal for both mildly ill-posed and severely ill-posed cases.

Many of the ideas in this paper can be easily adapted to treat other kinds of ill-posed inverse problems in econometrics. For instance, when the problem is mildly ill-posed, Horowitz and Lee (2007) show that their kernel based Tikhonov regularized estimator of nonparametric quantile instrumental variables (IV) regression reaches the minimax rate under conditions very similar to those imposed in Hall and Horowitz (2005) for NPIV regression. Similarly, one could show that the penalized SMD estimator proposed in Chen and Pouzo (2007) for nonlinear and possibly nonsmooth nonparametric conditional moment models is also rate-optimal, as their estimator achieves the minimax risk lower bounds established in our paper for the NPIV regression model.

Once this is established, the intriguing open problem remains how to choose the regularization parameters adaptively from the data, not knowing the true regularity, and even to select among the different proposed procedures (e.g. generated by different operators B) in a data-driven way.

References

[1] Ai, C. and X. Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795-1844.

[2] Bissantz, N., T. Hohage, A. Munk and F. Ruymgaart (2007). Convergence rates of general regularization methods for statistical inverse problems and applications. *SIAM J. Numerical Analysis*, to appear.

[3] Blundell, R., X. Chen and D. Kristensen (2007): "Semi-nonparametric IV Estimation of Shape-Invariant Engel Curves," forthcoming in *Econometrica*.

[4] Blundell, R. and J.L. Powell (2003): "Endogeneity in Semiparametric and Nonparametric Regression Models," In: *Advances in Economics and Econometrics: Theory and Applications*, Dewatripont, M., Hansen, L.P., and Turnovsky, S.J., eds, vol. 2, pp. 312-357. Cambridge, UK: Cambridge University Press.

[5] Carrasco, M., J.-P. Florens and E. Renault (2007): "Linear Inverse Problems in Structural Econometrics Estimation Based on Spectral Decomposition and Regularization", in J.J. Heckman and E.E. Leamer (eds.), *The Handbook of Econometrics*, vol. 6. North-Holland, Amsterdam, forthcoming.

[6] Chen, X. and D. Pouzo (2007): "On Estimation of Semi/nonparametric conditional moment models with possibly nonsmooth moments," manuscript, Yale University and New York University, Dept. of Economics.

[7] Chernozhukov, V. and C. Hansen (2005): "An IV Model of Quantile Treatment Effects, *Econometrica*, 73, 245-261.

[8] Chernozhukov, V., G. Imbens, and W. Newey (2007): "Instrumental Variable Estimation of Nonseparable Models," forthcoming in *Journal of Econometrics*.

[9] Cohen, A., I. Daubechies, and P. Vial (1993). Wavelets on the interval and fast wavelet transforms. *Appl. Comp. Harm. Anal.* **1**(1), 54–81.

[10] Cohen, A., M. Hoffmann, and M. Reiss (2004): "Adaptive wavelet Galerkin methods for linear inverse problems," *SIAM J. Numer. Anal.* 42, 1479-1501.

[11] Darolles, S., J.-P. Florens and E. Renault (2002): "Nonparametric Instrumental Regression," mimeo, GREMAQ, University of Toulouse.

[12] Donoho, D., R. Liu and B. MacGibbon (1990): "Minimax Risk Over Hyperrectangles, and Implications," *Annals of Statistics.* 18, 1416-1437.

[13] Edmunds, D. and W. Evans (1987): *Spectral Theory and Differential Operators*, Oxford University Press: Oxford.

[14] Efromovich, S. and V. Koltchinskii (2001): "On Inverse Problems with Unknown Operators," *IEEE Trans. on Information Theory*, 47, 2876-2893.

[15] Engl, H., M. Hanke and A. Neubauer (1996): *Regularization of Inverse Problems*, Kluwer Academic Publishers: London.

[16] Florens, J.-P. (2003): "Inverse Problems and Structural Econometrics: the Example of Instrumental Variables," In: *Advances in Economics and Econometrics: Theory and Applications*,

Dewatripont, M., Hansen, L.P., and Turnovsky, S.J., eds, vol. 2, pp. 284-311. Cambridge, UK: Cambridge University Press.

[17] Florens, J.-P., J. Johannes and S. Van Bellegem (2007): "Identification and Estimation by Penalization in Nonparametric Instrumental Regression," mimeo, GREMAQ, University of Toulouse.

[18] Gagliardini, P. and O. Scaillet (2006):"Tikhonov Regularisation for Functional Minimum Distance Estimation," mimeo

[19] Hall, P. and J. Horowitz (2005): "Nonparametric Methods for Inference in the Presence of Instrumental Variables," *Annals of Statistics*, 33, 2904-2929.

[20] Hoffmann, M. and M. Reiss (2007): "Nonlinear estimation for linear inverse problems with error in the operator," *Annals of Statistics*, to appear.

[21] Horowitz, J. and S. Lee (2007): "Nonparametric Instrumental Variables Estimation of a Quantile Regression Model," forthcoming in *Econometrica*.

[22] Korostelev, A. and A. Tsybakov (1993): *Minimax Theory of Image Reconstruction*, Lecture Notes in Statistics, v.82. Springer: New York.

[23] Le Cam, L. and G. Lo Yang (2000): *Asymptotics in statistics. Some basic concepts.* 2nd ed., Springer, New York.

[24] Nair, M., S. Pereverzev and U. Tautenhahn (2005): "Regularization in Hilbert scales under general smoothing conditions," *Inverse Problems*, 21, 1851-1869.

[25] Newey, W.K. and J. Powell (2003): "Instrumental Variables Estimation for Nonparametric Models," *Econometrica*, 71, 1565-1578.

[26] Yang, Y. and A. Barron (1999): Information-theoretic determination of minimax rates of convergence, *Ann. Stat.* 27(5), 1564–1599.

Appendix: Proofs

Proof of Lemma 3.4. Let $\hat{h}_n = \hat{h}_n(\{(X_i, Y_i, W_i)\}_{i=1}^n)$ be an estimator for the NPIV model. Knowing the operator K amounts to knowing the conditional law of X_i given W_i. Let us call the observations in the NPIR model $\{(Y_i', W_i')\}_{i=1}^n$ for some $(\mathscr{L}_W, \sigma(\bullet), h) \in \mathscr{C}_0$. We then generate artificially i.i.d. observations X_i' according to the conditional law $\mathscr{L}_{X|W=w}$ with $w = W_i'$. Then the observations $\{(X_i', Y_i', W_i')\}_{i=1}^n$ follow the law of some

$(\mathscr{L}_{UWX}, h) \in \mathscr{C}$ because $Y_i' = h(X_i') + U_i'$ holds with $U_i' = (Kh)(W_i') - h(X_i') + V_i'$ satisfying $\mathbb{E}[U_i' \mid W_i'] = 0$ and $\mathscr{L}_{V'|W'=w} = N(0, \sigma^2(w))$. Consequently, the (randomized) estimator $\tilde{h}_n(\{(Y_i', W_i')\}_{i=1}^n) := \hat{h}_n(\{(X_i', Y_i', W_i')\}_{i=1}^n)$ has the same risk under $(\mathscr{L}_W, \sigma(\bullet), h) \in \mathscr{C}_0$ as $\hat{h}_n$ has under $(\mathscr{L}_{U'W'X'}, h) \in \mathscr{C}$, and is thus not larger than the maximal risk over $\mathscr{C}$. □

Proof of Theorem 3.5. We consider for $\vartheta = (\vartheta_k)$ with $\vartheta_k \in \{-1, +1\}$ and a sequence (γ_k), to be specified below, the following functions in L_X^2:

$$h_\vartheta := \sum_{k=1}^m \vartheta_k \gamma_k u_k.$$

The property $h_\vartheta \in H_R^r$ yields the following constraint on m and (γ_k):

$$\|h_\vartheta\|_r^2 = \sum_{k=1}^m \nu_k^{2r} \gamma_k^2 \leqslant R^2.$$

For $\ell = 1, \dots, m$ and each ϑ introduce $\vartheta^{(\ell)}$ by $\vartheta_k^{(\ell)} = \vartheta_k$ for $k \neq \ell$ and $\vartheta_\ell^{(\ell)} = -\vartheta_\ell$. Then because of the Gaussianity of the V_i given W_i the log-likelihood of $\mathbb{P}_{\vartheta^{(\ell)}}$ w.r.t. $\mathbb{P}_\vartheta$ is

$$\log\Big(\frac{d\,\mathbb{P}_{\vartheta^{(\ell)}}}{d\,\mathbb{P}_\vartheta}\Big) = \sum_{i=1}^n \pm \frac{2\gamma_\ell (Ku_\ell)(W_i)}{\sigma^2(W_i)} V_i - \frac{1}{2}\sum_{i=1}^n \Big(\frac{2\gamma_\ell (Ku_\ell)(W_i)}{\sigma(W_i)}\Big)^2.$$

Its expectation satisfies

$$\begin{aligned}
\mathbb{E}_\vartheta\Big[\log\Big(\frac{d\,\mathbb{P}_{\vartheta^{(\ell)}}}{d\,\mathbb{P}_\vartheta}\Big)\Big] &= -2\gamma_\ell^2 n \|(Ku_\ell)\sigma^{-1}\|_W^2 \\
&\geqslant -2M\sigma_0^{-2}\gamma_\ell^2 n \|[\varphi(B^{-2})]^{1/2} u_\ell\|_X^2 \\
&= -2M\sigma_0^{-2}\gamma_\ell^2 n \varphi(\nu_\ell^{-2}) =: \mu_\ell.
\end{aligned}$$

In terms of the Kullback-Leibler divergence this means $\mathrm{KL}(\mathbb{P}_{\vartheta^{(\ell)}}, \mathbb{P}_\vartheta) \leqslant -\mu_\ell$. More explicitly, we obtain by Markov's inequality

$$\mathbb{P}_\vartheta\Big(-\frac{1}{2}\sum_{i=1}^n \Big(\frac{2\gamma_\ell (Ku_\ell)(W_i)}{\sigma(W_i)}\Big)^2 \leqslant -2\mu_\ell\Big) \leqslant \frac{-\mu_\ell}{-2\mu_\ell} = \frac{1}{2}.$$

Using the symmetry of the distribution of V_i given W_i, we infer by conditioning on $(W_i)_{1 \leqslant i \leqslant n}$

$$\mathbb{P}_\vartheta\Big(\frac{d\,\mathbb{P}_{\vartheta^{(\ell)}}}{d\,\mathbb{P}_\vartheta} \geqslant \exp(2\mu_\ell)\Big) = \mathbb{E}_\vartheta\Big[\mathbb{P}_\vartheta\Big(\log\Big(\frac{d\,\mathbb{P}_{\vartheta^{(\ell)}}}{d\,\mathbb{P}_\vartheta}\Big) \geqslant 2\mu_\ell \,\Big|\, (W_i)_{1 \leqslant i \leqslant n}\Big)\Big] \geqslant \frac{1}{2}.$$

We calculate for each estimator $\hat{h}_n$:

$$
\begin{aligned}
&\sup_{h \in H_R^r} \mathbb{E}_h[\|\hat{h}_n - h\|_X^2] \\
&\geqslant \sup_{\vartheta \in \{-1,+1\}^m} \mathbb{E}_\vartheta[\|\hat{h}_n - h_\vartheta\|_X^2] \\
&\geqslant 2^{-m} \sum_{\vartheta \in \{-1,+1\}^m} \sum_{k=1}^m \mathbb{E}_\vartheta[\langle \hat{h}_n - h_\vartheta, u_k\rangle_X^2] \\
&= \sum_{k=1}^m 2^{-m} \sum_{\vartheta \in \{-1,+1\}^m} \frac{1}{2}\Big(\mathbb{E}_\vartheta[\langle \hat{h}_n - h_\vartheta, u_k\rangle_X^2] + \mathbb{E}_{\vartheta^{(k)}}[\langle \hat{h}_n - h_{\vartheta^{(k)}}, u_k\rangle_X^2]\Big) \\
&= \sum_{k=1}^m 2^{-m} \sum_{\vartheta \in \{-1,+1\}^m} \frac{1}{2}\, \mathbb{E}_\vartheta \Big[\langle \hat{h}_n - h_\vartheta, u_k\rangle_X^2 + \langle \hat{h}_n - h_{\vartheta^{(k)}}, u_k\rangle_X^2 \frac{d\,\mathbb{P}_{\vartheta^{(k)}}}{d\,\mathbb{P}_\vartheta}\Big] \\
&\geqslant \sum_{k=1}^m 2^{-m} \sum_{\vartheta \in \{-1,+1\}^m} \frac{\exp(2\mu_k)}{2}\, \mathbb{E}_\vartheta \Big[\Big(\langle \hat{h}_n - h_\vartheta, u_k\rangle_X^2 + \langle \hat{h}_n - h_{\vartheta^{(k)}}, u_k\rangle_X^2\Big) \times \\
&\qquad \times \mathbf{1}\Big\{\tfrac{d\,\mathbb{P}_{\vartheta^{(k)}}}{d\,\mathbb{P}_\vartheta} \geqslant \exp(2\mu_k)\Big\}\Big] \\
&\geqslant \sum_{k=1}^m 2^{-m} \sum_{\vartheta \in \{-1,+1\}^m} \frac{\exp(2\mu_k)}{8} \langle h_\vartheta - h_{\vartheta^{(k)}}, u_k\rangle_X^2\, \mathbb{P}_\vartheta\Big(\frac{d\,\mathbb{P}_{\vartheta^{(k)}}}{d\,\mathbb{P}_\vartheta} \geqslant \exp(2\mu_k)\Big) \\
&\geqslant \sum_{k=1}^m \frac{\exp(2\mu_k)}{4} \gamma_k^2.
\end{aligned}
$$

We choose $\gamma_k = \sigma_0 n^{-1/2}[\varphi(\nu_k^{-2})]^{-1/2}$ such that $\mu_k = -2M$ and then pick the largest $m \geqslant 1$ such that $\sum_{k=1}^m \nu_k^{2r}\gamma_k^2 \leqslant R^2$.

This gives the lower bound

$$
\inf_{\hat{h}_n} \sup_{h \in \mathscr{H}(r,R)} \mathbb{E}_h[\|\hat{h}_n - h\|_X^2] \geqslant \inf_{\hat{h}_n} \sup_{h \in H_R^r} \mathbb{E}_h[\|\hat{h}_n - h\|_X^2] \geqslant \tfrac{\sigma_0^2}{4\exp(4M)} n^{-1} \sum_{k=1}^m [\varphi(\nu_k^{-2})]^{-1}
$$

where m is largest possible with $\sum_{k=1}^m \nu_k^{2r}\gamma_k^2 \leqslant R^2$, i.e.

$$
\sigma_0^2 n^{-1} \sum_{k=1}^m \nu_k^{2r}[\varphi(\nu_k^{-2})]^{-1} \leqslant R^2.
$$

(1) (mildly ill-posed case): When $\varphi(t) = t^a$ and $\nu_k \asymp k^\epsilon$ for some $a, \epsilon > 0$, we have asymptotically as $n \to \infty$:

$$
n^{-1} \sum_{k=1}^m \nu_k^{2r}[\varphi(\nu_k^{-2})]^{-1} = n^{-1} \sum_{k=1}^m k^{2\epsilon r + 2\epsilon a} \asymp n^{-1} m^{2\epsilon r + 2\epsilon a + 1}.
$$

Hence, choosing $m \asymp n^{1/(2\epsilon r + 2\epsilon a + 1)}$ we obtain the asymptotic lower bound

$$
\delta_n \asymp n^{-1} \sum_{k=1}^m [\varphi(\nu_k^{-2})]^{-1} \asymp n^{-1} m^{2\epsilon a + 1} \asymp n^{-2r/(2r + 2a + \epsilon^{-1})}.
$$

(2) (severely ill-posed case): When $\varphi(t) = \exp(-t^{-a/2})$, $\nu_k \asymp k^\epsilon$ for some $a, \epsilon > 0$, we have

$$n^{-1} \sum_{k=1}^{m} \nu_k^{2r} [\varphi(\nu_k^{-2})]^{-1} = n^{-1} \sum_{k=1}^{m} k^{2\epsilon r} \exp(k^{a\epsilon}) \asymp n^{-1} m^{2\epsilon r} \exp(m^{a\epsilon})$$

means that we have to choose $m = c \log(n)^{1/a\epsilon}$ with a sufficiently small $c > 0$. The resulting lower bound is

$$\delta_n \asymp n^{-1} \sum_{k=1}^{m} [\varphi(\nu_k^{-2})]^{-1} \asymp n^{-1} \exp(m^{a\epsilon}) \asymp m^{-2\epsilon r} \asymp (\log n)^{-2r/a}.$$

□

Proof of Proposition 4.3. We have

$$\mathbb{E}[\hat{\eta}_k] = \mathbb{E}[(Kh)(W_i)((K^*)^{-1}u_k)(W_i)] = \langle Kh, (K^*)^{-1}u_k \rangle_W = \langle h, u_k \rangle_X$$

and

$$\begin{aligned} \operatorname{Var}(\hat{\eta}_k) &= \frac{1}{n} \operatorname{Var}\Big((Kh)(W_i)((K^*)^{-1}u_k)(W_i) + V_i((K^*)^{-1}u_k)(W_i) \Big) \\ &\leqslant 2n^{-1} \Big(\|Kh\|_\infty^2 \, \mathbb{E}[((K^*)^{-1}u_k)^2(W_i)] + \mathbb{E}[V_i^2] \, \mathbb{E}[((K^*)^{-1}u_k)^2(W_i)] \Big) \\ &\leqslant 2n^{-1}(S^2 + \sigma_1^2) \|(K^*)^{-1}u_k\|_W^2. \end{aligned}$$

¿From $\|Kg\|_W \geqslant c \|[\varphi(B^{-2})]^{1/2} g\|_X$ for all $g \in L_X^2$ we infer by duality $\|(K^*)^{-1}g\|_W \leqslant c^{-1} \|[\varphi(B^{-2})]^{-1/2} g\|_X$ for all $g \in \operatorname{ran}(K^*)$. Hence,

$$\mathbb{E}_h[\|\hat{h}_n - h\|_X^2] \leqslant 2n^{-1}(S^2 + \sigma_1^2)c^{-2} \sum_{k=1}^{m} [\varphi(\nu_k^{-2})]^{-1} + \sum_{k=m+1}^{\infty} \langle h, u_k \rangle_X^2.$$

¿From $h \in \mathscr{H}(r, R)$ we have the bias estimate

$$\sum_{k=m+1}^{\infty} \langle h, u_k \rangle_X^2 \leqslant \nu_{m+1}^{-2r} R^2.$$

When choosing m as for the lower bound, then the variance term matches the lower bound in order and the estimator $\hat{h}_n$ attains the minimax-rate provided the bias term is not of larger order. This is equivalent to requiring for some uniform constant $c > 0$ that $\nu_{m+1}^{2r} n^{-1} \sum_{k=1}^{m} [\varphi(\nu_k^{-2})]^{-1} \geqslant c$, which in turn follows from $\nu_{m+1} \geqslant \nu_k$ for $k \leqslant m$ and $n^{-1} \sum_{k=1}^{m} \nu_k^{2r} [\varphi(\nu_k^{-2})]^{-1} \asymp 1$.

(1) For mildly ill-posed case with $\varphi(t) = t^a$, $\nu_k \asymp k^\epsilon$, we have

$$n^{-1} \sum_{k=1}^{m} [\varphi(\nu_k^{-2})]^{-1} = n^{-1} \sum_{k=1}^{m} k^{2a\epsilon} \asymp n^{-1} m^{2a\epsilon+1} \asymp m^{-2r\epsilon}$$

by setting $m \asymp n^{1/(2\epsilon r+2\epsilon a+1)}$. Thus we obtain the upper bound: $\delta_n \asymp m^{-2r\epsilon} \asymp n^{-2r/(2r+2a+\epsilon^{-1})}$.

(2) For severely ill-posed case with $\varphi(t) = \exp(-t^{-a/2})$, $\nu_k \asymp k^{\epsilon}$, we have

$$n^{-1}\sum_{k=1}^{m}[\varphi(\nu_k^{-2})]^{-1} = n^{-1}\sum_{k=1}^{m}\exp(k^{a\epsilon}) \asymp n^{-1}\exp(m^{a\epsilon}) \asymp m^{-2r\epsilon}$$

by setting $m = c\log(n)^{1/a\epsilon}$ with a sufficiently small $c > 0$. Thus we obtain the upper bound: $\delta_n \asymp m^{-2r\epsilon} \asymp (\log n)^{-2r/a}$. □

Proof of Theorem 5.3. Given Assumption 5.1 ($\{\psi_k\}$ is a Riesz basis associated with the operator B), there is a bounded invertible operator $\bar{B}$ on L_X^2 such that $\bar{B}\psi_k = u_k$ for all k. This implies that $\mathcal{H}_{m(n)} = \text{span}\{u_1, ..., u_{m(n)}\}$. Denote $\Pi_{m(n)}(h)$ as the projection of $h \in \mathscr{H}(r, R)$ onto $\mathcal{H}_{m(n)}$. Then

$$\|\hat{h}_n - h\|_X^2 \leqslant 2\{\|\Pi_{m(n)}(h) - h\|_X^2 + \|\hat{h}_n - \Pi_{m(n)}(h)\|_X^2\}.$$

As in Blundell, Chen and Kristensen (2007), we define τ_n as a *sieve measure of ill-posedness*:

$$\tau_n := \sup_{h\in\mathcal{H}_{m(n)}:h\neq 0}\frac{\|h\|_X}{\|Kh\|_W} = \sup_{h\in\text{span}\{u_1,\ldots,u_{m(n)}\}:h\neq 0}\frac{\|h\|_X}{\|Kh\|_W},$$

which is well defined under the conditions for identification. Then

$$\|\hat{h}_n - \Pi_{m(n)}(h)\|_X \leqslant \tau_n \times \|K[\hat{h}_n - \Pi_{m(n)}(h)]\|_W.$$

Under Assumption 5.2, by the definition of $\hat{h}_n$ and applying Claims 2 and 3 in Blundell, Chen and Kristensen (2007), we have:

$$\|\hat{h}_n - \Pi_{m(n)}(h)\|_X \leqslant \tau_n \times \{O_p(J_n^{-r_K} + \sqrt{(J/n)} + \|K[h - \Pi_{m(n)}(h)]\|_W)\},$$

where the $O_p()$ holds uniformly over $h \in \mathscr{H}(r, R)$.

By definition of τ_n we have:

$$\tau_n^2 \leqslant \sup_{h\in\text{span}\{u_1,\ldots,u_{m(n)}\}:h\neq 0}\frac{\|h\|_X^2}{\|[\varphi(B^{-2})]^{1/2}h\|_X^2} \leqslant [\varphi(\nu_{m(n)}^{-2})]^{-1},$$

where the first inequality is due to Assumption 4.2 (the reverse link condition), and the second inequality holds because ν_k is increasing in k and $\varphi(t)$ is non-decreasing function in $t \geqslant 0$.

By definition of τ_n we have under Assumptions 2.1, 2.2, 4.2 and $\lim_n \frac{J_n}{m(n)} = c \in (1, \infty)$ and $J_n > m(n)$, we obtain:

$$\tau_n^2\|K[h - \Pi_{m(n)}(h)]\|_W^2 \leqslant \|h - \Pi_{m(n)}(h)\|_X^2 \leqslant R^2\nu_{m(n)+1}^{-2r},$$

thus

$$\|\hat{h}_n - h\|_X^2 \leqslant C' \max\left\{\nu_{m(n)+1}^{-2r},\ \frac{J_n}{n}\tau_n^2\right\} \leqslant C \max\left\{\nu_{m(n)+1}^{-2r},\ \frac{m(n)}{n}[\varphi(\nu_{m(n)}^{-2})]^{-1}\right\}$$

uniformly over $h \in \mathscr{H}(r, R)$ except on an event whose probability tends to zero as $n \uparrow \infty$. $\square$