WORKING PAPERS

N° IDEI - 767

March 2013

Regularizing Priors for Linear Inverse Problems

Jean-Pierre Florens

Anna Simoni

Institut d'Economie Industrielle

Toulouse School of Economics

Regularizing Priors for Linear Inverse Problems*

Jean-Pierre Florens[†] Anna Simoni[‡]

Toulouse School of Economics *CNRS and THEMA*

This draft: March 2013

Abstract

This paper proposes a new Bayesian approach for estimating, nonparametrically, parameters in econometric models that are characterized as the solution of a linear inverse problem. By using a Gaussian process prior distribution we propose the posterior mean as an estimator and prove consistency, in the frequentist sense, of the posterior distribution. Consistency of the posterior distribution provides a frequentist validation of our Bayesian procedure. We show that the minimax rate of contraction of the posterior distribution can be obtained provided that either the regularity of the prior matches the regularity of the true parameter or the prior is scaled at an appropriate rate. The scaling parameter of the prior distribution plays the role of a regularization parameter. We propose a new, and easy-to-implement, data-driven method for optimally selecting in practice this regularization parameter. Moreover, we make clear that the posterior mean, in a conjugate-Gaussian setting, is equal to a Tikhonov-type estimator in a frequentist setting so that our data-driven method can be used in frequentist estimation as well. Finally, we apply our general methodology to two leading examples in econometrics: instrumental regression and functional regression estimation.

Key words: nonparametric estimation, Bayesian inverse problems, Gaussian processes, posterior consistency, data-driven method

JEL code: C13, C11, C14

*First draft: September 2008. A preliminary version of this paper circulated under the title *On the Regularization Power of the Prior Distribution in Linear ill-Posed Inverse Problems*. We thank Joel Horowitz, Enno Mammen, Andrew Stuart, and Sébastien Van Bellegem for interesting discussions. We also acknowledge helpful comments and suggestions from participants to workshops and seminars in Banff (2009), Bressanone (2009), Stats in the Château (Paris 2009), Toulouse (2009), CIRM-Marseille (2009), Brown (2010), Mannheim (2010), Northwestern (2010), World Congress of the Econometric Society (Shangai 2010), Warwick (2011), Cowles Foundation Econometric Summer conference (Yale 2011), Princeton (2012). Anna Simoni gratefully acknowledges financial support from the University of Mannheim through the DFG-SNF Research Group FOR916 and hospitality from CREST.

[†]Toulouse School of Economics - 21, allée de Brienne - 31000 Toulouse (France). florens@cict.fr

[‡]French National Center of Scientific Research and THEMA, Université de Cergy-Pontoise - 33, boulevard du Port, 95011 Cergy-Pontoise (France). Email: simoni.anna@gmail.com (*corresponding author*).

1 Introduction

In the last decade, econometric theory has shown an increasing interest in the theory of *stochastic inverse problems* as a fundamental tool for functional estimation of structural as well as reduced form models. The purpose of this paper is to develop an encompassing Bayesian approach to stochastic linear inverse problems for nonparametric estimation of econometric models.

We construct a Gaussian process prior and show that the corresponding posterior mean is a consistent estimator, in the frequentist sense, of the functional parameter of interest. Under mild conditions, we prove that our Bayes estimator is equal to a Tikhonov-type estimator in the frequentist setting. This result enables us to construct a data-driven method, based on a Bayes procedure, for selecting the regularization parameter. Such a parameter is necessary in order to implement in practice Tikhonov-type estimators and these estimators are very sensitive to the value of this parameter. We show that the value selected by our data-driven method is optimal in a minimax sense if the prior distribution is sufficiently smooth.

Stochastic inverse problems theory has recently been used in many subfields of econometrics to construct new estimation methods. First of all, the theory of inverse problems has been shown to be fundamental for nonparametric estimation of an instrumental regression function, see *e.g.* Florens (2003), Newey and Powell (2003), Hall and Horowitz (2005), Blundell *et al.* (2007), Darolles *et al.* (2011), Florens and Simoni (2012a). More generally, inverse problems theory has been used for semiparametric estimation under moment restrictions, see *e.g.* Carrasco and Florens (2000), Ai and Chen (2003), Chen and Pouzo (2012). In addition, it has been exploited for inference in econometric models with heterogeneity – *e.g.* Gautier and Kitamura (2012), Hoderlein *et al.* (2012) – for inference in auction models – *e.g.* Florens and Sbaï (2010) – and for frontier estimation for productivity analysis – *e.g.* Daouia *et al.* (2009). Functional estimation of reduced form econometric and statistical models based on inverse problem theory has been developed, to name only a few, by Hall and Horowitz (2007) and Johannes (2008). This large and incomplete list of references highlights the importance that inverse problem theory has gained in econometrics. We refer to Carrasco *et al.* (2007) and references therein for a general overview of inverse problems in econometrics.

The general framework, which accommodates many functional estimation problems in econometrics, is the following. Let $\mathcal{X}$ and $\mathcal{Y}$ be infinite dimensional separable Hilbert spaces over $\mathbb{R}$ and denote by $x \in \mathcal{X}$ the functional parameter that we want to estimate. For instance, x can be an Engel curve or the probability density function of the unobserved heterogeneity. The estimating equation characterizes x as the solution of the functional equation

$$y^{\delta} = Kx + U, \qquad x \in \mathcal{X}, \quad y^{\delta} \in \mathcal{Y}, \tag{1}$$

where y^δ is an observable function, $K : \mathcal{X} \to \mathcal{Y}$ is a known, bounded, linear operator and U is an error term with values in $\mathcal{Y}$. Thus, estimating x is an inverse problem. More precisely, y^δ is a transformation of a n-sample of finite dimensional objects and the parameter $\delta^{-1} > 0$ represents the "level of information" of the sample, so that $\delta \to 0$ as $n \to \infty$ and for $\delta = 0$ (*i.e.* perfect information) we have $y^0 = Kx$. Usually, we have that $\delta = \frac{1}{n}$ and (1) is a set of moment equations with y^δ an empirical moment. In this paper we propose a Bayesian procedure for nonparametric estimation of a large class of econometric models that write under the form (1) like moment equality models, asset pricing functional estimation in equilibrium models, density estimation in structural models that account for heterogeneity, deconvolution in structural models with measurement errors. We illustrate two leading examples in econometrics that can be estimated by using our method.

Example 1 (Instrumental variable (IV) regression estimation)**.** Let (Y, Z, W) be an observable real random vector and $x(Z)$ be the IV regression defined through the moment condition $\mathbf{E}(Y|W) = \mathbf{E}(x|W)$. Suppose that the distribution of (Z, W) is confined to the unit square $[0,1]^2$ and admits a density f_{ZW}. The moment restriction implies that x is a solution to

$$\mathbf{E}_W\left[\mathbf{E}(Y|w)a(w,v)\right](v) = \int_0^1\int_0^1 x(z)a(w,v)f_{ZW}(z,w)dwdz$$

where $a(w,v) \in L^2[0,1]^2$ is a known and symmetric function, $\mathbf{E}_W$ denotes the expectation with respect to the marginal density f_W of W and $L^2[0,1]^2$ denotes the space of square integrable functions on $[0,1]^2$. This transformation of the original moment condition is appealing because in this way its empirical counterpart is asymptotically Gaussian (as required by our Bayesian approach). Assume that $x \in \mathcal{X} \equiv L^2[0,1]$. By replacing the true distribution of (Y, Z, W) with a nonparametric estimator we obtain (1) with

$$y^\delta = \hat{\mathbf{E}}_W\left[\hat{\mathbf{E}}(Y|w)a(w,v)\right] \quad \text{and} \quad Kx = \int_0^1\int_0^1 x(z)a(w,v)\hat{f}_{ZW}(z,w)dwdz. \tag{2}$$

□

Example 2 (Functional Linear Regression Estimation)**.** The model is the following:

$$\xi = \int_0^1 x(s)Z(s)ds + \varepsilon, \quad \mathbf{E}(\varepsilon Z(s)) = 0, \quad Z, x \in L^2([0,1]), \quad \mathbf{E} < Z, Z >< \infty \tag{3}$$

and $\varepsilon|Z,\tau^2 \sim \mathcal{N}(0,\tau^2)$, with $< \cdot, \cdot >$ denotes the inner product in $L^2[0,1]$. We want to recover the functional regression x. Assuming that Z is a centered random function with covariance operator of trace-class, the most popular approach consists in multiply both sides of the first equation in (3) by $Z(s)$ and then take the expectation: $\mathbf{E}(\xi Z(t)) = \int_0^1 x(s)Cov(Z(s), Z(t))ds$, for $t \in [0,1]$. If we dispose of independent and identically distributed data $(\xi_1, Z_1), \ldots, (\xi_n, Z_n)$ we can estimate the unknown moments in the previous equation. Thus, x is solution of (1) with $y^\delta := \frac{1}{n}\sum_i \xi_i Z_i(t)$,

$U = \frac{1}{n}\sum_i \varepsilon_i Z_i(t)$ and $\forall \varphi \in L^2([0,1]) \mapsto K\varphi := \frac{1}{n}\sum_i < Z_i, \varphi > Z_i(t)$. $\square$

A Bayesian approach to stochastic inverse problems allows to answer two important questions. A first question concerns the way to incorporate in the estimation procedure the prior information about the functional parameter which is often provided by economic theory or beliefs of experts. This prior information may be particularly valuable in functional estimation since often the data available are concentrated only in a region of the graph of the functional parameter so that some parts of the function can not be recovered from the data. Therefore, the prior information allows to "*identify*" the parts of the curve for which we have no data.

A second question concerns the selection in practice of the tuning parameter that is necessary in order to compute the nonparametric estimator. In estimation based on inverse problem techniques this parameter is known as the *regularization parameter*. The value of such a parameter affects considerably the estimation result and it is very important to have an accurate way to choose it.

The Bayesian approach combines the prior and sampling information and proposes the posterior distribution as solution to the inverse problem. It allows to incorporate prior information in a natural way through the prior distribution as well as to select a value for the tuning parameter from its posterior distribution by using a hierarchical structure. Unfortunately, many of the Bayesian approaches to stochastic inverse problems proposed so far can not be applied to functional estimation in econometrics. This is due to two main reasons. First, the majority of Bayesian approaches do not work for functional observations and parameters because they consider a finite dimensional projection of (1) and recover x only on a grid of points, see Chapter 5 in Kaipio and Somersalo (2004) and Helin (2009), Lassas *et al.* (2009), Hofinger and Pikkarainen (2007), Hofinger and Pikkarainen (2007), Neubauer and Pikkarainen (2008). Second, the existing Bayesian approaches to inverse problems that consider models for functional estimation do not allow to accommodate the econometric model of interest because of the different definition of the model's error term U, see Knapik *et al.* (2011).

Since the econometric model (1) differs from the stochastic inverse problems addressed by the existing Bayesian inverse problems literature a specific Bayesian procedure must be developed. In particular, we have to deal with the problem that the exact posterior mean of x in (1) (when a conjugate-Gaussian setting is used) can be defined only in a complicated way as a measurable linear transformation (see the explanation below) and it is not possible to compute it in practice. To overcome this problem Florens and Simoni (2012b) construct a *regularized posterior distribution* for the solution of an inverse problem. The mean of this distribution is easy to implement and works well in practice but lacks a pure Bayesian interpretation. In fact, the regularization of the posterior distribution is introduced *ad hoc* and cannot be justified by any prior-to-posterior transformation. Therefore, it is not possible to construct a data-driven method based on a Bayes procedure for selecting the *regularization parameter* and the regularized posterior mean differs from a frequentist

Tikhonov-type estimator. The current paper proposes a pure Bayesian solution to this problem as we explain in the next section.

1.1 Our contribution

Our estimation procedure is based on a conjugate-Gaussian setting which is suggested by the linearity of problem (1). On one hand, such a setting is appealing because the corresponding posterior distribution can be computed analytically without using any MCMC algorithm which, even if very powerful, slows down the estimate computation. On the other hand, a conjugate-Gaussian Bayesian inverse problem has the drawback that the exact mean of the posterior distribution is, in general, not defined as a linear estimator but as a *measurable linear transformation* (*mlt*) which is a weaker notion, see Mandelbaum (1984). In particular, there is no explicit form for the *mlt* estimator and so it is unclear how we can construct the Bayes estimator of x in practice. Moreover, whether consistency of the *mlt* estimator holds or not is still an open question.

The first contribution of our paper is to provide a sufficient condition under which the exact posterior mean, in a conjugate-Gaussian setting, has a closed-form and thus can be easily computed and used as an estimator for x (as it is justified for a broad class of loss functions). We assume a Gaussian prior distribution for x, with mean function $x_0 \in \mathcal{X}$ and covariance operator $\Omega_0 : \mathcal{X} \to \mathcal{X}$. Then, in the case when $\mathcal{X}$ and $\mathcal{Y}$ are finite-dimensional and x and y^δ are jointly Gaussian, the posterior mean of x is the linear estimator $[x_0 + \Omega_0 K^* Var(y^\delta)^{-1}(y^\delta - Kx_0)]$ provided $Var(y^\delta)$ is invertible, where $Var(y^\delta)$ denotes the marginal covariance operator of y^δ. However, when the dimension of $\mathcal{X}$ and $\mathcal{Y}$ is infinite, the linear operator $\Omega_0 K^* Var(y^\delta)^{-1}$ is only defined on a dense subspace of $\mathcal{Y}$ of measure zero and is typically non-continuous (*i.e.* unbounded). This paper gives a sufficient condition that guarantees that $Var(y^\delta)^{-1}$ is continuous (and defined) on the whole $\mathcal{Y}$ and shows that this condition is in general satisfied in many econometric models. Hence, we provide a closed-form for the posterior mean of x that is implementable in practice and prove that it is a continuous and linear (thus consistent) estimator defined on $\mathcal{Y}$. Under this condition, the prior-to-posterior transformation can be interpreted as a regularization scheme and we do not need to introduce an *ad hoc* regularization scheme as in Florens and Simoni (2012b).

Our second contribution consists in the study of frequentist asymptotic properties of the conjugate-Gaussian Bayesian estimation of (1). For that, we admit the existence of a true x that generates the data. We establish that the posterior mean estimator and posterior distribution have good frequentist asymptotic properties for $\delta \to 0$. *Frequentist posterior consistency* is defined as the convergence of the posterior distribution towards a Dirac mass at the true value of x almost surely with respect to the sampling distribution, see *e.g.* Diaconis and Freedman (1986, 1998). This property provides the frequentist validation of our Bayesian procedure.

Besides proving frequentist consistency we also recover the rate of contraction of the risk associated with the posterior mean and of the posterior distribution. This rate depends on the smoothness and scale of the prior as well as on the smoothness of the true x. Depending on the specification of the prior this rate may be *minimax* over a Sobolev ellipsoid. In particular, *(i)* when the regularity of the prior matches the regularity of the true x, the minimax rate of convergence is obtained with a fixed prior covariance; *(ii)* when the prior is rougher or smoother at any degree than the truth, the minimax rate can still be obtained if the prior is scaled at an appropriate rate depending on the unknown regularity of the true x.

Our third contribution consists in proposing a new method for optimally selecting the *regularization parameter* α. This parameter enters the prior distribution as a scaling hyperparameter of the prior covariance and we construct an adaptive data-driven method for selecting it which is based on an empirical Bayes (EB) approach. Because the posterior mean is, under our assumptions, equal to a Tikhonov-type estimator for problem (1), our EB approach for selecting the regularization parameter is valid and can be used also for frequentist estimators based on inverse problems techniques.[1] Finally, the EB-selected regularization parameter is plugged into the prior distribution of x and for the corresponding EB-posterior distribution we prove frequentist posterior consistency.

In the following, the Bayesian approach and the asymptotic results are presented for general models of the form (1); then, we develop further results that apply to the specific examples 1 and 2. In Section 2 we set the Bayesian model associated with (1) and the main assumptions. In Section 3 the posterior distribution of x is computed and its frequentist asymptotic properties are analyzed. Section 4 focuses on the mildly ill-posed case. The EB method is developed in Section 5. Section 6 shows numerical implementations and Section 7 concludes. All the proofs are in the Appendix.

2 Model

Let $\mathcal{X}$ and $\mathcal{Y}$ be infinite dimensional separable Hilbert spaces over $\mathbb{R}$ with norm $||\cdot||$ induced by the inner product $<\cdot,\cdot>$. Let $\mathfrak{B}(\mathcal{X})$ and $\mathfrak{B}(\mathcal{Y})$ be the Borel σ-fields generated by the open sets of $\mathcal{X}$ and $\mathcal{Y}$, respectively. We consider the inverse problem of estimating the function $x \in \mathcal{X}$ which is linked to the data y^δ through the linear relation

$$y^\delta = Kx + U, \qquad x \in \mathcal{X}, \quad y^\delta \in \mathcal{Y} \tag{4}$$

where y^δ is an observable function and $K : \mathcal{X} \to \mathcal{Y}$ is a known, bounded, linear operator (we refer to Carrasco *et al.* (2007) for definition of terminology from functional analysis). The elements y^δ

[1] Notice that in general the posterior mean in a conjugate-Gaussian problem stated in infinite-dimensional Hilbert spaces cannot be equal to the Tikhonov solution of (1). This is due to the particular structure of the covariance operator of the error term U and it will be detailed later.

and U are Hilbert space-valued random variables (H-r.v.), that is, for a complete probability space $(S, \mathcal{S}, \mathbb{P})$, U (resp. y^δ) defines a measurable map $U : (S, \mathcal{S}, \mathbb{P}) \to (\mathcal{Y}, \mathfrak{B}(\mathcal{Y}))$, see *e.g.* Kuo (1975). Realizations of y^δ are functional transformations of the observed data. The true value of x that generates the data is denoted by x_*.

We assume a mean-zero Gaussian distribution on $\mathfrak{B}(\mathcal{Y})$ for U: $U \sim \mathcal{N}(0, \delta\Sigma)$ where $\delta > 0$ is the noise level and $\Sigma : \mathcal{Y} \to \mathcal{Y}$ is a covariance operator, that is, Σ is such that $< \delta\Sigma\phi_1, \phi_2 >= \mathbf{E}\left(< U, \phi_1 >< U, \phi_2 >\right)$ for all $\phi_1, \phi_2 \in \mathcal{Y}$. Therefore, Σ is a one-to-one, linear, positive definite, self-adjoint and trace-class operator. Because Σ is one-to-one the support of U is all $\mathcal{Y}$, see Kuo (1975) and Ito (1970). A *trace-class* operator is a compact operator with eigenvalues that are summable. This property rules out a covariance Σ proportional to the identity operator I and this is a key difference between our model and the model used in a large part of the statistical inverse problem literature, see *e.g.* Cavalier and Tsybakov (2002), Bissantz *et al.* (2007) and Knapik *et al.* (2011). On the other side, a covariance Σ different from I naturally arises in econometric problems since the structure of the estimating equation (1) does not allow for an identity (or proportional to identity) covariance operator, see examples 1 and 2.

Under model (1) and the assumption that $U \sim \mathcal{N}(0, \delta\Sigma)$ the sampling distribution P^x, *i.e.* the conditional distribution of y^δ given x, is a Gaussian distribution on $\mathfrak{B}(\mathcal{Y})$:

$$y^\delta | x \sim P^x = \mathcal{N}(Kx, \delta\Sigma). \tag{5}$$

with $\delta > 0$ such that $\delta \downarrow 0$. Hereafter, $\mathbf{E}_x(\cdot)$ will denote the expectation taken with respect to P^x.

Remark 2.1. The assumption of Gaussianity of the error term U in the econometric model (1) is not necessary and only made in order to construct (and give a Bayesian interpretation to) the estimator. The proofs of our results of frequency consistency do not rely on the normality of U. In particular, asymptotic normality of $y^\delta | x$ – as in example 1 – is enough for our estimation procedure and also for our EB data-driven method for choosing the regularization parameter. □

Remark 2.2. All the results in the paper are given for the general case where K and Σ are fixed and known. This choice is made in order to keep our presentation as simple as possible. We discuss how our results apply to the case with unknown K and Σ through examples 1 and 2. □

2.1 Notation

We set up some notational convention used in the paper. For positive quantities M_δ and N_δ depending on a discrete or continuous index δ, we write $M_\delta \asymp N_\delta$ to mean that the ratio M_δ / N_δ is bounded away from zero and infinity. We write $M_\delta = \mathcal{O}(N_\delta)$ if M_δ is at most of the same order as N_δ. For an H-r.v. W we write $W \sim \mathcal{N}$ for denoting that W is a Gaussian process. We denote by $\mathcal{R}(\cdot)$ the range of an operator and by $\mathcal{D}(\cdot)$ its domain. For an operator $B : \mathcal{X} \to \mathcal{Y}$, $\overline{\mathcal{R}(B)}$ denotes the closure in $\mathcal{Y}$ of the range of B. For a bounded operator $A : \mathcal{Y} \to \mathcal{X}$, we denote by A^* its adjoint, *i.e.* $A^* : \mathcal{X} \to \mathcal{Y}$ is such that $< A\psi, \varphi >=< \psi, A^*\varphi >$, $\forall \varphi \in \mathcal{X}$, $\psi \in \mathcal{Y}$. The operator norm is defined as $||A|| := \sup_{||\phi||=1} ||A\phi|| = \min\{C \geq 0 \,;\, ||A\phi|| \leq C||\phi|| \text{ for all } \phi \in \mathcal{Y}\}$. For a subset $\mathcal{Y}_1 \subset \mathcal{Y}$, $A|_{\mathcal{Y}_1} : \mathcal{Y}_1 \to \mathcal{X}$ denotes the restriction of A to the domain $\mathcal{Y}_1$. The operator I denotes the identity operator on both spaces $\mathcal{X}$ and $\mathcal{Y}$, *i.e.* $\forall \psi \in \mathcal{X}$, $\varphi \in \mathcal{Y}$, $I\psi = \psi$ and $I\varphi = \varphi$.

Let $\{\varphi_j\}_j$ denote an orthonormal basis of $\mathcal{X}$. The trace of a bounded linear operator $A : \mathcal{Y} \to \mathcal{X}$ is defined as $tr(A) := \sum_{j=1}^{\infty} < (A^*A)^{\frac{1}{2}}\varphi_j, \varphi_j >$ independently of the basis $\{\varphi_j\}_j$. If A is compact then its trace writes $tr(A) = \sum_{j=1}^{\infty} \lambda_j$, where $\{\lambda_j\}$ are the singular values of A. The Hilbert-Schmidt norm of a bounded linear operator $A : \mathcal{Y} \to \mathcal{X}$ is denoted by $||A||_{HS}$ and defined as $||A||_{HS}^2 = tr(A^*A)$, see Kato (1995).

2.2 Prior measure and main assumptions

In this section we introduce the prior distribution and two sets of assumptions. *(i)* The first set of assumptions (A.2 and B below) will be used for establishing the rate of contraction of the posterior distribution and concerns the smoothness of the operator $\Sigma^{-1/2}K$ and of the true value x_*. *(ii)* The assumptions in the second set (A.1 and A.3 below) are new in the literature and are sufficient for having a posterior mean that is a continuous linear operator defined on $\mathcal{Y}$. The detection of the latter assumptions is an important contribution because, as remarked in Luschgy (1995) and Mandelbaum (1984), in the Gaussian infinite-dimensional model the posterior mean is in general only defined as a *mlt* which is a weaker notion than that one of a continuous linear operator. Therefore, in general the posterior mean has not an explicit form and may be an inconsistent estimator in the frequentist sense while our assumptions A.1 and A.3 guarantee a closed-form (easy to compute) and consistency for the posterior mean.

Assumption A.1. $\mathcal{R}(K) \subset \mathcal{D}(\Sigma^{-1/2})$.

Since K and Σ are integral operators, $\Sigma^{-1/2}$ is a differential operator and Assumption A.1 demands that the functions in $\mathcal{R}(K)$ are at least as smooth as the functions in $\mathcal{R}(\Sigma^{1/2})$. A.1 ensures that $\Sigma^{-1/2}$ is defined on $\mathcal{R}(K)$ so that $\Sigma^{-1/2}K$, which is used in Assumption A.2 below, exists.

Assumption A.2. *There exists an unbounded, self-adjoint, densely defined operator L in the Hilbert space $\mathcal{X}$ for which $\exists\, \eta > 0$ such that $< L\psi, \psi > \geq\ \eta||\psi||^2$, $\forall \psi \in \mathcal{D}(L)$, and that satisfies*

$$\underline{m}||L^{-a}x|| \leq ||\Sigma^{-1/2}Kx|| \leq \overline{m}||L^{-a}x|| \tag{6}$$

on $\mathcal{X}$ for some $a > 0$ and $0 < \underline{m} \leq \overline{m} < \infty$. Moreover, L^{-2s} is trace-class for some $s > a$.

Assumption A.2 means that $\Sigma^{-1/2}K$ regularizes at least as much as L^{-a}. Because $\Sigma^{-1/2}K$ must satisfy (6) it is necessarily an injective operator.

We turn now to the construction of the prior distribution of x. Our proposal is to use the operator L^{-2s} to construct the prior covariance operator. We assume a Gaussian *prior distribution* μ on $\mathfrak{B}(\mathcal{X})$:

$$x|\alpha, s \sim \mu = \mathcal{N}\Big(x_0, \frac{\delta}{\alpha}\Omega_0\Big), \qquad x_0 \in \mathcal{X}, \quad \Omega_0 := L^{-2s},\ s > a \tag{7}$$

with $\alpha > 0$ such that $\alpha \to 0$. The parameter α describes a class of prior distributions and it may be viewed as an hyperparameter. We provide in Section 5 an Empirical Bayes approach for selecting it.

By definition of L, the operator $\Omega_0 : \mathcal{X} \to \mathcal{X}$ is linear, bounded, positive-definite, self-adjoint, compact and trace-class. It results evident from Assumption A.2 that such a choice for the prior covariance is aimed at linking the prior distribution to the sampling model. A similar idea was proposed by Zellner (1986) for linear regression models for which he constructed a class of prior called g-prior. Our prior (7) is an extension of the Zellner's g-prior and we call it *extended g-prior*.

The distribution μ (resp. P^x) is realizable as a proper random element in $\mathcal{X}$ (resp. $\mathcal{Y}$) if and only if Ω_0 (resp. Σ) is trace-class. Thus, neither Σ nor Ω_0 can be proportional to I so that, in general, in infinite-dimensional inverse problems, the posterior mean cannot be equal to the Tikhonov regularized estimator $x_\alpha^T := (\alpha I + K^*K)^{-1}K^*y^\delta$. However, we show in this paper that under A.1, A.2 and A.3, the posterior mean equates the Tikhonov regularized solution in the *Hilbert Scale* generated by L. We give later the definition of Hilbert Scale.

The following assumption ties further the prior to the sampling distribution by linking the smoothing properties of Σ, K and $\Omega_0^{\frac{1}{2}}$.

Assumption A.3. $\mathcal{R}(K\Omega_0^{\frac{1}{2}}) \subset \mathcal{D}(\Sigma^{-1})$.

Hereafter, we denote $B = \Sigma^{-1/2}K\Omega_0^{\frac{1}{2}}$. Assumption A.3 guarantees that B and $\Sigma^{-1/2}B$ exist.

We now discuss the link existing between Assumption A.2, which quantifies the smoothness of $\Sigma^{-1/2}K$, and Assumption B below, which quantifies the smoothness of the true value x_*. In order to explain these assumptions and their link we will: *(i)* introduce the definition of *Hilbert scale*, *(ii)* explain the meaning of the parameter a in (6), *(iii)* discuss the regularity conditions of $\Sigma^{-1/2}K$ and of the true value of x.

(i) The operator L in Assumption A.2 is a generating operator of the Hilbert scale $(\mathcal{X}_t)_{t\in\mathbb{R}}$ where $\forall t \in \mathbb{R}$, $\mathcal{X}_t$ is the completion of $\bigcap_{k\in\mathbb{R}} \mathcal{D}(L^k)$ with respect to the norm $||x||_t := ||L^t x||$ and is a Hilbert space, see Definition 8.18 in Engl *et al.* (2000), Goldenshluger and Pereverzev (2003) or Krein and Petunin (1966). For $t > 0$ the space $\mathcal{X}_t \subset \mathcal{X}$ is the domain of definition of L^t: $\mathcal{X}_t = \mathcal{D}(L^t)$. Typical examples of $\mathcal{X}_t$ are Sobolev spaces of various kinds.

(ii) We refer to the parameter a in A.2 as the "degree of ill-posedness" of the econometric model under study and a is determined by the rate of decreasing of the spectrum of $\Sigma^{-1/2}K$ (and not only by that one of K as it would be in a classical inverse problems framework for (1)). Since the spectrum of $\Sigma^{-1/2}K$ is decreasing slower than that one of K we have to control for less ill-

posedness than if we used the classical approach.

(iii) In inverse problems theory it is natural to impose conditions on the regularity of x_* by relating it to the regularity of the operator that characterizes the inverse problem (that is, the operator $\Sigma^{-1/2}K$ in our case). A possible implementation of this consists in introducing a Hilbert Scale and expressing the regularity of both x_* and $\Sigma^{-1/2}K$ with respect to this common Hilbert Scale. This is the meaning of - and the link between - Assumptions A.2 and B where we use the Hilbert Scale $(\mathcal{X}_t)_{t\in\mathbb{R}}$ generated by L. We refer to Chen and Reiss (2011) and Johannes *et al.* (2011) for an explanation of the relationship between Hilbert Scale and regularity conditions. The following assumption expresses the regularity of x_* according to $\mathcal{X}_t$.

Assumption B. *For some* $0 \leq \beta$, $(x_* - x_0) \in \mathcal{X}_\beta$, *that is, there exists a* $\rho_* \in \mathcal{X}$ *such that* $(x_* - x_0) = L^{-\beta}\rho_*$ $(\equiv \Omega_0^{\frac{\beta}{2s}}\rho_*)$.

The parameter β characterizes the "regularity" of the centered true function $(x_* - x_0)$ and is generally unknown. Assumption B is satisfied by regular functions x_*. In principle, it could be satisfied also by irregular x_* if we were able to decompose x_* in the sum of a regular part plus an irregular part and to choose x_0 such that it takes all the irregularity of x_*. This is clearly infeasible in practice as x_* is unknown. On the contrary, we could take x_0 very smooth so that Assumption B would be less demanding about the regularity of x_*. When $\mathcal{X}_\beta$ is the scale of Sobolev spaces, Assumption B is equivalent to assume that $(x_* - x_0)$ has at least β square integrable derivatives.

Remark 2.3. Assumption B is classical in inverse problems literature, see *e.g.* Chen and Reiss (2011) and Nair *et al.* (2005), and is closely related to the so-called *source condition* which expresses the regularity of the function x_* according to the spectral representation of the operator K^*K defining the inverse problem, see Engl *et al.* (2000) and Carrasco *et al.* (2007). In our case, the regularity of $(x_* - x_0)$ is expressed according to the spectral representation of L. □

Remark 2.4. Assumption A.2 covers not only the mildly ill-posed but also the severely ill-posed case if $(x_* - x_0)$ in assumption B is infinitely smooth. In the *mildly ill-posed* case the singular values of $\Sigma^{-1/2}K$ decay slowly to zero (typically at a geometric rate) which means that the kernel of $\Sigma^{-1/2}K$ is finitely smooth. In this case the operator L is generally some differential operator so that L^{-1} is finitely smooth. In the *severely ill-posed* case the singular values of $\Sigma^{-1/2}K$ decay very rapidly (typically at an exponential rate). Assumption A.2 covers also this case if $(x_* - x_0)$ is very smooth. This is because when the singular values of $\Sigma^{-1/2}K$ decay exponentially, Assumption A.2 is satisfied if L^{-1} has an exponentially decreasing spectrum too. On the other hand, L^{-1} is used to describe the regularity of $(x_* - x_0)$, so that in the severely ill-posed case, Assumption B can be satisfied only if $(x_* - x_0)$ is infinitely smooth. In this case we could for instance take $L = (K^*\Sigma^{-1}K)^{-\frac{1}{2}}$ which implies $a = 1$. We could make Assumption A.2 more general, as in Chen and Reiss (2011), in order to cover the *severely ill-posed* case even when $(x_* - x_0)$ is not infinitely smooth. Since computations to find the rate would become more cumbersome (even if still possible) we do not pursue this direction here. □

Remark 2.5. The specification of the prior covariance operator can be generalized as $\frac{\delta}{\alpha}\Omega_0 = \frac{\delta}{\alpha}QL^{-2s}Q^*$, for some bounded operator Q not necessarily compact. Then, the previous case is a particular case of this one for $Q = I$. In this setting, Assumptions A.1 and A.3 are replaced by the weaker assumptions $\mathcal{R}(KQ) \subset \mathcal{D}(\Sigma^{-1/2})$ and $\mathcal{R}(KQL^{-s}) \subset \mathcal{D}(\Sigma^{-1})$, respectively. In Assumption A.2 the operator $\Sigma^{-1/2}K$ must be replaced by $\Sigma^{-1/2}KQ$ and Assumption B becomes: there exists $\tilde{\rho}_* \in \mathcal{X}$ such that $(x_* - x_0) = QL^{-\beta}\tilde{\rho}_*$. □

Example 1 (Instrumental variable (IV) regression estimation *(continued)*)**.** Let us consider the integral equation (4), with y^δ and K defined as in (2), that characterizes the IV regression x. Suppose to use the kernel smoothing approach to estimate f_{YW} and f_{ZW}, where f_{YW} denotes the density of the distribution of (Y, W) with respect to the Lebesgue measure. For simplicity we assume that (Z, W) is a bivariate random vector. Let $K_{Z,h}$ and $K_{W,h}$ denote two univariate kernel functions in $L^2[0, 1]$, h be the bandwidth and let $(y_i, w_i, z_i)_{i=1}^n$ be the n-observed random sample. Denote $\Lambda : L^2[0, 1] \to L^2[0, 1]$ the operator $\Lambda\varphi = \int a(w, v)\varphi(w)dw$, with $a(w, v)$ a known function, and $\tilde{K} : L^2[0, 1] \to L^2[0, 1]$ the operator $\tilde{K}\phi = \frac{1}{n}\sum_{i=1}^n \frac{K_{W,h}(w_i - w)}{h} < \phi(z), \frac{K_{Z,h}(z_i - z)}{h} >$. Therefore, $K = \Lambda\tilde{K}$ so that the quantities in (2) can be rewritten as

$$y^\delta = \Lambda\left[\hat{\mathbf{E}}(Y|W = w)\hat{f}_W\right](v) = \int a(w, v)\frac{1}{nh}\sum_{i=1}^n y_i K_{W,h}(w_i - w)dw$$

$$\text{and} \quad Kx = \int a(w, v)\frac{1}{n}\sum_{i=1}^n \frac{K_{W,h}(w_i - w)}{h}\int x(z)\frac{K_{Z,h}(z_i - z)}{h}dzdw \tag{8}$$

Remark that $\lim_{n\to\infty}\tilde{K}\phi = f_W(w)\mathbf{E}(\phi|w) = M_f\mathbf{E}(\phi|w)$ where M_f denotes the multiplication operator by f_W. If $a = f_{WZ}$ then $\Lambda\lim_{n\to\infty}\tilde{K}$ is the same as the integral operator in Hall and Horowitz (2005).

In this example, the assumption that $U \sim \mathcal{N}(0, \delta\Sigma)$ (where $U = y^\delta - Kx$) holds asymptotically and the transformation of the model through Λ is necessary in order to guarantee such a convergence of U towards a zero-mean Gaussian process. We explain this fact by extending Ruymgaart (1998). It is possible to show that the covariance operator $\tilde{\Sigma}_h$ of $\frac{\sqrt{n}}{n}\sum_{i=1}^n\left(y_i - < x, \frac{K_{Z,h}(z_i - z)}{h} >\right)\frac{K_{W,h}(w_i - w)}{h}$ satisfies

$$< \phi_1, \tilde{\Sigma}_h\phi_2 > \quad \longrightarrow \quad < \phi_1, \tilde{\Sigma}\phi_2 >, \qquad \text{as } h \to 0, \quad \forall\phi_1, \phi_2 \in L^2[0, 1]$$

where $\tilde{\Sigma}\phi_2 = \sigma^2 f_W(v)\phi_2(v) = \sigma^2 M_f\phi_2(v)$ under the assumption $\mathbf{E}[(Y - x(Z))^2|W] = \sigma^2 < \infty$. Unfortunately, because $\tilde{\Sigma}$ has not finite trace, it is incompatible with the covariance structure of a Gaussian limiting probability measure. The result is even worst, since Ruymgaart (1998) shows that there are no scaling factors, with $h = n^{-r}$, for $0 < r < 1$, such that $n^{-r}\sum_{i=1}^{n}\left(y_i - < x, \frac{K_{Z,h}(z_i - z)}{h} >\right)\times \frac{K_{W,h}(w_i - w)}{h}$ converges weakly in $L^2[0,1]$ to a Gaussian distribution (unless this distribution is degenerate at the zero function). However, if we choose $a(w,v)$ appropriately so that Λ is a compact operator and $\Lambda^*\Lambda$ has finite trace, then $\sqrt{n}\left(y^\delta - Kx\right) \Rightarrow \mathcal{N}(0, \Lambda\tilde{\Sigma}\Lambda^*)$, in $L^2[0,1]$ as $n \to \infty$, where '$\Rightarrow$' denotes weak convergence and $\Lambda\tilde{\Sigma}\Lambda^* =: \Sigma$. The adjoint operator $\Lambda^* : L^2[0,1] \to L^2[0,1]$ is defined as: $\forall \varphi \in L^2[0,1]$, $\Lambda^*\varphi = \int a(w,v)\varphi(v)dv$ and $\Lambda = \Lambda^*$ if $a(w,v)$ is symmetric in w and v. The operator Σ is unknown and can be estimated by $\hat{\Sigma} = \Lambda\frac{\hat{\sigma}^2}{n}\sum_{i=1}^{n}\frac{K_{W,h}(w_i - w)}{h}\Lambda^*$.

We now discuss Assumptions A.1, A.2 and A.3. While A.1 and A.3 need to hold both in finite sample and for $n \to \infty$, A.2 only has to hold for $n \to \infty$. We start by checking assumption A.1. In large sample, the operator K converges to $\Lambda M_f \mathbf{E}(\phi|w)$ and it is trivial to verify that $\mathcal{D}(\Sigma^{-1/2}) = \mathcal{R}(\Lambda M_f^{\frac{1}{2}}) \supset \mathcal{R}(\Lambda M_f) \supset \mathcal{R}(\Lambda M_f \mathbf{E}(\cdot|w))$. In finite sample, the same holds with M_f and $\mathbf{E}(\cdot|w)$ replaced by their empirical counterparts. Next, we check the validity of assumption A.2 for $n \to \infty$. Remark that the operator $\Sigma^{1/2}$ may be equivalently defined in two ways. It may be a self-adjoint operator, that is $\Sigma^{1/2} = (\Sigma^{1/2})^* = (\Lambda M_f \Lambda^*)^{1/2}$, such that $\Sigma = \Sigma^{1/2}\Sigma^{1/2}$ or it may be defined as $\Sigma^{1/2} = \Lambda M_f^{1/2}$ so that $\Sigma = \Sigma^{1/2}(\Sigma^{1/2})^*$ where $(\Sigma^{1/2})^* = M_f^{\frac{1}{2}}\Lambda^*$. Thus, by using the second definition, we obtain that $\Sigma^{-1/2}K = (\Lambda M_f^{\frac{1}{2}})^{-1}\Lambda\tilde{K} = f_W^{-\frac{1}{2}}\Lambda^{-1}\Lambda\tilde{K} = f_W^{-\frac{1}{2}}\tilde{K}$ and $\forall x \in \mathcal{X}$, $||\lim_{n\to\infty}\Sigma^{-1/2}Kx|| = ||\mathbf{E}(x|w)||_W^2$, where $||\phi||_W^2 := \int[\phi(w)]^2 f_W(w)dw$. This shows that, in the IV case, assumption A.2 is a particular case of assumptions 2.2 and 4.2 in Chen and Reiss (2011).

Finally, we check assumption A.3 for both $n \to \infty$ and finite sample. In finite sample this assumption holds trivially since $\mathcal{R}(\hat{\Sigma})$ $(= \mathcal{D}(\hat{\Sigma}^{-1}))$ and $\mathcal{R}(K\Omega_0^{\frac{1}{2}})$ have finite ranks. Suppose that the conditions for the application of the Dominated Convergence Theorem hold, then Assumption A.3 is satisfied asymptotically if and only if $||\Sigma^{-1}\Lambda\lim_{n\to\infty}\tilde{K}\Omega_0^{\frac{1}{2}}||^2 < \infty$. This holds if Ω_0 is appropriately chosen. One possibility could be to set $\Omega_0 = T^*\Lambda^*\Lambda T$, where $T : L^2[0,1] \to L^2[0,1]$ is a trace-class integral operator $T\phi = \int \omega(w,z)\phi(z)dz$ for a known function ω and T^* is its adjoint. Define $\Omega_0^{\frac{1}{2}} = T^*\Lambda^*$. Then,[2]

$$\begin{aligned} ||\Sigma^{-1}\Lambda\lim_{n\to\infty}\tilde{K}\Omega_0^{\frac{1}{2}}||^2 &= ||\Sigma^{-1}\Lambda\lim_{n\to\infty}\tilde{K}T^*\Lambda^*||^2 \leq ||\Sigma^{-1}\Lambda\lim_{n\to\infty}\tilde{K}T^*\Lambda^*||_{HS}^2 \leq tr(\Sigma^{-1}\Lambda\lim_{n\to\infty}\tilde{K}T^*\Lambda^*) \\ &= tr(\Lambda^*\Sigma^{-1}\Lambda\lim_{n\to\infty}\tilde{K}T^*) = tr(\mathbf{E}(T^*\cdot|w)) \leq tr(T^*)||\mathbf{E}(\cdot|W)|| < \infty. \end{aligned}$$

□

2.2.1 Covariance operators proportional to K

A particular situation often encountered in applications is the case where the sampling covariance operator has the form $\Sigma = (KK^*)^r$, for some $r \in \mathbb{R}_+$, and is related to the classical example of g-priors given in Zellner (1986). In this situation it is convenient to choose $L = (K^*K)^{-\frac{1}{2}}$ so

[2] This is because for a compact operator $A : L^2[0,1] \to L^2[0,1]$ and by denoting $|A| = (A^*A)^{1/2}$, we have: $||A||^2 \leq ||A||_{HS}^2 = ||\,|A|\,||_{HS}^2 \leq tr(|A|) = tr(A)$.

that $\Omega_0 = (K^*K)^s$, for $s \in \mathbb{R}_+$. Because $(KK^*)^r$ and $(K^*K)^s$ are proper covariance operators only if they are trace-class then K is necessarily compact. Assumptions A.1 and A.2 hold for $r \leq 1$ and $a = 1 - r$, respectively. Assumption A.3 holds for $s \geq 2r - 1$.

Example 2 (Functional Linear Regression Estimation *(continued)*)**.** Let us consider model (3) and the associated integral equation $\mathbf{E}(\xi Z(t)) = \int_0^1 x(s)Cov(Z(s), Z(t))ds$, for $t \in [0, 1]$. If we dispose of *i.i.d.* data $(\xi_1, Z_1), \ldots, (\xi_n, Z_n)$ the unknown moments in this equation can be estimated. Thus, x is solution of (4) with $y^\delta := \frac{1}{n}\sum_i \xi_i Z_i(t)$, $U = \frac{1}{n}\sum_i \varepsilon_i Z_i(t)$ and $\forall \varphi \in L^2([0, 1]) \mapsto K\varphi := \frac{1}{n}\sum_i < Z_i, \varphi > Z_i(t)$. The operator $K : L^2([0, 1]) \to L^2([0, 1])$ is self-adjoint, *i.e.* $K = K^*$. Moreover, conditional on Z, the error term U is exactly a Gaussian process with covariance operator $\delta\Sigma = \delta\tau^2 K$ with $\delta = \frac{1}{n}$ which is trace-class since its range has finite dimension. Thus, we can write $\delta\Sigma = \frac{1}{n}\tau^2(KK^*)^r$ with $r = \frac{1}{2}$. Assumption A.1 is trivially satisfied in finite sample as well as for $n \to \infty$. We discuss later on how to choose L in order to satisfy assumptions A.2 and A.3. □

3 Main results

The posterior distribution of x, denoted by μ_δ^Y, is the Bayesian solution of the inverse problem (1). Because a separable Hilbert space is Polish, there exists a regular version of the posterior distribution μ_δ^Y, that is, a conditional probability characterizing μ_δ^Y. In many applications $\mathcal{X}$ and $\mathcal{Y}$ are L^2 spaces and L^2 spaces are Polish if they are defined on a separable metric space. In the next Theorem we characterize the joint distribution of (x, y^δ) and the posterior distribution μ_δ^Y of x. The notation $\mathfrak{B}(\mathcal{X}) \otimes \mathfrak{B}(\mathcal{Y})$ means the Borel σ-field generated by the product topology.

Theorem 1. *Consider two separable infinite dimensional Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$. Let $x|\alpha, s$ and $y^\delta|x$ be two Gaussian H-r.v. on $\mathcal{X}$ and $\mathcal{Y}$ as in (7) and (5), respectively. Then,*

(i) (x, y^δ) is a measurable map from $(S, \mathcal{S}, \mathbb{P})$ to $(\mathcal{X} \times \mathcal{Y}, \mathfrak{B}(\mathcal{X}) \otimes \mathfrak{B}(\mathcal{Y}))$ and has a Gaussian distribution: $(x, y^\delta)|\alpha, s \sim \mathcal{N}((x_0, Kx_0), \Upsilon)$, where Υ is a trace-class covariance operator defined as $\Upsilon(\varphi, \psi) = \left(\frac{\delta}{\alpha}\Omega_0\varphi + \frac{\delta}{\alpha}\Omega_0 K^*\psi, \frac{\delta}{\alpha}K\Omega_0\varphi + (\delta\Sigma + \frac{\delta}{\alpha}K\Omega_0K^*)\psi\right)$ for all $(\varphi, \psi) \in \mathcal{X} \times \mathcal{Y}$. The marginal sampling distribution of $y^\delta|\alpha, s$ is $P_\alpha \sim \mathcal{N}(Kx_0, (\delta\Sigma + \frac{\delta}{\alpha}K\Omega_0K^*))$.

Moreover, let $A : \mathcal{Y} \to \mathcal{X}$ be a P_α-measurable linear transformation (P_α-mlt), that is, $\forall \phi \in \mathcal{Y}$, $A\phi$ is a P_α-almost sure limit, in the norm topology, of $A_k\phi$ as $k \to \infty$, where $A_k : \mathcal{Y} \to \mathcal{X}$ is a sequence of continuous linear operators. Then,

(ii) the conditional distribution μ_δ^Y of x given y^δ exists, is regular and almost surely unique. It is Gaussian with mean $\mathbf{E}(x|y^\delta, \alpha, s) = A(y^\delta - Kx_0) + x_0$ and trace-class covariance operator $Var(x|y^\delta, \alpha, s) = \frac{\delta}{\alpha}[\Omega_0 - AK\Omega_0] : \mathcal{X} \to \mathcal{X}$. Furthermore, $A = \Omega_0 K^*(\alpha\Sigma + K\Omega_0K^*)^{-1}$ on $\mathcal{R}((\delta\Sigma + \frac{\delta}{\alpha}K\Omega_0K^*)^{\frac{1}{2}})$.

(iii) Under Assumptions A.1 and A.3, the operator A characterizing μ_{δ}^{Y} is a continuous linear operator on $\mathcal{Y}$ and can be written as

$$A = \Omega_0^{\frac{1}{2}}(\alpha I + B^*B)^{-1}(\Sigma^{-1/2}B)^* : \mathcal{Y} \to \mathcal{X} \tag{9}$$

with $B = \Sigma^{-1/2}K\Omega_0^{\frac{1}{2}}$.

Point *(ii)* in the theorem is an immediate application of the results of Mandelbaum (1984), we refer to this paper for the proof and for a rigorous definition of P_α-mlt. As stated above, the quantity Ay^δ is defined as a P_α-mlt, which is a weaker notion than that of a linear and continuous operator and A is in general not continuous. In fact, since Ay^δ is a P_α-almost sure limit of $A_k y^\delta$, for $k \to \infty$, the null set where this convergence is not satisfied depends on y^δ and we do not have an almost sure convergence of A_k to A. Moreover, in general, A takes the form $A = \Omega_0 K^*(\alpha\Sigma + K\Omega_0 K^*)^{-1}$ only on a dense subspace of $\mathcal{Y}$ of P_α-probability measure 0. Outside of this subspace, A is defined as the unique extension of $\Omega_0 K^*(\alpha\Sigma + K\Omega_0 K^*)^{-1}$ to $\mathcal{Y}$ for which we do not have an explicit expression. This means that in general it is not possible to construct a feasible estimator for x.

On the contrary, point *(iii)* of the theorem shows that, under A.1 and A.3, A is defined as a continuous linear operator on the whole $\mathcal{Y}$. This is the first important contribution of our paper since A.1 and A.3 permit to construct a point estimator for x – equal to the posterior mean – that is implementable in practice. Thus, our result *(iii)* make operational the Bayesian approach for linear statistical inverse problems in econometrics. When assumptions A.1 and A.3 do not hold then we can use a quasi Bayesian approach as proposed in Florens and Simoni (2012a,b). Summarizing, under a quadratic loss function, the Bayes estimator for a functional parameter x characterized by (1) is

$$\hat{x}_\alpha = \Omega_0^{\frac{1}{2}}(\alpha I + B^*B)^{-1}(\Sigma^{-1/2}B)^*(y^\delta - Kx_0), \quad \text{with} \quad B = \Sigma^{-1/2}K\Omega^{1/2}.$$

Remark 3.1. Under our assumptions it is possible to show the existence of a close relationship between Bayesian and frequentist approach to statistical inverse problems. In fact, the posterior mean $\hat{x}_\alpha$ is equal to the Tikhonov regularized solution in the Hilbert scale $(\mathcal{X}_s)_{s\in\mathbb{R}}$ generated by L of the equation $\Sigma^{-1/2}y^\delta = \Sigma^{-1/2}Kx + \Sigma^{-1/2}U$. The existence of $\Sigma^{-1/2}K$ is guaranteed by A.1. Since $\mathbf{E}(x|y^\delta, \alpha, s) = A(y^\delta - Kx_0) + x_0$, we have, under A.1 and A.3:

$$\begin{aligned} \mathbf{E}(x|y^\delta, \alpha, s) &= L^{-s}(\alpha I + L^{-s}K^*\Sigma^{-1}KL^{-s})^{-1}L^{-s}K^*\Sigma^{-1/2}\Sigma^{-1/2}(y^\delta - Kx_0) + x_0 \\ &= (\alpha L^{2s} + T^*T)^{-1}T^*(\tilde{y}^\delta - Tx_0) + x_0, \qquad T = T = \Sigma^{-1/2}K, \quad \tilde{y}^\delta = \Sigma^{-1/2}y^\delta \end{aligned}$$

and it is equal to the minimizer, with respect to x, of the Tikhonov functional

$$||\tilde{y}^\delta - Tx||^2 + \alpha||x - x_0||_s^2.$$

The model $\Sigma^{-1/2}y^{\delta} = Tx + \Sigma^{-1/2}U$ is the transformation of (1) through the operator $\Sigma^{-1/2}$. The quantities $\Sigma^{-1/2}y^{\delta}$ and $\Sigma^{-1/2}U$ can not be interpreted as H-r.v. but have to be interpreted in process form (as in Bissantz *et al.* (2007)), *i.e.* in the sense of weak distributions, see Kuo (1975). More precisely, $Z := \Sigma^{-1/2}U$ is a Hilbert space process in $\mathcal{Y}$ if $Z(\varphi) :=< Z, \varphi >: \mathcal{Y} \to L^2(S, \mathcal{S}, \mathbb{P})$ is a random variable with zero mean and $Cov_Z = I$, where $Cov_Z : \mathcal{Y} \to \mathcal{Y}$ is the covariance operator characterized by $< Cov_Z\varphi, \psi >= \mathbf{E}(Z(\varphi)Z(\psi))$, $\forall \varphi, \psi \in \mathcal{Y}$.

In the IV regression estimation, the equation $\Sigma^{-1/2}y^{\delta} = Tx + \Sigma^{-1/2}U$ writes $\hat{\mathbf{E}}(Y|W)\hat{f}_W^{\frac{1}{2}} = \hat{\mathbf{E}}(x|W)\hat{f}_W^{\frac{1}{2}} + error$ and the posterior mean estimator writes

$$\hat{x}_{\alpha} = \left(\alpha L^{2s} + \tilde{K}^* \frac{1}{\hat{f_W}} \tilde{K}\right)^{-1} \tilde{K}^* \left(\hat{E}(Y - x_0|W)\right) = \left(\alpha L^{2s} + \hat{\mathbf{E}}\left[\hat{\mathbf{E}}(\cdot|W)|Z\right]\right)^{-1} \hat{\mathbf{E}}\left[\hat{\mathbf{E}}(Y - x_0|W)|Z\right].$$

For $x_0 = 0$ this is, in the framework of Darolles *et al.* (2011) or Florens *et al.* (2011), the Tikhonov estimator in the Hilbert scale generated by $\frac{1}{f_Z}L^{2s}$. $\Box$

3.1 Asymptotic analysis

We analyze now frequentist asymptotic properties of the posterior distribution μ_{δ}^{Y} of x. The asymptotic analysis is for $\delta \to 0$. Let P^{x_*} denote the sampling distribution (5) with $x = x_*$, we remind the definition of *posterior consistency*, see Diaconis and Freedman (1986) or Ghosh and Ramamoorthi (2003):

Definition 1. *The posterior distribution is* consistent *at x_* with respect to P^{x_*} if it weakly converges towards the Dirac measure δ_{x_*} at the point x_*, i.e. if, for every neighborhood $\mathcal{U}$ of x_*, $\mu_{\delta}^{Y}(\mathcal{U}|y^{\delta}, \alpha, s) \to 1$ in P^{x_*}-probability or P^{x_*}-a.s. as $\delta \to 0$.*

Posterior consistency provides the basic frequentist validation of our Bayesian procedure because it ensures that with a sufficiently large amount of data, it is almost possible to recover the truth accurately. Lack of consistency is extremely undesirable, and one should not use a Bayesian procedure if the corresponding posterior distribution is inconsistent. Our asymptotic analysis is organized as follows. First, we take the posterior mean $\hat{x}_{\alpha}$ as an estimator for the solution of (1) and study the rate of convergence of the associated risk. Second, we state posterior consistency and recover the rate of contraction of the posterior distribution. We denote the risk (MISE) associated with $\hat{x}_{\alpha}$ by $\mathbf{E}_{x_*}||\hat{x}_{\alpha} - x_*||^2$ where $\mathbf{E}_{x_*}$ denotes the expectation taken with respect to P^{x_*}. Let $\{\lambda_{jL}\}$ denote the eigenvalues of L^{-1}, we define:

$$\gamma := \inf\{\tilde{\gamma} \in (0, 1]\, ;\, \sum_{j=1}^{\infty} \lambda_{jL}^{2\tilde{\gamma}(a+s)} < \infty\} \equiv \inf\{\tilde{\gamma} \in (0, 1]\, ;\, tr(L^{-2\tilde{\gamma}(a+s)}) < \infty\}. \quad (10)$$

We point out that γ is known since it depends on L. For instance, $\gamma = \frac{1}{2(a+s)}$ means that either L^{-1} is trace-class but $L^{-(1-\omega)}$ is not trace-class for every $\omega \in \mathbb{R}_+$ or that $tr\left(L^{-(1+\omega)}\right) < \infty$ $\forall \omega \in \mathbb{R}_+$ but

L^{-1} is not trace-class. Since under A.2 the operator L^{-2s} is trace-class, the parameter γ cannot be larger than $\frac{s}{a+s}$. Thus, if $\gamma = \frac{1}{2(a+s)}$ this implies that $s \geq \frac{1}{2}$ since $\frac{1}{2(a+s)}$ must be less than or equal to $\frac{s}{(a+s)}$. Remark that the smaller the γ is and the smaller the eigenvalues of L^{-1} are. Furthermore, we denote by $\mathcal{X}_\beta(\Gamma)$ the ellipsoid of the type

$$\mathcal{X}_\beta(\Gamma) := \{\varphi \in \mathcal{X}\,;\ ||\varphi||_\beta^2 \leq \Gamma\}, \qquad 0 < \Gamma < \infty \tag{11}$$

where $||\varphi||_\beta := ||L^\beta \varphi||$. Our asymptotic results will be valid uniformly on $\mathcal{X}_\beta(\Gamma)$. The following theorem gives the asymptotic behavior of $\hat{x}_\alpha$.

Theorem 2. *Let us consider the observational model (4) with x_* being the true value of x that generates the data. Under Assumptions A.1-A.3 and B, we have*

$$\sup_{(x_*-x_0)\in\mathcal{X}_\beta(\Gamma)} \mathbf{E}_{x_*}||\hat{x}_\alpha - x_*||^2 = \mathcal{O}\left(\alpha^{\frac{\tilde{\beta}}{a+s}} + \delta\alpha^{-\frac{a+\gamma(a+s)}{a+s}}\right)$$

with $\tilde{\beta} = \min(\beta, a+2s)$.

The value $a+2s$ plays the role of a *qualification* in a classical regularization scheme, that is, it limits to $a+2s$ the regularity of x_* that can be exploited in order to improve the rate. It is equal to the qualification of a Tikhonov regularization in the Hilbert scale $(\mathcal{X}_s)_{s\in\mathbb{R}}$, see *e.g.* Engl *et al.* (2000) Section 8.5.

The value of α that minimizes the rate given in theorem 2 is: $\alpha^{\min} := \kappa\delta^{\frac{a+s}{\tilde{\beta}+a+\gamma(a+s)}}$, for some constant $\kappa > 0$. When α is set equal to $\alpha^{\min}$ (*i.e.* the prior is scaling), then $\sup_{(x_*-x_0)\in\mathcal{X}_\beta(\Gamma)} \mathbf{E}_{x_*}||\hat{x}_{\alpha^{\min}} - x_*||^2 = \mathcal{O}\left(\delta^{\frac{\tilde{\beta}}{\tilde{\beta}+a+\gamma(a+s)}}\right)$ and this rate is equal to the minimax rate $\delta^{\frac{\beta}{\beta+a+\frac{1}{2}}}$ if $\beta \leq a+2s$ and $\gamma = \frac{1}{2(a+s)}$ (which is possible only if $s \geq \frac{1}{2}$). If the regularity of the prior matches the regularity of the truth through the relation $s = \beta + \frac{1}{2}$ then a prior with $\alpha = \alpha^{\min} = \kappa\delta$ is non-scaling and still provides the minimax rate of convergence if $\gamma = \frac{1}{2(a+s)}$ (in this case $\tilde{\beta} = \beta$ since $\beta = s - \frac{1}{2} < a+2s$). When $s \neq \beta + \frac{1}{2}$ (and $\gamma = \frac{1}{2(a+s)}$) the prior must be scaled in order to achieve a minimax rate of convergence: if $s > \beta + \frac{1}{2}$ the prior has to spread out (to become rougher) while if $s < \beta + \frac{1}{2}$ the prior must shrink (to become smoother).[3]

In all the other cases where $\alpha \neq \alpha^{\min}$ the rate is slower than $\delta^{\frac{\beta}{\beta+a+\frac{1}{2}}}$ but we still have consistency provided that we set $\alpha \asymp \delta^\epsilon$ for $0 < \epsilon < \frac{a+s}{a+\gamma(a+s)}$. Remark that since $tr(L^{-2s}) < \infty$ under A.2 then $\gamma \leq \frac{s}{a+s}$ so that $\frac{a+s}{a+\gamma(a+s)} \geq 1$. Thus, 1 is a possible value for ϵ which implies that consistency is always obtained with a non-scaling prior even if the minimax rate is obtained only in particular cases.

[3]Remark that a γ smaller than $\frac{1}{2(a+s)}$ implies that $tr(L^{-1}) < tr(L^{1/g}) < \infty$ for some $g > 1$. This means that L^{-1} is very smooth and its spectrum is decreasing fast. Thus, if $(x_* - x_0)$ is not very smooth then assumption B will be satisfied only with a β very small. A small β will decrease the rate of convergence.

The same discussion can be made concerning the rate of contraction of the posterior distribution which is given in the next theorem.

Theorem 3. *Let the assumptions of Theorem 2 be satisfied. For any sequence $M_\delta \to \infty$ the posterior distribution satisfies*

$$\mu_\delta^Y\{x \in \mathcal{X} : ||x - x_*|| > \varepsilon_\delta M_\delta\} \to 0$$

in P^{x_*}-probability as $\delta \to 0$, where $\varepsilon_\delta = \left(\alpha^{\frac{\tilde{\beta}}{2(a+s)}} + \delta^{\frac{1}{2}}\alpha^{-\frac{a+\gamma(a+s)}{2(a+s)}}\right)$ *and* $\tilde{\beta} = \min(\beta, a+2s)$.

We refer to ε_δ as the rate of contraction of the posterior distribution. If the prior is fixed, that is, $\alpha \asymp \delta$, then $\varepsilon_\delta = \delta^{\frac{\beta\wedge(s-\gamma(a+s))}{2(a+s)}}$. If α is chosen such that the two terms in ε_δ are balanced, that is, $\alpha \asymp \alpha^{\min}$, then $\varepsilon_\delta = \delta^{\frac{\tilde{\beta}}{2\tilde{\beta}+2a+2\gamma(a+s)}}$. The minimax rate $\delta^{\frac{\beta}{2\beta+2a+1}}$ is obtained when $\beta \leq a+2s$, $\gamma = \frac{1}{2(a+s)}$ and we set $\alpha = \alpha^{\min}$. In this case, the prior is either fixed or scaling depending whether s equates $\beta + \frac{1}{2}$ or not.

3.2 Example 1: Instrumental variable (IV) regression estimation (continued)

In this section we explicit the rate of theorem 2 for the IV regression estimation. Remark that $B^*B = \Omega_0^{\frac{1}{2}}\tilde{K}^*[\sigma^2\hat{f}_W]^{-1}\tilde{K}\Omega_0^{\frac{1}{2}}$ where $\hat{f}_W = \frac{1}{n}\sum_{i=1}^n \frac{K_{W,h}(w_i-w)}{h}$, $\tilde{K}$ has been defined before display (8) and $\tilde{K}^* : L^2[0,1] \to L^2[0,1]$, is the adjoint operator of $\tilde{K}$ that takes the form: $\tilde{K}^*\phi = \frac{1}{n}\sum_{i=1}^n \frac{K_{Z,h}(z_i-z)}{h} < \phi, \frac{K_{W,h}(w_i-w)}{h} >$, $\forall\phi \in L^2[0,1]$. The Bayesian estimator of the IV regression is $\hat{x}_\alpha = \Omega_0^{\frac{1}{2}}(\alpha I + B^*B)^{-1}(\Sigma^{-1/2}B)^*(y^\delta - \Lambda\tilde{K}x_0)$. We assume that the true IV regression x_* that generates the data satisfies Assumption B. In order to determine the rate of the MISE associated with $\hat{x}_\alpha$ the proof of theorem 2 must be slightly modified. This is because the covariance operator of U and K in the inverse problem associated with the IV regression estimation are changing with n. Therefore, the rate of the MISE must incorporate the rate of convergence of these operators towards their limits. The crucial issue in order to establish the rate of the MISE associated with $\hat{x}_\alpha$ is the rate of convergence of B^*B towards $\Omega_0^{\frac{1}{2}}\mathfrak{K}^*\frac{1}{\sigma^2 f_W}\mathfrak{K}\Omega_0^{\frac{1}{2}}$ where $\mathfrak{K} = \lim_{n\to\infty}\tilde{K} = f_W\mathbf{E}(\cdot|W)$ and $\mathfrak{K}^* = \lim_{n\to\infty}\tilde{K}^* = f_Z\mathbf{E}(\cdot|Z)$. This rate is specified by Assumption (HS) below and we refer to Darolles *et al.* (2011, Appendix B) for a set of sufficient conditions that justify this assumption.

Assumption HS. *There exists $\rho \geq 2$ such that:*

(i). $\mathbf{E}||B^*B - \Omega_0^{\frac{1}{2}}\mathfrak{K}^*\frac{1}{\sigma^2 f_W}\mathfrak{K}\Omega_0^{\frac{1}{2}}||^2 = \mathcal{O}\left(n^{-1} + h^{2\rho}\right)$;

(ii). $\mathbf{E}||\Omega_0^{\frac{1}{2}}(\hat{\mathcal{T}}^* - \mathcal{T}^*)||^2 = \mathcal{O}\left(\frac{1}{n} + h^{2\rho}\right)$, *where* $\mathcal{T}^* = \mathbf{E}(\cdot|Z)$ *and* $\hat{\mathcal{T}}^* = \hat{\mathbf{E}}(\cdot|Z)$.

To get rid of σ^2 it is sufficient to specify Ω_0 as $\Omega_0\sigma^2$ so that B^*B does not depend on σ^2 anymore and we do not need to estimate it to get the estimate of x. The next corollary to theorem 2 gives the rate of the MISE of $\hat{x}_\alpha$ under this assumption.

Corollary 1. *Let us consider the observational model $y^\delta = Kx_* + U$, with y^δ and K defined as in (8) and x_* being the true value of x. Under Assumptions A.1-A.3, B and HS:*

$$\sup_{(x_*-x_0)\in\mathcal{X}_\beta(\Gamma)} \mathbf{E}_{x_*}||\hat{x}_\alpha - x_*||^2 = \mathcal{O}\Big(\left(\alpha^{\frac{\tilde{\beta}}{a+s}} + n^{-1}\alpha^{-\frac{a+\gamma(a+s)}{a+s}}\right)(1+\alpha^{-2}\left(\frac{1}{n}+h^{2\rho}\right)) + \alpha^{-2}\left(\frac{1}{n}+h^{2\rho}\right)\left(\frac{1}{nh}+h^{2\rho}\right)\Big)$$

with $\tilde{\beta} = \min(\beta, a+2s)$.

In this example we have to set two tuning parameters: the bandwidth h and α. We can set h such that $h^{2\rho}$ goes to 0 at least as fast as n^{-1} and $\alpha = \alpha^{\min} \propto n^{-\frac{a+s}{\tilde{\beta}+a+\gamma(a+s)}}$. With this choice, the rate of convergence of B^*B towards $\Omega_0^{\frac{1}{2}}\mathfrak{K}^*\frac{1}{f_W}\mathfrak{K}\Omega_0^{\frac{1}{2}}$ and of $\Omega_0^{\frac{1}{2}}\hat{\mathcal{T}}^*$ towards $\Omega_0^{\frac{1}{2}}\mathcal{T}^*$ will not affect the rate in the MISE (that is, the rate will be the same as the rate given in theorem 2) if $\tilde{\beta} \geq 2s + a - 2\gamma(a+s)$ and $\rho > \frac{\tilde{\beta}+a+\gamma(a+s)}{\tilde{\beta}+2\gamma(a+s)-2s}$. We remark that when the prior is not scaling, *i.e.* $\alpha \asymp n^{-1}$, the condition $[\alpha^2 n]^{-1} = \mathcal{O}(1)$ is not satisfied. The rate of corollary 1, with $\alpha = \alpha^{\min}$ and $h = \mathcal{O}(n^{-1/(2\rho)})$, is minimax when $a + 2s \geq \beta \geq a + 2s - \frac{1}{2}$ and $\gamma = \frac{1}{2(a+s)}$.

4 Operators with geometric spectra

We analyze now the important case where the inverse problem (1) is *mildly ill-posed*. We denote with λ_{jK} the singular values of K and with $\lambda_{j\Sigma}$ and λ_{jL} the eigenvalues of Σ and L^{-1}, respectively. Assumption C states that the operators $\Sigma^{-1/2}$ and KK^* (resp. $K^*\Sigma^{-1}K$ and L^{-1}) are diagonalizable in the same eigenbasis and have polynomially decreasing spectra.

Assumption C. *The operator $\Sigma^{-1/2}$ (resp. $K^*\Sigma^{-1}K$) has the same eigenfunctions $\{\varphi_j\}$ as KK^* (resp. $\{\psi_j\}$ as L^{-1}). Moreover, the eigenvalues of KK^*, Σ and L^{-1} satisfy*

$$\underline{a}j^{-2a_0} \leq \lambda_{jK}^2 \leq \overline{a}j^{-2a_0}, \quad \underline{c}j^{-c_0} \leq \lambda_{j\Sigma} \leq \overline{c}j^{-c_0} \quad \textit{and} \quad \underline{l}j^{-1} \leq \lambda_{jL} \leq \overline{l}j^{-1}, \quad j = 1, 2, \ldots$$

with $a_0 \geq 0, c_0 > 1$ and $\underline{a}, \overline{a}, \underline{c}, \overline{c}, \underline{l}, \overline{l} > 0$.

This assumption implies that KK^* and K^*K are strictly positive definite. For the setting described by Assumption C we give in this section the exact rate attained by $\hat{x}_\alpha$.

Assumption C is standard in statistical inverse problems literature (see *e.g.* assumption A.3 in Hall and Horowitz (2005) or assumption B3 in Cavalier and Tsybakov (2002)) and, by using the notation defined in 2.1, it may be rewritten as $\lambda_{jK} \asymp j^{-a_0}$, $\lambda_{j\Sigma} \asymp j^{-c_0}$ and $\lambda_{jL} \asymp j^{-1}$. Under Assumption C we may rewrite $\mathcal{X}_\beta(\Gamma)$ as $\mathcal{X}_\beta(\Gamma) := \{\varphi \in \mathcal{X}\ ;\ \sum_j j^{2\beta} < \varphi, \psi_j >^2 \leq \Gamma\}$ and we may explicitly compute γ: it is equal to $\frac{1}{2(a+s)}$ so that this value, and not γ, will appear in the rate. The following proposition provides necessary and sufficient conditions for A.1, A.2 and A.3, when C is satisfied.

Proposition 1. *Under Assumption C:*

- *A.1 is satisfied if and only if* $a_0 \geq \frac{c_0}{2}$*;*
- *A.2 is satisfied if and only if* $a = a_0 - \frac{c_0}{2} > 0$ *and* $s > \frac{1}{2}$*;*
- *A.3 is satisfied if and only if* $a_0 \geq c_0 - s$*.*

The following Proposition gives the minimax rate attained by $\hat{x}_\alpha$. Then,

Proposition 2. *Let B, C hold,* $a = a_0 - \frac{c_0}{2} > 0$*,* $s > \frac{1}{2}$ *and* $a_0 \geq c_0 - s$*. Then we have*

$$\sup_{(x_* - x_0) \in \mathcal{X}_\beta(\Gamma)} \mathbf{E}_{x_*} ||\hat{x}_\alpha - x_*||^2 \asymp \alpha^{\frac{\tilde{\beta}}{a+s}} + \delta \alpha^{-\frac{2a+1}{2(a+s)}}, \tag{12}$$

with $\tilde{\beta} = \min(\beta, 2(a+s))$*. Moreover, (i) for* $\alpha \asymp \delta$ *(fixed prior),*

$$\sup_{(x_* - x_0) \in \mathcal{X}_\beta(\Gamma)} \mathbf{E}_{x_*} ||\hat{x}_\alpha - x_*||^2 \asymp \delta^{\frac{\beta \wedge \left(s - \frac{1}{2}\right)}{a+s}};$$

(ii) for $\alpha \asymp \delta^{\frac{a+s}{\beta + a + \frac{1}{2}}}$ *(optimal rate of* α*),*

$$\sup_{(x_* - x_0) \in \mathcal{X}_\beta(\Gamma)} \mathbf{E}_{x_*} ||\hat{x}_\alpha - x_*||^2 \asymp \delta^{\frac{\tilde{\beta}}{\beta + a + \frac{1}{2}}}.$$

The minimax rate of convergence over a Sobolev ellipsoid $\mathcal{X}_\beta(\Gamma)$ is of the order $\delta^{\frac{\beta}{\beta + a + \frac{1}{2}}}$. By the results of the proposition, the uniform rate of the MISE associated with $\hat{x}_\alpha$ is minimax if the parameter s of the prior is chosen such that $s = \beta + \frac{1}{2}$ and the prior is fixed (case *(i)*) or if $\beta \leq 2(a+s)$ and the prior is scaling at the optimal rate (case *(ii)*). In all the other cases the rate is slower than the minimax rate but consistency is still verified provided that $\alpha \asymp \delta^\epsilon$ for $0 < \epsilon < \frac{a+s}{a+\frac{1}{2}}$. Remark that since $s > \frac{1}{2}$ then a fixed prior ($\epsilon = 1$) always guarantees consistency (even if the rate is not always minimax).

This result, similar to that one in Theorem 2, means that when the prior is "correctly specified" ("correct" in the sense that the regularity $s - \frac{1}{2}$ of the trajectories generated by the prior is the same as the regularity of x_*) we obtain a minimax rate without scaling the prior covariance. On the other hand, if $s < \beta + \frac{1}{2}$, *i.e.* the prior is "undersmoothing", the minimax rate can still be obtained as soon as $\beta \leq 2(a+s)$ and the prior is shrinking at the optimal rate. When the prior is "oversmoothing", *i.e.* $s > \beta + \frac{1}{2}$, the minimax rate can be obtained if the prior distribution spreads out at the optimal α (the prior has to be more and more dispersed in order to become rougher).

In many cases it is reasonable to assume that the functional parameter x_* has generalized Fourier coefficients (in the basis made of the eigenfunctions $\{\psi_j\}$ of L^{-s}) that are geometrically decreasing, see *e.g.* assumption A.3 in Hall and Horowitz (2005) and theorem 4.1 in Van Rooij and Ruymgaart (1999). Thus, we may consider the following assumption.

Assumption B'. *For some* $b_0 > \frac{1}{2}$ *and* $\{\psi_j\}$ *defined in Assumption C,* $< (x_* - x_0), \psi_j > \asymp j^{-b_0}$.

Assumption B' is often encountered in statistical inverse problems literature. Assumption B is more general than Assumption B' since it allows to consider the important case of exponentially declining Fourier coefficients $< (x_* - x_0), \psi_j >$. We use Assumption B' to show sharp adaptiveness for our Empirical Bayes procedure. If B' holds then $\exists \Gamma < \infty$ such that assumption B holds for some $0 \leq \beta < b_0 - \frac{1}{2}$. The following result gives the rate of the MISE when assumption B' holds.

Proposition 3. *Let B', C hold with* $a = a_0 - \frac{c_0}{2} > 0$, $s > \frac{1}{2}$ *and* $a_0 \geq c_0 - s$*. Then,*

$$\alpha^{\frac{2b_0-1}{2(a+s)}} c_1 + \delta\alpha^{-\frac{2a+1}{2(a+s)}} c_2(\bar{t}) \leq \mathbf{E}_{x_*}||\hat{x}_\alpha - x_*||^2 \leq \alpha^2 I(b_0 \geq (2a+2s)) + \alpha^{\frac{2b_0-1}{2(a+s)}} \tilde{c}_1 + \delta\alpha^{-\frac{2a+1}{2(a+s)}} \tilde{c}_2(\bar{t}) \quad (13)$$

where c_1, $\tilde{c}_1$, c_2, $\tilde{c}_2$ *and* $\bar{t}$ *are positive constants. Moreover, for* $\alpha \asymp \delta^{\frac{a+s}{\tilde{b}_0+a}} \equiv \alpha_*$ *and* $\tilde{b}_0 = \min(b_0, 2a + 2s + 1/2)$,

$$\mathbf{E}_{x_*}||\hat{x}_{\alpha_*} - x_*||^2 \asymp \delta^{\frac{2\tilde{b}_0-1}{2(\tilde{b}_0+a)}}. \quad (14)$$

When $b_0 \leq 2a+2s+1/2$ the rate of the lower bound $\alpha^{\frac{2b_0-1}{2(a+s)}} + \delta\alpha^{-\frac{2a+1}{2(a+s)}}$ given in (13) provides, up to a constant, a lower and an upper bound for the rate of the estimator $\hat{x}_\alpha$ and so it is optimal. Thus, the minimax-optimal rate $\delta^{\frac{2b_0-1}{2(b_0+a)}}$ is obtained when we set $\alpha \asymp \delta^{\frac{a+s}{b_0+a}}$ if: either $s = b_0$ (fixed prior with $\alpha \asymp \delta$), or $s < b_0 \leq 2a + 2s + 1/2$ (shrinking prior), or $s > b_0$ (spreading out prior). When $s < b_0$ (resp. $s > b_0$) the trajectories generated by the prior are less smooth (resp. smoother) than x_* and so the support of the prior must be shrunk (resp. spread out). When $\delta = n^{-1}$ the rate $n^{-\frac{(2b_0-1)}{2(b_0+a)}}$ is shown to be minimax in Hall and Horowitz (2007).

Moreover, if we set $\beta = \sup\left\{\beta \geq 0;\ (x_* - x_0) \text{ satisfies B' and } \sum_j j^{2\beta} < (x_* - x_0), \psi_j >^2 < \infty\right\}$ then $\beta = b_0 - \frac{1}{2}$ and the rate $\delta^{\frac{2\tilde{b}_0-1}{2(\tilde{b}_0+a)}}$ is uniform in x_* over $\mathcal{X}_\beta(\Gamma)$ and equal to the optimal rate of proposition 2.

In the following theorem we give the rate of contraction of the posterior distribution for the *mildly ill-posed case.*

Theorem 4. *Let the assumptions of Proposition 2 be satisfied. For any sequence* $M_\delta \to \infty$ *the posterior probability satisfies*

$$\mu_\delta^Y\{x \in \mathcal{X} : ||x - x_*|| > \varepsilon_\delta M_\delta\} \to 0$$

in P^{x_*}*-probability as* $\delta \to 0$*, where* $\varepsilon_\delta = \left(\alpha^{\frac{2\tilde{b}_0-1}{4(a+s)}} + \delta^{\frac{1}{2}}\alpha^{-\frac{2a+1}{4(a+s)}}\right)$, $\alpha > 0$, $\alpha \to 0$ *and* $\tilde{b}_0 = \min(\beta, 2(a+s) + 1/2)$*. Moreover, (i) for* $\alpha \asymp \delta$ *(fixed prior):*

$$\varepsilon_\delta = \delta^{\frac{2(b_0 \wedge s)-1}{4(a+s)}}$$

and (ii) for $\alpha \asymp \delta^{\frac{a+s}{\tilde{b}_0+a}}$ *(optimal rate of* α*):*

$$\varepsilon_\delta = \delta^{\frac{2\tilde{b}_0-1}{4(\tilde{b}_0+a)}}.$$

The rate of contraction is minimax if $b_0 \leq 2(a+s)+1/2$ and $\alpha \asymp \delta^{\frac{a+s}{b_0+a}}$. Depending on the relation between s and b_0 the corresponding prior μ is shrinking, spreading out or fixed, see comments after proposition 3.

4.1 Example 2: Functional Linear Regression Estimation (continued)

We develop a little further Example 2. Here, the covariance operator Σ is proportional to K (as shown in section 2.2.1): $\delta\Sigma = \frac{\tau^2}{n}K$. The operator $K : L^2([0,1]) \to L^2([0,1])$, defined as $\forall \varphi \in L^2([0,1])$, $K\varphi := \frac{1}{n}\sum_i < Z_i, \varphi > Z_i(t)$, is self-adjoint, *i.e.* $K = K^*$, and depends on n. It converges to the operator $\tilde{K}$, defined as $\forall \varphi \in L^2([0,1])$, $\tilde{K}\varphi = \int_0^1 \varphi(s)Cov(Z(s), Z(t))ds$, which is trace-class since $\mathbf{E}||Z_i||^2 < \infty$. Choose $L = \tilde{K}^{-1}$, and suppose that the spectrum of $\tilde{K}$ declines at the rate j^{-a_0}, $a_0 \geq 0$, as in Assumption C (for instance, $\tilde{K}$ could be the covariance operator of a Brownian motion). Then, $j^{-a_0} = j^{-c_0}$, $a_0 > 1$ and $a_0 = c_0$. Moreover, we set $\Omega_0 = K^{2\tilde{s}}$, for some $\tilde{s} > \frac{1}{2a_0}$ (so that $\lim_{n\to\infty} \Omega_0 = \tilde{K}^{2\tilde{s}}$), and assume that x_* satisfies Assumption B'. The posterior mean estimator takes the form: $\hat{x}_\alpha = K^{\tilde{s}}(\alpha I + K^{2\tilde{s}+1})^{-1}K^{\tilde{s}}(y^\delta - Kx_0) + x_0$, for which the following lemma holds.

Lemma 1. *Let* $\tilde{K} : L^2([0,1]) \to L^2([0,1])$ *have eigenvalues* $\{\lambda_{jK}\}$ *that satisfy* $\underline{a}j^{-a_0} \leq \lambda_{jK} \leq \overline{a}j^{-a_0}$*, for* $\underline{a}, \overline{a} > 0$ *and* $a_0 \geq 0$*. Assume that* $\mathbf{E}||Z_i||^4 < \infty$*. Then, under Assumption B' with* $b_0 > \max\{a_0, a_0\tilde{s}\}$, $\tilde{s} > \frac{1}{2a_0}$ *and if* $\alpha \asymp \alpha_* = n^{-\frac{a_0(2\tilde{s}+1)}{2b_0+a_0}}$*, we have*

$$\mathbf{E}_{x_*}||\hat{x}_{\alpha_*} - x_*||^2 = \mathcal{O}\Big(n^{-\frac{2b_0-1}{2b_0+a_0}}\Big).$$

The rate and the assumptions in the Lemma are the same as in Hall and Horowitz (2007).

5 An adaptive selection of α through an empirical Bayes approach

As shown by theorem 1 *(iii)* and Remark 3.1, the parameter α of the prior plays the role of a *regularization parameter* and $\{\hat{x}_\alpha\}_{\alpha\geq 0}$ defines a class of estimators for x, which are equal to a Tikhonov-type frequentist estimator. We have shown that for α decreasing at a convenient rate, this estimator converges at the minimax rate. However, this rate and the corresponding value for α are unknown in practice since they depend on the regularity of x_* which is unknown. Thus, it is very important to have an adaptive data-driven method for selecting α since a suitable value for α is crucial for the implementation of the estimation procedure. A data-driven method selects a suitable value $\hat{\alpha}$ for α if the posterior distribution of x computed by using this $\hat{\alpha}$ still satisfies consistency in a frequentist sense at an (almost) minimax rate.

We propose in sections 5.1 and 5.2 a data-driven method based on an *Empirical Bayes* procedure for selecting $\hat{\alpha}$. This procedure can be easily implemented for general operators K, Σ and L satisfying Assumptions A.1 and A.3.

5.1 Characterization of the likelihood

The marginal distribution of $y^\delta|\alpha, s$ is

$$y^\delta|\alpha, s \sim P_\alpha, \quad P_\alpha = \mathcal{N}\bigg(Kx_0, \delta\Sigma + \frac{\delta}{\alpha}K\Omega_0 K^*\bigg). \tag{15}$$

The marginal distribution P_α is obtained by marginalizing P^x with respect to the prior distribution of x. This requires the implicit assumption that, conditionally on x, y^δ is independent of α since α is considered here as a random variable (hyperparameter). The following theorem, which is an application of Theorem 3.3 in Kuo (1975), characterizes a probability measure P_0 which is equivalent to P_α for every $\alpha > 0$ and characterizes the likelihood of P_α with respect to P_0.

Theorem 5. *Let P_0 be a Gaussian measure with mean Kx_0 and covariance operator $\delta\Sigma$, that is, $P_0 = \mathcal{N}(Kx_0, \delta\Sigma)$. Under Assumptions A.1 and A.3, the Gaussian measure P_α defined in (15) is equivalent to P_0. Moreover, the Radon-Nikodym derivative is given by*

$$\frac{dP_\alpha}{dP_0}(\{z_j\}) = \prod_{j=1}^{\infty}\sqrt{\frac{\alpha}{\lambda_j^2+\alpha}e^{\frac{\lambda_j^2}{2(\lambda_j^2+\alpha)}z_j^2}}, \tag{16}$$

where $\{\varphi_j, \lambda_j^2\}$ are the eigenfunctions and eigenvalues of BB^*, respectively.

In our setting: $z_j = \frac{<y^\delta - Kx_0, \Sigma^{-1/2}\varphi_j>}{\sqrt{\delta}}$ and $\Sigma^{-1/2}\varphi_j$ is defined under assumption A.3.

5.2 Adaptive Empirical Bayes (EB) procedure

Let ν denote a prior distribution for α such that $\frac{d\log\nu(\alpha)}{\alpha} = \nu_1\alpha^{-1} + \nu_2$ for two constants $\nu_1 > 0$, $\nu_2 < 0$. An EB procedure consists in plugging in the prior distribution of $x|\alpha, s$ a value for α selected from the data y^δ. We define the *marginal maximum a posteriori estimator* $\hat{\alpha}$ of α to be the maximizer of the marginal log-posterior of $\alpha|y^\delta$ which is proportional to $\log\left[\frac{dP_\alpha}{dP_0}\nu(\alpha)\right]$:

$$\begin{aligned}
\hat{\alpha} &:= arg\max_\alpha \bar{S}(\alpha, y^\delta) && (17)\\
\bar{S}(\alpha, y^\delta) &:= \log\left(\frac{dP_\alpha}{dP_0}(\{z_j\})\nu(\alpha)\right) && (18)\\
&= \frac{1}{2}\sum_{j=1}^{\infty}\Big[\log\Big(\frac{\alpha}{\alpha+\lambda_j^2}\Big) + \frac{\lambda_j^2}{\alpha+\lambda_j^2}\frac{< y^\delta - Kx_0, \Sigma^{-1/2}\varphi_j >^2}{\delta}\Big] + \log\nu(\alpha).
\end{aligned}$$

In an equivalent way, $\hat{\alpha}$ is defined as the solution of the first order condition $\frac{\partial}{\partial\alpha}\bar{S}(\alpha, y^\delta) = 0$. We denote $S_{y^\delta}(\alpha) := \frac{\partial}{\partial\alpha}\bar{S}(\alpha, y^\delta)$ where

$$S_{y^\delta}(\alpha) := \frac{1}{2}\Big(\sum_{j=1}^{\infty}\frac{\lambda_j^2}{\alpha(\alpha+\lambda_j^2)} - \sum_{j=1}^{\infty}\frac{\lambda_j^2 < y^\delta - Kx_0, \Sigma^{-1/2}\varphi_j >^2}{\delta(\alpha+\lambda_j^2)^2}\Big) + \frac{d\log\nu(\alpha)}{d\alpha} \quad \text{and} \quad S_{y^\delta}(\hat{\alpha}) = 0. \tag{19}$$

Our strategy will then be to plug the value $\hat{\alpha}$, found in this way, back into the prior of $x|\alpha, s$ and then compute the posterior distribution $\mu^Y_{\delta,\hat{\alpha}}$ and the posterior mean estimator $\hat{x}_{\hat{\alpha}}$ using this value of α. We refer to $\hat{x}_{\hat{\alpha}}$ as the EB-estimator and to $\mu^Y_{\delta,\hat{\alpha}}$ as the EB-posterior. Examples of suitable priors for α are: either a gamma distribution or a beta distribution on $(0,1)$.

5.3 Posterior consistency for the EB-posterior (mildly ill-posed case)

In this section we study existence of $\hat{\alpha}$ and the rate at which $\hat{\alpha}$ decreases to 0 and show posterior consistency of $\mu^Y_{\delta,\hat{\alpha}}$.[4] When the true x_* is not too smooth, with respect to the degree of ill-posedness and the smoothness of the prior, then $\hat{\alpha}$ is of the same order as α_* ($\equiv \delta^{\frac{a+s}{b_0+1}}$) with probability that approaches 1 as $\delta \to 0$. Moreover, we state that the EB-posterior distribution $\mu^Y_{\delta,\hat{\alpha}}$ concentrates around the true x_* in probability.

Theorem 6. *Let B' and C hold with $a = a_0 - \frac{c_0}{2} > 0$, $s > \frac{1}{2}$ and $a_0 \geq c_0 - s$. Let ν be a prior distribution for α such that $\frac{d\log\nu(\alpha)}{\alpha} = \nu_1\alpha^{-1} + \nu_2$ for two constants $\nu_1 > 0$, $\nu_2 < 0$. Then, with probability approaching 1, there exists $\hat{\alpha}$ such that $S_{y^\delta}(\hat{\alpha}) = 0$ and $\hat{\alpha} \asymp \delta^{\frac{a+s}{b_0+a+\tilde{\eta}}}$ for $\tilde{\eta} = \eta I(b_0 - a - 2s - 1/2 > 0)$ and any $(b_0 + a) > \eta > \max\{b_0 - s - 1/4, 0\}$. If in addition $\eta > 1/2$, then for any sequence $M_\delta \to \infty$ the EB-posterior distribution satisfies*

$$\mu^Y_{\delta,\hat{\alpha}}\{x \in \mathcal{X} : ||x - x_*|| > \varepsilon_\delta M_\delta\} \to 0 \tag{20}$$

in P^{x_*}-probability as $\delta \to 0$ where $\varepsilon_\delta = \delta^{\frac{\tilde{b}_0 - \frac{1}{2}}{2(b_0+a+\tilde{\eta})}}$ and $\tilde{b}_0 = \min(b_0, 2a + 2s + 1/2)$.

The consistency of the EB-estimator $\hat{x}_{\hat{\alpha}}$ follows from posterior consistency of $\mu^Y_{\delta,\hat{\alpha}}$. The theorem says that the posterior contraction rate of the EB-posterior distribution is minimax optimal when $b_0 \leq a + 2s + 1/2$. In all the other cases the rate is slower. To get a minimax contraction rate also for the case where $b_0 > a + 2s + 1/2$ we should specify the prior ν for α depending on b_0 and a in some convenient way. However, this prior would be unfeasible in practice since b_0 is never known. For this reason we do not pursue this analysis since it would have an interest only from a theoretical point of view while the main motivation for this section is the practical implementation

[4]For simplicity of exposition we limit this analysis to the case where K, Σ and L have geometric spectra (*mildly ill-posed case*, see section 4). It is possible to extend the result of theorem 6 to the general case at the price of complicate much more the proof and the notation. For this reason we do not show the general result here.

of our estimator.

Remark 5.1. While this theorem is stated and proved for a Normal error term U, this result holds also for the more general case where U is only asymptotically Gaussian. In appendix A we give some hints about how the proof should be modified in this case. Therefore, our EB-approach works also in the case where U is only approximately Gaussian.

6 Numerical Implementation

6.1 Instrumental variable regression estimation

This section shows the implementation of our proposed estimation method for the IV regression example 1 and its finite sample properties. We simulate $n = 1000$ observations from the following model, which involves only one endogenous covariate Z and two instrumental variables $W = (W_1, W_2)$,

$$W_i = \begin{pmatrix} W_{1,i} \\ W_{2,i} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix} \right)$$

$$\begin{aligned} v_i &\sim \mathcal{N}(0, \sigma_v^2), & Z_i &= 0.1 w_{i,1} + 0.1 w_{i,2} + v_i \\ \varepsilon_i &\sim \mathcal{N}(0, (0.4)^2), & u_i &= -0.5 v_i + \varepsilon_i \\ y_i &= x_*(Z_i) + u_i \end{aligned}$$

for $i = 1, \ldots, n$. Endogeneity is caused by correlation between u_i and the error term v_i affecting the covariates. The true x_* is the parabola $x_*(Z) = Z^2$. In all the simulations we have fixed $\sigma_v = 0.27$. We do not transform the data to the interval $[0,1]$ and the spaces of reference are $\mathcal{X} = \mathcal{Y} = L^2(Z)$, where $L^2(Z)$ denotes the space of square integrable functions of Z with respect to its marginal distribution. Moreover, the function $a(w, v)$ is chosen such that $K : L^2(Z) \to L^2(Z)$ is the (estimated) double conditional expectation operator, that is, $\forall \varphi \in L^2(Z)$, $K\varphi = \hat{\mathbf{E}}(\hat{\mathbf{E}}(\varphi|W)|Z)$ and the functional observation y^δ takes the form $y^\delta(Z) := \hat{\mathbf{E}}(\hat{\mathbf{E}}(Y|W)|Z)$. Here, $\hat{\mathbf{E}}$ denotes the estimated expectation that we compute through a Gaussian kernel smoothing estimator of the joint density of (Y, Z, W). The bandwidth for Z has been set equal to $\sqrt{\widehat{Var}(Z)}n^{-1/5}$ and in a similar way the bandwidths for Y and W.

The sampling covariance operator is $\Sigma\varphi = \sigma^2 \mathbf{E}(\mathbf{E}(\varphi|W)|Z)$, $\forall \varphi \in L^2(Z)$, under the assumption $\mathbf{E}[(Y - x_*(Z))^2|W] = \sigma^2$, and it has been replaced by its empirical counterpart. Following the discussion in section 2.2.1 we specify the prior covariance operator as $\Omega_0 = \omega_0 \sigma^2 K^s$ for $s \geq 1$ and ω_0 a fixed parameter. In this way the conditional variance σ^2 in Σ and Ω_0 simplifies and does not need to be estimated.

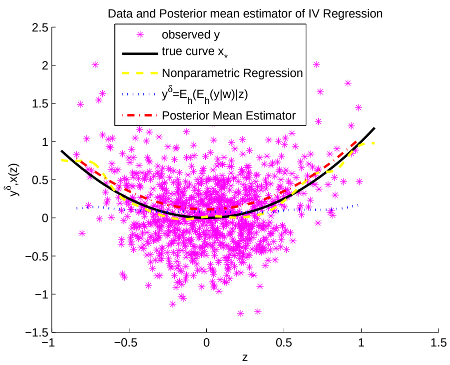
We have performed simulations for two specifications of x_0, ω_0 and s (where x_0 denotes the prior mean function): Figure 1 refers to $x_0(Z) = 0.95Z^2 + 0.25$, $\omega_0 = 1$ and $s = 2$ while Figure 2 refers to $x_0(Z) = 0$, $\omega_0 = 2$ and $s = 15$. We have first performed simulations for a fixed value of α (we have fixed $\alpha = 0.9$ to obtain Figures 1a-1b and Figures 2a-2b) and in a second simulation we have used the α selected through our EB method.

Graphs 1a and 2a represent: the n observed Y's (magenta asterisks), the corresponding y^δ (dotted blue line) obtained from the observed sample of (Y, Z, W), the true x_* (black solid line), the nonparametric estimation of the regression function $\mathbf{E}(Y|Z)$ (yellow dashed line), and our posterior mean estimator $\hat{x}_\alpha$ (dotted-dashed red line). We show the estimator of $\mathbf{E}(Y|Z)$ with the purpose of making clear the bias due to endogeneity. Graphs 1b and 2b represent: the prior mean function x_0 (magenta dashed line), the observed function y^δ (dotted blue line) obtained from the observed sample of (Y, Z, W), the true x_* (black solid line), and our posterior mean estimator $\hat{x}_\alpha$ (dotted-dashed red line).

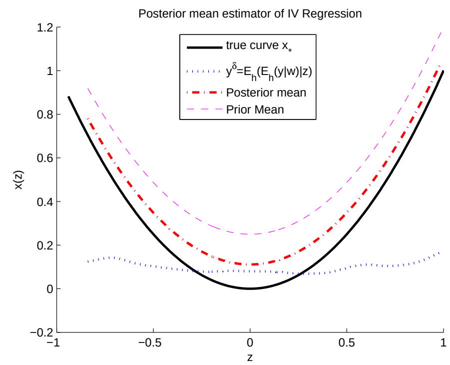
Graphs 1c and 2c draw the log-posterior $\log\left[\frac{dP_\alpha}{dP_0}\nu(\alpha)\right]$ against α and show the value of the maximum a posteriori $\hat{\alpha}$. We have specified an exponential prior for α: $\nu(\alpha) = 11e^{-11\alpha}$, $\forall\alpha \geq 0$. Finally, graphs 1d and 2d represent our EB-posterior mean estimator $\hat{x}_{\hat{\alpha}}$ (dotted-dashed red line) – obtained by using the $\hat{\alpha}$ selected with the EB-procedure – together with the prior mean function x_0 (magenta dashed line), the observed function y^δ (dotted blue line) and the true x_* (black solid line).

6.2 Geometric Spectrum case

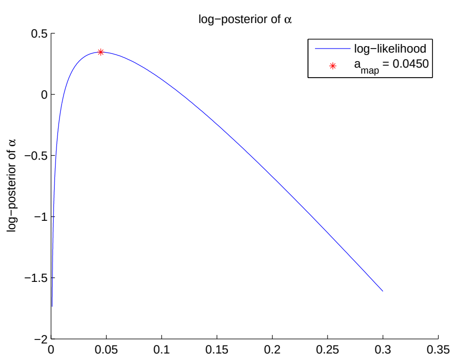
In this simulation we assume $\mathcal{X} = \mathcal{Y} = L^2(\mathbb{R})$ with respect to the measure $e^{-u^2/2}$ so that the operator K is self-adjoint. We use the Hermite polynomials as common eigenbasis for the operators K, Σ and L. The Hermite polynomials $\{H_j\}_{j\geq 0}$ form an orthogonal basis of $L^2(\mathbb{R})$ with respect to the measure $e^{-u^2/2}$. The first few Hermite polynomials are $\{1, u, (u^2-1), (u^3-3u), \ldots\}$ and an important property of these polynomials is that they are orthogonal with respect to $e^{-u^2/2}$: $\int_{\mathbb{R}} H_l(u)H_j(u)e^{-u^2/2}du = \sqrt{n}n!\delta_{lj}$, where δ_{lj} is equal to 1 if $l = j$ and 0 otherwise. Moreover, they satisfy the recursion $H_{j+1}(u) = uH_j(u) - jH_{j-1}(u)$ which is used in our simulation. We fix:



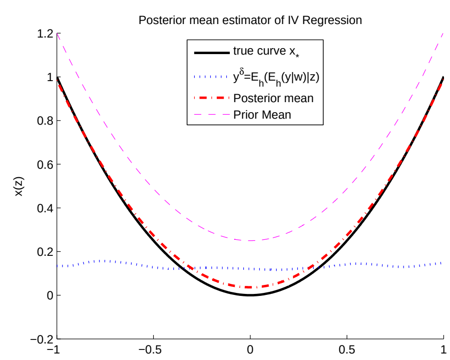
(a) Data and posterior mean estimator for $\alpha = 0.9$



(b) Posterior mean estimator for $\alpha = 0.9$

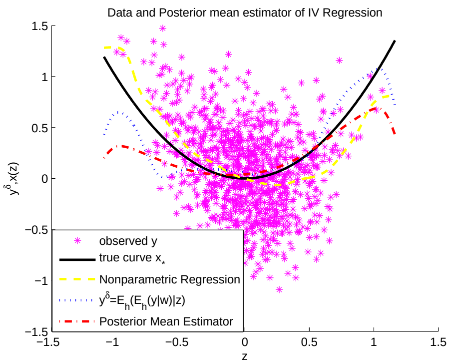


(c) α choice

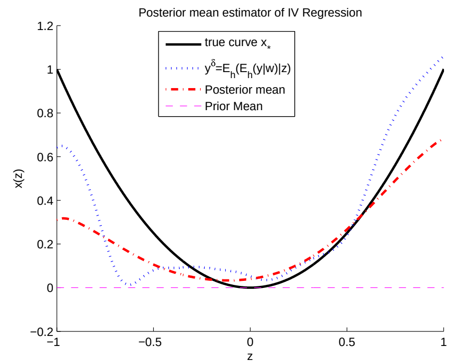


(d) Posterior mean estimator for $\alpha = 0.0450$

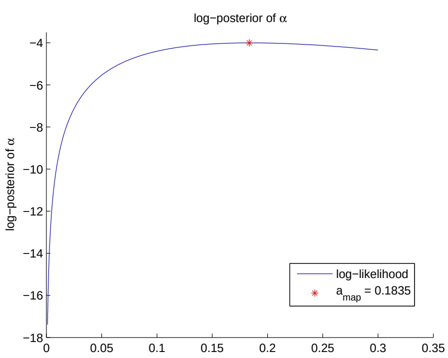
Figure 1: *Posterior mean estimator for smooth* x_*. Graph for: $x_0(Z) = 0.95Z^2 + 0.25$, $\omega_0 = 1$, $s = 2$.



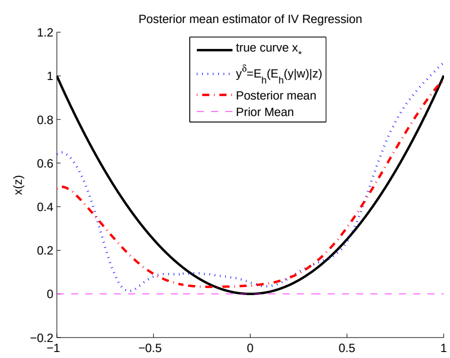
(a) Data and posterior mean estimator for $\alpha = 0.9$



(b) Posterior mean estimator for $\alpha = 0.9$



(c) α choice



(d) Posterior mean estimator for $\alpha = 0.1835$

Figure 2: *Posterior mean estimator for smooth* x_*. Graph for: $x_0(Z) = 0$, $\omega_0 = 2$, $s = 15$.

$\delta = 1/n$, $a_0 = 1$, $c_0 = 1.2$ and $s = 1$, thus the simulation design is:

$$\begin{aligned} \Sigma\cdot &= \tau \sum_{j=0}^{\infty} \frac{j^{-c_0}}{\sqrt{2\pi}(n!)} < H_j, \cdot > H_j \\ K\cdot &= \sum_{j=0}^{\infty} \frac{j^{-a_0}}{\sqrt{2\pi}(n!)} < H_j, \cdot > H_j \\ \Omega\cdot &= \omega_0 \sum_{j=0}^{\infty} \frac{j^{-2s}}{\sqrt{2\pi}(n!)} < H_j, \cdot > H_j \\ y^{\delta} &= Kx_* + U \end{aligned}$$

with $x_*(u) = u^2$ and $U = \frac{1}{\sqrt{n}}\mathcal{N}(0, \Sigma)$. Moreover, we fix: $\tau = 10$ and $\omega_0 = 5$. The inner product is approximated by discretizing the integral $\int_{\mathbb{R}} H_j(u) \cdot e^{-u^2/2} du$ with 1000 discretization points uniformly generated between -3 and 3. The infinite sums are truncated at $j = 200$.

We have first performed simulations for a fixed value of α (we have fixed $\alpha = 0.9$ to obtain Figure 3a) and in a second simulation we have used the α selected through our EB method.

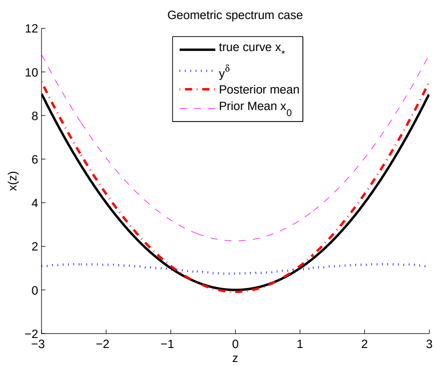
Graph 3a represents: the prior mean function x_0 (magenta dashed line), the observed function y^{δ} (dotted blue line), the true x_* (black solid line), and our posterior mean estimator $\hat{x}_{\alpha}$ (dotted-dashed red line).

Graph 3b draws the log-posterior log $\left[\frac{dP_{\alpha}}{dP_0}\nu(\alpha)\right]$ against α and shows the value of the maximum a posteriori $\hat{\alpha}$. We have specified a Gamma prior for α: $\nu(\alpha) \propto \alpha^{11}e^{-10\alpha}$, $\forall \alpha \geq 0$. Finally, graph 3c represents our EB-posterior mean estimator $\hat{x}_{\hat{\alpha}}$ (dotted-dashed red line) – obtained by using the $\hat{\alpha}$ selected with the EB-procedure – together with the prior mean function x_0 (magenta dashed line), the observed function y^{δ} (dotted blue line) and the true x_* (black solid line).

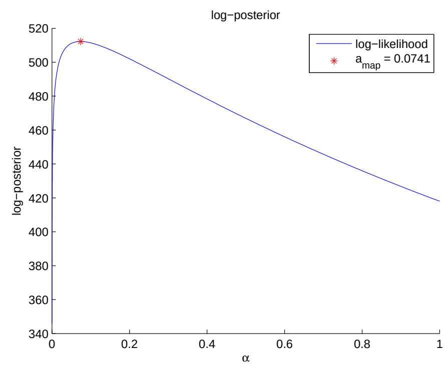
7 Conclusion

This paper develops a Bayesian approach for nonparametric estimation of parameters in econometric models that are characterized as the solution of an inverse problem. We consider a conjugate-Gaussian setting where the "likelihood" is only required to be asymptotically Gaussian. For "likelihood" we mean the sampling distribution of a functional transformation of the sample.

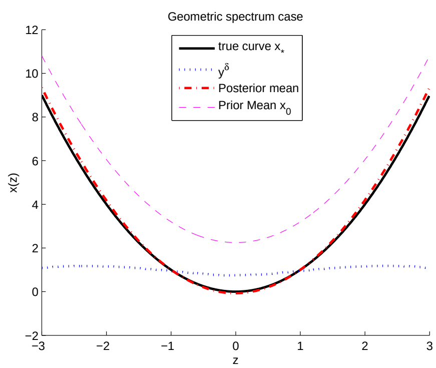
We first provide a point estimator – the posterior mean – that: (i) has a closed-form, (ii) is easy to implement in practice, (iii) has a pure Bayesian interpretation and (iv) is consistent in a frequentist sense. Second, we provide an adaptive data-driven method to select the regularization parameter. This method, while constructed for a Bayesian model, can be used for selecting the regularization parameter for frequentist estimators. This is due to the fact that, under mild conditions, the posterior mean estimator in a conjugate-Gaussian setting is shown to be the same as a Tikhonov-type estimator.



(a) Data and posterior mean estimator for $\alpha = 0.9$



(b) α choice



(c) Posterior mean estimator for $\alpha = 0.0741$

Figure 3: *Posterior mean estimator for smooth* x_*. Graph for: $x_0(Z) = 0.95Z^2 + 0.25$, $\omega_0 = 5$, $s = 1$ and $\tau = 10$.

Appendix

In all the proofs we use the notation $(\lambda_j, \varphi_j, \psi_j)_j$ to denote the singular value decomposition of B (or equivalently of B^*), that is, $B\psi_j = \lambda_j \varphi_j$ and $B^*\varphi_j = \lambda_j \psi_j$ where φ_j and ψ_j are of norm 1, $\forall j$. We also use the notation $I(A)$ to denote the indicator function of an event A and the notation $=^d$ to mean "equal in distribution". In order to prove several results we make use of Corollary 8.22 in Engl *et al.* (2000). We give here a simplified version of it adapted to our framework and we refer to Engl *et al.* (2000) for the proof of it.

Corollary 2. *Let $(\mathcal{X}_t)$, $t \in \mathbb{R}$ be a Hilbert scale generated by L and let $\Sigma^{-1/2}K : \mathcal{X} \to \mathcal{Y}$ be a bounded operator satisfying Assumption A.2, $\forall x \in \mathcal{X}$ and for some $a > 0$. Then, for $B = \Sigma^{-1/2}KL^{-s}$, $s \geq 0$ and $|\nu| \leq 1$*

$$\underline{c}(\nu)||L^{-\nu(a+s)}x|| \leq ||(B^*B)^{\frac{\nu}{2}}x|| \leq \bar{c}(\nu)||L^{-\nu(a+s)}x|| \tag{21}$$

holds on $\mathcal{D}((B^*B)^{\frac{\nu}{2}})$ with $\underline{c}(\nu) = \min(\underline{m}^\nu, \overline{m}^\nu)$ and $\bar{c}(\nu) = \max(\underline{m}^\nu, \overline{m}^\nu)$. Moreover, $\mathcal{R}((B^*B)^{\frac{\nu}{2}}) = \mathcal{X}_{\nu(a+s)} \equiv \mathcal{D}(L^{\nu(a+s)})$, where $(B^*B)^{\frac{\nu}{2}}$ has to be replaced by its extension to $\mathcal{X}$ if $\nu < 0$.

A Preliminary Lemmas

Lemma 2. *Let $\mathcal{X} = \mathcal{Y} = L^2([0,1])$ and K and $\tilde{K}$ be operators from $\mathcal{X}$ to $\mathcal{Y}$ such that $K\varphi = \frac{1}{n}\sum_{i=1}^n < Z_i, \varphi > Z_i(t)$ and $\tilde{K}\varphi = \int_0^1 \varphi(s)Cov(Z(s), Z(t))ds$, $\forall \varphi \in \mathcal{X}$, where $Z \in L^2([0,1])$ is a centered random function. Then, if $\mathbf{E}||Z||^4 < \infty$ we have*

$$\mathbf{E}_{x_*}||K - \tilde{K}||^2 = \mathcal{O}\Big(\frac{1}{n}\Big).$$

A.1 Proof of Lemma 2

Let $||\cdot||_{HS}$ denote the Hilbert-Schmidt norm. Since $Cov(Z(s), Z(t)) = \mathbf{E}(Z(s)Z(t))$ we have

$$||K - \tilde{K}||^2 \leq ||K - \tilde{K}||^2_{HS} := \int_0^1 \int_0^1 \Big[\frac{1}{n}\sum_{i=1}^n Z_i(s)Z_i(t) - \mathbf{E}(Z(s)Z(t))\Big]^2 dsdt$$

so that

$$\begin{aligned}
\mathbf{E}_{x_*}||K - \tilde{K}||^2 &\leq \int_0^1 \int_0^1 \mathbf{E}_{x_*}\Big[\frac{1}{n}\sum_{i=1}^n Z_i(s)Z_i(t) - \mathbf{E}(Z(s)Z(t))\Big]^2 dsdt \\
&= \int_0^1 \int_0^1 Var\Big[\frac{1}{n}\sum_{i=1}^n Z_i(s)Z_i(t)\Big]dsdt \\
&= \int_0^1 \int_0^1 \Big[\frac{1}{n^2}\sum_{i=1}^n Var(Z_i(s)Z_i(t)) + \frac{2n(n-1)}{n^2}\sum_{i>j} Cov(Z_i(s)Z_i(t), Z_j(s)Z_j(t))\Big]dsdt \\
&= \int_0^1 \int_0^1 \frac{1}{n}\mathbf{E}(Z^2(s)Z^2(t))dsdt = \frac{1}{n}\mathbf{E}\int_0^1 \int_0^1 (Z^2(s)Z^2(t))dsdt = \frac{1}{n}\mathbf{E}||Z||^4
\end{aligned}$$

since $Cov(Z_i(s)Z_i(t), Z_j(s)Z_j(t)) = 0$.

Lemma 3. *Let the assumptions of Theorem 6 be satisfied. Then,*

$$S_{y^\delta}(\alpha) \quad = \quad \mathcal{S}_1 - \mathcal{S}_2 - \mathcal{S}_3 - (\mathcal{S}_{4a} + \mathcal{S}_{4b}) + \nu_1 \frac{1}{\alpha} + \nu_2$$

where

$$\begin{aligned} \frac{1}{2\delta}\alpha^{\frac{b_0 - s - \frac{1}{2}}{a+s}} c_4 &\leq \mathcal{S}_2 \leq \frac{1}{2}\delta^{-1} I(b_0 \geq s) + \frac{1}{2\delta}\alpha^{\frac{b_0 - s - \frac{1}{2}}{a+s}} \tilde{c}_4, \\ \frac{1}{2}\alpha^{-\frac{2(a+s)+1}{2(a+s)}} \tilde{c}_3 &\leq (\mathcal{S}_1 - \mathcal{S}_{4b}) \leq \frac{1}{2}\alpha^{-\frac{2(a+s)+1}{2(a+s)}} \tilde{c}_3 + \frac{1}{2\alpha}, \end{aligned}$$

$$\mathcal{S}_3 \sim \xi \frac{1}{\sqrt{\delta}} \left(\sum_{j=1}^{\infty} \frac{\lambda_j^4}{(\alpha + \lambda_j^2)^4} j^{-2(a+b_0)} \right)^{\frac{1}{2}}, \qquad \mathcal{S}_{4a} \sim \xi \frac{1}{\sqrt{2}} \left(\sum_{j=1}^{\infty} \frac{\lambda_j^4}{(\alpha + \lambda_j^2)^4} \right)^{\frac{1}{2}}$$

with $\xi \sim \mathcal{N}(0,1)$ and $\tilde{c}_3$, c_4, $\tilde{c}_4$ are positive constants.

A.2 Proof of Lemma 3

We develop $S_{y^\delta}(\alpha)$ by using the fact that under Assumption C there exist $\underline{\lambda}, \bar{\lambda} > 0$ such that $\underline{\lambda} j^{-(a+s)} \leq \lambda_j \leq \bar{\lambda} j^{-(a+s)}$ for $j = 1, 2, \ldots$. Then,

$$\begin{aligned} S_{y^\delta}(\alpha) \quad = \quad & \frac{1}{2}\sum_{j=1}^{\infty} \frac{\lambda_j^2}{\alpha(\alpha + \lambda_j^2)} - \frac{1}{2}\sum_{j=1}^{\infty} \frac{\lambda_j^2}{\delta(\alpha + \lambda_j^2)^2 \lambda_{j\Sigma}} < K(x_* - x_0), \varphi_j >^2 \\ & - \frac{1}{2}\sum_{j=1}^{\infty} \frac{2\lambda_j^2}{\delta(\alpha + \lambda_j^2)^2 \lambda_{j\Sigma}} < K(x_* - x_0), \varphi_j >< U, \varphi_j > \\ & - \frac{1}{2}\sum_{j=1}^{\infty} \frac{\lambda_j^2}{\delta(\alpha + \lambda_j^2)^2 \lambda_{j\Sigma}} < U, \varphi_j >^2 := \mathcal{S}_1 - \mathcal{S}_2 - \mathcal{S}_3 - \mathcal{S}_4. \end{aligned}$$

Let us start by computing $\mathcal{S}_2$:

$$\begin{aligned} \mathcal{S}_2 &= \frac{1}{2}\sum_{j=1}^{\infty} \frac{\lambda_j^2}{\delta(\alpha + \lambda_j^2)^2 \lambda_{j\Sigma}} < K(x_* - x_0), \varphi_j >^2 \\ &\asymp \frac{1}{2\delta}\sum_{j=1}^{\infty} \frac{j^{-2(a+s)-2a-2b_0}}{(\alpha + j^{-2(a+s)})^2} = \frac{1}{2\delta}\sum_{j=1}^{\infty} \frac{j^{-4a-2s-2b_0}}{(\alpha + j^{-2(a+s)})^2} \end{aligned}$$

The function $f(j) = \frac{j^{-4a-2s-2b_0}}{(\alpha + j^{-2(a+s)})^2}$ defined on $\mathbb{R}_+$ is always decreasing in j when $b_0 \geq s$. When $b_0 < s$, $f(j)$ is increasing for $j < \bar{j}$ and decreasing for $j > \bar{j}$ where $\bar{j} = \left(\frac{2a+s+b_0}{s-b_0}\alpha\right)^{-\frac{1}{2(a+s)}}$. Therefore, to upper and lower bound $\mathcal{S}_2$ we have to consider these two cases separately. First, if $b_0 < s$:

$$\begin{aligned}
\frac{1}{2\delta}\sum_{j=1}^{\infty}\frac{j^{-4a-2s-2b_0}}{(\alpha+j^{-2(a+s)})^2} &\geq \frac{1}{2\delta}\sum_{j=\bar{j}}^{\infty}\frac{j^{-4a-2s-2b_0}}{(\alpha+j^{-2(a+s)})^2} \geq \frac{1}{2\delta}\alpha^{\frac{b_0-s-\frac{1}{2}}{a+s}}\int_{\underline{u}}^{\infty}\frac{u^{2(s-b_0)}}{(u^{2(a+s)}+1)^2}du := \frac{1}{2\delta}\alpha^{\frac{b_0-s-\frac{1}{2}}{a+s}}c_4(\underline{u});\\
\frac{1}{2\delta}\sum_{j=1}^{\infty}\frac{j^{-4a-2s-2b_0}}{(\alpha+j^{-2(a+s)})^2} &\leq \frac{1}{2\delta}\left(\sum_{j=1}^{\bar{j}}\frac{j^{-4a-2s-2b_0}}{(\alpha+j^{-2(a+s)})^2}+\alpha^{\frac{b_0-s-\frac{1}{2}}{a+s}}\int_{\underline{u}}^{\infty}\frac{u^{2(s-b_0)}}{(u^{2(a+s)}+1)^2}du\right)\\
&\leq \frac{1}{2\delta}\left(\bar{j}\frac{\bar{j}^{-4a-2s-2b_0}}{(\alpha+\bar{j}^{-2(a+s)})^2}+\alpha^{\frac{b_0-s-\frac{1}{2}}{a+s}}\int_{\underline{u}}^{\infty}\frac{u^{2(s-b_0)}}{(u^{2(a+s)}+1)^2}du\right)\\
&= \frac{1}{2\delta}\left(\frac{\left(\alpha^{\frac{2a+s+b_0}{(s-b_0)}}\right)^{-\frac{(1-4a-2s-2b_0)}{2(a+s)}}}{(\alpha+\alpha^{\frac{2a+s+b_0}{s-b_0}})^2}+\alpha^{\frac{b_0-s-\frac{1}{2}}{a+s}}c_4(\underline{u})\right) = \frac{1}{2\delta}\alpha^{\frac{b_0-s-1/2}{(a+s)}}\tilde{c}_4(\underline{u}),
\end{aligned}$$

where $\underline{u} = \left(\frac{2a+s+b_0}{s-b_0}\right)^{-\frac{1}{2(a+s)}}$, $c_4(\underline{u}) = \int_{\underline{u}}^{\infty}\frac{u^{2(s-b_0)}}{(u^{2(a+s)}+1)^2}du$ and $\tilde{c}_4(\underline{u}) := (2a+s+b_0)^{-\frac{1-4a-2s-2b_0}{2(a+s)}}\frac{(s-b_0)^{\frac{1+2s-2b_0}{2(a+s)}}}{4(a+s)^2} + c_4(\underline{u})$. Second, if $b_0 \geq s$:

$$\begin{aligned}
\frac{1}{2\delta}\sum_{j=1}^{\infty}\frac{j^{-4a-2s-2b_0}}{(\alpha+j^{-2(a+s)})^2} &\geq \frac{1}{2\delta}\alpha^{\frac{b_0-s-\frac{1}{2}}{a+s}}\int_{\bar{u}}^{\infty}\frac{u^{2(s-b_0)}}{(u^{2(a+s)}+1)^2}du =: \frac{1}{2\delta}\alpha^{\frac{b_0-s-\frac{1}{2}}{a+s}}c_4(\bar{u});\\
\frac{1}{2\delta}\sum_{j=1}^{\infty}\frac{j^{-4a-2s-2b_0}}{(\alpha+j^{-2(a+s)})^2} &\leq \frac{1}{2\delta}\left(\left.\frac{j^{-4a-2s-2b_0}}{(\alpha+j^{-2(a+s)})^2}\right|_{j=1}+\alpha^{\frac{b_0-s-\frac{1}{2}}{a+s}}\int_{\bar{u}}^{\infty}\frac{u^{2(s-b_0)}}{(u^{2(a+s)}+1)^2}du+\right)\\
&\leq \frac{1}{2\delta}\alpha^{\frac{b_0-s-1/2}{a+s}}c_4(\bar{u})+\frac{1}{2\delta}.
\end{aligned}$$

where $\bar{u} = \alpha^{\frac{1}{2(a+s)}}$. By defining $c_4 = c_4(\underline{u})I(b_0 < s)+c_4(\bar{u})I(b_0 \geq s)$ and $\tilde{c}_4 = \tilde{c}_4(\underline{u})I(b_0 < s)+c_4(\bar{u})I(b_0 \geq s)$, the first line of inequalities of the lemma is proved.
We analyze now $\mathcal{S}_3$. Let $\{\xi_j\}$ denote a sequence of independent $\mathcal{N}(0,1)$ random variables; we rewrite $\mathcal{S}_3$ as

$$\begin{aligned}
\mathcal{S}_3 &= \sum_{j=1}^{\infty}\frac{\lambda_j^2}{\sqrt{\delta}(\alpha+\lambda_j^2)^2\sqrt{\lambda_{j\Sigma}}} < K(x_*-x_0),\varphi_j > \frac{< U,\varphi_j >}{\sqrt{\delta\lambda_{j\Sigma}}}\\
&= \frac{1}{\sqrt{\delta}}\sum_{j=1}^{\infty}\frac{\lambda_j^2}{(\alpha+\lambda_j^2)^2\sqrt{\lambda_{j\Sigma}}} < K(x_*-x_0),\varphi_j > \xi_j.
\end{aligned}$$

This series is convergent if, for a fixed α, $\mathbf{E}||\mathcal{S}_3||^2 < \infty$. This is always verified because:

$$\begin{aligned}
\mathbf{E}||\mathcal{S}_3||^2 &\asymp \frac{1}{\delta}\sum_{j=1}^{\infty}\frac{\lambda_j^4}{(\alpha+\lambda_j^2)^4}j^{-2(a+b_0)}\\
&\leq \frac{1}{\delta}\left(\sup_j\frac{\lambda_j^2}{(\alpha+\lambda_j^2)^2}\right)^2\sum_{j=1}^{\infty}j^{-2(a+b_0)} = \frac{1}{\delta\alpha}\sum_{j=1}^{\infty}j^{-2(a+b_0)}
\end{aligned}$$

and it is finite if and only if $b_0 > \frac{1}{2}-a$ which is verified by assumption. Therefore, $\mathcal{S}_3$ is equal in distribution to a Gaussian random variable with 0 mean and variance $\frac{1}{\delta}\sum_{j=1}^{\infty}\frac{\lambda_j^4}{(\alpha+\lambda_j^2)^4}j^{-2(a+b_0)}$. [5]

[5] In the case where U is only asymptotically $\mathcal{N}(0,\delta\Sigma)$ then the analysis of $\mathcal{S}_3$ requires to use the Lyapunov Central Limit theorem. In particular we have to check the Lyapunov condition: for some $\varrho > 0$

Consider now term $\mathcal{S}_4$. Since $\frac{<U,\varphi_j>^2}{\delta\lambda_{j\Sigma}} \sim$ *i.i.d.* χ_1^2, we center term $\mathcal{S}_4$ around its mean and apply the Lyapunov Central Limit Theorem to term $\mathcal{S}_{4a}$ below:

$$\mathcal{S}_4 = \frac{1}{2}\sum_{j=1}^{\infty}\frac{\lambda_j^2}{(\alpha+\lambda_j^2)^2}\Big(\frac{<U,\varphi_j>^2}{\delta\lambda_{j\Sigma}}-1\Big)+\frac{1}{2}\sum_{j=1}^{\infty}\frac{\lambda_j^2}{(\alpha+\lambda_j^2)^2} := \mathcal{S}_{4a}+\mathcal{S}_{4b}.$$

For that, we verify the Lyapunov condition. Denote $\vartheta := \left(\frac{<U,\varphi_j>^2}{\delta\lambda_{j\Sigma}}-1\right)$ and it is easy to show that $\mathbf{E}|\vartheta|^3 = 8.6916$, thus $\sum_{j=1}^{\infty}\mathbf{E}\left|\frac{\lambda_j^2\vartheta}{(\alpha+\lambda_j^2)^2}\right|^3 = 8.6916\sum_{j=1}^{\infty}\frac{j^{-6(a+s)}}{(\alpha+j^{-2(a+s)})^6}$ which is upper bounded by

$$\begin{aligned}\sum_{j=1}^{\infty}\mathbf{E}\left|\frac{\lambda_j^2\vartheta}{(\alpha+\lambda_j^2)^2}\right|^3 &\leq 8.6916\left[\sum_{j=1}^{\bar{j}}\frac{j^{-6(a+s)}}{(\alpha+j^{-2(a+s)})^6}+\sum_{j=\bar{j}}^{\infty}\frac{j^{-6(a+s)}}{(\alpha+j^{-2(a+s)})^6}\right]\\ &= 8.6916\left[\frac{\bar{j}^{-6(a+s)+1}}{(\alpha+\bar{j}^{-2(a+s)})^6}+\alpha^{-3-\frac{1}{2(a+s)}}\int_1^{-\infty}\frac{u^{6(a+s)}}{(u^{2(a+s)}+1)^6}du\right]\\ &= 8.6916\alpha^{-3-\frac{1}{2(a+s)}}\left[2^{-6}+\int_1^{-\infty}\frac{u^{6(a+s)}}{(u^{2(a+s)}+1)^6}du\right]\end{aligned}$$

since the function $f(j)=\frac{j^{-6(a+s)}}{(\alpha+j^{-2(a+s)})^6}$ is increasing in j for $0<j<\bar{j}:=\alpha^{-\frac{1}{2(a+s)}}$ and decreasing for $j>\bar{j}$. By using the lower bound of $Var(\mathcal{S}_{4a})$ given in Lemma 4 below we obtain:

$$\begin{aligned}(Var(\mathcal{S}_{4a}))^{-3/2}\sum_{j=1}^{\infty}\mathbf{E}\left|\frac{\lambda_j^2\vartheta}{(\alpha+\lambda_j^2)^2}\right|^3 &\leq 8.6916\left(\frac{1}{2}\alpha^{-\frac{4(a+s)+1}{2(a+s)}}c_6\right)^{-3/2}\alpha^{-3-\frac{1}{2(a+s)}}\left[2^{-6}+\int_1^{-\infty}\frac{u^{6(a+s)}}{(u^{2(a+s)}+1)^6}du\right]\\ &\asymp \alpha^{\frac{1}{4(a+s)}}\end{aligned}$$

which converges to 0 so that the Lyapunov condition is satisfied and $\mathcal{S}_{4a} =^d \frac{1}{2}\left(2\sum_{j=1}^{\infty}\frac{\lambda_j^4}{(\alpha+\lambda_j^2)^4}\right)^{\frac{1}{2}}\xi$ where $\xi\sim\mathcal{N}(0,1)$. [6]

Term $\mathcal{S}_{4b}$ is non random and we subtract it from $\mathcal{S}_1$ to obtain:

$$\begin{aligned}\mathcal{S}_1-\mathcal{S}_{4b} &= \frac{1}{2}\Big[\sum_{j=1}^{\infty}\frac{\lambda_j^2}{\alpha(\alpha+\lambda_j^2)}-\sum_{j=1}^{\infty}\frac{\lambda_j^2}{(\alpha+\lambda_j^2)^2}\Big]=\frac{1}{2\alpha}\sum_{j=1}^{\infty}\frac{\lambda_j^4}{(\alpha+\lambda_j^2)^2}\\ &\asymp \frac{1}{2\alpha}\sum_{j=1}^{\infty}\frac{j^{-4(a+s)}}{(\alpha+j^{-2(a+s)})^2}=\frac{1}{2\alpha}\sum_{j=1}^{\infty}\frac{1}{(\alpha j^{2(a+s)}+1)^2}.\end{aligned}$$

$$(Var(\mathcal{S}_3))^{-(2+\varrho)/2}\sum_{i=1}^{\infty}\left(\frac{\lambda_j^2}{\delta(\alpha+\lambda_j^2)^2\lambda_{j\Sigma}}<K(x_*-x_0),\varphi_j>\right)^{2+\varrho}\mathbf{E}\left|<U,\varphi_j>-\mathbf{E}(<U,\varphi_j>)\right|^{2+\varrho}=0.$$

If this condition is satisfied then $\mathcal{S}_3$ is equal in distribution to a Gaussian random variable with mean $\sum_{j=1}^{\infty}\frac{j^{-2(2a+s-b_0)}}{\delta(\alpha+j^{-2(a+s)})^2}\mathbf{E}<U,\varphi_j>$ and variance $\frac{1}{\delta}\sum_{j=1}^{\infty}\frac{\lambda_j^4}{(\alpha+\lambda_j^2)^4}j^{-2(a+b_0)+c_0}Var\left(\frac{<U,\varphi_j>}{\sqrt{\delta}}\right)$. Asymptotically, this mean is 0 and the variance is equal to the expression for $\mathbf{E}||\mathcal{S}_3||^2$ given above.

[6] In the case where U is only asymptotically $\mathcal{N}(0,\delta\Sigma)$ then a remark similar to the one for term $\mathcal{S}_3$ applies.

Since $\frac{j^{-4(a+s)}}{(\alpha+j^{-2(a+s)})^2}$ is a decreasing function of j we have

$$\frac{1}{2\alpha}\int_1^\infty \frac{1}{(\alpha t^{2(a+s)}+1)^2}dt \le \frac{1}{2\alpha}\sum_{j=1}^\infty \frac{1}{(\alpha j^{2(a+s)}+1)^2} \le \frac{1}{2\alpha}\int_1^\infty \frac{1}{(\alpha t^{2(a+s)}+1)^2}dt + \frac{1}{2\alpha(\alpha+1)^2}.$$

By denoting $c_3(\bar{u}) = \int_{\bar{u}}^\infty \frac{1}{(u^{2(a+s)}+1)^2}du$, with $\bar{u}$ defined above, and since $\frac{1}{(\alpha+1)^2} < 1$ for a fixed α, after some algebra we conclude that $\frac{1}{2}\alpha^{-\frac{2(a+s)+1}{2(a+s)}}c_3(\bar{u}) \le (\mathcal{S}_1 - \mathcal{S}_{4b}) \le \frac{1}{2}\alpha^{-\frac{2(a+s)+1}{2(a+s)}}c_3(\bar{u}) + \frac{1}{2\alpha}$. By defining $\tilde{c}_3 = c_3(\bar{u})$, the second line of inequalities of the lemma is proved.

Lemma 4. *Let the assumptions of Theorem 6 be satisfied and $\mathcal{S}_3$, $\mathcal{S}_{4a}$ be as defined in Lemma 3. Then, for α and δ fixed, we have*

$$\frac{1}{\delta}\alpha^{-\frac{a+2s-b_0+\frac{1}{2}}{(a+s)}}c_5 \le Var(\mathcal{S}_3) \le \frac{1}{\delta}I(b_0 \ge a+2s) + \frac{1}{\delta}\alpha^{-\frac{a+2s-b_0+\frac{1}{2}}{(a+s)}}\tilde{c}_5, \tag{22}$$

$$\frac{1}{2}\alpha^{-\frac{4(a+s)+1}{2(a+s)}}c_6 \le Var(\mathcal{S}_{4a}) \le \frac{17\tilde{c}_6}{32}\alpha^{-\frac{4(a+s)+1}{2(a+s)}} \tag{23}$$

where c_5, $\tilde{c}_5$, c_6 and $\tilde{c}_6$ are positive constants.

A.3 Proof of Lemma 4

We start by considering $Var(\mathcal{S}_3) = \frac{1}{\delta}\sum_{j=1}^\infty \frac{\lambda_j^4}{(\alpha+\lambda_j^2)^4}j^{-2(a+b_0)}$. Under assumption C, we can rewrite $Var(\mathcal{S}_3) \asymp \frac{1}{\delta}\sum_{j=1}^\infty \frac{j^{-4(a+s)-2a-2b_0}}{(\alpha+j^{-2(a+s)})^4}$. The function $f(j) = \frac{j^{-4(a+s)-2a-2b_0}}{(\alpha+j^{-2(a+s)})^4}$ defined on $\mathbb{R}_+$ is always decreasing in j when $b_0 \ge a+2s$. When $b_0 < a+2s$, $f(j)$ is increasing for $j < \bar{j}$ and decreasing for $j > \bar{j}$ where $\bar{j} = \left(\frac{3a+2s+b_0}{a+2s+b_0}\alpha\right)^{-\frac{1}{2(a+s)}}$. Therefore, in order to find an upper and a lower bound for $Var(\mathcal{S}_3)$ we consider these two cases separately. Let us start with the case $b_0 < a+2s$:

$$\begin{aligned}
Var(\mathcal{S}_3) &\ge \frac{1}{\delta}\sum_{j=\bar{j}}^\infty \frac{j^{-4(a+s)-2a-2b_0}}{(\alpha+j^{-2(a+s)})^4} \ge \frac{1}{\delta}\alpha^{-\frac{a+2s-b_0+\frac{1}{2}}{(a+s)}}\int_{\underline{u}}^\infty \frac{u^{2a+4s-2b_0}}{(u^{2(a+s)}+1)^4}du =: \frac{1}{\delta}\alpha^{-\frac{a+2s-b_0+\frac{1}{2}}{(a+s)}}c_5(\underline{u});\\
Var(\mathcal{S}_3) &\le \frac{1}{\delta}\sum_{j=1}^{\bar{j}} \frac{j^{-4(a+s)-2a-2b_0}}{(\alpha+j^{-2(a+s)})^4} + \frac{1}{\delta}\alpha^{-\frac{a+2s-b_0+\frac{1}{2}}{(a+s)}}c_5(\underline{u}) \le \frac{1}{\delta}\bar{j}\frac{\bar{j}^{-4(a+s)-2a-2b_0}}{(\alpha+\bar{j}^{-2(a+s)})^4} + \frac{1}{\delta}\alpha^{-\frac{a+2s-b_0+\frac{1}{2}}{(a+s)}}c_5(\underline{u})\\
&= \frac{1}{\delta}\alpha^{-\frac{(1+2a+4s-2b_0)}{2(a+s)}}\left(\underline{u}^{(1+2a+4s-2b_0)}\left(\frac{a+2s+b_0}{4(a+s)}\right)+c_5(\underline{u})\right) =: \frac{1}{\delta}\alpha^{-\frac{(1+2a+4s-2b_0)}{2(a+s)}}\tilde{c}_5(\underline{u})
\end{aligned}$$

where $\underline{u} = \left(\frac{3a+2s+b_0}{a+2s+b_0}\right)^{-\frac{1}{2(a+s)}}$. Next, we consider the case $b_0 \ge a+2s$:

$$\begin{aligned}
Var(\mathcal{S}_3) &\ge \frac{1}{\delta}\alpha^{-\frac{a+2s-b_0+\frac{1}{2}}{(a+s)}}\int_{\bar{u}}^\infty \frac{u^{2a+4s-2b_0}}{(u^{2(a+s)}+1)^4}du =: \frac{1}{\delta}\alpha^{-\frac{a+2s-b_0+\frac{1}{2}}{(a+s)}}c_5(\bar{u});\\
Var(\mathcal{S}_3) &\le \frac{1}{\delta}\left(\left.\frac{j^{-4(a+s)-2a-2b_0}}{(\alpha+j^{-2(a+s)})^4}\right|_{j=1} + \alpha^{-\frac{a+2s-b_0+\frac{1}{2}}{(a+s)}}\int_{\bar{u}}^\infty \frac{u^{2a+4s-2b_0}}{(u^{2(a+s)}+1)^4}du\right)\\
&\le \frac{1}{\delta} + \frac{1}{\delta}\alpha^{-\frac{(1+2a+4s-2b_0)}{2(a+s)}}c_5(\bar{u})
\end{aligned}$$

where $\bar{u} = \alpha^{\frac{1}{2(a+s)}}$. By defining $c_5 = c_5(\underline{u})I(b_0 < a+2s) + c_5(\bar{u})I(b_0 \geq a+2s)$ and $\tilde{c}_5 = \tilde{c}_5(\underline{u})I(b_0 < a+2s) + c_5(\bar{u})I(b_0 \geq a+2s)$, the first inequalities in (22) is proved.

Under assumption C, the variance of $\mathcal{S}_{4a}$ rewrites as $Var(\mathcal{S}_{4a}) \asymp \frac{1}{2}\sum_{j=1}^{\infty} \frac{j^{-4(a+s)}}{(\alpha+j^{-2(a+s)})^4}$. The function $f(j) = \frac{j^{-4(a+s)}}{(\alpha+j^{-2(a+s)})^4}$ defined on $\mathbb{R}_+$ is increasing in j for $j < \bar{j}$ and decreasing for $j > \bar{j}$ where $\bar{j} = \alpha^{-\frac{1}{2(a+s)}}$. We can lower and upper bound $Var(\mathcal{S}_{4a})$ as follows:

$$\begin{aligned}
Var(\mathcal{S}_{4a}) &\geq \frac{1}{2}\sum_{j=\bar{j}}^{\infty} \frac{j^{-4(a+s)}}{(\alpha+j^{-2(a+s)})^4} \geq \frac{1}{2}\alpha^{-2-\frac{1}{2(a+s)}}\int_1^{\infty} \frac{u^{4(a+s)}}{(u^{2(a+s)}+1)^4}du =: \frac{1}{2}\alpha^{-\frac{4(a+s)+1}{2(a+s)}}c_6,\\
Var(\mathcal{S}_{4a}) &\leq \frac{1}{2}\sum_{j=1}^{\bar{j}} \frac{j^{-4(a+s)}}{(\alpha+j^{-2(a+s)})^4} + \frac{1}{2}\alpha^{-\frac{4(a+s)+1}{2(a+s)}}c_6 \leq \frac{1}{2}\bar{j}\frac{\bar{j}^{-4(a+s)}}{(\alpha+\bar{j}^{-2(a+s)})^4} + \frac{1}{2}\alpha^{-\frac{4(a+s)+1}{2(a+s)}}c_6\\
&= \frac{1}{32}\alpha^{-\frac{1+4a+4s}{2(a+s)}} + \frac{c_6}{2}\alpha^{-\frac{1+4a+4s}{2(a+s)}} = \frac{17c_6+1}{32}\alpha^{-\frac{1+4(a+s)}{2(a+s)}} =: \frac{17\tilde{c}_6}{32}\alpha^{-\frac{1+4(a+s)}{2(a+s)}}.
\end{aligned}$$

B Proofs for Section 3

B.1 Proof of Theorem 1

(i) See the proof of Theorem 1 *(*i) and *(*ii) in Florens and Simoni (2012b).
(ii) See Theorem 2 and Corollary 2 in Mandelbaum (1984) and their proofs in Sections 3.4 and 3.5, page 392.
(iii) The P_α-mlt A is defined as $A := \Omega_0 K^*(\alpha\Sigma + K\Omega_0 K^*)^{-1}$ on $\mathcal{R}((\delta\Sigma + \frac{\delta}{\alpha}K\Omega_0 K^*)^{\frac{1}{2}})$, see Luschgy (1995). Under A.3, the operator $\Omega_0 K^*(\alpha\Sigma + K\Omega_0 K^*)^{-1}$ can equivalently be rewritten as

$$\begin{aligned}
&\Omega_0^{\frac{1}{2}}\Omega_0^{\frac{1}{2}}K^*\Sigma^{-1/2}(\alpha I + \Sigma^{-1/2}K\Omega_0 K^*\Sigma^{-1/2})^{-1}\Sigma^{-1/2} = \Omega_0^{\frac{1}{2}}B^*(\alpha I + BB^*)^{-1}\Sigma^{-1/2}\\
=\ &\Omega_0^{\frac{1}{2}}(\alpha I + B^*B)^{-1}(\Sigma^{-1/2}B)^* + \Omega_0^{\frac{1}{2}}[B^*(\alpha I + BB^*)^{-1} - (\alpha I + B^*B)^{-1}B^*]\Sigma^{-1/2}\\
=\ &\Omega_0^{\frac{1}{2}}(\alpha I + B^*B)^{-1}(\Sigma^{-1/2}B)^* \qquad (24)
\end{aligned}$$

since $[B^*(\alpha I + BB^*)^{-1} - (\alpha I + B^*B)^{-1}B^*]$ is equal to

$$(\alpha I + B^*B)^{-1}[(\alpha I + B^*B)B^* - B^*(\alpha I + BB^*)](\alpha I + BB^*)^{-1}$$

which is zero. By using expression (24) for $\Omega_0 K^*(\alpha\Sigma + K\Omega_0 K^*)^{-1}$ we show that $\Omega_0 K^*(\alpha\Sigma + K\Omega_0 K^*)^{-1}$ is bounded and continuous on $\mathcal{Y}$. By Assumption A.3 the operators B and $\Sigma^{-1/2}B$ exist and are bounded. Under A.1 the operator B is compact because it is the product of a compact and a bounded operator. This implies that $B^*B : \mathcal{X} \to \mathcal{X}$ is compact. Because $< (\alpha I + B^*B)\phi, \phi > \geq \alpha||\phi||^2$, $\forall \phi \in \mathcal{X}$ we conclude that $(\alpha I + B^*B)$ is injective for $\alpha > 0$. Then, from the Riesz Theorem 3.4 in Kress (1999) it follows that the inverse $(\alpha I + B^*B)^{-1}$ is bounded.

Finally, since the product of bounded linear operators is bounded, $\Omega_0 K^*(\alpha\Sigma + K\Omega_0 K^*)^{-1}$ is bounded, *i.e.* there exists a constant C such that for all $\varphi \in \mathcal{Y}$: $||\Omega_0 K^*(\alpha\Sigma + K\Omega_0 K^*)^{-1}\varphi|| \leq C||\varphi||$. Since $\Omega_0 K^*(\alpha\Sigma + K\Omega_0 K^*)^{-1}$ is linear and bounded on $\mathcal{Y}$, it is continuous on the whole $\mathcal{Y}$ and $A := \Omega_0 K^*(\alpha\Sigma + K\Omega_0 K^*)^{-1} = \Omega_0^{\frac{1}{2}}(\alpha I + B^*B)^{-1}(\Sigma^{-1/2}B)^*$ on $\mathcal{Y}$.

B.2 Proof of Theorem 2

The difference $(\hat{x}_\alpha - x_*)$ is re-written as

$$\hat{x}_\alpha - x_* = -(I - AK)(x_* - x_0) + AU := \mathcal{C}_1 + \mathcal{C}_2,$$

where the expression of A is given in (9). We consider the MISE associated with $\hat{x}_\alpha$: $\mathbf{E}_{x_*}||\hat{x}_\alpha - x_*||^2 = ||\mathcal{C}_1||^2 + \mathbf{E}_{x_*}||\mathcal{C}_2||^2$ and we start by considering the first term. Our proof follows Natterer (1984). Since $0 < \frac{s}{a+s} < 1$ we can use the left part of inequality (21) in Corollary (2) with $\nu = \frac{s}{a+s}$ and $x = (\alpha I + B^*B)^{-1}L^s(x_* - x_0)$ to obtain (25) below:

$$\begin{aligned}
||\mathcal{C}_1||^2 &= ||[I - \Omega_0^{\frac{1}{2}}(\alpha I + B^*B)^{-1}(\Sigma^{-1/2}B)^*K](x_* - x_0)||^2 \\
&= ||[I - L^{-s}(\alpha I + B^*B)^{-1}(\Sigma^{-1/2}B)^*K]L^{-s}L^s(x_* - x_0)||^2 \\
&= ||L^{-s}[I - (\alpha I + B^*B)^{-1}(\Sigma^{-1/2}B)^*KL^{-s}]L^s(x_* - x_0)||^2 \\
&\leq \alpha^2 \underline{c}^{-2}\Big(\frac{s}{a+s}\Big)||(B^*B)^{\frac{s}{2(a+s)}}(\alpha I + B^*B)^{-1}L^s(x_* - x_0)||^2 \qquad (25) \\
&= \alpha^2 \underline{c}^{-2}\Big(\frac{s}{a+s}\Big)||(B^*B)^{\frac{s}{2(a+s)}}(\alpha I + B^*B)^{-1}(B^*B)^{\frac{\tilde{\beta}-s}{2(a+s)}}(B^*B)^{\frac{s-\tilde{\beta}}{2(a+s)}}L^s(x_* - x_0)||^2 \\
&= \alpha^2 \underline{c}^{-2}\Big(\frac{s}{a+s}\Big)||(B^*B)^{\frac{\tilde{\beta}}{2(a+s)}}(\alpha I + B^*B)^{-1}w||^2 \qquad (26)
\end{aligned}$$

where $\tilde{\beta} = \min(\beta, a + 2s)$ and $w := (B^*B)^{\frac{s-\tilde{\beta}}{2(a+s)}}L^s(x_* - x_0)$. Now, by using the right part of inequality (21) in Corollary (2) with $\nu = \frac{s-\tilde{\beta}}{a+s}$ (remark that $|\nu| \leq 1$ is verified) and $x = L^s(x_* - x_0)$ we have that

$$\begin{aligned}
\sup_{(x_* - x_0)\in\mathcal{X}_\beta(\Gamma)} ||w|| &\leq \overline{c}\Big(\frac{s-\tilde{\beta}}{a+s}\Big) \sup_{(x_* - x_0)\in\mathcal{X}_\beta(\Gamma)} ||L^{\tilde{\beta}-s}L^s(x_* - x_0)|| \\
&= \overline{c}\Big(\frac{s-\tilde{\beta}}{a+s}\Big) \sup_{(x_* - x_0)\in\mathcal{X}_\beta(\Gamma)} ||(x_* - x_0)||_{\tilde{\beta}} \\
&\leq \overline{c}\Big(\frac{s-\tilde{\beta}}{a+s}\Big) \sup_{(x_* - x_0)\in\mathcal{X}_\beta(\Gamma)} ||(x_* - x_0)||_\beta ||L^{\tilde{\beta}-\beta}|| \\
&= \overline{c}\Big(\frac{s-\tilde{\beta}}{a+s}\Big)\Gamma^{\frac{1}{2}}||L^{\tilde{\beta}-\beta}|| \qquad (27)
\end{aligned}$$

where in the penultimate line the equality holds for $\tilde{\beta} = \beta$. Remark that $||L^{\tilde{\beta}-\beta}|| = 1$ if $\tilde{\beta} = \beta$ and is bounded if $\tilde{\beta} < \beta$. Finally,

$$\begin{aligned} \sup_{(x_*-x_0)\in\mathcal{X}_\beta(\Gamma)} ||(B^*B)^{\frac{\tilde{\beta}}{2(a+s)}}(\alpha I + B^*B)^{-1}w||^2 &= \sup_{(x_*-x_0)\in\mathcal{X}_\beta(\Gamma)} \sum_{j=1}^{\infty} \frac{\lambda_j^{\frac{2\tilde{\beta}}{a+s}}}{(\alpha+\lambda_j^2)^2} < w, \psi_j >^2 \\ &\leq \Big(\sup_{\lambda_j\geq 0} \frac{\lambda_j^{\frac{\tilde{\beta}}{a+s}}}{(\alpha+\lambda_j^2)}\Big)^2 \sup_{(x_*-x_0)\in\mathcal{X}_\beta(\Gamma)} ||w||^2 \qquad (28) \end{aligned}$$

and combining (26)-(28) with the fact that $(\sup_{\lambda\geq 0}\lambda^{2b}(\alpha+\lambda^2)^{-1})^2 = \alpha^{2(b-1)}b^{2b}(1-b)^{2(1-b)}$ for $0 \leq b \leq 1$ (and in our case $b = \frac{\tilde{\beta}}{2(a+s)}$), we get the result

$$\sup_{(x_*-x_0)\in\mathcal{X}_\beta(\Gamma)} ||\mathcal{C}_1||^2 \leq \alpha^{\frac{\tilde{\beta}}{a+s}} b^{2b}(1-b)^{2(1-b)}\underline{c}^{-2}\Big(\frac{s}{a+s}\Big)\overline{c}^2\Big(\frac{s-\tilde{\beta}}{a+s}\Big)||L^{\tilde{\beta}-\beta}||^2\Gamma = \mathcal{O}\left(\alpha^{\tilde{\beta}/(a+s)}\right).$$

Next, we address the second term of the MISE. To obtain (29) below, we use the left part of inequality (21) in Corollary (2) with $\nu = \frac{s}{a+s}$ and $x = (\alpha I + B^*B)^{-1}(\Sigma^{-1/2}B)^*U$:

$$\begin{aligned} \mathbf{E}_{x_*}||\mathcal{C}_2||^2 &= \mathbf{E}_{x_*}||AU||^2 = \mathbf{E}_{x_*}||\Omega_0^{\frac{1}{2}}(\alpha I + B^*B)^{-1}(\Sigma^{-1/2}B)^*U||^2 \\ &= \mathbf{E}_{x_*}||L^{-s}(\alpha I + B^*B)^{-1}(\Sigma^{-1/2}B)^*U||^2 \\ &\leq \underline{c}^{-2}\Big(\frac{s}{a+s}\Big)\mathbf{E}_{x_*}||(B^*B)^{\frac{s}{2(a+s)}}(\alpha I + B^*B)^{-1}(\Sigma^{-1/2}B)^*U||^2 \qquad (29) \\ &= \delta\underline{c}^{-2}\Big(\frac{s}{a+s}\Big)tr\Big((B^*B)^{\frac{s}{2(a+s)}}(\alpha I + B^*B)^{-1}B^*B(\alpha I + B^*B)^{-1}(B^*B)^{\frac{s}{2(a+s)}}\Big) \\ &\leq \delta\underline{c}^{-2}\Big(\frac{s}{a+s}\Big)\left(\sup_j \frac{\lambda_j^{\frac{s}{(a+s)}+1-\gamma}}{(\alpha+\lambda_j^2)}\right)^2 \sum_{j=1}^{\infty} < (B^*B)^{\gamma}\psi_j, \psi_j > \\ &\leq \delta\underline{c}^{-2}\Big(\frac{s}{a+s}\Big)\alpha^{2(d-1)}d^{2d}(1-d)^{2(1-d)}\overline{c}(\gamma)tr(L^{-2\gamma(a+s)}) \qquad (30) \end{aligned}$$

where $d = \frac{a+2s-\gamma(a+s)}{2(a+s)}$ and which is bounded since $tr(L^{-2\gamma(a+s)}) < \infty$ by definition of γ. The last inequality has been obtained by applying the right part of Corollary 2 with $\nu = \gamma$ to $\sum_{j=1}^{\infty} < (B^*B)^{\gamma}\psi_j, \psi_j >$. In fact, $\sum_{j=1}^{\infty} < (B^*B)^{\gamma}\psi_j, \psi_j > = \sum_{j=1}^{\infty} < (B^*B)^{\frac{\gamma}{2}}\psi_j, (B^*B)^{\frac{\gamma}{2}}\psi_j > = \sum_{j=1}^{\infty} ||(B^*B)^{\frac{\gamma}{2}}\psi_j||^2 \leq \overline{c}(\gamma)\sum_j ||L^{-\gamma(a+s)}\psi_j||^2 = \overline{c}(\gamma)\sum_j < L^{-\gamma(a+s)}\psi_j, L^{-\gamma(a+s)}\psi_j > = \overline{c}(\gamma)tr(L^{-2\gamma(a+s)})$.

Therefore, $\mathbf{E}_{x_*}||\mathcal{C}_2||^2 = \mathcal{O}\Big(\delta\alpha^{2(d-1)}\Big)$ and $\sup_{(x_*-x_0)\in\mathcal{X}_\beta(\Gamma)} \mathbf{E}_{x_*}||\hat{x}_\alpha - x_*||^2 = \mathcal{O}\left(\alpha^{\frac{\tilde{\beta}}{a+s}} + \delta\alpha^{-\frac{a+\gamma(a+s)}{(a+s)}}\right)$.

B.3 Proof of theorem 3

Let $\mathbf{E}_\delta^Y$ be the expectation taken with respect to the posterior μ_δ^Y. By the Chebishev's inequality, for $\varepsilon_\delta > 0$ small enough and $M_\delta \to \infty$:

$$\begin{aligned} \mu_\delta^Y\{x \in \mathcal{X} : ||x - x_*|| > \varepsilon_\delta M_\delta\} &\leq \frac{1}{\varepsilon_\delta^2 M_\delta^2}\mathbf{E}_\delta^Y||x - x_*||^2 \\ &= \frac{1}{\varepsilon_\delta^2 M_\delta^2}\Big(||\mathbf{E}(x|y^\delta,\alpha,s) - x_*||^2 + trVar(x|y^\delta,\alpha,s)\Big) \end{aligned}$$

and we have to determine the rate of $Var(x|y^\delta,\alpha,s)$. Since $Var(x|y^\delta,\alpha,s) = \frac{\delta}{\alpha}[\Omega_0 - AK\Omega_0] = \delta\Omega_0^{\frac{1}{2}}(\alpha I + B^*B)^{-1}\Omega_0^{\frac{1}{2}}$ we have: $trVar(x|y^\delta,\alpha,s) = \delta tr(L^{-s}(\alpha I + B^*B)^{-1}L^{-s}) = \delta tr(RR^*)$ with $R = L^{-s}(\alpha I + B^*B)^{-\frac{1}{2}}$. Let $||\cdot||_{HS}$ denote the Hilbert-Schmidt norm. By using the left part of Corollary 2 with $\nu = \frac{s}{a+s}$ we get

$$\begin{aligned}
trVar(x|y^\delta,\alpha,s) &= \delta tr(R^*R) = \delta ||R||_{HS}^2 = \delta\sum_{j=1}^{\infty}||R\varphi_j||^2 \\
&\leq \delta\underline{c}^{-2}\Big(\frac{s}{a+s}\Big)\sum_{j=1}^{\infty}||(B^*B)^{\frac{s}{2(a+s)}}(\alpha I + B^*B)^{-\frac{1}{2}}\psi_j||^2 \\
&= \delta\underline{c}^{-2}\left(\frac{s}{a+s}\right)\sum_{j=1}^{\infty} < (B^*B)^{\frac{s}{2(a+s)}}(\alpha I + B^*B)^{-\frac{1}{2}}\psi_j, (B^*B)^{\frac{s}{2(a+s)}}(\alpha I + B^*B)^{-\frac{1}{2}}\psi_j > \\
&= \delta\underline{c}^{-2}\Big(\frac{s}{a+s}\Big)\sum_{j=1}^{\infty}\frac{\lambda_j^{\frac{2s}{a+s}-2\gamma}}{\alpha+\lambda_j^2} < (B^*B)^{\gamma}\psi_j,\psi_j > \\
&\leq \delta\underline{c}^{-2}\Big(\frac{s}{a+s}\Big)\sup_j\frac{\lambda_j^{\frac{2s}{a+s}-2\gamma}}{\alpha+\lambda_j^2}\sum_{j=1}^{\infty} < (B^*B)^{\gamma}\psi_j,\psi_j > \\
&= \delta\underline{c}^{-2}\Big(\frac{s}{a+s}\Big)\sup_j\frac{\lambda_j^{\frac{2s}{a+s}-2\gamma}}{\alpha+\lambda_j^2}\sum_{j=1}^{\infty}\Big|\Big|(B^*B)^{\gamma/2}\psi_j\Big|\Big|^2 \\
&\leq \delta\underline{c}^{-2}\Big(\frac{s}{a+s}\Big)\alpha^{v-1}v^v(1-v)^{1-v}\bar{c}(\gamma)tr(L^{-2\gamma(a+s)}), \quad v := \frac{s}{(a+s)} - \gamma
\end{aligned}$$

which is finite because $tr(L^{-2\gamma(a+s)}) < \infty$. The last inequality has been obtained by applying the right part of Corollary 2. Now, by this and the result of Theorem 2:

$$\mu_\delta^Y\{x\in\mathcal{X} : ||x-x_*|| > \varepsilon_\delta M_\delta\} \;=\; \mathcal{O}_p\Big(\frac{1}{\varepsilon_\delta^2 M_\delta^2}(\alpha^{\frac{\tilde{\beta}}{a+s}} + \delta\alpha^{-\frac{a+\gamma(a+s)}{(a+s)}})\Big), \quad \tilde{\beta} = \min(\beta, a+2s)$$

in P^{x_*}-probability. Hence, for $\varepsilon_\delta = (\alpha^{\frac{\tilde{\beta}}{2(a+s)}} + \delta^{\frac{1}{2}}\alpha^{-\frac{a+\gamma(a+s)}{2(a+s)}})$ we have $\mu_\delta^Y\{x\in\mathcal{X} : ||x-x_*|| > \varepsilon_\delta M_\delta\} \to 0$.

B.4 Proof of Corollary 1

Let $\widetilde{B^*B} = \Omega_0^{\frac{1}{2}}\mathfrak{K}^*\frac{1}{f_W}\mathfrak{K}\Omega_0^{\frac{1}{2}}$, $R_\alpha = (\alpha I + \widetilde{B^*B})^{-1}$, $\hat{R}_\alpha = (\alpha I + B^*B)^{-1}$, $\Sigma^{-1/2}B = (\Lambda^*)^{-1}\frac{1}{\hat{f}_W}\tilde{K}\Omega_0^{\frac{1}{2}}$, $\widetilde{\Sigma^{-1/2}B} = (\Lambda^*)^{-1}\frac{1}{f_W}\mathfrak{K}\Omega_0^{\frac{1}{2}}$, $\mathfrak{K} = f_W\mathbf{E}(\cdot|W) = \lim_{n\to\infty}\tilde{K}$ and $\mathfrak{K}^* = f_Z\mathbf{E}(\cdot|Z) = \lim_{n\to\infty}\tilde{K}^*$. Moreover, we define $\tilde{y}^\delta = \Lambda\mathfrak{K}x_* + U$ and $\tilde{x}_\alpha = \Omega_0^{\frac{1}{2}}R_\alpha(\widetilde{\Sigma^{-1/2}B})^*(\tilde{y}^\delta - \Lambda\mathfrak{K}x_0)$. We decompose $(\hat{x}_\alpha - x_*)$ as

$$\begin{aligned}
(\hat{x}_\alpha - x_*) &= (\tilde{x}_\alpha - x_*) + \Omega_0^{\frac{1}{2}}\hat{R}_\alpha(\Sigma^{-1/2}B)^*(y^\delta - \Lambda\tilde{K}x_0) - \Omega_0^{\frac{1}{2}}R_\alpha(\widetilde{\Sigma^{-1/2}B})^*(\tilde{y}^\delta - \Lambda\mathfrak{K}x_0) \\
&= (\tilde{x}_\alpha - x_*) + \Omega_0^{\frac{1}{2}}\left[\hat{R}_\alpha(\Sigma^{-1/2}B)^*\Lambda\tilde{K} - R_\alpha(\widetilde{\Sigma^{-1/2}B})^*\Lambda\mathfrak{K}\right](x_* - x_0) \\
&\quad + \Omega_0^{\frac{1}{2}}\left[\hat{R}_\alpha(\Sigma^{-1/2}B)^* - R_\alpha(\widetilde{\Sigma^{-1/2}B})^*\right]U \\
&:= (\tilde{x}_\alpha - x_*) + \mathfrak{A}_1 + \mathfrak{A}_2
\end{aligned}$$

I) Convergence of $\mathbf{E}||(\tilde{x}_\alpha - x_*)||^2$. This rate is given in theorem 2 (since this term depends on operators that do not vary with n).

II) Convergence of $\mathfrak{A}_1$.

$$\begin{aligned}
\mathbf{E}||\mathfrak{A}_1||^2 &= \mathbf{E}||\Omega_0^{\frac{1}{2}}\left(\hat{R}_\alpha\left[(\Sigma^{-1/2}B)^*\Lambda\tilde{K}L^{-s} - (\widetilde{\Sigma^{-1/2}B})^*\Lambda\mathfrak{K}L^{-s}\right] + (\hat{R}_\alpha - R_\alpha)(\widetilde{\Sigma^{-1/2}B})^*\Lambda\mathfrak{K}L^{-s}\right)L^s(x_* - x_0)||^2 \\
&= \mathbf{E}||\Omega_0^{\frac{1}{2}}\left(\hat{R}_\alpha\left[B^*B - \widetilde{B^*B}\right] + \hat{R}_\alpha(\widetilde{B^*B} - B^*B)R_\alpha\widetilde{B^*B}\right)L^s(x_* - x_0)||^2 \\
&= \mathbf{E}||\Omega_0^{\frac{1}{2}}\hat{R}_\alpha\left[B^*B - \widetilde{B^*B}\right]\alpha R_\alpha L^s(x_* - x_0)||^2 \\
&\leq \mathbf{E}||\Omega_0^{\frac{1}{2}}\hat{R}_\alpha||^2\mathbf{E}||\Omega_0^{\frac{1}{2}}(\tilde{K}^*[\hat{f}_W]^{-1}\tilde{K} - \mathfrak{K}^*[f_W]^{-1}\mathfrak{K})||^2||L^{-s}R_\alpha L^s(x_* - x_0)||^2\alpha^2.
\end{aligned}$$

The last term $||L^{-s}R_\alpha L^s(x_* - x_0)||^2\alpha^2$ is equal to term $\mathcal{C}_1$ in the proof of theorem 2 while $\mathbf{E}||\Omega_0^{\frac{1}{2}}\hat{R}_\alpha||^2 = \mathcal{O}(\alpha^{-2})$ and $\mathbf{E}||\Omega_0^{\frac{1}{2}}(\tilde{K}^*[\hat{f}_W]^{-1}\tilde{K} - \mathfrak{K}^*[f_W]^{-1}\mathfrak{K})||^2 = \mathcal{O}(n^{-1}+h^{2\rho})$ under assumption HS since $\mathbf{E}||\Omega_0^{\frac{1}{2}}(\tilde{K}^*[\hat{f}_W]^{-1}\tilde{K} - \mathfrak{K}^*[f_W]^{-1}\mathfrak{K})||^2 \asymp \mathbf{E}||B^*B - \widetilde{B^*B}||^2$. Therefore, $\mathbf{E}||\mathfrak{A}_1||^2 = \mathcal{O}\left(\alpha^{-2}(n^{-1} + h^{2\rho})\alpha^{\frac{\tilde{\beta}}{a+s}}\right)$.

III) Convergence of $\mathfrak{A}_2$.

$$\begin{aligned}
\mathbf{E}||\mathfrak{A}_2||^2 &= \mathbf{E}||\Omega_0^{\frac{1}{2}}\left(\hat{R}_\alpha[(\Sigma^{-1/2}B)^* - (\widetilde{\Sigma^{-1/2}B})^*] + (\hat{R}_\alpha - R_\alpha)(\widetilde{\Sigma^{-1/2}B})^*\right)U||^2 \\
&\leq 2\mathbf{E}||\Omega_0^{\frac{1}{2}}\hat{R}_\alpha[(\Sigma^{-1/2}B)^* - (\widetilde{\Sigma^{-1/2}B})^*]U||^2 + 2\mathbf{E}||L^{-s}\hat{R}_\alpha(\widetilde{B^*B} - B^*B)R_\alpha(\widetilde{\Sigma^{-1/2}B})^*U||^2 \\
&:= 2\mathfrak{A}_{2,1} + 2\mathfrak{A}_{2,2}.
\end{aligned}$$

We start with the analysis of term $\mathfrak{A}_{2,1}$ where we use the notation $\mathcal{T} = \mathbf{E}(\cdot|W)$, $\mathcal{T}^* = \mathbf{E}(\cdot|Z)$, $\hat{\mathcal{T}} = \hat{\mathbf{E}}(\cdot|W)$ and $\hat{\mathcal{T}}^* = \hat{\mathbf{E}}(\cdot|Z)$:

$$\begin{aligned}
\mathfrak{A}_{2,1} &\leq \mathbf{E}||\Omega_0^{\frac{1}{2}}\hat{R}_\alpha||^2\mathbf{E}||\left((\Sigma^{-1/2}B)^* - (\widetilde{\Sigma^{-1/2}B})^*\right)U||^2 \\
&= \mathbf{E}||\Omega_0^{\frac{1}{2}}\hat{R}_\alpha||^2\mathbf{E}||\left((\Lambda^*)^{-1}(\hat{\mathcal{T}} - \mathcal{T})\Omega_0^{\frac{1}{2}}\right)^*U||^2 \\
&= \mathbf{E}||\Omega_0^{\frac{1}{2}}\hat{R}_\alpha||^2\mathbf{E}||\Omega_0^{\frac{1}{2}}(\hat{\mathcal{T}} - \mathcal{T})^*\left[\frac{1}{nh}\sum_i(y_i - <x_*, \frac{K_{Z,h}(z_i - z)}{h}>)K_{W,h}(w_i - w)\right]||^2 \\
&\leq \mathbf{E}||\Omega_0^{\frac{1}{2}}\hat{R}_\alpha||^2\mathbf{E}||\Omega_0^{\frac{1}{2}}(\hat{\mathcal{T}} - \mathcal{T})^*||^2\mathbf{E}||\left[\frac{1}{nh}\sum_i(y_i - <x_*, \frac{K_{Z,h}(z_i - z)}{h}>)K_{W,h}(w_i - w)\right]||^2 \\
&= \mathcal{O}\left(\alpha^{-2}(n^{-1} + h^{2\rho})((nh)^{-1} + h^{2\rho})\right)
\end{aligned}$$

since $\mathbf{E}||\Omega_0^{\frac{1}{2}}(\hat{\mathcal{T}} - \mathcal{T})^*||^2 = \mathcal{O}(n^{-1}+h^{2\rho})$ under assumption HS. Finally, term $\mathfrak{A}_{2,2}$ can be developed as follows:

$$\begin{aligned}
\mathfrak{A}_{2,2} &= \mathbf{E}||\Omega_0^{\frac{1}{2}}\hat{R}_\alpha\Omega_0^{\frac{1}{2}}\left(\tilde{K}^*\frac{1}{\hat{f}_W}\tilde{K} - \mathfrak{K}^*\frac{1}{f_W}\mathfrak{K}\right)L^{-s}R_\alpha(\widetilde{\Sigma^{-1/2}B})^*U||^2 \\
&\leq \mathbf{E}||\Omega_0^{\frac{1}{2}}\hat{R}_\alpha||^2\mathbf{E}||\Omega_0^{\frac{1}{2}}\left(\tilde{K}^*\frac{1}{\hat{f}_W}\tilde{K} - \mathfrak{K}^*\frac{1}{f_W}\mathfrak{K}\right)||^2\mathbf{E}||L^{-s}R_\alpha(\widetilde{\Sigma^{-1/2}B})^*U||^2 \\
&= \mathcal{O}\left(\alpha^{-2}(n^{-1} + h^{2\rho})n^{-1}\alpha^{-\frac{a+\gamma(a+s)}{a+s}}\right)
\end{aligned}$$

since the last term is equal to term $\mathcal{C}_2$ in the proof of theorem 2 and $\mathbf{E}||\Omega_0^{\frac{1}{2}}\left(\tilde{K}^*\frac{1}{f_W}\tilde{K}-\mathfrak{K}^*\frac{1}{f_W}\mathfrak{K}\right)||^2 = \mathcal{O}(n^{-1}+h^{2\rho})$ under assumption HS. By writing $\mathbf{E}||\hat{x}_\alpha - x_*||^2 \leq 2\mathbf{E}||\tilde{x}_\alpha - x_*||^2 + 2\mathbf{E}||\mathfrak{A}_1+\mathfrak{A}_2||^2$ and by putting all these results together we get the result.

C Proofs for Section 4

C.1 Proof of Proposition 2

Let us consider the singular system $\{\lambda_j,\varphi_j,\psi_j\}$ associated with B. Under Assumption C there exist $\underline{\lambda},\bar{\lambda}>0$ such that $\underline{\lambda}j^{-(a+s)}\leq\lambda_j\leq\bar{\lambda}j^{-(a+s)}$. Therefore, the risk associated with $\hat{x}_\alpha$ can be rewritten as:

$$\mathbf{E}_{x_*}||\hat{x}_\alpha - x_*||^2 = \sum_j \frac{\alpha^2}{(\alpha+j^{-2(a+s)})^2} < (x_*-x_0),\psi_j >^2 + \sum_j \frac{\delta j^{-2s-2(a+s)}}{(\alpha+j^{-2(a+s)})^2} =: \mathcal{A}1+\mathcal{A}2.$$

We have that $\sup_{(x_*-x_0)\in\mathcal{X}_\beta(\Gamma)}\mathcal{A}1 = \sup_{\sum_j j^{2\beta}<(x_*-x_0),\psi_j>^2\leq\Gamma}\mathcal{A}1$ and

$$\begin{aligned}
\sup_{\sum_j j^{2\beta}<(x_*-x_0),\psi_j>^2\leq\Gamma}\mathcal{A}1 &= \alpha^2 \sup_{\sum_j j^{2\beta}<(x_*-x_0),\psi_j>^2\leq\Gamma}\sum_j \frac{1}{(\alpha+j^{-2(a+s)})^2} < (x_*-x_0),\psi_j>^2 \\
&= \alpha^2 \sup_{\sum_j j^{2\beta}<(x_*-x_0),\psi_j>^2\leq\Gamma}\sum_j \frac{j^{-2\beta}}{(\alpha+j^{-2(a+s)})^2} j^{2\beta} < (x_*-x_0),\psi_j>^2 \\
&= \alpha^2 \sup_j \frac{j^{-2\beta}}{(\alpha+j^{-2(a+s)})^2}\Gamma.
\end{aligned}$$

The supremum is attained at $j=\alpha^{-\frac{1}{2(a+s)}}\left(\frac{\beta}{2(a+s)-\beta}\right)^{-\frac{1}{2(a+s)}}$ as long as $\beta\leq 2(a+s)$. If $\beta>2(a+s)$ then $\sup_j\frac{j^{-2\beta}}{(\alpha+j^{-2(a+s)})^2}=1$. Consequently,

$$\sup_{(x_*-x_0)\in\mathcal{X}_\beta(\Gamma)}\mathcal{A}1 = \alpha^{\frac{\tilde{\beta}}{a+s}}b^{2b}(1-b)^{2(1-b)}\Gamma, \quad b=\frac{\tilde{\beta}}{2(a+s)}, \quad \tilde{\beta}=\min(\beta,2(a+s))$$

by using the convention $(1-b)^{2(1-b)}=1$ if $b=1$. In order to analyze term $\mathcal{A}2$ we first remark that the summand function $f(j):=\frac{j^{-2s-2(a+s)}}{(\alpha+j^{-2(a+s)})^2}$ defined on $\mathbb{R}_+$ is increasing for $j<\bar{j}$ and decreasing for $j>\bar{j}$ where $\bar{j}=\left(\frac{a+2s}{a}\alpha\right)^{-\frac{1}{2(a+s)}}$. Thus,

$$\begin{aligned}
&\delta\int_{\bar{j}}^{\infty}\frac{j^{-2s-2(a+s)}}{(\alpha+j^{-2(a+s)})^2}dj \leq \mathcal{A}_2 \leq \delta\sum_{j=1}^{\bar{j}}\frac{j^{-2s-2(a+s)}}{(\alpha+j^{-2(a+s)})^2} + \delta\sum_{j=\bar{j}}^{\infty}\frac{j^{-2s-2(a+s)}}{(\alpha+j^{-2(a+s)})^2} \\
\Leftrightarrow\quad &\delta\alpha^{-\frac{2a+1}{2(a+s)}}\int_{\bar{t}}^{\infty}\frac{1}{[t^{-a}(t^{2(a+s)}+1)]^2}dt \leq \mathcal{A}_2 \leq \delta\frac{\bar{j}^{1-2s-2(a+s)}}{(\alpha+\bar{j}^{-2(a+s)})^2} + \delta\alpha^{-\frac{2a+1}{2(a+s)}}\int_{\bar{t}}^{\infty}\frac{1}{[t^{-a}(t^{2(a+s)}+1)]^2}dt
\end{aligned}$$

where $\bar{t}=\left(\frac{a+2s}{a}\right)^{-\frac{1}{2(a+s)}}$. Denote $c_2(\bar{t})=\int_{\bar{t}}^{\infty}\frac{1}{[t^{-a}(t^{2(a+s)}+1)]^2}dt$ and replace $\bar{j}$ by its value to obtain

$$\delta\alpha^{-\frac{2a+1}{2(a+s)}}c_2(\bar{t}) \leq \mathcal{A}_2 \leq \delta\alpha^{-\frac{2a+1}{2(a+s)}}\left(c_2(\bar{t})+\bar{t}^{(1-2s-2(a+s))}\frac{a^2}{4(a+s)^2}\right) =: \delta\alpha^{-\frac{2a+1}{2(a+s)}}\tilde{c}_2(\bar{t}).$$

Therefore, $\mathcal{A}_2 \asymp \delta\alpha^{-\frac{2a+1}{2(a+s)}}$. Finally, let $G = b^{2b}(1-b)^{2(1-b)}\Gamma$,

$$\alpha^{\frac{\tilde{\beta}}{(a+s)}}G + \delta\alpha^{-\frac{2a+1}{2(a+s)}}c_2(\bar{t}) \leq \sup_{(x_*-x_0)\in\mathcal{X}_\beta(\Gamma)} \mathbf{E}_{x_*}||\hat{x}_\alpha - x_*||^2 \quad \leq \quad \alpha^{\frac{\tilde{\beta}}{(a+s)}}G + \delta\alpha^{-\frac{2a+1}{2(a+s)}}\tilde{c}_2(\bar{t}),$$

this proves (12). By replacing $\alpha \asymp \delta$ (resp. $\alpha \asymp \delta^{\frac{a+s}{\beta+a+1/2}}$) in this expression we get *(i)* (resp. *(ii)*).

C.2 Proof of Proposition 3

The proof proceeds similar to the proof of Proposition 2, so we just sketch it. The risk $\mathbf{E}_{x_*}||\hat{x}_\alpha - x_*||^2$ associated with $\hat{x}_\alpha$ rewrites as the risk in section C.1. The only term that we need to analyze is $\mathcal{A}1$ since the analysis of $\mathcal{A}2$ is the same as in C.1. By using Assumption B'

$$\mathcal{A}1 \quad = \quad \alpha^2 \sum_j \frac{1}{(\alpha + j^{-2(a+s)})^2} < (x_* - x_0), \psi_j >^2 = \alpha^2 \sum_j \frac{j^{-2b_0}}{(\alpha + j^{-2(a+s)})^2}$$

and the function $f(j) := \frac{j^{-2b_0}}{(\alpha+j^{-2(a+s)})^2}$ defined on $\mathbb{R}_+$ is decreasing in j if $b_0 \geq 2a+2s$. If $b_0 < 2a+2s$ then $f(j)$ is increasing for $j < \bar{j}$ and decreasing for $j > \bar{j}$ where $\bar{j} = \alpha^{-\frac{1}{2(a+s)}}\left(\frac{b_0}{2a+2s-b_0}\right)^{-\frac{1}{2(a+s)}}$. Therefore, to upper and lower bound $\mathcal{A}_1$ we have to consider these two cases separately. If $b_0 < 2a+2s$

$$\alpha^2 \int_{\bar{j}}^{\infty} \frac{j^{-2b_0}}{(\alpha + j^{-2(a+s)})^2} dj \leq \mathcal{A}_1 \leq \alpha^2 \sum_{j=1}^{\bar{j}} \frac{j^{-2b_0}}{(\alpha + j^{-2(a+s)})^2} + \alpha^2 \sum_{j=\bar{j}}^{\infty} \frac{j^{-2b_0}}{(\alpha + j^{-2(a+s)})^2}$$
$$\Leftrightarrow \quad \alpha^{\frac{2b_0-1}{2(a+s)}} \int_{\underline{u}}^{\infty} \frac{u^{4(a+s)-2b_0}}{(u^{2(a+s)}+1)^2} du \leq \mathcal{A}_1 \leq \alpha^2 \frac{\bar{j}^{1-2b_0}}{(\alpha + \bar{j}^{-2(a+s)})^2} + \alpha^{\frac{2b_0-1}{2(a+s)}} \int_{\underline{u}}^{\infty} \frac{u^{4(a+s)-2b_0}}{(u^{2(a+s)}+1)^2} du$$

where $\underline{u} = \left(\frac{b_0}{2a+2s-b_0}\right)^{-\frac{1}{2(a+s)}}$. Denote $c_1(\underline{u}) = \int_{\underline{u}}^{\infty} \frac{u^{4(a+s)-2b_0}}{(u^{2(a+s)}+1)^2} du$ and replace $\bar{j}$ by its value. Then,

$$\alpha^{\frac{2b_0-1}{2(a+s)}} c_1(\underline{u}) \leq \mathcal{A}_1 \leq \alpha^{\frac{2b_0-1}{2(a+s)}} \left(c_1(\underline{u}) + \underline{u}^{1-2b_0} \left(\frac{2a+2s-b_0}{2(a+s)} \right)^2 \right) =: \alpha^{\frac{2b_0-1}{2(a+s)}} \tilde{c}_1(\underline{u}).$$

If $b_0 \geq 2a+2s$:

$$\alpha^2 \int_1^{\infty} \frac{j^{-2b_0}}{(\alpha + j^{-2(a+s)})^2} dj \leq \mathcal{A}_1 \leq \alpha^2 \left. \frac{j^{-2b_0}}{(\alpha + j^{-2(a+s)})^2} \right|_{j=1} + \alpha^2 \int_1^{\infty} \frac{j^{-2b_0}}{(\alpha + j^{-2(a+s)})^2} dj$$
$$\Leftrightarrow \quad \alpha^{\frac{2b_0-1}{2(a+s)}} \int_{\bar{u}}^{\infty} \frac{u^{4(a+s)-2b_0}}{(u^{2(a+s)}+1)^2} du \leq \mathcal{A}_1 \leq \alpha^2 + \alpha^{\frac{2b_0-1}{2(a+s)}} \int_{\bar{u}}^{\infty} \frac{u^{4(a+s)-2b_0}}{(u^{2(a+s)}+1)^2} du$$

where $\bar{u} = \alpha^{\frac{1}{2(a+s)}}$. By using the notation defined above we obtain:

$$\alpha^{\frac{2b_0-1}{2(a+s)}} c_1(\bar{u}) \leq \mathcal{A}_1 \leq \alpha^2 + \alpha^{\frac{2b_0-1}{2(a+s)}} c_1(\bar{u})$$

Since the integral in $c_1(\bar{u})$ is convergent, the upper bound is of order $\alpha^{\frac{2\tilde{b}_0-1}{2(a+s)}}$ where $\tilde{b}_0 = \min(b_0, 2a+2s+1/2)$. Summarizing the two cases and by defining: $c_1 = c_1(\underline{u})I(b_0 < 2(a+s)) + c_1(\bar{u})I(b_0 \geq 2(a+s))$ and

$\tilde{c}_1 = \tilde{c}_1(\underline{u})I(b_0 < 2(a+s)) + c_1(\bar{u})I(b_0 \geq 2(a+s))$ we have

$$\alpha^{\frac{2b_0-1}{2(a+s)}} c_1 \leq \mathcal{A}_1 \leq \quad \alpha^2 I(b_0 \geq (2a+2s)) + \alpha^{\frac{2b_0-1}{2(a+s)}} \tilde{c}_1.$$

By using the upper and lower bounds for $\mathcal{A}_2$ given in the proof of Proposition 2 we get the expression in (13):

$$\alpha^{\frac{2b_0-1}{2(a+s)}} c_1 + \delta\alpha^{-\frac{2a+1}{2(a+s)}} c_2(\bar{t}) \leq \mathbf{E}_{x_*}||\hat{x}_\alpha - x_*||^2 \leq \alpha^2 I(b_0 \geq (2a+2s)) + \alpha^{\frac{2b_0-1}{2(a+s)}} \tilde{c}_1 + \delta\alpha^{-\frac{2a+1}{2(a+s)}} \tilde{c}_2(\bar{t}).$$

By replacing $\alpha \asymp \delta^{\frac{a+s}{b_0+a}}$ we obtain (14).

C.3 Proof of Theorem 4

We use the same strategy used for the proof of theorem 3. The trace of the posterior covariance operator is:

$$tr[Var(x|y^\delta, \alpha, s)] = \delta \sum_{j=1}^{\infty} < \Omega_0(\alpha I + B^*B)^{-1}\psi_j, \psi_j >= \delta \sum_{j=1}^{\infty} \frac{j^{-2s}}{\alpha + j^{-2(a+s)}}.$$

Since $f(j) := \frac{j^{-2s}}{\alpha+j^{-2(a+s)}}$ is increasing in j for $j < \bar{j}$ and decreasing for $j > \bar{j}$ where $\bar{j} = (\alpha s/a)^{-\frac{1}{2(a+s)}}$ then we can upper and lower bound the trace of the posterior covariance operator as:

$$\begin{aligned}
\delta \sum_{j=\bar{j}}^{\infty} f(j) \quad &\leq \quad tr[Var(x|y^\delta, \alpha, s)] \leq \delta \sum_{j=1}^{\bar{j}} f(j) + \delta \sum_{j=\bar{j}}^{\infty} f(j) \\
\Leftrightarrow \delta\alpha^{-\frac{2a+1}{2(a+s)}} \int_{\bar{t}}^{\infty} \frac{t^{2a}}{t^{2(a+s)}+1} dt \quad &\leq \quad tr[Var(x|y^\delta, \alpha, s)] \leq \delta \frac{\bar{j}^{1-2s}}{\alpha + \bar{j}^{-2(a+s)}} + \delta\alpha^{-\frac{2a+1}{2(a+s)}} \int_{\bar{t}}^{\infty} \frac{t^{2a}}{t^{2(a+s)}+1} dt \\
\Leftrightarrow \delta\alpha^{-\frac{2a+1}{2(a+s)}} \kappa(\bar{t}) \quad &\leq \quad tr[Var(x|y^\delta, \alpha, s)] \leq \delta\alpha^{-\frac{2a+1}{2(a+s)}} \bar{t}^{-\frac{(1-2s)}{2(a+s)}} \frac{a}{a+s} + \delta\alpha^{-\frac{2a+1}{2(a+s)}} \kappa_2(\bar{t})
\end{aligned}$$

where $\bar{t} = (s/a)^{-\frac{1}{2(a+s)}}$ and $\kappa_2(\bar{t}) = \int_{\bar{t}}^{\infty} \frac{t^{2a}}{t^{2(a+s)}+1} dt$. Thus, $tr[Var(x|y^\delta, \alpha, s)] \asymp \delta\alpha^{-\frac{2a+1}{2(a+s)}}$. By the Chebishev's inequality and Proposition 3, for $\varepsilon_\delta > 0$ small enough and $M_\delta \to \infty$

$$\begin{aligned}
\mu_\delta^Y\{x \in \mathcal{X} : ||x - x_*|| > \varepsilon_\delta M_\delta\} \quad &\leq \quad \frac{1}{\varepsilon_\delta^2 M_\delta^2}\Big(||\mathbf{E}(x|y^\delta, \alpha, s) - x_*||^2 + trVar(x|y^\delta, \alpha, s)\Big) \\
&= \quad \mathcal{O}_p\Big(\frac{1}{\varepsilon_\delta^2 M_\delta^2}(\alpha^{\frac{2\tilde{b}_0-1}{2(a+s)}} + \delta\alpha^{-\frac{2a+1}{2(a+s)}})\Big), \quad \tilde{b}_0 = \min(b_0, 2(a+s)+1/2)
\end{aligned}$$

in P^{x_*}-probability. Hence, for $\varepsilon_\delta = (\alpha^{\frac{2\tilde{b}_0-1}{4(a+s)}} + \delta^{\frac{1}{2}}\alpha^{-\frac{2a+1}{4(a+s)}})$ we conclude that $\mu_\delta^Y\{x \in \mathcal{X} : ||x - x_*|| > \varepsilon_\delta M_\delta\} \to 0$. For $\alpha \asymp \delta$ we obtain *(i)*. To obtain *(ii)*: $\arg\inf_\alpha \varepsilon_\delta = \delta^{\frac{a+s}{\tilde{b}_0+a}}$ and $\inf_\alpha \varepsilon_\delta = \delta^{\frac{2\tilde{b}_0-1}{4(\tilde{b}_0+a)}}$.

C.4 Proof of Lemma 1

Let us denote: $s = a_0\tilde{s}$, $\tilde{R}_\alpha = (\alpha I + \tilde{K}^{1+2\tilde{s}})^{-1}$, $R_\alpha = (\alpha I + K^{1+2\tilde{s}})^{-1}$, $\hat{x}_\alpha = K^{\tilde{s}} R_\alpha K^{\tilde{s}}(y^\delta - Kx_0) + x_0$. Then,

$$\begin{aligned}
\mathbf{E}_{x_*}||\hat{x}_\alpha - x_*||^2 &= \mathbf{E}_{x_*}||\alpha R_\alpha(x_* - x_0) + K^{\tilde{s}} R_\alpha K^{\tilde{s}} U||^2 \\
&= \mathbf{E}_{x_*}||\alpha \tilde{R}_\alpha(x_* - x_0) + \alpha(R_\alpha - \tilde{R}_\alpha)(x_* - x_0) + \tilde{R}_\alpha \tilde{K}^{2\tilde{s}} U + (R_\alpha K^{2\tilde{s}} - \tilde{R}_\alpha \tilde{K}^{2\tilde{s}})U||^2 \\
&\leq 2\mathbf{E}_{x_*}||\alpha \tilde{R}_\alpha(x_* - x_0) + \tilde{R}_\alpha \tilde{K}^{2\tilde{s}} U||^2 + 4\mathbf{E}_{x_*}||\alpha(R_\alpha - \tilde{R}_\alpha)(x_* - x_0)||^2 \\
&\quad +4\mathbf{E}_{x_*}||(R_\alpha K^{2\tilde{s}} - \tilde{R}_\alpha \tilde{K}^{2\tilde{s}})U||^2 =: 2\mathcal{A}_1 + 4\mathcal{A}_2 + 4\mathcal{A}_3.
\end{aligned}$$

The rate of term $\mathcal{A}_1$ is given by Proposition 3 with $\delta = n^{-1}$, $a = \frac{a_0}{2}$ and $s = a_0\tilde{s}$: $\mathcal{A}_1 \asymp \left(\alpha^{\frac{2b_0-1}{2s+a_0}} + n^{-1}\alpha^{-\frac{a_0+1}{2s+a_0}}\right)$. By using the result of Lemma 2 and a Taylor expansion of the second order of the function $\tilde{K}^{2\tilde{s}+1}$ around K (that is, $\tilde{K}^{2\tilde{s}+1} = K^{2\tilde{s}+1} + (2\tilde{s}+1)K^{2\tilde{s}}(|\tilde{K} - K|) + \mathcal{O}(|\tilde{K} - K|^2))$ we obtain

$$\begin{aligned}
\mathcal{A}_2 &= \mathbf{E}_{x_*}||\alpha(R_\alpha - \tilde{R}_\alpha)(x_* - x_0)||^2 = \mathbf{E}_{x_*}||R_\alpha(\tilde{K}^{2\tilde{s}+1} - K^{2\tilde{s}+1})\alpha\tilde{R}_\alpha)(x_* - x_0)||^2 \\
&= \mathbf{E}_{x_*}||R_\alpha\left((2\tilde{s}+1)K^{2s}(|\tilde{K} - K|) + \mathcal{O}(|\tilde{K} - K|^2)\right)\alpha\tilde{R}_\alpha)(x_* - x_0)||^2 \\
&\leq \mathcal{O}\left(\mathbf{E}_{x_*}||R_\alpha K^{2s}||^2 \mathbf{E}_{x_*}||\tilde{K} - K||^2 \mathbf{E}_{x_*}||\alpha\tilde{R}_\alpha(x_* - x_0)||^2\right) \\
&\quad +\mathcal{O}\left(\mathbf{E}_{x_*}||R_\alpha||^2 \mathbf{E}_{x_*}||(\tilde{K} - K)^2||^2 \mathbf{E}_{x_*}||\alpha\tilde{R}_\alpha(x_* - x_0)||^2\right) \\
&= \mathcal{O}\left(n^{-1}\alpha^{-\frac{2a_0}{2s+a_0}}\alpha^{\frac{2b_0-1}{2s+a_0}}\right) + \mathcal{O}\left(n^{-2}\alpha^{-2}\alpha^{\frac{2b_0-1}{2s+a_0}}\right)
\end{aligned}$$

where we have used the facts that: $\mathbf{E}_{x_*}||R_\alpha K^{2s}||^2 = \mathcal{O}\left(\alpha^{-\frac{2a_0}{2s+a_0}}\right)$ and $\mathbf{E}_{x_*}||(\tilde{K} - K)^2||^2 = \mathcal{O}\left(n^{-2}\right)$. The latter rate is obtained as follows: for some basis $\{\varphi_j\}$:

$$\begin{aligned}
||(\tilde{K} - K)^2||^2 &:= \sup_{||\phi||\leq 1}\sum_{j=1}^{\infty} < (\tilde{K} - K)^2\phi, \varphi_j >^2 = \sup_{||\phi||\leq 1}\sum_{j=1}^{\infty} < (\tilde{K} - K)\phi, (\tilde{K} - K)\varphi_j >^2 \\
&\leq \sup_{||\phi||\leq 1}\sum_{j=1}^{\infty} ||(\tilde{K} - K)\phi||^2 ||(\tilde{K} - K)\varphi_j||^2 \\
&= ||(\tilde{K} - K)||^2 \sum_{j=1}^{\infty} ||(\tilde{K} - K)\varphi_j||^2 = ||(\tilde{K} - K)||^2 ||(\tilde{K} - K)||^2_{HS}.
\end{aligned}$$

Thus, the proof of Lemma 2 implies: $\mathbf{E}_{x_*}||(\tilde{K} - K)^2||^2 \leq \left(\mathbf{E}_{x_*}||(\tilde{K} - K)^2||^2_{HS}\right) = \mathcal{O}\left(n^{-2}\right)$. Next,

$$\begin{aligned}
\mathcal{A}_3 &= \mathbf{E}_{x_*}||(R_\alpha K^{2\tilde{s}} - \tilde{R}_\alpha \tilde{K}^{2\tilde{s}})U||^2 = \mathbf{E}_{x_*}||(R_\alpha K^{2\tilde{s}} - \tilde{K}^{2\tilde{s}}\tilde{R}_\alpha)U||^2 \\
&= \mathbf{E}_{x_*}||R_\alpha(\alpha K^{2\tilde{s}} + K^{2\tilde{s}}\tilde{K}^{2\tilde{s}+1} - \alpha\tilde{K}^{2\tilde{s}} - K^{2\tilde{s}+1}\tilde{K}^{2\tilde{s}})\tilde{R}_\alpha U||^2 \\
&= \mathbf{E}_{x_*}||R_\alpha\left[\alpha(K^{2\tilde{s}} - \tilde{K}^{2\tilde{s}}) + K^{2\tilde{s}}(\tilde{K} - K)\tilde{K}^{2\tilde{s}}\right]\tilde{R}_\alpha U||^2 \\
&\leq 2\mathbf{E}_{x_*}||R_\alpha\alpha(K^{2\tilde{s}} - \tilde{K}^{2\tilde{s}})\tilde{R}_\alpha U||^2 + 2\mathbf{E}_{x_*}||R_\alpha K^{2\tilde{s}}(\tilde{K} - K)\tilde{K}^{2\tilde{s}}\tilde{R}_\alpha U||^2
\end{aligned}$$

and $2\mathbf{E}_{x_*}||R_\alpha K^{2\tilde{s}}(\tilde{K} - K)\tilde{K}^{2\tilde{s}}\tilde{R}_\alpha U||^2 = \mathcal{O}\left(\alpha^{-\frac{2a_0}{2s+a_0}}n^{-2}\alpha^{-\frac{a_0+1}{2s+a_0}}\right)$ by lemma 2 and since $\mathbf{E}_{x_*}||R_\alpha K^{2\tilde{s}}||^2 = \mathcal{O}\left(\alpha^{-\frac{2a_0}{2s+a_0}}\right)$. To obtain the rate for the first term we use a Taylor expansion of the second order of the function $K^{2\tilde{s}}$ around $\tilde{K}$ (that is, $K^{2\tilde{s}} = \tilde{K}^{2\tilde{s}} + 2\tilde{s}\tilde{K}^{2\tilde{s}-1}(|K-\tilde{K}|) + \mathcal{O}(|K-\tilde{K}|^2)$ and the $\mathcal{O}$ term is negligible as shown above):

$$\mathbf{E}_{x_*}||R_\alpha \alpha (K^{2\tilde{s}} - \tilde{K}^{2\tilde{s}})\tilde{R}_\alpha U||^2 = \mathcal{O}\left(\mathbf{E}_{x_*}||R_\alpha \alpha||\mathbf{E}_{x_*}||K-\tilde{K}||^2 \mathbf{E}_{x_*}||\tilde{K}^{2\tilde{s}-1}\tilde{R}_\alpha U||^2\right) = \mathcal{O}\left(n^{-1}\alpha^{-\frac{4a_0+1}{a_0(2\tilde{s}+1)}} n^{-1}\right)$$

by the result of lemma 2 and since $\mathbf{E}_{x_*}||\alpha R_\alpha||^2 = \mathcal{O}(1)$ and $\mathbf{E}_{x_*}||\tilde{K}^{2\tilde{s}-1}\tilde{R}_\alpha U||^2 = \mathcal{O}\left(n^{-1}\alpha^{-\frac{4a_0+1}{2s+a_0}}\right)$. Remark that to recover the rate for $\mathbf{E}_{x_*}||\tilde{K}^{2\tilde{s}-1}\tilde{R}_\alpha U||^2$ a procedure similar to the one for recovering the rate for term $\mathcal{A}_2$ in proposition 2 has been used. By putting all these results together we obtain

$$\begin{aligned}\mathbf{E}_{x_*}||\hat{x}_\alpha - x_*||^2 &= \mathcal{O}\Big(\alpha^{\frac{2b_0-1}{a_0+2s}} + n^{-1}\alpha^{-\frac{a_0+1}{a_0+2s}} + n^{-1}\alpha^{-\frac{2a_0}{a_0+2s}}\alpha^{\frac{2b_0-1}{2s+a_0}} + n^{-2}\alpha^{-2}\alpha^{\frac{2b_0-1}{2s+a_0}} \\ &\quad + n^{-1}\alpha^{-\frac{4a_0+1}{a_0(2\tilde{s}+1)}} n^{-1} + \alpha^{-\frac{2a_0}{2s+a_0}} n^{-2}\alpha^{-\frac{a_0+1}{a_0+2s}}\Big).\end{aligned}$$

Finally, by replacing α with $\alpha_* \asymp n^{-\frac{a_0+2s}{2b_0+a_0}}$, the third to sixth terms are negligible with respect to the first and second terms if $b_0 > s$ and $b_0 > a_0$. This concludes the proof.

D Proofs for Section 5

D.1 Proof of Theorem 5

To prove Theorem 5 we use Theorem 3.3 p.123 in Kuo (1975). We first rewrite this theorem and then show that the conditions of this theorem are verified in our case. The proof of theorem 7 is given in Kuo (1975).

Theorem 7. *Let P_2 be a Gaussian measure on $\mathcal{Y}$ with mean m and covariance operator S_2 and P_1 be another Gaussian measure on the same space with mean m and covariance operator S_1. If there exists a positive definite, bounded, continuously invertible operator $\mathcal{H}$ such that $S_2 = S_1^{\frac{1}{2}}\mathcal{H}S_1^{\frac{1}{2}}$ and $\mathcal{H} - I$ is Hilbert-Schmidt, then P_2 is equivalent to P_1. Moreover, the Radon-Nikodym derivative is given by*

$$\frac{dP_2}{dP_1}(\{z_j\}) = \prod_{j=1}^{\infty}\sqrt{\frac{\alpha}{\lambda_j^2+\alpha}}e^{\frac{\lambda_j^2}{2(\lambda_j^2+\alpha)}z_j^2}, \tag{31}$$

with $\frac{\lambda_j^2}{\alpha}$ the eigenvalues of $\mathcal{H} - I$ and z_j a sequence of real numbers.

In our case: $P_2 = P_\alpha$, $m = Kx_0$, $S_2 = \delta\Sigma + \frac{\delta}{\alpha}K\Omega_0 K^*$, $P_1 = P_0$ and $S_1 = \delta\Sigma$. We rewrite S_2 as

$$S_2 \equiv \left(\delta\Sigma + \frac{\delta}{\alpha}K\Omega_0 K^*\right) = \sqrt{\delta}\Sigma^{\frac{1}{2}}\left[I + \frac{1}{\alpha}\Sigma^{-1/2}K\Omega_0 K^*\Sigma^{-1/2}\right]\Sigma^{\frac{1}{2}}\sqrt{\delta} = S_1^{\frac{1}{2}}\mathcal{H}S_1^{\frac{1}{2}}$$

with $\mathcal{H} = [I + \frac{1}{\alpha}\Sigma^{-1/2}K\Omega_0 K^*\Sigma^{-1/2}] = (I + \frac{1}{\alpha}BB^*)$. In the following four points we show that $\mathcal{H}$ satisfies all the properties required in Theorem 7.

1) $\mathcal{H}$ is *positive definite.* In fact, $(I + \frac{1}{\alpha}BB^*)$ is self-adjoint, *i.e.* $(I + \frac{1}{\alpha}BB^*)^* = (I + \frac{1}{\alpha}BB^*)$ and $\forall\varphi \in \mathcal{Y}$, $\varphi \neq 0$

$$< (I + \frac{1}{\alpha}BB^*)\varphi, \varphi > = < \varphi, \varphi > + \frac{1}{\alpha} < B^*\varphi, B^*\varphi > = ||\varphi||^2 + \frac{1}{\alpha}||B^*\varphi||^2 > 0.$$

2) $\mathcal{H}$ is *bounded.* The operators B and B^* are bounded if Assumption A.3 holds, the operator I is bounded by definition and a linear combination of bounded operators is bounded, see Remark 2.7 in Kress (1999).

3) $\mathcal{H}$ is *continuously invertible.* To show this, we first recall that $(I + \frac{1}{\alpha}BB^*)$ is continuously invertible if its inverse is bounded, *i.e.* there exists a positive number C such that $||(I + \frac{1}{\alpha}BB^*)^{-1}\varphi|| \leq C||\varphi||$, $\forall \varphi \in \mathcal{Y}$. We have $||(I + \frac{1}{\alpha}BB^*)^{-1}\varphi|| \leq (\sup_j \frac{\alpha}{\alpha+\lambda_j^2})||\varphi|| = ||\varphi||$, $\forall \varphi \in \mathcal{Y}$.

4) $(\mathcal{H} - I)$ is *Hilbert-Schmidt.* To show this we have to consider the Hilbert-Schmidt norm $||\frac{1}{\alpha}BB^*||_{HS} = \frac{1}{\alpha}\sqrt{tr((BB^*)^2)}$. Now, $tr((BB^*)^2) = tr(\Omega_0 T^*T\Omega_0 T^*T) \leq tr(\Omega_0)||T^*T\Omega_0 T^*T|| < \infty$ since $T := \Sigma^{-1/2}K$ has a bounded norm under Assumption A.1.

D.2 Proof of Theorem 6

We start by showing the first statement of the theorem and then we proceed with the proof of (20). Let $\tilde{c}_3$, c_4 and $\tilde{c}_4$ be the constants defined in Lemma 3. Fix $\epsilon_\delta = \delta^{\frac{1-r}{4(b_0+a)}}$ for $0 < r < 1$, $\alpha_1 = \left(\frac{\tilde{c}_3}{c_4}(1+\epsilon_\delta)\right)^{\frac{a+s}{b_0+a}}\delta^{\tilde{p}}$, $\alpha_2 = \left(\frac{\tilde{c}_3}{\tilde{c}_4}(1-\epsilon_\delta)\right)^{\frac{a+s}{b_0+a}}\delta^{p}$ where: $\tilde{p} = \frac{a+s}{b_0+a+\tilde{\eta}}$, $p = \frac{a+s}{b_0+a-\eta}$, $\tilde{\eta} = \eta I\{b_0 - a - 2s - 1/2 > 0\}$ and $(b_0 + a) > \eta > \max\{b_0 - s - 1/4, 0\}$. Remark that, since $b_0 - s - 1/4 > \max\{(b_0 - 2s - a - 1), (b_0 - s - 1/2)\}$, $b_0 - s - 1/4 > \frac{1}{4}(b_0 - 2s - a - 1/2)$ when $b_0 > a + 2s + 1/2$ and $0 < \epsilon_\delta < 1$ then the assumptions of Lemmas 5 and 6 are satisfied. Moreover, $\tilde{c}_4 \geq c_4$, see the proof of Lemma 3. Because $S_{y^\delta}(\cdot)$ is continuous on $[\alpha_2, \alpha_1]$, in order to prove the existence of an $\hat{\alpha}$ such that $S_{y^\delta}(\hat{\alpha}) = 0$ it is sufficient to prove that $P\{\, S_{y^\delta}(\alpha_2) > 0 \text{ and } S_{y^\delta}(\alpha_1) < 0\} \to 1$ when $\delta \to 0$. In fact,

$$P\{\,\exists\hat{\alpha}\,;\, S_{y^\delta}(\hat{\alpha}) = 0\} \geq P\{S_{y^\delta}(\hat{\alpha}) = 0 \,\cap\, \hat{\alpha} \in (\alpha_2, \alpha_1)\} \geq P\{\, S_{y^\delta}(\alpha_2) > 0 \text{ and } S_{y^\delta}(\alpha_1) < 0\}.$$

From Lemmas 5, 6 and 7 we conclude that as $\delta \to 0$: $P\{\, S_{y^\delta}(\alpha_2) > 0$ and $S_{y^\delta}(\alpha_1) < 0\} \to 1$ and so $\hat{\alpha} \in (\alpha_2, \alpha_1)$ with probability approaching 1. This implies that we can write $\hat{\alpha}$ as a (random) convex combination of α_2 and α_1: $\hat{\alpha} = \rho\alpha_2 + (1-\rho)\alpha_1$ for ρ a random variable with values in $(0, 1)$. Since $\alpha_2 \to 0$ faster than $\alpha_1 \to 0$ then:

$$\hat{\alpha} = o_p\left(\delta^{\frac{a+s}{b_0+a+\tilde{\eta}}}\right) + \delta^{\frac{a+s}{b_0+a+\tilde{\eta}}}(1-\rho)\left[\frac{\tilde{c}_3}{c_4}(1+\epsilon_\delta)\right]^{\frac{a+s}{b_0+a}} =: o_p\left(\delta^{\frac{a+s}{b_0+a+\tilde{\eta}}}\right) + \delta^{\frac{a+s}{b_0+a+\tilde{\eta}}}(1-\rho)\kappa_1. \tag{32}$$

Now, we proceed with the proof of (20). Define the event $G := \{\hat{\alpha} \in (\alpha_2, \alpha_1)\}$ and G^c its complement. By the Markov's inequality, to show that $\mu^Y_{\delta,\hat{\alpha}}\{||x - x_*|| > \varepsilon_\delta M_\delta\} \to 0$ in probability we can show that its expectation, with respect P^{x_*} converges to 0. Then,

$$\begin{aligned}
\mathbf{E}_{x_*}\left(\mu^Y_{\delta,\hat{\alpha}}\{||x - x_*|| > \varepsilon_\delta M_\delta\}\right) &= \mathbf{E}_{x_*}\left(\mu^Y_{\delta,\hat{\alpha}}\{||x - x_*|| > \varepsilon_\delta M_\delta\}(1 - I_G + I_G)\right)\\
&= \mathbf{E}_{x_*}\left(\mu^Y_{\delta,\hat{\alpha}}\{||x - x_*|| > \varepsilon_\delta M_\delta\}I_{G^c}\right) + \mathbf{E}_{x_*}\left(\mu^Y_{\delta,\hat{\alpha}}\{||x - x_*|| > \varepsilon_\delta M_\delta\}I_G\right)\\
&\leq \mathbf{E}_{x_*}(I_{G^c}) + \mathbf{E}_{x_*}\left(\frac{||\hat{x}_{\hat{\alpha}} - x_*||^2}{\varepsilon_\delta^2 M_\delta^2}I_G + \frac{trVar(x|y^\delta, \hat{\alpha}, s)}{\varepsilon_\delta^2 M_\delta^2}I_G\right)
\end{aligned} \tag{33}$$

where the last inequality follows from the Chebishev's inequality. If $\eta > 1/2$, by lemma 7: $\mathbf{E}(I_{G^c}) = P(G^c) \to 0$. We then analyze the second term of (33). Let α_* denote the optimum value of α given in proposition 3, that is, $\alpha_* = \delta^{\frac{a+s}{b_0+a}}$. By using the notation $B(\alpha) := (\alpha I + B^*B)^{-1}$ and $\tilde{B}(\alpha) := (\alpha I + \Omega_0 K^*\Sigma^{-1}K)^{-1}$ and by adding and subtracting $\Omega_0^{\frac{1}{2}} B(\alpha_*)(\Sigma^{-1/2}B)^*(y^\delta - Kx_0)$ we obtain:

$$\begin{aligned}
&\mathbf{E}_{x_*}\left(||\hat{x}_{\hat{\alpha}} - x_*||^2 I_G\right) = \mathbf{E}_{x_*}||\Omega_0^{\frac{1}{2}} B(\hat{\alpha})(\Sigma^{-1/2}B)^*(y^\delta - Kx_0) + x_0 - x_*||^2 I_G \\
&\le 2\mathbf{E}_{x_*}||\Omega_0^{\frac{1}{2}} B(\hat{\alpha})(\alpha_* - \hat{\alpha})B(\alpha_*)(\Sigma^{-1/2}B)^*(y^\delta - Kx_0)||^2 I_G \\
&\qquad + 2\mathbf{E}_{x_*}||\Omega_0^{\frac{1}{2}} B(\alpha_*)(\Sigma^{-1/2}B)^*(y^\delta - Kx_0) + x_0 - x_*||^2 I_G \\
&\le 4\mathbf{E}_{x_*}||\tilde{B}(\hat{\alpha})(\alpha_* - \hat{\alpha})\left[\Omega_0^{\frac{1}{2}} B(\alpha_*)(\Sigma^{-1/2}B)^*(y^\delta - Kx_0) + x_0 - x_*\right]||^2 I_G \\
&\qquad + 4\mathbf{E}_{x_*}||\tilde{B}(\hat{\alpha})(\alpha_* - \hat{\alpha})(x_0 - x_*)||^2 I_G + 2\mathbf{E}_{x_*}||\hat{x}_{\alpha_*} - x_*||^2 \\
&\le 4\mathbf{E}_{x_*}||\tilde{B}(\hat{\alpha})(\alpha_* - \hat{\alpha})||^2 I_G \mathbf{E}_{x_*}||\hat{x}_{\alpha_*} - x_*||^2 + 4\mathbf{E}_{x_*}(\alpha_* - \hat{\alpha})^2||\tilde{B}(\hat{\alpha})(x_0 - x_*)||^2 I_G + 2\mathbf{E}_{x_*}||\hat{x}_{\alpha_*} - x_*||^2,
\end{aligned}$$

where the last inequality is due to the Cauchy-Schwartz inequality. By the first part of this proof $\hat{\alpha}$ can be written as in (32) so that there exists a positive constant κ_1 such that:

$$\begin{aligned}
\mathbf{E}_{x_*}||\tilde{B}(\hat{\alpha})(\alpha_* - \hat{\alpha})||^2 I_G &= \mathbf{E}_{x_*}(\alpha_* - \hat{\alpha})^2 \sup_{||\phi||=1} \sum_{j=1}^\infty \frac{1}{(\hat{\alpha} + j^{-2(a+s)})^2} < \phi, \psi_j >^2 I_G \ \le\ \mathbf{E}\left(\frac{\alpha_* - \hat{\alpha}}{\hat{\alpha}} I_G\right)^2 \\
&= \mathbf{E}_{x_*}\left(\frac{\left(\delta^{\tilde{p}\frac{(b_0 - \tilde{b}_0 + \tilde{\eta})}{(a+\tilde{b}_0)}} - o_p(1) - (1-\rho)\kappa_1\right)^2}{\left(o_p(1) + (1-\rho)\kappa_1\right)^2}\right) = \mathcal{O}(1).
\end{aligned}$$

Moreover, $||\tilde{B}(\hat{\alpha})(x_0 - x_*)||^2 \asymp \sum_j \frac{j^{-2b_0}}{(\hat{\alpha} + j^{-2(a+s)})^2} = \hat{\alpha}^{-2}\mathcal{A}_1$ where $\mathcal{A}_1$ is the term defined in the proof of proposition 3 with the only difference that α must be replaced by $\hat{\alpha}$. With reference to the notation of that proof we have:

$$\hat{\alpha}^{\frac{2b_0-1}{2(a+s)}-2} c_1 \le ||\tilde{B}(\hat{\alpha})(x_0 - x_*)||^2 \le I(b_0 \ge 2(a+s)) + \hat{\alpha}^{\frac{2b_0-1}{2(a+s)}-2}\tilde{c}_1.$$

Therefore,

$$\begin{aligned}
\delta^{\frac{2b_0-1}{2(b_0+a+\tilde{\eta})}} \mathbf{E}_{x_*}[o_p(1) + (1-\rho)\kappa_1]^{\frac{2b_0-1}{2(a+s)}-2}[\delta^{\tilde{p}\frac{b_0 - \tilde{b}_0 + \tilde{\eta}}{(a+\tilde{b}_0)}} - o_p(1) - (1-\rho)\kappa_1]^2 c_1 \le \mathbf{E}_{x_*}(\alpha_* - \hat{\alpha})^2\ ||\tilde{B}(\hat{\alpha})(x_0 - x_*)||^2 I_G \\
\le \left(\delta^{\frac{2(a+s)}{b_0+a+\tilde{\eta}}} I(b_0 \ge 2(a+s)) + \delta^{\frac{2b_0-1}{2(b_0+a+\tilde{\eta})}} \mathbf{E}_{x_*}[o_p(1) + (1-\rho)\kappa_1]^{\frac{2b_0-1}{2(a+s)}-2}\tilde{c}_1\right) \times \\
[\delta^{\tilde{p}\frac{b_0 - \tilde{b}_0 + \tilde{\eta}}{(a+\tilde{b}_0)}} - o_p(1) - (1-\rho)\kappa_1]^2,
\end{aligned}$$

that is, $\mathbf{E}_{x_*}(\alpha_* - \hat{\alpha})^2\ ||\tilde{B}(\hat{\alpha})(x_0 - x_*)||^2 I_G \asymp \delta^{\frac{2\tilde{b}_0-1}{2(b_0+a+\tilde{\eta})}}$ then, we conclude that

$$\mathbf{E}_{x_*}\left(||\hat{x}_{\hat{\alpha}} - x_*||^2 I_G\right) \le \mathbf{E}_{x_*}||\hat{x}_{\hat{\alpha}} - x_*||^2(2 + 4\mathcal{O}(1)) + 4\mathcal{O}\left(\delta^{\frac{2\tilde{b}_0-1}{2(b_0+a+\tilde{\eta})}}\right). \qquad (34)$$

Let us turn to the analysis of the variance term:

$$\begin{aligned}
trVar(x|y^\delta,\hat\alpha,s) &= \delta tr\left[\Omega_0(\hat\alpha I+B^*B)^{-1}\right]\\
&= \delta tr\left[\Omega_0(\alpha_* I+B^*B)^{-1}+\Omega_0[(\hat\alpha I+B^*B)^{-1}-(\alpha_* I+B^*B)^{-1}]\right]\\
&= \delta tr\left[\Omega_0(\alpha_* I+B^*B)^{-1}\right]+\delta tr\left[(\hat\alpha I+B^*B)^{-1}(\alpha_*-\hat\alpha)(\alpha_* I+B^*B)^{-1}\Omega_0\right]\\
&= \delta tr\left[\Omega_0(\alpha_* I+B^*B)^{-1}\right]+\delta\sum_{j=1}^{\infty}\frac{(\alpha_*-\hat\alpha)}{(\hat\alpha+j^{-2(a+s)})}\frac{j^{-2s}}{(\alpha_*+j^{-2(a+s)})}\\
&\leq \delta\left(1+\frac{(\alpha_*-\hat\alpha)}{\hat\alpha}\right)\sum_{j=1}^{\infty}\frac{j^{-2s}}{(\alpha_*+j^{-2(a+s)})}.
\end{aligned}$$

by using the expression for the trace of the variance in the proof of theorem 4. Therefore,

$$\mathbf{E}_{x_*}trVar(x|y^\delta,\hat\alpha,s)I_G=\mathbf{E}_{x_*}\left(\frac{\alpha_*}{\hat\alpha}I_G\right)\,tr[Var(x|y^\delta,\alpha_*,s)]$$

and by using (32) and the upper and lower bounds for $tr[Var(x|y^\delta,\alpha_*,s)]$ derived in the proof of theorem 4 we conclude that $\mathbf{E}_{x_*}trVar(x|y^\delta,\hat\alpha,s)I_G\asymp\mathbf{E}_{x_*}\left(\frac{\delta^{\tilde p\frac{b_0-\tilde b_0+\tilde\eta}{\tilde b_0+a}}}{o_p(1)+(1-\rho)\kappa_1}\right)\delta\alpha_*^{-\frac{2a+1}{2(a+s)}}$. Denote m_δ the rate of $P(G)$ given in Lemma 7, we conclude that:

$$\begin{aligned}
\mathbf{E}_{x_*}\left(\mu^Y_{\delta,\hat\alpha}\{||x-x_*||>\varepsilon_\delta M_\delta\}\right) &\leq \mathcal{O}\left(m_\delta\right)+\mathbf{E}_{x_*}||\hat x_{\alpha_*}-x_*||^2\frac{(2+4\mathcal{O}(1))}{\varepsilon_\delta^2M_\delta^2}\\
&\quad+\frac{4\mathcal{O}\left(\delta^{\frac{2\tilde b_0-1}{2(b_0+a+\tilde\eta)}}\right)}{\varepsilon_\delta^2M_\delta^2}+\mathbf{E}_{x_*}\left(\frac{\delta^{\tilde p\frac{b_0-\tilde b_0+\tilde\eta}{\tilde b_0+a}}}{o_p(1)+(1-\rho)\kappa_1}\right)\frac{\delta\alpha_*^{-\frac{2a+1}{2(a+s)}}}{\varepsilon_\delta^2M_\delta}^2\\
&= \mathcal{O}\left(\frac{1}{\varepsilon_\delta^2M_\delta^2}\left(\delta^{\frac{2\tilde b_0-1}{2(\tilde b_0+a)}}+\delta^{\frac{2\tilde b_0-1}{2(b_0+a+\tilde\eta)}}\right)\right)
\end{aligned}$$

which converges to 0 for $\varepsilon_\delta\asymp\delta^{\frac{2\tilde b_0-1}{4(b_0+a+\tilde\eta)}}$ since $m_\delta\to 0$ under the conditions of the theorem.

Lemma 5. *Let the assumptions of Theorem 6 be satisfied and* $\alpha_1=\left(\frac{\tilde c_3}{c_4}(1+\epsilon_\delta)\right)^{\frac{a+s}{b_0+a}}\delta^{\tilde p}$ *where:* $\tilde p=\frac{a+s}{b_0+a+\tilde\eta}$ *for* $\tilde\eta=\eta I\{b_0-a-2s-1/2>0\}$, $\eta>\max\{(b_0-2s-a-1),(b_0-2s-a-1/2)\frac{1}{4}\}$, $\epsilon_\delta=\delta^{\frac{1-r}{4(b_0+a)}}$, *for every* $0<r<\frac{4(a+s)(b_0+a)}{b_0+a+\tilde\eta}$ *and* $\tilde c_3$, c_4 *be as in Lemma 3. Then,*

$$P(S_{y^\delta}(\alpha_1)<0)\to 1\quad as\quad \delta\to 0.$$

D.3 Proof of Lemma 5

By using the notation of lemma 3 we write

$$\mathcal{S}_{y^\delta}(\alpha):=-\mathcal{S}_2-\mathcal{S}_3-\mathcal{S}_{4a}+(\mathcal{S}_1-\mathcal{S}_{4b})+\frac{\nu_1}{\alpha}+\nu_2$$

where $\mathcal{S}_3$ and $\mathcal{S}_{4a}$ are independent zero-mean Gaussian random variables with variances equal to

$$Var(\mathcal{S}_3) = \frac{1}{\delta}\sum_{j=1}^{\infty}\frac{\lambda_j^4}{(\alpha+\lambda_j^2)^4}j^{-2(a+b_0)} \quad \text{and} \quad Var(\mathcal{S}_{4a}) = \frac{1}{2}\sum_{j=1}^{\infty}\frac{\lambda_j^4}{(\alpha+\lambda_j^2)^4},$$

respectively. By using the lower bound of $\mathcal{S}_2$ and the upper bound of $(\mathcal{S}_1 - \mathcal{S}_{4b})$ provided in lemma 3 and by denoting $D(\alpha) = Var(\mathcal{S}_3) + Var(\mathcal{S}_4)$ we obtain:

$$\begin{aligned} P(S_{y^\delta}(\alpha_1) < 0) &= P(\mathcal{S}_3 + \mathcal{S}_{4a} > -\mathcal{S}_2 + (\mathcal{S}_1 - \mathcal{S}_{4b}) + \frac{\nu_1}{\alpha} + \nu_2) \\ &\geq P\left(\xi > \frac{-\frac{1}{2}\delta^{-1}\alpha_1^{\frac{b_0-s-1/2}{a+s}}c_4 + \nu_2 + \alpha_1^{-\frac{2(a+s)+1}{2(a+s)}}\frac{1}{2}(\tilde{c}_3 + \alpha_1^{\frac{1}{2(a+s)}}(1+2\nu_1))}{[D(\alpha_1)]^{1/2}}\right) \end{aligned} \tag{35}$$

where ξ denotes a $\mathcal{N}(0,1)$ random variable. Moreover, let $D^u(\alpha)$ be an upper bound for $D(\alpha)$ and $D_l(\alpha)$ be a lower bound for $D(\alpha)$ for every α. By the result in lemma 4 we can take

$$D^u(\alpha) = \delta^{-1}I(b_0 \geq a+2s) + \tilde{c}_5\delta^{-1}\alpha^{-\frac{a+2s-b_0+1/2}{a+s}} + \frac{17}{32}\alpha^{-\frac{4(a+s)+1}{2(a+s)}}\tilde{c}_6 \tag{36}$$

$$D_l(\alpha) = \delta^{-1}\alpha^{-\frac{a+2s-b_0+1/2}{a+s}}c_5 + \frac{1}{2}\alpha^{-\frac{4(a+s)+1}{2(a+s)}}c_6 \tag{37}$$

and by replacing the value of α_1 and after some algebras we get:

$$\begin{aligned} D^u(\alpha_1) &= \delta^{-1}I(b_0 \geq a+2s) + \tilde{c}_5\delta^{-1}\delta^{-\frac{a+2s-b_0+1/2}{b_0+a+\tilde{\eta}}}\left[\frac{\tilde{c}_3}{c_4}(1+\epsilon_\delta)\right]^{-\frac{a+2s-b_0+1/2}{b_0+a}} + \frac{17\tilde{c}_6}{32}\delta^{-\frac{4(a+s)+1}{2(b_0+a+\tilde{\eta})}}\left[\frac{\tilde{c}_3}{c_4}(1+\epsilon_\delta)\right]^{-\frac{4(a+s)+1}{2(b_0+a)}}. \\ D_l(\alpha_1) &= \delta^{-1}\delta^{-\frac{a+2s-b_0+1/2}{b_0+a+\tilde{\eta}}}\left(\left[\frac{\tilde{c}_3}{c_4}(1+\epsilon_\delta)\right]^{-\frac{a+2s-b_0+1/2}{b_0+a}}c_5 + \frac{1}{2}\delta^{\frac{\tilde{\eta}}{(b_0+a+\tilde{\eta})}}\left[\frac{\tilde{c}_3}{c_4}(1+\epsilon_\delta)\right]^{-\frac{2(a+s)+1/2}{b_0+a}}c_6\right). \end{aligned}$$

Remark that: (1) $D^u(\alpha_1) = \mathcal{O}(\delta^{-\frac{4(a+s)+1}{2(b_0+a)}})$ when $b_0 \leq \frac{1}{2} + a + 2s$ and (2) $D^u(\alpha_1) = \mathcal{O}(\delta^{-1})$ when $b_0 > \frac{1}{2} + a + 2s$. Therefore, we analyze these two cases separately.

CASE I: $b_0 \leq \frac{1}{2} + a + 2s$. By substituting the value of α_1 in the numerator of (35), factorizing the term $\delta^{-\frac{2(a+s)+1}{2(b_0+a)}}$ (in the first and second term of the numerator) and after some algebra we obtain

$$\begin{aligned} P(S_{y^\delta}(\alpha_1) < 0) &\geq P\Big(\xi > \frac{-\frac{1}{2}\delta^{-\frac{2(a+s)+1}{2(b_0+a)}}(\tilde{c}_3^{\frac{b_0-s-1/2}{b_0+a}}c_4^{\frac{2(a+s)+1}{2(b_0+a)}})\epsilon_\delta(1+\epsilon_\delta)^{-\frac{s+a+1/2}{b_0+a}} + \nu_2}{[D(\alpha_1)]^{1/2}} + \\ &\qquad \frac{\frac{1}{2}\delta^{-\frac{a+s}{b_0+a}}\tilde{c}_3^{-\frac{a+s}{(b_0+a)}}c_4^{\frac{a+s}{b_0+a}}(1+\varepsilon_\delta)^{-\frac{a+s}{b_0+a}}(1+2\nu_1)}{[D(\alpha_1)]^{1/2}}\Big) =: P\left(\xi > \frac{N_1}{[D(\alpha_1)]^{\frac{1}{2}}} + \frac{N_2}{[D(\alpha_1)]^{\frac{1}{2}}}\right) \\ &\geq P\left(\xi > \frac{N_1}{[D^u(\alpha_1)]^{\frac{1}{2}}} + \frac{N_2}{[D_l(\alpha_1)]^{\frac{1}{2}}}\right) = 1 - \Phi\left(\frac{N_1}{[D^u(\alpha_1)]^{\frac{1}{2}}} + \frac{N_2}{[D_l(\alpha_1)]^{\frac{1}{2}}}\right) \end{aligned}$$

since $N_1 < 0$ and $N_2 > 0$, where $\Phi(\cdot)$ denotes the cumulative distribution function of a $\mathcal{N}(0,1)$ distribution. Therefore, $\frac{N_1}{[D^u(\alpha_1)]^{\frac{1}{2}}} + \frac{N_2}{[D_l(\alpha_1)]^{\frac{1}{2}}} \asymp -\delta^{-\frac{1}{4(b_0+a)}}\epsilon_\delta + \delta^{\frac{1}{4(b_0+a)}}$, which converges to $-\infty$ as $\delta \to 0$ if we choose $\epsilon_\delta = \delta^{\frac{1-r}{4(b_0+a)}}$, for every $r > 0$.

CASE II: $b_0 > \frac{1}{2} + a + 2s$. So, $\tilde{\eta} = \eta$. By substituting the value of α_1 in the numerator of (35) we obtain

$$\begin{aligned} P(S_{y^\delta}(\alpha_1) < 0) \quad &\geq \quad P\Big(\xi > \frac{-\frac{1}{2}\delta^{-\frac{(a+s+1/2+\eta)}{b_0+a+\eta}} \left[\tilde{c}_3(1+\epsilon_\delta)\right]^{\frac{b_0-s-1/2}{b_0+a}} c_4^{\frac{a+s+1/2}{b_0+a}} + \nu_2}{[D(\alpha_1)]^{1/2}} + \\ &\qquad \frac{\frac{1}{2}\delta^{-\frac{a+s+1/2}{b_0+a+\eta}} \left[\frac{\tilde{c}_3}{c_4}(1+\epsilon_\delta)\right]^{-\frac{a+s+1/2}{b_0+a+1/2}} \left(\tilde{c}_3 + \delta^{\frac{1}{2(b_0+a+\eta)}} \left[\frac{\tilde{c}_3}{c_4}(1+\epsilon_\delta)\right]^{\frac{1}{2(b_0+a)}}\right)(1+2\nu_1)}{[D(\alpha_1)]^{1/2}}\Big) \\ &=: \quad P\left(\xi > \frac{\tilde{N}_1}{[D(\alpha_1)]^{\frac{1}{2}}} + \frac{\tilde{N}_2}{[D(\alpha_1)]^{\frac{1}{2}}}\right) \\ &\geq \quad P\left(\xi > \frac{\tilde{N}_1}{[D^u(\alpha_1)]^{\frac{1}{2}}} + \frac{\tilde{N}_2}{[D_l(\alpha_1)]^{\frac{1}{2}}}\right) = 1 - \Phi\left(\frac{\tilde{N}_1}{[D^u(\alpha_1)]^{\frac{1}{2}}} + \frac{\tilde{N}_2}{[D_l(\alpha_1)]^{\frac{1}{2}}}\right) \end{aligned}$$

since $\tilde{N}_1 < 0$ and $\tilde{N}_2 > 0$, where $\Phi(\cdot)$ denotes the cumulative distribution function of a $\mathcal{N}(0,1)$ distribution. Remark that in this case we can rewrite $D_l(\alpha_1)$ as

$$D_l(\alpha_1) = \delta^{-1}\delta^{-\frac{a+2s-b_0+1/2}{2(b_0+a+\eta)}} \left(\left[\frac{\tilde{c}_3}{c_4}(1+\epsilon_\delta)\right]^{-\frac{a+2s-b_0+1/2}{b_0+a}} c_5 + \frac{1}{2}\delta^{1-\frac{b_0+a}{b_0+a+\eta}} \left[\frac{\tilde{c}_3}{c_4}(1+\epsilon_\delta)\right]^{-\frac{2(a+s)+1/2}{b_0+a}} c_6\right)$$

and therefore, $\frac{\tilde{N}_1}{[D^u(\alpha_1)]^{\frac{1}{2}}} + \frac{\tilde{N}_2}{[D_l(\alpha_1)]^{\frac{1}{2}}} \asymp -\delta^{-\frac{2s+a-b_0+1+\eta}{2(b_0+a+\eta)}} + \delta^{\frac{b_0-a-2s-3/2+2\eta}{4(b_0+a+\eta)}} = -\delta^{-\frac{2s+a-b_0+1+\eta}{2(b_0+a+\eta)}}(1-o(1))$ if $\tilde{\eta}$ is sufficiently big so that $\eta > \frac{1}{4}(b_0 - 2s - a - \frac{1}{2})$. This quantity converges to $-\infty$ if $\eta > b_0 - 2s - a - 1$. Therefore, the condition which guarantees convergence is: $\eta > \max\{(b_0 - 2s - a - 1), (b_0 - 2s - a - 1/2)\frac{1}{4}\}$.

Lemma 6. *Let the assumptions of Theorem 6 be satisfied and* $\alpha_2 = \left(\frac{\tilde{c}_3}{\tilde{c}_4}(1-\epsilon_\delta)\right)^{\frac{a+s}{b_0+a}} \delta^p$ *where:* $p = \frac{a+s}{b_0+a-\eta}$ *for* $(b_0 + a) > \eta > \max\{b_0 - s - 1/2, 0\}$, $0 < \epsilon_\delta < 1$, $\tilde{c}_3$, $\tilde{c}_4$ *be as defined in Lemma 3. Then,*

$$P(S_{y^\delta}(\alpha_2) > 0) \to 1 \quad \textit{as} \quad \delta \to 0.$$

D.4 Proof of Lemma 6

This proof follows the line of the proof of Lemma 5, so some details are omitted. By using the upper bound of $\mathcal{S}_2$ and the lower bound of $(\mathcal{S}_1 - \mathcal{S}_{4b})$ provided in Lemma 3 we obtain:

$$\begin{aligned} P(S_{y^\delta}(\alpha_2) > 0) \quad &= \quad P(\mathcal{S}_3 + \mathcal{S}_{4a} < -\mathcal{S}_2 + (\mathcal{S}_1 - \mathcal{S}_{4b}) + \frac{\nu_1}{\alpha} + \nu_2) \\ &\geq \quad P\left(\xi < \frac{-\frac{1}{2}\delta^{-1}\alpha_2^{\frac{b_0-s-1/2}{a+s}}\tilde{c}_4 + \nu_2 + \frac{\nu_1}{\alpha_2} + \alpha_2^{-\frac{2(a+s)+1}{2(a+s)}}\frac{1}{2}\tilde{c}_3 - \frac{1}{2}\delta^{-1}I(b_0 \geq s)}{[D(\alpha_2)]^{1/2}}\right) \end{aligned} \quad (38)$$

where ξ denotes a $\mathcal{N}(0,1)$ random variable. Moreover, $\forall\alpha$ let $D^u(\alpha)$ (resp. $D_l(\alpha)$) denote the upper (resp. the lower) bound for $D(\alpha)$ defined in (36)-(37). By replacing the value of α_2 and after some algebras we get:

$$\begin{aligned} D^u(\alpha_2) &= \delta^{-1}I(b_0 \geq a+2s) + \tilde{c}_5\delta^{-1-\frac{a+2s-b_0+1/2}{b_0+a-\eta}}\left[\frac{\tilde{c}_3}{\tilde{c}_4}(1-\epsilon_\delta)\right]^{-\frac{a+2s-b_0+1/2}{b_0+a}} + \frac{17\tilde{c}_6}{32}\delta^{-\frac{4(a+s)+1}{2(b_0+a-\eta)}}\left[\frac{\tilde{c}_3}{\tilde{c}_4}(1-\epsilon_\delta)\right]^{-\frac{4(a+s)+1}{2(b_0+a)}} \\ D_l(\alpha_2) &= \delta^{-\frac{2a+2s+1/2}{b_0+a-\eta}}\left[\frac{1}{2}c_6\left[\frac{\tilde{c}_3}{\tilde{c}_4}(1-\epsilon_\delta)\right]^{-\frac{2(a+s)+1/2}{b_0+a}} + \delta^{-1+\frac{b_0+a}{b_0+a-\eta}}\left[\frac{\tilde{c}_3}{\tilde{c}_4}(1-\epsilon_\delta)\right]^{-\frac{a+2s-b_0+1/2}{b_0+a}} c_5\right]. \end{aligned}$$

Remark that $D^u(\alpha_2) = \mathcal{O}(\delta^{-\frac{2(a+s)+1/2}{(b_0+a-\eta)}})$ if $b_0 \leq \frac{1}{2}+a+2s+\eta$. This condition is satisfied by assumption because $\eta > b_0 - s - \frac{1}{2} > b_0 - a - 2s - \frac{1}{2}$

By substituting the value of α_2 in the numerator and after some algebra we obtain

$$\begin{aligned} P(S_{y^\delta}(\alpha_2) > 0) &\geq P\Bigg(\xi < \frac{\frac{1}{2}\delta^{-\frac{a+s+1/2}{b_0+a-\eta}}\left[\frac{\tilde{c}_3}{\tilde{c}_4}(1-\epsilon_\delta)\right]^{-\frac{a+s+1/2}{b_0+a}}\tilde{c}_3 + \nu_1\left[\frac{\tilde{c}_3}{\tilde{c}_4}(1-\epsilon_\delta)\right]^{-\frac{a+s}{b_0+a}}\delta^{-\frac{a+s}{b_0+a-\eta}}}{[D(\alpha_2)]^{1/2}} + \\ &\qquad \frac{-\frac{1}{2}\delta^{-1}\left[I(b_0\geq s) + \tilde{c}_4\delta^{\frac{b_0-s-1/2}{b_0+a-\eta}}\left[\frac{\tilde{c}_3}{\tilde{c}_4}(1-\epsilon_\delta)\right]^{\frac{b_0-s-1/2}{b_0+a}}\right] + \nu_2}{[D(\alpha_2)]^{1/2}}\Bigg) \\ &=: P\left(\xi < \frac{N_3}{[D(\alpha_2)]^{\frac{1}{2}}} + \frac{N_4}{[D(\alpha_2)]^{\frac{1}{2}}}\right) \\ &\geq P\left(\xi < \frac{N_3}{[D^u(\alpha_2)]^{\frac{1}{2}}} + \frac{N_4}{[D_l(\alpha_2)]^{\frac{1}{2}}}\right) = \Phi\left(\frac{N_3}{[D^u(\alpha_2)]^{\frac{1}{2}}} + \frac{N_4}{[D_l(\alpha_2)]^{\frac{1}{2}}}\right) \end{aligned}$$

since $N_3 > 0$ and $N_4 < 0$. Term N_3 converges to $+\infty$, as $\delta \to 0$, since

$$\frac{N_3}{[D^u(\alpha_2)]^{\frac{1}{2}}} = \frac{\frac{1}{2}\delta^{-\frac{a+s+1/2}{b_0+a-\eta}+\frac{2a+2s+1/2}{2(b_0+a-\eta)}}\left[\frac{\tilde{c}_3}{\tilde{c}_4}(1-\epsilon_\delta)\right]^{-\frac{a+s+1/2}{b_0+a}}\left(\tilde{c}_3 + 2\left[\frac{\tilde{c}_3}{\tilde{c}_4}(1-\epsilon_\delta)\right]^{\frac{a+s+1/2}{b_0+a}}\delta^{\frac{1}{2(b_0+a-\eta)}}\right)}{\left[o(1)I(b_0\geq a+2s) + \left[\frac{\tilde{c}_3}{\tilde{c}_4}(1-\epsilon_\delta)\right]^{-\frac{2(a+s)+1/2}{b_0+a}}(\frac{17\tilde{c}_6}{32}+o(1))\right]^{1/2}} \asymp \delta^{-\frac{1}{4(b_0+a-\eta)}}.$$

The asymptotic behavior of N_4 is different depending on the sign of $(b_0 - s - 1/2)$. So, we treat the two cases separately.

CASE I: $b_0 < s + 1/2$.

$$\begin{aligned} \frac{N_4}{[D_l(\alpha_2)]^{\frac{1}{2}}} &= \frac{-\frac{1}{2}\delta^{\frac{2(a+s)+1/2}{2(b_0+a-\eta)}}\left(\delta^{-1+\frac{b_0-s-1/2}{b_0+a-\eta}}\left[\left[\frac{\tilde{c}_3}{\tilde{c}_4}(1-\epsilon_\delta)\right]^{\frac{b_0-s-1/2}{b_0+a}}\tilde{c}_4 + \delta^{-\frac{(b_0-s-1/2)}{b_0+a-\eta}}I(b_0\geq s)\right] - \nu_2\right)}{\left[o(1) + \left[\frac{\tilde{c}_3}{\tilde{c}_4}(1-\epsilon_\delta)\right]^{-\frac{2(a+s)+1/2}{b_0+a}}\frac{c_6}{2}\right]^{1/2}} \\ &\asymp -\delta^{-1+\frac{2a+2b_0-1/2}{2(b_0+a-\eta)}} \end{aligned}$$

which is bounded if $\eta \geq 1/4$. If $\eta < 1/4$ then $\frac{N_4}{[D_l(\alpha_2)]^{\frac{1}{2}}} \to -\infty$ but slower than $\frac{N_3}{[D^u(\alpha_2)]^{\frac{1}{2}}} \to \infty$.

CASE II: $b_0 \geq s + 1/2$.

$$\frac{N_4}{[D_l(\alpha_2)]^{\frac{1}{2}}} = \frac{-\frac{1}{2}\left(\delta^{-1+\frac{2(a+s)+1/2}{2(b_0+a-\eta)}}\left[I(b_0 \geq s) + \delta^{\frac{b_0-s-1/2}{b_0+a-\eta}}\left[\frac{\tilde{c}_3}{\tilde{c}_4}(1-\epsilon_\delta)\right]^{\frac{b_0-s-1/2}{b_0+a}}\tilde{c}_4 - \nu_2\delta\right]\right)}{\left[o(1) + \left[\frac{\tilde{c}_3}{\tilde{c}_4}(1-\epsilon_\delta)\right]^{-\frac{2(a+s)+1/2}{b_0+a}}\frac{c_6}{2}\right]^{1/2}} \asymp -\delta^{-1+\frac{2(a+s)+1/2}{2(b_0+a-\eta)}}$$

which converges to $-\infty$ if $b_0 > s+\eta+1/4$. However, in this case, $\frac{N_4}{[D_l(\alpha_2)]^{\frac{1}{2}}} \to -\infty$ slower than $\frac{N_3}{[D^u(\alpha_3)]^{\frac{1}{2}}} \to \infty$ if $b_0 < s + 1/2 + \eta$.

Lemma 7. *Let the assumptions of Theorem 6 be satisfied and α_1, α_2, $\tilde{N}_1$, $\tilde{N}_2$, N_3, N_4, D^u and D_l be defined as in lemmas 5 and 6. Let $\tilde{\iota}_4 = I\left(\frac{N_3}{[D^u(\alpha_2)]^{1/2}} + \frac{N_4}{[D_l(\alpha_2)]^{1/2}} < 0\right)$ and $\tilde{\iota}_2 = I\left(\frac{\tilde{N}_1}{[D^u(\alpha_1)]^{1/2}} + \frac{\tilde{N}_2}{[D_l(\alpha_1)]^{1/2}} > 0\right)$. Then,*

$$\begin{aligned} P\left(\hat{\alpha} \notin (\alpha_2, \alpha_1)\right) &= \mathcal{O}\Big(\exp\{-\delta^{-\frac{1}{2(b_0+a-\eta)}}(1 + o(1)(1-\tilde{\iota}_4))\} - \delta^{\frac{\eta-b_0-1/4+(b_0\wedge s+1/2)}{b_0+a-\eta}}\tilde{\iota}_4 + \\ &\quad \left(\exp\{-\delta^{-\frac{2s+a-b_0+\eta+1}{b_0+a+\eta}}(1+o(1)(1-\tilde{\iota}_2))\} - \delta^{\frac{b_0-2s-a+2\eta-3/2}{4(b_0+a+\eta)}}\tilde{\iota}_2\right) I\left(b_0 > a + 2s + \frac{1}{2}\right)\Big). \end{aligned}$$

D.5 Proof of Lemma 7

The notation that we use in this proof has been defined in lemmas 5 and 6. We upper bound $P\left(\hat{\alpha} \notin (\alpha_2, \alpha_1)\right)$ by the probability of $\{\hat{\alpha} \notin (\alpha_2, \alpha_1)\} \cap \{S_{y^\delta}(\hat{\alpha}) = 0\}$ which is equal to the probability of

$$\begin{aligned} &\Big(\{S_{y^\delta}(\alpha_1) < 0\} \cap \{S_{y^\delta}(\alpha_2) < 0\} \cap \{S_{y^\delta}(\alpha) < 0,\ \forall \alpha \in (\alpha_2, \alpha_1)\}\Big) \cup \\ &\qquad \Big(\{S_{y^\delta}(\alpha_1) > 0\} \cap \{S_{y^\delta}(\alpha_2) > 0\} \cap \{S_{y^\delta}(\alpha) > 0,\ \forall \alpha \in (\alpha_2, \alpha_1)\}\Big). \end{aligned}$$

This probability is upper bounded by

$$\begin{aligned} & P\Big(\{S_{y^\delta}(\alpha_1) < 0\} \cap \{S_{y^\delta}(\alpha_2) < 0\} \cup \{S_{y^\delta}(\alpha_1) > 0\} \cap \{S_{y^\delta}(\alpha_2) > 0\}\Big) \\ = \; & P\left(\{S_{y^\delta}(\alpha_1) < 0 \cap S_{y^\delta}(\alpha_2) < 0\}\right) + P\left(\{S_{y^\delta}(\alpha_1) > 0 \cap S_{y^\delta}(\alpha_2) > 0\}\right) \\ \leq \; & P\left(\{S_{y^\delta}(\alpha_2) < 0\}\right) + P\left(\{S_{y^\delta}(\alpha_1) > 0\}\right) \\ \leq \; & \Phi\left(-\frac{N_3}{[D^u(\alpha_2)]^{\frac{1}{2}}} - \frac{N_4}{[D_l(\alpha_2)]^{\frac{1}{2}}}\right) + \Phi\left(\frac{N_1}{[D^u(\alpha_1)]^{\frac{1}{2}}} + \frac{N_2}{[D_l(\alpha_1)]^{\frac{1}{2}}}\right) I\left(b_0 \leq a + 2s + \frac{1}{2}\right) \\ & \qquad + \Phi\left(\frac{\tilde{N}_1}{[D^u(\alpha_1)]^{\frac{1}{2}}} + \frac{\tilde{N}_2}{[D_l(\alpha_1)]^{\frac{1}{2}}}\right) I\left(b_0 > a + 2s + \frac{1}{2}\right). \end{aligned} \tag{39}$$

Let us start with the first term and denote $\tilde{\iota}_4 = I\left(\frac{N_3}{[D^u(\alpha_2)]^{1/2}} + \frac{N_4}{[D_l(\alpha_2)]^{1/2}} < 0\right)$. Since $\frac{N_3}{[D^u(\alpha_2)]^{\frac{1}{2}}} > 0$ and $\frac{N_4}{[D_l(\alpha_2)]^{\frac{1}{2}}} < 0$ and since $2\Phi(a) \leq e^{-a^2/2}$, $\forall a < 0$:

$$\begin{aligned}
\Phi\left(-\frac{N_3}{[D^u(\alpha_2)]^{\frac{1}{2}}} - \frac{N_4}{[D_l(\alpha_2)]^{\frac{1}{2}}}\right) &= \Phi\left(-\frac{N_3}{[D^u(\alpha_2)]^{\frac{1}{2}}} - \frac{N_4(1-\tilde{\iota}_4)}{[D_l(\alpha_2)]^{1/2}}\right) + \int_{-\frac{N_3}{[D^u(\alpha_2)]^{1/2}}}^{-\frac{N_3}{[D^u(\alpha_2)]^{1/2}} - \frac{N_4\tilde{\iota}_4}{[D_l(\alpha_2)]^{1/2}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \\
&\leq \frac{1}{2}\exp\left\{-\frac{1}{2}\left(-\frac{N_3}{[D^u(\alpha_2)]^{1/2}} - \frac{N_4(1-\tilde{\iota}_4)}{[D_l(\alpha_2)]^{1/2}}\right)^2\right\} + e^{-\frac{u^2}{2}}\Big|_{u=0}\left(-\frac{N_4\tilde{\iota}_4}{[D_l(\alpha_2)]^{1/2}}\right) \\
&\leq \frac{1}{2}\exp\left\{-\frac{1}{2}\left(\frac{N_3^2}{D^u(\alpha_2)} + 2\frac{N_3N_4(1-\tilde{\iota}_4)}{[D^u(\alpha_2)D_l(\alpha_2)]^{1/2}}\right)\right\} - \frac{N_4\tilde{\iota}_4}{[D_l(\alpha_2)]^{1/2}} \\
&= \mathcal{O}\left(\exp\{-\delta^{-\frac{1}{2(b_0+a-\eta)}}(1+o(1)(1-\tilde{\iota}_4))\} - \delta^{\frac{\eta-b_0-1/4+(b_0\wedge s+1/2)}{b_0+a-\eta}}\tilde{\iota}_4\right)
\end{aligned}$$

where the $o(1)$ term converges at a different rate depending whether $b_0 \geq s+1/2$ or $b_0 < s+1/2$. The inequality in the third line is due to the fact that for two constants $a_1 > 0$ and $a_2 < 0$: $(-a_1-a_2)^2 \geq a_1^2 + 2a_1a_2$. By denoting $\tilde{\iota}_2 = I\left(\frac{\tilde{N}_1}{[D^u(\alpha_1)]^{1/2}} + \frac{\tilde{N}_2}{[D_l(\alpha_1)]^{1/2}} > 0\right)$, we can use a similar reasoning for the third term of (39) (since $\frac{\tilde{N}_1}{[D^u(\alpha_1)]^{1/2}} < 0$ and $\frac{\tilde{N}_2}{[D_l(\alpha_1)]^{1/2}} > 0$):

$$\begin{aligned}
\Phi\left(\frac{\tilde{N}_1}{[D^u(\alpha_1)]^{\frac{1}{2}}} + \frac{\tilde{N}_2}{[D_l(\alpha_1)]^{\frac{1}{2}}}\right) &= \Phi\left(-\frac{\tilde{N}_1}{[D^u(\alpha_1)]^{\frac{1}{2}}} - \frac{\tilde{N}_2(1-\tilde{\iota}_2)}{[D_l(\alpha_1)]^{1/2}}\right) + \int_{\frac{\tilde{N}_1}{[D^u(\alpha_1)]^{1/2}}}^{\frac{\tilde{N}_1}{[D^u(\alpha_1)]^{1/2}} - \frac{\tilde{N}_2\tilde{\iota}_2}{[D_l(\alpha_1)]^{1/2}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \\
&\leq \frac{1}{2}\exp\left\{-\frac{1}{2}\left(\frac{-\tilde{N}_1}{[D^u(\alpha_1)]^{1/2}} - \frac{\tilde{N}_2(1-\tilde{\iota}_2)}{[D_l(\alpha_1)]^{1/2}}\right)^2\right\} + e^{-\frac{u^2}{2}}\Big|_{u=0}\left(-\frac{\tilde{N}_2\tilde{\iota}_2}{[D_l(\alpha_1)]^{1/2}}\right) \\
&\leq \frac{1}{2}\exp\left\{-\frac{1}{2}\left(\frac{\tilde{N}_1^2}{D^u(\alpha_1)} + 2\frac{\tilde{N}_1\tilde{N}_2(1-\tilde{\iota}_2)}{[D^u(\alpha_1)D_l(\alpha_1)]^{1/2}}\right)\right\} - \frac{\tilde{N}_2\tilde{\iota}_2}{[D_l(\alpha_1)]^{1/2}} \\
&= \mathcal{O}\left(\exp\{-\delta^{-\frac{2s+a-b_0+\eta+1}{b_0+a+\eta}}(1+o(1)(1-\tilde{\iota}_2))\} - \delta^{\frac{b_0-2s-a+2\eta-3/2}{4(b_0+a+\eta)}}\tilde{\iota}_2\right).
\end{aligned}$$

To analyze the second term of (39) we use a Taylor expansion of $\Phi(a+b)$ around a, for $a = \frac{N_1}{[D^u(\alpha_1)]^{\frac{1}{2}}}$ and $b = \frac{N_2}{[D_l(\alpha_1)]^{\frac{1}{2}}}$: $\Phi(a+b) = \Phi(a) + \phi(a)b - \frac{\phi(a)ab^2}{2} + \mathcal{O}(b^3)$, where $\phi(\cdot)$ denotes the density function of a standard Normal distribution. Hence,

$$\Phi\left(\frac{N_1}{[D^u(\alpha_1)]^{\frac{1}{2}}} + \frac{N_2}{[D_l(\alpha_1)]^{\frac{1}{2}}}\right) = \mathcal{O}\left(\exp\{-\delta^{-\frac{1}{b_0+a}}\}(1+o(1))\right)$$

since $b = o(1)$. By retaining only the non-negligible terms we obtain the result.

References

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

Agapiou, S., Larsson, S. and A.M., Stuart (2012), Posterior contraction rates for the Bayesian approach to linear ill-posed inverse problems, arXiv:1203.5753.

Bissantz, N., Hohage, T., Munk,A. and F., Ruymgaart (2007), Convergence rates of general regularization methods for statistical inverse problems and applications, *SIAM J. Numer. Anal.*, **45**, 2610-2636.

Blundell, R., Chen, X. and D., Kristensen (2007), Semi-nonparamtric IV estimation of shape-invariant engel curves, *Econometrica*, **75**, 1613-1669.

Carrasco, M. and J.P., Florens (2000), Generalization of GMM to a continuum of moment conditions, *Econometric Theory*, **16**, 797-834.

Carrasco, M., Florens, J.P., and E., Renault (2005), Estimation based on spectral decomposition and regularization, *Handbook of Econometrics*, J.J. Heckman and E. Leamer, eds., **6**, Elsevier, North Holland.

Cavalier, L. and A.B., Tsybakov (2002), Sharp adaptation for inverse problems with random noise, *Proba. Theory and Related Fields*, **123**, 323–354.

Chen, X. and D., Pouzo (2012), Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals, *Econometrica*, **80**, 277-321.

Chen, X. and M., Reiss, (2011), On rate optimality for ill-posed inverse problems in econometrics, *Econometric Theory*, **27**, Special Issue 03, 497-521.

Daouia, A., Florens, J-P. and L. Simar (2009), Regularization of nonparametric frontier estimators, *Journal of Econometrics*, **168**, 285-299.

Darolles, S., Fan, Y., Florens, J.P., and E., Renault (2011), Nonparametric instrumental regression, *Econometrica*, **79**, 1541-1565.

Diaconis, P.W., and D., Freedman (1986), On the consistency of Bayes estimates, *Annals of Statistics*, **14**, 1-26.

Diaconis, P.W. and D., Freedman (1998), Consistency of Bayes estimates for nonparametric regression: normal theory, *Bernoulli*, **4**, 411-444.

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

Florens, J-P. (2003), Inverse problems and structural econometrics: the example of instrumental variables. Invited Lectures to the World Congress of the Econometric Society, Seattle 2000. In: M., Dewatripont, L.-P., Hansen, and S.J., Turnovsky, (Eds.), *Advances in Economics end econometrics: theory and applications*, Vol.II, 284-311. Cambridge University Press.

Florens, J-P., Johannes, J. and S., Van Bellegem (2011), Identification and estimation by penalization in nonparametric instrumental regression, *Econometric Theory*, **27**, 472-496.

Florens, J-P. and E. Sbaï (2010), Local identification in empirical games of incomplete information, *Econometric Theory*, **26**, 1638-1662.

Florens, J.P. and A., Simoni (2010), Regularizing priors for linear inverse problems, IDEI Working paper, n. 621.

Florens, J.P. and A. Simoni (2012a), Nonparametric estimation of an instrumental regression: a quasi-Bayesian approach based on regularized posterior, *Journal of Econometric*, **170**, 458-475.

Florens, J.P. and A., Simoni (2012b), Regularized posteriors in linear ill-posed inverse problems, *Scandinavian Journal of Statistics*, **39**, 214-235.

Gautier, E. and Y. Kitamura (2012), Nonparametric estimation in random coefficients binary choice models, *Econometrica*, forthcoming.

Ghosh, J.K. and R.V., Ramamoorthi (2003), Bayesian nonparametrics, *Springer Series in Statistics*.

Goldenshluger, A. and S.V., Pereverzev (2003), On adaptive inverse estimation of linear functionals in Hilbert scales, *Bernoulli*.

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

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

Helin, T. (2009), On infinite-dimensional hierarchical probability models in statistical inverse problems, *Inverse Probl. Imaging*, **3**, pp. 567-597.

Hoderlein, S., Nesheim, L. and A. Simoni (2012), Semiparametric estimation of random coefficients in structural economic models, CEMMAP working paper, CWP09/12.

Hofinger, A. and H.K., Pikkarainen (2007), Convergence rate for the Bayesian approach to linear inverse problems, *Inverse Problems*, **vol. 23**, 2469-2484.

Hofinger, A. and H.K., Pikkarainen (2009), Convergence rate for linear inverse problems in the presence of an additive normal noise, *Stochastic Analysis and Applications*, **vol. 27**, 240-257.

Ito, K. (1970), The topological support of Gauss measure on Hilbert space, *Nagoya Math. J.*, **38**, 181 - 183.

Johannes, J. (2008), Nonparametric estimation in functional linear model, in: *Functional and Operatorial Statistics*, S. Dabo-Niang and F. Ferraty, editors, 215-221, Physica-Verlag.

Johannes, J., Van Bellegem, S. and A., Vanhems (2011), Convergence rates for ill-posed inverse problems with an unknown operator, *Econometric Theory*, **27**, 522-545.

Kaipio, J., and E., Somersalo (2004), Statistical and computational inverse problems, *Applied Mathematical Series, vol.160, Springer, Berlin.*

Kato, T. (1995), Perturbation theory for linear operators, Springer.

Knapik, B.T., Van der Vaart, A.W. and J.H., Van Zanten (2011), Bayesian inverse problems, *Annals of Statistics*, **39**, 2626-2657.

Krein, S.G. and J.I., Petunin (1966), Scales of Banach spaces, *Russian Math. Surveys*, **21**, 85 - 160.

Kress, R. (1999), Linear integral equation, Springer.

Kuo, H.H. (1975), Gaussian measures in Banach spaces, Springer.

Lassas, M., Saksman, E. and S., Siltanen (2009), Discretization invariant Bayesian inversion and Besov space priors', *Inverse Probl. Imaging*, **3**, pp.87-122.

Lehtinen, M.S., Päivärinta, L. and E., Somersalo (1989), Linear inverse problems for generalised random variables, *Inverse Problems*, **5**, 599-612.

Luschgy, H. (1995), Linear estimators and radonifying operators, *Theory Probab. Appl.*, **40**, 167-175.

Mandelbaum, A. (1984), Linear estimators and measurable linear transformations on a Hilbert space, *Z. Wahrcheinlichkeitstheorie*, 3, 385-98.

Nair, M.T., Pereverzev, S.V. and U., Tautenhahn (2005), Regularization in Hilbert scales under general smoothing conditions, *Inverse Problems*, **21**, 1851-1969.

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

Newey, W.K. and J.L., Powell (2003), Instrumental variable estimation of nonparametric models. Econometrica, Vol.71, **5**, 1565-1578.

Neubauer, A. and H.K., Pikkarainen (2008), Convergence results for the Bayesian inversion theory, *J. Inverse Ill-posed Probl.*, **Vol. 16**, 601-613.

Nussbaum, M. (1996), Asymptotic equivalence of density estimation and gaussian white noise, *Annals of Statistics*, **24**, No.6, 2399–2430.

Ruymgaart, F.H. (1998), A note on weak convergence of density in hilbert spaces, *Statistics*, **30**, 331-343.

Tsybakov, A.B. (2004), Introduction à l'estimation non-paramétrique, Springer.

Van der Vaart, A.W. and J.A., Wellner (1996), Weak convergence and empirical processes, Springer, New York.

Van Rooij, A.C.M. and F.H., Ruymgaart (1999), On inverse estimation, in: Ghosh, S. (Ed) *Asymptotic, Nonparametric and Time Series*, 579-613 (Dekker).

Zellner, A. (1986), On Assessing Prior distributions and Bayesian regression analysis with g-prior distribution, in: Goel, P.K. and Zellner, A. (Eds) *Bayesian Inference and Decision Techniques: essays in honour of Bruno de Finetti*, pp. 233-243 (Amsterdam, North Holland).