Journal of Econometrics 170 (2012) 458–475

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Nonparametric estimation of an instrumental regression: A quasi-Bayesian approach based on regularized posterior☆

Jean-Pierre Florens [a], Anna Simoni [b,*]

[a] Toulouse School of Economics, Université de Toulouse 1 - Capitole, 21, Allée de Brienne, 31000 Toulouse, France

[b] Department of Decision Sciences and IGIER, Università Bocconi, Via Sarfatti, 25 - 20136 Milano, Italy

ARTICLE INFO

Article history:
Available online 2 June 2012

JEL classification:
C11
C14
C30

Keywords:
Instrumental regression
Nonparametric estimation
Posterior distribution
Tikhonov regularization
Posterior consistency

ABSTRACT

We propose a quasi-Bayesian nonparametric approach to estimating the structural relationship φ among endogenous variables when instruments are available. We show that the posterior distribution of φ is inconsistent in the frequentist sense. We interpret this fact as the ill-posedness of the Bayesian inverse problem defined by the relation that characterizes the structural function φ. To solve this problem, we construct a *regularized posterior distribution*, based on a Tikhonov regularization of the inverse of the marginal variance of the sample, which is justified by a penalized projection argument. This regularized posterior distribution is consistent in the frequentist sense and its mean can be interpreted as the mean of the exact posterior distribution resulting from a Gaussian prior distribution with a shrinking covariance operator.

1. Introduction

In structural econometrics an important question is the treatment of endogeneity. Economic analysis provides econometricians with theoretical models that specify a structural relationship $\varphi(\cdot)$ among variables: a response variable, denoted by Y, and a vector of explanatory variables, denoted by Z. In many cases, the variables in Z are exogenous, where exogeneity is defined by the property $\varphi(Z) = \mathbb{E}(Y|Z)$. However, very often in economic models the explanatory variables are endogenous and the structural relationship $\varphi(Z)$ is not the conditional expectation function $\mathbb{E}(Y|Z)$. In this paper we deal with this latter case and the structural model we consider is:

$$Y = \varphi(Z) + U, \qquad \mathbb{E}(U|Z) \neq 0 \tag{1}$$

under the assumption of additive separability of U. Function $\varphi(\cdot) : \mathbb{R}^p \to \mathbb{R}$, for some $p > 0$, is the link function we are interested in and U denotes a disturbance that, by (1), is nonindependent of the explanatory variables Z. This dependence could be due for instance to the fact that there are other variables that cause both Y and Z and that are omitted from the model. In order to characterize $\varphi(\cdot)$ we suppose that there exists a vector W of random variables, called instruments, that have a sufficiently strong dependence on Z and for which $\mathbb{E}(U|W) = 0$. Then,

$$\mathbb{E}(Y|W) = \mathbb{E}(\varphi|W) \tag{2}$$

and the function $\varphi(\cdot)$, defined as the solution of this moment restriction, is called *instrumental variable (IV) regression*. If the joint cumulative distribution function of (Y, Z, W) is characterized by its density with respect to the Lebesgue measure, Eq. (2) is an integral equation of the first kind and recovering its solution φ is an ill-posed inverse problem, see O'Sullivan (1986) and Kress (1999). Recently, theory and concepts typical of inverse problems literature, like *regularization of the solution*, *Hilbert Scale*, and *Source condition*, have become more and more popular in estimation of IV regression, see Florens (2003), Blundell and Powell (2003), Hall and Horowitz (2005), Darolles et al. (2011), Florens et al. (2011, forthcoming) and Gagliardini and Scaillet (2012), to name only a few. Other recent contributions to the literature on nonparametric estimation of IV regression, based on the finite dimensional sieve minimum distance estimator, are Newey and Powell (2003), Ai and Chen (2003) and Blundell et al. (2007).

☆ We acknowledge helpful comments from the editors Mehmet Caner, Marine Carrasco, Yuichi Kitamura and Eric Renault and from two anonymous referees. We also thank Joel Horowitz, Enno Mammen and participants in seminars and conferences in Marseille (2007, 2008), Yale (2008), Boulder (2008), ESEM (2008), CEMMAP (2010) and the "Inverse Problems" group of Toulouse. The usual disclaimer applies and all errors remain ours. Simoni is grateful for the hospitality of Toulouse School of Economics and University of Mannheim where part of this research was conducted. Financial support from Alexander-von-Humboldt chair at the University of Mannheim is gratefully acknowledged by the second author.

* Corresponding author.
E-mail address: simoni.anna@gmail.com (A. Simoni).

0304-4076/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2012.05.016

The existing literature linking IV regression estimation and inverse problems theory is based on frequentist techniques. Our aim is to develop a *quasi-Bayesian* nonparametric estimation of the IV regression based on the Bayesian inverse problems theory. Bayesian analysis of inverse problems has been developed by Franklin (1970), Mandelbaum (1984) and Lehtinen et al. (1989) and recently by Florens and Simoni (2011, 2012). We call our approach *quasi-Bayesian* because the posterior distribution that we recover is not the exact one and because asymptotic properties of it and of the posterior mean estimator of the IV regression are analyzed from a frequentist perspective, i.e. with respect to the sampling distribution.

The Bayesian estimation of φ that we develop in this paper is based on the reduced form model associated with (1) and (2):

$$Y = \mathbb{E}(\varphi|W) + \varepsilon, \quad \mathbb{E}(\varepsilon|W) = 0 \tag{3}$$

where the residual ε is defined as $\varepsilon = Y - \mathbb{E}(Y|W) = \varphi(Z) - \mathbb{E}(\varphi|W) + U$ and is supposed to be Gaussian conditionally on W and homoskedastic. The reduced form model (3), without the homoskedasticity assumption, has been also considered by Chen and Reiss (2011) under the name *nonparametric indirect regression* model and by Loubes and Marteau (2009). Model (3) is used to construct the sampling distribution of Y given φ. In the Bayesian philosophy the functional parameter of interest φ is not conceived as a given parameter, but it is conceived as a realization of a random process and the space of reference is the product space of the sampling and parameter space. We do not constrain φ to belong to some parametric space; we only require that it satisfies some regularity condition as is usual in nonparametric estimation. We specify a very general Gaussian prior distribution for φ, general in the sense that the prior covariance operator is not required to have any particular form or any relationship with the sampling model (3); the only requirement is that the prior covariance operator is of trace-class.

The Bayes estimation of φ, or equivalently the Bayes solution of the inverse problem, is the posterior distribution of φ. The Bayesian approach solves the original ill-posedness of an inverse problem by changing the nature of the problem in the following sense. The problem of finding the solution of an integral equation is replaced by the problem of finding the inverse decomposition of a joint probability measure constructed as the product of the prior and the sampling distribution. This means that we have to find the posterior distribution of φ and the marginal distribution of Y. Unfortunately, because the parameter φ is of infinite dimension, its Gaussian posterior distribution suffers another kind of ill-posedness. The posterior distribution, which is well defined in small sample size, may have a bad frequentist behavior as the sample size increases. In this paper we adopt a frequentist perspective and admit the existence of a true value of φ, denoted by φ_*, that characterizes the data generating process and that satisfies (2). Therefore, we study consistency of the posterior distribution in a frequentist sense. *Posterior*, or frequency, *consistency* means that the posterior distribution degenerates to a Dirac measure on the true value φ_*. We show that the posterior distribution is not consistent in a frequentist sense, even if it is consistent from a Bayesian point of view, i.e. with respect to the joint distribution of the sample and the parameter. The bad frequentist asymptotic behavior of the posterior distribution entails that, as the sample size increases, the posterior mean is no longer continuous in Y and it is an inconsistent estimator in the frequentist sense. This is due to the fact that in a conjugate Gaussian setting, the expression of the posterior mean involves the inverse of the marginal covariance operator of the sample which converges towards an unbounded operator.

To get rid of the problem of inconsistency of the Bayes estimator of the IV regression φ, we replace the posterior distribution by the *regularized posterior distribution* that we have introduced in Florens and Simoni (2012). This distribution is similar to the exact posterior distribution where the inverse of the marginal covariance operator of the sample, which characterizes the posterior mean and variance, is replaced by a regularized inverse through a Tikhonov regularization scheme. We point out that the model analyzed in this paper differs from the inverse problem studied in Florens and Simoni (2011, 2012). An important contribution of this paper, with respect to Florens and Simoni (2012), consists in providing a fully Bayesian interpretation for the mean of the regularized posterior distribution. It is the mean of the posterior distribution that would result if the prior covariance operator was specified as a shrinking operator depending on the sample size and on the regularization parameter α_n of the Tikhonov regularization. However, the variance of this posterior distribution slightly differs from the regularized posterior variance. This interpretation of the regularized posterior mean does not hold for a general inverse problem like that considered in Florens and Simoni (2012).

We assume homoskedasticity of the error term in (3) and our quasi-Bayesian approach is able to simultaneously estimate φ and the variance parameter of ε by specifying a prior distribution either conjugate or independent on these parameters.

The paper is organized as follows. The reduced form model for IV estimation is presented in Section 2. In Section 3 we present our Bayes estimator for φ—based on the regularized posterior distribution—and for the error variance parameter. Then, we discuss inconsistency of the exact posterior distribution of φ and state frequentist asymptotic properties of our estimator. The conditional distribution of Z given W is supposed to be known in Section 3. This assumption is relaxed in Section 4. Numerical simulations are presented in Section 5. Section 6 concludes. All the proofs are in Appendix.

2. The model

Let $S = (Y, Z, W)$ denote a random vector belonging to $\mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^q$ with distribution characterized by the cumulative distribution function F. We assume that F is absolutely continuous with respect to the Lebesgue measure with density f. We denote by f_z, f_w the marginal probability densities of Z and W, respectively, and by $f_{z,w}$ their joint density (Z, W). We introduce the real Hilbert space L^2_F of square integrable real functions of S with respect to F. We denote by $L^2_F(Z)$ and $L^2_F(W)$ the subspaces of L^2_F of square integrable functions of Z and of W, respectively. Hence, $L^2_F(Z) \subset L^2_F$ and $L^2_F(W) \subset L^2_F$. The inner product and the norm in these spaces are indistinctly denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively. We introduce the two following conditional expectation operators:

$$K : L^2_F(Z) \to L^2_F(W) \qquad K^* : L^2_F(W) \to L^2_F(Z)$$
$$h \mapsto \mathbb{E}(h|W) \qquad h \mapsto \mathbb{E}(h|Z).$$

The operator K^* is the adjoint of K with respect to the inner product in L^2_F. We assume that the IV regression φ, which satisfies model (3), is such that $\varphi \in L^2_F(Z)$.

The reduced form model (3) provides the sampling model for inference on φ and it is a conditional model, conditioned on W, that does not depend on Z. This is a consequence of the fact that the instrumental variable approach specifies a statistical model concerning (Y, W), and not concerning the whole vector (Y, Z, W) since the only information available is that $\mathbb{E}(U|W) = 0$. Nothing is specified about the joint distribution of (U, Z) and (Z, W) except that the dependence between Z and W must be sufficiently strong. It follows that if the conditional densities $f(Z|W)$ and $f(W|Z)$ are known, we need only a sample of (Y, W) and not of Z. However, we assume below that also a sample of Z is available since this will be used in Section 4 when $f(Z|W)$ and $f(W|Z)$ are unknown and must be estimated.

The ith observation of the random vector S is denoted by small letters: $s_i = (y_i, z_i', w_i')'$, where z_i and w_i are respectively $p \times 1$ and $q \times 1$ vectors. Boldface letters $\mathbf{z}$ and $\mathbf{w}$ denote the matrices where vectors z_i and w_i, $i = 1, \ldots, n$ have been stacked columnwise.

Assumption 1. We observe an i.i.d. sample $s_i = (y_i, z_i', w_i')'$, $i = 1, \ldots, n$ satisfying model (3).

Each observation satisfies the reduced form model: $y_i = \mathbb{E}(\varphi(Z)|w_i) + \varepsilon_i$, $\mathbb{E}(\varepsilon_i|\mathbf{w}) = 0$, for $i = 1, \ldots, n$, and Assumption 2 below. After having scaled every term in the reduced form by $\frac{1}{\sqrt{n}}$, we rewrite (3) for an n-sample in matrix form as

$$y_{(n)} = K_{(n)}\varphi + \varepsilon_{(n)}, \tag{4}$$

where

$$y_{(n)} = \frac{1}{\sqrt{n}}\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \qquad \varepsilon_{(n)} = \frac{1}{\sqrt{n}}\begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}, \quad y_{(n)}, \varepsilon_{(n)} \in \mathbb{R}^n$$

$$\forall \phi \in L^2_F(Z), \quad K_{(n)}\phi = \frac{1}{\sqrt{n}}\begin{pmatrix} \mathbb{E}(\phi(Z)|W = w_1) \\ \vdots \\ \mathbb{E}(\phi(Z)|W = w_n) \end{pmatrix},$$

$$K_{(n)} : L^2_F(Z) \to \mathbb{R}^n$$

$$\text{and} \quad \forall x \in \mathbb{R}^n, \quad K^*_{(n)}x = \frac{1}{\sqrt{n}}\sum_{i=1}^n x_i \frac{f_{z,w}(Z, w_i)}{f_z(Z)f_w(w_i)},$$

$$K^*_{(n)} : \mathbb{R}^n \to L^2_F(Z).$$

The set $\mathbb{R}^n$ is provided with its canonical Hilbert space structure where the scalar product and the norm are still denoted, by abuse of notation, by $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$. Operator $K^*_{(n)}$ is the adjoint of $K_{(n)}$, as can be easily verified by solving the equation $\langle K_{(n)}\phi, x\rangle = \langle\phi, K^*_{(n)}x\rangle$, for any $x \in \mathbb{R}^n$ and $\phi \in L^2_F(Z)$. Since $K_{(n)}$ and $K^*_{(n)}$ are finite rank operators they have only n singular values different than zero. We denote by $y^i_{(n)}$ and $\varepsilon^i_{(n)}$ the ith element of vectors $y_{(n)}$ and $\varepsilon_{(n)}$, respectively.

We use the notation $\mathcal{GP}$ to denote a Gaussian distribution either in finite or in infinite dimensional spaces. The residuals of Y given W in model (3) are assumed to be Gaussian and homoskedastic, thus we have the following assumption:

Assumption 2. The residuals of y_i given $\mathbf{w}$ are i.i.d. Gaussian: $\varepsilon_i|\sigma^2, \mathbf{w} \sim \text{i.i.d.}\mathcal{GP}(0, \sigma^2)$, $i = 1, \ldots, n$ and $\sigma^2 \in \mathbb{R}_+$.

It follows that $\varepsilon_{(n)}|\sigma^2, \mathbf{w} \sim \mathcal{GP}(0, \frac{\sigma^2}{n}I_n)$, where I_n is the identity matrix of order n. We only treat the homoskedastic case. Under the assumption of additive separability of the structural error term U and under Assumption 2, the conditional sampling distribution, conditioned on $\mathbf{w}$, is: $y_{(n)}|\varphi, \sigma^2, \mathbf{w} \sim \mathcal{GP}(K_{(n)}\varphi, \frac{\sigma^2}{n}I_n)$. We use the notation $P^{\sigma,\varphi,\mathbf{w}}$ to denote this distribution and P_i^{σ,φ,w_i} to denote the sampling distribution of $y^i_{(n)}$, conditioned on $W = w_i$, i.e. $P_i^{\sigma,\varphi,w_i} = \mathcal{GP}(\frac{1}{\sqrt{n}}\mathbb{E}(\varphi|W = w_i), \frac{1}{n}\sigma^2)$. We remark that elements $y^i_{(n)}$, $i = 1, \ldots, n$, represent n independent, but not identically distributed, random variables. This notation stresses the fact that φ and σ^2 are treated as random variables. When frequentist consistency is analyzed in the rest of the paper, we shall replace φ and σ^2 by their true values φ_* and σ^2_*, then the true sampling distribution will be denoted by $P^{\sigma_*,\varphi_*,\mathbf{w}}$.

Remark 1. The normality of errors in Assumption 2 is not restrictive. The proof of frequentist consistency of our IV estimator does not rely on this parametric restriction. Therefore, making Assumption 2 simply allows us to find a Bayesian justification for our estimator, but the estimator is well suited even if the normality assumption is violated. Hence, our approach is robust to normality assumption. On the other side, homoskedasticity of $\varepsilon_i|\mathbf{w}$ is crucial even if our approach may be extended to the heteroskedastic case.

3. Bayesian analysis

In this section we analyze the Bayesian experiment associated with the reduced form model (4) and we construct the Bayes estimator for (σ^2, φ). Let $\mathcal{F}_Y$ denote the Borel σ-field associated with the product sample space $\mathcal{Y} := \mathbb{R}^n$; we endow the measurable space $(\mathcal{Y}, \mathcal{F}_Y)$ with the sampling distribution $P^{\sigma,\varphi,\mathbf{w}}$ defined in the previous section.

This distribution, conditioned on the vector of instruments $\mathbf{w}$, depends on two parameters: the nuisance variance parameter σ^2 and the IV regression φ which represents the parameter of interest. Parameter $\sigma^2 \in \mathbb{R}_+$ is endowed with a probability measure, denoted by ν, on the Borel σ-field $\mathfrak{B}$ associated with $\mathbb{R}_+$. Parameter $\varphi(Z) \in L^2_F(Z)$ is endowed with a conditional probability measure, denoted by μ^σ and conditional on σ^2, on the Borel σ-field $\mathfrak{E}$ associated with $L^2_F(Z)$. The probability measure $\nu \times \mu^\sigma$ is the prior distribution on the parameter space $(\mathbb{R}_+ \times L^2_F(Z), \mathfrak{B} \otimes \mathfrak{E})$ and is specified in a conjugate way in the following assumption.

Assumption 3. (a) Let ν be an *Inverse Gamma* distribution on $(\mathbb{R}_+, \mathfrak{B})$ with parameters $\xi_0 \in \mathbb{R}_+$ and $s_0^2 \in \mathbb{R}_+$, i.e. $\nu \sim I\Gamma(\xi_0, s_0^2)$.

(b) Let μ^σ be a Gaussian measure on $(L^2_F(Z), \mathfrak{E})$ with a mean element $\varphi_0 \in L^2_F(Z)$ and a covariance operator $\sigma^2\Omega_0 : L^2_F(Z) \to L^2_F(Z)$ that is trace-class, i.e. $\varphi|\sigma^2 \sim \mathcal{GP}(\varphi_0, \sigma^2\Omega_0)$.

Notation $I\Gamma$ in part (a) of the previous assumption is used to denote the Inverse Gamma distribution. Parameter ξ_0 is the shape parameter and s_0^2 is the scale parameter. There exist different specifications of the density of an $I\Gamma$ distribution. We use in our study the form: $f(\sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^{\xi_0/2+1}\exp\left\{-\frac{1}{2}\frac{s_0^2}{\sigma^2}\right\}$ with $\mathbb{E}(\sigma^2) = \frac{s_0^2/2}{\xi_0/2-1} = \frac{s_0^2}{\xi_0-2}$ and $\text{Var}(\sigma^2) = \frac{s_0^4/4}{(\xi_0/2-1)^2(\xi_0/2-2)}$. Properties of the measure μ^σ specified in part (b) imply that $\mathbb{E}(\|\varphi\|^2) < \infty$ and that Ω_0 is linear, bounded, non-negative and self-adjoint. We give a brief reminder of the definition of covariance operator: Ω_0 is such that $\langle\sigma^2\Omega_0\delta, \phi\rangle = \mathbb{E}(\langle\varphi - \varphi_0, \delta\rangle\langle\varphi - \varphi_0, \phi\rangle|\sigma^2)$, for all δ, ϕ in $L^2_F(Z)$, see Chen and White (1998). The covariance operator Ω_0 needs to be trace-class in order that μ^σ generates almost surely trajectories belonging to $L^2_F(Z)$. A trace-class operator is a compact operator with eigenvalues that are summable. Therefore, Ω_0 cannot be proportional to the identity operator.

We introduce the *Reproducing Kernel Hilbert Space* ($\mathcal{R.K.H.S.}$ hereafter) associated with Ω_0 and denoted by $\mathcal{H}(\Omega_0)$. Let $\{\lambda_j^{\Omega_0}, \varphi_j^{\Omega_0}\}_j$ be the eigensystem of Ω_0, see Kress (1999, Section 15.4) for a definition of eigensystem and singular value decomposition of an operator. We define the space $\mathcal{H}(\Omega_0)$ embedded in $L^2_F(Z)$ as:

$$\mathcal{H}(\Omega_0) = \left\{h; h \in L^2_F(Z) \text{ and } \sum_{j=1}^\infty \frac{\langle h, \varphi_j^{\Omega_0}\rangle^2}{\lambda_j^{\Omega_0}} < \infty\right\} \tag{5}$$

and, by Proposition 3.6 in Carrasco et al. (2007), we have the relation $\mathcal{H}(\Omega_0) = \mathcal{R}(\Omega_0^{\frac{1}{2}})$, where $\mathcal{R}(\cdot)$ denotes the range of an operator.

The $\mathcal{R.K.H.S.}$ is a subset of $L^2_F(Z)$ that gives the geometry of the distribution of φ. The support of a centered Gaussian process, taking its values in $L^2_F(Z)$, is the closure in $L^2_F(Z)$ of the $\mathcal{R.K.H.S.}$ associated with the covariance operator of this process (denoted by $\overline{\mathcal{H}(\Omega_0)}$ in our case). Then $\mu^\sigma\{\varphi; (\varphi - \varphi_0) \in \overline{\mathcal{H}(\Omega_0)}\} = 1$ but it is well known that $\mu^\sigma\{\varphi; (\varphi - \varphi_0) \in \mathcal{H}(\Omega_0)\} = 0$, see Van der Vaart and Van Zanten (2008a).

From a frequentist point of view, there exists a true value $\varphi_* \in L^2_F(Z)$ that generates the data $y_{(n)}$ in model (4) and that satisfies the regularity assumption below:

Assumption 4. $(\varphi_* - \varphi_0) \in \mathcal{H}(\Omega_0)$, i.e. there exists $\delta_* \in L_F^2(Z)$ such that $(\varphi_* - \varphi_0) = \Omega_0^{\frac{1}{2}}\delta_*$.

This assumption may be discussed by the following remarks. First, let us notice that Ω_0 is an integral operator. Indeed, $\forall \delta, \phi \in L_F^2(Z)$ it is defined as

$$\begin{aligned}\langle \Omega_0\delta, \phi\rangle &= \frac{1}{\sigma^2}\mathbb{E}(\langle \varphi - \varphi_0, \delta\rangle\langle \varphi - \varphi_0, \phi\rangle|\sigma^2)\\ &= \frac{1}{\sigma^2}\mathbb{E}\left(\int (\varphi(z) - \varphi_0(z))\delta(z)f_z(z)dz\right.\\ &\quad \times \left.\int (\varphi(\zeta) - \varphi_0(\zeta))\phi(\zeta)f_z(\zeta)d\zeta\right|\sigma^2\Bigg)\\ &= \int \omega_0(z,\zeta)\delta(z)\phi(\zeta)f_z(z)f_z(\zeta)dzd\zeta\end{aligned}$$

where $\omega_0(z,\zeta) = \frac{1}{\sigma^2}\mathbb{E}[(\varphi(z)-\varphi_0(z))(\varphi(\zeta)-\varphi_0(\zeta))]$ is the kernel of the Ω_0 operator. Then, $\Omega_0\delta = \int \omega_0(z,\zeta)\delta(\zeta)f_z(\zeta)d\zeta$. If $\bar{\omega}_0$ satisfies the equation:

$$\omega_0(z,\zeta) = \int \bar{\omega}_0(z,t)\bar{\omega}_0(t,\zeta)f_z(t)dt$$

the operator $\Omega_0^{\frac{1}{2}}$ is also an integral operator with kernel $\bar{\omega}_0$, i.e.

$$\forall \delta \in L_F^2(Z), \quad \Omega_0^{\frac{1}{2}}\delta = \int \bar{\omega}_0(z,\zeta)\delta(\zeta)f_z(\zeta)d\zeta.$$

Assumption 4 can be rewritten:

$$\varphi_* - \varphi_0 = \int \bar{\omega}_0(z,\zeta)\delta_*(\zeta)f_z(\zeta)d\zeta$$

which is clearly a smoothing assumption on φ_*. This assumption may also be viewed as a hypothesis on the rate of decline of the Fourier coefficient of φ in the basis defined by the $\varphi_j^{\Omega_0}$s. Indeed, $(\varphi_* - \varphi_0) = \Omega_0^{\frac{1}{2}}\delta_*$ implies that $\|\delta_*\|^2 = \sum_{j=1}^{\infty}\frac{\langle\varphi_*-\varphi_0,\varphi_j^{\Omega_0}\rangle^2}{\lambda_j^{\Omega_0}}$ is bounded and, as $\lambda_j^{\Omega_0} \downarrow 0$ this implies that the Fourier coefficients $\langle \varphi_* - \varphi_0, \varphi_j^{\Omega_0}\rangle$ go to zero sufficiently fast or, intuitively, that $(\varphi_* - \varphi_0)$ may easily be approximated by a linear combination of the $\varphi_j^{\Omega_0}$s.

To give an idea of the smoothness of the functions in $\mathcal{H}(\Omega_0)$, consider for instance an operator Ω_0 with kernel the variance of a standard Brownian motion in $\mathcal{C}[0,1]$ (where $\mathcal{C}[0,1]$ denotes the space of continuously defined functions on [0, 1]), i.e. $\delta \in L_F^2(Z) \mapsto \Omega_0\delta = \int_0^1 (s \wedge t)\delta(s)ds$. The associated $\mathcal{R.K.H.S.}$ is the space of absolutely continuous functions h on [0, 1] with at least one square integrable derivative and such that $h(0) = 0$. Summarizing, according to our prior beliefs about the smoothness of φ_*, the operator Ω_0 must be specified in such a way that the corresponding $\mathcal{H}(\Omega_0)$ contains functions that satisfy such a smoothness and Assumption 4 is a way to impose a smoothness assumption on φ_*. We refer to Van der Vaart and Van Zanten (2008b, Section 10) for other examples of $\mathcal{R.K.H.S.}$ associated with the covariance operator of processes related to the Brownian motion.

Assumption 4 is closely related to the so-called "*source condition*" which expresses the smoothness (i.e. the regularity, for instance the number of square integrable derivatives) of the function φ_* according to smoothing properties of the operator K defining the inverse problem. More precisely, a source condition assumes that there exists a source $w \in L_F^2(Z)$ such that

$$\varphi_* = (K^*K)^{\mu}w, \qquad \|w\|^2 \le R, \quad R, \mu > 0.$$

Since for ill-posed problems K is usually a smoothing operator, the requirement for φ_* to belong to $\mathcal{R}(K^*K)^{\mu}$ can be considered as an (abstract) smoothness condition, see Engl et al. (2000, Section 3.2) and Carrasco et al. (2007).

Lastly, we remark that the fact that $\mu^{\sigma}\{\varphi; (\varphi-\varphi_0) \in \mathcal{H}(\Omega_0)\} = 0$ implies that the prior measure μ^{σ} is not able to generate trajectories of φ that satisfy Assumption 4. However, if Ω_0 is injective, then $\mathcal{H}(\Omega_0)$ is dense in $L_F^2(Z)$ so that the support of μ^{σ} is the whole space $L_F^2(Z)$ and the trajectories generated by μ^{σ} are as close as possible to φ_*. The incapability of the prior to generate the true parameter characterizing the data generation process is known in the literature as *prior inconsistency* and it is due to the fact that, because of the infinite dimensionality of the parameter space, the support of μ^{σ} can cover only a very small part of it.

Assumption 4 is sufficient in order to get the consistency result in Theorem 2 below because Ω_0 and $K\Omega_0^{\frac{1}{2}}$ do not necessarily have the same eigenfunctions and then they do not commute. If this were the case, then consistency of our estimator would be true even without Assumption 4.

3.1. Identification and overidentification

From a frequentist perspective, φ is identified in the IV model if the solution of Eq. (2) is unique. This is verified if K is one-to-one, i.e. $\mathcal{N}(K) = \{0\}$, where $\mathcal{N}(\cdot)$ denotes the kernel (or null space) of an operator. The existence of a solution of Eq. (2) is guaranteed if the regression function $\mathbb{E}(Y|W) \in \mathcal{R}(K)$. Non-existence of this solution characterizes a problem of *overidentification*. Henceforth, overidentified solutions come from equations with an operator that is not surjective and nonidentified solutions come from equations with an operator that is not one-to-one. Thus, existence and uniqueness of the classical solution depend on the properties of F.

The identification condition that we need in our problem is the following one:

Assumption 5. The operator $K\Omega_0^{\frac{1}{2}} : L_F^2(Z) \to L_F^2(W)$ is one-to-one on $L_F^2(Z)$.

This assumption is weaker than requiring K to be one to one since if $\Omega_0^{\frac{1}{2}}$ and $K\Omega_0^{\frac{1}{2}}$ are both one-to-one, this does not imply that K is one-to-one. This is due to the fact that we are working in spaces of infinite dimension. If we were in spaces of finite dimension and if the matrices $\Omega_0^{\frac{1}{2}}$ and $K\Omega_0^{\frac{1}{2}}$ were one-to-one, this would imply K was one-to-one. In reverse if $\Omega_0^{\frac{1}{2}}$ and K are one-to-one this does imply $K\Omega_0^{\frac{1}{2}}$ is one-to-one.

In order to understand the meaning of Assumption 5, it must be considered together with Assumption 4. Under Assumption 4, we can rewrite Eq. (2) as $\mathbb{E}(Y|W) = K\varphi_* = K\Omega_0^{\frac{1}{2}}\delta_*$, if $\varphi_0 = 0$. Then, Assumption 5 guarantees identification of the δ_* that corresponds to the true value φ_* satisfying Eq. (2). However, this assumption does not guarantee that the true value φ_* is the unique solution of (2) since it does not imply that $\mathcal{N}(K) = \{0\}$.

3.2. Regularized posterior distribution

Let $\Pi^{\mathbf{w}}$ denote the joint conditional distribution on the product space $(\mathbb{R}_+ \times L_F^2(Z) \times \mathcal{Y}, \mathfrak{B} \otimes \mathfrak{E} \otimes \mathcal{F}_Y)$, conditional on $\mathbf{w}$, that is $\Pi^{\mathbf{w}} = \nu \times \mu^{\sigma} \times P^{\sigma,\varphi,\mathbf{w}}$. We assume, in all the Section 3, that the density $f_{z,w}$, f_z and f_w are known. When this is not the case, the density f must be considered as a nuisance parameter to be incorporated in the model. Therefore, for completeness we should index the sampling probability with f: $P^{f,\sigma,\varphi,\mathbf{w}}$, but, for simplicity, we omit f when it is known.

Bayesian inference consists in finding the inverse decomposition of $\Pi^{\mathbf{w}}$ in the product of the posterior distributions of σ^2 and of φ conditionally on σ^2—denoted by $\nu_n^{y,\mathbf{w}} \times \mu_n^{\sigma,y,\mathbf{w}}$—and the marginal distribution $P^{\mathbf{w}}$ of $y_{(n)}$. After that, we recover the marginal posterior distribution of φ, $\mu_n^{y,\mathbf{w}}$, by integrating out σ^2 with respect to its posterior distribution. In the following, we simplify the notation by eliminating index $\mathbf{w}$ in the posterior distributions, so ν_n^y, $\mu_n^{\sigma,y}$ and μ_n^y must all be meant conditioned on $\mathbf{w}$. Summarizing, the joint distribution $\Pi^{\mathbf{w}}$ is:

$$\sigma^2 \sim I\Gamma(\xi_0, s_0^2)$$
$$\begin{pmatrix}\varphi \\ y_{(n)}\end{pmatrix}\Big|\sigma^2$$
$$\sim \mathcal{GP}\left(\begin{pmatrix}\varphi_0 \\ K_{(n)}\varphi_0\end{pmatrix}, \sigma^2\begin{pmatrix}\Omega_0 & \Omega_0 K_{(n)}^* \\ K_{(n)}\Omega_0 & \frac{1}{n}I_n + K_{(n)}\Omega_0 K_{(n)}^*\end{pmatrix}\right) \tag{6}$$

and the marginal distribution $P^{\sigma,\mathbf{w}}$ of $y_{(n)}$, obtained by marginalizing with respect to μ^σ, is $P^{\sigma,\mathbf{w}} \sim \mathcal{GP}(K_{(n)}\varphi_0, \sigma^2 C_n)$ with $C_n = (\frac{1}{n}I_n + K_{(n)}\Omega_0 K_{(n)}^*)$.

The marginal posterior distributions ν_n^y and μ_n^y will be analyzed in the next subsection; here we focus on $\mu_n^{\sigma,y}$. The conditional posterior distribution $\mu_n^{\sigma,y}$, conditional on σ^2, and more generally the posterior μ_n^y, are complicated objects in infinite dimensional spaces since the existence of a transition probability characterizing the conditional distribution of φ given $y_{(n)}$ (whether conditional or not on σ^2) is not always guaranteed, differently to the finite dimensional case. A discussion about this point can be found in Florens and Simoni (2012, Theorem 2). Here, we simply mention the fact that Polish spaces[1] guarantee the existence of such a transition probability (see the *Jirina Theorem* in Neveu (1965)) and both $\mathbb{R}^n$ and L_F^2, with the corresponding Borel σ-fields, are Polish spaces. The conditional posterior distribution $\mu_n^{\sigma,y}$, conditioned on σ^2, is Gaussian and $\mathbb{E}(\varphi|y_{(n)}, \sigma^2)$ exists, since $|\varphi|^2$ is integrable, and it is an affine transformation of $y_{(n)}$. We state the following theorem and we refer to Mandelbaum (1984) and Florens and Simoni (2012) for a proof of it.

Theorem 1. *Let $(\varphi, y_{(n)}) \in L_F^2(Z) \times \mathbb{R}^n$ be two Gaussian random elements jointly distributed as in (6), conditionally on σ^2. The conditional distribution $\mu_n^{\sigma,y}$ of φ given $(y_{(n)}, \sigma^2)$ is Gaussian with mean $Ay_{(n)} + b$ and covariance operator $\sigma^2\Omega_y = \sigma^2(\Omega_0 - AK_{(n)}\Omega_0)$, where*

$$A = \Omega_0 K_{(n)}^* C_n^{-1}, \qquad b = (I - AK_{(n)})\varphi_0 \tag{7}$$

and $I : L_F^2(Z) \to L_F^2(Z)$ is the identity operator.

Since we use a conjugate model, the variance parameter σ^2 affects the posterior distribution of φ only through the posterior covariance operator, so that $\mathbb{E}(\varphi|y_{(n)}, \sigma^2) = \mathbb{E}(\varphi|y_{(n)})$.

The posterior mean and variance are well defined for small n since C_n is an $n \times n$ matrix with n eigenvalues different than zero and then it is continuously invertible. Nevertheless, as $n \to \infty$, the operator $K_{(n)}\Omega_0 K_{(n)}^*$ in C_n converges towards the compact operator $K\Omega_0 K^*$ which has a countable number of eigenvalues accumulating at zero and which is not continuously invertible. Then, $K_{(n)}\Omega_0 K_{(n)}^*$ becomes not continuously invertible as $n \to \infty$. One could think that the operator $\frac{1}{n}I_n$ in C_n plays the role of a regularization operator and controls the ill-posedness of the inverse of the limit of $K_{(n)}\Omega_0 K_{(n)}^*$. This is not the case since $\frac{1}{n}$ converges to 0 too fast. Therefore, C_n^{-1} converges toward a noncontinuous operator that amplifies the measurement error in $y_{(n)}$ and $\mathbb{E}(\varphi|y_{(n)})$ is not consistent in the frequentist sense, that is, with respect to $P^{\sigma_*,\varphi_*,\mathbf{w}}$. This prevents the posterior distribution from being consistent in the frequentist sense and the posterior mean is not a suitable estimator. We discuss the inconsistency of the posterior distribution in more detail in Section 3.4 and we formally prove it in Lemma 2 below.

Remark 2. The IV model (4) describes an equation in finite dimensional spaces, but the parameter of interest is of infinite dimension so that the reduced form model can be seen as a projection of φ_* on a space of smaller dimension. If we solved (4) in a classical way, we would realize that some regularization scheme would be necessary also in the finite sample case since $\hat{\varphi} = (K_{(n)}^* K_{(n)})^{-1} K_{(n)}^* y_{(n)}$, but $K_{(n)}^* K_{(n)}$ is not full rank and then is not continuously invertible.

In order to solve the lack of continuity of C_n^{-1} we use the methodology proposed in Florens and Simoni (2012): we replace the exact posterior distribution with a *regularized posterior distribution*. This new distribution, denoted by $\mu_\alpha^{\sigma,y}$, is obtained by applying a Tikhonov regularization scheme to the inverse of C_n, so that we get $C_{n,\alpha}^{-1} = (\alpha_n I_n + \frac{1}{n}I_n + K_{(n)}\Omega_0 K_{(n)}^*)^{-1}$, where α_n is a regularization parameter. In practice, this consists in translating the eigenvalues of C_n far from 0 by a factor $\alpha_n > 0$. As $n \to \infty$, $\alpha_n \to 0$ at a suitable rate to ensure that operator $C_{n,\alpha}^{-1}$ stays well defined asymptotically.

Therefore, the *regularized conditional posterior distribution* (RCPD) $\mu_\alpha^{\sigma,y}$ is the conditional distribution on $\mathfrak{E}$, conditional on $(y_{(n)}, \sigma^2)$, defined in Theorem 1 with the operator A replaced by $A_\alpha := \Omega_0 K_{(n)}^* C_{n,\alpha}^{-1}$. The regularized conditional posterior mean and covariance operator are:

$$\hat{\varphi}_\alpha := \mathbb{E}_\alpha(\varphi|y_{(n)}, \sigma^2) = A_\alpha y_{(n)} + b_\alpha \tag{8}$$
$$\sigma^2\Omega_{y,\alpha} := \sigma^2(\Omega_0 - A_\alpha K_{(n)}\Omega_0)$$

with

$$A_\alpha = \Omega_0 K_{(n)}^*\left(\alpha_n I_n + \frac{1}{n}I_n + K_{(n)}\Omega_0 K_{(n)}^*\right)^{-1} \tag{9}$$
$$b_\alpha = (I - A_\alpha K_{(n)}\varphi_0)$$

and $\mathbb{E}_\alpha(\cdot|y_{(n)}, \sigma^2)$ denotes the expectation with respect to $\mu_\alpha^{\sigma,y}$.

We take the regularized posterior mean $\hat{\varphi}_\alpha$ as the point estimator for the IV regression. This estimator is justified as the minimizer of the penalized mean squared error obtained by approximating φ by a linear transformation of $y_{(n)}$. More clearly, by fixing $\varphi_0 = 0$ for simplicity, the bounded linear operator $A_\alpha : \mathbb{R}^n \to L_F^2(Z)$ is the unique solution to the problem:

$$A_\alpha = \arg\min_{\tilde{A}\in\mathcal{B}_2(\mathbb{R}^n, L_F^2(Z))} \mathbb{E}\|\tilde{A}y_{(n)} - \varphi\|^2 + \alpha_n\sigma^2\|\tilde{A}\|_{HS}^2 \tag{10}$$

where $\mathbb{E}(\cdot)$ denotes the expectation taken with respect to the conditional distribution $\mu^\sigma \times P^{\sigma,\varphi,\mathbf{w}}$ of $(\varphi, y_{(n)})$, given $(\sigma^2, \mathbf{w})$, $\|\tilde{A}\|_{HS}^2 := tr\tilde{A}^*\tilde{A}$ denotes the HS norm,[2] and $\mathcal{B}_2(\mathbb{R}^n, L_F^2(Z))$ is the set of all bounded operators from $\mathbb{R}^n$ to $L_F^2(Z)$ for which $\|A\|_{HS} < \infty$.

Even if we have constructed the RCPD through a Tikhonov regularization scheme and justified its mean as a penalized projection, we can derive the regularized posterior mean $\hat{\varphi}_\alpha$ as the mean of an exact Bayesian posterior. The mean $\hat{\varphi}_\alpha$ is the mean of the exact posterior distribution obtained from the sequence of prior probabilities, denoted by $\tilde{\mu}_n^\sigma$, of the form:

$$\varphi|\sigma^2 \sim \mathcal{GP}\left(\varphi_0, \frac{\sigma^2}{\alpha_n n + 1}\Omega_0\right)$$

1 A *Polish space* is a separable, completely metrizable topological space.

2 For a compact operator A, $tr(A)$ denotes the trace of A.

and from the sampling distribution $P^{\sigma,\varphi,\mathbf{w}} = \mathcal{GP}(K_{(n)}\varphi, \frac{\sigma^2}{n}I_n)$ (which is unchanged). With this sequence of prior probabilities, the posterior mean is:

$$\begin{aligned}\mathbb{E}(\varphi|y_{(n)},\sigma^2) &= \varphi_0 + \frac{\sigma^2}{\alpha_n n+1}\Omega_0K^*_{(n)}\left(\frac{\sigma^2}{n}I_n\right.\\ &\quad \left.+\ \frac{\sigma^2}{\alpha_n n+1}K_{(n)}\Omega_0K^*_{(n)}\right)^{-1}(y_{(n)}-K_{(n)}\varphi_0)\\ &= \varphi_0 + \Omega_0K^*_{(n)}\left(\frac{\alpha_n n+1}{n}I_n\right.\\ &\quad \left.+\ K_{(n)}\Omega_0K^*_{(n)}\right)^{-1}(y_{(n)}-K_{(n)}\varphi_0)\\ &= \varphi_0 + \Omega_0K^*_{(n)}\left(\alpha_nI_n + \frac{1}{n}I_n + K_{(n)}\Omega_0K^*_{(n)}\right)^{-1}\\ &\quad \times (y_{(n)}-K_{(n)}\varphi_0) \equiv \hat{\varphi}_\alpha.\end{aligned}$$

However, the posterior variance associated with this sequence of prior probabilities is different than the regularized conditional posterior variance:

$$\begin{aligned}\mathrm{Var}(\varphi|y_{(n)},\sigma^2) &= \frac{\sigma^2}{\alpha_n n+1}\Bigg[\Omega_0 - \Omega_0K^*_{(n)}\left(\alpha_nI_n + \frac{1}{n}I_n\right.\\ &\quad \left.+\ K_{(n)}\Omega_0K^*_{(n)}\right)^{-1}K_{(n)}\Omega_0\Bigg]\end{aligned}$$

and it converges faster than $\sigma^2\Omega_{y,\alpha}$. This is due to the fact that the prior covariance operator of $\tilde{\mu}^\sigma_n$ is linked to the sample size and to the regularization parameter α_n. Under the classical assumption $\alpha_n^2 n \to \infty$ (classical in inverse problems theory), this prior covariance operator is shrinking with the sample size and this speeds up the rate of $\mathrm{Var}(\varphi|y_{(n)},\sigma^2)$.

Such a particular feature of the prior covariance operator can make $\tilde{\mu}^\sigma_n$ a not desirable prior: first of all because a sequence of priors that become more and more precise requires that we are very sure about the value of the prior mean; secondly, because a prior that depends on the sample size is not acceptable for a subjective Bayesian. For these reasons, we prefer to construct $\hat{\varphi}_\alpha$ by starting from a prior distribution with a general covariance operator and by using a Tikhonov scheme even if our point estimator $\hat{\varphi}_\alpha$ can be equivalently derived with a fully Bayes rule.

3.3. The Student t process

We proceed now to compute the posterior distribution ν^y_n of σ^2. This distribution will be used in order to marginalize $\mu^{\sigma,y}_\alpha$ and is computed from the probability model $P^{\sigma,\mathbf{w}} \times \nu$, where $P^{\sigma,\mathbf{w}}$, defined in (6), is obtained by integrating out φ from $P^{\sigma,\varphi,\mathbf{w}}$ by using the prior μ^σ. The posterior distribution of σ^2 is an $I\Gamma$ distribution:

$$\begin{aligned}\sigma^2|y_{(n)} \sim \nu^y_n \propto \left(\frac{1}{\sigma^2}\right)^{\xi_0/2+n/2+1}\exp\Bigg\{-\frac{1}{2\sigma^2}[(y_{(n)}\\ -\ K_{(n)}\varphi_0)'C_n^{-1}(y_{(n)}-K_{(n)}\varphi_0)+s_0^2]\Bigg\}.\end{aligned} \tag{11}$$

Then, $\nu^y_n \sim I\Gamma(\xi_*, s^2_*)$ with $\xi_* = \xi_0 + n$,

$s^2_* = s^2_0 + (y_{(n)} - K_{(n)}\varphi_0)'C_n^{-1}(y_{(n)} - K_{(n)}\varphi_0)$

and we can take the posterior mean $\mathbb{E}(\sigma^2|y_{(n)}) = \frac{s_*}{(\xi_*-2)}$ as point estimator.

Since ν^y_n does not depend on φ it can be used for marginalizing the RCPD $\mu^{\sigma,y}_\alpha$ of φ, conditional on σ^2, with respect to σ^2. In the finite dimensional case, integrating a Gaussian process with respect to an Inverse Gamma distribution gives a *Student t* distribution. This suggests that we should find a similar result for infinite dimensional random variables and that $\varphi|y_{(n)}$ should be a process with a distribution equivalent to the Student t distribution, i.e. $\varphi|y_{(n)}$ should be a *Student t process* in $L^2_F(Z)$. This type of process has been used implicitly in the literature on Bayesian inference with Gaussian process priors in order to characterize the marginal posterior distribution of a functional parameter evaluated at a finite number of points, see e.g. O'Hagan et al. (1998) and Rasmussen and Williams (2006, Section 9.9). In these works this process is called a *Student process* simply because it generalizes the multivariate t distribution. Nevertheless, to the best of our knowledge, a formal definition of a Student t process in infinite dimensional Hilbert spaces has not been provided. In the next definition we give a formal definition of the *Student t process* ($\mathcal{S}t\mathcal{P}$) in an infinite dimensional Hilbert Space $\mathcal{X}$ by using the scalar product in $\mathcal{X}$.

Definition 1. Let $\mathcal{X}$ be an infinite dimensional Hilbert space with inner product $\langle\cdot,\cdot\rangle_{\mathcal{X}}$. We say that a random element x, with values in $\mathcal{X}$, is a *Student t process* with parameters $x_0 \in \mathcal{X}$, $\Omega_0 : \mathcal{X} \to \mathcal{X}$ and $\iota \in \mathbb{R}_+$, denoted $x \sim \mathcal{S}t\mathcal{P}(x_0, \Omega_0, \iota)$, if and only if $\forall\delta \in \mathcal{X}$,

$$\langle x,\delta\rangle_{\mathcal{X}} \sim t(\langle x_0,\delta\rangle_{\mathcal{X}}, \langle\Omega_0\delta,\delta\rangle_{\mathcal{X}}, \iota),$$

i.e. $\langle x,\delta\rangle_{\mathcal{X}}$ has a density proportional to

$$\left[\iota + \frac{(\langle x,\delta\rangle_{\mathcal{X}} - \langle x_0,\delta\rangle_{\mathcal{X}})^2}{\langle\Omega_0\delta,\delta\rangle_{\mathcal{X}}}\right]^{-\frac{\iota+1}{2}},$$

with mean and variance

$$\mathbb{E}(\langle x,\delta\rangle_{\mathcal{X}}) = \langle x_0,\delta\rangle_{\mathcal{X}}, \quad \text{if } \iota > 1$$
$$\mathrm{Var}(\langle x,\delta\rangle_{\mathcal{X}}) = \frac{\iota}{\iota-2}\langle\Omega_0\delta,\delta\rangle_{\mathcal{X}}, \quad \text{if } \iota > 2.$$

We admit the following Lemma, concerning the marginalization of a Gaussian process with respect to a scalar random variable distributed as an *Inverse Gamma*.

Lemma 1. *Let $\sigma^2 \in \mathbb{R}_+$ and x be a random function with value in the Hilbert space $\mathcal{X}$. If $\sigma^2 \sim I\Gamma(\xi, s^2)$ and $x|\sigma^2 \sim \mathcal{GP}(x_0, \sigma^2\Omega_0)$, with $\xi \in \mathbb{R}_+$, $s^2 \in \mathbb{R}_+$, $x_0 \in \mathcal{X}$ and $\Omega_0 : \mathcal{X} \to \mathcal{X}$, then*

$$x \sim \mathcal{S}t\mathcal{P}\left(x_0, \frac{s^2}{\xi}\Omega_0, \xi\right).$$

The proof of this lemma is trivial and follows immediately if we consider the scalar product $\langle x,\delta\rangle_{\mathcal{X}}$, $\forall\delta \in \mathcal{X}$, which is normally distributed on $\mathbb{R}$ conditioned on σ^2.

We apply this result to the IV regression process φ, so that by integrating out σ^2 in $\mu^{\sigma,y}_\alpha$, with respect to ν^y_n, we get

$$\varphi|y_{(n)} \sim \mathcal{S}t\mathcal{P}\left(\hat{\varphi}_\alpha, \frac{s^2_*}{\xi_*}\Omega_{y,\alpha}, \xi_*\right),$$

with marginal mean $\hat{\varphi}_\alpha$ and marginal variance $\frac{s^2_*}{\xi_*-2}\Omega_{y,\alpha}$. We call this distribution a *regularized posterior distribution* (RPD) and denote it by μ^y_α.

3.4. Asymptotic analysis

In this section we analyze asymptotic properties of ν^y_n, $\mu^{\sigma,y}_\alpha$ and μ^y_α from a frequentist perspective and we check that $\hat{\varphi}_\alpha$ and $\mathbb{E}(\sigma^2|y_{(n)})$ are consistent estimators for φ_* and σ^2_*, respectively (consistent in the frequentist sense). We say that the RCPD is consistent in the frequentist sense if the probability, taken with respect to $\mu^{\sigma,y}_\alpha$, of any complement of a neighborhood of φ_* converges to zero in $P^{\sigma_*,\varphi_*,\mathbf{w}}$-probability or $P^{\sigma_*,\varphi_*,\mathbf{w}}$-a.s. In other words, the pair $(\varphi_*, \mu^{\sigma,y}_\alpha)$ is consistent if for $P^{\sigma_*,\varphi_*,\mathbf{w}}$-almost all sequences $y_{(n)}$, the regularized posterior $\mu_\alpha^{\sigma,y}$ converges weakly to a Dirac measure on φ_*. Moreover, $\mu_\alpha^{\sigma,y}$ is consistent if $(\varphi_*, \mu_\alpha^{\sigma,y})$ is consistent for all φ_*. This concept of *regularized posterior consistency* is adapted from the concept of *posterior consistency* in the Bayesian literature, see for instance Diaconis and Freedman (1986), Definition 1.3.1 in Ghosh and Ramamoorthi (2003) and Van der Vaart and Van Zanten (2008a).

Posterior consistency is an important concept in the Bayesian nonparametric literature. The idea is that if there exists a true value of the parameter, the posterior should learn from the data and put more and more mass near this true value. The first to consider this idea was Laplace; Von Mises refers to posterior consistency as the second law of large numbers, see von Mises (1981) and Ghosh and Ramamoorthi (2003, Chapter 1). In Doob (1949) is published a fundamental result regarding consistency of Bayes estimators. Doob shows that, under weak measurability assumptions, for every prior distribution on the parameter space, the posterior mean estimator is a martingale which converges almost surely except possibly for a set of parameter values having prior measure zero. This convergence is with respect to the joint distribution of the sample and the parameter. A more general version of this theorem can be found in Florens et al. (1990, Chapters 4 and 7).

Doob's results have been extended by Breiman et al. (1964), Freedman (1963) and Schwartz (1965) extend Doob's theorem in a frequentist sense, that is, by considering a convergence with respect to the sampling distribution. Diaconis and Freedman (1986) point out the negative result that, in some infinite dimensional problems, inconsistency of the posterior distribution is the rule, see Freedman (1965).

We first analyze the inconsistency of the posterior distribution $\mu_n^{\sigma,y}$ defined in Theorem 1. Inconsistency of the posterior distribution represents the ill-posedness of the Bayesian inverse problem and it is stated in the following lemma:

Lemma 2. *Let $\varphi_* \in L_F^2(Z)$ be the true IV regression characterizing the data generating process $P^{\sigma_*,\varphi_*,\mathbf{w}}$. The pair $(\varphi_*, \mu_n^{\sigma,y})$ is inconsistent, i.e. $\mu_n^{\sigma,y}$ does not weakly converge to Dirac measure δ_{φ_*} centred on φ_* with probability one.*

This lemma shows that, in contrast to the finite dimensional case where the posterior distribution is consistent, in infinite dimensional problems we may have inconsistency. This is due to compactness of $K\Omega_0$ and to the fact that the sampling covariance operator shrinks at the rate n^{-1} which is too fast to control the ill-posedness.

On the other hand, we state in the following theorem that the RCPD $\mu_\alpha^{\sigma,y}$ and the regularized posterior mean $\hat{\varphi}_\alpha$ are consistent. For some $\beta > 0$, we denote by Φ_β the β-regularity space defined as

$$\Phi_\beta := \mathcal{R}(\Omega_0^{\frac{1}{2}} K^* K \Omega_0^{\frac{1}{2}})^{\frac{\beta}{2}}. \tag{12}$$

Theorem 2. *Let (σ_*^2, φ_*) be the true value of (σ^2, φ) that generates the data $y_{(n)}$ under model (4). Under Assumptions 4 and 5, if $\alpha_n \to 0$ and $\alpha_n^2 n \to \infty$, we have:*

(i) *$\|\hat{\varphi}_\alpha - \varphi_*\| \to 0$ in $P^{\sigma_*,\varphi_*,\mathbf{w}}$-probability, where $\hat{\varphi}_\alpha$ is defined in (8), and if $\delta_* \in \Phi_\beta$ for some $\beta > 0$,*

$$\|\hat{\varphi}_\alpha - \varphi_*\|^2 = \mathcal{O}_p\left(\alpha_n^\beta + \frac{1}{\alpha_n^2 n}\alpha_n^\beta + \frac{1}{\alpha_n^2 n}\right);$$

(ii) *if there exists a $\kappa > 0$ such that $\lim_{n\to\infty} \sum_{j=1}^n \frac{\langle \Omega_0 \varphi_{jn}, \varphi_{jn}\rangle}{\lambda_{jn}^{2\kappa}} < \infty$, where $\{\lambda_{jn}, \varphi_{jn}, \psi_{jn}\}_{j=1}^n$ is the singular value decomposition associated with $K_{(n)}\Omega_0^{\frac{1}{2}}$, then, for a sequence ϵ_n with $\epsilon_n \to 0$, $\mu_\alpha^{\sigma,y}\{\varphi \in L_F^2(Z); \|\varphi - \varphi_*\| \geq \epsilon_n\} \to 0$ in $P^{\sigma_*,\varphi_*,\mathbf{w}}$-probability, where $\mu_\alpha^{\sigma,y}$ is the RCPD. Moreover, if $\delta_* \in \Phi_\beta$ for some $\beta > 0$, it is of order*

$$\mu_\alpha^{\sigma,y}\{\varphi \in L_F^2(Z); \|\varphi - \varphi_*\| \geq \epsilon_n\} = \frac{1}{\epsilon_n^2}\mathcal{O}_p\left(\alpha_n^\beta + \frac{1}{\alpha_n^2 n}\alpha_n^\beta + \frac{1}{\alpha_n^2 n} + \alpha_n^\kappa\right).$$

(iii) *Lastly, $\forall \phi \in L_F^2(Z)$, $\|\sigma^2 \Omega_{y,\alpha}\phi\| \to 0$ in $P^{\sigma_*,\varphi_*,\mathbf{w}}$-probability and the restriction of $\Omega_{y,\alpha}$ to the set $\{\phi \in L_F^2(Z); \Omega_0^{\frac{1}{2}}\phi \in \Phi_\beta$, for some $\beta > 0\}$, is of order*

$$\|\Omega_{y,\alpha}\phi\|^2 = \mathcal{O}\left(\alpha_n^\beta + \frac{1}{\alpha_n^2 n}\alpha_n^\beta\right).$$

The condition $\delta_* \in \Phi_\beta$ required for δ_*, where δ_* is defined in Assumption 4, is just a regularity condition that is necessary for having convergence at a certain rate. It is a source condition on δ_* (see the discussion following Assumption 4) which expresses the regularity of δ_* in according to the smoothing properties of $K\Omega_0^{\frac{1}{2}}$. The larger β is, the smoother the function $\delta_* \in \Phi_\beta$ will be. However, with a Tikhonov regularization we have a *saturation effect* that implies that β cannot be greater than 2, see Engl et al. (2000, Section 4.2). Therefore, having a function δ_* with a degree of smoothness larger than 2 is useless with a Tikhonov regularization scheme.

The fastest global rate of convergence of $\hat{\varphi}_\alpha$ is obtained by equating α_n^β to $\frac{1}{\alpha_n^2 n}$; while the first rate α_n^β requires a regularization parameter α_n going to zero as fast as possible, the rate $\frac{1}{\alpha_n^2 n}$ requires an α_n decreasing to zero as slowly as possible. Hence, the optimal α_n, optimal for $\hat{\varphi}_\alpha$, is proportional to $\alpha_n^* \propto n^{-\frac{1}{\beta+2}}$ and the corresponding optimal rate for $\|\hat{\varphi}_\alpha - \varphi_*\|^2$ is proportional to $n^{-\frac{\beta}{\beta+2}}$. The optimal α_n for the RCPD $\mu_\alpha^{\sigma,y}$ is proportional to α_* if $\kappa \geq \beta$ and proportional to $n^{-\frac{1}{\kappa+2}}$ otherwise. Thus, the optimal rate of contraction of $\mu_\alpha^{\sigma,y}$ is $\epsilon_n \propto n^{-\frac{\beta\wedge\kappa}{(\beta\wedge\kappa)+2}}$.

Remark 3. From result (i) of Theorem 2 we can easily prove that the rate of contraction for the MISE $\mathbb{E}(\|\hat{\varphi}_\alpha - \varphi_*\|^2 | \sigma_*^2, \varphi_*, \mathbf{w})$ is the same as the rate for $\|\hat{\varphi}_\alpha - \varphi_*\|^2$.

Remark 4. Theorem 2 could be seen as a special case of Theorems 3–5 of Florens and Simoni (2012). However, the facts that $K_{(n)}$ and $K_{(n)}^*$ are finite rank operators and that the variance parameter σ^2 is treated as a random variable make the rates of convergence in Theorem 2 and strategy of its proof different from those of Theorems 2–4 in Florens and Simoni (2012).

Remark 5. The rate of convergence of the regularized posterior mean, given in Theorem 2 (i), can be improved if there exists a $0 < \tau < 1$ such that the operator $(TT^*)^\tau$ is trace-class for $T := K\Omega_0^{\frac{1}{2}}$; this is a condition on the joint density $f(Y, Z, W)$. If this assumption holds, the rate of the term depending on $\varepsilon_{(n)}$ in $\|\hat{\varphi}_\alpha - \varphi_*\|$ would be faster.

Next, we analyze consistency of $\mathbb{E}(\sigma^2 | y_{(n)})$ and of the posterior ν_n^y. If $\bar{\omega}_0(s, z)$ denotes the kernel of $\Omega_0^{\frac{1}{2}}$, we use the notation $g(Z, w_i) = \Omega_0^{\frac{1}{2}}\left(\frac{f_{z,w}(s,w_i)}{f_z(s)f_w(w_i)}\right)(Z) = \int \bar{\omega}_0(s, Z)\frac{f_{z,w}(s,w_i)}{f_z(s)f_w(w_i)} f_z(s)ds$, then $\Omega_0^{\frac{1}{2}} K_{(n)}^* \varepsilon_{(n)} = \frac{1}{n}\sum_{i=1}^n \varepsilon_i g(Z, w_i)$.

Theorem 3. *Let (σ_*^2, φ_*) be the true value of (σ^2, φ) that generates the data under model (4). Under Assumption 4, if there exists a $\gamma > 0$*

such that $\forall w, g(Z,w) \in \Phi_\gamma$ (with Φ_γ defined as in (12)), then

$$\sqrt{n^{\gamma\wedge 1}}\left(\mathbb{E}(\sigma^2|y_{(n)}) - \sigma_*^2\right) = \mathcal{O}_p(1).$$

It follows that, for a sequence ϵ_n such that $\epsilon_n \to 0$, $\nu_n^y\{\sigma^2 \in \mathbb{R}_+; |\sigma^2 - \sigma_*^2| \geq \epsilon_n\} \to 0$ in $P^{\sigma_*,\varphi_*,\mathbf{w}}$-probability where ν_n^y is the $I\Gamma(\xi_*, s_*^2)$ distribution on $\mathbb{R}_+$ described in (11).

The last assertion of the theorem shows that the posterior probability of the complement of any neighborhood of σ_*^2 converges to 0; then, ν^y is consistent in the frequentist sense.

We conclude this section by giving a result of joint posterior consistency, that is, the joint regularized posterior $\nu_n^y \times \mu_\alpha^{\sigma,y}$ degenerates toward a Dirac measure on (σ_*^2, φ_*).

Corollary 1. *Under conditions of Theorems 2 and 3, the joint posterior distribution*

$$\nu_n^y \times \mu_\alpha^{\sigma,y}\{(\sigma^2,\varphi) \in \mathbb{R}_+ \times L_F^2(Z); \|(\sigma^2,\varphi) - (\sigma_*^2,\varphi_*)\|_{\mathbb{R}_+\times L_F^2(Z)} \geq \epsilon_n\}$$

converges to zero in $P^{\sigma_*,\varphi_*,\mathbf{w}}$-probability.

3.5. Independent priors

We would like to briefly analyze an alternative specification of the prior distribution for φ. We replace the prior distribution μ^σ in Assumption 3(b) by a Gaussian distribution with a covariance operator not depending on σ^2. This distribution, denoted by μ, is independent of σ^2: $\varphi \sim \mu = \mathcal{GP}(\varphi_0, \Omega_0)$, with φ_0 and Ω_0 as in Assumption 3(b). Hence, the joint prior distribution on $\mathbb{R}_+ \times L_F^2(Z)$ is equal to the product of two independent distributions: $\nu \times \mu$, with ν specified as in Assumption 3(a). The sampling measure $P^{\sigma,\varphi,\mathbf{w}}$ remains unchanged.

The resulting posterior conditional expectation $\mathbb{E}(\varphi|y_{(n)},\sigma^2)$ depends now on σ^2 and the marginal posterior distribution of φ does not have a nice closed form. Since we have a closed form for the regularized conditional posterior distribution (RCPD) of φ, conditional on σ^2, $\mu_\alpha^{\sigma,y}$, and for the RCPD of σ^2, conditional on φ, $\nu_\alpha^{\varphi,y}$, we can use a *Gibbs sampling* algorithm to get a good approximation of the stationary laws represented by the desired regularized marginal posterior distributions μ_α^y and ν_α^y of φ and σ^2, respectively.

In this framework, the regularization scheme affects also the posterior distribution of σ^2, whether conditional or not. We explain this fact in the following way. The conditional posterior distribution of φ given σ^2 still suffers from a problem of inconsistency since it demands the inversion of the covariance operator $\left(\frac{\sigma^2}{n}I_n + K_{(n)}\Omega_0 K_{(n)}^*\right)$ of the distribution of $y_{(n)}|\sigma^2$ which, as $n \to \infty$, converges toward an operator with noncontinuous inverse. Therefore, we use a Tikhonov regularization scheme and obtain the RCPD for φ, still denoted by $\mu_\alpha^{\sigma,y}$. It is a Gaussian measure with mean $\mathbb{E}(\varphi|y_{(n)},\sigma^2) = A_\alpha^\sigma y_{(n)} + b_\alpha^\sigma$ and covariance operator $\Omega_{y,\alpha}^\sigma = \Omega_0 - A_\alpha^\sigma K_{(n)}\Omega_0$ where

$$A_\alpha^\sigma = \Omega_0 K_{(n)}^*\left(\alpha_n I_n + \frac{\sigma^2}{n}I_n + K_{(n)}\Omega_0 K_{(n)}^*\right)^{-1},$$

$$b_\alpha^\sigma = (I - A_\alpha^\sigma K_{(n)})\varphi_0$$

that are different from A_α and b_α in (9). For computing the posterior $\nu_\alpha^{\varphi,y}$ of σ^2, given φ, we use the homoskedastic model specified in Assumption 2 for the reduced form error term: $\varepsilon_{(n)}|\sigma^2, \mathbf{w} \sim i.i.d.\mathcal{GP}(0, \frac{\sigma^2}{n}I_n)$ with $\varepsilon_{(n)} = y_{(n)} - K_{(n)}\varphi$ and φ is drawn from $\mu_\alpha^{\sigma,y}$. Therefore, we talk about the *regularized error term* and it results that the regularization scheme plays a role also in the conditional posterior distribution of σ^2 through φ, so that we index this distribution with α_n: $\nu_\alpha^{\varphi,y}$. The distribution $\nu_\alpha^{\varphi,y}$ is an $I\Gamma(\xi_*, \tilde{s}_*^2)$, with $\xi_* = \xi_0 + n$, $\tilde{s}^2 = s_0^2 + n\sum_i (y_{(n)}^i - K_{(n)}^i\varphi)^2$ and $K_{(n)}^i$ denotes the ith component of $K_{(n)}$.

It is then possible to implement a Gibbs sampling algorithm by alternatively drawing from $\mu_\alpha^{\sigma,y}$ and $\nu_\alpha^{\varphi,y}$ with the initial values for σ^2 drawn from an overdispersed $I\Gamma$ distribution. The first J draws are discarded; we propose to determine the number J for instance by using the technique proposed in Gelman and Rubin (1992), which can be trivially adapted for an infinite dimensional parameter, see Simoni (2009, Section 4.3.3).

4. The unknown operator case

In this section the variance parameter σ^2 is considered as known, in order to simplify the setting, and we specify the prior for φ as in Assumption 3(b) with the difference that the prior covariance operator does not depend on σ^2, then $\mu \sim \mathcal{GP}(\varphi_0, \Omega_0)$.

4.1. Unknown infinite dimensional parameter

We consider the case in which the density $f_{z,w} := f(Z,W)$ is unknown and then operators $K_{(n)}$ and $K_{(n)}^*$ are also unknown. We do not use a Bayesian treatment for estimating $f_{z,w}$. The Bayesian estimation of all the parameters of our model $(f_{z,w}, \sigma^2, \varphi)$ is difficult for the following reason. Given $f_{z,w}$, the inference on φ and σ^2 may be concentrated on the conditional distribution of Y given W as we did before (note that we may assume that $Y|Z,W \sim Y|W$). In reverse, the inference on $f_{z,w}$ given φ and σ^2 may not be concentrated on the (Z,W)-distribution: the curve Y (given W) also contains some information on $f_{z,w}$.

In order to bypass these problems we propose to use another technique that does not appear among Bayesian methods. We propose to substitute the true $f_{z,w}$ in $K_{(n)}$ and $K_{(n)}^*$ with a nonparametric classical estimator $\hat{f}_{z,w}$ and to redefine the IV regression φ as the solution of the estimated reduced form equation

$$y_{(n)} = \hat{K}_{(n)}\varphi + \eta_{(n)} + \varepsilon_{(n)} \tag{13}$$

where $\hat{K}_{(n)}$ and $\hat{K}_{(n)}^*$ denote the corresponding estimated operators. We have two error terms: $\varepsilon_{(n)}$ is the error term of the reduced form model (4) and $\eta_{(n)}$ accounts for the estimation error of operator $K_{(n)}$, i.e. $\eta_i = \frac{1}{\sqrt{n}}(K_{(n)}^i\varphi_* - \hat{K}_{(n)}^i\varphi_*)$ and $\eta_{(n)} = (\eta_1,\ldots,\eta_n)'$. If model (4) is true, then also (13) is true and characterizes φ_*.

We estimate $f_{z,w}$ by a kernel smoothing. Let L be a kernel function satisfying the usual properties and ρ be the minimum between the order of L and the order of differentiability of f. We use the notation $L(u)$ for $L(\frac{u}{h})$ where h is the bandwidth used for kernel estimation such that $h \to 0$ as $n \to \infty$ (to lighten notation we have eliminated the dependence on n from h). We denote by L_w the kernel used for W and L_z the kernel used for Z. The estimated density function is

$$\hat{f}_{z,w} = \frac{1}{nh^{p+q}}\sum_{i=1}^n L_w(w_i - w)L_z(z_i - z).$$

The estimator of $K_{(n)}$ is the classical Nadaraya–Watson estimator and $K_{(n)}^*$ is estimated by plugging in the estimates $\hat{f}_{z,w}, \hat{f}_z$ and $\hat{f}_w$:

$$\hat{K}_{(n)}\varphi = \frac{1}{\sqrt{n}}\begin{pmatrix} \sum_j \varphi(z_j)\dfrac{L_w(w_1 - w_j)}{\sum_l L_w(w_1 - w_l)} \\ \vdots \\ \sum_j \varphi(z_j)\dfrac{L_w(w_n - w_j)}{\sum_l L_w(w_n - w_l)} \end{pmatrix}, \quad \varphi \in L_Z^2$$

$$\hat{K}^*_{(n)}x = \frac{1}{\sqrt{n}}\sum_i x_i \frac{\sum_j L_z(z-z_j)L_w(w_i-w_j)}{\sum_l L_z(z-z_l)\frac{1}{n}\sum_l L_w(w_i-w_l)}, \quad x\in\mathbb{R}^n$$

and

$$\hat{K}^*_{(n)}\hat{K}_{(n)}\varphi = \frac{1}{n}\sum_i\left(\sum_j \varphi(z_j)\frac{L_w(w_i-w_j)}{\sum_l L_w(w_i-w_l)}\right) \times \frac{\sum_j L_z(z-z_j)L_w(w_i-w_j)}{\sum_l L_z(z-z_l)\frac{1}{n}\sum_l L_w(w_i-w_l)}.$$

The element in brackets in the last expression converges to $\mathbb{E}(\varphi|w_i)$, the last ratio converges to $\frac{f(Z,w_i)}{f(Z)f(w_i)}$ and hence by the Law of Large Numbers $\hat{K}^*_{(n)}\hat{K}_{(n)}\varphi \to \mathbb{E}(\mathbb{E}(\varphi|w_i)|Z)$.

From asymptotic properties of the kernel estimator of a regression function we know that $\eta_{(n)} \Rightarrow \mathcal{N}_n(0, \frac{\sigma^2}{n^2h^q}D_{(n)})$ with $D_{(n)} = \mathrm{diag}(\frac{1}{f(w_i)}\int L_w^2(u)du)$ and $\Rightarrow$ denotes convergence in distribution. The asymptotic variance of $\eta_{(n)}$ is negligible with respect to $\mathrm{Var}(\varepsilon_{(n)}) \equiv \frac{\sigma^2}{n}I_n$ since, by definition, the bandwidth h is such that $nh^q \to \infty$. The same is true for the covariance between $\eta_{(n)}$ and $\varepsilon_{(n)}$. This implies that the conditional probability distribution of $(y_{(n)} - \hat{K}_{(n)}\varphi)$ conditional on $(\hat{f}_{z,w}, \varphi, \mathbf{w})$ is asymptotically Gaussian.

In our quasi-Bayesian approach the gaussianity of the sampling measure is used only in order to give a Bayesian justification to our estimator of the IV regression which is the regularized posterior mean. Gaussianity of the sampling measure is not used in the proof of frequentist consistency. For this reason, we can use the approximated sampling measure which is given by $y_{(n)}|\hat{f}_{z,w}, \varphi, \mathbf{w} \sim P^{\hat{f},\varphi,\mathbf{w}} \sim^a \mathcal{GP}(\hat{K}_{(n)}\varphi, \Sigma_n)$, where $\sim^a$ means "approximately distributed as", $\Sigma_n = \mathrm{Var}(\eta_{(n)} + \varepsilon_{(n)}) = (\frac{\sigma^2}{n} + o_p(\frac{1}{n}))I_n$ and for simplicity σ^2 is considered as known. The estimated density $\hat{f}_{z,w}$ affects the sampling measure through $\hat{K}_{(n)}$.

As in the basic case, the factor $\frac{1}{n}$ in Σ_n does not stabilize the inverse of the covariance operator $\hat{C}_n := (\Sigma_n + \hat{K}_{(n)}\Omega_0\hat{K}^*_{(n)})$: it converges to zero too fast to compensate the decline towards 0 of the spectrum of the limits of the operator $\hat{K}_{(n)}\Omega_0\hat{K}^*_{(n)}$. Therefore, to guarantee consistency of the posterior distribution a regularization parameter $\alpha_n > 0$ that goes to 0 more slowly than n^{-1} must be introduced. The regularized posterior distribution that results is called the *estimated regularized posterior distribution* since now it depends on $\hat{K}_{(n)}$ instead of on $K_{(n)}$. It is denoted by $\hat{\mu}^y_\alpha$, it is Gaussian with mean $\hat{\mathbb{E}}_\alpha(\varphi|y_{(n)})$ and covariance operator $\hat{\Omega}_{y,\alpha}$ given by

$$\hat{\mathbb{E}}_\alpha(\varphi|y_{(n)}) = \varphi_0 + \overbrace{\Omega_0\hat{K}^*_{(n)}(\alpha_nI_n + \Sigma_n + \hat{K}_{(n)}\Omega_0\hat{K}^*_{(n)})^{-1}}^{\hat{A}_\alpha} \times (y_{(n)} - \hat{K}_{(n)}\varphi_0)$$

$$\hat{\Omega}_{y,\alpha} = \Omega_0 - \Omega_0\hat{K}^*_{(n)}(\alpha_nI_n + \Sigma_n + \hat{K}_{(n)}\Omega_0\hat{K}^*_{(n)})^{-1}\hat{K}_{(n)}\Omega_0. \tag{14}$$

Asymptotic properties of the posterior distribution for the case with unknown $f_{z,w}$ are very similar to ones shown in Theorem 2. In fact, the estimation error associated with $\hat{K}_{(n)}$ is negligible with respect to the other terms in the bias and variance. In the following theorem we focus on the consistency of $\hat{\mathbb{E}}_\alpha(\varphi|y_{(n)})$; consistency of $\hat{\mu}^y_\alpha$ and convergence to 0 of $\hat{\Omega}_{y,\alpha}$ may be easily derived from consistency of $\hat{\mathbb{E}}_\alpha(\varphi|y_{(n)})$ and Theorem 2. Darolles et al. (2011) provides primitive assumptions that guarantee $\|\Omega_0^{\frac{1}{2}}\hat{K}^*_{(n)}\hat{K}_{(n)} - \Omega_0^{\frac{1}{2}}K^*K\|^2 = \mathcal{O}_p(\frac{1}{n} + h^{2\rho})$. We implicitly assume in the following theorem (and in Lemma 4 below) that the regularity Assumptions B.1–B.5 of Darolles et al. (2011) are satisfied.

Theorem 4. *Let φ_* be the true value of φ that generates the data $y_{(n)}$ under model (4). Under Assumptions 4 and 5, if $\alpha_n \to 0$ and $\alpha_n^2 n \to \infty$, we have $\|\hat{\mathbb{E}}_\alpha(\varphi|y_{(n)}) - \varphi_*\|^2 \to 0$ in $P^{\hat{f},\varphi_*,\mathbf{w}}$-probability and if $\delta_* \in \Phi_\beta$, for some $\beta > 0$, then*

$$\|\hat{\mathbb{E}}_\alpha(\varphi|y_{(n)}) - \varphi_*\|^2 = \mathcal{O}_p\left(\alpha_n^\beta + \frac{1}{\alpha_n^2 n} + \frac{1}{\alpha_n^2}\left(\frac{1}{n} + h^{2\rho}\right)\frac{1}{\alpha_n^2 n}\right).$$

If the bandwidth h is chosen in such a way to guarantee that $\frac{1}{\alpha_n^2}(\frac{1}{n} + h^{2\rho}) = \mathcal{O}(\frac{1}{\alpha_n^2 n})$, the optimal speed of convergence is obtained by equating $\alpha_n^\beta = \frac{1}{\alpha_n^2 n}$. Hence, we set $h \propto n^{-\frac{1}{2\rho}}$ and we get the optimal regularization parameter $\alpha_n^* \propto n^{-\frac{1}{\beta+2}}$ and the optimal rate of convergence of $\|\hat{\mathbb{E}}_\alpha(\varphi|y_{(n)}) - \varphi_*\|^2$ proportional to $n^{-\frac{\beta}{\beta+2}}$. The rate is the same as for the case with $f_{z,w}$ known.

5. Numerical implementation

In this section we summarize the results of a numerical investigation of the finite sample performance of the regularized posterior mean estimator in both the known (Cases I and II below) and unknown operator case (Case III below). More figures concerning this simulation can be found in an additional appendix available at http://sites.google.com/site/simonianna/research.

We simulate $n = 1000$ observations from the following model, which involves only one endogenous covariate and two instrumental variables,[3]

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

$$v_i \sim \mathcal{N}(0,\sigma_v^2), \qquad z_i = 0.1w_{i,1} + 0.1w_{i,2} + v_i$$

$$\varepsilon_i \sim \mathcal{N}(0,(0.5)^2), \qquad u_i = \mathbb{E}(\varphi_*(z_i)|w_i) - \varphi_*(z_i) + \varepsilon_i$$

$$y_i = \varphi_*(z_i) + u_i.$$

We consider two alternative specifications for the true value of the IV regression: a smooth function $\varphi_*(Z) = Z^2$ and an irregular one $\varphi_*(Z) = \exp(-|Z|)$. Therefore, the structural error u_i takes the form $u_i = \sigma_v^2 - v_i^2 - 0.2v_i(w_{1,i} + w_{2,1}) + \varepsilon_i$ in the smooth case and the form $u_i = \exp(\frac{1}{2}\sigma_v^2)[e^{-\gamma}(1 - \Phi(\sigma_v - \frac{\gamma}{\sigma_v})) + e^\gamma\Phi(\sigma_v + \frac{\gamma}{\sigma_v})] - e^{-|z_i|} + \varepsilon_i$ in the irregular case, where $\Phi(\cdot)$ denotes the *cdf* of a $\mathcal{N}(0,1)$ distribution and $\gamma = 0.1w_{i,1} + 0.1w_{i,2}$. This mechanism of generation entails that $\mathbb{E}(u_i|w_i) = 0$; moreover, w_i, v_i and ε_i are mutually independent for every i. The joint density $f_{z,w}$ is

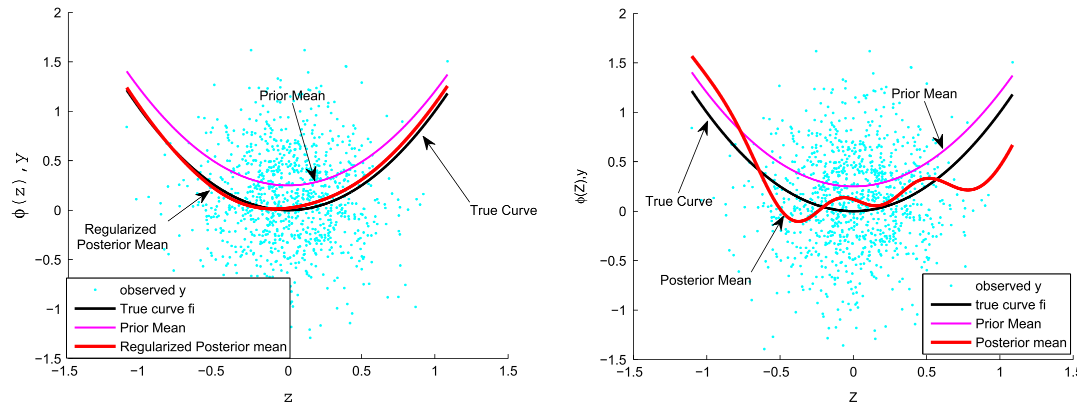
$$\begin{pmatrix} Z\\ W_1\\ W_2\end{pmatrix} \sim \mathcal{N}_3\left(\begin{pmatrix}0\\0\\0\end{pmatrix}, \begin{pmatrix}(0.026+\sigma_v^2) & 0.13 & 0.13\\ 0.13 & 1 & 0.3\\ 0.13 & 0.3 & 1\end{pmatrix}\right).$$

Endogeneity is caused by correlation between u_i and the error term v_i affecting the covariates. In all the simulations below we fix $\sigma_v = 0.27$ and select an α_n producing a good estimation by letting α_n vary in a large range of values. In the next section we present a data-driven method for selecting α_n.

Case I. *Conjugate Model with $f_{z,w}$ known and smooth φ_**

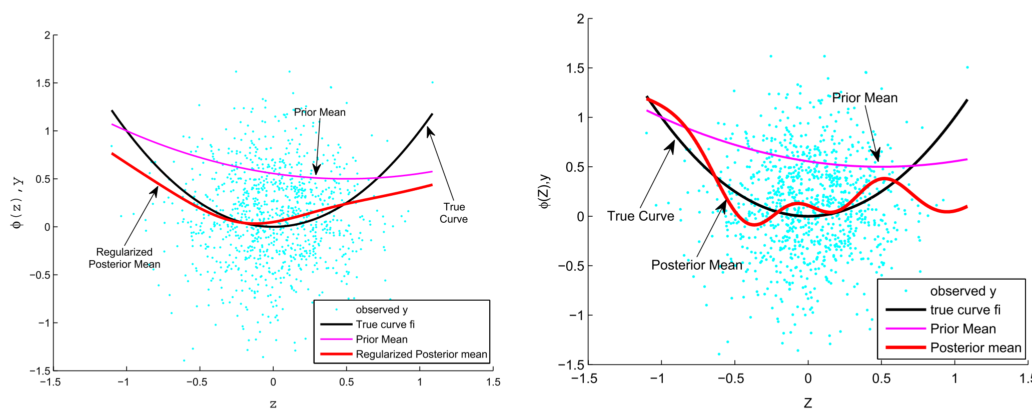
The true value of the IV regression is $\varphi_*(Z) = Z^2$. We use the following prior specification: $\sigma^2 \sim I\Gamma(6,1)$, $\varphi \sim \mathcal{GP}(\varphi_0, \sigma^2\Omega_0)$ with covariance operator $(\Omega_0\delta)(Z) = \sigma_0\int\exp(-(s-Z)^2)\delta(s) f_z(s)ds$, where $\sigma_0 = 200$ and $\delta \in L_F^2(Z)$. We have performed

[3] This data generating process is borrowed from Example 3.2 in Chen and Reiss (2011).



(a) Regularized posterior mean estimate with $\alpha_n = 0.3$. (b) Posterior mean estimate with $\alpha_n = 0$.

Fig. 1. Case I. *Conjugate Model with $f_{z,w}$ known and smooth φ_**. Graphs for $\varphi_0(Z) = 0.95Z^2 + 0.25$ and $\sigma_0 = 200$.



(a) Regularized posterior mean estimate with $\alpha_n = 0.3$. (b) Posterior mean estimate with $\alpha_n = 0$.

Fig. 2. Case I. *Conjugate Model with $f_{z,w}$ known and smooth φ_**. Graphs for $\varphi_0(Z) = \frac{2}{9}Z^2 - \frac{2}{9}Z + \frac{5}{9}$ and $\sigma_0 = 200$.

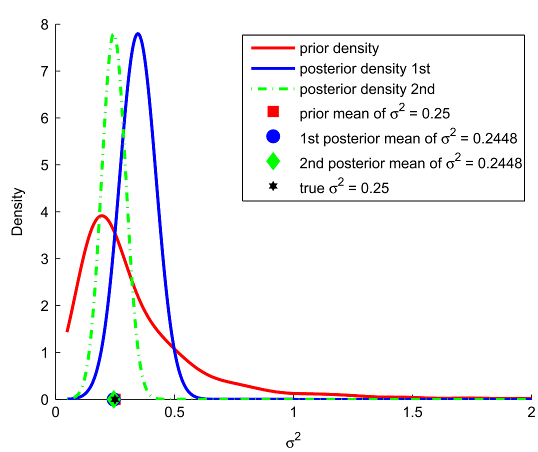


Fig. 3. Case I. *Conjugate Model with $f_{z,w}$ known and smooth φ_**. Prior and posterior distributions of σ^2. The label '1st' refers to the simulation with $\varphi_0(Z) = 0.95Z^2 + 0.25$, while '2nd' refers to the simulation with $\varphi_0(Z) = \frac{2}{9}Z^2 - \frac{2}{9}Z + \frac{5}{9}$.

simulations for two specifications of φ_0: Fig. 1 refers to $\varphi_0(Z) = 0.95Z^2 + 0.25$ while Fig. 2 refers to $\varphi_0(Z) = \frac{2}{9}Z^2 - \frac{2}{9}Z + \frac{5}{9}$.

We show in the first graph of both figures (graphs 1a and 2a) the estimation result for $\alpha_n = 0.3$: the magenta curve is the prior mean curve while the black curve is the true φ_* and the red curve is the regularized posterior mean $\hat{\varphi}_\alpha$. The second graph of both figures (graphs 1b and 2b) represents the posterior mean of φ with $\alpha = 0$, i.e. the mean of the non regularized posterior distribution $\mu_n^{\sigma,y}$.

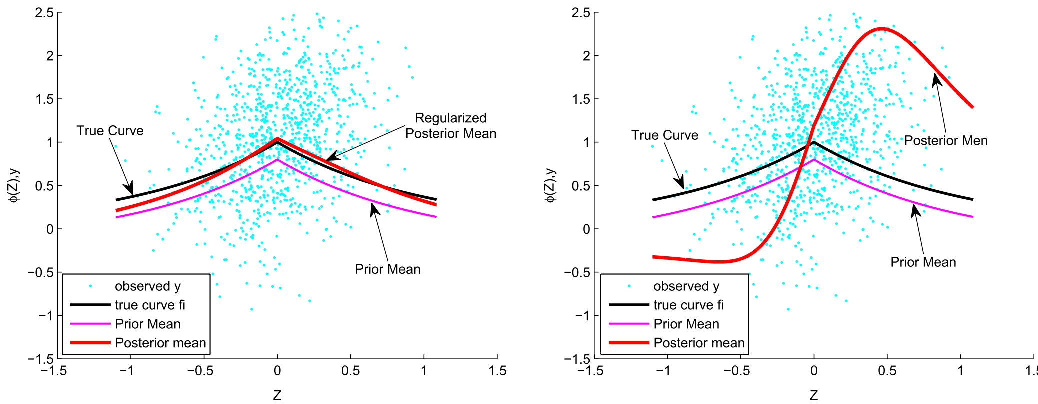
Fig. 3 represents the kernel smoothing estimators of the prior and posterior densities of σ^2. We have used a standard Gaussian kernel and a bandwidth equal to 0.05. The red curve is the prior density, while with the blue and the dashed–dotted green line we represent the posterior densities corresponding to the prior means $\varphi_0(Z) = 0.95Z^2 + 0.25$ (called 'posterior density 1st' in the graph) and $\varphi_0(Z) = \frac{2}{9}Z^2 - \frac{2}{9}Z + \frac{5}{9}$ (called 'posterior density 2nd' in the graph), respectively. The true value σ_*^2, the prior and posterior means are also shown.

Case II. *Conjugate Model with $f_{z,w}$ known and irregular φ_**

The true value of the IV regression is $\varphi_*(Z) = \exp(-|Z|)$. The prior distributions for σ^2 and φ are specified as in Case I but the variance parameter is $\sigma_0 = 2$ and the prior mean φ_0 is alternatively specified as $\varphi_0(Z) = \exp(-|Z|) - 0.2$ or $\varphi_0(Z) = 0$. The results concerning $\varphi_0(Z) = \exp(-|Z|) - 0.2$ and $\alpha_n = 0.4$ are reported in Fig. 4 while the results for $\varphi_0(Z) = 0$ and $\alpha_n = 0.3$ are in Fig. 5. The kernel estimators of the prior and posterior distributions of σ^2, together with its posterior mean estimator, are shown in Fig. 6. The interpretation of the graphs in each figure is the same as in Case I.

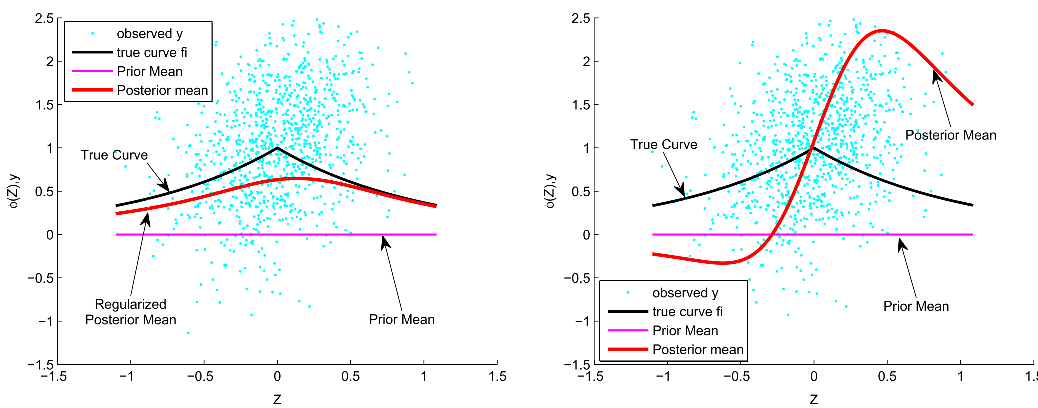
Case III. *$f_{z,w}$ unknown, σ^2 known and smooth φ_**

In this simulation we have specified a prior only on φ since σ^2 is supposed to be known. The prior distribution for φ is specified as in Case I with the same φ_0 and $\sigma_0 = 20$. We show in Figs. 7 the results obtained by using a kernel estimator for $f_{z,w}$ as described in Section 4. We have used a multivariate Gaussian kernel and a bandwidth equal to 0.1.



(a) Regularized posterior mean estimate with $\alpha_n = 0.4$. (b) Posterior mean estimate with $\alpha_n = 0$.

Fig. 4. Case II. *Conjugate Model with $f_{z,w}$ known and irregular φ_**. Graphs for $\varphi_0(Z) = \exp(-|Z|) - 0.2$ and $\sigma_0 = 2$.



(a) Regularized posterior mean estimate with $\alpha_n = 0.3$. (b) Posterior mean estimate with $\alpha_n = 0$.

Fig. 5. Case II. *Conjugate Model with $f_{z,w}$ known and irregular φ_**. Graphs for $\varphi_0(Z) = 0$ and $\sigma_0 = 2$.

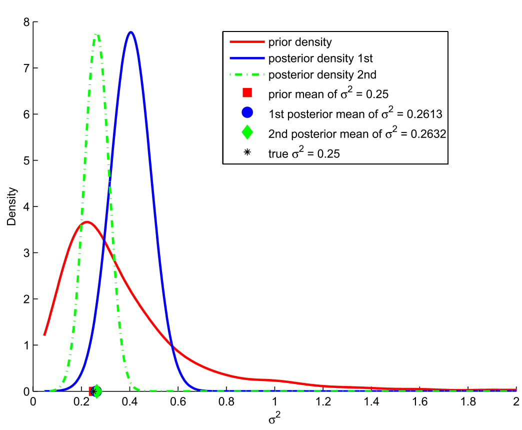


Fig. 6. Case II. *Conjugate Model with $f_{z,w}$ known and irregular φ_**. Prior and posterior distributions of σ^2. The label '1st' refers to the simulation with $\varphi_0(Z) = \exp(-|Z|) - 0.2$, while '2nd' refers to the simulation with $\varphi_0(Z) = 0$.

5.1. Data driven method for choosing α

In inverse problem theory there exist several parameter choice rules which determine the regularization parameter α_n on the basis of the performance of the regularization method under consideration. These techniques are often known as *error free* and we refer to Engl et al. (2000, Section 4.5) and the references therein for a review of them. We propose in this section a data-driven method that rests upon a slight modification of the estimation residuals derived when the regularized posterior mean $\hat{\varphi}_\alpha$ is used as a point estimator of the IV regression. Our method is a variation of the *error free* technique presented by Engl et al. (2000, p. 101).

The use of residuals instead of the estimation error $\|\hat{\varphi}_\alpha - \varphi_*\|$ is justified only if the residuals are adjusted in order to preserve the same rate of convergence as the estimation error. In particular, as noted in Engl et al. (2000), there exists a relation between the estimation error and the residuals rescaled by a convenient power of $\frac{1}{\alpha_n}$. Let ϑ_α denote the residual we are considering; we have to find the value d such that asymptotically

$$\frac{\|\vartheta_\alpha\|}{\alpha_n^d} \sim \|\hat{\varphi}_\alpha - \varphi_*\|,$$

where "$\sim$" means "of the same order as". Hence, it makes sense to take $\frac{\|\vartheta_\alpha\|}{\alpha_n^d}$ as the error estimator and to select the optimal α_n as the one that minimizes the ratio:

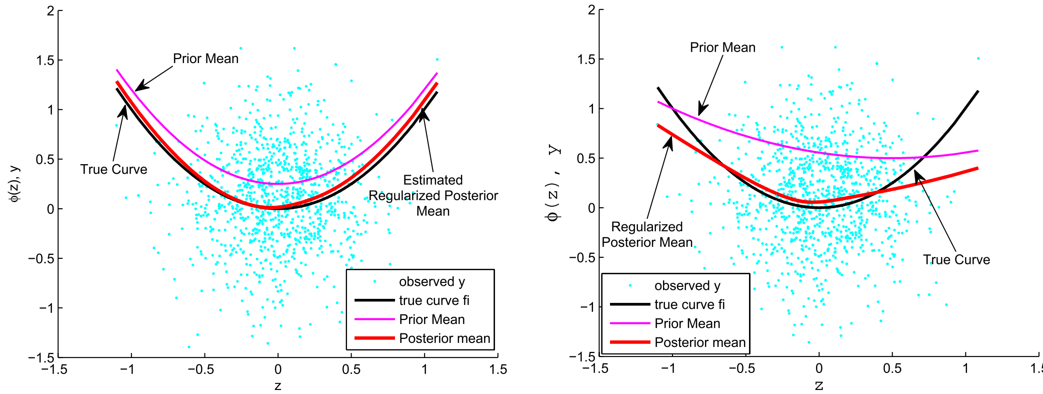
$$\hat{\alpha}_n^* = \arg\min \frac{\|\vartheta_\alpha\|}{\alpha_n^d}.$$

In the light of this argument, even if the classical residual $y_{(n)} - K_{(n)}\hat{\varphi}_\alpha$ would seem the natural choice, it is not acceptable since it does not converge to zero at a good rate. On the other hand, convergence is satisfied by the *projected residuals* defined as

$$\vartheta_\alpha = \Omega_0^{\frac{1}{2}} K_{(n)}^* y_{(n)} - \Omega_0^{\frac{1}{2}} K_{(n)}^* K_{(n)} \hat{\varphi}_\alpha$$

which for simplicity we rewrite as $\vartheta_\alpha = T_{(n)}^* y_{(n)} - T_{(n)}^* K_{(n)} \hat{\varphi}_\alpha$, using the notation $T_{(n)}^* = \Omega_0^{\frac{1}{2}} K_{(n)}^*$.

In order to explain our data-driven method we have to introduce the notion of *qualification* of a regularization method.



(a) Estimated regularized posterior mean for $\varphi_0(Z) = 0.95Z^2 + 0.25, \sigma_0 = 20$ and $\alpha_n = 0.3$.

(b) Estimated regularized posterior mean for $\varphi_0(Z) = \frac{2}{9}Z^2 - \frac{2}{9}Z + \frac{5}{9}, \sigma_0 = 20$ and $\alpha_n = 0.3$.

Fig. 7. Case III. *Conjugate Model with $f_{z,w}$ unknown and smooth φ_*.*

Under the assumption $\varphi_* \in \Phi_\beta$, we call the qualification β_0 of the regularization method the largest value of β such that $\|A_\alpha K_{(n)}\varphi_* + b_\alpha - \varphi_*\|^2 = \mathcal{O}_p(\alpha^\beta)$ for $0 < \beta < \beta_0$. While the qualification of the Tikhonov regularization is $\beta_0 = 2$, see Engl et al. (2000, Sections 4.1, 4.2 and 5.1), the data-driven method that we use requires that the qualification of the regularization scheme used be at least equal to $\beta + 2$, which is impossible for a Tikhonov regularization. Therefore, we have to substitute the Tikhonov regularization scheme, used to construct $\hat{\varphi}_\alpha$, with an *iterated Tikhonov* scheme, see e.g. Engl et al. (2000). In our case, it is enough to iterate only twice, so that the qualification will be 4, the resulting operator $A_\alpha^{(2)}$ takes the form: $A_\alpha^{(2)} = (\alpha\Omega_0 K_{(n)}^* C_{n,\alpha}^{-1} + \Omega_0 K_{(n)}^*)C_{n,\alpha}^{-1}$ and it replaces A_α in (8). We denote by $\hat{\varphi}_\alpha^{(2)}$ the regularized posterior mean obtained by using operator $A_\alpha^{(2)}$ and with $\vartheta_\alpha^{(2)}$ the corresponding projected residuals. Then, we have the following lemma.

Lemma 3. *Let $\hat{\varphi}_\alpha^{(2)} = A_\alpha^{(2)}(y_{(n)} - K_{(n)}\varphi_0) + \varphi_0$ and $\vartheta_\alpha^{(2)} = T_{(n)}^*(y_{(n)} - K_{(n)}\hat{\varphi}_\alpha^{(2)})$. Under Assumptions 4 and 5, if $\alpha_n \to 0$, $\alpha_n^2 n \to \infty$ and $\delta_* \in \Phi_\beta$ for some $\beta > 0$, then*

$$\|\vartheta_\alpha^{(2)}\|^2 = \mathcal{O}_p\left(\alpha_n^{\min(\beta+2,4)} + \frac{1}{n}\right).$$

The rate of convergence given in Lemma 3 can be made equivalent, up to negligible terms, to the rate given in Theorem 2 (i) by dividing $\|\vartheta_\alpha^{(2)}\|^2$ by α_n^2. Hence, once we have performed estimation for a given sample, we construct the curve $\frac{\|\vartheta_\alpha^{(2)}\|^2}{\alpha_n^2}$, as a function of α_n, and we select the value of the regularization parameter which minimizes it. To simplify the graphical representation of this curve, we take a logarithmic transformation of the ratio $\frac{\|\vartheta_\alpha^{(2)}\|^2}{\alpha_n^2}$.

A result similar to Lemma 3 can be derived when the density $f_{z,w}$ is unknown and the nonparametric method described in Section 4.1 is applied. In this case we denote by $\hat{T}_{(n)}^* = \Omega_0^{\frac{1}{2}}\hat{K}_{(n)}^*$ the estimates of the corresponding $T_{(n)}^*$ and we define the *estimated projected residual* as: $\hat{\vartheta}_\alpha^{(2)} = \hat{T}_{(n)}^*(y_{(n)} - \hat{K}_{(n)}\hat{\mathbb{E}}_\alpha^{(2)}(\varphi|y_{(n)}))$, where $\hat{\mathbb{E}}_\alpha^{(2)}(\varphi|y_{(n)})$ has been obtained by using a twice iterated Tikhonov scheme for constructing $\hat{A}_\alpha^{(2)} = \left(\alpha_n\Omega_0\hat{K}_{(n)}^*\hat{C}_{n,\alpha}^{-1} + \Omega_0\hat{K}_{(n)}^*\right)\hat{C}_{n,\alpha}^{-1}$ with $\hat{C}_{n,\alpha} = \left((\alpha_n + n^{-1})I_n + \hat{K}_{(n)}\Omega_0\hat{K}_{(n)}^*\right)$. We obtain the following result:

Lemma 4. *Let $\hat{\mathbb{E}}_\alpha^{(2)}(\varphi|y_{(n)}) = \hat{A}_\alpha^{(2)}(y_{(n)} - \hat{K}_{(n)}\varphi_0) + \varphi_0$ and $\hat{\vartheta}_\alpha^{(2)} = \hat{T}_{(n)}^*\left(y_{(n)} - \hat{K}_{(n)}\hat{\mathbb{E}}_\alpha^{(2)}(\varphi|y_{(n)})\right)$. Under Assumptions 4 and 5, if $\alpha_n \to 0$, $\alpha_n^2 n \to \infty$ and $\delta_* \in \Phi_\beta$ for some $\beta > 0$, then*

$$\|\hat{\vartheta}_\alpha^{(2)}\|^2 = \mathcal{O}_p\left(\alpha_n^{\beta+2} + \left(\frac{1}{n} + h^{2\rho}\right)\left(\alpha_n^\beta + \frac{1}{\alpha_n^2}\left(\frac{1}{n} + h^{2\rho}\right) + \frac{1}{\alpha_n^2 n}\right) + \frac{1}{n}\right).$$

In the lemma we implicitly assume that Assumptions B.1–B.5 of Darolles et al. (2011) are satisfied. It is necessary to rescale the residual by α_n^{-2} to obtain the same rate of convergence given in Theorem 4.

The graphical results of a numerical implementation concerning our data-driven method can be found in an additional appendix available at http://sites.google.com/site/simonianna/research.

6. Conclusions

We have proposed in this paper a new quasi-Bayesian method to make inference on an IV regression φ defined through a structural econometric model. The main feature of our method is that it does not require any specification of the functional form for φ, though it allows us to incorporate all the prior information available. A deeper analysis of the role played by the prior distribution is an important issue for future research.

Our estimator for φ is the mean of a slightly modified posterior distribution whose moments have been regularized through a Tikhonov scheme. We show that this estimator can be interpreted as the mean of an exact posterior distribution obtained with a sequence of Gaussian prior distributions for φ that shrink as $\alpha_n n$ increases. Alternatively, we motivate the regularized posterior mean estimator as the minimizer of a penalized mean squared error.

Frequentist asymptotic properties are analyzed; consistency of the regularized posterior distribution and of the regularized posterior mean estimator are stated.

Several possible extensions of our model can be developed. First of all, it would be interesting to consider other regularization methods, different from the Tikhonov scheme, and to analyze the way in which the regularized posterior mean is affected. We could also consider Sobolev spaces with regularization methods that use differential norms.

Appendix. Proofs

In all the proofs that follow we use the notation:

– (σ_*^2, φ_*) is the true parameter that generates the data according to model (4);
– $\mathcal{H}(\Omega_0) = \mathcal{R.K.H.S}(\Omega_0)$;

- if $(\varphi_* - \varphi_0) \in \mathcal{H}(\Omega_0)$, we write $(\varphi_* - \varphi) = \Omega_0^{\frac{1}{2}} \delta_*$, $\delta_* \in L_F^2(Z)$;
- I_n is the identity matrix of order n;
- $I : L_F^2(Z) \to L_F^2(Z)$ is the identity operator defined as $\varphi \in L_F^2(Z) \mapsto I\varphi = \varphi$;
- $T = K\Omega_0^{\frac{1}{2}}, T : L_F^2(Z) \to L_F^2(W)$;
- $T_{(n)} = K_{(n)}\Omega_0^{\frac{1}{2}}, T_{(n)} : L_F^2(Z) \to \mathbb{R}^n$;
- $\hat{T}_{(n)} = \hat{K}_{(n)}\Omega_0^{\frac{1}{2}}, \hat{T}_{(n)} : L_F^2(Z) \to \mathbb{R}^n$;
- $T^* = \Omega_0^{\frac{1}{2}} K^*, T^* : L_F^2(W) \to L_F^2(Z)$;
- $T_{(n)}^* = \Omega_0^{\frac{1}{2}} K_{(n)}^*, T_{(n)}^* : \mathbb{R}^n \to L_F^2(Z)$;
- $\hat{T}_{(n)}^* = \Omega_0^{\frac{1}{2}} \hat{K}_{(n)}^*, \hat{T}_{(n)}^* : \mathbb{R}^n \to L_F^2(Z)$;
- $\Omega_0^{\frac{1}{2}} = \int_{\mathbb{R}^p} \bar{\omega}_0(s, Z) f_z(s) ds$;
- $g(Z, w_i) = \int_{\mathbb{R}^p} \bar{\omega}_0(s, Z) \frac{f_{z,w}(s,w_i)}{f_z(s) f_w(w_i)} f_z(s) ds$;
- $\Phi_\beta = \mathcal{R}(T^*T)^{\frac{\beta}{2}}$ and $\Phi_\gamma = \mathcal{R}(T^*T)^{\frac{\gamma}{2}}$ for $\beta, \gamma > 0$;
- $\{\lambda_{jn}, \varphi_{jn}, \psi_{jn}\}_{j=1}^n$ is the singular value decomposition (SVD) of $T_{(n)}$, that is, $\{\lambda_{jn}^2\}_{j=1}^n$ are the nonzero eigenvalues of the selfadjoint operator $T_{(n)}T_{(n)}^*$ (and also of $T_{(n)}^*T_{(n)}$) written in decreasing order, $\lambda_{jn} > 0$ and the following formulas hold
$$T_{(n)}\varphi_{jn} = \lambda_{jn}\psi_{jn} \quad \text{and} \quad T_{(n)}^*\psi_{jn} = \lambda_{jn}\varphi_{jn}, \quad j = 1, \ldots, n \tag{15}$$
see e.g. Engl et al. (2000, Section 2.2);
- $C_n = (\frac{1}{n} I_n + T_{(n)}T_{(n)}^*)$.

A.1. Proof of Lemma 2

In this proof the limits are taken for $n \to \infty$. We say that the sequence of probability measures $\mu_n^{\sigma,y}$ on a Hilbert space $L_F^2(Z)$, endowed with the Borel σ-field $\mathfrak{E}$, converges weakly to a probability measure δ_{φ_*} if
$$\left\| \int a(\varphi)\mu_n^{\sigma,y}(d\varphi) - \int a(\varphi)\delta_{\varphi_*}(d\varphi) \right\| \to 0,$$
$P^{\sigma_*,\varphi_*,\mathbf{w}}$-a.s. (or in $P^{\sigma_*,\varphi_*,\mathbf{w}}$-probability)

for every bounded and continuous functional $a : L_F^2(Z) \to L_F^2(Z)$. The probability measure δ_{φ_*} denotes the Dirac measure on φ_*.

We prove that this convergence is not satisfied at least for one functional a. We consider the identity functional $a : \phi \mapsto \phi$, $\forall \phi \in L_F^2(Z)$, so that we have to check convergence of the posterior mean. For simplicity, we set $\varphi_0 = 0$, then the posterior mean is
$$\mathbb{E}(\varphi|y_{(n)}) = \Omega_0 K_{(n)}^* \left(\frac{1}{n} I_n + K_{(n)}\Omega_0 K_{(n)}^* \right)^{-1} y_{(n)}$$
and we have to prove that the L_F^2-norm $\|\mathbb{E}(\varphi|y_{(n)}) - \varphi_*\| \nrightarrow 0$ $P^{\sigma_*,\varphi_*,\mathbf{w}}$-a.s. By decomposing
$$\mathbb{E}(\varphi|y_{(n)}) - \varphi_* = \overbrace{\Omega_0 K_{(n)}^* \left(\frac{1}{n} I_n + K_{(n)}\Omega_0 K_{(n)}^* \right)^{-1} \varepsilon_{(n)}}^{\mathcal{A}_n} \underbrace{- \left(I - \Omega_0 K_{(n)}^* \left(\frac{1}{n} I_n + K_{(n)}\Omega_0 K_{(n)}^* \right)^{-1} K_{(n)}\right)\varphi_*}_{\mathcal{B}_n}$$
we get the lower bound: $\|\mathbb{E}(\varphi|y_{(n)}) - \varphi_*\| \geq |\|\mathcal{A}_n\| - \|\mathcal{B}_n\||$. Let us suppose that $\Omega_0^{\frac{1}{2}}$ has the same eigenfunctions $\{\varphi_{jn}\}$ as $K_{(n)}^*K_{(n)}$ where the first n eigenfunctions $\{\varphi_{jn}\}_{j=1}^n$ correspond to the nonzero eigenvalues λ_{jn}^2 of $K_{(n)}^*K_{(n)}$ and the remaining $\{\varphi_{jn}\}_{j>n}$ correspond to the zero eigenvalues $\lambda_{jn}^2 = 0$. Moreover, we assume that the eigenvalues of Ω_0 are of order j^{-2c}, for $c > \frac{1}{2}$ and the eigenvalues of $K_{(n)}^*K_{(n)}$ are of order j^{-2a} for $a > 0$, so that the eigenvalues of $T_{(n)}^*T_{(n)}$ are of order $j^{-2(c+a)}$. For this particular case we will prove that $\|\mathcal{A}_n\| \to \infty$ and $\|\mathcal{B}_n\| = \mathcal{O}(1)$ as $n \to \infty$.

We expand $\|\mathcal{A}_n\|$ by using the SVD of $T_{(n)}$ in the following way
$$\|\mathcal{A}_n\| = \left(\sum_{j=1}^n \left\langle \Omega_0^{\frac{1}{2}} T_{(n)}^* \left(\frac{1}{n} I_n + T_{(n)}T_{(n)}^* \right)^{-1} \varepsilon_{(n)}, \varphi_{jn} \right\rangle^2 + \sum_{j>n} \left\langle \Omega_0^{\frac{1}{2}} T_{(n)}^* \left(\frac{1}{n} I_n + T_{(n)}T_{(n)}^* \right)^{-1} \varepsilon_{(n)}, \varphi_{jn} \right\rangle^2 \right)^{\frac{1}{2}} \tag{16}$$
$$= \left(\sum_{j=1}^n \frac{j^{-2(2c+a)}}{\left(\frac{1}{n} + j^{-2(c+a)}\right)^2} \langle \varepsilon_{(n)}, \psi_{jn} \rangle^2 \right)^{\frac{1}{2}} \tag{17}$$
$$= \left(\frac{\sigma_*^2}{n} \sum_{j=1}^n \frac{j^{-2(2c+a)}}{\left(\frac{1}{n} + j^{-2(c+a)}\right)^2} \xi_j^2 \right)^{\frac{1}{2}} \tag{18}$$
$$= \left(\sigma_*^2 n \sum_{j=1}^n \frac{j^{-2(2c+a)}}{(1 + j^{-2(c+a)}n)^2} \xi_j^2 \right)^{\frac{1}{2}}$$
$$\geq \sigma_* \sqrt{n} \left(\sum_{j=1}^n \frac{j^{-2(2c+a)}}{(1 + j^{-2(c+a)}n)^2} (\xi_j^2 - 1) \right)^{\frac{1}{2}} \tag{19}$$
where the equality in (17) follows because the eigenvalues of $T_{(n)}^*T_{(n)}$ are zero for $j > n$. The random elements $\{\xi_j, j \geq 1\}$ are independent Gaussian random variables with $\mathbf{E}(\xi_j^2) = 1$ and $\mathrm{Var}(\xi_j^2) = 2$; the equality in (18) is a consequence of Assumption 2.

By the Khintchine–Kolmogorov Convergence Theorem, see e.g. Chow and Teicher (1997, p. 113), the sum in (19) converges with probability 1, as $n \uparrow \infty$ if its second moment $2\sum_{j=1}^n \frac{j^{-4(2c+a)}}{(1+j^{-2(c+a)}n)^4}$ is finite for $n \uparrow \infty$. In order to prove this, we remark that $(1 + j^{-2(c+a)}n)^{-4} \leq (1 + n^{-2(c+a)}n)^{-4}$. Then,
$$2\sum_{j=1}^n \frac{j^{-4(2c+a)}}{(1 + j^{-2(c+a)}n)^4} \leq 2\sum_{j=1}^n \frac{j^{-4(2c+a)}}{(1 + n^{1-2(c+a)})^4} \leq \frac{2}{(1 + n^{1-2(c+a)})^4} \sum_{j=1}^n j^{-4(2c+a)}$$
that is convergent for $n \uparrow \infty$ since $c > \frac{1}{2}$ guarantees that $4(2c+a) > 1$ and that $n^{1-2(c+a)} = n^{-b}$ for some $b > 0$ so that it converges to 0. Therefore, $\|\mathcal{A}_n\| \geq \sqrt{n}\sigma_*^2(\kappa_n)^{1/2}$ where κ_n is a random element that is convergent with probability 1. So, $\|\mathcal{A}_n\| \to \infty$ with probability 1. Next, we consider term $\mathcal{B}_n$:
$$\|\mathcal{B}_n\| = \left(\sum_{j=1}^\infty \left\langle \left(I - \Omega_0 K_{(n)}^* \left(\frac{1}{n} I_n + T_{(n)}^*T_{(n)} \right)^{-1} K_{(n)} \right) \varphi_*, \varphi_{jn} \right\rangle^2 \right)^{\frac{1}{2}}$$
$$= \left(\sum_{j=1}^n \left(1 - \frac{j^{-2(c+a)}}{\frac{1}{n} + j^{-2(a+c)}} \right)^2 \langle \varphi_*, \varphi_j \rangle^2 + \sum_{j>n} \langle \varphi_*, \varphi_j \rangle^2 \right)^{\frac{1}{2}}$$
$$= \left(\sum_{j=1}^n \frac{n^{-2}}{(n^{-1} + j^{-2(a+c)})^2} \langle \varphi_*, \varphi_j \rangle^2 + \sum_{j>n} \langle \varphi_*, \varphi_j \rangle^2 \right)^{\frac{1}{2}}$$
$$\leq \|\varphi_*\| < \infty.$$
This shows that, as $n \to \infty$, $\|\mathcal{A}_n\| - \|\mathcal{B}_n\| > 0$ with probability 1. We conclude that, with probability 1, $\|\mathbb{E}(\varphi|y_{(n)}) - \varphi_*\| \geq \|\mathcal{A}_n\| - \|\mathcal{B}_n\|$ which is increasing to ∞ as $n \to \infty$.

A.2. Proof of Theorem 2

(i) We develop $\hat{\varphi}_\alpha - \varphi_*$ in two terms:

$$\hat{\varphi}_\alpha - \varphi_*$$
$$= -\overbrace{\left(I - \Omega_0 K^*_{(n)}\left(\alpha_n I_n + \frac{1}{n}I_n + K_{(n)}\Omega_0 K^*_{(n)}\right)^{-1} K_{(n)}\right)(\varphi_* - \varphi_0)}^{\mathcal{A}}$$
$$+ \underbrace{\Omega_0 K^*_{(n)}\left(\alpha_n I_n + \frac{1}{n}I_n + K_{(n)}\Omega_0 K^*_{(n)}\right)^{-1}\varepsilon_{(n)}}_{\mathcal{B}}.$$

Under Assumption 4 we have the equations in Box I, since if $\delta_* \in \Phi_\beta$ and Assumption 5 holds, then $\|\alpha_n(\alpha_n I + T^*T)^{-1}\delta_*\| = \mathcal{O}(\alpha_n^{\beta/2})$, see Carrasco et al. (2007) and $\|T^*_{(n)}T_{(n)} - T^*T\|^2 \leq \mathbb{E}(\|T^*_{(n)}T_{(n)} - T^*T\|^2) = \mathcal{O}_p(n^{-1})$, where $\mathbb{E}(\cdot)$ is the expectation taken with respect to $f_w(w_i)$, because $\mathbb{E}(T^*_{(n)}T_{(n)}) = T^*T$ and $Var(T^*_{(n)}T_{(n)})$ is of order n^{-1}.

Next, we rewrite $\|\mathcal{A}2\| = \|\Omega_0^{\frac{1}{2}}(\alpha_n I + \frac{1}{n}I + T^*_{(n)}T_{(n)})^{-1}\frac{1}{n}T^*_{(n)}T_{(n)}(\alpha_n I + T^*_{(n)}T_{(n)})^{-1}\delta_*\|$ and by using similar developments as for $\mathcal{A}1$ we get $\|\mathcal{A}2\|^2 = \mathcal{O}(\frac{1}{\alpha_n^4 n^2}(\alpha_n^\beta + \frac{1}{\alpha_n^2 n}\alpha_n^\beta))$ which is negligible with respect to $\|\mathcal{A}1\|^2$.

Let us consider term $\mathcal{B}$. A similar decomposition as for $\mathcal{A}$ gives the inequalities in Box II and $T^*_{(n)}\varepsilon_{(n)} = \frac{1}{\sqrt{n}}\left[\frac{1}{\sqrt{n}}\sum_i \varepsilon_i g(Z, w_i)\right] = \frac{1}{\sqrt{n}}\mathcal{O}_p(1)$ because, by the Central Limit Theorem (CLT) the term in squared brackets converges toward a Gaussian random variable. Then $\|\mathcal{B}1\|^2 = \mathcal{O}_p(\frac{1}{\alpha_n^2 n})$. Lastly, $\|\mathcal{B}2\|^2 = \mathcal{O}_p(\frac{1}{\alpha_n^2 n^2}\frac{1}{\alpha_n^2 n})$ and since n^{-1} converges to zero faster than α_n, it is negligible with respect to $\|\mathcal{B}1\|^2$. Summarizing, $\|\hat{\varphi}_\alpha - \varphi_*\|^2 = \mathcal{O}_p((\alpha_n^\beta + \frac{1}{\alpha_n^2 n}\alpha_n^\beta)(1 + \frac{1}{\alpha_n^4 n^2}) + \frac{1}{\alpha_n^2 n}(1 + \frac{1}{\alpha_n^2 n^2}))$ that, after simplification of the terms that are negligible, becomes $\mathcal{O}_p(\alpha_n^\beta + \frac{1}{\alpha_n^2 n}\alpha_n^\beta + \frac{1}{\alpha_n^2 n})$ and then $\|\hat{\varphi}_\alpha - \varphi_*\|^2$ goes to zero if $\alpha_n \to 0$ and $\alpha_n^2 n \to \infty$.

To prove the intuition in Remark 3 we simply have to replace $\|\mathcal{B}\|^2$ with $\mathbb{E}\|\mathcal{B}\|^2$ so that $\|T^*_{(n)}\varepsilon_{(n)}\|^2$ is replaced by $\mathbb{E}\|T^*_{(n)}\varepsilon_{(n)}\|^2$ which is of order $\frac{1}{n}$ too.

(ii) By the Chebishev's Inequality, for a sequence ϵ_n with $\epsilon_n \to 0$,

$$\mu_\alpha^{\sigma,y}\{\varphi \in L^2_F(Z); \|\varphi - \varphi_*\| \geq \epsilon_n\}$$
$$\leq \frac{\mathbb{E}_\alpha(\|\varphi - \varphi_*\|^2 | y_{(n)}, \sigma^2)}{\epsilon_n^2}$$
$$= \frac{1}{\epsilon_n^2}\left(\|\hat{\varphi}_\alpha - \varphi_*\|^2 + \sigma^2 tr\Omega_{y,\alpha}\right)$$

where $\mathbb{E}_\alpha(\cdot|y_{(n)}, \sigma^2)$ denotes the expectation taken with respect to $\mu_\alpha^{\sigma,y}$. Since the equation in Box III holds, then, $tr(\Omega_{y,\alpha}) = tr(\mathcal{C}) + tr(\mathcal{D})$. By using properties and the definition of the trace function, we get

$$\lim_{n\to\infty} tr(\mathcal{C}) = \lim_{n\to\infty} tr[\alpha_n(\alpha_n I + T^*_{(n)}T_{(n)})^{-1}\Omega_0]$$
$$= \lim_{n\to\infty}\sum_{j=1}^n \frac{\alpha_n}{\alpha_n + \lambda_{jn}^2}\langle\Omega_0\varphi_{jn}, \varphi_{jn}\rangle$$
$$= \lim_{n\to\infty}\sum_{j=1}^n \frac{\alpha_n\lambda_{jn}^{2\kappa}}{\alpha_n + \lambda_{jn}^2}\frac{\langle\Omega_0\varphi_{jn}, \varphi_{jn}\rangle}{\lambda_{jn}^{2\kappa}}$$
$$\leq \alpha_n^\kappa \lim_{n\to\infty}\sum_{j=1}^n \frac{\langle\Omega_0\varphi_{jn}, \varphi_{jn}\rangle}{\lambda_{jn}^{2\kappa}}$$

which is an $\mathcal{O}_p(\alpha_n^\kappa)$ under the assumption that $\lim_{n\to\infty}\sum_{j=1}^n \frac{\langle\Omega_0\varphi_{jn},\varphi_{jn}\rangle}{\lambda_{jn}^{2\kappa}} < \infty$. Then, $tr(\mathcal{C}) \to 0$ as $\alpha_n \to 0$. The $tr(\mathcal{D})$ is less than or equal to $tr[T_{(n)}\Omega_0 T^*_{(n)}(\alpha_n I_n + T_{(n)}T^*_{(n)})^{-1}]$ and in a similar way as for term $tr(\mathcal{C})$, it is easy to prove that $tr(\mathcal{D}) = \mathcal{O}(\alpha^\kappa\frac{1}{n})$. By the Kolmogorov's Theorem, $\sigma^2 = \mathcal{O}_p(1)$ since $\mathbb{E}[\sigma^2|y_{(n)}] = \mathcal{O}_p(1)$ by Theorem 3. Then, $\sigma^2 tr(\Omega_{y,\alpha}) \to 0$ and by using the result on convergence of $\|\hat{\varphi}_\alpha - \varphi_*\|$ in (i) we can conclude.

(iii) We use the decomposition (20) (where the first term does not include $\frac{1}{n}I_n$ and the second one does). We have to consider the squared norm in $L^2_F(Z)$ of $\sigma^2\Omega_{y,\alpha}\phi$: $\|\sigma^2\Omega_{y,\alpha}\phi\| \leq |\sigma^2|(\|\mathcal{C}\phi\| + \|\mathcal{D}\phi\|)$. By the Kolmogorov's Theorem $|\sigma^2| = \mathcal{O}_p(1)$ if and only if $\mathbb{E}[(\sigma^2)^2|y_{(n)}] = \mathcal{O}_p(1)$. Since the second moment of σ^2 is $\mathbb{E}[(\sigma^2)^2|y_{(n)}] = Var(\sigma^2|y_{(n)}) + \mathbb{E}^2(\sigma^2|y_{(n)})$, it follows from Theorem 3 that $|\sigma^2|^2 = \mathcal{O}_p(1)$. Moreover,

$$\|\mathcal{C}\phi\|^2 \leq \|\Omega_0^{\frac{1}{2}}\|^2\|[I - (\alpha_n I + T^*_{(n)}T_{(n)})^{-1}T^*_{(n)}T_{(n)}]\Omega_0^{\frac{1}{2}}\phi\|^2$$
$$= \|\Omega_0^{\frac{1}{2}}\|^2\|\alpha_n(\alpha_n I + T^*_{(n)}T_{(n)})^{-1}\Omega_0^{\frac{1}{2}}\phi\|^2$$
$$\leq \|\Omega_0^{\frac{1}{2}}\|^2(\|\alpha_n(\alpha_n I + T^*T)^{-1}\Omega_0^{\frac{1}{2}}\phi\|^2$$
$$+ \|\alpha_n[(\alpha_n I + T^*_{(n)}T_{(n)})^{-1} - (\alpha_n I + T^*T)^{-1}]\Omega_0^{\frac{1}{2}}\phi\|^2)$$

and $\|\alpha_n(\alpha_n I + T^*T)^{-1}\Omega_0^{\frac{1}{2}}\phi\|^2 = \mathcal{O}(\alpha_n^\beta)$ if $\Omega_0^{\frac{1}{2}}\phi \in \Phi_\beta$ and T is one-to-one on $L^2_F(Z)$. Moreover, the second term in brackets is an $\mathcal{O}_p(\frac{1}{\alpha_n^2 n}\alpha_n^\beta)$ and $\|\Omega_0^{\frac{1}{2}}\|^2 = \mathcal{O}(1)$ since Ω_0 is a compact operator so we get $\|\mathcal{C}\phi\|^2 = \mathcal{O}\left(\alpha_n^\beta + \frac{1}{\alpha_n^2 n}\alpha_n^\beta\right)$.

Term $\|\mathcal{D}\phi\|^2$ is equivalent to term $\|\mathcal{A}2\|^2$ in point (i) except that δ_* is replaced by $\Omega_0^{\frac{1}{2}}\phi$, but this does not modify the rate of convergence since both these two elements belong to the β-regularity space Φ_β. Hence, $\|\mathcal{D}\phi\|^2 = \mathcal{O}(\frac{1}{\alpha_n^4 n^2}(\alpha_n^\beta + \frac{1}{\alpha_n^2 n}\alpha_n^\beta))$. Summarizing, $\|\Omega_{y,\alpha}\phi\|^2 = \mathcal{O}_p((1 + \frac{1}{\alpha_n^4 n^2})(\alpha_n^\beta + \frac{1}{\alpha_n^2 n}\alpha_n^\beta))$ which becomes $\mathcal{O}(\alpha_n^\beta + \frac{1}{\alpha_n^2 n}\alpha_n^\beta)$ once the fastest terms are neglected and which implies that $\|\sigma^2\Omega_{y,\alpha}\phi\| \to 0$ in $P^{\sigma_*,\varphi_*,\mathbf{w}}$-probability.

A.3. Proof of Theorem 3

The posterior mean $\mathbb{E}(\sigma^2|y_{(n)})$ is asymptotically equal to

$$\mathbb{E}(\sigma^2|y_{(n)}) \approx \frac{1}{n}(y_{(n)} - K_{(n)}\varphi_0)'C_n^{-1}(y_{(n)} - K_{(n)}\varphi_0)$$
$$= \overbrace{\frac{1}{n}(K_{(n)}(\varphi_* - \varphi_0))'C_n^{-1}(K_{(n)}(\varphi_* - \varphi_0))}^{\mathcal{A}}$$
$$+ \underbrace{\frac{2}{n}(K_{(n)}(\varphi_* - \varphi_0))'C_n^{-1}\varepsilon_{(n)}}_{\mathcal{B}} + \underbrace{\frac{1}{n}\varepsilon'_{(n)}C_n^{-1}\varepsilon_{(n)}}_{\mathcal{C}}.$$

Under Assumption 4,

$$\mathcal{A} = \frac{1}{n}\langle K_{(n)}\Omega_0^{\frac{1}{2}}\delta_*, C_n^{-1}K_{(n)}\Omega_0^{\frac{1}{2}}\delta_*\rangle$$
$$= \frac{1}{n}\langle\delta_*, T^*_{(n)}C_n^{-1}T_{(n)}\delta_*\rangle$$
$$\leq \frac{1}{n}\|\delta_*\|\left\|\left(\frac{1}{n}I + T^*_{(n)}T_{(n)}\right)^{-1}T^*_{(n)}T_{(n)}\right\|\|\delta_*\| = \mathcal{O}\left(\frac{1}{n}\right)$$

since $\|(\frac{1}{n}I + T^*_{(n)}T_{(n)})^{-1}T^*_{(n)}T_{(n)}\| = \mathcal{O}(1)$.

$$\|\mathcal{A}\| \le \left\|\overbrace{(I-\Omega_0^{\frac{1}{2}}T^*_{(n)}(\alpha_n I_n + T_{(n)}T^*_{(n)})^{-1}K_{(n)})\Omega_0^{\frac{1}{2}}\delta_*}^{\mathcal{A}1}\right\| + \left\|\underbrace{\Omega_0^{\frac{1}{2}}T^*_{(n)}\left(\alpha_n I_n + \frac{1}{n}I_n + T_{(n)}T^*_{(n)}\right)^{-1}\frac{1}{n}I_n(\alpha_n I_n + T_{(n)}T^*_{(n)})^{-1}T_{(n)}\delta_*}_{\mathcal{A}2}\right\|$$

$$\|\mathcal{A}1\| = \left\|\Omega_0^{\frac{1}{2}}\left[\alpha_n(\alpha_n I + T^*T)^{-1}\delta_* + \alpha_n[(\alpha_n I + T^*_{(n)}T_{(n)})^{-1} - (\alpha_n I + T^*T)]\delta_*\right]\right\|$$
$$\le \|\Omega_0^{\frac{1}{2}}\|\left(\|\alpha_n(\alpha_n I + T^*T)^{-1}\delta_*\| + \|(\alpha_n I + T^*_{(n)}T_{(n)})^{-1}\|\|T^*_{(n)}T_{(n)} - T^*T\|\|\alpha_n(\alpha_n I + T^*T)^{-1}\delta_*\|\right)$$
$$\|\mathcal{A}1\|^2 = \mathcal{O}_p\left(\alpha_n^\beta + \frac{1}{\alpha_n^2 n}\alpha_n^\beta\right)$$

Box I.

$$\|\mathcal{B}\|^2 \le \|\Omega_0^{\frac{1}{2}}\|^2\left(\left\|\underbrace{T^*_{(n)}(\alpha_n I_n + T_{(n)}T^*_{(n)})^{-1}\varepsilon_{(n)}}_{\mathcal{B}1}\right\|^2 + \left\|\underbrace{T^*_{(n)}\left(\alpha_n I_n + \frac{1}{n}I_n + T_{(n)}T^*_{(n)}\right)^{-1}\frac{1}{n}I_n(\alpha_n I_n + T_{(n)}T^*_{(n)})^{-1}\varepsilon_{(n)}}_{\mathcal{B}2}\right\|^2\right)$$
$$\|\mathcal{B}1\|^2 \le \|(\alpha_n I + T^*_{(n)}T_{(n)})^{-1}\|^2\|T^*_{(n)}\varepsilon_{(n)}\|^2$$

Box II.

$$\Omega_{y,\alpha} = \overbrace{\Omega_0^{\frac{1}{2}}[I - T^*_{(n)}(\alpha_n I_n + T_{(n)}T^*_{(n)})^{-1}T_{(n)}]\Omega_0^{\frac{1}{2}}}^{\mathcal{C}} + \underbrace{\Omega_0^{\frac{1}{2}}T^*_{(n)}[(\alpha_n I_n + T_{(n)}T^*_{(n)})^{-1} - (\alpha_n I_n + \frac{1}{n}I_n + T_{(n)}T^*_{(n)})^{-1}]T_{(n)}\Omega_0^{\frac{1}{2}}}_{\mathcal{D}} \quad (20)$$

Box III.

Term $\mathcal{C}$ requires a little bit more computation. First we have to remark that, by the Binomial Inverse Theorem, $C_n^{-1} = nI_n - n^2T_{(n)}(I + nT^*_{(n)}T_{(n)})^{-1}T^*_{(n)}$; hence,

$$\mathcal{C} = \varepsilon'_{(n)}\varepsilon_{(n)} - n\varepsilon'_{(n)}T_{(n)}(I + nT^*_{(n)}T_{(n)})^{-1}T^*_{(n)}\varepsilon_{(n)} \quad (21)$$
$$\mathcal{C} - \sigma_*^2 \le (\varepsilon'_{(n)}\varepsilon_{(n)} - \sigma_*^2) + n\varepsilon'_{(n)}T_{(n)}(I + nT^*_{(n)}T_{(n)})^{-1}T^*_{(n)}\varepsilon_{(n)}.$$

It is easy to see that $\varepsilon'_{(n)}\varepsilon_{(n)} - \sigma_*^2 = \mathcal{O}_p(\frac{1}{\sqrt{n}})$ and that

$$T^*_{(n)}\varepsilon_{(n)} = \frac{1}{n}\sum_i \varepsilon_i g(Z, w_i)$$
$$n\left(I + nT^*_{(n)}T_{(n)}\right)^{-1}T^*_{(n)}\varepsilon_{(n)} = \frac{1}{n}\sum_i \varepsilon_i\left(\left(\frac{1}{n}I + T^*_{(n)}T_{(n)}\right)^{-1}g(Z, w_i)\right).$$

The second term in (21) becomes

$$n\varepsilon'_{(n)}T_{(n)}(I + nT^*_{(n)}T_{(n)})^{-1}T^*_{(n)}\varepsilon_{(n)} = \left\langle T^*_{(n)}\varepsilon_{(n)}, \left(\frac{1}{n}I + T^*_{(n)}T_{(n)}\right)^{-1}T^*_{(n)}\varepsilon_{(n)}\right\rangle$$
$$\le \|T^*_{(n)}\varepsilon_{(n)}\|\left\|\left(\frac{1}{n}I + T^*_{(n)}T_{(n)}\right)^{-1}T^*_{(n)}\varepsilon_{(n)}\right\|.$$

The first norm is an $\mathcal{O}_p(\frac{1}{\sqrt{n}})$ since $\|T^*_{(n)}\varepsilon_{(n)}\| \le \frac{1}{\sqrt{n}}[\frac{1}{\sqrt{n}}\sum_i \varepsilon_i\|g(Z, w_i)\|]$ and the term in squared brackets is an $\mathcal{O}_p(1)$ because it converges toward a Gaussian random variable (by the CLT).

If $g(Z, w_i) \in \Phi_\gamma$, for $\gamma > 1$, then there exists a function $h(Z, w_i) \in L^2_F(Z)$ such that $g = (T^*T)^{\frac{\gamma}{2}}h(Z, w_i)$ and hence

$$\left\|\left(\frac{1}{n}I + T^*_{(n)}T_{(n)}\right)^{-1}T^*_{(n)}\varepsilon_{(n)}\right\| = \left\|\overbrace{\frac{1}{n}\sum_i \varepsilon_i\left(\left(\frac{1}{n}I + T^*T\right)^{-1}(T^*T)^{\frac{\gamma}{2}}h(Z, w_i)\right)}^{\mathcal{C}1}\right\| + \left\|\underbrace{\frac{1}{n}\sum_i \varepsilon_i\left(\left[\left(\frac{1}{n}I + T^*_{(n)}T_{(n)}\right)^{-1} - \left(\frac{1}{n}I + T^*T\right)^{-1}\right]g(Z, w_i)\right)}_{\mathcal{C}2}\right\|$$

$$\|\mathcal{C}1\| \le \frac{n}{\sqrt{n}}\frac{1}{\sqrt{n}} \times \sum_i |\varepsilon_i|\underbrace{\left\|\frac{1}{n}\left(\frac{1}{n}I + T^*T\right)^{-1}(T^*T)^{\frac{\gamma}{2}}\right\|}_{=\mathcal{O}_p\left(n^{-\frac{\gamma}{2}}\right)}\|h(Z, w_i)\|$$
$$= \mathcal{O}_p(\sqrt{n}n^{-\frac{\gamma}{2}}) = \mathcal{O}_p\left(\left(\frac{1}{\sqrt{n}}\right)^{\gamma-1}\right)$$

$$\|\mathcal{C}2\| \leq \frac{1}{n}\sum_i |\varepsilon_i| \left\| \left(\frac{1}{n}I + T^*_{(n)}T_{(n)}\right)^{-1} \right\| \|T^*_{(n)}T_{(n)} - T^*T\| \times \left\| \left(\frac{1}{n}I + T^*T\right)^{-1} (T^*T)^{\frac{\gamma}{2}} \right\| \|h(Z, w_i)\| = \mathcal{O}_p\left(\left(\frac{1}{\sqrt{n}}\right)^{\gamma-1}\right)$$

which converges at the same rate as $\|\mathcal{C}1\|$. Hence, $(C - \sigma_*^2) = \mathcal{O}_p(\frac{1}{\sqrt{n}} + (\frac{1}{n})^{\frac{\gamma}{2}})$. Finally,

$$\mathcal{B} = \frac{2}{n}\langle \varepsilon_{(n)}, C_n^{-1}T_{(n)}\delta_*\rangle = \frac{2}{n}\left\langle T^*_{(n)}C_n^{-1}\varepsilon_{(n)}, \delta_*\right\rangle \leq \frac{2}{n}\|\delta_*\|\|T^*_{(n)}C_n^{-1}\varepsilon_{(n)}\| = \mathcal{O}_p\left(\left(\frac{1}{n}\right)^{\frac{\gamma+1}{2}}\right)$$

since $\|T^*_{(n)}C_n^{-1}\varepsilon_{(n)}\| = \|(\frac{1}{n}I + T^*_{(n)}T_{(n)})^{-1}T^*_{(n)}\varepsilon_{(n)}\|$ and its rate has been computed for term $\mathcal{C}$.

Therefore, $\mathbb{E}(\sigma^2|y_{(n)}) - \sigma_*^2 = \mathcal{O}_p\left(\frac{1}{n} + \frac{1}{\sqrt{n}} + \left(\frac{1}{\sqrt{n}}\right)^{\gamma+1} + \left(\frac{1}{\sqrt{n}}\right)^{\gamma}\right) = \mathcal{O}_p((\frac{1}{\sqrt{n}})^{\gamma\wedge 1})$.

By the Chebishev's Inequality,

$$\nu_n^y\{\sigma \in \mathbb{R}_+; |\sigma^2 - \sigma_*^2| \geq \epsilon_n\} \leq \mathbb{E}[(\sigma^2 - \sigma_*^2)^2|y_{(n)}]\frac{1}{\epsilon_n^2} = \frac{1}{\epsilon_n^2}\left[\mathrm{Var}(\sigma^2|y_{(n)}) + (\mathbb{E}(\sigma^2|y_{(n)}) - \sigma_*^2)^2\right].$$

Term $(\mathbb{E}(\sigma^2|y_{(n)}) - \sigma_*^2)^2$ converges to 0 and it is of order $(\frac{1}{n})^{\gamma\wedge 1}$; the variance is $\mathrm{Var}(\sigma^2|y_{(n)}) = 2\mathbb{E}(\sigma^2|y_{(n)})\frac{1}{\xi_0+n-2}$ and it goes to 0 faster than the squared bias. Then, the posterior probability of the complement of any neighborhood of σ_*^2 converges to 0.

A.4. Proof of Corollary 1

Let us remark that

$$\begin{aligned}
\|(\sigma^2, \varphi) - (\sigma_*^2, \varphi_*)\|_{\mathbb{R}_+\times L_F^2(Z)} &= \|(\sigma^2 - \sigma_*^2, \varphi - \varphi_*)\|_{\mathbb{R}_+\times L_F^2(Z)} \\
&= \sqrt{\langle(\sigma^2 - \sigma_*^2, \varphi - \varphi_*), (\sigma^2 - \sigma_*^2, \varphi - \varphi_*)\rangle_{\mathbb{R}_+\times L_F^2(Z)}} \\
&= \sqrt{\langle(\sigma^2 - \sigma_*^2), (\sigma^2 - \sigma_*^2)\rangle_{\mathbb{R}_+} + \langle(\varphi - \varphi_*), (\varphi - \varphi_*)\rangle_{L_F^2(Z)}} \\
&= (\|\sigma^2 - \sigma_*^2\|^2_{\mathbb{R}_+} + \|\varphi - \varphi_*\|^2_{L_F^2(Z)})^{\frac{1}{2}} \\
&\leq \left((\|\sigma^2 - \sigma_*^2\|_{\mathbb{R}_+} + \|\varphi - \varphi_*\|_{L_F^2(Z)})^2\right)^{\frac{1}{2}} \\
&= \|\sigma^2 - \sigma_*^2\|_{\mathbb{R}_+} + \|\varphi - \varphi_*\|_{L_F^2(Z)}
\end{aligned}$$

where we have specified the space to which each norm refers. Then,

$$\begin{aligned}
&\nu_n^y \times \mu_\alpha^{\sigma,y}\{(\sigma^2, \varphi) \in \mathbb{R}_+ \times L_F^2(Z), \|(\sigma^2, \varphi) - (\sigma_*^2, \varphi_*)\|_{\mathbb{R}_+\times L_F^2(Z)} > \epsilon_n\} \\
&\leq \nu_n^y \times \mu_\alpha^{\sigma,y}\{(\sigma^2, \varphi) \in \mathbb{R}_+ \times L_F^2(Z), \|\sigma^2 - \sigma_*^2\|_{\mathbb{R}_+} + \|\varphi - \varphi_*\|_{L_F^2(Z)} > \epsilon_n\} \\
&= \mathbb{E}^y(\mu_\alpha^{\sigma,y}\{\varphi \in L_F^2(Z); \|\varphi - \varphi_*\|_{L_F^2(Z)} > \epsilon_n - \|\sigma^2 - \sigma_*^2\|_{\mathbb{R}_+}\}|y_{(n)}),
\end{aligned}$$

with $\mathbb{E}^y(\cdot|y_{(n)})$ denoting the expectation taken with respect to ν_n^y. Since $\mu_\alpha^{\sigma,y}$ is a bounded and continuous function of σ^2, by definition of weak convergence of a probability measure and by Theorem 3, this expectation converges in $\mathbb{R}_+$-norm toward $\mu_\alpha^{\sigma_*,y}\{\varphi \in L_F^2(Z); \|\varphi - \varphi_*\|_{L_F^2(Z)} > \epsilon_n\}$ which converges to 0 by Theorem 2.

A.5. Proof of Theorem 4

The proof is very similar to that one for Theorem 2(i), then we shorten it as much as possible. We use the decomposition in Box IV, where we have used Assumptions 4 and 5 and $\delta_* \in \mathcal{R}(T^*T)^{\frac{\beta}{2}}$. From the results in Appendix B in Darolles et al. (2011) and the smoothing effect of operator $\Omega_0^{\frac{1}{2}}$ in $\hat{T}^*_{(n)}$ and T^*, it follows that $\|\hat{T}^*_{(n)}\hat{T}_{(n)} - T^*T\|^2 = \mathcal{O}_p(\frac{1}{n} + h^{2\rho})$. This implies that $\|\mathcal{A}\|^2 = \mathcal{O}_p(\alpha_n^\beta + \alpha_n^{\beta-2}(\frac{1}{n} + h^{2\rho}))$ and $\|\mathcal{B}\|^2 = \mathcal{O}_p(\frac{1}{\alpha_n^2 n}\alpha_n^\beta + \frac{1}{\alpha_n^2 n}(\frac{1}{n} + h^{2\rho})\alpha_n^{\beta-2})$.

Lastly, term $\|\mathcal{C}\|^2$ can be rewritten as

$$\|\mathcal{C}\| \leq \|\Omega_0^{\frac{1}{2}}\| \left\|\overbrace{\hat{T}^*_{(n)}(\alpha_n I_n + \hat{T}_{(n)}\hat{T}^*_{(n)})^{-1}(\eta_{(n)} + \varepsilon_{(n)})}^{\mathcal{C}1}\right\| + \left\|\overbrace{\hat{T}^*_{(n)}(\alpha_n I_n + \Sigma_n + \hat{T}_{(n)}\hat{T}^*_{(n)})^{-1}\Sigma_n(\alpha_n I_n + \hat{T}_{(n)}\hat{T}^*_{(n)})^{-1}(\eta_{(n)} + \varepsilon_{(n)})}^{\mathcal{C}2}\right\|$$

$$\begin{aligned}
\|\mathcal{C}1\|^2 &\leq \|(\alpha_n I + \hat{T}^*_{(n)}\hat{T}_{(n)})^{-1}\|^2\|\hat{T}^*_{(n)}(\eta_{(n)} + \varepsilon_{(n)})\|^2 \\
&= (\|(\alpha_n I + T^*T)^{-1}\| + \|(\alpha_n I + \hat{T}^*_{(n)}\hat{T}_{(n)})^{-1}(\hat{T}^*_{(n)}\hat{T}_{(n)} - T^*T)(\alpha_n I + T^*T)\|)^2 \\
&\quad \times \|T^*_{(n)}(\eta_{(n)} + \varepsilon_{(n)}) + (\hat{T}^*_{(n)} - T^*_{(n)})(\eta_{(n)} + \varepsilon_{(n)})\|^2 \\
&= \mathcal{O}_p\left(\frac{1}{\alpha_n^2 n} + \frac{1}{\alpha^2}\left(\frac{1}{n} + h^{2\rho}\right)\frac{1}{\alpha_n^2 n}\right)
\end{aligned}$$

since $\|\hat{T}^*_{(n)} - T^*_{(n)}\|^2 \sim \|\hat{T}^*_{(n)}\hat{T}_{(n)} - T^*T\|^2 = \mathcal{O}_p(\frac{1}{n} + h^{2\rho})$. Term $\mathcal{C}2$ is developed as

$$\|\mathcal{C}2\|^2 \leq \|\Omega_0^{\frac{1}{2}}\|^2 \left\|\left(\alpha_n I + \left(\frac{\sigma^2}{n} + o_p\left(\frac{1}{n}\right)\right)I + \hat{T}^*_{(n)}\hat{T}_{(n)}\right)^{-1}\right\|^2 \left\|\left(\frac{\sigma^2}{n} + o_p\left(\frac{1}{n}\right)\right)I\right\|^2 \times \|\hat{T}^*_{(n)}(\alpha_n I_n + \hat{T}_{(n)}\hat{T}^*_{(n)})^{-1}(\eta_{(n)} + \varepsilon_{(n)})\|^2$$

where the last norm is the same as term $\mathcal{C}1$. Hence, $\|\mathcal{C}2\|^2 = \mathcal{O}_p(\frac{1}{\alpha_n^2 n} + \frac{1}{\alpha_n^2}(\frac{1}{n} + h^{2\rho})\frac{1}{\alpha_n^2 n})$ and $\|\hat{\mathbb{E}}_\alpha(\varphi|y_{(n)}) - \varphi_*\|^2 = \mathcal{O}_p(\alpha_n^\beta + \frac{1}{\alpha_n^2 n} + \frac{1}{\alpha_n^2}\left(\frac{1}{n} + h^{2\rho}\right)\frac{1}{\alpha_n^2 n})$ after having eliminated the negligible terms.

A.6. Proof of Lemma 3

We give a brief sketch of the proof and we refer to Fève and Florens (2010) for a more detailed proof. Let $R^\alpha = (\alpha_n I_n + T_{(n)}T^*_{(n)})^{-1}$ and $R^\alpha_{(n)} = (\alpha_n I_n + \frac{1}{n}I_n + T_{(n)}T^*_{(n)})^{-1}$. We decompose the residual as in Box V.

Standard computations similar to those used in previous proofs allow us to show that: $\|\mathcal{A}\|^2 = \mathcal{O}_p(\alpha_n^{\beta+2} + \frac{1}{n})$, $\|\mathcal{B}\|^2 = \mathcal{O}_p(\frac{1}{n^2} + \frac{1}{\alpha_n^2 n^2} + \frac{\alpha_n^2}{n})$, $\|\mathcal{C}\|^2 = \mathcal{O}_p(\frac{1}{n} + \frac{1}{\alpha_n^2 n^2})$, $\|\mathcal{D}\|^2 = \mathcal{O}_p(\frac{1}{\alpha_n^2 n^3} + \frac{1}{\alpha_n^4 n^3})$.

A.7. Proof of Lemma 4

We give a brief sketch of the proof and we refer to Fève and Florens (2010) for a more detailed proof. It is the same as the proof of Lemma 3 with $T_{(n)}$, $T^*_{(n)}$, $K_{(n)}$ and $K^*_{(n)}$ replaced by $\hat{T}_{(n)}$, $\hat{T}^*_{(n)}$,

$$\begin{aligned}
\hat{\mathbb{E}}_\alpha(\varphi|y_{(n)}) - \varphi_* \;=\; & -\overbrace{(I - \Omega_0^{\frac{1}{2}}\hat{T}^*_{(n)}(\alpha_n I_n + \hat{T}_{(n)}\hat{T}^*_{(n)})^{-1}\hat{K}_{(n)})(\varphi_* - \varphi_0)}^{\mathcal{A}} \\
& + \underbrace{\Omega_0^{\frac{1}{2}}\hat{T}^*_{(n)}[(\alpha_n I_n + \Sigma_n + \hat{T}_{(n)}\hat{T}^*_{(n)})^{-1} - (\alpha_n I_n + \hat{T}_{(n)}\hat{T}^*_{(n)})^{-1}]\hat{K}_{(n)}(\varphi_* - \varphi_0)}_{\mathcal{B}} \\
& + \underbrace{\Omega_0^{\frac{1}{2}}\hat{T}^*_{(n)}(\alpha_n I_n + \Sigma_n + \hat{T}_{(n)}\hat{T}^*_{(n)})^{-1}(\eta_{(n)} + \varepsilon_{(n)})}_{\mathcal{C}}
\end{aligned}$$

$$\begin{aligned}
\|\mathcal{A}\|^2 \;\leq\; & \|\Omega_0^{\frac{1}{2}}\|^2 \|\alpha_n(\alpha_n I + \hat{T}^*_{(n)}\hat{T}_{(n)})^{-1}\delta_*\|^2 \\
\leq\; & \|\Omega_0^{\frac{1}{2}}\|^2 \left(\|\alpha_n(\alpha_n I + T^*T)^{-1}\delta_*\| + \|\alpha_n(\alpha_n I + \hat{T}^*_{(n)}\hat{T}_{(n)})^{-1}(T^*T - \hat{T}^*_{(n)}\hat{T}_{(n)})(\alpha_n I + T^*T)^{-1}\delta_*\|\right)^2 \\
=\; & \mathcal{O}_p(\alpha_n^\beta + \alpha_n^{\beta-2}\|\hat{T}^*_{(n)}\hat{T}_{(n)} - T^*T\|^2)
\end{aligned}$$

$$\|\mathcal{B}\|^2 \leq \|\Omega_0^{\frac{1}{2}}\|^2 \left\|\left(\alpha_n I + \left(\frac{\sigma^2}{n} + o_p\left(\frac{1}{n}\right)\right) I + \hat{T}^*_{(n)}\hat{T}_{(n)}\right)^{-1}\right\|^2 \left\|\left(\frac{\sigma^2}{n} + o_p\left(\frac{1}{n}\right)\right) I\right\|^2 \left\|\underbrace{\hat{T}^*_{(n)}(\alpha_n I_n + \hat{T}_{(n)}\hat{T}^*_{(n)})^{-1}\hat{T}_{(n)}\delta_*}_{\mathcal{B}1}\right\|^2$$

$$\begin{aligned}
\mathcal{B}1 \;=\; & (\alpha_n I + \hat{T}^*_{(n)}\hat{T}_{(n)})^{-1}\hat{T}^*_{(n)}\hat{T}_{(n)}\delta_* \\
=\; & (\alpha_n I + T^*T)^{-1}T^*T\delta_* + [(\alpha_n I + \hat{T}^*_{(n)}\hat{T}_{(n)})^{-1}\hat{T}^*_{(n)}\hat{T}_{(n)} - (\alpha_n I + T^*T)^{-1}T^*T]\delta_* \\
=\; & (\alpha_n I + T^*T)^{-1}T^*T\delta_* + (\alpha_n I + \hat{T}^*_{(n)}\hat{T}_{(n)})^{-1}(\hat{T}^*_{(n)}\hat{T}_{(n)} - T^*T)\alpha_n(\alpha_n I + T^*T)^{-1}\delta_*
\end{aligned}$$

$$\|\mathcal{B}1\|^2 = \mathcal{O}_p\left(\alpha_n^\beta + \frac{1}{\alpha_n^2}\|\hat{T}^*_{(n)}\hat{T}_{(n)} - T^*T\|^2\alpha_n^\beta\right)$$

Box IV.

$$\begin{aligned}
\vartheta_\alpha^{(2)} =\; & \overbrace{T^*_{(n)}[I - (\alpha_n T_{(n)}T^*_{(n)}R^\alpha + T_{(n)}T^*_{(n)})R^\alpha]K_{(n)}(\varphi_* - \varphi_0)}^{\mathcal{A}} \\
& + \overbrace{T^*_{(n)}[(\alpha_n T_{(n)}T^*_{(n)}R^\alpha + T_{(n)}T^*_{(n)})R^\alpha - (\alpha_n T_{(n)}T^*_{(n)}R^\alpha_{(n)} + T_{(n)}T^*_{(n)})R^\alpha_{(n)}]K_{(n)}(\varphi_* - \varphi_0)}^{\mathcal{B}} \\
& + \overbrace{T^*_{(n)}[I - (\alpha_n T_{(n)}T^*_{(n)}R^\alpha + T_{(n)}T^*_{(n)})R^\alpha]\varepsilon_{(n)}}^{\mathcal{C}} \\
& + \underbrace{T^*_{(n)}[(\alpha_n T_{(n)}T^*_{(n)}R^\alpha + T_{(n)}T^*_{(n)})R^\alpha - (\alpha_n T_{(n)}T^*_{(n)}R^\alpha_{(n)} + T_{(n)}T^*_{(n)})R^\alpha_{(n)}]\varepsilon_{(n)}}_{\mathcal{D}}.
\end{aligned}$$

Box V.

$\hat{K}_{(n)}$ and $\hat{K}^*_{(n)}$. Then, we have the same decomposition and we get: $\|\mathcal{A}\|^2 = \mathcal{O}_p(\alpha_n^{\beta+2} + (\frac{1}{n} + h^{2\rho}))$, $\|\mathcal{B}\|^2 = \mathcal{O}_p(\alpha_n^{\beta+2} + (\frac{1}{n} + h^{2\rho})\alpha_n^\beta)$, $\|\mathcal{C}\|^2 = \mathcal{O}_p(\frac{1}{\alpha_n^2 n}(\frac{1}{n} + h^{2\rho}))$, $\|\mathcal{D}\|^2 = \mathcal{O}_p(\frac{1}{n} + \frac{1}{\alpha_n^2 n}(\frac{1}{n} + h^{2\rho}))$.

References

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

Blundell, R., Chen, X., Kristensen, D., 2007. Semi-nonparametric IV estimation of shape-invariant Engel curves. Econometrica 75, 1613–1669.

Blundell, R., Powell, J.L., 2003. Endogeneity in nonparametric and semiparametric regression models. In: Dewatripont, M., Hansen, L.-P., Turnovsky, D.J. (Eds.), Advances in Economics and Econometrics: Theory and Applications, vol. 2. UK, Cambridge University Press, Cambridge, pp. 312–357.

Breiman, L., Le Cam, L., Schwartz, L., 1964. Consistent estimates and zero–one sets. Annals of Mathematical Statistics 35, 157–161.

Carrasco, M., Florens, J.P., Renault, E., 2007. Linear inverse problems in structural econometrics: estimation based on spectral decomposition and regularization. In: Heckman, J., Leamer, E. (Eds.), Handbook of Econometrics, Vol. 6B. Elsevier, North Holland, pp. 5633–5751.

Chen, X., Reiss, M., 2011. On rate optimality for ill-posed inverse problems in econometrics. Econometric Theory 27, 497–521.

Chen, X., White, H., 1998. Central limit and functional central limit theorems for Hilbert-valued dependent heterogeneous arrays with applications. Econometric Theory 14, 260–284.

Chow, Y.S., Teicher, H., 1997. Probability Theory. Springer-Verlag, New York.

Darolles, S., Fan, Y., Florens, J.P., Renault, E., 2011. Nonparametric instrumental regression. Econometrica 79, 1541–1565.

Diaconis, F., Freedman, D., 1986. On the consistency of Bayes estimates. Annals of Statistics 14, 1–26.

Doob, J.L., 1949. Application of the theory of martingales. In: Le Calcul des Probabilités et ses Applications. In: Colloques Internationaux du Centre National de la Recherche Scientifique, Centre National de la Recherche Scientifique, Paris, pp. 23–27.

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

Fève, F., Florens, J.P., 2010. The practice of nonparametric estimation by solving inverse problems: the example of transformation models. The Econometrics Journal 13, S1–S27.

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

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

Florens, J.P., Johannes, J., Van Bellegem, S., 2012, Instrumental regression in partially linear models, The Econometrics Journal (forthcoming).

Florens, J.P., Mouchart, M., Rolin, J.M., 1990. Elements of Bayesian Statistics. Dekker, New York.

Florens, J.P., Simoni, A., 2011, Regularizing priors for linear inverse problems, IDEI Working Paper, n. 621, 2010, Available at http://sites.google.com/site/simonianna/research.

Florens, J.P., Simoni, A., 2012. Regularized posteriors in linear ill-posed inverse problems. Scandinavian Journal of Statistics 39, 214–235.
Franklin, J.N., 1970. Well-posed stochastic extension of ill-posed linear problems. Journal of Mathematical Analysis and Applications 31, 682–716.
Freedman, D., 1963. On the asymptotic behavior of Bayes estimates in the discrete case. The Annals of Mathematical Statistics 34, 1386–1403.
Freedman, D., 1965. On the asymptotic behavior of Bayes estimates in the discrete case II. The Annals of Mathematical Statistics 36, 454–456.
Gagliardini, P., Scaillet, O., 2012. Tikhonov regularization for nonparametric instrumental variable estimators. Journal of Econometrics 167, 61–75.
Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7, 457–472.
Ghosh, J.K., Ramamoorthi, R.V., 2003. Bayesian Nonparametrics. Springer Series in Statistics.
Hall, P., Horowitz, J., 2005. Nonparametric methods for inference in the presence of instrumental variables. Annals of Statistics 33, 2904–2929.
Kress, R., 1999. Linear Integral Equation. Springer.
Lehtinen, M.S., Päivärinta, L., Somersalo, E., 1989. Linear inverse problems for generalised random variables. Inverse Problems 5, 599–612.
Loubes, J.M., Marteau, C., 2009, Oracle inequality for instrumental variable regression, preprint.
Mandelbaum, A., 1984. Linear estimators and measurable linear transformations on a Hilbert space. Z. Wahrcheinlichkeitstheorie 3, 385–398.
Neveu, J., 1965. Mathematical Foundations of the Calculus of Probability. Holden-Day, San Francisco.
Newey, W.K., Powell, J.L., 2003. Instrumental variable estimation of nonparametric models. Econometrica 71, 1565–1578. 5.
O'Hagan, A., Kennedy, M.C., Oakley, J.E., 1998. Uncertainty analysis and other inference tools for complex computer codes. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics, Vol. 6. Oxford University Press, pp. 503–524.
O'Sullivan, F., 1986. A statistical perspective on ill-posed inverse problems. Statistical Science 1, 502–527.
Rasmussen, C.E., Williams, C.K.I., 2006. Gaussian Processes for Machine Learning. MIT Press.
Schwartz, L., 1965. On Bayes procedures. Probability Theory and Related Fields 4, 10–26.
Simoni, A., 2009, Bayesian analysis of linear inverse problems with applications in economics and finance, Ph.D. Dissertation—Université de Science Sociales, Toulouse, Available at http://simoni.anna.googlepages.com/PhD_Dissertation.pdf.
Van der Vaart, A.W., Van Zanten, J.H., 2008a. Rates of contraction of posterior distributions based on gaussian process priors. Annals of Statistics 36, 1435–1463.
Van der Vaart, A.W., Van Zanten, J.H., 2008b. Reproducing kernel Hilbert spaces of gaussian priors. In: Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh. IMS Collections, vol. 3. Institute of Mathematical Statistics, pp. 200–222.
von Mises, R., 1981. Probability, Statistics and Truth, english ed.. Dover Publications Inc., New York.