TSE working paper series

10-176

Research Group: *Econometrics and Statistics*

March, 2010

Nonparametric Estimation of an Instrumental Regression: A Quasi-Bayesian Approach Based on Regularized Posterior

JEAN-PIERRE FLORENS AND ANNA SIMONI

Ecole d'économie de Toulouse (TSE) - Manufacture des Tabacs
Aile Jean-Jacques Laffont - 21, allée de Brienne - 31000 TOULOUSE
Tél : +33 (0) 5 61 12 85 89 - Fax : +33 (0) 5 61 12 86 37 - **www.TSE-fr.eu** - contact@TSE-fr.eu

Nonparametric Estimation of an Instrumental Regression: a Quasi-Bayesian Approach Based on Regularized Posterior*

Jean-Pierre Florens
Toulouse School of Economics
(GREMAQ - Université Toulouse 1 Capitole)

Anna Simoni
Department of Decision Sciences
Università Bocconi

March 22, 2010

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.

JEL codes: C11, C14, C30.
Keywords: Instrumental Regression, Nonparametric Estimation, Posterior distribution, Tikhonov Regularization, Posterior Consistency.

*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 to conferences in Marseille, Yale, Boulder, ESEM-2008 and to the "Inverse Problems" group of Toulouse. The usual disclaimer applies and all errors remain ours.

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 with Y, and a vector of explanatory variables, denoted with 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 non-independent 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 with 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, equation (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*, *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.* (2003), Florens *et al.* (2005) and (2010), Gagliardini and Scaillet (2009), to name only a few. Other recent contributions to the literature on nonparametric estimation of IV regression, based on finite dimensional sieve minimum distance estimator, are Newey and Powell (2003), Ai and Chen (2003) and Blundell *et al.* (2007).

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), Lehtinen *et al.* (1989) and recently by Florens and Simoni (2009a,b). 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 considers the reduced form model associated with (1) and (2):

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

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 (2007) under the name *nonparametric indirect regression* model and by Loubès 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 it 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 trace-class.

The Bayes estimation of φ, or equivalently the Bayes solution of the inverse problem, is the posterior distribution of φ. It results that the Bayesian approach solves the original ill-posedness of an inverse problem by changing the nature of the problem: 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 distributions, that is, we have to find the posterior distribution of φ and the marginal distribution of Y. However, because the parameter φ is of infinite dimension, its posterior distribution suffers of another kind of ill-posedness. The posterior distribution, which is well-defined in small sample size, has a bad frequentist behavior as the sample size increases. More specifically, as the sample size increases, the posterior mean is no longer continuous in Y and becomes an inconsistent estimator in the frequentist sense. This is due to the fact that its expression involves the inverse of the marginal covariance operator of the sample and this operator converges towards an operator with unbounded inverse. Henceforth, the posterior distribution is not consistent in a frequentist sense, even if it stays consistent from a Bayesian point of view, *i.e.* with respect to the joint distribution of the sample and the parameter.

In this paper we adopt a frequentist perspective, therefore we admit the existence of a true value of φ, denoted by φ_*, that characterizes the distribution having generated the data and that satisfies (2). We study consistency of the posterior distribution. *Posterior*, or frequency, *consistency* means that the posterior distribution degenerates to a Dirac measure on the true value φ_*.

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 (2009a). This distribution is like the exact posterior distribution but the mean and variance are replaced by new moments in which the

inverse of the marginal covariance operator of the sample has been regularized by using a Tikhonov regularization scheme. An important contribution of this paper, with respect to Florens and Simoni (2009a), 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 one considered in Florens and Simoni (2009a).

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 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 A.

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 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 $< \cdot, \cdot >$ and $|| \cdot ||$, respectively. We introduce the two following conditional expectation operators:

$$\begin{array}{rccc} K: & L^2_F(Z) & \to & L^2_F(W) \\ & h & \mapsto & \mathbb{E}(h|W) \end{array} \qquad \begin{array}{rccc} K^*: & L^2_F(W) & \to & L^2_F(Z) \\ & h & \mapsto & \mathbb{E}(h|Z) \end{array}$$

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 i-th observation of the random vector S is denoted with 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 the sample of (3) in matrix form as

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

where

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

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

$$\text{and} \quad \forall x \in \mathbb{R}^n, \qquad K^*_{(n)}x = \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i \frac{f(Z, w_i)}{f(Z)f(w_i)}, \qquad K^*_{(n)} : \mathbb{R}^n \to L^2_F(Z).$$

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 $< \cdot, \cdot >$ and $|| \cdot ||$. Operator $K^*_{(n)}$ is the adjoint of $K_{(n)}$, as it can be easily verified by solving the equation $< K_{(n)}\phi, x > = < \phi, K^*_{(n)}x >$, $\forall x \in \mathbb{R}^n$ 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 with $y^i_{(n)}$ and $\varepsilon^i_{(n)}$ the i-th element of vectors $y_{(n)}$ and $\varepsilon_{(n)}$, respectively.

We use the notation $\mathcal{GP}$ for denoting 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: $\varepsilon_i|\sigma^2, \mathbf{w} \sim i.i.d.\mathcal{GP}(0, \sigma^2)$, $i = 1, \ldots, n$ 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. In this notation, φ and σ^2 are treated as random variables. When frequentist consistency will be analyzed in the following 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 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 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 $Var(\sigma^2) = \frac{s_0^4/4}{(\xi_0/2-1)^2(\xi_0/2-2)}$. Properties of the measurement μ^σ specified in part *(b)* imply that $\mathbb{E}(||\varphi||^2) < \infty$ and that Ω_0 is linear, bounded, nonnegative and self-adjoint. We give a brief reminder of the definition of covariance operator: Ω_0 is such that $< \sigma^2\Omega_0\delta, \phi >= \mathbb{E}(< \varphi - \varphi_0, \delta >< \varphi - \varphi_0, \phi > |\sigma^2)$, for all δ, ϕ in $L_F^2(Z)$, see Chen and White (1998). The covariance operator Ω_0 needs to be trace-class in order that μ^σ generates with probability 1 trajectories belonging to $L_F^2(Z)$. Therefore, Ω_0 cannot be proportional to the identity operator. The fact that Ω_0 is trace-class entails that $\Omega_0^{\frac{1}{2}}$ is Hilbert-Schmidt (HS, hereafter), see Kato (1995) Section 10.1.3. HS operators are compact and compactness of $\Omega_0^{\frac{1}{2}}$ implies compactness of Ω_0.

We introduce the *Reproducing Kernel Hilbert Space* ($\mathcal{R}.\mathcal{K}.\mathcal{H}.\mathcal{S}.$ hereafter) associated with Ω_0 and denoted with $\mathcal{H}(\Omega_0)$. 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 $L_F^2(Z)$ as:

$$\mathcal{H}(\Omega_0) = \Big\{h; h \in L_F^2(Z) \quad \text{and} \quad \sum_{j=1}^{\infty} \frac{< h, \varphi_j^{\Omega_0} >^2}{\lambda_j^{\Omega_0}} < \infty\Big\} \tag{5}$$

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 $\mathcal{R}.\mathcal{K}.\mathcal{H}.\mathcal{S}.$ is a subset of $L_F^2(Z)$ that gives the geometry of the distribution of φ. The support of a centered gaussian process, taking its values in $L_F^2(Z)$, is the closure in $L_F^2(Z)$ of the $\mathcal{R}.\mathcal{K}.\mathcal{H}.\mathcal{S}.$ associated with the covariance operator of this process (denoted with $\overline{\mathcal{H}(\Omega_0)}$ 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 classical point of view, there exists a true value φ_* that has generated the data $y_{(n)}$ in model (4) and that satisfies the 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 note that Ω_0 is an integral operator. Indeed, $\forall \delta, \phi \in L_F^2(Z)$ it is defined as

$$\begin{aligned} < \Omega_0\delta, \phi > &= \frac{1}{\sigma^2}\mathbb{E}(< \varphi - \varphi_0, \delta >< \varphi - \varphi_0, \phi > |\sigma^2) \\ &= \frac{1}{\sigma^2}\mathbb{E}\Big(\int (\varphi(z) - \varphi_0(z))\delta(z)f_z(z)dz \int (\varphi(\zeta) - \varphi_0(\zeta))\phi(\zeta)f_z(\zeta)d\zeta \Big| \sigma^2\Big) \\ &= \int \omega_0(z,\zeta)\delta(z)\phi(\zeta)f_z(z)f_z(\zeta)dzd\zeta \end{aligned}$$

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 an 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 $||\delta_*||^2 = \sum_{j=1}^{\infty} \frac{<\varphi_* - \varphi_0, \varphi_j^{\Omega_0}>^2}{\lambda_j^{\Omega_0}}$ is bounded and, as $\lambda_j^{\Omega_0} \downarrow 0$ this implies that the Fourier coefficients $< \varphi_* - \varphi_0, \varphi_j^{\Omega_0} >$ 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 $\mathcal{R}.\mathcal{K}.\mathcal{H}.\mathcal{S}.$ is the space of absolutely continuous functions h on $[0,1]$ with at least one square integrable derivative and such that $h(0) = 0$, see Carrasco and Florens (2000). Summarizing, in 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 $\mathcal{R}.\mathcal{K}.\mathcal{H}.\mathcal{S}.$ 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

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

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).

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.

We need Assumption 4 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 would be 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 equation (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. Existence of a solution of equation (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 non-identified 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 is 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 then K would be implied to be 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 equation (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 equation (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 lighten 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:

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

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 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 $\mu_n^{\sigma,y}$, conditionally 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 (2009a). 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 $(\mathbb{R}^n, \mathfrak{B}(\mathbb{R}^n))$ and the space L_F^2 on $(\mathbb{R}^n, \mathfrak{B}(\mathbb{R}^n), F)$, with $\mathfrak{B}(\mathbb{R}^n)$ denoting the Borel σ-field on $\mathbb{R}^n$, are Polish because $(\mathbb{R}^n, \mathfrak{B}(\mathbb{R}^n))$ is a separable metric space. 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 (2009a) 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 non-continuous operator that amplifies the measurement error in $y_{(n)}$ and $\mathbb{E}(\varphi|y_{(n)})$ is not consistent in

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

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. We discuss the inconsistency of the posterior distribution in more detail in subsection 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 that we have proposed in Florens and Simoni (2009a): we replace the exact posterior distribution with a *regularized posterior distribution*. This new distribution, denoted with $\mu_\alpha^{\sigma,y}$, 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 but 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:

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

with

$$\begin{aligned} A_\alpha &= \Omega_0 K^*_{(n)}\Big(\alpha_n I_n + \frac{1}{n}I_n + K_{(n)}\Omega_0 K^*_{(n)}\Big)^{-1} \\ b_\alpha &= (I - A_\alpha K_{(n)}\varphi_0) \end{aligned} \tag{9}$$

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:

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

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 σ^2, $||\tilde{A}||^2_{HS} := tr\tilde{A}^*\tilde{A}$ denotes the HS norm, $\mathcal{B}_2(\mathbb{R}^n, L_F^2(Z))$ is the set of all bounded operators on $\mathbb{R}^n$ 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 with $\tilde{\mu}_n^\sigma$, of the form:

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

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:

$$\begin{aligned}
\mathbb{E}(\varphi|y_{(n)},\sigma^2) &= \varphi_0 + \frac{\sigma^2}{\alpha_n n+1}\Omega_0 K_{(n)}^*\Big(\frac{\sigma^2}{n}I_n + \frac{\sigma^2}{\alpha_n n+1}K_{(n)}\Omega_0 K_{(n)}^*\Big)^{-1}(y_{(n)} - K_{(n)}\varphi_0) \\
&= \varphi_0 + \Omega_0 K_{(n)}^*\Big(\frac{\alpha_n n+1}{n}I_n + K_{(n)}\Omega_0 K_{(n)}^*\Big)^{-1}(y_{(n)} - K_{(n)}\varphi_0) \\
&= \varphi_0 + \Omega_0 K_{(n)}^*\Big(\alpha_n I_n + \frac{1}{n}I_n + K_{(n)}\Omega_0 K_{(n)}^*\Big)^{-1}(y_{(n)} - K_{(n)}\varphi_0) \equiv \hat{\varphi}_\alpha.
\end{aligned}$$

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

$$Var(\varphi|y_{(n)},\sigma^2) = \frac{\sigma^2}{\alpha_n n+1}\Big[\Omega_0 - \Omega_0 K_{(n)}^*(\alpha_n I_n + \frac{1}{n}I_n + K_{(n)}\Omega_0 K_{(n)}^*)^{-1}K_{(n)}\Omega_0\Big]$$

and it converges faster than $\sigma^2\Omega_{y,\alpha}$. This is due to the fact that the prior covariance operator of $\tilde{\mu}_n^\sigma$ 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 $Var(\varphi|y_{(n)},\sigma^2)$. Such a particular feature of the prior covariance operator can make $\tilde{\mu}_n^\sigma$ 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, but we want to stress that 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_n^y of σ^2. This distribution will be used in order to marginalize $\mu_\alpha^{\sigma,y}$.

Since we have a conjugate model, we integrate out φ from the sampling probability $P^{\sigma,\varphi,\mathbf{W}}$ by using the prior μ^σ and we use the probability model $P^{\sigma,\mathbf{W}} \times \nu$ to make inference on σ^2, with $P^{\sigma,\mathbf{W}}$ defined in (6). The posterior distribution of σ^2 has the kernel of an $I\Gamma$ distribution:

$$\sigma^2|y_{(n)} \sim \nu^{\mathcal{F}} \propto \Big(\frac{1}{\sigma^2}\Big)^{\xi_0/2+n/2+1}\exp\{-\frac{1}{2\sigma^2}[(y_{(n)}-K_{(n)}\varphi_0)'C_n^{-1}(y_{(n)}-K_{(n)}\varphi_0)+s_0^2]\}. \quad (11)$$

Then, $\nu_n^y \sim I\Gamma(\xi_*, s_*^2)$ with $\xi_* = \xi_0 + n$,

$$s_*^2 \;=\; s_0^2 + (y_{(n)} - K_{(n)}\varphi_0)' C_n^{-1} (y_{(n)} - K_{(n)}\varphi_0)$$

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

Since ν_n^y does not depend on φ it can be used for marginalizing the RCPD $\mu_\alpha^{\sigma,y}$ of φ, conditional on σ^2, by integrating out 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 $L_F^2(Z)$. 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 *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 $< \cdot, \cdot >_\mathcal{X}$. 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 $\forall \delta \in \mathcal{X}$,*

$$< x, \delta >_\mathcal{X} \sim\; t(< x_0, \delta >_\mathcal{X}, < \Omega_0\delta, \delta >_\mathcal{X}, \iota),$$

i.e. $< x, \delta >_\mathcal{X}$ has a density proportional to

$$\Big[\iota + \frac{(< x, \delta >_\mathcal{X} - < x_0, \delta >_\mathcal{X})^2}{< \Omega_0\delta, \delta >_\mathcal{X}}\Big]^{-\frac{\iota+1}{2}},$$

with mean and variance

$$\begin{aligned} \mathbb{E}(< x, \delta >_\mathcal{X}) &= < x_0, \delta >_\mathcal{X}, \qquad \text{if } \iota > 1 \\ Var(< x, \delta >_\mathcal{X}) &= \frac{\iota}{\iota - 2} < \Omega_0\delta, \delta >_\mathcal{X}, \qquad \text{if } \iota > 2. \end{aligned}$$

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*

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

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

We apply this result to the IV regression process φ, so that if we integrate out σ^2 in $\mu_\alpha^{\sigma,y}$, with respect to ν_n^y, we get

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

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

3.4 Asymptotic Analysis

In this section we analyze asymptotic properties of ν_n^y, $\mu_\alpha^{\sigma,y}$ and μ_α^y 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 $\mu_\alpha^{\sigma,y}$, 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 $(\varphi_*, \mu_\alpha^{\sigma,y})$ 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), 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 1949 Doob publishes 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), Chapter 4 and 7.

Doob's results have been extended by Breiman *et al.* (1964); Freedman (1963) and Schwartz (1965) extended Doob's theorem in a frequentist sense, that is, by considering a convergence with respect to the sampling distribution. Let θ be the finite dimensional parameter of interest and P^θ denote the sampling distribution; they prove that the posterior mean of θ converges P^θ-almost surely to θ, for θ belonging to the support of the prior distribution, if and only if θ has finite dimension and if P^θ is smooth with respect to θ. Diaconis and Freedman (1986) point out that the assumption of finite dimensionality of θ is really needed, so 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, contrarily to the finite dimensional case where the posterior distribution is consistent, in infinite dimensional problems the prior-to-posterior transformation does not solve the problem of ill-posedness. This is due to compactness of $K\Omega_0$ and to the fact that the sampling covariance operator shrinks at the rate $\frac{1}{n}$ which is too fast to control the ill-posedness.

In the reverse, we state in the following theorem that the regularized posterior distribution $\mu_\alpha^{\sigma,y}$ and the regularized posterior mean $\hat{\varphi}_\alpha$ are consistent. For some $\beta > 0$, we denote with 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) having generated the data $y_{(n)}$ under model (4) and $\mu_\alpha^{\sigma,y}$ be a gaussian random measure on $L_F^2(Z)$ with mean $\hat{\varphi}_\alpha = A_\alpha y_{(n)} + b_\alpha$ and covariance operator $\sigma^2\Omega_{y,\alpha}$ defined in (8) and (9). 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 and if $\delta_* \in \Phi_\beta$ for some $\beta > 0$,

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

(ii) if there exists a $\kappa > 0$ such that $\lim_{n\to\infty} \sum_{j=1}^n \frac{<\Omega_0\varphi_{jn}, \varphi_{jn}>}{\lambda_{jn}^{2\kappa}} < \infty$, 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. Moreover, if $\delta_* \in \Phi_\beta$ for some $\beta > 0$, it is of order

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

(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

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

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 slow 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 $n^{-\frac{\beta}{\beta+2}}$.

When $\alpha_n = \alpha_n^*$, then $||\Omega_{y,\alpha}\phi||^2 \sim n^{-\frac{\beta}{\beta+2}}$, $\forall \phi$ such that $\Omega_0^{\frac{1}{2}}\phi \in \Phi_\beta$. The optimal α_n for the RCPD $\mu_\alpha^{\sigma,y}$ is given by α_* if $\kappa \geq \beta$ and by $n^{-\frac{1}{\kappa+2}}$ 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 *We point out that Theorem 2 can be obtained as a special case of Theorems 2, 3 and 4 of Florens and Simoni (2009a). However, the fact that operators* $K_{(n)}$ *and* $K_{(n)}^*$ *are finite rank and the variance parameter* σ^2 *is treated as random variable make the rates of convergence in Theorem 2 and strategy of its proof different than those ones of Theorems 2, 3 and 4 in Florens and Simoni (2009a).*

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

Next, we analyze consistency of $\mathbb{E}(\sigma^2|y_{(n)})$ and of the posterior ν_n^y for a true value σ_*^2 having generated data in model (4). If $\bar{\omega}_0(s,z)$ denotes the kernel of $\Omega_0^{\frac{1}{2}}$, we use the notation $g(Z,w_i) = \Omega_0^{\frac{1}{2}}(\frac{f(s,w_i)}{f(s)f(w_i)})(Z) = \int \bar{\omega}_0(s,Z)\frac{f(s,w_i)}{f(s)f(w_i)}f(s)ds$, 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) *having generated the data under model (4) and* ν_n^y *be the* $I\Gamma(\xi_*, s_*^2)$ *distribution on* $\mathbb{R}_+$ *described in (11). Under Assumption 4, if there exists a* γ *such that* $\forall w$, $g(Z,w) \in \Phi_\gamma$ *(with* Φ_γ *defined as in (12)), then*

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

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.*

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 with 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 has not 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 of a problem of inconsistency since it demands the inversion of the covariance operator $(\frac{\sigma^2}{n} I_n + K_{(n)}\Omega_0 K_{(n)}^*)$ of the distribution of $y_{(n)}|\sigma^2$ which, as $n \to \infty$, converges toward an operator with non-continuous inverse. Therefore, we use a Tikhonov regularization scheme and obtain the RCPD for φ, still denoted with $\mu_\alpha^{\sigma,y}$. 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

$$\begin{aligned} A_\alpha^\sigma &= \Omega_0 K_{(n)}^* \Big(\alpha_n I_n + \frac{\sigma^2}{n} I_n + K_{(n)}\Omega_0 K_{(n)}^*\Big)^{-1}, \\ b_\alpha^\sigma &= (I - A_\alpha^\sigma K_{(n)})\varphi_0 \end{aligned}$$

that must not be confused with 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: $\varepsilon_{(n)}|\sigma^2, \mathbf{w} \sim i.i.d.\, \mathcal{N}(0, \frac{\sigma^2}{n} I_n)$ with $\varepsilon_{(n)} = y_{(n)} - K_{(n)}\varphi$ and φ is drawn from $\mu_\alpha^{\sigma,y}$. Therefore, we talk about *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

$\xi_* = \xi_0 + n$, $\tilde{s}^2 = s_0^2 + n\sum_i (y^i_{(n)} - K^i_{(n)}\varphi)^2$ and $K^i_{(n)}$ denotes the i-th 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.* $\eta_i = \frac{1}{\sqrt{n}}(K^i_{(n)}\varphi_* - \hat{K}^i_{(n)}\varphi_*)$ 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 $n \to \infty$ (for lightening notation we have eliminated the dependence on n from h). We denote 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 $\hat{f}_w$:

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

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

and

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

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 Number $\hat{K}^*_{(n)}\hat{K}_{(n)}\varphi \to \mathbb{E}(\mathbb{E}(\varphi|w_i)|Z)$.

From asymptotic properties of the kernel estimator of a regression function we know that $\eta_{(n)} \Rightarrow \mathcal{N}_n(0, \frac{\sigma^2}{n^2 h^q} D_{(n)})$ with $D_{(n)} = diag(\frac{1}{f(w_i)} \int L^2_w(u)du)$ and $\Rightarrow$ denotes convergence in distribution. The asymptotic variance of $\eta_{(n)}$ is negligible with respect to $Var(\varepsilon_{(n)}) \equiv \frac{\sigma^2}{n} I_n$ 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 probability distribution of $(y_{(n)} - \hat{K}_{(n)}\varphi)|\hat{f}_{z,w}, \varphi, \mathbf{w}$ is asymptotically gaussian.

In our Quasi-Bayesian approach the gaussianity of the sampling measure is used only in order to construct the posterior distribution and the regularized posterior mean, which is our Bayes estimator of the IV regression. Gaussianity of the sampling measure is not used neither in the proof of frequentist consistency of the regularized posterior distribution nor in that one of the regularized posterior mean. For this reason, we can approximate the sampling measure by its asymptotic limit, so that $y_{(n)}|\hat{f}_{z,w}, \varphi, \mathbf{w} \sim P^{\hat{f},\varphi,\mathbf{w}} \sim^a \mathcal{GP}(\hat{K}_{(n)}\varphi, \Sigma_n)$, where $\sim^a$ means "approximately distributed as", $\Sigma_n = Var(\eta_{(n)} + \varepsilon_{(n)}) = (\frac{\sigma^2}{n} + o_p(\frac{1}{n}))I_n$ and for simplicity σ^2 is considered as known. The estimated density $\hat{f}_{z,w}$ affects the sampling measure through $\hat{K}_{(n)}$, which converges to $K_{(n)}$.

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 it must be introduced a regularization parameter $\alpha_n > 0$ that goes to 0 slower than $\frac{1}{n}$. The regularized posterior distribution that results is called *estimated regularized posterior distribution* since now it depends on $\hat{K}_{(n)}$ instead of on $K_{(n)}$. It is denoted with $\hat{\mu}^y_\alpha$, it is gaussian with mean $\hat{\mathbb{E}}_\alpha(\varphi|y_{(n)})$ and covariance operator $\hat{\Omega}_{y,\alpha}$ given by

$$\begin{aligned} \hat{\mathbb{E}}_\alpha(\varphi|y_{(n)}) &= \varphi_0 + \overbrace{\Omega_0\hat{K}^*_{(n)}(\alpha_n I_n + \Sigma_n + \hat{K}_{(n)}\Omega_0\hat{K}^*_{(n)})^{-1}}^{\hat{A}_\alpha}(y_{(n)} - \hat{K}_{(n)}\varphi_0) \\ \hat{\Omega}_{y,\alpha} &= \Omega_0 - \Omega_0\hat{K}^*_{(n)}(\alpha_n I_n + \Sigma_n + \hat{K}_{(n)}\Omega_0\hat{K}^*_{(n)})^{-1}\hat{K}_{(n)}\Omega_0. \end{aligned} \tag{14}$$

Asymptotic properties of the posterior distribution for the case with unknown $f_{z,w}$ are very similar to those 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 $\hat{\mu}_\alpha^y$ and 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.* (2003) provides regularity conditions in order to get $||\int\int \varphi(z)\hat{f}(z|w_i)dz\frac{\hat{f}(z,w_i)}{\hat{f}(z)}dw_i - \mathbb{E}(\mathbb{E}(\varphi|W)|Z)||^2 = \mathcal{O}_p(\frac{1}{nh^p} + h^{2\rho})$. We implicitly assume in the following theorem (and in Lemma 4 below) that the regularity Assumptions B.1-B.5 of Darolles *et al.* (2003) are satisfied.

Theorem 4 *Let φ_* be the true value having generated the data $y_{(n)}$ under model (4) and $\hat{\mu}_\alpha^y$ be a gaussian measure on $L_F^2(Z)$ with mean and covariance operator defined in (14). Under Assumptions 4 and 5, if $\alpha_n \to 0$ and $\alpha_n^2 n \to \infty$, we have*

$$||\hat{\mathbb{E}}_\alpha(\varphi|y_{(n)}) - \varphi_*||^2 \to 0 \text{ in } P^{\hat{f},\varphi_*,\mathbf{W}}\text{-probability and if } \delta_* \in \Phi_\beta, \text{ for some } \beta > 0,$$

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

If the bandwidth h is chosen in such a way to guarantee that $\frac{1}{\alpha_n^2}(\frac{1}{n} + h^{2\rho}) = \mathcal{O}_p(\frac{1}{\alpha_n^2 n})$, 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 speed of convergence of $||\hat{\mathbb{E}}_\alpha(\varphi|y_{(n)}) - \varphi_*||^2$ proportional to $n^{-\frac{\beta}{\beta+2}}$. We have the same speed 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 (Case I and Case II below) and unknown operator case (Case III below). More figures concerning this simulation can be found in an additional appendix available at
`http://didattica.unibocconi.it/mypage/index.php?IdUte=107247&idr=11421&lingua=ita/`.

We simulate $n = 1000$ observations from the following model, which involves only one endogenous covariate and two instrumental variables[2],

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

$$\begin{aligned} v_i \sim \mathcal{N}(0, \sigma_v^2), \qquad & z_i = 0.1w_{i,1} + 0.1w_{i,2} + v_i \\ \varepsilon_i \sim \mathcal{N}(0, (0.5)^2), \qquad & u_i = \mathbb{E}(\varphi_*(z_i)|w_i) - \varphi_*(z_i) + \varepsilon_i \\ y_i = \varphi_*(z_i) + u_i. & \end{aligned}$$

[2]This data generating process is borrowed from Example 3.2 in Chen and Reiss (2007).

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

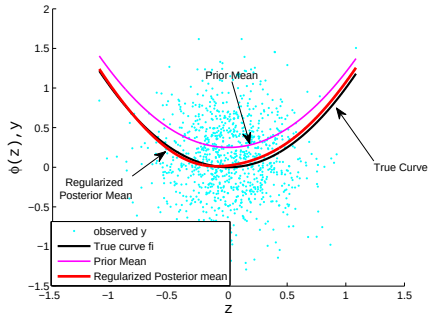
$$\begin{pmatrix} Z \\ W_1 \\ W_2 \end{pmatrix} \sim \mathcal{N}_3 \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} (0.026 + \sigma_v^2) & 0.13 & 0.13 \\ 0.13 & 1 & 0.3 \\ 0.13 & 0.3 & 1 \end{pmatrix} \right).$$

Endogeneity is caused by correlation between u_i and the error term v_i affecting the covariates. For all the simulations below we fix $\sigma_v = 0.27$ and α_n is fixed at a value determined by letting α_n vary in a large range of values and selecting by hand that one producing a good estimation. We present in the next section a data-driven method for selecting α_n.

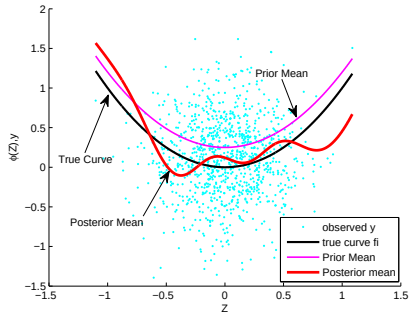
Case I. *Conjugate Model with $f_{z,w}$ known and smooth φ_*.*

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 $\delta \in L^2_F(Z)$. We have performed simulations for two specifications of φ_0: Figure 1 refers to $\varphi_0(Z) = 0.95Z^2 + 0.25$ while Figure 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 $\mu_n^{\sigma,y}$.



(a) Regularized Posterior Mean Estimate with $\alpha_n = 0.3$.



(b) Posterior Mean Estimate with $\alpha_n = 0$.

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

Figure 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. In red is drawn 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 Figure 4 while the results for $\varphi_0(Z) = 0$ and $\alpha_n = 0.3$ are in Figure 5. The kernel estimators of the prior and posterior distributions of σ^2, together with its posterior mean estimator, are shown in Figure 6. The interpretation of the graphs in each figure is the same as in CASE I.

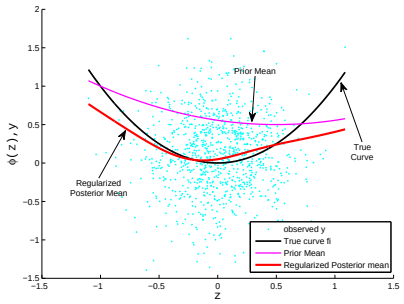
CASE III. *$f_{z,w}$ unknown, σ^2 known and smooth φ_*.*
In this simulation we have specified a prior only on φ since σ^2 is supposed to be known. The prior distribution for φ is specified as in CASE I with same φ_0's and $\sigma_0 = 20$. We show in Figures 7 the results obtained by using a kernel estimator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

5.1 Data driven method for choosing α

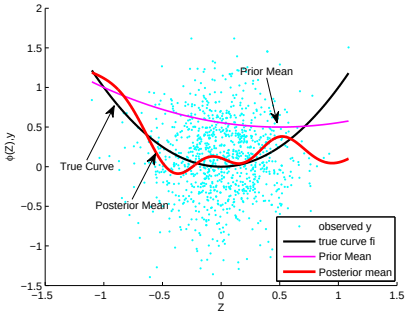
In inverse problem theory there exist several parameter choice rules which determine the regularization parameter α_n on the basis of the performance of the regularization method under consideration. These techniques are often known under the name of *error free* and we refer to Engl *et al.* (2000) 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 speed of convergence as the estimation error. In particular, as it is 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

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



(a) Regularized Posterior Mean Estimate with $\alpha_n = 0.3$.



(b) Posterior Mean Estimate with $\alpha_n = 0$.

Figure 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 $\sigma_0 = 200$.

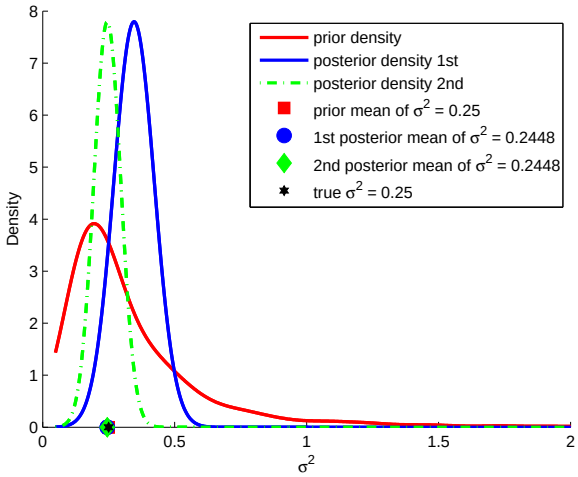
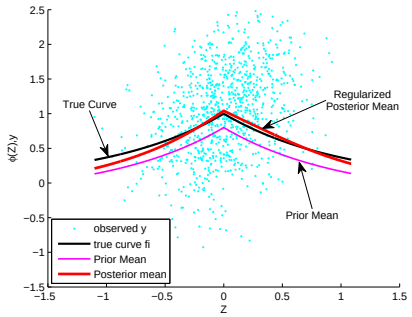
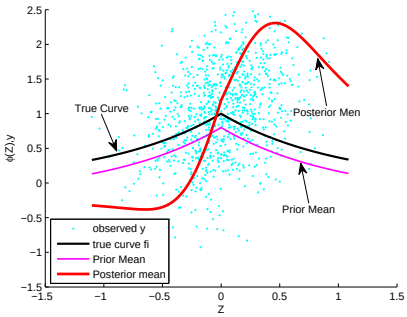


Figure 3: CASE I. *Conjugate Model with $f_{z,w}$ known and smooth φ_*.* Prior and posterior distributions of σ^2. The label '1st' refers to the simulation with $\varphi_0(Z) = 0.95Z^2 + 0.25$, while '2nd' refers to the simulation with $\varphi_0(Z) = \frac{2}{9}Z^2 - \frac{2}{9}Z + \frac{5}{9}$.

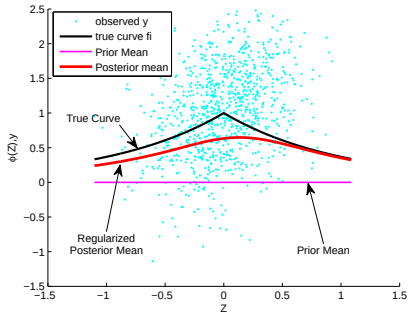


(a) Regularized Posterior Mean Estimate with $\alpha_n = 0.4$.

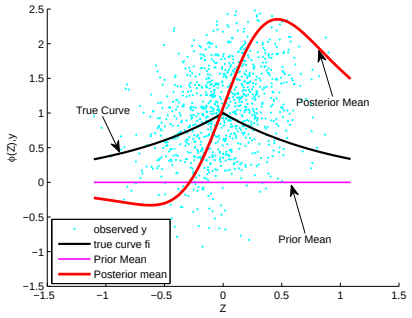


(b) Posterior Mean Estimate with $\alpha_n = 0$.

Figure 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 $\alpha_n = 0$.

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

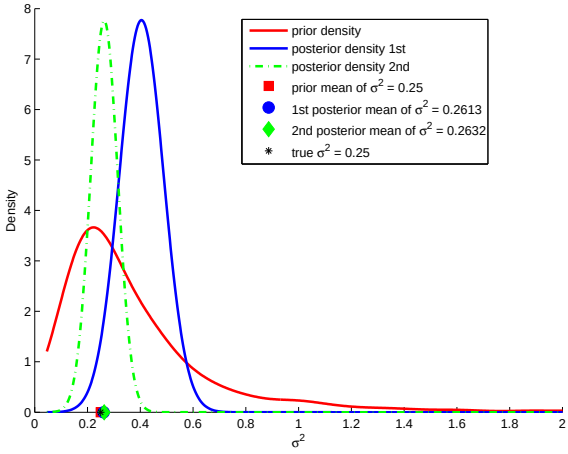
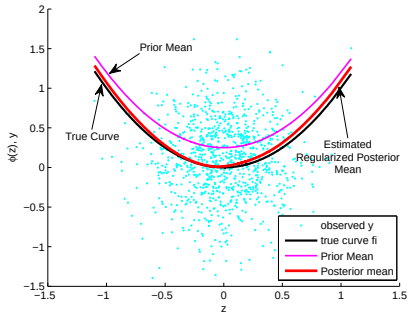
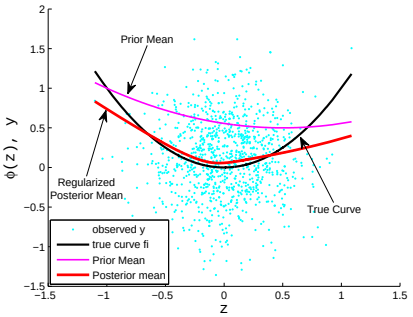


Figure 6: CASE II. *Conjugate Model with $f_{z,w}$ known and irregular φ_*.* Prior and posterior distributions of σ^2. The label '1st' refers to the simulation with $\varphi_0(Z) = \exp(-|Z|) - 0.2$, while '2nd' refers to the simulation with $\varphi_0(Z) = 0$.



(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 $\alpha_n = 0.3$.

Figure 7: CASE III. *Conjugate Model with $f_{z,w}$ unknown and smooth φ_*.*

where "$\sim$" means "of the same order of". Hence, it makes sense to take $\frac{||\vartheta_\alpha||}{\alpha^d}$ as 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 the good rate. In reverse, 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 $T_{(n)}^* = \Omega_0^{\frac{1}{2}} K_{(n)}^*$ and $T_{(n)} = K_{(n)} \Omega_0^{\frac{1}{2}}$.

In order to explain our data-driven method we have to introduce the notion of *qualification* of a regularization method. Under the assumption $\varphi_* \in \Phi_\beta$, we call the qualification β_0 of the regularization method the largest value of β such that $||\hat{\varphi}_\alpha - \varphi_*||^2 = \mathcal{O}_p(\alpha^\beta)$ for $0 < \beta < \beta_0$; the qualification of 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 be at least equal to $\beta_0 \geq \beta + 2$, which is impossible for a Tikhonov regularization. We have to substitute the Tikhonov regularization scheme, used to construct $\hat{\varphi}_\alpha$, with an *iterated Tikhonov* scheme. In our case, it is enough to iterate only two times, 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 with $\hat{\varphi}_\alpha^{(2)}$ 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 $\hat{\varphi}_\alpha^{(2)}$ be the regularized posterior mean obtained through a two-times-iterated Tikhonov scheme in the conjugate case described in Assumption 3 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*

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

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. The minimization program does not change if we take an increasing transformation of this ratio, for instance we have considered the logarithmic transformation. This simplifies the graphical representation of the curve.

A result similar to Lemma 3 can be derived when the density $f_{z,w}$ is unknown and the nonparametric method described in subsection 4.1 is applied. In this case we denote $\hat{T}_{(n)}^* = \Omega_0^{\frac{1}{2}} \hat{K}_{(n)}^*$ 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 two-times iterated Tikhonov scheme for constructing $\hat{A}_\alpha^{(2)}$. We obtain the following result:

Lemma 4 *Let $\hat{\mathbb{E}}_\alpha^{(2)}(\varphi|y_{(n)})$ be the estimated regularized posterior mean obtained through a two-times-iterated Tikhonov scheme in the unknown operator case described in Section 4.1 and $\hat{\vartheta}_\alpha^{(2)} = \hat{T}_{(n)}^*(y_{(n)} - \hat{K}_{(n)}\hat{\mathbb{E}}_\alpha^{(2)}(\varphi|y_{(n)}))$. 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*

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

In the previous Lemma we have implicitly assumed that Assumptions B.1-B.5 of Darolles *et al.* (2003) are satisfied. It is necessary to rescale the residual by $\frac{1}{\alpha_n^2}$ to reach the same speed 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://didattica.unibocconi.it/mypage/index.php?IdUte=107247&idr=11421&lingua=ita/`.

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 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 the 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 than Tikhonov scheme, and to analyze the way in which the regularized posterior mean is affected. We could also consider Sobolev spaces, instead of Hilbert spaces, with regularization methods that use differential norms.

APPENDIX

A Proofs

In all the proofs that follow we use the following notation:

- (σ_*^2, φ_*) is the true parameter having generated the data according to model (4);
- $\mathcal{H}(\Omega_0) = \mathcal{R}.\mathcal{K}.\mathcal{H}.\mathcal{S}(\Omega_0)$;
- 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 $\varphi \in L_F^2(Z) \mapsto I\varphi = \varphi$;
- $T = K\Omega_0^{\frac{1}{2}}$, $T : L_F^2(Z) \to L_F^2(W)$;
- $T_{(n)} = K_{(n)}\Omega_0^{\frac{1}{2}}$, $T_{(n)} : L_F^2(Z) \to \mathbb{R}^n$;
- $\hat{T}_{(n)} = \hat{K}_{(n)}\Omega_0^{\frac{1}{2}}$, $\hat{T}_{(n)} : L_F^2(Z) \to \mathbb{R}^n$;
- $T^* = \Omega_0^{\frac{1}{2}}K^*$, $T^* : L_F^2(W) \to L_F^2(Z)$;
- $T_{(n)}^* = \Omega_0^{\frac{1}{2}}K_{(n)}^*$, $T_{(n)}^* : \mathbb{R}^n \to L_F^2(Z)$;
- $\hat{T}_{(n)}^* = \Omega_0^{\frac{1}{2}}\hat{K}_{(n)}^*$, $\hat{T}_{(n)}^* : \mathbb{R}^n \to L_F^2(Z)$;
- $\Omega_0^{\frac{1}{2}} = \int_{\mathbb{R}^p} \bar{\omega}_0(s, Z)f(s)ds$;
- $g(Z, w_i) = \int_{\mathbb{R}^p} \bar{\omega}_0(s, Z)\frac{f(s,w_i)}{f(s)f(w_i)}f(s)ds$;
- $\Phi_\beta = \mathcal{R}(T^*T)^{\frac{\beta}{2}}$ 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 an Hilbert space $L_F^2(Z)$, endowed with the Borel σ-field $\mathfrak{E}$, converges weakly to a probability measure δ_{φ_*} if

$$|| \int a(\varphi)\mu_n^{\sigma,y}(d\varphi) - \int a(\varphi)\delta_{\varphi_*}(d\varphi)|| \to 0, \quad P^{\sigma_*,\varphi_*,\mathbf{W}} - a.s. \text{ (or in } P^{\sigma_*,\varphi_*,\mathbf{W}}\text{-probability)}$$

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 $a : \phi \mapsto \phi$, $\forall \phi \in L^2_F(Z)$, so that we have to check convergence of the posterior mean. For simplicity, we set $\varphi_0 = 0$, then the posterior mean is

$$\mathbb{E}(\varphi|y_{(n)}) = \Omega_0 K^*_{(n)} \Big(\frac{1}{n} I_n + K_{(n)} \Omega_0 K^*_{(n)}\Big)^{-1} y_{(n)}$$

and we have to prove that the L^2_F-norm $||\mathbb{E}(\varphi|y_{(n)}) - \varphi_*|| \to 0$ $P^{\sigma_*,\varphi_*,\mathbf{w}}$-a.s. By decomposing

$$\begin{aligned}\mathbb{E}(\varphi|y_{(n)}) - \varphi_* &= \overbrace{\Omega_0 K^*_{(n)} \Big(\frac{1}{n} I_n + K_{(n)} \Omega_0 K^*_{(n)}\Big)^{-1} \varepsilon_{(n)}}^{\mathcal{A}} \\ &\quad \underbrace{- (I - \Omega_0 K^*_{(n)} \Big(\frac{1}{n} I_n + K_{(n)} \Omega_0 K^*_{(n)}\Big)^{-1} K_{(n)}) \varphi_*}_{\mathcal{B}}.\end{aligned}$$

we get the lower bound: $||\mathbb{E}(\varphi|y_{(n)}) - \varphi_*|| \geq \Big|||\mathcal{A}|| - ||\mathcal{B}||\Big|$. We will prove that $||\mathcal{A}|| \to \infty$ and $||\mathcal{B}|| \to 0$. We start by considering $||\mathcal{A}||$ and we prove that it is not convergent by contradiction.

First, we remark that, by the Cauchy-Schwarz inequality, $\forall \varphi \in L^2_F(Z)$

$$||\mathcal{A}||||\varphi|| \geq < \Omega_0^{\frac{1}{2}} T^*_{(n)} \Big(\frac{1}{n} I_n + T_{(n)} T^*_{(n)}\Big)^{-1} \varepsilon_{(n)}, \varphi > = < T^*_{(n)} \Big(\frac{1}{n} I_n + T_{(n)} T^*_{(n)}\Big)^{-1} \varepsilon_{(n)}, \Omega_0^{\frac{1}{2}} \varphi >$$

and, without loss of generality, we can take φ such that $||\varphi|| = 1$, so that

$$||\mathcal{A}|| \geq < T^*_{(n)} \Big(\frac{1}{n} I_n + T_{(n)} T^*_{(n)}\Big)^{-1} \varepsilon_{(n)}, \Omega_0^{\frac{1}{2}} \varphi > := < \mathcal{A}_1, \Omega_0^{\frac{1}{2}} \varphi > . \tag{16}$$

Next, we study the convergence to zero of $\mathcal{A}$ and we expand $\mathcal{A}_1$ by using the SVD of $T_{(n)}$ in the following way

$$\begin{aligned}< \mathcal{A}_1, \Omega_0^{\frac{1}{2}} \varphi > &= \sum_{j=1}^n < T^*_{(n)} (\frac{1}{n} I_n + T_{(n)} T^*_{(n)})^{-1} \varepsilon_{(n)}, \varphi_{jn} >< \varphi_{jn}, \Omega_0^{\frac{1}{2}} \varphi > \\ &= \sum_{j=1}^n < \varepsilon_{(n)}, (\frac{1}{n} I_n + T_{(n)} T^*_{(n)})^{-1} T_{(n)} \varphi_{jn} >< \varphi_{jn}, \Omega_0^{\frac{1}{2}} \varphi > \\ &= \sum_{j=1}^n \frac{\lambda_{jn}}{\frac{1}{n} + \lambda^2_{jn}} < \varepsilon_{(n)}, \psi_{jn} >< \varphi_{jn}, \Omega_0^{\frac{1}{2}} \varphi >\end{aligned}$$

as it results from (15). Let us suppose that $\frac{1}{n}$ plays the role of a regularization parameter and call it α_n; by definition of *regularization scheme*, see e.g. Kress (1999) Definition 15.7 p. 270, $\mathcal{A}$ should converge to 0 with probability 1 as $n \to \infty$. Let $\{\xi_j, j \geq 1\}$ be independent random variables with $\mathbf{E}(\xi_j) = 0$ and $Var(\xi_j) = 1$, $j \geq 1$. Under Assumption 2, $< \varepsilon_{(n)}, \psi_{jn} >= \frac{\sigma_*}{\sqrt{n}} \xi_j$ since $\mathbf{E}(< \varepsilon_{(n)}, \psi_{jn} >) = 0$ and $cov(< \varepsilon_{(n)}, \psi_{jn} >, < \varepsilon_{(n)}, \psi_{kn} >) = \frac{\sigma_*}{\sqrt{n}} < \psi_{jn}, \psi_{kn} >$ which is equal to 0 for $j \neq k$ and equal to $\frac{\sigma^2_*}{n}$ for $j = k$. Then,

$$\begin{aligned}\lim_{n\to\infty} < \mathcal{A}_1, \Omega_0^{\frac{1}{2}} \varphi > &= \lim_{n\to\infty} \sum_{j=1}^n \frac{\lambda_{jn} < \varepsilon_{(n)}, \psi_{jn} >}{(\alpha_n + \lambda^2_{jn})} < \varphi_{jn}, \Omega_0^{\frac{1}{2}} \varphi > \qquad (17) \\ &= \lim_{n\to\infty} \sum_{j=1}^n \frac{\lambda_{jn} \frac{\sigma_*}{\sqrt{n}} \xi_j}{(\alpha_n + \lambda^2_{jn})} < \varphi_{jn}, \Omega_0^{\frac{1}{2}} \varphi > \\ &\geq \lim_{n\to\infty} \frac{\sigma_*}{\sqrt{n}} \sum_{j=1}^n \frac{\lambda_{jn} \xi_j}{\alpha_n + \lambda_{1n}} < \varphi_{jn}, \Omega_0^{\frac{1}{2}} \varphi > \qquad (18) \\ &= \lim_{n\to\infty} \frac{\sigma_*}{\sqrt{n}(\alpha_n + \lambda_{1n})} \sum_{j=1}^n \lambda_{jn} \xi_j < \varphi_{jn}, \Omega_0^{\frac{1}{2}} \varphi > \qquad (19)\end{aligned}$$

since $\lambda_{1n} \geq \lambda_{jn}$, $j \geq 1$. We remark that $\{\lambda_{jn}\xi_j < \varphi_{jn}, \Omega_0^{\frac{1}{2}}\varphi >, j \geq 1\}$ are independent random variables with mean 0, finite second moment and such that $\sum_{j=1}^{\infty} \lambda_{jn}^2 \mathbb{E}(\xi_j^2) < \varphi_{jn}, \Omega_0^{\frac{1}{2}}\varphi >^2 < \infty$. The last convergence follows from the fact that, as $n \to \infty$, λ_{jn}^2 converges to the eigenvalues of $K\Omega_0 K^*$ and $K\Omega_0 K^*$ is trace-class, that is, $tr(K\Omega_0 K^*) := \sum_{j=1}^{\infty} \lambda_j^2 < \infty$. The operator $K\Omega_0 K^*$ is trace-class because Ω_0 is trace-class, K is bounded and $tr(K\Omega_0 K^*) \leq ||K||tr(\Omega_0)||K^*||$, see Kato (1995) p. 522. Moreover, $||\Omega_0^{\frac{1}{2}}\varphi||^2 < \infty$. From the Khintchine-Kolmogorov Convergence Theorem, see e.g. Chow and Teicher (1997) p.113, it follows that $\sum_{j=1}^{\infty} \lambda_{jn}\xi_j < \varphi_{jn}, \Omega_0^{\frac{1}{2}}\varphi >< \infty$ with probability 1. Then, $< \mathcal{A}_1, \Omega_0^{\frac{1}{2}}\varphi > \to 0$ if and only if $\sqrt{n}\alpha_n \to \infty$, *i.e.* if and only if $\alpha_n \to 0$ slower than $\sqrt{n}$. This implies that α_n cannot be equal to $\frac{1}{n}$ and if it is equal to $\frac{1}{n}$ the term (19) diverges and then $\lim_{n\to\infty} < \mathcal{A}_1, \Omega_0^{\frac{1}{2}}\varphi >$ diverges with probability 1. Inequality (16) allows to conclude that $||\mathcal{A}|| \to \infty$ with probability 1.
Next, let consider term $\mathcal{B}$: $\mathcal{B} = (I - \Omega_0^{\frac{1}{2}}(\frac{1}{n}I + T_{(n)}^* T_{(n)})^{-1} T_{(n)}^* K_{(n)})\Omega_0^{\frac{1}{2}}\delta_*$. Then,

$$\begin{aligned} ||\mathcal{B}|| &\leq ||\Omega_0^{\frac{1}{2}}||||\frac{1}{n}(\frac{1}{n}I + T_{(n)}^* T_{(n)})^{-1}\delta_*|| \\ &= ||\Omega_0^{\frac{1}{2}}||\frac{1}{n}\Big(\sum_j \frac{< \delta_*, \varphi_{jn} >^2}{(\frac{1}{n} + \lambda_{jn}^2)^2}\Big)^{\frac{1}{2}} \end{aligned}$$

which converges to 0. This concludes the proof.

A.2 Proof of Theorem 2

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

$$\begin{aligned} \hat{\varphi}_\alpha - \varphi_* &= \overbrace{-(I - \Omega_0 K_{(n)}^* \Big(\alpha_n I_n + \frac{1}{n}I_n + K_{(n)}\Omega_0 K_{(n)}^*\Big)^{-1} K_{(n)})(\varphi_* - \varphi_0)}^{\mathcal{A}} \\ &\quad \underbrace{+ \Omega_0 K_{(n)}^* \Big(\alpha_n I_n + \frac{1}{n}I_n + K_{(n)}\Omega_0 K_{(n)}^*\Big)^{-1} \varepsilon_{(n)}}_{\mathcal{B}}. \end{aligned}$$

Under Assumption 4

$$\begin{aligned} ||\mathcal{A}|| &\leq \overbrace{||\,(I - \Omega_0^{\frac{1}{2}} T_{(n)}^* (\alpha_n I_n + T_{(n)} T_{(n)}^*)^{-1} K_{(n)})\Omega_0^{\frac{1}{2}}\delta_*\,||}^{\mathcal{A}1} \\ &\quad \underbrace{+||\,\Omega_0^{\frac{1}{2}} T_{(n)}^* (\alpha_n I_n + \frac{1}{n}I_n + T_{(n)} T_{(n)}^*)^{-1} \frac{1}{n} I_n (\alpha_n I_n + T_{(n)} T_{(n)}^*)^{-1} T_{(n)}\delta_*\,||}_{\mathcal{A}2} \\ ||\mathcal{A}1|| &= ||\Omega_0^{\frac{1}{2}}\Big[\alpha_n(\alpha_n I + T^*T)^{-1}\delta_* + \alpha_n[(\alpha_n I + T_{(n)}^* T_{(n)})^{-1} - (\alpha_n I + T^*T)]\delta_*\Big]|| \\ &\leq ||\Omega_0^{\frac{1}{2}}||\Big(||\alpha_n(\alpha_n I + T^*T)^{-1}\delta_*|| + ||(\alpha_n I + T_{(n)}^* T_{(n)})^{-1}||||T_{(n)}^* T_{(n)} - T^*T||||\alpha_n(\alpha_n I + T^*T)^{-1}\delta_*||\Big) \\ ||\mathcal{A}1||^2 &= \mathcal{O}(\alpha_n^\beta + \frac{1}{\alpha_n^2 n}\alpha_n^\beta) \end{aligned}$$

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

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 consider term $\mathcal{B}$. A similar decomposition as for $\mathcal{A}$ gives

$$\begin{aligned}
||\mathcal{B}||^2 &\leq ||\Omega_0^{\frac{1}{2}}||^2\Big(||\underbrace{T^*_{(n)}(\alpha_n I_n + T_{(n)}T^*_{(n)})^{-1}\varepsilon_{(n)}}_{\mathcal{B}1}||^2 \\
&\quad + ||\underbrace{T^*_{(n)}(\alpha_n I_n + \frac{1}{n}I_n + T_{(n)}T^*_{(n)})^{-1}(\frac{1}{n}I_n)(\alpha_n I_n + T_{(n)}T^*_{(n)})^{-1}\varepsilon_{(n)}}_{\mathcal{B}2}||^2\Big) \\
||\mathcal{B}1||^2 &\leq ||(\alpha_n I + T^*_{(n)}T_{(n)})^{-1}||^2||T^*_{(n)}\varepsilon_{(n)}||^2
\end{aligned}$$

and $T^*_{(n)}\varepsilon_{(n)} = \frac{1}{\sqrt{n}}\Big[\frac{1}{\sqrt{n}}\sum_i \varepsilon_i g(Z,w_i)\Big] = \frac{1}{\sqrt{n}}\mathcal{O}_p(1)$ 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 $\frac{1}{n}$ 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, simplifying the term 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 $\epsilon_n \to 0$,

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

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

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

then, $tr(\Omega_{y,\alpha}) = tr(\mathcal{C}) + tr(\mathcal{D})$. By using properties and the definition of the trace function, we get

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

which is an $\mathcal{O}_p(\alpha_n^\kappa)$ under the assumption that $\lim_{n\to\infty}\sum_{j=1}^n \frac{<\Omega_0\varphi_{jn},\varphi_{jn}>}{\lambda_{jn}^{2\kappa}} < \infty$. Then, $tr(\mathcal{C}) \to 0$ as $\alpha_n \to 0$. The $tr(\mathcal{D})$ is less or equal than $tr[T_{(n)}\Omega_0 T^*_{(n)}(\alpha_n I_n + T_{(n)}T^*_{(n)})^{-1}]$ 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 $L_F^2(Z)$ 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,

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

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 $L_F^2(Z)$. Moreover, the second term in brackets is an $\mathcal{O}(\frac{1}{\alpha_n^2 n}\alpha_n^\beta)$ and $||\Omega_0^{\frac{1}{2}}||^2 = \mathcal{O}(1)$ since Ω_0 is a compact operator, so we get $||\mathcal{C}\phi||^2 = \mathcal{O}(\alpha_n^\beta + \frac{1}{\alpha_n^2 n}\alpha_n^\beta)$.

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 speed 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 $\mathcal{O}_p(\alpha_n^\beta + \frac{1}{\alpha_n^2 n}\alpha_n^\beta)$ 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

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

Under Assumption 4,

$$\begin{aligned}\mathcal{A} &= \frac{1}{n} < K_{(n)}\Omega_0^{\frac{1}{2}}\delta_*, C_n^{-1}K_{(n)}\Omega_0^{\frac{1}{2}}\delta_* > = \frac{1}{n} < \delta_*, T_{(n)}^* C_n^{-1} T_{(n)}\delta_* > \\ &\leq \frac{1}{n}||\delta_*||||(\frac{1}{n}I + T_{(n)}^* T_{(n)})^{-1}T_{(n)}^* T_{(n)}||||\delta_*|| = \mathcal{O}_p\Big(\frac{1}{n}\Big)\end{aligned}$$

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

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

$$\begin{aligned}\mathcal{C} &= \varepsilon_{(n)}'\varepsilon_{(n)} - n\varepsilon_{(n)}' T_{(n)}(I + nT_{(n)}^* T_{(n)})^{-1}T_{(n)}^*\varepsilon_{(n)} \qquad (21) \\ \mathcal{C} - \sigma_*^2 &\leq (\varepsilon_{(n)}'\varepsilon_{(n)} - \sigma_*^2) + n\varepsilon_{(n)}' T_{(n)}(I + nT_{(n)}^* T_{(n)})^{-1}T_{(n)}^*\varepsilon_{(n)}.\end{aligned}$$

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

$$\begin{aligned}T_{(n)}^*\varepsilon_{(n)} &= \frac{1}{n}\sum_i \varepsilon_i g(Z, w_i) \\ n(I + nT_{(n)}^* T_{(n)})^{-1}T_{(n)}^*\varepsilon_{(n)} &= \frac{1}{n}\sum_i \varepsilon_i\Big((\frac{1}{n}I + T_{(n)}^* T_{(n)})^{-1}g(Z, w_i)\Big).\end{aligned}$$

The second term in (21) becomes

$$\begin{aligned}n\varepsilon_{(n)}' T_{(n)}(I + nT_{(n)}^* T_{(n)})^{-1}T_{(n)}^*\varepsilon_{(n)} &= < T_{(n)}^*\varepsilon_{(n)}, (\frac{1}{n}I + T_{(n)}^* T_{(n)})^{-1}T_{(n)}^*\varepsilon_{(n)} > \\ &\leq ||T_{(n)}^*\varepsilon_{(n)}||\Big|\Big|(\frac{1}{n}I + T_{(n)}^* T_{(n)})^{-1}T_{(n)}^*\varepsilon_{(n)}\Big|\Big|.\end{aligned}$$

The first norm is an $\mathcal{O}_p(\frac{1}{\sqrt{n}})$ since $||T^*_{(n)}\varepsilon_{(n)}|| \leq \frac{1}{\sqrt{n}}\Big[\frac{1}{\sqrt{n}}\sum_i \varepsilon_i||g(Z,w_i)||\Big]$ 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

$$\begin{aligned}
||(\frac{1}{n}I + T^*_{(n)}T_{(n)})^{-1}T^*_{(n)}\varepsilon_{(n)}|| &= ||\overbrace{\frac{1}{n}\sum_i \varepsilon_i\Big((\frac{1}{n}I + T^*T)^{-1}(T^*T)^{\frac{\gamma}{2}}h(Z,w_i)\Big)}^{\mathcal{C}1}|| \\
&\quad + \underbrace{||\frac{1}{n}\sum_i \varepsilon_i\Big([(\frac{1}{n}I + T^*_{(n)}T_{(n)})^{-1} - (\frac{1}{n}I + T^*T)^{-1}]g(Z,w_i)\Big)||}_{\mathcal{C}2} \\
||\mathcal{C}1|| &\leq \frac{n}{\sqrt{n}}\frac{1}{\sqrt{n}}\sum_i |\varepsilon_i|\underbrace{||\frac{1}{n}(\frac{1}{n}I + T^*T)^{-1}(T^*T)^{\frac{\gamma}{2}}||}_{=\mathcal{O}_p(n^{-\frac{\gamma}{2}})}||h(Z,w_i)|| \\
&= \mathcal{O}_p(\sqrt{n}n^{-\frac{\gamma}{2}}) = \mathcal{O}_p\Big(\Big(\frac{1}{\sqrt{n}}\Big)^{\gamma-1}\Big).
\end{aligned}$$

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

which converges faster than $||\mathcal{C}1||$. Hence, $(C - \sigma^2_*) = \mathcal{O}_p(\frac{1}{\sqrt{n}} + (\frac{1}{n})^{\frac{\gamma}{2}})$. Finally,

$$\begin{aligned}
\mathcal{B} &= \frac{2}{n} < \varepsilon_{(n)}, C_n^{-1}T_{(n)}\delta_* > = \frac{2}{n} < T^*_{(n)}C_n^{-1}\varepsilon_{(n)}, \delta_* > \\
&\leq \frac{2}{n}||\delta_*||||T^*_{(n)}C_n^{-1}\varepsilon_{(n)}|| = \mathcal{O}_p\Big(\Big(\frac{1}{n}\Big)^{\frac{\gamma+1}{2}}\Big).
\end{aligned}$$

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

By the Chebishev's Inequality,

$$\begin{aligned}
\nu^y_n\{\sigma \in \mathbb{R}_+; |\sigma^2 - \sigma^2_*| \geq \epsilon_n\} &\leq \mathbb{E}[(\sigma^2 - \sigma^2_*)|y_{(n)}]\frac{1}{\epsilon^2_n} \\
&= \frac{1}{\epsilon^2_n}\Big[Var(\sigma^2|y_{(n)}) + (\mathbb{E}(\sigma^2|y_{(n)}) - \sigma^2_*)^2\Big].
\end{aligned}$$

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 $Var(\sigma^2|y_{(n)}) = 2\mathbb{E}(\sigma^2|y_{(n)})\frac{1}{\xi_0+n-2}$ 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 remark that

$$\begin{aligned}
||(\sigma^2,\varphi) - (\sigma^2_*,\varphi_*)||_{\mathbb{R}_+\times L^2_F(Z)} &= ||(\sigma^2 - \sigma^2_*, \varphi - \varphi_*)||_{\mathbb{R}_+\times L^2_F(Z)} \\
&= \sqrt{< (\sigma^2 - \sigma^2_*, \varphi - \varphi_*), (\sigma^2 - \sigma^2_*, \varphi - \varphi_*) >_{\mathbb{R}_+\times L^2_F(Z)}} \\
&= \sqrt{< (\sigma^2 - \sigma^2_*), (\sigma^2 - \sigma^2_*) >_{\mathbb{R}_+} + < (\varphi - \varphi_*), (\varphi - \varphi_*) >_{L^2_F(Z)}} \\
&= (||\sigma^2 - \sigma^2_*||^2_{\mathbb{R}_+} + ||\varphi - \varphi_*||^2_{L^2_F(Z)})^{\frac{1}{2}}
\end{aligned}$$

$$
\begin{aligned}
&\leq \Big((||\sigma^2-\sigma_*^2||_{\mathbb{R}_+}+||\varphi-\varphi_*||_{L^2_F(Z)})^2\Big)^{\frac{1}{2}}\\
&= ||\sigma^2-\sigma_*^2||_{\mathbb{R}_+}+||\varphi-\varphi_*||_{L^2_F(Z)}
\end{aligned}
$$

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

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

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

$$
\mu_\alpha^{\sigma_*,y}\{\varphi\in L^2_F(Z); ||\varphi-\varphi_*||_{L^2_F(Z)}>\epsilon_n\}
$$

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 following decomposition:

$$
\begin{aligned}
\hat{\mathbb{E}}_\alpha(\varphi|y_{(n)})-\varphi_* &= \overbrace{-(I-\Omega_0^{\frac{1}{2}}\hat{T}^*_{(n)}(\alpha_nI_n+\hat{T}_{(n)}\hat{T}^*_{(n)})^{-1}\hat{K}_{(n)})(\varphi_*-\varphi_0)}^{\mathcal{A}}\\
&\quad+\underbrace{\Omega_0^{\frac{1}{2}}\hat{T}^*_{(n)}[(\alpha_nI_n+\Sigma_n+\hat{T}_{(n)}\hat{T}^*_{(n)})^{-1}-(\alpha_nI_n+\hat{T}_{(n)}\hat{T}^*_{(n)})^{-1}]\hat{K}_{(n)}(\varphi_*-\varphi_0)}_{\mathcal{B}}\\
&\quad+\underbrace{\Omega_0^{\frac{1}{2}}\hat{T}^*_{(n)}(\alpha_nI_n+\Sigma_n+\hat{T}_{(n)}\hat{T}^*_{(n)})^{-1}(\eta_{(n)}+\varepsilon_{(n)})}_{\mathcal{C}}\\
||\mathcal{A}||^2 &\leq ||\Omega_0^{\frac{1}{2}}||^2||\alpha_n(\alpha_nI+\hat{T}^*_{(n)}\hat{T}_{(n)})^{-1}\delta_*||^2\\
&\leq ||\Omega_0^{\frac{1}{2}}||^2\Big(||\alpha_n(\alpha_nI+T^*T)^{-1}\delta_*||+\\
&\quad||\alpha_n(\alpha_nI+\hat{T}^*_{(n)}\hat{T}_{(n)})^{-1}(T^*T-\hat{T}^*_{(n)}\hat{T}_{(n)})(\alpha_nI+T^*T)^{-1}\delta_*||\Big)^2\\
&= \mathcal{O}_p(\alpha_n^\beta+\alpha_n^{\beta-2}||\hat{T}^*_{(n)}\hat{T}_{(n)}-T^*T||^2)\\
||\mathcal{B}||^2 &\leq ||\Omega_0^{\frac{1}{2}}||^2\Big|\Big|\Big(\alpha_nI+\Big(\frac{\sigma^2}{n}+o_p(\frac{1}{n})\Big)I+\hat{T}^*_{(n)}\hat{T}_{(n)}\Big)^{-1}\Big|\Big|^2\Big|\Big|\Big(\frac{\sigma^2}{n}+o_p(\frac{1}{n})\Big)I||^2\\
&\quad||\underbrace{\hat{T}^*_{(n)}(\alpha_nI_n+\hat{T}_{(n)}\hat{T}^*_{(n)})^{-1}\hat{T}_{(n)}\delta_*}_{\mathcal{B}1}||^2\\
\mathcal{B}1 &= (\alpha_nI+\hat{T}^*_{(n)}\hat{T}_{(n)})^{-1}\hat{T}^*_{(n)}\hat{T}_{(n)}\delta_*\\
&= (\alpha_nI+T^*T)^{-1}T^*T\delta_*+[(\alpha_nI+\hat{T}^*_{(n)}\hat{T}_{(n)})^{-1}\hat{T}^*_{(n)}\hat{T}_{(n)}-(\alpha_nI+T^*T)^{-1}T^*T]\delta_*\\
&= (\alpha_nI+T^*T)^{-1}T^*T\delta_*+(\alpha_nI+\hat{T}^*_{(n)}\hat{T}_{(n)})^{-1}(\hat{T}^*_{(n)}\hat{T}_{(n)}-T^*T)\alpha_n(\alpha_nI+T^*T)^{-1}\delta_*\\
||\mathcal{B}1||^2 &= \mathcal{O}_p\Big(\alpha_n^\beta+\frac{1}{\alpha_n^2}||\hat{T}^*_{(n)}\hat{T}_{(n)}-T^*T||^2\alpha_n^\beta\Big)
\end{aligned}
$$

where we have used Assumptions 4, 5 and $\delta_*\in\mathcal{R}(T^*T)^{\frac{\beta}{2}}$. Next, we prove that $||\hat{T}^*_{(n)}\hat{T}_{(n)}-T^*T||^2=\mathcal{O}_p(\frac{1}{n}+h^{2\rho})$. For this, we notice that $\hat{K}^*_{(n)}\hat{K}_{(n)}\varphi$ has the same asymptotic behavior of $\int\int\varphi(z)\hat{f}(z|w_i)dz\frac{\hat{f}(z,w_i)}{\hat{f}(z)}dw_i$. In Darolles *et al.* (2003, Appendix B, under Assumptions B.1-B.5) it

is proved that $||\int\int \varphi(z)\hat{f}(z|w_i)dz\frac{\hat{f}(z,w_i)}{\hat{f}(z)}dw_i - \mathbb{E}(\mathbb{E}(\varphi|W)|Z)||^2 = \mathcal{O}_p(\frac{1}{nh^p}+h^{2\rho})$ and it follows that $\hat{K}^*_{(n)}\hat{K}_{(n)}$ is of the same order. Then, operator $\Omega_0^{\frac{1}{2}}$ in $\hat{T}^*_n$ has a smoothing effect on the variance term of the MISE of $\hat{K}^*_{(n)}\hat{K}_{(n)}\varphi$ which becomes of order $\frac{1}{n}$. This prove the results and 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

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

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

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

$$\begin{aligned}
||\mathcal{C}2||^2 &\leq ||\Omega_0^{\frac{1}{2}}||^2||(\alpha_n I + (\frac{\sigma^2}{n} + o_p(\frac{1}{n}))I + \hat{T}^*_{(n)}\hat{T}_{(n)})^{-1}||^2||(\frac{\sigma^2}{n} + o_p(\frac{1}{n}))I||^2 \\
&\quad ||\hat{T}^*_{(n)}(\alpha_n I_n + \hat{T}_{(n)}\hat{T}^*_{(n)})^{-1}(\eta_{(n)}+\varepsilon_{(n)})||^2
\end{aligned}$$

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 $||\hat{\mathbb{E}}_\alpha(\varphi|y_{(n)}) - \varphi_*||^2 = \mathcal{O}_p(\alpha_n^\beta + \frac{1}{\alpha_n^2 n}\frac{1}{\alpha_n^2}(\frac{1}{n}+h^{2\rho})\frac{1}{\alpha_n^2 n})$ 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 $R^\alpha = (\alpha I_n + T_{(n)}T^*_{(n)})^{-1}$ and $R^\alpha_{(n)} = (\alpha I_n + \frac{1}{n}I_n + T_{(n)}T^*_{(n)})^{-1}$. We decompose the residual as

$$\begin{aligned}
\vartheta^{(2)}_\alpha &= \overbrace{T^*_{(n)}[I - (\alpha K_{(n)}\Omega_0 K^*_{(n)}R^\alpha + K_{(n)}\Omega_0 K^*_{(n)})R^\alpha]K_{(n)}(\varphi_* - \varphi_0)}^{\mathcal{A}} \\
&\quad + \overbrace{T^*_{(n)}[(\alpha K_{(n)}\Omega_0 K^*_{(n)}R^\alpha + K_{(n)}\Omega_0 K^*_{(n)})R^\alpha - (\alpha K_{(n)}\Omega_0 K^*_{(n)}R^\alpha_{(n)} + K_{(n)}\Omega_0 K^*_{(n)})R^\alpha_{(n)}]K_{(n)}(\varphi_* - \varphi_0)}^{\mathcal{B}} \\
&\quad + \overbrace{T^*_{(n)}[I - (\alpha K_{(n)}\Omega_0 K^*_{(n)}R^\alpha + K_{(n)}\Omega_0 K^*_{(n)})R^\alpha]\varepsilon_{(n)}}^{\mathcal{C}} \\
&\quad + \underbrace{T^*_{(n)}[(\alpha K_{(n)}\Omega_0 K^*_{(n)}R^\alpha + K_{(n)}\Omega_0 K^*_{(n)})R^\alpha - (\alpha K_{(n)}\Omega_0 K^*_{(n)}R^\alpha_{(n)} + K_{(n)}\Omega_0 K^*_{(n)})R^\alpha_{(n)}]\varepsilon_{(n)}}_{\mathcal{D}}.
\end{aligned}$$

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

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. The same as the Proof of Lemma 3 with $T_{(n)}$, $T^*_{(n)}$, $K_{(n)}$ and $K^*_{(n)}$ replaced by $\hat{T}_{(n)}$, $\hat{T}^*_{(n)}$, $\hat{K}_{(n)}$ and $\hat{K}^*_{(n)}$. Then, we have the same decomposition and we get: $||\mathcal{A}||^2 = \mathcal{O}_p(\alpha^{\beta+2} + (\frac{1}{n}+h^{2\rho}))$, $||\mathcal{B}||^2 = \mathcal{O}_p(\alpha^{\beta+2} + (\frac{1}{n}+h^{2\rho})\alpha^\beta)$, $||\mathcal{C}||^2 = \mathcal{O}_p(\frac{1}{\alpha^2 n}(\frac{1}{n}+h^{2\rho}))$, $||\mathcal{D}||^2 = \mathcal{O}_p(\frac{1}{n} + \frac{1}{\alpha^2 n}(\frac{1}{n}+h^{2\rho}))$.

References

[1] Blundell, R. and J.L., Powell, 2003, Endogeneity in nonparametric and semiparametric regression models, in: M., Dewatripont, L.-P., Hansen and D.J., Turnovsky, (Eds.), Advances in economics and econometrics: theory and applications, Vol.2, pp. 312-357. Cambridge, UK:Cambridge University Press.

[2] Breiman, L., Le Cam, L. and L., Schwartz, 1964, Consistent estimates and zero-one sets. Annals of Mathematical Statistics, **35**, 157 - 161.

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

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

[5] Chen, X. and M., Reiss, 2007, On rate optimality for ill-posed inverse problems in econometrics, working paper.

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

[7] Chow, Y.S. and H. Teicher, 1997, Probability Theory, Springer-Verlag New York.

[8] Darolles, S., Florens, J.P., and E., Renault, 2003, Nonparametric instrumental regression. Working paper.

[9] Diaconis, F., and D., Freedman, 1986, On the consistency of Bayes estimates. Annals of Statistics **14**, 1-26.

[10] Doob, J. L., 1949, Application of the theory of martingales, in: Le calcul des probabilités et ses applications, pages 23 - 27. Centre National de la Recherche Scientifique. Paris, 1949. Colloques Internationaux du Centre National de la Recherche Scientifique, no. 13.

[11] Engl, H.W., Hanke, M. and A., Neubauer, 2000, Regularization of inverse problems, Kluwer Academic, Dordrecht.

[12] Fève, F. and J.P., Florens, (2010), The practice of nonparametric estimation by solving inverse problems: the example of transformation models. The Econometric Journal, (EC)2 Special Issue.

[13] Florens, J.P., 2003, Inverse problems and structural econometrics: the example of instrumental variables. Invited Lectures to the World Congress of the Econometric Society, Seattle 2000. In: M., Dewatripont, L.-P., Hansen, and S.J., Turnovsky, (Eds.),

Advances in Economics end econometrics: theory and applications, Vol.II, pp. 284-311. Cambridge University Press.

[14] Florens, J.P., Johannes, J. and S., Van Bellegem, 2005, Instrumental regression in partially linear models. Discussion Paper # 0537, Institut de statistique, Université catholique de Louvain.

[15] Florens, J.P., Johannes, J. and S., Van Bellegem, 2007, Identification and estimation by penalization in nonparametric instrumental regression. Discussion Paper # 0721, Institut de statistique, Université catholique de Louvain.

[16] Florens, J.P., Mouchart, M., and J.M., Rolin, 1990, Elements of Bayesian statistics, Dekker, New York.

[17] Florens, J.P., and A., Simoni, 2009a, Regularized posteriors in linear ill-posed inverse problems. Preprint. Available at
`http://simoni.anna.googlepages.com/Regularized_Posterior_Florens_Simoni.pdf`

[18] Florens, J.P., and A., Simoni, 2009b, Regularizing priors for linear inverse problems. Preprint. Available at
`http://simoni.anna.googlepages.com/Regularizing_Priors.pdf`

[19] Franklin, J.N., 1970, Well-posed stochastic extension of ill-posed linear problems. Journal of Mathematical Analysis and Applications **31**, 682 - 716.

[20] Freedman, D., 1963, On the asymptotic behavior of Bayes estimates in the discrete case. The Annals of Mathematical Statistics, **34**, 1386-1403.

[21] Freedman, D., 1965, On the asymptotic behavior of Bayes estimates in the discrete case II. The Annals of Mathematical Statistics, **36**, 454-456.

[22] Gagliardini, P. and O., Scaillet, 2009, Tikhonov regularization for nonparametric instrumental variable estimators, preprint.

[23] Gelman, A. and D.B., Rubin, 1992, Inference from iterative simulation using multiple sequences. Statistical Science, **7**, 457 - 472.

[24] Ghosh, J.K and R.V. Ramamoorthi, 2003, Bayesian nonparametrics, Springer Series in Statistics.

[25] Ghosal, S., A review of consistency and convergence rates of posterior distribution, in Proceedings of Varanashi Symposium in Bayesian Inference, Banaras Hindu University.

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

[27] Kato, T., 1995, Perturbation theory for linear operators, Springer.

[28] Kress, R., 1999, Linear integral equation, Springer.

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

[30] Loubes, J.M. and C., Marteau, 2009, Oracle inequality for instrumenatl variable regression. Available on Arxiv.

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

[32] Neveu, J., 1965, Mathematical foundations of the calculus of probability, San Francisco: Holden-Day.

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

[34] O'Hagan, A., Kennedy, M.C. and J.E., Oakley, 1998, Uncertainty analysis and other inference tools for complex computer codes, in: J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (Eds.), Bayesian Statistics 6, Oxford University Press, pp. 503 - 524.

[35] O'Sullivan, F., 1996, A statistical perspective on ill-posed inverse problems. Statistical Science 1, 502-527.

[36] Rasmussen, C.E. and C.K.I., Williams, 2006, Gaussian processes for machine learning, MIT Press.

[37] Schwartz, L., 1965, On Bayes procedures. Probability Theory and Related Fields, **4**, 10-26.

[38] Simoni, A., 2009, Bayesian analysis of linear inverse problems with applications in economics and finance, PhD Dissertation - Université de Science Sociales, Toulouse. Available at `http://simoni.anna.googlepages.com/PhD_Dissertation.pdf`

[39] Van der Vaart, A.W. and Van Zanten, J.H., 2008a. Rates of contraction of posterior distributions based on gaussian process priors. Annals of Statistics, **36**, 1435-1463.

[40] Van der Vaart, A.W. and Van Zanten, J.H., 2008b. Reproducing kernel Hilbert spaces of gaussian priors. *Pushing the limits of contemporary statistics: contributions in honor of Jayanta K. Ghosh. IMS Collections*, **3**, 200 - 222. Institute of Mathematical Statistics.

[41] von Mises, R., 1981, Probability, statistics and truth. Dover Publications Inc., New York, english edition.