Série des Documents de Travail

n° 2015-09
Gaussian processes and Bayesian moment estimation

J-P.Florens[1]
A.Simoni[2]

[1] Toulouse School of Economics, E-mail : jean-pierre.florens@tse-fr.eu

[2] CREST, E-mail : simoni.anna@gmail.com

Gaussian processes and Bayesian moment estimation*

Jean-Pierre Florens† Anna Simoni‡

Toulouse School of Economics CREST-Ensae and CNRS

This version: November 2015

Abstract

Given a set of moment restrictions that characterize a parameter θ, we investigate a semiparametric Bayesian approach for estimation of θ that imposes these moment restrictions in the nonparametric prior for the data distribution. As main contribution, we construct a degenerate Gaussian process prior for the density function associated with the data distribution F that imposes overidentifying restrictions. We show that this prior is computationally convenient. Since the likelihood function is not specified by the model we construct it based on a linear functional transformation of F that has an asymptotically Gaussian empirical counterpart. This likelihood is used to construct the posterior distribution. We provide a frequentist validation of our procedure by showing: consistency of the maximum a posteriori estimator for θ, consistency and asymptotic normality of the posterior distribution of θ.

Key words: Moment restrictions, Gaussian processes, overidentification, posterior consistency, functional equation.

JEL code: C11, C14, C13

*First version: February 2012. The authors especially thank Yuichi Kitamura for insightful discussions. The authors are grateful to Christoph Breunig, Laurent Davezies, Xavier D'Haultfoeuille, Taisuke Otsu and to participants to seminars and conferences in: Berlin, Boston College, Bristol, Carlos III - Madrid, CREST, Northwestern, AFSE 2012 in Paris, SBIES 2015, ICEEE 2015, NASM 2012, Toulouse. We thank financial support from ANR-13-BSH1-0004 (IPANEMA). Anna Simoni gratefully acknowledges financial support from Labex ECODEC (ANR - 11-LABEX-0047), SFB-884 and hospitality from the University of Mannheim and Boston College.

†Toulouse School of Economics - 21, allée de Brienne - 31000 Toulouse (France). Email: jean-pierre.florens@tse-fr.eu

‡CREST, 15 Boulevard Gabriel Péri, 92240 Malakoff (France). Email: simoni.anna@gmail.com (*corresponding author*).

1 Introduction

Econometric models are often formulated in terms of moment restrictions that hinge on economic restrictions. These restrictions provide the only information available about the parameter of interest θ and the data distribution. Given a set of moment restrictions that characterize θ, this paper builds a semiparametric Bayesian inference procedure for θ that imposes these moment restrictions in the nonparametric prior distribution for the data distribution and that is computationally convenient. Apart from these moment restrictions, the data distribution is left unrestricted.

A main advantage of Bayesian inference consists in providing a well-defined posterior distribution that is important for many decision problems and for predictive analysis. On the other hand, constructing Bayesian inference procedures for moment restrictions-based models presents two difficulties. A first difficulty is due to the fact that a likelihood is not available. A second difficulty arises because imposing overidentifying moment restrictions on the prior distribution for the data distribution is challenging. The contribution of this paper is to propose an elegant approach that allows to deal with these two difficulties. As a by-product we show that the quasi-likelihood of some Laplace-type procedures arises as the limit of our Bayesian procedure.

The model we consider is as follows. Let x be an observable random element in $\mathbb{R}^m$ with distribution F and $x_1, \ldots, x_n$ be an *i.i.d.* sample of x. The parameter $\theta \in \Theta \subset \mathbb{R}^p$ is linked to the data generating process (DGP) F through the moment restrictions

$$\mathbf{E}^F \left[h(\theta, x)\right] = 0, \tag{1.1}$$

where $h(\theta, x) = (h_1(\theta, x), \ldots, h_d(\theta, x))^T$ and the functions $h_j(\theta, x)$, $j = 1, \ldots, d$ are real-valued and known. We assume $d \geq p$ and our main interest is the case where $d > p$, which is in general more challenging than the case $d = p$. Apart from (1.1), F is completely unrestricted. The Bayesian procedure proposed in this paper constructs a nonparametric prior for F with support equal to the subset of distributions that satisfy the moment restrictions for a given θ. Because the moment restrictions are imposed in the prior for F, the distributions generated from the prior satisfy (1.1) by construction.

Imposing moment restrictions via semiparametric priors may be challenging depending on the relationship existing between θ and F. More precisely, when the model is just-identified (*i.e.* $p = d$), and under mild conditions, (1.1) characterizes θ as an explicit function of F: $\theta = b(F)$, where b is a function defined on the set of probability distributions. Thus, for a particular transformation b, the prior of θ may be recovered from an unrestricted nonparametric prior on F, and the (θ, F)s generated by this prior automatically satisfy the constraints.

On the contrary, in an overidentified model where $d > p$, θ cannot be expressed as an explicit function of F. Indeed, (1.1) imposes constraints on F and the existence of a solution θ to (1.1) is guaranteed only for a subset of distributions F. Therefore, a restricted nonparametric prior on F must be specified conditionally on θ and the support of this prior is a proper subset of the set of probability distributions. It turns out that incorporating overidentifying moment restrictions in a semiparametric prior for (θ, F) is not straightforward. In this paper we propose a way to construct a semiparametric prior that incorporates the overidentifying moment restrictions.

Our strategy is based on a degenerate Gaussian process ($\mathcal{GP}$) prior with restricted support which is easy to deal with and that works as follows. The DGP F is assumed to admit a density function f with respect to some positive measure Π chosen by the econometrician (for instance the Lebesgue measure). Then, we endow f with a $\mathcal{GP}$ prior conditional on θ. The (overidentifying) moment restrictions are incorporated by constraining the prior mean and prior covariance of this $\mathcal{GP}$ in an appropriate way. Because this prior imposes the moment restrictions, it will be degenerate on a proper subset of the set of probability density functions. The reason for the appropriateness of a $\mathcal{GP}$ prior in such a framework is due to the fact that the moment equations in (1.1) are linear in f and the linearity of the model matches extremely well with a $\mathcal{GP}$ prior. An advantage of our method is that, in both the just-identified and overidentified cases, the moment restrictions are imposed directly through the $\mathcal{GP}$ prior of f given θ without requiring a second step projection over the set of density functions satisfying the moment restrictions. To the best of our knowledge a $\mathcal{GP}$ prior has not been used yet in the moment estimation framework.

Our Bayesian procedure, that we call the $\mathcal{GP}$-approach, is constructed as follows. In the overidentified case we first specify a prior on θ and then a $\mathcal{GP}$ prior on f conditional on θ. In the just-identified case we may either proceed as in the overidentified case or specify an unrestricted $\mathcal{GP}$ prior on f and then deduce from it the prior for θ through the explicit relationship $\theta = b(f)$. We circumvent the difficulty of the likelihood function specification, which is not available, by constructing a linear functional transformation of the DGP F such that its empirical counterpart, say r_n, has an asymptotic Gaussian distribution. This will be used as the sampling model. Therefore, our model is approximately conjugate and allows easy computations while being nonparametric in F.

We provide a closed-form expression for the marginal posterior distribution of θ (obtained by integrating out f) and propose the maximum of this distribution as an estimator for θ. The maximum a posteriori of θ is usually not available in closed-form but can be easily computed via drawn from the marginal posterior. We show that the quasi-likelihood function (also called limited information likelihood) used, among others, by Kim [2002] and Chernozhukov and Hong [2003], can be obtained as the limit of the marginal posterior distribution for θ when the $\mathcal{GP}$ for f is allowed to become diffuse. In addition, when the prior for f becomes noninformative, the marginal posterior distribution for θ becomes the same (up to constants) as the GEL objective function with quadratic criterion and is a monotonic transformation of the continuous updating GMM objective function (Hansen et al. [1996]).

Finally, we provide a frequentist validation of our method by showing: *(i)* frequentist consistency of the maximum a posteriori estimator, *(ii)* posterior consistency and *(iii)* asymptotic normality of the posterior distribution of θ.

Related literature. Estimation of a parameter by exploiting the only information contained in moment restrictions is at the core of econometrics and statistical literature. Since Hansen [1982] and Hansen and Singleton [1982], the generalized method of moments (GMM) estimator and its variants have been extensively applied in econometrics. Alternative frequentist estimators to the GMM and the continuous updating GMM estimators includes the empirical likelihood (EL), exponential tilting, exponentially tilted EL and generalized empirical likelihood (GEL) estimators (*e.g.* Owen [1988], Qin and Lawless [1994], Smith [1997], Kitamura and Stutzer [1997], Kitamura [1997], Imbens et al. [1998], Newey and Smith [2004], Schennach [2007], Kitamura [2007]).

Since the works of Florens and Rolin [1994] and Zellner [1996], much attention has been devoted to construct posterior distributions for Bayesian inference and predictive analysis in presence of moment restrictions. There are two ways to construct a semiparametric Bayesian procedure to make inference on θ by only using the information contained in the moment restrictions (1.1). The first way consists in constructing a quasi-likelihood by exponentiating the generalized method of moments (GMM) criterion function. The corresponding approach is quasi-Bayesian and has been investigated *e.g.* by Kwan [1999], Kim [2002], Chernozhukov and Hong [2003], Liao and Jiang [2011], Gallant [2015] and Gallant et al. [2015] among others. Our paper shows that the quasi-likelihood used in this type of approach arises as the limit of our $\mathcal{GP}$ prior as it becomes diffuse. We provide thus a pure Bayesian justification to this approach.

The second way is purely Bayesian and consists in imposing the moment restrictions in the prior for (θ, F) while leaving the likelihood completely unrestricted. The approach proposed in this paper is of this type and constructs a constrained prior distribution that is different with respect to the priors proposed so far. Previous contributions include Chamberlain and Imbens [2003] who use a Dirichlet prior, Lazar [2003] who studies the validity of EL as the basis for Bayesian inference, and Schennach [2005] who proposes a Bayesian exponentially tilted EL which relies on a non-informative prior, different from a Dirichlet process, on the space of distributions. Recent contributions are Kitamura and Otsu [2011] and Shin [2014]. They propose to first specify an unrestricted Dirichlet process mixture (DPM) prior for F and a mixture of Dirichlet Process prior, respectively. Then, in a second step they select the distribution that, among all the distributions satisfying the moment restrictions, minimizes the Kullback-Leibler divergence to the F generated by the Dirichlet prior. A nonparametric prior constructed by minimizing the Kullback-Leibler divergence is also proposed by Ragusa [2007]. Finally, Bornn et al. [2015] use Hausdorff measures to build probability tools for dealing with moment estimation.

The paper is organized as follows. The $\mathcal{GP}$-approach is described in section 2, which contains our main contribution. In section 3 we analyze asymptotic properties of the posterior distribution of θ and of the maximum a posteriori estimator. In section 4 we show the link existing between our approach and some frequentist approaches to moment estimations. In section 5 we detail how to implement our method for both the just identified and the overidentified case through simulation studies. All the proofs are gathered in the Appendix.

2 The Gaussian Process ($\mathcal{GP}$) -approach

Let x be a continuous random element in $S \subseteq \mathbb{R}^m$ with distribution F and $x_1, \ldots, x_n$ be an i.i.d. sample of x. Assume that F is absolutely continuous with respect to some positive measure Π (*e.g.* the Lebesgue measure) with density function f. In other words, conditionally on f the data are drawn from F: $x_1, \ldots, x_n | f \sim F$. The set of probability density functions on S with respect to Π is denoted by M.

Let $\theta \in \Theta \subseteq \mathbb{R}^p$ be the parameter of interest characterized by (1.1). By adopting a frequentist point of view, we denote, throughout the paper, the true value of θ by θ_*, the true DGP by F_* and its density with respect to Π by f_*. The model is assumed to be well-specified, that is, $\mathbf{E}^{F_*}(h(\theta_*, x)) = 0$ holds. We endow $S \subseteq \mathbb{R}^m$ with the trace of the Borelian σ-field $\mathfrak{B}_S$ and specify Π as a positive measure on this subset. We denote by $\mathcal{E} = L^2(S, \mathfrak{B}_S, \Pi)$ the Hilbert space of square integrable functions on S with respect to Π and by $\mathfrak{B}_{\mathcal{E}}$ the Borel σ-field generated by the open sets of $\mathcal{E}$. The scalar product and norm on this space are defined in the usual way and denoted by $\langle \cdot, \cdot \rangle$ and $|| \cdot ||$, respectively.

The parameters of the model are (θ, f), where f is the nuisance parameter, and the parameter space is

$$\Lambda = \left\{ (\theta, f) \in \Theta \times \mathcal{E}_M; \int h(\theta, x) f(x) \Pi(dx) = 0 \right\}, \qquad \mathcal{E}_M := \mathcal{E} \cap M,$$

where $h : \Theta \times \mathbb{R}^m \to \mathbb{R}^d$ is a known function. In the following of the paper we maintain the following assumption.

Assumption 2.1. *(i) The true f_* satisfies $f_* \in \mathcal{E}_M := \mathcal{E} \cap M$; (ii) the moment function $h(\theta, \cdot)$ is such that $h_i(\theta, \cdot) \in \mathcal{E}$ for every $i = 1, \ldots, d$ and for every $\theta \in \Theta$, where h_i denotes* *the i-th component of h; (iii)* $d \geq p$.

Assumption 2.1 *(i)* restricts f_* to be square integrable with respect to Π and is for instance verified if f_* is bounded and Π is a bounded measure. The model is made up of three elements that we detail in the next two subsections: a prior on θ, denoted by $\mu(\theta)$, a conditional prior on f given θ, denoted by $\mu(f|\theta)$ and the sampling model. In the following, we shorten "almost surely" by "a.s." and omit the probability which "a.s." refers to. We denote by $\mathbf{E}^F$ the expectation taken with respect to F and by $\mathbf{E}^*$ the expectation taken with respect to F_*.

2.1 Prior distribution

We specify a prior probability measure μ for (θ, f) of the form $\mu(\theta, f) = \mu(\theta)\mu(f|\theta)$. By abuse of notation, $\mu(\theta)$ will also denote the density of the prior distribution of θ with respect to the Lebesgue measure in the case it admits it. The prior $\mu(\theta)$ may either be flat (non-informative) or incorporate any additional information available to the econometrician about θ.

Given a value for θ, the conditional prior $\mu(f|\theta)$ is specified such that its support equals the subset of probability density functions in $\mathcal{E}_M$ that satisfy (1.1) for this particular value of θ. At the best of our knowledge, the construction of such a conditional prior $\mu(f|\theta)$ is new in the literature and we now explain it in detail.

Construction of the conditional prior $\mu(f|\theta)$. We construct the conditional prior distribution $\mu(f|\theta)$ of f, given θ, as a $\mathcal{GP}$ on $\mathfrak{B}_{\mathcal{E}}$ with mean function $f_{0\theta} \in \mathcal{E}_M$ and covariance operator $\Omega_{0\theta} : \mathcal{E} \to \mathcal{E}$. We restrict $f_{0\theta}$ and $\Omega_{0\theta}$ to guarantee that the trajectories f generated by $\mu(f|\theta)$ are such that the corresponding F (which is given by $F = f\Pi$) integrates to 1 and satisfies equation (1.1) with probability 1. The two sets of restrictions that we impose are the following (one on $f_{0\theta}$ and one on $\Omega_{0\theta}$):

Restriction 1 (Restriction on $f_{0\theta}$)**.** The prior mean function $f_{0\theta} \in \mathcal{E}_M$ is chosen such that

$$\int h(\theta, x) f_{0\theta}(x)\Pi(dx) = 0. \tag{2.1}$$

Restriction 2 (Restriction on $\Omega_{0\theta}$)**.** The prior covariance operator $\Omega_{0\theta} : \mathcal{E} \to \mathcal{E}$ is chosen such that

$$\begin{cases} \Omega_{0\theta}^{1/2} h(\theta, x) &= 0 \\ \Omega_{0\theta}^{1/2} 1 &= 0 \end{cases} \tag{2.2}$$

where $\Omega_{0\theta}^{1/2} : \mathcal{E} \to \mathcal{E}$ and $\Omega_{0\theta} = \Omega_{0\theta}^{1/2}\Omega_{0\theta}^{1/2}$.

The covariance operator $\Omega_{0\theta}$ is linear, self-adjoint and trace-class.[1] Due to Restriction 2, $\Omega_{0\theta}$ is not injective. In fact, the null space of $\Omega_{0\theta}$, denoted by $\mathfrak{N}(\Omega_{0\theta})$, is not trivial and contains effectively the constant 1 – which implies that the trajectories f generated by the prior integrate to 1 a.s. (with respect to Π) – and the function $h(\theta, x)$ – which implies that the trajectories f satisfy the moment conditions a.s. This means that $\Omega_{0\theta}$ is degenerate in the directions along which we want that the corresponding projections of f and $f_{0\theta}$ are equal. Therefore, the support of $\mu(f|\theta)$ is a proper subset of $\mathcal{E}$. equal. This is the meaning of the next lemma.

Lemma 2.1. *The conditional $\mathcal{GP}$ prior $\mu(f|\theta)$, with mean function $f_{0\theta}$ and covariance operator $\Omega_{0\theta}$ satisfying Restrictions 1 and 2, generates trajectories f that satisfy $\mu(f|\theta)$-a.s. the conditions*

$$\int f(x)\Pi(dx) = 1 \qquad \text{and} \qquad \int h(\theta, x)f(x)\Pi(dx) = 0.$$

Remark 2.1. Restrictions 1 and 2 imply that the trajectories generated by $\mu(f|\theta)$ integrates to 1 (with respect to Π) and satisfy (1.1) a.s. but they do not guarantee non-negativity of the trajectories. Thus, the support of $\mu(f|\theta)$ is smaller than $\mathcal{E}$ but bigger than $\mathcal{E}_M$. To impose non-negativity we could: (i) either project the prior on the space of non-negative functions or (ii) write $f = g^2$, $g \in \mathcal{E}$, and specify a conditional prior distribution, given θ, for g instead of for f. The resulting prior distribution would not be Gaussian anymore and the resulting posterior for θ would not be available in closed form which is instead one of the main advantages of our procedure. Because our goal is to make inference on θ, and f is a nuisance parameter, failing to impose the non-negativity constraint is not an issue as long as our procedure is shown to be consistent for θ (which we show in section 3).

From a practical implementation point of view, a covariance operator satisfying Restriction 2 may be constructed as follows. Let $(\lambda_j)_{j\in\mathbb{N}}$ be a decreasing sequence of non-negative numbers accumulating at 0 such that $\sum_j \lambda_j < \infty$, and $(\varphi_j)_{j\in\mathbb{N}}$ be a basis for $\mathcal{E}$. Then, $\forall \phi \in \mathcal{E}$: $\Omega_{0\theta}\phi = \sum_{j=0}^{\infty} \lambda_j \langle \phi, \varphi_j \rangle \varphi_j$. Remark that $(\lambda_j)_{j\in\mathbb{N}}$ and $(\varphi_j)_{j\in\mathbb{N}}$ correspond to the eigenvalues and eigenfunctions of $\Omega_{0\theta}$, respectively.

Since the null space $\mathfrak{N}(\Omega_{0\theta}) \subset \mathcal{E}$ is spanned by $\{1, h_1(\theta, \cdot), \ldots, h_d(\theta, \cdot)\}$, we can set the first eigenfunctions of $\Omega_{0\theta}$ equal to the elements of any basis of $\mathfrak{N}(\Omega_{0\theta})$. Restriction 2 is then fulfilled by setting the corresponding eigenvalues equal to 0. For instance, if $\{1, h_1(\theta, \cdot), \ldots, h_d(\theta, \cdot)\}$ are orthonormal as elements of $\mathcal{E}$, then $\mathfrak{N}(\Omega_{0\theta})$ has dimension $d + 1$, the first eigenvalues are $(\varphi_0, \varphi_1, \ldots, \varphi_d)^T = (1, h^T)^T$ and the corresponding eigenvalues are $\lambda_j = 0$, $\forall j = 0, 1, \ldots, d$. Remark that in this case, necessarily, $\int \Pi(dx) = 1$. If

[1] A trace-class operator is a compact operator with eigenvalues that are summable. Remark that this guarantees that the trajectories f generated by $\mu(f|\theta)$ satisfy $\int f^2 d\Pi < \infty$ a.s.

$\{1, h_1(\theta,\cdot), \ldots, h_d(\theta,\cdot)\}$ are not orthonormal then one can use their orthonormalized counterparts as the first eigenfunctions of $\Omega_{0\theta}$. The latter is the method we use to implement our procedure. The remaining components $(\varphi_j)_{j>d}$ are chosen such that $(\varphi_j)_{j\geq 0}$ forms an orthonormal basis of $\mathcal{E}$ and $(\lambda_j)_{j>d}$ are chosen such that $\sum_{j>d}\lambda_j < \infty$. Hence,

$$\forall \phi \in \mathcal{E}, \quad \Omega_{0\theta}\phi = \sum_{j=d+1}^{\infty} \lambda_j \langle \phi, \varphi_j \rangle \varphi_j.$$

Examples of choices for $(\lambda_j)_{j>d}$ are, for some constant $c > 0$: *(i)* $\lambda_j = cj^{-a}$ with $a > 1$, *(ii)* $\lambda_j = ce^{-j}$. In section 5 we provide some examples that clarify the construction of $\Omega_{0\theta}$.

Remark 2.2. In the just-identified case where $d = p$ and the moment restrictions (1.1) can be solved explicitly for θ (that is, $\theta = b(f)$, for some functional b), then the prior for (θ, f) may be constructed in an alternative way: one can first specify a prior for f and then recover from it the prior for θ. When b is a linear functional and θ can take any value in $\mathbb{R}^p$, one can specify a $\mathcal{GP}$ prior $\mu(f)$ for f (independent of θ) with a mean function f_0 restricted only to be a *pdf* and a covariance operator Ω_0 restricted only to satisfy $\Omega_0^{1/2}1 = 0$. Then, the prior for θ is obtained through the transformation $b(\cdot)$ and will be Gaussian. Because the support of this prior is $\mathbb{R}^p$, then this approach is feasible if every value in $\mathbb{R}^p$ is plausible for θ. For example, if $\theta = \mathbf{E}^F(x)$ and the support of x is $\mathbb{R}^p$, then $b(f) = \langle f, \iota\rangle$ and $\mu(\theta) = \mathcal{N}(\langle f_0, \iota\rangle, \langle \Omega_0\iota, \iota\rangle)$, where $\iota \in \mathcal{E}$ denotes the identity function $\iota(x) = x$.

2.2 The sampling model

Given the observed *i.i.d.* sample $(x_1, \ldots, x_n)$, the likelihood function is $\prod_{i=1}^n f(x_i)$. While apparently simple, using this likelihood for Bayesian inference on θ makes the analysis of the posterior distribution complicated. This is because to compute the posterior for θ one has to marginalize out f. Since a $\mathcal{GP}$ prior is not a natural conjugate of the *i.i.d.* model then, marginalization of f has to be carried out through numerical, or Monte Carlo, integration on a functional space, which may be computationally costly. To avoid this difficulty, we propose an alternative and original way to construct the sampling model that allows for a conjugate analysis and prevents from numerical integration. Our approach is based on a functional transformation r_n of the sample $x_1, \ldots, x_n$.

This transformation r_n is chosen by the researcher and must have the following characteristics: *(I)* r_n is an observable element of an infinite-dimensional Hilbert space $\mathcal{F}$ (to be defined below), for instance a L^2-space; *(II)* r_n converges weakly towards a Gaussian process in $\mathcal{F}$; *(III)* the expectation of r_n, conditional on f, defines a linear operator $K : \mathcal{E} \to \mathcal{F}$ such that $\mathbf{E}^F(r_n) = Kf$, where $\mathcal{F}$ is an infinite-dimensional separable Hilbert space. Moreover, $r_n \in \mathcal{F}$ is a Hilbert space-valued random variable (H-r.v.). We recall that, for a complete probability space $(Z, \mathcal{Z}, \mathbb{P})$, r_n is a H-r.v. if it defines a measurable map $r_n : (Z, \mathcal{Z}, \mathbb{P}) \to (\mathcal{F}, \mathfrak{B}_{\mathcal{F}})$, where $\mathfrak{B}_{\mathcal{F}}$ denotes the Borel σ-field generated by the open sets of $\mathcal{F}$.

Construction of r_n. Let $\mathfrak{T} \subseteq \mathbb{R}^l$, $l > 0$. To construct r_n we first select a function $k(t,x) : \mathfrak{T} \times S \to \mathbb{R}$ (or in $\mathbb{C}$) that is measurable in x for every $t \in \mathfrak{T}$ and that is non-constant in (t,x). The transformation r_n is then taken to be the expectation of $k(t,\cdot)$ under the empirical measure:

$$r_n(t) = \frac{1}{n}\sum_{i=1}^{n} k(t, x_i), \qquad \forall t \in \mathfrak{T}.$$

Define $\mathcal{F} = L^2(\mathfrak{T}, \mathfrak{B}_{\mathfrak{T}}, \rho)$ where ρ is a measure on $\mathfrak{T}$ and $\mathfrak{B}_{\mathfrak{T}}$ denotes the Borel σ-field generated by the open sets of $\mathfrak{T}$. The scalar product and norm on $\mathcal{F}$ are defined in the usual way and denoted by $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$, respectively, with the same notation as for the inner product and norm in $\mathcal{E}$. The function $k(t,x)$ defines also a bounded operator $K : \mathcal{E} \to \mathcal{F}$ and must be such that, for every $\varphi \in \mathcal{E}$, $K\varphi \in \mathcal{F}$ and r_n is an H-r.v. with realizations in $\mathcal{F}$. Hence,

$$\begin{array}{rcl} K : \mathcal{E} & \to & \mathcal{F} \\ \varphi & \mapsto & \int k(t,x)\varphi(x)\Pi(dx). \end{array} \tag{2.3}$$

For every $f \in \mathcal{E}_M$, Kf is the expectation of $k(t,\cdot)$ under F: $(Kf)(t) = \mathbf{E}^F(k(t,x))$. Under the true distribution F^* the expectation of r_n is Kf_* and the covariance function of r_n is: $\forall s,t \in \mathfrak{T}$,

$$\frac{1}{n}\sigma(t,s) = \mathbf{E}^* r_n(t) r_n(s) = \frac{1}{n}\left[\mathbf{E}^*\left(k(t,x)k(s,x)\right) - \mathbf{E}^*(k(t,x))\mathbf{E}^*(k(s,x))\right].$$

If the class of functions $\{k(t,\cdot), t \in \mathfrak{T}\}$ is Donsker then, as $n \to \infty$, the conditional distribution of $\sqrt{n}(r_n - Kf_*)$ weakly converges to a $\mathcal{GP}$ with covariance operator $\Sigma : \mathcal{F} \to \mathcal{F}$ defined as

$$\forall \psi \in \mathcal{F}, \qquad (\Sigma\psi)(t) = \int \sigma(t,s)\psi(s)\rho(ds) \tag{2.4}$$

which is one-to-one, linear, positive definite, self-adjoint and trace-class. In the following we assume that $\{k(t,\cdot), t \in \mathfrak{T}\}$ is Donsker such that r_n is approximately Gaussian: $r_n \sim \mathcal{GP}(Kf_*, \Sigma_n)$ where $\Sigma_n = \frac{1}{n}\Sigma$. In our analysis we treat f_* as the realization of the random parameter f and Σ_n as known. Therefore, the sampling distribution of $r_n|f$ is $P^f = \mathcal{GP}(Kf, \Sigma_n)$ and we construct the posterior distribution based on it. In practice, Σ_n must be replaced by its empirical counterpart. In finite sample, P^f is an approximation of the true sampling distribution but the approximation error vanishes as $n \to \infty$.

Example 2.1 (Empirical cumulative distribution function (cdf).)**.** Let $(x_1, \ldots, x_n)$ be an *i.i.d.* sample of $x \in \mathbb{R}$. A possible choice for $k(t,x)$ is $k(t,x) = 1\{x \le t\}$, where $1\{A\}$ denotes the indicator function of the event A. In this case, $r_n(t) = F_n(t) := \frac{1}{n}\sum_{i=1}^n 1\{x_i \le t\}$ is the empirical *cdf* and the operator K is $(K\varphi)(t) = \int_S 1\{s \le t\}\varphi(s)\Pi(ds)$, $\forall \varphi \in \mathcal{E}$. By the Donsker's theorem, $F_n(\cdot)$ is asymptotically Gaussian with mean the true cdf $F_*(\cdot)$ and covariance operator characterized by the kernel: $\frac{1}{n}(F_*(s \wedge t) - F_*(s)F_*(t))$.

Example 2.2 (Empirical characteristic function)**.** Let $(x_1, \ldots, x_n)$ be an *i.i.d.* sample of $x \in \mathbb{R}$. Let $k(t,x) = e^{itx}$, so that $r_n(t) = c_n(t) := \frac{1}{n}\sum_{j=1}^n e^{itx_j}$ is the empirical characteristic function. In this case, the operator K is $(K\varphi)(t) = \int_S e^{its}\varphi(s)\Pi(ds)$, $\forall \varphi \in \mathcal{E}$. By the Donsker's theorem, $c_n(\cdot)$ is asymptotically a Gaussian process with mean the true characteristic function $c(\cdot) \equiv \mathbf{E}^*[e^{itx}]$ and covariance operator characterized by the kernel: $\frac{1}{n}(c(s+t) - c(s)c(t))$.

The following lemma gives an useful characterization of the operator Σ in terms of K and its adjoint K^*. We recall that the adjoint K^* of a bounded and linear operator $K : \mathcal{E} \to \mathcal{F}$ is defined as the operator from $\mathcal{F}$ to $\mathcal{E}$ that satisfies $\langle K\varphi, \psi\rangle = \langle \varphi, K^*\psi\rangle$, $\forall \varphi \in \mathcal{E}$ and $\forall \psi \in \mathcal{F}$. In our case, an elementary computation shows that $(K^*\psi)(t) = \int_{\mathfrak{T}} k(t,x)\psi(t)\rho(dt)$, $\forall \psi \in \mathcal{F}$.

Lemma 2.2. *Let $K : \mathcal{E} \to \mathcal{F}$ be a bounded and linear operator defined as in* (2.3) *and $K^* : \mathcal{F} \to \mathcal{E}$ be its adjoint, that is, $(K^*\psi)(t) = \int_{\mathfrak{T}} k(t,x)\psi(t)\rho(dt)$, $\forall \psi \in \mathcal{F}$. The operator $\Sigma_n = \frac{1}{n}\Sigma$, with $\Sigma : \mathcal{F} \to \mathcal{F}$ defined in* (2.4) *takes the form*

$$\forall \psi \in \mathcal{F}, \quad \Sigma\psi = KM_fK^*\psi - (KM_f1)\langle M_f, K^*\psi\rangle \tag{2.5}$$

where $M_f : \mathcal{E} \to \mathcal{E}$ is the multiplication operator $M_f\varphi = f_*\varphi$, $\forall \varphi \in \mathcal{E}$.

We denote by $\mathfrak{D}$ the subset of $\mathcal{E}$ whose elements integrate to 0 with respect to Π:

$$\mathfrak{D} := \left\{ g \in \mathcal{E};\ \int_S g(x)\Pi(dx) = 0 \right\}.$$

Remark that $\mathfrak{D}$ contains the functions in $\mathcal{E}$ that are the difference of *pdf*s of F with respect to Π. Moreover, $\mathcal{R}(\Omega_{0\theta}^{1/2}) \subset \mathfrak{D}$, where $\mathcal{R}(\cdot)$ denotes the range of an operator, and

$$\mathcal{R}(\Omega_{0\theta}^{1/2}) \subseteq \left\{ \varphi \in \mathcal{E};\ \int_S \varphi(h)h(\theta,x)\Pi(dx) = 0 \text{ and } \int_S \varphi(x)\Pi(dx) = 0 \right\}.$$

Remark that the equality holds when the sequence $(\lambda_j)_{j>d}$, used to construct $\Omega_{0\theta}$, is strictly positive. Let $\Sigma^{1/2} : \mathcal{F} \to \mathcal{F}$ be such that $\Sigma = \Sigma^{1/2}\Sigma^{1/2}$. The following lemma states the relationship between the range of $\Sigma^{1/2}$ and the range of K restricted to $\mathfrak{D}$.

Lemma 2.3. *Let $K : \mathcal{E} \to \mathcal{F}$ be the operator defined in (2.3), and denote by $K|_{\mathfrak{D}}$ the operator K restricted to $\mathfrak{D} \subset \mathcal{E}$. Then, if $K|_{\mathfrak{D}}$ is injective we have*

$$\mathcal{R}(K|_{\mathfrak{D}}) = \mathcal{D}(\Sigma^{-\frac{1}{2}}).$$

2.3 Posterior distribution

The Bayesian model defines a joint distribution on (the Borel σ-field) of Λ and can be summarized in the following way:

$$\begin{aligned}
\theta \sim\ & \mu(\theta) \\
f|\theta \sim\ & \mu(f|\theta) = \mathcal{GP}(f_{0\theta}, \Omega_{0\theta}), \qquad \int h(\theta, x) f_{0\theta}(x)\Pi(dx) = 0 \quad \text{and} \quad \Omega_{0\theta}^{\frac{1}{2}}(1, h(\theta, \cdot)^T)^T = 0 \\
r_n|f, \theta \sim\ & r_n|f \sim P^f = \mathcal{GP}(Kf, \Sigma_n)
\end{aligned} \tag{2.6}$$

where we use the $\mathcal{GP}$ approximation P^f. Theorem 1 in Florens and Simoni [2012] shows that the joint distribution of (f, r_n), conditional on θ, is:

$$\left.\begin{pmatrix} f \\ r_n \end{pmatrix}\right| \theta \sim \mathcal{GP}\left(\begin{pmatrix} f_{0\theta} \\ Kf_{0\theta} \end{pmatrix}, \begin{pmatrix} \Omega_{0\theta} & \Omega_{0\theta}K^* \\ K\Omega_{0\theta} & \Sigma_n + K\Omega_{0\theta}K^* \end{pmatrix}\right) \tag{2.7}$$

where $(\Sigma_n + K\Omega_{0\theta}K^*) : \mathcal{F} \to \mathcal{F}$, $\Omega_{0\theta}K^* : \mathcal{F} \to \mathcal{E}$ and $K\Omega_{0\theta} : \mathcal{E} \to \mathcal{F}$. The marginal sampling distribution of r_n conditional on θ, obtained by integrating out f, is:

$$r_n|\theta \sim P_n^\theta \sim \mathcal{GP}(Kf_{0\theta}, \Sigma_n + K\Omega_{0\theta}K^*). \tag{2.8}$$

We now discuss the posterior distribution, denoted by $\mu(\cdot|r_n)$. Recovering the posterior distribution of f is an ill-posed inverse problem. Since f is a nuisance parameter we discuss in the main text only the posterior distribution of the parameter of interest θ. We postpone to Appendix A the discussion about the conditional posterior distribution $\mu(f|r_n, \theta)$ of f given θ.

2.3.1 Posterior distribution of θ

The marginal posterior for θ, denoted by $\mu(\theta|r_n)$, is obtained by using the marginal sampling distribution P_n^θ given in (2.8). We first have to characterize the likelihood of P_n^θ with respect to an appropriate dominating measure that will be denoted by P_n^0. The following theorem characterizes a probability measure P_n^0 which is equivalent to P_n^θ as well as the corresponding likelihood of P_n^θ with respect to P_n^0. Denote by $(l_{j\theta}, \rho_j(\theta), \psi_j(\theta))_{j\geq 0}$ the singular value decomposition of the operator $\Sigma^{-1/2}K\Omega_{0\theta}^{1/2}$. Remark that by the result in Lemma 2.3 this operator is well defined.

Theorem 2.1. *Let P_n^0 be a Gaussian measure with mean Kf_* and covariance operator $n^{-1}\Sigma$, i.e. $P_n^0 = \mathcal{GP}(Kf_*, n^{-1}\Sigma)$ with Σ defined in (2.4). For n fixed, if $K|_{\mathfrak{D}}$ is injective, then P_n^0 and P_n^θ are equivalent. Moreover, assume that $\forall j \geq 0$ and $\forall \theta \in \Theta$, $\psi_j(\theta) \in \mathcal{R}(\Sigma^{1/2})$. Then, the Radon-Nikodym derivative is given by*

$$\begin{aligned} p_{n\theta}(r_n;\theta) &:= \frac{dP_n^\theta}{dP_n^0}(r_n) \qquad (2.9) \\ &= \prod_{j=0}^{\infty} \sqrt{\frac{n^{-1}}{n^{-1}+l_{j\theta}^2}} \exp\left\{ -\frac{1}{2}\sum_{j=0}^{\infty} \frac{\left(Z_j - \langle \sqrt{n}K(f_{0\theta}-f_*), \Sigma^{-1/2}\psi_j(\theta)\rangle\right)^2}{1+nl_{j\theta}^2} \right\} e^{\{\frac{1}{2}\|Z\|_\Sigma^2\}} \end{aligned}$$

where $Z := \sqrt{n}(r_n - Kf_*)$, $Z_j := \langle Z, \Sigma^{-1/2}\psi_j(\theta)\rangle$ for all $j \geq 0$, and $\|Z\|_\Sigma := \|\Sigma^{-1/2}Z\|$.

The quantity $\|\Sigma^{-1/2}Z\|^2$ is defined as the limit in $\mathcal{F}$ of the series $\sum_{j=0}^m \sigma_j^{-2}\langle Z_j, \phi_j\rangle^2$ as $m \to \infty$ (where $\{\sigma_j^2, \phi_j\}_{j=0}^\infty$ is the eigensystem of Σ). By using (2.9), the (marginal) posterior distribution of θ takes the form (after simplifying the terms that do not depend on θ):

$$\begin{aligned} \mu(\theta|r_n) &= \frac{p_{n\theta}(r_n;\theta)\mu(\theta)}{\int_\Theta p_{n\theta}(r_n;\theta)\mu(\theta)d\theta} \qquad (2.10) \\ &= \frac{\prod_{j=0}^\infty \sqrt{\frac{1}{n^{-1}+l_{j\theta}^2}} \exp\left\{-\frac{1}{2}\sum_{j=0}^\infty \frac{\left(Z_j - \langle\sqrt{n}K(f_{0\theta}-f_*),\Sigma^{-1/2}\psi_j(\theta)\rangle\right)^2}{1+nl_{j\theta}^2}\right\}\mu(\theta)}{\int_\Theta \prod_{j=0}^\infty \sqrt{\frac{1}{n^{-1}+l_{j\theta}^2}} \exp\left\{-\frac{1}{2}\sum_{j=0}^\infty \frac{\left(Z_j - \langle\sqrt{n}K(f_{0\theta}-f_*),\Sigma^{-1/2}\psi_j(\theta)\rangle\right)^2}{1+nl_{j\theta}^2}\right\}\mu(\theta)d\theta} \end{aligned}$$

and can be used to compute a point estimator of θ. We propose to use the maximum a posterior (MAP) estimator θ_n defined as

$$\begin{aligned} \theta_n &= \arg\max_{\theta\in\Theta} \mu(\theta|r_n) \\ &= \arg\max_{\theta\in\Theta} \prod_{j=0}^\infty \sqrt{\frac{1}{n^{-1}+l_{j\theta}^2}} \exp\left\{ -\frac{1}{2}\sum_{j=0}^\infty \frac{\left(Z_j - \langle\sqrt{n}K(f_{0\theta}-f_*), \Sigma^{-1/2}\psi_j(\theta)\rangle\right)^2}{1+nl_{j\theta}^2} \right\} \mu(\theta) \\ &= \arg\max_{\theta\in\Theta} \prod_{j=0}^\infty \sqrt{\frac{1}{n^{-1}+l_{j\theta}^2}} \exp\left\{ -\frac{1}{2}\sum_{j=0}^\infty \frac{\left(\langle\sqrt{n}(r_n - Kf_{0\theta}), \Sigma^{-1/2}\psi_j(\theta)\rangle\right)^2}{1+nl_{j\theta}^2} \right\} \mu(\theta) \qquad (2.11) \end{aligned}$$

or the posterior mean estimator $\mathbf{E}(\theta|r_n) := \int_\Theta \theta\mu(\theta|r_n)d\theta$.

Remark 2.3. We have already discussed in Remark 2.2 the possibility of using a different prior scheme when we are in the just-identified case and θ can be written as a linear functional of f. If one uses this different prior scheme, then given a $\mathcal{GP}$ prior for f as described in Remark 2.2, the posterior distribution for θ is recovered from the $\mathcal{GP}$ posterior of f through the transformation $b(f)$.

2.3.2 Properties of the posterior distribution of θ

Before concluding this section, we show two important results. The first one establishes that expression (2.9) is invariant to the choice of Π and therefore the marginal posterior of θ is invariant to the choice of Π. More precisely the following proposition holds.

Proposition 2.1. *For a positive measure Π_1 on S, let $\mathcal{E}_{\Pi_1} = L^2(S, \mathfrak{B}_S, \Pi_1)$ and $\mathfrak{z} = \frac{d\Pi}{d\Pi_1}$. Let $\varphi : \mathcal{E} \to \mathcal{E}_{\Pi_1}$ be the transformation $\varphi(f) = f\mathfrak{z}$ and Φ be the set of the measurable transformations defined as*

$$\Phi := \left\{\varphi : \mathcal{E} \to \mathcal{E}_{\Pi_1};\ \varphi(f) = f\mathfrak{z},\ \Pi_1 \textit{ is a positive measure and } \sup_{x\in S}\frac{d\Pi_1(x)}{d\Pi(x)} < \infty\right\}.$$

Then, the marginal posterior distribution $\mu(\theta|r_n)$ of θ is Φ-invariant.

This result shows that, once we integrate out the nuisance parameter f, the posterior distribution of θ is not affected by the choice of the dominating measure Π which only causes a transformation of the nuisance parameter. In particular, if $\sup_{x\in S}\frac{dF_*(x)}{d\Pi(x)} < \infty$ then the singular values $l_{j\theta}$s in (2.9) are equal to the $\lambda_j^{1/2}$s used to construct the prior covariance operator $\Omega_{0\theta}$ which simplifies the expression for $\mu(\theta|r_n)$ to:

$$\mu(\theta|r_n) \quad = \quad \frac{\exp\left\{-\frac{1}{2}\sum_{j=0}^\infty \frac{\langle\sqrt{n}(r_n - Kf_{0\theta}), \Sigma^{-1/2}\psi_j(\theta)\rangle^2}{1+n\lambda_j}\right\}\mu(\theta)}{\int_\Theta \exp\left\{-\frac{1}{2}\sum_{j=0}^\infty \frac{\langle\sqrt{n}(r_n - Kf_{0\theta}), \Sigma^{-1/2}\psi_j(\theta)\rangle^2}{1+n\lambda_j}\right\}\mu(\theta)d\theta}. \tag{2.12}$$

Moreover, given the result in Proposition 2.1, to show properties of $\mu(\theta|r_n)$ we may use a positive measure different from Π as long as the induced transformation belongs to Φ.

The second result we are going to show[2] establishes a link between our Bayesian procedure, GEL estimators with quadratic criterion and the continuous updating GMM estimator. This relationship, given in Theorem 2.2 below, holds when the $\mathcal{GP}$ prior for $f|\theta$ is allowed to become diffuse. More precisely, let us rescale the prior covariance operator of

[2]We thank Yuichi Kitamura for having suggested this research question.

$f|\theta$ by a positive scalar c so that the prior of $f|\theta$ may be written

$$\mu(f|\theta,c) \sim \mathcal{GP}(f_{0\theta}, c\Omega_{0\theta}), \qquad \int h(\theta,x)f_{0\theta}(x)\Pi(dx) = 0, \quad \Omega_{0\theta}^{1/2}(1, h(\theta,\cdot)^T)^T = 0, \quad c \in \mathbb{R}_+.$$

Theorem 2.2. *Assume that* $\sup_{x\in S}\frac{dF_*(x)}{d\Pi(x)} < \infty$, $h_j(\theta,x) \in \mathcal{R}(K^*)$, $\forall j = 1,\dots,d$ *and* $\forall\theta \in \Theta$, *and that* $\mathbf{E}^*[h(\theta,x_i)h(\theta,x_i)^T]$ *is nonsingular* $\forall\theta\in\Theta$. *Let* $\mu(f|\theta,c) \sim \mathcal{GP}(f_{0\theta}, c\Omega_{0\theta})$, *with* $f_{0\theta}$ *and* $\Omega_{0\theta}$ *satisfying Restrictions 1 and 2, and* $c \in \mathbb{R}_+$. *Let* $\mu(\theta|r_n,c)$ *denote the (marginal) posterior of* θ *obtained by integrating out* f *from* P^f *with respect to* $\mu(f|\theta,c)$. *Then,*

$$\lim_{c\to\infty}\mu(\theta|r_n,c) \propto \exp\left\{-\frac{1}{2}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n h(\theta,x_i)\right)^T V_n(\theta)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n h(\theta,x_i)\right)\right\}\mu(\theta)$$

where $V_n(\theta) = \frac{1}{n}\sum_{i=1}^n h(\theta,x_i)h(\theta,x_i)^T$.

Remarks that in the theorem the limit $c\to\infty$ is taken after f has been marginalized out. The result in the theorem deserves some comments. First, it shows that, as the (conditional) prior on f becomes more and more diffuse, our marginal likelihood becomes the quasi-likelihood function (also called limited information likelihood in the literature) that has been used often in the literature, for instance by Chernozhukov and Hong [2003] and Kim [2002]. Therefore, the quasi-likelihood naturally arises from a nonparametric Bayesian procedure, which places a Gaussian Process prior on the set of probability density functions, as the nonparametric prior becomes noninformative.

Second, Theorem 2.2 shows that, as the prior on f becomes noninformative, the MAP objective function is the same (up to constants) as the GEL objective function with quadratic criterion, see the proof of Theorem 2.1 in Newey and Smith [2004]. Moreover, as it can be deduced from Newey and Smith [2004, Theorem 2.1], the MAP objective function becomes a monotonic transformation of the continuous updating GMM objective function.

3 Asymptotic Analysis

In this section we focus on the frequentist asymptotic properties of our approach for $n\to\infty$. For this analysis we use the true probability measure P^* which corresponds to the true DGP F_*. We analyze three issues: *(i)* frequentist consistency of the MAP estimator θ_n (Theorem 3.1), *(ii)* consistency of the posterior of θ (Theorem 3.2), *(iii)* convergence in Total Variation distance of $\mu(\theta|r_n)$ towards a normal distribution (section 3.2). In the following, for every $\tilde{\theta}\in\Theta$ and $\delta > 0$ we denote by $B(\tilde{\theta},\delta)$ the closed ball centered in $\tilde{\theta}$ with radius δ, that is, $B(\tilde{\theta},\delta) = \{\theta\in\Theta;\ \|\theta-\tilde{\theta}\|\le\delta\}$, where here $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^p$. Moreover, denote $\delta_n = n^{-1/2}$ and

$$l_n(\theta) = \sum_{j=0}^{\infty} \log \sqrt{\frac{1}{n^{-1} + l_{j\theta}^2}} - \frac{1}{2} \sum_{j=0}^{\infty} \left[\frac{\left(Z_j - \langle \sqrt{n} K(f_{0\theta} - f_*), \Sigma^{-1/2} \psi_j(\theta) \rangle \right)^2}{1 + n l_{j\theta}^2} \right].$$

3.1 Frequentist Consistency

In this section we first establish frequentist consistency of the MAP estimator θ_n in Theorem 3.1. For this, we need the following assumptions.

A1. The true parameter θ_* belongs to the interior of a compact convex subset Θ of $\mathbb{R}^d$ and is the unique solution of $E^*[h(\theta, x)] = 0$.

A2. The singular functions $\{\psi_j(\theta), \rho_j(\theta)\}$ and singular values $\{l_{j\theta}\}$ are continuous functions of θ.

A3. The prior mean function $f_{0\theta}$ is continuous in θ.

A4. At least one of the following holds: (i) the eigenvalues $\{l_{j\theta}^2\}$ do not depend on θ and the prior $\mu(\theta)$ is flat or (ii) the eigenvalues $\{l_{j\theta}^2\}$ do depend on θ and $\mu(\theta)$ is chosen such that $\prod_{j=0}^{\infty} \frac{1}{\sqrt{n^{-1} + l_{j\theta}^2}} \mu(\theta) \to 1$ as $n \to \infty$.

Assumption A1 is a standard assumption in the literature on moment estimation. Assumptions A2 and A3 can be easily satisfied since $f_{0\theta}$ and the operators $\Omega_{0\theta}$, K and Σ are chosen by the econometrician. Assumption A4 (ii) is verified for instance if we set $\mu(\theta) \propto \prod_{j=0}^{\infty} l_{j\theta}$.

Theorem 3.1. *Under Assumptions A1-A4:*

$$\theta_n \xrightarrow{p} \theta_*$$

in P^*-probability as $n \to \infty$.

The second result of this section establishes consistency of the posterior distribution of θ. For that, we introduce the following assumptions:

B1. There exists a constant $C > 0$ such that for any sequence $M_n \to \infty$,

$$P^* \left(\sup_{\theta \in B(\theta_*, \delta_n M_n)^c} [l_n(\theta) - l_n(\theta_*)] \leq -C M_n^2 \right) \to 1 \qquad \text{as } n \to \infty.$$

B2. There exists a constant $C > 0$ such that for any sequence $M_n \to \infty$,

$$P^* \left(\int_{\Theta} e^{l_n(\theta) - l_n(\theta_*)} \mu(\theta) d\theta \leq e^{-C M_n^2 / 2} \right) \to 0 \qquad \text{as } n \to \infty.$$

Assumption B1 is a standard identifiability condition that controls the behavior of the likelihood at a distance from θ_*, see *e.g.* Lehmann and Casella [1998, Condition (B.3) of Theorem 6.8.2] and Bickel and Kleijn [2012, Lemma 6.1]. Assumption B2 is satisfied if $l_n(\theta)$ is continuous in a suitable neighborhood of θ_* and the prior assigns enough mass to this neighborhood. Lemma D.1 in Appendix D provides primitive conditions for Assumption B2. The next theorem gives concentration of the posterior distribution around θ_* and around θ_n.

Theorem 3.2. *Let Assumptions B1-B2 be satisfied, then for any prior $\mu(\theta)$ thick at θ_* and any sequence $M_n \to \infty$,*

$$\mu\left(\sqrt{n}\|\theta - \theta_*\| > M_n | r_n\right) \to 0 \tag{3.1}$$

in P^*-probability as $n \to 0$, for any $M_n \to \infty$. Moreover, under the assumptions of Theorem 3.1

$$\mu\left(\sqrt{n}\|\theta - \theta_n\| > M_n | r_n\right) \to 0 \tag{3.2}$$

in P^*-probability as $n \to 0$, for any $M_n \to \infty$.

3.2 Asymptotic Normality

In this section we first establish asymptotic normality of $\mu(\theta|r_n)$ for the Bayesian model described in (2.6). We refer to it as the overidentified case to stress that this result applies to the case $d > p$ (which is our main interest), but of course it applies also to the just-identified case. Then, in section 3.2.2 we establish asymptotic normality of $\mu(\theta|r_n)$ for the Bayesian model described in Remarks 2.2 and 2.3 where the prior for θ is deduced from the prior for f.

3.2.1 Convergence in Total Variation: the overidentified case

For some $\tau \in \mathbb{R}^p$ let

$$s_n(\tau) = p_{n,\theta_*+\delta_n\tau}(r_n; \theta_* + \delta_n\tau).$$

We assume that there exist a random vector $\widetilde{\ell}_*$ and a nonsingular matrix $\widetilde{I}_*^{-1}$ (that depend on the true θ_* and f_*) such that the sequence $\widetilde{\ell}_*$ is bounded in probability, and satisfy

$$\log\frac{s_n(\tau)}{s_n(0)} = \frac{1}{\sqrt{n}}\tau^T\widetilde{I}_*\widetilde{\ell}_* - \frac{1}{2}\tau^T\widetilde{I}_*\tau + o_p(1) \tag{3.3}$$

for every random sequence τ which is bounded in P^*-probability. Condition (3.3) is known as the integral local asymptotic normality assumption which is used to prove asymptotic normality of semiparametric Bayes procedures, see e.g. Bickel and Kleijn [2012]. In Appendix D.3 we prove that, if $\sup_{x\in S}\frac{dF_*(x)}{d\Pi(x)}<\infty$, then equation (3.3) holds with

$$\widetilde{I}_* = -\mathbf{E}^*\left[\frac{\partial h(\theta_*,x)}{\partial\theta}\right]\left[\mathbf{E}^* h(\theta_*,x)h(\theta_*,x)^T\right]^{-1}\mathbf{E}^*\left[\frac{\partial h(\theta_*,x)}{\partial\theta^T}\right]$$

if $\left[\mathbf{E}^* h(\theta_*,x)h(\theta_*,x)^T\right]$ is nonsingular. For two probability measures P_1 and P_2 absolutely continuous with respect to a positive measure Q, define the total variation (TV) distance as

$$||P_1 - P_2||_{TV} = \frac{1}{2}\int |f_1 - f_2| dQ$$

where f_1 and f_2 are the Radon-Nikodym derivatives of P_1 and P_2, respectively, with respect to Q. The following theorem shows that under (3.3) the posterior distribution of $\sqrt{n}(\theta-\theta_*)$ converges in the TV distance to a Normal distribution with mean $\Delta_* := \frac{1}{\sqrt{n}}\widetilde{\ell}_*$ and variance $\widetilde{I}_*^{-1}$.

Theorem 3.3. *Assume that A1-A3,* (3.1) *and* (3.3) *hold and that the prior* $\mu(\theta)$ *puts enough mass in a neighborhood of* θ_**. If* $\mu(\sqrt{n}(\theta-\theta_*)|r_n)$ *denotes the posterior of* $\sqrt{n}(\theta-\theta_*)$*, then:*

$$\|\mu(\sqrt{n}(\theta-\theta_*)|r_n) - \mathcal{N}(\Delta_*, \widetilde{I}_*^{-1})\|_{TV} \to 0 \tag{3.4}$$

in P^**-probability as* $n\to\infty$*.*

3.2.2 Convergence in Total Variation for linear functionals: the just-identified case

In this section we consider the just-identified case where: $d = p$, the moment restriction (1.1) can be solved explicitly for θ, that is, $\theta = b(f)$, and $b : \mathcal{E} \to \mathbb{R}^p$ is a bounded linear functional. Denote by $\mathcal{E}^p$ the cartesian product $\Pi_{i=1}^p \mathcal{E}$. Hence, by the Riesz theorem, there exists a unique $g \in \mathcal{E}^p$ such that:

$$\theta = b(f) = \int_S g(x)f(x)\Pi(dx), \qquad \forall f \in \mathcal{E}.$$

If θ can take any value in $\mathbb{R}^p$, then the prior distribution of θ can be deduced from the $\mathcal{GP}$ prior of f as described in Remark 2.2: $\theta \sim \mu(\theta) = \mathcal{N}(\langle f_0, g\rangle, \langle\Omega_0 g, g\rangle)$ with $\Omega_0^{1/2}1 = 0$ and all the eigenvalues of Ω_0 but the first one are different from 0. In this section we consider this type of prior. The posterior distribution of θ is then given by

$$\begin{aligned}&\mu(\theta|r_n) = \mathcal{N}\left(\theta^{r_n}, \Omega_n\right),\\ &\text{where } \theta^{r_n} = \langle f_0 + A(r_n - Kf_0), g\rangle, \quad \text{and} \quad \Omega_n = \langle(\Omega_0 - AK\Omega_0)g, g\rangle\end{aligned} \tag{3.5}$$

and $A : \mathcal{F} \to \mathcal{E}$ is a continuous and linear operator whose expression is given in Lemma A.1 in the Appendix. In the following, we implicitly assume that the conditions of Lemma A.1 are satisfied. When this is not the case, then the asymptotic result of Theorem 3.4 below is still valid, under minor modifications, if we replace the exact posterior $\mu(f|\theta, r_n)$ with the *regularized posterior distribution* discussed in Remark A.3 in the Appendix and introduced by Florens and Simoni [2012].

By using the usual notation for empirical processes, we denote by $\hat{\theta} = b(\mathbb{P}_n) := n^{-1}\sum_{i=1}^n g(x_i)$ the method of moments estimator and by V the variance of $\sqrt{n}b(\mathbb{P}_n)$ under F_*. The efficient influence function $\tilde{g} : \mathcal{S} \to \mathbb{R}^p$ takes the form $\widetilde{g} = g - \mathbf{E}^* g$ and then $V = \mathbf{E}^*(\widetilde{g}\widetilde{g}^T)$. The next theorem states that the TV distance between $\mu\left(\sqrt{n}(\theta - \hat{\theta})\middle| r_n\right)$ and $\mathcal{N}(0, V)$ converges to 0 in probability.

Theorem 3.4. *Let $\theta = b(f) = \langle g, f\rangle$, $\hat{\theta} = b(\mathbb{P}_n) := n^{-1}\sum_{i=1}^n g(x_i)$ and consider the Gaussian model* (2.7) *without θ: $r_n|f \sim \mathcal{GP}(Kf, \Sigma_n + K\Omega_0K^*)$ and $f \sim \mathcal{GP}(f_0, \Omega_0)$ where f_0 is a pdf, $\Omega_0^{1/2}1 = 0$ and all the eigenvalues of Ω_0 but the first one are different from* 0. *If $f_*^{-1/2} \in \mathcal{R}(K^*)$, V is nonsingular and $g \in \mathcal{C}^\infty$, then*

$$\left\|\mu\left(\sqrt{n}(\theta - \hat{\theta})\middle| r_n\right) - \mathcal{N}(0, V)\right\|_{TV} \to 0$$

in P^*-probability as $n \to \infty$.

The result of this theorem, while similar to the result of Theorem 3.3, is obtained by using a proof different from the one used to obtain Theorem 3.3 and that works only in $d = p$ case.

4 The case with $span\{1, h_1(\theta,\cdot), \ldots, h_d(\theta,\cdot)\}$ independent of θ

In this section, we consider the particular case where the space spanned by $\{1, h_1(\theta,\cdot), \ldots, h_d(\theta,\cdot)\}$, namely $\mathfrak{N}(\Omega_{0\theta})$, does not depend on θ. This arises for instance when the moment functions $\{h_j(\theta, x)\}_{j=1}^d$ are separable in θ and x. In this case, one can choose any orthonormal basis (o.n.b.) with respect to Π that spans $\mathfrak{N}(\Omega_{0\theta})$ and that does not depend on θ. Denote this basis by $\{\varphi_j\}_{j=0}^d$, where we assume that $\mathfrak{N}(\Omega_{0\theta})$ has dimension $d+1$. The orthogonal space $\mathfrak{N}(\Omega_{0\theta})^\perp$ is also independent of θ and is spanned by an o.n.b. $\{\varphi_j\}_{j>d}$ that is independent of θ as well. Thus, the prior covariance operator $\Omega_{0\theta}$ does not depend on θ and writes:

$$\forall\phi \in \mathcal{E}, \quad \Omega_{0\theta}\phi = \Omega_0\phi = \sum_{j>d}\lambda_j\langle\phi, \varphi_j\rangle\varphi_j.$$

On the other hand, the prior mean function $f_{0\theta}$ does depend on θ. An example where $span\{1, h_1(\theta,\cdot),\ldots,h_d(\theta,\cdot)\}$ does not depend on θ is the case where $h(\theta,x)$, after normalization, is of the form: $h(\theta,x) = a(x) - b(\theta)$ for some vector-valued functions $a(x) = (a_1(x),\ldots,a_d(x))^T$ and $b(\theta) = (b_1(\theta),\ldots,b_d(\theta))^T$.

Let $\{\psi_j\}_{j\geq 0}$ be an o.n.b. in $\mathcal{F}$ and $\{\lambda_{jK}\}_{j\geq 0}$ be a square-summable sequence of positive real numbers. We can then construct the operator K and the transformation r_n as:

$$\begin{aligned}
\forall\phi\in\mathcal{E},\quad (K\phi)(t) &= \sum_{j=0}^{\infty}\lambda_{jK}\langle\phi,\varphi_j\rangle\psi_j(t) = \int\sum_{j=0}^{\infty}\lambda_{jK}\phi(x)\varphi_j(x)\psi_j(t)\Pi(dx)\\
r_n &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=0}^{\infty}\lambda_{jK}\varphi_j(x_i)\psi_j(t).
\end{aligned}$$

Hence, the kernel $k(x,t)$ characterizing the operator K writes: $k(x,t) = \sum_{j=0}^{\infty}\lambda_{jK}\varphi_j(x)\psi_j(t)$. Remark that this describes a Donsker class if, for instance, $\sum_{j=0}^{\infty}\lambda_{jK}^2 \leq 1$, see van der Vaart and Wellner [1996, Theorem 2.13.2]. The adjoint of K, denoted by K^*, writes: $\forall\phi\in\mathcal{F}$, $(K^*\phi)(x) = \sum_{j=0}^{\infty}\lambda_{jK}\langle\phi,\psi_j\rangle\varphi_j(x)$. Remark that K, K^*, r_n and Σ do not depend on θ.

By Proposition 2.1, our inference procedure is invariant to the choice of Π. Then, if $\sup_{x\in S} f_*(x) < \infty$ we can fix $\Pi = F_*$ so that $\mathcal{E} = L^2(S,\mathfrak{B}_S,F_*)$ and $f_* = 1$. Therefore, $\{\varphi_j\}_{j\geq 0}$ is an o.n.b. with respect to F_* and $Var(h(\theta,x)) = I_d$, where I_d denotes the d-dimensional identity matrix.

The operator Σ has eigenfunction $\{\psi_j\}_{j\geq 0}$ and eigenvalues $\{0,\lambda_{jK}^2, j\geq 1\}$, that is: $\Sigma\psi_j = \lambda_{jK}^2\psi_j$ for $j\geq 1$ and $\Sigma\psi_0 = 0$. To see this, let us write Σ in the form given in Lemma 2.2: $\Sigma\cdot = KM_fK^*\cdot - (KM_f1)\langle M_f, K^*\cdot\rangle$, and since $f_* = 1$ we have:

$$\begin{aligned}
\forall j\neq 0,\quad \Sigma\psi_j = KK^*\psi_j - (K1)\langle K1,\psi_j\rangle &= \lambda_{jK}K\varphi_j - \lambda_{0K}^2\psi_0\langle\psi_0,\psi_j\rangle\\
&= \lambda_{jK}^2\psi_j\\
\Sigma\psi_0 = \lambda_{0K}^2\psi_0 - (\lambda_{0K}\psi_0)\langle K1,\psi_0\rangle &= \lambda_{0K}^2 - \lambda_{0K}^2 = 0.
\end{aligned}$$

This result is obtained by using the fact that $\langle K1,\psi_j\rangle = \lambda_{0K}\langle\psi_0,\psi_j\rangle = 0$ for $j\geq 1$ and $\langle\psi_0,\psi_0\rangle = 1$.

Remark 4.1. The intuition behind the fact that Σ has an eigenvalue equal to 0 comes from Lemma 2.3. We know from this lemma that when $K|_{\mathfrak{D}}$ is injective, $\mathcal{R}(\Sigma^{1/2})$ is equal to the range of the restriction of K to $\mathfrak{D}$, that is $\mathcal{R}(K|_{\mathfrak{D}})$. Since φ_0 is orthogonal to $\mathfrak{D}$ because $\varphi_0 = 1$, the eigenfunction ψ_0, which is the transformation of φ_0 through K, does not belong to $\mathcal{R}(K|_{\mathfrak{D}})$ and so neither to $\mathcal{R}(\Sigma)$ (since $\mathcal{R}(\Sigma)\subset\mathcal{R}(\Sigma^{1/2})$). Therefore, it must be that the eigenvalue corresponding to ψ_j be zero because $\overline{\mathcal{R}(\Sigma)}$ is spanned by the eigenfunctions corresponding to the nonzero eigenvalues, that is $\{\psi_j\}_{j\geq 1}$. Finally, because

Σ has one eigenvalue equal to 0 it is not injective.

Trivial computations show that, in this particular case, the eigenvalues $l_{j\theta}^2$ and eigenfunctions $\psi_j(\theta)$ in (2.9) are as follows:

$$l_{j\theta}^2 = \begin{cases} 0 & \text{for } j = 0, \dots, d \\ \lambda_j & \text{for } j > d \end{cases} \qquad \text{and} \qquad \psi_j(\theta) = \psi_j \quad \text{for } j = 0, 1, \dots. \tag{4.1}$$

It follows that the likelihood in (2.9) can be simplified and the MAP writes as:

$$\begin{aligned}
\theta_n &= \arg\max_{\theta\in\Theta} \mu(\theta|r_n) = \arg\max_{\theta\in\Theta} \left(\log p_{n\theta}(r_n;\theta) + \log\mu(\theta)\right) \\
&= \arg\max_{\theta\in\Theta} \Big(-\sum_{j=1}^{d} \frac{1}{\lambda_{jK}^2}\langle\sqrt{n}(r_n - Kf_{0\theta}), \psi_j\rangle^2 - \sum_{j>d}\frac{1}{\lambda_{jK}^2}\langle\sqrt{n}(r_n - Kf_{0\theta}), \psi_j\rangle^2 \frac{1}{1+n\lambda_j} \\
&\qquad - \sum_{j>d}\log(n\lambda_j + 1) + 2\log\mu(\theta)\Big)\frac{1}{2} \\
&= \arg\min_{\theta\in\Theta}\left(\sum_{j=1}^{d}\frac{n}{\lambda_{jK}^2}\langle r_n - Kf_{0\theta}, \psi_j\rangle^2 + \sum_{j>d}\frac{n}{\lambda_{jK}^2}\langle r_n - Kf_{0\theta}, \psi_j\rangle^2\frac{1}{1+n\lambda_j} - 2\log\mu(\theta)\right) \\
&= \arg\min_{\theta\in\Theta}\Big(\sum_{j=1}^{d}\left(\frac{1}{n}\sum_{i=1}^{n}\varphi_j(x_i) - \langle f_{0\theta}, \varphi_j\rangle\right)^2 \\
&\qquad + \sum_{j>d}\left(\frac{1}{n}\sum_{i=1}^{n}\varphi_j(x_i) - \langle f_{0\theta}, \varphi_j\rangle\right)^2\frac{1}{1+n\lambda_j} - \log\mu(\theta)\Big)
\end{aligned} \tag{4.2}$$

where we have eliminated the terms that do not depend on θ and we have used the fact that $\frac{1}{\lambda_{jK}^2}\langle r_n - Kf_{0\theta}, \psi_j\rangle^2 = \left(\frac{1}{n}\sum_{i=1}^{n}\varphi_j(x_i) - \langle f_{0\theta}, \varphi_j\rangle\right)^2$. According to Assumption A4, in the particular case considered in this section the prior can be chosen independent of θ. Equation (4.2) is quite useful and allows to emphasize several aspects of our methodology.

I. The first term in (4.2) accounts for the moment restrictions. Minimization of this term corresponds to the classical GMM. In fact, by construction $\int\varphi_j(x)f_{0\theta}(x)\Pi(dx)$ is not 0 because we are using transformations of the moment functions. Thus, $\left[\frac{1}{n}\sum_{i=1}^{n}\varphi_j(x_i) - \int\varphi_j(x)f_{0\theta}(x)\Pi(dx)\right]^2$ depends on θ through $f_{0\theta}$. Remark that if we do the inverse transformation from $\{\varphi_j\}_{j=1}^{d}$ to $\{h_j(x,\theta)\}_{j=1}^{d}$ then the term involving $f_{0\theta}$ will be zero and the term $\varphi_j(x_i)$ will be written in terms of θ. For instance, in the separable case where $h_j(\theta, x) = a_j(x) - b_j(\theta)$: $\int\varphi_j(x)f_{0\theta}(x)\Pi(dx) = b_j(\theta)$ and so $\left[\frac{1}{n}\sum_{i=1}^{n}\varphi_j(x_i) - \int\varphi_j(x)f_{0\theta}(x)\Pi(dx)\right]^2 = \left[\frac{1}{n}\sum_{i=1}^{n}h_j(\theta, x_i)\right]^2$.

II. The second term in (4.2) accounts for the extra information that we have, namely, the information contained in the subspace of $\mathcal{E}$ orthogonal to $span\{1, h_1(\theta,\cdot),\ldots,h_d(\theta,\cdot)\}$. This information, which is in general not exploited by the classical GMM estimation, can be exploited thanks to the prior distribution and the prior mean $f_{0\theta}$ if the prior is not fixed but varies with n (see comment III below). On the contrary, if the prior is fixed then, as $n\to\infty$, the second term of (4.2) converges to 0 since $n^{-1}\sum_{i=1}^{n}\varphi_j(x_i)\to E^*[\varphi_j(X)]$ a.s. and $E^*[\varphi_j(X)]=0$ because φ_j is orthogonal to 1 for $j>d$ and since $(1+n\lambda_j)^{-1}\to 0$.

III. Expression (4.2) makes an explicit connection between the parametric case (infinite number of moment restrictions) and the semiparametric case (when only the first d moment conditions hold). The semiparametric case corresponds to the classical GMM approach while the parametric case corresponds to the maximum likelihood estimator (MLE). Indeed, the prior distribution for f specifies a parametric model for $f_{0\theta}$ which satisfies the d moment restrictions and eventually other “extra” moment restrictions. The eigenvalues λ_j of the prior covariance operator play the role of weights of the “extra” moment restrictions and represent our “beliefs” concerning these restrictions. When we are very confident about these “extra” conditions, or equivalently we believe that $f_{0\theta}$ is close to f_*, then the λ_js are close to zero or converge to 0 faster than n^{-1} as $n\to\infty$. So, the prior distribution for f is degenerate on $f_{0\theta}$ (as n increases) when the parametric model is the true one. In that case, the MAP estimator will essentially be equivalent to the MLE that we would obtain if we use the prior mean function $f_{0\theta}$ as the likelihood. When we are very uncertain about $f_{0\theta}$ then the λ_js are very large and may tend to $+\infty$ (uninformative prior). In this case the MAP estimator will be close to the GMM estimator (up to a prior on θ).

4.1 Testing and moment selection procedures

Remark III in section 4 is important if one is interested in constructing testing procedures or doing moment selection. We are not going to develop a formal test/selection procedure here as this will make the object of a separated paper, but we would like to point out that our procedure suggests an easy way to test a parametric model against a semiparametric one characterized by a finite number of moment restrictions. We can deal with the two following situations:

1. We know that the distribution of the data satisfies d moment restrictions and we want to test that it has a particular parametric form. In this case, for a given *pdf* $g\in\mathcal{E}_M$ such that $\int h(\theta,x)g(x)\Pi(dx)=0$ for a known vector of functions $h(\theta,x)$, the null hypothesis is $H_0: f_*=g$. An example is the univariate linear regression model:

$Y = Z\theta + \varepsilon$, where f_* is the true joint *pdf* of $X := (Y, Z)^T$ and $\mathbf{E}^*(Y|Z) = Z\theta$. We may want to test that f_* belongs to a particular parametric class.

2. There are d moment restrictions of which we are sure and we want to test the validity of the other moment restrictions. The null hypothesis writes $H_0 : \mathbf{E}^*(h_j(\theta, x)) = 0$ for some $j > d$.

To treat the first situation, we have to specify $f_{0\theta} = g$. Then, for both the situations, the natural approach would be to treat the λ_js corresponding to the extra conditions (namely, the λ_j for $j > d$) as hyperparameters for which a prior distribution is specified. The null hypothesis, in both the cases above, writes as $H_0 : \lambda_j = 0$ for all (or for some) $j > d$. Then, the posterior distribution of λ_j may be used to draw a conclusion on the test: either by considering posterior odds ratio or by constructing encompassing tests.

To construct a prior for the λ_js let us write: $\lambda_j = c\rho_j$ where $c = tr\Omega_0$ and $\sum_{j=0}^{\infty} \rho_j = 1$. We propose two alternatives priors.

Dirichlet prior. Suppose that we want to test the nullity of some λ_js, say λ_j for $d < j < J < \infty$. Then we specify a Dirichlet prior for $(\rho_{d+1}, \ldots, \rho_{J-1})$:

$$\mu_\rho(\rho_{d+1}, \ldots, \rho_{J-1}|\nu) \propto \prod_{j=d+1}^{J-1} \rho_j^{\nu_j - 1} \left(1 - \sum_{j=d+1}^{J-1} \rho_j\right)^{\nu_J - 1} \prod_{j=d+1}^{J-1} I(\rho_j \geq 0) I\left(\sum_{j=1}^{J-1} \rho_j \leq 1\right)$$

where $\nu = (\nu_{d+1}, \ldots, \nu_J)$.

Prior on $c > 0$. Suppose that we want to test that all the moment restrictions are true (that is, test of a parametric model against a semiparametric one). Thus, the null hypothesis is $H_0 : c = 0$. Remark that the $\{\lambda_j\}_{j=1}^d$ corresponding to the first d moment restrictions do not affect the trace of Ω_0 since they are equal to 0. A prior for c will be any distribution with support contained in the positive real semi-axis, for example an inverse gamma distribution.

5 Implementation

In this section we show, through the illustration of several examples, how our method can be implemented in practice. We start with a toy example. The interest in using a $\mathcal{GP}$ prior will be made evident in the more complicated examples where there are overidentifying restrictions which we show can be easily dealt with by using Gaussian priors.

5.1 Just identification and prior on θ through $\mu(f)$

Let the parameter θ of interest be the population mean with respect to f, that is, $\theta = \int x f(x)dx$ and $h(\theta, x) = (\theta - x)$. This example considers the just identified case where θ is a linear functional of f that can take every value in $\mathbb{R}$ and the prior of θ is deduced from the prior of f, denoted by $\mu(f)$. The prior $\mu(f)$ is a $\mathcal{GP}$ which is unrestricted except for the fact that it must generate trajectories that integrate to 1 a.s., namely, $\mu(f) \sim \mathcal{GP}(f_0, \Omega_0)$ where f_0 is a *pdf* and Ω_0 is such that $\Omega_0^{1/2}1 = 0$. Therefore, the prior distribution of θ is Gaussian with mean $\langle f_0, \iota\rangle$ and variance $\langle \Omega_0\iota, \iota\rangle$. The posterior distribution of θ is

$$\theta|r_n \sim \mathcal{N}(\langle f_0, \iota\rangle + \langle \Omega_0 K^* C_n^{-1}(r_n - Kf_0), \iota\rangle, \langle [\Omega_0 - \Omega_0 K^* C_n^{-1} K \Omega_0]\iota, \iota\rangle)$$

where $C_n^{-1} = \left(n^{-1}\Sigma + K\Omega_0 K^*\right)^{-1}$ and ι denotes the identity functional, that is, $\iota(x) = x$.

We illustrate now how to construct in practice the covariance operator Ω_0 in this case where the support of F_* is $\mathbb{R}$, so that θ can take every value in $\mathbb{R}$. Let $S = \mathbb{R}$; the Hermite polynomials $\{H_j\}_{j\geq 0}$ form an orthogonal basis of $L^2(\mathbb{R}, \mathfrak{B}, \Pi)$ for $d\Pi(x) = e^{-x^2/2}dx$ and can be used to construct the eigenfunctions of Ω_0. The first few Hermite polynomials are $\{1, x, x^2 - 1, (x^3 - 3x), \ldots\}$ and an important property of these polynomials is that they are orthogonal with respect to Π: $\int_{\mathbb{R}} H_l(x)H_j(x)e^{-x^2/2}dx = \sqrt{2\pi}n!\delta_{lj}$, where δ_{lj} is equal to 1 if $l = j$ and to 0 otherwise. The operator Ω_0 is constructed as

$$\Omega_0 \cdot \quad = \quad \sigma_0 \sum_{j=0}^{\infty} \lambda_j \frac{1}{\sqrt{2\pi}n!}\langle H_j, \cdot\rangle H_j$$

where $H_{j+1}(x) = xH_j(x) - jH_{j-1}(x)$, $\lambda_0 = 0$ and $\{\lambda_j,\ j \geq 1\} = \{a^j,\ j \geq 1\}$ with $a < 1$.

In our simulation exercise we generate $n = 1000$ *i.i.d.* observations $(x_1, \ldots, x_n)$ from a $\mathcal{N}(1,1)$ distribution and construct the function $r_n = n^{-1}\sum_{i=1}^n e^{tx_i}$ as the empirical Laplace transform. Therefore, $f_*(x) = \frac{1}{\sqrt{2\pi}}e^{-(1-2x)/2}$ and $\theta_* = 1$. We set $\mathfrak{T} = \mathbb{R}$ and $\rho = \Pi$. Thus, the operators K and K^* take the form

$$\forall \phi \in \mathcal{E},\ \ K\phi = \int_{\mathbb{R}} e^{tx}\phi(x)e^{-x^2/2}dx \quad \text{and } \forall \psi \in \mathcal{F}, \qquad K^*\psi = \int_{\mathbb{R}} e^{tx}\psi(t)e^{-t^2/2}dt.$$

The prior mean function f_0 is set equal to a $\mathcal{N}(\varrho, 1)$ distribution. We show in Figure 1 the prior and posterior distribution of θ. We also show the prior mean (magenta asterisk), the posterior mean (blue asterisk) and the MAP (red asterisk) of θ. The posterior mean of θ is computed by discretizing the inner product $\langle \mathbf{E}(f|r_n), \iota\rangle$. The pictures are obtained for $n = 1000$, $f_0(x) = \frac{1}{\sqrt{2\pi}}e^{-(\varrho^2 - 2\varrho x)/2}$, $\varrho = 2$, $a = 0.3$ and $\sigma_0 = 1$. The number of discretization points, used to approximate the integrals, is equal to 1000 for all the simulation schemes.

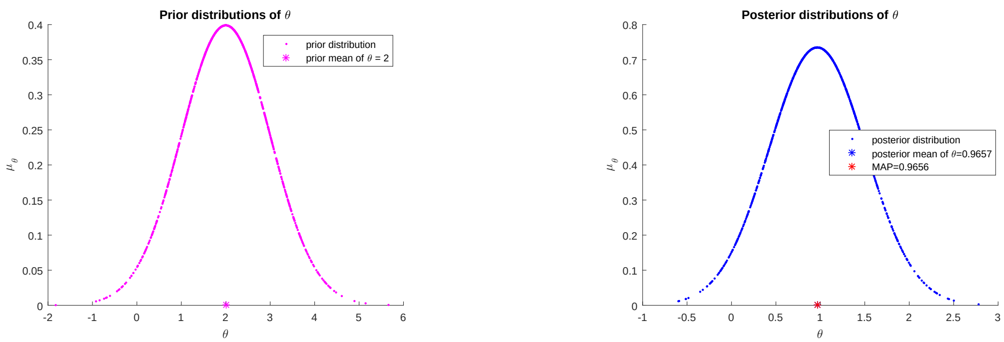


Figure 1: Prior and posterior distributions and means of θ. The true value of θ is $\theta_* = 1$.

5.2 Just identification and prior on θ

We consider the same framework as in the previous example where the parameter θ of interest is the population mean: $\theta = \int x f(x)dx$ and $h(\theta, x) = (\theta - x)$, but now we specify a joint proper prior distribution on (θ, f). We specify a marginal prior $\mu(\theta)$ on θ and a conditional prior on f given θ. While the first one can be arbitrarily chosen, the latter is specified as a $\mathcal{GP}$ constrained to generate functions that integrate to 1 and that have mean equal to θ a.s., as described in section 2.1.

Compared to the approach in section 5.1, this approach allows to easily incorporate any prior information that one may have about θ. In fact, incorporating the information on θ through the prior distribution of f is complicated while to incorporate such an information directly in the prior distribution of θ results to be very simple. In particular, the approach of this section works even when θ takes values only in a compact subset of $\mathbb{R}^p$, while the approach of section 5.1 does not work in this case.

Let us suppose that $m = 1$, $S = [-1, 1]$ and let Π and ρ be the Lebesgue measure. Then, the covariance operator $\Omega_{0\theta}$ can be constructed by using Legendre polynomials since the second Legendre polynomial $P_1(x) = x$ allows to implement the constraint on θ. Because the moment function is separable in θ and x, the prior covariance operator does not depend on θ (see section 4), so that we denote it by Ω_0. The first few Legendre polynomials are $\{1, x, (3x^2 - 1)/2, (5x^3 - 3x)/2, \ldots\}$ and an important property of these polynomials is that they are orthogonal with respect to the L^2 inner product on $[-1, 1]$: $\int_{-1}^{1} P_l(x)P_j(x)dx = 2/(2j + 1)\delta_{lj}$, where δ_{lj} is equal to 1 if $l = j$ and to 0 otherwise. Moreover, the Legendre polynomial obey the recurrence relation $(j + 1)P_{j+1}(x) = (2j + 1)xP_j(x) - jP_{j-1}(x)$ which is useful for computing Ω_0 in practice. The normalized Legendre polynomials form a basis for $L^2[-1, 1]$ so that we can construct the operator Ω_0 as

$$\Omega_0 \cdot = \sigma_0 \sum_{j=2}^{\infty} \lambda_j \frac{2j + 1}{2} \langle P_j, \cdot \rangle P_j$$

where we have set $\lambda_0 = \lambda_1 = 0$ in order to implement the constraints. The remaining λ_j, $j \geq 2$ can be chosen in an arbitrary way provided that $\sum_{j\geq 2} \lambda_j < \infty$. The constant σ_0 can be set to an arbitrary value and has the purpose of tuning the size of the prior covariance.

Many orthogonal polynomials are suitable for the construction of $\Omega_{0\theta}$ and they may be used to treat cases where S is different from $[-1, 1]$.

We perform two simulations exercises: the first one makes use of the empirical cumulative distribution function to construct r_n: $r_n(t) = F_n(t) := n^{-1} \sum_{i=1}^n 1\{x_i \leq t\}$ and the second one uses the empirical moment generating function $r_n(t) = n^{-1} \sum_{i=1}^n e^{tx_i}$. In both the simulations we use Legendre polynomials and we generate $n = 1000$ *i.i.d.* observations $(x_1, \ldots, x_n)$ from a $\mathcal{N}(0, 1)$ distribution truncated to the interval $[-1, 1]$. The prior distribution for θ is uniform over the interval $[-1, 1]$. The prior mean function $f_{0\theta}$ is taken equal to the *pdf* of a Beta distribution with parameters p_θ and q and with support $[-1, 1]$:

$$f_{0\theta}(x) = \frac{(x+1)^{p_\theta - 1}(1-x)^{q-1}}{B(p_\theta, q) 2^{p_\theta + q - 1}}. \tag{5.1}$$

We use the notation p_θ to stress the dependence on θ of this shape parameter. We fix $q = 2$ and recover p_θ such that $\int_{-1}^1 x f_{0\theta}(x)dx = \theta$. It is easy to see that for our Beta distribution: $\int_{-1}^1 x f_{0\theta}(x)dx = \frac{p_\theta - q}{p_\theta + q}$. The covariance operator $\Omega_{0\theta}$ is constructed by using the Legendre polynomials, $\lambda_j = j^{-1.7}$ and $\sigma_0 = 5$.

Since the posterior distribution $\mu(\theta|r_n)$ can not be computed in a closed-form we simulate from it by using a *Metropolis-Hastings algorithm*, see for instance Robert [2002]. To implement this algorithm we have to selected an auxiliary *pdf* g_a. We summarize the simulation schemes for the two cases.

1. Draw a n *i.i.d.* sample $(x_1, \ldots, x_n)$ from f_* (where f_* is a $\mathcal{N}(0, 1)$ truncated to $[-1, 1]$);

2. compute $r_n = F_n$ or $r_n = n^{-1} \sum_{i=1}^n e^{tx_i}$;

3. draw $\theta \sim \mathcal{U}[-1, 1]$ and denote it $\tilde{\theta}$;

4. compute p_θ as $p_\theta = \frac{(\tilde{\theta}+1)2}{1-\tilde{\theta}}$ (where we have fixed $q = 2$);

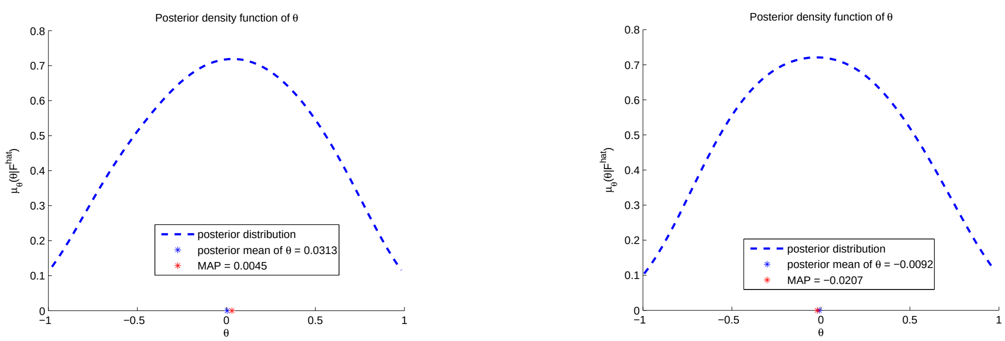
5. compute $f_{0\theta}$ as in (5.1) with parameters $(p_\theta, q = 2)$;

6. draw θ from the marginal posterior distribution of θ by using a *Metropolis-Hasting algorithm* with the following auxiliary *pdf* (triangular distribution):

$$g_a(\xi; \theta) = \frac{\xi + 1}{\theta + 1} I\{\xi \in [-1, \theta)\} + \frac{1 - \xi}{1 - \theta} I\{\xi \in [\theta, 1]\}.$$

We draw 10000 values and discard the first 5000. The initial value for the algorithm is $\theta = 0.5$.

We represent in Figure 2a the results for the simulation with $r_n(t) = F_n(t)$ and in Figure 2b the results for $r_n = n^{-1}\sum_{i=1}^{n} e^{tx_i}$: the blue asterisk represent the posterior mean estimate while the red asterisk represents the MAP estimate. These figures also show the marginal posterior distribution of θ (dashed blue line) approximated by using a kernel smoothing and 5000 drawings from the posterior. In both the simulations, $n = 1000$ and the number of discretization points, used to approximate the integrals, is equal to 1000.



(a) $r_n(t) = F_n(t)$ (b) $r_n = n^{-1}\sum_{i=1}^{n} e^{tx_i}$

Figure 2: Estimations of θ based on the posterior distribution: posterior mean and MAP. The true value of θ is $\theta_* = 0$ and $n = 1000$.

5.3 Overidentified case

Let us consider the case in which x is univariate and the one-dimensional parameter of interest θ is characterized by the moment conditions $\mathbf{E}^F(h(\theta, x)) = 0$ with $h(\theta, x) = (x - \theta, 2\theta^2 - x^2)^T$. For instance, this arises when the true data generating process F is an exponential distribution with parameter θ. The prior $\mu(\theta)$ is specified as a $\mathcal{U}[\theta_* - 1, \theta_* + 1]$.

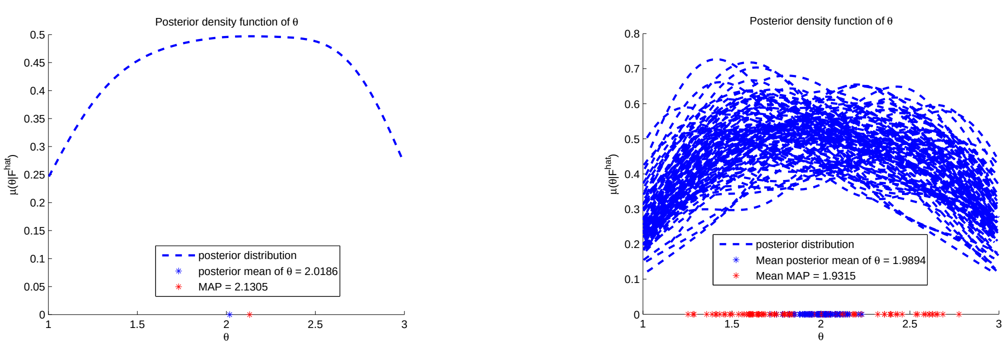
The moment conditions are incorporated in the prior $\mu(f|\theta)$ for f as described in section 2.1. We chose $\Pi(dx) = e^{-x}dx$ and the empirical cumulative distribution function to construct r_n, that is, $r_n(t) = F_n(t)$. We first orthonormalize the moment functions 1, $x - \theta$, $2\theta^2 - x^2$ with respect to Π and then complete the bases by using the Gram-Schmidt orthonormalization process. The inner products in $\mathcal{E}$ are approximating by using the trapezoidal rule on equally spaced subintervals of the interval $[\min x_i - 1, \max x_i + 1]$. We use polynomially decreasing eigenvalues for $\Omega_{0\theta}$: $\lambda_j = j^{-1.7}$. Finally, to construct $\Omega_{0\theta}$ we truncate the series at $J = 300$ since after that the value of λ_j is of the order 10^{-5} and then can be considered zero.

In our simulation, we generate $n = 500$ observations $x_1, \ldots, x_n$ from an exponential distribution with parameter $\theta_* = 2$. Operators K and K^* are approximated by using the trapezoidal rule on equally spaced subintervals of the intervals: $[\min x_i - 1, \max x_i + 1]$ for K and $[\min x_i, \max x_i]$ for K^*. The measure $\rho(dt)$, necessary to construct K^*, is taken equal to the Lebesgue measure. The operator Σ is approximated in a similar way. Because of this discretization, the operator Σ is ill-conditioned and hence we regularize it by adding to it the identity matrix scaled by n^{-1}.

The prior mean function $f_{0\theta}$ is chosen by using a two-step procedure where in the first step we compute $\tilde{f} = (0.1I + K^*K)^{-1}K^*r_n$ and in the second step we project it on $\Lambda(\theta)$ for a given θ.

To draw from the posterior distribution of θ, we use a *Metropolis-Hastings algorithm*. To implement this algorithm we use, as auxiliary distribution, a $\chi^2_{\lceil\theta\rceil}$ distribution. The posterior distribution, its mean and its mode obtained in this simulation are plotted in Figure 3a. The posterior density function has been obtained by kernel smoothing with a Gaussian kernel and a bandwidth equal to 0.3.

Finally, we have repeated the same Monte Carlo simulation 100 times and have computed the average of the posterior mean estimators and MAP estimators. We report the results in Figure 3b together with the posterior density, mean and MAP obtained in each simulation.



(a) Result for one simulation.

(b) Average of posterior means and MAP estimators over 100 Monte Carlo iterations.

Figure 3: Overidentified case. Posterior distributions of θ, mean and MAP estimators. $r_n = F_n$ and the true value of θ is $\theta_* = 2$.

Appendix

A Conditional posterior distribution of f, given θ

The conditional posterior distribution of f given (r_n, θ), $\mu(f|r_n, \theta)$, is a Gaussian process, see Florens and Simoni [2014, Theorem 1]. The conditional posterior mean and variance of $f|r_n, \theta$, in general, rise problems due to the infinite dimension of f. While this point has been broadly discussed in [Florens and Simoni, 2012, 2014] and references therein, in this section we analyze this problem in the particular case considered in the paper where the operators take a specific form.

Intuitively, the problem encountered in the computation of the moments of the Gaussian posterior distribution $\mu(f|r_n, \theta)$ is the following. The moments of a conditional Gaussian distribution involve the inversion of the covariance operator of the conditioning variable r_n, that is $(\Sigma_n + K\Omega_{0\theta}K^*)$ in our case. The problem arises because the inverse operator $(\Sigma_n + K\Omega_{0\theta}K^*)^{-1}$ is in general defined only on a subset of $\mathcal{F}$ of measure zero. Therefore, in general there is no closed-form available for the mean and variance of $\mu(f|r_n, \theta)$ which implies that they cannot be computed.

However, for the framework under consideration we determine mild conditions under which there exists a closed-form for the mean and variance of $\mu(f|r_n, \theta)$. We illustrate these conditions in the lemmas below where we use the notation $\mathfrak{B}$ for the Borel σ-field generated by the open sets of $\mathbb{R}^p$.

Lemma A.1. *Consider the Gaussian distribution (2.7) on $\mathfrak{B}_{\mathcal{E}} \times \mathfrak{B}_{\mathcal{F}}$ and assume that $f_*^{-1/2} \in \mathcal{R}(K^*)$. Then, the conditional distribution on $\mathfrak{B}_{\mathcal{E}}$ conditional on $\mathfrak{B}_{\mathcal{F}} \times \mathfrak{B}$, denoted by $\mu(f|r_n, \theta)$, exists, is regular and a.s. unique. It is Gaussian with mean*

$$\mathbf{E}[f|r_n] = f_{0\theta} + A(r_n - Kf_{0\theta}) \tag{A.1}$$

and trace class covariance operator

$$Var[f|r_n] = \Omega_{0\theta} - AK\Omega_{0\theta} : \mathcal{E} \to \mathcal{E} \tag{A.2}$$

where $A := \Omega_{0\theta}M_f^{-1/2}\left(\frac{1}{n}I - \frac{1}{n}M_f^{1/2}\langle M_f^{1/2}, \cdot\rangle + M_f^{-1/2}\Omega_{0\theta}M_f^{-1/2}\right)^{-1}((K^*)^{-1}M_f^{-1/2})^*$ and $M_f : \mathcal{E} \to \mathcal{E}$ is the multiplication operator $M_f\varphi = f_*\varphi$, $\forall \varphi \in \mathcal{E}$. If in addition: either (i) $f_*^{-1/2} \in \mathcal{E} \cap \mathcal{C}^\infty$ or (ii) the domain $\mathcal{D}\left(M_f^{-\frac{1}{2}}\Omega_{0\theta}M_f^{-\frac{1}{2}}\right)$ is dense in $\mathcal{E}$ and
$\inf_{x \in S}\left(M_f^{-\frac{1}{2}}\Omega_{0\theta}M_f^{-\frac{1}{2}}\right)(x) \geq C$ for a constant $C > 0$ then, A is a continuous and linear operator from $\mathcal{F}$ to $\mathcal{E}$.

Proof. The first part of the theorem follows from Theorem 1 (ii) in Florens and Simoni

[2014]. From this result, since $\Sigma_n = \frac{1}{n}\Sigma$, where $\Sigma : \mathcal{F} \to \mathcal{F}$ is defined in Lemma 2.2, we know that $\mathbf{E}[f|r_n] = f_{0\theta} + \Omega_{0\theta}K^*(\frac{1}{n}\Sigma + K\Omega_{0\theta}K^*)^{-1}(r_n - Kf_{0\theta})$ and $Var[f|r_n] = \Omega_{0\theta} - \Omega_{0\theta}K^*(\frac{1}{n}\Sigma + K\Omega_{0\theta}K^*)^{-1}K\Omega_{0\theta}$. Hence, we have to show that $\Omega_{0\theta}K^*(\frac{1}{n}\Sigma+K\Omega_{0\theta}K^*)^{-1} = A$ and that A is continuous and linear. Denote $\widetilde{M}\cdot = \left(\frac{1}{n}I\cdot - \frac{1}{n}M_f^{\frac{1}{2}}\langle M_f^{\frac{1}{2}}, \cdot\rangle + M_f^{-\frac{1}{2}}\Omega_{0\theta}M_f^{-\frac{1}{2}}\cdot\right)^{-1}$ and

$$\breve{M}\cdot = \left(\frac{1}{n}KM_fK^*\cdot - \frac{1}{n}(KM_f1)\langle M_f, K^*\cdot\rangle + K\Omega_{0\theta}K^*\cdot\right)^{-1}.$$

By using the result of Lemma 2.2, we can write: $\Omega_{0\theta}K^*(\frac{1}{n}\Sigma + K\Omega_{0\theta}K^*)^{-1} = \Omega_{0\theta}K^*\breve{M}$ and then

$$\begin{aligned}\Omega_{0\theta}K^*\left(\frac{1}{n}\Sigma + K\Omega_{0\theta}K^*\right)^{-1} &= \Omega_{0\theta}M_f^{-\frac{1}{2}}\widetilde{M}((K^*)^{-1}M_f^{-\frac{1}{2}})^* \\ &\quad + \Omega_{0\theta}\left[K^*\breve{M} - M_f^{-\frac{1}{2}}\widetilde{M}((K^*)^{-1}M_f^{-\frac{1}{2}})^*\right] \\ &= \Omega_{0\theta}M_f^{-\frac{1}{2}}\widetilde{M}((K^*)^{-1}M_f^{-\frac{1}{2}})^*\end{aligned}$$

where the second equality follows because

$$\begin{aligned}\left[K^*\breve{M} - M_f^{-\frac{1}{2}}\widetilde{M}((K^*)^{-1}M_f^{-\frac{1}{2}})^*\right] &= \left[K^* - M_f^{-\frac{1}{2}}\widetilde{M}((K^*)^{-1}M_f^{-\frac{1}{2}})^*\breve{M}^{-1}\right]\breve{M} \\ = M_f^{-\frac{1}{2}}\widetilde{M}&\left[\left(\frac{1}{n}M_f^{\frac{1}{2}} - \frac{1}{n}M_f^{\frac{1}{2}}\langle M_f, \cdot\rangle + M_f^{-\frac{1}{2}}\Omega_{0\theta}\right)K^* - ((K^*)^{-1}M_f^{-\frac{1}{2}})^*\breve{M}^{-1}\right]\breve{M} \\ &= 0.\end{aligned}$$

This establishes that $\Omega_{0\theta}K^*(\frac{1}{n}\Sigma+K\Omega_{0\theta}K^*)^{-1}$ is equal to A. We now show that the operator A is continuous and linear on $\mathcal{F}$. First, remark that the assumption $f_*^{-\frac{1}{2}} \in \mathcal{R}(K^*)$ ensures that $(K^*)^{-1}M_f^{-\frac{1}{2}}$ exists and is bounded and that $\Omega_{0\theta}f_*^{-1/2}$ is bounded. By construction, $\Omega_{0\theta}$ is trace class. This means that $\Omega_{0\theta}^{\frac{1}{2}}$ is Hilbert-Schmidt, which is a compact operator. Therefore, since the product of two bounded and compact operators is compact, it follows that $\Omega_{0\theta}$ is compact.

Consider the case where *(i)* holds. Hence, $M_f^{-\frac{1}{2}}\Omega_{0\theta}M_f^{-\frac{1}{2}}$ is compact. Moreover, it is easy to show that the operator $\frac{1}{n}M_f^{\frac{1}{2}}\langle M_f^{\frac{1}{2}}, \cdot\rangle : \mathcal{E} \to \mathcal{E}$ is compact since its Hilbert-Schmidt norm is equal to 1. In particular this operator has rank equal to $1/n$ since it has only one eigenvalue different from 0 and which is equal to 1. This eigenvalue corresponds to the eigenfunction $f_*^{\frac{1}{2}}$. Therefore, the operator $(\frac{1}{n}M_f^{\frac{1}{2}}\langle M_f^{\frac{1}{2}}, \cdot\rangle - M_f^{-\frac{1}{2}}\Omega_{0\theta}M_f^{-\frac{1}{2}})$ is compact. By the Cauchy-Schwartz inequality we have

$$\forall \phi \in \mathcal{E}, \quad \langle \tilde{M}^{-1}\phi, \phi\rangle = \frac{1}{n}||\phi||^2 - \frac{1}{n}\langle f_*^{\frac{1}{2}}, \phi\rangle^2 + \langle \Omega_{0\theta}^{\frac{1}{2}}f_*^{-\frac{1}{2}}\phi, \Omega_{0\theta}^{\frac{1}{2}}f_*^{-\frac{1}{2}}\phi\rangle$$

$$
\begin{aligned}
&\geq \frac{1}{n}||\phi||^2 - \frac{1}{n}||f_*^{\frac{1}{2}}||^2||\phi||^2 + ||\Omega_{0\theta}^{\frac{1}{2}} f_*^{-\frac{1}{2}}\phi||^2 \\
&\geq ||\Omega_{0\theta}^{\frac{1}{2}} f_*^{-\frac{1}{2}}\phi||^2 \geq 0 \qquad \text{(A.3)}
\end{aligned}
$$

since $||f_*^{\frac{1}{2}}||^2 = 1$. Therefore, we conclude that $\tilde{M}$ is injective. Then, from the Riesz Theorem 3.4 in Kress [1999] it follows that the operator $\tilde{M} : \mathcal{E} \to \mathcal{E}$ is bounded.

Consider the case where *(ii)* holds. Hence, by using (A.3) for every $\phi \in \mathcal{D}(M_f^{-\frac{1}{2}}\Omega_{0\theta}M_f^{-\frac{1}{2}})$, it follows that $\tilde{M} : \mathcal{E} \to \mathcal{E}$ exists and is bounded.

Finally, the operator A is bounded and linear since it is the product of bounded linear operators. We conclude that A is a continuous operator from $\mathcal{F}$ to $\mathcal{E}$.

$\square$

Remark A.1. If $f_*^{-1} \in \mathcal{R}(K^*)$ then the operator $A : \mathcal{F} \to \mathcal{E}$ of Lemma (A.1) may be written in an equivalent way as: $\forall \varphi \in \mathcal{F}$

$$
A\varphi = \Omega_{0\theta}\left(\frac{1}{n}I - \frac{1}{n}\langle f_*, \cdot\rangle + f_*^{-1}\Omega_{0\theta}\right)^{-1}((K^*)^{-1}f_*^{-1})^*. \qquad \text{(A.4)}
$$

In addition, if either *(i)* $f_*^{-1} \in \mathcal{E} \cap \mathcal{C}^\infty$ or *(ii)* the domain $\mathcal{D}(M_f^{-1}\Omega_{0\theta})$ is dense in $\mathcal{E}$ and $\inf_{x \in S}\left(M_f^{-\frac{1}{2}}\Omega_{0\theta}M_f^{-\frac{1}{2}}\right)(x) \geq C$ for a constant $C > 0$ then, A is a continuous and linear operator from $\mathcal{F}$ to $\mathcal{E}$.

The trajectories of f generated by the conditional posterior distribution $\mu(f|r_n, \theta)$ verify a.s. the moment conditions and integrate to 1. To see this, first remark that the posterior covariance operator satisfies the moment restrictions:

$$
[\Omega_{0\theta} - AK\Omega_{0\theta}]^{1/2}(1, h^T(\theta, \cdot))^T = [I - AK]^{1/2}\Omega_{0\theta}^{1/2}(1, h^T(\theta, \cdot))^T = 0
$$

where we have factorized out $\Omega_{0\theta}$ on the right and used (2.2). Second, a trajectory f drawn from the posterior $\mu(f|\theta, r_n)$ is such that $(f - f_{0\theta}) \in \overline{\mathcal{R}\left((\Omega_{0\theta} - AK\Omega_{0\theta})^{1/2}\right)}$, a.s. Now, for any $\phi \in \mathcal{R}\left((\Omega_{0\theta} - AK\Omega_{0\theta})^{1/2}\right)$ we have $\langle \phi, h(\theta, \cdot)\rangle = \langle [\Omega_{0\theta} - AK\Omega_{0\theta}]^{1/2}\psi, h(\theta, \cdot)\rangle = \langle [I - \tilde{A}K\Omega_{0\theta}]^{1/2}\psi, \Omega_{0\theta}^{\frac{1}{2}}h(\theta, \cdot)\rangle = 0$, for some $\psi \in \mathcal{E}$ where $\tilde{A} = K^*(\Sigma_n + K\Omega_{0\theta}K^*)^{-1}$, and $\langle \phi, 1\rangle = 0$ by a similar argument. This shows that

$$
\mathcal{R}\left((\Omega_{0\theta} - AK\Omega_{0\theta})^{1/2}\right) \subset \left\{\phi \in \mathcal{E}\ ;\ \int \phi(x)h(\theta, x)\Pi(dx) = 0 \text{ and } \int \phi(x)\Pi(dx) = 0\right\}
$$

and since the set on the right of this inclusion is closed we have

$$
\overline{\mathcal{R}\left((\Omega_{0\theta} - AK\Omega_{0\theta})^{1/2}\right)} \subset \left\{\phi \in \mathcal{E}\ ;\ \int \phi(x)h(\theta, x)\Pi(dx) = 0 \text{ and } \int \phi(x)\Pi(dx) = 0\right\}.
$$

Therefore, a.s. a trajectory f drawn from $\mu(f|\theta, r_n)$ is such $\int(f - f_{0\theta})(x)\Pi(dx) = 0$ and $\int(f - f_{0\theta})(x)h(\theta, x)\Pi(dx) = 0$ which implies: $\int f(x)\Pi(dx) = 1$ and $\int f(x)h(\theta, x)\Pi(dx) = 0$.

Remark A.2. The posterior distribution of f conditional on θ revises the prior on f except in the directions given by the constant and the moment functions h which remain unchanged.

Remark A.3. [Regularized Posterior Distribution] When neither the conditions of Lemma A.1 nor the conditions of Remark A.1 are satisfied then we cannot use the exact posterior distribution $\mu(f|\theta, r_n)$. Instead, we can use the *regularized posterior distribution* denoted by $\mu(f|\tau, \theta, r_n)$, where $\tau > 0$ is a regularization parameter that must be suitably chosen and that converges to 0 with n. This distribution has been proposed by Florens and Simoni [2012] and we refer to this paper for a complete description of it. Here, we only give its expression: $\mu(f|\tau, \theta, r_n)$ is a Gaussian distribution with mean function

$$\mathbf{E}[f|r_n, \tau] = f_{0\theta} + A_\tau(r_n - Kf_{0\theta})$$

and covariance operator $Var[f|r_n, \tau] = \Omega_{0\theta} - A_\tau K\Omega_{0\theta} : \mathcal{E} \to \mathcal{E}$ where

$$A_\tau := \Omega_{0\theta}K^*\left(\tau I + \frac{1}{n}I + K\Omega_{0\theta}K^*\right)^{-1} : \mathcal{E} \to \mathcal{E}$$

and $I : \mathcal{F} \to \mathcal{F}$ denotes the identity operator.

B Proofs for Section 2

Proof of Lemma 2.1

Let $\mathcal{H}(\Omega_{0\theta})$ denote the reproducing kernel Hilbert space associated with $\Omega_{0\theta}$ and embedded in $\mathcal{E}$ and $\overline{\mathcal{H}(\Omega_{0\theta})}$ denote its closure. Because $f|\theta \sim \mathcal{GP}(f_{0\theta}, \Omega_{0\theta})$ then $(f - f_{0\theta}) \in \overline{\mathcal{H}(\Omega_{0\theta})}$ a.s. Moreover, $\mathcal{H}(\Omega_{0\theta}) = \mathcal{D}(\Omega_{0\theta}^{-1/2}) = \mathcal{R}(\Omega_{0\theta}^{1/2})$ where $\mathcal{D}$ and $\mathcal{R}$ denote the domain and the range of an operator, respectively. This means that $\forall\phi \in \mathcal{H}(\Omega_{0\theta})$ there exists $\psi \in \mathcal{E}$ such that $\phi = \Omega_{0\theta}^{\frac{1}{2}}\psi$. Moreover, for any $\phi \in \mathcal{H}(\Omega_{0\theta})$ we have $\langle\phi, h(\theta, \cdot)\rangle = \int \phi(x)h(\theta, x)\Pi(dx) = \langle\Omega_{0\theta}^{\frac{1}{2}}\psi, h(\theta, \cdot)\rangle = \langle\psi, \Omega_{0\theta}^{\frac{1}{2}}h(\theta, \cdot)\rangle = 0$ and $\langle\phi, 1\rangle = 0$ by a similar argument. Hence,

$$\mathcal{H}(\Omega_{0\theta}) \subset \left\{\phi \in \mathcal{E}\ ;\ \int \phi(x)h(\theta, x)\Pi(dx) = 0 \text{ and } \int \phi(x)\Pi(dx) = 0\right\}. \tag{B.1}$$

Since the set on the right of this inclusion is closed we have

$$\overline{\mathcal{H}(\Omega_{0\theta})} \subset \left\{\phi \in \mathcal{E}\ ;\ \int \phi(x)h(\theta, x)\Pi(dx) = 0 \text{ and } \int \phi(x)\Pi(dx) = 0\right\}.$$

We deduce that,

$$\int (f - f_{0\theta})(x)\Pi(dx) = 0 \quad \text{and} \quad \int (f - f_{0\theta})(x)h(\theta, x)\Pi(dx) = 0 \qquad \text{a.s.}$$

Condition (2.1) and the fact that $f_{0\theta}$ is a *pdf* imply the results of the lemma.

□

Proof of Lemma 2.2

The result follows trivially from the definition of the covariance operator $\Sigma_n : \mathcal{F} \to \mathcal{F}$ and from the Fubini's Theorem: $\forall \psi \in \mathcal{F}$,

$$\begin{aligned}
\Sigma_n \psi = \frac{1}{n}\Big[\int_{\mathfrak{T}} \int_S & (k(t,x)k(s,x))\, f_*(x)\Pi(dx)\psi(t)\rho(dt) \\
& - \int_{\mathfrak{T}} \int_S k(t,x) f_*(x)\Pi(dx) \left(\int_S k(s,x) f_*(x)\Pi(dx) \right) \psi(t)\rho(dt) \Big] \\
&= \frac{1}{n}\Big[\int_S k(s,x) f_*(x) \int_{\mathfrak{T}} k(t,x)\psi(t)\rho(dt)\Pi(dx) \\
& - \int_S k(s,x) f_*(x)\Pi(dx) \left(\int_S \int_{\mathfrak{T}} k(t,x)\psi(t)\rho(dt) f_*(x)\Pi(dx) \right) \Big] \\
&= \frac{1}{n} \left[KM_f K^* \psi - (KM_f 1)\langle M_f, K^*\psi \rangle \right]
\end{aligned}$$

where the second equality has been obtained by using the Fubini's theorem.

□

Proof of Lemma 2.3

We can rewrite Σ as

$$\begin{aligned}
\forall \psi \in \mathcal{F}, \quad \Sigma\psi &= \int_{\mathfrak{T}} \mathbf{E}^* \left(v(x,t)v(x,s) \right) \psi(t)\rho(dt) \\
&= \int_{\mathfrak{T}} \int_S \left(v(x,t)v(x,s) \right) f_*(x)\Pi(dx)\psi(t)\rho(dt)
\end{aligned}$$

where $v(x,t) = [k(x,t) - \mathbf{E}(k(x,t))]$. Then, $\forall \psi \in \mathcal{F}$ we can write $\Sigma\psi = RM_f R^*\psi$ where $R : \mathcal{E} \to \mathcal{F}$, $M_f : \mathcal{E} \to \mathcal{E}$ and $R^* : \mathcal{F} \to \mathcal{E}$ are the operators defined as

$$\begin{aligned}
\forall \psi \in \mathcal{F}, \quad R^*\psi &= \int_{\mathfrak{T}} v(x,t)\psi(t)\rho(dt) \\
\forall \varphi \in \mathcal{E}, \quad M_f \varphi &= f_*(x)\varphi(x) \\
\forall \varphi \in \mathcal{E}, \quad R\varphi &= \int_S v(x,t)\varphi(x)\Pi(dx).
\end{aligned}$$

Moreover, we have $\mathcal{D}(\Sigma^{-\frac{1}{2}}) = \mathcal{R}(\Sigma^{\frac{1}{2}}) = \mathcal{R}((RM_fR^*)^{\frac{1}{2}}) = \mathcal{R}(RM_f^{1/2})$.

Let $h \in \mathcal{R}(K)$, namely, there exists a $g \in \mathcal{E}$ such that $h(t) = \int_S k(t,x)g(x)\Pi(dx)$. Then $h \in \mathcal{D}(\Sigma^{-\frac{1}{2}})$ if there exists an element $\nu \in \mathcal{E}$ such that $h(t) = \int_S v(x,t)f_*^{\frac{1}{2}}(x)\nu(x)\Pi(dx)$. By developing this equality, the element ν has to satisfy

$$\begin{aligned}
&\int_S k(t,x)g(x)\Pi(dx) = \int_S v(x,t)f_*^{\frac{1}{2}}(x)\nu(x)\Pi(dx)\\
\Leftrightarrow &\int_S k(t,x)g(x)\Pi(dx) = \int_S \left[k(x,t) - \left(\int_S k(x,t)f_*(x)\Pi(dx)\right)\right] f_*^{\frac{1}{2}}(x)\nu(x)\Pi(dx)\\
\Leftrightarrow &\int_S k(t,x)g(x)\Pi(dx) = \int_S k(x,t)\left[f_*^{\frac{1}{2}}(x)\nu(x) - f_*(x)\left(\int_S f_*^{\frac{1}{2}}(x)\nu(x)\Pi(dx)\right)\right]\Pi(dx).
\end{aligned}$$

If K is injective it follows that such an element ν must satisfy

$$g(x) = f_*^{\frac{1}{2}}\nu(x) - f_*(x)\left(\int_S f_*^{\frac{1}{2}}(x)\nu(x)\Pi(dx)\right)$$

which in turn implies that $\int_S g(x)\Pi(dx) = 0$, *i.e.* that $h \in \mathcal{R}(K|_{\mathfrak{D}})$. Therefore, one solution is $\nu(x) = f_*^{-\frac{1}{2}}g(x)$ which proves that the range of the truncated operator $K|_{\mathfrak{D}}$ in contained in $\mathcal{D}(\Sigma^{-\frac{1}{2}})$. On the other side, let $h \in \mathcal{D}(\Sigma^{-\frac{1}{2}})$, then there exists a $\nu \in \mathcal{E}$ such that $h = \int_S v(x,t)f_*^{\frac{1}{2}}(x)\nu(x)\Pi(dx)$. By the previous argument and under the assumption that $K|_{\mathfrak{D}}$ is injective, this implies that $h \in \mathcal{R}(K|_{\mathfrak{D}})$ since there exists $g \in \mathfrak{D}$ such that $g(x) = f_*^{\frac{1}{2}}\nu(x) - f_*(x)\left(\int_S f_*^{\frac{1}{2}}(x)\nu(x)\Pi(dx)\right)$. This shows the inclusion of $\mathcal{D}(\Sigma^{-\frac{1}{2}})$ in $\mathcal{R}(K|_{\mathfrak{D}})$ and concludes the proof.

$\square$

Proof of Theorem 2.1

In this proof we denote $B = \Sigma^{-1/2}K\Omega_{0\theta}^{1/2}$. To prove that P_n^θ and P_n^0 are equivalent we first rewrite the covariance operator of P_n^θ as

$$\left(n^{-1}\Sigma + K\Omega_{0\theta}K^*\right) = \frac{1}{n}\Sigma^{\frac{1}{2}}\left[I + n\Sigma^{-\frac{1}{2}}K\Omega_{0\theta}K^*\Sigma^{-\frac{1}{2}}\right]\Sigma^{\frac{1}{2}}.$$

Then, according to Corollary 3.1, Theorem 3.3 and Theorem 3.4 p.125 in Kuo [1975], P_n^θ and P_n^0 are equivalent if $K(f_{0\theta} - f_*) \in \mathcal{R}(\Sigma^{1/2})$ and if $\left[I + n\Sigma^{-\frac{1}{2}}K\Omega_{0\theta}K^*\Sigma^{-\frac{1}{2}}\right]$ is positive definite, bounded, invertible with BB^* Hilbert Schmidt, where B^* denotes the adjoint of B. We now verify these conditions.

1) Since $(f_{0\theta} - f_*) \in \mathfrak{D}$ and since $K|_{\mathfrak{D}}$ is injective then, by Lemma 2.3, $K(f_{0\theta} - f_*) \in \mathcal{R}(\Sigma^{1/2})$.

2) *Positive definiteness.* It is trivial to show that the operator $(I + nBB^*)$ is self-adjoint,

i.e. $(I + nBB^*)^* = (I + nBB^*)$. Moreover, $\forall \varphi \in \mathcal{F}$, $\varphi \neq 0$

$$\langle (I + nBB^*)\varphi, \varphi\rangle = \langle \varphi, \varphi\rangle + n\langle B^*\varphi, B^*\varphi\rangle = ||\varphi||^2 + n||B^*\varphi|| > 0.$$

3) *Boundedness.* By Lemma 2.3, if $K|_{\mathfrak{D}}$ is injective, the operators B and B^* are bounded; the operator I is bounded by definition and a linear combination of bounded operators is bounded, see Remark 2.7 in Kress [1999].

4) *Continuously invertible.* The operator $(I+nBB^*)$ is continuously invertible if its inverse is bounded, *i.e.* there exists a positive number C such that $||(I + nBB^*)^{-1}\varphi|| \leq C||\varphi||$, $\forall \varphi \in \mathcal{F}$. We have $||(I + nBB^*)^{-1}\varphi|| \leq ||\varphi|| < \infty$, $\forall \varphi \in \mathcal{F}$.

5) *Hilbert-Schmidt.* We consider the Hilbert-Schmidt norm $||nBB^*||_{HS} = n\sqrt{tr((BB^*)^2)}$. Now, $tr((BB^*)^2) = tr(\Omega_{0\theta}\widetilde{B}^*\widetilde{B}\Omega_{0\theta}\widetilde{B}^*\widetilde{B}) \leq tr(\Omega_{0\theta})||\widetilde{B}^*\widetilde{B}\Omega_{0\theta}\widetilde{B}^*\widetilde{B}|| < \infty$ since the operator $\widetilde{B} := \Sigma^{-\frac{1}{2}}K|_{\mathcal{R}(\Omega_{0\theta}^{1/2})}$ has a bounded norm by Lemma 2.3. This shows that P_n^θ and P_n^0 are equivalent.

Next, we derive (2.9). Remark that P_n^θ and P_0 are the distributions of the stochastic process r_n. In an equivalent way, $Z := \sqrt{n}(r_n - Kf_*)$ is distributed as a $\mathcal{GP}(\sqrt{n}K(f_{0\theta} - f_*), (\Sigma + nK\Omega_{0\theta}K^*))$ according to P_n^θ and as a $\mathcal{GP}(0, \Sigma)$ according to P_0. Let $Z_j := \langle\sqrt{n}(r_n - Kf_*), \Sigma^{-1/2}\psi_j(\theta)\rangle$ for $j \geq 0$. This variable is defined under the further assumption that $\psi_j(\theta) \in \mathcal{R}(\Sigma^{1/2})$, $\forall j \geq 0$ and $\forall \theta \in \Theta$. By Theorem 2.1 in Kuo [1975, page 116]:

$$\frac{dP_n^\theta}{dP_n^0} = \prod_{j=0}^{\infty} \frac{d\nu_j}{d\mu_j}$$

where ν_j denotes the distribution of Z_j under P_n^θ (namely,

$$\nu_j = \mathcal{N}\Big(\langle\sqrt{n}K(f_{0\theta} - f_*), \Sigma^{-1/2}\psi_j(\theta)\rangle, (1 + nl_{j\theta}^2)\Big))$$

and μ_j denotes the distribution of Z_j under P_n^0 (namely, $\mu_j = \mathcal{N}(0, 1)$). By writing down the likelihoods of ν_j and μ_j with respect to the Lebesgue measure and after simplifications we obtain

$$\frac{dP_n^\theta}{dP_n^0}(r_n) = \prod_{j=0}^{\infty}\sqrt{\frac{n^{-1}}{n^{-1} + l_{j\theta}^2}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{\infty}\frac{\left(Z_j - \langle\sqrt{n}K(f_{0\theta} - f_*), \Sigma^{-1/2}\psi_j(\theta)\rangle\right)^2}{1 + nl_{j\theta}^2}\right\} e^{\{\frac{1}{2}\|Z\|_\Sigma^2\}}$$

where $\|Z\|_\Sigma^2 := \|\Sigma^{-1/2}Z\|^2$ and it is defined as the $\mathcal{F}$ limit of the series $\sum_{j=0}^m \sigma_j^{-2}\langle Z_j, \phi_j\rangle^2$ as $m \to \infty$ (where $\{\sigma_j^2, \phi_j\}_{j=0}^\infty$ is the eigensystem of Σ).

□

Proof of Proposition 2.1

Let $\mathcal{E}_{\Pi_1} = L^2(S, \mathfrak{B}_S, \Pi_1)$ with $\langle \cdot, \cdot \rangle_{\Pi_1}$ the inner product in $\mathcal{E}_{\Pi_1}$. Let $\mathfrak{z} = \frac{d\Pi}{d\Pi_1}$ and consider the transformation φ and its inverse:

$$\begin{array}{rccc} \varphi : & \mathcal{E} & \rightarrow & \mathcal{E}_{\Pi_1} \\ & f & \mapsto & f\mathfrak{z} \end{array} ; \qquad \begin{array}{rccc} \varphi^{-1} : & \mathcal{E}_{\Pi_1} & \rightarrow & \mathcal{E} \\ & g & \mapsto & g\mathfrak{z}^{-1} \end{array} .$$

If $\sup_{x \in S} \frac{d\Pi_1(x)}{d\Pi(x)} < \infty$ then, $\varphi^{-1}(\mathcal{E}_{\Pi_1}) \subset \mathcal{E}$ which means that $\mathcal{E}_{\Pi_1}$ is φ-stable (see Florens et al. [1990, Definition 8.2.14]). For every positive measure Π_1 such that $\sup_{x \in S} \frac{d\Pi_1(x)}{d\Pi(x)} < \infty$ we can define a transformation φ. Let Φ be the set of measurable functions defined in Proposition 2.1. Every $\varphi \in \Phi$ induces a transformation on the parameter space: $\overline{\varphi} : \Theta \times \mathcal{E} \rightarrow \Theta \times \mathcal{E}_{\Pi_1}$ such that $\overline{\varphi}(\theta, f) = (\theta, f\mathfrak{z})$. Moreover, define

$$\begin{array}{rcl} K_1 : \mathcal{E}_{\Pi_1} & \rightarrow & \mathcal{F} \\ g & \mapsto & \int k(t,x) g(x) \Pi_1(dx). \end{array} \tag{B.2}$$

For every $\varphi \in \Phi$, the sampling distribution conditional on the transformed parameter $\varphi(f)$ is $P^{\varphi(f)} = \mathcal{GP}(K_1 \varphi(f), \Sigma_n)$ and it is easy to see that $P^{\varphi(f)} = P^f$ since $K_1 \varphi(f) = Kf$. The rest of the proof has to be meant for a generic element $\varphi \in \Phi$.

Let us look at the prior distribution of the transformed parameter. The prior of θ does not change as θ is not affected by this transformation. On the other hand, the conditional prior of $\varphi(f)$, given θ, is a Gaussian measure on the Borel σ-field of $\mathcal{E}_{\Pi_1}$ induced by the prior distribution of f. That is, $\mu(\varphi(f)|\theta) = \mathcal{GP}(\varphi(f_{0\theta})\mathfrak{z}, \Omega_{\varphi\theta})$ where $\varphi(f_{0\theta}) = f_{0\theta}\mathfrak{z} \in \mathcal{E}_{\Pi_1}$ and $\Omega_{\varphi\theta} : \mathcal{E}_{\Pi_1} \rightarrow \mathcal{E}_{\Pi_1}$ is such that $\forall \delta_1, \delta_2 \in \mathcal{E}_{\Pi_1}$, $\langle \Omega_{\varphi\theta}\delta_1, \delta_2 \rangle = \mathbf{E}^{f|\theta}(\langle f\mathfrak{z}, \delta_1 \rangle_{\Pi_1} \langle f\mathfrak{z}, \delta_2 \rangle_{\Pi_1})$ where $\mathbf{E}^{f|\theta}$ denotes the expectation with respect to $\mu(f|\theta)$. Hence,

$$\begin{aligned} \mathbf{E}^{f|\theta}(\langle f\mathfrak{z}, \delta_1 \rangle_{\Pi_1} \langle f\mathfrak{z}, \delta_2 \rangle_{\Pi_1}) &= \int_S \mathfrak{z}(x) \int_S \mathbf{E}^{f|\theta}\left(f(x)f(s)\right) \mathfrak{z}(s)\delta_1(s)\Pi_1(ds)\delta_2(x)\Pi_1(x) \\ &= \int_S \mathfrak{z}(x) \int_S \sum_{j>d} \lambda_j \varphi_{j\theta}(x)\varphi_{j\theta}(s)\mathfrak{z}(s)\delta_1(s)\Pi_1(ds)\delta_2(x)\Pi_1(x) \\ &= \langle \mathfrak{z} \sum_{j>d} \lambda_j \varphi_{j\theta} \langle \varphi_{j\theta}\mathfrak{z}, \delta_1 \rangle_{\Pi_1}, \delta_2 \rangle_{\Pi_1} \end{aligned}$$

which implies that $\Omega_{\varphi\theta} \cdot = \sum_{j>d} \lambda_j \langle \varphi_{j\theta}\mathfrak{z}, \cdot \rangle_{\Pi_1} \mathfrak{z}\varphi_{j\theta}$.
Therefore, by using the new parametrization, the Bayesian experiment is (by using the notation as in Florens et al. [1990]): $(\Theta \times \mathcal{E}_{\Pi_1} \times \mathcal{F}, \mathfrak{B}_\Theta \otimes \mathfrak{B}_{\mathcal{E}_{\Pi_1}} \otimes \mathfrak{B}_\mathcal{F}, \mu(\theta) \otimes \mu(\varphi(f)|\theta) \otimes P^{\varphi(f)})$ where $\mathfrak{B}_\Theta$ (resp. $\mathfrak{B}_{\mathcal{E}_{\Pi_1}}$, $\mathfrak{B}_\mathcal{F}$) denotes the Borel σ-field generated by the open sets of $\mathbb{R}^p$

(resp. $\mathcal{E}_{\Pi_1}$, $\mathcal{F}$). The moment conditions restrict the parameter space to

$$\Lambda_1 := \left\{(\theta, g) \in \Theta \times \mathcal{E}_{M_1}; \int_S h(\theta, x)g(x)\Pi_1(dx) = 0\right\}, \qquad \mathcal{E}_{M_1} = \mathcal{E}_{\Pi_1} \cap M.$$

The marginal Bayesian experiment, obtained by integrating out the parameter $\varphi(f)$ with respect to $\mu(\varphi(f)|\theta)$ is $(\Theta \times \mathcal{F}, \mathfrak{B}_\Theta \otimes \mathfrak{B}_\mathcal{F}, \mu(\theta) \otimes P_{n,1}^\theta)$ where

$$P_{n,1}^\theta := \mathcal{GP}(K_1 f_{0\theta\mathfrak{z}}, \Sigma_n + K_1 \Omega_{\varphi\theta} K_1^*)$$

and $K_1^* : \mathcal{F} \to \mathcal{E}_{\Pi_1}$ is such that $\forall g \in \mathcal{E}_{\Pi_1}$ and $\forall \psi \in \mathcal{F}$, $\langle K_1 g, \psi\rangle = \langle g, K_1^* \psi\rangle$. The adjoint operator K_1^* has the same kernel as K^*, that is, $\forall \psi \in \mathcal{F}$, $K_1^*\psi = \int_\mathfrak{T} k(t,x)\psi(t)\rho(dt)$. We now show that $K_1\Omega_{\varphi\theta}K_1^* = K\Omega_{0\theta}K^*$: by using the fact that $\mathfrak{z} = \frac{d\Pi}{d\Pi_1}$ we obtain

$$\begin{aligned} K_1\Omega_{\varphi\theta}K_1^* \cdot &= \int_S k(t,x) \sum_{j>d} \lambda_j \langle \varphi_{j\theta\mathfrak{z}}, K_1^* \cdot\rangle_{\Pi_1} \mathfrak{z}(x)\varphi_{j\theta}(x)\Pi_1(dx) \\ &= \int_S k(t,x) \sum_{j>d} \lambda_j \langle \varphi_{j\theta}, K_1^* \cdot\rangle \varphi_{j\theta}(x)\Pi(dx) \\ &= \sum_{j>d} \lambda_j \langle \varphi_{j\theta}, K^* \cdot\rangle K\varphi_{j\theta} = K\Omega_{0\theta}K^* \cdot . \end{aligned}$$

This implies that $P_{n,1}^\theta = P_n^\theta$ and $\Sigma^{-1/2}K_1\Omega_{\varphi\theta}K_1^*\Sigma^{-1/2} = \Sigma^{-1/2}K\Omega_{0\theta}K^*\Sigma^{-1/2}$ so that these two operators have the same eigensystem $(l_{j\theta}, \psi_j(\theta))_{j\in\mathbb{N}}$. This and the fact that $Kf_* = K_1 f_{*\mathfrak{z}}$ imply that Theorem 2.1 applies with P_n^0 replaced by $P_{n,1}^0 := \mathcal{GP}(K_1 f_{*\mathfrak{z}}, n^{-1}\Sigma)$ and

$$\begin{aligned} p_{n\theta,1}(r_n;\theta) &:= \frac{dP_{n,1}^\theta}{dP_{n,1}^0}(r_n) \\ &= \prod_{j=0}^\infty \sqrt{\frac{n^{-1}}{n^{-1} + l_{j\theta}^2}} \exp\left\{-\frac{1}{2}\sum_{j=0}^\infty \frac{\left(Z_{1,j} - \langle \sqrt{n}K_1(f_{0\theta} - f_*)\mathfrak{z}, \Sigma^{-1/2}\psi_j(\theta)\rangle\right)^2}{1 + nl_{j\theta}^2}\right\} e^{\left\{\frac{1}{2}\|Z_1\|_\Sigma^2\right\}} \end{aligned} \tag{B.3}$$

where $Z_1 := \sqrt{n}(r_n - K_1 f_{*\mathfrak{z}})$, $Z_{1,j} := \langle Z_1, \Sigma^{-1/2}\psi_j(\theta)\rangle$ for all $j \geq 0$, and $\|Z\|_\Sigma := \|\Sigma^{-1/2}Z_1\|$. The previous results imply that $p_{n\theta,1}(r_n;\theta) = p_{n\theta}(r_n;\theta)$, $\forall (r_n, \theta) \in \mathcal{F} \times \Theta$, where $p_{n\theta}(r_n;\theta)$ is the expression given in (2.9). Hence, the marginal posterior distribution of θ is φ-invariant. Since the same reasoning carries on for every $\varphi \in \Phi$ we conclude that the marginal posterior distribution of θ is Φ-invariant.

□

Proof of Theorem 2.2

Remark that we may write: $\Sigma = \Sigma_{1/2}\Sigma^*_{1/2}$ where

$$\begin{array}{rccl} \Sigma_{1/2}: & \mathcal{E} & \to & \mathcal{F} \\ & \varphi & \mapsto & Kf_*^{1/2}\varphi - (Kf_*)\langle f_*^{1/2}, \varphi\rangle \\ \Sigma^*_{1/2}: & \mathcal{F} & \to & \mathcal{E} \\ & \psi & \mapsto & f_*^{1/2}K^*\psi + f_*^{1/2}\langle Kf_*, \psi\rangle. \end{array} \tag{B.4}$$

Remark that, despite the notation, $\Sigma^*_{1/2}$ is not the adjoint of $\Sigma_{1/2}$. Let $(\Sigma^*_{1/2})^{-1} = (f_*^{1/2}K^*)^{-1} - \frac{1}{2}(K^*)^{-1}\langle f_*^{1/2}, \cdot\rangle$, it is easy to verify that $(\Sigma^*_{1/2})^{-1}\Sigma^*_{1/2} = I$, so that in the following we use the notation $\Sigma^{-1/2} = (\Sigma^*_{1/2})^{-1} : \mathcal{E} \to \mathcal{F}$. Hence, $\Omega_{0\theta}^{1/2}K^*(\Sigma^*_{1/2})^{-1}f_*^{1/2}\varphi_l = \lambda_l^{1/2}\varphi_l$ for every $l > d$ and $\Omega_{0\theta}^{1/2}K^*(\Sigma^*_{1/2})^{-1}f_*^{1/2}\varphi_l = 0$ for every $l \leq d$. We want to determine the functions $\psi_j(\theta)$ used in (2.9). These are the eigenfunctions of

$$[\Omega_{0\theta}^{1/2}K^*(\Sigma^*_{1/2})^{-1}]^*\Omega_{0\theta}^{1/2}K^*(\Sigma^*_{1/2})^{-1}$$

where $[\Omega_{0\theta}^{1/2}K^*(\Sigma^*_{1/2})^{-1}]^*$ denotes the adjoint of the operator $\Omega_{0\theta}^{1/2}K^*(\Sigma^*_{1/2})^{-1}$ which exists since $\Omega_{0\theta}^{1/2}K^*(\Sigma^*_{1/2})^{-1}$ is bounded if $f_*^{1/2}$ is bounded away from 0 (this is the case if we set $\Pi = F_*$). For every $\phi_1, \phi_2 \in \mathcal{E}$:

$$\langle\Omega_{0\theta}^{1/2}K^*(\Sigma^*_{1/2})^{-1}\phi_1, \phi_2\rangle = \langle\phi_1, \sum_{j>d}\lambda_j^{1/2}\langle\varphi_j, \phi_2\rangle f_*^{-1/2}\varphi_j\rangle$$

which gives an expression for $[\Omega_{0\theta}^{1/2}K^*(\Sigma^*_{1/2})^{-1}]^*$. By using this result it is easy to check that for every $l > d$:

$$[\Omega_{0\theta}^{1/2}K^*(\Sigma^*_{1/2})^{-1}]^*\Omega_{0\theta}^{1/2}K^*(\Sigma^*_{1/2})^{-1}f_*^{1/2}\varphi_l = \lambda_l f_*^{-1/2}\varphi_l.$$

By Proposition 2.1, our inference procedure is invariant to the choice of Π. Then, if $\sup_{x\in S} f_*(x) < \infty$ we can fix $\Pi = F_*$ so that $f_* = 1$ and the eigenfunctions $\{\psi_j(\theta)\}_{j\geq 0}$ are given by $\{\varphi_j\}_{j\geq 0}$ (which depend on θ) with corresponding eigenvalues $\{l_{j\theta}\}_{j\geq 0} = \{\lambda_j 1\{j > d\}\}_{j\geq 0}$ (which do not depend on θ). Remark that, since $\Pi = F_*$, the $\{\varphi_j\}_{j=1}^d$ denote here the moment functions orthonormalized with respect to F_*, that is, $(\varphi_1, \ldots, \varphi_d)^T(x) = V_*(\theta)^{-1/2}h(\theta, x)$ for every $\theta \in \Theta$ and every $x \in S$ and where $V_*(\theta) = \mathbf{E}^*[h(\theta, x)h(\theta, x)^T]$. It follows that if we replace $\Omega_{0\theta}$ by $c\Omega_{0\theta}$, then $\{l_{j\theta}\}_{j\geq 0}$ have to be multiplied by c as well, so that:

$$p_{n\theta}(r_n;\theta) = e^{-\{\frac{1}{2}\langle\sqrt{n}(r_n-Kf_{0\theta}),(\Sigma^*_{1/2})^{-1}1\rangle^2\}} exp\{-\frac{1}{2}\sum_{j=1}^{d}\langle\sqrt{n}(r_n-Kf_{0\theta}),(\Sigma^*_{1/2})^{-1}\varphi_j\rangle^2\}\times$$

$$\prod_{j>d}\sqrt{\frac{n^{-1}}{n^{-1}+cl^2_{j\theta}}}\exp\left\{-\frac{1}{2}\sum_{j>d}\frac{\langle\sqrt{n}(r_n-Kf_{0\theta}),(\Sigma^*_{1/2})^{-1}\varphi_j\rangle^2}{1+cnl^2_{j\theta}}\right\}e^{\{\frac{1}{2}\|Z\|^2_{\Sigma}\}}$$

and

$$\mu(\theta|r_n) =$$

$$e^{-\left\{\frac{1}{2}\sum_{j=0}^{d}\langle\sqrt{n}(r_n-Kf_{0\theta}),(\Sigma^*_{1/2})^{-1}\varphi_j\rangle^2\right\}}\prod_{j>d}\sqrt{\frac{n^{-1}}{\frac{1}{cn}+l^2_{j\theta}}}e^{-\left\{\frac{1}{2}\sum_{j>d}\frac{\langle\sqrt{n}(r_n-Kf_{0\theta}),(\Sigma^*_{1/2})^{-1}\varphi_j\rangle^2}{1+cnl^2_{j\theta}}\right\}}\mu(\theta)\times$$

$$\left[\int e^{-\{\frac{1}{2}\sum_{j=0}^{d}\langle\sqrt{n}(r_n-Kf_{0\theta}),(\Sigma^*_{1/2})^{-1}\varphi_j\rangle^2\}}\prod_{j>d}\sqrt{\frac{n^{-1}}{\frac{1}{cn}+l^2_{j\theta}}}e^{-\left\{\sum_{j>d}\frac{\langle\sqrt{n}(r_n-Kf_{0\theta}),(\Sigma^*_{1/2})^{-1}\varphi_j\rangle^2}{2(1+cnl^2_{j\theta})}\right\}}\right.$$

$$\left.\times\,\mu(\theta)d\theta\right]^{-1}.$$

By taking the limit for $c\to\infty$ we obtain

$$\mu(\theta|r_n) \to \frac{e^{-\left\{\frac{1}{2}\sum_{j=0}^{d}\langle\sqrt{n}(r_n-Kf_{0\theta}),(\Sigma^*_{1/2})^{-1}\varphi_j\rangle^2\right\}}\mu(\theta)}{\int e^{-\{\frac{1}{2}\sum_{j=0}^{d}\langle\sqrt{n}(r_n-Kf_{0\theta}),(\Sigma^*_{1/2})^{-1}\varphi_j\rangle^2\}}\mu(\theta)d\theta}.$$

Next, remark that for $\Pi = F_*$ and for every $j = 1,\ldots,d$: $K^*(\Sigma^*_{1/2})^{-1}\varphi_j = \varphi_j$ and $(\Sigma^*_{1/2})^{-1}\varphi_j = (K^*)^{-1}\varphi_j$ which is well defined under the assumption of the theorem. Therefore, for $\varphi := (\varphi_1,\ldots,\varphi_d)^T$, $\sqrt{n}\langle Kf_{0\theta},(\Sigma^*_{1/2})^{-1}\varphi\rangle = \sqrt{n}\langle f_{0\theta}(x),V(\theta)^{-1/2}h(\theta,x)\rangle = 0$ by construction of $f_{0\theta}$, and

$$\begin{aligned}\sqrt{n}\langle r_n,(\Sigma^*_{1/2})^{-1}\varphi\rangle &= \sqrt{n}\langle r_n,(K^*)^{-1}\varphi\rangle\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(K^*(K^*)^{-1}\varphi)(x_i) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}V(\theta)^{-1/2}h(\theta,x_i).\end{aligned}$$

Moreover, for $j=0$:

$$\exp\left\{-\frac{1}{2}\langle\sqrt{n}(r_n-Kf_{0\theta}),(\Sigma^*_{1/2})^{-1}\varphi_0\rangle^2\right\} = \exp\left\{-\frac{1}{2}\left(\langle\sqrt{n}r_n,(\Sigma^*_{1/2})^{-1}\varphi_0\rangle-\frac{1}{2}\right)^2\right\}$$

which, since it does not depend on θ, simplifies with the denominator. By putting all these results together, we obtain:

$$\lim_{c\to\infty}\mu(\theta|r_n,c)=\lim_{c\to\infty}\frac{p_{n\theta}(r_n;\theta)\mu(\theta)}{\int p_{n\theta}(r_n;\theta)\mu(\theta)d\theta}\propto\exp\left\{-\frac{1}{2}\sum_{j=1}^{d}n\left(\frac{1}{n}\sum_{i=1}^{n}\varphi_j(x_i)\right)^2\right\}\mu(\theta)$$
$$=\exp\left\{-\frac{1}{2}n\left(\frac{1}{n}\sum_{i=1}^{n}h(\theta,x_i)\right)^T V_*(\theta)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}h(\theta,x_i)\right)\right\}\mu(\theta). \quad \text{(B.5)}$$

Since in practice F_* is unknown, the matrix V_* has to be replaced by its empirical counterpart $V_n(\theta)$.

□

C Proofs for Section 3

Proof of Theorem 3.1

Let Assumption A4 (ii) be verified (for the case where Assumption A4 (i) is satisfied the proof is similar and then omitted). Let $l(\theta)=-\frac{1}{2}\sum_{j=0}^{\infty}\langle K(f_{0\theta}-f_*),\Sigma^{-1/2}\psi_j(\theta)\rangle^2\frac{1}{l_{j\theta}^2}$, $Z_j(\theta):=\langle\sqrt{n}(r_n-Kf_*),\Sigma^{-1/2}\psi_j(\theta)\rangle$ for $j\geq 0$ where $\psi_j(\theta)$ is as defined in Theorem 2.1. We make the following decomposition:

$$|l_n(\theta)-\mu(\theta)-l(\theta)|=\Bigg|\log\mu(\theta)-\frac{1}{2}\sum_{j=0}^{\infty}\log(n^{-1}+l_{j\theta}^2)-\frac{1}{2}\sum_{j=0}^{\infty}Z_j^2(\theta)\frac{1}{1+nl_{j\theta}^2}$$
$$-\frac{1}{2}\sum_{j=0}^{\infty}\langle\sqrt{n}K(f_{0\theta}-f_*),\Sigma^{-1/2}\psi_j(\theta)\rangle^2\frac{1}{1+nl_{j\theta}^2}$$
$$+\sum_{j=0}^{\infty}\langle\sqrt{n}K(f_{0\theta}-f_*),\Sigma^{-1/2}\psi_j(\theta)\rangle Z_j(\theta)\frac{1}{1+nl_{j\theta}^2}-l(\theta)\Bigg|.$$

Because $Z_j=O_p(1)$, $\psi_j(\theta)$ and $l_{j\theta}$ are continuous functions of θ, Θ is compact, then by the Continuous Mapping Theorem applied to (continuous) functions of $Z_j(\theta)$ it follows that: $\sum_{j=0}^{\infty}Z_j^2(\theta)\frac{1}{1+nl_{j\theta}^2}=O_p(n^{-1})$ and $\sum_{j=0}^{\infty}\langle\sqrt{n}K(f_{0\theta}-f_*),\Sigma^{-1/2}\psi_j(\theta)\rangle Z_j(\theta)\frac{1}{1+nl_{j\theta}^2}=O_p(n^{-1/2})$ uniformly in θ. By Assumptions A2, A3, A4 (ii) and by compactness of Θ: $\log\mu(\theta)-\frac{1}{2}\sum_{j=0}^{\infty}\log(n^{-1}+l_{j\theta}^2)\to 0$ uniformly in θ. Moreover,

$$\sum_{j=0}^{\infty}\langle\sqrt{n}K(f_{0\theta}-f_*),\Sigma^{-1/2}\psi_j(\theta)\rangle^2\frac{1}{1+nl_{j\theta}^2}=\sum_{j=0}^{\infty}\langle K(f_{0\theta}-f_*),\Sigma^{-1/2}\psi_j(\theta)\rangle^2\frac{n}{1+nl_{j\theta}^2}$$

and it converges to $\sum_{j=0}^{\infty}\langle K(f_{0\theta}-f_*), \Sigma^{-1/2}\psi_j(\theta)\rangle^2 \frac{1}{l_{j\theta}^2} =: -l(\theta)$ uniformly in θ. This shows that the maximizer of $l_n(\theta)-\mu(\theta)$ converges in probability to the maximizer of $l(\theta)$.

We now show that the maximizer of $l(\theta)$ is θ_*. Let $B = \Sigma^{-1/2}K\Omega_{0\theta}^{1/2}$ and B^* be its adjoint (which exists because B is bounded). First, remark that maximizing $l(\theta)$ is equivalent to minimize $\sum_{j=0}^{\infty}\langle \Sigma^{-1/2}K(f_{0\theta}-f_*), (B^*)^{-1}\rho_j(\theta)\rangle^2$ which in turn is equivalent to find the value of θ such that this expression exists, *i.e.* such that

$$\|B^{-1}\Sigma^{-1/2}K(f_{0\theta}-f_*)\|^2 < \infty. \tag{C.1}$$

Because $B^{-1}\Sigma^{-1/2}K = \Omega_{0\theta}^{-1/2}$, (C.1) is verified if and only if $(f_{0\theta}-f_*) \in \mathcal{R}(\Omega_{0\theta}^{1/2})$. Finally, we have to show that $(f_{0\theta}-f_*) \in \mathcal{R}(\Omega_{0\theta}^{1/2})$ if and only if $\theta = \theta_*$. For this, remark that since $\Omega_{0\theta}^{1/2}$ is a bounded and linear operator:

$$\begin{aligned}\overline{\mathcal{R}(\Omega_{0\theta}^{1/2})} &= \mathfrak{N}(\Omega_{0\theta}^{1/2})^{\perp} \\ &= \left\{\varphi \in \mathcal{E}; \int \varphi(x)h(x,\theta)\Pi(dx) = 0 \text{ and } \int \varphi(x)\Pi(dx) = 0\right\}.\end{aligned}$$

Clearly, $\int(f_{0\theta}-f_*)(x)\Pi(dx) = 0$, but $\int(f_{0\theta}-f_*)(x)h(x,\theta)\Pi(dx) = 0$ if and only if $\int f_*(x)h(x,\theta)\Pi(dx) = 0$. By the identifiability assumption A1, the last equality is satisfied if and only if $\theta = \theta_*$. This shows that the maximizer of $l(\theta)$ is the true θ_*.

□

Proof of Theorem 3.2

Define the events $A_1 := \left\{\sup_{\theta\in B(\theta_*,\delta_n)^c}[l_n(\theta)-l_n(\theta_*)] \leq -CM_n^2\right\}$ and

$$B_1 := \left\{\int_\Theta e^{l_n(\theta)-l_n(\theta_*)}\mu(\theta)d\theta > e^{-CM_n^2/2}\right\}$$

for some sequence $M_n \to \infty$ and a constant $C > 0$ as in Assumptions B1-B2. By these assumptions $P^*(A_1) \to 1$ and $P^*(B_1) \to 1$ as $n \to \infty$. Hence,

$$\begin{aligned}\mathbf{E}^*\int_{B(\theta_*,\delta_nM_n)^c}\mu(\theta|r_n)d\theta &= \mathbf{E}^*\left[\int_{B(\theta_*,\delta_nM_n)^c}\mu(\theta|r_n)d\theta\middle| A_1\right]P^*(A_1) \\ &\quad + \mathbf{E}^*\left[\int_{B(\theta_*,\delta_nM_n)^c}\mu(\theta|r_n)d\theta\middle| A_1^c\right]P^*(A_1^c) \\ &\leq \mathbf{E}^*\left[\frac{\int_{B(\theta_*,\delta_nM_n)^c}\exp[l_n(\theta)-l_n(\theta_*)]\mu(\theta)d\theta}{\int_\Theta \exp[l_n(\theta)-l_n(\theta_*)]\mu(\theta)d\theta}\middle| A_1\right]P^*(A_1) + o(1)\end{aligned}$$

$$\leq \mathbf{E}^* \left[\frac{\int_{B(\theta_*,\delta_n M_n)^c} \exp[l_n(\theta) - l_n(\theta_*)]\mu(\theta)d\theta}{\int_\Theta \exp[l_n(\theta) - l_n(\theta_*)]\mu(\theta)d\theta} \middle| A_1 \cap B_1 \right] P^*(A_1)P^*(B_1) + o(1)$$
$$\leq e^{-CM_n^2} e^{CM_n^2/2} \int_{B(\theta_*,\delta_n M_n)^c} \mu(\theta)d\theta + o(1) = o(1).$$

This shows (3.1). To prove (3.2), first remark that

$$P^* \left(B(\theta_n, \delta_n M_n) \cap B(\theta_*, \delta_n M_n) \neq \varnothing\right) = P^* \left(\|\theta_n - \theta_*\| \leq 2\delta_n M_n\right)$$

which converges to 1, as $n \to \infty$, by Theorem 3.1. Then,

$$\begin{aligned}
\mathbf{E}^*[\mu(B(\theta_n, \delta_n M_n)^c | r_n)] = \mathbf{E}^*[&\int_{B(\theta_n,\delta_n M_n)^c \cap B(\theta_*,\delta_n M_n)} \mu(\theta|r_n)d\theta] \\
&+ \mathbf{E}^*[\int_{B(\theta_n,\delta_n M_n)^c \cap B(\theta_*,\delta_n M_n)^c} \mu(\theta|r_n)d\theta] \\
\leq &\int\int 1\{\theta \in B(\theta_n, \delta_n M_n)^c \cap B(\theta_*, \delta_n M_n)\} dP^*(r_n)\mu(\theta|r_n)d\theta \\
&+ E^*[\mu(B(\theta_*, \delta_n M_n)^c | r_n)] \\
\leq &P^*(B(\theta_n, \delta_n M_n)^c \cap B(\theta_*, \delta_n M_n) \neq \varnothing) + o(1) \\
= &P^*(B(\theta_*, \delta_n M_n) \backslash (B(\theta_n, \delta_n M_n) \cap B(\theta_*, \delta_n M_n)) + o(1) \\
= &P^*(B(\theta_*, \delta_n M_n)) - P^*(B(\theta_n, \delta_n M_n) \cap B(\theta_*, \delta_n M_n)) + o(1) \\
= &o_p(1)
\end{aligned}$$

by the first result of this proof.

□

Proof of Theorem 3.3

This proof follows the proof of Theorem 2.1 in Kleijn and van der Vaart [2012]. For completeness reasons we give here the main steps.

Let $\mathfrak{K} \subset \mathbb{R}^p$ be an arbitrary compact set and $\mu_\tau(\cdot)$ be the prior for the random sequence $\tau = \sqrt{n}(\theta - \theta_*)$ and $\mu_\tau(\cdot|r_n)$ be its posterior. Moreover, we denote by Φ the $\mathcal{N}(\Delta_*, \tilde{I}_*^{-1})$ distribution, by ϕ its density with respect to Lebesgue, by $\Phi^{\mathfrak{K}}$ (resp. $\mu_\tau^{\mathfrak{K}}(\cdot|r_n)$) the conditional version of Φ (resp. $\mu_\tau(\cdot|r_n)$) conditioned on $\mathfrak{K}$. The proof proceeds as follows: we first prove that $\mu_\tau^{\mathfrak{K}}(\cdot|r_n)$ converges to $\Phi^{\mathfrak{K}}$ in total variation and then we use this result to prove (3.4).

Let $U \subset \Theta$ be a neighborhood of θ_*. Then, $\forall U \subset \Theta$ there exists N such that $\forall n > N$:

$\theta_* + \mathfrak{K}n^{-1/2} \subset U$. Let $G_n : \mathfrak{K} \times \mathfrak{K} \to \mathbb{R}$ be the function

$$G_n(\tau, g) := \left(1 - \frac{\phi(\tau)s_n(g)\mu_\tau(g)}{\phi(g)s_n(\tau)\mu_\tau(\tau)}\right)_+$$

which is well-defined $\forall n > N$ since θ_* is an interior point of Θ. By assumption (3.3) we have that, for every random sequence (τ_n) and (g_n), $\mu_\tau(g_n)\mu(\tau_n) \to 1$ as $n \to \infty$ and

$$\log \frac{\phi(\tau)s_n(g)\mu_\tau(g)}{\phi(g)s_n(\tau)\mu_\tau(\tau)} = o_p(1).$$

Hence, $G_n(\tau_n, g_n) \to 0$.

By continuity of G_n in (τ, g) and because Θ is compact: $\sup_{\tau,g\in\mathfrak{K}} G_n(\tau, g) \to 0$ as $n \to 0$. Let $\mathfrak{K}$ contain a neighborhood of 0, so that $\Phi(\mathfrak{K}) > 0$, and define the events $\mathcal{A}_1 := \{\mu_\tau(\mathfrak{K}|r_n) > 0\}$ and $\mathcal{A}_2 := \{\sup_{\tau,g\in\mathfrak{K}} G_n(\tau, g) \leq \eta\}$. Hence, since the TV distance is upper bounded by 2:

$$\begin{aligned}
\mathbf{E}^*\|\mu_\tau^{\mathfrak{K}}(\tau|r_n) - \Phi^{\mathfrak{K}}\|_{TV} 1_{\mathcal{A}_1} &\leq \mathbf{E}^*\|\mu_\tau^{\mathfrak{K}}(\tau|r_n) - \Phi^{\mathfrak{K}}\|_{TV} 1_{\mathcal{A}_1\cap\mathcal{A}_2} + 2P^*(\mathcal{A}_1 \backslash \mathcal{A}_2) \\
&= 2\mathbf{E}^* \int \left(1 - \int_{\mathfrak{K}} \frac{\phi(\tau)s_n(g)\mu_\tau(g)}{\phi(g)s_n(\tau)\mu_\tau(\tau)} d\Phi^{\mathfrak{K}}(g)\right)_+ \mu_\tau^{\mathfrak{K}}(\tau|r_n)d\tau 1_{\mathcal{A}_1\cap\mathcal{A}_1} + o(1) \\
&\leq 2\mathbf{E}^* \int \int_{\mathfrak{K}} G_n(\tau, g) d\Phi^{\mathfrak{K}}(g)\mu_\tau^{\mathfrak{K}}(\tau|r_n)d\tau 1_{\mathcal{A}_1\cap\mathcal{A}_1} + o(1) \\
&\leq 2\mathbf{E}^* \int \sup_{\tau,g\in\mathfrak{K}} G_n(\tau, g) d\Phi^{\mathfrak{K}}(g)\mu_\tau^{\mathfrak{K}}(\tau|r_n)d\tau 1_{\mathcal{A}_1\cap\mathcal{A}_1} + o(1) = o(1)
\end{aligned}$$

where we have used the fact that the function $x \mapsto (1 - x)_+$ is convex and the Jensen's inequality. This concludes the first part of the proof.

Next, we use this result to show (3.4). Let $\mathfrak{K}_n$ be a closed ball centered at 0 with radius $M_n \to \infty$. The corresponding event $\mathcal{A}_1 := \{\mu_\tau(\mathfrak{K}_n|r_n) > 0\}$ has P^*-probability converging to 1 and so, if $M_n \to \infty$ slow enough, $\mathbf{E}^*\|\mu_\tau^{\mathfrak{K}_n}(\tau|r_n) - \Phi^{\mathfrak{K}_n}\|_{TV} 1_{\mathcal{A}_1} \to 0$. Moreover, for this M_n, $\mu_\tau(\mathfrak{K}_n^c|r_n) \to 0$. Finally,

$$\begin{aligned}
&\mathbf{E}^*\|\mu_\tau(\tau|r_n) - \mathcal{N}(\Delta_*, \tilde{I}_*^{-1})\|_{TV} \\
&\quad \leq \mathbf{E}^*\|\mu(\tau|r_n) - \mu_\tau^{\mathfrak{K}_n}(\tau|r_n)\|_{TV} + \mathbf{E}^*\|\mu_\tau^{\mathfrak{K}_n}(\tau|r_n) - \Phi^{\mathfrak{K}_n}\|_{TV} + \mathbf{E}^*\|\Phi^{\mathfrak{K}_n} - \Phi\|_{TV} \\
&\quad \leq 2\mathbf{E}^*\|\mu_\tau(\mathfrak{K}_n^c|r_n)\|_{TV} + \mathbf{E}^*\|\mu_\tau^{\mathfrak{K}_n}(\tau|r_n) - \Phi^{\mathfrak{K}_n}\|_{TV} + 2\mathbf{E}^*\|\Phi(\mathfrak{K}_n^c)\|_{TV} = o(1).
\end{aligned}$$

$\square$

Proof of Theorem 3.4

Denote $s_n := \sqrt{n}(\theta - \hat{\theta})$. Let $\theta^{r_n} = \langle \mathbf{E}[f|r_n], g\rangle$, $\Omega_n = \langle Var[f|r_n]g, g\rangle$ and $\widehat{s}_n := \sqrt{n}(\theta^{r_n} - \hat{\theta})$. Remark that $\mu(s|r_n)$ is a $\mathcal{N}(\sqrt{n}(\theta^{r_n} - \hat{\theta}), n\Omega_n)$ distribution. We can upper bound the TV distance by

$$\begin{aligned}\left\|\mu(s|r_n) - \mathcal{N}(\hat{\theta}, V)\right\|_{TV} &\leq \|\mu(s|r_n) - \mathcal{N}(0, n\Omega_n)\|_{TV} + \|\mathcal{N}(0, n\Omega_n) - \mathcal{N}(0, V)\|_{TV} \\ &=: \mathcal{A} + \mathcal{B}.\end{aligned}$$

We start by considering $\mathcal{A}$. By doing a change of variable it is easy to show that $\mathcal{A} = \left\|\mathcal{N}(0, I_k) - \mathcal{N}(-(n\Omega_n)^{-1/2}\widehat{s}_n, I_k)\right\|_{TV}$. An application of Lemma D.2 and then of Lemmas D.3 and D.4 allows to conclude that $\mathcal{A} \leq \frac{1}{\sqrt{2\pi}}\|(n\Omega_n)^{-1/2}\widehat{s}_n\| = o_p(1)$.
To bound $\mathcal{B}$ we introduce the Kullback-Leibler distance between two probability measures P_1 and P_2, denoted by $K(P_1, P_2)$, and satisfying $K(P_1, P_2) = \int \log \frac{p_1}{p_2} dP_1$ where p_1 and p_2 are the densities of P_1 and P_2, respectively, with respect to the Lebesgue measure. We get

$$\begin{aligned}\mathcal{B} &\leq \sqrt{K\left(\mathcal{N}(0, V), \mathcal{N}(0, n\Omega_n)\right)} \\ &= \sqrt{\int \log\left(\frac{|V|^{-1/2}e^{\{-\frac{1}{2}s^tV^{-1}s\}}}{|n\Omega_n|^{-1/2}e^{\{-\frac{1}{2}s^t(n\Omega_n)^{-1}s\}}}\right)\frac{1}{(2\pi)^{k/2}}|V|^{-1/2}e^{\{-\frac{1}{2}s^tV^{-1}s\}}ds} \\ &= \sqrt{\frac{1}{2}\int\left(\log\frac{|n\Omega_n|}{|V|} - s^t[V^{-1} - (n\Omega_n)^{-1}]s\right)\frac{1}{(2\pi)^{k/2}}|V|^{-1/2}e^{\{-\frac{1}{2}s^tV^{-1}s\}}ds} \\ &= \sqrt{\frac{1}{2}\log\frac{|n\Omega_n|}{|V|} - \frac{1}{2}trV[V^{-1} - (n\Omega_n)^{-1}]}\end{aligned}$$

that converges to zero by the result of Lemma D.3.

□

D Technical Appendix

D.1 Proof of (2.12)

If $\sup_{x\in S}\frac{dF_*(x)}{d\Pi(x)} < \infty$ then, by the invariance property established in Proposition 2.1 we can fix $\Pi = F_*$. Therefore, $f_* = 1$ and we may write: $\Sigma = \Sigma_{1/2}\Sigma_{1/2}^*$ where

$$\begin{aligned}\Sigma_{1/2}: \ &\mathcal{E} \to \mathcal{F} \qquad\qquad & \Sigma_{1/2}^*: \ &\mathcal{F} \to \mathcal{E} \\ &\varphi \mapsto K\varphi - (K1)\langle 1, \varphi\rangle & &\psi \mapsto K^*\psi + \langle K1, \psi\rangle.\end{aligned}$$

Let $(\Sigma_{1/2}^*)^{-1} = (K^*)^{-1} - \frac{1}{2}(K^*)^{-1}\langle 1, \cdot\rangle$, it is easy to verify that $(\Sigma_{1/2}^*)^{-1}\Sigma_{1/2}^* = I$, so that in the following we use the notation $\Sigma^{-1/2} = (\Sigma_{1/2}^*)^{-1} : \mathcal{E} \to \mathcal{F}$. Hence, $\Omega_{0\theta}^{1/2} K^*(\Sigma_{1/2}^*)^{-1}\varphi_l = \lambda_l^{1/2}\varphi_l$ for every $l > d$ and $\Omega_{0\theta}^{1/2} K^*(\Sigma_{1/2}^*)^{-1}\varphi_l = 0$ for every $l \leq d$. This shows that $l_{j\theta}^2 = \lambda_j$, $\forall j \geq 0$. By replacing this in (2.9) and simplifying the terms that do not depend on θ gives the result.

□

D.2 Primitive conditions for Assumption B2

Lemma D.1. *Let Assumption A2 be satisfied and $\sum_{j=1}^{\infty} |dl_{j\theta}/d\theta| < \infty$. Then, there exists a constant $C > 0$ such that for any sequence $M_n \to \infty$,*

$$P^*\left(\int_\Theta e^{l_n(\theta)-l_n(\theta_*)}\mu(\theta)d\theta \leq e^{-CM_n^2/2}\right) \to 0 \qquad \text{as } n \to \infty.$$

Proof. Let μ_τ be the prior for the random sequence $\tau = \sqrt{n}(\theta - \theta_*)$ with support $\mathcal{T}$ and define $S_n(\tau) := \exp\{l_n(\theta_* + n^{-1/2}\tau) - l_n(\theta_*)\}$. A second order Taylor expansion around $\tau = 0$ gives:

$$\log S_n(\tau) = \frac{\dot{S}_n(0)}{S_n(0)}^T \tau - \frac{1}{2}\tau^T \frac{[\dot{S}_n(0)]^2 - \ddot{S}_n(0)S_n(0)}{S_n^2(0)}\tau + o(\|\tau\|) \tag{D.1}$$

where $\dot{S}_n(0)$ (resp. $\ddot{S}_n(0)$) denote the first (resp. the second) derivative of S_n evaluated at $\tau = 0$. Simple algebra allows to show that $\frac{[\dot{S}_n(0)]^2 - \ddot{S}_n(0)S_n(0)}{S_n^2(0)} = -\left.\frac{d^2 l_n(\theta)}{d\theta d\theta^T}\right|_{\theta=\theta_*} n^{-1}$. By the Markov inequality:

$$P^*\left(\left.\frac{d^2 l_n(\theta)}{d\theta d\theta^T}\right|_{\theta=\theta_*} n^{-1} > \frac{M_n^2}{2}\right) \leq \frac{2}{nM_n^2}\mathbf{E}^*\left|\left.\frac{d^2 l_n(\theta)}{d\theta d\theta^T}\right|_{\theta=\theta_*}\right| \tag{D.2}$$

which converges to 0 as $n \to \infty$. This implies: $-\tau^T \frac{[\dot{S}_n(0)]^2 - \ddot{S}_n(0)S_n(0)}{S_n^2(0)}\tau \geq -\|\tau\|^2 \frac{M_n^2}{2}$. By defining $\mathcal{T}_C := \{\tau; \|\tau\| \leq \sqrt{C}\}$ we have:

$$\begin{aligned}
P^*\left(\int_\Theta \exp\{l_n(\theta) - l_n(\theta_*)\}\mu(\theta)d\theta \leq e^{-CM_n^2/2}\right) &= P^*\left(\int_{\mathcal{T}} S_n(\tau)\mu_\tau(\tau)d\tau \leq e^{-CM_n^2/2}\right) \\
&\leq P^*\left(\int_{\mathcal{T}_C} \exp\left\{\frac{\dot{S}_n(0)}{S_n(0)}^T \tau - CM_n^2/4\right\}\mu_\tau(\tau)d\tau \leq e^{-CM_n^2/2}\right) \\
\leq P^*\left(\exp\left\{\int_{\mathcal{T}_C} \frac{\dot{S}_n(0)}{S_n(0)}^T \tau\mu_\tau(\tau)d\tau\right\} \leq e^{-CM_n^2/4}\right) &= P^*\left(\frac{\dot{S}_n(0)}{S_n(0)}^T \int_{\mathcal{T}_C} \tau\mu_\tau(\tau)d\tau \leq -\frac{CM_n^2}{4}\right)
\end{aligned}$$

$$\leq \frac{4}{CM_n^2}\mathbf{E}^*\left\|\frac{\dot{S}_n(0)}{S_n(0)}\right\|\left|\int_{\mathcal{T}_C}\tau\mu_\tau(\tau)d\tau\right| \quad \text{(D.3)}$$

where the inequality in the third line follows from (D.1) and (D.2), the inequality in the forth line follows from the Jensen's inequality and the inequality in the last line follows from the Chebishev's inequality. Simple algebra allows to show that $\mathbf{E}^*\left\|\frac{\dot{S}_n(0)}{S_n(0)}\right\| = \mathbf{E}^*\|dl_n(\theta_*)/d\theta\|n^{-1/2} \to 0$ as $n \to \infty$ if $\sum_j |dl_{j\theta}/d\theta| < \infty$. Moreover, $\left|\int_{\mathcal{T}_C}\tau\mu_\tau(\tau)d\tau\right|^2 = O(n^{-1/2})$ so that the last term of (D.3) converges to 0.

□

D.3 Proof of (3.3)

In this section we prove the integral local asymptotic normality (3.3). Remark that we may write: $\Sigma = \Sigma_{1/2}\Sigma_{1/2}^*$ where $\Sigma_{1/2} : \mathcal{E} \to \mathcal{F}$ and $\Sigma_{1/2}^* : \mathcal{F} \to \mathcal{E}$ are as defined in (B.4). Remark that, despite the notation, $\Sigma_{1/2}^*$ is not the adjoint of $\Sigma_{1/2}$. Let $(\Sigma_{1/2}^*)^{-1} = (f_*^{1/2}K^*)^{-1} - \frac{1}{2}(K^*)^{-1}\langle f_*^{1/2}, \cdot\rangle$, it is easy to verify that $(\Sigma_{1/2}^*)^{-1}\Sigma_{1/2}^* = I$, so that in the following we use the notation $\Sigma^{-1/2} = (\Sigma_{1/2}^*)^{-1} : \mathcal{E} \to \mathcal{F}$. Hence, $\Omega_{0\theta}^{1/2}K^*(\Sigma_{1/2}^*)^{-1}f_*^{1/2}\varphi_l = \lambda_l^{1/2}\varphi_l$ for every $l > d$ and $\Omega_{0\theta}^{1/2}K^*(\Sigma_{1/2}^*)^{-1}f_*^{1/2}\varphi_l = 0$ for every $l \leq d$. We want to determine the functions $\psi_j(\theta)$ used in (2.9). These are the eigenfunctions of

$$[\Omega_{0\theta}^{1/2}K^*(\Sigma_{1/2}^*)^{-1}]^*\Omega_{0\theta}^{1/2}K^*(\Sigma_{1/2}^*)^{-1}$$

where $[\Omega_{0\theta}^{1/2}K^*(\Sigma_{1/2}^*)^{-1}]^*$ denotes the adjoint of the operator $\Omega_{0\theta}^{1/2}K^*(\Sigma_{1/2}^*)^{-1}$ which exists since $\Omega_{0\theta}^{1/2}K^*(\Sigma_{1/2}^*)^{-1}$ is bounded if $f_*^{1/2}$ is bounded away from 0. For every $\phi_1, \phi_2 \in \mathcal{E}$:

$$\langle\Omega_{0\theta}^{1/2}K^*(\Sigma_{1/2}^*)^{-1}\phi_1, \phi_2\rangle = \langle\phi_1, \sum_{j>d}\lambda_j^{1/2}\langle\varphi_j, \phi_2\rangle f_*^{-1/2}\varphi_j\rangle$$

which gives an expression for $[\Omega_{0\theta}^{1/2}K^*(\Sigma_{1/2}^*)^{-1}]^*$. By using this result it is easy to check that for every $l > d$: $[\Omega_{0\theta}^{1/2}K^*(\Sigma_{1/2}^*)^{-1}]^*\Omega_{0\theta}^{1/2}K^*(\Sigma_{1/2}^*)^{-1}f_*^{1/2}\varphi_l = \lambda_l f_*^{-1/2}\varphi_l$. By Proposition 2.1, our inference procedure is invariant to the choice of Π. Then, if $\sup_{x\in S} f_*(x) < \infty$ we can fix $\Pi = F_*$ so that $f_* = 1$ and the eigenfunctions $\{\psi_j(\theta)\}_{j\geq 1}$ are given by $\{\varphi_j\}_{j\geq 1}$ (which depend on θ) with corresponding eigenvalues $\{l_{j\theta}\}_{j\geq 1} = \{\lambda_j 1\{j > d\}\}_{j\geq 1}$ (which do not depend on θ).

Let us consider the function $s_n(\tau) = p_{n,\theta_*+\delta_n\tau}(r_n; \theta_* + \delta_n\tau)$ which is the localized integrated likelihood:

$$s_n(\tau) = \int \frac{dP^f(\sqrt{n}r_n)}{dP^{f_*}(\sqrt{n}r_n)}d\mu(f|\theta_* + n^{-1/2}\tau).$$

Its logarithm is equal to (by using (2.9))

$$\log s_n(\tau) = -\frac{1}{2}\sum_{j>d}\log(1+n\lambda_j^2) - \frac{1}{2}\sum_{j=1}^{d}\langle\sqrt{n}(r_n - Kf_{0(\theta_*+\delta_n\tau)}),(\Sigma_{1/2}^*)^{-1}\varphi_j\rangle^2$$
$$-\frac{1}{2}\sum_{j>d}\langle\sqrt{n}(r_n - Kf_{0(\theta_*+\delta_n\tau)}),(\Sigma_{1/2}^*)^{-1}\varphi_j\rangle^2\frac{1}{1+n\lambda_j^2} + \frac{1}{2}\|\sqrt{n}(r_n - Kf_*)\|_\Sigma^2$$

where we have left implicit the dependence of φ_j on τ, and

$$\log s_n(0) = -\frac{1}{2}\sum_{j>d}\log(1+n\lambda_j^2) - \frac{1}{2}\sum_{j=1}^{d}\langle\sqrt{n}(r_n - Kf_{0\theta_*}),(\Sigma_{1/2}^*)^{-1}\varphi_j\rangle^2$$
$$-\frac{1}{2}\sum_{j>d}\langle\sqrt{n}(r_n - Kf_{0\theta_*}),(\Sigma_{1/2}^*)^{-1}\varphi_j\rangle^2\frac{1}{1+n\lambda_j^2} + \frac{1}{2}\|\sqrt{n}(r_n - Kf_*)\|_\Sigma^2.$$

A second order Taylor expansion of $\log s_n(\tau)$ around $\tau = 0$ gives (recall that $\delta_n = 1/\sqrt{n}$):

$$\log\frac{s_n(\tau)}{s_n(0)} = \frac{1}{2}\frac{\partial\sum_{j=1}^{d}\langle\sqrt{n}(r_n - Kf_{0(\theta_*+\delta_n\tau)}),(\Sigma_{1/2}^*)^{-1}\varphi_j\rangle^2}{\partial\theta^T}\Bigg|_{\tau=0}\tau\frac{1}{\sqrt{n}}$$
$$\frac{1}{2}\tau^T\left[\frac{\partial^2\sum_{j=1}^{d}\langle\sqrt{n}(r_n - Kf_{0(\theta_*+\delta_n\tau)}),(\Sigma_{1/2}^*)^{-1}\varphi_j\rangle^2}{\partial\theta\partial\theta^T}\Bigg|_{\tau=0}\right]\tau\frac{1}{n} + o_p(1) \quad \text{(D.4)}$$

where the first and second derivatives of $\sum_{j>d}\langle\sqrt{n}(r_n - Kf_{0(\theta_*+\delta_n\tau)}),(\Sigma_{1/2}^*)^{-1}\varphi_j\rangle^2\frac{1}{1+n\lambda_j^2}$ with respect to θ converges to 0 in probability. Remark that $f_{0\theta}$ depends on θ implicitly through the equation $\int h_j(\theta,x)f_{0\theta}\Pi(dx) = 0$. To compute the derivative we have first to compute the Gâteaux derivative in the direction of f_* and then compute the derivative with respect to θ. Hence,

$$\frac{d\sum_{j=1}^{d}\langle\sqrt{n}(r_n - Kf_{0(\theta_*+\delta_n\tau)}),(\Sigma_{1/2}^*)^{-1}\varphi_j\rangle^2}{d\theta^T} =$$
$$\frac{d}{d\theta}\frac{d\sum_{j=1}^{d}\langle\sqrt{n}(r_n - K(f_* + \gamma(f_{0(\theta_*+\delta_n\tau)} - f_*))),(\Sigma_{1/2}^*)^{-1}\varphi_j\rangle^2}{d\gamma}\Bigg|_{\gamma=0}$$
$$+2\sum_{j=1}^{d}\langle\sqrt{n}(r_n - Kf_*),(\Sigma_{1/2}^*)^{-1}\varphi_j\rangle\langle\sqrt{n}(r_n - Kf_*),(\Sigma_{1/2}^*)^{-1}d\varphi_j/d\theta\rangle$$
$$= 2\sum_{j=1}^{d}\langle\sqrt{n}(r_n - Kf_*),(\Sigma_{1/2}^*)^{-1}\varphi_j\rangle\langle\sqrt{n}K\left[(f_{0\theta_*} - f_*)\dot{f}_{0\theta_*}\right]^{\perp},(\Sigma_{1/2}^*)^{-1}\varphi_j\rangle + o_p(\sqrt{n}) \quad \text{(D.5)}$$

where $\dot{f}_{0\theta_*} = df_{0\theta}/d\theta|_{\theta=\theta_*}$, $\left[(f_{0\theta_*} - f_*)\dot{f}_{0\theta_*}\right]^\perp \in \mathcal{R}(\Omega_{0\theta_*})^\perp \cap \mathfrak{D}$ and

$$\left.\frac{d^2 \sum_{j=1}^d \langle \sqrt{n}(r_n - Kf_{0(\theta_*+\delta_n\tau)}), (\Sigma_{1/2}^*)^{-1}\varphi_j\rangle^2}{d\theta d\theta^T}\right|_{\tau=0}$$
$$= -n\sum_{j=1}^d \left\langle \frac{(f_{0\theta_*} - f_*)\dot{f}_{0\theta_*}}{f_*^{1/2}}, \varphi_j \right\rangle \left\langle \frac{(f_{0\theta_*} - f_*)\dot{f}_{0\theta_*}^T}{f_*^{1/2}}, \varphi_j \right\rangle + o_p(n)$$
$$= -n\sum_{j=1}^d \left\langle \frac{[(f_{0\theta_*} - f_*)\dot{f}_{0\theta_*}]^\perp}{f_*^{1/2}}, \varphi_j \right\rangle \left\langle \frac{[(f_{0\theta_*} - f_*)\dot{f}_{0\theta_*}^T]^\perp}{f_*^{1/2}}, \varphi_j \right\rangle + o_p(n) =: -n\tilde{I}_* + o_p(n). \tag{D.6}$$

Remark that in (D.5) and (D.6) we have used the fact that since $(f_{0\theta_*} - f_*)\dot{f}_{0\theta_*} \in \mathfrak{D}$ and $(f_{0\theta_*} - f_*)\dot{f}_{0\theta_*} = [(f_{0\theta_*} - f_*)\dot{f}_{0\theta_*}]^\perp + [(f_{0\theta_*} - f_*)\dot{f}_{0\theta_*}]^\circ$ with $[(f_{0\theta_*} - f_*)\dot{f}_{0\theta_*}]^\circ \in \mathcal{R}(\Omega_{0\theta_*})$ then for $f_* = 1$:

$$\langle K(f_{0\theta_*} - f_*)\dot{f}_{0\theta_*}, (\Sigma_{1/2}^*)^{-1}\varphi_j\rangle = \left\langle \frac{\left[(f_{0\theta_*} - f_*)\dot{f}_{0\theta_*}\right]^\perp}{f_*^{1/2}}, \varphi_j \right\rangle, \qquad j = 1, \ldots, d.$$

By replacing (D.5) and (D.6) in (D.4) we get

$$\log\frac{s_n(\tau)}{s_n(0)} = \tau^T \sum_{j=1}^d \langle \sqrt{n}(r_n - Kf_*), (\Sigma_{1/2}^*)^{-1}\varphi_j\rangle \langle f_*^{-1/2}\left[(f_{0\theta_*} - f_*)\dot{f}_{0\theta_*}\right]^\perp, \varphi_j\rangle - \tau^T\tilde{I}_*\tau + o_p(1). \tag{D.7}$$

To show that $\tilde{I}_*$ is equal to the inverse of the asymptotic variance of the GMM estimator remark that the derivative of the moment restriction $\int h_j(\theta, x)f_{0\theta}(x)\Pi(dx) = 0$ with respect to θ gives for every $j = 1, \ldots, d$:

$$\int \frac{\partial h_j(\theta, x)}{\partial\theta}\Big(f_*(x) + \gamma(f_{0\theta}(x) - f_*(x))\Big)\Pi(dx) +$$
$$\left.\int h_j(\theta, x)(f_{0\theta}(x) - f_*(x))\dot{f}_{0\theta}(x)\Pi(dx)\right|_{\gamma=0} = 0$$
$$\leftrightarrow \quad \int \frac{\partial h_j(\theta, x)}{\partial\theta} f_*(x)\Pi(dx) = -\int h_j(\theta, x)(f_{0\theta}(x) - f_*(x))\dot{f}_{0\theta}(x)\Pi(dx)$$
$$\leftrightarrow \quad \int \frac{\partial h_j(\theta, x)}{\partial\theta} f_*(x)\Pi(dx) = -\int h_j(\theta, x)\left[(f_{0\theta}(x) - f_*(x))\dot{f}_{0\theta}(x)\right]^\perp \Pi(dx) \tag{D.8}$$

by using the Gâteau derivative in the direction of f_*. Therefore, from (D.6), (D.8), $\Pi = F_*$ and $\varphi_j = h_j$ for $j = 1, \dots d$, it follows that

$$\begin{aligned} -\tilde{I}_* &:= -\sum_{j=1}^{d} \left\langle \frac{[(f_{0\theta_*} - f_*)\dot{f}_{0\theta_*}]^{\perp}}{f_*^{1/2}}, \varphi_j \right\rangle \left\langle \frac{[(f_{0\theta_*} - f_*)\dot{f}_{0\theta_*}^{T}]^{\perp}}{f_*^{1/2}}, \varphi_j \right\rangle \\ &= -\mathbf{E}^* \left[\frac{\partial h(\theta_*, x)}{\partial \theta} \right] \left[\mathbf{E}^* h(\theta_*, x) h(\theta_*, x)^T \right]^{-1} \mathbf{E}^* \left[\frac{\partial h(\theta_*, x)}{\partial \theta^T} \right]. \end{aligned}$$

□

D.4 Technical Lemmas

The next lemmas apply to the just identified case described in Remark 2.2 where the prior covariance operator does not depend on θ and for which we use the notation Ω_0.

Lemma D.2. *Let $\Omega_0 : \mathcal{E} \to \mathcal{E}$ be a covariance operator of a $\mathcal{GP}$ on $\mathfrak{B}_{\mathcal{E}}$ such that $\Omega_0^{1/2} 1 = 0$ and all the eigenvalues of Ω_0 but the first one are different from 0. Let $\mathfrak{D} \in \mathcal{E}$ be defined as $\mathfrak{D} := \left\{ g \in \mathcal{E};\ \int g(x) \Pi(dx) = 0 \right\}$. Then, $\mathcal{R}(\Omega_0) = \mathfrak{D}$.*

Proof. Because $\Omega_0 = \Omega_0^{1/2} \Omega_0^{1/2}$ and because $\Omega_0^{1/2} 1 = 0$ then $\Omega_0 1 = 0$, that is, $1 \in \mathfrak{N}(\Omega_0)$. Hence, if $g \in \mathcal{R}(\Omega_0)$ then $\exists \nu \in \mathcal{E}$ such that $g = \Omega_0 \nu$ and so

$$\int g(x) \Pi(dx) \quad = \quad \int \Omega_0 \nu \Pi(dx) =< \nu, \Omega_0 1 >= 0.$$

This shows that $\mathcal{R}(\Omega_0) \subset \mathfrak{D}$. Now, take $g \in \mathfrak{D}$ (so, $< g, 1 >= 0$) and suppose that $g \notin \mathcal{R}(\Omega_0)$. Then, $\forall h \in \mathcal{R}(\Omega_0)$: $< g, h >= 0 =< g, 1 >$ and $< g, h - 1 >= 0$. Because the same reasoning holds for every $g \in \mathfrak{D}$, then: $< g, h - 1 >= 0$ for every $g \in \mathfrak{D}$ and for every $h \in \mathcal{R}(\Omega_0)$. Hence, it must be $h = 1$ but this is impossible since $1 \notin \mathcal{R}(\Omega_0)$. Therefore, g must belong to $\mathcal{R}(\Omega_0)$ and so $\mathfrak{D} \subset \mathcal{R}(\Omega_0)$.

□

Lemma D.3. *Let Ω_n be defined as in (3.5) and $V = \mathbf{E}^*[(g - \mathbf{E}^*(g))(g - \mathbf{E}^*(g))^T]$. Then, under the assumptions of Theorem 3.4: (i) $n\Omega_n = O(1)$ and (ii) $n\Omega_n \to V$, as $n \to \infty$.*

Proof. Under the conditions of the theorem the posterior variance writes as in (A.2) with the operator A defined in Lemma A.1. By using this expression:

$$\begin{aligned} \Omega_n &= \langle \Omega_0 - \Omega_0 f_*^{-1/2} \left(\frac{1}{n} I - \frac{1}{n} f_*^{1/2} \langle f_*^{1/2}, \cdot \rangle + f_*^{-1/2} \Omega_0 f_*^{-1/2} \right)^{-1} f_*^{-1/2} \Omega_0 g, g^t \rangle \\ &= \left\langle \left[I - \Omega_0 f_*^{-1/2} \left(\frac{1}{n} f_*^{1/2} - \frac{1}{n} f_* \langle f_*^{1/2}, \cdot \rangle + \Omega_0 f_*^{-1/2} \right)^{-1} \right] \Omega_0 g, g^t \right\rangle \end{aligned}$$

$$= \left\langle \left[\frac{1}{n}f_*^{1/2} - \frac{1}{n}f_*\langle f_*^{1/2},\cdot\rangle\right]\left(\frac{1}{n}f_*^{1/2} - \frac{1}{n}f_*\langle f_*^{1/2},\cdot\rangle + \Omega_0 f_*^{-1/2}\right)^{-1}\Omega_0 g, g^t\right\rangle$$

$$= \left\langle \left[\frac{1}{n}f_* - \frac{1}{n}f_*\langle f_*,\cdot\rangle\right]\left(\frac{1}{n}f_* - \frac{1}{n}f_*\langle f_*,\cdot\rangle + \Omega_0\right)^{-1}\Omega_0 g, g^t\right\rangle$$

$$= \frac{1}{n}\left\langle T\left(\frac{1}{n}T + \Omega_0\right)^{-1}\Omega_0 g, g^t\right\rangle$$

where $T : \mathcal{E} \to \mathcal{E}$ is the self-adjoint operator defined as $T\phi = f_*(\phi - \mathbf{E}^*\phi)$, $\forall\phi \in \mathcal{E}$. This shows (i). To show (ii), by using the previous expression for Ω_n and the definition of T, we write

$$\begin{aligned} V - n\Omega_n &= -\langle T[(\frac{1}{n}T + \Omega_0)^{-1}\Omega_0 - I]g, g^t\rangle \\ &= -\langle \Omega_0^{-1/2}(\frac{1}{n}\Omega_0^{-1/2}T\Omega_0^{-1/2} + I)^{-1}\frac{1}{n}\Omega_0^{-1/2}Tg, Tg^t\rangle \\ &= \frac{1}{n}\langle(\frac{1}{n}\Omega_0^{-1/2}T\Omega_0^{-1/2} + I)^{-1}\nu, \nu^t\rangle \end{aligned}$$

since T is self-adjoint and since there exists $\nu \in \mathcal{E}$ such that $Tg = \Omega_0^{1/2}\nu$. This is because $Tg = f_*(g - \mathbf{E}_* g) \in \mathcal{R}(\Omega_0^{1/2})$. Finally, because $(\frac{1}{n}\Omega_0^{-1/2}T\Omega_0^{-1/2} + I)^{-1}$ is bounded then $n\Omega_n \to V$ as $n \to \infty$.

□

Lemma D.4. *Let $\hat\theta$ be as defined in Theorem 3.4 and θ^{r_n} be as defined in* (3.5). *Then, under the assumptions of Theorem 3.4: $\sqrt{n}(\theta^{r_n} - \hat\theta) = o_p(1)$.*

Proof. We want to show that $\sqrt{n}(\theta^{r_n} - \hat\theta) \to 0$. Define $T : \mathcal{E} \to \mathcal{E}$ as the operator $T\phi = f_*(\phi - \mathbf{E}^*\phi)$, $\forall\phi \in \mathcal{E}$. Remark that T is self-adjoint and that $\mathcal{R}(T) \subset \mathcal{R}(\Omega_0)$ so that $\Omega_0^{-1}T$ is well defined.

$$\begin{aligned} \sqrt{n}(\theta^{r_n} - \hat\theta) = \sqrt{n}\Big(&-b(\mathbb{P}_n) \\ &+ \left\langle f_0 + \Omega_0 f_*^{-1/2}\left(\frac{1}{n}I - \frac{1}{n}f_*^{1/2}\langle f_*^{1/2},\cdot\rangle + f_*^{-1/2}\Omega_0 f_*^{-1/2}\right)^{-1} f_*^{-1/2}K^{-1}(r_n - Kf_0), g\right\rangle\Big) \\ &= \sqrt{n}\left\langle\left(I - \Omega_0\left(\frac{1}{n}T + \Omega_0\right)^{-1}\right)f_0, g\right\rangle \\ &+ \sqrt{n}\left(\langle\Omega_0\left(\frac{1}{n}T + \Omega_0\right)^{-1}K^{-1}r_n, g\rangle - b(\mathbb{P}_n)\right). \end{aligned} \quad (D.9)$$

We consider these two terms separately. Consider the first term:

$$\sqrt{n}\langle\left(I-\Omega_0\left(\frac{1}{n}T+\Omega_0\right)^{-1}\right)f_0,g\rangle=\sqrt{n}\langle\frac{1}{n}T\left(\frac{1}{n}T+\Omega_0\right)^{-1}f_0,g\rangle$$
$$=\frac{1}{\sqrt{n}}\left\langle\left(\frac{1}{n}(\Omega_0^{-1}T)^*+I\right)^{-1}f_0,\Omega_0^{-1}Tg\right\rangle=O(n^{-1/2})$$

since $(\Omega_0^{-1}T)^*$ exists because $\Omega_0^{-1}T$ is bounded and $\langle\left(\frac{1}{n}(\Omega_0^{-1}T)^*+I\right)^{-1}f_0,\Omega_0^{-1}Tg\rangle=O(1)$.

Consider now the second term in (D.9) and remark that $b(\mathbb{P}_n)=\frac{1}{n}\sum_{i=1}^n[(\frac{1}{n}T+\Omega_0)^{-1}\Omega_0 g](x_i)+\frac{1}{n}\sum_{i=1}^n(g(x_i)-[(\frac{1}{n}T+\Omega_0)^{-1}\Omega_0 g](x_i))$. Let δ_{x_i} be the Dirac measure that assigns a unit mass to the point x_i. Hence, it is possible to identify such a measure with a distribution D_i, namely, a linear functional D_i defined on $\mathcal{C}^\infty$ and continuous with respect to the supremum norm: $g\mapsto g(x_i)=D_i(g)$, see Schwartz [1966]. We get

$$\frac{1}{n}\sum_{i=1}^n(g(x_i)-[(\frac{1}{n}T+\Omega_0)^{-1}\Omega_0 g](x_i))=\frac{1}{n}\sum_{i=1}D_i\left(g-\left(\frac{1}{n}T+\Omega_0\right)^{-1}\Omega_0 g\right)$$
$$=\frac{1}{n}\sum_{i=1}D_i\left(\frac{1}{n}\left(\frac{1}{n}T+\Omega_0\right)^{-1}Tg\right)$$
$$=\frac{1}{n^2}\sum_{i=1}D_i\left(\left(\frac{1}{n}\Omega_0^{-1}T+I\right)^{-1}\Omega_0^{-1}Tg\right)=O_p(n^{-1})$$

since $\Omega_0^{-1}T$ is bounded. By using the expression for the operator A given in Lemma A.1 we can write the second term in (D.9) as

$$\sqrt{n}\left(\langle\Omega_0\left(\frac{1}{n}T+\Omega_0\right)^{-1}K^{-1}r_n,g\rangle-b(\mathbb{P}_n)\right)=\sqrt{n}\Big(\langle Ar_n,g\rangle$$
$$-\frac{1}{n}\sum_{i=1}^n\left[\left(\frac{1}{n}T+\Omega_0\right)^{-1}\Omega_0 g\right](x_i)-\frac{1}{n}\sum_{i=1}^n\left(g(x_i)-[\left(\frac{1}{n}T+\Omega_0\right)^{-1}\Omega_0 g](x_i)\right)\Big)$$
$$=\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n\left(\langle k(t,x_i),A^*g\rangle-[\left(\frac{1}{n}T+\Omega_0\right)^{-1}\Omega_0 g](x_i)\right)\right)+O_p(n^{-1/2})$$

because A is bounded by Lemma A.1 and by the result in the previous display. Remark that

$$\langle k(t,x_i),A^*g\rangle=(K^*A^*g)(x_i)=((AK)^*g)(x_i)$$
$$=\left(\left(\Omega_0\left(\frac{1}{n}T+\Omega_0\right)^{-1}\right)^*g\right)(x_i)=\left(\left(\left(\frac{1}{n}(\Omega_0^{-1}T)^*+I\right)^{-1}\right)^*g\right)(x_i)$$

$$= \left(\left(\frac{1}{n} \Omega_0^{-1} T + I \right)^{-1} g \right)(x_i).$$

By replacing this result in the previous expression we get:

$$\sqrt{n} \left(\langle \Omega_0 \left(\frac{1}{n} T + \Omega_0 \right)^{-1} K^{-1} r_n, g \rangle - b(\mathbb{P}_n) \right)$$
$$= \sqrt{n} \Big(\frac{1}{n} \sum_{i=1}^{n} \left(\left(\left(\frac{1}{n} \Omega_0^{-1} T + I \right)^{-1} g \right)(x_i) - \left(\left(\frac{1}{n} \Omega_0^{-1} T + I \right)^{-1} g \right)(x_i) \right) \Big) + O_p(n^{-1/2})$$
$$= 0 + O_p(n^{-1/2}).$$

□

References

P. J. Bickel and B. J. K. Kleijn. The semiparametric bernstein–von mises theorem. *The Annals of Statistics*, 40(1):206–237, 2012.

L. Bornn, N. Shephard, and R. Solgi. Moment conditions and bayesian nonparametrics. Technical report, arXiv:1507.08645, 2015.

G. Chamberlain and G. W. Imbens. Nonparametric applications of bayesian inference. *Journal of Business & Economic Statistics*, 21(1):12–18, 2003.

V. Chernozhukov and H. Hong. An mcmc approach to classical estimation. *Journal of Econometrics*, 115(2):293–346, 2003.

J.-P. Florens and J.-M. Rolin. Bayes, bootstrap, moments. Technical report, Université Catholique de Louvain, Institut de Statistique DP9413, 1994.

J.-P. Florens and A. Simoni. Regularized posteriors in linear ill-posed inverse problems. *Scandinavian Journal of Statistics*, 39(2):214–235, 2012.

J.-P. Florens and A. Simoni. Regularizing priors for linear inverse problems. *Econometric Theory*, FirstView, 11 2014.

J.-P. Florens, M. Mouchart, and J.-M. Rolin. *Elements of Bayesian statistics*. Dekker, 1990.

A. R. Gallant. Reflections on the probability space induced by moment conditions with implications for bayesian inference. *Journal of Financial Econometrics*, 2015.

A. R. Gallant, R. Giacomini, and G. Ragusa. Bayesian estimation of state space models using moment conditions. Technical report, 2015.

L. P. Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50:1029–1054, 1982.

L. P. Hansen and K. J. Singleton. Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica*, 50(5):1269–1286, 1982.

L. P. Hansen, J. Heaton, and A. Yaron. Finite-sample properties of some alternative gmm estimators. *Journal of Business & Economic Statistics*, 14(3):262–280, 1996.

G. W. Imbens, R. H. Spady, and P. Johnson. Information theoretic approaches to inference in moment condition models. *Econometrica*, 66(2):333–357, 1998.

J.-Y. Kim. Limited information likelihood and bayesian analysis. *Journal of Econometrics*, 107(1-2):175–193, 2002.

Y. Kitamura. Empirical likelihood methods with weakly dependent processes. *Ann. Statist.*, 25(5):2084–2102, 1997.

Y. Kitamura. Empirical likelihood methods in econometrics: Theory and practice. In R. Blundell, W. Newey, and T. Persson, editors, *Advances in Economics and Econometrics*, volume 3, pages 174–237. Cambridge University Press, 2007. Cambridge Books Online.

Y. Kitamura and T. Otsu. Bayesian analysis of moment condition models using nonparametric priors. Technical report, Department of Economics: Yale University, 2011.

Y. Kitamura and M. Stutzer. An information-theoretic alternative to generalized method of moments estimation. *Econometrica*, 65(4):861–874, 1997.

B. Kleijn and A. van der Vaart. The bernstein-von-mises theorem under misspecification. *Electron. J. Statist.*, 6:354–381, 2012.

R. Kress. *Linear Integral Equations*. Springer New York, 1999.

H.-H. Kuo. *Gaussian Measures in Banach Spaces*. Springer Berlin Heidelberg, 1975.

Y. K. Kwan. Asymptotic bayesian analysis based on a limited information estimator. *Journal of Econometrics*, 88(1):99–121, 1999.

N. A. Lazar. Bayesian empirical likelihood. *Biometrika*, 90(2):319–326, 2003.

E. Lehmann and G. Casella. *Theory of point estimation*. Springer New York, 2nd edition, 1998.

Y. Liao and W. Jiang. Posterior consistency of nonparametric conditional moment restricted models. *The Annals of Statistics*, 39(6):pp. 3003–3031, 2011.

W. K. Newey and R. J. Smith. Higher order properties of gmm and generalized empirical likelihood estimators. *Econometrica*, 72(1):219–255, 2004.

A. B. Owen. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2):237–249, 1988.

J. Qin and J. Lawless. Empirical likelihood and general estimating equations. *The Annals of Statistics*, 22(1):300–325, 03 1994.

G. Ragusa. Bayesian likelihoods for moment condition models. Working Papers 060714, University of California-Irvine, Department of Economics, 2007.

C. Robert. Méthodes de calcul en analyse bayésienne. In J. F. J.J. Droesbeke and G. Saporta, editors, *Méthodes Bayésiennes en statistique*, pages 149 – 181. TECHNIP, 2002.

S. M. Schennach. Bayesian exponentially tilted empirical likelihood. *Biometrika*, 92(1): 31–46, 2005.

S. M. Schennach. Point estimation with exponentially tilted empirical likelihood. *The Annals of Statistics*, 35(2):634–672, 2007.

L. Schwartz. *Théorie des distributions*. Hermann Éditeurs, 1966.

M. Shin. Bayesian GMM. Technical report, University of Pennsylvania, 2014.

R. J. Smith. Alternative semi-parametric likelihood approaches to generalised method of moments estimation. *The Economic Journal*, 107(441):503–519, 1997.

A. van der Vaart and J. Wellner. *Weak convergence and empirical processes*. Springer, the first edition edition, 1996.

A. Zellner. Bayesian method of moments (BMOM) analysis of mean and regression models. In J. C. Lee, W. O. Johnson, and A. Zellner, editors, *Modelling and Prediction Honoring Seymour Geisser*, pages 61–72. Springer New York, 1996.