Regularized Posteriors for Asset Pricing Functionals in Rational Expectation Models

Anna Simoni
Toulouse School of Economics
GREMAQ

October 15, 2008

Job Market Paper

Abstract

In this paper we recover the posterior distribution of the equilibrium asset pricing functional p in a completely nonparametric way. We consider the rational expectation model of Lucas (1978) that characterizes the function p as the solution of an integral equation of second kind. We adopt a Bayesian procedure since it allows to incorporate all the prior information we have and mimics the Bayesian learning process of economic agents that leads to form rational expectations. The Bayesian approach reformulates the problem of solving an integral equation as an estimation problem in an Hilbert space.

The infinite dimension of the space and of the parameter of interest causes a problem of non continuity of the posterior distribution for p. Therefore, we propose to use a regularized version of the posterior distribution in order to guarantee consistency in the sampling sense of the Bayesian estimator for p.

Some of the elements of the Lucas' functional equation are unknown, hence in a first step we estimate the Markov state density by a kernel smoothing, and individual preference parameters by a GMM procedure. In the second step, we use the measurement error caused by the first step estimation to define the sampling probability.

Finally, frequentist asymptotic properties of the regularized posterior distribution are established.

JEL codes: C11, C14, E21, E44.

Keywords: Rational Expectations, Markov Process, Tikhonov Regularization, Hilbert scale, Posterior Consistency.

1 Introduction

Dynamic rational expectation models have been extensively studied in economic and econometric theory. In these models economic agents are supposed to face an intertemporal choice problem in which they have to determine their consumption and investment plans through a maximization of an infinite horizon expected utility function under budget and positiviy constraints. The result is a model for general equilibrium assets pricing where the assumption of *rational expectations* is fundamental. In fact, it is assumed that the market clearing price, implied by consumer behavior, is the same as the price on which consumer decisions are based.

This paper exploits the equilibrium characterization provided by such kind of models in order to analyze the performance of the Bayesian nonparametric approach, that we have introduced in Florens *et al.* (2008) [13], to estimate the equilibrium asset pricing functional. We introduced the nonparametric bayesian approach in a general setting where the object of interest was the solution of a linear functional equation. In dynamic rational expectation models we find a functional equation characterizing the asset pricing functional. Estimate the pricing functional is useful in practice since it allows to make forecast, to evaluate the efficiency of financial markets or to analyze if a financial asset is over- or under-priced.

The Bayesian approach is appropriate to analyze rational expectation models since the way in which economic agents form rational expectations is driven by a Bayesian learning process. The theory of rational expectations was introduced by Muth (1961) [29] and applied to the economy as a whole by Lucas during the 1970s, see Lucas (1976) [26] and Lucas (1978) [27]. This theory revolutionized macroeconomics and economic thinking. It is based on the belief that economic agents make their economic choices by taking into account their previous experiences and their rational expectations of the result of those choices. So, as Lucas (1978) [27] points out, the hypothesis of rational expectation "is not *behavioral*: it does not describe the way agents think about their environment, how they learn...It is rather a properly likely to be (approximately) possessed by the outcome of this unspecified process of learning and adapting".

The econometric analysis of dynamic rational expectation models is widely developed. Lucas (1976) [26] and Hansen *et al.* (1980) [17] observed that, instead of estimating the parameter of agents' decision rules, we should estimate the parameters of agents' objective functions and the random process they face as decision makers. This is enough for enabling the econometricians to predict how agents' decision rules change over time across alterations in their stochastic environment.

On the basis of the nature of the optimization problem solved by economic agents, in this kind of models, it is possible to find two veins of econometric literature. The first one considers quadratic optimization problems subject to linear constraints where it is possible to completely characterize the equilibrium time paths of the variables of interest. Econometric analysis of this case can be found in Hansen *et al.* (1980) [17], Hansen *et al.* (1981) [18] and Sargent (1981) [33].

In the second vein the linear-quadratic framework is replaced with a nonquadratic objective function; this causes dynamic rational expectations models no more yield representations for the variable of interest that are easy to handle from an econometric point of view. However, the dynamic optimization problem of economic agents provide a set of stochastic Euler equations that must be satisfied in equilibrium. These Euler equations, in turn, imply a set of population orthogonality conditions that can be exploited to estimate the parameters of interest. Several authors have proposed to use Euler equations to estimate parameters, see Hayashi (1980) [21], Hansen *et al.* (1982) [19] or Fair *et al.* (1980) [10].
An other branch of econometric literature, concerning dynamic rational expectation models, is interested in directly recovering the equilibrium asset pricing functional and it considers as general dynamic equilibrium model the rational expectations model proposed by Lucas (1978) [27]. Our paper gets into this literature. By considering a one-good, pure exchange economy with identical consumers, the equilibrium asset vector price is characterized as a functional $p(\cdot)$ of the Markov state of the economy solution of an integral equation of second kind. The functional equation is of the form $(I - K)p = r$, where I and K are two operators (the identity and an integral operator, respectively) onto an infinite dimensional Hilbert space and r is a known element of this Hilbert space [1]. Such characterization is particularly useful since it allows to recover the equilibrium asset prices without imposing any parametric restriction on them by using the theory on inverse problems. Only regularity and smoothness conditions will be imposed.
Carrasco *et al.* (2007) [4] propose a classical method for estimating the asset price in Lucas' model based on an estimation of r and K and on the simple inversion of operator $(I - K)$. The inverse problem is well-posed so that no regularization technique is demanded for solving it. Alternatively, numerical procedures have been proposed. Tauchen *et al.* (1991) [34] compute a discrete state space solution method for the pricing functional based on numerical quadrature approximation of the integral operator K. Rust *et al.* (2002) [32] use the observation that operator $K + r$ is a *quasi linear contraction* and compute a pointwise ε-approximation of its fixed point. This approximation is shown to converge at rate close to T^{-1}. An other approach for obtaining the price-dividend ratio is given by Chen *et al.* (2008) [5].
The new approach that we propose to estimate the asset pricing functional is different than the previous ones first of all because it it bayesian. Our approach restates the integral equation in a larger space of probability distributions so that each quantity in it (p and r in our case) are re-interpreted as random functions. This reformulation of an inverse problem as a parameter estimation is due to Franklin (1970) [14]. Hence, from a Bayesian point of view, the solution to an inverse problem is the *posterior distribution* of the quantity of interest p.
A bayesian analysis is interesting for many reasons. First, in computing the estimator of

[1]An integral equation of second kind is a particular type of inverse problem and it can be ill-posed or well-posed according to the fact that the integral operator K in it has an eigenvalue equal to one or not. Methods for treating integral equations of second kind are extensively treated in Kress (1999) [23] and Carrasco *et al.* (2007) [4].

the price functional, it allows to exploit the prior information we could have. In financial markets it is usual to possess this kind of information and hence it is efficient to use it for improving forecasting. Moreover, the Bayesian method that we propose for recovering solution of integral equations extend Bayesian nonparametric estimation out of the Dirichlet process framework. In this paper we are able to stay completely nonparametric by using a gaussian process prior.

Some element of the integral equation defining the asset pricing functional is unknown and requires to be estimated, so that we obtain an approximation of the integral equation: $\hat{r} \approx (I - \hat{K})p$. Some elements in $\hat{r}$ and in $\hat{K}$ are estimated parametrically and some other nonparamettrically. To derive the sampling probability associated to this functional equation, only the nonparametrically estimated part matters. The exact sampling distribution is not computable and derivation of the asymptotic one requires to transform the model as $\hat{K}^*\hat{r} = \hat{K}^*(I - \hat{K})p$, where $\hat{K}^*$ denotes the estimation of the adjoint of $\hat{K}$. Finally, we end up with an integral equation of first kind that is solvable through the technique we have proposed in Florens *et al.* (2008) [13]. Hence, even if both the classical and the bayesian approaches start with the same functional equation, they finally solve two substantially different, though linked, functional equations.

The infinite dimension of the pricing functional inverse problem makes the posterior distribution not well defined due to lack of continuity of its mean function. Hence, the posterior mean, and consequently the posterior distribution, is prevented from being consistent in the frequentist sense. This is an interesting example of frequentist inconsistency in Bayesian nonparametric estimation, see Diaconis *et al.* (1986) [7]. If p_* denotes the true value of the pricing functional having generated the data, the posterior distribution is said to be consistent in the frequentist sense if it degenerates, with respect to the sampling distribution, towards a point mass in p_* as more and more observations are collected.

Previous literature on Bayesian analysis of integral equations, see Franklin (1970) [14] and Mandelbaum (1984) [28], has solved this problem of non-continuity by restricting the space of definition of the observable element (r in our case). However, this technique is not always applicable, above all with real data.

The strategy that we propose consists in getting rid of the lack of continuity by applying a regularization scheme in the computation of the posterior distribution. We propose two alternative regularization schemes: a classical Tikhonov scheme and a Tikhonov regularization in the Hilbert scale induced by the prior covariance operator. The posterior distribution that we get is slightly modified and it is called *Regularized Posterior distribution* to highlight the role played by the regularization scheme. We take as punctual Bayesian estimator the mean of this distribution. Under some regularity condition on the true pricing functional p_*, our bayesian estimator converges towards p_* faster, in L^2-norm and in the sampling probability, than the classical estimator proposed in Carrasco *et al.* (2007) [4].

Finally, we study a particular prior distribution that is itself able to introduce the regularization scheme necessary for making the posterior distribution consistent.

The paper is organized as follows. In Section 2 we briefly remind the rational expectation general equilibrium model of Lucas (1978) and we explicit the functional equation in equilibrium asset price as an integral equation of second kind. We properly define the Hilbert space we are working in and the integral operator K. The Bayesian approach will be explained and adapted to this particular inverse problem in Section 3. In this section we compute the regularized posterior distribution by using the two alternative regularization scheme. In Section 4, posterior consistency of the regularized posterior distribution of the asset price p is proved. Section 5 presents a particular prior distribution for the pricing functional that is able to regularize. We develop an extension of our model in Section 6 where the parameter in the sampling covariance operator is unknown. Section 7 concludes. All the proofs and some numerical simulation can be found in the Appendix.

2 Rational Expectations Asset Pricing Model

2.1 Lucas' (1978) Model

Lucas (1978) [27] constructed the equilibrium in an exchange economy under the assumption of *rational expectations*. The first-order conditions for attaining the optimum define a functional equation in the vector of equilibrium prices of financial assets which is solved for price as a function of the physical state of the economy.

We consider a one-good pure exchange economy with a single consumer interpreted as representative of a large number of identical consumers. The consumer faces the intertemporal choice problem between consumption and trading in financial assets and she/he maximizes the expectation of a time-separable utility function:

$$\mathbb{E}_t\Big[\sum_{j=0}^{\infty}\beta^j U(C_{t+j})\Big] \tag{1}$$

where $\mathbb{E}_t$ denotes the conditional expectation operator conditional on the information set $\mathcal{F}_t$ available in t, $\beta \in (0,1)$ is the time discount factor, $U(\cdot)$ is a current period strictly concave utility function and C_{t+j} is a stochastic process representing the consumption of a single good at time $t+j$. Since expectations are supposed to be formed rationally, $\mathbb{E}_t$ denotes both the mathematical conditional expectation and the agents' subjective expectations at time t.

In this economy there exist n distinct productive units (denoted with $i = 1,\dots,n$) each one producing a quantity Y_{it} of the consumption good in period t. The production $Y_t = (Y_{1t},\dots,Y_{nt})$ is assumed to be entirely *exogenous* and to follow a Markov process defined by its transition distribution function $F(y_{t+1}|y_t) = \mathbb{P}\{Y_{t+1} \leq y_{t+1}|Y_t = y_t\}$. Moreover, since the produced output is perishable, feasible consumption levels are those which satisfy $0 \leq C_t \leq \sum_{i=1}^n Y_{it}$. Each productive unit has outstanding one perfectly divisible equity share held by the representative consumer and traded at a competitively determined price vector $p_t = (p_{1t},\dots,p_{nt})$. We denote with $z_t = (z_{1t},\dots,z_{nt})$ the consumer's share holding at the beginning of period t, *i.e.* z_{it} is the period t share holding in the i-th productive unit.

Definition of the equilibrium of this economy requires to determine the equilibrium *quantities* of consumption and asset holdings and the equilibrium *price vector* p. As Lucas stresses, the equilibrium quantities of consumption and asset holdings are easily determined since all output will be consumed and all shares will be held, then

$$C_t = \sum_{i=1}^{n} Y_{it}, \qquad z_t = (1, \dots, 1), \quad \forall t. \tag{2}$$

The feasible equilibrium consumption and investment plans must satisfy, at each period t, the budget constraint

$$C_t + p_t z_t \leq Y_t z_{t-1} + p_t z_{t-1}, \quad C_t \geq 0 \;\; z_t \geq 0. \tag{3}$$

The important economic variable whose equilibrium value remains to be determined is the asset price. Equilibrium prices are set by the asset market by solving a problem of the same form each period, so that it seems natural to express them as some fixed function $p(\cdot)$ of the state of the economy: $p_t = p(Y_t)$, where the i-th coordinate $p_i(Y_t)$ is the price of a share of unit i when the economy is in the state Y_t.
The first order conditions for maximizing (1) subject to (3), once equilibrium conditions (2) have been incorporated, gives a functional equation in the equilibrium price vector, or equivalently, n functional equations:

$$p_i(Y_t) = \beta \int \frac{U'(\sum_i y_{i,t+1})}{U'(\sum_i Y_{i,t})} \left(y_{i,t+1} + p_i(y_{t+1})\right) d\, F(y_{t+1}|Y_t), \tag{4}$$

for $i = 1, \dots, n$, where the conditional expectation $\mathbb{E}_t$ in (1) has been explicited. This equilibrium asset-pricing relation equates current price of the i-th security to its expected discounted future payoff, discounted using the *stochastic discount factor* $M_{t+1}(Y_t, Y_{t+1}) = \beta \frac{U'(\sum_i Y_{i,t+1})}{U'(\sum_i Y_{i,t})}$. The *stochastic discount factor* is expressed as a function of the vectorial Markov state $\{Y_t\}$ instead of consumption process $\{C_t\}$. In the following of the paper, sometimes we shall denote it, at time $t+1$, simply by M_{t+1}, by neglecting its arguments. The object of interest in this paper will be the determination of the vector of pricing functionals $p(\cdot)$. Since equilibrium prices are a fixed function of the state of the economy, once the transition function $F(y_{t+1}|y_t)$ is known or estimated, this will be sufficient to determine the stochastic process of prices p_t.

2.2 Martingale Property

The equilibrium asset-pricing relation (4) says that $p_i(Y_t) = \mathbb{E}[M_{t+1}(Y_{i,t+1} + p_{i,t+1})|Y_t]$. Therefore, we can write:

$$M_{t+1}(Y_{i,t+1} + p_{i,t+1}) = p_i(Y_t) + \varepsilon_{t+1}. \tag{5}$$

The variable ε_{t+1} is a noise satisfying the following assumption that will turn out useful in determining the covariance operator of the sampling distribution in the Bayesian experiment.

Assumption 1 $\{\varepsilon_{t+1}\}$ *is a weak white noise with variance* σ^2 *that is constant for all* t.

The fact that error terms are serially uncorrelated prevents problems of endogeneity of the regressors.

Lucas (1978) [27] stresses that *"asset prices themselves do not possess the Martingale property"*, but that asset prices properly corrected for dividends and for the stochastic discount factor β possess this property, how can be seen from equation (4). This observation confirms the finding of Leroy (1973) [25] that the martingale property is neither a necessary nor sufficient condition for rationally determined asset prices. However, it is possible to show that there exists a probability, known as *risk-neutral probability* (or *equivalent martingale measure* - EMM) under which the discounted price process corrected for dividends is a martingale. To show this, note that relation (4), divided by the value of the function $p_i(Y_t)$, gives for a risk-free security

$$1 = (1 + r_f)\mathbb{E}_F(M_{t+1}|Y_t),$$

where r_f denotes the risk-free rate compounded once in period $[t, t+1]$. We make the following assumption concerning the transition distribution function of the Markov state

Assumption 2 *The transition distribution function* $F(y_{t+1}|y_t)$ *is absolutely continuous with respect to the Lebesgue measure and there exists a positive function* f *such that* $\frac{dF(y_{t+1}|y_t)}{dy_{t+1}} = f(y_{t+1}|y_t)$.

Hence, under this hypothesis, we have $\forall i = 1, \ldots, n$

$$\begin{aligned} p_i(Y_t) &= \int \frac{Y_{i,t+1} + p_i(Y_{i,t+1})}{1 + r_f} \frac{M_{t+1}(Y_t, Y_{t+1})}{\mathbb{E}(M_{t+1}|Y_t)} f(Y_{t+1}|Y_t) dY_{t+1} \\ &= \int \frac{Y_{i,t+1} + p_i(Y_{i,t+1})}{1 + r_f} f^*(Y_{t+1}|Y_t) dY_{t+1}, \end{aligned}$$

where $f^*(Y_{t+1}|Y_t) = \frac{M_{t+1}}{\mathbb{E}(M_{t+1}|Y_t)} f(Y_{t+1}|Y_t)$ is the equivalent martingale measure. In the following we denote with $\mathbb{E}^*$ the expectation taken with respect to this probability.

2.3 Integral Equations of Second Kind and Characterization of the Operator

We study in this subsection mathematical properties of functional equation (4), meant as a functional equation in $p_i(\cdot)$ and we properly characterize all the elements appearing in it. If Assumption 2 holds we can restate equation (4) in a more general form:

$$p_i(Y_t) - \int M_{t+1}(Y_t, Y_{t+1}) p_i(Y_{t+1}) f(Y_{t+1}|Y_t) dY_{t+1} = \int M_{t+1}(Y_t, Y_{t+1}) b_i(Y_{t+1}) f(Y_{t+1}|Y_t) dY_{t+1}, \tag{6}$$

for $i = 1, \ldots, n$. Function b_i is the coordinate function associating vector Y_{t+1} to its i-th component. $\{Y_t\}$ is an n-dimensional stationary stochastic process that satisfies Markov property with stationary distribution Π, *i.e.* Π is the unique solution to

$$\Pi(Y_{t+1}) = \int F(Y_{t+1}|Y_t) d\Pi(Y_t).$$

We denote with π the density function associated to Π.

Let $\mathcal{X}$ be the space of square integrable functions of one realization of $\{Y_t\}$ with respect to the stationary distribution Π endowed with the scalar product $< \cdot, \cdot >$ inducing the norm $|| \cdot ||$. We assume that $p \in \mathcal{X}$ and we define an operator K acting on this space as:

$$\forall \phi \in \mathcal{X}, \qquad K\phi(Y_t) = \mathbb{E}_F(M_{t+1}(Y_t, Y_{t+1})\phi(Y_{t+1})|Y_t),$$

where the conditional expectation is taken with respect to the transition distribution $F(Y_{t+1}|Y_t)$. Operator K is a contraction operator with norm less then 1. The contraction property can be easily proved by using Theorem 5 in Blackwell (1965) [3] or directly through the definition of contraction operator. In particular, $||K|| := \sup_{\phi:||\phi||\leq 1} ||K\phi|| \leq \frac{1}{1+r_f} \sup_{\phi:||\phi||\leq 1} ||\mathbb{E}^*(\phi|Y_t)|| < 1$ since the conditional operator has norm equal to 1 and $\frac{1}{1+r_f} < 1$.
The adjoint K^* of this operator is defined through the equality $< K\phi, \psi > = < \phi, K^*\psi >$, $\forall \phi, \psi \in \mathcal{X}$, so that $K^*\psi = \mathbb{E}_F(M_{t+1}(Y_t, Y_{t+1})\psi(Y_t)|Y_{t+1}) = \int \beta \frac{U'(Y_{t+1})}{U'(y_t)} \psi(y_t) f(y_t|y_{t+1}) dy_t$. Although $F(Y_t|Y_{t+1}) = F(Y_{t+1}|Y_t)$, the two operators K and K^* are substantially different due to the fact that M_{t+1} is not symmetric in its arguments. Thus, $K\phi$ coincides, up to a constant, with the conditional expectation of the product of ϕ and the marginal utility function whereas $K^*\phi$ is proportional to the conditional expectation of the ratio $\frac{\phi}{U'}$.

We call $r_i(Y_t)$, or simply r_i, the right hand side of equation (6), so that we rewrite the equilibrium model as

$$\begin{aligned} r_i(Y_t) &= (I - K)p_i(Y_t), \qquad i = 1, \ldots, n \qquad (7) \\ r_i(Y_t) &= \mathbb{E}_F(M_{t+1}(Y_t, Y_{t+1})b_i(Y_{t+1})|Y_t), \qquad i = 1, \ldots, n \end{aligned}$$

where I is the identity operator onto $\mathcal{X}$. In the following we eliminate the subscript i in the price, b_i and r_i functions and it will be implied that the functional equation $(I - K)p = r$ refers to a single security.

We will now introduce an assumption, that is only a regularity assumption but that is useful to guarantee compacity of operator K.

Assumption 3 *The Equivalent Martingale Measure $f^*(Y_{t+1}|Y_t)$ is dominated by the marginal distribution of Y_{t+1} and its density is square integrable with respect to the product of margins of Y_{t+1} and Y_t.*

Exploiting this assumption it is possible to show that K is an Hilbert-Schmidt operator. Let $k(Y_t, Y_{t+1}) = M_{t+1}\frac{f(Y_{t+1}|Y_t)}{\pi(Y_{t+1})}$ be the kernel characterizing operator K, K is an Hilbert-Schmidt operator if the Hilbert-Schmidt norm $||\cdot||_{HS}$ is finite:

$$\begin{aligned}
||K||^2_{HS} &= \int |k(Y_t, Y_{t+1})|^2 \pi(Y_t)\pi(Y_{t+1})dY_t dY_{t+1} \\
&\leq (1+r_f)^2 \int (M_{t+1}\frac{f(Y_{t+1}|Y_t)}{\pi(Y_{t+1})})^2 \pi(Y_t)\pi(Y_{t+1})dY_t dY_{t+1} \\
&= \int (\frac{M_{t+1}}{\mathbb{E}(M_{t+1}|Y_t)}\frac{f(Y_{t+1}|Y_t)}{\pi(Y_{t+1})})^2 \pi(Y_t)\pi(Y_{t+1})dY_t dY_{t+1} \\
&= \int (g^*(Y_{t+1}|Y_t))^2 \pi(Y_t)\pi(Y_{t+1})dY_t dY_{t+1} < \infty
\end{aligned}$$

where the second line follows from the fact that $(1+r_f)^2 \geq 1$ and g^* is the density of the EMM f^* with respect to $\pi(Y_{t+1})$, *i.e.* $\frac{dF^*(Y_{t+1}|Y_t)}{d\Pi(Y_{t+1})} = g^*(Y_{t+1|Y_t})$.
Hilbert-Schmidt operators are compact; this is a very attractive property since every compact operator is the limit of a sequence of operators with finite dimensional range. Hence, when operator K has to be estimated it can be approached by a sequence of finite dimensional operators. Furthermore, a compact operator has peculiar spectral properties. The eigenvectors of a self-adjoint compact operator can be orthonormalized, the set of its eigenvalues $\{\lambda_j^2\}$ is at most countable and if there are infinitely many eigenvalues they accumulate only at 0. For a compact operator that is non self-adjoint, like K, we consider its *singular values* that are defined to be the square roots of the eigenvalues of the nonnegative self-adjoint compact operator K^*K. Then, there exist orthonormal sequences $\{\varphi_j\}$ and $\{\psi_j\}$ of $\mathcal{X}$ such that

$$K\varphi_j = \lambda_j\psi_j, \quad K^*\psi_j = \lambda_j\varphi_j.$$

Assumption 3 also implies that $r(Y_t) \in \mathcal{X}$ and $\mathcal{R}(K) \subseteq \mathcal{X}$, then $K : \mathcal{X} \to \mathcal{X}$.
Functional equation (6) is an integral equation of second kind and its properties are well known in the literature (see Kress (1999) [23]). While K is compact, $(I-K)$ is not compact. Moreover, 1 is not an eigenvalue of K so that $(I-K)$ is one-to-one and its inverse is bounded. Therefore, the inverse problem defined by (6) is well-posed in the sense the it satisfies Hadamard's conditions, see Engl *et al.* [9].

3 Bayesian Econometric Analysis

The aim moving our econometric analysis is the characterization and estimation of the price process $\{p_t\}$. The price process can be expressed at each period t as a fixed function $p(\cdot)$ of the state of the economy: $p_t = p(Y_t)$. Therefore, once function $p(\cdot)$ is known, knowledge of the transition function $F(y_{t+1}|y_t)$ is enough to determine the stochastic character of the price process. While the transition function will be approximated in a classical nonparametric way (*e.g.* with a kernel method) the price function $p(\cdot)$ will be the object of a Bayesian analysis.

Our estimation choice is justified by the argument that prices are economic variables that economic agents have to take into consideration when they make their economic decisions and on which they performs a Bayesian learning through a continuous updating of the prior distribution. Hence, it seems natural to consider a similar learning process for the econometrician. On the contrary, the transition probability of the state of the economy is exogenous to the learning process of the economic agents and so it does not seem suitable to treat it in a Bayesian way. Roughly speaking, we could consider $F(y_{t+1}|y_t)$ as a nuisance parameter.

In the same way, we shall consider all the preference parameters as nuisance parameter and they will not enter the bayesian problem but will be estimated by a classical GMM or Maximum likelihood procedure. For this purpose, we could specify $U(C_t)$ as a *constant relative risk aversion* (CRRA) utility function and use the estimation procedure illustrated in Hansen *et al.* (1983) [20] and in Hansen *et al.* (1982) [19]. However, our method is valid for any form of the utility function and we will not limit us to this case.

Because of the fastest parametric speed of convergence of the GMM estimator with respect to the nonparametric rate that characterizes convergence of the kernel estimated transition density and bayesian estimated pricing functional, the estimated preference parameters will not play any role in the asymptotic analysis of our price estimator. For this reason, without loosing any generality, we can consider these parameters as known.

3.1 Nonparametric Estimation of the Transition Density

The transition density function $f(Y_{t+1}|Y_t)$ is usually unknown. In this subsection, it will be briefly reviewed construction and properties of the kernel density estimation considered in [31].

With abuse of notation, we use f to denote both the transition density and the two-dimensional joint density of the Markov process $\{Y_t\}$ with respect to Lebeasgue measure. It is assumed that π is strictly positive on $\mathbb{R}_+$. Then, the transition density of the process is written as $\frac{f(Y_t,Y_{t+1})}{\pi(Y_t)}$. We state the following assumption where small letters denote realization of the random variable Y_t.

Assumption 4 *We dispose of a $(T+1)$ sample $(y_1,\ldots,y_{T+1})$ from the weakly stationary Markov process $\{Y_t\}$.*

We follow the original setup of Lucas (1978) which assumes stationarity of dividends levels, so we take Y_t as the aggregate consumption process.

In some case, data may not confirm the hypothesis of stationarity of the consumption process. When this is the case, it is sufficient to rewrite the basic asset pricing equation (4) to express it in terms of consumption growth rates, which is shown to be stationary and Markov by empirical evidence. Then, Y_t will denote either the consumption growth rate process or a stationary state variable whose the consumption growth rate is a transformation, see Chen *et al.*. The slightly modified asset pricing equation can be rewritten as

$$v_i(Y_t) = \mathbb{E}(m(Y_{t+1}, Y_t)[1 + v_i(Y_{t+1})]\frac{Y_{t+1}}{Y_t}|Y_t) \quad (8)$$

where v_i denotes the i-th asset's price-dividend ratio, $m(Y_{t+1}, Y_t) = \beta\frac{U'(C_{t+1})}{U'(C_t)}$, under the hypothesis of homogeneous utility function, and $\frac{Y_{t+1}}{Y_t}$ is the dividend growth variable.
For clarity and simplicity of exposition we consider the basic Lucas setting where the stationarity assumption is taken. All the results in the following can be trivially adapted to the functional equation (8) with only minor modifications.

Let $L : (\mathbb{R}^n) \to \mathbb{R}$ be a measurable function satisfying properties:

$$|L(u)| \leq M_1(< \infty), \quad u \in \mathbb{R}^n; \qquad \int |K(u)|du < \infty,$$
$$||u||^m|K(u)| \to 0, \text{ as } ||u|| \to \infty; \qquad \int K(u)du = 1,$$

$h = h(T)$ be a function of T such that $h \to 0$ as $T \to \infty$ and $L_h(u)$ stands for $L(\frac{u}{h_T})$. Then, the kernel transition density estimation is obtained as the ratio of the kernel density estimation of the joint f and of π, $\hat{f}(Y_{t+1}|Y_t) = \frac{\hat{f}(Y_t,Y_{t+1})}{\hat{\pi}(Y_t)}$:

$$\hat{f}(Y_{t+1}|Y_t) = \frac{\frac{1}{Th^2}\sum_{j=1}^{T} L_h(Y_t - y_j)L_h(Y_{t+1} - y_{j+1})}{\frac{1}{Th}\sum_{l=1}^{T} L_h(Y_t - y_l)}.$$

We plug this estimator in the operator K and in r:

$$\hat{K}p(Y_t) = \int M_{t+1}(Y_t, Y_{t+1})p(Y_{t+1})\hat{f}(Y_{t+1}|Y_t)dY_{t+1}$$
$$\hat{r}(Y_t) = \int M_{t+1}(Y_t, Y_{t+1})b(Y_{t+1})\hat{f}(Y_{t+1}|Y_t)dY_{t+1}.$$

We assume that $\hat{K}$ defines an operator from $\mathcal{X}$ into $\mathcal{X}$ and $\hat{r}$ an element of $\mathcal{X}$. These assumptions are actually integrality assumptions on the kernel function L. The change of variable $\frac{Y_{t+1}-y_{j+1}}{h} = u$ and a Taylor expansion at the first order, allows to simplify these expressions as

$$\hat{K}p = \frac{\frac{1}{Th}\sum_{j=1}^{T} M_{t+1}(Y_t, y_{j+1})p(y_{j+1})L_h(Y_t - y_j)}{\frac{1}{Th}\sum_{l=1}^{T} L_h(Y_t - y_l)}$$
$$\hat{r} = \frac{\frac{1}{Th}\sum_{j=1}^{T} M_{t+1}(Y_t, y_{j+1})b(y_{j+1})L_h(Y_t - y_j)}{\frac{1}{Th}\sum_{l=1}^{T} L_h(Y_t - y_l)}.$$

$\hat{K}$ is a degenerate operator with T-dimensional range, then there are at most T nonzero singular values $\hat{\lambda}_j$. They are estimators of the singular values $\{\lambda_j\}$ of the true operator K and they accumulates to 0 as $\hat{K} \to K$.
Asymptotic properties of this kernel estimator will affect the asymptotic properties of the Bayesian estimator for p. Note that the use of these estimated quantities implies that the equality in (7) is now only approximately true: $\hat{r} \approx (I - \hat{K})p$.

3.2 Construction of the Bayesian experiment

We concentrate in this paragraph on the characterization of the Bayesian experiment associated to (7). Given the reasons discussed at the beginning of Section 3 and in order to simplify notation we take no notice of the fact that preference parameters and β are estimated: they will be assumed as known. Hence, we do not have to use notation $\hat{M}_{t+1}$ for the stochastic discount factor. On the contrary, the transition density will be substituted with the kernel estimator previously described.

3.2.1 Prior Distribution

The first step in order to well define the Bayesian experiment is the characterization of a *prior probability* μ induced by the price process p on the parameter space $\mathcal{X}$. We endow the parameter space with the σ-field $\mathcal{E}$ and we assume that μ is a gaussian measure.

Assumption 5 *Let μ be a probability measure on $(\mathcal{X}, \mathcal{E})$ such that $\mathbb{E}(||p||^2) < \infty$, with $\mathbb{E}$ the expectation taken with respect to μ. μ is a Gaussian measure that defines a mean element $p_0 \in \mathcal{X}$ and a covariance operator $\Omega_0 : \mathcal{X} \to \mathcal{X}$.*

μ is gaussian if the probability distribution on the Borel sets of $\mathbb{R}$ induced from μ by every bounded linear functional on $\mathcal{X}$ is gaussian. More clearly, μ gaussian means that $\forall B \in \mathcal{B}(\mathbb{R})$

$$\mathbb{P}(B) = \mu\{p; < p, \varphi > \in B\}$$

is gaussian for all $\varphi \in \mathcal{X}$, see Baker (1973) [2]. The mean element p_0 in $\mathcal{X}$ is defined by

$$< p_0, \varphi > = \int_{\mathcal{X}} < p, \varphi > d\mu(p)$$

and the operator Ω_0 by

$$< \Omega_0 \varphi_1, \varphi_2 > = \int_{\mathcal{X}} < p - p_0, \varphi_1 > < p - p_0, \varphi_2 > d\mu(p)$$

for every $\varphi_1, \varphi_2 \in \mathcal{X}$. Let $\mathcal{S}(\mathcal{X})$ denote the set of all linear, bounded, self-adjoint, positive semi-definite and trace-class operators onto $\mathcal{X}$. In particular, $\mathcal{S}(\mathcal{X})$ is the set of all covariance operators of Gaussian measure on $\mathcal{X}$. On the basis of Assumption 5, Ω_0 is correctly specified as a covariance operator in the sense that it belongs to $\mathcal{S}(\mathcal{X})$. A covariance operator needs to be trace-class in order the associated measure be able to generate trajectories in the well suited space. Indeed, by Kolmogorov's inequality a realization of the random function p is in $\mathcal{X}$ if $\mathbb{E}(||p||^2)$ is finite[2]. Since $\mathbb{E}(||p||^2) = \sum_j \lambda_j^{\Omega_0}$, this is guaranteed if Ω_0 is trace-class, that is if $\sum_j \lambda_j^{\Omega_0} < \infty$, with $\{\lambda_j^{\Omega_0}\}$ the eigenvalues associated to Ω_0 and $\mathbb{E}(\cdot)$ the expectation taken with respect to μ.
Since the eigenvalues of $\Omega_0^{\frac{1}{2}}$ are the square roots of the eigenvalues of Ω_0 the fact to be

[2]Namely, following Kolmogorov's inequality $\mathbb{P}(||p|| > \epsilon_n) \sim \mathcal{O}_p(1)$ if and only if $\mathbb{E}(||p||^2)$ is finite.

trace-class entails that $\Omega_0^{\frac{1}{2}}$ is Hilbert-Schmidt. Hilbert-Schmidt operators are compact and the adjoint is still Hilbert-Schmidt. Compacity of $\Omega_0^{\frac{1}{2}}$ implies compacity of Ω_0.
This specification for the prior measure is suitable in the sense that its support is the closure of the *Reproducing Kernel Hilbert Space* associated to Ω_0, ($\overline{\mathcal{H}(\Omega_0)}$ in the following), that is dense in $\mathcal{X}$ if Ω_0 is one to one. Let $\{\lambda_j^{\Omega_0}, \varphi_j^{\Omega_0}\}$ be the eigensystem of Ω_0. We define the space $\mathcal{H}(\Omega_0)$ embedded in $\mathcal{X}$ as

$$\mathcal{H}(\Omega_0) = \{\varphi : \varphi \in \mathcal{X} \quad \text{and} \quad \sum_{j=1}^{\infty} \frac{| < \varphi, \varphi_j^{\Omega_0} > |^2}{\lambda_j^{\Omega_0}} < \infty\} \tag{9}$$

and, following Proposition 3.6 in Carrasco *et al.* (2007) [4], we have the relation $\mathcal{H}(\Omega_0) = \mathcal{R}(\Omega_0^{\frac{1}{2}})$. It results evident how the choice of the covariance operator can modify the support of a gaussian measure. In particular, if Ω_0 is injective then the support of μ is the whole space $\mathcal{X}$, otherwise, the support is any subset of $\mathcal{X}$; henceforth, a particular choice of the covariance operator allows to incorporate in the prior distribution constraints on the parameter of interest.
Let p_* denote the true value of the price functional having generated data $\hat{r}$, we assume that

Assumption 6 *$(p_* - p_0) \in \mathcal{H}(\Omega_0)$, i.e. there exists $\delta_* \in \mathcal{X}$ such that $(p_* - p_0) = \Omega_0^{\frac{1}{2}}\delta_*$.*

In other words, we are supposing there exists a function $\delta_* \in \mathcal{X}$ such that the centered true value of the pricing functional is the image of it through operator $\Omega_0^{\frac{1}{2}}$. This assumption is only a regularity condition on p_* and will be exploited for proving asymptotic results.

3.2.2 Sampling Distribution

In our model both the parameter and the sample space coincide with $\mathcal{X}$. We denote with Q^p the sampling probability on $\mathcal{X}$, namely the conditional probability of the observations given p. An exact conditional distribution of the process $\hat{r} - (I - \hat{K})p$ given p is impossible to compute. Hence, we need to compute its asymptotic distribution. However, the nonparametric estimator used for obtaining $\hat{K}$ and $\hat{r}$ prevents us to find convergence of $\hat{r} - (I - \hat{K})p$ to a well defined continuous process (like a gaussian process). In order to obtain weakly convergence of this process it is necessary to smooth its trajectories. Hence, we need to modify the basic functional equation $\hat{r} - (I - \hat{K})p$ and we redefine p as the solution of the estimated integral equation

$$\hat{K}^*\hat{r} = \hat{K}^*(I - \hat{K})p + U. \tag{10}$$

We introduce the notation $\hat{R}$ for denoting $\hat{K}^*\hat{r}$ and $\hat{H}$ for denoting $\hat{K}^*(I-\hat{K})$ so that $\hat{R} = \hat{H}p + U$. It is implicit in this new model that the estimated operator $\hat{H}$ becomes the true operator defining the functional equation for p and p is now solution of an integral equation of first kind. The error term process can be rewritten as $U = \hat{K}^*((\hat{r} + \hat{K}p) - (r + Kp))$ and the following theorem shows that it is asymptotically gaussian.

Theorem 1 *There exists a random element $\vartheta \in \mathcal{X}$ such that $\sqrt{T}\hat{K}^*((\hat{r}+\hat{K}p)-(r+Kp))$ is asymptotically equivalent to*

$$\frac{\sqrt{T}}{T}\sum_j M_{t+1}(y_j,Y_{t+1})[M_{t+1}(y_j,y_{j+1})(b(y_{j+1})+p(y_{j+1}))-p(y_j)]\frac{f(y_j,Y_{t+1})}{\pi(y_j)\pi(Y_{t+1})}+h^{\rho}\vartheta.$$

Moreover, $\sqrt{T}\hat{K}^*((\hat{r}+\hat{K}p)-(r+Kp)) \Rightarrow \mathcal{GP}(0,\sigma^2K^*K)$ (weak convergence in $\mathcal{X}$) and K^*K is a trace-class operator.

It will be proved in the Appendix that the first term of the above equality and ϑ weakly converge to a gaussian element in $\mathcal{X}$, but that the second term becomes negligible after having been scaled by $h \to 0$.
The sampling distribution Q^p of $\hat{R}$ given p is characterized by the transition probability $\mathbb{P}(\cdot|p)$ that associates to each p a probability measure on $(\mathcal{X},\mathcal{F})$: $Q^p = \mathbb{P}(\hat{R} \in B|p)$, for all $B \in \mathcal{F}$, where $\mathcal{F}$ is the σ-field associated to the sample space. This probability is deduced from the above theorem, thus Q^p is approximately gaussian with mean $\hat{H}p$ and covariance operator $\Sigma_T = \frac{\sigma^2}{T}K^*K$. Because K is unknown, operator Σ_T is replaced by the estimator $\hat{\Sigma}_T = \frac{\sigma^2}{T}\hat{K}^*\hat{K}$ (under the assumption that σ^2 is known, we consider in Section 6 the case with σ^2 unknown). We should substitute Q^p with the estimated sampling measure $\hat{Q}^p$. All the probabilities characterizing the Bayesian experiment become estimated, anyway for simplicity of exposition we shall omit the corresponding notation but the fact that they are all estimated must be understood.
After this clarification some remarks are in order. First, the fact that the sampling probability is only asymptotically gaussian does not affect properties of our estimator. Indeed, we need normality only to construct the estimator of p and it is not used at all to prove consistency (that is the argument that justifies the proposed estimator).
Second, in order to recover the sampling probability, we have considered the estimated projected model instead of the more natural one $\hat{r} = (I-\hat{K})p + U$. This is because such error term does not weakly converge to any well-defined stochastic process since kernel estimation produces an empirical process converging to a process with discontinuous trajectories. Projecting the model through a further application of operator K^* allows to smooth trajectories and to increase the speed of convergence.
Third, $\Sigma_T \in \mathcal{S}(\mathcal{X})$, thus it possesses all the properties that characterize a covariance operator.

3.2.3 Identification

A model, and the corresponding parameter of interest, is identified in a Bayesian sense if the posterior distribution completely revises the prior distribution. For such a kind of identification we do not have to introduce strong assumptions, see Florens *et al.* (1990) [11] Section 4.6 for an exhaustive explanation of this concept. Anyway, this paper is not only concerned with the computation of the posterior distribution but mainly with the frequentist consistency of it. We will give in Section 4 the definition of *frequentist* *consistency*, also called *posterior consistency* or consistency in the sampling sense. For this type of consistency be verified we need the following assumption for identification.

Assumption 7 *The operator* $K^*(I-K)\Omega_0^{\frac{1}{2}} : \mathcal{X} \to \mathcal{Y}$ *is one-to-one on* $\mathcal{X}$*.*

This assumption guarantees continuity of the *regularized posterior mean* that we shall define below, so that posterior consistency is satisfied.
Some comments about this hypothesis are in order. If we used the classical model $r = (I-K)p$ and a classical (non bayesian) procedure to recover p then no identification condition would be required since operator $(I-K)$ is one-to-one (due to the fact that 1 is not an eigenvalue of K). In reality, we are using the projected model $K^*r = K^*(I-K)p$, so that if a classical resolution method is used the identification of p would require injectivity of $K^*(I-K)$ that is not guaranteed by injectivity of $(I-K)$. If we compare Assumption 7 to this last one, we see that it is weaker in the sense that if $\Omega_0^{\frac{1}{2}}$ is one-to-one then $K^*(I-K)\Omega_0^{\frac{1}{2}}$ injective does not imply $K^*(I-K)$ injective while the reverse is true.

3.2.4 Joint Probability Distribution

With relevant space we refer to the product of the sample and parameter space, associated to model (10), endowed with the associated σ-field $\mathcal{E} \otimes \mathcal{F}$ and with the joint measure determined by recomposing the prior and sampling distributions. We define the product space $\mathcal{X} \times \mathcal{X}$ as the set

$$\mathcal{X} \times \mathcal{X} := \{(\phi, \psi); \phi, \psi \in \mathcal{X}\}$$

with addition and scalar multiplication defined by $(\phi_1, \psi_1) + (\phi_2, \psi_2) = (\phi_1 + \phi_2, \psi_1 + \psi_2)$ and $h(\phi_1, \psi_1) = (h\phi_1, h\psi_1)$, $h \in \mathbb{R}$. $\mathcal{X} \times \mathcal{X}$ is a separable Hilbert space under the norm induced by the scalar product defined as

$$< (\phi_1, \psi_1), (\phi_2, \psi_2) > := < \phi_1, \phi_2 > + < (\psi_1, \psi_2) >, \quad \forall (\phi_i, \psi_i) \in \mathcal{X} \times \mathcal{X},\ i = 1, 2.$$

The joint probability measure on $\mathcal{X} \times \mathcal{X}$, denoted with Λ, is constructed by recomposing the prior μ and the sampling distribution Q^p in the following way:

$$\Lambda(A \times B) = \int_A Q^p(B)\mu(dp), \qquad A, B \in \mathcal{X}.$$

After that, function Λ is extended to $\mathcal{E} \otimes \mathcal{F}$. Following Florens and Simoni (2008) [13], it is trivial to prove that $(\hat{R}, p)$ are (asymptotically) jointly distributed as a gaussian process:

$$\begin{pmatrix} \hat{R} \\ p \end{pmatrix} \sim \mathcal{GP}\Big(\begin{pmatrix} \hat{H}p_0 \\ p_0 \end{pmatrix}, \begin{pmatrix} \hat{\Sigma}_T + \hat{H}\Omega_0\hat{H}^* & \hat{H}\Omega_0 \\ \Omega_0\hat{H}^* & \Omega_0 \end{pmatrix}\Big) \tag{11}$$

The marginal distribution induced by $\hat{R}$, denoted with Q, is gaussian with mean $\hat{H}p_0$ and covariance $C_T = \hat{\Sigma}_T + \hat{H}\Omega_0\hat{H}^*$ that is trace class. It should be noted that $\hat{H}$ and H are

compact operators since they are the product of a bounded and a compact operator, see Theorem 2.16 in Kress [23]. While $\hat{H}$ has a finite number of non-zero singular values, H has a countable number of singular values only accumulating to 0.
Summarizing, the bayesian experiment associated to model (4) can be written as

$$\Xi = (\mathcal{X} \times \mathcal{X}, \mathcal{E} \otimes \mathcal{F}, \Lambda = \mu \otimes Q^p).$$

Bayesian inference consists in finding the inverse decomposition of Λ in the product of the posterior distribution $\mu^{\mathcal{F}}$ and the predictive measure Q.

3.3 Analysis of the Posterior Distribution

The infinite dimension of the Bayesian experiment makes application of Bayes theorem not evident and in defining and computing the posterior distribution we should care about three points: *(i)* existence of a *regular* version of the conditional probability on $\mathcal{E}$ given $\mathcal{F}$, *(ii)* the fact that it is a gaussian measure and *(iii)* its continuity. The conditional probability $\mu^{\mathcal{F}}$, given $\hat{R}$, is said *regular* if a transition probability characterizing it exists, *i.e.* there exists a probability $\mathbb{P}(\cdot|\mathcal{F})$ such that $\mathbb{P}(A|\mathcal{F}) = \mu^{\mathcal{F}}(A)$, $\forall A \in \mathcal{E}$. The next theorem answers to the first two questions:

Theorem 2 *(i) Let $(\mathcal{X} \times \mathcal{X}, \mathcal{E} \otimes \mathcal{F}, \Lambda)$ a probability space that is Polish*[3]*, then there exists at least one regular conditional probability $\mathbb{P}(\cdot|\mathcal{F})$ such that $\mathbb{P}(A|\mathcal{F}) = \mu^{\mathcal{F}}(A)$, $\forall A \in \mathcal{E}$.*

(ii) The probability $\mu^{\mathcal{F}}$ is characterized by the characteristic function

$$\mathbb{E}(e^{i<p,h>}|\hat{Y}) = e^{i<A\hat{R}+b,h>-\frac{1}{2}<(\Omega_0 - A\hat{H}\Omega_0)h,h>}, \qquad h \in \mathcal{X},$$

where $A : \mathcal{X} \to \mathcal{X}$ and $b \in \mathcal{X}$. Then it is gaussian with mean $A\hat{R} + b$ and covariance operator $(\Omega_0 - A\hat{H}\Omega_0)$.

A proof of this theorem can be found in Florens *et al.* (2008) [13], here we only make some remarks. The first point of the theorem is an application of *Jirina theorem*, see [30], and in our application the condition in it is met since the space $\mathcal{X}$ we are considering is Polish, see [22]. Concerning the second part of the theorem, the characteristic function takes the form of the characteristic function of a gaussian random variable. The posterior mean is $A\hat{R}+b$ and the posterior variance is $\Omega_0 - A\hat{H}\Omega_0$. The deterministic function b has the following form: $b = (I - A\hat{H})p_0$ and operator A is determined through the equality between the two expressions for the covariance operator:

$$\begin{aligned}
\forall \phi, \psi \in \mathcal{X}, \quad Cov(< p, \phi >, < \hat{R}, \psi >) &= Cov(< \mathbb{E}(p|\hat{R}), \phi >, < \hat{R}, \psi >) \\
&= Cov(< A\hat{R}, \phi >, < \hat{R}, \psi >) \\
&= Cov(< \hat{R}, A^*\phi >, < \hat{R}, \psi >) \\
&= < (\hat{\Sigma}_T + \hat{H}\Omega_0\hat{H}^*)A^*\phi, \psi >)
\end{aligned}$$

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

and from (11)

$$Cov(< p, \phi >, < \hat{R}, \psi >) \quad = \quad < \hat{H}\Omega_0\phi, \psi > .$$

Therefore, by equating these two terms, A is defined as the solution of the functional equation:

$$(\hat{\Sigma}_T + \hat{H}\Omega_0\hat{H}^*)A^*\phi = \hat{H}\Omega_0\phi \qquad \forall\phi \in \mathcal{X}. \tag{12}$$

However, with F replaced by the estimator $\hat{F}$, which is of finite rank, the null set of operator $\hat{H}$ is not reduced to zero. So that A^* cannot be identified. Moreover, as $T \to \infty$ C_T^{-1} becomes unbounded, since the limit operator $(\Sigma_T + H\Omega_0 H^*)$ has a countable set of eigenvalues accumulating only at zero, causing A to be unbounded and the posterior mean to not be continuous in $\hat{R}$. This entails that small measurement errors in $\hat{R}$ will have a severe impact on the posterior mean of p that consequently will be prevented to be a consistent estimator (in the sampling sense).

The computation of the posterior distribution in infinite dimensional spaces requires to solve the further inverse problem (12) that is ill-posed. In the following two sections we propose two solutions based on two different regularization of the inverse of operator $(\hat{\Sigma}_T + \hat{H}\Omega_0\hat{H}^*)$ in (12); the first one uses a classical Tikhonov regularization scheme and the second one uses a Tikhonov regularization in the Hilbert scale induced by the prior.

3.4 Tikhonov Regularized Posterior Distribution

We solve the problem of unboundedness of operator A in the posterior mean function by applying a Tikhonov regularization scheme, see [23], to the inverse of operator $(\hat{\Sigma}_T + \hat{H}\Omega_0\hat{H}^*)$. We define the regularized operator A_α as:

$$A_\alpha\phi = \Omega_0\hat{H}^*(\alpha I + \hat{\Sigma}_T + \hat{H}\Omega_0\hat{H}^*)^{-1}\phi \tag{13}$$

where $\alpha > 0$ is a regularization parameter that is function of the sample size T, $\alpha = \alpha(T)$, and it is such that $\alpha \to 0$ as $T \to \infty$. This parameter must be chosen in order to balance the trade-off between the bias due to the regularization and the variance due to the instability of the inversion. Operator $(\alpha I + \hat{\Sigma}_T + \hat{H}\Omega_0\hat{H}^*)$ is surjective and then injective, so that it has a bounded inverse.

The regularized operator A_α is used to construct a new posterior distribution that we denote with $\mu_\alpha^{\mathcal{F}}$ and that we guess is the solution to the transformed pricing functional equation (10). Asymptotic arguments will justify this choice as far as it is proved, see Section 4, that $\mu_\alpha^{\mathcal{F}}$ weakly converges to the Dirac measure concentrated in p_*, where p_* is the true value of the pricing functional. We will call this convergence *posterior convergence* and it is a convergence in the sense of the sampling probability.

The *regularized posterior distribution* $\mu_\alpha^{\mathcal{F}}$ is a conditional gaussian measure on the σ-field $\mathcal{E}$ given $\mathcal{F}$, with mean and variance

$$\mathbb{E}_{\alpha}(p|\hat{R}) = A_{\alpha}(\hat{R} - \hat{H}p_0) + p_0$$
$$\Omega_{\alpha,R} = \Omega_0 - A_{\alpha}\hat{H}\Omega_0.$$

This probability measure is characterized by the estimated operator $\hat{K}$, therefore it must be meant as an estimation of the corresponding regularized posterior distribution with true K. We select as punctual estimator of the equilibrium price function the regularized posterior mean $\mathbb{E}_{\alpha}(p|\hat{R})$, as it is suggested by a quadratic loss function. This estimator is a continuous function of $\hat{R}$ and then it is consistent.

Tikhonov regularization is a stabilization procedure and it is the equivalent, in inverse problem theory, of shrinkage estimators in statistics and econometrics. These estimators are defined by adding a bias in order to stabilize the inversion. One example of shrinkage estimator is the well-known ridge regression. In particular, in finite dimensional Bayesian inverse problem, for particular choices of the prior and sampling variance, the posterior mean and the Tikhonov regularized solution coincides.

Tikhonov regularization is easy to implement but in certain situations the rate of convergence of the regularized solution, toward the true value p_*, is not optimal. More properly, when the true pricing functional p_* is highly regular, Tikhonov regularization does not permit to exploit all its regularity to reach a faster rate of convergence. This is what is called *saturation* or *qualification* effect.

3.5 Tikhonov regularization in the Prior Variance Hilbert scale

Different methods for better exploiting the regularity of function p_* have been proposed in literature. One of these is the Tikhonov regularization in Hilbert Scale, see [9] for general theory of regularization in Hilbert scale.

In this subsection, we recover A by applying a *Tikhonov regularization in the Hilbert scale induced by the inverse of the prior covariance operator.* Let $L = \Omega_0^{-\frac{1}{2}}$ be a densely defined, unbounded, self-adjoint, strictly positive operator in the Hilbert space $\mathcal{X}$ [4]. The norm $||\cdot||_s$ is defined as $||x||_s := ||L^s x||$. We define the Hilbert Scale $\mathcal{X}_s$ induced by L as the completion of the domain of L^s, $\mathcal{D}(L^s)$, with respect to the norm $||\cdot||_s$ previously defined; moreover $\mathcal{X}_s \subseteq \mathcal{X}_{s'}$ if $s' \leq s$, $\forall s \in \mathbb{R}$. Usually, when a regularization scheme in Hilbert Scale is adopted, the operator L, and consequently the Hilbert Scale, is created *ad hoc*. In the Bayesian case this regularization scheme results to be very interesting since the Hilbert Scale is not created ad-hoc but is suggested by the prior information we have and this represents a big difference and advantage with respect to the standard methods. Hence, the regularization scheme is strictly linked to the prior distribution. The following assumption is necessary in order the theory of regularization in Hilbert scale works and gives suitable rates of convergence.

Assumption 8 *(i)* $||H\Omega_0^{\frac{1}{2}}x|| \sim ||\Omega_0^{\frac{a}{2}}x||$, $\forall x \in \mathcal{X}$;

[4] More clearly, $L = \Omega_0^{-\frac{1}{2}}$ is a closed operator in $\mathcal{X}$ satisfying: $\mathcal{D}(L) = \mathcal{D}(L^*)$ is dense in $\mathcal{X}$, $< Lx, y > = < x, Ly >$ for all $x, y \in \mathcal{D}(L)$, and there exists $\gamma > 0$ such that $< Lx, x > \geq \gamma ||x||^2$ for all $x \in \mathcal{D}(L)$.

(ii) $(p_* - p_0) \in \mathcal{X}_{\beta+1}$*, i.e.* $\exists\, \rho_* \in \mathcal{X}$ *such that* $(p_* - p_0) = \Omega_0^{\frac{\beta+1}{2}} \rho_*$

(iii) $a \leq s \leq \beta + 1 \leq 2s + a$.

Some remarks about this assumption are in order. Assumption *(i)* is equivalent to say that in specifying the prior distribution we take into account the sampling model, hence the prior variance is linked to the sampling model (10) we are studying and, in particular, to operator H. This kind of prior specification is not new in Bayesian literature since it is similar to the Zellner's *g-prior*, see Zellner (1986) [36] or Agliari *et al.* (1988) [1]. The link between the prior covariance $\Omega_0^{\frac{1}{2}}$ and operator H is affected by parameter a that can be interpreted as a *degree of ill-posedness*. Therefore, the prior is specified not only by taking into account the sampling model but also the degree of ill-posedness of the problem.
Assumption *(ii)* is known as a *source condition* and is formulated in order to reach a certain speed of convergence of the regularized solution. Given Assumption 6, it says that $\delta_* \in \mathcal{R}(\Omega_0^{\frac{\beta}{2}})$, hence $\mathcal{X}_{\beta+1} \equiv \mathcal{R}(\Omega_0^{\frac{\beta+1}{2}}) \equiv \mathcal{D}(L^{\beta+1})$. The meaning of such an assumption is that the prior distribution contains information about the regularity of the true value of p. In fact, parameter β is interpreted as the *regularity parameter*. These two remarks stress the fact that we are not taking whatever Hilbert Scale, but the Hilbert Scale linked to the prior. Either we first choose the Hilbert Scale and then we use the information contained in it to specify the prior distribution or we use the information contained in the prior distribution to specify the Hilbert Scale.
The restriction $\beta + 1 \geq s$ means that the centered value of the true value p_* has to be at least an element of $\mathcal{X}_s$ and it guarantees that the norm $||L^s x||$ exists $\forall x \in \mathcal{X}_{\beta+1}$.
Under such assumptions the regularized solution in $\mathcal{X}_s$ to equation (12) is:

$$A_s = \Omega_0 \hat{H}^* (\alpha L^{2s} + \hat{\Sigma}_T + \hat{H}\Omega_0\hat{H}^*)^{-1}. \tag{14}$$

The regularized posterior distribution is thus defined similarly as in Section 3.4 with A_α substituted by A_s and is denoted with $\mu_s^{\mathcal{F}}$. The regularized posterior mean and variance are

$$\begin{aligned} \mathbb{E}_s(p|\hat{R}) &= A_s\hat{R} + (I - A_s\hat{H})p_0 \\ \Omega_{s,R} &= \Omega_0 - A_s\hat{H}\Omega_0. \end{aligned} \tag{15}$$

A classical Tikhonov regularization method allows to obtain a rate of convergence to zero of the regularization bias that is at most of order 2; on the contrary with a Tikhonov scheme in an Hilbert Scale the smoother the function p_* is, the faster the rate of convergence to zero of the regularization bias will be. An other regularization method that allows to obtain a similar improvement over the classical Tikhonov scheme is the iterate Tikhonov regularization scheme.

4 Asymptotic Analysis

A very important result, due to Doob (1949), see [8] and [11], states that for any prior, the posterior distribution is consistent in the sense that it converges to a point mass at the unknown parameter that is outside a set of prior mass zero. Actually, no one can be so certain about the prior and values of the parameter for which consistency is not verified may be obtained. To move around this problem it is customary to use a frequentist notion of consistency, the idea of which consists in thinking the data as generated from a distribution characterized by the true value of the parameter and in checking the accumulation of the posterior distribution in a neighborhoods of this true value.

Hence, in according to the "classical bayesian" point of view, we assume there exists a true value of the price functional, already denoted with p_*, and we check that the regularized posterior distribution becomes more and more accurate and precise, around p_*, as the number of observed data increases indefinitely. Thus, it is a convergence in the sampling probability sense and it is known as *consistency of the posterior distribution.*

Following Diaconis *et al.* (1986) [7] we give the following definition of *posterior consistency* (or consistency in the sampling sense):

Definition 1 *The pair $(p, \mu^{\mathcal{F}})$ is consistent if $\mu^{\mathcal{F}}$ converges weakly to δ_p as $T \to \infty$ under Q^p-probability or Q^p-a.s., where δ_p is the Dirac measure in p.*
The posterior probability $\mu^{\mathcal{F}}$ is consistent if $(p, \mu^{\mathcal{F}})$ is consistent for all p.

If $(p, \mu^{\mathcal{F}})$ is consistent in the previous sense, the Bayes estimate for p, for instance the posterior mean for a quadratic loss function, is consistent too.

The meaning of this definition is that, for any neighborhood U of the true parameter p_*, the posterior probability of the complement of U converges toward zero when $T \to \infty$: $\mu^{\mathcal{F}}(U^c) \to 0$ in Q^p-probability, or Q^p-*a.s.* Therefore, since distribution expresses one's knowledge about the parameter, consistency stands for convergence of knowledge towards the perfect knowledge with increasing amount of data.

It is appropriate to separate Bayesians into two groups: "classical" and "subjectivist". Classical bayesians believes there exists a true value of the parameter that has generated the data, therefore they care for, as data set becomes large, the posterior converging to a point mass at the true parameter. In point of fact, consistency is interesting also for subjective Bayesian for different reasons (e.g. "intersubjective agreement" or to check if the posterior is a correct representation of the updated prior, see Florens *et al.* (1990) [11]).

Having a posterior distribution, and hence a bayesian estimator, that is consistent in the sampling sense justifies, also from a classical non-bayesian point of view, our estimator obtained with a bayesian approach. On the basis of this argument we are persuaded about the importance of studying posterior consistency and in this section we apply this concept of consistency to the regularized posterior distribution. By *Chebyshev's Inequality* in L^2 space we have, for any sequence $M_n \to \infty$:

$$\begin{aligned}\mu_\alpha^{\mathcal{F}}\{p : ||p - p_*|| \geq M_n\varepsilon_n\} &\leq \frac{\mathbb{E}_\alpha(||p - p_*||^2|\hat{R})}{(M_n\varepsilon_n)^2} \\ &= \frac{1}{(M_n\varepsilon_n)^2}[< \Omega_{\alpha,R}1, 1 > + ||\mathbb{E}_\alpha(p|\hat{R}) - p_*||^2] \\ &\leq \frac{||\Omega_{\alpha,R}|| + ||\mathbb{E}_\alpha(p|\hat{R}) - p_*||^2}{(M_n\varepsilon_n)^2}. \end{aligned} \tag{16}$$

The same inequality is valid for $\mu_s^{\mathcal{F}}$.

4.1 Speed of convergence with classical Tikhonov regularization

We begins by checking posterior consistency of the regularized posterior $\mu_\alpha^{\mathcal{F}}$ computed with the classical Tikhonov, namely we check accumulation of $\mu_\alpha^{\mathcal{F}}$ to the point mass δ_{p_*}. The main results are contained in the following theorem.

Theorem 3 *Let p_* be the true value of the asset pricing functional and $\mu_\alpha^{\mathcal{F}}$ a gaussian measure on $\mathcal{X}$ with mean $A_\alpha(\hat{R} - \hat{H}p_0) + p_0$ and covariance operator $\Omega_{\alpha,R}$. Under Assumption 6, and if $\alpha \to 0$, $\alpha^2 T \to \infty$, then:*

(i) $\mu_\alpha^{\mathcal{F}}$ weakly converges towards a point mass δ_{p_*} in p_*;

(ii) if moreover $\delta_* \in \mathcal{R}(\Omega_0^{\frac{1}{2}} H^* H \Omega_0^{\frac{1}{2}})^{\frac{\beta}{2}}$ for some $\beta > 0$, then

$$\begin{aligned}\mu_\alpha^{\mathcal{F}}\{p : ||p - p_*|| \geq \varepsilon_T\} \sim\ & \mathcal{O}_p(\alpha^{\frac{\beta}{2}} + \frac{1}{\alpha_T T} + \frac{1}{\alpha}\Big(\frac{1}{T} + h^{2\rho}\Big)^{\frac{1}{2}}\alpha^{\frac{\beta}{2}} + \frac{1}{\alpha^2 T}\frac{1}{\alpha}\Big(\frac{1}{Th^n} + h^{2\rho}\Big)^{\frac{1}{2}} \\ & + \frac{1}{\alpha^2 T}\alpha^{\frac{(\beta+1)}{2}\wedge 1}).\end{aligned}$$

It should be noted that the condition for the second part of the theorem is only a regularity condition that is necessary for having convergence at a certain speed. The condition that really matters is the fact that the centered true parameter must belong to the *Reproducing Kernel Hilbert Space* associated to Ω_0, *i.e.* $(p_* - p_0) \in \mathcal{H}(\Omega_0)$.
The support of a centered gaussian process, taking its value in an Hilbert space $\mathcal{X}$, is the closure in $\mathcal{X}$ of the *Reproducing Kernel Hilbert Space* associated with the covariance operator of this process, see [35]. Then, for the prior distribution, $(p - p_0) \in \overline{\mathcal{H}(\Omega_0)}$ with μ-probability 1, but with μ-probability 1, $p - p_0$ is not in $\mathcal{H}(\Omega_0)$. Hence, the prior distribution is not able to generate trajectories that satisfy Assumption 6, or in other words, the true value of the price functional p_*. This concept is known in literature as *prior inconsistency* and it refers to a prior that is unable to generate the true parameter characterizing the data generating process. This problem is present only for infinite dimensional parameter sets and it is due to the fact that it is difficult to be sure about a prior on an infinite dimensional parameter space so that it can happen that the true value of the parameter is not in the support of the prior, see e.g. [15] or [16].
Anyway, if Ω_0 is one-to-one, $\mathcal{H}(\Omega_0)$ is dense in $\mathcal{X}$ and since the support of μ is the closure

$\overline{\mathcal{H}(\Omega_0)}$, this measure is able to generate trajectories as close as possible to the true one. The next corollary states consistency of the regularized posterior mean and convergence to zero of the regularized posterior variance; it provides the necessary results for having Theorem 3.

Corollary 1 *Under Assumption 6, and if $\alpha \to 0$, $\alpha^2 T \to \infty$, then:*

(i) $||\hat{\mathbb{E}}_\alpha(p|\hat{R}) - p_*|| \to 0$ *in Q^{p_*}-probability and if $\Omega_0^{-\frac{1}{2}}(p_* - p_0) \in \mathcal{R}(\Omega_0^{\frac{1}{2}} H^* H \Omega_0^{\frac{1}{2}})^{\frac{\beta}{2}}$ for some $\beta > 0$,*

$$\begin{aligned}||\hat{\mathbb{E}}_\alpha(p|\hat{R}) - p_*||^2 \quad &\sim \quad \mathcal{O}_p(\alpha^\beta + \frac{1}{(\alpha^2 T)^2}\alpha^{(\beta+1)\wedge 2} + \frac{1}{\alpha T} + \\ &\frac{1}{\alpha^2}\Big(\frac{1}{T} + h^{2\rho}\Big)\alpha^\beta + \frac{1}{\alpha^2 T}\frac{1}{T}\frac{1}{\alpha^2}\Big(\frac{1}{Th^n} + h^{2\rho}\Big)).\end{aligned}$$

(ii) $||\Omega_{\alpha,R}|| \to 0$ *in P^{φ_*}-probability and $\forall \phi \in \mathcal{X}$ such that $\Omega_{\alpha,R}\phi \in \mathcal{R}(\Omega_0^{\frac{1}{2}} H^* H \Omega_0^{\frac{1}{2}})^{\frac{\beta}{2}}$ for some $\beta > 0$,*

$$||\Omega_{\alpha,R}\phi||^2 \sim \mathcal{O}_p(\alpha^\beta + \frac{1}{\alpha^2}\Big(\frac{1}{T} + h^{2\rho}\Big)\alpha^\beta + \frac{1}{(\alpha^2 T)^2}\frac{1}{\alpha^2}\Big(\frac{1}{Th^n} + h^{2\rho}\Big) + \frac{1}{(\alpha^2 T)^2}\alpha^{(\beta+1)\wedge 2}).$$

It should be noted that we have introduced the qualification 2 since, with Tikhonov regularization, the rate of convergence cannot exceed α^2.
The rate of convergence to zero of the posterior variance is negligible with respect to the rate in the bias, so that the optimal parameter of regularization will be chosen by taking into account the rate of the squared norm of the bias. Concerning this rate, only the first and third terms matter, being the other three terms negligible for particular choices of β and of the bandwidth h. While the first rate α^β requires a regularization parameter α going to zero as fast as possible, the third one requires an α going to zero as slow as possible. In choosing the regularization parameter we should take into account this trade-off, hence, the optimal regularization parameter α_* will be obtained when the two rates are equated: $\alpha^\beta = \frac{1}{\alpha T}$, so that

$$\alpha^* \propto T^{-\frac{1}{\beta+1}}.$$

The optimal rate of convergence of the squared norm of the regularized posterior mean and variance is $T^{-\frac{\beta}{\beta+1}}$, while the optimal rate of the regularized posterior distribution is $T^{-\frac{\beta}{2(\beta+1)}}$. Let analyze conditions on β and h to guarantee convergence to zero of the other rates in the bias. A sufficient condition for $\frac{1}{(\alpha^2 T)^2}\alpha^{(\beta+1)\wedge 2}$ converging to zero is that $\frac{1}{(\alpha^2 T)} \sim \mathcal{O}_p(1)$, *i.e.* $\alpha^2 \sim \mathcal{O}_p\Big(\frac{1}{T}\Big)$. With α replaced by its optimal value, this condition is met for $\beta \geq 1$. For $\frac{1}{\alpha_T^2}\Big(\frac{1}{T} + h^{2\rho}\Big)\alpha_T^\beta$ being negligible we have to choose h in such a way that $h^{2\rho} \sim \mathcal{O}_p(\frac{1}{T})$, *i.e.*

$$h \propto \Big(\frac{1}{T}\Big)^{\frac{1}{2\rho}}$$

For the last rate $\frac{1}{\alpha^4 T^2}(\frac{1}{Th^n}+h^{2\rho})$ converging to zero we simply have to check that $\frac{1}{\alpha^2 Th} \sim \mathcal{O}_p(1)$ since the second term is $o_p(1)$ due to the choice of h and to the fact that $\frac{1}{(\alpha T)^2} \sim o_p(1)$. Then, $\frac{1}{\alpha^4 T^2}\frac{1}{Th} = (\frac{1}{T})^{\frac{-4}{\beta+1}+3-\frac{n}{2\rho}}$ and it goes to zero if $\beta > \frac{2\rho+n}{6\rho-n}$. This constraint is bounding, with respect to the constraint $\beta \geq 1$, when $\frac{n}{6} \leq \rho \leq \frac{n}{2}$; otherwise, the only constraint guaranteing decay of all the rates is $\beta \geq 1$.
Lastly, it should be noted that the forth rate of the squared norm of the regularized variance operator can be rewritten as $\frac{1}{(\alpha^2 T)}\frac{1}{(\alpha^4 T)}\left(\frac{1}{Th^n}+h^{2\rho}\right)$ that goes to zero if conditions for ensuring converges to zero of the terms in the bias are satisfied.

4.2 Speed of convergence with Tikhonov regularization in the Prior Variance Hilbert Scale

We compute in this subsection the speed of convergence for $\mu_s^{\mathcal{F}}$. The speed obtained in this case is faster than that one with a simple Tikhonov regularization scheme. In this section we suppose Assumption 8 holds, the attainable speed of convergence is given in the following theorem, the proof of which is provided in Appendix 8.4.

Theorem 4 *Let $\mathbb{E}_s(x|\hat{Y})$ and V_s be as in (15). Under Assumptions 6, 7 and 8*

$$\begin{aligned}||\mathbb{E}_s(p|\hat{R}) - p_*||^2 \quad \sim \quad & \mathcal{O}_p(\alpha^{\frac{\beta+1}{a+s}} + \alpha^{\frac{1-a}{a+s}}\frac{1}{T} + \frac{1}{\alpha^4}\frac{1}{T^2}\alpha^{\frac{a+\beta+2s}{a+s}} + \alpha^{\frac{\beta+1}{a+s}}\frac{1}{\alpha^2}\Big(\frac{1}{T}+h^{2\rho}\Big) \\ & + \frac{1}{\alpha^3}\frac{1}{T^2} + \alpha^{\frac{1-2a-s}{a+s}}\Big(\frac{1}{T}+h^{2\rho}\Big)\frac{1}{\alpha T} + \alpha^{\frac{1-2a-s}{a+s}}\Big(\frac{1}{T}+h^{2\rho}\Big)\frac{1}{T}.\end{aligned}$$

Moreover, if the covariance operator V_s is applied to elements $\varphi \in \mathcal{X}$ such that $\Omega_0^{\frac{1}{2}}\varphi \in \mathcal{R}(\Omega_0^{\frac{\beta}{2}})$, then

$$||\Omega_{s,R}\varphi||^2 \sim \mathcal{O}_p(\alpha^{\frac{\beta+1}{a+s}} + \frac{1}{\alpha^4 T^2}\alpha^{\frac{2s+a+\beta}{a+s}} + \alpha^{\frac{\beta+1}{a+s}}\frac{1}{\alpha^2}\Big(\frac{1}{T}+h^{2\rho}\Big) + \frac{1}{\alpha^3 T^2}).$$

The optimal α_n is obtained by equating the first two rates of convergence of the posterior mean:

$$\alpha_n^* \propto (\frac{1}{T})^{\frac{a+s}{a+\beta}}$$

and the optimal bandwidth is determined in the same way as before, hence $h = c_1(\frac{1}{T})^{\frac{1}{2\rho}}$, with c_1 a given constant. With this optimal choice of the regularization parameter, for guaranteing the other terms in the bias and variance are of order $o_p(1)$, we have to restrict the values of β. In particular, if $a+2s \geq 3$ then the regularity parameter must satisfy $\frac{2s+a-1}{2} < \beta < 2s+a-1$; otherwise $\frac{2s+a}{3} < \beta < 2s+a-1$. The corresponding optimal speed of the squared bias and variance is proportional to $(\frac{1}{T})^{\frac{\beta+1}{a+\beta}}$, while the regularized posterior distribution $\mu_s^{\mathcal{F}}$ is of order $\mathcal{O}_p((\frac{1}{T})^{\frac{\beta+1}{2(a+\beta)}})$. It should be noted that parameter s characterizing the norm in the Hilbert scale does not play any role on the speed of convergence.
An advantage of the Tikhonov regularization in Hilbert Scale is that we can even obtain a

rate of convergence for other norms, namely $||\cdot||_r$ for $-a \leq r \leq \beta+1 \leq a+2s$. Actually, the speed of convergence of these norms gives the speed of convergence of the estimate of the r-th derivative of the parameter of interest p.

Tikhonov regularization in Hilbert scale improves the speed of convergence of the regularized posterior distribution with respect to the classical Tikhonov regularization. Let call γ the regularity parameter of function $(p_* - p_0)$, instead of β as was used in subsection 4.1. This is for differentiating with respect to the regularity parameter in Hilbert scale that will continue to be denoted with β. If Assumption 8 *(i)* holds, it implies the equivalence $||(\Omega_0^{\frac{1}{2}} H^* H \Omega_0^{\frac{1}{2}})^{\frac{\gamma}{2}} v|| \sim ||\Omega_0^{\frac{a\gamma}{2}} v||$, for some $v \in \mathcal{X}$. Then, $||\Omega_0^{\frac{\beta}{2}} v|| \sim ||\Omega_0^{\frac{a\gamma}{2}} v||$ if and only if $\beta = a\gamma$. The optimal bayesian speed of convergence with an Hilbert scale is $\left(\frac{1}{T}\right)^{\frac{a\gamma+1}{a(1+\gamma)}}$ that is fastest than the bayesian speed of convergence with a classical Tikhonov: $\left(\frac{1}{T}\right)^{\frac{\gamma}{\gamma+1}}$, $\forall \gamma > 0$.

4.3 Comparison with the classical estimation of the pricing functional

We develop in this paragraph a comparison between the bayesian method we have proposed in this paper for recovering the asset pricing functional and the classical solution to the integral equation (6) computed in Carrasco *et al.* (2007) [4]. The classical solution does not require the use of any regularization scheme since the operator $(I-K)$ is continuously invertible. Since K is unknown it is substituted by $\hat{K}$ as defined in subsection 3.1, the estimated pricing functional $\hat{p}$ is

$$\hat{p} = (I - \hat{K})^{-1}\hat{r},$$

with $\hat{r}$ defined in subsection 3.1. By applying Theorem 7.2 in Carrasco *et al.* [4] the squared norm of the asymptotic bias is of order

$$||\hat{p} - p_*||^2 \sim \mathcal{O}_p\Big(\frac{1}{Th^n} + h^{2\rho}\Big).$$

The optimal speed of convergence is obtained when $\frac{1}{Th^n} = h^{2\rho}$, that is when $h = c_1(\frac{1}{T})^{\frac{1}{2\rho+n}}$. With this optimal choice of bandwidth the classical estimator $\hat{p}$ converges at the rate of $(\frac{1}{T})^{\frac{2\rho}{2\rho+n}}$: $||\hat{p} - p_*||^2 \sim \mathcal{O}_p((\frac{1}{T})^{\frac{2\rho}{2\rho+n}})$.

We compare this rate of convergence with the rate of the estimated regularized posterior mean obtained when the optimal α is used: $||\hat{\mathbb{E}}_\alpha(p|\hat{R}) - p_*||^2 \sim \mathcal{O}_p((\frac{1}{T})^{\frac{\beta}{\beta+1}})$. Our solution converges faster if $\beta > \frac{2\rho}{n}$. This condition is more likely to be satisfied when the parameter ρ (that is a measure of regularity of the transition density function) is small or equivalently, for a given value of ρ, when the dimension of Y_t, *i.e.* the number of conditioning variables in the transition probability, increases.

Anyway, with Tikhonov regularization the qualification matters so that we can only exploit a regularity β of the function p that is less or equal than 2. Therefore, in order condition $\beta > \frac{2\rho}{n}$ is satisfied, it must be $\frac{2\rho}{n} \leq 2$, that holds when $\rho \leq n$.

Let us consider the regularized posterior mean obtained through a Tikhonov scheme in

Hilbert scale. In this case and with the optimal regularization parameter α_* the rate of convergence is $||\mathbb{E}_s(p|\hat{R}) - p_*||^2 \sim \mathcal{O}_p((\frac{1}{T})^{\frac{\beta+1}{a+\beta}})$ and it is faster than the rate of convergence with classical solution if $\beta > \frac{2\rho(a-1)}{n} - 1$. When $a > 2$ and $\rho < \frac{n}{2(a-2)}$, this condition is less stringent than condition $\beta > \frac{2\rho}{n}$, demanded for Tikhonov regularized posterior mean converging faster than the classical estimator $\hat{p}$. When the degree of ill-posedness a is less than 2, then the condition $\beta > \frac{2\rho(a-1)}{n} - 1$ is less stringent than condition $\beta > \frac{2\rho}{n}$ if $\rho > \frac{n}{2(a-2)}$.
Summarizing, under some condition on the regularity of the function p_*, in particular if the price function is highly smooth, or if n is high or ρ is small, our Bayesian estimator converges faster than the classical one. The price to pay for having this fastest speed of convergence is to impose a regularity assumption on the price functional that we do not impose with the classical resolution method.

5 A g-prior with Regularizing Power

We have shown in preceding sections that, in general, the prior distribution does not regularize and we need to artificially introduce a regularization scheme in order to obtain consistency of the posterior distribution.
Nevertheless, there exists some kind of prior distribution, known as g-prior, that, under some condition, has a regularizing power so that the prior-to-posterior transformation has the same effect as the application of a regularization scheme. The regularizing power of this kind of prior distribution is studied in Florens *et al.* (2008). Let suppose that the prior measure specified in 3.2.1 is replaced by a g-prior of the type introduced by Zellner (1986) [36]:

$$p \sim \mathcal{GP}\Big(x_0, \frac{\sigma^2}{g}(K^*K)^s\Big), \qquad \text{for some } s > 0 \tag{17}$$

with $g = g(T)$ a function of the sample size T such that $g \to \infty$ with T. Let $\alpha = \frac{1}{T}g$ be a positive parameter going to zero with T and such that $\alpha^2 T \to \infty$. For this being true g must goes to infinity faster than $\sqrt{T}$ and slower than T.
Equation (12) implies an operator $A = (K^*K)^s\hat{H}^*(\alpha(K^*K) + \hat{H}(K^*K)^s\hat{H}^*)^{-1}$ that, as $T \to \infty$, is well-defined if it is applied to $(\hat{R} - \hat{H}p_0)$. The fact that (K^*K) multiplying α can be factorized out allows to directly obtain a regularization of the inverse of the limit of $(K^*K)^{-\frac{1}{2}}\hat{H}(K^*K)^s\hat{H}^*(K^*K)^{\frac{1}{2}}$. Using equation (12) for defining A we have

$$\begin{aligned} A &= \frac{\sigma^2}{g}(K^*K)^s\hat{H}^*(\hat{\Sigma}_T + \frac{\sigma^2}{g}\hat{H}(K^*K)^s\hat{H}^*)^{-1} \\ &= ((K^*K)^{-\frac{1}{2}}\hat{H}(K^*K)^s)^*(\alpha I + (K^*K)^{-\frac{1}{2}}\hat{H}(K^*K)^s\hat{H}^*(K^*K)^{-\frac{1}{2}})^{-1}(K^*K)^{-\frac{1}{2}} \end{aligned}$$

that is a continuous operator even in the limit for $T \to \infty$ if it is applied to $(\hat{R} - \hat{H}p_0)$. This is due to the fact that $\mathcal{R}(K^*K) \subset \mathcal{R}(K) = \mathcal{D}((K^*)^{-1}) \subset \mathcal{D}((K^*K)^{-\frac{1}{2}})$. The posterior mean and variance are $\mathbb{E}^g(p|\hat{R}) = A(\hat{R} - \hat{H}p_0) + p_0$ and $Var^g(p|\hat{R}) = (K^*K)^s - A\hat{H}(K^*K)^s$.

Because operators K and K^* are unknown, it follows that they must be substituted by their consistent estimators in all the previous expressions. We denote with $\hat{\mathbb{E}}^g(p|\hat{R})$ and $\widehat{Var}^g(p|\hat{R})$ the corresponding estimated mean and variance.
Study of asymptotic behavior of the posterior distribution is based on the decompositions:

$$\begin{aligned}\hat{\mathbb{E}}^g(p|\hat{R}) - p_* &= [\hat{\mathbb{E}}^g(p|\hat{R}) - \tilde{\mathbb{E}}^g(p|\hat{R})] + [\tilde{\mathbb{E}}^g(p|\hat{R}) - \mathbb{E}^g(p|\hat{R})] + [\mathbb{E}^g(p|\hat{R}) - p_*] \\ \widehat{Var}^g(p|\hat{R}) &= [\widehat{Var}^g(p|\hat{R}) - \widetilde{Var}^g(p|\hat{R})] + [\widetilde{Var}^g(p|\hat{R}) - Var^g(p|\hat{R})] + Var^g(p|\hat{R}).\end{aligned}$$

The only difference between $\hat{\mathbb{E}}^g(p|\hat{R})$ and $\tilde{\mathbb{E}}^g(p|\hat{R})$ is that in the first one the prior covariance operator is estimated while in the latter it is known. The same difference characterizes $\widehat{Var}^g(p|\hat{R})$ and $\widetilde{Var}^g(p|\hat{R})$. Hence, the first term of both the two decompositions above is due to estimation of Ω_0, the second error is due to estimation of all the other operators and the last one is the quantity we have to check when operators are known.
We show in the following theorem that the posterior distribution corresponding to the g-prior is consistent. This is guaranteed by convergence to zero of the bias an the posterior variance.

Theorem 5 *Let (17) be the prior distribution for the functional p in the sampling equation (10). If, for some $\gamma > 0$, $(K^*K)^\gamma$ is trace class and if $(p_* - p_0) \in \mathcal{R}(\Omega_0^{\frac{\beta}{2s}})$ then $||\mathbb{E}^g(p|\hat{R}) - p_*||^2$ converges to zero with respect to the sampling probability at the speed*

$$\begin{aligned}||\hat{\mathbb{E}}^g(p|\hat{R}) - p_*||^2 \sim\ & \mathcal{O}_p\Big(\alpha^{\frac{2\beta}{2s+1}} + \frac{1}{T}\alpha^{\frac{-2(1+\gamma)}{2s+1}} + \frac{1}{\alpha^2}\Big(\frac{1}{Th^n} + h^{2\rho}\Big)(\alpha^{\frac{2\beta-1+4s}{2s+1}} + \frac{1}{T}\alpha^{\frac{2(s-\gamma-1)}{2s+1}}) \\ & + \Big(\frac{1}{T} + h^{2\rho}\Big)\Big(\alpha^{\frac{\beta-3s-2}{2s+1}} + \frac{1}{T}\alpha^{\frac{2(s-2-\gamma 2)}{2s+1}}\Big) + \frac{1}{T}\alpha^{-\frac{4}{2s+1}}\Big).\end{aligned}$$

Furthermore, if $\alpha = c_1(\frac{1}{T})^{\frac{2s+1}{2(\beta+\gamma+1)}}$, $h = c_2(\frac{1}{T})^{\frac{1}{2\rho}}$ for some constants c_1 and c_2,

$$T^{\frac{\beta}{\beta+\gamma+1}}||\mathbb{E}(p|\hat{R}) - p_*||^2 \sim \mathcal{O}_p(1)$$

if $s \geq 2$, $\beta > \frac{s-\gamma}{2}$, $\beta > \frac{3s-2\gamma}{3}$ and $\beta > \frac{(\gamma+1)n-2\rho(\gamma-s)}{4\rho-n}$ if $4\rho > n$ (or $\beta < \frac{(\gamma+1)n-2\rho(\gamma-s)}{4\rho-n}$ if $4\rho < n$).

The optimal regularization parameter has been obtained by equating the first two rates of the posterior mean, since if $\gamma > 1$, $\alpha^{\frac{-2(1+\gamma)}{2s+1}}$ explodes faster than $\alpha^{-\frac{4}{2s+1}}$.

Theorem 6 *Let (17) be the prior distribution for the functional p in the sampling equation (10). If $s \geq 2$ then $||\widehat{Var}^g(p|\hat{R}) - p_*||^2$ converges to zero with respect to the sampling probability. Moreover, $\forall\phi \in \mathcal{X}$ such that $\Omega_0^{\frac{1}{2}}\phi \in \mathcal{R}(\Omega_0^{\frac{\beta-s}{2s}})$, the posterior variance converges at the speed*

$$\begin{aligned}||\widehat{Var}^g(p|\hat{R}) - p_*||^2 \sim\ & \mathcal{O}_p\Big(\alpha^{\frac{2\beta}{2s+1}} + \Big(\frac{1}{Th^n} + h^{2\rho}\Big)\alpha^{\frac{2\beta-1+4s}{2s+1}} + \Big(\frac{1}{T} + h^{2\rho}\Big)\alpha^{\frac{2\beta-1+4s}{2s+1}} \\ & + \alpha^{\frac{2\beta}{2s+1}}\Big).\end{aligned}$$

When α is set equal to the optimal one, $\alpha = c_1(\frac{1}{T})^{\frac{2s+1}{2(\beta+\gamma+1)}}$, the posterior variance converges to zero if $s > \frac{n-2\rho}{2\rho}(\beta+\gamma+1)$ and $\beta > \frac{2s-\gamma}{2}$.
The value of g corresponding to the optimal α is: $g = (\frac{1}{T})^{\frac{2(s-\beta-\gamma)-1}{2(\beta+\gamma+1)}}$. It converges at infinite faster than $\sqrt{T}$ and slower than T if $\beta > 2s - \gamma$. In particular, converges at a slower rate than T is always guaranteed.

6 Prior on the Variance Parameter

Until now we have considered the variance parameter σ^2 in the covariance operator of the sampling measure as known. This parameter is the variance of the white noise in the regression model (5) defined by the Lucas' equilibrium model. In the reality this parameter is often unknown and need to be estimated. In this section, we redefine the Bayesian experiment in order to incorporate the parameter space of definition of the variance parameter σ^2: $(\mathbb{R}_+, \mathcal{B}, \nu)$, with $\mathcal{B}$ the Borel σ-field and ν a measure on it.
There exists two possibilities to specify the probability measure on the parameter space. The traditional approach calls for a conjugate model with a joint distributions on the sample space that is separable in a marginal on $\mathbb{R}_+$ and a conditional μ^σ, given $\mathcal{B}$, on $\mathcal{X}$. New developments in Bayesian literature propose more and more models in which the prior distribution on the parameter space is the product of two marginal independent distributions. In the next section we only develop the first case since in this case it is possible to define a closed form for the marginal posterior distribution of both the parameters.

6.1 Conjugate model

We start by analyzing the more simple approach, simpler at least for numerical implementation since it gives explicit forms for posterior distributions. The modified Bayesian experiment is

$$\Xi_\sigma = (\mathbb{R}_+ \times \mathcal{X} \times \mathcal{X},\, \mathcal{B} \otimes \mathcal{E} \otimes \mathcal{F},\, \Pi = \nu \times \mu^\sigma \times Q^{\sigma,p}).$$

μ^σ represents the conditional prior distribution for p conditioned on σ^2: $\mu^\sigma \sim \mathcal{GP}(p_0, \sigma^2\Omega_0)$.
$Q^{\sigma,p}$ denotes the sampling distribution conditional on both the parameters, then the covariance operator associated to it is $\frac{\sigma^2}{T}\hat{K}\hat{K}^*$.
We take, as prior distribution for the variance parameter σ^2, an *Inverse Gamma* distribution: $\sigma^2 \sim \Gamma^{-1}(v_0, s_0^2)$, with v_0 and s_0^2 two known parameters.
A conjugate model allows to easily integrate out p from the sampling distribution by using the prior μ^σ so that we obtain a sampling measure Q^σ depending only on σ^2:

$$\begin{aligned} \sigma^2 &\sim \Gamma^{-1}(v_0, s_0^2) \\ \hat{R}|\sigma^2 &\sim \mathcal{GP}(\hat{H}p_0, \sigma^2(\frac{1}{T}\hat{K}\hat{K}^* + \hat{H}\Omega_0\hat{H}^*)). \end{aligned}$$

Anyway, computation of the posterior of σ^2 is not trivial due to the fact that, because $\hat{R}$ is finite dimensional, we do not have a likelihood function. We make up for this lack by using the projected observations $\hat{R}$ projected by using the eigenfunction associated to the covariance operator $(\frac{1}{T}\hat{K}\hat{K}^* + \hat{H}\Omega_0\hat{H}^*)$. Let $\{\hat{\lambda}_j, \hat{\varphi}_j\}_{j=1}^J$ be the eigensystem associated to this operator; this eigensystem is actually an estimation of the eigensystem associated to the true covariance operator $(\frac{1}{T}KK^* + H\Omega_0 H^*)$ that we had if K was known. Moreover, the convergence $||(\frac{1}{T}\hat{K}\hat{K}^* + \hat{H}\Omega_0\hat{H}^*) - (\frac{1}{T}KK^* + H\Omega_0 H^*)|| \to 0$ implies that the eigensystem $\{\hat{\lambda}_j, \hat{\varphi}_j\}$ converges uniformly to the $\{\lambda_j, \varphi_j\}$. Thus, when the sample size is finite, we only have a finite number of eigenvalues $\hat{\lambda}_j$ different than 0. The projected observation $< \hat{R}, \hat{\varphi}_j >$ is normally distributed with mean and variance

$$
\begin{aligned}
\mathbb{E}(< \hat{R}, \hat{\varphi}_j > |\sigma^2) &= < \mathbb{E}(\hat{R}|\sigma^2), \hat{\varphi}_j > \\
&= < \hat{H}p_0, \hat{\varphi}_j > \\
Var(< \hat{R}, \hat{\varphi}_j > |\sigma^2) &= < Var(\hat{R}|\sigma^2), \hat{\varphi}_j > \\
&= \sigma^2 < (\frac{1}{T}\hat{K}\hat{K}^* + \hat{H}\Omega_0\hat{H}^*)\hat{\varphi}_j, \hat{\varphi}_j > \\
&= \sigma^2\hat{\lambda}_j,
\end{aligned}
$$

and $< \hat{R}, \hat{\varphi}_j >$ is independent of $< \hat{R}, \hat{\varphi}_i >$, $\forall j \neq i$ due to orthogonality between eigenfunctions. It should be noted that if operator K was known we would know all its eigensystem and then we would know the variance parameter σ^2, in fact $\frac{<\hat{R}-Hp_0,\varphi_j>^2}{\lambda_j}|\sigma^2 \sim \sigma^2\chi_1^2$ with mean equal to σ^2. Then, $\frac{1}{J}\sum_{j=1}^J \frac{<\hat{R}-Hp_0,\varphi_j>^2}{\lambda_j} \to \sigma^2$ and we know the limit since we know all the eigenvalues.
From classical computations we obtain the posterior distribution $\nu^{\mathcal{F}}$ of σ^2 given the sample $< \hat{R}, \hat{\varphi}_1 >, \dots, < \hat{R}, \hat{\varphi}_J >$:

$$
\nu(\sigma^2|\{< \hat{R}, \hat{\varphi}_j >\}_{j=1}^J) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{v_0+J}{2}+1} \exp\Big\{ -\frac{1}{2\sigma^2}[s_0^2 + \sum_{j=1}^J \frac{1}{\hat{\lambda}_j}(< \hat{R} - \hat{H}p_0, \hat{\varphi}_j >)^2]\Big\}
$$

then

$$
\begin{aligned}
\sigma^2|\{< \hat{R}, \hat{\varphi}_j >\}_{j=1}^J &\sim \Gamma^{-1}(v_*, s_*^2), \\
v_* = v_0 + J, &\qquad s_*^2 = s_0^2 + \sum_{j=1}^J \frac{1}{\hat{\lambda}_j}(< \hat{R} - \hat{H}p_0, \hat{\varphi}_j >)^2 \\
\mathbb{E}(\sigma^2|\{< \hat{R}, \hat{\varphi}_j >\}_{j=1}^J) &= \frac{s_*^2}{v_0 + J - 2}, \qquad Var(\sigma^2|\{< \hat{R}, \hat{\varphi}_j >\}_{j=1}^J) = \frac{\frac{s_*^4}{4}}{(\frac{v_*}{2} - 1)^2(\frac{v_*}{2} - 2)}.
\end{aligned}
$$

In order to compute the posterior distribution for p we first need to compute the conditional posterior distribution of p given σ^2, denoted with $\mu^{\mathcal{F},\sigma}$ and then to integrate out σ^2 by using its posterior distribution.
Also in this case, problems of continuity of $\mu^{\mathcal{F},\sigma}$ require some technique of regularization.

For simplicity, we consider only a classical Tikhonov regularization scheme. Extension to other regularization schemes is immediate. The regularized conditional posterior distribution, denoted with $\mu_\alpha^{\mathcal{F},\sigma}$ is a gaussian process with mean function and covariance operator given by:

$$\begin{aligned}\mathbb{E}_\alpha(p|\hat{R},\sigma^2) &= \Omega_0\hat{H}^*(\alpha I + \frac{1}{T}\hat{K}\hat{K}^* + \hat{H}\Omega_0\hat{H}^*)^{-1}(\hat{R} - \hat{H}p_0) + p_0 \\ Var_\alpha(p|\hat{R},\sigma^2) &= \sigma^2[\Omega_0 - \Omega_0\hat{H}^*(\alpha I + \frac{1}{T}\hat{K}\hat{K}^* + \hat{H}\Omega_0\hat{H}^*)^{-1}\hat{H}\Omega_0],\end{aligned}$$

where α still denotes the regularization parameter. While the regularized conditional posterior mean does not depend on σ^2, so that $\mathbb{E}_\alpha(p|\hat{R},\sigma^2) = \mathbb{E}_\alpha(p|\hat{R})$, the regularized conditional posterior variance does and then we need to integrate out σ^2 with respect to $\nu^{\mathcal{F}}$. With analogy to the finite dimensional case, this integration transform the posterior of p in a *Student process*. We refer to Florens and Simoni (2007) [12] for a definition of this process. Thus the marginal regularized posterior distribution $\mu_\alpha^{\mathcal{F}}$ for p is Student with parameters v_*, $\mathbb{E}_\alpha(p|\hat{R})$ and $\frac{s_*^2}{v_0+J}[\Omega_0 - A_\alpha\hat{H}\Omega_0]$:

$$\begin{aligned}p|\hat{R} &\sim \mathcal{S}t\mathcal{P}(\mathbb{E}_\alpha(p|\hat{R}), \frac{s_*^2}{v_*}[\Omega_0 - A_\alpha\hat{H}\Omega_0], v_*) \\ \mathbb{E}_\alpha(p|\hat{R}) &= \Omega_0\hat{H}^*(\alpha I + \frac{1}{T}\hat{K}\hat{K}^* + \hat{H}\Omega_0\hat{H}^*)^{-1}(\hat{R} - \hat{H}p_0) + p_0 \\ Var_\alpha(p|\hat{R}) &= \frac{s_*^2}{v_* - 2}[\Omega_0 - \Omega_0\hat{H}^*(\alpha I + \frac{1}{T}\hat{K}\hat{K}^* + \hat{H}\Omega_0\hat{H}^*)^{-1}\hat{H}\Omega_0].\end{aligned}$$

Analysis of posterior consistency of the regularized posterior distribution for p is specular to analysis performed in Section 4 and Corollary 1 holds with $\Omega_{\alpha,R}$ replaced by $Var_\alpha(p|\hat{R},\sigma^2)$.

Concerning the posterior distribution of σ^2, its posterior mean $\mathbb{E}(\sigma^2|\{<\hat{R},\hat{\varphi}_j>\}_{j=1}^J)$ is asymptotically equivalent to $\frac{1}{J}\sum_{j=1}^J\frac{1}{\hat{\lambda}_j}(<\hat{R}-\hat{H}p_0,\hat{\varphi}_j>)^2$ and its posterior variance is asymptotically equivalent to $\frac{1}{J}(\frac{s_*^2}{J})^2$. As $T\to\infty$, $\hat{K}\to K$ and the number J of eigenfunctions becomes large. Then, $Var(\sigma^2|\{<\hat{R},\hat{\varphi}_j>\}_{j=1}^J)$ converges to 0 and $\frac{1}{J}\sum_{j=1}^J\frac{1}{\hat{\lambda}_j}(<\hat{R}-\hat{H}p_0,\hat{\varphi}_j>)^2 \to \mathbb{E}(\frac{1}{\hat{\lambda}_j}(<\hat{R}-\hat{H}p_0,\hat{\varphi}_j>)^2) = \sigma^2$ at the parametric rate. Chebyshev's inequality implies consistency of $\nu^{\mathcal{F}}$.

Computation of eigenvalues and eigenfunction is not an easy task but computations can be considerably simplified by noting that for computing posterior distribution we need to know the quantities $<\hat{R},\hat{\varphi}_j>$, $j=1,\dots,J$ instead of the eigenfunctions $\{\hat{\varphi}_j\}$. Kernel estimation provide us with the following approximations:

$$\begin{aligned}\hat{R} &= \sum_i\sum_j M(y_i,Y_{t+1})M(y_i,y_{j+1})y_{j+1}\frac{L_h(y_i-y_j)L_h(Y_{t+1}-y_{i+1})}{\sum_l L_h(y_i-y_l)\sum_l L_h(Y_{t+1}-y_{l+1})} \\ \hat{H}p_0 &= \sum_i M(y_i,Y_{t+1})p_0(y_i)\frac{L_h(Y_{t+1}-y_{i+1})}{\sum_l L_h(Y_{t+1}-y_{l+1})} \\ &\quad -\sum_i\sum_j M(y_i,Y_{t+1})M(y_i,y_{j+1})p_0(y_{j+1})\frac{L_h(y_i-y_j)L_h(Y_{t+1}-y_{i+1})}{\sum_l L_h(y_i-y_l)\sum_l L_h(Y_{t+1}-y_{l+1})},\end{aligned}$$

where, for simplicity, we have eliminated the index $t+1$ in function M. Then,

$$
\begin{aligned}
< \hat{R} - \hat{H}p_0, \hat{\varphi}_j > &= \int (\hat{R} - \hat{H}p_0)(Y_{t+1})\hat{\varphi}_j(Y_{t+1})\pi(Y_{t+1})dY_{t+1} \\
&= \sum_i \sum_j [M(y_i, y_{j+1})(y_{j+1} + p_0(y_{j+1}))\frac{L_h(y_i - y_j)}{\sum_l L_h(y_i - y_l)} - p_0(y_i)] \\
&\quad \int M(y_i, Y_{t+1})\frac{L_h(Y_{t+1} - y_{i+1})}{\sum_l L_h(Y_{t+1} - y_{l+1})}\hat{\varphi}_j(Y_{t+1})\pi(Y_{t+1})dY_{t+1} \\
&= \sum_i \sum_j \phi_j(y_i, y_{i+1})[M(y_i, y_{j+1})(y_{j+1} + p_0(y_{j+1}))\frac{L_h(y_i - y_j)}{\sum_l L_h(y_i - y_l)} - p_0(y_i)]
\end{aligned}
$$

with $\phi_j(y_i, y_{i+1}) = \int M(y_i, Y_{t+1})\frac{L_h(Y_{t+1}-y_{i+1})}{\sum_l L_h(Y_{t+1}-y_{l+1})}\hat{\varphi}_j(Y_{t+1})\pi(Y_{t+1})dY_{t+1}$. Finally, by expliciting the stochastic discount function we get

$$
\begin{aligned}
\phi_j(y_i, y_{i+1}) &= \beta \frac{1}{U'(y_i)}\bar{\phi}_j(y_{i+1}), \\
\bar{\phi}_j(y_{i+1}) &= \int U'(Y_{t+1})\frac{L_h(Y_{t+1} - y_{i+1})}{\sum_l L_h(Y_{t+1} - y_{l+1})}\hat{\varphi}_j(Y_{t+1})\pi(Y_{t+1})dY_{t+1}.
\end{aligned}
$$

Henceforth, we only need to compute $(\lambda_j, \bar{\phi}_j)$, $j = 1, \ldots, J$ that is an easier task. $\bar{\phi}_j$ is a T dimensional vector and it is the jth eigenvector of the $T \times T$ matrix $\mathcal{A}$ with (k,t)th element

$$
\begin{aligned}
\mathcal{A}(k,t) &= \sum_i \frac{\beta}{U'(y_i)}\Big[T\int M(y_i, Y)g(Y, y_{k+1})L(x_i, x_t, Y, y_{i+1})\pi(Y)dY + \\
&\quad \sum_{i'} (\int \bar{b}(y_{i'}, Y, y_i)g(Y, y_{k+1})L(y_i, y_t, Y, y_{i'+1})\pi(Y)dY) + \\
&\quad \sum_l \sum_{l'} \int c(y_{l'}, y_{l+1}, Y)\bar{L}(y_{l'}, y_t, Y, y_{l'+1})g(Y, y_{k+1})\pi(Y)dY W(y_i, y_t, y_{i+1}, y_{l+1}) - \\
&\quad \sum_m \sum_{m'} \bar{b}(y_{m'}, y_{m+1}, y_i)\frac{L_h(y_i - y_t)}{\sum_m L_h(y_i - y_m)}\int M(y_{m'}, Y)g(Y, y_{k+1})\bar{L}(y_{m'}, y_t, Y, y_{m'+1})\pi(Y)dY - \\
&\quad T\sum_{k'} W(y_i, y_t, y_{i+1}, y_{k'+1})\int M(y_{k'+1}, Y)g(Y, y_{k+1})\frac{L_h(Y - y_{k'+1})}{\sum_l L_h(Y - y_{l+1})}\pi(Y)dY\Big],
\end{aligned}
$$

with $\bar{b}(y_{i'}, Y, y_i) = M(y_{i'}, Y)\omega(Y, y_i)$, $\omega(\cdot,\cdot)$ is the kernel of the prior covariance operator, $c(y_{l'}, y_{l+1}, Y) = M(y_{l'}, y_{l+1})M(y_{l'}, Y)$, $g(Y, y_l) = U'(Y)\frac{L_h(Y-y_l)}{\sum_t L_h(Y-y_{t+1})}$, $\bar{L}(y_i, y_t, Y, y_{i+1}) = \frac{L_h(y_i-y_t)L_h(Y-y_{i+1})}{\sum_t L_h(y_i-y_t)\sum_t L_h(Y-y_{t+1})}$ and

$$
W(y_i, y_t, y_{i+1}, y_{l+1}) = \int \bar{b}(y_i, Y, y_{l+1})\bar{L}(y_i, y_t, Y, y_{i+1})\pi(Y)dY.
$$

Proof for obtaining this matrix are provided in the Appendix.

7 Conclusions

In this paper we have exploited the dynamic rational expectation model of Lucas (1978) for recovering an estimation of the asset pricing functional of equilibrium. Due to the fact that it does not exists an explicit solution for such functional we have to solve an integral equation of second kind.

Our contribution is to propose a Bayesian nonparametric approach to solve the integral equation. A Bayesian approach allows to exploit all the prior information we have on the price functional parameter.

The Bayesian solution to this problem is the posterior distribution of p. However this distribution is not consistent in the sampling sense. Hence, we propose to substitute this distribution by a regularized version of it that has good frequentist asymptotic properties. Our estimator converges towards the true value p_* with a faster rate than the classical estimator, obtained by simply inverting operator $(I-K)$ in (7), does.

8 Appendix A: Proofs

8.1 Proof of Theorem 1

Let $T(\hat{F})$ denote the functional in the estimated transition distribution function $F(y_{t+1}|y_t)$ of the Markov process $\{Y_t\}$:

$$T(\hat{F}) = \int M_{t+1}(y_t, Y_{t+1})[M_{t+1}(y_t, y_{t+1})(b(y_{t+1}) + p(y_{t+1})) - p(y_t)]d\hat{F}(y_{t+1}|y_t)d\hat{F}(y_t|Y_{t+1}).$$

Note that $T(\hat{F})$ coincides with the error term U since $r + Kp = p$ and that $T(F) = 0$. We make a first order Taylor expansion of $T(\hat{F})$ around the true value F: $T(\hat{F}) - T(F) = d_1T(F; \hat{F} - F) + R_{1T}$, where d_1 denotes the Gâteaux differential of T at F in the direction of $\hat{F}$ and R_{1T} is the rest. Let λ be a scalar and $\xi(y_t, y_{t+1}, Y_{t+1}) = M_{t+1}(y_t, Y_{t+1})[M_{t+1}(y_t, y_{t+1})(b(y_{t+1}) + p(y_{t+1})) - p(y_t)]$, then

$$\begin{aligned} d_1T(F; \hat{F} - F) &= \frac{d}{d\lambda} T(F + \lambda(\hat{F} - F))\Big|_{\lambda=0} \\ &= \int \xi(y_t, y_{t+1}, Y_{t+1})\hat{F}(dY_{t+1}|Y_t)F(dY_t|y_{t+1}) + \\ &\quad \int \xi(y_t, y_{t+1}, Y_{t+1})F(dY_{t+1}|Y_t)\hat{F}(dY_t|y_{t+1}) \\ &\quad -2\int \xi(y_t, y_{t+1}, Y_{t+1})F(dY_{t+1}|Y_t)F(dY_t|y_{t+1}). \end{aligned}$$

Since the last two terms are null and $T(F) = 0$, we obtain that $T(\hat{F})$, and then U, is asymptotically equivalent to $\int M_{t+1}(y_t, Y_{t+1}) \int [M_{t+1}(y_t, y_{t+1})(b(y_{t+1}) + p(y_{t+1})) - p(y_t)]\hat{f}(y_{t+1}|y_t)dy_{t+1}f(y_t|Y_{t+1})dy_t$. The central integral can be approximated through a first order Taylor expansion around the true value of F as: $\frac{1}{\pi(y_t)}[\int M_{t+1}(y_t, y_{t+1})(b(y_{t+1}) + p(y_{t+1}))\hat{f}(y_{t+1}, y_t)dy_{t+1} - p(y_t)\int \hat{\pi}(y_t)dy_t]$. Then, by substituting $\hat{f}$ and $\hat{\pi}$ with the expression for their kernel estimations we obtain:

$$U \approx \int M_{t+1}(y_t, Y_{t+1}) \frac{1}{Th_t} \sum_{j=1}^{T} [M_{t+1}(y_t, y_{j+1})(b(y_{j+1}) + p(y_{j+1})) - p(y_t)] L_h(y_t - y_j) \frac{f(y_t|Y_{t+1})}{\pi(y_t)} dy_t$$

$$\approx \frac{1}{T}\sum_{j=1}^{T} M_{t+1}(y_j,Y_{t+1})[M_{t+1}(y_j,y_{j+1})(b(y_{j+1})+p(y_{j+1}))-p(y_j)]\frac{f(y_j|Y_{t+1})}{\pi(y_j)} +$$
$$\frac{1}{T}\sum_{j=1}^{T}\int\sum_{i=1}^{\rho}\Big[\frac{\partial^i}{\partial Y_t^i}M_{t+1}(Y_t,Y_{t+1})M_{t+1}(Y_t,y_{j+1})\frac{f(Y_t|Y_{t+1})}{\pi(Y_t)}\Big|_{Y_t=y_j}(b(y_{j+1})+p(y_{j+1}))$$
$$-\frac{\partial^i}{\partial Y_t^i}M_{t+1}(Y_t,Y_{t+1})p(Y_t)\frac{f(Y_y|Y_{t+1})}{\pi(Y_t)}\Big|_{Y_t=y_j}\Big]h^i u.$$

The second equality is obtained by making the change of variable $\frac{y_t-y_j}{h_T}=u$ and a Taylor expansion at order ρ around y_t, where ρ is the minimum among the order of the kernel, the order of differentiability of the utility function, of the transition and of the stationary density. By simplifying, we get

$$\sqrt{T}U(Y_{t+1}) \approx \frac{\sqrt{T}}{T}\sum_{j=1}^{T} M_{t+1}(y_j,Y_{t+1})[M_{t+1}(y_j,y_{j+1})(b(y_{j+1})+p(y_{j+1}))-p(y_j)]\frac{f(y_j|Y_{t+1})}{\pi(y_j)} + h_T^{\rho}\vartheta,$$

that is the expression in the theorem with ϑ deducible from the previous expression. Note that all the terms corresponding to h^i, with $i<\rho$ are null since they integrate to 0. When $T\to\infty$, $h\to 0$ then we can neglect the second term in $\sqrt{T}U$ and rewrite the scaled error term as $\sqrt{T}U = T^{-\frac{1}{2}}\sum_{j=1}^{T}\theta_j(Y_{t+1})$, with

$$\theta_j(Y_{t+1}) = M_{t+1}(y_j,Y_{t+1})[M_{t+1}(y_j,y_{j+1})(b(y_{j+1})+p(y_{j+1}))-p(y_j)]\frac{f(y_j|Y_{t+1})}{\pi(y_j)}.$$

where $\theta_j(Y_{t+1})$ is a sequence of stationary Hilbert random element such that $||\theta_j(Y_{t+1})||$ is bounded with probability 1 since

$$\mathbb{E}||\theta_j(Y_{t+1})|| = \sigma^2\int M_{t+1}^2(y_j,Y_{t+1})\frac{f^2(Y_{t+1}|y_j)}{\pi^2(Y_{t+1})}\pi(Y_{t+1})\pi(y_j)dY_{t+1}dy_j < \infty.$$

This guarantees that $\sqrt{T}U$ weakly converges toward a Gaussian process, see Theorem 2.46 in [4]. Its expectation is equal to 0 since

$$\begin{aligned}\sqrt{T}\mathbb{E}(U(Y_{t+1})) &= \int M_{t+1}(y_j,Y_{t+1})[M_{t+1}(y_j,y_{j+1})(b(y_{j+1})+p(y_{j+1}))-p(y_j)]\frac{f(y_j|Y_{t+1})}{\pi(y_j)}f(y_j,y_{j+1})dy_jdy_{j+1}\\ &= \int M_{t+1}(y_j,Y_{t+1})\mathbb{E}[M_{t+1}(y_j,y_{j+1})(b(y_{j+1})+p(y_{j+1}))-p(y_j)|y_j]\frac{f(y_j|Y_{t+1})}{d}y_j\\ &= 0\end{aligned}$$

and the kernel $\varpi(Y_{t+1},\tilde{Y}_{t+1})$ of its covariance operator is computed as

$$\begin{aligned}\varpi(Y_{t+1},\tilde{Y}_{t+1}) &= \frac{1}{T}Cov(\sum_{j=1}^{T}\theta_j(Y_{t+1}),\sum_{j=1}^{T}\theta_j(\tilde{Y}_{t+1}))\\ &= Cov(\theta_j(Y_{t+1}),\theta_j(\tilde{Y}_{t+1})) + \frac{2}{T}\sum_{l>j}Cov(\theta_j(Y_{t+1}),\theta_l(\tilde{Y}_{t+1})).\end{aligned}$$

By exploiting equality (6), the second term is null. Then,

$$\varpi(Y_{t+1},\tilde{Y}_{t+1}) = \int M_{t+1}(y_j,Y_{t+1})M_{t+1}(y_j,\tilde{Y}_{t+1})[M_{t+1}(y_j,y_{j+1})(b(y_{j+1})+p(y_{j+1}))-p(y_j)]^2$$

$$\frac{f(y_j|Y_{t+1})f(y_j|\tilde{Y}_{t+1})}{\pi^2(y_j)}f(y_j,y_{j+1})dy_jdy_{j+1}$$

$$= \int M_{t+1}(y_j,Y_{t+1})M_{t+1}(y_j,\tilde{Y}_{t+1})Var[M_{t+1}(y_j,y_{j+1})(b(y_{j+1})+p(y_{j+1}))-p(y_j)|y_j]$$

$$\frac{f(y_j|Y_{t+1})f(y_j|\tilde{Y}_{t+1})}{\pi(y_j)}f(y_j,y_{j+1})dy_j$$

$$= \sigma^2\int M_{t+1}(y_j,Y_{t+1})M_{t+1}(y_j,\tilde{Y}_{t+1})\frac{f(y_j|Y_{t+1})f(y_j|\tilde{Y}_{t+1})}{\pi(y_j)}f(y_j,y_{j+1})dy_j.$$

The factor scaled by σ^2 is the kernel of the operator K^*K. Then, the asymptotic covariance operator associated to $\sqrt{T}U$ is asymptotically equal to σ^2K^*K. Then, $\sqrt{T}U \Rightarrow \mathcal{GP}(0,\sigma^2K^*K)$.

8.2 Proof of Corollary 1

The bias associated to $\mu_\alpha^{\mathcal{F}}$ can be decomposed in two terms:

$$\hat{\mathbb{E}}_\alpha(p|\hat{R})-p_* = (\hat{\mathbb{E}}_\alpha(p|\hat{R})-\mathbb{E}_\alpha(p|\tilde{R}))+(\mathbb{E}_\alpha(p|\tilde{R})-p_*),$$

where $\mathbb{E}_\alpha(p|\tilde{R}) = \Omega_0H^*(\alpha_TI+\Sigma_T+K\Omega_0K^*)^{-1}(\tilde{R}-Hp_0)+p_0$ and $\tilde{R} = Hp_*+U$. The first term represent the estimation error of the operators and the second one stands for the error due to approximate the true value p_* of the asset price with the regularized posterior mean. We begin the analysis by the second term that we rewrite as:

$$\mathbb{E}_\alpha(p|\tilde{R})-p_* = -\overbrace{[I-\Omega_0H^*(\alpha_TI+\Sigma_T+H\Omega_0H^*)^{-1}H](p_*-p_0)}^{I} + \underbrace{\Omega_0H^*(\alpha_TI+\Sigma_T+H\Omega_0H^*)^{-1}U}_{II}.$$

The first term can still be decomposed into two terms, in order to isolate the effect of the covariance operator Σ_T:

$$I = \overbrace{[I-\Omega_0H^*(\alpha I+H\Omega_0H^*)^{-1}H](p_*-p_0)}^{IA} + \underbrace{[\Omega_0H^*(\alpha I+\Sigma_T+H\Omega_0H^*)^{-1}H-\Omega_0H^*(\alpha I+H\Omega_0H^*)^{-1}H](p_*-p_0)}_{IB}$$

and term IA looks very similar to the regularization bias of the solution of a functional equation. More properly, to obtain such a kind of object we use the assumption that $(p_*-p_0)\in\mathcal{H}(\Omega_0)$, *i.e.* there exists a δ_* belonging to the domain of $\Omega_0^{\frac{1}{2}}$ such that we can write $(p_*-p_0) = \Omega_0^{\frac{1}{2}}\delta$. Therefore,

$$\begin{aligned} IA &= [I-\Omega_0H^*(\alpha I+H\Omega_0H^*)^{-1}H]\Omega_0^{\frac{1}{2}}\delta_* \\ &= [\Omega_0^{\frac{1}{2}}-\Omega_0H^*(\alpha I+H\Omega_0H^*)^{-1}H\Omega_0^{\frac{1}{2}}]\delta_* \\ &= \Omega_0^{\frac{1}{2}}[I-\Omega_0^{\frac{1}{2}}H^*(\alpha I+H\Omega_0H^*)^{-1}H\Omega_0^{\frac{1}{2}}]\delta_*, \end{aligned}$$

where in the last equality we have used the fact that, since Ω_0 is positive definite and self-adjoint, it can be rewritten as $\Omega_0 = \Omega_0^{\frac{1}{2}}\Omega_0^{\frac{1}{2}}$. Let $B = H\Omega_0^{\frac{1}{2}}$ we take the norm in $\mathcal{X}$ of IA and after commutation of operators:

$$||IA||^2 \leq ||\Omega_0^{\frac{1}{2}}||^2||(I-(\alpha I+B^*B)^{-1}B^*B)\delta_*||^2.$$

The second norm in the right hand side of the previous expression is equal to $||\alpha(\alpha I+B^*B)^{-1}\delta_*||^2$ and it appears as the regularization bias associated to the regularized solution of the ill-posed inverse problem $B\delta_* = r$ computed using Tikhonov regularization scheme. It converges to zero when the regularization parameter α goes to zero and therefore also $||IA||^2$ converges to zero. This way to rewrite the above operator justifies the identification condition. Injectivity of $H\Omega_0^{\frac{1}{2}}$ ensures that the solution of $B\delta_* = r$ is identified and therefore, if $\Omega_0^{\frac{1}{2}}$ is injective, that (p_*-p_0) is identified and that the convergence of the regularized posterior mean is towards the right true value.
The speed of convergence to zero of $||(I-(\alpha I+B^*B)^{-1}B^*B)||^2$ depends on the regularity of δ_*, and consequently of (p_*-p_0). If the true solution δ_* lies in the β-regularity space Φ_β of the operator B, *i.e.* $\delta_* \in \mathcal{R}(\Omega_0^{\frac{1}{2}}H^*H\Omega_0^{\frac{1}{2}})^{\frac{\beta}{2}}$, the squared regularization bias is at most of order α_T^β and then $||IA||^2 = \mathcal{O}_p(\alpha_T^\beta)$. We refer to Carrasco *et al.* (2007) [4] and Kress (1999) [23] for a proof of it.
The larger β is, the smoother the function $\delta_* \in \Phi_\beta$ will be and the faster the regularization bias will converge to zero. However, since for Tikhonov regularization scheme, β cannot be grater than 2 we implicitly assume that $\delta_* \in \Phi_\beta$ for $\beta \leq 2$.
Now, let us consider term IB:

$$\begin{aligned}||IB||^2 &\leq ||\Omega_0H^*||^2(\alpha I+\Sigma_T+H\Omega_0H^*)^{-1}||^2||\Sigma_T||^2||(\alpha I+H\Omega_0H^*)^{-1}H(p_*-p_0)||^2\\ &\sim \mathcal{O}_p\Big(\frac{1}{\alpha^2}||\Sigma_T||^2||(\alpha I+H\Omega_0H^*)^{-1}H(p_*-p_0)||^2\Big).\end{aligned}$$

Since $\Sigma_T = \frac{\sigma^2}{T}K^*K$, its squared norm is $||\Sigma_T||^2 \sim \mathcal{O}_p(\frac{1}{T^2})$. Moreover, by using the regularity condition $\delta_* \in \mathcal{R}((\Omega_0^{\frac{1}{2}}H^*H\Omega_0^{\frac{1}{2}})^{\frac{\beta}{2}}) \equiv \mathcal{R}((B^*B)^{\frac{\beta}{2}})$

$$\begin{aligned}||(\alpha_T I+H\Omega_0H^*)^{-1}H(p_*-p_0)||^2 &\sim ||(\alpha I+B^*B)^{-1}B\delta_*||^2\\ &\sim ||(\alpha I+B^*B)^{-1}(B^*B)^{\frac{\beta+1}{2}}\rho||^2\\ &\sim \frac{1}{\alpha^2}||\alpha(\alpha I+B^*B)^{-1}(B^*B)^{\frac{\beta+1}{2}}\rho||^2\\ &\sim \mathcal{O}_p\Big(\frac{1}{\alpha^2}\alpha^{(\beta+1)\wedge 2}\Big),\end{aligned}$$

since $B=(B^*B)^{\frac{1}{2}}$. Thus $||IB||^2 \sim \mathcal{O}_p\Big(\frac{1}{\alpha^4T^2}\alpha^{(\beta+1)\wedge 2}\Big)$.
To find speed of convergence of term II we decompose it in the following equivalent way:

$$\begin{aligned}II &= \overbrace{\Omega_0^{\frac{1}{2}}B^*(\alpha I+BB^*)^{-1}U}^{IIA}\\ &\quad +\underbrace{\Omega_0H^*[(\alpha I+\Sigma_T+H\Omega_0H^*)^{-1}-(\alpha I+H\Omega_0H^*)^{-1}]U}_{IIB}+\\ ||IIA||^2 &\leq ||\Omega_0^{\frac{1}{2}}||^2||(\alpha I+B^*B)^{-1}B^*||^2||U||^2\\ ||IIB||^2 &\leq ||\Omega_0^{\frac{1}{2}}||^2||B^*(\alpha I+BB^*)^{-1}||^2||\Sigma_T||^2||(\alpha I+\Sigma_T+BB^*)^{-1}||^2||U||^2.\end{aligned}$$

By Kolmogorov theorem, $||U||^2$ is bounded in probability if $\mathbb{E}||U||^2 < \infty$ and $\mathbb{E}||U||^2 = tr\Sigma_n$. Then, $||IIA||^2 \sim \mathcal{O}_p(\frac{1}{\alpha}tr\Sigma_T)$ and $||IIB||^2 \sim \mathcal{O}_p(\frac{1}{\alpha^3}||\Sigma_T||^2tr\Sigma_T)$. Since $tr\Sigma_T \sim \mathcal{O}_p(\frac{1}{T})$ and

$||\Sigma_T||^2 \sim \mathcal{O}_p(\frac{1}{T^2})$ we conclude that $||II||^2 \sim \mathcal{O}_p(\frac{1}{\alpha T} + \frac{1}{\alpha^3 T^3}) \sim \mathcal{O}_p(\frac{1}{\alpha T})$ because the second rate is negligible with respect to the first one.

Let consider now the term $(\hat{\mathbb{E}}_\alpha(p|\hat{R}) - \mathbb{E}_\alpha(p|\tilde{R}))$ due to the estimation error. We make a decomposition similar to that done before:

$$\begin{aligned}
\hat{\mathbb{E}}_\alpha(p|\hat{R}) - \mathbb{E}_\alpha(p|\tilde{R}) &= \overbrace{\Omega_0[\hat{H}^*(\alpha I + \Sigma_T + \hat{H}\Omega_0\hat{H}^*)^{-1}\hat{H} - H^*(\alpha I + \Sigma_T + H\Omega_0 H^*)^{-1}H](p_* - p_0)}^{A} \\
&\quad + \underbrace{\Omega_0[\hat{H}^*(\alpha I + \Sigma_T + \hat{H}\Omega_0\hat{H}^*)^{-1} - H^*(\alpha I + \Sigma_T + H\Omega_0 H^*)^{-1}]U}_{B}, \\
A &= \overbrace{\Omega_0^{\frac{1}{2}}[\hat{B}^*(\alpha I + \hat{B}\hat{B}^*)^{-1}\hat{B} - B^*(\alpha I + BB^*)^{-1}B]\delta_*}^{A1} \\
&\quad + \underbrace{\Omega_0^{\frac{1}{2}}[\hat{B}^*(\alpha I + \Sigma_T + \hat{B}\hat{B}^*)^{-1}\hat{B} - [\hat{B}^*(\alpha I + \hat{B}\hat{B}^*)^{-1}\hat{B}]\delta_*}_{A2} \\
&\quad - \underbrace{\Omega_0^{\frac{1}{2}}[B^*(\alpha I + \Sigma_T + BB^*)^{-1}B - B^*(\alpha I + BB^*)^{-1}B]\delta_*}_{A3}, \\
B &= \Omega_0^{\frac{1}{2}}[\hat{B}^*(\alpha I + \hat{B}\hat{B}^*)^{-1} - B^*(\alpha I + BB^*)^{-1}]U \\
&\quad + \Omega_0^{\frac{1}{2}}[\hat{B}^*(\alpha I + \Sigma_T + \hat{B}\hat{B}^*)^{-1} - [\hat{B}^*(\alpha I + \hat{B}\hat{B}^*)^{-1}]U \\
&\quad - \Omega_0^{\frac{1}{2}}[B^*(\alpha I + \Sigma_T + BB^*)^{-1} - B^*(\alpha I + BB^*)^{-1}]U.
\end{aligned}$$

The norm $||A3||^2$ is equal to $||IB||^2$. Note that $||\hat{B}^*\hat{B} - B^*B||^2 \sim \mathcal{O}_p(\frac{1}{T} + h^{2\rho})$ and $||\hat{B}\hat{B}^* - BB^*||^2 \sim \mathcal{O}_p(\frac{1}{Th^n} + h^{2\rho})$, see Darolles *et al.* (2007) [6]. By using methods similar to those one used before and a Taylor expansion of $(\alpha I + \hat{B}^*\hat{B})$ around the true operator B, we get

$$\begin{aligned}
||A1||^2 &\sim \mathcal{O}_p((\frac{1}{\alpha^2} + \frac{1}{\alpha^4}\Big(\frac{1}{T} + h^{2\rho}\Big))\Big(\frac{1}{T} + h^{2\rho}\Big)\alpha^\beta) \\
||A2||^2 &\sim \mathcal{O}_p(\frac{1}{T^2\alpha^4}(1 + \frac{1}{\alpha^2(\Big(\frac{1}{Th^n} + h^{2\rho}\Big))})(\alpha^{(\beta+1)\wedge 2} + \Big(\frac{1}{Th^n} + h^{2\rho}\Big))(1 + \frac{1}{\alpha^2}\Big(\frac{1}{Th^n} + h^{2\rho}\Big))).
\end{aligned}$$

In a similar way we obtain

$$||B||^2 \sim \mathcal{O}_p(\frac{1}{\alpha^4 T^3}(1+\frac{1}{\alpha^2}\Big(\frac{1}{Th^n}+h^{2\rho}\Big))(1+\Big(\frac{1}{Th^n}+h^{2\rho}\Big))+\frac{1}{\alpha T}\Big(\frac{1}{Th^n}+h^{2\rho}\Big)(\frac{1}{\alpha}+\frac{1}{\alpha^3}\Big(\frac{1}{Th^n}+h^{2\rho}\Big))+\frac{1}{\alpha^3 T^3}).$$

Elimination of the negligible terms allows to conclude.
The procedure to obtain the rate of convergence of $\Omega_{\alpha,R}$ is equivalent so that in the proof we only show the fundamental decomposition that we have to perform:

$$\begin{aligned}
\Omega_{\alpha,R} &= -\Omega_0^{\frac{1}{2}}[\hat{B}^*(\alpha I + \hat{\Sigma}_T + \hat{B}\hat{B}^*)^{-1}\hat{B} - B^*(\alpha I + \Sigma_T + BB^*)^{-1}B]\Omega_0^{\frac{1}{2}} \\
&\quad - \Omega_0^{\frac{1}{2}}B^*(\alpha I + \Sigma_T + BB^*)^{-1}B]\Omega_0^{\frac{1}{2}}.
\end{aligned}$$

8.3 Proof of Theorem 3

Point *(i)* follows from Chebyshev's Inequality (16) and results in Corollary 1.
Point *(ii)* can be obtained by Chebishev's Inequality (16) and by keeping the non negligible rates in $||\hat{\mathbb{E}}_\alpha(p|\hat{R}) - p_*||^2$ and in $||\Omega_{\alpha,R}||$.

8.4 Proof of Theorem 4

Write the bias $(\mathbb{E}_s(p|\hat{R}) - p_*)$ as

$$\begin{aligned}
\mathbb{E}_s(p|\hat{R}) - p_* &= (\mathbb{E}_s(p|\hat{R}) - \mathbb{E}_s(p|\tilde{R})) + (\mathbb{E}_s(p|\tilde{R}) - p_*),\\
\mathbb{E}_s(p|\hat{R}) - \mathbb{E}_s(p|\tilde{R}) &= [\Omega_0\hat{H}^*(\alpha L^{2s} + \hat{\Sigma}_T + \hat{H}\Omega_0\hat{H}^*)^{-1}\hat{H} - \Omega_0 H^*(\alpha L^{2s} + \Sigma + H\Omega_0 H^*)^{-1}H](p_* - p_0)\\
&\quad + [\Omega_0\hat{H}^*(\alpha L^{2s} + \hat{\Sigma}_T + \hat{H}\Omega_0\hat{H}^*)^{-1} - \Omega_0 H^*(\alpha L^{2s} + \Sigma + H\Omega_0 H^*)^{-1}]U,\\
\mathbb{E}_s(p|\tilde{R}) - p_* &= -[I - \Omega_0 H^*(\alpha L^{2s} + \Sigma + H\Omega_0 H^*)^{-1}H](p_* - p_0)\\
&\quad + \Omega_0 H^*(\alpha L^{2s} + \Sigma + H\Omega_0 H^*)^{-1}U.
\end{aligned}$$

We omit computation of the rate of convergence of $(\mathbb{E}_s(p|\tilde{R}) - p_*)$ since it is given in the proof of Theorem 5 in [13]. The obtained rate is:

$$||\mathbb{E}_s(p|\tilde{R}) - p_*||^2 \sim \mathcal{O}_p(\alpha^{\frac{\beta+1}{a+s}} + \alpha^{\frac{1-a}{a+s}} tr\Sigma_T + \frac{1}{\alpha^4}||\Sigma_T||^2\alpha^{\frac{a+\beta+2s}{a+s}} + \frac{1}{\alpha^2}||\Sigma||^2\alpha^{\frac{1-a}{a+s}} tr\Sigma_T).$$

Consider the estimation error $(\mathbb{E}_s(p|\hat{R}) - \mathbb{E}_s(p|\tilde{R}))$, denote $H\Omega_0^{\frac{1}{2}} = T$, the first term in it can be rewritten as:

$$\begin{aligned}
&\overbrace{\Omega_0^{\frac{1}{2}}\Big([\hat{T}^*(\alpha\Omega_0^{-s} + \hat{T}\hat{T}^*)^{-1}\hat{T} - T^*(\alpha\Omega_0^{-s} + TT^*)^{-1}T]\delta_*}^{A1}\\
+ &\underbrace{[\hat{T}^*(\alpha\Omega_0^{-s} + \hat{\Sigma}_T + \hat{T}\hat{T}^*)^{-1}\hat{T} - \hat{T}^*(\alpha\Omega_0^{-s} + \hat{T}\hat{T}^*)^{-1}\hat{T}]\delta_*}_{A2}\\
- &\underbrace{[T^*(\alpha\Omega_0^{-s} + \Sigma_T + TT^*)^{-1}T - T^*(\alpha\Omega_0^{-s} + TT^*)^{-1}T]\delta_*\Big)}_{A3}.
\end{aligned}$$

Let $B = T\Omega_0^{\frac{s}{2}} = H\Omega_0^{\frac{s+1}{2}}$ By commuting operators and factorizing $\Omega_0^{\frac{s}{2}}$ we get

$$\begin{aligned}
||A1|| &= ||\Omega_0^{\frac{s+1}{2}}[(\alpha I + \hat{B}^*\hat{B})^{-1}\hat{B}^*\hat{B} - (\alpha I + B^*B)^{-1}B^*B]\Omega_0^{\frac{\beta-s}{2}}\rho_*||\\
&= ||\Omega_0^{\frac{s+1}{2}}\Big(-[I - (\alpha I + \hat{B}^*\hat{B})^{-1}\hat{B}^*\hat{B}] + [I - (\alpha I + B^*B)^{-1}B^*B]\Big)\Omega_0^{\frac{\beta-s}{2}}\rho_*||\\
&= ||\Omega_0^{\frac{s+1}{2}}\Big(-\alpha(\alpha I + \hat{B}^*\hat{B})^{-1} + \alpha(\alpha I + B^*B)^{-1}\Big)\Omega_0^{\frac{\beta-s}{2}}\rho_*||\\
&= ||\Omega_0^{\frac{s+1}{2}}\alpha(\alpha I + \hat{B}^*\hat{B})^{-1}(\hat{B}^*\hat{B} - B^*B)(\alpha I + B^*B)^{-1}\Omega_0^{\frac{\beta-s}{2}}\rho_*||\\
&\leq ||(\alpha I + \hat{B}^*\hat{B})^{-1}||_{-(s+1)}||\hat{B}^*\hat{B} - B^*B)||||(\alpha I + B^*B)^{-1}\Omega_0^{\frac{\beta-s}{2}}\rho_*||.
\end{aligned}$$

The last norm is an $\mathcal{O}_p(\alpha^{\frac{\beta-s}{2(a+s)}})$; moreover $(\alpha I + \hat{B}^*\hat{B})^{-1} = (\alpha I + B^*B)^{-1} - (\alpha I + B^*B)^{-1}(\hat{B}^*\hat{B} - B^*B)(\alpha I + \hat{B}^*\hat{B})^{-1}$. Then, by using the Corollary 8.22 in [9]

$$\begin{aligned}
||(\alpha I + \hat{B}^*\hat{B})^{-1}||_{-(s+1)} &\leq ||(B^*B)^{\frac{s+1}{2(a+s)}}(\alpha I + \hat{B}^*\hat{B})^{-1}|| + ||(B^*B)^{\frac{s+1}{2(a+s)}}(\alpha I + B^*B)^{-1}(\hat{B}^*\hat{B} - B^*B)(\alpha I + \hat{B}^*\hat{B})^{-1}||\\
&\sim \mathcal{O}_p(\alpha^{\frac{1-2a-s}{2(a+s)}}).
\end{aligned}$$

since the second norm is negligible once multiplied by the remaining terms of $||A1||$. It follows that $||A1||^2 \sim \mathcal{O}_p(\alpha^{\frac{\beta+1}{a+s}}\frac{1}{\alpha^2}||\hat{B}^*\hat{B} - B^*B||^2)$. Following the same logic, term $A2$ is rewritten

$$\Omega_0^{\frac{1}{2}}\hat{B}^*(\alpha I + \Omega_0^{\frac{s}{2}}(\hat{\Sigma}_T + \hat{T}\hat{T}^*)\Omega_0^{\frac{s}{2}})^{-1}\Sigma_T(\alpha I + \hat{B}\hat{B}^*)^{-1}\hat{B}]\delta_*$$

that has norm of order $\mathcal{O}_p(\frac{1}{\alpha^3}||\Sigma_T||^2)$. Lastly,

$$\begin{aligned}
||A3|| &\leq ||\Omega_0^{\frac{1}{2}} B^*(\alpha I + \Omega_0^{\frac{s}{2}}(\Sigma_T + TT^*)\Omega_0^{\frac{s}{2}})^{-1}\Omega_0^{\frac{s}{2}}|| ||\Sigma_T|| ||(\alpha\Omega_0^{-s} + TT^*)^{-1}T\delta_*||,\\
||(\alpha\Omega_0^{-s} + TT^*)^{-1}T\delta_*|| &= ||T(\alpha\Omega_0^{-s} + T^*T)^{-1}\Omega_0^{\frac{\beta}{2}}\rho_*||\\
&= ||T\Omega_0^{\frac{s}{2}}(\alpha I + \Omega_0^{\frac{s}{2}}T^*T\Omega_0^{\frac{s}{2}})^{-1}\Omega_0^{\frac{\beta+s}{2}}\rho_*||\\
&= ||(B^*B)^{\frac{1}{2}}(\alpha I + B^*B)^{-1}\Omega_0^{\frac{\beta+s}{2}}\rho_*||\\
&= ||(B^*B)^{\frac{1}{2}}(\alpha I + B^*B)^{-1}(B^*B)^{\frac{\beta+s}{2(a+s)}}v||\\
&\sim \mathcal{O}_p(\alpha^{\frac{\beta-a}{2(a+s)}}),
\end{aligned}$$

for some v such that $\Omega_0^{\frac{\beta+s}{2}}\rho_* = (B^*B)^{\frac{\beta+s}{2(a+s)}}v$. Such v exists since, under Assumption 8, $\mathcal{R}(\Omega_0^{a+s}) = \mathcal{R}(B^*B)$. Then, $||A3||^2 \sim \mathcal{O}_p(\frac{1}{\alpha^4}||\Sigma_T||^2\alpha^{\frac{a+\beta+2s}{a+s}})$.
The second term of $(\mathbb{E}_s(p|\hat{R}) - \mathbb{E}_s(p|\tilde{R}))$ is rewritten

$$\begin{aligned}
&\overbrace{\Omega_0^{\frac{1}{2}}\Big([\hat{T}^*(\alpha\Omega_0^{-s} + \hat{T}\hat{T}^*)^{-1} - T^*(\alpha\Omega_0^{-s} + TT^*)^{-1}]U}^{A4}\\
+ &\underbrace{[\hat{T}^*(\alpha\Omega_0^{-s} + \hat{\Sigma}_T + \hat{T}\hat{T}^*)^{-1} - \hat{T}^*(\alpha\Omega_0^{-s} + \hat{T}\hat{T}^*)^{-1}]U}_{A5}\\
- &\underbrace{[T^*(\alpha\Omega_0^{-s} + \Sigma_T + TT^*)^{-1} - T^*(\alpha\Omega_0^{-s} + TT^*)^{-1}]U\Big)}_{A6}.
\end{aligned}$$

Then,

$$\begin{aligned}
||A4||^2 &= ||\Omega_0^{\frac{1}{2}}(\alpha\Omega_0^{-s} + \hat{T}^*\hat{T})^{-1}\hat{T}^* - (\alpha\Omega_0^{-s} + T^*T)^{-1}T^*]U||^2\\
&\leq ||\Omega_0^{\frac{s+1}{2}}(\alpha I + B^*B)^{-1}||^2\Big(||\hat{B}^*\hat{B} - B^*B||^2||(\alpha I + \hat{B}^*\hat{B})^{-1}\hat{B}^*||^2 + ||\hat{B}^* - B^*||^2\Big)||U||^2\\
&\sim \mathcal{O}_p(\alpha^{\frac{1-2a-s}{2(a+s)}}||\hat{B}^*\hat{B} - B^*B||^2\frac{1}{\alpha}tr\Sigma_T + \alpha^{\frac{1-2a-s}{2(a+s)}}||\hat{B}^* - B^*||^2 tr\Sigma_T)\\
||A5||^2 &\leq ||\Omega_0^{\frac{1}{2}}||^2||(\alpha I + \hat{B}^*\hat{B})^{-1}\hat{B}^*\Omega_0^{\frac{s}{2}}||^2||\hat{\Sigma}_T||^2||\Omega_0^{\frac{s}{2}}(\alpha I + \Omega_0^{\frac{s}{2}}(\hat{\Sigma}_T + \hat{T}\hat{T}^*)\Omega_0^{\frac{s}{2}})^{-1}||^2||U||^2\\
&\sim (\frac{1}{\alpha^3}||\hat{\Sigma}_T||^2 tr\Sigma_T)\\
||A6||^2 &\leq ||\Omega_0^{\frac{1}{2}}T^*(\alpha\Omega_0^{-s} + TT^*)||^2||\Sigma_T||^2||(\alpha\Omega_0^{-s} + \Sigma_T + TT^*)|| ||U||^2\\
&\sim \mathcal{O}_p(\frac{1}{\alpha^2}||\Sigma_T||^2 tr\Sigma_T\alpha^{\frac{1-a}{a+s}}).
\end{aligned}$$

Elimination of negligible terms allows to get the result.
The rate of convergence of $||\Omega_{s,R}||^2$ is based on specular methods and on the decomposition

$$\begin{aligned}
\Omega_{s,R} &= -\Omega_0[\hat{H}^*(\alpha L^{2s} + \hat{\Sigma}_T + \hat{H}\Omega_0\hat{H}^*)^{-1}\hat{H} - H^*(\alpha I + \Sigma_T + H\Omega_0 H^*)^{-1}H]\Omega_0\\
&\quad -\Omega_0 H^*(\alpha I + \Sigma_T + H\Omega_0 H^*)^{-1}H]\Omega_0.
\end{aligned}$$

8.5 Computation of the Eigensystem for Section 6

In this appendix we prove that the eigensystem $\{\lambda_j, \bar{\phi}_j\}$, necessary for obtaining the posterior distribution in Section 6, can be computed as the eigensystem associated to matrix $\mathcal{A}$.

We start by explicitating the estimated elements of $(\frac{1}{T}\hat{K}\hat{K}^* + \hat{H}\Omega_0\hat{H}^*)$. Note that $\hat{K}\hat{\varphi}_j \approx \int M(y_i,Y)\hat{\varphi}_j\frac{\hat{f}(y_i,Y)}{\pi(\hat{y}_i),\pi(\hat{Y})}\pi(Y)dY$. By remembering the definition of ϕ_j, we have:

$$\begin{aligned}
\hat{K}\hat{\varphi}_j &= T\sum_t \phi_j(y_i,y_{t+1})\frac{L_h(y_i-y_t)}{\sum_t L_h(y_i-y_t)}\\
\hat{K}^*\hat{K}\hat{\varphi}_j &= T\sum_t\sum_i M(y_i,Y)\phi_j(y_i,y_{t+1})\bar{L}(y_i,y_t,Y,y_{i+1})\\
\hat{H}\Omega_0\hat{H}^* &= \hat{K}^*\Omega_0\hat{K} + \hat{K}^*\hat{K}\Omega_0\hat{K}^*\hat{K} - \hat{K}^*\hat{K}\Omega_0\hat{K} - \hat{K}^*\Omega_0\hat{K}^*\hat{K}\\
\hat{K}^*\Omega_0\hat{K}\hat{\varphi}_j &= T\sum_t\sum_i\sum_{i'} M(y_{i'},Y)\omega(y_i,Y)\phi_j(y_i,y_{t+1})\bar{L}(y_i,y_t,Y,y_{i'+1})\pi(y)dy\\
\hat{K}^*\hat{K}\Omega_0\hat{K}^*\hat{K} &= \sum_t\sum_i\sum_l\sum_{l'} M(y_{l'},y_{l+1})M(y_{l'},Y)\bar{L}(y_{l'},y_t,Y,y_{l'+1})\phi_j(y_i,y_{t+1})\\
&\quad \int M(y_i,y)\omega(y,y_{l+1})\bar{L}(y_i,y_t,y,y_{i+1})\pi(y)dy\\
\hat{K}^*\hat{K}\Omega_0\hat{K} &= \sum_t\sum_i\sum_m\sum_{m'} M(y_{m'},y_{m+1})\omega(y_{m+1},y_i)\frac{L_h(y_i-y_t)}{\sum_m L_h(y_i-y_m)}M(y_{m'},Y)\\
&\quad \bar{L}(y_{m'},y_t,Y,y_{m'+1})\phi_j(y_i,y_{t+1})\\
\hat{K}^*\Omega_0\hat{K}^*\hat{K} &= \sum_t\sum_i\sum_{k'} M(y_{k'+1},Y)\frac{L_h(Y-y_{k'+1})}{\sum_l L_h(Y-y_{l+1})}\int M(y_i,y)\omega(y,y_{k'+1})\\
&\quad \bar{L}(y_i,y_t,y,y_{i+1})\pi(y)dy\phi_j(y_i,y_{t+1}).
\end{aligned}$$

Then, $(\frac{1}{T}\hat{K}\hat{K}^*+\hat{H}\Omega_0\hat{H}^*)\hat{\varphi}_j = \hat{\lambda}_j\hat{\varphi}_j$. By taking the integral $\int U'(Y)\frac{L_h(Y-y_{k+1})}{\sum_k L_h(Y-y_{k+1})}\pi(Y)dY$ on both sides of this equality, and developing $\phi_j(y_i,y_{y+1}) = \beta\frac{1}{U'(y_i)}\bar{\phi}_j(y_{t+1})$, we get $\mathcal{A}_k\varphi_j = \hat{\lambda}_j\bar{\phi}_j(y_{k+1})$, where $\mathcal{A}_k$ denotes the $(k+1)$-th row of $\mathcal{A}$, for $k=0,\ldots,T-1$.

8.6 Proof of Theorem 5

Consider the decomposition

$$\hat{\mathbb{E}}^g(p|\hat{R}) - p_* = \overbrace{[\hat{\mathbb{E}}^g(p|\hat{R}) - \tilde{\mathbb{E}}^g(p|\hat{R})]}^{I} + \overbrace{[\tilde{\mathbb{E}}^g(p|\hat{R}) - \mathbb{E}^g(p|\hat{R})]}^{II} + \overbrace{[\mathbb{E}^g(p|\hat{R}) - p_*]}^{III}.$$

Let $W = (\hat{K}^*\hat{K})^{-\frac{1}{2}}\hat{H}(K^*K)^{\frac{s}{2}}$ and $\hat{W} = (\hat{K}^*\hat{K})^{-\frac{1}{2}}\hat{H}(\hat{K}^*\hat{K})^{\frac{s}{2}}$. Then,

$$\begin{aligned}
||I||^2 &\leq \overbrace{||[(\hat{K}^*\hat{K})^{\frac{s}{2}}\hat{W}^*(\alpha I+\hat{W}\hat{W}^*)^{-1}(\hat{K}^*\hat{K})^{-\frac{1}{2}} - (K^*K)^{\frac{s}{2}}W^*(\alpha I+WW^*)^{-1}W(\hat{K}^*\hat{K})^{-\frac{1}{2}}]\hat{H}(p_*-p_0)||^2}^{IA}\\
&\quad + \underbrace{||[(\hat{K}^*\hat{K})^{\frac{s}{2}}\hat{W}^*(\alpha I+\hat{W}\hat{W}^*)^{-1}(\hat{K}^*\hat{K})^{-\frac{1}{2}} - (K^*K)^{\frac{s}{2}}W^*(\alpha I+WW^*)^{-1}W(\hat{K}^*\hat{K})^{-\frac{1}{2}}]U}_{IB}\\
||IA||^2 &\leq ||(\hat{K}^*\hat{K})^{\frac{s}{2}}(\alpha I+\hat{W}\hat{W}^*)^{-1}||^2\Big(||\alpha(\hat{W}^*-W^*)||^2 + ||\hat{W}*||2||\hat{W}-W||^2||W^*||^2\Big)\\
&\quad ||W(\alpha I+W^*W)^{-1}(K^*K)^{\frac{\beta-s}{2}}\rho_*||^2 + ||(\hat{K}^*\hat{K})^{\frac{s}{2}} - (K^*K)^{\frac{s}{2}}||^2||W^*(\alpha I+WW^*)^{-1}W(K^*K)^{\frac{\beta-s}{2}}\rho_*||^2\\
&\sim \mathcal{O}_p\Big(\frac{1}{\alpha^2}(\frac{1}{Th^n}+h^{2\rho})(\alpha^{\frac{2(\beta-s+1)}{2s+1}+2} + \alpha^{\frac{2\beta-1+4s}{2s+1}} + \alpha^{\frac{2\beta-1}{2s+1}+2})\Big)\\
&\sim \mathcal{O}_p\Big(\frac{1}{\alpha^2}(\frac{1}{Th^n}+h^{2\rho})\alpha^{\frac{2\beta-1+4s}{2s+1}}\Big),
\end{aligned}$$

since the first and third rates are negligible with respect to the second one. It has been used the assumption that $(p_*-p_0)\in\mathcal{R}(\Omega_0^{\frac{\beta}{2s}})$, *i.e.* $\exists\rho_*\in\mathcal{X}$ such that $(p_*-p_0) = (K^*K)^{\frac{\beta}{2}}\rho_*$.

$$
\begin{aligned}
||IB||^2 &\leq ||(\hat{K}^*\hat{K})^{\frac{s}{2}}[\hat{W}^*(\alpha I + \hat{W}\hat{W}^*)^{-1} - W^*(\alpha I + WW^*)^{-1}](K^*K)^{-\frac{1}{2}}U||^2 \\
&\quad + ||(\hat{K}^*\hat{K})^{\frac{s}{2}} - (K^*K)^{\frac{s}{2}}||^2 ||W^*(\alpha I + WW^*)^{-1}(\hat{K}\hat{K}^*)^{-\frac{1}{2}}U||^2 \\
&\sim \mathcal{O}_p\Big(\frac{1}{\alpha^2 T}(\frac{1}{Th^n} + h^{2\rho})\alpha^{\frac{2(s-\gamma-1)}{2s+1}}\Big).
\end{aligned}
$$

Hence,

$$
||[\hat{\mathbb{E}}^g(p|\hat{R}) - \tilde{\mathbb{E}}^g(p|\hat{R})]||^2 \sim \mathcal{O}_p\Big(\frac{1}{\alpha^2}(\frac{1}{Th^n} + h^{2\rho})(\alpha^{\frac{2\beta-1+4s}{2s+1}}) + \frac{1}{\alpha^2 T}(\frac{1}{Th^n} + h^{2\rho})\alpha^{\frac{2(s-\gamma-1)}{2s+1}}\Big).
$$

Let $B = (K^*K)^{-\frac{1}{2}}H(K^*K)^{\frac{s}{2}}$ and $\hat{B} = (\hat{K}^*\hat{K})^{-\frac{1}{2}}H(K^*K)^{\frac{s}{2}}$, the second error is rewritten as:

$$
\begin{aligned}
||II||^2 &\leq \overbrace{||(K^*K)^{\frac{s}{2}}[\hat{B}^*(\alpha I + \hat{B}\hat{B}^*)^{-1}\hat{B} - B^*(\alpha I + BB^*)^{-1}B](K^*K)^{\frac{\beta-s}{2}}\rho_*||^2}^{IIA} \\
&\quad + \underbrace{||(K^*K)^{\frac{s}{2}}[\hat{B}^*(\alpha I + \hat{B}\hat{B}^*)^{-1} - B^*(\alpha I + BB^*)^{-1}](K^*K)^{-\frac{1}{2}}U||^2}_{IIB} \\
||IIA||^2 &= ||(K^*K)^{\frac{s}{2}}(\alpha I + \hat{B}^*\hat{B})^{-1}(\hat{B}\hat{B}^* - BB^*)\alpha(\alpha I + B^*B)^{-1}(K^*K)^{\frac{\beta-s}{2}}\rho_*||^2 \\
&\sim \mathcal{O}_p(\alpha^{\frac{\beta-3s-2}{2s+1}}(\frac{1}{T} + h^{2\rho})). \\
||IIB||^2 &\leq ||(K^*K)^{\frac{s}{2}}[\hat{B}^*(\alpha I + \hat{B}\hat{B}^*)^{-1} - B^*(\alpha I + BB^*)^{-1}](\hat{K}^*\hat{K})^{-\frac{1}{2}}U||^2 \\
&\quad + ||(K^*K)^{\frac{s}{2}}\hat{B}^*(\alpha I + \hat{B}\hat{B}^*)^{-1}[(K^*K)^{-\frac{1}{2}} - (\hat{K}^*\hat{K})^{-\frac{1}{2}}]U||^2 \\
&\sim \mathcal{O}_p\Big((\frac{1}{T} + h^{2\rho})\alpha^{-\frac{2\gamma-1}{2s+1}}\frac{1}{T} + (\frac{1}{T} + h^{2\rho})\frac{1}{T}\alpha^{-\frac{2(s+2+\gamma)}{2s+1}}\Big) + \mathcal{O}_p\Big(\frac{1}{T}\alpha^{-\frac{4}{2s+1}}\Big).
\end{aligned}
$$

Then, $||\tilde{\mathbb{E}}^g(p|\hat{R}) - \mathbb{E}^g(p|\hat{R})||^2 \sim \mathcal{O}_p((\frac{1}{T} + h^{2\rho})(\alpha^{\frac{\beta-3s-2}{2s+1}} + \frac{1}{T}\alpha^{-\frac{2(s+2+\gamma)}{2s+1}} + \frac{1}{T}\alpha^{-\frac{4}{2s+1}}))$. Lastly,

$$
\begin{aligned}
||\mathbb{E}^g(p|\hat{R}) - p_*||^2 &\leq \overbrace{||[I - (K^*K)^s H^*(\alpha(K^*K) + H(K^*K)^s H^*)^{-1}H(p_* - p_0)]||^2}^{IIIA} \\
&\quad + \underbrace{||(K^*K)^s H^*(\alpha(K^*K) + H(K^*K)^s H^*)^{-1}U||^2}_{IIIB} \\
||IIIA||^2 &= \Big(\sup_j(\lambda_j - \frac{\lambda_j^{2s+\beta}(1-\lambda_j)^2}{\alpha + (1-\lambda_j)(\lambda_j^{2s} - \lambda_j^{2s+1})})\Big)^2 \\
&\leq \Big(\sup_j \frac{\alpha\lambda_j^{\beta}}{\alpha - 2\lambda_j^{2s+1}}\Big)^2 \\
&\sim \mathcal{O}_p\Big(\alpha^{\frac{2\beta}{2s+1}}\Big) \\
||IIIB||^2 &\leq tr(Var((K^*K)^s H^*(\alpha(K^*K) + H(K^*K)^s H^*)^{-1}U)) \\
&= \frac{\sigma^2}{T}\sum_j \frac{\lambda_j^{4s}(1-\lambda_j)^2}{(\alpha + (1-\lambda_j)^2\lambda_j^{2s})^2} \\
&= \frac{\sigma^2}{T}\sum_j \frac{\lambda_j^{2(2s-\gamma)}(1-\lambda_j)^2}{(\alpha + (1-\lambda_j)^2\lambda_j^{2s})^2}\lambda_j^{2\gamma} \\
&\leq \frac{\sigma^2}{T}\Big(\sup_j \frac{\lambda_j^{2(2s-\gamma)}}{\alpha - 2\lambda_j^{2s+1}}\Big)\sum_j \lambda_j^{2\gamma} \\
&\sim \mathcal{O}_p\Big(\frac{1}{T}\alpha^{-\frac{2(\gamma+1)}{2s+1}}\Big).
\end{aligned}
$$

The optimal α is obtaining by equating the two rates of $||\mathbb{E}^g(p|\hat{R})-p_*||^2$. Then, $\alpha^{\frac{2\beta}{2s+1}} = \frac{1}{T}\alpha^{-\frac{2(\gamma+1)}{2s+1}}$ if $\alpha \propto (\frac{1}{T})^{\frac{2s+1}{2(\beta+\gamma+1)}}$. The corresponding optimal speed of convergence is proportional to $(\frac{1}{T})^{\frac{\beta}{\beta+\gamma+1}}$. When α is set equal to the optimal one, the terms I and II go to zero if $\beta > \frac{s}{2} - \frac{1+2\rho}{4}$, $\beta > \frac{(\gamma+1)n-2\rho(\gamma-s)}{4\rho-n}$ when $\rho > \frac{n}{4}$ (or $\beta < \frac{(\gamma+1)n-2\rho(\gamma-s)}{4\rho-n}$ when $\rho < \frac{n}{4}$), $\beta > \frac{s-\gamma}{2}$ and $\beta > \frac{3s-2\gamma}{3}$. Condition $\beta > \frac{s}{2} - \frac{1+2\rho}{4}$ is weaker than $\beta > \frac{s-\gamma}{2}$.
Moreover, $||(\hat{K}^*\hat{K})^{\frac{s}{2}} - (K^*K)^{\frac{s}{2}}||^2 \sim \mathcal{O}_p(\frac{1}{Th^n+h^{2\rho}})$ if $s \geq 2$.

8.7 Proof of Theorem 6

We consider the posterior variance applied to an element $\phi \in \mathcal{X}$ and its decomposition

$$\widehat{Var}^g(p|\hat{R})\phi = \overbrace{[\widehat{Var}^g(p|\hat{R}) - \widetilde{Var}^g(p|\hat{R})]\phi}^{I} + \overbrace{[\widetilde{Var}^g(p|\hat{R}) - Var^g(p|\hat{R})]\phi}^{II} + \overbrace{Var^g(p|\hat{R})\phi}^{III}.$$

Let $W = (\hat{K}^*\hat{K})^{-\frac{1}{2}}\hat{H}(K^*K)^{\frac{s}{2}}$ and $\hat{W} = (\hat{K}^*\hat{K})^{-\frac{1}{2}}\hat{H}(\hat{K}^*\hat{K})^{\frac{s}{2}}$. Then, for any $v \in \mathcal{X}$ such that $(K^*K)^{\frac{s}{2}}\phi = (K^*K)^{\frac{\beta-s}{2}}v$

$$\begin{aligned}
||I||^2 &= ||(\hat{K}^*\hat{K})^s\phi - (\hat{K}^*\hat{K})^{\frac{s}{2}}\hat{W}^*(\alpha I + \hat{W}\hat{W}^*)^{-1}\hat{W}(\hat{K}^*\hat{K})^{\frac{s}{2}}\phi \\
&\quad -(K^*K)^s\phi + (K^*K)^{\frac{s}{2}}W^*(\alpha i + WW^*)^{-1}W(K^*K)^{\frac{\beta-s}{2}}v||^2 \\
&= ||(\hat{K}^*\hat{K})^{\frac{s}{2}}[I - \hat{W}^*(\alpha I + \hat{W}\hat{W}^*)^{-1}\hat{W}](\hat{K}^*\hat{K})^{\frac{s}{2}}\phi \\
&\quad -(K^*K)^{\frac{s}{2}}[I - W^*(\alpha I + WW^*)^{-1}W](K^*K)^{\frac{s}{2}}\phi||^2 \\
&\leq ||(\hat{K}^*\hat{K})^{\frac{s}{2}}\alpha(\alpha I + \hat{W}\hat{W}^*)^{-1}[(\hat{K}^*\hat{K})^{\frac{s}{2}} - (K^*K)^{\frac{s}{2}}]\phi||^2 \\
&\quad +||(\hat{K}^*\hat{K})^{\frac{s}{2}}[\alpha(\alpha I + \hat{W}^*\hat{W})^{-1} - \alpha(\alpha I + W^*W)^{-1}](K^*K)^{\frac{\beta-s}{2}}v||^2 \\
&\quad +||[(\hat{K}^*\hat{K})^{\frac{s}{2}} - (K^*K)^{\frac{s}{2}}]\alpha(\alpha I + W^*W)^{-1}(K^*K)^{\frac{\beta-s}{2}}v||^2 \\
&\sim \mathcal{O}_p\Big((\frac{1}{Th^*} + h^{2\rho})(\alpha^{\frac{2s}{2s+1}} + \alpha^{2(\beta-2s-1)}2s+1) + \alpha^{\frac{2(\beta-s)}{2s+1}}\Big) \\
&\sim \mathcal{O}_p\Big((\frac{1}{Th^*} + h^{2\rho})(\alpha^{\frac{2s}{2s+1}} + \alpha^{\frac{2(\beta-2s-1)}{2s+1}})\Big).
\end{aligned}$$

Let $B = (K^*K)^{-\frac{1}{2}}H(K^*K)^{\frac{s}{2}}$ and $\hat{B} = (\hat{K}^*\hat{K})^{-\frac{1}{2}}H(K^*K)^{\frac{s}{2}}$, term II is:

$$\begin{aligned}
||II||^2 &\leq ||(K^*K)^s\phi - (K^*K)^{\frac{s}{2}}\hat{B}(\alpha I + \hat{B}\hat{B}^*)^{-1}\hat{B}(K^*K)^{\frac{s}{2}}\phi||^2 \\
&\quad +||(K^*K)^s\phi - (K^*K)^{\frac{s}{2}}B(\alpha I + BB^*)^{-1}B(K^*K)^{\frac{s}{2}}\phi||^2 \\
&\sim \mathcal{O}_p\Big((\frac{1}{T} + h^{2\rho})\alpha^{\frac{2(\beta-2s-1)}{2s+1}}\Big).
\end{aligned}$$

Lastly, $||III||^2 = ||(K^*K)^{\frac{s}{2}}\alpha(\alpha I + B^*B)^{-1}(K^*K)^{\frac{\beta-s}{2}}v||^2$ that is of order $\mathcal{O}_p(\alpha^{\frac{2\beta}{2s+1}})$.

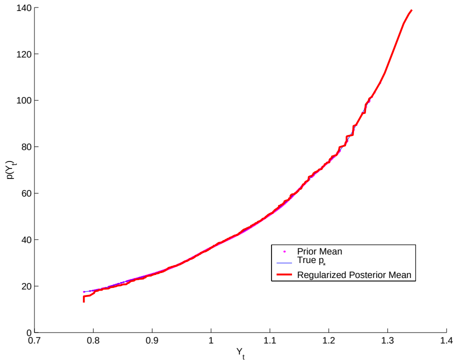
9 Appendix B: Numerical Implementation

We present in this subsection a numerical simulation able to show the good properties of our estimator. For simplicity, we take $n = 1$, so that only 1 consumption good is present in the economy. The law of motion for the relevant state variable Y_t is

$$\ln Y_t = a + b\ln Y_{t-1} + \epsilon,$$

where ϵ is a normal random variable with variance 0.01. The agent's per-period utility function is of CRR type: $U(Y_t) = \frac{Y_t^{(1-\gamma)}}{1-\gamma}$, with $\gamma = 0.30$. We chose the agent's subjective

discount factor $\beta = 0.97$.
The true value of the pricing functional is taken as the function satisfying equation (7). The transition density of the state variable is estimated through a kernel smoothing with a gaussian kernel and a bandwidth $h = 0.1$. The prior distribution is specified as a gaussian measure with mean alternatively equal to the true value p_* or to $p_0 = 2.35Y_t^2 + 22 - 04Y_t + 10.5$. The prior covariance operator is $\Omega_0 = \sigma_0 \int (\tilde{Y} \wedge Y_t)^2 \pi(\tilde{Y}) d\tilde{Y}$. We show the results of the simulation in Figure 1 .



(a) Prior Mean: $p_0 = p_*$

Figure 1: Asset Pricing functional estimation.

References

[1] Agliari, A. and C.C. Parisetti (1988), *A g-reference Informative Prior: a Note on Zellner's g-prior*, The Statistician, Vol.37, **3**, 271 - 275.

[2] Baker, C.R. (1973), *Joint Measures and Cross-Covariance Operators*, Transactions of the American Mathematical Society, **186**, 273-289.

[3] Blackwell, D. (1965), *Discounted Dynamic Programming*, The Annals of Mathematical Statistics, **Vol.36**, No. 1, 226-235.

[4] Carrasco, M., Florens, J.P., and E., Renault (2007), *Linear Inverse Problems in Structural Econometrics Estimation Based on Spectral Decomposition and Regularization*, Handbook of Econometrics, **Vol.6**, Part 2, 5633-5751.

[5] Chen, X., Favilukis, J. and S.C., Ludvigson (2008), *An Estimation of Economic Models with Recursive Preferences*, working paper.

[6] Darolles, S., Florens, J.P., and E., Renault (2006), *Nonparametric Instrumental Regression*, Econometrica (forthcoming).

[7] Diaconis, F., and D., Freedman (1986), *On the Consistency of Bayes Estimates*, Annals of Statistics, **14**, 1-26.

[8] Doob, J.L. (1949), *Application of the Theory of Martingales*, in Le Calcul des Probabilités et ses Applications, Colloques Internationaux du Centre National de la Recherche Scientifique, **13**, 23-27.

[9] Engl, H.W., Hanke, M. and A., Neubauer (2000), *Regularization of Inverse Problems*, Kluwer Academic, Dordrecht.

[10] Fair, R.C. and J.B., Taylor (1980), *Solution and Maximum Likelihood Estimation of Dynamic Nonlinear Rational Expectation Models*, Cowles Foundation Discussion Paper No. 564.

[11] Florens, J.P., Mouchart, M., and J.M., Rolin (1990), *Elements of Bayesian Statistics*, Dekker, New York.

[12] Florens, J.P., and A., Simoni (2007), *Nonparametric Estimation of Instrumental Regression: a Bayesian Approach Based on Regularized Posterior*, mimeo.

[13] Florens, J.P., and A., Simoni (2008), *Regularizzed Posteriors in Linear Ill-Posed Inverse Problems*, working paper.

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

[15] Freedman, D. (1965), *On the Asymptotic Behavior of Bayes Estimates in the Discrete Case II.*, Ann. Math. Statist., **36**, 454-456.

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

[17] Hansen, L.P., and T.J., Sargent (1980), *Formulating and Estimating Dynamuc Linear Rational Expectations Models*, Journal of Economic Dynamics and Control, **2**, 7 - 46.

[18] Hansen, L.P., and T.J., Sargent (1981), *Linear Rational Expectations Models for Dynamically Interrelated Variables*, in Rational Expectation and Economic Practice, ed. by R.E. Lucas, Jr. and T.J. Sargent. Minneapolis: University of Minnesota Press, 1981.

[19] Hansen, L.P., and K.J., Singleton (1982), *Generalized Instrumental Variables Estimation on Nonlinear Rational Expectations Models*, Econometrica, vol. 50, **5**, 1269 - 1286.

[20] Hansen, L.P., and K.J., Singleton (1983), *Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns*, the Journal of Political Economy, vol. 91, **2**, 249 - 265.

[21] Hayashi, F. (1980), *Estimation Macroeconomic Models Under Rational Expectations: A Survey*, Manuscript, Northwestern University, 1980.

[22] Hiroshi, S. and O., Yoshiaki (1975), *Separabilities of a Gaussian Measure*, Annales de l'I.H.P., section B, tome 11, **3**, 287 - 298.

[23] Kress, R. (1999), *Linear Integral Equation*, Springer.

[24] Lehtinen, M.S., Päivärinta, L. and E., Somersalo (1989), *Linear Inverse Problems for Generalised Random Variables*, Inverse Problems, **5**, 599-612.

[25] LeRoy, S.F. (1973), *Risk Aversion and the Martingale Property of Stock Prices*, International Economic Review, **Vol.14**, No.2, 436-446.

[26] Lucas, R.E. (1976), *Econometric policy evaluation: A critique*, in: K. Brunner and A.H. Meltzer, eds., The Phillips curve and labor markets, Carnegie-Rochester conferences on public policy 1 (North Holland, Amsterdam).

[27] Lucas, R.E. (1978), *Asset Prices in An Exchange Economy*, Econometrica, **Vol.46**, No.6, 1429-1445.

[28] Mandelbaum, A. (1984), *Linear Estimators and Measurable Linear Transformations on a Hilbert Space*, Z. Wahrcheinlichkeitstheorie, **3**, 385-98.

[29] Muth, J.F. (1961), *Rational Expectations and the Theory of Price Movements*, Econometrica, **29**, 41 - 49.

[30] Neveu, J. (1965), *Mathematical Fundations of the Calculus of Probability*, San Francisco: Holden-Day.

[31] Roussas, G.G. (1967), *Nonparametric Estimation in Markov Processes*, Ann. Inst. Stat. Math., **21**, 73 - 87.

[32] Rust, J.B., Traub, J.F. and H., Wozniakowski (2002), *Is There a Curse of Dimensionality for Contraction Fixed Points in the Worst Case?*, Econometrica, **70**, No.1, 285 - 329.

[33] Sargent, T.J. (1981), *Interpreting Economic Time Series*, Journal of Political Economy, **89**, 213 - 248.

[34] Tauchen, G. and R., Hussey (1991), *Quadrature Based Methods for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models*, Econometrica, **59**, 371 - 396.

[35] Van der Vaart, A.W., and J.H., Van Zanten (2000), *Rates of Contraction of Posterior Distributions Based on Gaussian Process Priors*, Working paper.

[36] Zellner, A. (1986), *On Assessing Prior Distributions and Bayesian Regression Analysis with g-prior Distribution*, in: Goel, P.K. and Zellner, A. (Eds) Bayesian Inference and Decision Techniques: essays in honour of Bruno de Finetti, pp. 233-243 (Amsterdam, North Holland).