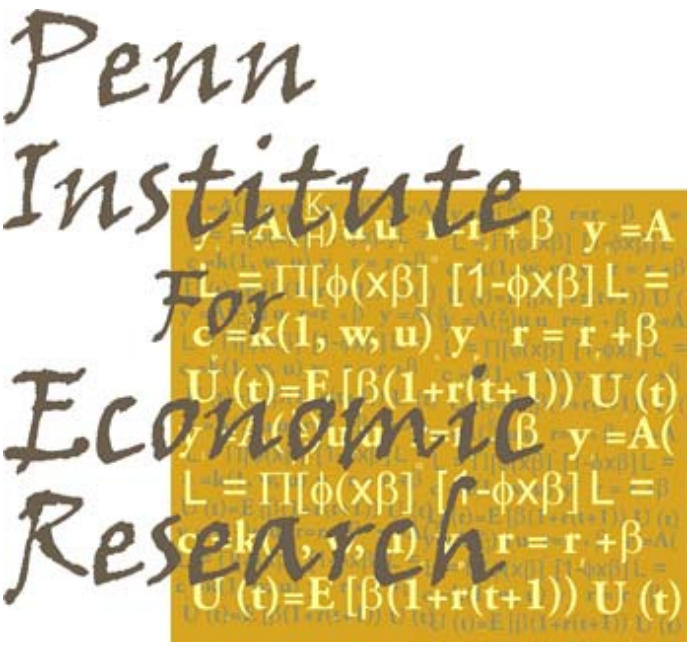


Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
http://www.econ.upenn.edu/pier

PIER Working Paper 01-037

"Comparing Dynamic Equilibrium Models to Data"

by

Jesús Fernández-Villaverde and Juan Francisco Rubio Ramírez

http://papers.ssrn.com/sol3/papers.cfm?abstract_id=288332

Comparing Dynamic Equilibrium Models to Data*†

Jesús Fernandez-Villaverde
University of Pennsylvania

Juan Francisco Rubio-Ramírez
Federal Reserve Bank of Atlanta

October 16, 2001

Abstract

This paper studies the properties of the Bayesian approach to estimation and comparison of dynamic equilibrium economies. Both tasks can be performed even if the models are nonnested, misspecified and nonlinear. First, we show that Bayesian methods have a classical interpretation: asymptotically the parameter point estimates converge to their pseudotrue values and the best model under the Kullback-Leibler distance will have the highest posterior probability. Second, we illustrate the strong small sample behavior of the approach using a well-known application: the U.S. cattle cycle. Bayesian estimates outperform Maximum Likelihood results and the proposed model is easily compared with a set of BVARs.

Keywords: Bayesian inference, asymptotics, cattle cycle. *JEL classification Numbers*: C11, C15, C51, C52.

†Corresponding Author: Jesús Fernández-Villaverde, Department of Economics, 160 McNeil Building, University of Pennsylvania, Philadelphia, PA 19104. E-mail: jesusfv@econ.upenn.edu. Thanks to A. Atkeson, J. Geweke, W. McCausland, E. McGrattan, L. Ohanian, T. Sargent, C. Sims and H. Uhlig an participants at several seminars for useful comments. Beyond the usual disclaimer, we must notice that any views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Atlanta or of the Federal Reserve System.

1. Introduction

Over the last two decades, Lucas' (1980) call to economists to concentrate in the building of fully articulated, artificial economies has become a reality. Dynamic equilibrium models are now the standard instrument to study a variety of issues in economics, from Business Cycles and Economic Growth to Public Finance and Demographics. This class of models present two main challenges for econometric practice: a) how to select appropriate values for the "deep" parameters of the model (*i.e.* those describing technology, preferences and so on), specially since by construction the model is false and b) how to compare models that can be very different.

Figure 1 illustrates these two questions. Assume we want to account for some observed data using two different models. Let A be the set of all possible distributions that generate the data. Let P be the true distribution. Each point in the set *Model 1* represents the distribution implied by the model for a particular choice of parameter values. A similar description applies for set *Model 2*. In this figure, the first of our questions (how to select parameters) is solved picking a point in each of the sets *Model 1* and *Model 2*. The second question (how to compare models) is solved learning which of the two sets is closer to P under a certain metric. Note that in Figure 1 we make two assumptions: first, that the models are false (none of the sets includes P) and second, that there is no choice of parameter values such that one model is an special case of the other (the intersection of these two models is the empty set). However, none of these assumptions is required.

Bayesian econometrics gives both a procedure to select parameters and a criterium for model comparison. Parameter choice is undertaken by the usual computation of posteriors while model comparison is performed through the use of posterior odds ratios. This approach is, of course, rather old. Parameters inference follows directly from the Bayes' Theorem while model comparison through posterior odds was introduced by Jeffreys (1961) (in the slightly different form of hypothesis testing) and recently revived by the work of Gelfand and Key (1994), Geweke (1998), Landon-Lane (1999) and Schorfheide (1999), among others.

Our work follows this long tradition. In particular this paper makes two main contributions. First we show that the Bayesian approach to model estimation and comparison has a classical interpretation: asymptotically the parameter point estimates converge to their pseudotrue values and the best model under the Kullback-Leibler distance will have the highest posterior probability- both results holding even for misspecified and/or nonnested models. Second, we illustrate the strong small sample behavior of Bayesian methods using a well-known application: the U.S. cattle cycle. Bayesian estimates outperform Maximum Likelihood results and the proposed model is easily compared with a set of Bayesian Vector Autoregressions. An additional contribution- how to evaluate the likelihood of nonlinear representations of dynamic equilibrium models with Monte Carlo filtering- is described in detail in a companion paper (Fernández-Villaverde and Rubio-Ramírez (2001)).

There are several reasons to justify our "Bayes choice." First, Bayesian inference builds on the basic insight that models are false and is ready to deal with this issue in a natural way. Since a dynamic equilibrium economy is an artificial construction, the model will *always* be false. Estimation moves from being a process of discovery of some "true" value of a parameter to being a selection device in the parameter space that maximizes our ability to use the model as a language in which to express the regular features of the data (Rissanen (1986)).

Second, the Bayesian approach is conceptually simple yet general. Issues such as stationarity do not require specific methods as needed in classical inference (Sims and Uhlig (1991)). Third, there is an asymptotic justification of the Bayes procedure. As mentioned before, we prove consistency of both the point estimates and the posterior odds ratio. Fourth, also as shown in the paper, the small sample performance of Bayesian estimates tends to outperform classical ones even when evaluated by frequentist criteria (for similar findings see Jacquier, Polson and Rossi (1994) or Geweke, Keane and Runkle (1997)). Fifth, the recent advent of powerful Markov chain Monte Carlo techniques has removed the need for analytically suitable expressions for likelihoods and priors. A quite general set of models and priors can be used and robustness analysis is a simple extension of the needed computations.

This paper relates with previous Frequentist and Bayesian work on model comparison. Frequentist literature has concentrated on the use of nonnested hypothesis testing (for a review see Gourieux and Monfort (1998)). In particular, Vuong (1989) and Kitamura (1998) have developed tests for nonnested and misspecified models. We see our contributions as very similar in spirit to these two papers.

In the Bayesian literature, DeJong, Ingram and Whiteman (2000) pioneered the Bayesian estimation of Real Business Cycles models using importance sampling; Landon-Lane (1999) and Otrok (2001) first applied the *Metropolis-Hastings algorithm* to the estimation problem; while in the area of dynamic equilibrium models comparison Landon-Lane (1999) has studied one dimensional-linear processes, and Schorfheide (1999) has compared the impulse-response functions of linearized models.

We advance with respect to these papers in several aspects. First, we pose the problem in very general terms, not limiting ourselves to linearized Real Business Cycles models. Second, our use of State Space representations allows us to deal with high dimensional vectors. Third, we can study nonlinear models. Fourth, we develop the asymptotic properties of the procedure. Fifth, we document the performance of Bayesian estimation in small samples and compare the marginal likelihood of the model against a set of alternatives.

The rest of the paper is organized as follows. Section 2 presents the asymptotic properties of the Bayesian approach to model estimation and comparison. Section 3 develops a dynamic equilibrium economy: the cattle cycle model. Section 4 estimates the model and section 5 compares it with a set of Bayesian Vector Autoregressions. Section 6 concludes.

2. Asymptotic Properties of the Bayesian Approach

This section develops the asymptotic properties of Bayesian inference when models are possibly misspecified and/or nonnested. We will prove that the posterior distribution of the parameters collapses to their pseudotrue values and that posterior odds ratio of any model over the best model under the Kullback-Leibler distance will approach zero as the sample size goes to infinity. The novelty of these two results is that we do not need to assume that the models are well-specified and/or nested as the existing literature requires. Dispensing with these requirements is key when dealing with a set of artificial economies that are false by construction and that may have very different structures. This situation is the most common for economists. Subsection A presents the notation, subsection B explains the Bayesian model comparison, and subsection C shows the above-mentioned two theorems. Subsection

D briefly discusses numerical implementation.

2.1. Notation

Assume that the observed data is a realization of the real-valued stochastic process $Y \equiv \{Y_t : \Omega \to \Re^m,\ m \in \mathcal{N},\ t = 1, 2, ...\}$ defined on a complete probability space $(\Omega, \Im, P_0)$ where $\Omega = \Re^{m\times\infty} \equiv \lim_{T\to\infty} \otimes_{t=0}^{T} \Re^m$ and $\Im \equiv \lim_{T\to\infty} \Im^T \equiv \lim_{T\to\infty} \otimes_{t=0}^{T} \mathcal{B}(\Re^m) \equiv \mathcal{B}(\Re^{m\times\infty})$ is just the Borel σ-algebra generated by the measurable finite-dimensional product cylinders. Define a $T-$segment as $Y^T \equiv (Y_1', ..., Y_T')'$ with $Y^0 = \{\emptyset\}$ and a realization of that segment as $y^T \equiv (y_1', ..., y_T')'$. Also define $P_0^T(B) \equiv P_0(B)\,|\Im^T \equiv P_0\left(Y^T \in B\right),\ \forall B \in \Im^T$ to be the restriction of P_0 to $\Im^T$. The structure of Ω is important only to the extent this allows for a sufficiently rich behavior in Y. For convenience, we have chosen $\Omega = \Re^{m\times\infty}$. In this case, Y_t is the projection operator that selects y_t, the tth coordinate of ω, so that $Y_t(\omega) = y_t$. With $\Im \equiv \mathcal{B}(\Re^{m\times\infty})$ the projection operator is measurable and Y is indeed a stochastic process.

It is often more convenient to work with densities rather than measures. As a consequence, we will assume there exists a measure ν^T on $\left(\Re^{m\times T}, \mathcal{B}(\Re^{m\times T})\right)$ for $T = 1, 2, ...$ such that $P_0^T \ll \nu^T$ (where "$\ll$" stands for "absolute continuity with respect to"). We will call the Radon-Nykodym derivatives of P_0^T with respect to ν^T, the probability density function $p_0^T(\cdot)$ for $\forall T$.

Let M be a finite subset of $\aleph$. Now we can define a model i as the collection $S(i) \equiv \{f(\theta, i), \pi(\theta|i), \Theta_i\}$ where $f(\theta, i) \equiv \{f^n(\cdot|\theta, i) : \Re^{m\times n} \times \Theta_i \to \Re, n = 1, 2, 3...\}$ is the set of densities $f^n(\cdot|\theta, i)$ on $(\Re^{m\times n}, \mathcal{B}(\Re^{m\times n}))$, $\pi(\theta|i)$ is a prior density on $(\Theta_i, \mathcal{B}(\Theta_i))$ and θ is a k_i-dimensional vector of unknown parameters such that $\theta \subseteq \Theta_i \subseteq \Re^{k_i}\ \forall i \in M$. Each family of parametrized probability densities comprises different candidates to account for the observations while the prior probability densities embodies the previous knowledge about the parameter values. Now we can define $S \equiv \{S(i), i \in M\}$ as the set of considered models. We can think about S in a very general way: it can contain nested as well as nonnested models. For example, it can include models derived directly from economic theory (as a Real Business Cycle) and/or pure statistical models (as an unrestricted Vector Autoregression).

The function $f^T\left(y^T|\theta, i\right)$ is usually called the pseudo-likelihood function of the data. Note that we are never assuming that there exists a value θ^* such that $f^T\left(y^T|\theta^*, i\right) = p_0^T\left(y^T\right)$. Statistically this means that the model may be misspecified. Far more importantly, from an economic perspective, this is a direct consequence of the fact that the model is false.

Often we will find it more convenient to write, for $\forall\theta \in \Theta_i$:

$$f^T(y^T, \theta|i) = f^T\left(y^T|\theta, i\right)\pi(\theta|i)$$

With this notation and using conditional probabilities, we can write the posterior of the parameters as $\pi\left(\theta|y^T, i\right) \propto f^T\left(y^T|\theta, i\right)\pi(\theta|i)$ and its marginal likelihood as:

$$f^T\left(y^T|i\right) = E_i\left(f^T\left(y^T|\theta, i\right)\right) = \int_{\Theta_i} f^T\left(y^T|\theta, i\right)\pi(\theta|i)\,d\theta = \int_{\Theta_i} f^T(y^T, \theta|i)d\theta \qquad (1)$$

This marginal likelihood is the probability that the model assigns having observed the data. This interpretation relates the marginal likelihood with the pseudo-likelihood evaluated at the pseudo-maximum likelihood point estimate (PMLE). In this case, the parameters are

integrated out through maximization using a measure that puts all the mass at the PMLE while, in the marginal likelihood, they are integrated out using the prior (herein we are assuming that we built our densities from a probability measure and, as a consequence, $\pi(\theta|i)$ is always proper).

Usually we will be in the situation where $f^T\left(y^T|\theta,i\right)$ can be factorized in the following way: $f^T\left(y^T|\theta,i\right)=\prod_{t=1}^{T} f_t\left(y_t|y^{t-1},\theta,i\right)$ where $f_t\left(\cdot|y^{t-1},\theta,i\right):\Re^{m\times t}\times\Theta_i\to\Re^+$ is $\mathcal{B}(\Re^{m\times t})$-measurable for each $\theta\in\Theta_i$. This factorization turns out to be important both theoretically (for instance to interpret the marginal likelihood as a measure of with-in sample forecasting performance) and computationally (to evaluate pseudo-likelihoods recursively).

We can define the expectation of the logpseudo-likelihood divided by T with respect to the true density:

$$\begin{aligned}\overline{L}(\theta,i) &= \int_{\Re^{m\times T}} T^{-1}\log f^T\left(Y^T|\theta,i\right)p_0^T\left(Y^T\right)d\nu^T = \\ &\int_{\Re^{m\times T}} T^{-1}\left[\sum_{t=1}^{T}\log f_t(Y_t|Y^{t-1},\theta,i)\right]p_0^T\left(Y^T\right)d\nu^T\end{aligned}$$

and two associated parameter values: the "pseudo-true" value $\widehat{\theta}_T^*(i)\equiv\arg\max_{\theta\in\Theta_i}\overline{L}(\theta,i)$ and the PMLE point $\widehat{\theta}_T\left(i,y^T\right)\equiv\arg\max_{\theta\in\Theta_i}\log f^T\left(y^T|\theta,i\right)$. From now on we will assume that these values are unique. This assumption is the fundamental identificability condition in our context of false models.

The pseudotrue value selects the member of the parametric family that is "closest" to P_{0T} in some appropriate sense. We define the Kullback-Leibler measure distance:

$$K\left(f^T(\cdot|\theta,i);p_0^T(\cdot)\right)=\int_{\Re^{m\times T}}\log\left(\frac{p_0^T\left(Y^T\right)}{f^T\left(Y^T|\theta,i\right)}\right)p_0^T\left(Y^T\right)d\nu^T$$

The intuition of this closeness concept is simple: it evaluates the average surprise with respect to the true measure that the researcher using $f^T(\cdot|\theta,i)$ will suffer if suddenly he learns that the true density is $p_0^T(\cdot)$. Clearly $\widehat{\theta}_T^*(i)$ minimizes $K\left(f^T(\cdot|\theta,i);p_0^T(\cdot)\right)$, and it is the point in Figure 1 that minimizes the distance between the set i and the true distribution. We omit a discussion of the decision-choice foundations of the Kullback-Leibler measure distance. Complete axiomatic foundations of this measure are presented in Shore and Johnson (1980) and Csiszar (1990) among others.

2.2. Model Comparison

First, define the measurable space $(M,P(M),\Pi)$ where $P(M)$ is the power set of M and Π is a measure that assigns a probability π_i to each element of M. This measure tries to reflect the previous knowledge of the researcher about the different models being considered.

Model comparison is a straightforward application of the Bayes' Theorem. The posterior probabilities of each model are given by:

$$\widehat{\pi_k}=\frac{f^T(y^T|i)\pi_k}{\sum_M f^T(y^T|i)\pi_i} \tag{2}$$

The division of any two posteriors produces the Posterior Odds Ratio:

$$POR_{i,j|Y_T} = \frac{\widehat{\pi}_i}{\widehat{\pi}_j} = \frac{f^T(y^T|i)\pi_i}{f^T(y^T|j)\pi_j}$$

that can be intuitively factored between the Bayes Factor:

$$B_{i,j|Y_T} = \frac{f^T(y^T|i)}{f^T(y^T|j)} \tag{3}$$

and the ratio of priors $\frac{\pi_i}{\pi_j}$ as:

$$POR_{i,j|Y_T} = B_{i,j|Y_T}\frac{\pi_i}{\pi_j} \tag{4}$$

The Bayes Factor is the ratio of probabilities from having observed the data given each model and represents by how much we should change our beliefs about the probability of each model given the empirical evidence. In other words, the Bayes Factor is a summary of the evidence provided by the data in favor of one model as opposed to the other, and it is our chosen approach to model comparison.

It is important to note that model comparison is a somewhat related but different task than the decision-theory problem of selecting one model among a set of alternatives since the latter requires the specification of a loss function.

Also, in the same way the marginal likelihood is related with the likelihood value at the PMLE, the Bayes Factor is closely related to the Likelihood Ratio (LR), where maximization substitutes integration. The Bayes Factor enjoys three clear advantages. First, LR tests may simultaneously reject or accept different nulls because of the asymmetric treatment of the two hypothesis. In comparison, the Bayes Factor states clearly which of the two models fits the data better. Second, no arbitrary choice of a significance level is needed. Third, when both models are false (the normal case in economics), the LR tests do not imply an asymptotic distribution of the ratio (for an exception see Voung (1989)).

This last point raises the first important question of the paper: how does the Bayes Factor perform when the sample becomes larger?

2.3. Convergence Theorems

In this subsection we will prove two theorems. First, we will show that the posterior distribution of the parameters collapses to their pseudotrue values. Second, we will demonstrate that posterior odds ratio of any model over the best model under the Kullback-Leibler distance will approach zero. As mentioned before, these results are important because we do not require models to be well-specified and/or nested as the results existing in the literature. With these two theorems we follow the recent literature on the asymptotic properties of Bayesian inference. Examples include Phillips and Ploberger (1996), Phillips (1996) and Kim (1998).

The structure of this subsection is as follows. First, we prove lemmas 1 and 2. The first lemma states the asymptotic concentration of the posterior around the PMLE, and the second states the consistency of PMLE to the pseudo-true value. These two lemmas imply the first of the theorems: the posterior concentrates asymptotically around the pseudo-true value. Then, we prove the second theorem stating that the Bayes Factor of any other model over

the model closest to P_0^T under the Kullback-Leibler distance will asymptotically approach zero.

Let us begin making three technical assumptions:

Condition 1. *For* $\forall\, i \in M$ *and* $\forall\, \theta \in \Theta_i$*:*

$$\lim_{T\to\infty} P_0^T \left(T^{-1} \log f^T(Y^T|\theta, i) < \infty\right) = 1 \tag{5}$$

Condition 2. *For* $\forall\, i \in M$*:*

$$\lim_{T\to\infty} P_0^T \left(T^{-1} \log f^T(Y^T|\widehat{\theta}_T^*(i), i) > -\infty\right) = 1 \tag{6}$$

Condition 3. *For* $\forall\, i \in M$*:*

$$\lim_{T\to\infty} P_0^T \left(\left|f^T\left(Y^T|i\right)\right| = 0\right) = 0 \tag{7}$$

$$\lim_{T\to\infty} P_0^T \left(f^T\left(Y^T|\widehat{\theta}_T^*(i), i\right)\pi\left(\widehat{\theta}_T^*(i)\,|i\right) = 0\right) = 0 \tag{8}$$

Conditions 1 and 2 bound the loglikelihood while Condition 3 precludes priors without support on the pseudo-true value (as dogmatic priors except in quite remarkable cases when the dirac point is exactly $\widehat{\theta}_T^*(i)$).

Following Chen (1985) and Kim (1998), we will approach the analysis of the posterior behavior defining a "shrinking neighborhood system" in the parameter space:

Definition 4. *For* $\forall\ a \in \Theta_i \subseteq R^{k_i}$ *and* $\forall\, i \in M$*, a shrinking neighborhood system is a collection of* k_i-*dimensional ellipsoids* $\{\mathcal{E}(a, \delta_j(i)),\ j = 1, 2, ...\}$ *such that:*

$$\mathcal{E}(a, \delta_j(i)) \equiv \left\{\theta \in \Theta_i : |a_1 - \theta_1|^2 + ... + \left|a_{k(i)} - \theta_{k(i)}\right|^2 < \delta_j(i)\right\} \tag{9}$$

where $\delta_j(i) \in R$, $j = 1, 2, ...$

The idea behind this system is to look at the parameter values close enough to some k_i-dimensional point a, making the values of $\delta_j(i)$ smaller as $T \nearrow \infty$. We will accomplish this task requiring the likelihood function to progressively concentrate around a point:

Condition 5. *For* $\forall\, i \in M$ *and* $\forall\, \theta \in \Theta_i$*, let* $\{\delta_t(i)\}_{t=1}^{\infty}$ *such that* $\mathcal{E}(a, \delta_t(i)) \subseteq \mathcal{E}(a, \delta_{t-1}(i))$ *and* $\cap_{t=1}^{\infty}\mathcal{E}(a, \delta_t(i)) = \{a\}$*. Then, there exists a sequence of nonincreasing positive functions* $\{k_T(\delta_T(i), i), T = 1, 2, ...\}$ *such that* $Tk_T(\delta_T(i), i) \nearrow \infty$ *and*

$$\lim_{T\to\infty} \inf P_0^T \left(\sup_{\theta\in\Theta_i\setminus\mathcal{E}\left(\widehat{\theta}_T(i,Y^T),\delta_T(i)\right)} \frac{\log f^T(Y^T|\theta, i) - \log f^T(Y^T|\widehat{\theta}_T^*(i), i)}{T} \le -k_T(\delta_T(i), i)\right) = 1 \tag{10}$$

Given these conditions, we are ready to prove the following lemma:

Lemma 6. *Under Conditions (1)-(4),* $\int_{\theta\in\Theta_i\backslash\mathcal{E}\left(\widehat{\theta}_T(i,Y^T),\delta_T(i)\right)}\pi(\theta|Y^T,i)d\theta\to 0$ *as* $T\to\infty$ *in* P_0*-probability* $\forall\, i\in M$.

Proof. See appendix ■

For the second lemma we need some further technical assumptions:

Condition 7.

$$\lim_{T\to\infty}P_0^T\left(\left|\frac{f^T\left(Y^T|\widehat{\theta}_T\left(i,Y^T\right),i\right)}{f^T\left(Y^T|i\right)}\right|=0\right)=0 \tag{11}$$

Condition 8. $\forall\eta>0$

$$\lim_{T\to\infty}P_0^T\left(\int_{\mathcal{B}(\widehat{\theta}_T^*(i),\eta)}\pi\left(\theta,i\right)\exp\left[\log f^T(Y^T|\theta,i)-\log f^T(Y^T|\widehat{\theta}_T^*\left(i\right),i)\right]d\theta>0\right)=1 \tag{12}$$

where $\mathcal{B}(a,\eta)\equiv\{\theta:|\theta-a|<\eta\}$

Lemma 9. *Under Conditions (1)-(7),* $\widehat{\theta}_T^*\left(i\right)-\widehat{\theta}^*\left(i\right)\to 0$ *as* $T\to\infty$ *in* P_0*-probability.*

Proof. See appendix ■

With these two lemmas, it can be shown:

Theorem 10. *The Bayes point estimator converges to the pseudo-true value of the parameter.*

Proof. This follows directly from lemmas 1 and 2. ■

Note that Theorem 1 does not ask for any specific loss function. Any sensible loss function will choose the only point with positive posterior as $T\nearrow\infty$.

Now, to prove the second of the theorems we need first some notation for the measure 0 sets where Lemma 2 does not hold.

Definition 11. $\forall\, i\in M$ *let* $\Omega_0\left(i\right)\subset\Omega$ *such that Lemma 2 does not hold. Also let* $\Omega_0\equiv\cup_{i\in M}\Omega_0\left(i\right)$.

Condition 12. $\forall\ \{y_t\}_{t=1}^{\infty}\in\Omega\backslash\Omega_0$, $\sup_T f^T\left(y^T|\widehat{\theta}_T\left(i,y^T\right),i\right)<\infty$

Condition 13. $\forall\, i\in M$ *and* $\forall\,\theta\in\Theta_i$, $\int_{\Re^{m\times T}}f^T\left(Y^T|\theta,i\right)p_0^T\left(Y^T\right)d\nu^T$ *exists and it is finite for* $T=1,2,3,...$

Condition 14. $\forall\, i\in M$, $\int_{\Re^{m\times T}}f^T\left(Y^T|\theta,i\right)p_0^T\left(Y^T\right)d\nu^T$ *is continuous on* Θ_i *for* $T=1,2,3,...$

Condition 15. $\forall\, i\in M$, $\left\{f^T\left(Y^T|\theta,i\right)\right\}_{t=0}^{\infty}$ *obeys a strong uniform law of the large numbers*[1].

[1] Andrews (1988) proves laws of large numbers for L^1-mixingales. We proved, but we do not include, that an exponential density family, $\{\log f^t\left(Y^t|\theta,i\right)\}_{t=1}^{\infty}$ is a L^1-mixingale.

Condition 16. $\exists\, i \in M$ *such that* $\exists T_0$ *such that* $\forall\, T \geq T_0$

$$\int_{\Re^{m\times T}} T^{-1} \log f^T\left(Y^T|\widehat{\theta}_T^*(i), i\right) p_0^T\left(Y^T\right) d\nu^T > \int_{\Re^{m\times T}} T^{-1} \log f^T\left(Y^T|\widehat{\theta}_T^*(j), j\right) p_0^T\left(Y^T\right) d\nu^T$$

$\forall\, j \in M \backslash \{i\}$

Condition 17. $\forall\, i \in M$ *and* $\forall\ T$, $0 < \pi\left(\widehat{\theta}_T^*(i)\,|i\right) < \infty$

Of this long list of conditions we only see Condition 11 as slightly restrictive and even with respect to this one, the results in Andrews (1988) make it quite general. The other conditions basically only require the model comparison to be a meaningful task.

Finally, we are ready to prove the main result in this paper, i.e. that the Bayes factor will select the model closest to the data regardless of the priors used.

Theorem 18. *Under Conditions (1)-(13),* $\lim_{T\to\infty} P_{0T}\left(B_{j,i|Y^T} = 0\right) = 1$.

Proof. See appendix ■

The second theorem is closely related to the asymptotic justification of the Schwarz Information Criterion (Kass and Raftery (1995)) and the Posterior Information Criterion (Phillips and Ploberger (1996)). Both criteria had been proposed as simple ways to choose among competing models. We think, however, that computing the Bayes factor is the appropriate choice. Even if these other criteria are easy to compute, in general we will know relatively little about their small sample properties. The Bayes factor, in comparison, is well understood regardless of the sample size, and we can always check its robustness against different priors.

Finally, we conjecture, based in similar arguments to Phillips (1996) and Kim (1998), the asymptotic normality of the posterior. We do not seek to use asymptotic approximations to the posteriors because the use of Markov chain Monte Carlo method allows pseudo-exact (up to a simulation error) Bayesian computations. As a consequence, we do not see normality as a very interesting result in our context.

2.4. Numerical Implementation

From our previous description, it should be clear that the actual implementation of Bayesian inference requires two conditions: being able to evaluate the likelihood function for arbitrary parameter values and being able to compute the marginal likelihood.

The first task can be accomplished using a State Space representation of the economy. If this representation is linear (a Linear Quadratic Approximation or loglinearization of the Euler Conditions can achieve this objective), the *Kalman Filter* provides an efficient procedure to evaluate the likelihood. If this representation is nonlinear, more involved procedures are required. In a companion paper, Fernández-Villaverde and Rubio-Ramírez (2001) show how to use *Monte Carlo filters* to evaluate the likelihood function. State Space representations also allow the use of different possible solutions to a common problem in dynamic equilibrium economies: their stochastic singularity. Since the number of stochastic innovations in the models is usually lower than the dimensions of the data we are studying, their variance-covariance matrix is singular. These solutions include augmenting the sources of

randomness in the model (Leeper and Sims (1994)), or introducing measurement errors in some observable. In this paper we are agnostic about how to solve this singularity: we merely point out how State Space representations may potentially deal with this problem.

For the second task we can use Markov chain Monte Carlo methods to both approximate the posterior distribution of the parameters of the model and compute the marginal likelihood of the model.

3. A Dynamic Equilibrium Model: the cattle cycle

Once we have shown the asymptotic properties of the Bayesian approach to inference and model comparison of dynamic equilibrium economies, the rest of the paper explores the small sample behavior of the procedure. To do so, we first need a model for its posterior econometric analysis. This section fills that need by presenting a model of the cattle cycle.

3.1. The cattle cycle

As pointed out numerous times, cattle stocks are among the most periodic time series in economics. The standard model to account for this behavior is based on Rosen, Murphy and Scheinkman (1994) as modified by Hansen, McGrattan and Sargent (1994)[2].

Two reasons suggest the choice of this application. First, despite its relative simplicity, this model delivers a rich and easily tractable dynamic that has been argued to be able to account for the observed data (Rosen, Murphy and Scheinkman (1994)). Second, and more importantly, a number of different estimation procedures have been performed with basically the same model and data and under the more or less explicit assumption that the model is misspecified. For instance, Rosen, Murphy and Scheinkman (1994) mix calibration and ARMA estimation; Hansen, McGrattan and Sargent (1994) use Maximum Likelihood Methods; and Diebold, Ohanian and Berkowitz (1998) minimize the spectral distance between the data and the model. All these procedures give us a benchmark set to assess the performance of Bayesian method: we will know that any surprising or different result will come from the econometric approach and not from the model itself.

3.2. The Model

There is a representative farmer who breeds cattle and slaughters it for the market. Adult stocks are either held for breeding or slaughtered. After one year, each animal in the breeding stock, x_t, gives birth to g calves. Calves became part of the adult stock after two cycles. Therefore, given an exponential death rate δ for the breeding stock and a slaughtering rate c_t, x_t is given by $x_t = (1-\delta)\, x_{t-1} + g x_{t-3} - c_t$ and the total head count of cattle (the sum of adults, yearlings and calves) is $s_t = x_t + g x_{t-1} + g x_{t-2}$.

The price of freshly slaughtered beef is p_t (we assume no difference in the quality of beef depending on age). There are two types of cost for the farmer. The first type includes the feeding cost of preparing an animal for slaughter, m_t, the one period cost of holding an

[2] A second version of this paper has been published as Anderson, Hansen, McGrattan and Sargent (1996). This second version omits however important details in the description of the Cattle Cycles model. The first version is freely available at `http://woodrow.mpls.frb.fed.us/research/sr/sr182.html`

adult, h_t, of holding a yearling, $\gamma_0 h_t$ and of holding a calf, $\gamma_1 h_t$. These costs are exogenous, autoregressive, stochastic stationary processes:

$$\begin{aligned} h_{t+1} &= (1-\rho_h)\,\mu_h + h_t + \varepsilon_{ht} \text{ where } \varepsilon_{ht} \sim N\left(0, \sigma_h^2\right) && (13) \\ m_{t+1} &= (1-\rho_h)\,\mu_m + m_t + \varepsilon_{mt} \text{ where } \varepsilon_{mt} \sim N\left(0, \sigma_m^2\right) && (14) \end{aligned}$$

The second type of cost is associated with the holding and slaughtering of cattle and has a quadratic structure $\Psi = \frac{\psi_1}{2}x_t^2 + \frac{\psi_2}{2}x_{t-1}^2 + \frac{\psi_3}{2}x_{t-2}^2 + \frac{\psi_4}{2}c_t^2$ where ψ_i are small, positive parameters.

A representative farmer, taking as given the vector sequence $\{p_t, h_t, m_t\}_{t=0}^{t=\infty}$, solves the following maximization problem:

$$\max_{\{c_t\}_{t=0}^{t=\infty}} E_0 \sum_{t=0}^{\infty} \beta^t \left\{ \begin{array}{c} (p_t - m_t)\,c_t - h_t x_t - \gamma_0 h_t g x_{t-1} - \gamma_1 h_t g x_{t-2} \\ -\frac{\psi_1}{2}x_t^2 - \frac{\psi_2}{2}x_{t-1}^2 - \frac{\psi_3}{2}x_{t-2}^2 - \frac{\psi_4}{2}c_t^2 \end{array} \right\} \tag{15}$$

$$\begin{aligned} s.t.\ x_t &= (1-\delta)\,x_{t-1} + g x_{t-3} - c_t \\ &\quad \{x_{-1}, x_{-2}, x_{-3}\} \text{ fixed} \end{aligned}$$

The quadratic costs can be rewritten in a more convenient way. Set $g_{1t} = f_1 x_t + f_2 h_t$, $g_{2t} = f_3 x_{t-1} + f_4 h_t$, $g_{3t} = f_5 x_{t-1} + f_6 h_t$ and $g_{4t} = f_7 c_t + f_8 m_t$. Notice that:

$$\begin{aligned} g_{1t}^2 &= f_1^2 x_t + f_2^2 h_t + 2 f_1 f_2 x_t h_t \\ g_{2t}^2 &= f_3^2 x_{t-1} + f_4^2 h_t + 2 f_3 f_4 x_{t-1} h_t \\ g_{3t}^2 &= f_5^2 x_{t-1} + f_6^2 h_t + 2 f_5 f_6 x_{t-2} h_t \\ g_{4t} &= f_7 c_t + f_8 m_t + 2 f_7 f_8 c_t m_t \end{aligned}$$

and then, using the problem of the farmer, we can find:

$$\begin{aligned} f_1^2 &= \frac{\psi_1}{2},\ f_2^2 = \frac{\psi_2}{2},\ f_3^2 = \frac{\psi_3}{2},\ f_7^2 = \frac{\psi_4}{2} && (16) \\ 2 f_1 f_2 &= 1,\ 2 f_3 f_4 = g\gamma_1,\ 2 f_5 f_6 = g\gamma_0,\ 2 f_7 f_8 = 1 && (17) \end{aligned}$$

From these equations, the ψ_i's and four of the f's can be found given the other four f's.

The model is closed with a demand function $c_t = \alpha_0 - \alpha_1 p_t + d_t$ where α_0, $\alpha_1 > 0$ are the parameters of the demand and d_t is a stochastic, autoregressive, stationary, demand shifter with zero mean, $d_{t+1} = \rho_d d_t + \varepsilon_{dt}$ where $\varepsilon_{dt} \sim N\left(0, \sigma_d^2\right)$.

Finally, we will assume that there is a measurement error in the total stock of cattle, s_t and the slaughter rate, c_t, such that the observed rates are given by:

$$\begin{aligned} \widetilde{s_t} &= s_t + \varepsilon_{yt} \text{ where } \varepsilon_{st} \sim N\left(0, \sigma_s^2\right) && (18) \\ \widetilde{c_t} &= c_t + \varepsilon_{ct} \text{ where } \varepsilon_{ct} \sim N\left(0, \sigma_c^2\right) && (19) \end{aligned}$$

We are ready now to define a competitive equilibrium for this economy:

Definition 19. *A Competitive Equilibrium for the Cattle Industry is a sequence of beef consumptions* $\{c_t\}_{t=0}^{\infty}$*, cattle stocks* $\{s_t\}_{t=0}^{\infty}$*, breeding stocks* $\{x_t\}_{t=0}^{\infty}$*, prices* $\{p_t\}_{t=0}^{\infty}$*, exogenous stochastic processes* $\{h_t, m_t, d_t\}_{t=0}^{\infty}$ *and initial conditions* $\{x_{-1}, x_{-2}, x_{-3}\}$ *such that:*

1. *Given prices, the stochastic processes and initial conditions, the representative farmer solves its problem:*

$$\max_{\{c_t\}_{t=0}^{t=\infty}} E_0 \sum_{t=0}^{\infty} \beta^t \left\{ \begin{array}{c} (p_t - m_t)\, c_t - h_t x_t - \gamma_0 h_t g x_{t-1} - \gamma_1 h_t g x_{t-2} \\ -f_1^2 x_t^2 - \frac{1}{4f_1^2} x_{t-1}^2 - f_5^2 x_{t-2}^2 - f_7^2 c_t^2 \end{array} \right\} \tag{20}$$

$$s.t.\ x_t = (1-\delta)\, x_{t-1} + g x_{t-3} - c_t$$

2. *Demand is given by* $c_t = \alpha_0 - \alpha_1 p_t$.

3. *Stocks evolve given by* $x_t = (1-\delta)\, x_{t-1} + g x_{t-3} - c_t$ *and* $s_t = x_t + g x_{t-1} + g x_{t-2}$.

4. *Stochastic Processes follow:*

$$h_{t+1} = (1-\rho_h)\,\mu_h + h_t + \varepsilon_{ht} \text{ where } \varepsilon_{ht} \sim N\left(0, \sigma_h^2\right) \tag{21}$$

$$m_{t+1} = (1-\rho_h)\,\mu_m + m_t + \varepsilon_{mt} \text{ where } \varepsilon_{mt} \sim N\left(0, \sigma_m^2\right) \tag{22}$$

$$d_{t+1} = \rho_d d_t + \varepsilon_{dt} \text{ where } \varepsilon_{dt} \sim N\left(0, \sigma_d^2\right) \tag{23}$$

4. A Structural Estimation of the cattle cycle Model

In this section, we estimate the structural parameters (the parameters that determine the technology and preferences) of the cattle cycle model and its associated marginal likelihood, using the annual measured total stock of beef, measured slaughter rate and price of slaughter beef for 1900-1990 (Bureau of Census (1975) and (1989)). First, we will specify priors over these structural parameters. Second, using the *Metropolis-Hastings* algorithm and the Kalman Filter, we will find the posterior distributions and moments of the parameters. To check the accuracy of our computations, we will also present estimates of our numerical errors and convergence assessment of our Markov chain Monte Carlo. In addition, we will study the robustness of the results to different priors. Finally, assuming a quadratic loss function, we will compare our point estimates with the results of Maximum Likelihood estimation (MLE).

4.1. Specifying the Priors

The parameters of the cattle cycle model described above are collected in a 21 dimensional vector $\theta = \{\beta, \delta, \alpha_0, \alpha_1, \gamma_0, \gamma_1, g, \rho_h, \rho_m, \rho_d, \mu_h, \mu_m, \sigma_h, \sigma_m, \sigma_s,, \sigma_c, \sigma_d, f_1, f_2, f_3, f_4\}$. We will impose dogmatic priors on 10 parameters[3]. This restriction plays two different roles. First, since it reduces the dimensionality of this problem by half, the computational burden is greatly diminished. Second, since this the same restriction used in Hansen, McGrattan and Sargent (1994), it increases the comparability of our results to previous estimations. We will set $\beta = 0.96$, $\delta = 0$, $f_1 = f_3 = f_5 = f_7 = 0.0001$, $\rho_d = \sigma_h = 0$, $\mu_h = 37$, $\mu_m = 63$. The first restriction pins down the discount factor, a difficult parameter to estimate in this type of models, to a commonly used value. The second one rules out deaths in the breeding stock. The value for the f's is a small number that creates the quadratic costs and it is basically irrelevant.

[3]Formally, our prior over these parameters will be a dirac function that implies a dirac posterior.

The last restrictions make demand deterministic and fix the mean value of the processes to the observed means. The remaining vector is then $\theta' = \{\alpha_0, \alpha_1, \gamma_0, \gamma_1, g, \rho_h, \rho_m, \sigma_h, \sigma_m, \sigma_{s,}, \sigma_c\}$.

We adopt standard priors for these parameters. The independent term of the demand function follows a normal distribution with mean 146 and variance 35, the point MLE. The next three parameters follow a Gamma distribution with hyperparameters 2 and 0.5, that imply a mean of 1 and variance of 0.5. This choice gives support to all positives values of those parameters. That means that, in the case of α_1, we only impose the condition that the good is not *giffen* (we are not aware of any evidence supporting the hypothesis that beef is a *giffen* good). The mean of 1 is a focal point for the effect of changes of prices on beef consumption. A not very tight variance of 0.5 spreads the density enough around this value. For the case of γ_0 and γ_1 we require that both costs of raising beef are positive. Setting the mean to 1 is intuitive (different types of cattle should not have very different relative holding costs) and the variance to 0.5 shows that we are relatively unsure about that guess. The growth factor is set to obey a normal centered at 1: the number of births per animal in stock is one per year with a small variance. Biological constraints justify this choice. The autoregressive terms follow a beta with mean 0.6 and variance 0.04, i.e. the process is stationary, with positive autocorrelation and with mean skewed to the right in a somehow imprecise way. For the four variances of the innovations terms we choose gamma distributions to stay in the positive reals. The parameters 2,1 reflect an (imprecise) opinion in favor of large variances (mean and variance of 2). Table 4.1 summarizes the previous discussion.

Table 4.1: Priors for the Parameters of the cattle cycle Model

Parameters	Distribution	Hyperparameters
α_0	Normal	146,35
α_1	Gamma	2,0.5
γ_0	Gamma	2,0.5
γ_1	Gamma	2,0.5
g	Normal	1,0.1
ρ_h	Beta	3,2
ρ_m	Beta	3,2
σ_h	Gamma	2,1
σ_m	Gamma	2,1
σ_s	Gamma	2,1
σ_c	Gamma	2,1

4.2. Results

As previously discussed, to solve for the lack of tractable expressions for the likelihood and posterior distributions of model parameters, we use the Kalman Filter to evaluate the likelihood of the model with parameters values generated by a *Random-Walk Metropolis-Hastings Algorithm* (see Robert and Casella (1999)). This procedure produces a Markov chain $\{\theta_1, \theta_2, ...\}$ of size m of parameter values such that the distribution of these values converges to the true posterior implied by the likelihood and the prior. The empirical histograms of the parameters in our estimation are included as Figure 2.

Given this Markov chain and a function of interest $g(\cdot)$ defined over some aspect of the

simulation output θ_i, the expectation of such function, $\mu = E(g(\theta))$ can be approximated by a strong law of large numbers by $\widehat{\mu} = \frac{1}{m}\sum_{i=1}^{m} g(\theta_i)$. Then, using appropriate indicators functions, we can approximate the different moments of the distribution or compute quantiles. Also, an appropriate Central Limit Theorem assures that $\sqrt{m}(\widehat{\mu} - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma_\mu^2\right)$, allowing us to evaluate the accuracy and stability of the estimates and to build probability intervals statements.

We simulate a chain of size 10^6 that passes all the requirements of convergence (more details below). Table 4.2 reports the expectation and standard deviation for the parameters.

Table 4.2. Parameters Statistics

Parameters	Expectation	s.d.
α_0	146.23	20.62
α_1	1.27	0.20
γ_0	1.02	0.52
γ_1	1.36	0.54
g	0.95	0.04
ρ_h	0.93	0.03
ρ_m	0.70	0.03
σ_h	5.30	1.31
σ_m	4.05	0.68
σ_s	0.33	0.10
σ_c	4.54	0.58

The computation of the marginal likelihood is done using the method proposed by Gelfand and Dey (1994). For any k-dimensional probability density $h(\cdot)$ with support contained in Θ, Gelfand and Dey noted that:

$$E\left[\left.\frac{h(\theta)}{f^T(Y^T|\theta,i)\,\pi(\theta)}\right| Y_T, i\right] = \int_\Theta \frac{h(\theta)}{f^T(Y^T|\theta,i)\,\pi(\theta)} f^T\left(\theta|Y^T,i\right)d\theta =$$

$$= \int_\Theta \frac{h(\theta)}{f^T(Y^T|\theta,i)\,\pi(\theta)} \frac{f^T(Y_T|\theta,i)\,\pi(\theta)}{\int_\Theta f^T(Y_T|\theta,i)\,\pi(\theta)\,d\theta} d\theta = \frac{\int_\Theta h(\theta)\,d\theta}{\int_\Theta f^T(Y^T|\theta,i)\,\pi(\theta)\,d\theta} = f^T\left(Y^T|i\right)^{-1} \tag{24}$$

This expression is an unbiased and consistent estimator of the marginal likelihood and satisfies a Central Limit Theorem if $\frac{\int_\Theta h^2(\theta)d\theta}{\int_\Theta f^T(Y^T|\theta,i)\pi(\theta)d\theta} < \infty$. Then, from the m draws of the simulation and applying a Strong Law of Large Numbers, we can compute:

$$f^T\left(Y^T|i\right)^{-1} = \frac{1}{m}\sum_{i=1}^{m} \frac{h(\theta)}{f^T(Y^T|\theta,i)\,\pi(\theta)} \tag{25}$$

As a choice of h we modify Geweke's (1998) proposal. First, from the output of the simulation define $\widehat{\theta_M} = \frac{1}{m}\sum_{i=1}^{m}\theta$ and $\widehat{\Sigma_m} = \frac{1}{m}\sum_{i=1}^{m}\left(\theta - \widehat{\theta}\right)\left(\theta - \widehat{\theta}\right)'$. Then, for a given $p \in (0,1)$ define the set $\Theta_M = \left\{\theta : \left(\theta - \widehat{\theta}\right)\widehat{\Sigma_m}^{-1}\left(\theta - \widehat{\theta}\right)' \leq \chi^2_{1-p}(11)\right\}$ where $\chi^2_{1-p}(\cdot)$ is a

chi-squared distribution with degrees of freedom equal to the number of parameters. Letting $I_{\Theta_M \cap \Theta}(\cdot)$ be the indicator function of a vector of parameters belonging to the intersection $\Theta_M \cap \Theta$, we can take a truncated multivariate normal as our h function:

$$h(\theta) = \frac{1}{\widehat{p}(2\pi)^{\frac{k}{2}}} \left|\widehat{\Sigma_m}\right|^{\frac{1}{2}} \exp\left[-0.5\left(\theta - \widehat{\theta}\right)\widehat{\Sigma_m}^{-1}\left(\theta - \widehat{\theta}\right)'\right] I_{\Theta_M \cap \Theta}(\theta) \tag{26}$$

where $\widehat{p}$ is an appropriate normalizing constant. With this choice, if the posterior density is uniformly bounded away from zero on every compact subset of Θ, our computation approximates the marginal likelihood.

With the output of the Markov chain Monte Carlo, the estimation of the marginal likelihood is then rather direct: we use the computed values of $f^T\left(Y^T|\theta, i\right)\pi(\theta)$ for each point in the Markov chain and we find its harmonic mean using the function h as a weight. Following this procedure, our estimated marginal likelihood value is $\exp(-647.5281)$.

4.3. Computation of the Numerical Standard Error

The convergence of $\widehat{\mu}$ to its true value established by the Central Limit Theorem is of little use without the estimation of the asymptotic variance or, its square root, the numerical standard error (NSE). This estimation is complicated by the lack of independent sampling in the simulated Markov chain. Different methods have been proposed to overcome this problem. We follow here a simple suggestion by Hannan (1970). Assuming that the function of interest $g(\cdot)$ has a spectral density $S_g(\omega)$ continuous at the origin[4], we can estimate the NSE as $\left(\frac{1}{m}\widehat{S_g}(0)\right)^{\frac{1}{2}}$ (Corollary 4, page 208 in Hannan (1970)). We computed the required power spectral density using a Welch's averaged, modified periodogram method. All the estimated NSEs were less than 0.5% of the mean value of the parameter, suggesting tight estimations and confirming the evidence from repeated simulations that systematically generated nearly identical values for the means.

4.4. Assessing Convergence

Maybe the most important issue in the empirical implementation of a Markov chain Monte Carlo is to assess the convergence of the simulation (see Mengersen, Robert and Guihenneuc-Jouyaux (1999)). Theorems to this respect require conditions difficult to check in practice. As a response, the use of informal methods to check convergence has been quite common. As an example of these informal methods, we simulated 10 chains of size 10^5 and one of size 10^6. All of them generated very similar results and their draws seemed to follow a stationary process. However, informal methods can hide subtle nonconvergence problems.

To address this issue, we implement the convergence test proposed by Geweke (1992). We take the first p_A and the last p_B vectors of the simulation and compute $\widehat{\mu_1} = \frac{1}{p_A}\sum_{i=1}^{p_a} g(\theta_i)$

[4] A sufficient condition for continuity is given by the strict stationarity of the simulation (Corollary 1, page 205, Hannan (1970)) as it is the case if the conditions for consistency of section 2 hold. In practice strict stationarity can be checked using standard tests.

and $\widehat{\mu_2} = \frac{1}{p_B}\sum_{i=m-p_B+1}^{m} g(\theta_i)$. Then, as $m \to \infty$,

$$\frac{(\widehat{\mu_1} - \widehat{\mu_2})}{\left[\frac{\widehat{S_g^A}(0)}{p_A} + \frac{\widehat{S_g^B}(0)}{p_B}\right]^{\frac{1}{2}}} \Rightarrow \mathcal{N}(0,1)$$

The computed values of the test for each first moment were all less than $|0.7 * 10^{-4}|$, strongly supporting that, as previously suggested, our simulation converges.

4.5. Robustness Analysis

The subjective character of the Bayesian paradigm calls for an indication of how the posterior expectations differ with changes in the prior distribution. In that way we can avoid spurious findings in favor of one model purely based on "strategically" chosen priors.

Methods to undertake robustness analysis have been presented in Geweke (1999). These methods allow modifications of the priors in a generic and fast way. A general approach defines, for any prior density $\pi^*(\theta)$ with support included in our prior $\pi(\theta)$ support[5], the weighting function $w(\theta) = \frac{\pi^*(\theta)}{\pi(\theta)}$ and finds the new posterior functions of interest as $\widehat{\mu} = \frac{1}{m}\frac{\sum_{i=1}^{m} w(\theta) g(\theta_i)}{\sum_{i=1}^{m} w(\theta)}$.

An extensive prior set was tested without altering substantially the reported results. We attribute that to the fact that the sample size is big enough to swamp the prior (we can think of the prior loosely as an additional "dummy observation" without too much weight when the sample consists of 91 periods). However, our robustness checks may be quite different from what the reader desires. As a consequence, upon request, we will electronically deliver the simulator output matrices and required documentation. These simulation matrices include the draws from the posterior, θ_i, the likelihood times the prior $f^T\left(Y^T|\theta_i, i\right)\pi(\theta)$, and the prior values $\pi(\theta_i)$ $i = 1, ..., m$, for each of the different models described in the paper. With these matrices, the application of a reweighting scheme will allow third parties to quickly recompute both the moments of interest and the marginal likelihood with any desired prior that satisfies the support condition.

4.6. Comparison with Other Results

One of the reasons for the choice of the cattle cycle model as an application was the existence of previous econometric estimations of the model we could use as benchmarks to assess the performance of the Bayesian procedure.

We will only discuss in detail the closest existing estimation- the one in Hansen, McGrattan and Sargent (1994) that estimated the same model with the same parametric restrictions and data using MLE. We successfully reproduced their point and standard error estimation (table 4.3).

Comparison with table 4.2 highlights two main points. First, the MLE with low standard error (*precise* estimates) are closely matched (α_1 equals to 1.27 against 1.27, ρ_m equal 0.70

[5]Note that the only restriction to the support in our model has a strong theory base: i.e. cost should be positive and so on.

against 0.70, etc.). Second, for those parameters imprecisely estimated, as γ_0 and γ_1 (the relative holding costs of cattle according to its age), the Bayes estimate is both more precise and closer to our intuition of relatively homogenous costs of holding differently aged cattle. Figure 2 explains the result. While the posteriors of α_1 or α_0 are well-behaved and unimodal, the posteriors of γ_0 and $\gamma_1$are multimodal and relatively flat over a long range of values. Given these shapes, the MLE will tend to find one of the local maxima, where the *Metropolis-Hastings* algorithm stays longer because the likelihood is higher (the point estimates are local maxima in the densities) while the flatness of the likelihood will turn out very high standard errors. The Bayes estimate overcomes these difficulties and gives a much more accurate finite sample view of the plausible parameter values[6]. We interpret this result as a strong endorsement of the small sample properties of Bayesian estimation. This result is also similar to other frequentist evaluations of the small sample performance of Bayesian methods, as in Jacquier, Polson and Rossi (1994) and Geweke, Keane and Runkle (1997).

Table 4.3. ML estimation for cattle cycle

Parameters	Estimates	s.e.
α_0	146	33.4
α_1	1.27	0.323
γ_0	0.65	11.5
γ_1	1.77	12
g	0.94	0.0222
ρ_h	0.89	0.115
ρ_m	0.70	0.0417
σ_h	6.82	10.6
σ_m	4.04	1.05
σ_s	0.27	0.0383
σ_c	4.82	0.531

Once we have estimated the cattle cycle model, the next question to address is to explore how it compares with alternatives accounts of the data, i.e. with competing models. We perform this model comparison in the next section.

5. Comparing Models: the cattle cycle vs. BVARs

In this section we will compare the cattle cycle model with a set of Bayesian Vector Autoregressions (BVARs). This choice is motivated by our desire to compare a dynamic equilibrium model against a pure and powerful statistical competitor. Vector Autoregressions models, a simple linear statistical representation of the dynamic relations among variables, have a proven forecasting record (Litterman (1986)) and have been often proposed as alternatives to a more structural modelling of time series (Sims (1980)[7]. We will describe first the Vector Autoregression specification, then the priors used and finally we will show the result of the comparison of models.

[6]Through robustness analysis, we checked that this higher precision is not spuriously induced by the prior.

[7]Note that, however, these BVARs are not competely nonnested with the cattle cycle model since the latter has a restricted vector autoregression representation. We thank Tom Sargent for this comment.

5.1. A Vector Autoregression Specification

We will define nine versions of a three variables BVAR, indexed by the number of lags (1, 2 and 3) and by three different priors. Let y_t be the row vector of three observed variables at time t. The p lags BVAR can be written as:

$$y_t = \sum_{i=1}^{p} y_{t-i} A_i + C + u_t \qquad \forall t \in \{1, ..., T\},\ u_t \sim \mathcal{N}(0, \Psi) \tag{27}$$

where A_i and C are parameter matrices of dimension 3×3 and 3×1 respectively.

A useful way to rewrite (27) is as follows. Define $y_t = z_t \Gamma + u_t$ where $z_t = (I, y_{t-1}, ..., y_{t-p})$ and $\Gamma = \left(C', A_1', ..., A_p'\right)'$. Stacking the row vectors y_t, z_t and u_t in Y, Z and U such that $Y = Z\Gamma + U$ and letting the i subscript denote the ith column vector, we will have $y_i = Z\gamma_i + u_i$. Stacking now the column vectors y_i, γ_i and u_i in y, γ and u, we finally get the much more convenient form $y = (I \otimes Z)\gamma + u$, where $u \sim \mathcal{N}(0, \Psi \otimes I)$. The likelihood function is given then by:

$$f^T(\gamma|\Psi) \propto |\Psi|^{-\frac{T}{2}} \exp\left\{-tr\left[(Y - Z\Gamma)'\Psi^{-1}(Y - Z\Gamma)\right]/2\right\} \tag{28}$$

5.2. Prior Distributions

In order to show the power of model comparison, we will use three different priors, each more general than the previous one: a modified *Minnesota prior*, a *Normal-Wishart prior* and a *Hierarchial prior* (see also Kadiyala and Karlsson (1997) and Sims and Zha (1998)).

5.2.1. Minnesota prior

Litterman (1980) defined the often-called *Minnesota prior*. The basic feature of this prior is that the prior mean for the parameter on the first own lag is set to unit and the prior mean of the remaining parameters in γ_i are set to zero, i.e. the mean of the regression in each variable is specified as a random walk.

To win further flexibility, we will modify two aspects of this prior. First, we will let the prior variances decrease slower with the lags. Litterman used a rate $\frac{1}{k^2}$ while we use $\frac{1}{k}$. Second, we will not restrict the variance-covariance matrix to be diagonal since, thanks to the use of simulation methods, we are not looking for a closed form for the posterior distributions.

In more detail, our version of the *Minnesota prior* for p lags is:

1. The prior mean for the parameter on the first own lag is set to unit and the prior mean of the remaining parameters are set to zero, i.e. the mean of γ_s for $s \in \{1, 2, 3\}$ is $\underline{\mu_s} = \left(0, \chi_{\{1\}}(s-1), \chi_{\{1\}}(s-2), \chi_{\{1\}}(s-3), 0, ..., 0\right)'_{(1+3p)x1}$.

2. The variance of γ_s for $s \in \{1, 2, 3\}$ is equal to:

$$\underline{\Sigma_s} = \begin{pmatrix} \pi_3 \sigma_s^2 & 0 & \cdots & 0 \\ 0 & \widetilde{\pi}_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \widetilde{\pi}_p \end{pmatrix}_{(1+3p)x(1+3p)} \tag{29}$$

where σ_i is a scale factor accounting for the variability of the different variables and $\widetilde{\pi}_1 = \pi(\chi_{\{1\}}(s-1))\frac{\sigma_s^2}{\sigma_1^2}$, $\widetilde{\pi}_2 = \pi(\chi_{\{1\}}(s-2))\frac{\sigma_s^2}{\sigma_2^2}$, $\widetilde{\pi}_3 = \pi(\chi_{\{1\}}(s-3))\frac{\sigma_s^2}{\sigma_3^2}$ and $\widetilde{\pi}_p = \frac{\pi(\chi_{\{1\}}(s-3))}{p}\frac{\sigma_s^2}{\sigma_p^2}$.

3. For $s \in \{1,2,3\}$, $\gamma_s \sim \mathcal{N}(\underline{\mu_s}, \underline{\Sigma_s})$
4. The variance-covariance matrix, Ψ, is fixed and equal to its point MLE.

5.2.2. Normal-Wishart prior

This last characteristic of our *Minnesota prior* seems counterintuitive since it implies an extraordinarily precise knowledge of the variances of innovations. A simple, and more plausible alternative, is to assume that Ψ is Wishart distributed. Thus, we define the prior distributions $\gamma|\Psi \sim N(\underline{\mu}, \Psi \otimes \underline{\Sigma})$ and $\Psi \sim iW(\underline{\Psi}, \alpha)$ where $\gamma = (\gamma_1, \gamma_2, \gamma_3)'$, $E(\gamma) = \underline{\mu} = (\underline{\mu_1}, \underline{\mu_2}, \underline{\mu_3})'$, s_i^2 is the maximum likelihood estimate of the variance of the residuals for each of the n variables of the model, $\underline{\Sigma}$ is determined such that $var(\gamma_s) = \underline{\Sigma}_s$, $\forall s \in \{1,2,3\}$ as before and $\underline{\Psi}$ is a diagonal matrix with entries $\{(\alpha-n-1)s_1^2, (\alpha-n-1)s_2^2, (\alpha-n-1)s_3^2\}$.

Note that since $Var(\gamma) = \frac{1}{(\alpha-n-1)}\underline{\Psi}\otimes\underline{\Sigma}$, for $\forall s, j \in \{1,2,3\}$ is the case that $var(\gamma_s) = \left(\sigma_s^2/\sigma_j^2\right) var(\gamma_j)$. However, since we want $var(\gamma_s) = \underline{\Sigma}_s$ for $\forall s \in \{1,2,3\}$, $\underline{\Sigma}_s = \left(\sigma_s^2/\sigma_j^2\right)\underline{\Sigma}_j$ and thus, $\pi(0) = \pi(1)$. Also, the Kronecker structure implies that all priors, conditional on s_s^2, are equally informative. This last restriction imposes the uncomfortable restriction that information assumptions have to be symmetric across equations.

5.2.3. Hierarchial prior

Finally, we can relax the basic *Minnesota prior* assumption: forcing the prior mean for the parameter on the first own lag to one and the prior mean of the remaining parameters to zero. Using an *Hierarchial prior*, the prior mean of the parameters will follow a normal distribution with the above-remarked mean. Formally, $\gamma|\Psi, \mu \sim N(\mu, \Psi \otimes \underline{\Sigma})$, $\Psi \sim iW(\underline{\Psi}, \alpha)$ and $\mu \sim N(\underline{\mu}, \delta I)$.

5.3. Results

We estimate the nine different BVARs and use the output of the *Metropolis-Hastings* simulation, and we compute the marginal likelihoods as reported in table 5.1 in log terms[8]. This table summarizes then the evidence in favor of one model against the others.

We learn two main lessons from this table. The first is that, despite how well the cattle cycle model comes to match some aspects of the data, it is not even close to the performance of a BVAR with *Minnesota prior* and two lags. The log difference in favor of the BVAR is 43.46. How big is, intuitively, this difference? We will provide two measures. First, we will

[8] Each BVAR is called by the name of its prior and, in parenthesis, by the number of lags. For each BVAR, we computed the moments of the posterior and we assessed convergence using the same methods described in the previous section.

note that this difference means that the empirical evidence overcomes any prior ratio lower 7.4892e+018 in favor of the cattle cycle. Second, this difference is substantially bigger than 7, a bound for DNA testing in forensic science, often accepted by courts of law as evidence beyond reasonable doubt (Evett (1991)). This difference does not mean by itself, however, that we must disregard the model. This decision is a different task than its comparison with alternative models. We may still keep it as the best available alternative within the class of models with substantive economic content, or we can use it to perform welfare analysis or forecasting under changing policy regimes beyond the capabilities of BVARs.

Table 5.1: LogMarginal Likelihoods	
cattle cycle	−647.5281
Minnesota (1)	−615.4347
Minnesota (2)	−604.0657
Minnesota (3)	−618.9883
Wishart (1)	−791.4154
Wishart (2)	−779.1833
Wishart (3)	−808.9510
Hierar. (1)	−715.9167
Hierar. (2)	−732.1339
Hierar. (3)	−782.9960

Also, we should note that the *Minnesota prior* has the variance fixed at the MLE. Allowing the data to enter into the prior in this way gives a tremendous boost to any model and makes the model comparison unfair. If we restrict our comparison to the other six BVARs, the cattle cycle model performs quite well- a remarkable result in our view.

Our second lesson is that more flexible priors or longer lags are not always preferable. The reason is simple: richer models have many more hyperparameters and the Bayes Factor discriminates against these[9]. We see this "built-in" Ockam's razor as a final and attractive feature of the Bayes Factor: it embodies a strong preference for parsimonious modelling.

6. Conclusions

In this paper we have studied some properties of the Bayesian estimation and comparison of dynamic equilibrium models. Not only is this framework general, flexible, robust and simple to apply, but also its shown properties have a clear intuitive appeal. Asymptotically our convergence theorems show how the priors are irrelevant under appropriate technical conditions. On small samples, the prior is a way to achieve exact inference and, given the evidence in our paper, not inferior to the use of classical asymptotic approximations. Some parallel research (Fernández-Villaverde and Rubio-Ramírez (2001)) tries to further advance the Bayesian approach, solving the numerical problems associated with the evaluation of the likelihood of nonlinear representations of a dynamic equilibrium model that have so far limited its application.

[9]This discrimination can be easily seen in the Schwarz criterion (an asymptotic approximation of the log Bayes Factor) that explicitely penalizes the difference in the dimensionality of the parameter space.

7. Appendix

This appendix presents the omitted proofs from the text and offers some additional details about the computational procedures.

7.1. Proofs

[Lemma 1] Let $i \in M$. We can rewrite $f^T(Y^T|\theta,i)$ as:

$$f^T(Y^T|\theta,i) = f^T(Y^T|\widehat{\theta}_T\left(i,Y^T\right),i)\exp\left[\log f^T(Y^T|\theta,i) - \log f^T(Y^T|\widehat{\theta}_T\left(i,Y^T\right),i)\right] =$$
$$= f^T(Y^T|\widehat{\theta}_T\left(i,Y^T\right),i)\exp\left[\log f^T(Y^T|\widehat{\theta}^*_T\left(i\right),i) - \log f^T(Y^T||\widehat{\theta}_T\left(i\right),i)\right] \times$$
$$\exp\left[\log f^T(Y^T|\theta,i) - \log f^T(Y^T|\widehat{\theta}^*_T\left(i\right),i)\right]$$

Then:

$$\int_{\theta\in\Theta_i\setminus\mathcal{E}\left(\widehat{\theta}_T(i,Y^T),\delta_T(i)\right)} \pi(\theta|Y^T,i)d\theta =$$

$$f^T\left(Y^T,i\right)^{-1} f^T(Y^T|\widehat{\theta}_T\left(i,Y^T\right),i)\exp\left[\log f^T(Y^T|\widehat{\theta}^*_T\left(i\right),i) - \log f^T(Y^T|\widehat{\theta}_T\left(i,Y^T\right),i)\right]$$
$$\times \int_{\theta\in\Theta_i\setminus\mathcal{E}\left(\widehat{\theta}_T(i,Y^T),\delta_T(i)\right)} \pi\left(\theta,i\right)\exp\left[\log f^T(Y^T|\theta,i) - \log f^T(Y^T|\widehat{\theta}^*_T\left(i\right),i)\right]d\theta \quad (30)$$

but by (10)

$$\lim_{T\to\infty} P_0^T\left(\left[\exp\left[\log f^T(Y^T|\widehat{\theta}^*_T\left(i\right),i) - \log f^T(Y^T|\widehat{\theta}_T\left(i,Y^T\right),i)\right] \leq \exp\left[-k_T\left(\delta_T,i\right)T\right]\right]\right) = 1$$

which implies that $\exp\left[\log f^T(Y^T|\widehat{\theta}^*_T\left(i\right),i) - \log f^T(Y^T|\widehat{\theta}_T\left(i,Y^T\right),i)\right] = O_p\left(1\right)$ as $T\to\infty$ in P_0-probability.

With this last statement, we only need to check that

$$f^T\left(Y^T,i\right)^{-1}\int_{\theta\in\Theta_i\setminus\mathcal{E}\left(\widehat{\theta}_T(i,Y^T),\delta_T(i)\right)} \pi\left(\theta,i\right)\exp\left[\log f^T(Y^T|\theta,i) - \log f^T(Y^T|\widehat{\theta}^*_T\left(i\right),i)\right]d\theta \to 0$$

as $T\to\infty$ in P_0-probability. Since:

$$f^T\left(Y^T,i\right)^{-1}\int_{\theta\in\Theta_i\setminus\mathcal{E}\left(\widehat{\theta}_T(i,Y^T),\delta_T(i)\right)} \pi\left(\theta\right)\exp\left[\log f^T(Y^T|\theta,i) - \log f^T(Y^T|\widehat{\theta}^*_T\left(i\right),i)\right]d\theta$$

by (10), for T large enough,

$$f^T\left(Y^T,i\right)^{-1}\int_{\theta\in\Theta_i\setminus\mathcal{E}\left(\widehat{\theta}_T(i,Y^T),\delta_T(i)\right)} \pi\left(\theta\right)\exp\left[\log f^T(Y^T|\theta,i) - \log f^T(Y^T|\widehat{\theta}^*_T\left(i\right),i)\right]d\theta \leq$$

$$\leq \exp\left[-k_T T\right] f^T\left(Y^T,i\right)^{-1}\int_{\theta\in\Theta_i\setminus\mathcal{E}\left(\widehat{\theta}_T(i,Y^T),\delta_T(i)\right)} \pi\left(\theta,i\right)d\theta \leq \exp\left[-k_T\left(\delta_T,i\right)T\right] f^T\left(Y^T,i\right)^{-1}$$

but, (10) also implies that $\exp\left[-k_T\left(\delta_T,i\right)T\right]\to 0$ as $T\to\infty$ in P_0-probability and the results follow.

[Lemma 2] Assume Lemma 2 is not true. Then $\exists\,\gamma>0$ such that

$$\lim_{T\to\infty} P_0^T\left(\left|\widehat{\theta}_T^*\left(i\right)-\widehat{\theta}^*\left(i\right)\right|>\gamma\right)>0$$

and $\exists\,\eta>0$ such that

$$\lim_{T\to\infty} P_0^T\left(\mathcal{B}(\widehat{\theta}_T^*\left(i\right),\eta)\cap\mathcal{E}\left(\widehat{\theta}_T\left(i,Y^T\right),\delta_T(i)\right)=\emptyset\right)>0$$

since $\delta_T(i)\searrow 0$. But since $\mathcal{B}(\widehat{\theta}_T^*\left(i\right),\eta)\cap\mathcal{E}\left(\widehat{\theta}_T\left(i,Y^T\right),\delta_T(i)\right)=\emptyset\Longrightarrow\mathcal{B}(\widehat{\theta}_T^*\left(i\right),\eta)\subseteq\Theta_i\backslash\mathcal{E}\left(\widehat{\theta}_T\left(i,Y^T\right),\delta_T(i)\right)$

$$\int_{\Theta_i\backslash\mathcal{E}\left(\widehat{\theta}_T(i,Y^T),\delta_T(i)\right)}\pi\left(\theta,i\right)\exp\left[\log f^T(Y^T|\theta,i)-\log f^T(Y^T|\widehat{\theta}_T^*\left(i\right),i)\right]d\theta>$$

$$\int_{\mathcal{B}(\widehat{\theta}_T^*(i),\eta)}\pi\left(\theta,i\right)\exp\left[\log f^T(Y^T|\theta,i)-\log f^T(Y^T|\widehat{\theta}_T^*\left(i\right),i)\right]d\theta \tag{31}$$

but 12 implies that the right hand side is bigger than zero in P_0-probability. Then 6, 11 and (30) imply:

$$\int_{\theta\in\Theta_i\backslash\mathcal{E}\left(\widehat{\theta}_T(i,Y^T),\delta_T(i)\right)}\pi(\theta|Y^T,i)d\theta>0$$

as $T\nearrow\infty$ in P_0-probability, that contradicts Lemma 1.

[Theorem 2] First note that $B_{j,i|Y^T}=\frac{f^T(Y^T|j)}{f^T(Y^T|i)}$ and

$$f^T\left(Y^T|i\right)=\int_{\Theta_i}f^T\left(Y^T|\theta,i\right)\pi\left(\theta|i\right)d\theta=$$

$$=\int_{\mathcal{E}\left(\widehat{\theta}_T(i,Y^T),\delta_T(i)\right)}f^T\left(Y^T|\theta,i\right)\pi\left(\theta|i\right)d\theta+\int_{\Theta_i\backslash\mathcal{E}\left(\widehat{\theta}_T(i,Y^T),\delta_T(i)\right)}f^T\left(Y^T|\theta,i\right)\pi\left(\theta|i\right)d\theta$$

Since

$$\int_{\Theta_i\backslash\mathcal{E}\left(\widehat{\theta}_T(i,Y^T),\delta_T(i)\right)}f^T\left(Y^T|\theta,i\right)\pi\left(\theta|i\right)d\theta=f^T(Y^T,i)^{-1}\int_{\Theta_i\backslash\mathcal{E}\left(\widehat{\theta}_T(i,Y^T),\delta_T(i)\right)}\pi\left(Y^T|\theta,i\right)d\theta$$

Lemma 1 and 7 imply $\lim_{T\to\infty}P_0^T\left(\int_{\Theta_i\backslash\mathcal{E}\left(\widehat{\theta}_T(i,Y^T),\delta_T(i)\right)}f^T\left(Y^T|\theta,i\right)\pi\left(\theta|i\right)d\theta=0\right)=1$. We can write

$$\int_{\mathcal{E}\left(\widehat{\theta}_T(i,Y^T),\delta_T(i)\right)}f^T\left(Y^T|\theta,i\right)\pi\left(\theta|i\right)d\theta=\int_{\Theta_i}\chi\left(\theta\right)_{\left\{\mathcal{E}\left(\widehat{\theta}_T(i,Y^T),\delta_T(i)\right)\right\}}f^T\left(Y^T|\theta,i\right)\pi\left(\theta|i\right)d\theta$$

Let$\{Y_t\}_{t=1}^{\infty} \in \Omega \backslash \Omega_0$. Using Lemma 2, we can construct a sequence $\left\{\mathcal{E}\left(\widehat{\theta}_T\left(i, Y^T\right), \delta_T(i)\right)\right\}_{i=1}^{\infty}$ such that $\widehat{\theta}_T^*(i) \in \mathcal{E}\left(\widehat{\theta}_T\left(i, Y^T\right), \delta_T(i)\right) \ \forall\ T$ and $\forall\, i \in M$ that makes

$$\chi(\theta)_{\left\{\mathcal{E}\left(\widehat{\theta}_T(i,Y^T),\delta_T(i)\right)\right\}} f^T\left(Y^T|\theta, i\right)\pi(\theta|i) - \chi(\theta)_{\left\{\widehat{\theta}_T^*(i)\right\}} f^T\left(Y^T|\theta, i\right)\pi(\theta|i) \to 0$$

pointwise as $T \to \infty$. At the same time,

$$\chi(\theta)_{\left\{\mathcal{E}\left(\widehat{\theta}_T(i,Y^T),\delta_T(i)\right)\right\}} f^T\left(Y^T|\theta, i\right)\pi(\theta|i) - \chi(\theta)_{\left\{\widehat{\theta}_T^*(i)\right\}} f^T\left(Y^T|\theta, i\right)\pi(\theta|i) =$$

$$= \chi(\theta)_{\left\{\mathcal{E}\left(\widehat{\theta}_T(i,Y^T),\delta_T(i)\right)\right\}} f^T\left(Y^T|\theta, i\right)\pi(\theta|i)$$

a.s. in Lebesgue measure. Then

$$\chi(\theta)_{\left\{\mathcal{E}\left(\widehat{\theta}_T(i,Y^T),\delta_T(i)\right)\right\}} f^T\left(Y^T|\theta, i\right)\pi(\theta|i) - \chi(\theta)_{\left\{\widehat{\theta}_T^*(i)\right\}} f^T\left(Y^T|\theta, i\right)\pi(\theta|i) \leq$$

$$\leq \chi(\theta)_{\left\{\mathcal{E}\left(\widehat{\theta}_T(i,Y^T),\delta_T(i)\right)\right\}} f^T\left(Y^T|\widehat{\theta}_T\left(i, Y^T\right), i\right)\pi(\theta|i) \leq \sup_T f^T\left(Y^T|\widehat{\theta}_T\left(i, Y^T\right), i\right)\pi(\theta|i)$$

also a.s. in Lebesgue measure. Using Condition 7:

$$\int_{\Theta_i} \pi(\theta|i)\, d\theta = \sup_T f^T\left(Y^T|\widehat{\theta}_T\left(i, Y^T\right), i\right) < \infty$$

Then, we can apply the Dominated Converge Theorem to conclude that

$$\lim_{T\to\infty} P_{0T}\left(f^T\left(Y^T|i\right) - f^T\left(Y^T|\widehat{\theta}_T^*(i), i\right)\pi\left(\widehat{\theta}_T^*(i)\,|i\right) = 0\right) = 1$$

and find:

$$\lim_{T\to\infty} P_{0T}\left(\frac{f^T\left(Y^T|i\right)}{f^T\left(Y^T|j\right)} - \frac{f^T\left(Y^T|\widehat{\theta}_T^*(i), i\right)\pi\left(\widehat{\theta}_T^*(i)\,|i\right)}{f^T\left(Y^T|\widehat{\theta}_T^*(j), j\right)\pi\left(\widehat{\theta}_T^*(j)\,|j\right)} = 0\right) = 1 \tag{32}$$

Now, to prove that $\lim_{T\to\infty} P_{0T}\left(\frac{f^T\left(Y^T|\widehat{\theta}_T^*(i),i\right)\pi\left(\widehat{\theta}_T^*(i)|i\right)}{f^T\left(Y^T|\widehat{\theta}_T^*(j),j\right)\pi\left(\widehat{\theta}_T^*(j)|j\right)} = 0\right) = 1$ and since

$$\begin{aligned}
&\left[\begin{array}{c} \frac{1}{T}\log f^T\left(Y^T|\widehat{\theta}_T^*(i), i\right)\pi\left(\widehat{\theta}_T^*(i)\,|i\right) - \\ -\frac{1}{T}\log f^T\left(Y^T|\widehat{\theta}_T^*(j), j\right)\pi\left(\widehat{\theta}_T^*(j)\,|j\right) = -\infty \end{array}\right] \\
\subseteq\ &\left[\frac{f^T\left(Y^T|\widehat{\theta}_T^*(i), i\right)\pi\left(\widehat{\theta}_T^*(i)\,|i\right)}{f^T\left(Y^T|\widehat{\theta}_T^*(j), j\right)\pi\left(\widehat{\theta}_T^*(j)\,|j\right)} = 0\right]
\end{aligned}$$

we only need to show

$$\lim_{T\to\infty} P_{0T}\left(\begin{array}{c} \frac{1}{T}\log f^T\left(Y^T|\widehat{\theta}_T^*(i), i\right)\pi\left(\widehat{\theta}_T^*(i)\,|i\right) - \\ -\frac{1}{T}\log f^T\left(Y^T|\widehat{\theta}_T^*(j), j\right)\pi\left(\widehat{\theta}_T^*(j)\,|j\right) = -\infty \end{array}\right) = 1 \tag{33}$$

Now, using the factorization $\log f^T\left(Y^T|\widehat{\theta}_T^*(i),i\right)=\sum_{t=1}^T\log f_t\left(Y^t|\widehat{\theta}_T^*(i),i\right)$ we can rewrite (33) as

$$\lim_{T\to\infty}P_{0T}\left(\begin{array}{c}\frac{1}{T}\sum_{t=1}^T\log f_t\left(Y^t|\widehat{\theta}_T^*(i),i\right)+\log\pi\left(\widehat{\theta}_T^*(i)\,|i\right)- \\ -\frac{1}{T}\sum_{t=1}^T\log f^T\left(Y^T|\widehat{\theta}_T^*(j),j\right)-\log\pi\left(\widehat{\theta}_T^*(j)\,|j\right)=-\infty\end{array}\right)=1 \qquad (34)$$

Conditions (10)-(12) allow us to use an argument similar to Wald (1949) to prove (33) and use (32) and (34) to finish the proof.

7.2. Some Computational Details

The cattle cycle model was computed using Vaughan's eigenvector method to solve the associated Algebraic Riccati equation to the representative farmer problem. This method exploits the linear restrictions that stability imposes among multipliers and the state vector, resulting in an efficient algorithm feasible for constant revaluation.

With respect to the *Metropolis-Hasting algorithm*, its success depends on the fulfillment of a number of technical conditions. In practice, however, the main issue is to assess the convergence of the simulated chain to the ergodic density. In addition to the more formal tests of convergence discussed in the text, it is key to adjust the parameters of the transition density (in the case of the random walk, the variance of the innovation term) to get an appropriate acceptance rate (the percentage of times when the chain changes position). If the acceptance rate is very small, the chain will not visit a set large enough in any reasonable number of iterations. If the acceptance rate is very high, the chain will not stay enough time in high probability regions. Gelman, Roberts and Gilks (1996) suggest that a 20% acceptance rate tends to give the best performance. We found that an acceptance rate of around 40% outperformed different alternatives.

The code for the evaluation of all the likelihoods and all the simulations was written in Matlab 5.3 and compiled when feasible with MCC 1.2. Theoretically, it should be portable to any machine equipped with the Matlab interpreter. The code was run on a Sun Workstation Ultra-2 with SunOS 5.6.

All the programs and their corresponding documentation, the simulation output (including additional empirical distributions, time series graphs, trial runs and additional convergence assessments) are available upon request from the corresponding author.

References

[1] ANDERSON, E.W., HANSEN, L.P., McGRATTAN, E.R. and SARGENT, T.J. (1996). "On the Mechanics of Forming and Estimating Dynamic Linear Economies", in H.M. Amman ,*et al.* (eds.) *Handbook of Computational Economics*, Elsevier.

[2] ANDREW, D.W.K. (1988). "Laws of Large Numbers for Dependent Non-identically Distributed Random Variables", *Econometric Theory,* **4**, 458-467.

[3] BUREAU OF THE CENSUS (1975), *Historial Statistics of the United States, Colonial Times to 1970.* U.S. Department of Commerce.

[4] BUREAU OF THE CENSUS (1989), *Agricultural Statistics*. U.S. Department of Commerce.

[5] CHEN, C.F. (1985). "On Asymptotic Normality of Limiting Density Functions with Bayesian Implications", *Journal of the Royal Statistical Society, Series B*, **47**, 540-546.

[6] CSISZAR, I. (1991). "Why Least Squares and Maximum Entropy? an Axiomatic Approach to Inference for Linear Inverse Problems", *The Annals of Statistics*, **19**, 2032-2066.

[7] DeJONG, D.N., INGRAM, B.F. and WHITEMAN, C.H. (2000). "A Bayesian Approach to Dynamic Macroeconomics", *Journal of Econometrics,* **98**, 203-223.

[8] DIEBOLD, F.X., OHANIAN, L.E. and BERKOWITZ, J. (1998). "Dynamic Equilibrium Economies: A Framework for Comparing Model and Data", *Review of Economic Studies* **65**, 433-51.

[9] EVETT, I.W. (1991). "Implementing Bayesian Methods in Forensic Science", paper presented at the Fourth Valencia International Meeting on Bayesian Statistics.

[10] FERNÁNDEZ-VILLAVERDE, J. and RUBIO-RAMÍREZ, J.F. (2000). "Estimating Nonlinear Dynamic Equilibrium Economies: A Likelihood Approach", mimeo, University of Minnesota.

[11] GELFAND, A.E. and DEY, D.K. (1994). "Bayesian Model Choice: Asymptotics and Exact Calculations", *Journal of the Royal Statistical Society serie B*, **56**, 501-514.

[12] GELMAN, A., ROBERTS, G.O. and GILKS, W.R. (1996). "Efficiente Metropolis Jumping Rules", in J.O. Berger *et al.* (eds.) *Bayesian Statistics 5*, Oxford University Press.

[13] GEWEKE, J. (1992). "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments", in J.M. Bernardo *et al.* (eds), *Bayesian Statistics 4*, pp. 169-193. Oxford University Press.

[14] GEWEKE, J. (1998). "Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication", *Staff Report* 249, Federal Reserve Bak of Minneapolis.

[15] GEWEKE, J. (1999). "Simulation methods for Model Criticism and Robustness Analysis" in Bernardo, J.M. *et al.* (eds.), *Bayesian Statistics 6*. Oxford University Press.

[16] GEWEKE, J., KEANE, M. and RUNKLE, D. (1997). "Statistical Inference in the Multinomial Multiperiod Probit Model", *Journal of Econometrics*, **80**, 125-165.

[17] GOURIEROUX, C. and MONFORT, A. (1998). "Testing Non-Nested Hypotheses" in R. F. Engle and D. L. McFadden (eds.), *Handbook of Econometrics, volume 4* , Elsevier.

[18] HANNAN, E.J. (1970). *Multiple Time Series*, John Wiley and Sons.

[19] HANSEN, L.P., McGRATTAN, E.R. and SARGENT, T.J. (1994). "Mechanics of Forming and Estimating Dynamic Linear Economies", *Staff Report* 182, Federal Reserve Bank of Minneapolis.

[20] JACQUIER, E., POLSON, N.G. and ROSSI, P. (1994). "Bayesian Analysis of Stochastic Volatility Models", *Journal of Business and Economic Statistics,* **12**, 371-89.

[21] JEFFREYS, H. (1961). *The Theory of Probability* (3rd ed). Oxford University Press.

[22] KADIYALA, K.R. and KARLSSON, S. (1997). "Numerical Methods for Estimation and Inference in Bayesian VAR-models", *Journal of Applied Econometrics*, **12**, 99-132.

[23] KASS, R.E. and RAFTERY, A.E. (1995). "Bayes Factors", *Journal of the American Statistical Association*, **90**, 773-795.

[24] KIM, JAE-YONG (1998). "Large Sample Properties of Posterior Densities, Bayesian Information Criterion and the Likelihood Principle in Nonstationary Time Series Models", *Econometrica,* **66**, 359-380.

[25] KITAMURA, Y. (1998). "Comparing Misspecified Dynamic Econometric Models Using Nonparametric Likelihood", mimeo, University of Wisconsin-Madison.

[26] LANDON-LANE, J. (1999). "Bayesian Comparison of Dynamic Macroeconomic Models", Ph. D. Thesis, Univeristy of Minnesota.

[27] LEEPER, E.M. and SIMS, C.A. (1994). "Toward a Modern Macroeconomic Model Usable for Policy Analysis", *NBER Macroeconomics Annual,* **1994**, 81-118.

[28] LITTERMAN, R.B. (1980), "A Bayesian Procedure for Forecasting with Vector Autoregressions", mimeo, University of Minnesota.

[29] LITTERMAN, R.B. (1986). "Forecasting with Bayesian vector autoregressions-five years of experience", Journal of Bussines & Economic Statistics, **4**, 1-4.

[30] LUCAS, R.E. (1980). "Methods and Problems in Business Cycle Theory", *Journal of Money, Credit and Banking*, **12**, 696-715.

[31] MENGERSEN, K.L., ROBERT, C.P. and GUIHENNEUC-JOUYAUX, C. (1999). "MCMC Convergence Diagnostics: a 'reviewww"', in J. Berger *et al.* (eds.), *Bayesian Statistics 6*, Oxford Sciences Publications.

[32] OTROK, C. (2001). "On Measuring the Welfare Cost of Business Cycles", Journal of Monetary Economics, **47**, 61-92.

[33] PHILLIPS, P.C.B. (1996). "Econometric Model Determination". *Econometrica*, **64**, 763-812.

[34] PHILLIPS, P.C.B. and PLOBERGER, W. (1996). "An Asymptotic Theory of Bayesian Inference for Time Series", *Econometrica*, **64**, 381-412.

[35] RISSANEN, J. (1986). "Stochastic Complexity and Modeling", *The Annales of Statistics* **14**, 1080-1100.

[36] ROBERT, C.P. and CASELLA, G. (1999), *Monte Carlo Statistical Methods*. Springer-Verlag.

[37] ROSEN, S., K.M. MURPHY and SCHEINKMAN, J.A. (1994). "cattle cycle", *Journal of Political Economy*, **102**, 468-492.

[38] SIMS, C.A. (1980). "Macroeconomics and Reality", *Econometrica,* **48**, 1-48.

[39] SIMS, C.A. and UHLIG, H. (1991). "Understanding Unit Rooters: A Helicopter Tour", *Econometrica*, **59**, 1591-1599.

[40] SIMS, C.A. and ZHA, T. (1998). "Bayesian Methods for Dynamic Multivariate Models", *International Economic Review*, **39**, 949-968.

[41] SCHORFHEIDE, F. (1999). "A Unified Econometric Framework for the Evaluation of DGSE Models", mimeo, University of Pennsylvania.

[42] SHORE, J.E. and JOHNSON, R.W (1980). "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy", *IEEE Transactions on Information Theory*, **IT-26**, 26-37.

[43] VUONG, Q.H. (1989). "Likelihood Ratio Test for Model Selection and Non-Nested Hypotheses", *Econometrica,* **57**, 307-333.

[44] WALD, A. (1949). "Note on the Consistency of the Maximum Likelihood Estimate", *Annales of Mathematical Statistics,* **20**, 595-601.

Figure 1: Model Estimation and Comparison

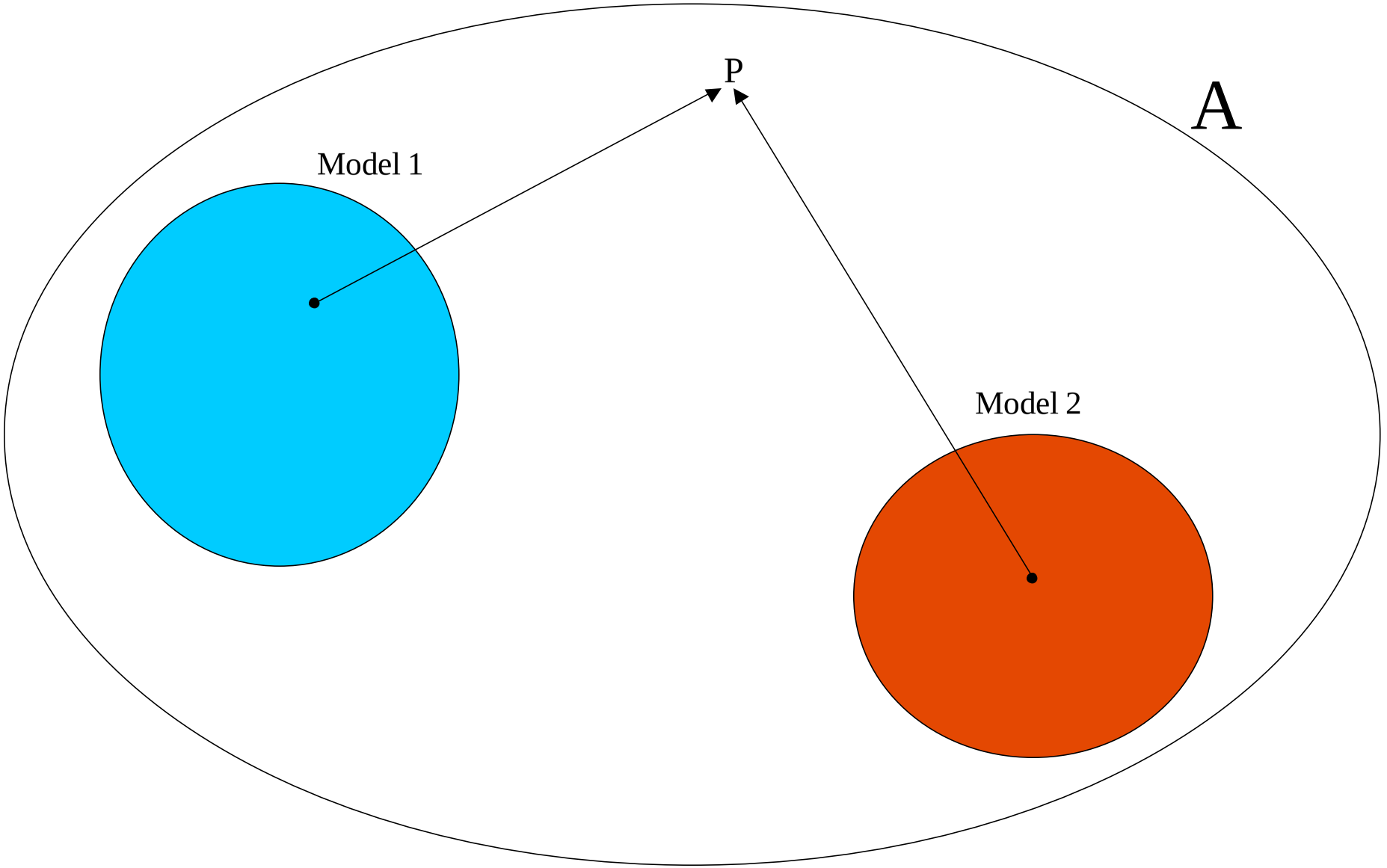
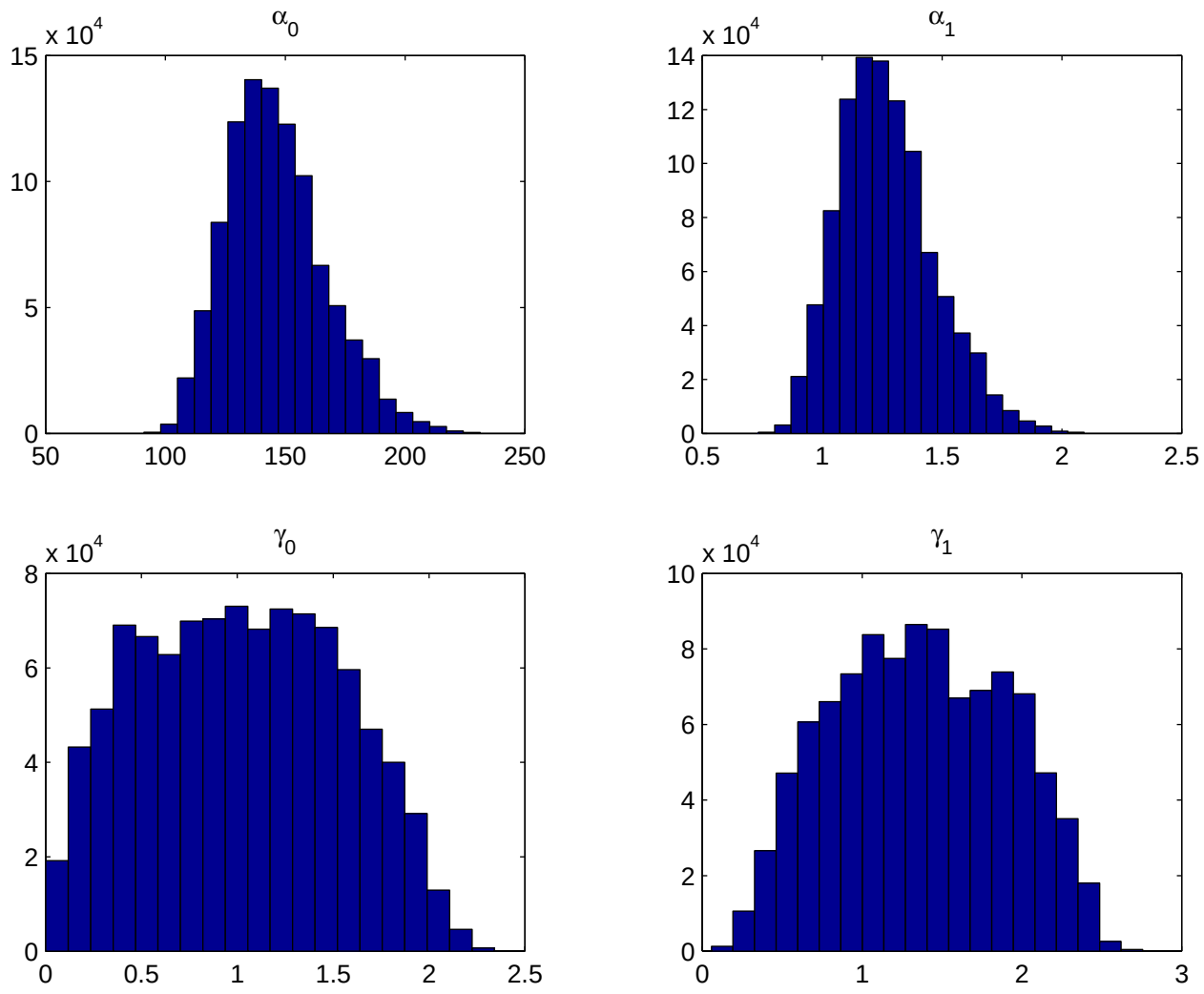
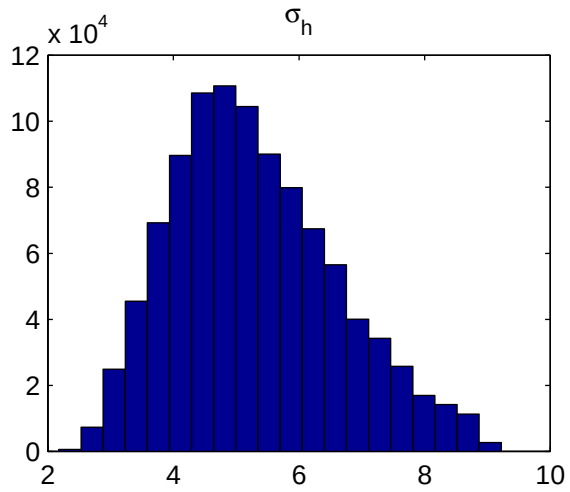
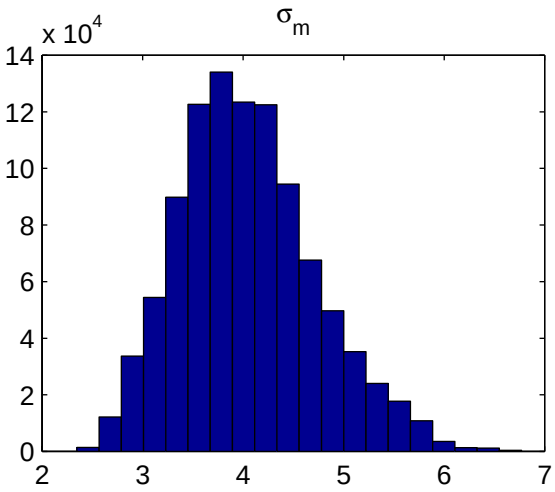
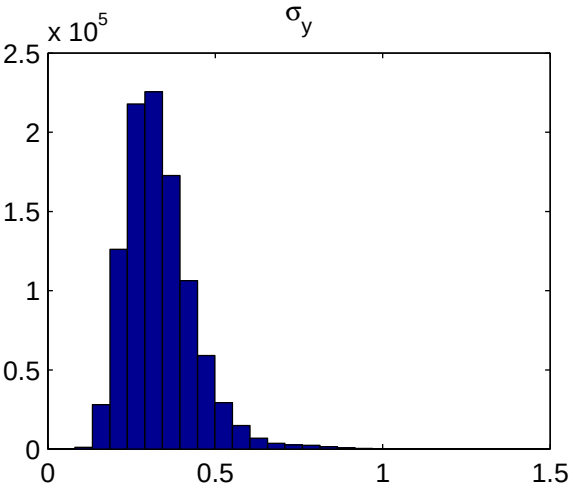


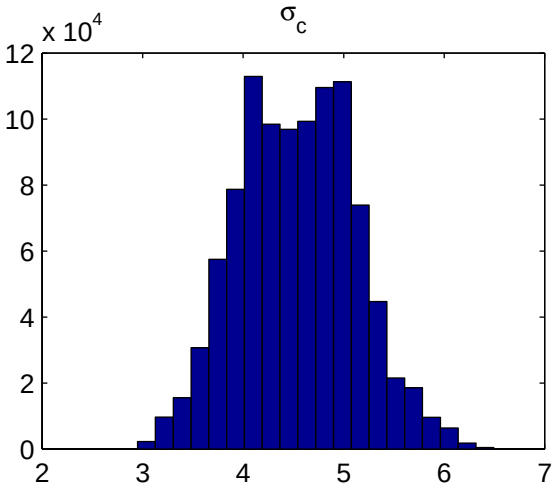
Figure 2 : Empirical Distribution of the Posterior

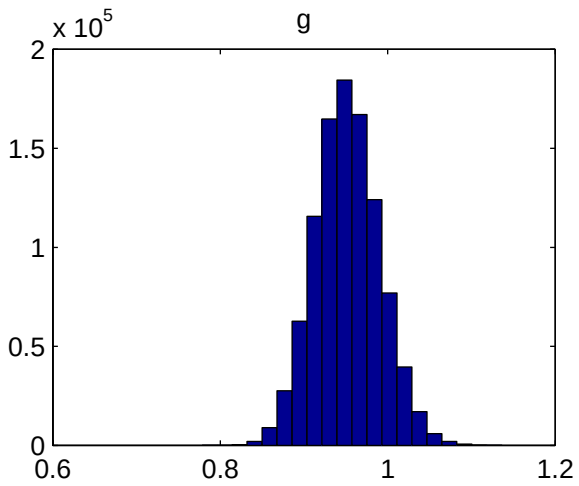


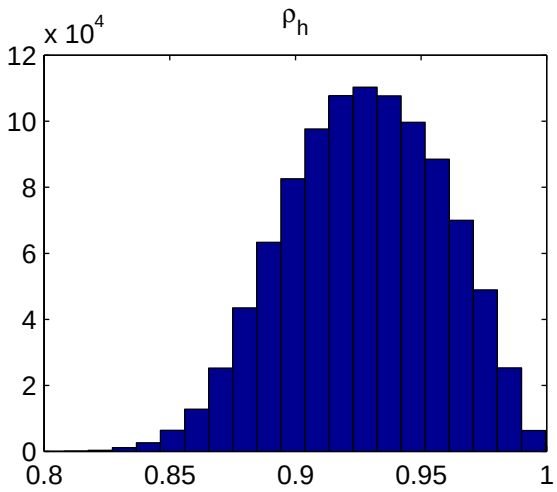

σh
x 10^4
12
10
8
6
4
2
0
2
4
6
8
10

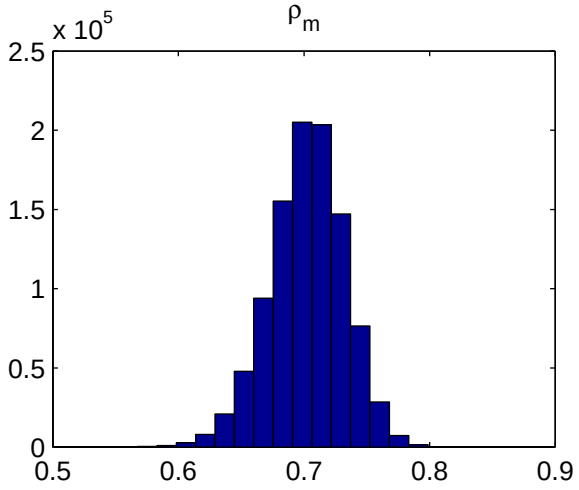

σm
x 10^4
14
12
10
8
6
4
2
0
2
3
4
5
6
7


σy
x 10^5
2.5
2
1.5
1
0.5
0
0
0.5
1
1.5


σc
x 10^4
12
10
8
6
4
2
0
2
3
4
5
6
7


g
x 10^5
2
1.5
1
0.5
0
0.6
0.8
1
1.2


ρh
x 10^4
12
10
8
6
4
2
0
0.8
0.85
0.9
0.95
1


ρm
x 10^5
2.5
2
1.5
1
0.5
0
0.5
0.6
0.7
0.8
0.9