FEDERAL
RESERVE
BANK
of ATLANTA

Comparing Dynamic Equilibrium Models to Data:
A Bayesian Approach

Jesus Fernández-Villaverde and Juan F. Rubio-Ramírez

Working Paper 2001-23a
February 2003

Working Paper Series

Federal Reserve Bank of Atlanta
Working Paper 2001-23a
February 2003

Comparing Dynamic Equilibrium Models to Data: A Bayesian Approach

Jesus Fernández-Villaverde, University of Pennsylvania
Juan F. Rubio-Ramírez, Federal Reserve Bank of Atlanta

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 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.

JEL classification: C11, C15, C51, C52

Key words: Bayesian inference, asymptotics, cattle cycle

The authors thank A. Atkeson, J. Geweke, W. McCausland, E. McGrattan, L. Ohanian, T. Sargent, C. Sims, H. Uhlig, and participants at several seminars for useful comments. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Any remaining errors are the authors' responsibility.

Please address questions regarding content to Jesus Fernández-Villaverde, assistant professor, University of Pennsylvania, Department of Economics, 160 McNeil Building, 3718 Locust Walk, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6297, 215-898-1504, 215-573-2057 (fax), jesusfv@econ.upenn.edu, or Juan F. Rubio-Ramírez, research economist and assistant policy adviser, Federal Reserve Bank of Atlanta, Research Department, 1000 Peachtree Street, N.E., Atlanta, Georgia 30309-4470, 404-498-8057, 404-498-8956 (fax), juan.rubio@atl.frb.org

Comparing Dynamic Equilibrium Models to Data: A Bayesian Approach

Over the last two decades, dynamic equilibrium models have become a standard instrument to study a variety of issues in economics, from Business Cycles and Growth Theory to Public Finance, International Trade, Industrial Organization and Labor Economics. Since a dynamic equilibrium economy is an artificial construction, these models will *always* be false. This fact presents two main challenges for econometric practice: first, how to select appropriate values for the "deep" parameters of the model (i.e. those describing technology, preferences, and so on) and, second, how to compare two or more misspecified models that might be nonnested.

Bayesian econometrics addresses these two challenges by suggesting both a procedure to select parameters and a criterion for model comparison. Parameter choice is undertaken by the computation of posteriors while model comparison is performed through the use of posterior odds ratios. The bayesian approach is, of course, well known. Inference about parameter values follows directly from Bayes' Theorem while model comparison through posterior odds was introduced by Jeffreys (1961) (in the form of hypothesis testing) and recently revived by Gelfand and Dey (1994), Geweke (1998), Landon-Lane (1999), and Schorfheide (1999), among others.

Our work follows this tradition. In particular, this paper makes two 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 measure 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 compared with a set of Bayesian Vector Autoregressions.[1]

These contributions are important for two reasons. Our first point helps to remove one of the main criticism of bayesian methods, the possible impact of priors in our reading of the data, since they imply that, as the sample grows, the priors disappear. The convergence of the posterior odds ratio toward the Kullback-Leibler preferred model is attractive because there is a complete axiomatic foundation that justifies why this measure is precisely the criterion a rational agent should use to choose between models. Details of this axiomatic foundation are presented in Shore and Johnson (1980) and Csiszar (1990). Our second point shows how, in real life applications, a bayesian approach delivers a very strong performance when applied to dynamic equilibrium models.

There are several reasons to justify our "Bayes choice." First, Bayesian inference builds on the insight that models are false and is ready to deal with this issue in a natural way. 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 nonstationarity 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 advent of Markov chain

[1] An additional contribution-how to evaluate the likelihood of nonlinear representations of dynamic equilibrium models using Sequential Monte Carlo filtering-is described in detail in a companion paper (Fernández-Villaverde and Rubio-Ramírez (2002)).

Monte Carlo techniques has removed the need for suitable expressions for likelihoods and priors (in fact, simulation methods like the Sequential Monte Carlo even allow dealing with models without closed-form likelihood functions). Sixth, it is computationally straightforward to conduct a robustness analysis of the results.

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 Gourieroux and Monfort (1998)). In particular, Vuong (1989) and Kitamura (1998) have developed tests for nonnested and misspecified models, and Aguirre-Torres and Gallant (1999) have proposed the use of the EMM for such a purpose. We see our contributions as very similar in spirit to these papers.

In the Bayesian literature, DeJong, Ingram and Whiteman (2000) pioneered the Bayesian estimation of Real Business Cycles models using importance sampling. Otrok (2001) first applied the Metropolis-Hastings algorithm to the estimation problem. 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, the use of State-Space representations allows us to deal with high dimensional vectors and to study a general class of (possibly nonlinear) models. Third, we develop the asymptotic properties of the procedure. Fourth, 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 measure will approach zero as the sample size goes to infinity. The novelty of these two results is that we do not assume that the models are well-specified and/or nested as the existing literature requires (see for instance Chen (1985) or Gelfand and Dey (1994)). After presenting the notation, we explain the Bayesian model comparison, and we proved the two theorems mentioned above. Finally we discuss the numerical implementation of the bayesian approach.

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 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 that it allows for a sufficiently rich behavior in Y. For convenience, we choose $\Omega = \Re^{m\times\infty}$. 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 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 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 $\theta \subseteq \Theta_i \subseteq \Re^{k_i}$ $\forall i \in M$. We assume that $f^n(\cdot|\theta, i)$ is measurable with respect to P_0^n $\forall n, i$. Each family of parameterized probability densities comprises different candidates to account for the observations while the prior probability densities embody the previous knowledge about the parameter values. We 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 models derived directly from economic theory (as the stochastic neoclassical growth model) and/or pure statistical models (as an unrestricted Vector Autoregression).

The function $f^T\left(y^T|\theta, i\right)$ is called the pseudo-likelihood function of the data. Define the pseudo-maximum likelihood point estimate (PMLE) as $\widehat{\theta}_T\left(i, y^T\right) \equiv \arg\max_{\theta\in\Theta_i} \log f^T\left(y^T|\theta, i\right)$. Note that we do not assume 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 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 $\pi\left(\theta|y^T,i\right) \propto f^T\left(y^T|\theta,i\right)\pi\left(\theta|i\right)$ and its marginal likelihood as:

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

The marginal likelihood is the probability that the model assigns to having observed the data. This interpretation relates the marginal likelihood with the pseudo-likelihood evaluated at the 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, $\pi\left(\theta|i\right)$ 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).

Now we define the Kullback-Leibler measure as:

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

The intuition of this closeness concept is simple: it evaluates the average surprise with respect to the true measure that the researcher using $f^T\left(\cdot|\theta,i\right)$ suffers if he suddenly learns that the true density is $p_0^T\left(\cdot\right)$. As mentioned in the introduction, the Kullback-Leibler measure is particularly attractive because of its sound foundations on decision-choice theory. Because of space constraints we omit a thorough discussion of its complete axiomatic foundations and

we refer the interested reader to Shore and Johnson (1980) and Csiszar (1990).

We define the pseudotrue value as $\theta_T^*(i) \equiv \arg\min_{\theta\in\Theta_i} K\left(f^T(\cdot|\theta,i);p_0^T(\cdot)\right)$, i.e. the parameter values that select the member of the parametric family that is "closest" to P_{0T} in the the Kullback-Leibler sense. Also we define $\theta^*(i) = \lim_{T\to\infty}\theta_T^*(i)$. Finally, we assume that $\theta_T^*(i)$ and the PMLE are unique. This assumption is the fundamental identification condition in our context of misspecified models.

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 reflects the previous knowledge of the researcher about the different models being considered.

Model comparison is an application of 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

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

which 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 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.[2]

In the same way the marginal likelihood is related to the likelihood value at the PMLE, the Bayes Factor is related to the Likelihood Ratio (LR). The Bayes Factor enjoys three 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 a way to deal with this problem, see Vuong (1989)).

2.3. Convergence Theorems

In this subsection we prove two new theorems. First, we show that the posterior distribution of the parameters collapses to their pseudotrue values. Second, we demonstrate that the Bayes Factor of any model over the best model under the Kullback-Leibler measure approaches zero. With these two theorems we build on the recent literature on the asymptotic properties of Bayesian inference. Examples include Phillips and Ploberger (1996), Phillips (1996) and Kim (1998) among others.

The contribution embodied in the theorems is important for several reasons. First, we

[2]Note that model comparison is a 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.

assure that, even when the models are misspecified, the priors are irrelevant as the sample size grows. Second, if we want to choose our best model to satisfy the Kullback-Leibler criterion (and, as we argued before, there are axiomatic systems that tell us that this is the right thing to do), our results indicate that the Bayes factor is a consistent selection device even when the models are missespecified and/or nonnested.

The structure of this subsection is as follows. First, after stating some technical conditions, 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 lemma 3. This lemma characterizes the asymptotic behavior of the marginal likelihood, and it is an intermediate step to prove the second of the theorems: the Bayes Factor of any other model over the model closest to P_0^T under the Kullback-Leibler measure asymptotically approaches zero.

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

Definition 1. *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:*

$$\mathcal{E}(a, \delta_j(i)) \equiv \left\{ \theta \in \Theta_i : \frac{|a_1 - \theta_1|^2}{\delta_{j1}^2(i)} + ... + \frac{|a_{k_i} - \theta_{k_i}|^2}{\delta_{jk_i}^2(i)} < 1 \right\} \tag{5}$$

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

The idea behind this system is to look at the parameter values closely enough to some k_i-dimensional point a, making the values of $\delta_j(i)$ smaller as $T \nearrow \infty$. In general, this

point a will be the PMLE $\widehat{\theta}_T\left(i, y^T\right)$. This system allows for different rates of accumulation along different dimensions. Kim (1998) shows that this feature of the "shrinking neighborhood system" makes the theory relevant to work with many nonstationary processes that otherwise could not be analyzed.

Now we introduce some conditions that we would need to prove lemmas 1 and 2.

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

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

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

$$\lim_{T\to\infty} P_0^T\left(T^{-1}\log f^T(Y^T|\theta_T^*(i), i) > -\infty\right) = 1 \tag{7}$$

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

$$\lim_{T\to\infty} P_0^T\left(\left|f^T\left(Y^T|i\right)\right| = 0\right) = 0 \tag{8}$$

$$\lim_{T\to\infty} P_0^T\left(\left|\frac{f^T\left(Y^T|\widehat{\theta}_T\left(i, Y^T\right), i\right)}{f^T\left(Y^T|i\right)}\right| = 0\right) = 0 \tag{9}$$

Condition 4. *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*

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

Conditions 1 and 2 bound the log likelihood. These conditions only mean that the likelihood is informative, i.e., that there is a chance that we can learn from the data. Condition 3 precludes priors without support of the pseudo-true value. This condition implies that we can write the posterior distribution as the ratio of the integral of the prior times the likelihood over the marginal likelihood. This condition is not strictly needed. We could prove the following lemmas using much weaker conditions than 3. However, for clarity of exposition and since the proof with weaker conditions does not provide further insight into the logic of the reasoning, we prefer to use this slightly stronger condition. Condition 4 is an adaptation for the case of misspecified models of condition D2 (ii) in Kim (1998). It requires that the difference between the log likelihood evaluated at the pseudo-true value and the best of the candidates outside of the shrinking neighborhood $\mathcal{E}\left(\widehat{\theta}_T\left(i, Y^T\right), \delta_T(i)\right)$ goes to infinity with probability one when T goes to infinity. In other words, we require that the tails of the log likelihood function decrease sufficiently fast as more information arrives. This condition plays an important role in the proof of lemma 1 since it allows us to bound the posterior distribution on $\Theta_i \backslash \mathcal{E}\left(\widehat{\theta}_T\left(i, Y^T\right), \delta_T(i)\right)$ by $\exp\left[-k_T\left(\delta_T, i\right) T\right] \downarrow 0$. As in the case of condition 3, we could substitute it for a weaker one by paying the cost of a more cumbersome proof. In addition, as Kim (1998) remarks, if we assume we know the underlying true process (although in general we do not do so), this condition is, in some cases, verifiable.

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

Lemma 1. *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^T*-probability* $\forall\, i \in M$.

Proof. See Appendix. ■

It is important to emphasize that with lemma 1 we have shown a result that is often

directly imposed as a condition in the literature (see, for instance, condition C6 in Phillips and Ploberger (1996)).

In order to prove lemma 2, we need an additional condition:

Condition 5. $\forall \eta > 0$, $\exists\ \epsilon > 0$ *such that*

$$\lim_{T\to\infty} P_0^T \left(\int_{\mathcal{B}(\theta_T^*(i),\eta)} \pi(\theta, i) \exp\left[\log f^T(Y^T|\theta, i) - \log f^T(Y^T|\theta_T^*(i), i)\right] d\theta > \epsilon \right) = 1 \qquad (11)$$

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

Condition 5 implies that the posterior does not vanish around the pseudo-true value. If this is true, and since lemma 1 implies that the posterior concentrates around the PMLE when $T \to \infty$, it should be the case that $\theta_T^*(i) \in \mathcal{E}\left(\widehat{\theta}_T\left(i, Y^T\right), \delta_T(i)\right)$ for T big enough in P_0^T-probability. We formalize this argument in the next lemma:

Lemma 2. *Under Conditions 1-4,* $\widehat{\theta}_T\left(i, Y^T\right) - \theta^*(i) \to 0$ *as* $T \to \infty$ *in* P_0^T*-probability.*

Proof. See Appendix. ■

With these two lemmas, it can be shown that:

Theorem 1. *The posterior distribution of the parameters collapses to the pseudo-true value of the parameter.*

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

Now, let us now prove the main theorem of this paper. Before proving the theorem, we need a previous step in the form of lemma 3. This result extends the lemma 2.1. in Chen (1985) when models are misspecified. In order to prove this lemma, we need to make the following definitions and conditions:

Condition 6. $\forall\, i \in M$, $\log f^T(y^T|\theta, i)$ *is twice differentiable.*

Definition 2. *Define:*

$$\begin{aligned} L'_T\left(y^T|\theta,i\right) &\equiv \frac{\partial \log f^T(y^T|\theta,i)}{\partial\theta} \\ L''_T\left(y^T|\theta,i\right) &\equiv \frac{\partial^2 \log f^T(y^T|\theta,i)}{\partial\theta\partial\theta'} \\ \Sigma_T\left(y^T|i\right) &\equiv \left[-L''_T\left(y^T|\widehat{\theta}_T\left(i,y^T\right),i\right)\right]^{-1} \end{aligned}$$

where $\Sigma_T\left(y^T|i\right)$ *is the Cramér-Rao bound.*

Condition 7. $\forall\, i \in M$, *for any* $\varepsilon_T \downarrow 0$ *there exist* $\delta_T(i) \downarrow 0$ *such that*

$$\lim_{T\to\infty} P_0^T\left(I - A_T(\varepsilon_T) \leq -L''_T\left(y^T|\theta,i\right)\Sigma_T\left(y^T|i\right) \leq I + A_T(\varepsilon_T)\right) = 1$$

$\forall\theta \in \mathcal{E}\left(\widehat{\theta}_T\left(i,Y^T\right),\delta_T(i)\right)$, *where* $A_T(\varepsilon_T)$ *is a semidefinite positive symmetric whose largest eigenvalue goes to zero as* $\varepsilon_T \downarrow 0$.

Condition 8. $\forall\, i \in M$, *for any* $\delta_T(i)$ *such that 7 holds,* $\lambda_{\min}\left(\Sigma_T\left(y^T|i\right)^{-1}\right)\delta_T(i) \to \infty$ *as* $T \to \infty$ *in* P_0^T*-probability.*

Condition 9. $\forall\, i \in M$, $0 < \pi\left(\theta^*(i)\,|i\right) < \infty$ *and* $\pi(\theta|i)$ *are continuous.*

Condition 6 is a common assumption in the literature. Condition 7 is also quite common and imposes a smoothing condition on the second derivative of the log likelihood. Condition 8 implies that the pace at which data provides information about the process is fast enough. We use these last two conditions to bound the log likelihood in an intermediate step to prove lemma 3. Condition 9 is just a technical assumption to simplify the proof, and it can easily be eliminated, making the argument slightly longer.

Now we are ready to state the following lemma:

Lemma 3. *Under Conditions 1-9,*

$$f^T\left(Y^T|i\right) = \left|\Sigma_T\left(Y^T|i\right)\right|^{-\frac{1}{2}}(2\pi)^{-\frac{k_i}{p}}\pi\left(\theta^*\left(i\right)|i\right)\log f^T(Y^T|\widehat{\theta}\left(i,Y^T\right),i)$$

as $T\to\infty$ *in* P_0^T*-probability.*

Proof. See Appendix. ■

Before we move on, we introduce some additional conditions.

Condition 10. $\forall\, i\in M$, $\left\{f^T\left(Y^T|\theta,i\right)\right\}_{t=0}^{\infty}$ *obeys a strong uniform law of large numbers.*

Condition 11. $\exists\, j\in M$ *such that* $\exists T_0$ *such that* $\forall\, T\geq T_0$

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

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

Condition 12. *For the same* $j\in M$ *of Condition 11* $\exists T_1$ *such that* $\forall\, T\geq T_1$

$$\int_{\Re^{m\times T}} T^{-1}\log\left|\Sigma_T\left(Y^T|j\right)\right|^{-\frac{1}{2}} f^T\left(Y^T|\theta_T^*\left(j\right),j\right)p_0^T\left(Y^T\right)d\nu^T >$$
$$\int_{\Re^{m\times T}} T^{-1}\log\left|\Sigma_T\left(Y^T|i\right)\right|^{-\frac{1}{2}} f^T\left(Y^T|\theta_T^*\left(i\right),i\right)p_0^T\left(Y^T\right)d\nu^T$$

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

Condition 10 is only slightly restrictive, and the results in Andrew (1988) assure that a large class of models satisfy it.[3] Condition 11 requires the model comparison to be a

[3] Andrew (1988) proves laws of large numbers for L^1-mixingales. We proved, 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.

meaningful task by asking one of the models to be the closest to the "true" one under the Kullback-Leibler measure.

At this point it is important to remember that the Cramér-Rao bound is directly related to the speed at which we learn about the parameter as the sample size grows. Condition 12 precludes the pathological case of a model that is further away in the Kullback-Leibler measure than the closest one yet has such a high learning speed that it overcomes the effects of condition 11. Note that if we had not assumed condition 12 we would need to modify the Bayes factor by the ratio of Cramér-Rao bounds to assure consistency.[4] Since for stationary models condition 12 holds, we prefer to show the theorem under this condition.

Finally, we are ready to prove our main theoretical result, i.e., that the Bayes factor selects the model closest to the data regardless of the priors used.

Theorem 2. *Under Conditions 1-12,* $\lim_{T\to\infty} P_{0T}\left(\frac{f^T(Y^T|i)}{f^T(Y^T|j)} = 0\right) = 1.$

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 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 on similar arguments in Chen (1985), Phillips (1996) and

[4] We can show, but do not include because space considerations, that, if condition 12 does not hold, then the result below changes to $\lim_{T\to\infty} P_{0T}\left(\frac{|\Sigma_T(Y^T|i)|^{\frac{1}{2}} f^T(Y^T|i)}{|\Sigma_T(Y^T|j)|^{\frac{1}{2}} f^T(Y^T|j)} = 0\right) = 1.$

Kim (1998), the asymptotic normality of the posterior. We do not seek to use asymptotic approximations to the posteriors because the use of the Markov chain Monte Carlo method allows exact (up to a simulation error) Bayesian computations.

2.4. Numerical Implementation

From our previous description, it is clear that the 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 (or if we use a Linear Quadratic Approximation of the objective function or a loglinearization of the Euler Conditions), the *Kalman Filter* provides an efficient procedure to evaluate the likelihood. If this representation is nonlinear, Fernández-Villaverde and Rubio-Ramírez (2002) show how to use Sequential Monte Carlo methods to evaluate the likelihood function of a general class of nonlinear dynamic equilibrium models.

State Space representations also allow the use of different solutions to a common problem in dynamic equilibrium economies: their stochastic singularity. Since the number of stochastic innovations specified by the theory 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)), introducing measurement errors, using principal components analysis (Landon-Lane (1999)) and others. In this paper we are agnostic about how to solve this singularity, and we merely point out how State Space representations may deal with this problem.

For the second task of computing the marginal likelihood, and since we will not have in general exact analytic expressions, we can use Markov chain Monte Carlo methods as

described, for instance, in Geweke (1998).

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 present an example of a dynamic equilibrium model, the cattle cycle, for its econometric analysis.

3.1. The Cattle Cycle

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) and modified by Anderson, Hansen, McGrattan, and Sargent (1996).

Two reasons suggest the choice of this application. First, despite its relative simplicity, the 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. For instance, Rosen, Murphy, and Scheinkman (1994) mix calibration and ARMA estimation; Anderson, Hansen, McGrattan, and Sargent (1996) use Maximum Likelihood Methods; and Diebold, Ohanian, and Berkowitz (1998) minimize the spectral distance between the data and the model. These procedures give us a benchmark set to assess the performance of the Bayesian approach to model estimation and comparison.

3.2. The Model

There is a representative farmer who breeds cattle and slaughters them 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 adult, h_t, of holding a calf, $\gamma_0 h_t$, and of holding a yearling, $\gamma_1 h_t$. These costs are exogenous, autoregressive, stochastic stationary processes:

$$h_{t+1} = (1-\rho_h)\,\mu_h + h_t + \varepsilon_{ht} \text{ where } \varepsilon_{ht} \sim N\left(0, \sigma_h^2\right) \tag{12}$$

$$m_{t+1} = (1-\rho_h)\,\mu_m + m_t + \varepsilon_{mt} \text{ where } \varepsilon_{mt} \sim N\left(0, \sigma_m^2\right) \tag{13}$$

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

A representative farmer solves the following maximization problem:

$$\max_{\{c_t\}_{t=0}^{\infty}} E_0 \sum_{t=0}^{\infty} \beta^t \left\{ \begin{array}{c} (p_t - m_t)\, c_t - h_t x_t - \gamma_0 h_t g x_{t-1} - \gamma_1 h_t g x_{t-2} \\ -\frac{\psi_1}{2} x_t^2 - \frac{\psi_2}{2} x_{t-1}^2 - \frac{\psi_3}{2} x_{t-2}^2 - \frac{\psi_4}{2} c_t^2 \end{array} \right\} \tag{14}$$

$$s.t.\ x_t = (1-\delta)\, x_{t-1} + g x_{t-3} - c_t \tag{15}$$

$$\{x_{-1}, x_{-2}, x_{-3}\} \text{ fixed} \tag{16}$$

To simplify the model we assume that the quadratic component of the cost is common for all activities:

$$\epsilon^2 = \frac{\psi_1}{2} = \frac{\psi_2}{2} = \frac{\psi_3}{2} = \frac{\psi_4}{2} \tag{17}$$

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 $\varepsilon_{dt} \sim N(0, \sigma_d^2)$.

Finally, we 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:

$$\widetilde{s_t} = s_t + \varepsilon_{yt} \text{ where } \varepsilon_{st} \sim N\left(0, \sigma_s^2\right) \tag{18}$$

$$\widetilde{c_t} = c_t + \varepsilon_{ct} \text{ where } \varepsilon_{ct} \sim N\left(0, \sigma_c^2\right) \tag{19}$$

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

Definition 3. *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 the stochastic processes and initial conditions, the representative farmer solves her problem:*

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

$$s.t.\ x_t = (1-\delta)\, x_{t-1} + g x_{t-3} - c_t \tag{21}$$

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 are given by:*

$$h_{t+1} = (1-\rho_h)\, \mu_h + h_t + \varepsilon_{ht} \text{ where } \varepsilon_{ht} \sim N\left(0, \sigma_h^2\right) \tag{22}$$

$$m_{t+1} = (1-\rho_h)\, \mu_m + m_t + \varepsilon_{mt} \text{ where } \varepsilon_{mt} \sim N\left(0, \sigma_m^2\right) \tag{23}$$

$$d_{t+1} = \rho_d d_t + \varepsilon_{dt} \text{ where } \varepsilon_{dt} \sim N\left(0, \sigma_d^2\right) \tag{24}$$

4. A Structural Estimation of the Cattle Cycle Model

In this section, we estimate the structural parameters of the cattle cycle model and its associated marginal likelihood using the annual measured total stock of beef, the measured slaughter rate, and the price of slaughtered beef for 1900-1990 (Bureau of the Census (1975), and (1989)). First, we specify priors over these structural parameters. Second, using the Metropolis-Hastings algorithm and the Kalman Filter, we find the posterior distributions and moments of the parameters. To check the accuracy of our computations, we present estimates of our numerical errors and convergence assessment of our Markov chain Monte

Carlo. In addition, we study the robustness of the results to different priors. Finally, assuming a quadratic loss function, we 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 an eighteen-dimensional vector $\theta = \{\beta, \delta, \alpha_0, \alpha_1, \gamma_0, \gamma_1, g, \rho_h, \rho_m, \rho_d, \mu_h, \mu_m, \sigma_h, \sigma_m, \sigma_{s,}, \sigma_c, \sigma_d, \epsilon\}$. We will impose dogmatic priors on six parameters. 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 the same restriction is used in Anderson, Hansen, McGrattan, and Sargent (1996), it increases the comparability of our results to previous estimations. We will set $\beta = 0.96$, $\delta = 0$, $\epsilon = 0.0001$, $\rho_d = \sigma_h = 0$, $\mu_h = 37$, and $\mu_m = 63$. The first restriction pins down the discount factor, a difficult parameter to estimate in this type of model, to a commonly used value. The second one rules out deaths in the breeding stock. The value for ϵ is a small number that creates the quadratic costs, and it is basically irrelevant. The last restrictions make demand deterministic and fix the mean value of the processes to the observed means. The remaining vector is then $\theta' = \{\alpha_0, \alpha_1, \gamma_0, \gamma_1, g, \rho_h, \rho_m, \sigma_h, \sigma_m, \sigma_{s,}, \sigma_c\}$.

Table 4.1 presents our priors. 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 positive 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 a mean skewed to the right in a somehow imprecise way. For the four variances of the innovation 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: 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

To solve for the lack of tractable expressions for the likelihood and posterior distributions of the parameters, we use the Kalman Filter and the Random-Walk Metropolis-Hastings to produce a Markov chain $\{\theta_1, \theta_2, ...\theta_m\}$ of parameter values. The empirical histograms of the parameters are included as Figure 1.

Given this Markov chain and a function of interest $g(\cdot)$, 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 indicator functions, we can find the different moments of the distribution or compute quantiles. We simulate a chain of size 10^6 that passes all the requirements of convergence. Table 4.2 reports the expectation and standard deviation of the posterior of 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, note that:

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

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

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

$$f^T\left(Y^T|i\right)^{-1} = \frac{1}{m}\sum_{i=1}^{m} \frac{h(\theta)}{f^T(Y^T|\theta,i)\,\pi(\theta)} \tag{26}$$

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:

$$h(\theta) = \frac{1}{\widehat{p}\,(2\pi)^{\frac{k}{2}}}\left|\widehat{\Sigma}_m\right|^{\frac{1}{2}} \exp\left[-0.5\left(\theta-\widehat{\theta}\right)\widehat{\Sigma}_m^{-1}\left(\theta-\widehat{\theta}\right)'\right] I_{\Theta_M\cap\Theta}(\theta) \tag{27}$$

where $\widehat{p}$ is a normalizing constant. Then, if the posterior is uniformly bounded away from zero on every compact subset of Θ, our computation approximates the marginal likelihood.

Hence, with the output of the Markov chain Monte Carlo, the estimation of the marginal likelihood is 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. Numerical Standard Error of Posterior Moments

A 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 of the estimates. However, the estimation of σ_{μ}^{2} 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 the suggestion by Hannan (1970). Assuming that the function of interest $g(\cdot)$ has a spectral density $S_g(\omega)$ continuous at the origin,[5] we can estimate the NSE as $\left(\frac{1}{m}\widehat{S_g}(0)\right)^{\frac{1}{2}}$ (Hannan (1970) corollary 4, page 208). 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)). Since there is serial correlation in the Markov chain, we need to assure that the simulation is long enough so that the results do not depend on the initial conditions. Theorems of this type require conditions difficult to check in practice.

To overcome this problem, we followed two routes. First, as common in the literature, we

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

compared several chains. Among other things, we simulated ten chains of size 10^5 and one of size 10^6. All of them generated very similar results and their draws followed a stationary process. Second, since informal methods can hide subtle nonconvergence problems, we implemented the convergence test proposed by Geweke (1992). We took the first p_A and the last p_B vectors of the simulation and computed the partial means $\widehat{\mu}_1 = \frac{1}{p_A}\sum_{i=1}^{p_a} g(\theta_i)$ and $\widehat{\mu}_2 = \frac{1}{p_B}\sum_{i=m-p_B+1}^{m} g(\theta_i)$. Then, under the null hypothesis that both means are equal, as $m \to \infty$ we know that

$$\frac{(\widehat{\mu}_1 - \widehat{\mu}_2)}{\left[\frac{\widehat{S_g^A}(0)}{p_A} + \frac{\widehat{S_g^B}(0)}{p_B}\right]^{\frac{1}{2}}} \Rightarrow \mathcal{N}(0,1)$$

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

4.5. Robustness Analysis

The subjective character of the Bayesian paradigm calls for an indication of how the posterior expectations vary with changes in the prior distribution. Methods to undertake this robustness analysis have been presented in Geweke (1999). Given any prior density $\pi^*(\theta)$ with support included in our prior $\pi(\theta)$ support, we can define the weighting function $w(\theta) = \frac{\pi^*(\theta)}{\pi(\theta)}$ and find the new posterior function of interest as $\widehat{\mu} = \frac{1}{m}\left[\sum_{i=1}^{m} w(\theta) g(\theta_i) / \sum_{i=1}^{m} w(\theta)\right]$.

An extensive prior set was tested without altering the reported results. We attribute that to the fact that the sample size is big enough (ninety one observations) to swamp the prior. 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 Anderson, Hansen, McGrattan, and Sargent (1996) 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).

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

Comparison with table 4.2 highlights two points. First, the MLE with low standard errors (*precise* estimates) are closely matched (α_1 equals to 1.27 against 1.27, ρ_m equal 0.70 against 0.70, etc.). Second, for those parameters imprecisely estimated, as γ_0 and γ_1 (the relative holding costs of cattle according to their 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 γ_1 are multimodal and relatively flat over a long range of values. Given these shapes, the MLE will find one of the local maxima, and the flatness of the likelihood around these points will turn out very high standard errors. The Bayes estimate overcomes these difficulties and gives a more accurate finite sample view of the plausible parameter values. It is important to emphasize that, through robustness analysis, we checked that this higher precision is *not* spuriously induced by the prior but by the use of the whole likelihood shape that bayesian procedures imply instead of just one point (and its neighborhood) as in MLE. 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).

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

5. Comparing Models: The Cattle Cycle versus BVARs

In this section we 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 Autoregression models, simple linear statistical representations of the dynamic relations among variables, have a proven forecasting record (Litterman (1986)) and have been proposed as alternatives to a more structural modeling of time series (Sims (1980)).[6] We first describe the Vector Autoregression specifications, then the priors and finally the results of the models comparison.

5.1. A Vector Autoregression Specification

We define nine versions of a three-variable 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:

$$y_t = \sum_{i=1}^{p} y_{t-i} A_i + C + u_t \qquad \forall t \in \{1, ..., T\},\ u_t \sim \mathcal{N}(0, \Psi) \tag{28}$$

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

A useful way to rewrite (28) 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 have $y_i = Z\gamma_i + u_i$. Stacking the column vectors y_i, γ_i, and u_i in y, γ, and u, we get $y = (I \otimes Z)\gamma + u$, where

[6] Note that, however, these BVARs are not completely nonnested with the cattle cycle model since the latter has a restricted vector autoregression representation.

$u \sim \mathcal{N}(0, \Psi \otimes I)$. The likelihood function is:

$$f^T(\gamma|\Psi) \propto |\Psi|^{-\frac{T}{2}} \exp\left\{-tr\left[(Y - Z\Gamma)'\Psi^{-1}(Y - Z\Gamma)\right]/2\right\} \tag{29}$$

5.2. Prior Distributions

We use three different priors, each one more general than the previous one: a modified *Minnesota prior*, a *Normal-Wishart prior,* and a *Hierarchical Prior* (see Kadiyala and Karlsson (1997) and Sims and Zha (1998)).

5.2.1. Minnesota Prior

Litterman (1980) defined the Minnesota prior. Its basic feature is that the prior mean implies that each variable follows a random walk. To win further flexibility, we modify two aspects of the original prior. First, we let the prior variances decrease slowly with the lags. Litterman used a rate $1/k^2$ while we use $1/k$. Second, we do not restrict the variance-covariance matrix to be diagonal.

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 one, and the prior means of the remaining parameters are set to zero, i.e., the mean of γ_s for $s \in \{1, 2, 3\}$ is $\underline{\mu_s} = \left(0, \chi_{\{1\}}(s-1), \chi_{\{1\}}(s-2), \chi_{\{1\}}(s-3), 0, ..., 0\right)'$.

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

$$\underline{\Sigma_s} = \begin{pmatrix} \pi_3\sigma_s^2 & 0 & \cdots & 0 \\ 0 & \widetilde{\pi}_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \widetilde{\pi}_p \end{pmatrix} \tag{30}$$

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

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 the MLE.

5.2.2. Normal-Wishart Prior

The last characteristic of our Minnesota prior is restrictive since it implies an extraordinarily precise knowledge of the variances of innovations. An alternative is to assume that Ψ is Wishart distributed. We define the prior distributions $\gamma|\Psi \sim N(\underline{\mu}, \Psi \otimes \underline{\Sigma})$ and $\Psi \sim iW(\underline{\Psi}, \alpha)$ where $\gamma = (\gamma_1, \gamma_2, \gamma_3)'$, $E(\gamma) = \underline{\mu} = (\underline{\mu_1}, \underline{\mu_2}, \underline{\mu_3})'$, and $var(\gamma_s) = \underline{\Sigma}_s$, $\forall s \in \{1,2,3\}$. If we let s_i^2 be the MLE of the variances of the innovations, then $\underline{\Psi}$ is diagonal with entries $\{(\alpha - n - 1)s_1^2, (\alpha - n - 1)s_2^2, (\alpha - n - 1)s_3^2\}$.

5.2.3. Hierarchical 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 a Hierarchical 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 to compute their marginal likelihoods. We report our finding in table 5.1.[7] We learn two lessons. First, 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 a Minnesota prior and two lags. The log difference in favor of the BVAR is 43.46. How big is this difference intuitively? We will provide two measures. First, we will note that this difference means that the empirical evidence overcomes any prior ratio lower than 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. Also, it may be argued that the model is designed to capture only certain characteristics of the data (as, for example, in the stochastic growth model, the business cycles fluctuations). It is an open question how to use the marginal likelihood to extract how well the model accounts for particular aspects of the data we may

[7] 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 assessed convergence using the same methods described in the previous section.

be interested in.

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
Hierarchical (1)	−715.9167
Hierarchical (2)	−732.1339
Hierarchical (3)	−782.9960

Finally, 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 outperforms them.

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.[8] 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 modeling.

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

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

simple to apply, but also its shown properties have an 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, possibly superior to the use of classical asymptotic approximations. Some parallel research (Fernández-Villaverde and Rubio-Ramírez (2002)) 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 models.

7. Appendix

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

7.1. Proofs

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

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

Then:

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

$$\begin{aligned} &f^T\left(Y^T,i\right)^{-1}f^T(Y^T|\widehat{\theta}_T\left(i,Y^T\right),i)\exp\left[\log f^T(Y^T|\theta_T^*\left(i\right),i)-\log f^T(Y^T|\widehat{\theta}_T\left(i,Y^T\right),i)\right] \\ &\times\int_{\theta\in\Theta_i\setminus\mathcal{E}\left(\widehat{\theta}_T(i,Y^T),\delta_T(i)\right)}\pi\left(\theta,i\right)\exp\left[\log f^T(Y^T|\theta,i)-\log f^T(Y^T|\theta_T^*\left(i\right),i)\right]d\theta \end{aligned} \tag{31}$$

but (1) and (2) imply that $\exp\left[\log f^T(Y^T|\theta_T^*\left(i\right),i)-\log f^T(Y^T|\widehat{\theta}_T\left(i,Y^T\right),i)\right]=O_p\left(1\right)$ as $T\to\infty$ in P_0^T-probability.

With this last statement, we only need to check that

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

as $T \to \infty$ in P_0^T-probability.

Then, by (10), for T large enough,

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

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

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^T-probability and the results follow. ■

Proof of Lemma 2. Assume Lemma 2 is not true. Since $\theta_T^*\left(i\right)\to\theta^*\left(i\right)$, then $\exists\,\gamma>0$ such that

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

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

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

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

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

$$\int_{\mathcal{B}(\theta_T^*(i),\eta)}\pi\left(\theta,i\right)\exp\left[\log f^T(Y^T|\theta,i)-\log f^T(Y^T|\theta_T^*\left(i\right),i)\right]d\theta \tag{32}$$

but (11) implies that the right-hand side is bigger than zero in P_0^T-probability. Then:

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

as $T\nearrow\infty$ in P_0^T-probability, which contradicts Lemma 1. ■

Proof of Lemma 3. Let $\varepsilon_T\searrow 0$ and choose $\delta_T(i)$ such that

$$\left|\pi\left(\theta|i\right)-\pi\left(\theta^*\left(i\right)|i\right)\right|\leq\varepsilon_T\pi\left(\theta^*\left(i\right)|i\right)$$

$$I-A_T\left(\varepsilon_T\right)\leq-L_T''\left(Y^T|\theta,i\right)\Sigma_T\left(Y^T|i\right)\leq I+A_T\left(\varepsilon_T\right)$$

$\forall\theta\in\mathcal{E}\left(\widehat{\theta}_T\left(i,Y^T\right),\delta_T(i)\right)$ as $T\to\infty$ in P_0^T-probability.

Note that we can write $f^T\left(Y^T|i\right)$ as:

$$f^T\left(Y^T|i\right)=\int_{\Theta_i\setminus\mathcal{E}\left(\widehat{\theta}_T(i,Y^T),\delta_T(i)\right)}+\int_{\mathcal{E}\left(\widehat{\theta}_T(i,Y^T),\delta_T(i)\right)}\pi\left(\theta|i\right)\exp\left(\log f^T(Y^T|\theta,i)\right)d\theta=I_{1,T}+I_{2,T}$$

Since we know by Lemma 1 that $I_{1,T}\to 0$ as $T\to\infty$ in P_0^T-probability, we need to concentrate only on the asymptotic behavior of $I_{2,T}$.

Then

$$I_{2,T}=\int_{\mathcal{E}\left(\widehat{\theta}_T(i,Y^T),\delta_T(i)\right)}\pi\left(\theta|i\right)\exp\left(\log f^T(Y^T|\theta,i)\right)d\theta=$$

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

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

where

$$\Xi\left(\theta, Y^T\right) = -\frac{1}{2}\left(\theta - \widehat{\theta}_T\left(i, Y^T\right)\right)' \left(I + R_T\left(Y^T|i\right)\right) \Sigma_T\left(Y^T|i\right)^{-1} \left(\theta - \widehat{\theta}_T\left(i, Y^T\right)\right)$$

and $R_T\left(Y^T|i\right) = -L_T''\left(y^T|\widetilde{\theta}, i\right) \Sigma_T\left(y^T|i\right) - I$, where $\widetilde{\theta}$ lies between θ and $\widehat{\theta}_T\left(i, Y^T\right)$.

Then we can bound $I_{2,T}$ in the following way

$$\left(1 - \varepsilon_T\right) I_{3,T} \leq \frac{I_{2,T}}{f^T(Y^T|\widehat{\theta}_T\left(i, Y^T\right), i)\pi\left(\theta^*\left(i\right)|i\right)} \leq \left(1 + \varepsilon_T\right) I_{3,T}$$

where $I_{3,T} = \int_{\mathcal{E}\left(\widehat{\theta}_T(i,Y^T),\delta_T(i)\right)} \exp\left(\Xi(\theta, Y^T)\right) d\theta$.

Let $\overline{\vartheta}_T = \delta_T(i)\sqrt{\left(1 + \overline{\varrho}\left(\varepsilon_T\right)\right)/\underline{\lambda_T}}$ and $\underline{\vartheta}_T = \delta_T(i)\sqrt{\left(1 - \overline{\varrho}\left(\varepsilon_T\right)\right)/\overline{\lambda_T}}$, where $\overline{\varrho}\left(\varepsilon_T\right)$ and $\underline{\varrho}\left(\varepsilon_T\right)$ are the largest and the smallest eigenvalues of $A_T\left(\varepsilon_T\right)$ and $\overline{\lambda_T}$ and $\underline{\lambda_T}$ are the largest and smallest eigenvalues of Σ_T and note that $\left\{z; z'\left(I + A_T\left(\varepsilon_T\right)\right)\Sigma_T^{-1} z < \underline{\vartheta}_T\right\} \subseteq \mathcal{E}\left(0, \frac{\underline{\vartheta}_T}{\sqrt{\frac{1-\overline{\varrho}\left(\varepsilon_T\right)}{\underline{\lambda_T}}}}\right)$ and $\mathcal{E}\left(0, \frac{\overline{\vartheta}_T}{\sqrt{\frac{1+\overline{\varrho}\left(\varepsilon_T\right)}{\underline{\lambda_T}}}}\right) \subseteq \left\{z; z'\left(I - A_T\left(\varepsilon_T\right)\right)\Sigma_T\left(Y^T|i\right)^{-1} z < \overline{\vartheta}_T\right\}$.

Thus,

$$\left|I + A_T\left(\varepsilon_T\right)\right|^{-\frac{1}{2}} \left|\Sigma_T\left(Y^T|i\right)\right|^{\frac{1}{2}} \int_{\mathcal{E}\left(0,\underline{\vartheta}_T\right)} \exp\left(-\frac{1}{2}z'z\right) dz \leq$$

$$\int_{\mathcal{E}\left(\widehat{\theta}_T(i,Y^T),\delta_T(i)\right)} \exp\left(-\frac{1}{2}\left(\theta - \widehat{\theta}_T\left(i, Y^T\right)\right)' \left(I + A_T\left(\varepsilon_T\right)\right)\Sigma_T\left(Y^T|i\right)^{-1}\left(\theta - \widehat{\theta}_T\left(i, Y^T\right)\right)\right) d\theta \leq$$

$$I_{3,T} \leq$$

$$\int_{\mathcal{E}\left(\widehat{\theta}_T(i,Y^T),\delta_T(i)\right)} \exp\left(-\frac{1}{2}\left(\theta - \widehat{\theta}_T\left(i, Y^T\right)\right)' \left(I - A_T\left(\varepsilon_T\right)\right)\Sigma_T\left(Y^T|i\right)^{-1}\left(\theta - \widehat{\theta}_T\left(i, Y^T\right)\right)\right) d\theta \leq$$

$$\left|I - A_T\left(\varepsilon_T\right)\right|^{-\frac{1}{2}} \left|\Sigma_T\left(Y^T|i\right)\right|^{\frac{1}{2}} \int_{\mathcal{E}\left(0,\overline{\vartheta}_T\right)} \exp\left(-\frac{1}{2}z'z\right) dz$$

as $T \to \infty$ in P_0^T-probability.

Since $\overline{\varrho}(\varepsilon_T) \downarrow 0$, Condition 8 implies that $\overline{\vartheta}_T \to \infty$ as $T \to \infty$ in P_0^T-probability, then

$$|I + A_T(\varepsilon_T)|^{-\frac{1}{2}} \left|\Sigma_T\left(Y^T|i\right)\right|^{\frac{1}{2}} (2\pi)^{\frac{k_i}{2}} \leq I_{3,T} \leq |I - A_T(\varepsilon_T)|^{-\frac{1}{2}} \left|\Sigma_T\left(Y^T|i\right)\right|^{\frac{1}{2}} (2\pi)^{\frac{k_i}{2}}$$

as $T \to \infty$ in P_0^T-probability, which implies the result of the lemma

$$f^T\left(Y^T|i\right) = I_{2,T} = \left|\Sigma_T\left(Y^T|i\right)\right|^{-\frac{1}{2}} (2\pi)^{-\frac{k_i}{2}} f^T(Y^T|\widehat{\theta}_T\left(i, Y^T\right), i)\pi\left(\theta^*(i)|i\right)$$

as $T \to \infty$ in P_0^T-probability. ■

Proof of Theorem 2. From lemma 3 we can write

$$\lim_{T\to\infty} P_{0T}\left(\frac{f^T\left(Y^T|i\right)}{f^T\left(Y^T|j\right)} = \frac{(2\pi)^{-\frac{k_i}{2}}\left|\Sigma_T\left(Y^T|i\right)\right|^{-\frac{1}{2}} f^T\left(Y^T|\theta_T^*(i), i\right)\pi\left(\theta^*(i)|i\right)}{(2\pi)^{-\frac{k_j}{2}}\left|\Sigma_T\left(Y^T|j\right)\right|^{-\frac{1}{2}} f^T\left(Y^T|\theta_T^*(j), j\right)\pi\left(\theta^*(j)|j\right)}\right) = 1 \quad (33)$$

Now, to prove that $\lim_{T\to\infty} P_{0T}\left(\frac{(2\pi)^{-\frac{k_i}{2}}\left|\Sigma_T\left(Y^T|i\right)\right|^{-\frac{1}{2}} f^T\left(Y^T|\theta_T^*(i),i\right)\pi(\theta^*(i)|i)}{(2\pi)^{-\frac{k_j}{2}}|\Sigma_T(Y^T|j)|^{-\frac{1}{2}} f^T\left(Y^T|\theta_T^*(j),j\right)\pi(\theta^*(j)|j)} = 0\right) = 1$ and since

$$\begin{aligned}
&\left[\begin{array}{c}\frac{1}{T}\log(2\pi)^{-\frac{k_i}{2}}\left|\Sigma_T\left(Y^T|i\right)\right|^{-\frac{1}{2}} f^T\left(Y^T|\theta_T^*(i), i\right)\pi\left(\theta^*(i)|i\right) - \\ -\frac{1}{T}\log(2\pi)^{-\frac{k_j}{2}}\left|\Sigma_T\left(Y^T|j\right)\right|^{-\frac{1}{2}} f^T\left(Y^T|\theta_T^*(j), j\right)\pi\left(\theta^*(j)|j\right) = -\infty\end{array}\right] \\
\subseteq &\left[\frac{(2\pi)^{-\frac{k_i}{2}}\left|\Sigma_T\left(Y^T|i\right)\right|^{-\frac{1}{2}} f^T\left(Y^T|\theta_T^*(i), i\right)\pi\left(\theta^*(i)|i\right)}{(2\pi)^{-\frac{k_j}{2}}|\Sigma_T(Y^T|j)|^{-\frac{1}{2}} f^T\left(Y^T|\theta_T^*(j), j\right)\pi\left(\theta^*(j)|j\right)} = 0\right]
\end{aligned}$$

we only need to show

$$\lim_{T\to\infty} P_{0T}\left(\begin{array}{c}\frac{1}{T}\log(2\pi)^{-\frac{k_i}{2}}\left|\Sigma_T\left(Y^T|i\right)\right|^{-\frac{1}{2}} f^T\left(Y^T|\theta_T^*(i), i\right)\pi\left(\theta^*(i)|i\right) - \\ -\frac{1}{T}\log(2\pi)^{-\frac{k_j}{2}}\left|\Sigma_T\left(Y^T|j\right)\right|^{-\frac{1}{2}} f^T\left(Y^T|\theta_T^*(j), j\right)\pi\left(\theta^*(j)|j\right) = -\infty\end{array}\right) = 1 \quad (34)$$

Conditions (9)-(10) allow us to use an argument similar to Wald (1949) to prove (34) and the result from lemma 3 to finish the proof. ■

7.2. Some Computational Details

The cattle cycle model was computed using Vaughan's eigenvector method to solve the Algebraic Riccati equation associated with 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. As suggested by Anderson, Hansen, McGrattan, and Sargent (1996), we checked the robustness of the Vaughan's eigenvector method comparing our results with those implied by alternative algorithms (Schur, generalized Schur, and Matrix Sign) since Vaughan's algorithm, although fast, may provide inaccurate answers when we have nearly repeated eigenvalues. This possibility may arise in our estimation procedure as we travel regions of the parameter space far away from the MLE. We found, however, that the results using these different methods were nearly identical to the ones with Vaughan's procedure. As mentioned before, Anderson, Hansen, McGrattan, and Sargent (1996) provide further details on this issue.

The Metropolis-Hastings 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 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. If the acceptance rate is very small, the chain will not visit a large enough set in a reasonable number of iterations. If the acceptance rate is very high, the chain will not stay enough time in the 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.

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] AGUIRRE-TORRES, V. and A. RONALD GALLANT (1999). "Testing Separate Dynamic Nonlinear Econometric Models". Mimeo, University of North Carolina.

[2] 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.

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

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

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

[6] 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.

[7] 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.

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

[9] 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.

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

[11] FERNÁNDEZ-VILLAVERDE, J. and RUBIO-RAMÍREZ, J.F. (2002). "Estimating Nonlinear Dynamic Equilibrium Economies: A Likelihood Approach". Mimeo, University of Pennsylvania.

[12] 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.

[13] 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.

[14] 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.

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

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

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

[18] 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.

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

[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] 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.

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

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

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

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

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

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

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

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

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

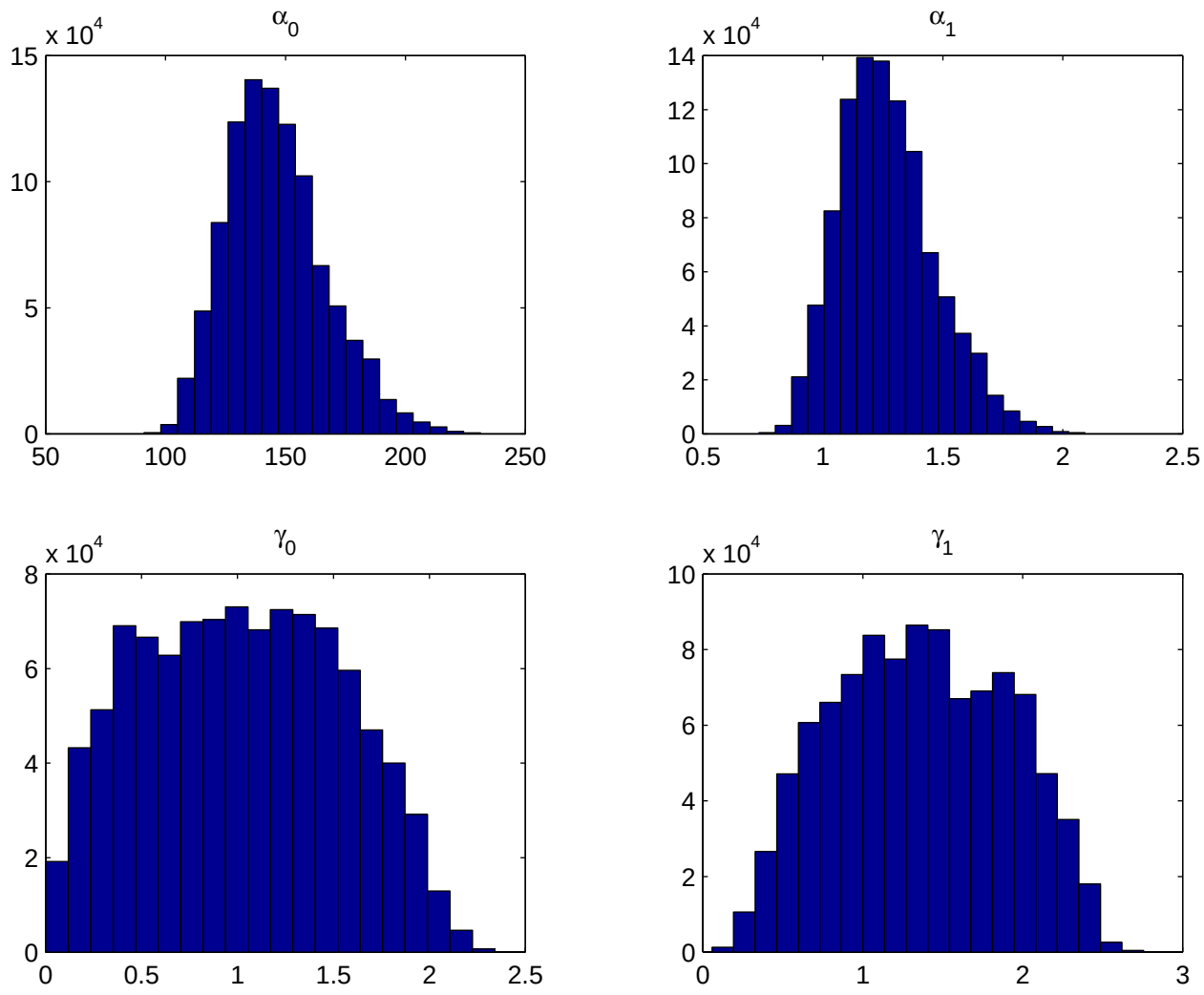
[40] SCHORFHEIDE, F. (1999). "A Unified Econometric Framework for the Evaluation of DGSE Models". Mimeo, University of Pennsylvania.

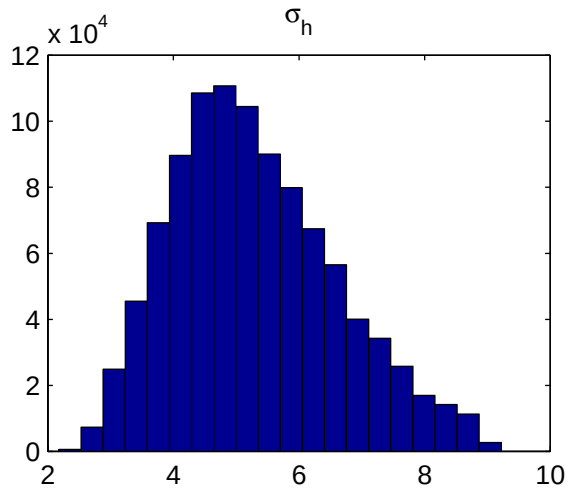
[41] SHORE, J.E. and JOHNSON, R.W (1980). "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy". *IEEE Transactions on Information Theory*, **IT-26**, 26-37.

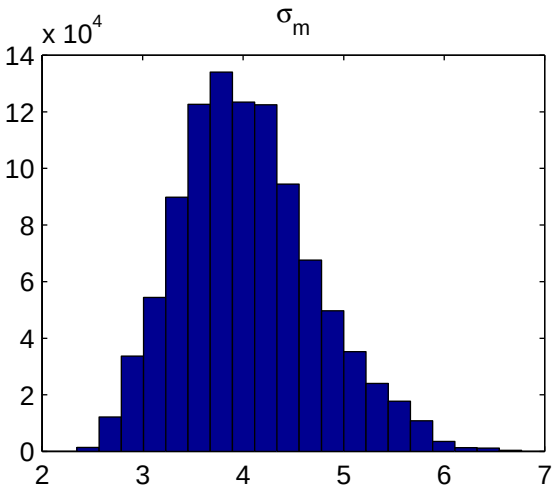
[42] VUONG, Q.H. (1989). "Likelihood Ratio Test for Model Selection and Non-Nested Hypotheses". *Econometrica,* **57**, 307-333.

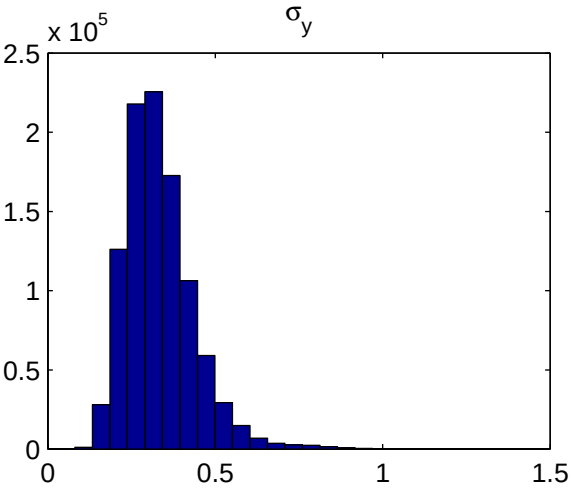
[43] WALD, A. (1949). "Note on the Consistency of the Maximum Likelihood Estimate". *Annales of Mathematical Statistics,* **20**, 595-601.

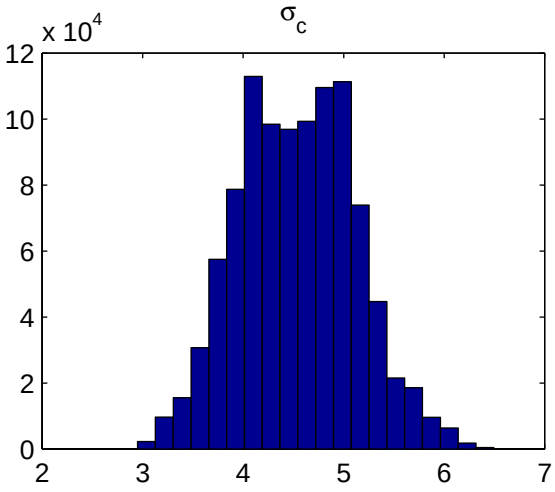
Figure 1 : Empirical Distribution of the Posterior

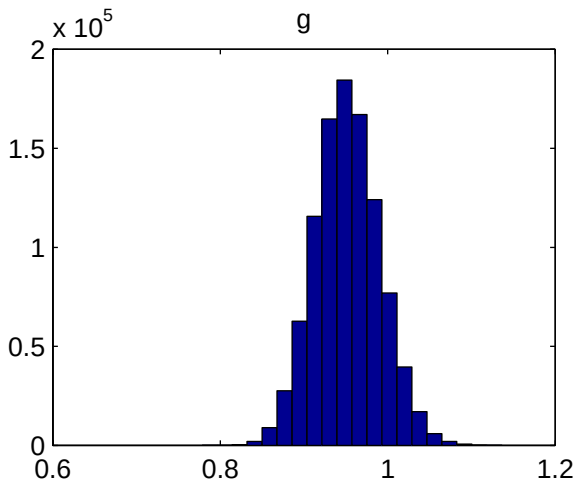


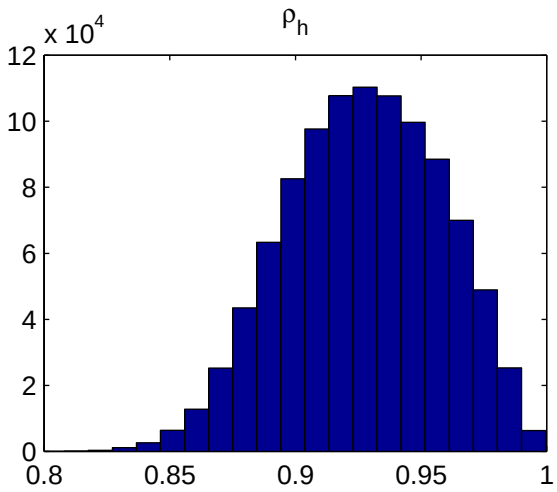

σh
x 10^4
12
10
8
6
4
2
0
2
4
6
8
10

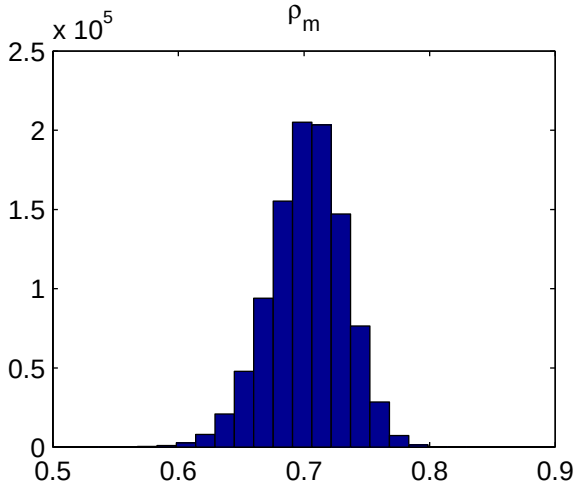

σm
x 10^4
14
12
10
8
6
4
2
0
2
3
4
5
6
7


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