



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Journal of Econometrics 123 (2004) 153–187

JOURNAL OF
Econometrics

www.elsevier.com/locate/econbase

Comparing dynamic equilibrium models to data: a Bayesian approach

Jesús Fernández-Villaverde^{a,*}, Juan Francisco Rubio-Ramírez^b

^a*Department of Economics, 160 McNeil Building, 3718 Locust Walk, University of Pennsylvania, Philadelphia, PA 19104, USA*

^b*Department of Research, 1000 Peachtree St NE, Federal Reserve Bank of Atlanta, Atlanta, GA 30309, USA*

Received 21 July 2003; accepted 22 October 2003

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.

© 2003 Elsevier B.V. All rights reserved.

JEL classification: C11; C15; C51; C52

Keywords: Bayesian inference; Bayesian asymptotics; Bayes factors; Dynamic equilibrium models; Cattle cycle

1. Introduction

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

* Corresponding author. Tel.: 215-898-15-04; fax: 215-573-20-57.

E-mail address: jesusfv@econ.upenn.edu (J. Fernández-Villaverde).

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 (2000), 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 criticisms 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 (1991). 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 and flexible. Issues such as nonstationarity do not require specific methods as needed in classical inference (Sims and Uhlig, 1991). Also from the parameters posterior distribution we can build point estimates to match different objects of interest as impulse-response functions to shocks estimated from the data (Schorfheide, 2000) or moment conditions (Geweke, 1999b). Third, there is an asymptotic justification of the Bayes procedure. As mentioned before, we prove consistency of both the point

¹An 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)).

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 et al., 1994](#); or [Geweke et al., 1997](#)). Fifth, the Bayes' Theorem is an optimal information processing rule in the sense of [Zellner \(1988\)](#): it uses efficiently all the available information in the data (both in small and large samples) and it does not add extraneous information. Sixth, the advent of Markov chain 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). Seventh, 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 \(2001\)](#) 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 et al. \(2000\)](#) 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 \(2000\)](#) 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 and discusses further extensions. 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 $Y \equiv \{Y_t : \Omega \rightarrow \mathfrak{R}^m, m \in \mathcal{N}, t = 1, 2, \dots\}$, defined on a complete probability space $(\Omega, \mathfrak{I}, P_0)$, where $\Omega = \mathfrak{R}^{m \times \infty} \equiv \lim_{T \rightarrow \infty} \otimes_{t=0}^T \mathfrak{R}^m$ and $\mathfrak{I} \equiv \lim_{T \rightarrow \infty} \mathfrak{I}^T \equiv \lim_{T \rightarrow \infty} \otimes_{t=0}^T \mathcal{B}(\mathfrak{R}^m) \equiv \mathcal{B}(\mathfrak{R}^{m \times \infty})$ is the Borel σ -algebra generated by the measurable finite-dimensional product cylinders. Define a T segment as $Y^T \equiv (Y'_1, \dots, Y'_T)'$ with $Y^0 = \{\emptyset\}$ and a realization of that segment as $y^T \equiv (y'_1, \dots, y'_T)'$. Also define $P_0^T(B) \equiv P_0(B) | \mathfrak{I}^T \equiv P_0(Y^T \in B), \forall B \in \mathfrak{I}^T$ to be the restriction of P_0 to \mathfrak{I}^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 $\Omega = \mathfrak{R}^{m \times \infty}$. In this case, Y_t is the projection operator that selects y_t , the t th coordinate of ω , so that $Y_t(\omega) = y_t$. With $\mathfrak{I} \equiv \mathcal{B}(\mathfrak{R}^{m \times \infty})$, 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 $(\mathfrak{R}^{m \times T}, \mathcal{B}(\mathfrak{R}^{m \times T}))$ for $T = 1, 2, \dots$ 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 $M \equiv \{1, 2, \dots, H\}$. We define a model i as the collection $S(i) \equiv \{f(\theta, i), \pi(\theta | i), \Theta_i\}$, where $f(\theta, i) \equiv \{f^n(\cdot | \theta, i) : \mathfrak{R}^{m \times n} \times \Theta_i \rightarrow \mathfrak{R}, n = 1, 2, 3, \dots\}$ is the set of densities $f^n(\cdot | \theta, i)$ on $(\mathfrak{R}^{m \times n}, \mathcal{B}(\mathfrak{R}^{m \times n}))$, $\pi(\theta | i)$ is a prior density on $(\Theta_i, \mathcal{B}(\Theta_i))$, and θ is a k_i -dimensional vector of unknown parameters such that $\theta \subseteq \Theta_i \subseteq \mathfrak{R}^{k_i} \forall i \in M$. 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 a Vector Autoregression).

The function $f^T(y^T | \theta, i)$ is called the pseudo-likelihood function of the data. Define the pseudo-maximum likelihood point estimate (PMLE) as $\hat{\theta}_T(i, y^T) \equiv \operatorname{argmax}_{\theta \in \Theta_i} \log f^T(y^T | \theta, i)$. Note that we do not assume that there exists a value θ^* such that $f^T(y^T | \theta^*, i) = p_0^T(y^T)$. 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 $\forall \theta \in \Theta_i, f^T(y^T, \theta | i) = f^T(y^T | \theta, i) \pi(\theta | i)$. With this notation and using conditional probabilities, we can write the posterior of the parameters as $\pi(\theta | y^T, i) \propto f^T(y^T | \theta, i) \pi(\theta | i)$ and its marginal likelihood as

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

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, $\pi(\theta|i)$ is always proper).

Usually we will be in the situation where $f^T(y^T|\theta, i)$ can be factorized in the following way: $f^T(y^T|\theta, i) = \prod_{t=1}^T f_t(y_t|y^{t-1}, \theta, i)$ where $f_t(\cdot|y^{t-1}, \theta, i) : \mathfrak{R}^{m \times t} \times \Theta_i \rightarrow \mathfrak{R}^+$ is $\mathcal{B}(\mathfrak{R}^{m \times t})$ -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 within sample forecasting performance) and computationally (to evaluate pseudo-likelihoods recursively).

Now we define the Kullback–Leibler measure as

$$K(f^T(\cdot|\theta, i); p_0^T(\cdot)) = \int_{\mathfrak{R}^{m \times T}} \log \left(\frac{p_0^T(Y^T)}{f^T(Y^T|\theta, i)} \right) p_0^T(Y^T) dv^T.$$

The intuition of this closeness concept is simple: it evaluates the average surprise with respect to the true measure that the researcher using $f^T(\cdot|\theta, i)$ suffers if he suddenly learns that the true density is $p_0^T(\cdot)$. 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. We refer the interested reader to [Shore and Johnson \(1980\)](#) and [Csiszar \(1991\)](#).

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

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

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

which can be intuitively factored between the Bayes Factor

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

and the ratio of priors π_i/π_j as

$$POR_{i,j|Y_T} = B_{i,j|Y_T} \frac{\pi_i}{\pi_j}. \quad (4)$$

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

The Bayes factor also has an interpretation as an optimal information processing rule in small and in large samples. Zellner (1988) shows how the Bayes' theorem is a 100% efficient processing information rule in the sense that it uses all the existing evidence and it does not add extraneous information. Since the comparison of marginal likelihoods is nothing more than the application of the Bayes' theorem, Zellner's result provides a strong information-based justification of the Bayes factor in addition to the standard axiomatic foundations of Bayesian theory.

In the same way the marginal likelihood is related to the likelihood value at the PMLE, the Bayes factor is linked with 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).

The contribution embodied in the theorems is important for several reasons. First, we assure that, even when the models are misspecified, the priors are irrelevant as the

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.

sample size grows. This result is not surprising since it is the natural counterparts from the bayesian perspective of the results in White (1994). However its formal derivation is a useful addition to the literature. Also is a required step to prove the next result. 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 misspecified and/or nonnested. This result, the main theoretical point of the paper, is to the best of our knowledge new.

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 “neighborhood system” in the parameter space.

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

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

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

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 $\hat{\theta}_T(i, y^T)$. This system allows for different rates of accumulation along different dimensions. Kim (1998) shows that this feature of the “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 $\forall \theta \in \Theta_i, \lim_{T \rightarrow \infty} P_0^T(T^{-1} \log f^T(Y^T | \theta, i) < \infty) = 1$.

Condition 2. For $\forall i \in M, \lim_{T \rightarrow \infty} P_0^T(T^{-1} \log f^T(Y^T | \theta_T^*(i), i) > -\infty) = 1$.

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

$$\lim_{T \rightarrow \infty} P_0^T(|f^T(Y^T | i) = 0) = 0, \tag{6}$$

$$\lim_{T \rightarrow \infty} P_0^T \left(\left| \frac{f^T(Y^T | \hat{\theta}_T(i, Y^T), i)}{f^T(Y^T | i)} \right| = 0 \right) = 0. \tag{7}$$

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 $\bigcap_{t=1}^\infty \mathcal{E}(a, \delta_t(i)) = \{a\}$. Then, there exists a sequence of nonincreasing positive functions $\{k_T(\delta_T(i), i), T = 1, 2, \dots\}$ such that $Tk_T(\delta_T(i), i) \nearrow \infty$ and

$$\lim_{T \rightarrow \infty} \inf P_0^T \left(\sup_{\theta \in \Theta_i \setminus \mathcal{E}(\hat{\theta}_T(i, Y^T), \delta_T(i))} \frac{\log f^T(Y^T | \theta, i) - \log f^T(Y^T | \theta_T^*(i), i)}{T} \leq -k_T(\delta_T(i), i) \right) = 1. \tag{8}$$

Conditions 1–4 are common on the literature and not very restrictive. 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 neighborhood $\mathcal{E}(\hat{\theta}_T(i, Y^T), \delta_T(i))$ 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 $\Theta_i \setminus \mathcal{E}(\hat{\theta}_T(i, Y^T), \delta_T(i))$ by $\exp[-k_T(\delta_T, i)T] \downarrow 0$. 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.

With these conditions we can prove the following lemma:

Lemma 1. Under Conditions 1–4, $\int_{\theta \in \Theta_i \setminus \mathcal{E}(\hat{\theta}_T(i, Y^T), \delta_T(i))} \pi(\theta | Y^T, i) d\theta \rightarrow 0$ as $T \rightarrow \infty$ in P_0^T -probability $\forall i \in M$.

Proof. See Appendix. \square

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 on the asymptotic behavior of the series $\{f^T(Y^T | \theta, i)\}_{t=0}^\infty$:

Condition 5. $\forall i \in M, \{f^T(Y^T | \theta, i)\}_{t=0}^\infty$ obeys a strong uniform law of large numbers.

Condition 5 is only slightly restrictive, and the results in Andrews (1988) and Davidson (1994) assure that a very large class of models satisfy it.³

Lemma 2. *Under Conditions 1–5, $\hat{\theta}_T(i, Y^T) - \theta^*(i) \rightarrow 0$ as $T \rightarrow \infty$ in P_0^T -probability.*

Proof. See Appendix. \square

With these two lemmas, it can be shown that the posterior distribution of the parameters collapses to the pseudo-true value of the parameter. Formally:

Theorem 1. *Under Conditions 1–5, $\int_{\theta \in \Theta_i \setminus \mathcal{E}(\theta^*(i), \delta_T(i))} \pi(\theta | Y^T, i) d\theta \rightarrow 0$ as $T \rightarrow \infty$ in P_0^T -probability $\forall i \in M$.*

Proof. This proof follows directly from Lemmas 1 and 2. \square

Now we prove the main theorem of this paper. Before doing so, 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 the lemma, we 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

$$L'_T(y^T | \theta, i) \equiv \frac{\partial \log f^T(y^T | \theta, i)}{\partial \theta},$$

$$L''_T(y^T | \theta, i) \equiv \frac{\partial^2 \log f^T(y^T | \theta, i)}{\partial \theta \partial \theta'}$$

$$\Sigma_T(y^T | i) \equiv [-L''_T(y^T | \hat{\theta}_T(i, y^T), i)]^{-1},$$

where $\Sigma_T(y^T | i)$ is the Cramér–Rao bound.

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

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

$\forall \theta \in \mathcal{E}(\hat{\theta}_T(i, Y^T), \delta_T(i))$, where $A(\varepsilon_T)$ is a semidefinite positive symmetric matrix 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, $\lambda_{\min}(\Sigma_T(y^T | i)^{-1}) \|\delta_T(i)\|_\infty \rightarrow \infty$ as $T \rightarrow \infty$ in P_0^T -probability.

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

³ Andrews (1988) proves laws of large numbers for L^1 -mixingales. We proved, we do not include, that an exponential density family, $\{\log f^t(Y^t | \theta, i)\}_{t=1}^\infty$, is a L^1 -mixingale. Davidson (1994) presents even more general results for Near-epoch dependent and mixing processes.

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,*

$$f^T(Y^T|i) = |\Sigma_T(Y^T|i)|^{-1/2} (2\pi)^{-k_i/p} f^T(Y^T|\hat{\theta}(i, Y^T), i) \pi(\theta^*(i)|i)$$

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

Proof. See Appendix. \square

Before we move on, we introduce two final conditions.

Condition 10. $\exists j \in M$ such that $\exists T_0$ such that $\forall T \geq T_0$

$$\begin{aligned} & \int_{\mathfrak{R}^{m \times T}} T^{-1} \log f^T(Y^T|\theta_T^*(j), j) p_0^T(Y^T) \, dv^T \\ & > \int_{\mathfrak{R}^{m \times T}} T^{-1} \log f^T(Y^T|\theta_T^*(i), i) p_0^T(Y^T) \, dv^T \end{aligned}$$

$\forall i \in M \setminus \{j\}$.

Condition 11. For the same $j \in M$ of Condition 10 $\exists T_1$ such that $\forall T \geq T_1$

$$\begin{aligned} & \int_{\mathfrak{R}^{m \times T}} T^{-1} \log |\Sigma_T(Y^T|j)|^{-1/2} f^T(Y^T|\theta_T^*(j), j) p_0^T(Y^T) \, dv^T \\ & > \int_{\mathfrak{R}^{m \times T}} T^{-1} \log |\Sigma_T(Y^T|i)|^{-1/2} f^T(Y^T|\theta_T^*(i), i) p_0^T(Y^T) \, dv^T, \end{aligned}$$

$\forall i \in M \setminus \{j\}$.

Condition 10 requires the model comparison to be a meaningful task by asking one of the models to be the closest to the “true” one under the Kullback–Leibler measure. Condition 11 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 10. Note that if we had not assumed Condition 11 we would need to modify the Bayes factor by the ratio of Cramér–Rao bounds to assure consistency.⁴ Since for stationary models condition 11 holds, we prefer to show the theorem under this condition.

⁴ The Cramér–Rao bound is directly related to the speed at which we learn about the parameter as the sample size grows. We can show, but do not include because space of considerations, that, if Condition 11 does not hold, then the result below changes to $\lim_{T \rightarrow \infty} P_{0T}(|\Sigma_T(Y^T|i)|^{1/2} f^T(Y^T|i) / |\Sigma_T(Y^T|j)|^{1/2} f^T(Y^T|j) = 0) = 1$.

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

Theorem 2. *Under Conditions 1–11, $\lim_{T \rightarrow \infty} P_{0T}(f^T(Y^T|i)/f^T(Y^T|j) = 0) = 1$.*

Proof. See Appendix. \square

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 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 linearization 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.

Even if we need to work with stochastically singular models (because for instance we are reluctant to artificially modify the theory with measurement errors), Bayesian methods allow for the recovery of useful information. We can always build likelihood functions with the same dimensionality than the sources of uncertainty in the model.

Those likelihoods may be very informative (Landon-Lane, 1999). Also with the likelihoods we can find posteriors for all the parameters of the model and perform point estimates as the matching of impulse-response functions of all the relevant variables of the model to shocks (beyond the ones used in the building of the likelihood), moment evaluation or spectral analysis. The results in the paper apply for this case and help to understand the properties of these exercises.

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 et al. (1994) and modified by Anderson et al. (1996).

Two reasons suggest the choice of this application. First we want to provide an “exhibit” of what one can learn and what can go wrong when applying the methods proposed in the paper to a simple yet rich and easily tractable dynamic equilibrium model.⁵ The cattle cycle has been chosen in the past by numerous authors as their favorite display for the purpose of illustrating how to estimate dynamic equilibrium models. Following their choice makes our paper comparable with the literature. Second, and as a consequence of the last point, a number of different estimation procedures have been performed with basically the same model and data. For instance, Rosen et al. (1994) mix calibration and ARMA estimation; Anderson et al. (1996) use Maximum Likelihood Methods; and Diebold et al. (1998) minimize the spectral distance between the data and the model. These procedures give us a benchmark to assess the performance of the Bayesian approach to model estimation and comparison. Given the previous findings we will know that any surprising or different result will come from our econometric approach and not from the model itself.

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 become part of the adult stock

⁵ We thank a referee for this point.

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 $x_t = (1 - \delta)x_{t-1} + gx_{t-3} - c_t$ and the total head count of cattle (the sum of adults, yearlings, and calves) is $s_t = x_t + gx_{t-1} + gx_{t-2}$.

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:

$$h_{t+1} = (1 - \rho_h)\mu_h + h_t + \varepsilon_{ht} \text{ where } \varepsilon_{ht} \sim N(0, \sigma_h^2) \tag{9}$$

$$m_{t+1} = (1 - \rho_m)\mu_m + m_t + \varepsilon_{mt} \text{ where } \varepsilon_{mt} \sim N(0, \sigma_m^2). \tag{10}$$

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

A representative farmer solves the following maximization problem:

$$\max_{\{c_t\}_{t=0}^{\infty}} E_0 \sum_{t=0}^{\infty} \beta^t \left\{ \begin{aligned} &(p_t - m_t)c_t - h_t x_t - \gamma_0 h_t g x_{t-1} - \gamma_1 h_t g x_{t-2} \\ &- \frac{\psi_1}{2} x_t^2 - \frac{\psi_2}{2} x_{t-1}^2 - \frac{\psi_3}{2} x_{t-2}^2 - \frac{\psi_4}{2} c_t^2 \end{aligned} \right\} \tag{11}$$

$$\text{s.t. } x_t = (1 - \delta)x_{t-1} + gx_{t-3} - c_t \tag{12}$$

$$\{x_{-1}, x_{-2}, x_{-3}\} \text{ fixed.} \tag{13}$$

To simplify we assume that the quadratic component of the cost is common for all activities $\varepsilon^2 = \psi_1/2 = \psi_2/2 = \psi_3/2 = \psi_4/2$ and trivially small.⁶

The model is closed with a demand function $c_t = \alpha_0 - \alpha_1 p_t + d_t$, where $\alpha_0, \alpha_1 > 0$ 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

$$\tilde{s}_t = s_t + \varepsilon_{st} \text{ where } \varepsilon_{st} \sim N(0, \sigma_s^2), \tag{14}$$

$$\tilde{c}_t = c_t + \varepsilon_{ct} \text{ where } \varepsilon_{ct} \sim N(0, \sigma_c^2). \tag{15}$$

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 $\{p_t\}_{t=0}^{\infty}$,

⁶ In this way and since we have (nearly) constant returns to scale and competitive markets, the model does not need to deal with the entry and exit of farmers in the market. In general models of an industry need to be careful regarding entry and exit in response to profits and losses à la Marshall (see [Veloce and Zellner, 1984, 1985](#)). We thanks the editor for pointing out this issue to us.

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:

$$\max_{\{c_t\}_{t=0}^\infty} E_0 \sum_{t=0}^\infty \beta^t \left\{ \begin{array}{l} (p_t - m_t)c_t - h_t x_t - \gamma_0 h_t g x_{t-1} - \gamma_1 h_t g x_{t-2} \\ - \varepsilon^2 x_t^2 - \varepsilon^2 x_{t-1}^2 - \varepsilon^2 x_{t-2}^2 - \varepsilon^2 c_t^2 \end{array} \right\} \quad (16)$$

$$\text{s.t. } x_t = (1 - \delta)x_{t-1} + g x_{t-3} - c_t \quad (17)$$

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

$$h_{t+1} = (1 - \rho_h)\mu_h + h_t + \varepsilon_{ht} \text{ where } \varepsilon_{ht} \sim N(0, \sigma_h^2), \quad (18)$$

$$m_{t+1} = (1 - \rho_m)\mu_m + m_t + \varepsilon_{mt} \text{ where } \varepsilon_{mt} \sim N(0, \sigma_m^2), \quad (19)$$

$$d_{t+1} = \rho_d d_t + \varepsilon_{dt} \text{ where } \varepsilon_{dt} \sim N(0, \sigma_d^2). \quad (20)$$

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, 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 $\theta = \{\beta, \delta, \alpha_0, \alpha_1, \gamma_0, \gamma_1, g, \rho_h, \rho_m, \rho_d, \mu_h, \mu_m, \sigma_h, \sigma_m, \sigma_s, \sigma_c, \sigma_d, \varepsilon\}$. 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 et al. (1996), it increases the comparability of our results to previous estimations. We will set $\beta = 0.96$, $\delta = 0$, $\varepsilon = 0.0001$, $\rho_d = \sigma_h = 0$, $\mu_h = 37$, 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,

Table 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

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 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 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 $\{\theta_1, \theta_2, \dots, \theta_m\}$ of parameter values. The empirical histograms of the parameters are included as Fig. 1.

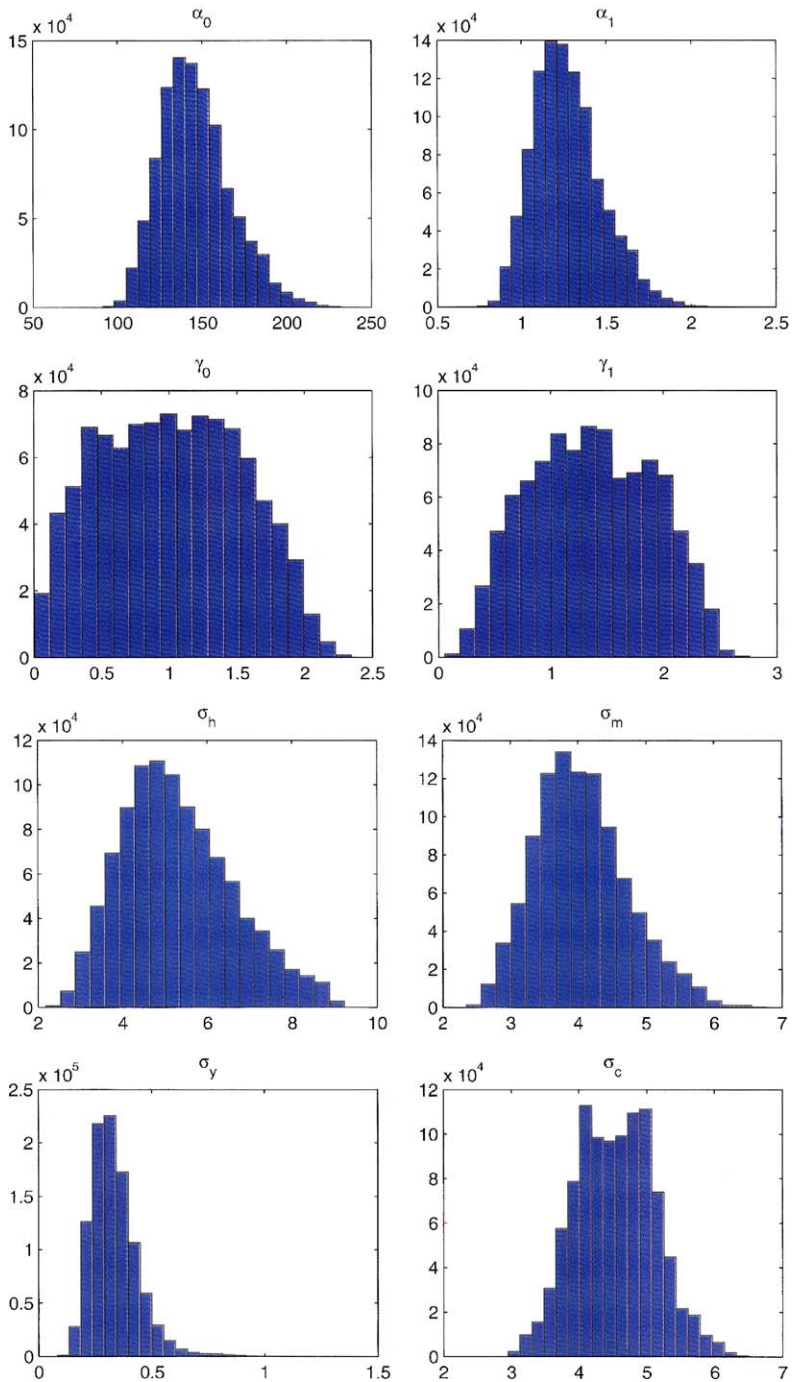


Fig. 1. Empirical distribution of the posterior.

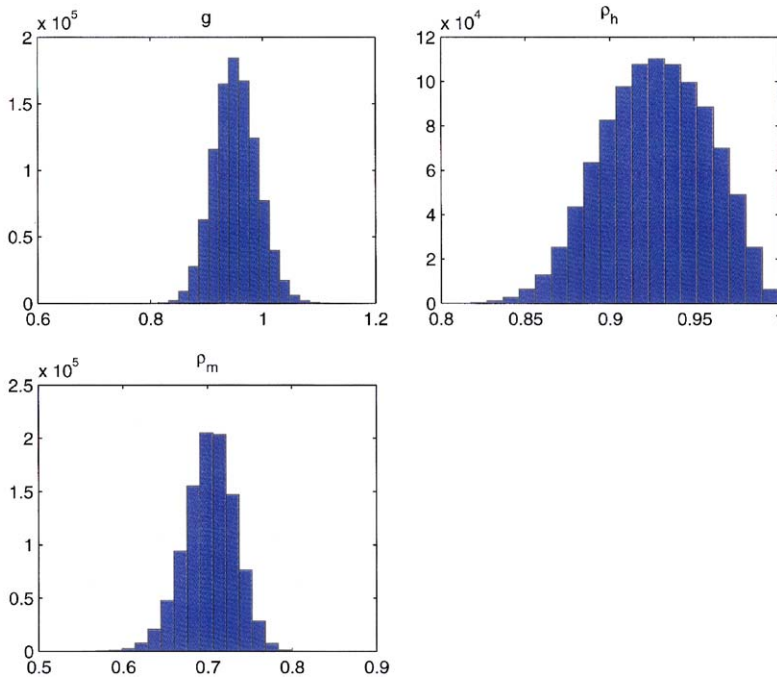


Fig. 1. *continued.*

Given this Markov chain and a function of interest $g(\cdot)$, its expectation $\mu = E(g(\theta))$ can be approximated using a strong law of large numbers by $\hat{\mu} = 1/m \sum_{i=1}^m g(\theta_i)$. Then, with 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 2 reports the expectation and standard deviation of the posterior of the parameters.⁷

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:

$$\begin{aligned}
 E \left[\frac{h(\theta)}{f^T(Y^T|\theta, i)\pi(\theta)} \middle| Y_T, i \right] &= \int_{\Theta} \frac{h(\theta)}{f^T(Y^T|\theta, i)\pi(\theta)} f^T(\theta|Y^T, i) d\theta \\
 &= \int_{\Theta} \frac{h(\theta)}{f^T(Y^T|\theta, i)\pi(\theta)} \frac{f^T(Y_T|\theta, i)\pi(\theta)}{\int_{\Theta} f^T(Y_T|\theta, i)\pi(\theta) d\theta} d\theta = \frac{\int_{\Theta} h(\theta) d\theta}{\int_{\Theta} f^T(Y^T|\theta, i)\pi(\theta) d\theta} \\
 &= f^T(Y^T|i)^{-1}.
 \end{aligned}
 \tag{21}$$

⁷ Under a quadratic loss function the expectation of the posterior is the point estimate. However more flexible loss functions are possible. For example an important literature in macroeconomics has focused on the ability of dynamic equilibrium models to match impulse response functions to shocks estimated from the data (Christiano et al., 2001). If this is the object of interest, the researcher can specify a loss function to obtain point estimates that minimize the distance between the estimated impulse response functions from the data and the impulse response functions generated by the model (Schorfheide, 2000).

Table 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

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

$$f^T(Y^T|i)^{-1} = \frac{1}{m} \sum_{i=1}^m \frac{h(\theta)}{f^T(Y^T|\theta, i)\pi(\theta)}. \tag{22}$$

As a choice of h we modify Geweke’s (1998) proposal. First, from the output of the simulation, define $\hat{\theta}_M = 1/m \sum_{i=1}^m \theta$ and $\hat{\Sigma}_m = 1/m \sum_{i=1}^m (\theta - \hat{\theta})(\theta - \hat{\theta})'$. Then, for a given $p \in (0, 1)$, define the set $\Theta_M = \{\theta: (\theta - \hat{\theta})\hat{\Sigma}_m^{-1}(\theta - \hat{\theta})' \leq \chi_{1-p}^2(11)\}$ where $\chi_{1-p}^2(\cdot)$ 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:

$$h(\theta) = \frac{1}{\hat{p}(2\pi)^{k/2} |\hat{\Sigma}_m|^{1/2}} \exp[-0.5(\theta - \hat{\theta})\hat{\Sigma}_m^{-1}(\theta - \hat{\theta})'] I_{\Theta_M \cap \Theta}(\theta), \tag{23}$$

where \hat{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 $f^T(Y^T|\theta, i)\pi(\theta)$ 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 $\sqrt{m}(\hat{\mu} - \mu) \xrightarrow{D} \mathcal{N}(0, \sigma_{\mu}^2)$, 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,⁸ we can estimate the NSE as $((1/m)\widehat{S}_g(0))^{1/2}$ (Hannan, 1970, Corollary 4, p. 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 et al., 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 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 $\hat{\mu}_1 = 1/p_A \sum_{i=1}^{p_A} g(\theta_i)$ and $\hat{\mu}_2 = 1/p_B \sum_{i=m-p_B+1}^m g(\theta_i)$. Then, under the null hypothesis that both means are equal, as $m \rightarrow \infty$ we know that

$$\frac{(\hat{\mu}_1 - \hat{\mu}_2)}{[\widehat{S}_g^A(0)/p_A + \widehat{S}_g^B(0)/p_B]^{1/2}} \Rightarrow \mathcal{N}(0, 1).$$

The computed values of the test for each first moment were all less than $|0.7 \times 10^{-4}|$, 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 (1999a). Given any prior density $\pi^*(\theta)$ with support included in our prior $\pi(\theta)$ support, we can define the

⁸ A sufficient condition for continuity is given by the strict stationarity of the simulation (Hannan, 1970, Corollary 1, p. 205) as is the case if the conditions for consistency of Section 2 hold. In practice, strict stationarity can be checked using standard tests.

Table 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

weighting function $w(\theta) = \pi^*(\theta)/\pi(\theta)$ and find the new posterior function of interest as $\hat{\mu} = 1/m[\sum_{i=1}^m w(\theta)g(\theta_i)/\sum_{i=1}^m w(\theta)]$.

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 $f^T(Y^T|\theta_i, i)\pi(\theta)$, and the prior values $\pi(\theta_i)$ $i = 1, \dots, m$, 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 et al. (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 3).

Comparison with Table 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. Fig. 1 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 et al. (1994) and Geweke et al. (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, linear statistical representations of the dynamic relations among variables, have been proposed as an alternative to structural modeling of time series that can provide a good forecasting performance (see Sims, 1980; Litterman, 1986).⁹ However VARs have problems of their own. Critics have pointed out their low observation/parameter ratio in comparison with alternatives like the structural-econometric-modeling-time-series approach (see Zellner, 1982), their inability to account for structural instability in time series relations (Stock and Watson, 1996) or their missing of nonlinearities in the data, specially around turning points (Kim and Nelson, 1988). As a consequence VARs may forecast poorly (Stock and Watson, 2001) and the outcome of the model comparison is not 100% assured.

We first describe the vector autoregression specifications, then the priors and finally the results of the models comparison. We finish discussing the issue of parameter stability and the estimation of a time-varying parameters model.

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:

$$y_t = \sum_{i=1}^p y_{t-i}A_i + C + u_t \quad \forall t \in \{1, \dots, T\}, \quad u_t \sim \mathcal{N}(0, \Psi) \tag{24}$$

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

⁹ Note that, however, these BVARs are not completely nonnested with the cattle cycle model since the latter has a restricted vector autoregression representation.

A useful way to rewrite (24) is as follows. Define $y_t = z_t\Gamma + u_t$ where $z_t = (I, y_{t-1}, \dots, y_{t-p})$ and $\Gamma = (C', A'_1, \dots, A'_p)'$. 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 i th column vector, we have $y_i = Z\gamma_i + u_i$. Stacking the column vectors $y_i, \gamma_i,$ and u_i in $y, \gamma,$ and $u,$ we get $y = (I \otimes Z)\gamma + u,$ where $u \sim \mathcal{N}(0, \Psi \otimes I)$. The likelihood function is

$$f^T(\gamma|\Psi) \propto |\Psi|^{-T/2} \exp\{-tr[(Y - Z\Gamma)' \Psi^{-1}(Y - Z\Gamma)]/2\} \tag{25}$$

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; Sims and Zha, 1998). In this way we can show the power of model comparison for different specifications of the models and assess how useful some of the recently proposed priors are for empirical analysis.

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 $\underline{\mu}_s = (0, \chi_{\{0\}}(s - 1), \chi_{\{0\}}(s - 2), \chi_{\{0\}}(s - 3), 0, \dots, 0)'$.
2. The variance of γ_s for $s \in \{1, 2, 3\}$ is equal to

$$\underline{\Sigma}_s = \begin{pmatrix} \pi(3)\sigma_s^2 & 0 & \dots & 0 \\ 0 & \tilde{\pi}_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \tilde{\pi}_{3p} \end{pmatrix} \tag{26}$$

where $\sigma_i = E(y_{i,t} - E y_{i,t})^2$ is a scale factor accounting for the variability of the different variables and $\tilde{\pi}_1 = \pi(2 - \chi_{\{0\}}(s - 1))\sigma_s^2/\sigma_1^2, \tilde{\pi}_2 = \pi(2 - \chi_{\{0\}}(s - 2))\sigma_s^2/\sigma_2^2, \tilde{\pi}_3 = \pi(2 - \chi_{\{0\}}(s - 3))\sigma_s^2/\sigma_3^2, \tilde{\pi}_i = 3\tilde{\pi}_1/[(i - 1)/3 + 1]$ for $i \in \{4, 7, \dots, 3p - 2\}, \tilde{\pi}_i = 3\tilde{\pi}_2/[(i - 2)/3 + 1]$ for $i \in \{5, 8, \dots, 3p - 1\}$ and $\tilde{\pi}_i = 3\tilde{\pi}_3/[(i - 3)/3 + 1]$ for $i \in \{6, 9, \dots, 3p\}$.

3. For $s \in \{1, 2, 3\}, \gamma_s \sim \mathcal{N}(\underline{\mu}_s, \underline{\Sigma}_s)$.
4. The variance-covariance matrix, $\Psi,$ is fixed and equal to the MLE.

Table 4
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

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 inverse-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. Formally, $\gamma | \Psi, \mu \sim N(\underline{\mu}, \Psi \otimes \underline{\Sigma})$, $\Psi \sim iW(\underline{\Psi}, \alpha)$, 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.¹⁰ We report our finding in Table 4.¹¹ 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

¹⁰ We can use these marginal likelihoods to combine models and obtain efficient forecasts. Zellner (1989) extends model combining to groups of models that are not an exhaustive set.

¹¹ 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.

bigger than 7, a bound for DNA testing in forensic science, often accepted by courts of law as evidence beyond reasonable doubt (Evetts, 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 be interested in.

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.¹² 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.

5.4. Structural stability and time-varying parameters models

In this subsection we study in more detail why the Cattle cycle model performs poorly in comparison with a BVAR with Minnesota prior. One possible reason could be that we have chosen bad parametric functions for the equilibrium model. Some exploration suggested to us that playing with these forms does not improve the performance of the model. A second possibility is that the lack of structural stability of the parameters burdens the estimation.

To explore that route we divided the sample into several intervals and repeated the estimation exercise for each of those. After trying different splitting strategies, our favorite choice was a division in two sub-samples: 1900:1946 and 1947:1990.¹³ As we will see below a time-varying parameters model also suggests a change occurs around that time.

Table 5 presents the means of the posterior for each parameter using the whole sample and only each of the two parts. Most parameters posterior means are stable across samples, within one standard deviation of the posterior of the whole sample. We only find an interesting difference in α_0 and α_1 where there is evidence that the mean demand for meat has decreased and has become less elastic over time. The γ_0 and γ_1 ’s, where we also see some differences, where the two parameters that were difficult

¹² 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.

¹³ Further results are available upon request. We do not include them because of space considerations.

Table 5
Means of the posterior

Parameter	1900–1991	1900–1946	1946–1991
α_0	146.23	186.53	130.81
α_1	1.27	1.66	1.10
γ_0	1.02	1.19	1.37
γ_1	1.36	1.13	1.32
g	0.95	0.92	0.99
ρ_h	0.93	0.89	0.94
ρ_m	0.70	0.64	0.74
σ_h	5.30	3.85	6.49
σ_m	4.05	3.17	4.33
σ_s	0.33	0.43	0.34
σ_c	4.54	4.30	4.32

to pin down in the benchmark estimation. Splitting the sample moves the mean but still within the main body of the (flat) posterior for the whole sample.

The increased flexibility of a smaller sample does not save the Cattle cycle: its marginal likelihood still lags behind the one for BVAR with Minnesota Prior (in log terms, -421.15 for Cattle Cycle versus -333.80 for Minnesota (2) in the 1900:46 sample and -470.46 versus -366.76 in the 1947:1990 sample). If anything the distance between the two models increases.

The evidence from the previous discussion leads us to estimate a more flexible model that can take account of the time-varying nature of the parameters. Again bayesian methods easily deal with the estimation problem. We follow Zellner et al. (1991) and Cogley and Sargent (2001) and estimate a BVAR with drifting parameters.¹⁴

Let a_i^m be the m th column of matrix A_i and c^m the m th column of a matrix C . We rewrite Eq. (24) for $p=1$ as $y_t' = I_{3 \times 3} \otimes [1 \ y_{t-1}]A + u_t'$ where $A = [c^1 a_1^1 c^2 a_1^2 c^3 a_1^3]'$.¹⁵ If we let A to be time-varying, we can write $A_t = [c_t^1 a_{1,t}^1 c_t^2 a_{1,t}^2 c_t^3 a_{1,t}^3]'$.

We treat the coefficients in A_t as a hidden state that evolves following a driftless random walk $A_t = A_{t-1} + v_t$ where v_t is an i.i.d. Gaussian process with mean 0 and covariance Q . Also $(u_t, v_t)'$ are i.i.d. Gaussian process with mean 0 and covariance matrix

$$E_t \left[\begin{pmatrix} u_t' \\ v_t \end{pmatrix} \begin{pmatrix} u_t & v_t' \end{pmatrix} \right] = V = \begin{pmatrix} \Psi & C' \\ C & Q \end{pmatrix}.$$

Our priors about the initial state A_0 and V are as follows. We assume that the initial state A_0 is distributed as a normal with mean \bar{A} and covariance \bar{P} while V is an inverse-Wishart $IW(\bar{V}^{-1}, T)$. Following Cogley and Sargent (2001) we set the

¹⁴ We cannot estimate a dynamic equilibrium model with time-varying parameters since such task involves in general non-linear cross-equation restrictions that require non-linear filtering methods not well understood. See Cogley and Sargent (2001) for an explanation.

¹⁵ Time-varying parameters models are richly parametrized. Higher order BVARs with only 273 data points perform worse than our specification.

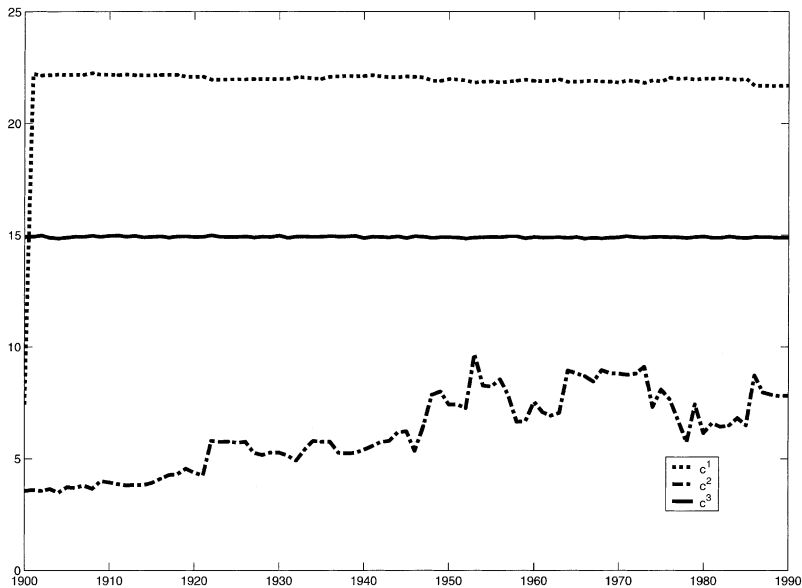


Fig. 2. Posterior mean of the time-varying parameters, constants.

prior mean of A_0 , \bar{A} , is equal to the MLE point estimates, \bar{P} to its asymptotic covariance matrix, $\bar{\Psi}$ equal to the estimated innovation covariance matrix, $\bar{C} = 0$ and $\bar{Q} = (0.01)^2 \bar{\Psi}$.¹⁶

Given $A^T = [A'_1, A'_2, \dots, A'_T]'$ our goal is to summarize the posterior density $\pi(A^T, V|Y^T)$. Following Zellner et al. (1991) we use the Gibbs sampler to simulate draws from $\pi(A^T, V|Y^T)$ in two steps. First, conditional on Y^T and V , we draw from $\pi(A^T|Y^T, V)$. Then, conditional on Y^T and A^T , we draw from $\pi(V|Y^T, A^T)$.

Figs. 2–5 report the posterior mean for A^T . We ordered the variables in the BVAR as stock of beef, slaughter rate and the price of beef. In Fig. 2 we see how the means of posteriors for the constant parameters are constant over time with the partial exception of some positive drift for the price. Figs. 3–5 show the autoregressives on each of the three variables. The most interesting finding is in Fig. 4 that shows how the autoregressive component of the slaughter rate on itself has a clear change around 1947 (right where we split our sample above) within a negative drift. A way to interpret this finding is to note that the number of sacrificed animals has increased relative to the stock of beef due to a move towards younger animals being slaughtered. In that context the previous slaughter rates are less and less important to account for the dynamics of the variable.

¹⁶ As sensitivity analysis we tried Litterman (1980) prior and stochastic volatility schemes (Cogley and Sargent, 2003) with limited successes.

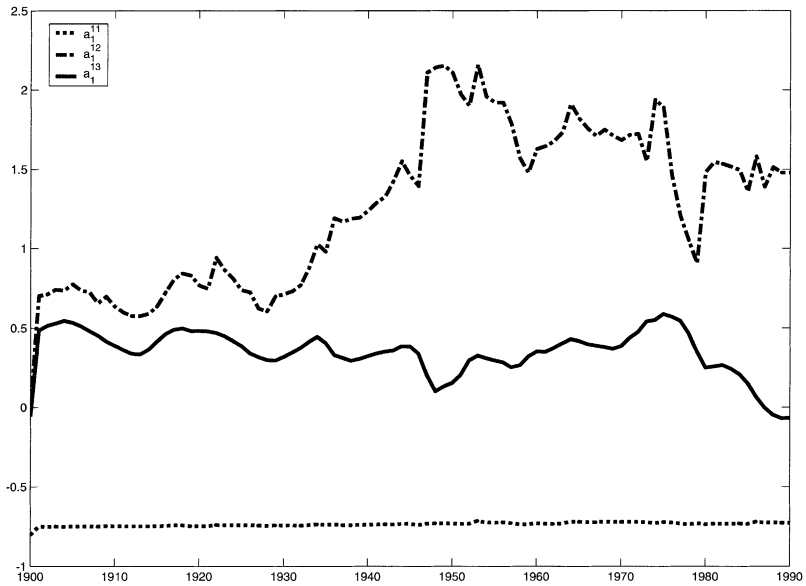


Fig. 3. Posterior mean of the time-varying parameters, stock of beef.

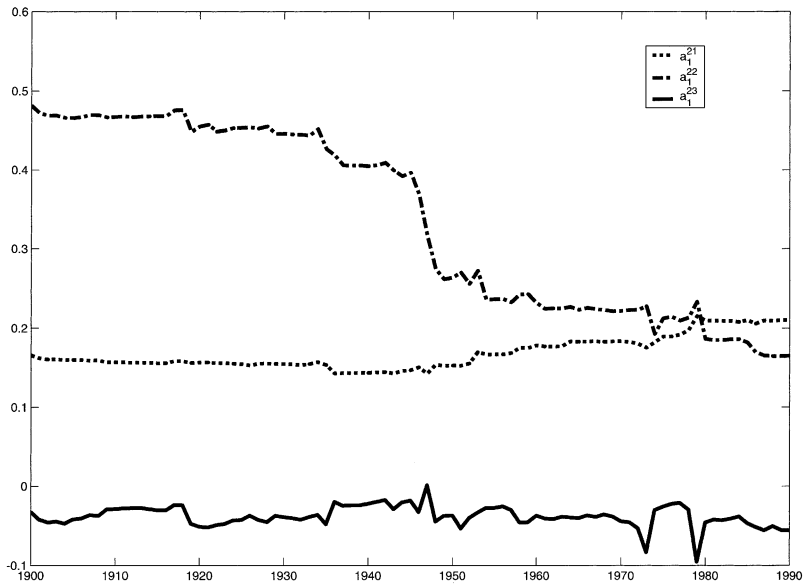


Fig. 4. Posterior mean of the time-varying parameters, slaughter rate.

The main conclusion from this part is that there is evidence of the changing parameters over time and that this may be an important reason behind the poor performance of the Cattle cycle model in comparison with a simple BVAR.

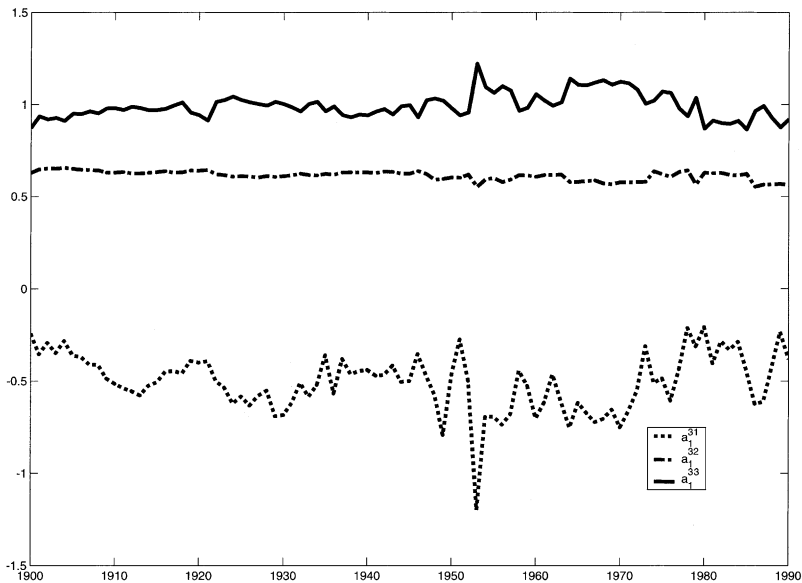


Fig. 5. Posterior mean of the time-varying parameters, price.

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 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.

Acknowledgements

Thanks to A. Atkeson, A. Bauer, M. Boldrin, J. Geweke, W. McCausland, E. McGrattan, L. Ohanian, T. Sargent, C. Sims, E. Tallman, H. Uhlig and T. Zha, the participants at several seminars, the Editor, the Associate Editor and the Referees for useful comments. Both authors thanks the Supercomputing Institute of the University of Minnesota. Juan Francisco Rubio-Ramírez acknowledge funding from the Bank of Spain. Beyond the usual disclaimer, we must note 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.

Appendix A.

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

A.1. Proofs

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

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

Then

$$\begin{aligned} &\int_{\theta \in \Theta_i \setminus \mathcal{E}(\hat{\theta}_T(i, Y^T), \delta_T(i))} \pi(\theta|Y^T, i) d\theta \\ &= f^T(Y^T, i)^{-1} f^T(Y^T|\hat{\theta}_T(i, Y^T), i) \exp[\log f^T(Y^T|\theta_T^*(i), i) \\ &\quad - \log f^T(Y^T|\hat{\theta}_T(i, Y^T), i)] \int_{\theta \in \Theta_i \setminus \mathcal{E}(\hat{\theta}_T(i, Y^T), \delta_T(i))} \pi(\theta, i) \exp[\log f^T(Y^T|\theta, i) \\ &\quad - \log f^T(Y^T|\theta_T^*(i), i)] d\theta \end{aligned} \tag{27}$$

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

With this last statement, we only need to check that

$$\begin{aligned} &f^T(Y^T, i)^{-1} \int_{\theta \in \Theta_i \setminus \mathcal{E}(\hat{\theta}_T(i, Y^T), \delta_T(i))} \pi(\theta, i) \exp[\log f^T(Y^T|\theta, i) \\ &\quad - \log f^T(Y^T|\theta_T^*(i), i)] d\theta \rightarrow 0 \end{aligned}$$

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

Then, by (8), for T large enough,

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

but (8) also implies that $\exp[-k_T(\delta_T, i)T] \rightarrow 0$ as $T \rightarrow \infty$ in P_0^T -probability and the results follow. \square

Proof of Lemma 2. By the application of the law of large numbers

$$\frac{1}{T} \log f^T(Y^T|\theta, i) \rightarrow \int_{R^m \times T} \frac{1}{T} \log f^T(Y^T|\theta, i) p_0^T(Y^T) dv^T$$

as $T \rightarrow \infty$ in P_0^T -probability. Lemma 2.2 in White (1980) proves that, if $\theta_T^*(i)$ is unique, $\hat{\theta}_T(i, Y^T) \rightarrow \theta_T^*(i)$ as $T \rightarrow \infty$ in P_0^T -probability. Since, $\theta_T^*(i) \rightarrow \theta^*(i)$ the lemma follows. \square

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

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

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

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

Note that we can write $f^T(Y^T|i)$ as

$$\begin{aligned} f^T(Y^T|i) &= \left(\int_{\Theta \setminus \mathcal{E}(\hat{\theta}_T(i, Y^T), \delta_T(i))} + \int_{\mathcal{E}(\hat{\theta}_T(i, Y^T), \delta_T(i))} \right) \pi(\theta|i) \exp(\log f^T(Y^T|\theta, i)) d\theta \\ &= I_{1,T} + I_{2,T}. \end{aligned}$$

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

Then

$$\begin{aligned} I_{2,T} &= \int_{\mathcal{E}(\hat{\theta}_T(i, Y^T), \delta_T(i))} \pi(\theta|i) \exp(\log f^T(Y^T|\theta, i)) d\theta \\ &= f^T(Y^T|\hat{\theta}_T(i, Y^T), i) \int_{\mathcal{E}(\hat{\theta}_T(i, Y^T), \delta_T(i))} \pi(\theta|i) \exp(\log f^T(Y^T|\theta, i) \\ &\quad - \log f^T(Y^T|\hat{\theta}_T(i, Y^T), i)) d\theta \\ &= f^T(Y^T|\hat{\theta}_T(i, Y^T), i) \int_{\mathcal{E}(\hat{\theta}_T(i, Y^T), \delta_T(i))} \pi(\theta|i) \exp(\Xi(\theta, Y^T)) d\theta \end{aligned}$$

where

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

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

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

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

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

Let $\bar{\vartheta}_T = \|\delta_T(i)\|_\infty \sqrt{(1 + \bar{\varrho}(\varepsilon_T))/\underline{\lambda}_T}$ and $\underline{\vartheta}_T = \|\delta_T(i)\|_\infty \sqrt{(1 - \bar{\varrho}(\varepsilon_T))/\underline{\lambda}_T}$, where $\bar{\varrho}(\varepsilon_T)$ and $\underline{\varrho}(\varepsilon_T)$ are the largest and the smallest eigenvalues of $A(\varepsilon_T)$ and $\bar{\lambda}_T$ and $\underline{\lambda}_T$ are the largest and smallest eigenvalues of Σ_T and note that

$$\{z; z'(I + A(\varepsilon_T))\Sigma_T^{-1}z < \underline{\vartheta}_T\} \subseteq \mathcal{E} \left(0, \frac{\underline{\vartheta}_T}{\sqrt{1 - \bar{\varrho}(\varepsilon_T)/\underline{\lambda}_T}} \right)$$

and

$$\mathcal{E} \left(0, \frac{\bar{\vartheta}_T}{\sqrt{(1 + \bar{\varrho}(\varepsilon_T))/\underline{\lambda}_T}} \right) \subseteq \{z; z'(I - A(\varepsilon_T))\Sigma_T(Y^T|i)^{-1}z < \bar{\vartheta}_T\}.$$

Thus,

$$\begin{aligned} & |I + A(\varepsilon_T)|^{-1/2} |\Sigma_T(Y^T|i)|^{1/2} \int_{\mathcal{E}(0, \underline{\vartheta}_T)} \exp\left(-\frac{1}{2} z'z\right) dz \\ & \leq \int_{\mathcal{E}(\hat{\theta}_T(i, Y^T), \delta_T(i))} \exp\left(-\frac{1}{2}(\theta - \hat{\theta}_T(i, Y^T))'(I + A(\varepsilon_T))\Sigma_T(Y^T|i)^{-1}\right. \\ & \quad \left. \times (\theta - \hat{\theta}_T(i, Y^T))\right) d\theta \leq I_{3,T} \\ & \leq \int_{\mathcal{E}(\hat{\theta}_T(i, Y^T), \delta_T(i))} \exp\left(-\frac{1}{2}(\theta - \hat{\theta}_T(i, Y^T))'(I - A(\varepsilon_T))\Sigma_T(Y^T|i)^{-1}\right. \\ & \quad \left. \times (\theta - \hat{\theta}_T(i, Y^T))\right) d\theta \\ & \leq |I - A(\varepsilon_T)|^{-1/2} |\Sigma_T(Y^T|i)|^{1/2} \int_{\mathcal{E}(0, \bar{\vartheta}_T)} \exp\left(-\frac{1}{2} z'z\right) dz \end{aligned}$$

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

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

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

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

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

as $T \rightarrow \infty$ in P_0^T -probability. \square

Proof of Theorem 2. From Lemma 3 we can write

$$\lim_{T \rightarrow \infty} P_{0T} \left(\frac{f^T(Y^T|i)}{f^T(Y^T|j)} = \frac{(2\pi)^{-k_i/2} |\Sigma_T(Y^T|i)|^{-1/2} f^T(Y^T|\theta_T^*(i), i)\pi(\theta^*(i)|i)}{(2\pi)^{-k_j/2} |\Sigma_T(Y^T|j)|^{-1/2} f^T(Y^T|\theta_T^*(j), j)\pi(\theta^*(j)|j)} \right) = 1. \tag{28}$$

Now, to prove that

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

and since

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

we only need to show

$$\lim_{T \rightarrow \infty} P_{0T} \left(\begin{aligned} & \frac{1}{T} \log(2\pi)^{-(k_i/2)} |\Sigma_T(Y^T|i)|^{-(1/2)} f^T(Y^T|\theta_T^*(i), i) \pi(\theta^*(i)|i) \\ & - \frac{1}{T} \log(2\pi)^{-(k_j/2)} |\Sigma_T(Y^T|j)|^{-(1/2)} f^T(Y^T|\theta_T^*(j), j) \pi(\theta^*(j)|j) \\ & = -\infty \end{aligned} \right) = 1. \tag{29}$$

Conditions (5) and (9) allow us to use an argument similar to Wald (1949) to prove (29) and the result from Lemma 3 to finish the proof. □

A.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 et al. (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 et al. (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 et al. (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

- Aguirre-Torres, V., Gallant, A.R., 2001. Testing separate dynamic nonlinear econometric models. In: Schueler, G.I., Spanos, P.D. (Eds.), *Monte Carlo Simulation*. Balkema, Rotterdam, pp. 423–430.
- Anderson, E.W., Hansen, L.P., McGrattan, E.R., Sargent, T.J., 1996. On the mechanics of forming and estimating dynamic linear economies. In: Amman, H.M., Kendrick, D.A., Rust, J. (Eds.), *Handbook of Computational Economics*, Vol. 1. Elsevier, Amsterdam.
- Andrews, D.W.K., 1988. Laws of large numbers for dependent non-identically distributed random variables. *Econometric Theory* 4, 458–467.
- Bureau of the Census, 1975. *Historical Statistics of the United States. Colonial Times to 1970*, U.S. Department of Commerce.
- Bureau of the Census, 1989. *Agricultural Statistics*. U.S. Department of Commerce.
- 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.
- Christiano, L., Eichenbaum, M., Evans, C., 2001. Nominal rigidities and the dynamic effects of a shock to monetary policy. Working paper 01–07, Federal Reserve Bank of Cleveland.
- Cogley, T., Sargent, T.J., 2001. Evolving post-world war II U.S. Inflation dynamics. In: Bernanke, B.S., Rogoff, K. (Eds.), *NBER Macroeconomics Annual*, Vol. 16. MIT Press, Cambridge, pp. 331–372.
- Cogley, T., Sargent, T.J., 2003. Drifts and volatilities: monetary policies and outcomes in the Post World War II U.S., Working paper, Arizona State University.
- 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.
- Davidson, J., 1994. *Stochastic Limit Theory*. Oxford University Press, Oxford.
- DeJong, D.N., Ingam, B.F., Whiteman, C.H., 2000. A Bayesian approach to dynamic macroeconomics. *Journal of Econometrics* 98, 203–223.
- Diebold, F.X., Ohanian, L.E., Berkowitz, J., 1998. Dynamic equilibrium economies: a framework for comparing model and data. *Review of Economic Studies* 65, 433–451.
- Evett, I.W., 1991. Implementing Bayesian methods in forensic science. Paper presented at the Fourth Valencia International Meeting on Bayesian Statistics.
- Fernández-Villaverde, J., Rubio-Ramírez, J.F., 2002. Estimating nonlinear dynamic equilibrium economies: a likelihood approach. Mimeo, University of Pennsylvania.
- Gelfand, A.E., Dey, D.K., 1994. Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society (Series B)* 56, 501–514.
- Gelman, A., Roberts, G.O., Gilks, W.R., 1996. Efficient metropolis jumping rules. In: Berger, J.O., Bernardo, J.M., David, A.P., Smith, A.F.M. (Eds.), *Bayesian Statistics*, Vol. 5. Oxford University Press, Oxford, pp. 599–607.
- Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Berger, J.O., Bernardo, J.M., David, A.P., Smith, A.F.M. (Eds.), *Bayesian Statistics*, Vol. 4. Oxford University Press, Oxford, pp. 169–193.
- Geweke, J., 1998. Using simulation methods for bayesian econometric models: inference, development and communication. Staff Report 249, Federal Reserve Bank of Minneapolis.
- Geweke, J., 1999a. Simulation methods for model criticism and robustness analysis. In: Berger, J.O., Bernardo, J.M., David, A.P., Smith, A.F.M. (Eds.), *Bayesian Statistics*, Vol. 6. Oxford University Press, Oxford, pp. 275–299.

- Geweke, J., 1999b. Computational experiments and reality. Mimeo, University of Iowa.
- Geweke, J.M., Keane, M., Runkle, D., 1997. Statistical inference in the multinomial multiperiod probit model. *Journal of Econometrics* 80, 125–165.
- Gourieroux, C., Monfort, A., 1998. Testing non-nested hypotheses. In: Engle, R.F., McFadden, D.L. (Eds.), *Handbook of Econometrics*, Vol. 4. Elsevier, Amsterdam, pp. 2583–2637.
- Hannan, E.J., 1970. *Multiple Time Series*. Wiley, Brisbane.
- Jacquier, E., Polson, N.G., Rossi, P., 1994. Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics* 12, 371–389.
- Jeffreys, H., 1961. *The Theory of Probability*, 3rd Edition. Oxford University Press, Oxford.
- Kadiyala, K.R., Karlsson, S., 1997. Numerical methods for estimation and inference in Bayesian VAR-models. *Journal of Applied Econometrics* 12, 99–132.
- Kass, R.E., Raftery, A.E., 1995. Bayes factors. *Journal of the American Statistical Association* 90, 773–795.
- Kim, J., 1998. Large sample properties of posterior densities, bayesian information criterion and the likelihood principle in nonstationary time series models. *Econometrica* 66, 359–380.
- Kim, C.J., Nelson, C.R., 1988. Business cycle turning points: a new coincident index and test of duration dependence based on a dynamic factor model with regime switching. *Review of Economics and Statistics* 80, 188–201.
- Kitamura, Y., 1998. Comparing misspecified dynamic econometric models using nonparametric likelihood. Mimeo, University of Wisconsin-Madison.
- Landon-Lane, J., 1999. Bayesian comparison of dynamic macroeconomic models. Ph.D. Thesis, University of Minnesota.
- Leeper, E.M., Sims, C.A., 1994. Toward a modern macroeconomic model usable for policy analysis. In: Stanley, F., Rotemberg, J.J. (Eds.), *NBER Macroeconomics Annual*, Vol. 9. MIT Press, Cambridge, pp. 81–118.
- Litterman, R.B., 1980. A bayesian procedure for forecasting with vector autoregressions. Mimeo, University of Minnesota.
- Litterman, R.B., 1986. Forecasting with Bayesian vector autoregressions-five years of experience. *Journal of Business and Economic Statistics* 4, 1–4.
- Mengersen, K.L., Robert, C.P., Guihenneuc-Jouyaux, C., 1999. MCMC convergence diagnostics: a 'reviewww'. In: Berger, J.O., Bernardo, J.M., David, A.P., Smith, A.F.M. (Eds.), *Bayesian Statistics*, Vol. 6. Oxford University Press, Oxford, pp. 415–440.
- Otrok, C., 2001. On measuring the welfare cost of business cycles. *Journal of Monetary Economics* 47, 61–92.
- Phillips, P.C.B., 1996. Econometric model determination. *Econometrica* 64, 763–812.
- Phillips, P.C.B., Ploberger, W., 1996. An asymptotic theory of Bayesian inference for time series. *Econometrica* 64, 381–412.
- Rissanen, J., 1986. Stochastic complexity and modeling. *The Annals of Statistics* 14, 1080–1100.
- Rosen, S., Murphy, K.M., Scheinkman, J.A., 1994. Cattle cycle. *Journal of Political Economy* 102, 468–492.
- Sims, C.A., 1980. Macroeconomics and reality. *Econometrica* 48, 1–48.
- Sims, C.A., Uhlig, H., 1991. Understanding unit rooters: a helicopter tour. *Econometrica* 59, 1591–1599.
- Sims, C.A., Zha, T., 1998. Bayesian methods for dynamic multivariate models. *International Economic Review* 39, 949–968.
- Schorfheide, F., 2000. Loss function-based evaluation of DGSE models. *Journal of Applied Econometrics* 15, 645–670.
- Shore, J.E., 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.
- Stock, J.H., Watson, M.W., 1996. Evidence on structural instability in macroeconomic time series relations. *Journal of Business Economics and Statistics* 14, 11–30.
- Stock, J.H., Watson, M.W., 2001. Macroeconomic forecasting using many predictors. Mimeo, Princeton University.
- Veloce, W., Zellner, A., 1984. Modeling a competitive industry with entry. *Economic Letters* 16, 71–75.
- Veloce, W., Zellner, A., 1985. Entry and empirical demand and supply analysis for competitive industries. *Journal of Econometrics* 30, 459–471.

- Vuong, Q.H., 1989. Likelihood ratio test for model selection and non-nested hypotheses. *Econometrica* 57, 307–333.
- Wald, A., 1949. Note on the consistency of the maximum likelihood estimate. *Annales of Mathematical Statistics* 20, 595–601.
- White, H., 1980. Nonlinear regression on cross-section data. *Econometrica* 48, 721–746.
- White, H., 1994. *Estimate, Inference and Specification Analysis*. Cambridge University Press, Cambridge.
- Zellner, A., 1982. Basic issues in econometrics, past and present. *The American Economist* 26, 5–10.
- Zellner, A., 1988. Optimal information processing and Bayes' theorem. *American Statistician* 42, 278–284.
- Zellner, A., 1989. Bayesian and non-Bayesian methods to combining models and forecast. Mimeo, University of Chicago.
- Zellner, A., Hong, C., Min, C.K., 1991. Forecasting turning points in international output growth rates using Bayesian exponentially weighted autoregression. Time-varying parameter, and pooling techniques. *Journal of Econometrics* 49, 275–304.