Landon-Lane, John S.

Working Paper

Evaluating dynamic stochastic general equilibrium models using likelihood methods

Working Paper, No. 2002-11

Provided in Cooperation with:
Department of Economics, Rutgers University

Suggested Citation: Landon-Lane, John S. (2002) : Evaluating dynamic stochastic general equilibrium models using likelihood methods, Working Paper, No. 2002-11, Rutgers University, Department of Economics, New Brunswick, NJ

This Version is available at:
https://hdl.handle.net/10419/79169

Evaluating Dynamic Stochastic General Equilibrium Models using Likelihood Methods

John Landon-Lane *

Rutgers University

First version: October, 2000

This Version: June, 2002

*Contact Information: Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA. E-mail: lane@econ.rutgers.edu . The author would like to thank John Geweke, Michael Keane, Lee Ohanian and Yuichi Kitamura for many constructive comments. All remaining errors are my own.

Abstract

This paper develops a method that uses a likelihood approach to directly compare two or more non-nested dynamic, stochastic general equilibrium (DSGE) models. It is shown how DSGE models can be compared across the whole sample and how this measure can be decomposed across individual observations thus allowing models to be compared across any sub-sample of the data. The method is applied to the problem of determining whether the technology shock process in a standard Real Business Cycle model should consist of permanent or temporary, albeit persistent, shocks. Overall, a permanent shock model has a better prediction performance than the temporary shock model. However, the model with the temporary shock performs much better for the part of the sample that includes the most of the 1980's and the 1990's.

J.E.L. Classification C11, C52, E32

Keywords: Real Business Cycles, Model Evaluation, Markov chain Monte Carlo, Bayes Factor.

1 Introduction

Dynamic, stochastic general equilibrium (DSGE) models are widely used in the macroeconomic literature, especially in the Real Business Cycle (RBC) literature. Beginning with the work of Long and Plosser (1983), Kydland and Prescott (1982), and Hansen (1985) it is now common to investigate the business cycle using general equilibrium models in which cycles occur through individual agents making optimal decisions in the face random fluctuations. The development of the literature has been to develop models that aim to model an economy; to use these models to ask and answer questions regarding observed economic phenomena; and to use these models to conduct experiments on how the economy reacts to various changes to the characteristics of these economies (Kydland and Prescott 1996).

The methods that have been used to validate the use of the models in the RBC literature have come under increasing criticism. The most common method of determining whether a model does a "good" job of mimicking an observed economy has been criticized for being too informal and for not being likelihood based (Hansen and Heckman 1996), and for not directly comparing models in the RBC literature (Stadler 1994) . Various alternative methods have been proposed to evaluate the ability of models with respect to observations on the economy (for example DeJong, Ingram and Whiteman (1996), Diebold, Ohanian and Berkowitz (1998), and Cogley and Nason (1994)). However, these alternative methods evaluate the performance of a model with respect to the observed data rather than evaluating a model with respect to other competing models in the literature.

Kydland and Prescott (1996) attempt to deflect this criticism by arguing that the models used in the RBC literature should not be expected to predict the observed economy that well, which implies that standard statistical techniques would almost always reject the models used. This could lead to the rejection of models that could nevertheless lead

to increased understanding of the observations at hand. Their view is that, while the models are not good at predicting observations, the models used are useful in helping to understand the observations from the economy. If this is the case, then it would be important to be able to distinguish which model of the type used in the RBC literature is the "best". Therefore, a method that is able to distinguish among the class of models that is used in the RBC literature is needed.

This paper develops a method that is able to directly compare the types of competing DSGE models used in the RBC literature. Importantly, the method is likelihood based, thus evaluating, potentially non-nested, models over the full dimension of the data. One of the issues with working with the DSGE models that are common in the RBC literature is that there are usually less shocks in the model than there are variables that are being modelled. For example, it is common for an RBC model, such as the one described in Hansen (1985), to try to explain observed fluctuations through only a technology shock. This has implications for the likelihood function of such a model. The assumption of less shocks than variables implies that the model predicts that some of the variables of the model are related to each other by a deterministic function. Therefore, if this deterministic function is rejected by the data the likelihood of this model would be zero. In this case any attempt to estimate the parameters of the model by maximum likelihood would fail.

A number of papers, such as Sargent (1989), Anderson, Hansen, McGratten and Sargent (1996), and Ireland (1999), have attempted to solve this problem by assuming that the variables of the model are measured with error thus increasing the number of stochastic terms in the model. Other papers, such as DeJong, Ingram and Whiteman (2000) and Landon-Lane (1998) construct likelihoods based on a subset of the data. This paper uses principle component analysis to construct a likelihood function for the model, using information from all of the variables that are modelled, without adding any extra stochastic

terms. However, it should be noted that the method of model comparison that is used in this paper is not dependent on the method by which the likelihood is constructed.

Sims (1996) argues that the appropriate way to evaluate or compare models is via Bayesian methods (for example, DeJong et al. (1996)) as Bayesian methods treat parameter and model uncertainty in the same manner. In this paper, the comparison of two of more DSGE models in this paper are undertaken using Bayesian model comparison methods described in Geweke (1999). In particular, this method of model comparison allows for the comparison of models over subsets of the data.

The layout of the paper is as follows: Section 2 describes how the likelihood function is calculated, Section 3 describes the model estimation and model comparison techniques used for this paper, Section 4 contains an application of this method to the problem of determining the appropriate error structure for the technology shock process in the standard RBC model and finally Section 5 concludes.

2 Calculating the Likelihood Function of an RBC Model

In order to be able to apply the Bayesian model comparison literature to the problem of directly comparing two or more models from the RBC literature, likelihood functions for the relevant models are needed. There have been a number of attempts in the literature to construct an approximate likelihood function for models in the RBC literature.

One such attempt is the method of Smith (1993), where the data that is generated from a model is represented as a VAR. The likelihood function of the VAR is constructed and this likelihood function is used to approximate the likelihood function of the model. The parameters of the model are then estimated using simulated maximum likelihood

methods.

However, Kim and Pagan (1995) show that if an RBC model is completely linearized, that is the approximate solution to the model is linear, then the model can be represented as a vector-autoregression (VAR). However, as there are fewer stochastic elements than the number of variables in the model, the implied variance-covariance matrix of the VAR is singular and so the likelihood function for the VAR does not exist. Therefore, only subsets of the data can be used in the calculation of the likelihood function. However, this raises the question of which subsets to use in the calculation of the likelihood.

Beginning with Sargent (1989), full dimensional likelihood functions for RBC models have been constructed by adding extra stochastic terms to the system. Typically the extra stochastic components are in the form of data measurement errors[1]. However, if you assume that all variables are measured with error you have the case where the extended model has more stochastic terms than variables. In this case, without added restrictions being placed on the stochastic terms, it would be difficult to identify the individual stochastic terms of the original model. Therefore, as in Anderson et al. (1996), only a subset of the variables are assumed to be measured with error so that the total number of stochastic elements in the extended model are equal to the number of variables being modelled. This immediately raises the question of which variables to assume are measured with error.

A common approach in the calculation of the likelihood function for an RBC model in the literature is to utilize the state-space nature of the problem. Once in state-space form the innovations representation can be calculated using the Kalman[2] filter. This is the procedure used by authors such as Anderson et al. (1996), DeJong et al. (2000), and Ireland (1999). Note that DeJong et al. (2000) do not add measurement error shocks to the basic model but rather use sub-samples of the data to calculate the likelihood using

[1] see, for example, Anderson et al. (1996) and Ireland (1999)
[2] see Harvey (1989)

the Kalman Filter.

All of the above approaches are valid approaches to calculating a likelihood function for a DSGE model typically used in the RBC literature. The state-space representation of a DSGE model is also utilized in this paper. It is the case that almost all DSGE models used in the RBC literature do not have a closed form solution. A number of approaches have been suggested to find approximate solutions to the model as reported in Taylor and Uhlig (1990). Anderson et al. (1996) describe a method within a framework of dynamic linear-quadratic economies. The general setup is as follows:

Let $D_T = \{d_t\}_{t-1}^T$ be a series of $n \times 1$ data vectors that are to be modelled by a DSGE model and let ψ be an $m \times 1$ vector of structural parameters of the model. Let $S_T = \{s_t\}_{t=1}^T$ be a series of $n_s \times 1$ state vectors of the DSGE model that includes, among other variables, the shock terms of the model. Then an approximate solution to the model, using any of the methods described in Taylor and Uhlig (1990), in state-space form is

$$\begin{aligned} s_t &= f(s_{t-1}, \eta_t; \psi) \;\; \forall t = 1, \ldots T \quad s_0 \text{ given} \\ d_t &= g(s_t; \psi), \end{aligned} \tag{1}$$

where $f : \mathbf{R}^{n_s} \to \mathbf{R}^{n_s}$ and $g : \mathbf{R}^{n_s} \to \mathbf{R}^n$ are vector valued functions and η_t is a $n_s \times 1$ vector of innovations to the state variable. Typically the shock process for the RBC model is defined as a state variable of the model and so η_t includes the innovation to the shock process. Also, some of the elements of η_t may be zero. Combining the two equations in (1) gives

$$d_t = h(s_{t-1}, \eta_t; \psi), \tag{2}$$

where $h : \mathbf{R}_s^n \to \mathbf{R}^n$ and is given by $h = f \circ g$.

It is common in the literature to completely linearize the approximate solution to the

model in which case (1) becomes

$$\begin{aligned} s_t &= \mathbf{A}(\psi)s_{t-1} + \eta_t \;\; \forall t = 1, \ldots, T \;\; s_0 \text{ given} \\ d_t &= \mathbf{B}(\psi)s_t, \end{aligned} \tag{3}$$

where $\mathbf{A}$ and $\mathbf{B}$ are $(n_s \times n_s)$ and $(n \times n_s)$ dimensional matrices respectively.

Now suppose that there are n_e non-zero elements of η_t where $0 < n_b < n$. This is the case for a large number of models that are found in the RBC literature with the modal value of n_b being one. In this case the model is singular and the value of the shock cannot be uniquely determined from the observations D_T. As already noted, one method to solve this problem is to add enough extra stochastic terms so that the model is no longer singular. There are two related issues with this approach. First, there is no unique way to add extra shocks to the model. If one wanted to compare the performance of two competing models then this is an issue. One would have to compare the models across a number of possible extensions in order to be sure that the comparison is independent of the way in which the extra stochastic terms are added. The second issue is whether the added shocks would disguise some of the characteristics of the original model, therefore weakening the result of the comparison.

Therefore, in the context of directly comparing DSGE models with less shocks than variables, this paper constructs a likelihood function without adding extra stochastic components. There are two approaches to constructing the likelihood function with the addition of extra stochastic terms. The first is to use only a subset of n_b elements of d_t in the construction of the likelihood function. This is the approached used in DeJong et al. (2000) and Landon-Lane (1998).

Another way to handle the singularity problem is to use n_b linear combinations of the data, D_T in constructing the likelihood function. The obvious way to construct independent linear combinations of the data, D_T, is to use principle components analysis

(Johnson and Wichern 1988). The obvious choice for the n_b independent linear combinations are the n_b principle components associated with the n_b highest eigenvalues of the covariance matrix of the data. Let $\mathbf{P}$ be the $n \times n_b$ matrix of these principle components. Define $x_t = \mathbf{P}'d_t$ to be a $n_b \times 1$ vector of linear combinations of the data, y. Then it follows from (2) that

$$x_t = \mathbf{P}'h(s_{t-1}, \eta_t; \psi) \;\; \forall t = 1, \ldots, T, \;\; s_0 \text{ given.} \tag{4}$$

Thus we now have a, possibly non-linear, system relating the n_b stochastic terms of the model to the n_b most informative independent linear combinations of the data. Using (4) it is possible to uniquely determine the values of the stochastic elements of the model using information from all of the components of d_t. Therefore, conditional on the initial value of the state variable, s_0, it is possible to iteratively solve for the non-zero values of η_t for each $t = 1, \ldots, T$ implied by the observations D_T. The iterative procedure is as follows: For each $t = 1, \ldots, T$, given s_{t-1} and d_t, solve (4) for η_t noting that s_{t+1} is known once we know η_t though the first equation of (1). Let ω_t be the $n_b \times 1$ vector of the non-zero elements of η_t and let $p_\omega(\Omega_T)$ be the joint density function of $\Omega_T = \{\omega_t\}_{t=1}^T$. Then the likelihood function, or data density for the model defined in (1) is

$$p(D_T|\psi, s_0) = p_\omega(\hat{\Omega}_T(D_T))J \tag{5}$$

where J is

$$J = \left| \left(\frac{\partial(d_1, \ldots, d_T)}{\partial(s_0, \omega_1, \ldots, \omega_T)'} \right)^{-1} \right| \tag{6}$$

and $\hat{\Omega}_T(D_T)$ are the values of ω_t that are the solution to (4).

3 Estimating and Comparing RBC models

The purpose of this paper is to answer the criticisms of Hansen and Heckman (1996) and Stadler (1994) and develop a likelihood based method for directly comparing models found in the RBC literature. Canova and Ortega (1995) discuss various approaches to evaluating RBC models but only Ortega (1995) contains a method that allows for the direct comparison of two or more RBC models.

The standard practice in the literature is to evaluate the model by defining a metric and measuring the distance between the "model" and the observed data. Canova and Ortega (1995) discuss various approaches to the definition of the metric and the measurement of the distance between model and data. They make the distinction between formal and informal approaches. The standard approach in the literature, for example Hansen (1985), of measuring a model's performance by determining how well a model can match certain observed moments in the data falls into the second category. This approach to evaluating a model's performance was the approach that was criticised in Hansen and Heckman (1996) for being ad hoc and not evaluating the models over the full dimension of the data. Another example of this informal approach can be found in Farmer and Guo (1994) where they partly evaluate models using their implied impulse response functions. One common attribute of the methods that can be categorized in the informal category is that there is not a formal discussion of the metric used to measure the model's performance.

A number of authors have attempted to formalise the metric by which to measure the "distance" between the model and the observed data. Watson (1993) uses the autocovariance functions of the observed data and data generated from the models to define a metric, while Christiano and Eichenbaum (1992) use a variant of the J-test of Hansen (1982). Another approach is to use the spectral density to define a metric between the model and the data. Examples of this approach can be found in Diebold et al. (1998)

and Ortega (1995).

The method of model comparison used in this paper is the method of Bayesian model comparison described in Geweke (1995) and Geweke (1999). A Bayesian approach allows for the formal treatment of model uncertainty and prior beliefs. Suppose, for example, that two models are to be compared, one of which is a fully developed model from theory while the other is an econometric representation of the data with little theoretical underpinnings. A priori, one might believe that the model that was developed from theory is the better model. It is a simple procedure to explicitly include these prior beliefs using Bayesian methods. Sims (1996) argues that Bayesian methods are the most appropriate methods for the problem of comparing and evaluating models in the RBC literature.

Another reason for using Bayesian methods is that there is a fair degree of prior information used in the RBC literature. Models are calibrated to certain values based on the prior beliefs of the authors. Bayesian methods are the natural way to formally incorporate prior information with information from the observed data in the estimation of the models.

3.1 Model Comparison

The method of comparison is as follows: Let m_k be any model contained in a finite set of models $\mathcal{M}$ indexed by k that aim to model the observed data D_T. Let $\phi_k \in \Phi_k$ represent the unknown parameters of the data density $p(D_T|\phi_k, m_k)$ for model m_k, and let $p(\phi_k|m_k)$ be the joint prior distribution over ϕ_k. Then, according to Bayes' Theorem[3], the posterior distribution of ϕ_k is

$$p(\phi_k|D_T, m_k) \propto p(\phi_k|m_k)p(D_T|\phi_k, m_k). \tag{7}$$

[3]see, for example, Bernado and Smith (1994)

Then the marginal likelihood is defined as

$$p(D_T|m_k) = \int_{\Phi_k} p(\phi_k|m_k)p(D_T|\phi_k, m_k)d\phi_k \tag{8}$$

and is interpreted as the probability of observing D_T conditional on the model m_k.

The marginal likelihood is the object that will be used for model comparison. For any two models m_i and m_j in $\mathcal{M}$, the Bayes Factor in favor of model m_i over model m_j is defined to be the ratio of the marginal likelihoods of the respective models given by

$$B_{ij} = \frac{p(D_T|m_i)}{p(D_T|m_j)}. \tag{9}$$

If $B_{ij} > 1$ it is clear that model m_i is favored as this implies that $p(D_T|m_i) > p(D_T|m_j)$. Clearly, if a group of models are being compared then m_i, for some i, is the most favored model if $B_{ij} > 1$ for all $j \neq i$, $m_j \in \mathcal{M}$. Another way of stating this is to say that the most preferred model is the one that has the highest marginal likelihood.

Prior beliefs over the relative strengths of the set of models under consideration can be explicitly introduced using the posterior odds ratio defined as

$$POR_{ij} = \frac{p_i}{p_j} B_{ij} \tag{10}$$

where p_i and p_j are subjective prior probabilities for each model. Thus one way to interpret the Bayes Factor is that it reflects how much more prior probability has to placed on a model in order for it to have a higher posterior odds ratio than another model. For example, if $B_{ij} = 4$ then $p_j > 4p_i$ for model m_j to have a higher posterior odds ratio.

Geweke (1995) describes how the Bayes Factor can be decomposed across sub-samples of

the data. That is

$$B_{ij,1}^{T} = \prod_{t=1}^{T} B_{ij,d_{t-1}}^{d_t} \tag{11}$$

where

$$B_{ij,u}^{v} = \frac{p((d_{u+1}, \ldots, d_v)'|m_i)}{p((d_{u+1}, \ldots, d_v)'|m_j)} \tag{12}$$

is the ratio of marginal likelihoods for the observations $d_{u+1}, \ldots, d_v$. A full derivation of (11) can be found in Section A.1 of the Appendix.

Observations or periods of observations that make large contributions to the overall Bayes factor in favor of model i over model j can be identified using (11). Unusually low values of the marginal likelihood for an observation or for a group of observations would indicate that the model did not do a good job of predicting that particular observation or group of observations. By breaking up the Bayes factor up into a product of intermediate Bayes factors, as in (11), we are able to see what observations or group of observations have the biggest contribution to the overall Bayes factor. There may be observations that are surprising to both models, but the decomposition allows us to see which model handles the surprising event the best. The question of how models handle a rare but potentially important event could be extremely useful in the evaluation of that model. Also, it may be the case that two competing models have different strengths and therefore perform better in different sub-samples of the data. It would be useful to know this as you would not want to discard a model if it is able to explain some characteristics of the data, even if the majority of the observations are better explained by another model.

3.2 Estimating RBC Models and Calculating the Bayes Factor

Estimation of a typical DSGE model found in the RBC literature will revolve around the posterior distribution defined in (7) where the likelihood function for the model is calculated using the method described in Section 2. Using the notation developed in

Section 2, the posterior distribution for a model m_j is

$$p(\psi, s_0|D_T, m_j) \propto p(\psi, s_0|m_j)p(D_T|\psi, s_0, m_j). \tag{13}$$

where $p(\psi, s_0|m_j)$ is the joint prior placed over the structural parameters, ψ, and the initial state variables, s_0 conditional on model m_j.

The object is to make inferences about the marginal distribution, $p(\psi|D_T, m_j)$. To do that we first obtain a series of N draws, $\{\phi^1, \ldots, \phi^N)\}$ from (13) where $\phi = (\psi, s_0)'$. Once we have these draws we also have draws from the marginal distribution $p(\psi|D_T, m_j)$, namely $\{\psi^1, \ldots, \psi^N\}$.

There are a number of ways to make draws from (13)[4] using Markov chain Monte Carlo (MCMC) techniques. The nature of the way in which the likelihood function is calculated leads to the use of the random-walk Metropolis-Hastings (RWMH) algorithm in making draws from (13). This algorithm only requires that (13) is able to be evaluated at any point in the parameter space $\Phi = \Psi \cup \mathcal{S}_0$. Other MCMC methods require knowledge of the exact distribution of (13) or at least knowledge of the conditional distributions of (13). This is difficult in this case as there is a complicated non-linear relationship between (13) and the parameters, $\phi = (\psi, s_0)'$, through the approximate solution to the model given in (1). However, given a value of ϕ it is simple to calculate (1) and in turn easy to calculate the value for (13) using (1).

The RWMH algorithm for this problem is as follows: Choose an initial draw from (13), $\phi^1 \in \Phi$. The easiest way to guarantee that the initial draw is in the domain of the posterior is to pick a value from the prior, $p(\psi, s_0|m_j)$. Then, for each $l = 1, \ldots, N$ calculate a candidate draw, ϕ^c, from (13) by setting $\phi^c = \phi^{l-1} + \nu$ where ν is drawn from a fixed distribution with mean zero and constant variance-covariance matrix $\mathbf{V}$. In practice drawing ν from a multivariate Gaussian distribution with mean $\mathbf{0}$ and variance-

[4]see Geweke (1999) for an overview of the various sampling methods employed.

covariance matrix $\mathbf{V}$ is usually successful. Tierney (1994) describes the exact properties of the RWMH algorithm and in particular shows that the candidate draw, ϕ^c should be accepted with probability $\alpha(\phi^c, \phi^{l-1}|m_j)$ where

$$\alpha(\phi^c, \phi^{l-1}|m_j) = \min\left\{\frac{p(\phi^c \mid D_T, m_j)}{p(\phi^{l-1} \mid D_T, m_j)}, 1\right\}. \tag{14}$$

Thus, any candidate draw for which $p(\phi^c|D_T, m_j) \geq p(\phi^{l-1}|D_T, m_j)$ is accepted with probability one and any candidate draw with $p(\phi^c|D_T, m_j) < p(\phi^{l-1}|D_T, m_j)$ is accepted with probability $p(\phi^c \mid D_T, m_j)/p(\phi^{l-1} \mid D_T, m_j)$. Therefore, for each $l = 1, \ldots, N$

$$\phi^l = \begin{cases} \phi^{l-1} + \nu & \text{with probability } \alpha(\phi^c, \phi^{l-1}) \\ \phi^{l-1} & \text{else} \end{cases}. \tag{15}$$

The variance-covariance matrix $\mathbf{V}$ is chosen to tune the algorithm so that the draws, $\{\phi^1, \ldots, \phi^N\}$, from (13) have desirable time series properties[5].

Once draws are obtained from (13) an implementation of the method of Gelfand and Dey (1994) as suggested by Geweke (1999) is used to calculate the value of the marginal likelihood, $p(D_T|m_j)$, for each model.

4 Determining the Error Structure for the Shock Process of an RBC Model

A large part of the RBC literature has attempted to explain observed aggregate fluctuations via a single technology shock. There have been two approaches to how the technology shock is modelled in the literature. The first approach is to model the technology

[5]see Geweke (1999) and Tierney (1994) for a description of the desirable properties of the output from a MCMC algorithm.

shock process as a stationary AR(1) process with a high degree of persistence. This is the approach in papers such as Kydland and Prescott (1982) and Hansen (1985). Another approach is to assume that the technology process contains a unit root as in King, Plosser, Stock and Watson (1991). Hansen (1997) performs a comparison of variants of a standard DSGE model with a number of different assumptions on the technology shock process. The comparison is undertaken using the informal approach that is common in the literature of matching second moments of the data.

The aim of this section is to apply the method described in Section 3 to the issue of whether the technology shock process in the model should be stochastic or not. Another question that will be answered is what the degree of persistence should be if the shock process is not stochastic. The model that this application will be based on is the simple one-sector indivisible labor model that was used in Hansen (1985) and also used in Hansen (1997).

The model is as follows: The representative agent chooses $\{c_t, h_t, i_t, k_{t+1}\}_{t=1}^{\infty}$ to solve the following dynamic problem

$$\begin{aligned} &\text{Max } E\sum_{t=1}^{\infty}\rho^t(\log c_t - A\, h_t), \;\; 0 \le \rho \le 1, \;\; A > 0 \\ &\text{subject to} \\ &c_t + i_t \le z_t k_t^{1-\beta} h_t^{\beta}, \;\; 0 \le \beta \le 1 \\ &k_{t+1} = (1-\delta)\, k_t + i_t, \;\; 0 \le \delta \le 1 \\ &z_{t+1} = z_t^{\theta}\epsilon_t, \;\; 0 \le \theta \le 1, \forall t \ge 2 \\ &\epsilon \sim (0, \sigma_{\epsilon}^2) \;\text{ and } z_1 \text{ and } k_1 \text{ given,} \end{aligned} \tag{16}$$

where c_t, h_t, i_t, and k_t represent consumption, hours supplied, investment and beginning of period capital stock respectively. Also, $y_t = z_t k_t^{1-\beta} h_t^{\beta}$ is defined to be output with z_t being the period t value of the technology shock. Note that in this model growth has

been abstracted.

Farmer (1993) describes how to approximate a model of this type. In particular, for any variable x_t define

$$\hat{x}_t = \frac{x_t - x_t^*}{x_t^*} \tag{17}$$

where x_t^* represents either the steady state value of x or the balanced growth path for x. In the case of no growth then $x_t^* = \overline{x}$ for all t. Then taking a first order Taylor's series approximation to the first order conditions of (16) one gets the following difference equation

$$\hat{s}_{t+1} = \mathbf{J}\hat{s}_t + \hat{\eta}_t \tag{18}$$

where $s_t = (k_t, z_t)'$ and $\hat{\eta}_t = (0, \epsilon_t)'$. Farmer (1993) also show that it is possible to express the data variables of the model, $d_t = (c_t, h_t, i_t, y_t)'$, as a function of the state variables, s_t in the following way:

$$\hat{d}_t = \mathbf{B}\hat{s}_t. \tag{19}$$

Therefore the state-space representation of (16) is given by (18) together with (19). Therefore the likelihood function, $p(D_T|\theta, s_0, m_j)$ for the model given in (16) is calculated using the method described in Section 3 using the state-space representation defined above in (18) and (19).

There are two models that are studied. The first model, m_1, is given by (16) with $0 < \theta < 1$. Therefore the first model has a technology shock process that is given by a stationary AR(1) process. The second model, m_2, is given by (16) with the restriction that $\theta = 1$. That is, the technology shock process is a random walk. These two models will be compared using the methods described in Section 3.

4.1 Data

The data used in this application consists of seasonally adjusted US quarterly time series, for the period 1959:1 until 1999:4, on real consumption of non-durables and services (c_t), total hours for non-agricultural industries (h_t), real gross investment (i_t), and real GNP (y_t). All series apart form the total hours series were obtained from the FRED[6] database. The total hours series was obtained from the Bureau of Labor Statistics LABSTAT[7] database. Two series were used in the construction of the total-hours series used. They were average hours supplied[8] and number employed[9] according to the Household Labor Survey. One aspect of the Current Population Survey is that for the years of 1959, 1964, 1970, 1981, 1987, and 1992, Labor Day fell in the survey week. For those years the average hours supplied for September was artificially low, as the reference week only contained four days. A dummy variable approach was used to account for these outliers.

As the model described in (16) abstracts from growth the data were then detrended using the Hodrick and Prescott (1997)(HP) filter with smoothing parameter set equal to 1600. Therefore for each variable , x_t, the long run trend, x_t^*, was set equal to the smoothed trend obtained from the HP filter. This was used to calculate $\hat{d}_t = (\hat{c}_t, \hat{h}_t, \hat{i}_t, \hat{y}_t)'$ using (17). There is only one stochastic component in the specification of (16). Therefore principle components were used to create a composite variable, $\hat{a}_t = \alpha' \hat{d}_t$, where the vector α is chosen so that $\hat{a}_t$ contains the maximum amount of the information present in $\hat{d}_t$. In all cases α is identified by choosing vectors that satisfy $\alpha'\alpha = 1$.

[6] http://www.stls.frb.org/fred/
[7] http://stats.bls.gov:80/datahome.htm
[8] The series ID for average hours supplied is lfu1231040000000
[9] The series ID for number employed is lfs11104010000

4.2 Prior Specifications

The structural parameters in (16) are $\psi = (\beta, \delta, \rho, \theta, A, \sigma_\epsilon^2)'$ and the state variables are $s_0 = (\hat{k}_0, \hat{z}_0)'$. Then, the parameters of the model are $\phi = (\psi, s_0) \in \Psi \cup \mathcal{S}_0$. The prior, $p(\phi_j|m_j)$, for model m_j is defined to be

$$p(\phi|m_j) = p(\psi|m_j)\ p(s_0|m_j). \tag{20}$$

The support for the prior, $p(\psi|m_j)$ for j=1,2, is $\Psi = (0,1)^4 \times (0,\infty)^2$ as the parameters β, δ, ρ, and θ are all constrained to be in the interval (0,1), while the parameters A and σ_ϵ^2 are constrained to be positive. The joint prior distribution is assumed to be the product of independent prior distributions with

$$p(\phi|m_1) = p(\beta|m_1)p(\delta|m_1)p(\rho|m_1)p(\theta|m_1)p(A|m_1)p(\sigma_\epsilon^2|m_1) \tag{21}$$

and

$$p(\phi|m_2) = p(\beta|m_2)p(\delta|m_2)p(\rho|m_2)p(A|m_2)p(\sigma_\epsilon^2|m_2) \tag{22}$$

where the prior for θ for model m_2 is defined to be

$$p(\theta|m_2) = \begin{cases} 1 & \text{if } \theta = 1 \\ 0 & \text{else} \end{cases}. \tag{23}$$

There are a number of approaches one can take in defining a prior distribution constrained on the interval (0,1). One approach is to use a truncated prior such as a truncated Gaussian prior. However, a truncated prior assigns positive prior weight to the endpoints of the distribution and this has some drawbacks. The approach in this paper is to define

a prior on the Real line and then use the transformation

$$x^t = \frac{e^x}{1+e^x} \tag{24}$$

to transform the prior to the interval (0,1). In practice a Gaussian distribution is used for the prior defined on $\mathbf{R}$. A summary of the prior specifications for model m_1 is contained in Table 1.

Table 1: Prior Specification for model m_1

	Prior mode	95% Highest Prior Density Region
β	0.64	[0.538,0.737]
δ	0.025	[0.013,0.047]
ρ	0.99	[0.985,0.994]
θ	0.95	[0.915,0.975]
A	2.82	[1.84,4.09]
σ_ϵ	0.007	[0.0049,0.0091]

A graphical representation of the priors can be found in Figure 1 at the end of this section. The prior variances were chosen to reflect a reasonably large degree of uncertainty over the values of the parameters. However they were chosen so that the 95% prior highest density intervals lay in a region of the parameter space that were not too unreasonable with respect to the literature. For example, the 95% prior coverage interval for β, labor's share of income, was set to be equal to [0.538, 0.737]. Authors have suggested a variety of values for β. DeJong et al. (1996) calibrate β to be 0.54 and use a prior for β of [0.48, 0.68]. Hansen (1985) and Kydland and Prescott (1982) calibrate β to be 0.64. However, depending on how capital is defined and measured, other authors have suggested higher values for β. For example, Prescott (1986) suggests a value for β of 0.75 while Christiano (1988) suggests a value of 0.66.

The prior for δ, the depreciation rate for capital, covers a range of about 5% per annum to about 17.5% per annum with a prior mean at 10% per annum. Most studies calibrate

δ to be 0.025. The prior for ρ, the time discount factor, has a 95% prior coverage interval of [0.985,0.994]. This implies an economy with a real interest rate ranging from 2.5% to about 7% with a prior mode of about 4%. The 95% prior coverage interval for the AR(1) parameter θ is [0.915,0.975]. This implies a wide range of persistence in the productivity shock process. Hansen (1985) calibrates θ to be 0.95 while DeJong et al. (1996)) use a value of 0.90. A recent paper by Hansen (1997) finds evidence to suggest that the value of θ is lower than 0.95 and closer to 0.90. Finally, the 95% prior coverage interval for σ_η is [0.0049,0.0091]. Note that in this case the variance to the shock process will reflect the variances of all of the data as opposed to the practice of matching the variance of the output series.

The prior for the initial state variables has support on $\mathcal{S}_0 = \mathbf{R}^2$ and is defined to be

$$p(s_0|m_j) = p(\hat{k}_0|m_j)\ p(\hat{z}_0|m_j) \tag{25}$$

where both $p(\hat{k}_0|m_j)$and $p(\hat{z}_0|m_j)$ are defined to be Gaussian with mean zero and standard deviation 0.008. This prior reflects that both the technology shock and the capital stock are expected to deviate around their balanced growth path or steady state. The variance is chosen to reflect a reasonable amount of uncertainty over the values of the initial state variables.

The prior specification for model m_2 is the same as for model m_1 except that the prior for θ is a point mass prior concentrated at $\theta = 1$.

4.3 Results

The posterior distributions of the two variants of (16) were formed as described in Section 3.1 and 50,000 draws from these distributions were obtained using the RWMH al-

gorithm outlined in Section 3.2. In all cases the numerical standard errors[10] were less than 10% of the calculated posterior standard deviations. The posterior moments are reported in Tables 2 and 3 and the marginal distributions for each parameter are reported in Figures 1 and 2 at the end of this section.

Table 2: Posterior Moments for Model 1: Transitory Shock Process

	Mean	Mode	95% Highest Posterior Density Region
β	0.4657	0.4637	[0.3569,0.7990]
δ	0.0580	0.0524	[0.0242,0.1997]
ρ	0.9913	0.9918	[0.9871, 0.9987]
θ	0.9614	0.9664	[0.8212,0.9871]
A	2.7910	2.7831	[1.6354,4.0441]
σ_ϵ	8.403×10^{-6}	8.045×10^{-6}	$[4.876 \times 10^{-6}, 9.984 \times 10^{-5}]$

Table 3: Posterior Moments for Model 2: Permanent Shock Process

	Mean	Mode	95% Highest Posterior Density Region
β	0.5124	0.5119	[0.4118,0.7990]
δ	0.0428	0.0391	[0.0186,0.1996]
ρ	0.9919	0.9923	[0.9882,0.9989]
A	2.8005	2.8376	[1.6244,4.0692]
σ_ϵ	1.053×10^{-5}	1.015×10^{-5}	$[6.299 \times 10^{-6}, 9.984 \times 10^{-5}]$

The first observation one can make from the results reported in Tables 2 and 3 is that the posterior distribution for some parameters are significantly different, in location, than the prior distributions. In particular, the posterior distribution for the depreciation rate of capital, δ, and the variance to the shock process are quite different from their prior distributions. There is also some disparity in the posterior distributions between the two models.

The general results of the estimation are as follows: There is evidence that the labor share of income is smaller than the 64% usually used in the RBC literature, there is

[10]Computations reported in this paper were undertaken [in part] using the Bayesian Analysis, Computation and Communication software (http://www.econ.umn.edu/ bacc) described in Geweke (1999)

Table 4: Marginal likelihoods and Bayes Factors in favor of m_2 over m_1

	Model m_1 (Temporary)	Model m_2 (Permanent)
Log marginal likelihoods	575.5818	582.3541
	(0.1096)	(0.0280)
Log Bayes factor	6.7723	-
	(0.1116)	

evidence that the depreciation rate of capital is higher than the 2.5% per quarter that is used in the RBC literature, the variance to the innovations of the technology shock process is significantly smaller than the values typically used, and that, for the transitory shock model, there is evidence that the technology shock process is more persistent than what is typically used.

The posterior distributions for the intertemporal rate of time preference, ρ, and the dis-utility of labor parameter, A, are very similar to the prior distribution. This would suggest that there is little information in the data with respect to those parameters.

In estimating (16) under the two error assumptions we obtained a set of draws, $\{\phi_j^1, \ldots, \phi_j^N\}$ for each model, m_1 and m_2. These draws can be used to calculate the marginal likelihoods, $p(D_T|\phi_j, s_0, m_j) \quad j = 1, 2$, and hence the Bayes factors and the posterior odds ratios in favor of model m_2 over model m_1 using the method of Gelfand and Dey (1994) as suggested in Geweke (1999). The marginal likelihoods for each model and the subsequent Bayes factor in favor of model m_2, the model with the permanent shock process, over model m_1, the model with the temporary shock process is given in Table 4. Note that the standard errors for the estimates are reported in parentheses below each estimate.

Note that Table 4 reports log marginal likelihoods and log Bayes factors. Hence the Bayes factor in favor of the permanent shock model over the temporary shock model is 873.32 which implies that one would have to put 873.32 as much prior weight on the temporary shock model than the permanent shock model for the posterior odds ration

to favor the temporary shock model. Thus, it is fair to state that there is overwhelming evidence in favor of the permanent shock model over the temporary shock model. This result is opposite to the result found in Hansen (1997) where he concludes that the shock process should be less persistent than than what is typically used. Even if we restrict ourselves to the case of temporary shocks it is found that a higher degree of persistence is preferred. The posterior mean of θ for model m_1 is 0.9614 which is greater than the value of 0.95, which is commonly used in the literature.

There may be a number of reasons for why the results are different to those of Hansen (1997). First is that the results reported here are obtained using likelihood methods, and hence uses all of the information in the data, while Hansen (1997) concentrates on certain moments of the data and implied impulse response functions of the model in making his conclusions. Second is that Hansen (1997) only allows for the parameters that affect the shock process to vary in his study whereas in this study all of the parameters are allowed to vary. It can be seen that the posterior distribution of the labor share parameter, β, and the depreciation of capital parameter, δ, are quite sensitive to the choice of shock process.

The above results are for the data set as a whole. It would be wise to check to see if this overall result is mirrored across the whole sample. To do this the log Bayes factor is decomposed across every observation using the method described in Section 3.1. The particular, using the priors defined above, the first five observations are used to make draws from $p(\phi, s_0|D_5, m_j)$ for each model. Then for each $t = 5, \ldots, T-1$ draws are made from the distribution $p(\log(\hat{p}_t^{t+1})|D_t, m_j)$ for each model. This distribution is the distribution of the log predictive likelihood for d_{t+1} given information up to and including period t. The log Bayes factor in favor of model m_2 over model m_1, for each observation is the difference in the log predictive likelihoods. The means of the log Bayes factor for 1960:2 until 1999:4 are reported in Figure 3 at the end of this section. The cumulative

log Bayes factor is reported in Figure 4.

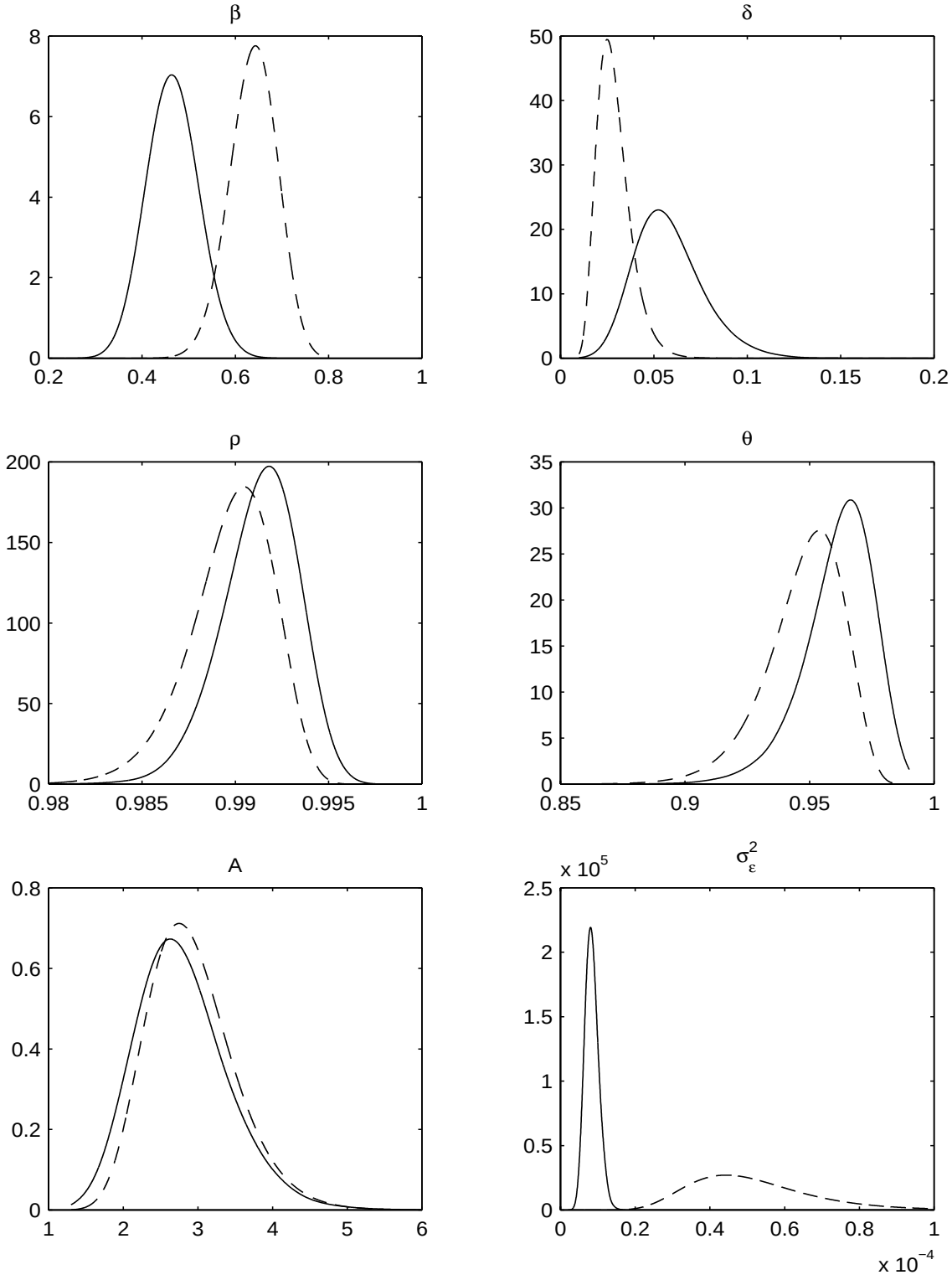
It should be noted that the cumulative log Bayes factor represents the sum of the means of the individual Bayes factors. Hence there needs to be a confidence band reported around this line. It is not expected that the final value of the cumulative log Bayes factor equals exactly the log Bayes Factor reported in Table 4. In fact, one would expect the confidence band to be quite large as there would be a large uncertainty over the values of the structural parameters when they are drawn from the posterior in the case when T is small. The most important information from the graphs of the decomposed log Bayes factor and the cumulative Bayes factor is the general shape.

Periods where the decomposed log Bayes factor is less than zero are periods where the transitory shock model is favored. This also implies that the cumulative log Bayes factor graph would be declining. Conversely, periods where the log Bayes factor is greater than zero or periods where the cumulative log Bayes factor has a positive slope implies that the permanent shock model is favored.[11]

It is clear that the permanent shock model is favored in the early periods of the sample. However it is also clear that there are periods when the transitory shock model is more favored and from the middle 1980's onwards it appears that the two models are hard to distinguish from each other. Thus, while there is strong evidence for the permanent shock model using all of the data from 1959:1 to 1999:4 it is also clear that the data from 1959:1 to the early 70's is where the permanent shock model is most favored. From the early 1970's onwards there is not much between the two models and this would suggest that a model that had a shock process with both temporary and permanent components might be optimal.

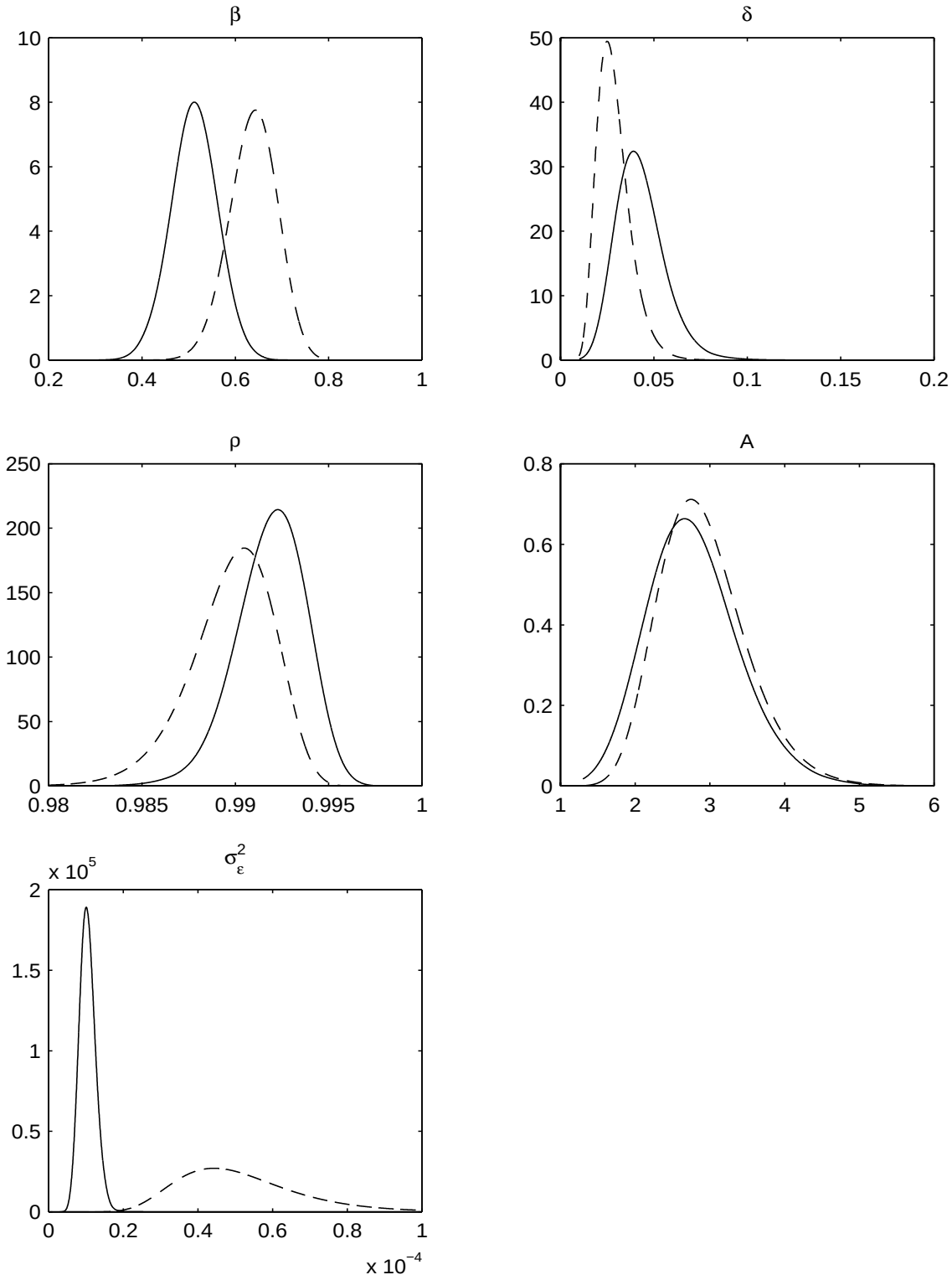
[11]The log Bayes factor that is referred to is the log Bayes factor in favor of the permanent shock model over the transitory shock model. If the ordering is reversed then the interpretation of the slopes of the cumulative log Bayes factor would also be reversed.

Figure 1: Posterior and Prior Distributions for Model 1: Transitory Shock Process[a]



[a] Notes: The prior distribution is represented by the dashed line while the posterior distribution is represented by the solid line.

Figure 2: Posterior and Prior Distributions for Model 2: Permanent Shock Process[b]



[b] Notes: The prior distribution is represented by the dashed line while the posterior distribution is represented by the solid line.

Figure 3: Decomposed Log Bayes Factor

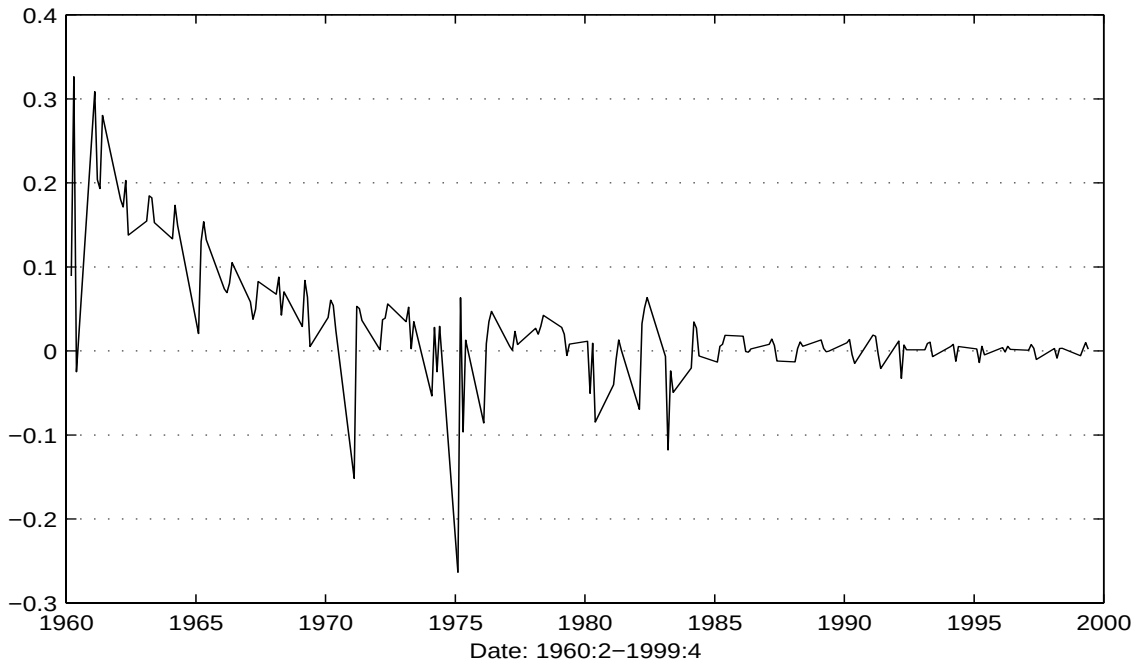
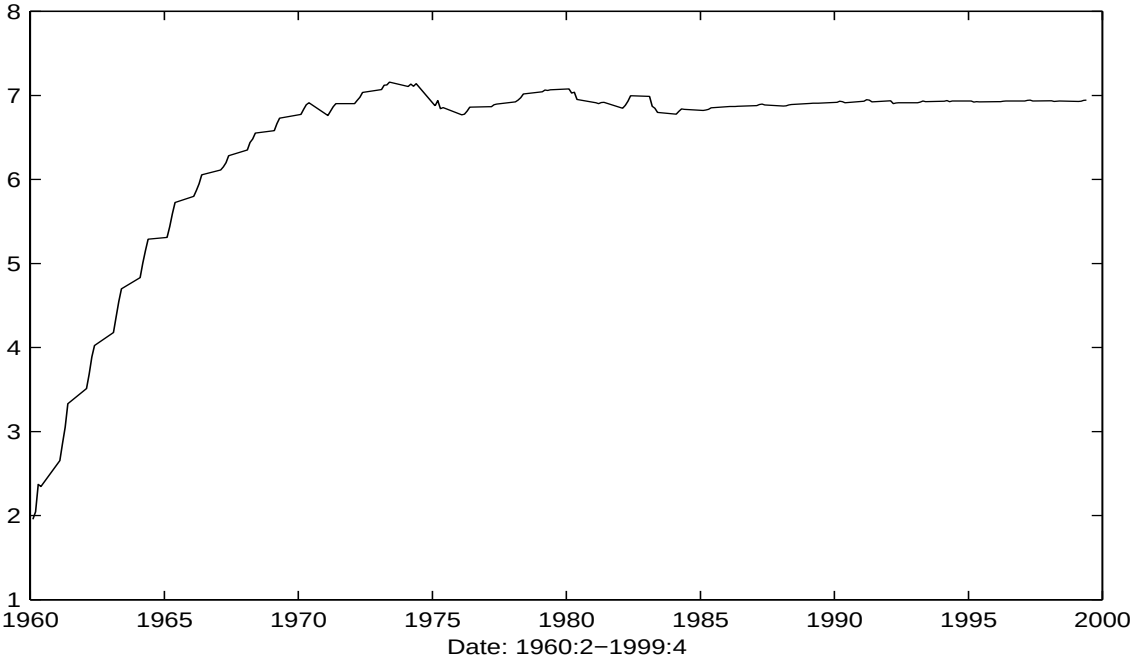


Figure 4: Cumulative Log Bayes Factor



5 Conclusion

The aim of this paper was develop a method that could directly compare two or more DSGE models used in the RBC literature using likelihood methods. The reason for this was to answer the criticisms of Hansen and Heckman (1996) that most of the methods used to evaluate models in the RBC literature did evaluate models over the full dimension of the data.

The method developed in this paper constructs a likelihood function for a DSGE model using the approximate state-space representation of the model. The problem of calculating a likelihood for a model that contained more observables than stochastic components was solved using the method of principle components rather than adding additional stochastic terms which is common in the literature. This allows the method of comparison to compare models based solely on the description of the model without any additional components. Bayesian methods were used to estimate and compare models. These methods allowed for the use of prior information with information from the data in a formal way. Hence, RBC models were estimated and compared using all of the information contained in the data. It was also shown how models could be compared across different sub-samples of the data in a formal and consistent way.

The method was applied to the problem of determining the appropriate technology shock mechanism for the standard one sector growth model of Hansen (1985). Two types of shock processes were modelled with one having persistent but temporary technology shocks while the other model had permanent technology shocks. It was found that the permanent shock model was preferred for the whole data set. This result contradicted the result reported in Hansen (1997). However, when the comparison was decomposed across the whole sample it was found that the permanent shock model was strongly favored only in the first part of the data. In the second part of the data the two shock process were very similar in terms of their out of sample prediction performance with the transitory

shock process being favored for some periods and the permanent model being favored for others. Thus the final conclusion on what is the appropriate shock process to use is quite different depending on what part of the data is being studied.

References

Anderson, E.W., L.P. Hansen, E.R. McGratten, and T.J. Sargent (1996) 'Mechanics of forming and estimating dynamic linear economies.' In *Handbook of Computational Economics,* ed. H.M. Amman, D.A. Kendrick, and J. Rust, vol. I (Elsevier Science B.V.)

Bernado, J.M., and A.F.M. Smith (1994) *Bayesian Theory* (Chicester: John Wiley and Sons)

Canova, F., and E. Ortega (1995) 'Testing calibrated general equilibrium models.' unpublished manuscript

Christiano, L.H. (1988) 'Why does inventory investment fluctuate so much?' *Journal of Monetary Economics* 28, 247–80

Christiano, L.H., and M. Eichenbaum (1992) 'Current business cycle theories and aggregate labor market fluctuations.' *American Economic Review* 82(3), 430–50

Cogley, T., and J.M. Nason (1994) 'Testing the implications of long-run neutrality for monetary business cycle models.' *Journal of Applied Econometrics* 9(0), S37–70

DeJong, D., B. Ingram, and C. Whiteman (1996) 'A Bayesian approach to calibration.' *Journal of Business and Economic Statistics* 14(1), 1–9

——— (2000) 'A Bayesian approach to dynamic macroeconomics.' *Journal of Econometrics* 98(2), 203–223

Diebold, F., L. Ohanian, and J. Berkowitz (1998) 'Dynamic equilibrium economies: A framework for comparing models and data.' *Review of Economic Studies* 65(3), 433–51

Farmer, R.E.A. (1993) *The Macroeconomics of Self-Fulfilling Prophecies* (Cambridge, Massachsetts: M.I.T. Press)

Farmer, R.E.A., and J-T. Guo (1994) 'Real business cycles and the animal spirits hypothesis.' *Journal of Economic Theory* 63, 42–72

Gelfand, A.E., and D.K. Dey (1994) 'Bayesian model choice: Asymptotics and exact calculations.' *Journal of the Royal Statistical Society Series B*

Geweke, J.F. (1995) 'Bayesian comparison of econometric models.' University of Minnesota Working Paper.

—— (1999) 'Using simulation methods for Bayesian econometric models: Inference, development and communication.' *Econometric Reviews* 18(1), 1–73

Hansen, G.D. (1985) 'Indivisible labor and the business cycle.' *Journal of Monetary Economics* 16(3), 309–27

—— (1997) 'Technical progress and aggregate fluctuations.' *Journal of Economic Dynamics and Control* 21(6), 1005–1023

Hansen, L.P. (1982) 'Large sample properties of generalized method of moments estimators.' *Econometrica* 50, 1029–54

Hansen, L.P., and J.J. Heckman (1996) 'The empirical foundations of calibration.' *The Journal of Economic Perspectives* 10(1), 87–104

Harvey, A.C. (1989) *Forecasting, Structural Time Series Models, and the Kalman Filter* (Cambridge: Cambridge University Press)

Hodrick, R.J., and E.C. Prescott (1997) 'Post-war U.S. business cycles: An empirical investigation.' *Journal of Money, Credit and Banking* 29(1), 1–16

Ireland, P.N. (1999) 'A method for taking models to data.' Department of Economcis Working Paper 421, Boston College.

Johnson, R. A., and D. W. Wichern (1988) *Applied Multivariate Statistical Analysis,* second ed. (Prentice-Hall of Australia Pty. Limited, Sydney: Prentice-Hall International)

Kim, K, and A.R. Pagan (1995) 'The econometric analysis of calibrated macroeconomic models.' In *Handbook of Applied Econometrics,* ed. H.H. Pesaran and Mike Wickens (Blackwell Press)

King, R.G., C.I. Plosser, J. Stock, and M. Watson (1991) 'Stochastic trends and economic fluctuations.' *American Economic Review* 81, 819–40

Kydland, F.E., and E.C. Prescott (1982) 'Time to build and aggregate fluctuations.' *Econometrica* 50(6), 1345–70

—— (1996) 'The computational experiment: An econometric tool.' *Journal of Economic Perspectives* 10(1), 69–85

Landon-Lane, J.S. (1998) 'Bayesian comparison of dynamic macroeconomic models.' PhD dissertation, University of Minnesota

Long, J.B. Jr., and C.I. Plosser (1983) 'Real business cycles.' *Journal of Political Economy* 91, 39–69

Ortega, E. (1995) 'Assessing and comparing computable dynamic general equilibrium models.' Unpublished manuscript, European University Institute

Prescott, E.C (1986) 'Theory ahead of business cycle measurement.' *Federal Reserve Bank of Minneapolis Quarterly Review* 10(4), 9–22

Sargent, T.J. (1989) 'Two models of measurements and the investment accelerator.' *Journal of Political Economy* 97, 251–287

Sims, C.A. (1996) 'Macroeconomics and methodology.' *The Journal of Economic Perspectives* 10(1), 105–20

Smith, A.A. (1993) 'Estimating non-linear time series models using simulated VAR's.' *Journal of Applied Econometrics* 8, S63–84

Stadler, G.W. (1994) 'Real business cycles.' *Journal of Economic Literature* 32(4), 1750–83

Taylor, J.B., and H. Uhlig (1990) 'Solving non-linear stochastic growth models: A comparison of alternative solution methods.' *Journal of Business and Economics Statistics* 8(1), 1–17

Tierney, L. (1994) 'Markov chains for exploring posterior distributions.' *Annals of Statistics* 22(4), 1701–1762

Watson, M. (1993) 'Measures of fit for calibrated models.' *Journal of Political Economy* 101, 1011–41

A Appendix

A.1 Decomposition of Bayes Factor

Suppose that we have data $D_T = \{d_t\}_{t=0}^T$ and we wish to predict the values of $d_{T+1}, \ldots, d_{T+m}$. The predictive density of $d_{T+1}, \ldots, d_{T+m}$ conditional on model $m_k \in \mathcal{M}$ and data D_T is

$$p(d_{T+1}, \ldots, d_{T+m}|D_T, m_k) = \int_{\Phi_K} p(\phi_k \mid D_T, m_k) \prod_{t=T+1}^{T+m} p(d_t \mid D_{t-1}, \phi_k, m_k) d\phi_k. \quad \text{(A.1)}$$

The predictive density applies prior to observing the data and as usual we can define the analogous predictive likelihood function,

$$\hat{p}_T^{T+m}(m_k) = \int_{\Phi_K} p(\phi_k \mid D_T, m_k) \prod_{t=T+1}^{T+m} p(d_t \mid D_{t-1}, \phi_k, m_k) d\phi_k. \quad \text{(A.2)}$$

Note that $\hat{p}_1^T(m_k) = p(D_T|m_k)$. Given any $0 < u < v$ it follows that

$$\begin{aligned}
\hat{p}_u^v(m_k) &= \int_{\Phi_k} p(\phi_k \mid D_T, m_k) \prod_{t=u+1}^{v} p(d_t \mid D_{t-1}, \phi_k, m_k) d\phi_k \\
&= \int_{\Phi_k} \frac{p(\phi_k|m_k) \prod_{t=0}^{u} p(d_t \mid D_{t-1}, \phi_k, m_k)}{\int_{\Phi_k} p(\phi_k|m_k) p(D_u|\phi_k, m_k) d\phi_k} \prod_{t=u}^{v} p(d_t \mid D_{t-1}, \phi_k, m_k) d\phi_k \\
&= \frac{p(\phi_k|m_k) \prod_{t=0}^{v} p(d_t \mid D_{t-1}, \phi_k, m_k) d\phi_k}{p(\phi_k|m_k) \prod_{t=0}^{u} p(d_t \mid D_{t-1}, \phi_k, m_k) d\phi_k} \\
&= \frac{p(D_v|m_k)}{p(D_u|m_k)}.
\end{aligned} \quad \text{(A.3)}$$

Thus the predictive likelihood function for observations u+1 through v is just the ratio

of the marginal likelihood's for the sample of observations 0 through v and 0 through u respectively.

The predictive likelihood can then be decomposed using (A.3). Consider any sequence of numbers such that $0 \leq u = s_0 < s_1 < \ldots < s_q = v$. Then it follows from (A.3) that

$$\hat{p}_u^v(m_k) = \frac{p(D_{s_1}|m_k)}{p(D_{s_0})} \cdots \frac{p(D_{s_q}|m_k)}{p(D_{s_{q-1}}|m_k)} = \prod_{\tau=1}^{q} \hat{p}_{s_{\tau-1}}^{s_\tau}(m_k). \tag{A.4}$$

As the marginal likelihood can be decomposed into the product of ratios of predictive likelihoods, we can see that the marginal likelihood represents the out of sample prediction performance of the model. Thus, by using the Bayes factor, which is just the ratio of the respective marginal likelihood's for each model, we are comparing models on their ability to predict out of sample.

Another application for the decomposition given in (A.4) is for direct model diagnostics. Consider the complete decomposition in which u=0 and v=T and where $s_i - s_{i-1} = 1$. A relatively low value of $\hat{p}_{ks_{i-1}}^{s_i}$ would indicate that the observation indexed by s_i was surprising given observation s_{i-1} and model k. Thus, one could use this decomposition to evaluate the performance of models in regard to predicting large movements in the data. For example, a large movement may be surprising to all models but some may do better than others. The decomposition is also useful in getting some insight into the Bayes factor. Using the decomposition of the marginal likelihood in A.4, one can do the same for the Bayes factor. That is,

$$B_{ij,u}^v = \frac{\hat{p}_u^v(m_i)}{\hat{p}_u^v(m_j)} = \prod_{\tau=1}^{q} \left(\frac{\hat{p}_{s_{\tau-1}}^{s_\tau}(m_i)}{\hat{p}_{s_{\tau-1}}^{s_\tau}(m_j)} \right) = \prod_{\tau=1}^{q} B_{ij,s_{\tau-1}}^{s_\tau} \tag{A.5}$$