# Assessing Markov Chain Approximations: A Minimal Econometric Approach

Pedro Silos*

Dept. of Economics

University of Iowa

October 18, 2004

**Abstract**

Markov chain approximations of continuous state-space processes are common in dynamic economic models. Increasing the dimensions of the approximating state space is costly; this paper develops a procedure to evaluate the tradeoff between the number of dimensions devoted to modelling dynamics and those devoted to modelling the contemporaneous state space for the variable. The methodology borrows from a previous literature which formalizes statistical inference within calibrated general equilibrium models. As a particular application, Markov chain approximations for post-war-real-per-capita-US consumption growth are compared. Standard business cycle theory is used to generate needed information regarding state transition probabilities. In this application it is useful to trade some accuracy in defining the state space for more realistic dynamics.

*Keywords:* Markov Chain approximations, Bayesian model comparison
*JEL Classification:* C11, C15, C52

*I thank John Geweke, Beth Ingram and especially Charles Whiteman for their help. I also thank Grace Chan, Simon Lee, Linnea Polgreen and seminar participants at the University of Iowa for useful comments. Finally, I thank Chris Otrok for providing me with the computer code for the multivariate version of Tauchen's procedure.
Address: Dept. of Economics, W210 PBAB, University of Iowa, Iowa City, IA, 52245, tel.: (319)-335-0920, fax: (319)-335-1956, email: pedro-silos@uiowa.edu

# 1 Introduction

Markov chain approximations of macroeconomic variables have been used frequently in dynamic stochastic general equilibrium models. Examples include approximations of income processes as in Aiyagari(1994), who uses a seven state Markov chain to approximate labor endowment. Also, many papers in the asset pricing literature approximate consumption growth, from Mehra and Prescott (1985) to Otrok, Ravikumar and Whiteman (2001). The reasons for using such approximations are usually computational, in that subsequent calculations are more convenient to carry out on a discrete grid. For example, in consumption-based asset pricing models, some preference specifications are simple enough that returns and prices can be calculated even if consumption growth is described as a continuous state autoregression; for recursive preferences, discretization decreases the computational burden.

A well-known method for computing Markov chain approximations of continuous state space univariate or vector first order autoregressions is due to Tauchen (1986). Although not the only one available (see for example Tauchen and Hussey (1991)) this method is popular due to its simplicity.

If a given time series can be modelled as an autoregressive process there are two choices to be made when approximating it as a Markov chain. One pertains to the choice of the lag length: the optimal model might have $p$ lags, versus a model with $p+1$ lags or $p-1$ lags. The other choice is the dimension of the state space for the variable. Let $N$ denote the number of states the variable can take values on. Choices of $N$ and $p$ imply a transition probability matrix of dimension $N^p$, which can be quite large even for moderate $N$ if $p > 1$.

Table 1 and Figures 1 and 2 illustrate the tradeoffs involved in modelling U.S. consumption growth. The first column in Table 1 shows the sample values for the mean, variance and the first three autocorrelations for the observed consumption growth series. The second and third columns show the same moments obtained by simulating a long time series from a transition matrix calculated for US consumption growth using Tauchen's method for the $p = 2, N = 3$

case, (second column), and the $p = 1, N = 9$ case, (third column). Figure 1 plots the unconditional densities (left axis) for a continuous state space AR(3) model, with the unconditional mass function (right axis) obtained from the transition matrices for the $p = 2, N = 3$ Markov chain approximation, while figure 2 compares the same continuous state space density, with a Markov chain approximation where $p = 1, N = 9$. Although the number of states ($N$) is rather large in the second case, the model does not match the second and third autocorrelations closely. However, with a clearly worse approximation to the unconditional density, the Markov chain model with $p = 2$ is closer to matching the second and third autocorrelation, but to keep the dimension of the transition matrix fixed, $N$ has to be decreased.

Table 1: Consumption Growth Moments: Models and Data.

| | Actual data | $p = 2, N = 3$ | $p = 1, N = 9$ |
|---|---|---|---|
| Mean | 1.005 | 1.005 | 1.005 |
| Variance | $2.9 \times 10^{-5}$ | $3.5 \times 10^{-5}$ | $3 \times 10^{-5}$ |
| First Autocorr. | 0.22 | 0.19 | 0.22 |
| Second Autocorr. | 0.14 | 0.1 | 0.05 |
| Third Autocorr. | 0.16 | 0.03 | 0.01 |

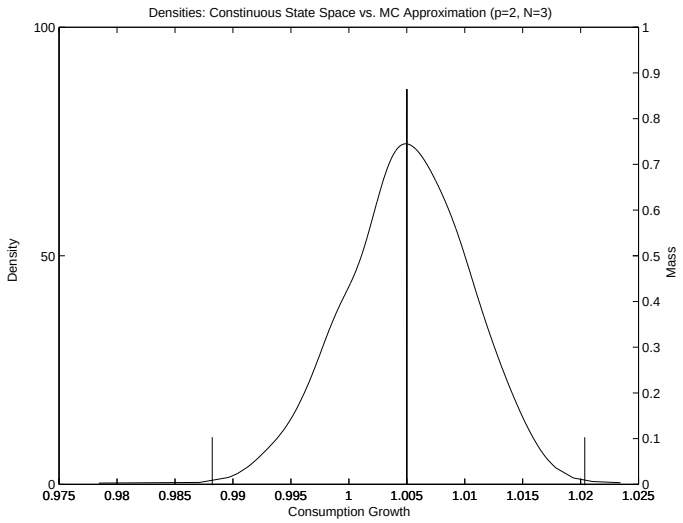


Figure 1: $p = 2, N = 3$ vs. Continuous State Space AR(3)

With $N^p$ limited by computational considerations, how should one choose $N$ and $p$? Clearly, this depends upon the economic questions being addressed.

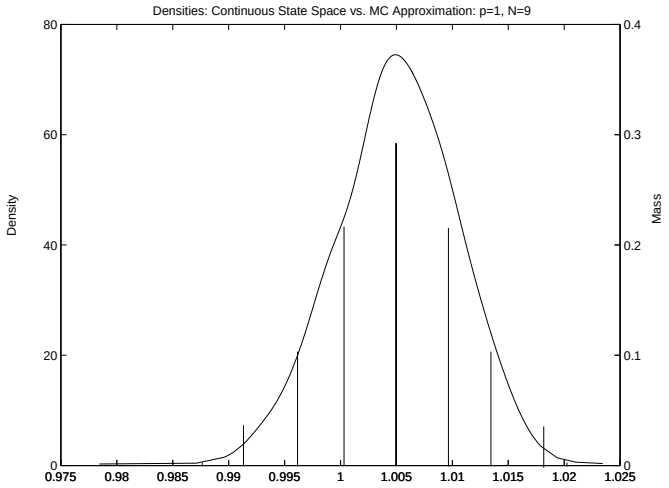


Figure 2: $p = 1, N = 9$ vs. Continuous State Space AR(3)

More detail in the unconditional marginal distribution for the random variable can be characterized with greater $N$; greater $p$ is necessary to capture more features of the dynamics of the process. Inasmuch as both volatility (unconditional variance) and persistence (first and higher order dynamics) are often of interest in dynamic macroeconomics, it is useful to have a formal procedure for assessing the tradeoffs involved. The classical statistical procedure for doing this would involve assuming that one of the candidate approximations represents the "correct" specification, and comparing the maximized value of the likelihood under that specification to those under other specifications. Such a procedure is not feasible here: none of approximations under consideration could be considered the correct specification, and further, not even the obvious candidate for a likelihood function, the economic model whose solution requires the discretization, is regarded as correct.

This state of affairs is common in the calibrated model literature, and several methods have been proposed recently for introducing formalism into model assessment procedures (see, for example, the survey by Canova and Ortega). Traditionally, studies that have used calibrated models to examine economic issues have been informal in assessing the fit of those models, and usually their conclusions are based on some implicit measure of distance between some of

the moments of model-simulated data and the corresponding sample moments from the observed data (a prototypical example is Cooley and Prescott (1995)). Following the notation in Canova and Ortega (2000), let $y_t$ represent the data and $x_t = f(\gamma, z_t)$ represent a simulated value of the corresponding vector from the model, which is a function of a set of parameters $\gamma$ and a driving process $z_t$. Besides the informal approach when comparing moments, examples in the literature are Watson (1993), who focuses on the properties of the implicit "residual term" between the observed variable $y_t$ and the simulated $x_t$ values; Gregory and Smith (1993), who take sampling uncertainty into account in their treatment of $z_t$, but treat the data as given; Cechetti, Lam and Mark (1994), who base their estimation and calibration method in the sampling variability in the data, treating $\gamma$ and the parameters of the $z_t$ process as known; and finally a Bayesian approach developed by DeJong, Ingram and Whiteman (1996) and Geweke (1999) who treat model and data uncertainty symmetrically.

The "minimal econometric approach" developed by Geweke does permit comparison of multiple models, none of which are regarded as "correct", and under conditions when the likelihood is unknown. This methodology only requires being able to simulate from a given model, i.e. to generate a sequence of random draws from the implied distribution of a model. The approach, described in more detail Section 2 of the paper, is implemented in Section 3 to assess alternative Markov chain approximations for the consumption growth series, which is used routinely in asset pricing models, and thus the assessment provided here may be of broader use.

## 2 Model Comparison

Suppose a researcher wants to compare two asset pricing models that differ on, say, preferences or the endowment process. The researcher will need a formal method to carry out the comparison between the two economic models. This is analogous to what is being done in this paper, but instead of economic models, I am comparing two Markov chain approximations. The remaining part of the section describes formally the comparison methodology.

The comparison is based on a set of moments of interest. For comparing two economic models, such moments might include average ratios of key economic variables, standard deviations of some of those variables, etc. In the Markov chain example, I will focus on intercepts, slope coefficients, and variances in the implied continuous-state autoregression. The competing models are used to simulate large samples of model output, and these samples are used to estimate, nonparametrically, the joint distribution of the objects of interest. These distributions are judged in light of the analogous joint distribution produced from a reduced-form, statistical model of the actual observed data. Specifically, the models are compared via a "Bayes factor": one model "beats" another if the expected value of its joint distribution, taken with respect to the data-induced distribution, is larger than the other's.

To present the details of the comparison procedure, I borrow heavily from Geweke (1999). Let $A_1$ and $A_2$ be any two models to be compared. These can be general equilibrium models or Markov chain approximations. What is common between these models is that they do not describe the data in an exact way: exact relations between variables in a dynamic stochastic models do not hold in the data, and variables that are sometimes approximated as Markov chains can take more than a finite number of states. Let $\theta$ denote the vector of moments (or functions of moments) that these models claim to describe, and hence how the models perform when describing $\theta$ will be the basis of comparison between $A_1$ and $A_2$. This vector of moments will be a function of the model's underlying parameters, and by specifying prior distributions over these parameters a density over $\theta$ is induced. Denoting this density $p(\theta \mid A_j)$ within model $A_j$, realizations from this distribution are obtained via simulation. Let model D (D for data), be a statistical model, a reduced form model independent of $A_1$ and $A_2$ that will serve as a link between these and the data. It will also be endowed with a prior distribution $p(\theta \mid D)$, and the assumption of normality in the errors will imply a certain likelihood function for the data, denoted as $p(\mathbf{y} \mid \theta, D)$. The distribution of interest for the vector $\theta$ in the context of model D will be the posterior distribution $p(\theta \mid \mathbf{y}, D)$ given by:

$$p\left(\theta \mid \mathbf{y}, D\right) \propto p\left(\mathbf{y} \mid \theta, D\right) p(\theta \mid D) \tag{1}$$

Simulated values of $\theta$ from this distribution will be generated via a posterior simulator described below. The final goal is to obtain a numerical measure that allows the comparison between any two given models. First, I need the following two assumptions:

**Assumption 1**: $p\left(\mathbf{y} \mid \theta, A_1, D\right) = p\left(\mathbf{y} \mid \theta, A_2, D\right) = p\left(\mathbf{y} \mid \theta, D\right)$

This assumption implies that $A_1$ and $A_2$ bring no new information about the observables $\mathbf{y}$ if $\theta$ is known in the context of model $D$, since the purpose of those models is precisely to describe $\theta$. A straightforward result from assumption 1 is:

$$p\left(A_1 \mid \theta, \mathbf{y}, D\right) = \frac{p\left(\mathbf{y} \mid \theta, A_1, D\right) p\left(A_1 \mid \theta, D\right)}{p\left(\mathbf{y} \mid \theta, D\right)} = p\left(A_1 \mid \theta, D\right) \tag{2}$$

An equivalent result holds for model $A_2$. The sequence of observables $\mathbf{y}$ is not needed to make any comparison between $A_1$ and $A_2$ if the vector $\theta$ is already known in the context of model $D$. That is, given model D, $\theta$ is sufficient to distinguishing $A_1$ and $A_2$.

The second assumption needed is:

**Assumption 2**: $p\left(\theta \mid D\right) \propto constant, p\left(\theta \mid A_1, D\right) = p\left(\theta \mid A_1\right), and$
$p\left(\theta \mid A_2, D\right) = p\left(\theta \mid A_2\right)$.

This assumption means that prior to observing data, model $D$ brings no information about $\theta$, either directly through the prior distribution of $\theta$ given model $D$, or indirectly through either $A_1$ or $A_2$. We are interested in $\frac{p\left(A_1 \mid \mathbf{y}, D\right)}{p\left(A_2 \mid \mathbf{y}, D\right)}$, which is the ratio model probabilities given the observed data and the model

D that we are using to conduct the comparison. Now note that under Assumptions 1 and 2, we have

$$\begin{aligned} & p\left(A_1 \mid \mathbf{y}, D\right) \\ &= \int p\left(A_1 \mid \theta, \mathbf{y}, D\right) p\left(\theta \mid \mathbf{y}, D\right) d\theta \\ &= \int p\left(A_1 \mid \theta, D\right) p\left(\theta \mid \mathbf{y}, D\right) d\theta \\ &= \int \frac{p\left(\theta \mid A_1, D\right) p\left(A_1 \mid D\right)}{p\left(\theta \mid D\right)} p\left(\theta \mid \mathbf{y}, D\right) d\theta \\ &= \frac{p\left(A_1 \mid D\right)}{p\left(\theta \mid D\right)} \int p\left(\theta \mid A_1\right) p\left(\theta \mid \mathbf{y}, D\right) d\theta \end{aligned} \tag{3}$$

The last integral, $\int p\left(\theta \mid A_1\right) p\left(\theta \mid \mathbf{y}, D\right) d\theta$, is the expectation of $p\left(\theta \mid A_1\right)$ under the posterior distribution $p\left(\theta \mid \mathbf{y}, D\right)$, which is proportional to the likelihood because of Assumption 2. Although the support for both densities is the same, this quantity will be small whenever simulations from model $D$ and simulations from model $A_1$ congregate in different regions of that support. The farther away these regions are from one another, the smaller the quantity will be. The greater the overlap between these regions, the greater the quantity. An analogous expression can be derived for model $A_2$, and therefore a posterior odds ratio can be constructed:

$$\frac{p\left(A_1 \mid \mathbf{y}, D\right)}{p\left(A_2 \mid \mathbf{y}, D\right)} = \frac{p\left(A_1 \mid D\right)}{p\left(A_2 \mid D\right)} \frac{p\left(\mathbf{y} \mid A_1, D\right)}{p\left(\mathbf{y} \mid A_2, D\right)} \tag{4}$$

which, by virtue of (3), becomes

$$\frac{p\left(A_1 \mid \mathbf{y}, D\right)}{p\left(A_2 \mid \mathbf{y}, D\right)} = \frac{p\left(A_1 \mid D\right)}{p\left(A_2 \mid D\right)} \frac{\int p\left(\theta \mid A_1\right) p\left(\theta \mid \mathbf{y}, D\right) d\theta}{\int p\left(\theta \mid A_2\right) p\left(\theta \mid \mathbf{y}, D\right) d\theta} \tag{5}$$

Comparing (5) and (4), the key ingredient in the model comparison is the Bayes factor, a quotient of marginal likelihoods:

$$\frac{p\left(\mathbf{y} \mid A_1, D\right)}{p\left(\mathbf{y} \mid A_2, D\right)} = \frac{\int p\left(\theta \mid A_1\right) p(\theta \mid \mathbf{y}, D) d\theta}{\int p\left(\theta \mid A_2\right) p\left(\theta \mid \mathbf{y}, D\right) d\theta} \tag{6}$$

Given that the ratio of prior odds will be set to one throughout applications in the paper, this ratio will provide a numerical value for comparing models $A_1$ and $A_2$. The higher the Bayes factor, the closer simulations from model $A_1$ are to simulations from model $D$, relative to model $A_2$ and vice-versa. Given a sequence of simulated values $\theta^{(m)}$; $m = 1, \ldots, M$ from model $D$, I can estimate the Bayes factor as:

$$\frac{p\left(\mathbf{y} \mid A_1, D\right)}{p\left(\mathbf{y} \mid A_2, D\right)} \approx \frac{\frac{1}{M} \sum_{m=1}^{M} p\left(\theta^{(m)} \mid A_1\right)}{\frac{1}{M} \sum_{m=1}^{M} p\left(\theta^{(m)} \mid A_2\right)} \tag{7}$$

If functional forms for $p\left(\theta \mid A_1\right)$ and $p\left(\theta \mid A_2\right)$ were known, the computation of Bayes factors would be straightforward. Here these functional forms are not known. However, as noted above, the approach here only relies on being able to generate random draws from the distributions of interest. These draws are used to estimate $p\left(\theta \mid A_1\right)$ and $p\left(\theta \mid A_2\right)$ using standard non-parametric techniques. Given simulations $\{\theta^{(s_1)}\}_{s_1=1}^{M_1}$ and $\{\theta^{(s_2)}\}_{s_2=1}^{M_2}$ from models $A_1$ and $A_2$ respectively, and the simulation $\theta^{(m)}{}_{m=1}^{M}$ from model $D$; using a kernel density estimator, an approximate Bayes factor will be:

$$\frac{p\left(\mathbf{y} \mid A_1, D\right)}{p\left(\mathbf{y} \mid A_2, D\right)} \approx \frac{\frac{1}{M} \sum_{m=1}^{M} (\frac{1}{M_1} \sum_{s_1=1}^{M_1} K(\theta^{(s_1)}; \theta^{(m)}))}{\frac{1}{M} \sum_{m=1}^{M} (\frac{1}{M_2} \sum_{s_2=1}^{M_2} K(\theta^{(s_2)}; \theta^{(m)}))} \tag{8}$$

In the above expression K is the kernel function; in the computations below this is a Gaussian density.

## 3 Comparing Markov Chain Approximations

The "minimal econometric" approach described in section 2 will be used to explore the significance of the tradeoff between the number of lags in the original autoregressive model and the number of gridpoints that defines the state space

for the variable. The main elements will be two models that the researcher wants to compare ($A_1$ and $A_2$), an auxiliary model denoted $D$ and a vector $\theta$, considered the "object of interest" that models claim to describe. Denoting $y_t$ the variable of interest I assume that it can be modelled as[1]:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \epsilon_t \ ; \ \epsilon_t \sim N(0, \sigma^2) \tag{9}$$

This implies that $y_t$ has a continuous state space, and furthermore $y_t$ is not yet a Markov process. However, defining the vector $X_t = \{y_{t-1}, \ldots, y_{t-p}\}$ ,it is clear then that $X_t$ will depend only on $X_{t-1}$. The details of going from a continuous state space autoregression to a Markov chain are described in the Appendix, but the main idea is to define a grid of $N$ points, $\{\bar{y}^1, \bar{y}^2, \ldots, \bar{y}^N\}$, which are the states that the variable $y_t$ can take. A simple way of doing this is to center this set at the mean of the variable and let its range be a multiple of the standard deviation of $y_t$. The state space for $y_t$ will be the same as the state space for any lag of $y_t$ and hence the total dimension of the "extended" state space for $X_t$ will be $N^p$. Using the notation from the previous section, models $A_1$ and $A_2$ will be any two Markov chain models that approximately describe the above process and that will differ in the choices of $p$ and $N$. The "extended" state space will be described by a transition matrix of dimension $N^p$ with typical element $\varrho_{ik}$, denoting the probability of transiting from state $i$ to state $k$. This transition matrix will satisfy the conditions for a limiting (stationary) distribution $\pi_i$, $i = 1, \ldots, N^*$ to exist, with each $\pi_i$ being a function of the $\varrho_{ik}$. With both the transition probabilities and the limiting distribution we can find any population moment: mean, variance, $jth$ order autocovariance, etc. As an example, denoting the state space for the discretized variable $y_t$ as $\{\bar{y}^1, \bar{y}^2, \ldots, \bar{y}^N\}$ the mean $m$, variance $\gamma_0$ and first autocovariance $\gamma_1$ will be:

$$m = \sum_{i=1}^{N} \pi_i \bar{y}^i \tag{10}$$

$$\gamma_0 = \sum_{i=1}^{N} \pi_i (\bar{y}^i - m)^2 \tag{11}$$

[1]The process in (9) represents the starting point for applying Tauchen's method. In other words, it represents the choice of $p$ for a Markov chain approximation. As seems natural and as it will become clear below, model D, the auxiliary model, will have the same structure: it will be an AR process with lags $p_D$.

$$\gamma_1 = \sum_{i=1}^{N} \pi_i \{ \sum_{k=1}^{N} \varrho_{ik} (\bar{y}^i - m)(\bar{y}^k - m) \} \tag{12}$$

I now need to determine what type of econometric model I will use to link any two Markov chains (i.e. I need to choose model $D$) and choose the vector of moments that will be the basis of comparison between models $A_1$ and $A_2$.

The econometric model will be a continuous state space autoregression of order $p_D$, since one would want the Markov chain to represent such a process.

An initial and sensible choice for $\theta$ would be a vector whose elements were population moments of the model , say, mean, variance and first autocorrelation. However, given that the prior distribution for the elements of $\theta$ is "flat" (see Assumption 2) it will be necessary to work with an alternative choice, namely $\theta = \{\phi_0, \phi_1, \phi_2, \ldots, \phi_{p_D}, \sigma^2\}'$[2]. These are the intercept, slopes and variance of the error term in an AR($p_D$) model with normal disturbances. Note that the elements of $\theta$ are functions of moments. Given a transition probability matrix, the relationship between consumption population moments and $\theta$ is given by:

$$\gamma_j = \begin{cases} \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \ldots + \phi_{p_D} \gamma_{j-p_D} & \text{if } j = 1, 2, \ldots \\ \phi_1 \gamma_1 + \phi_2 \gamma_2 + \ldots + \phi_{p_D} \gamma_{p_D} + \sigma^2 & \text{if } j = 0 \end{cases} \tag{13}$$

where $\gamma_j$ is the $j^{th}$ order autocovariance and,

$$\phi_0 = m\,(1 - \phi_1 - \ldots - \phi_{p_D}) \tag{14}$$

where $m$ is the population mean of the series.

The mapping from the $\varrho_{ik}$, the elements of the transition matrix, to $\theta$ should be clear now. Equations (10)-(12) and their extensions to higher order autocovariances provide the link between the $\varrho_{ik}$ and $\pi_i$, and the population moments in the Markov chain model, while the inverse Yule walker equations (13)-(14), map these population moments into elements of $\theta$. Hence, by specifying a prior distribution on the transition probabilities for a given model $A_j$ a

[2] As noted below, due to computational constraints I will not be able to work with the entire vector $\{\phi_0, \phi_1, \phi_2, \ldots, \phi_{p_D}, \sigma^2\}'$ and I will report results with different subsets of it.

prior distribution is induced on $\theta$ through the stationary distribution and the previous equations (inverse Yule Walker equations). This distribution will be the density $p(\theta \mid A_j)$. The description of where these priors come from is the purpose of the next section.

### 3.1 Priors

It has already been mentioned that $p(\theta \mid D) \propto constant$, and therefore the choice of a prior distribution affects only the parameters involved within the two Markov chain models $A_1$ and $A_2$. The final goal is to induce priors on the elements of the object of interest $\theta$ through the elements of the transition matrix $\varrho_{ik}$, to generate draws (i.e. to simulate) from the density $p(\theta \mid A_j)$. The way I proceed is to begin with a dynamic economic model that replicates certain features of the observed consumption process. I will use it to induce priors on the elements of the transition matrix and hence induce a density on the elements of $\theta$. The model chosen is a standard (Hansen (1985)) real business cycle model. This model is well known, having been studied by several economists such as Campbell (1994) and Uhlig(1999), and I will just give a short introduction, to fix notation and to understand the mapping from this model to the transition matrix. This description follows Uhlig (1999) closely.

The model has a representative agent that maximizes lifetime utility over consumption and leisure. The agent makes the standard two choices in simple macroeconomic models: how to allocate his endowment of time between labor and leisure, and what proportion of the physical good to invest and how much to consume. More formally, the maximization problem can be written as:

$$max_{\{c_t,k_t,n_t\}_{t=0}^{\infty}} E_0 \sum_{t=0}^{\infty} \beta^t \left(\frac{c_t^{1-\gamma}-1}{1-\gamma} + H(1-n_t)\right) \,,\ \beta \in (0,1)\,,\ \gamma, H > 0$$

subject to:

$$i_t + c_t = z_t k_{t-1}^{\alpha} n_t^{1-\alpha} \ ,\ 0 < \alpha < 1 \tag{15}$$

$$k_t = (1-\delta)k_{t-1} + i_t \ ,0 < \delta < 1 \tag{16}$$

$$l_t + n_t = 1 \tag{17}$$

$$lnz_{t+1} = \rho\, lnz_t + \epsilon_{t+1} \;\; \epsilon_{t+1} \sim N(0, \sigma_\epsilon^2) \;,\; 0 < \rho < 1 \tag{18}$$

$$k_0 > 0 \;\; given \tag{19}$$

The notation is standard: $c_t$, $l_t$ and $n_t$ denote consumption, leisure and employment at time $t$; labor and capital, $k_{t-1}$, are used in the production process which is also affected by a productivity shock $z_t$ whose logarithm follows a first order autoregressive process. Due to the uncertainty about the true values of the parameters $\beta$, $\gamma$, $\delta$, $\alpha$, $\rho$ and $\sigma_\epsilon^2$, I will specify prior distributions for these parameters instead of specifying fixed values. Generating draws from these distributions and solving the model for each of the draws, these priors will induce a distribution for the $\varrho_{ij}$, the parameters in the transition matrix.

To derive decision rules (policy functions) for consumption, investment and labor, I will solve the log-linearized version of the problem; that is, instead of solving for the decision rules in the original variables $c_t, i_t, n_t,$ etc.. the policy functions will be constrained to be linear in their logarithms, and will be obtained in terms of deviations from the steady state. For a variable $x_t$, let $\hat{x}_t$ denote its percentage deviation from its steady state value $\bar{x}$, i.e. $\hat{x}_t = logx_t - log\bar{x}$. In log-linear form, the resource constraint, Euler equation and law of motion for the technology shock are:

$$0 = A\hat{k}_t + B\hat{k}_{t-1} + C\hat{w}_t + D\hat{z}_t \tag{20}$$

$$0 = E_t[J\hat{w}_{t+1} + K\hat{w}_t] \tag{21}$$

$$\hat{z}_{t+1} = \rho\hat{z}_t + \epsilon_{t+1},\; \epsilon_{t+1} \sim N(0, \sigma_\epsilon^2). \tag{22}$$

The matrices in this system will be functions only of the parameters in the model. In this simple problem there is only one endogenous state variable $k_{t-1}$, one exogenous state variable $z_t$ and a vector $w_t$ of three control variables with elements $i_t$, $c_t$ and $n_t$. The solution is a set of two (matrix) equations that describe the equilibrium laws of motion for the steady state deviations for the endogenous state variable and the control variables:

$$\hat{k}_t = P\hat{k}_{t-1} + Q\hat{z}_t \tag{23}$$

$$\hat{w}_t = R\hat{k}_{t-1} + S\hat{z}_t \tag{24}$$

The solutions for the matrices $P$, $Q$, $R$ and $S$ are given using methods developed by Blanchard and Kahn (1980) and Sims (2000), also briefly described by Uhlig (1999). These matrices will be also functions of the behavioral and technology parameters $\beta, \gamma, \delta, \alpha, \rho$ and $\sigma_\epsilon^2$, and hence given a value for the initial deviation of the capital stock from its steady state value, realizations for the technology shock $z_t$, and a value for the parameters, simulation from the system (23) and (24) is straightforward.

Before discussing the prior for $\beta, \gamma, \delta, \alpha, \rho$ and $\sigma_\epsilon^2$, and the computational procedure in detail, it is important to notice that the previous model has no growth in it. By definition the mean value of the steady state deviations for any of the variables in the above system is zero. However, there is a positive trend in US post-war consumption: the mean value for annual growth is about 2 %. To deal with this problem, once a sequence $\{c_t\}_{t=0}^T$ is obtained from the model and growth rates are calculated, a value of about 0.5% per quarter is added to each realization to force the model to match the observed mean value. This process will shift the state space upward by 0.5 percent but will have no impact in the transition probabilities. [3].

To specify prior distributions for the set of parameters, I obtained means drawn from previous work and also some prior predictive analysis. I want to ensure that the model roughly matches consumption growth moments, without being too concerned how the model did in terms of matching observations for investment, employment, interest rates, etc. All parameters were endowed with (truncated) normal prior distributions, and the exact values for mean and standard deviations are given in the following table[4] :

[3] An equivalent approach would be to de-mean the time series obtained from the data, fit model D to these de-meaned values and do the entire procedure with the transformed series, which would indeed have mean zero.

[4] The value of $H$ was determined so that in a steady state the proportion of time that an individual spends working is 1/3.

Table 2

| Parameter | Prior |
|---|---|
| $\gamma$ | $N(1.1, (0.05)^2) I_{[0,5]}(\gamma)$ |
| $\beta$ | $N(0.99, (0.01)^2) I_{[0,1]}(\beta)$ |
| $\delta$ | $N(0.025, (0.002)^2) I_{[0,1]}(\delta)$ |
| $\alpha$ | $N(0.36, (0.02)^2) I_{[0,1]}(\alpha)$ |
| $\rho$ | $N(0.92, (0.01)^2) I_{[-1,1]}(\rho)$ |
| $\sigma_\epsilon$ | $N(0.012, (0.001)^2) I_{[0,\infty]}(\sigma_\epsilon)$ |

The process of simulating one draw from the density $p(\theta \mid A_j)$ can be summarized in the following steps:

- *Step 1*: Draw one value for the parameters of the business cycle model from the prior distribution given in the previous table. Solve the model and generate a time series of length 1,000 for consumption growth.

- *Step 2*: Fit an autoregression of order $p$ by least squares to this time series and take it to be the population autoregression for the induced process for consumption growth in the real business cycle model.

- *Step3*: Specify a value for $N$ (number of states in the Markov chain) and the two endpoints of the grid for consumption growth (fixed throughout simulations to be [-1%,2%],i.e. states are added in the center of the distribution), and apply Tauchen's method to the autoregression fit in Step 2, yielding a transition matrix of order $N^* = N^p$, with typical element $\varrho_{ik}$. Impose restrictions on this matrix to yield a symmetric unconditional distribution. The choice of these two elements, the number of states $N$ and the number of lags $p$ is what defines a particular Markov chain approximation $A_j$.

- *Step 4*: From the transition matrix and the grid for consumption growth obtained in Step 2, calculate population moments such as $m$, $\gamma_0$, $\gamma_1$, etc...as in equations (10)-(12), and from these, retrieve the elements of $\theta$ from the inverse Yule-Walker equations (13)-(14).

These four steps result in one draw from the density $p(\theta \mid A_j)$. The procedure is repeated $M$ times to obtain the shape of the entire density. Figure 3 shows

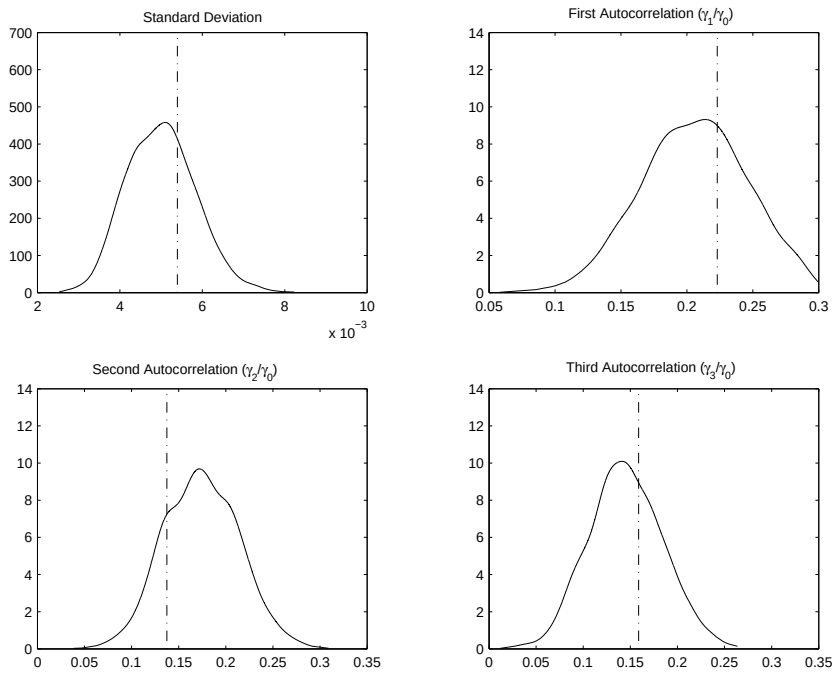


Figure 3: Priors for moments of consumption growth (solid line). Observed data value (dotted line)

how the model does (with the priors for parameters specified above) in terms of matching certain moments of consumption growth for 2,500 simulations, along with values obtained from the actual post-war consumption growth series[5]. It is important to note that results will be sensitive to the mean values in these distributions. The implied consumption growth process would look very different with substantial changes in parameters such as the discount factor, the coefficient of relative risk aversion and the variance and autocorrelation of the productivity shock.

### 3.2 Posterior Simulation

As mentioned, the posterior density $p(\theta \mid \mathbf{y}, D)$ is the density to which simulations from a given model will be compared, providing a link between two models and enabling the computation of a Bayes factor. A kernel for this

[5]When simulating the model, the first autocorrelation was constrained to be smaller than 0.3, not to have too much persistence in the series. The standard deviation was also bounded below at a value of 0.0009. These constraints can be seen as additional properties of my prior distribution for consumption growth within the RBC model.

density is given by Bayes theorem, a formula that I rewrite again here for convenience:

$$p\left(\theta \mid \mathbf{y}, D\right) \propto p\left(\mathbf{y} \mid \theta, D\right) p\left(\theta \mid D\right) \tag{25}$$

Assumption 2 ensures that this posterior will be proportional just to the likelihood given that the prior is flat.

For the comparison of Markov chain approximations it is convenient to specify an autoregression as my model $D$. In this case, I can make use of the results in Zellner (1971) and generate draws from the posterior in the following way. Suppose I specify model $D$ to be:

$$y_t = \phi_0 + \phi_1 \, y_{t-1} + \ldots + \phi_{p_D} \, y_{t-p_D} + \epsilon_t; \;\; \epsilon_t \sim N(0, \sigma^2) \tag{26}$$

Denote $\beta = (\phi_0, \phi_1, \ldots, \phi_{p_D})'$, then $\theta = (\beta, \sigma^2)'$. Given Assumption 2, in which the prior distribution for $\theta$ is constrained to be proportional to a constant, the posterior distribution is then proportional to the likelihood:

$$p\left(\beta, \sigma^2 \mid \mathbf{y}, D\right) \propto \left(\frac{1}{\sigma^2}\right)^{-T/2} \exp\left\{-\frac{1}{2\sigma^2}\left(S^2 + (\beta - \hat{\beta})' X' X (\beta - \hat{\beta})\right)\right\} I_{(s)}(\beta) \tag{27}$$

To clarify the notation in this expression, here $T$ denotes the total number of observations, $\hat{\beta}$ denotes the ordinary least squares estimate of $\beta$, $S^2$ is the total sum of squares $(\mathbf{y} - X\hat{\beta})'(\mathbf{y} - X\hat{\beta})$, $y$ is a $T$x1 vector with all the observations, $X$ is the data with $p_D$ lagged values of $y_t$ for all observations $T$, and $I_{(s)}(\beta)$ is an indicator function that assigns a value of zero to the density whenever any eigenvalue of the autoregression is outside the unit circle. From the previous expression is clear that:

$$\beta \mid \sigma^2, \mathbf{y}, D \sim MVN(\hat{\beta}, (X'X)^{-1}\sigma^2) I_{(s)}(\beta). \tag{28}$$

Since usually $\sigma^2$ is unknown I need the marginal posterior density for $\sigma^2$,

$p(\sigma^2 \mid \mathbf{y}, D)$, which is obtained by integrating out $\beta$ from the posterior:

$$p(\sigma^2 \mid \mathbf{y}, D) = \int p(\beta, \sigma^2 \mid \mathbf{y}, D) d\beta \tag{29}$$

Which yields,

$$\sigma^2 \mid \mathbf{y} \sim Inverse - \chi^2(S^2, T - d - 2)$$

In this expression $d$ is the dimension of the vector $\beta$. Hence to simulate from this posterior a draw from $\sigma^2 \mid (\mathbf{y}, D)$ is made, and conditional on this draw, $\beta$ is generated from $\beta \mid (\sigma^2, \mathbf{y}, D)$ [6].

# 4 Results

After discussing how to obtain simulations from models $D$ and $A_j$, and explaining how to obtain a marginal likelihood, I present some results in this section.

The definition of consumption used here is total real personal consumption expenditures, which is transformed to per capita terms through dividing by the total non-institutionalized population 16 years or older. Real consumption is taken from the DRI database (original source: Bureau of Economic Analysis) and population is taken from the Bureau of Labor Statistics. The sample covers the period 1947 to 2001; there are 212 quarterly observations.

A continuous state space AR of order $p_D = 3$ has been chosen as model $D$. The reason for this choice is that according to Otrok, Ravikumar and Whiteman (2001) this seems to be the best representation for post-war real consumption growth when using autoregressive models with the number of lags varying between 1 and 5.

$$\Delta c_t = \phi_0 + \phi_1 \Delta c_{t-1} + \phi_2 \Delta c_{t-2} + \phi_3 \Delta c_{t-3} + \epsilon_t \ ; \ \epsilon_t \sim N(0, \sigma^2) \tag{30}$$

[6] With this approach, each draw comes from the posterior distribution and there is no convergence (burn-in) phase, as is the case with Gibbs sampling, for example.

Throughout, the number of draws from the posterior distribution of this autoregression, and the number of simulations from the prior for the real business cycle model were both set at 2,500. [7] Posterior means and standard deviations from model D are:

Table 3

| | Mean | Standard Deviation |
|---|---|---|
| $\phi_0$ | 0.7229 | 0.1045 |
| $\phi_1$ | 0.1053 | 0.0678 |
| $\phi_2$ | 0.0843 | 0.0686 |
| $\phi_3$ | 0.0915 | 0.0678 |
| $\sigma^2$ | $4.2\text{x}10^{-5}$ | $4.2\text{x}10^{-6}$ |

It would be optimal to consider $\theta$ to be this set of five parameters. However, I need to restrict $d$ (the dimension of $\theta$) to be at most three, because kernel smoothing in higher dimensions becomes infeasible. I present results for two different $\theta$'s, one that includes only the first slope ($\phi_1$) and the variance ($\sigma^2$), and a second $\theta$ that adds also the constant term ($\phi_0$) to the previous vector. Two different bandwidths [8] $h$ are used: 0.2 and 0.5, to see how sensitive the results are to the choice of this parameter.

All results are presented in Table 4. The auxiliary model $D$ was always an autoregression of order $p_D = 3$, and only one simulation from this model was used in all calculations. Instead of presenting Bayes factors, the table shows marginal likelihoods[9] (along with numerical standard errors) for several choices of $p$, the order of the population autoregression in the Markov chain model, and $N$, the number of gridpoints in the state space.

[7] Clearly the higher the number of simulations, the more stable the kernel density estimator will be, and results will show that in some cases 2,500 is small. However, due to the amount of computation time involved, significantly higher number of draws were infeasible.

[8] The bandwidth parameter $h$ controls the degree of smoothing in the non-parametric function evaluation.

[9] The Bayes factor between any two models is the ratio of the marginal likelihoods.

Table 4: Marginal likelihoods (standard errors in parentheses)

Model D: AR(3)

| | $\theta = (\phi_1, \sigma^2)'$ | | $\theta = (\phi_0, \phi_1, \sigma^2)'$ | |
|---|---|---|---|---|
| | $h$=0.2 | $h$=0.5 | $h$=0.2 | $h$=0.5 |
| $p = 1$,N=2 | 0 | 0 | 0 | 0 |
| $p = 1$,N=3 | 131,309 (2,039) | 118,807 (1,708) | 29,113,468 (5,957,950) | 34,541,805 (5,144,584) |
| $p = 1$,N=4 | 210,530 (3,888) | 186,843 (2,804) | 63,965,657 (9,609,284) | 53,952,989 (7,038,213) |
| $p = 1$,N=5 | 234,977 (3,187) | 211,497 (2,649) | 66,445,142 (8,838,618) | 59,514,744 (7,399,384) |
| $p = 1$,N=6 | 237,108 (3,185) | 216,489 (2,711) | 65,506,560 (8,835,662) | 60,031,604 (7,404,775) |
| $p = 1$,N=7 | 237,450 (3,273) | 209,288 (2,636) | 61,595,363 (391,976) | 58,785,430 (7,281,454) |
| $p = 1$,N=8 | 227,544 (3,181) | 209,901 (2,650) | 61,226,171 (8,200,143) | 57,845,737 (7,178,380) |
| $p = 1$,N=9 | 228,329 (3,202) | 205,495 (2,574) | 60,692,195 (8,206,808) | 56,770,676 (7,024,697) |
| $p = 1$,N=10 | 223,906 (3,109) | 205,324 (2,578) | 60,210,750 (8,100,610) | 55,498,947 (6,851,060) |
| $p = 1$,N=15 | 216,040 (3,014) | 199,674 (2,516) | 58,845,009 (7,939,074) | 54,812,891 (6,773,737) |
| $p = 1$,N=20 | 215,014 (3,012) | 199,217 (2,558) | 58,114,564 (13,055,294) | 54,472,088 (6,733,999) |
| $p = 2$,N=2 | 3.9852 (0.7440) | 2.5320 (0.3894) | 0 | 0 |
| $p = 2$,N=3 | 116,698 (2,447) | 107,175 (2,126) | 58,942,434 (3,041,061) | 55,152,114 (2,539,151) |
| $p = 2$,N=4 | 255,657 (4,486) | 226,304 (3,437) | 167,479,330 (6,092,917) | 143,340,286 (4,536,992) |
| $p = 2$,N=5 | 257,694 (4,170) | 230,861 (3,398) | 156,006,072 (6,825,142) | 141,468,311 (4,639,943) |
| $p = 3$,N=2 | 267.70 (13.33) | 254.54 (12.62) | 0 | 0 |
| $p = 3$,N=3 | 103,809 (2,435) | 97,689 (2,153) | 63,496,448 (7,339,432) | 59,132,788 (2,808,938) |
| $p = 3$,N=4 | 254,414 (4,451) | 227,864 (3,561) | 158,063,706 (14,620,417) | 121,703,904 (4,641,582) |
| $p = 3$,N=5 | 248,024 (4,314) | 221,270 (3,522) | 175,380,996 (15,827,566) | 141,033,422 (4,752,142) |

For the case of $h = 0.2$, the maximum of all models considered is the combination $p = 2, N = 5$, with a marginal likelihood of $2.57 \times 10^5$ although this model is indistinguishable from the $p = 3, N = 4$ case, given the standard errors. Restricting $p$ to be 1, the maximum is attained at $N = 7$. The only case with a marginal likelihood that is zero is $p = 1, N = 2$, the reason being that since the endpoints of the state space are fixed throughout simulations, the model does not match the variance very well for lower values of $N$ (notice the small values of marginal likelihoods when $N = 2$ when $p > 1$). In addition to this, it seems that the steepest increases occur when going from $N = 2$ to $N = 4$, with further increases or decreases being fairly small.

These results are fairly robust to increasing the bandwidth to 0.5: all marginal likelihoods are smaller, however most rankings of models in terms of $p$ and $N$ remain unchanged. The only exception is the case $p = 1, N = 7$; it has the highest marginal likelihood when $h = 0.2$, but only the third highest marginal likelihood when $h = 0.5$. It might seem surprising that the highest marginal likelihood is not achieved at the highest $N$. There are two possible explanations for this: First, the dynamic model used as a prior distribution does not represent certain aspects of consumption growth correctly and this misspecification might be acting in such a way that the optimal Markov chain approximation in the RBC model (the one that has the highest $N$) is not the best when compared with model D. The second explanation might be related to the choice of $\theta$: A model might be improving when $N$ is increased, but along dimensions that are not in $\theta$, therefore giving the impression that the model is actually doing worse. As stated above, due to technical reasons the dimension of $\theta$ needs to be kept low, and to further study these issues an alternative method for comparing the resulting densities should be used.

It is interesting to look at the optimal choice among different models when the dimension of the transition matrix $N^* = N^p$ is fixed. For $N^p$=4, there are two choices: either $p = 1, N = 4$ or $p = 2, N = 2$. Clearly the choice favors $p = 1$ with a Bayes factor in the order of $5 \times 10^4$. Increasing $N^*$ to about 8 or 9, gives three possibilities: $p = 1, N = 8$, $p = 2, N = 3$ and $p = 3, N = 2$. Again

the marginal likelihood is maximized at $p = 1$, but with a Bayes factor over the $p = 2$ case of only 2. For higher number of states, however, the optimal choice leads to values of $p = 2$ or $p = 3$, but with increases in marginal likelihoods that are not too large.

What happens when $\phi_0$ is added to $\theta$? First, no model with $N = 2$ should be chosen, since now all marginal likelihoods are zero. Increasing the dimension of $\theta$ increases the errors in the estimation so that now, within any choice of $p$ models with $N \geq 4$ are always within two standard deviations from each other (with one exception: when $h = 0.5$, $p = 3, N = 5$ is clearly better than $p = 3, N = 4$) . As an example, when $p = 1$, the choice of $N$ that maximizes the marginal likelihood is 5, but its value is well within even one standard deviation of the marginal likelihood when $N = 20$.

The implications for the optimal choice of $p$ and $N$ for a given dimension of the transition matrix can be studied in an analogous way. Using the same cases as before, when $N^p$=4 the result is unchanged: $p = 1, N = 4$ performs better than $p = 2, N = 2$. When $N^*$ is increased to 8 or 9, again $p = 1$ dominates $p = 2$ or $p = 3$; and for higher number of states an increase in $p$ is needed. Overall it seems that modelling consumption growth as an autoregression of order 2 prior to its discrete state approximation yields satisfactory results with only 3 states and a total dimension for the transition probability matrix of 9.

## 5 Conclusion

Although Markov chain approximations have been used extensively in economics, the tradeoff between the two fundamental issues when modelling a time series as a Markov chain, namely, the number of lags ($p$) in the original model, and the dimension of the state space ($N$) has been (to my knowledge) ignored. In this paper I have approached this issue using a methodology that has been recently used to compare calibrated models in dynamic macroeconomics, and amounts to measuring the degree of overlap between multivariate densities of a vector of interest that the models claim to describe. The particular series chosen to address the question was US post-war-real-per-capita consumption

growth. For two different vectors of interest, optimal choices for $N$ and $p$ were calculated for several dimensions of the final transition matrix, and sensitivity analysis to the bandwidth parameter in the kernel smoothing was done. In this case, it turned out to be advantageous to sacrifice some detail in the state space ($N$) in order to obtain a better representation of the dynamics. However, further research is needed to overcome technical difficulties pertaining to the way multivariate densities are compared, and an alternative to kernel smoothing would be desirable to increase the dimension of the vector of interest and be able to compare models along a higher number of dimensions.

## References


[1] Aiyagari, R (1994),"Uninsured Idiosyncratic Risk and Aggregate Saving", *Quarterly Journal of Economics*, Vol. 109, No. 3, pp. 659-684.

[2] Blanchard,O.J and Kahn, C.M , " The Solution of Linear Difference Models under Rational Expectations " *Econometrica*, Vol. 48, No. 5. (Jul.,1980), pp. 1305-1312.

[3] Campbell, J. Y.(1994), "Inspecting the Mechanism: An Analytical Approach to the Stochastic Growth Model", *Journal of Monetary Economics*, **33**, no. 3 (June): 463-506.

[4] Canova, F. and Ortega, E. (2000), "Testing Calibrated General Equilibrium Models",in *Simulation-Based Inference in Econometrics*, Mariano, R., Schuerman, T., and Weeks, M,(eds.), Cambridge University Press.

[5] Cechetti, S., Lam,P. and Mark, N., (1994), "The Equity Premium and the Risk Free Rate: Matching the Moments", *Journal of Monetary Economics*, **31**, 21-45.

[6] Cooley, T. and Prescott, E. (1995), "Economic Growth and Business Cycles", in *Frontiers of Business Cycle Research*, Cooley, T. (ed.), Princeton University Press.

[7] DeJong, D., Ingram, B. and Whiteman, C. (1996), "A Bayesian Approach to Calibration", *Journal of Business and Economics Statistics*, **14**:1-10.

[8] Geweke, J. (1999), "Computational Experiments and Reality", Federal Reserve Bank of Minneapolis Staff Report.

[9] Geweke, J. (2002),*Contemporary Bayesian Econometrics and Statistics*, in press.

[10] Gregory A., and Smith (1991), "Calibration as Testing: Inference in Simulated Macroeconomic Models", *Journal of Business and Economics Statistics*.

[11] Mehra, R. and Prescott, E. (1985),"The Equity Premium: A Puzzle", *Journal of Monetary Economics*, 10, pp. 335-359.

[12] Otrok, C. Ravikumar, B., and Whiteman, C. (2001), "Evaluating Asset Pricing Models Using the Hansen-Jagannathan Bound: A Monte Carlo Investigation", *Journal of Applied Econometrics*, forthcoming.

[13] Silverman, B. (1986), *Density Estimation for Statistics and Data Analysis*, Monographs in Statistics and Applied Probability, Chapman and Hall.

[14] Sims, C. (2000), "Solving Linear Rational Expectations Models", manuscript, Princeton University.

[15] Tauchen, G. (1986), "Finite State Markov-Chain Approximations to Univariate and Vector Autoregressions", *Economic Letters* **20**:177-181.

[16] Tauchen, G. and Hussey, (1991),"Quadrature-Based Methods for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models", *Econometrica*,Vol. 59, No. 2.

[17] Uhlig, H. (1999), "A toolkit for analysing nonlinear dynamic stochastic models easily," in Ramon Marimon and Andrew Scott, eds, *Computa-*

*tional Methods for the Study of Dynamic Economies*, Oxford University Press, Oxford, 30-61.

[18] Zellner, A. (1971), "An Introduction to Bayesian Inference in Econometrics", John Wiley and Sons.