Title
Methods to Estimate Dynamic Stochastic General Equilibrium Models

Author
Ruge-Murcia, Francisco J.

Publication Date
2002-10-01

Methods to Estimate Dynamic Stochastic General Equilibrium Models*

Francisco J. Ruge-Murcia

Département de sciences économiques and C.I.R.E.Q.,
Université de Montréal

First Draft: August 2002
This Copy: October 2002

Abstract

This paper employs the one-sector Real Business Cycle model as a testing ground for four different procedures to estimate Dynamic Stochastic General Equilibrium (DSGE) models. The procedures are: 1) Maximum Likelihood (with and without measurement errors and incorporating priors), 2) Generalized Method of Moments, 3) Simulated Method of Moments, and 4) the Extended Method of Simulated Moments proposed by Smith (1993). Monte Carlo analysis shows that although all procedures deliver reasonably good estimates, there are substantial differences in statistical and computational efficiency in the small samples currently available to estimate DSGE models. The implications of the singularity of DSGE models for each estimation procedure are fully discussed.

JEL Classification: E13, C11, C13, C15, C32
Key Words: DSGE models, estimation methods, Monte Carlo analysis, singularity

*This paper represents work in progress and is circulated for comments only. I have benefitted from conversations with Rui Castro, Wouter den Haan, and Jim Hamilton. This project was completed while I was Visiting Professor at the University of California at San Diego. I wish to thank the Department of Economics at UCSD for its hospitality. Financial support from the Social Sciences and Humanities Research Council and the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche is gratefully acknowledged. *Correspondence:* Départemente de sciences économiques, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal (Québec) H3C 3J7, Canada. E-mail: francisco.ruge-murcia@umontreal.ca

1 Introduction

Dynamic Stochastic General Equilibrium (DSGE) models have become a standard tool in various fields of Economics, most notably in Macroeconomics and International Economics. DSGE models are attractive because they explicitly specify the objectives and constraints faced by households and firms, and then determine the prices and allocations that result from their market interaction in an uncertain environment.

To date, calibration is by far the most common approach in the literature to examine the empirical properties of DSGE models. In calibration, the value of the structural parameters is fixed to those estimated in previous microeconometric studies and/or those computed using long-run averages of aggregate data. Then, the model is simulated using a synthetic series of shocks, and the unconditional moments of the simulated economic series are computed and compared with the ones of actual data. The model is usually evaluated in terms of the distance between these two set of moments. This comparison can be casual or based on measures of fit like the ones proposed, for example, by Gregory and Smith (1991), Watson (1993), and DeJong, Ingram, and Whiteman (1996). Impulse-response analysis and variance decomposition are also used to examine, respectively, the model's behavior following exogenous shocks and to assess the relative importance of these shocks in explaining the conditional and unconditional variances of the variables.

Although calibration is a very useful tool for understanding the dynamic properties of DSGE models, there are some advantages in their fully-fledged econometric estimation. *First,* parameter estimates are obtained by imposing on the data the restrictions of the model of interest. This addresses the concern that the assumptions of the DSGE model might be inconsistent with the assumptions employed by the micro studies that produced the parameter estimates used in calibration. *Second,* the estimation of the DSGE model allows one to obtain estimates of parameters that might be hard to estimate using disaggregated data alone. *Third,* parameter uncertainty can be explicitly incorporated in impulse-response analysis using, for example, bootstrap techniques to construct confidence intervals for the model's response to a shock. *Finally,* standard tools of model selection and evaluation can be readily applied. For example, one can test the residuals for serial correlation and neglected Autoregressive Conditional Heteroskedasticity, compare the Root Mean Square Error of the DSGE model with that of another DSGE model or a Vector Autoregression, perform tests of parameter stability or directly test some of the model's identification assumptions. All this is valuable information in the construction of more realistic economic models.[1]

[1] See Hansen and Heckman (1996), and Browning, Hansen, and Heckman (1999) for additional discussion. For a defense of the merits of calibration, see Kydland and Prescott (1996).

The estimation procedures studied here are Maximum Likelihood (ML), Generalized Method of Moments (GMM), Simulated Method of Moments (SMM), and the Extended Method of Simulated Moments (EMSM) proposed by Smith (1993). These procedures are standard and their asymptotic properties are well know. The goals of this paper are to describe in a pedagogical manner their application to the estimation of DSGE models, to study their small sample properties, to compare their computational costs, and to discuss fully the implications of the singularity of DSGE models for each estimation procedure. The intention here is not to perform a "horse race" between different estimation strategies. Instead, the more constructive goal is to evaluate their relative strengths and weaknesses in the context of a simple, but economically interesting model.

An important feature of DSGE models that has implications for all estimation procedures is their singularity. DSGE models are usually singular because they generate predictions about a larger number of observable endogenous variables than exogenous shocks are used to feed the model. This means that there are linear combinations of the variables that hold without noise. These restrictions of the theoretical model arise from a particular form of misspecification: the model assumes a smaller number of shocks than are present in the real world.

Singularity limits the number of variables/moments that can be exploited for the estimation of the model. For the particular DSGE model studied here, it is shown that singularity affects more severely the method of Maximum Likelihood than the methods of moments. The model cannot be estimated by ML using more than one variable, unless additional errors are added. However, one can use data on up to two endogenous variables for the methods of moments. This means that the methods of moments might yield more precise estimates of the structural parameters than ML, despite the fact that they are limited information procedures. It is also shown that singularity imposes restrictions on both the order and the number of variables in a VAR representation of the data generated by a DSGE model.

The paper also studies the effect of adding measurement errors and incorporating priors on ML estimation. Adding measurement error provides a less stringent platform to assess the theory and allows the researcher to exploit information on a larger set of variables to estimate the model parameters. Monte Carlo results indicate that this strategy delivers more efficient parameter estimates than alternative procedures. The use of priors allows the research to exploit information from previous microeconometric studies, long-run averages of aggregate data, and/or economic theory. Priors are incorporated here into the Maximum Likelihood framework using the mixed estimation strategy in Theil and Goldberger (1961) and are shown to yield sharper estimates than those obtained by the classic ML estimator.

The paper is organized as follows. Section 2 describes the DSGE model that will be used as backdrop for the estimation procedures. Section 3 describes the estimation procedures and their application to DSGE models. Section 4 presents the Monte Carlo design and report its results. Section 5 concludes.

2 The Artificial Economy

The discussion of the different estimation procedures is best made in the context of a specific economic model. This paper employs a version of the well-known one-sector Real Business Cycle model with indivisible labor [see Hansen (1985), and King, Plosser and Rebelo (1988)].[2] The representative agent maximizes expected lifetime utility defined by:

$$U_t = E_t \sum_{i=t}^{\infty} \beta^{i-t} \left[\ln(c_i) + \psi \ell_i \right],$$

where $\beta \in (0,1)$ is the subjective discount factor, c_t is consumption, ℓ_t is leisure, and ψ is the weight of leisure in the instantaneous utility function. There is no population growth. Without loss of generality, the population size and time endowment are normalized to one. Hence,

$$n_t = 1 - \ell_t,$$

where n_t is hours worked. The agent's income consists of wages and rents received from selling labor and renting capital to firms, and is allocated to consumption and investment:

$$c_t + x_t = w_t n_t + r_t k_t,$$

where x_t is investment, w_t is the real wage, r_t is the real rental rate of capital, and k_t is the capital stock. The prices w_t and r_t are expressed in terms of units of the consumption good. Investment increases the stock of capital according to:

$$k_{t+1} = (1-\delta)k_t + x_t,$$

where $\delta \in (0,1)$ is the depreciation rate. In addition to the transversality condition, the first-order necessary conditions associated with the optimal choice of c_t, n_t,and k_{t+1} for this problem are:

$$\begin{aligned} 1/c_t &= \beta E_t[(1/c_{t+1})(1 + r_t - \delta)], \\ \psi c_t &= w_t. \end{aligned}$$

[2] In preliminary work, I performed a limited number of Monte-Carlo experiments using a more complicated DSGE model with monopolistic competition, price rigidity, adjustment costs to capital, and money in the utility function. Conclusions are qualitatively similar to those reported here, but the complexity of the model obscures some of the points illustrated below.

The single, perishable good in this economy is produced by perfectly competitive firms using a constant returns to scale technology. Since in this setup the number of firms in equilibrium is indeterminate, it is convenient to focus on a representative firm. The representative firm rents labor and capital from the agent and combines them according to:

$$y_t = z_t(k_t)^{\alpha}(n_t)^{1-\alpha},$$

where $\alpha \in (0,1)$, y_t is output and z_t is a technology shock. The technology shock follows the exogenous stochastic process:

$$\ln z_{t+1} = \rho \ln z_t + \epsilon_t,$$

where $\rho \in (-1,1)$ and ϵ_t is an innovation assumed to be independently, identically, and Normally distributed with zero mean and variance σ^2. In every period, the firm chooses input levels to maximize profits and equates the marginal product of labor (capital) to the real wage (rental rate). Due to the assumptions of perfect competition and constant returns to scale, firms make zero profits in equilibrium.

The competitive equilibrium for this economy is the sequence of prices $\{w_t,\, r_t\}_{t=0}^{\infty}$ and allocations $\{c_t,\, n_t,\, x_t,\, k_{t+1},\, y_t\}_{t=0}^{\infty}$ such that firms maximize profits, agents maximize utility, and all markets clear. In particular, goods market clearing requires that aggregate output be equal to aggregate demand:

$$y_t = c_t + x_t.$$

The usual strategy to solve DSGE models involves the linearization of first-order conditions and constraints by means of a first-order Taylor series expansion around the model's steady state. The linearized equations for this model are presented in the Appendix. These equations form a dynamic system that determines the path of the six variables in the model, namely consumption, capital, output, investment, hours worked, and the technology shock.[3] Using the circumflex to denote percentage deviation from steady state and after some manipulations, it is possible to write:

$$\begin{bmatrix} \hat{k}_{t+1} \\ E_t\hat{c}_{t+1} \end{bmatrix} = \mathbf{A} \begin{bmatrix} \hat{k}_t \\ \hat{c}_t \end{bmatrix} + \mathbf{B}\hat{z}_t,$$

where

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{bmatrix} = \begin{bmatrix} 1+\delta\gamma/(1-\gamma) & -\delta(1+\alpha\gamma-\alpha)/(\alpha-\alpha\gamma) \\ 0 & \alpha/(\varsigma+\alpha-\alpha\varsigma) \end{bmatrix}, \\ \mathbf{B} &= \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \delta/(\alpha-\alpha\gamma) \\ \varsigma\rho/(\varsigma+\alpha-\alpha\varsigma) \end{bmatrix}, \end{aligned}$$

[3] For convenience and without loss of generality, I have substituted out the wage and rental rate by equating them to the marginal products of labor and capital, respectively.

$\varsigma = \alpha\beta(k/n)^{\alpha-1}$, $k/n = ((1/\beta + \delta - 1)/\alpha)^{1/(\alpha-1)}$ is the steady-state capital-labor ratio, $\gamma = 1 - \delta(k/n)^{1-\alpha}$ is the steady-state consumption-output ratio, and variables without time subscript denote steady state values. The rational expectations solution of this system can be found using, for example, the approaches proposed by Blanchard and Kahn (1980) and Sims (1997) to obtain:

$$\hat{k}_{t+1} = a_{11}\hat{k}_t + a_{12}\hat{c}_t + b_1\hat{z}_t, \tag{1}$$

$$\hat{c}_t = \phi_{ck}\hat{k}_t + \phi_{cz}\hat{z}_t, \tag{2}$$

where ϕ_{ck} and ϕ_{cz} are combinations of the eigenvectors and eigenvalues of the matrix $\mathbf{A}$ and, consequently, depend nonlinearly on the structural parameters.

In what follows, it is convenient to define the 2×1 vector $\boldsymbol{\xi}_t = (\hat{k}_t, \hat{z}_t)'$ that collects the state variables of the system, and the 3×1 vector $\mathbf{s}_t = (\hat{y}_t, \hat{n}_t, \hat{c}_t)'$ that contains the observable variables that the researcher will use later in the estimation of the model. Using the linearized equations of the model, it is possible to write the components of $\mathbf{s}_t$ as functions of the capital stock and technology shock alone:

$$\mathbf{s}_t = \begin{bmatrix} \hat{y}_t \\ \hat{n}_t \\ \hat{c}_t \end{bmatrix} = \boldsymbol{\Phi}\boldsymbol{\xi}_t = \begin{bmatrix} \phi_{yk} & \phi_{yz} \\ \phi_{nk} & \phi_{nz} \\ \phi_{ck} & \phi_{cz} \end{bmatrix} \begin{bmatrix} \hat{k}_t \\ \hat{z}_t \end{bmatrix}. \tag{3}$$

The last row of $\mathbf{s}_t$ (*i.e.,* the expression for consumption) reproduces exactly equation (2). Finally, notice that the elements of the 3×2 matrix $\boldsymbol{\Phi}$ are nonlinear functions of the structural parameters of the model.

This model takes as input the predetermined level of capital and one exogenous shock, and generates predictions about (at least) three observable endogenous variables, namely output, consumption and hours worked. Since the number of shocks is less than the number of endogenous variables, there are linear combinations of these variables that are predicted to hold without noise. For example, one can eliminate both $\hat{k}_t$ and $\hat{z}_t$ from (3) to obtain:

$$(\phi_{yk}\phi_{cz} - \phi_{yz}\phi_{ck})\hat{n}_t + (\phi_{nz}\phi_{ck} - \phi_{nk}\phi_{cz})\hat{y}_t - (\phi_{nz}\phi_{yk} - \phi_{yz}\phi_{nk})\hat{c}_t = 0. \tag{4}$$

Similarly, using the equations for $\hat{y}_t$ and $\hat{c}_t$ in (3) and the linearized law of motion for capital, it is possible to write

$$\begin{array}{l} [\phi_{yz} + \delta\gamma(\phi_{yz}\phi_{ck} - \phi_{yk}\phi_{cz})/(1-\gamma)]\hat{c}_t - (1-\delta)\phi_{yz}\hat{c}_{t-1} \\ -[\phi_{cz} + \delta(\phi_{yz}\phi_{ck} - \phi_{yk}\phi_{cz})/(1-\gamma)]\hat{y}_t + (1-\delta)\phi_{cz}\hat{y}_{t-1} = 0. \end{array} \tag{5}$$

Combining (4) and (5), it is easy to show that the systems ($\hat{y}_t$, $\hat{n}_t$, $\hat{y}_{t-1}$, $\hat{n}_{t-1}$) and ($\hat{n}_t$, $\hat{c}_t$, $\hat{n}_{t-1}$, $\hat{c}_{t-1}$) are also singular. That is, for any sample size, the sample variance-covariance matrix of these systems are singular. We will see below that the singularity of DSGE models has different and nontrivial implications for each estimation procedure.

3 Estimation Methods

3.1 Maximum Likelihood

The Maximum Likelihood (ML) estimation of DSGE models is complicated by the fact that they frequently involve unobserved or poorly measured state variables (*e.g,* the capital stock). However, it is possible to exploit the recursive nature of the model and its fully-specified laws of motion to apply filtering techniques like the one proposed by Kalman (1960). The Kalman filter allows the construction of inferences about the unobserved state vector and permits the evaluation of the joint likelihood function of observable endogenous variables. In turn, the maximization of this likelihood function yields consistent and asymptotically normal estimates of the structural parameters of the model. This approach has been employed by, among others, McGrattan (1994), Hall (1996), McGrattan, Rogerson, and Wright (1997), Ireland (1999, 2001), Kim (2000), Bouakez, Cardia, and Ruge-Murcia (2001), and Dib and Phaneuf (2001).[4]

The state-space representation of the DSGE model above consists of the following state and observation equations. The state equation is constructed by substituting (2) into (1) to obtain the law of motion of $\hat{k}_{t+1}$ in terms of $\hat{k}_t$ and $\hat{z}_t$ only, and by using the linearized process of the technology shock (see the last equation in the Appendix). Then, it is possible to write:

$$\boldsymbol{\xi}_{t+1} = \mathbf{F}\boldsymbol{\xi}_t + \mathbf{v}_{t+1},$$

where

$$\mathbf{F} = \begin{bmatrix} a_{11} + a_{12}\phi_{ck} & a_{12}\phi_{cz} + b_1 \\ 0 & \rho \end{bmatrix},$$

is a 2×2 matrix and $\mathbf{v}_t = (0, \boldsymbol{\epsilon}_t)'$ is a 2×1 vector. The observation equation consists of the process of one of the observable endogenous variables in (3):

$$\mathbf{x}_t = \mathbf{h}\mathbf{s}_t = \mathbf{h}\boldsymbol{\Phi}\boldsymbol{\xi}_t = \mathbf{H}\boldsymbol{\xi}_t,$$

where $\mathbf{h}$ is a 1×3 selection vector.[5] For example, in the case where the model is estimated using output data alone, $\mathbf{h} = (1, 0, 0)$ and $\mathbf{x}_t = \hat{y}_t$.

Let us collect the structural parameters of the model in the $q \times 1$ vector $\boldsymbol{\theta}$. Denote by $\aleph_{t-1} = (\mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \ldots, \mathbf{x}_1)$ the set of past observations of $\mathbf{x}_t$, by $\tilde{\boldsymbol{\xi}}_{t|t-1}$ the time $t-1$ forecast

[4] Hansen and Sargent (1980) propose a ML procedure for the estimation of dynamic linear rational expectations models. However, their procedure is not designed to deal with unobserved state variables and was originally proposed for the estimation of partial equilibrium models. Christiano (1988), Altug (1989), and Bencivenga (1992) employ variants of Sargent and Hansen's method to estimate general equilibrium models. Fernández-Villaverde and Rubio (2002) propose the use of a nonlinear Sequential Monte Carlo filter to evaluate the log likelihood function of the nonlinearized DSGE model.

[5] The reason why $\mathbf{x}_t$ contains only one variable will become clearer below.

of $\boldsymbol{\xi}_t$ constructed on the basis of $\aleph_{t-1}$, and by $\mathbf{P}_{t|t-1}$ the Mean Square Error of this forecast. Then, under the assumption that the technology innovation, ϵ_t, is normally distributed, the density of $\mathbf{x}_t$ conditional on $\aleph_{t-1}$ is

$$f(\mathbf{x}_t|\aleph_{t-1};\boldsymbol{\theta}) = N(\mathbf{H}\tilde{\boldsymbol{\xi}}_{t|t-1}, \mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}').$$

The Maximum Likelihood estimator of $\boldsymbol{\theta}$ is

$$\tilde{\boldsymbol{\theta}}_{ml} = \max_{\{\boldsymbol{\theta}\}} L(\boldsymbol{\theta}), \tag{6}$$

where $L(\boldsymbol{\theta})$ denotes the log likelihood function:

$$\begin{aligned} L(\boldsymbol{\theta}) = & -(T/2)\ln(2\pi) - (1/2)\ln|\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}'| \\ & -(1/2)\sum_{i=1}^{T}(\mathbf{x}_t - \mathbf{H}\tilde{\boldsymbol{\xi}}_{t|t-1})'(\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}')^{-1}(\mathbf{x}_t - \mathbf{H}\tilde{\boldsymbol{\xi}}_{t|t-1}), \end{aligned}$$

and T is the sample size.

Since the process of $\boldsymbol{\xi}_t$ is stationary by construction, the Kalman filter recursion can be started with the unconditional moments $\tilde{\boldsymbol{\xi}}_{1|0} = E(\boldsymbol{\xi}_t) = (0,0)'$ and $\mathbf{P}_{1|0} = E(\boldsymbol{\xi}_t\boldsymbol{\xi}_t')$. The subsequent forecasting and updating of $\boldsymbol{\xi}_t$ and the computation of the MSE of $\tilde{\boldsymbol{\xi}}_{t|t-1}$ are obtained using the Kalman algorithm described, for example, in Hamilton (1994, ch. 13). Under standard regularity conditions [see Judge *et al.* (1985, p. 178)], the ML estimator is consistent and asymptotically normal:

$$\sqrt{T}(\tilde{\boldsymbol{\theta}}_{ml} - \boldsymbol{\theta}) \rightarrow N(\mathbf{0}, (\Im/T)^{-1}),$$

where $\Im = -E(\partial^2 L(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\ \partial\boldsymbol{\theta}')$ is the information matrix. In the Monte Carlo, $\Im$ is estimated using the numerically computed Hessian of the log likelihood function at the optimum.

Since there is only one random shock in this model, any attempt to estimate it by Maximum Likelihood using data on more than one variable yields a singular $\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}'$ matrix. To see why, write the innovation in $\mathbf{x}_t$:

$$\mathbf{x}_t - E(\mathbf{x}_t|\aleph_{t-1}) = \mathbf{H}(\boldsymbol{\xi}_t - \tilde{\boldsymbol{\xi}}_{t|t-1}),$$

and note that when $\mathbf{x}_t$ contains more than one variable, the innovations to the variables in $\mathbf{x}_t$ (say, consumption and output) are perfectly correlated and proportional to the technology shock. More generally, DSGE models cannot be estimated by ML using more observable variables than structural shocks are specified in the model.

While this DSGE model cannot be estimated by ML using more than one variable, we will see below that for the methods of moments examined here it is possible to use data on

up to two endogenous variables. In this sense, the singularity of DSGE models affects the method of Maximum Likelihood more severely than the methods of moments.

There are two strategies to deal with the singularity of DSGE models in the Maximum Likelihood framework. *First,* one can estimate the model using just as many observable variables as structural shocks. This strategy is followed by Kim (2000), Ireland (2001), Bouakez, Cardia, and Ruge-Murcia (2001), and Dib and Phaneuf (2001). *Second,* one can add error terms to the observation equation of the state-space representation as in McGrattan (1994), Hall (1996), McGrattan, Rogerson, and Wright (1997) and Ireland (1999).

3.1.1 Adding Measurement Errors

Adding extra error terms to the observation equation of the state-space representation of the DSGE model yields:

$$\mathbf{x}_t = \mathbf{h}\mathbf{s}_t + \mathbf{u}_t = \mathbf{h}\mathbf{\Phi}\boldsymbol{\xi}_t + \mathbf{u}_t = \mathbf{H}\boldsymbol{\xi}_t + \mathbf{u}_t,$$

where $\mathbf{x}_t$ is now a $d \times 1$ vector, $d = 1, 2,$ or 3 is the number of observable variables used to estimate the model, $\mathbf{h}$ is a $d \times 3$ selection matrix, and $\mathbf{u}_t$ is a $d \times 1$ vector of shocks assumed independently, identically, and Normally distributed with zero mean and variance-covariance matrix $E(\mathbf{u}_t\mathbf{u}_t') = \mathbf{R}$. Based on Sargent (1989), it is customary to interpret $\mathbf{u}_t$ as measurement error. It is further assumed that $E(\mathbf{v}_i\mathbf{u}_j') = \mathbf{0}$ for all $i \geq j$, meaning that measurement error contains no information about current or future technology shocks. The extension to serially correlated errors is straightforward and is discussed in Hansen and Sargent (1998, ch. 8).

As before, the ML estimator of $\boldsymbol{\theta}$ is

$$\tilde{\boldsymbol{\theta}}_{ml} = \max_{\{\boldsymbol{\theta}\}} \; L(\boldsymbol{\theta}), \tag{7}$$

but now the log likelihood function is:

$$\begin{aligned} L(\boldsymbol{\theta}) = \; & -(T/2)\ln(2\pi) - (1/2)\ln|\mathbf{H}\mathbf{P}'_{t|t-1}\mathbf{H}' + \mathbf{R}| \\ & -(1/2)\sum_{i=1}^{T}(\mathbf{x}_t - \mathbf{H}\boldsymbol{\xi}_t)'(\mathbf{H}\mathbf{P}'_{t|t-1}\mathbf{H}' + \mathbf{R})^{-1}(\mathbf{x}_t - \mathbf{H}\boldsymbol{\xi}_t). \end{aligned}$$

Notice that since (by assumption) $\mathbf{R}$ is full rank, then $\mathbf{H}\mathbf{P}'_{t|t-1}\mathbf{H}' + \mathbf{R}$ will no longer be singular when the number of observable variables included in $\mathbf{x}_t$ is larger than the number of structural shocks.

One drawback of this approach is that measurement error lacks a truly structural interpretation and it essentially represents specification error. One can think of the singularity of DSGE models as arising from a particular form of misspecification: in the real world

there are many more types of shocks than the ones assumed by the model. This is the main specification error that is captured by the term $\mathbf{u}_t$. When $\mathbf{u}_t$ is modeled as serially correlated [for example as in Ireland (1999)], other forms of specification error can be also captured by this term.

On the other hand, it can be argued that adding extra error terms is just a simple strategy to deal with misspecification. Relationships like (4) and (5) might hold approximately in the data, even if not without noise as predicted by the DSGE model. Thus, adding measurement error provides a less stringent platform to assess the theory. On the empirical side, adding measurement error allows one to exploit information on a larger set of variables to estimate $\boldsymbol{\theta}$ and, as we will see below, delivers more efficient parameter estimates than alternative procedures.

3.1.2 Incorporating Priors

Economic theory, previous microeconometric studies, and long-run averages of aggregate data can be informative about the parameter values in structural macroeconomic models. This prior knowledge about $\boldsymbol{\theta}$ can be represented in a prior density and combined with aggregate time series data to obtain a posterior density of $\boldsymbol{\theta}$. The posterior density summarizes our knowledge about $\boldsymbol{\theta}$ after observing the data and is the basis of probabilistic statements regarding the structural parameters. There is a sense in which calibration can be interpreted as a Bayesian procedure where the prior density of $\boldsymbol{\theta}$ is degenerate and concentrated on a single numerical value. With such a strong prior, observations of the data series contribute nothing to our knowledge of the parameter values and the posterior density coincides with the prior one.

A simple way to incorporate priors into the Maximum Likelihood framework is based on the mixed estimation strategy in Theil and Goldberger (1961). This approach was originally developed for the linear regression model and leads to a Generalized Least Squares (GLS) estimator that incorporates optimally prior information regarding the parameter values. It is easy to show that the mean and variance of this GLS estimator corresponds exactly to mean and variance of the Bayesian posterior distribution [see Hamilton (1994, p. 359)]. Stone (1954) gives a Maximum Likelihood interpretation to the same estimator. DeJong, Ingram, and Whiteman (2000) and Chang, Gomes, and Schorfheide (2002) incorporate priors in the estimation of DSGE models and use (respectively) importance sampling and the Metropolis-Hastings algorithm to compute numerically the moments of the posterior distribution.

For the mixed estimation strategy, write the prior distribution of the parameters as

$$\boldsymbol{\mu} = \mathbf{K}\boldsymbol{\theta} + \mathbf{e}, \tag{8}$$

where $\boldsymbol{\mu}$ is $q \times 1$ vector, $\mathbf{K}$ is a known $q \times q$ nonsingular matrix, and $\mathbf{e}$ is $q \times 1$ vector of random errors assumed Normally distributed with zero mean, variance-covariance matrix $\boldsymbol{\Sigma}$, and independent of $\mathbf{v}_t$ and $\mathbf{u}_t$. The matrix $\boldsymbol{\Sigma}$ is assumed known and represents the researcher's uncertainty about the prior information. This specification of the prior distribution is general in that it allows the characterization of the priors in terms of linear combinations of the parameters, and permits correlations across priors in the form of nonzero elements in the off-diagonal of $\boldsymbol{\Sigma}$. In the special case where $\mathbf{K}$ is diagonal, the prior would take the familiar form $f(\boldsymbol{\theta}) = N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

The mixed estimation strategy interprets the prior information as a set of q additional observations of $\boldsymbol{\mu}$ and combines them with the sample of T observations of the data $\aleph_T = (\mathbf{x}_T, \mathbf{x}_{T-1}, \mathbf{x}_{T-2}, \ldots, \mathbf{x}_1)$ to obtain an estimate of $\boldsymbol{\theta}$ as in:

$$\tilde{\boldsymbol{\theta}}_{qb} = \max_{\{\boldsymbol{\theta}\}} \; L(\boldsymbol{\theta}) + L(\boldsymbol{\mu}(\boldsymbol{\theta})). \tag{9}$$

where $L(\boldsymbol{\mu}(\boldsymbol{\theta}))$ is the log of the density of $\boldsymbol{\mu}$ in (8) and $L(\boldsymbol{\theta})$ was defined above. For the Monte Carlo, I will treat this quasi-Bayesian estimate of $\boldsymbol{\theta}$ as asymptotically normally distributed with variance-covariance matrix $(\Im)^{-1}$ and estimate the information matrix using the numerically computed Hessian at the optimum.

Notice that the estimator defined in (9) corresponds to the mode of the log of the posterior distribution $f(\boldsymbol{\theta}|\aleph_T)$. However, under the assumption of Gaussianity, the mode corresponds exactly to the mean and, consequently, the point estimate of $\boldsymbol{\theta}$ obtained here is the same as the one obtained using the approach in DeJong, Ingram, and Whiteman (2000). In the special case where the prior is diffuse or improper, the estimator $\tilde{\boldsymbol{\theta}}_{qb}$ converges to the classical ML estimator. When the prior is proper, the quasi-Bayesian estimate of $\boldsymbol{\theta}$ can be interpreted as the one obtained by the maximization of a penalized log likelihood function. The penalty $L(\boldsymbol{\mu}(\boldsymbol{\theta}))$ depends on the strength of the researcher's prior about $\boldsymbol{\theta}$ and has the effect of "pulling" the estimator towards the mean of the prior density.

3.2 Simulated Method of Moments

In calibration, the researcher computes the unconditional moments of synthetic series simulated using given parameter values and then compares them with the unconditional moments of the data. The Simulated Method of Moments (SMM) estimator pursues this idea further by updating the parameter values in a manner that reduces the distance between these two sets moments. SMM estimators have been proposed by McFadden (1989) and Pakes and Pollard (1989) to estimate discrete-choice problems and by Lee and Ingram (1991) and Duffie and Singleton (1993) to estimate time-series models.

Let us define $\mathbf{m}_t$ to be a $p \times 1$ vector of empirical observations on variables whose moments are of interest. Elements of $\mathbf{m}_t$ could include, for example, $\hat{y}_t\hat{c}_t$, $\hat{y}_t\hat{y}_{t-1}$, etc. Define $\mathbf{m}_i(\boldsymbol{\theta})$ to be the synthetic counterpart of $\mathbf{m}_t$ whose elements are computed on the basis of artificial data generated by the DSGE model using parameter values $\boldsymbol{\theta}$. The sample size is denoted by T and the number of observations in the artificial time series is τT. The SMM estimator, $\tilde{\boldsymbol{\theta}}_{smm}$, is the value of $\boldsymbol{\theta}$ that solves

$$\min_{\{\boldsymbol{\theta}\}} \quad \mathbf{G}(\boldsymbol{\theta})'\mathbf{W}\mathbf{G}(\boldsymbol{\theta}), \tag{10}$$

where

$$\mathbf{G}(\boldsymbol{\theta}) = (1/T)\sum_{t=1}^{T}\mathbf{m}_t - (1/\tau T)\sum_{i=1}^{\tau T}\mathbf{m}_i(\boldsymbol{\theta}),$$

is a $p \times 1$ vector, and $\mathbf{W}$ is the optimal weighting matrix

$$\mathbf{W} = \lim_{T \to \infty} Var\left((1/\sqrt{T})\sum_{t=1}^{T}\mathbf{m}_t\right)^{-1}. \tag{11}$$

Notice that, as usual, by using the optimal weighting matrix in minimizing $\mathbf{G}(\boldsymbol{\theta})'\mathbf{W}\mathbf{G}(\boldsymbol{\theta})$, a larger weight is given to the moments, or the linear combinations of moments, that are most informative.

Under the regularity conditions in Duffie and Singleton (1993),

$$\sqrt{T}(\tilde{\boldsymbol{\theta}}_{smm} - \boldsymbol{\theta}) \to N(\mathbf{0},(1+1/\tau)(\mathbf{D}'\mathbf{W}^{-1}\mathbf{D})^{-1}),$$

where $\mathbf{D} = E(\partial\mathbf{m}_i(\boldsymbol{\theta})/\partial\boldsymbol{\theta})$ is a $q \times p$ matrix assumed to be finite and of full rank. In the Monte Carlo, the derivatives $\partial\mathbf{m}_i(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ are computed numerically and the expectation approximated by the average over the simulated τT data points. $\mathbf{W}$ is computed using the Newey-West estimator with a Barlett kernel.

An advantage of method of moments estimators is that a general specification test of the DSGE model can be constructed using the chi-square test proposed by Hansen (1982). As it is well known, this test can be applied provided that the model be overidentified, meaning $p > q$. In the case of SMM:

$$T(1+1/\tau)[\mathbf{G}(\tilde{\boldsymbol{\theta}}_{smm})'\mathbf{W}\mathbf{G}(\tilde{\boldsymbol{\theta}}_{smm})] \to \chi^2(p-q),$$

where $\mathbf{G}(\tilde{\boldsymbol{\theta}}_{smm})'\mathbf{W}\mathbf{G}(\tilde{\boldsymbol{\theta}}_{smm})$ is the value of the objective function at the optimum [see Lee and Ingram (1991, p. 204)].

Although it is not immediately obvious, the singularity of DSGE models also affects their estimation by SMM. In particular, singularity restricts the variables that can be included

in the vector $\mathbf{m}$. For example, multiplying (4) by $\hat{y}_t$ shows that $\hat{n}_t\hat{y}_t$, $\hat{y}_t^2$, and $\hat{c}_t\hat{y}_t$ are not linearly independent. Thus, if one were to include these three variables as elements of $\mathbf{m}$, the matrix $\mathbf{D}$ would not be of full rank. Similarly, multiplying (5) by $\hat{y}_t$ shows that $\hat{y}_t^2$, $\hat{c}_t\hat{y}_t$, $\hat{y}_{t-1}\hat{y}_t$, and $\hat{c}_{t-1}\hat{y}_t$ are not linearly independent either. Hence, the DSGE model imposes restrictions on the moments that can be exploited for its estimation by SMM.

For the DSGE model studied here, a set of linearly independent objects is ($\hat{y}_t^2$, $\hat{c}_t^2$, $\hat{c}_t\hat{y}_t$, $\hat{c}_t\hat{c}_{t-1}$, $\hat{y}_t\hat{y}_{t-1}$). These variables are linearly independent despite relations (4) and (5) because $\hat{n}_t\hat{y}_t$, $\hat{y}_t\hat{c}_{t-1}$ and $\hat{c}_t\hat{y}_{t-1}$ are not included in $\mathbf{m}_i$. For the Monte Carlo experiments below, I also use $\mathbf{m}_i = (\hat{y}_t^2,\, \hat{n}_t^2,\, \hat{n}_t\hat{y}_t,\, \hat{n}_t\hat{n}_{t-1},\, \hat{y}_t\hat{y}_{t-1})'$ and $\mathbf{m}_i = (\hat{n}_t^2,\, \hat{c}_t^2,\, \hat{c}_t\hat{n}_t,\, \hat{c}_t\hat{c}_{t-1},\, \hat{n}_t\hat{n}_{t-1})'$, in order to examine the sensitivity of the results to the variables and moments employed.

The singularity of the DSGE model has different implications for the SMM and Maximum Likelihood estimation methods. *First,* singularity means that (unless measurement errors are added) one can use only one observable variable to estimate the model by ML, but information on up to two variables to estimate the model by SMM. This means that SMM might be more efficient than ML, despite the fact that the former is a limited information procedure. *Second,* SMM requires at least as many moments as parameters are to be estimated (*i.e.,* $p \geq q$) but only linearly independent objects can be included in $\mathbf{m}$. Thus, singularity might impose limits on the number of structural parameters that can by identified by SMM.

3.3 Generalized Method of Moments

For the DSGE model here, it is possible to compute analytical expressions for the unconditional moments of the variables as a function of the structural parameters. This means that in the objective function (10), the simulation-based estimate $(1/\tau T)\sum_{i=1}^{\tau T}\mathbf{m}_i(\boldsymbol{\theta})$ could be replaced with its analytical counterpart $E(\mathbf{m}(\boldsymbol{\theta}))$. Then, a Generalized Method of Moments (GMM) estimator of $\boldsymbol{\theta}$ can be obtained by minimizing the distance between the empirical moments of the data and the theoretical moments predicted by the model. This approach has been followed, among others, by Christiano and Eichenbaum (1992), and Ambler, Guay, and Phaneuf (1999). GMM is also used by Braun (1994) to estimate the Euler equations of a DSGE model. Christiano and den Haan (1996) examine the effects of alternative detrending methods and estimates of the weighting matrix for the small sample properties of GMM. Although the analytical computation of the moments can be algebraically tedious, GMM estimates are computationally and statistically more efficient than SMM.

The GMM estimator is defined by

$$\tilde{\boldsymbol{\theta}}_{gmm} = \min_{\{\boldsymbol{\theta}\}} \mathbf{G}(\boldsymbol{\theta})'\mathbf{W}\mathbf{G}(\boldsymbol{\theta}), \tag{12}$$

where

$$\mathbf{G}(\boldsymbol{\theta}) = (1/T)\sum_{t=1}^{T}\mathbf{m}_t - E(\mathbf{m}(\boldsymbol{\theta})),$$

and $\mathbf{W}$ is the $q \times q$ optimal weighting matrix defined in (11). Under standard regularity conditions

$$\sqrt{T}(\tilde{\boldsymbol{\theta}}_{gmm} - \boldsymbol{\theta}) \rightarrow N(\mathbf{0},(\mathbf{D}'\mathbf{W}^{-1}\mathbf{D})^{-1}),$$

where $\mathbf{D} = \partial E(\mathbf{m}(\boldsymbol{\theta}))/\partial\boldsymbol{\theta}$ is a $q \times p$ matrix assumed to be finite and of full rank. In principle, one could obtain $\partial E(\mathbf{m}(\boldsymbol{\theta}))/\partial\boldsymbol{\theta}$ analytically using the expressions for the unconditional moments $E(\mathbf{m}(\boldsymbol{\theta}))$. Note, however, that these derivatives need to be computed only once, when standard errors are calculated.[6] Thus, for the Monte Carlo, I follow the simpler route of computing $\partial E(\mathbf{m}(\boldsymbol{\theta}))/\partial\boldsymbol{\theta}$ numerically prior to the calculation of the standard errors. As before, the optimal weighting matrix, $\mathbf{W}$, is computed using the Newey-West estimator with a Barlett kernel.

Comparing the asymptotic variance-covariance matrix of SMM and GMM estimates, note that they differ by the term $(1 + 1/\tau)$ that premultiplies $(\mathbf{D}'\mathbf{W}^{-1}\mathbf{D})^{-1}$ in the former case. Since $\mathbf{W}$ depends only on the data, and the simulated moments converge to the analytical ones as $\tau T \rightarrow \infty$, then the difference in the standard errors of both estimates is primarily due to the term $(1+1/\tau)$, rather than to differences in the estimates of $\mathbf{W}$ and $\mathbf{D}$. The term $(1 + 1/\tau)$ can be thought of as a measure of the increase in sample uncertainty due to the use of simulation to compute the population moments. Since $(1 + 1/\tau) \rightarrow 1$ as $\tau \rightarrow \infty$, the efficiency of SMM converges to that of GMM as the length of the simulated series increases. Moreover, since τ is chosen by the researcher, the effect of simulation on sample uncertainty can be reduced by selecting an appropriately large value of τ.

It is easy to see that the effect of simulation on sample uncertainty decreases rapidly as τ increases. For $\tau = 5$, the standard errors of SMM estimates are $(1 + 1/5)^{1/2} \approx 1.10$ times larger than the ones obtained using GMM. For $\tau = 10$ and 20, they are approximately 1.05 and 1.025 times larger than those obtained using GMM. Hence, it would appear that the difference in statistical efficiency between SMM and GMM can be made reasonably small by increasing τ. However, we will below that SMM is less computationally efficient than GMM and that the time per replication under SMM increases proportionally with τ.

[6] Also, one might want to compute $\partial E(\mathbf{m}(\boldsymbol{\theta}))/\partial\boldsymbol{\theta}$ analytically in order to supply an expression for the gradient in the optimization routine. As we will see below, this would make GMM even more computationally efficient than alternative estimation methods.

The implications of the singularity of DSGE models for GMM estimation are basically the same as those for SMM. *First,* singularity limits the unconditional moments that can be exploited by GMM to those that form a linearly independent set. *Second,* since identification requires at least as many moments as parameters are to be estimated, limits in the number of moments that can be employed can have implications for the number of structural parameters that can be identified. *Finally,* GMM might be more efficient than Maximum Likelihood (without measurement errors) because one can use information on more observable endogenous variables. For the model here, GMM can use moments computed using up to two variables, but one can only use one observable variable to estimate the model by ML without adding measurement errors.

Finally, a global specification test for overidentified DSGE models can be performed using the chi-square statistic proposed by Hansen (1982):

$$T[\mathbf{G}(\tilde{\boldsymbol{\theta}}_{gmm})'\mathbf{W}\mathbf{G}(\tilde{\boldsymbol{\theta}}_{gmm})] \rightarrow \chi^2(p-q),$$

where $\mathbf{G}(\tilde{\boldsymbol{\theta}}_{gmm})'\mathbf{W}\mathbf{G}(\tilde{\boldsymbol{\theta}}_{gmm})$ is the value of the objective function at the optimum.

3.4 Extended Method of Simulated Moments

The Extended Method of Simulated Moments (EMSM) is a version of SMM proposed by Smith (1993). Under SMM, an estimate of $\boldsymbol{\theta}$ is constructed by minimizing the distance between the unconditional moments of the data and those of an artificial series simulated using given parameter values. Under EMSM, the estimate of $\boldsymbol{\theta}$ is constructed by minimizing the distance between the parameters of a Vector Autoregression (VAR) estimated from the data and those estimated from an artificial series simulated using given parameter values.

More formally, denote by $\boldsymbol{\eta}$ the $p \times 1$ vector with the estimates of a VAR representation of the data. Denote by $\boldsymbol{\eta}(\boldsymbol{\theta})$ the synthetic counterpart of $\boldsymbol{\eta}$ with the estimates of a VAR representation of artificial data generated by the DSGE model. As before, the sample size is denoted by T and the number of observations in the artificial time series is τT. Then, the EMSM estimator of $\boldsymbol{\theta}$, $\tilde{\boldsymbol{\theta}}_{emsm}$, is the value that solves

$$\min_{\{\boldsymbol{\theta}\}} \quad [\boldsymbol{\eta} - \boldsymbol{\eta}(\boldsymbol{\theta})]'\mathbf{V}[\boldsymbol{\eta} - \boldsymbol{\eta}(\boldsymbol{\theta})], \tag{13}$$

where $\mathbf{V}$ is the $p \times p$ optimal weighting matrix. In the case where the information matrix equality holds, Smith suggests using the inverse of the variance-covariance matrix of the estimate $\boldsymbol{\eta}$ as an estimator of $\mathbf{V}$.

Under the regularity conditions in Smith (1993),

$$\sqrt{T}(\tilde{\boldsymbol{\theta}}_{emsm} - \boldsymbol{\theta}) \rightarrow N(\mathbf{0},(1+1/\tau)(\mathbf{J}'\mathbf{V}^{-1}\mathbf{J})^{-1}),$$

where $\mathbf{J} = E(\partial \boldsymbol{\eta}(\boldsymbol{\theta})/\partial \boldsymbol{\theta})$ is a $q \times p$ matrix assumed to be finite and of full rank. In the Monte Carlo, the derivatives $\partial \boldsymbol{\eta}_t(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ are computed numerically and the expectation approximated by the average over the simulated τT data points.

Smith suggests a test based on Hansen's (1982) chi-square statistic as specification test of an overidentified DSGE model:

$$T(1+1/\tau)\{[\boldsymbol{\eta} - \boldsymbol{\eta}(\tilde{\boldsymbol{\theta}}_{emsm})]'\mathbf{V}[\boldsymbol{\eta} - \boldsymbol{\eta}(\tilde{\boldsymbol{\theta}}_{emsm})]\} \rightarrow \chi^2(p-q),$$

where $[\boldsymbol{\eta} - \boldsymbol{\eta}(\tilde{\boldsymbol{\theta}}_{emsm})]'\mathbf{V}[\boldsymbol{\eta} - \boldsymbol{\eta}(\tilde{\boldsymbol{\theta}}_{emsm})]$ is the value of the objective function at the optimum.

The singularity of the DSGE model has implications for both the order and the number of variables included in the synthetic VAR. More precisely, the artificial data generated by the DSGE model does not have an unconstrained VAR representation. In the case of the RBC model studied here, recall that the systems $(\hat{y}_t, \hat{n}_t, \hat{c}_t)$, $(\hat{y}_t, \hat{c}_t, \hat{y}_{t-1}, \hat{c}_{t-1})$, $(\hat{y}_t, \hat{n}_t, \hat{y}_{t-1}, \hat{n}_{t-1})$, and $(\hat{n}_t, \hat{c}_t, \hat{n}_{t-1}, \hat{c}_{t-1})$ are singular, meaning that their sample variance-covariance matrices are singular. This implies that the data generated by this simple DSGE model has only a bivariate VAR representation of order one.[7] Any attempt to estimate a VAR with the three observable variables and/or using more than one lag will fail because the matrix of explanatory variables is not of full rank.

Since in constructing the synthetic VAR, one is limited to two of the three observable endogenous variables and to only one of their lags, I use VARs of order one on $(\hat{y}_t, \hat{n}_t)$, $(\hat{y}_t, \hat{c}_t)$ and $(\hat{n}_t, \hat{c}_t)$ in the Monte Carlo experiments below in order to examine the sensitivity of the results to the variables used.

4 Monte Carlo Experiments

4.1 Design

The small sample properties of the estimation procedures above are studied here using Monte Carlo analysis. All experiments are based on 500 replications using a sample size of 200 observations. This sample size corresponds to, say quarterly observations of the series for a period of 50 years. In order to limit the effect of the starting values used to generate the series, 100 extra observations were generated in every replication. Then, for the estimation of the model, the initial 100 observations were discarded.

[7] Smith is able to specify a bivariate VAR of order two for a model similar to the one here because he assumes a second disturbance that affects the productivity of the new investment good. Since in this case, the linearized law of motion for capital has an error term, systems like $(\hat{y}_t, \hat{c}_t, \hat{y}_{t-1}, \hat{c}_{t-1})$ in (5) are not singular.

I focus on three observable variables, namely output, consumption and hours worked, and examine whether estimates are sensitive to using different combinations of these variables in the estimation procedures. In order to reduce the computational burden in the Monte Carlo experiments, I concentrate on three (of the five) model parameters, namely the subjective discount factor (β), the autocorrelation coefficient of the technology shock (ρ), and the standard deviation of the technology innovation (σ). Thus, $\boldsymbol{\theta} = (\beta, \rho, \sigma)'$ is a 3×1 vector. The data were generated using parameter values $(\beta, \rho, \sigma) = (.95, .85, .04)$. The share of capital in production (α) and the depreciation rate (δ) were fixed in all experiments to .36 and .025, respectively. This Monte Carlo design is realist in the sense that the true parameter values correspond to the ones typically found in empirical analysis.[8]

For the Monte Carlo experiments with measurement errors, $\mathbf{u}_t \sim N(\mathbf{0}, .01^2\mathbf{I}(d))$, where $\mathbf{I}(d)$ is the identity matrix of size d, and d is the number of observable variables used to estimate the model. For the Monte Carlo experiments with priors, I consider independent prior densities for β and ρ: $\beta \sim N(.95, .025^2)$, $\rho \sim N(.85, .07^2)$, and a diffuse prior for σ. In terms of the prior representation in (8), they correspond to

$$\begin{bmatrix} .95 \\ .85 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \beta \\ \rho \end{bmatrix} + \mathbf{e}, \text{ with } \mathbf{e} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} .025^2 & 0 \\ 0 & .07^2 \end{bmatrix} \right). \tag{14}$$

The priors for β and ρ mean that, before observing the data, the researcher believes that with a 95 percent probability their true values are in the intervals $(.901, .999)$ and $(.713, .987)$, respectively.

For SMM and EMSM, I used three different values of $\tau = 5$, 10, 20, meaning that the simulated series are, respectively, 5, 10, and 20 times larger than the sample size of 200 observations. Here, I also simulated 100 extra observations in every simulation and then discarded the initial 100 observations when computing the moments or the VAR parameters. For these simulation procedures, I fixed the seed in the random numbers generator in each replication and used the same draw for the model estimation. A problem with using blocks of random numbers is that they are *perforce* small [Ripley (1987, p. 138)]. However, in this case the use of common random draws is essential to calculate the numerical derivatives of the maximization algorithm. Otherwise, the objective function would be discontinuous and the optimization algorithm would be unable to distinguish a change in the objective function due to a changes in the parameters from a change in the random draw used to simulate the series.

[8] A slightly larger value of β, say $\beta = .99$, would more appropriate if one were to adhere to the interpretation of the series as quarterly observations of the variables. However, from the numerical perspective, it is convenient to work with a value of β that is close, but not too close, to the admissible boundary of 1. In unreported work, I performed a very limited set of Monte-Carlo experiments using the parameterization $(\beta, \rho, \sigma_\epsilon) = (.98, .95, .04)$ with similar results to the ones reported.

4.2 Results

Monte Carlo results are reported in Tables 1 through 6. In all tables, Mean is the average of the estimated parameter values. A.S.E. is the average asymptotic standard error. Both averages are taken over the 500 replications in each experiment. Median and S.D. are the median and standard deviation of the empirical parameter distribution. Comparing Median and Mean with the true parameter value, and S.D. with A.S.E., is informative about the small sample distribution of the estimates. For example, if the Mean is well below the true parameter value, this indicates the downward bias of the estimate. If Mean and Median are substantially different, this indicates that the small sample distribution of the estimates is skewed. If S.D. is much larger than A.S.E., this indicates that using the asymptotic formula to compute the standard error might understate the true variability of the estimate in small samples.

In all tables, Size is the proportion of times that the null hypothesis that the parameter takes its true value is rejected using a t-test with nominal size of 5 percent. In other words, Size denotes the empirical size of this t-test. S.E. is the standard error of this empirical size and is computed as the standard deviation of a Bernoulli variable. In ideal circumstances, the nominal and empirical sizes of the t-test would be close. More formally, the 95 percent confidence interval around the empirical size would contain the nominal size of 5 percent. However, we will see below that in practice there are substantial size distortions because the asymptotic standard error is not always a good approximation to the small-sample standard error of the estimates. In Tables 4 through 6, OI is the empirical size of the χ^2 test of the overidentification restrictions.

Table 1 reports results using the method of Maximum Likelihood without adding measurement errors/priors. Recall that singularity implies that the model cannot be estimated by ML using more than one observable variable. Experiments 1 through 3 refer to the experiments using output, consumption, and hours worked, respectively.

Despite the fact that the model is a general equilibrium one, not all variables appear to be equally informative about all structural parameters. In other words, the choice of variable(s) employed in Maximum Likelihood estimation might matter. (We will see below that this is also true for the methods of moments). In particular, the average estimate and A.S.E. of β vary substantially depending on whether one uses data on output, consumption, or hours worked. When the model is estimated using data on output alone, the average estimate of β is well below its true value and the A.S.E. is very large. When the model is estimated using data on either consumption or hours worked, the average estimate is very close to the true value and the A.S.E. is small. All this means that for the simple RBC

model studied here, a sharper estimate of the subjective discount factor can be obtained using consumption or hours worked, rather than output.

The effect of the choice of variable on the point estimates of the other structural parameters is minor. However, the standard error of the empirical distribution and the A.S.E. of ρ seem to vary with the variable employed. For example, the A.S.E. of ρ is 30 percent larger when consumption, rather than output, is used to estimate the model. Thus, the autoregressive coefficient of the technology shock is estimated more precisely using output than either consumption or hours worked.

The asymptotic standard errors approximate well the small-sample standard deviation of the estimates. However, the difference between them is large enough that in some cases there are size distortions for the t-test (computed using the A.S.E.) of the null hypothesis that the true parameter value is the one used to generate the data.

Table 2 reports results using Maximum Likelihood with measurement errors added to sidestep the singularity of the DSGE model. Parameters are estimated using combinations of two or the three observable endogenous variables. In all experiments, parameters are estimated very precisely and standard errors are by far the smallest among the procedures studied. The choice of variables employed has no effect on the point estimates but a limited effect on the standard errors. The empirical distributions of the parameters appear very close to the asymptotic one. The nominal size of the t-test that the parameters take their true values is well within their 95 percent confidence interval around the empirical size in all experiments.

Table 3 reports results using Maximum Likelihood without measurement errors but incorporating the priors specified in (14). As in Table 1, Experiments 1 through 3 refer (respectively) to the experiments using output, consumption, and hours worked. Because the prior about σ was uninformative, results regarding this parameter are basically the same as those reported in Table 1 without priors. Regarding β and ρ, the mean of the prior density was (by design) the same as the true value used to generate the sample. Recall that an interpretation of the prior is that of a penalty on the log likelihood function as the estimate deviates from the prior mean. As a result, estimates obtained using ML with priors tend to be numerically closer to their true values than ML estimates without priors.

An interesting case is the estimate of β obtained using output data. Results from Table 1 suggest that output is not very informative regarding the discount factor. Consequently when the prior and the output data are combined, the resulting posterior distribution of the estimate looks very similar to the prior density.

In most cases, the A.S.E. is larger than the standard deviation of the parameter estimates. Hence, asymptotic standard errors seem to overestimate the variability of the parameter

estimates in small samples. As a result, the t-test of the hypothesis that the parameter takes its true value has a smaller empirical than nominal size and tends to underreject the null hypothesis.

Tables 4 and 5 report results using the Simulated and Generalized Methods of Moments, respectively. In all cases, the mean and median of the estimated parameters are close to their true values. Standard errors are reasonably low, though larger than those obtained using ML with measurement errors added.[9]

Estimated standard errors vary with τ as predicted by the discussion in Section 3.4. That is, SMM standard errors based on simulated series with $\tau = 5, 10$ and 20 are roughly $1.1, 1.05$, and 1.025 times larger than those obtained using GMM, respectively. However, this variation in efficiency is of the same order of magnitude as that observed across moments employed in the estimation procedures. This means that the choice of the length of the simulated series in SMM might be as important as the choice of moments to match.[10]

As reported above for Maximum Likelihood, the choice of moments to match under SMM and GMM can have some effect on the precision of the parameter estimates. This is because not all moments are equally informative about all structural parameters. (See below the related discussion about EMSM). For example, in Table 4, the standard deviation of the estimate of β obtained using the moments of output and hours worked is roughly 30 percent larger than the one obtained using the moments of consumption and hours worked, for the same value of τ.

There are some differences between the A.S.E. and the standard deviations of the empirical distribution of the parameters. Since the t-statistics are computed using asymptotic standard errors, this translates into fairly large size distortions for the t-test. There is no obvious pattern for the size distortions, but they appear to vary more with the moments matched than with the length of the simulated series (in SMM).

The last column in Tables 4 and 5 report the empirical size of the χ^2 test of the overidentification restrictions of the model. Notice that in all cases the empirical size of the test is well below its nominal size of 5 percent. This means that a researcher comparing

[9] Recent research [see, for example, Fuhrer, Moore, and Schuh (1995)] shows that GMM can yield biased parameter estimates when applied to conditional moments of the data (*e.g.,* Euler equations). The reason is that in this case instruments are used to frame the model implications in terms of orthogonality conditions to which GMM is then applied. Their research finds that GMM can have very poor small sample properties when the instruments are weak. However, notice that GMM is applied here to unconditional moments of the data and no instruments are required. This means that the problems caused by weak instruments do arise here, but could arise potentially when one combines conditional and unconditional moments to estimate DSGE models by GMM.

[10] In practice, however, it might be hard to know *apriori* what are the most informative set of linearly independent moments.

the χ^2 statistic with the 5 percent critical value of the appropriate distribution is very likely to conclude that the overidentification restrictions of the DSGE model cannot be rejected. This is because rather than taking a 5 percent probability of rejecting a true model, the researcher is actually taking only a 1 percent probability (approximately). In some cases (see, for example, Experiments 7 and 9 in Table 4), the model's overidentification restrictions were never rejected by the χ^2 test in the 500 replications. The fact that Hansen's χ^2 easily fails to detect a misspecified model is well known in the literature [see, among others, Newey (1985)]. The results in this paper suggest that the severe size distortions of the χ^2 test also arise in the context of fully-specified DSGE models.

Finally, Table 6 reports results using the Extended Method of Simulated Moments (EMSM). In all experiments, the mean and median of the estimated parameters are close to their true values. The standard errors of the estimates of β and ρ are the largest among the procedures studied, but the standard deviation of the technology shock, σ, is estimated more precisely by EMSM than by any other method.

As in ML, SMM, and GMM, the choice of variables/moments to use in estimation has some effect on the precision (but not on the consistency) of the estimates. This can be traced back to the fact that not all variables/moments are equally informative about all structural parameters. For example, σ can only be identified under EMSM when $\boldsymbol{\eta}$ includes the standard deviation of the residuals of the Vector Autoregression. Thus, the autocorrelations and cross correlations of the variables do not seem informative regarding the standard deviation of technology shocks. In line with the finding that under ML, consumption and hours worked are more informative than output about the subjective discount rate, the standard deviation of the empirical distribution of β is smallest when the bivariate VAR consist of $\hat{c}_t$ and $\hat{n}_t$.

There are very large size distortions for both the t-test that the parameters take their true value and the χ^2 test of the overidentification restrictions. As for SMM/GMM, there is no clear pattern for these size distortions, but they seem to vary more with the moments matched than with the length of the simulated series.

It is enlightening to go beyond the summary statistics in these Tables and plot the empirical distribution of the parameter estimates obtained under the different methods. The frequency histograms for estimators of β, ρ, and σ are plotted in Figures 1 to 3, respectively. They correspond to results in Experiments 2 in Tables 1 and 2, Experiment 3 in Table 5, and Experiments 5 and 8 in Tables 4 and 6. The reason I focus on these experiments, is because they help illustrate more general results uncover by the Monte Carlo analysis.

Five conclusion can be drawn from these Figures. First, Maximum Likelihood estimates obtained adding measurement errors are by far the most efficient, with the empirical distri-

butions tightly packed around the true parameter values. This is explained by the fact, that once extra errors are added to sidestep the singularity of the model, ML can exploit information on more variables than alternative methods. Since ML is a full-information procedure, it imposes all model restrictions on the series employed leading to sharper estimates of the parameters.

Second, the methods of moments estimators can be more efficient than Maximum Likelihood with no measurement errors. This was anticipated from the earlier observation that singularity implies that the model can only be estimated by ML using one observable variable (without adding measurement errors), but using moments of up to two variables by the methods of moments. However, it is very likely that this finding is specific to the one-shock DSGE model examined here and might not carry over to the models with a larger number of structural shocks.[11]

Third, informative priors can be combined with sample data to sharpen the researcher's inferences regarding the structural parameters of DSGE model. This can be seen in Figures 1 and 2 by comparing the empirical distribution of the estimators of β and ρ obtained using ML with and without priors.

Fourth, the difference in efficiency between GMM and SMM does not appear to be very large, though the empirical distribution obtained by SMM with $\tau = 5$ (not reported) is somewhat more diffuse than the others. Hence, the effect of simulation on sample uncertainty and the precision of parameter estimates can be moderated by a suitable choice of τ.

Fifth, the choice of variables/moments employed can have some effect on the precision (but not on the consistency) of the estimators. This can be seen in Figures 1 and 2 by comparing the empirical distribution of the estimators of β and ρ obtained by SMM and EMSM using the moments of output/hours worked and consumption/hours worked. EMSM estimates based on the moments of output/hours worked have a very diffuse empirical distribution and a substantial number of outliers, though their mean is close to the true parameter value used to generate the sample.

One of the reason DSGE model are interesting is because they allow the researcher to examine the response of a model economy to shocks. An advantage of the full estimation of DSGE models is that parameter uncertainty can be incorporated to construct confidence intervals around the model's dynamic response to a shock. Since impulse-responses depend nonlinearly on the structural parameters, it is useful to examine how the precision of the estimates translates into less or more precise impulse-responses. Figures 4 to 11 plot the dynamic responses of consumption, output, hours worked, and the capital stock following a

[11] For example, in the preliminary work mentioned in footnote 2, where there were two structural shocks, ML and SMM were very similar in statistical efficiency.

technology shock. The dotted lines are the 90 percent confidence intervals for the response. The parameter estimates used to construct these figures come from the same experiments used to obtain Figures 1 to 3.

Figure 4 to 11 support four conclusions. First, the mean response does not depend on the estimation method because all methods deliver unbiased estimates of the parameters. However, there are some differences in the coverage probabilities of the estimated confidence intervals because not all estimation procedures are equally efficient. Second, the tightest confidence intervals around the impulse-response are obtained using Maximum Likelihood estimates (with added measurement errors) [see Figure 5]. Since we are fairly certain that these ML estimates are close to their true values, we can also be confident that the impulse-response to a shock based on these estimates is also close the true one. Third, confidence intervals for the response based on EMSM estimates obtained using output/hours worked moments are the widest [see Figure 10]. This reflects the diffuse empirical distributions of β and ρ reported in Figures 1 and 2. Fourth, aside from these two cases, there are no large differences in the confidence intervals obtained using different estimators of the structural parameters. This means that the differences in parameter efficiency across estimation methods reported above, does not translate necessarily into substantial differences in the coverage probabilities of their impulse responses.

The estimation of DSGE models can be computationally demanding because the model needs to be solved for each observation in each iteration of the optimization procedure that maximizes (or minimizes) the relevant objective function. Thus, an important goal of this paper is to compare the different estimation methods in terms of their computation time. Table 7 reports in the first column the average number of seconds taken to complete a replication, including the computation of standard errors. The average is taken over all replications for all experiments that employ the same estimation method. The second column reports the ratio of the number in column one to the corresponding one for GMM, that is taken as benchmark. For example, for ML this ratio is $3.44/.67 = 5.1$, and means that Maximum Likelihood takes on average 5 times longer than GMM to complete a replication.

From this Table is clear that GMM is by far the most efficient procedure computationally, followed by Maximum Likelihood. On the other hand, GMM requires the analytical calculation of the unconditional moments implied by the model. This task can be algebraically tedious and time-consuming for models more complicated than the one studied here. There is a large difference in computational efficiency between SMM/EMSM and GMM, and the time per iteration appears to increase proportionally with τ. The reason is that SMM and EMSM require the solution of the DSGE model and computation of the gradients using τ

times more observations than GMM.[12]

5 Summary

The Monte Carlo analysis in this paper shows that standard econometric techniques can be applied for the full estimation of DSGE models. Although singularity means that there are restrictions on the variables/moments that can be exploited for model estimation, it is possible to obtain unbiased and reasonably precise estimates of the structural parameters of the model using the methods of Maximum Likelihood or versions of the method of moments. Results here indicate that, despite the fact that the model is a general equilibrium one, not all variables/moments are equally informative about all structural parameters. This has no consequence for the point estimates, since all method yield unbiased and consistent parameter estimates, but it appears to have some moderate implications for the size of the estimated standard errors.

For the one sector Real Business Cycle model studies here, the method of Maximum Likelihood with measurement errors added, yields the most efficient parameter estimates among the procedures considered. The Simulated and Generalized Methods of Moments are roughly comparable in terms of statistical efficiency, but SMM is more computationally demanding. On the other hand, GMM requires the analytical computation of the unconditional moments, that can be a time-consuming task in more complicated models than the one studied here.

Both SMM and GMM are more efficient than Maximum Likelihood without added measurement errors. The reason is that the singularity of the model limits to one the number of variables that can be using in ML, but one can use moments of up to two variables in the methods of moments. However, this result is probably specific to one-shock DSGE models and might not carry on to the models with a larger number of structural shocks that are more representative of the current state of the literature.

[12]Results regarding EMSM need to be interpreted with caution. For all estimation methods, the maximization (or minimization) routines were started at the true parameter values in order the make the Monte Carlo experiment more efficient. However, I found that for EMSM, the algorithm would frequently blow up if the routine was started at the true value of σ. Hence, for EMSM, the minimization routine was started using a value for σ much larger than the one used to generate the sample. Just for this reason alone EMSM would take longer to converge than the other estimation methods. This means that the numbers in Table 7 most likely overstate the computational time required by EMSM.

Table 1. ML Results

Experiment #	Var.	β Mean / Median	β A.S.E. / S.D.	β Size / S.E.	ρ Mean / Median	ρ A.S.E. / S.D.	ρ Size / S.E.	σ Mean / Median	σ A.S.E. / S.D.	σ Size / S.E.
1	y_t	.7256	.4077	.1740	.8408	.0390	.0260	.0400	.0020	.0520
		.9313	.3747	.0170	.8429	.0343	.0071	.0400	.0019	.0099
2	c_t	.9440	.0235	.0480	.8367	.0517	.0560	.0398	.0020	.0660
		.9740	.0236	.0096	.8429	.0542	.0103	.0398	.0021	.0111
3	n_t	.9457	.0346	.0320	.8296	.0477	.0160	.0398	.0020	.0580
		.9541	.0432	.0079	.8366	.0368	.0056	.0397	.0019	.0105

Notes: The true values are $\beta = .95$, $\rho = .85$, and $\sigma = .04$. Mean is the arithmetic average of the estimated parameter values;, A.S.E. is the average asymptotic standard error; Median and S.D. are the median and standard deviation of the empirical parameter distribution; Size is an estimate of the actual size of the t-test with nominal size of 5 percent of the null hypothesis that the parameter takes its true value; and S.E. is the standard error of the actual test size. The experiments were based on 500 replications.

Table 2. ML Results with Measurement Errors

Experiment		β			ρ			σ		
		Mean	A.S.E.	Size	Mean	A.S.E.	Size	Mean	A.S.E	Size
#	Var.	Median	S.D.	S.E.	Median	S.D.	S.E.	Median	S.D.	S.E.
1	y_t, c_t	.9502	.0024	.0520	.8486	.0091	.0560	.0040	.0021	.0500
		.9502	.0024	.0099	.8493	.0099	.0103	.0400	.0021	.0097
2	y_t, n_t	.9500	.0023	.0420	.8492	.0079	.0520	.0399	.0020	.0640
		.9501	.0022	.0090	.8494	.0083	.0099	.0400	.0021	.0109
3	c_t, n_t	.9499	.0018	.0440	.8496	.0064	.0340	.0399	.0021	.0480
		.9499	.0018	.0092	.8500	.0065	.0081	.0400	.0021	.0096
4	y_t, c_t, n_t	.9499	.0015	.0520	.8497	.0056	.0620	.0399	.0020	.0660
		.9500	.0016	.0099	.8499	.0059	.0099	.0399	.0021	.0111

Notes: See notes to Table 1.

Table 3. ML Results Incorporating Priors

Experiment #	Var.	β Mean Median	β A.S.E. S.D.	β Size S.E.	ρ Mean Median	ρ A.S.E. S.D.	ρ Size S.E.	σ Mean Median	σ A.S.E S.D.	σ Size S.E.
1	y_t	.9500	.0250	0.0000	.8459	.0332	.0240	.0400	.0020	.0400
		.9500	.0008	–	.8472	.0279	.0068	.0400	.0019	.0088
2	c_t	.9498	.0164	.0180	.8445	.0402	.0060	.0397	.0020	.0580
		.9499	.0124	.0059	.8469	.0322	.0035	.0397	.0020	.0105
3	n_t	.9482	.0192	.0020	.8436	.0318	.0180	.0400	.0020	.0400
		.9493	.0094	.0020	.8467	.0247	.0059	.0400	.0021	.0640

Notes: See notes to Table 1. The priors used are: $\beta \sim N(.95, .025^2)$, $\rho \sim N(.85, .07^2)$, and a diffuse prior for σ.

Table 4. Results using Simulated Method of Moments

Experiment			β			ρ			σ			
			Mean	A.S.E.	Size	Mean	A.S.E.	Size	Mean	A.S.E.	Size	OI
#	Var.	τ	Median	S.D.	S.E.	Median	S.D.	S.E.	Median	S.D.	S.E.	S.E.
1	y_t, c_t	5	.9505	.0128	.0160	.8368	.0411	.1040	.0395	.0042	.0360	.0100
			.9511	.0089	.0056	.8466	.0504	.0137	.0396	.0032	.0083	.0044
2	y_t, c_t	10	.9503	.0122	.0300	.8370	.0392	.0840	.0393	.0040	.0240	.0120
			.9505	.0095	.0076	.8442	.0478	.0124	.0394	.0028	.0068	.0049
3	y_t, c_t	20	.9506	.0120	.0260	.8372	.0382	.0880	.0394	.0039	.0320	.0120
			.9509	.0088	.0071	.8426	.0443	.0127	.0394	.0029	.0079	.0049
4	y_t, n_t	5	.9497	.0138	.0860	.8400	.0314	.1880	.0395	.0037	.0820	.0060
			.9508	.0153	.0125	.8452	.0470	.0175	.0394	.0037	.0123	.0035
5	y_t, n_t	10	.9483	.0133	.0740	.8367	.0303	.1320	.0396	.0035	.0740	.0020
			.9497	.0148	.0117	.8414	.0417	.0151	.0394	.0036	.0117	.0020
6	y_t, n_t	20	.9496	.0131	.1000	.8428	.0291	.1420	.0398	.0035	.0720	.0020
			.9510	.0147	.0134	.8496	.0426	.0156	.0395	.0037	.0116	.0020
7	c_t, n_t	5	.9501	.0107	.0520	.8413	.0294	.1580	.0396	.0032	.0860	.0000
			.9502	.0099	.0099	.8444	.0409	.0163	.0396	.0034	.0125	–
8	c_t, n_t	10	.9510	.0101	.0520	.8436	.0277	.1640	.0397	.0030	.0920	.0020
			.9516	.0098	.0099	.8475	.0377	.0166	.0396	.0034	.0129	.0020
9	c_t, n_t	20	.9499	.0101	.0600	.8411	.0277	.1260	.0395	.0030	.0820	.0000
			.9502	.0098	.0106	.8458	.0376	.0148	.0820	.0033	.0123	–

Notes: See notes to Table 1. For Experiments 1 to 3, $\mathbf{m}_t = (\hat{y}_t^2,\ \hat{c}_t^2,\ \hat{c}_t\hat{y}_t,\ \hat{c}_t\hat{c}_{t-1},\ \hat{y}_t\hat{y}_{t-1})'$; for Experiments 4 to 6, $\mathbf{m}_t = (\hat{y}_t^2,\ \hat{n}_t^2,\ \hat{n}_t\hat{y}_t,\ \hat{n}_t\hat{n}_{t-1},\ \hat{y}_t\hat{y}_{t-1})'$; and Experiments 7 to 9, $\mathbf{m}_t = (\hat{n}_t^2,\ \hat{c}_t^2,\ \hat{c}_t\hat{n}_t,\ \hat{c}_t\hat{c}_{t-1},\ \hat{n}_t\hat{n}_{t-1})'$.

Table 5. Results using Generalized Method of Moments

Experiment		β			ρ			σ			
	Var.	Mean	A.S.E.	Size	Mean	A.S.E.	Size	Mean	A.S.E.	Size	OI
		Median	S.D.	S.E.	Median	S.D.	S.E.	Median	S.D.	S.E.	S.E.
1	y_t, c_t	.9501	.0118	.0100	.8369	.0375	.1020	.0395	.0039	.0180	.0040
		.9502	.0084	.0044	.8429	.0457	.0135	.0395	.0026	.0059	.0028
2	y_t, n_t	.9484	.0128	.0720	.8382	.0288	.1480	.0397	.0034	.0780	.0020
		.9496	.0137	.0116	.8418	.0400	.0159	.0396	.0035	.0120	.0020
3	c_t, n_t	.9499	.0099	.0540	.8380	.0272	.1360	.0395	.0029	.0740	.0000
		.9497	.0094	.0101	.8407	.0383	.0153	.0395	.0030	.0117	–

Notes: See notes to Tables 1 and 3.

Table 6. Results using the Extended Method of Simulated Moments

Experiment			β			ρ			σ			
			Mean	A.S.E.	Size	Mean	A.S.E.	Size	Mean	A.S.E.	Size	OI
#	Var.	τ	Median	S.D.	S.E.	Median	S.D.	S.E.	Median	S.D.	S.E.	S.E.
1	y_t, c_t	5	.9560	.0261	.2700	.8436	.0807	.3200	.0400	.0004	.7220	.0620
			.9543	.0295	.0199	.8652	.1134	.0209	.0400	.0021	.0200	.0108
2	y_t, c_t	10	.9566	.0252	.2500	.8446	.0744	.3160	.0398	.0004	.7500	.0540
			.9574	.0299	.0194	.8759	.1172	.0208	.0399	.0020	.0194	.0101
3	y_t, c_t	20	.9559	.0246	.3060	.8366	.0701	.3660	.0400	.0003	.7840	.0600
			.9564	.0327	.0206	.8724	.1283	.0215	.0400	.0020	.0184	.0106
4	y_t, n_t	5	.9499	.0254	.2380	.8120	.0915	.2400	.0401	.0002	.8480	.2340
			.9505	.0319	.0190	.8519	.1513	.0191	.0401	.0021	.0161	.0189
5	y_t, n_t	10	.9463	.0251	.2020	.8020	.0948	.1960	.0400	.0002	.8780	.1700
			.9478	.0291	.0180	.8420	.1484	.0178	.0399	.0021	.0146	.0168
6	y_t, n_t	20	.9448	.0254	.2040	.7854	.0970	.2000	.0398	.0002	.8680	.2280
			.9489	.0316	.0180	.8460	.1838	.0179	.0397	.0019	.0151	.0188
7	c_t, n_t	5	.9482	.0262	.0020	.8390	.0980	.0000	.0400	.0006	.6260	.0040
			.9488	.0108	.0020	.8457	.0426	–	.0399	.0022	.0216	.0028
8	c_t, n_t	10	.9484	.0255	.0040	.8405	.0945	.0000	.0400	.0005	.6160	.0020
			.9489	.0103	.0028	.8459	.0405	–	.0400	.0021	.0218	.0020
9	c_t, n_t	20	.9488	.0251	.0100	.8414	.0925	.0020	.0401	.0005	.5840	.0040
			.9487	.0109	.0044	.8451	.0433	.0020	.0401	.0020	.0220	.0028

Notes: See notes to Table 1. For Experiments 1 to 3, the VAR consists of $\hat{y}_t$ and $\hat{c}_t$; for Experiments 4 to 6, the VAR consists of $\hat{y}_t$ and $\hat{n}_t$; and for Experiments 7 to 9 the VAR consists of $\hat{n}_t$, and $\hat{c}_t$. In all cases a VAR of order one is used.

Table 7. Comparison in Terms of Computing Time

Method	τ	Seconds per Replication	Compared with GMM
ML	–	3.44	5.1
ML (with errors)	–	2.95	4.4
ML (with priors)	–	2.56	3.8
SMM	5	7.52	11.2
	10	14.74	22.0
	20	27.41	40.1
EMSM	5	27.89	41.6
	10	59.15	88.3
	20	90.96	135.8
GMM	–	0.67	1

Notes: The Monte Carlo was performed using GAUSS for Windows running in a Dell Inspiron 7500 with Pentium III processor.

A Appendix A : The Log-linearized Model

In what follows, variables without time subscript denote steady state values and the circumflex denotes percentage deviation from steady state. For example, $\hat{c}_t = (c_t - c)/c$ is the percentage deviation of consumption from its steady state at time t. For the model in Section 2, the linearized first-order conditions of the agent's problem are (notice that the marginal products of labor and capital have already been substituted out):

$$\begin{aligned} E_t\hat{c}_{t+1} &= \hat{c}_t + \varsigma(\alpha-1)E_t\hat{k}_{t+1} + \varsigma(1-\alpha)E_t\hat{n}_{t+1} + \varsigma E_t\hat{z}_{t+1}, \\ \hat{n}_t &= -(1/\alpha)\hat{c}_t + \hat{k}_t + (1/\alpha)\hat{z}_t. \end{aligned}$$

where $\varsigma = \alpha\beta(k/n)^{\alpha-1}$ and the steady-state capital-labor ratio $k/n = ((1/\beta+\delta-1)/\alpha)^{1/(\alpha-1)}$. The linearized production function and resource constraint are:

$$\begin{aligned} \hat{y}_t &= \alpha\hat{k}_t + (1-\alpha)\,\hat{n}_t + \hat{z}_t, \\ \hat{y}_t &= \gamma\hat{c}_t + (1-\gamma)\hat{x}_t, \end{aligned}$$

where γ is the consumption-output ratio in steady and equals $1 - \delta(k/n)^{1-\alpha}$. Finally, the linearized law of motions for capital and the technology shock are:

$$\begin{aligned} \hat{k}_{t+1} &= (1-\delta)\hat{k}_t + \delta\hat{x}_t, \\ \hat{z}_{t+1} &= \rho\hat{z}_t + \epsilon_t. \end{aligned}$$

References

[1] Altug, S. (1989), "Time-to-Build and Aggregate Fluctuations: Some New Evidence," *International Economic Review* 30, 889-920.

[2] Ambler, S., Guay, A., and Phaneuf, L. (1999), "Wage Contracts and Labor Adjustment Costs as Endogenous Propagation Mechanisms," Université du Québec à Montréal, Working Paper 69.

[3] Bencivenga, V. R. (1992), "An Econometric Study of Hours and Output Variation with Preference Shocks," *International Economic Review* 33, 449-471.

[4] Blanchard, O. J. and Kahn, C. M. (1980), "The Solution of Linear Difference Models under Rational Expectations," *Econometrica*, 48, 1305-1311.

[5] Bouakez, H., Cardia, E., and Ruge-Murcia, F. J. (2001), "Habit Formation and the Persistence of Monetary Shocks," Université de Montréal, *Mimeo.*

[6] Braun, R. A. (1994), "Tax Disturbances and Real Economic Activity in the Postwar United States?" *Journal of Monetary Economics* 33, 441-462.

[7] Browning, M., Hansen, L. P., and Heckman, J. J. (1999), "Micro Data and General Equilibrium Models," in *Handbook of Macroeconomics* edited by J. Taylor and M. Woodford. North-Holland: Amsterdam.

[8] Chang, Y., Gomes, J. F., and Schofheide, F. (2002), "Learning-by-Doing as a Propagation Mechanism," *American Economic Review*, forthcoming.

[9] Christiano, L. J. (1988), "Why Does Inventory Investment Fluctuate So Much?" *Journal of Monetary Economics* 21, 247-280.

[10] Christiano, L. J. and Eichenbaum, M. (1992), "Current Real-Business Cycle Theories and Aggregate Labor-Market Fluctuations," *American Economic Review*, 82, 430-450.

[11] Christiano, L. J. and den Haan, W. J. (1996), "Small-Sample Properties of GMM for Business-Cycle Analysis?" *Journal of Business and Economic Statistics* 14, 309327.

[12] DeJong, D. N., Ingram, B. F., and Whiteman, C. H. (1996), "A Bayesian Approach to Calibration," *Journal of Business and Economic Statistics* 14, 1-9.

[13] DeJong, D. N., Ingram, B. F., and Whiteman, C. H. (2000), "A Bayesian Approach to Dynamic Macroeconomics," *Journal of Econometrics* 98, 203-223.

[14] Dib, A. and Phaneuf, L. (2001), "An Econometric U.S. Business Cycle Model with Nominal and Real Rigidities," Université du Québec à Montréal, Working Paper 137.

[15] Duffie, D. and Singleton, K. J. (1993), "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica* 61, 929-952.

[16] Fernández-Villaverde, J. and Rubio, J. (2002), "Estimating Nonlinear Dynamic Equilibrium Economies: A Likelihood Approach," University of Pennsylvania, *Mimeo.*

[17] Fuhrer, J. C., Moore, G. R., and Schuh, S. D. (1995), "Estimating the Linear-Quadratic Inventory Model: Maximum Likelihood versus Generalized Method of Moments," *Journal of Monetary Economics* 35, 115-157.

[18] Gregory, A. and Smith, G. (1991), "Calibration as Testing: Inference in Simulated Macroeconomic Models," *Journal of Business and Economic Statistics* 9, 297-304.

[19] Hall, G. J. (1996), "Overtime, Effort, and the Propagation of Business Cycle Shocks," *Journal of Monetary Economics* 38, 139-160.

[20] Hamilton, J. D. (1994), *Time Series Analysis,* Princeton: Princeton University Press.

[21] Hansen, G. D. (1985), "Indivisible Labor and the Business Cycle," *Journal of Monetary Economics*, 16, 309-327.

[22] Hansen, L. P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* 50, 1929-1954.

[23] Hansen, L. P. and Sargent, T. J. (1980), "Formulating and Estimating Dynamic Linear Rational Expectations Models," *Journal of Economic Dynamics and Control* 2, 7-46.

[24] Hansen, L. P. and Heckman, J. J. (1996), "The Empirical Foundations of Calibration," *Journal of Economic Perspectives* 10, 87-104.

[25] Hansen, L. P. and Sargent, T. J. (1998), "Recursive Linear Models of Dynamic Economies," University of Chicago, *Mimeo.*

[26] Ireland, P. (1999), "A Method for Taking Models to the Data," Boston College, *Mimeo.*

[27] Ireland, P. (2001), "Sticky Price Models of the Business Cycle: Specification and Stability," *Journal of Monetary Economics* 47, 3-18.

[28] Judge, G. G., Griffiths, W. E., Carter Hill, R., Lütkepohl, H., and Lee, T.-C. (1985), *The Theory and Practice of Econometrics*, John Wiley and Sons: New York.

[29] Kalman, R. E. (1980), "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering, Transactions of the ASME Series D* 82, 35-45.

[30] Kydland, F. E., and Prescott, E. C. (1996), "The Computational Experiment: An Econometric Tool," *Journal of Economic Perspectives* 10, 69-85.

[31] Kim, J. (2000), "Constructing and Estimating a Realistic Optimizing Model of Monetary Policy," *Journal of Monetary Economics* 45, 329-359.

[32] King, R. G., Plosser, C. I., and Rebelo, S. (1988), "Production, Growth, and Bussines Cycles: The Basic Neoclassical Model," *Journal of Monetary Economics*, 21, 195-232.

[33] Lee, B.-S., and Ingram, B. F. (1991), "Simulation Estimation of Time-Series Models," *Journal of Econometrics* 47, 195-205.

[34] McFadden, D. (1986), "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica*, 57, 995-1026.

[35] McGrattan, E. R. (1994), "The Macroeconomic Effects of Distortionary Taxation," *Journal of Monetary Economics* 33, 573-601.

[36] McGrattan, E. R., Rogerson, R., and Wright, R. (1997), "An Equilibrium Model of the Business Cycle with Household Production and Fiscal Policy," *International Economic Review* 38, 267-290.

[37] Newey, W. K. (1985), "Generalized Method of Moments Specification Testing," *Journal of Econometrics* 29, 229-256.

[38] Pakes, A. and Pollard, D. (1989), "The Asymptotic Distribution of Simulation Experiments," *Econometrica*, 57, 1027-1057.

[39] Ripley, B. D. (1987), *Stochastic Simulation,* John Wiley & Sons: New York.

[40] Sargent, T. (1989), "Two Models of Measurement and the Investment Accelerator," *Journal of Political Economy* 97, 251-287.

[41] Sims, C. (1997), "Solving Linear Rational Expectations Models," Yale University, *Mimeo.*

[42] Smith, A. A. (1993), "Estimating Nonlinear Time-Series Models using Simulated Vector Autoregressions," *Journal of Applied Econometrics* 8, 63-84.

[43] Stone, J. R. N. (1954), *The Measurement of Consumers' Expenditure and Behaviour in the United Kingdom,* Cambridge University Press: Cambridge.

[44] Theil, H. and Goldberger, A. S. (1961), *International Economic Review* 2, 65-78.

[45] Watson, M. (1993), "Measures of Fit for Calibrated Models," *Journal of Political Economy* 101, 1011-1041.

Figure 1: Empirical Distribution of β

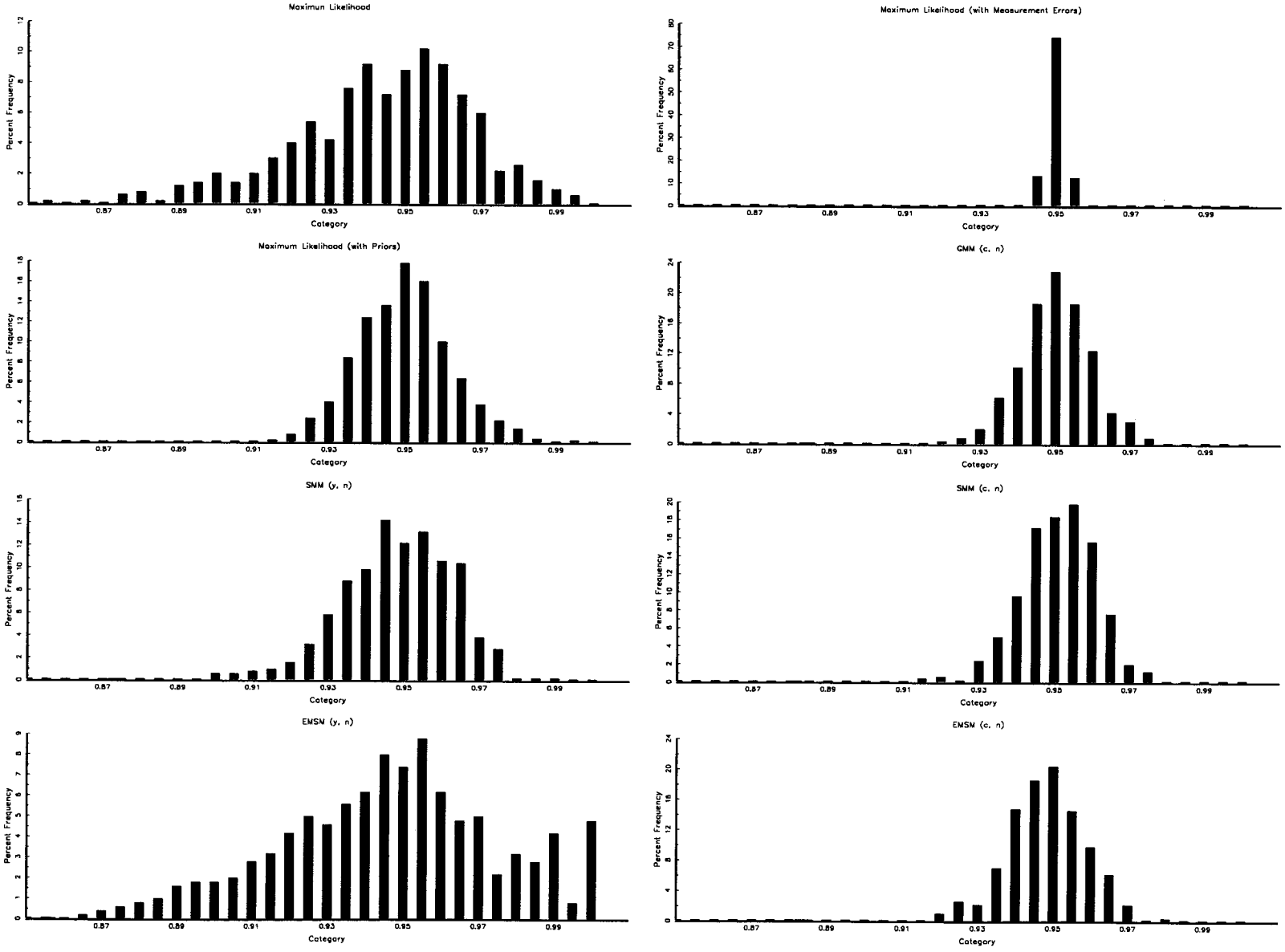

Figure 2: Empirical Distribution of ρ

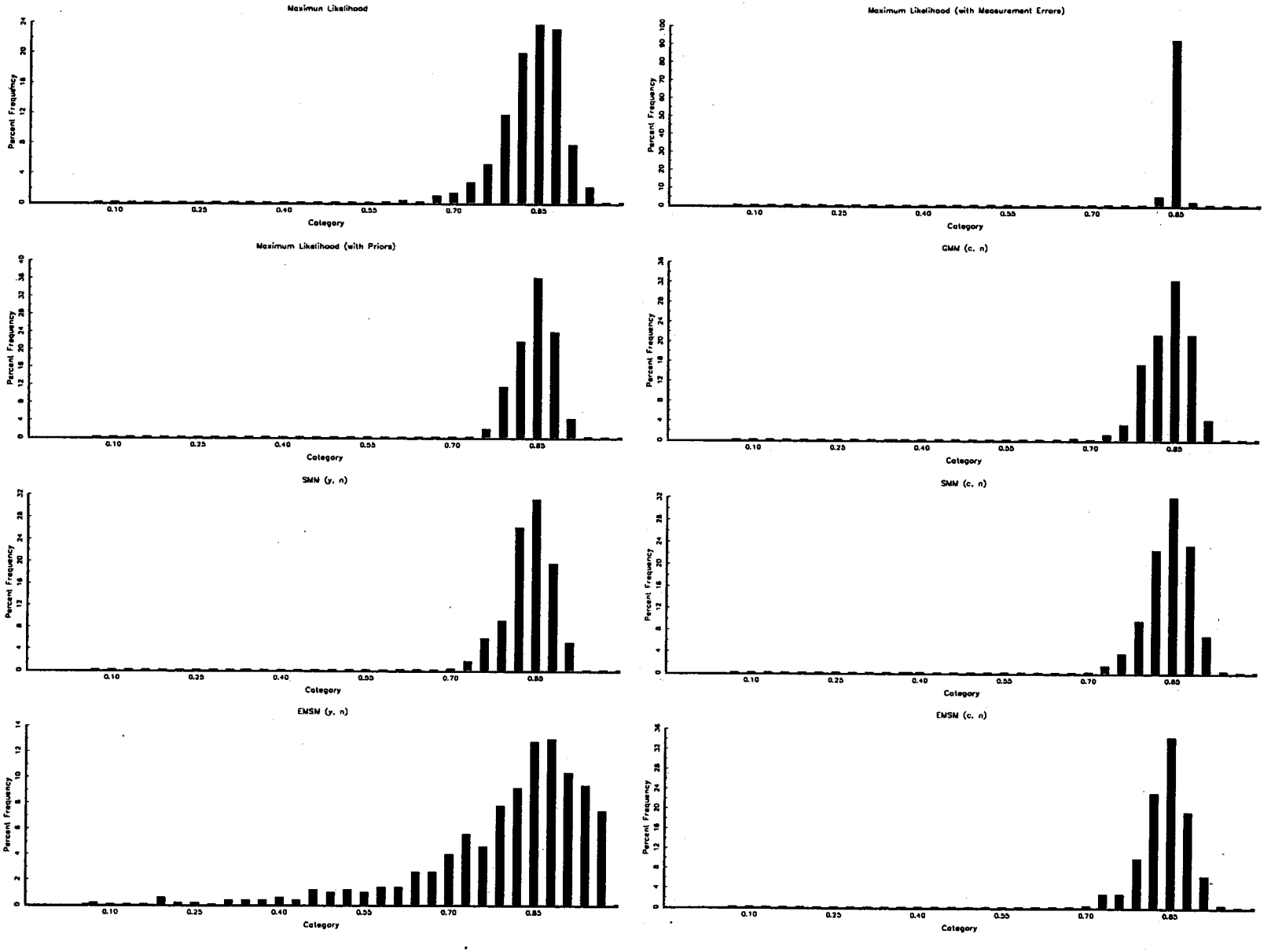

Figure 3: Empirical Distribution of σ

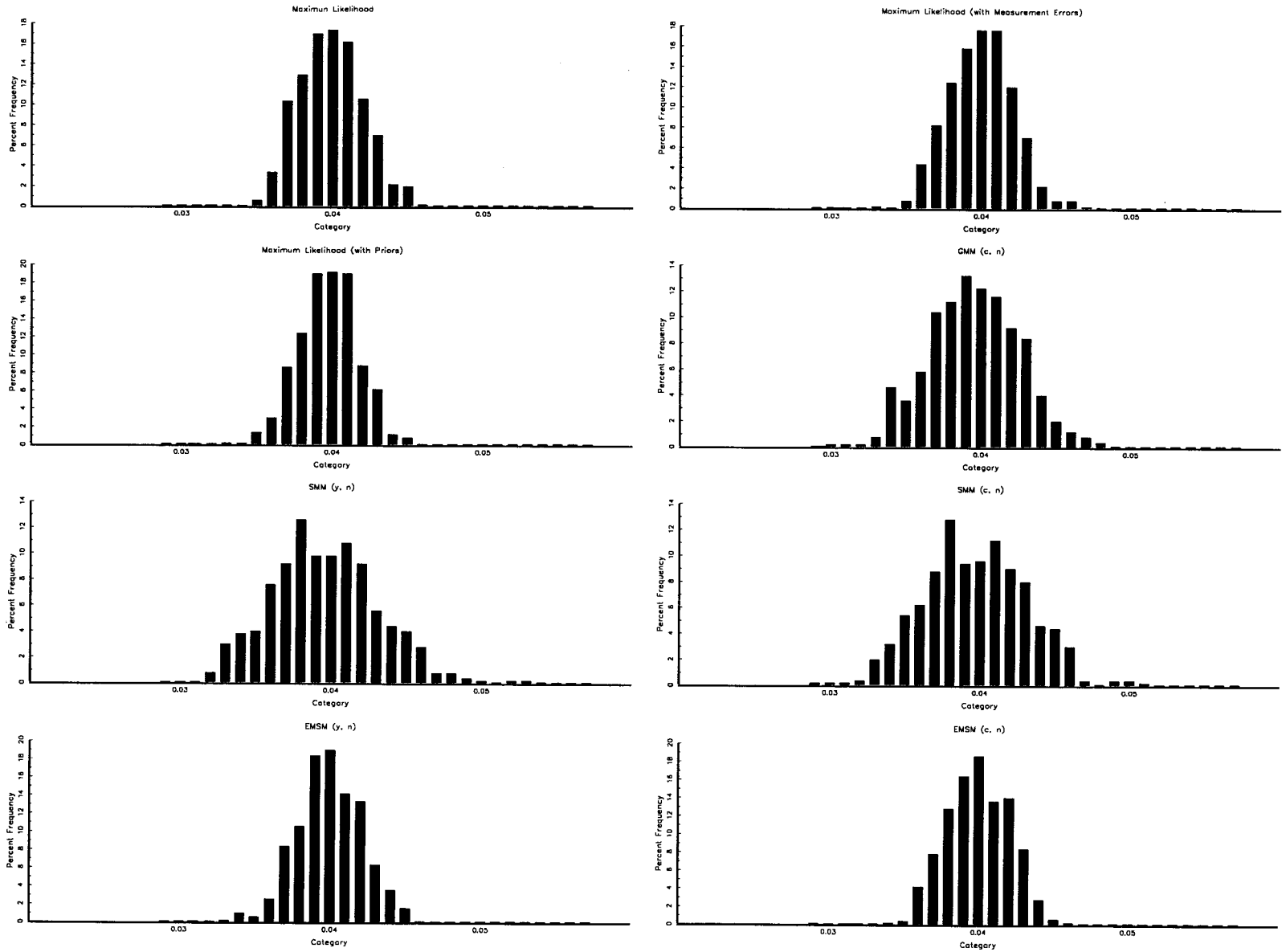

Figure 4: Impulse Response using ML Estimates

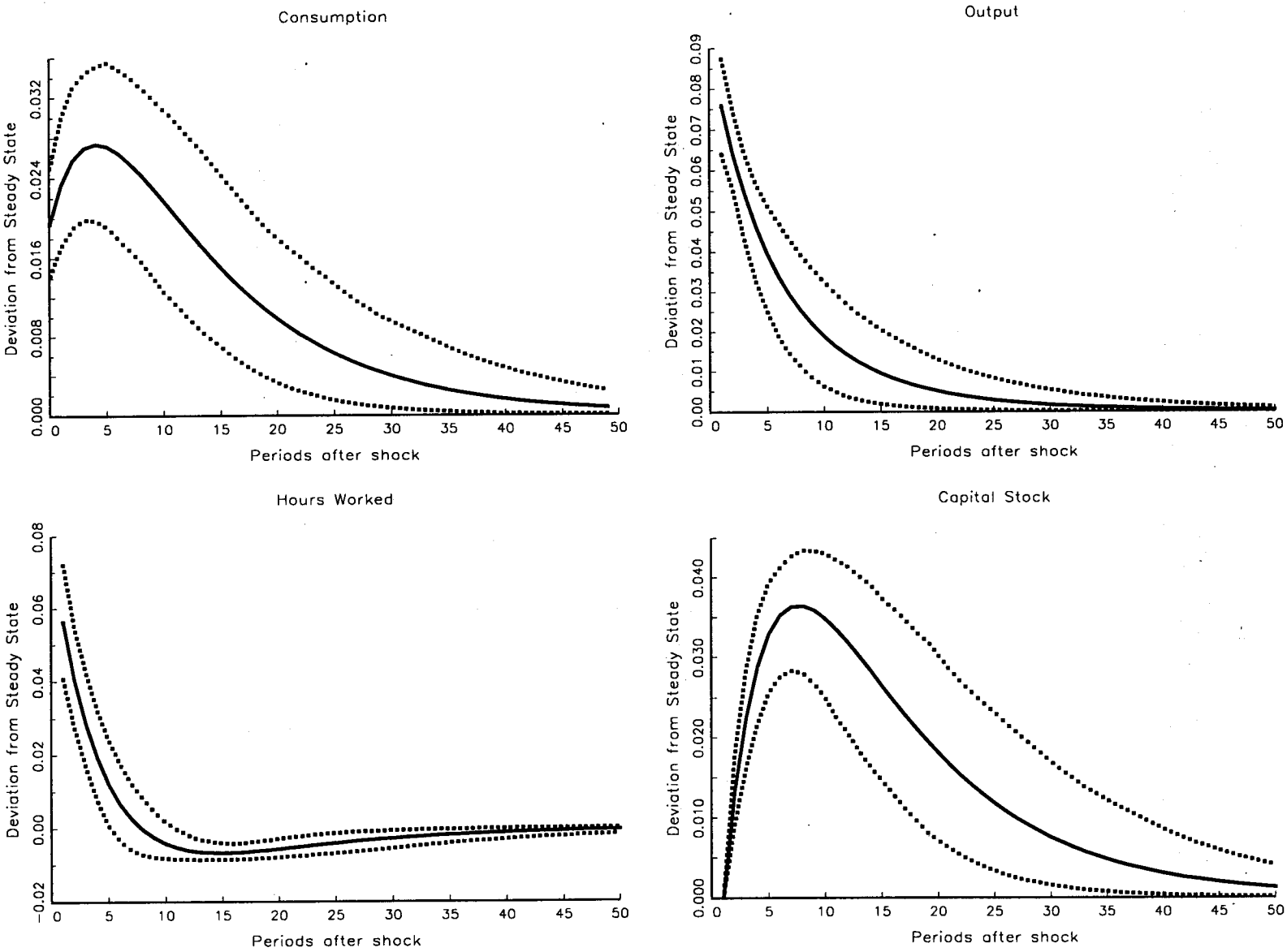

Figure 5: Impulse Response using ML Estimates (with Measurement Errors)

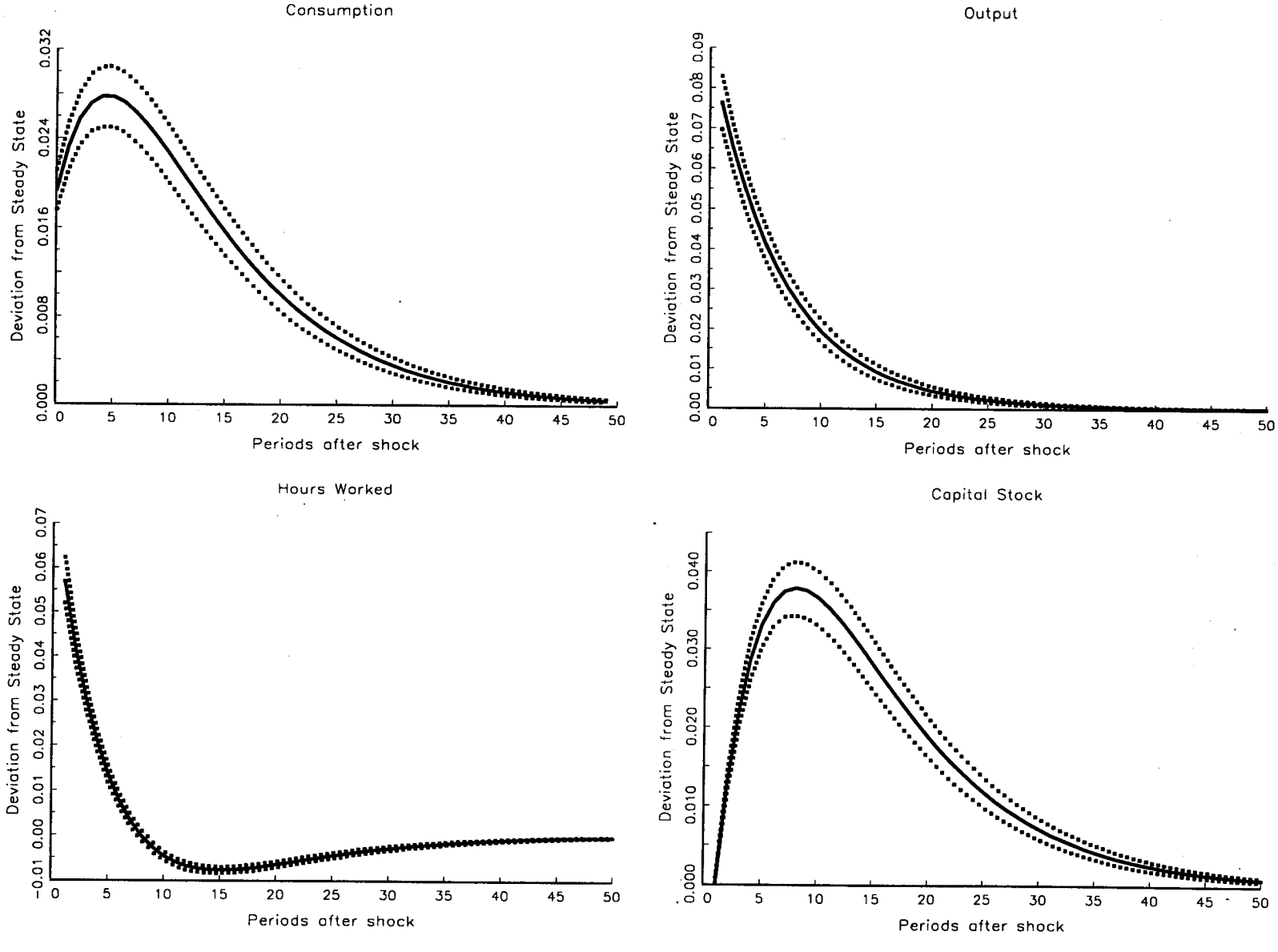

Figure 6: Impulse Response using ML Estimates (with Priors)

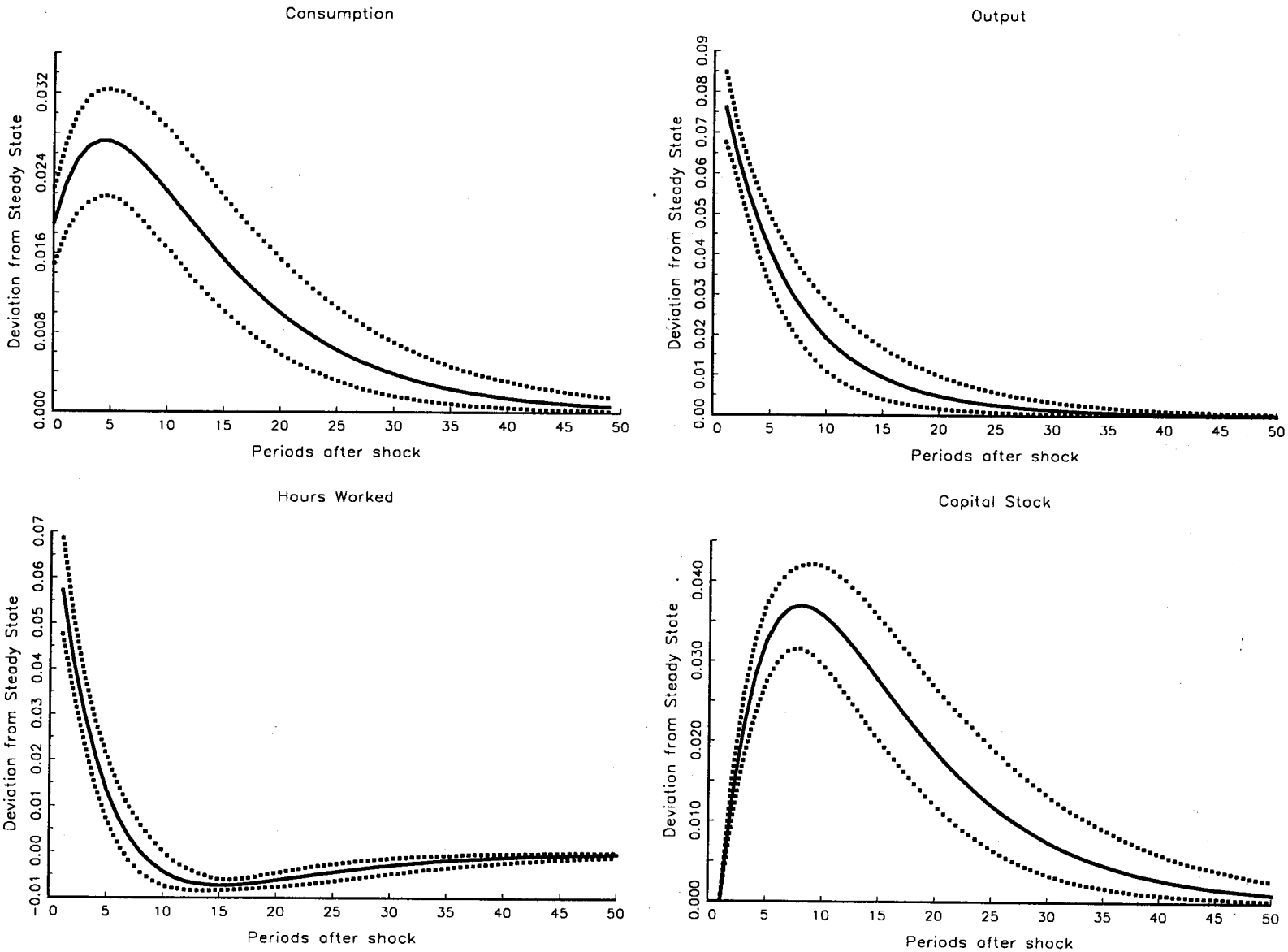

Figure 7: Impulse Response using GMM Estimates

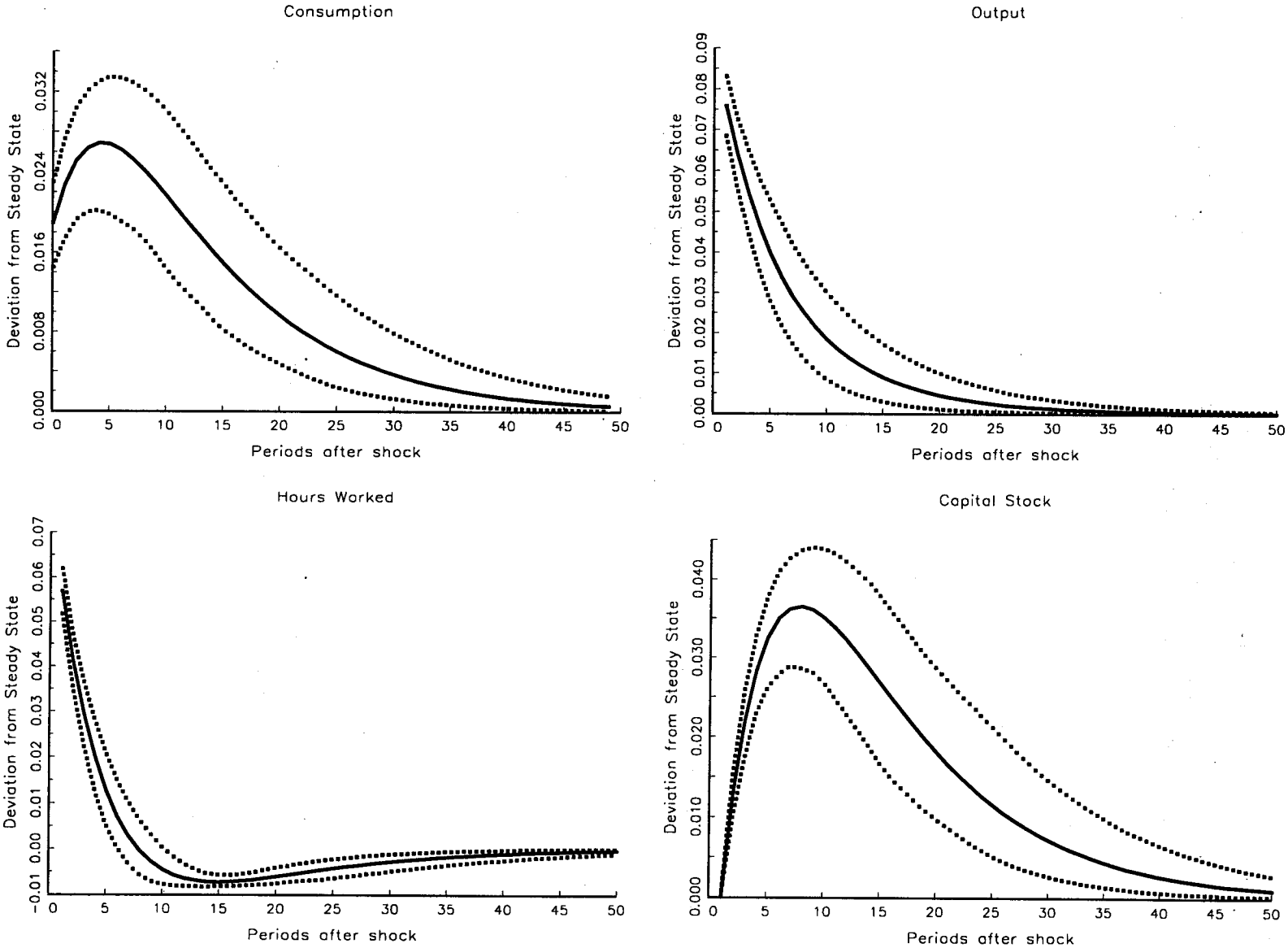

Figure 8: Impulse Response using SMM Estimates (y, n)

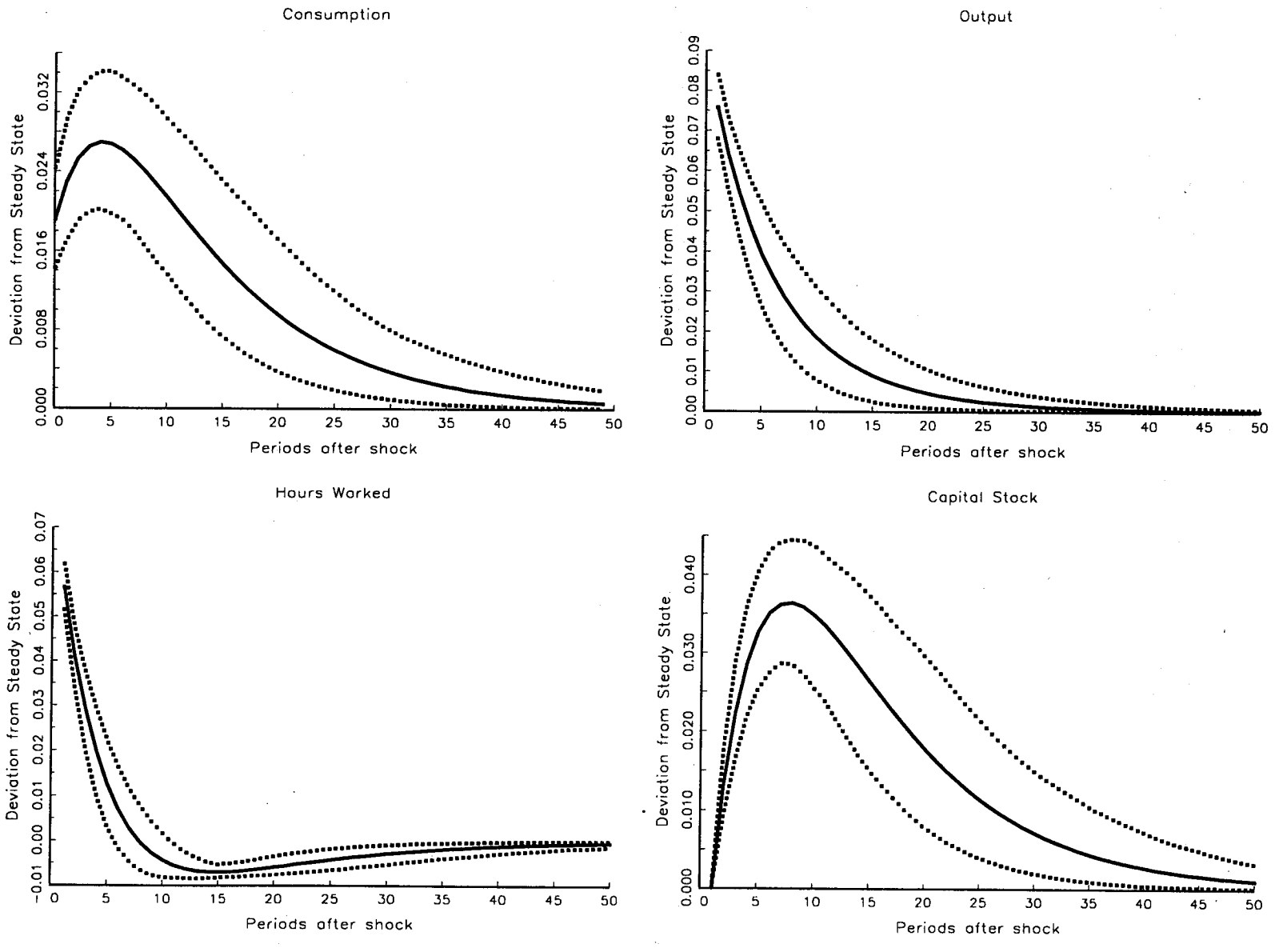

Figure 9: Impulse Response using SMM Estimates (c, n)

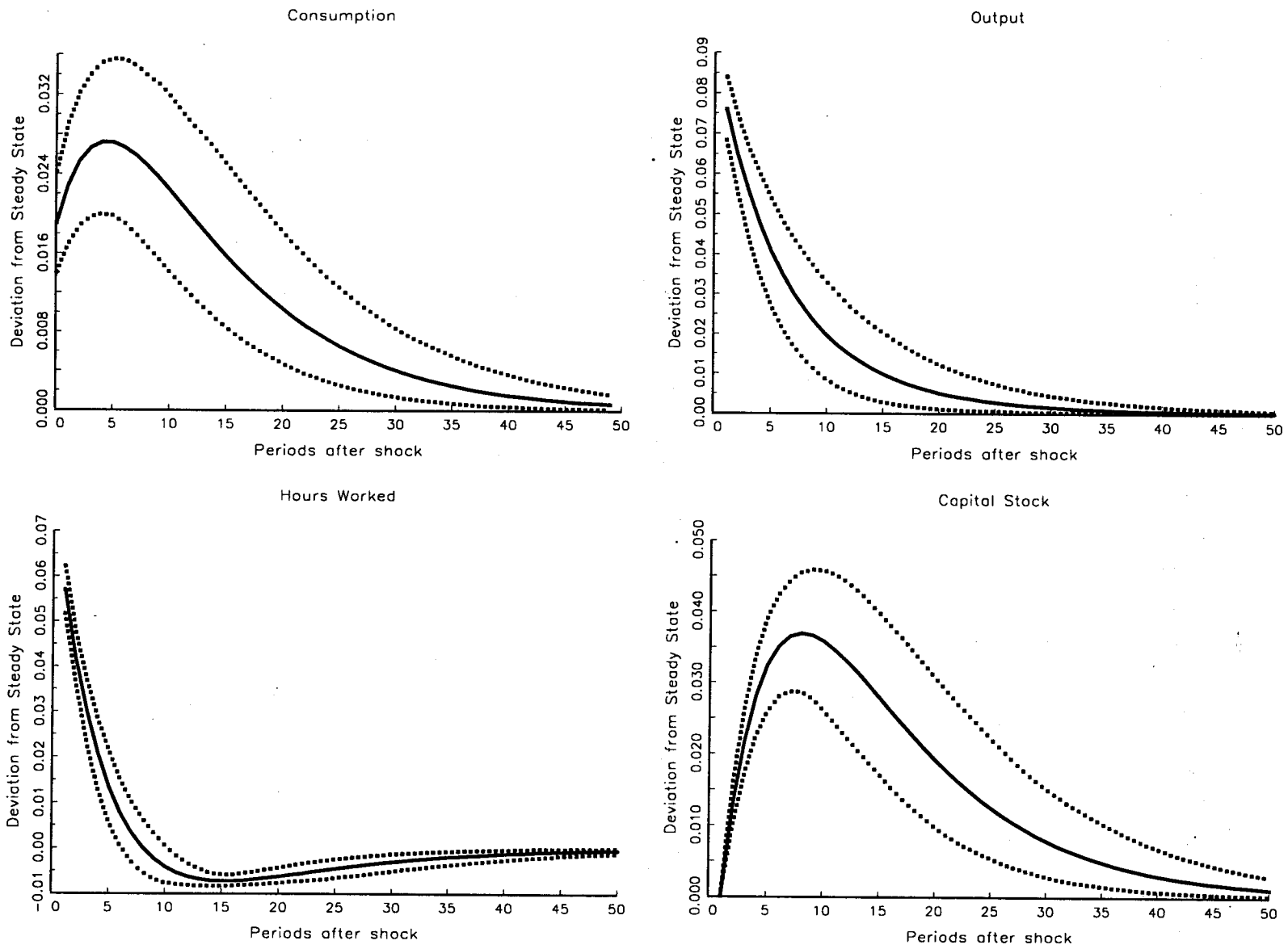

Figure 10: Impulse Response using EMSM Estimates (y, n)

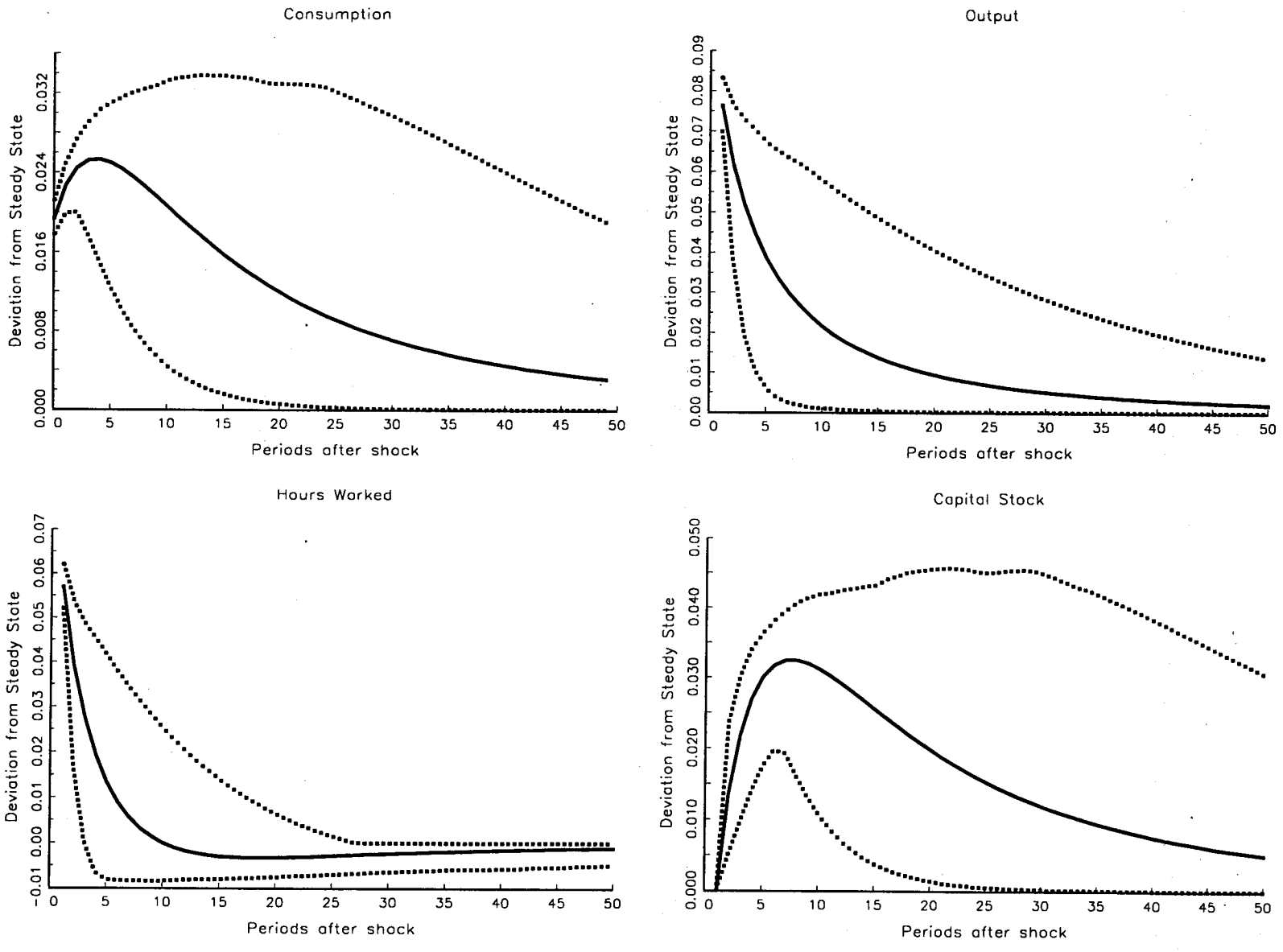

Figure 11: Impulse Response using EMSM Estimates (c, n)